

Информатика и её применения

Том 9 Выпуск 3 Год 2015

СОДЕРЖАНИЕ

Analysis of survey data containing rounded censoring intervals <i>Yu. K. Belyaev and B. Kriström</i>	2
Аналитическое моделирование распределений методом ортогональных разложений в нелинейных стохастических системах на многообразиях <i>И. Н. Синицын</i>	17
Совместное стационарное распределение числа заявок в m очередях в N -канальной системе обслуживания с переупорядочением заявок <i>А. В. Печинкин, Р. В. Разумчик</i>	25
Аппроксимация времени отклика системы облачных вычислений <i>А. В. Горбунова, И. С. Зарядов, С. И. Матюшенко, К. Е. Самуйлов, С. Я. Шоргин</i>	32
Экстремальные индексы в схеме серий и их приложения <i>А. В. Лебедев</i>	39
Пороговая стратегия ограничения доступа к ресурсам в системе массового обслуживания $M/D/1$ с функцией штрафов за несвоевременное обслуживание заявок <i>Я. М. Агаларов</i>	55
Пожар на конфигурационном графе со случайными переходами огня по ребрам <i>М. М. Лери</i>	65
Асимптотические разложения высокого порядка для несмещенных оценок и их дисперсий в модели однопараметрического экспоненциального семейства <i>В. В. Чичагов</i>	72
Реализуемость вероятностных реакций конечными вероятностными автоматами <i>А. М. Миронов</i>	85
Ссылочная целостность данных в корпоративных информационных системах <i>В. С. Зыкин</i>	97
Процессы целенаправленной генерации и развития кросс-языковых экспертных знаний: семиотические основания моделирования <i>И. М. Зацман</i>	106
Влияние морфологических операций на распознавание фигуры движущегося человека по потоку изображений <i>М. О. Абрамов, М. Ю. Катаев</i>	124
Об авторах	131
Правила подготовки рукописей	132
Requirements for manuscripts	135

ANALYSIS OF SURVEY DATA CONTAINING ROUNDED CENSORING INTERVALS

Yu. K. Belyaev¹ and B. Kriström²

Abstract: This paper makes a contribution towards the statistical analysis of data sets containing intervals, that naturally arises in survey contexts. The suggested approach is sufficiently general to cover most cases where interval data are used. Interval data appear in many contexts, such as in reliability studies and survival analysis, in medicine and economics, in opinion elicitation surveys, etc. There are several reasons for the extensive use of interval data, perhaps, the most common being one of necessity; exact values of the underlying observations are censored. The nature of the intervals analyzed here is somewhat unusual. The self-selected intervals (SeSeI) are (freely) chosen by the subjects. A generalization of the influential approach has been suggested to the statistical analysis of general censoring introduced by B. W. Turnbull. A key independence assumption in Turnbull's analysis has been explained and generalized. A sampling stopping rule based on the coverage probability has been suggested and the properties of a two-step estimator, based on the idea of asking two questions, where the second involves a way of fine-graining the information, has been discussed. This paper provides several informatics methods for SeSeI, targeting the problem of partial nonparametric identification. The properties of the suggested statistical models are stated, including a recursion for easy numerical calculations. An extensive simulation study, displaying, inter alia, the usefulness of the proposed resampling methods for the situation under study, completes the paper.

Keywords: elicitation surveys; random sampling; rounding; anchoring; coverage probability; likelihood; recursion; maximization; resampling

DOI: 10.14357/19922264150301

1 Introduction

Interval data appear in many contexts, such as in reliability studies, survival analysis in medicine, and in opinion surveys. There are several reasons for the extensive use of interval data, perhaps, the most common being one of necessity; exact values of the underlying observations are not observable. In many survey studies, missing information is a problem. Having an interval rather than a missing observation is valuable. It is known that the SeSeIs studied here, or, as they are also known, unfolding brackets, increase response-rates and help reducing a number of well-known problem as described in detail below. The authors focus on a case where the exact values are, in principle, observable; yet, for reasons just stated, the individual has difficulty pinning down an exact value.

Because an individual is asked to select any interval he or she finds most suitable, the data obtained is necessarily richer and, we argue, provide additional insights relative to points data. The SeSeIs studied here are just natural generalizations of a certain type of open ended survey questions. In some ways, the SeSeIs respond to a point made by Manski [1] who notes:

One problem is the fixation in the social sciences on point identification of parameters. . . Weaker and more

plausible assumptions often suffice to bound parameters in informative ways.

The idea of SeSeIs is, certainly, not new; the basic idea goes back to at least Morgan and Henrion [2], who suggested SeSeIs as a way of overcoming “overconfidence” and also to attack the anchoring problem (as discussed further below). There is also a connection to symbolic data analysis [3], in which intervals play an important role. Recent applications of SeSeIs include Manski and Molinari [4] and Johansson and Kriström [5].

The present authors' previous research on SeSeIs is reported in [6, 7], in which new statistical methods have been developed that cover SeSeIs in some detail. In particular, consistency of certain parametric models and measures of accuracy (using resampling methods) has been proved. In addition, a sampling stopping rule has been proposed and the properties of a 2-step estimator have been discussed based on the idea of asking two questions, where the second involves a way of fine-graining the information. This paper provides additional statistical results targeting the problem of nonparametric identification. A simple example might be useful to see more clearly what the underlying identification problem is. Assume that half of the respondents in a population state the interval $(0, 2]$ and that the other half $(1, 3]$ as

¹Department of Mathematics and Mathematical Statistics, Umeå University, Umeå SE-901 87, Sweden, yuri.belyaev@umu.se

²Center for Environmental and Resource Economics (CERE), Swedish University of Agricultural Sciences, Umeå SE-901 83, Sweden, bengt.kristrom@umu.se

their SeSeIs. Clearly, even if one knows the SeSeIs for all members of the population, there cannot be identified the 3 probabilities that the (unknown) exact value is in either of the three sets $\{(0, 1], (1, 2], (2, 3]\}$ without further information (which is proposed then to be obtained in the second step of data collection). A main contribution in this paper is to show how the nonparametric estimator generalizes Turnbull [8, 9] on nonparametric estimation in general cases of censoring. In particular, Turnbull’s estimator is based on a particular independency assumption not needed in the considered case. The obtained data also suggest that the independency assumption might be quite stringent. The authors introduce an assumption, of which Turnbull’s independence assumption can be viewed as a special case. The real world example data illustrate this idea in some detail. Finally, the authors note that the analysis of SeSeIs is closely related to the econometrics literature on partial identification [10].

To get some intuition and feel for the theoretical results, in particular, for Turnbull’s assumption, a pilot analysis of an illustrative data set is made in section 2. While the paper focuses on statistical modeling, section 3 discusses salient assumption made about the response process used by respondents in the survey situation studied. Section 4 introduces the basic statistical model, along with a discussion about the size of sampling and the properties of the likelihood in the simplest case. The present authors propose a solution to the identification problem that arises if one is interested in nonparametric models in this context in section 5. Here, a more general assumption about the response process has been also introduced, which has the Turnbull assumption as a special case. The properties of statistical models are stated, including a recursion for easy numerical calculations. A final section concludes.

2 Preliminary Analysis Using Real Data-Set

It will be useful before introducing the theoretical details to fix ideas by scrutinizing in some detail a real world data-set, where the SeSeIs have been used. The authors use contingent valuation, a widely used survey method to shed light on the value respondent attach to, for example, environmental improvements and other nonmarket goods (see [11]). While the present approach can be used in many settings, contingent valuation is a natural application, not the least because individuals are typically uncertain about how much they want to pay, say, for nonmarket priced quality improvements. The interval questions were used within a study of the costs and benefits of changing instead flow for wild salmon at

the Stornorrforss hydropower plant on the Vindel River, in northern Sweden [12]. The respondents were asked about their Willingness-To-Pay (WTP) for increasing the number of salmon that reach their spawning grounds in the river each year. The respondents were randomly sampled from a general register of the Swedish subpopulation (SPAR) older than 18 years. The sample was split into three subsets, each with a different formulation of the valuation question.

In the first sample, denoted S_1 , a standard open-ended question was used to obtain points, i. e., the WTP value for each respondent. In the second sample, S_2 , the authors asked for WTP-intervals, and in the third sample, S_3 , individuals were free to select either a point or any interval of choice. The data are summarized in the table and in Figs. 1 and 2.

These figures reveal, inter alia, that the survival distribution functions (s.d.f.s), of stated points in S_1 and the right ends intervals stated in S_2 , are nearly the same (see Fig. 1). This (unexpected) coincidence is hardly related to an underlying assumption in the Turnbull approach [8, 9]; the censored value is independent of the interval ends. In the present case, it appears as if respondents censored values are “to the right” in any given stated interval. Furthermore, Fig. 2 suggests that the union $S_2 \cup S_3$ can be considered as the basic interval-data set. This gives 241 intervals in the data collected by Håkansson [12]. There are 46 different intervals in $S_2 \cup S_3$ and it is useful to consider the choice respondents have made in further detail.

The results of the study of costs and benefits

Sample	# not answered	# of stated 0	# of stated intervals	# of stated points	Total #
S_1	97	76	0	72	245
S_2	97	88	58	0	243
S_3	527	334	183	148	1192

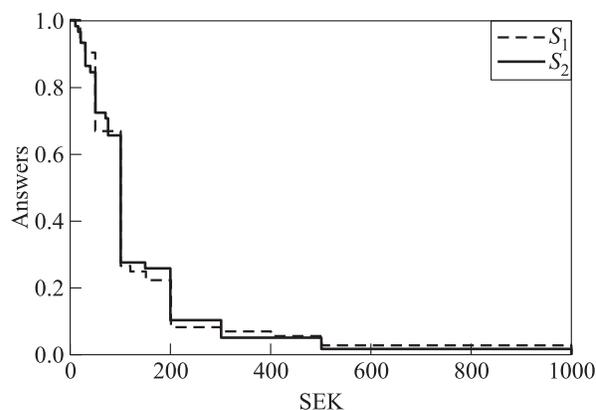


Figure 1 Two empirical distributions corresponding to the stated points in the sample S_1 and the stated right ends of WTP-intervals in sample S_2 . Both distributions are trimmed at 1000 SEK

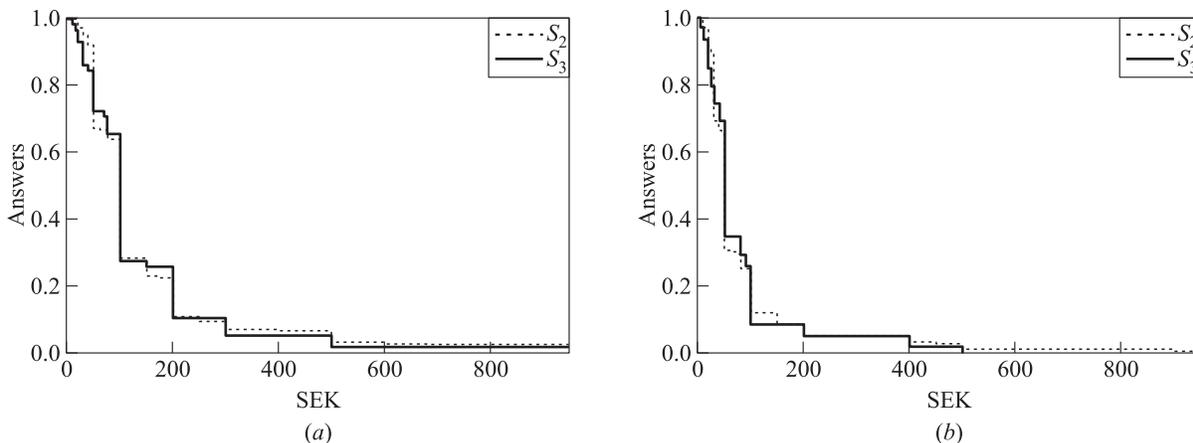


Figure 2 Two empirical distributions describing the right ends of the intervals (a) and the interval lengths (b) in S_2 and S_3 . Both distributions are trimmed at 950 SEK

First notice that there is considerable heaping on a certain set of intervals. Thus, 142 out of 241 respondents stated the following four WTP-intervals: (20, 50], (20, 100], (50, 100], and (100, 200]. These intervals were chosen by 39, 11, 69, and 23 individuals, respectively. Consequently, there are four “popular” intervals that together make up an important part of the data. Of the four “popular” intervals, three of them correspond to the face value of money bills in Sweden (the fourth has a multiple of a 100 SEK bill). Consequently, even though an individual can state any value (that he or she can afford), it is clearly the case that rounding is prevalent in the data. The 21 unique intervals in the data-set are also all possible to obtain by combining existing coin and bills in Sweden. In short, respondent invariably round their answers, seemingly using the existing bills and coins nominations in Sweden when reporting their valuations. Let now turn to some economic and statistical modeling details.

3 The Economic Model

Here, a detailed microeconomic model of the response process is not presented, because the focus is on the statistical issues. There are, however, a number of different existing models that can be reinterpreted to fit this particular case. For example, in the McFadden/Manski random utility maximization (RUM) model, the essential assumption is that the individual knows his/her utility function, but the analyst cannot observe it completely; hence, a random error term is added to the utility function. It is possible to reinterpret the SeSeIs in terms of the RUM, in the sense that one can assume that the individual considers the value $Y = \mu + \epsilon$, where Y is the point value and $\mu > 0$ is the constant and ϵ is the random errors. In this interpretation, the individual is unable to

pin down the exact value but reports instead a support of the distribution of Y . At any rate, this model will not be pursued further, given the focus of the present analysis. However, it is important to state key assumptions about the response process, to which let now turn.

Basic assumptions

The authors base their statistical models on three basic assumptions (implicitly, that the person is also assumed to reveal his/her WTP truthfully):

Assumption 1. *Each respondent might not be aware of the exact location of the true WTP-point. The respondents may freely choose SeSeIs containing their true WTP-points. The ends of stated intervals may be rounded, e. g., to simple sums of coins or paper values of money.*

Assumption 2. *The true WTP-points are independent of question mode, i. e., the structure of the valuation question does not change the true WTP-points in the SeSeI.*

Assumption 3. *The pairs of true WTP-points and the stated SeSeIs, corresponding to different sampled individuals, are the values of independent identically distributed (i.i.d.) random variables (r.v.s).*

The first part of Assumption 1 has ample support in the contingent valuation literature. For example, the standard approach in this literature is to use a payment card (with given brackets), where the individual is to state (for each interval) how certain he or she is about his/her WTP being in any of the brackets displayed in the card (see [13], a recent survey of 76 papers is in [14]). This is another way to cater for respondent uncertainty compared to what is suggested here.

In the second part of Assumption 1, the authors naturally assume (but do not explicitly state) that the individual chooses an interval such that WTP does not

exceed his or her income. The authors explicitly allow for rounding in the final part of Assumption 1.

Assumption 2 is important. From some perspectives, this assumption is strong, given the evidence that exists on the differing results between types of valuation questions. For example, it has been demonstrated (e.g., in the research by the Nobel Laureate Daniel Kahnemann) that anchoring is prevalent under uncertainty. Anchoring here means that an individual, in a situation of uncertainty, uses any information, whether relevant or not, to form his/her answer. In particular, it is apparent that the actual brackets being displayed in a payment card may affect the answers; if so, the answers become “anchored” around the numbers suggested. If we are just asking a SeSeI question this issue is moot, in the sense that no numbers are suggested that may be used as anchors. However, as it will be seen, for (partial) nonparametric identification, it is necessary to ask the second question that does involve suggesting numbers. Fortunately, if anchoring exists in the suggested two-step approach, it is likely to be less of a problem, since one has information about the “undistorted” interval, elicited in the first stage.

Assumption 3 is, typically innocuous, in that most contingent valuation applications are based on large significant national samples.

Let now turn to the statistical modeling in more detail. It is useful to begin with addressing the sampling issue first; how many individuals it is necessary to sample in order to be sure that we have, in a loose sense, “enough” information? The proposed solution is based on an estimator obtained by Good [15]. The estimation which is used in the derivation of a sample stopping rule, is related to what is called here a coverage probability. The end-result is a stochastic difference equation, which will be now derived.

4 Statistical Modeling of Self-Selected Intervals: Basic Ideas

Let begin the discussion of the statistical modeling with analyzing the sampling problem under SeSeIs, then turn to a basic building block called division intervals. Let then introduce similar Turnbull’s basic assumption, derive the likelihood and its properties, and select a convenient numerical scheme for the implied numerical maximization problem. The discussion of the basic statistical model will be ended by introducing an alternative to Turnbull’s assumption, essentially an assumption suggested by the data. This will lead one naturally to the two-step extension, introduced in subsequent section 5.

4.1 Sampling stopping rule

Consider n , randomly sampled respondents, from a population \mathfrak{P} of interest, that have stated SeSeIs $\mathbf{y}_1^n = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$, $\mathbf{y}_i = (y_{Li}, y_{Ri}]$. The intervals need not to be unique, because of rounding we do expect to see quite many repeated intervals. Yet, an important question is when to stop the sampling process; which n may be considered sufficiently large? The problem is related to the problem of estimating, e.g., the number of species in a certain area, or the number of words in a given language. It will be addressed using an idea by Good (1954), which will be generalized to a stochastic difference equation and to a “coverage probability.” Let first introduce some notation.

Let consider $\mathbf{y}_1^n = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ as a realization of a multinomial random process $\{\mathbf{Y}_i\}_{i \geq 1}$ with “time” parameter $i = 1, 2, \dots$. The r.v.s $\{\mathbf{Y}_i\}_{i \geq 1}$ are (as noted) i.i.d. and the set of their values is all SeSeIs $\mathcal{U}_{\text{all}} = \{\mathbf{u}_\alpha : p_\alpha = P[\mathbf{Y}_i = \mathbf{u}_\alpha] > 0, \alpha \in A\}$. The set \mathcal{U}_{all} and the discrete distribution $\{p_\alpha, \alpha \in A\}$ are not known. α is an integer index identifying \mathbf{u}_α , i.e., $\mathbf{u}_{\alpha'} \neq \mathbf{u}_{\alpha''}$ if $\alpha' \neq \alpha''$. All $m(n) \leq n$ different intervals in \mathbf{y}_1^n can be ordered by their endpoints. Let write $\mathbf{y}_{i_1} < \mathbf{y}_{i_2}$ if either $y_{Li_1} < y_{Li_2}$, or $y_{Li_1} = y_{Li_2}$ but $y_{Ri_1} < y_{Ri_2}$. Then different intervals $\mathbf{y}_{i'_1} < \mathbf{y}_{i'_2} < \dots < \mathbf{y}_{i'_{m(n)}}$ and let $\mathbf{u}_{h,n} = \mathbf{y}_{i'_h}$ for all \mathbf{y}_i identical with $\mathbf{y}_{i'_h}$ and $\mathbf{t}(n) = \{t_{1,n}, \dots, t_{m(n),n}\}$, $t_{h,n} = \sum_{i=1}^n I[\mathbf{y}_i = \mathbf{u}_{h,n}]$. The collected data can be written as the following list:

$$d\mathbf{1}_n = \{\dots, \{h, \{u_{Lh,n}, u_{Rh,n}\}, t_{h,n}\}, \dots\}$$

where ordering indexes $h = 1, \dots, m(n)$ of different intervals $\mathbf{u}_{h,n} = (u_{Lh,n}, u_{Rh,n}]$ depend on the collected data \mathbf{y}_1^n . Finally, let $\mathcal{U}_{m(n),n} = \{\mathbf{u}_{1,n}, \dots, \mathbf{u}_{m(n),n}\} \subseteq \mathcal{U}_{\text{all}}$ be the set of all different SeSeIs stated in \mathbf{y}_1^n .

Consider now the sampling stopping problem. It is interesting to estimate the fraction of individuals (in \mathfrak{P}) that will state WTP-intervals already observed to be in $\mathcal{U}_{m(n),n}$. Let $p_c(n)$ be probability of the event that the last WTP-interval, $\mathbf{y}_n = \mathbf{u}_{h_n,n}$, in \mathbf{y}_1^n , $t_{h_n,n} \geq 2$. Let H_n denote the r.v. that the n th individual states a WTP-interval $\mathbf{u}_{H_n,n}$ containing his/her WTP-value x_n . Since $E[I[t_{H_n,n} \geq 2]] = p_c(n)$, one can consider $I[t_{h_n,n} \geq 2]$ as a value of an unbiased estimator of $p_c(n)$. After averaging $I[t_{H_n,n} \geq 2]$ given the sufficient statistic $\mathbf{t}(n)$, one obtains an unbiased estimate of $p_c(n)$:

$$\hat{p}_c(n) = \frac{r(n)}{n}$$

where $r(n) = \sum_{h=1}^{m(n)} t_{h,n} I[t_{h,n} \geq 2]$, and $\hat{p}_c(n)$ denotes an estimate [16].

To proceed, let reduce the problem to a classical urn problem with $r(i)$ white and $i - r(i)$ black balls. Let

$$\begin{aligned} \mathcal{U}_{m(i),i}(1) &= \{\mathbf{u}_h : \mathbf{u}_h \in \mathcal{U}_{m(i),i}, t_{hi} = 1\}; \\ \mathcal{U}_{m(i),i}(2) &= \{\mathbf{u}_h : \mathbf{u}_h \in \mathcal{U}_{m(i),i}, t_{hi} \geq 2\}; \\ \mathcal{U}_{m(i),i} &= \mathcal{U}_{m(i),i}(1) \cup \mathcal{U}_{m(i),i}(2). \end{aligned}$$

Then

$$\Delta_c(i) = \begin{cases} \hat{p}_c(i+1) - \hat{p}_c(i) = \frac{i-r(i)}{i(i+1)} & \text{if } \mathbf{y}_{i+1} \in \mathcal{U}_{m(i),i}(2); \\ \frac{2i-r(i)}{i(i+1)} & \text{if } \mathbf{y}_{i+1} \in \mathcal{U}_{m(i),i}(1); \\ -\frac{r(i)}{i(i+1)} & \text{if } \mathbf{y}_{i+1} \notin \mathcal{U}_{m(i),i}. \end{cases}$$

The probability $p_c(n)$ is increasing in n . Hence, one can use $\hat{p}_c(n)$ as an approximation of a lower bound estimate for $p_c(n')$ for any $n' > n$. Then, one may interpret $\hat{p}_c(n)$ 100% as a lower bound estimate of the percentage of all individuals in $\mathfrak{P}_c \subset \mathfrak{P}$ who would claim an interval $\mathbf{u}_h \in \mathcal{U}_{m(n),n}$. As indicated above, $p_c(n)$ is called here a *coverage probability*.

By calculating $\hat{p}_c(i)$ for each $i \leq n$, one can observe the evolution of $\hat{p}_c(i)$ for some empirical data (Fig. 3).

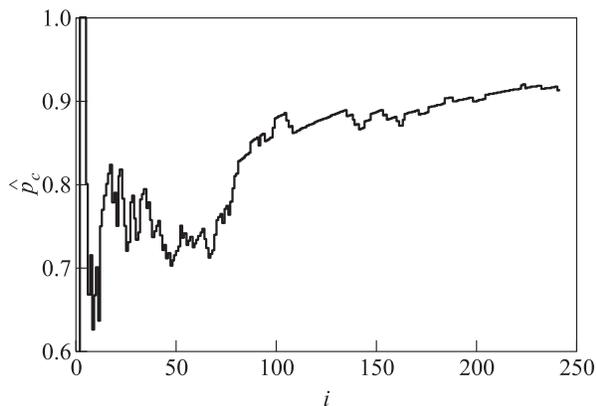


Figure 3 The dynamics of the coverage probability estimates $\hat{p}_c[i]$ as a function of i

The decision to stop collecting data on the first step can be determined by the value of $\hat{p}_c(n)$. If this value is not sufficiently close to 1 and if it is possible to extend data collection, then sampling continues. Note that if the collection of data has stopped with $n_1 \geq n$ and $\hat{p}_c(n_1) = r(n_1)/n_1$, then inference about the WTP-distribution will correspond to the subset of individuals \mathfrak{P}_c with WTP-intervals already collected in $\mathcal{U}_{m(n_1),n_1}$.

The next step is to introduce the basic building block of the statistical model, the so-called *division intervals*. The basic idea is to divide each interval into smaller pieces where those pieces are obtained from the responses of the respondents.

4.2 Division intervals

Henceforth, the number n_1 of randomly sampled respondents is fixed and it is suppressed it in the sequel; let write $\mathcal{U}_m = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ instead of $\mathcal{U}_{m(n_1),n_1} = \{\mathbf{u}_{1,n_1}, \dots, \mathbf{u}_{m(n_1),n_1}\}$ and t_h instead of t_{h,n_1} . Let simplify and use \mathcal{U}_m , \mathbf{u}_h , \mathbf{v}_j , \mathcal{C}_h , and \mathcal{D}_j instead of \mathcal{U}_{m,n,n_1} , \mathbf{u}_{hn_1} , \mathbf{v}_{jn_1} , \mathcal{C}_{hn_1} , and \mathcal{D}_{jn_1} .

Let $v_0 < v_1 < \dots < v_{k-1} < v_k$ be ordered values of the end points of all intervals $\mathbf{u}_h \in \mathcal{U}_m$. Let $v_{Lj} = v_{j-1}$ and $v_{Rj} = v_j$. Then, the intervals in the following collection $\mathcal{V}_k = \{\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_k\}$, $\mathbf{v}_j = (v_{Lj}, v_{Rj})$, $j = 1, \dots, k$, are called *division intervals*.

Each interval $\mathbf{u}_h \in \mathcal{U}_m$ is, thus, the union of disjoint division intervals

$$\mathbf{u}_h = \bigcup_{j \in \mathcal{C}_h} \mathbf{v}_j; \quad \mathcal{C}_h = \{j : \mathbf{v}_j \subseteq \mathbf{u}_h\}.$$

Below, the following sets of indices h will be also used:

$$\mathcal{D}_j = \{h : \mathbf{v}_j \subseteq \mathbf{u}_h\}, \quad j = 1, \dots, k.$$

For example, if there are used the data in $S_2 \cup S_3$, then $v_1 = (0, 5]$; $v_2 = (5, 10]$; $v_3 = (10, 15]$; $v_4 = (15, 20]$; $v_5 = (20, 25]$; $v_6 = (25, 30]$; $v_7 = (30, 40]$; $v_8 = (40, 50]$; $v_9 = (50, 60]$; $v_{10} = (60, 70]$; $v_{11} = (70, 75]$; $v_{12} = (75, 80]$; $v_{13} = (80, 100]$; $v_{14} = (100, 150]$; $v_{15} = (150, 170]$; $v_{16} = (170, 200]$; $v_{17} = (200, 250]$; $v_{18} = (250, 300]$; $v_{19} = (300, 400]$; $v_{20} = (400, 500]$; $v_{21} = (500, 600]$; $v_{22} = (600, 1000]$; $v_{23} = (1000, 1000]$. For $j = 21$, $\mathbf{v}_{21} = (500, 600]$. Then, $\mathcal{D}_{21} = \{38, 43, 45, 46\}$; $\mathbf{u}_{38} = (100, 1000]$; $\mathbf{u}_{43} = (300, 600]$; $\mathbf{u}_{45} = (500, 1000]$; $\mathbf{u}_{46} = (500, 2000]$. An informative picture of the data was obtained by considering how each stated interval can be constructed via properly selecting division intervals (Fig. 4).

Respondents' WTP-points $\{x_i\}$ by Assumption 3 are values of $\{X_i\}$ i.i.d. r.v.s. Let define even $\{H_i = h\} \subset \{X_i \in \mathbf{u}_h\}$, $w_h = P[\{H_i = h\} \leq p_h] = P[X_i \in \mathbf{u}_h]$, $h = 1, \dots, m$. Note that events $\{H_i = h\}$ are observable, but not all of $\{X_i \in \mathbf{u}_{h'}\}$, $h' \neq h$, $\mathbf{u}_{h'} \cap \mathbf{u}_h \neq \emptyset$ are observable. The probability, of the number of times $t_{1n_1}, \dots, t_{mn_1}$ that $\mathbf{u}_1, \dots, \mathbf{u}_m$ had appeared in \mathbf{y}_1^n is proportional to $\prod_{h=1}^m w_h^{t_{hn_1}}$, i.e., one has a multinomial distribution. The corresponding (normed by n_1) log likelihood (llik) is

$$\begin{aligned} \text{llik}[w_1, \dots, w_m | t_1, \dots, t_m] &= \sum_{h=1}^m \frac{t_{hn_1}}{n_1} \text{Log}[w_h]; \\ \sum_{h=1}^m t_{hn_1} &= n_1. \end{aligned}$$

The maximum of llik over $w_h \geq 0$, $\sum_{h=1}^m w_h = 1$, is attained at $\check{w}_h = t_{hn_1}/n_1$, $h = 1, \dots, m$. Note that $e[\hat{\mathbf{w}}_1^m] = -\sum_{h=1}^m \check{w}_h \text{Log}[\check{w}_h]$ is the empirical en-

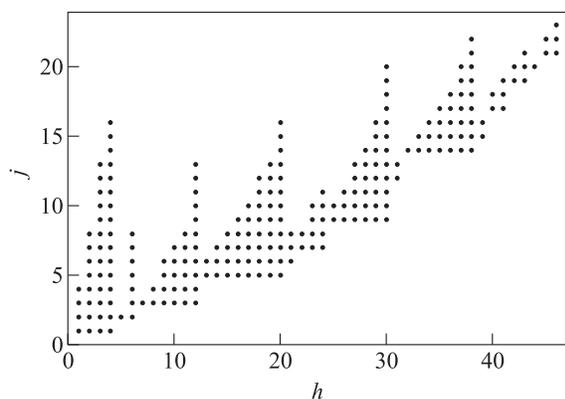


Figure 4 The set of all compatible indexes $h, j, \mathbf{v}_j \subseteq \mathbf{u}_h$, $h = 1, \dots, 46$, $j = 1, \dots, 23$ in the empirical data. The sets $\mathcal{C}_h = \{j, \mathbf{v}_j \subseteq \mathbf{u}_h\}$ and $\mathcal{D}_j = \{h, \mathbf{v}_j \subseteq \mathbf{u}_h\}$ are the h -cuts and j -cuts of the shown set, e. g., $\mathcal{C}_{10} = \{3, 4, 5, 6, 7\}$ and $\mathbf{u}_{10} = \cup_{j=3}^7 \mathbf{v}_j$

entropy of the multinomial distribution with probabilities $\hat{\mathbf{w}}_1^m = \{\hat{w}_1, \dots, \hat{w}_m\}$.

4.3 The Turnbull assumption

Let consider statistical models where there are finite numbers m of different SeSeIs \mathbf{y}_i , $i = 1, \dots, n_1$, each contains one unobserved point $X_i = x_i$. The target is nonparametric estimation of the underlying distribution function. As it is common in the survival analysis, let assume a kind of independence. Here, it means that an interval containing value x_i has been stated independently of the position of x_i in the interval. In statistical models with exogenously given intervals, these assumptions are called the *noninformative condition* (see, e. g., [17, 18]). In the considered context, this restrictive assumption on independence of positions (AIP) WTP-points included in SeSeIs can be stated as follows.

Assumption 4. AIP : For each pair of $\{h, j\}$, $h = 1, \dots, m$ and $j \in \mathcal{C}_h$ for events $\{H_i = h\}$ and $\{X_i \in \mathbf{v}_j\}$,

$$P_o[\{X_i \in \mathbf{v}_j\} \cap \{H_i = h\}] = q_{oj} w_{oh} I[j \in \mathcal{C}_h], \\ j \in \mathcal{C}_h, h = 1, \dots, m.$$

From this assumption, it follows that

$$P_o[\{H_i = h, X_i \in \mathbf{y}_i = \mathbf{u}_h = \cup_{j \in \mathcal{C}_h} \mathbf{v}_j\}] = w_{oh} p_{oh}$$

where P_o is the true probability; $w_{oh} = P_o[H_i = h]$; $p_{oh} = \sum_{j \in \mathcal{C}_h} q_{oj}$ with $q_{oj} = P_o[X_i \in \mathbf{v}_j]$; $\mathbf{v}_j \subseteq \mathbf{u}_h$.

This assumption is basic to Turnbull [8, 9] and corresponds to the case where all end points of censoring intervals are known before realizations of r.v.s X_i , and the conditional distribution of X_i , $i = 1, 2, \dots$, given these points, does not depend on the ends of these

intervals. Turnbull's assumption on conditional probabilities of $X_i \in u_h$, given $\{\mathbf{Y}_i = (u_{Lh}, u_{Rh})\}$, implies that $p_{oh} = P_o\{X_i \in \mathbf{u}_h\}$. This assumption is enough for obtaining a simplified version of the llik (see [17]).

By AIP, additional probabilities w_{oh} , $h = 1, 2, \dots, m$, have been introduced which are the true probabilities that the i th respondent selects $\mathbf{u}_h \in U_m$ such that $X_i \in \mathbf{u}_h$. Note that since $I[X_i \in \mathbf{u}_h] = 1$ and due to presence $\mathbf{u}_h \in \mathbf{y}_1^n$, the probabilities w_{oh} should be positive, $h = 1, \dots, m$. In addition, all probabilities $p_{oh} = P_o[X_i \in \mathbf{u}_h] > 0$, $h = 1, \dots, m$, are positive.

We do not know w_{oh} and parameters q_{oj} where $\{h, j\}$ are the compatible pairs. The natural consistent estimates of w_{oh} are the frequencies

$$\hat{w}_h(n_1) = \frac{t_{hn_1}}{n_1}, \quad t_{hn_1} = \sum_{i=1}^{n_1} I[h_i = h].$$

Such probabilities are the basic ingredients of the likelihood, to which let now turn.

4.4 The likelihood

The current problem is to find consistent estimates for points of the true s.d.f. $S_{ot}[\cdot]$. One has the following normed log likelihood (llik_{AIP}) with parameters \mathbf{q}_1^k and \mathbf{w}_1^m :

$$\text{llik}_{\text{AIP}}[\mathbf{q}_1^k, \mathbf{w}_1^m | \mathbf{y}_1^{n_1}] = \frac{1}{n_1} \text{Log} \left[\prod_{i=1}^{n_1} \left(w_{h_i} \sum_{j \in \mathcal{C}_{h_i}} q_j \right) \right] \\ = \frac{1}{n_1} \sum_{i=1}^{n_1} \text{Log} \left[w_{h_i} \sum_{j \in \mathcal{C}_{h_i}} q_j \right] = f_{ll}[\mathbf{q}_1^k] + \hat{e}[\mathbf{w}_1^m]$$

where

$$f_{ll}[\mathbf{q}_1^k] = \sum_{h=1}^m \hat{w}_h(n_1) \text{Log} \left[\sum_{j \in \mathcal{C}_h} q_j \right], \\ \mathbf{q}_1^k = \{q_1, \dots, q_k\}; \\ \hat{e}[\mathbf{w}_1^m] = \sum_{h=1}^m \hat{w}_h(n_1) \text{Log}[w_h], \quad \mathbf{w}_1^m = \{w_1, \dots, w_m\}.$$

To save on notation, let write llik_{AIP} instead of llik_{AIP}[\mathbf{q}_1^k, \mathbf{w}_1^m | \mathbf{y}_1^n]. Note that llik_{AIP} depends on \mathbf{q}_1^k only through $\mathbf{p}_1^m = \{p_1, \dots, p_m\}$. Hence, let consider llik_{AIP} as a function of \mathbf{p}_1^m and write it as llik_{AIP}[\mathbf{p}_1^m]. Due to presence of all $\mathbf{u}_h \in U_m$ in the data \mathbf{y}_1^n , there is a small positive number $\varepsilon > 0$ such that the global maximum of llik_{AIP} is contained in the following compact convex multidimensional polyhedron:

$$\mathcal{S}_{k-1} = \left\{ \mathbf{q}_1^k : 0 \leq q_j < 1, \sum_{j=1}^k q_j = 1, \right. \\ \left. p_h = \sum_{j \in \mathcal{C}_h} q_j \geq \varepsilon > 0, h = 1, \dots, m \right\}.$$

If $k = 4$, then $S_3 \subseteq \mathbb{R}^3$ is the main part of an equiedged pyramid with $k = 4$ vertices. The case when $\mathbf{q}_1^k \in S_{k-1}$ have zero components is not excluded, i. e., $q_j = 0$ for some of j .

Let now state an important property by means of the following lemma and then obtain the concavity of the lik.

Lemma 4.1. *fu[\mathbf{q}_1^k] is concave on S_{k-1} .*

Proof. For any pair of points $\mathbf{q}_1^k(i) \in S_{k-1}$, $i = 1, 2$, $\mathbf{q}_1^k(1) \neq \mathbf{q}_1^k(2)$, $\mathbf{q}_1^k[t] = (1-t)\mathbf{q}_1^k(1) + t\mathbf{q}_1^k(2) \in S_{k-1}$, $0 \leq t \leq 1$. Then, for $p_h[t] = \sum_{j \in C_h} q_j[t] \geq \varepsilon > 0$, $h = 1, 2, \dots, m$, one has:

$$\frac{d^2 \text{Log}[p_h[t]]}{dt^2} = - \sum_{j \in C_h} \left(\frac{q_j(2) - q_j(1)}{p_h[t]} \right)^2 < 0;$$

$$\frac{d^2 \text{lik}[\mathbf{q}_1^k[t]]}{dt^2} = \sum_{h=1}^m \hat{w}_h(n_1) \frac{d^2 \text{Log}[p_h[t]]}{dt^2} < 0$$

because $\hat{w}_h(n_1) > 0$, $h = 1, 2, \dots, m$.

It follows that for any $\mathbf{q}_1^k(i)$, $i = 1, 2, \dots$, on interval $I[q_1^k(1), q_1^k(2)]$ connecting these points, $fu[q_1^k]$ is concave.

Theorem 4.2.

- (i) *The lik_{AIP}[\mathbf{q}_1^k] is concave on the S_{k-1} .*
- (ii) *The lik_{AIP} has a stationary point $\check{\mathbf{q}}_1^k = \{\check{q}_1, \dots, \check{q}_k\} \in S_{k-1}$ where it attains its global maximum of lik_{AIP}[\mathbf{q}_1^k] on S_{k-1} and $\check{\mathbf{p}}_1^m = \{\check{p}_1, \dots, \check{p}_m\}$, $\check{p}_h = \sum_{j \in C_h} \check{q}_j$, is the maximum likelihood (ML) estimate of the true parameters \mathbf{p}_{o1}^m given $\hat{\mathbf{w}}_1^m(n_1)$.*

Proof. The part $e(\mathbf{w}_1^m)$ does not depend on \mathbf{q}_1^k ; hence, from Lemma 4.1, it follows that lik_{AIP} is concave on S_{k-1} . At a stationary point $s \in S_{k-1}$, all first-order partial derivatives have to be zero and there are some negative second order derivatives. To include constrains $q_j(s) \geq 0$, $\sum_{j=1}^k q_j(s) = 1$, the method with the following Lagrange function is used:

$$\varphi[\mathbf{q}_1^k(s)] = \text{lik}[\mathbf{q}_1^k(s)] + \sum_{j=1}^k q_j(s) (\mu_j - \mu_o) .$$

Here, μ_j and μ_o are the multipliers satisfying the Kuhn–Tucker conditions $\mu_j q_j(s) = 0$, $\mu_j \geq 0$, $j = 1, \dots, k$ [19, 20]. Then, one has k equations, $j = 1, \dots, k$:

$$\frac{\partial \varphi[q_1^k(s)]}{\partial q_j(s)} = \sum_{h=1}^m \hat{w}_h(n_1) \frac{I[j \in C_h]}{\sum_{j' \in C_h} q_{j'}(s)} + (\mu_j - \mu_o) = 0 .$$

By multiplying each of these equations by $q_j(s)$ and sum them using $\mu_j q_j(s) = 0$, $j = 1, \dots, k$, one has $\mu_o = 1$.

Then, for each j after multiplying both sides of these equations by $q_j(s)$, one has

$$\sum_{h=1}^m \hat{w}_h(n_1) \frac{I[j \in C_h] q_j(s)}{\sum_{j' \in C_h} q_{j'}(s)} - q_j(s) = 0, \quad j = 1, \dots, k.$$

If some of $q_j(s) = 0$, $j \in J_o = \{j' : q_{j'}(s) = 0\} \neq \emptyset$, then, to guarantee that the stationary point $\mathbf{q}_1^k(s) \in S_{k-1}$, it is necessary to check that all $p_h(s) = \sum_{j \in C_h} q_j(s) > 0$, $h = 1, \dots, m$. If so, then $\mathbf{q}_1^k(s)$ is placed in the side of S_{k-1} where $q_j = q_j(s) \equiv 0$, $j \in J_o$, and $q_j(s) > 0$, $j \in J_p = \{j' : q_{j'}(s) > 0\}$. Due to concavity, the stationary point $q_1^k(s)$ is unique and it attains the global maximum of the lik.

This likelihood can be calculated by a recursion next described.

4.5 Numerical recursive method

The authors suggest a method efficient for numerical calculation of a sequence of $\mathbf{q}_1^{(r)k}$, $r = 1, 2, \dots$, converging to the point $\mathbf{q}_1^k(s)$ of the global maximum lik_{AIP}[$\mathbf{q}_1^k(s)$] where $\mathbf{q}_1^k(s)$ is a stationary point of the lik. Henceforth, let write $q_j^{(r)}(n_1)$ and $q_j(s, n_1)$ instead of $q_j^{(r)}$ and $q_j(s)$ to underline dependence on n_1 . One can consider the following operator, sequentially transforming points $\{q_1^{(r)}(n_1), \dots, q_k^{(r)}(n_1)\}$ into points $\{q_1^{(r+1)}(n_1), \dots, q_k^{(r+1)}(n_1)\}$, $k = 1, 2, \dots$ in S_{k-1} :

$$q_j^{(r+1)}(n_1) = \sum_{h=1}^m \hat{w}_h(n_1) \frac{q_j^{(r)}(n_1) I[j \in C_h]}{\sum_{j' \in C_h} q_{j'}^{(r)}(n_1)} .$$

The set S_{k-1} is compact and $\sum_{j=1}^k q_j^{(r)}(n_1) = 1$ for any r . Hence, by the Banach theorem, S_{k-1} contains at least one limit point of the sequence $\mathbf{q}_1^{(r)k}(n_1) = \{q_1^{(r)}(n_1), \dots, q_k^{(r)}(n_1)\}$, $r = 1, 2, \dots$. Due to concavity of the lik, there is only one limit point $\mathbf{q}_1^{k(\infty)}(n_1) = \{q_1^{(\infty)}(n_1), \dots, q_k^{(\infty)}(n_1)\}$. The above equations are satisfied for this stationary point, i. e.,

$$\mathbf{q}_1^{k(\infty)}(n_1) = \mathbf{q}_1^k(s, n_1) .$$

The corresponding recursive method of sequential convergent calculation coordinates of the stationary point can be based on this relation. The calculation started with k initial values, e. g., $q_j^{(0)} = 1/k$, $j = 1, \dots, k$, then next k values $q_j^{(1)}(n_1)$ will be calculated and so on after k values $q_j^{(r)}(n_1)$, next k values $q_j^{(r+1)}(n_1)$ can be calculated. The process is stopped when maximal variation of $q_j^{(k+1)}(n_1) - q_j^{(k)}(n_1)$ is negligibly small. The plot in Fig. 5 illustrates convergence of the lik values at points $\mathbf{q}_1^{(r)k}$ to the global maximum.

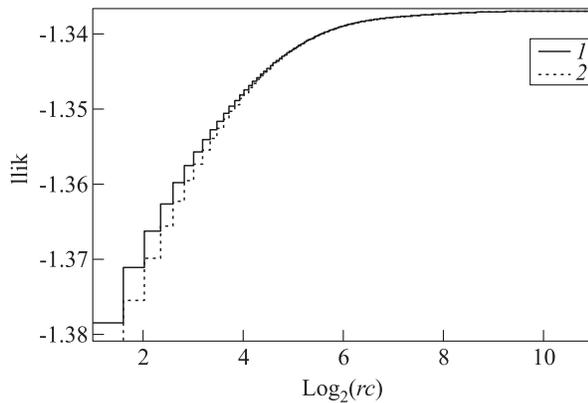


Figure 5 Dynamics of llik_{AIP} values as functions of recursion's iterations rc with two different starting points. The two starting points have q_j equal $1/23$ (1), and $j/276$ (2), $1 \leq j \leq k = 23$. Time needed for 2000 iterations is near 20 s (Intel processor 3.2 GHz)

All frequencies $\hat{w}_h(n_1) \rightarrow w_{oh}$, $h = 1, \dots, m$, a.s. as $n_1 \rightarrow \infty$. Hence, at any point $\mathbf{q}_1^k \in \mathcal{S}_{k-1}$, the $\text{llik}_{\text{AIP}}[\mathbf{q}_1^k, \hat{w}_1^m(n_1)]$ converges a.s. to the *limit* llik :

$$\text{llik}_\infty[\mathbf{q}_1^k] = \sum_{h=1}^m w_{oh} \text{Log} \left[\sum_{j \in \mathcal{C}_h} q_j \right] + e[\mathbf{w}_{o1}^m].$$

Due to compactness of \mathcal{S}_{k-1} , one has a.s. convergence in the uniform metric, i. e.,

$$\max_{\mathbf{q}_1^k \in \mathcal{S}_{k-1}} |\text{llik}_{\text{AIP}}[\mathbf{q}_1^k, \hat{w}_1^m(n_1)] - \text{llik}_\infty[\mathbf{q}_1^k]| \rightarrow 0$$

a.s. as $n_1 \rightarrow \infty$.

The concavity of the llik_{AIP} implies that

$$\begin{aligned} \text{llik}_{\text{AIP}}[\mathbf{q}_1^{k(\infty)}(n_1) | \hat{w}_1^m(n_1), \mathbf{y}_1^{n_1}] \\ = \max_{\mathbf{q}_1^k \in \mathcal{S}_{k-1}} \text{llik}_{\text{AIP}}[\mathbf{q}_1^k(n_1) | \hat{w}_1^m(n_1), \mathbf{y}_1^{n_1}]. \end{aligned}$$

If together with $r \rightarrow \infty$ also $n_1 \rightarrow \infty$, then, from the uniform convergence, one has:

$$\begin{aligned} \text{llik}_{\text{AIP}}[\mathbf{q}_1^k | \hat{w}_1^m(n_1), \mathbf{y}_1^{n_1}] \rightarrow \max_{\mathbf{q}_1^k} \text{llik}_\infty[\mathbf{q}_1^k]; \\ \sum_{j \in \mathcal{C}_h} q_j^{(r)}(n_1) \rightarrow p_{oh} = \sum_{j \in \mathcal{C}_h} q_{oj}, \text{ a.s. } r \rightarrow \infty, n_1 \rightarrow \infty. \end{aligned}$$

Note that here, it is not stated that $q_j^{(r)}(n_1) \rightarrow q_{oj}$ as r and n_1 are growing unboundedly but the sums $\sum_{j \in \mathcal{C}_h} q_{oj}$ are consistently estimable if AIP is valid. Note that the present proof of consistency is different from that suggested in [21].

One may consider $\hat{a}_{hj} = \hat{w}_h / \sum_{h' \in \mathcal{D}_j} \hat{w}_{h'}$ as an *empirical index of attractiveness* of interval \mathbf{u}_h with rounded ends. In the data discussed in section 2 for $h = 27$, this index is 0.740638; for $h = 32$, it is 0.203977; and for

$h = 2$, it is 0.015431. $a_{ohj} = w_{oh}/p_{oh}$ denote the *true index of attractiveness* of \mathbf{u}_h , $h = 1, \dots, m$.

It is possible to simulate data with n SeSeIs when AIP is valid. It is necessary to define a c.d.f. $F_o[x]$, $x > 0$, of i.i.d. WTP r.v.s X_1, \dots, X_n , and to define a collection $\mathcal{U}_m = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ with SeSeIs. Then, it is possible to find all division intervals $V_k = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ and all corresponding subsets of indices \mathcal{C}_h , $h = 1, \dots, m$, and \mathcal{D}_j , $j = 1, \dots, k$. For each i , $i = 1, \dots, n$, a value $X_i = x_i$ is simulated. If $X_i = x_i \in \mathbf{v}_{j_i}$, then an index $H_i = h_i$ is sampled randomly from \mathcal{D}_{j_i} with probability proportional to w_{h_i} , $h_i \in \mathcal{D}_{j_i}$. The indices H_1, \dots, H_n are sampled independently and the simulated data, $\mathbf{y}_1^n = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$, $\mathbf{y}_i = \mathbf{u}_{h_i}$, are obtained.

It is seen that at least for data in section 2, the AIP may not necessarily hold. Before turning to the more general assumption, let introduce an assumption on the response process which seems to fit data better.

4.6 Another behavioral assumption — on preference of the right division interval

The pilot statistical analysis (see Fig. 1) suggests that the last division intervals contain most of the values of interest x_i , $i = 1, \dots, 46$. In this case, both rounded ends and positions of x_i are essential for the individual's choice of SeSeI. If so, the AIP is not valid and another assumption is needed. To sharply outline this case, let introduce the following assumption on preference of the right division interval (ARDI).

Assumption 5. *ARDI: In each stated SeSeI $\mathbf{y}_i = (u_{Lh_i}, u_{Rh_i}) \in \mathcal{U}_m$, the value of interest x_i is in the right division interval $\mathbf{v}_{j_{h_i}\text{RD}} = (v_{Lj_{h_i}}, v_{Rh_i})$, $v_{Rj_{h_i}} = u_{Rh_i}$, $i = 1, \dots, n_1$.*

Here, one has $\{H_i = h\} \cap \{X_i \in \mathbf{u}_h\} = \{H_i = h\} \cap \{X_i \in \mathbf{v}_{j_{h}\text{RD}}\}$.

Then, $P[\{H_i = h\} \cap \{X_i \in \mathbf{v}_{j_{h}\text{RD}}\}] = w_h q_{j_{h}\text{RD}}$, $j_h \in \mathcal{C}_h$.

Note that in the empirical data, there are no intervals $\mathbf{u}_h \in \mathcal{U}_m(n_1)$ where \mathbf{v}_1 or \mathbf{v}_5 are the last division intervals. In Fig. 4, it is seen that there are no pairs $\{h, 1\}$ and $\{h, 5\}$ where \mathbf{v}_1 and \mathbf{v}_5 are the end intervals in \mathbf{u}_h , $h = 1, \dots, 46$. Here, the ML-estimates of q_{oj} , $j \neq 1, 5$, have been found.

Thus, the assumptions AIP and ARDI imply essentially different ML-estimates of several points in the estimated s.d.f. Due to consistency of these estimators, if both assumptions are true, one should expect that the estimators have to be similar. But the positions of points on s.d.f.s differ essentially. One may suspect that either at least one or even both assumptions are erroneous because the data correspond to the same sampled respondents. One possible way to eliminate this contradiction is an extension of the used questionnaire; more

collected empirical information is needed to overcome the identification problem.

5 Two-Step Intervals

To address the identification problem in a nonparametric setting, the authors propose the second-step approach. In addition, let introduce a more general assumption that has the Turnbull assumption as a special case. The data collected with a questionnaire extended in this way can be applied in obtaining consistent estimates of points on the s.d.f. of interest. The following assumption on the selection of a division interval (ASDI) containing our value of interest is proposed.

Assumption 6. ASDI: Probabilities $w_{hj}, j = 1, \dots, k, h \in \mathcal{D}_j$, to state an SeSeI \mathbf{u}_h , given $X_i \in \mathbf{v}_j \subseteq \mathbf{u}_h$, do not depend on the exact position $X_i \in \mathbf{v}_j \subseteq \mathbf{u}_h$.

Note that AIP and ARDI are the special cases of ASDI. In the case AIP, $w_h = w_{hj}, h \in \mathcal{D}_j$, and in the case ARDI, $w_h = w_{hjRD}$ where h_{jRD} corresponds to the last right division interval \mathbf{v}_{jRD} in the stated SeSeI \mathbf{u}_h . In the next subsection, mboxASDI is supposed to be valid.

5.1 The second step

Let now introduce the *extended second step* of data collection. The authors prolong random sampling of new (not yet sampled) individuals from the population \mathfrak{P} . First, each individual is to announce an interval containing his/her WTP-point. If the interval does not belong to \mathcal{U}_m , then it is not included in the collected data. If the interval \mathbf{u}_h belongs to \mathcal{U}_m , then this respondent is asked to select from the division \mathcal{V} an interval $\mathbf{v}_j \in \mathcal{V}_k, \mathbf{v}_j \subseteq \mathbf{u}_h$, containing his/her true WTP-point. The respondents may well abstain from answering this second question, it is just recorded this in the data. The collected data will be the list of triples

$$\mathbf{z}_i = \{i, \mathbf{u}_{h_i}, \text{NA}\};$$

or

$$\{i, \mathbf{u}_{h_i}, \mathbf{v}_{j_i}\}, \mathbf{d}_{2n.2} = \{\mathbf{z}_1, \dots, \mathbf{z}_{n.2}\}$$

where NA is “no answer” to the additional question. For each individual, we thus have an interval and, potentially, a selected division interval. These triples have been called *singles* and *pairs*., depending on whether the individual reported or do not reported a division interval.

The following notations have been used:

$$c_{pj} = \sum_{i=1}^{n.2} I[\mathbf{z}_i = \{i, \mathbf{u}_{h_i}, \mathbf{v}_j\}];$$

$$c_{phj} = \sum_{i=1}^{n.2} I[\mathbf{z}_i = \{i, \mathbf{u}_h, \mathbf{v}_j\}];$$

$$n_{s2} = \sum_{h=1}^m t_{sh}; \quad n_{p2} = \sum_{j=1}^k c_{pj};$$

$$t_{sh} = \sum_{i=1}^{n.2} I[\mathbf{z}_i = \{i, \mathbf{u}_h, \text{NA}\}]; \quad n.2 = n_{s2} + n_{p2}.$$

The subindexes s and p correspond to *singles* and *pairs*. Henceforth, it is supposed that all $q_j > 0, j = 1, \dots, k$, then for all sufficiently large $n.2, c_{pj} > 0$ for all $j = 1, \dots, k$.

The strongly consistent estimates of q_{oj} and w_{hj} in pairs are:

$$\check{q}_{pj} = \frac{c_{pj}}{n_{p2}} \rightarrow q_{oj};$$

$$\left\{ \hat{w}_{hj} = \frac{c_{phj}}{c_{pj}} \rightarrow w_{ohj}, \quad h \in \mathcal{D}_j, \quad j = 1, \dots, k \right\},$$

a.s. $j = 1, \dots, k$, as $n.2 \rightarrow \infty$.

5.2 The likelihood

Let use estimates \hat{w}_{hj} instead of $w_{hj}, h \in \mathcal{D}_j, j = 1, \dots, k$. The i th respondent contributes to the estimated llik in two different ways depending on whether or not the follow-up question was answered:

$$\begin{aligned} \text{ll}_i [\mathbf{q}_1^k | \mathbf{d}_{2n.2}] \\ = I[z_i = \{i, \mathbf{u}_{h_i}, \text{NA}\}] \text{Log} \left[\sum_{j \in \mathcal{C}_{h_i}} \hat{w}_{h_i j} q_j \right] \\ + I[z_i = \{i, \mathbf{u}_{h_i}, \mathbf{v}_{j_i}\}] \text{Log} [\hat{w}_{h_i j_i} q_{j_i}]. \end{aligned}$$

Let write the following, normed by $n.2 = n_{s2} + n_{p2}$, llik function corresponding to the data $\mathbf{d}_{2n.2}$ containing both singles and pairs:

$$\begin{aligned} \text{llik} [\mathbf{q}_1^k | \mathbf{d}_{2n.2}] &= \frac{1}{n.2} \sum_{i=1}^{n.2} \text{ll}_i [\mathbf{q}_1^k | \mathbf{d}_{2n.2}] \\ &= \frac{n_{s2}}{n.2} \sum_{h=1}^m \frac{t_{sh}}{n_{s2}} \text{Log} \left[\sum_{j \in \mathcal{C}_h} \hat{w}_{hj} q_j \right] \\ &\quad + \frac{n_{p2}}{n.2} \sum_{j=1}^k \frac{c_{pj}}{n_{p2}} \text{Log} [q_j] \\ &\quad + \frac{1}{n.2} \sum_{i=1}^{n.2} \text{Log} [\hat{w}_{h_i j_i}] I[\mathbf{z}_i = \{i, \mathbf{u}_{h_i}, \mathbf{v}_{j_i}\}]. \quad (1) \end{aligned}$$

This llik function of $\mathbf{q}_1^k = \{q_1, \dots, q_k\}$ corresponds to the all data, with pairs and singles, collected on the second step.

The following properties of this llik are collected in the following theorem and corollary:

Theorem 5.1. For every sufficiently large $n_{.2}$, the llik is concave on \mathcal{S}_{k-1} .

Corollary 5.1. The llik attains its maximum at a point $\check{\mathbf{q}}_M = \{\check{q}_{1M}, \dots, \check{q}_{kM}\} \in \mathcal{S}_{k-1}$ and $\check{\mathbf{q}}_M$ is its unique stationary point in \mathcal{S}_{k-1} .

It is interesting to find the stationary point $\check{\mathbf{q}}_M$ corresponding to the maximum value of the llik in the multidimensional set $\check{\mathcal{S}}_{k-1}$. Let suppose that $n_{.2}$ so large that all $c_{pj} > 0, j = 1, \dots, k$. Then, Lagrange method with one multiplier λ can be applied. Let consider the following Lagrange function

$$\varphi_L[\mathbf{q}_1^k, \lambda] = \text{llik}[\mathbf{q}_k | \mathbf{d}\mathbf{2}_{n_{.2}}] + \lambda(q_1 + \dots + q_k^k).$$

Components of a stationary point q_1, \dots, q_k are the solutions of the following equations:

$$q_j = \frac{n_{p2}}{n_{.2}} \frac{t_{pj}}{n_{p2}} + \frac{n_{s2}}{n_{.2}} \sum_{h \in \mathcal{D}_j} \frac{t_{sh}}{n_{s2}} \frac{\hat{w}_{hj} q_j}{\sum_{j' \in \mathcal{C}_h} \hat{w}_{hj'} q_{j'}}, j = 1, \dots, k. \quad (2)$$

Then, all first-order partial derivatives $\partial \text{llik}[\mathbf{q}_k] / \partial q_j = 0, j = 1, \dots, k$, and $\mathbf{q}_k = \{q_1, \dots, q_k\}$ together represent a stationary point. From Corollary 5.1, one has $\mathbf{q}_1^k = \check{\mathbf{q}}_M$.

The proofs of Theorem 5.1 and Corollary 5.1 are based on ideas similar to that used in Subsection 4.4.

5.3 Recursive method for numerical analysis

From (2), one may find the following recursion:

$$q_j^{(r+1)}(n_{.2}) = \frac{n_{p2}}{n_{.2}} q_j^{(1)} + \frac{n_{s2}}{n_{.2}} \sum_{h \in \mathcal{D}_j} \frac{t_{sh}}{n_{s2}} \frac{\hat{w}_{hj} q_j^{(r)}(n_{.2})}{\sum_{j' \in \mathcal{C}_h} \hat{w}_{hj'} q_{j'}^{(r)}(n_{.2})}, r = 1, 2, \dots; \quad (3)$$

$$q_j^{(1)}(n_{.2}) = \check{q}_{pj}; \quad (4)$$

$$\hat{w}_{sh}(n_{.2}) = \frac{1}{n_{s2}} \sum_{i=1}^{n_{.2}} I[z_i = \{i, \mathbf{u}_h, \text{NA}\}].$$

As in section 4.5, one has $\sum_{j=1}^k q_j^{(r)}(n_{.2}) = 1$ which implies the following Corollary.

Corollary 5.2. For any sufficiently large $n_{.2}$, this recursion generates the sequence of points $\mathbf{q}_1^{(r)k} = \{q_1^{(r)}, \dots, q_k^{(r)}\} \in \mathcal{S}_{k-1}, r = 1, 2, \dots$, which converge to the unique stationary point $\check{\mathbf{q}}_M = \{\check{q}_{1M}, \dots, \check{q}_{kM}\}$.

Note. For any $r \geq 1, q_j^{(r)} \geq (n_{p2}/n_{.2})q_j^{(1)}$ if $n_{.2}$ is sufficiently large.

If $n_{.2} \rightarrow \infty$, then all below following estimates a.s. converge to their limits:

$$\frac{n_{p2}}{n_{.2}} \rightarrow \alpha_{p2} > 0; \quad \frac{n_{s2}}{n_{.2}} \rightarrow \alpha_{s2} > 0;$$

$$\alpha_{p2} + \alpha_{s2} = 1; \quad \frac{t_{sh}}{n_{s2}} \rightarrow w_h;$$

$$\frac{t_{pj}}{n_{p2}} \rightarrow q_j; \quad \frac{t_{phj}}{n_{p2}} \rightarrow w_{hj} q_j, \quad j = 1, \dots, k, h \in \mathcal{D}_j.$$

Then, the llik on the compact set \mathcal{S}_{k-1} uniformly converges, as $n_{.2} \rightarrow \infty$, to the nonrandom function of \mathbf{q}_1^k :

$$\begin{aligned} \text{llik}^{(\infty)}[\mathbf{q}_1^k] &= \alpha_{p2} \sum_{j=1}^k q_j \text{Log}[q_j] \\ &+ \alpha_{s2} \sum_{h=1}^m w_h \text{Log} \left[\sum_{j \in \mathcal{C}_h} w_{hj} q_j \right] + C_p \end{aligned}$$

where the constant $C_p = \alpha_{p2} \sum_{j=1}^k \sum_{j \in \mathcal{D}_j} q_j \text{Log}[w_{hj}]$. Let call the concave function $\text{llik}^{(\infty)}[\mathbf{q}_1^k]$ a *limit llik*. Let summarize these results in the following theorem.

Theorem 5.2. Suppose that Assumptions 1–3 and 6 are valid and the sizes n_{p2} and n_{s2} of collected pairs and singles are growing unboundedly as $n_{.2} \rightarrow \infty$. Then, for any sufficiently large n_{p2} and n_{s2} , the ML-estimator $\check{\mathbf{q}}_1^k = \{\check{q}_1, \dots, \check{q}_k\}$ of $\check{\mathbf{q}}_M$ (based on the data $\mathbf{d}\mathbf{2}_{n_{.2}}$ with singles and pairs) of the llik $\text{llik}[\mathbf{q}_1^k | \mathbf{2}_{n_{.2}}]$ exists, is strongly consistent, and can be found as the stationary point of recursion (3).

Note that this inference can only be applied to the respondents in \mathfrak{P}_c who will choose $\mathbf{u}_h \in \mathcal{U}_m$.

6 Numerical Experiments

To get some insights into the properties of these estimators, some numerical experiments have been carried out. In these experiments, simulated data have been used with the same set $\mathcal{U}_m, m = 46$, of SeSeI intervals and the set $\mathcal{V}_k, k = 23$, of division intervals. The authors decided to take the WTP-distribution to be a p_{WE} -mixture WE of the Weibull distribution $W(a, b)$ and the Exponential distribution $E(m_1)$ with parameters $p_{\text{WE}} = 0.8160, a = 74.8992, b = 1.8374$, and $m_1 = 254.7344$:

$$\text{sf}_{\text{WTP}}[x] = p_{\text{WE}} e^{-(x/a)^b} + (1 - p_{\text{WE}}) e^{-x/m_1}. \quad (5)$$

The motivation for this choice is further explained in [7] where five “behavioral models” have been also introduced. Each of these provides one possible model of how individuals choose the intervals, a process which is essentially unknown. The authors have decided to use a behavior model described by the array $\mathbb{W}_{mk} = (w_{hj})$, for any $j = 1, \dots, k$, and $h \in \mathcal{D}_j, w_{hj} = t_h / (\sum_{h' \in \mathcal{D}_j} t_{h'})$

where t_h are the numbers of repetitions $\mathbf{u}_h \in \mathcal{U}_m$ in the first step of data collection, $k = 1, \dots, m$. t_h were taken as they were in the real data.

Let $\omega_1, \omega_2, \dots, \omega_{4i-3}, \omega_{4i-2}, \omega_{4i-1}, \omega_{4i}, \dots$ be a sequence of real values generated by a random number generator. Let use these values as follows. If $\omega_{4i-3} \leq p_{WE}$, then ω_{4i-2} is transformed to x_i which is a value of an r.v. X_i with the Weibull s.d.f. $e^{-(x/a)^b}$. Otherwise, if $\omega_{4i-3} > p_{WE}$, then ω_{4i-3} is transformed to r.v. X_i with the exponential s.d.f. e^{-x/m_1} . Further, it will be found which of divisions intervals contains x_i . Let $\mathbf{v}_j \ni x_i$, then $\mathcal{D}_j = \{h_1(j), h_2(j), \dots, h_{d_j}(j)\}$ where index $h_d(j)$ corresponds to probability w_{h_j} , and d_j is the number of all indexes in \mathcal{D}_j . Remind that $\sum_{h \in \mathcal{D}_j} w_{h_j} = 1$. Interval $(0, 1]$ is the union of disjoint intervals $(0, w_{h_1(j),j}]$, $(w_{h_1(j),j}, \sum_{c=1}^2 w_{h_c(j),j}]$, \dots , $(\sum_{c=1}^{d_j-1} w_{h_c(j),j}, 1]$. The event $\{\omega_{4i-1} \in (\sum_{c=1}^{d_j-1} w_{h_c(j),j}, \sum_{c=1}^{d_j} w_{h_c(j),j}]\}$ means that $\mathbf{u}_{h_c(j)} \supseteq \mathbf{v}_j$ has been stated as SeSeI interval, $h_c(j) \in \mathcal{D}_j$. At last, the event $\omega_{4i} \leq 2/3$ implies that the i th element in the data \mathbf{d}_{2tn_2} is the triplet $\{i, \mathbf{u}_h, \text{NA}\}$, i. e., the single triplet. If $\omega_{4i} > 2/3$, the i th data element is the triplet $\{i, \mathbf{u}_h, \mathbf{v}_j\}$ with the pair $\{\mathbf{u}_h, \mathbf{v}_j\}$, $h = h_c(j)$.

To illustrate consistency of the suggested estimates, the size of simulated data \mathbf{d}_{2tn_2} was selected rather large with $n_2 = 9000$ triplets. There were $n_{s2} = 5856$ singles and $n_{p2} = 3144$ pairs. Let add index t to underline that data are considered as “true” data. By using triplets with pairs $\{i, \mathbf{u}_{h_i}, \mathbf{v}_{j_i}\}$ in \mathbf{d}_{2tn_2} , let find estimates $\check{q}_{tj} > 0$ for the true probabilities $\check{q}_{oj} = P_o[X_i \in \mathbf{v}_j]$, $j = 1, \dots, 23$. To improve accuracy, one can find, by using all data set \mathbf{d}_{2tn_2} and by applying r ($r \leq 10$) iterations, ML-estimates $\check{\mathbf{q}}_t^{(10)} = \{\check{q}_{t1}^{(10)}, \dots, \check{q}_{t23}^{(10)}\}$, $k = 23$, corresponding the lik (1). Let use these estimates to illustrate consistency estimates of true points $\{v_{Rj}, s_{oj}^k\}$ on s.d.f. (5). Here, true and estimated points are $\{v_{Rj}, s_o^k\}$, $\{v_{Rj}, \check{s}_{tj}^k\}$ $s_o^k = \sum_{j'=j}^k q_{oj'}$; and $\check{s}_{tj}^k = \sum_{j'=j}^k \check{q}_{tj'}^{(10)}$. Here, the s.d.f. (5) can be used, because only stated $\mathbf{u}_h \in \mathcal{U}_m$ are considered. It means that only intervals in \mathcal{U}_m will be stated by any respondent in the investigated population \mathfrak{P} . In Fig. 6, it is seen that ML-estimates based on both pairs and singles are rather accurate. The usage of estimates based only on pairs is less accurate. To illustrate utility of usage of both pairs and singles, $C = 2000$ i.i.d. copies of data $\mathbf{d}_{2tn_2}^c$ have been simulated, each such copy had size $n_2 = 9000$. The same values of parameters p_w, a, b, m_1 , and p_{NA} were used.

The mean value is of central interest in contingent valuation surveys. If a particular parametric distribution is used, one can, of course, derive the estimate of the

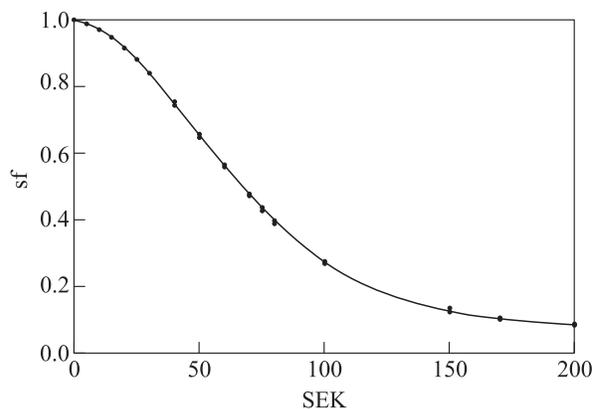


Figure 6 The true WTP-survival function is shown over interval $[0, 200]$ together with its true points $\{v_{Rj}^k, q_j^k\}$ and the approximated ML-estimated points $\{v_{Rj}^k, \check{q}_j^{(10)k}\}$, $1 \leq j \leq 16$, $n_2 = 9000$

mean by estimating parameters. In a nonparametric setting, let propose

$$mm_{o1} = v_{L1} + \sum_{j=1}^k (q_{oj+1}^k + \frac{1}{2} q_{oj}^k) (v_{Rj} - v_{Lj})$$

and call it as *medium mean value*. The medium mean value mm_{o1} can be considered as acceptable approximation to the mean value $m_{o1} = E[X_i]$ of the WTP-e.d.f.

One has the following strongly consistent and asymptotically unbiased ML-estimate of mm_{o1} :

$$\begin{aligned} \check{m}_{1ps} &= v_{L1} + \sum_{j=1}^k \left(\check{s}_{j+1,ps}^k + \frac{1}{2} \check{q}_{jps} \right) (v_{Rj} - v_{Lj}) \quad (6) \end{aligned}$$

where $\check{s}_{j+1,ps}^k = \sum_{i=j+1}^k \check{q}_{ips}$, $j = 1, \dots, k$, and $\check{q}_{k+1,ps} = 0$. Note that \check{m}_{1ps} is the area below the broken line, connected by intervals with end points $\{v_{Rj}, \check{s}_{j+1,ps}^k\}$, $j = 1, \dots, k - 1$, and $\{v_{Ro,1}\}$, $\{v_{Rk}, 0\}$ on the estimated WTP-survival function.

For each copy of data $\mathbf{d}_{2tn_2}^c$, the authors apply r iterations of recursion (3) and obtain sufficiently accurate approximation $\check{\mathbf{q}}_k^{c(r)}$ of the ML-estimate $\check{\mathbf{q}}_{kM}^c$. Here, all copies of data have been used, i. e., both pairs and singles. One can also find the ML-estimates \check{q}_{pk}^c of q_k by using (4) and the part of data $\mathbf{d}_{2n_2}^c$ with triples containing pairs (3). By substituting $\check{\mathbf{q}}_{pk}^c$ and $\check{\mathbf{q}}_k^{c(r)}$ in (6), one obtains estimates of WTP-medium mean values \check{m}_{1p}^c and $\check{m}_{1ps}^{c(r)}$, $c = 1, \dots, R$. To underline the gain of usage of the whole data $\mathbf{d}_{2n_2}^c$ compared with only the part of all pairs, two lists with deviations have been calculated:

$$\left. \begin{aligned} \check{d}v_{1ps}^{c(r)} &= \check{m}_{1ps}^{c(r)} - mm_{o1} ; \\ \check{d}v_{1p}^c &= \check{m}_{1p}^c - mm_{o1}, \quad c = 1, \dots, C. \end{aligned} \right\} \quad (7)$$

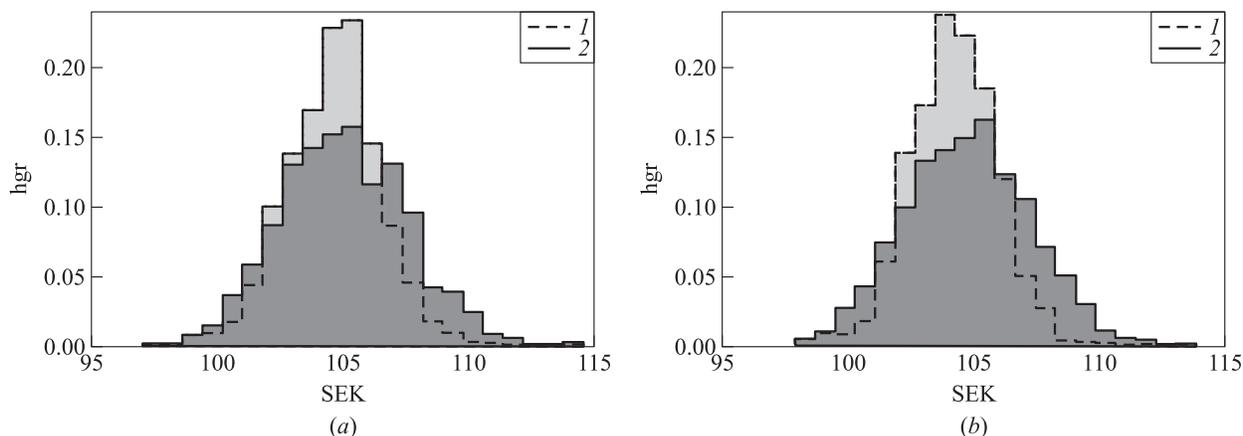


Figure 7 Two pairs of histograms, corresponding to $C = 2000$ deviations $\text{dev}_{1ps}^c = \check{m}m_{1ps}^c - mm_{o1}$ (1) and $\text{dev}_{1p}^c = \check{m}m_{1p}^c - mm_{o1}$ (2) (a) and $\text{dev}_{1ps}^{*c} = \check{m}m_{1ps}^{*c} - \check{m}m_{1ps}^c$ (1) and $\text{dev}_{1p}^{*c} = \check{m}m_{1p}^{*c} - \check{m}m_{1p}^c$ (2) (b): 1 — triples with both pairs and singles; and 2 — triples with only pairs, $n_2 = 9000$, $C = C^* = 2000$

The deviations (7) can be considered as values of i.i.d. r.v.s. Then, their standard deviation estimates are:

$$\check{std}_{1ps}^{(r)} = 1.780; \quad \check{std}_{1p} = 2.507.$$

Hence, the estimates of mm_{o1} based on both pairs and singles are more exact than based only on pairs.

It is useful to know the following empirical distribution functions (e.d.f.s) of deviations:

$$\left. \begin{aligned} \check{edf}_{1sp}^{(r)}[x] &= \frac{1}{C} \sum_{c=1}^C I \left[\check{dev}_{1sp}^{c(r)} \leq x \right]; \\ \check{edf}_{1p}[x] &= \frac{1}{C} \sum_{c=1}^C I \left[\check{dev}_{1p}^c \leq x \right]. \end{aligned} \right\} \quad (8)$$

The lists with deviations (8) can be presented as a paired histograms, composed from rectangles with low edges intervals $I_{r'} = (2r', 2(r'+1)]$, $r' = 0, 1, \dots$ on the x -axis. The length of $I_{r'}$ is 2 SEK. The rectangles side intervals are proportional to the numbers of $\check{m}m_{1ps}^{c(r)} \in I_{r'}$ and $\check{m}m_{1p}^c \in I_{r'}$, $c = 1, \dots, C$. Two corresponding components of paired histograms are shown with the dashed and solid contour lines, respectively, in Fig. 7a.

Note that total size of 2000 samples with size $n_2 = 9000$ is rather large. Nearly the same accuracy of pairs histogram one can obtain by using 2000 independently resampled copies from only one sample (!) with the same size 9000. Such resampled copies of $\mathbf{d2}_{n_2}^c$ is denoted $\mathbf{d2}_{n_2}^{*c}$, $c = 1, \dots, 2000$, and use them farther similarly as if they are common samples from a population \mathfrak{F} . Let apply recursion (3) with r sufficient iterations and find $\mathbf{q}_k^{*c(r)}$. By substituting $\check{\mathbf{q}}_{kps}^{*c}$ and $\check{\mathbf{q}}_{kps}^{*c(r)}$ in relation similar to (6), one obtains $\check{m}m_{1ps}^{*c(r)}$ and $\check{m}m_{1p}^c$. Their deviations are similar to that given in (7):

$$\begin{aligned} \text{dev}_{1ps}^{*c(r)} &= \check{m}m_{1ps}^{*c(r)} - \check{m}m_{1ps}^c; \\ \text{dev}_{1p}^{*c} &= \check{m}m_{1p}^{*c} - \check{m}m_{1p}^c. \end{aligned}$$

The pairs histogram obtained by resampled copies of data $\mathbf{d2}_{n_2}$ are shown in Fig. 7b. Both paired histograms are similar which reflects consistency of the resamplings method. Sufficient and necessary assumptions for consistency the resampling method were given in [22, 23].

7 Concluding Remarks

The present authors have provided an approach to the theory of SeSeIs obtained in surveys. These SeSeIs should have wide application, not the least in survey research, where they are already being used. The drawbacks of alternative elicitation mechanisms are well-documented in the literature and the SeSeIs can overcome some of them. Or so, the authors have argued. The new 2-step approach suggested has not been tested empirically. The present simulation exercise provides some indications that merit trying this approach in the field. A key assumption in the 2-step approach (an assumption currently always made when using any other elicitation mechanisms) is that the suggested interval/information in the second step does not have any impact on the censored variable. In other words, the authors assume that anchoring is not a problem in this context. While anchoring is less of a problem in the context compared to, say, standard bracketing approaches (because one has information from the “undistorted by anchoring” intervals), it is still an assumption that needs to be scrutinized. In future research, this assumption will be studied in lab and field experiments.

A recent bibliography on surveys targeting contingent valuation has about 7.500 studies from 130 countries [11]. Big data sets are also collected in surveys

health status of individuals, during aging and retirement. censoring intervals are common in diversity of biomedical data sets [24]. Hopefully, interval data can also be useful in expert assessments of economic data, in durations of individuals' unemployment, in usage of futures contracts at bourses, and in psychometrics evaluation of individuals' relation to different risks.

References

1. Manski, C. F. 1999. *Identification problems in the social sciences*. Harvard University Press. 194 p.
2. Morgan, M. G., and M. Henrion. 1990. *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press. 325 p.
3. Billard, L., and E. Diday. 2006. *Symbolic data analysis: Conceptual statistics and data mining*. Wiley ser. in computational statistics. Wiley. Vol. 636. 330 p.
4. Manski, C. F., and F. Molinari. 2010. Rounding probabilistic expectations in surveys. *J. Bus. Econ. Stat.* 28:219–231.
5. Johansson, P.-O., and B. Kriström. 2012. *The economics of evaluating water projects. Hydroelectricity versus other uses*. Heidelberg: Springer. 135 p.
6. Belyaev, Y., and B. Kriström. 2010. Approach to analysis of self-selected interval data. Umeå: CERE. Working Paper 2010:2. 1–34. Available at: <http://www.cere.se/se/forskning/publikationer/155-approach-to-analysis-of-self-selected-interval-data.html> (accessed September 8, 2015).
7. Belyaev, Y., and B. Kriström. 2012. Two-step approach to self-selected interval data in elicitation surveys. Umeå: CERE. Working Paper 2012:10. 1–46. Available at: <http://www.cere.se/se/forskning/publikationer/386-two-step-approach-to-self-selected-interval-data-in-elicitation-surveys.html> (accessed September 8, 2015).
8. Turnbull, B. W. 1974. Nonparametric estimation of a survivorship function with doubly censored data. *J. Am. Stat. Assoc.* 69:169–173.
9. Turnbull, B. W. 1976. The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. Roy. Stat. Soc. B* 38:290–295.
10. Manski, C. F. 2003. *Partial identification of probability distributions*. Springer ser. in statistics. Springer. 196 p.
11. Carson, R. 2012. *Contingent valuation: A comprehensive bibliography and history*. Edward Elgar Publishing. 464 p.
12. Håkansson, C. 2008. A new valuation question — analysis of and insights from interval open ended data in contingent valuation. *Environ. Resour. Econ.* 39(2):175–188.
13. Broberg, T., and R. Brännlund. 2008. An alternative interpretation of multiple bounded WTP data—certainty dependent payment card intervals. *Energy Resour. Econ.* 30:555–567.
14. Mahieu, P., P. Riera, B. Kriström, R. Brännlund, and M. Giergiczny. 2014. Exploring the determinants of uncertainty in contingent valuation surveys. *J. Environ. Econ. Policy* 3(2):186–200. Available at: <http://dx.doi.org/10.1080/21606544.2013.876941> (accessed September 3, 2015).
15. Good, I. J. 1953. The population frequencies of species and the estimation of population parameters. *Biometrika* 40(3-4):237–264.
16. Belyaev, Y., and B. Kriström. 2013. Analysis of contingent valuation data with self-selected rounded WTP-intervals collected by two-steps sampling plans. *9th Tartu Conference on Multivariate Statistics and 20th IWMS Proceedings*. Tartu: World Scientific. 48–60.
17. Gomez, J., M. Calle, and R. Oller. 2004. Frequentist and Bayesian approaches for interval-censored data. *Stat. Pap.* 45:139–173.
18. Gomez, J., M. Calle, R. Oller, and K. Langhor. 2009. Tutorial on methods for interval-censored data and their implementations in *R*. *Stat. Model.* 9(4):259–297.
19. Gentleman, R., and C. J. Geyer. 1994. Maximum likelihood for interval censored data: Consistency and computation. *Biometrika* 81(3):618–623.
20. Brinkhuis, J., and V. Tihomirov. 2005. *Optimization: Insights and applications*. Princeton—Oxford: Princeton University Press. 680 p.
21. Jammalamadaka, S. R., and V. Mangalam. 2003. Non-parametric estimation for middle-censored data. *J. Non-parametr. Stat.* 15:253–265.
22. Belyaev, Y. K., and L. Nilsson. 1997. Parametric maximum likelihood estimators. Department of Mathematical Statistics, Umeå University. Research Report 1997-15. 1–28.
23. Belyaev, Y. K. 2003. Necessary and sufficient conditions for consistency of resampling. Sweden: Centre of Biostochastics, Swedish University of Agricultural Sciences. Research Report 2003-1. 1–26. Available at: http://biostochastics.slu.se/publikationer/dokument/Report2003_1.pdf (accessed September 3, 2015).
24. Klein, J. P., and M. L. Moeschberger. 2003. *Survival analysis: Techniques for censored and truncated data*. New York, N.Y.: Springer-Verlag. 536 p.

Received June 30, 2015

Contributors

Belyaev Yuri K. (b. 1932) — Doctor of Science in physics and mathematics, professor, Umeå University, Petrus Laestadius väg, Umeå SE-901 87, Sweden; yuri.belyaev@umu.se

Kriström Bengt (b. 1960) — PhD in economics, Director, Department of Forest Economics, Center for Environmental and Resource Economics (CERE), Swedish University of Agricultural Sciences, Umeå SE-901 83, Sweden; bengt.kristrom@umu.se

АНАЛИЗ ОБЗОРНЫХ ОБСЛЕДОВАНИЙ, СОДЕРЖАЩИХ ЦЕНЗУРИРОВАННЫЕ ДАННЫЕ В ОКРУГЛЕННЫХ ИНТЕРВАЛАХ

Ю. К. Беляев¹, Б. Кристрём²

¹Университет Умеа, Умеа, Швеция

²Центр экономики природных ресурсов и окружающей среды Шведского университета сельскохозяйственных наук, Умеа, Швеция

Аннотация: Предложены методы статистического анализа данных, собираемых в обзорных обследованиях. Предложенный подход общий и применим в большинстве случаев, когда в собранных данных содержатся интервалы. Такие данные типичны во многих исследованиях: в анализе надежности изделий и систем, продолжительности жизни в демографии, в медицине и экономике, в обзорных обследованиях мнения населения и др. Имеются серьезные причины для интенсивного использования данных с интервалами. Наиболее общей причиной является невозможность наблюдения точных значений. Природа исследуемых интервалов необычна. Так называемые самовыбираемые интервалы без каких-либо ограничений свободно выбираются субъектами обследований. Концы таких интервалов могут быть округлены. Предлагается обобщение продуктивного подхода к статистическому анализу в общей схеме цензурирования, предложенной Б. В. Турнбуллом. Объяснено и обобщено основное условие независимости в анализе Турнбулла. Предложено правило останковки выборочного процесса на основе достигнутого значения вероятности покрытия. Введение дополнительного (второго) вопроса всем выбранным респондентам дает возможность получения более точной оценки характеристик искомого распределения. Дано обоснование методов информатики, применяемых для анализа статистических данных, содержащих самовыбираемые интервалы. Эти методы дают возможность частичной идентификации искомого непараметрических распределений. Дано описание статистических моделей данных, допускающих зависимость выбора интервалов от положения в них точных значений. Получены рекурсии быстрого вычисления оценок максимального правдоподобия для характеристик искомого распределений. Приведены результаты применения предлагаемых методов, подтверждающие их полезность в анализе смоделированных данных, содержащих самовыбираемые интервалы.

Ключевые слова: выявляющие выборочные обследования; случайный выбор; округление; якорность; вероятность покрытия; правдоподобие; рекурсия; максимизация; случайный перевыбор

DOI: 10.14357/19922264150301

Литература

1. *Manski C. F.* Identification problems in the social sciences. — Harvard University Press, 1999. 194 p.
2. *Morgan M. G., Henrion M.* Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis. — Cambridge University Press, 1990. 325 p.
3. *Billard L., Diday E.* Symbolic data analysis: Conceptual statistics and data mining. Wiley ser. in computational statistics. — Wiley, 2006. Vol. 636. 330 p.
4. *Manski C. F., Molinari F.* Rounding probabilistic expectations in surveys // *J. Bus. Econ. Stat.*, 2010. Vol. 28. P. 219–231.
5. *Johansson P.-O., Kriström B.* The economics of evaluating water projects. Hydroelectricity versus other uses. — Heidelberg: Springer, 2012. 135 p.
6. *Belyaev Y., Kriström B.* Approach to analysis of self-selected interval data. Umeå: CERE, 2010. Working Paper 2010:2. P. 1–34. <http://www.cere.se/se/forskning/publikationer/155-approach-to-analysis-of-self-selected-interval-data.html>.
7. *Belyaev Y., Kriström B.* Two-step approach to self-selected interval data in elicitation surveys. Umeå: CERE, 2012. Working Paper 2012:10. P. 1–46. <http://www.cere.se/se/forskning/publikationer/386-two-step-approach-to-self-selected-interval-data-in-elicitation-surveys.html>.
8. *Turnbull B. W.* Nonparametric estimation of a survivorship function with doubly censored data // *J. Am. Stat. Assoc.*, 1974. Vol. 69. P. 169–173.
9. *Turnbull B. W.* The empirical distribution function with arbitrarily grouped, censored and truncated data // *J. Roy. Stat. Soc. B*, 1976. Vol. 38. P. 290–295.
10. *Manski C. F.* Partial identification of probability distributions. Springer ser. in statistics. — Springer, 2003. 196 p.
11. *Carson, R.* Contingent valuation: A comprehensive bibliography and history. — Edward Elgar Publishing, 2012. 464 p.
12. *Håkansson C.* A new valuation question — analysis of and insights from interval open ended data in contingent valuation // *Environ. Resour. Econ.*, 2008. Vol. 39. No. 2. P. 175–188.
13. *Broberg T., Brännlund R.* An alternative interpretation of multiple bounded WTP data-certainty dependent payment card intervals // *Energy Resour. Econ.*, 2008. Vol. 30. P. 555–567.

14. Mahieu P., Riera P., Kriström B., Brännlund R., Giergiczy M. Exploring the determinants of uncertainty in contingent valuation surveys // *J. Environ. Econ. Policy*, 2014. <http://dx.doi.org/10.1080/21606544.2013.876941>.
15. Good I. J. The population frequencies of species and the estimation of population parameters // *Biometrika*, 1953. Vol. 40. Iss. 3-4. P. 237–264.
16. Belyaev Y., Kriström B. Analysis of contingent valuation data with self-selected rounded WTP-intervals collected by two-steps sampling plans // 9th Tartu Conference on Multivariate Statistics and 20th IWMS Proceedings. — Tartu: World Scientific, 2013. P. 48–60.
17. Gomez J., Calle M., Oller R. Frequentist and Bayesian approaches for interval-censored data // *Stat. Pap.*, 2004. Vol. 45. P. 139–173.
18. Gomez J., Calle M., Oller R., Langhor K. Tutorial on methods for interval-censored data and their implementations in *R* // *Stat. Model.*, 2009. Vol. 9. No. 4. P. 259–297.
19. Gentleman R., Geyer C. J. Maximum likelihood for interval censored data: Consistency and computation // *Biometrika*, 1994. Vol. 81. No. 3. P. 618–623.
20. Brinkhuis J., Tihomirov V. Optimization: Insights and applications. — Princeton—Oxford: Princeton University Press, 2005. 680 p.
21. Jammalamadaka S. R., Mangalam V. Non-parametric estimation for middle-censored data // *J. Nonparametr. Stat.*, 2003. Vol. 15. P. 253–265.
22. Belyaev Y. K., Nilsson L. Parametric maximum likelihood estimators. Department of Mathematical Statistics, Umeå University, 1997. Research Report 1997-15. P. 1–28.
23. Belyaev Y. K. Necessary and sufficient conditions for consistency of resampling. Sweden: Centre of Biostochastics, Swedish University of Agricultural Sciences, 2003. Research Report 2003-1. P. 1–26. http://biostochastics.slu.se/publikationer/dokument/Report2003_1.pdf.
24. Klein, J. P., M. L. Moeschberger. 2003. Survival analysis: Techniques for censored and truncated data. — New York, N.Y., USA: Springer-Verlag. 536 p.

Поступила в редакцию 30.06.2015

АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЙ МЕТОДОМ ОРТОГОНАЛЬНЫХ РАЗЛОЖЕНИЙ В НЕЛИНЕЙНЫХ СТОХАСТИЧЕСКИХ СИСТЕМАХ НА МНОГООБРАЗИЯХ*

И. Н. Синицын¹

Аннотация: Рассматриваются вопросы оценки точности и чувствительности алгоритмов параметрического на базе методов ортогональных разложений (МОР) и квазиоментов (МКМ) аналитического моделирования одномерных распределений в стохастических системах (СтС) на многообразиях (МСтС) с винеровскими и пуассоновскими шумами. На основе обобщенной формулы Ито выведены уравнения точности и чувствительности алгоритмов МОР для МСтС. Особое внимание уделено МКМ для нормальных МСтС. Полученные методические результаты положены в основу разрабатываемого инструментального символического программного обеспечения в среде MATLAB-MAPLE. Рассмотрены вопросы сокращения числа уравнений МОР и МКМ. В качестве иллюстративного примера изучена одномерная нелинейная МСтС с мультипликативным гауссовским (нормальным) белым шумом. Для типовых задач оценки надежности и безопасности технических систем предложены алгоритмы оценки точности и чувствительности. Сформулированы некоторые возможные обобщения.

Ключевые слова: метод аналитического моделирования (МАМ); метод квазиоментов (МКМ); метод ортогональных разложений (МОР); обобщенная формула Ито; плотность одномерного распределения; полиномы Эрмита; стохастическая система на многообразиях (МСтС); уравнения точности МОР и МКМ; уравнения чувствительности МОР и МКМ

DOI: 10.14357/19922264150302

1 Введение

Известные методы аналитического моделирования (МАМ) распределений процессов в СтС, описываемых дифференциальными стохастическими уравнениями Ито с винеровскими и пуассоновскими шумами, основанные на параметризации их распределений, подробно описаны в [1, 2]. Обобщение результатов [1, 2] на случай многоканальных круговых и сферических СтС выполнено в [3–12]. Статья [8] посвящена развитию дискретных методов параметрического статистического и аналитического моделирования в МСтС. В ней рассмотрены уравнения МСтС, приближенные методы статистического моделирования (МСМ) различной точности и МАМ, основанные на ортогональных разложениях. Подробно развита нелинейная корреляционная теория МСМ и МАМ. В [9–12] развиты методы и алгоритмы аналитического моделирования гауссовских (нормальных) процессов в МСтС.

Применительно к типовым задачам оценки надежности и безопасности технических систем рассмотрены вопросы оценки точности и чувствитель-

ности параметрических алгоритмов МАМ одномерных распределений процессов по МОР и МКМ в нелинейных негауссовских дифференциальных МСтС.

Статья включает в себя: введение, 4 раздела, заключение и список литературы. В разд. 2 получены уравнения МОР (теорема 1). Уравнения МКМ как частный случай получены из теоремы 1 при использовании многомерных полиномов Эрмита и составляют основу теоремы 2 разд. 3. В разд. 4 получены уравнения точности и чувствительности МОР и МКМ. Раздел 5 содержит иллюстративный пример. В заключении кратко сформулированы основные результаты и указаны некоторые возможные обобщения.

2 Метод ортогональных разложений

Как известно [1, 8], для дифференциальных СтС в конечномерных пространствах, в том числе и МСтС, используется дифференциальное стохастическое уравнение Ито вида

* Работа выполнена при финансовой поддержке РФФИ (проект 15–07–02244).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sinitsin@dol.ru

$$dY_t = a(Y_t, \Theta, t) dt + b(Y_t, \Theta, t) dW_0 + \int_{R_0^q} c(Y_t, \Theta, t, v) dP^0(\Theta, t, dv). \quad (1)$$

Здесь Y_t — p -мерный вектор состояния, $Y_t \in \Delta^y$ (Δ^y — многообразие состояний); Θ — вектор случайных параметров размерности p^Θ ; $a = a(y_t, \Theta, t)$ и $b = b(y_t, \Theta, t)$ — известные $(p \times 1)$ -мерная и $(p \times r)$ -мерная функции вектора Y_t и времени t ; $W_0 = W_0(\Theta, t)$ — r -мерный винеровский стохастический процесс интенсивности $\nu_0 = \nu_0(\Theta, t)$; $c(y_t, \Theta, t, v)$ — $(p \times 1)$ -мерная функция y_t, t и вспомогательного $(q \times 1)$ -мерного параметра v ; $\int dP^0(\Theta, t, A)$ — центрированная пуассоновская мера:

$$\int_{\Delta} dP^0(\Theta, t, A) = \int_{\Delta} dP(\Theta, t, A) - \int_{\Delta} \nu_P(\Theta, t, A) dt,$$

где $\int_{\Delta} dP(\Theta, t, A)$ — число скачков пуассоновского процесса в интервале времени Δ ; $\nu_P(\Theta, t, A)$ — интенсивность пуассоновского процесса $P(t, A)$; A — некоторое борелевское множество пространства R_0^q с выколотым началом координат. Интеграл (1) в общем случае распространяется на все пространство R_0^q с выколотым началом координат. Начальное значение Y_0 вектора Y_t представляет собой случайную величину, не зависящую от приращений винеровского процесса $W_0(\Theta, t)$ и пуассоновского процесса $P(\Theta, t, A)$ на интервалах времени $\Delta = (t_1, t_2]$, следующих за t_0 , $t_0 \leq t_1 \leq t_2$, для любого множества A .

Для вычисления вероятностей событий, связанных со случайными функциями, в прикладных задачах достаточно знания многомерных распределений. Поэтому центральной задачей теории МСтС является задача вероятностного анализа одномерных распределений процессов, удовлетворяющих дифференциальному стохастическому уравнению Ито (1) при соответствующих начальных условиях. В теории МСтС различают два принципиально разных подхода к вычислению распределений. Первый общий подход основан на статистическом моделировании, т. е. на прямом численном решении (1) с последующей статистической обработкой результатов. Второй общий подход основан на теории непрерывных марковских процессов и предполагает аналитическое моделирование, т. е. решение детерминированных уравнений в функциональных пространствах (уравнений Фоккера—Планка—Колмогорова, Феллера—Колмогорова, Пугачёва и др.) для одномерных распределений. В практических задачах часто используют и комбинированные методы.

При этом будем предполагать, что существуют одномерные плотности процессов в МСтС (1). Достаточные условия их существования можно найти, например, в [13].

Будем полагать, что, во-первых, одномерные плотности распределений существуют и, во-вторых, плотности можно параметризовать с помощью условных параметров $\vartheta^y = \vartheta^y(\Theta, t)$: вероятностных моментов, квазимоментов, семиинвариантов, коэффициентов ортогонального разложения плотности и др. [1, 2].

Следуя [2], представим уравнения МОР в виде следующего отрезка разложения:

$$f = f(y_t, \Theta, \vartheta^y, t) \approx w(y_t, \Theta) \left[1 + \sum_{l=3}^n \sum_{|\nu|=l} c_{\nu t}(\Theta, t) p_{\nu}(y_t) \right]. \quad (2)$$

Здесь $w = w(y_t; \Theta)$ — эталонная одномерная плотность, выбираемая из условия совпадения первых двух вероятностных моментов для w и f ; $\{p_{\nu}(y_t), q_{\nu}(y_t)\}$ — система биортонормальных полиномов с весом w , т. е. удовлетворяющих условию

$$M_{\Delta^y}^w [p_{\nu}(Y_t) q_{\mu}(Y_t)] = \delta_{\nu\mu} = \begin{cases} 0 & \text{при } \mu \neq \nu; \\ 1 & \text{при } \mu = \nu, \end{cases} \quad (3)$$

где $M_{\Delta^y}^w$ — символ математического ожидания в области Δ^y ; $\vartheta^y = \{m_t, K_t, c_{\nu t}\}$ — условные параметры МОР, т. е. вектор условного математического ожидания $m_t = m(\Theta, t)$, условная ковариационная матрица $K_t = K(\Theta, t)$ и условная матрица коэффициентов ортогонального разложения $c_{\nu t} = c_{\nu t}(\Theta, t)$, удовлетворяющая условию

$$c_{\nu t} = M_{\Delta^y}^f [q_{\nu}(Y_t)] = q_{\nu}(\alpha), \quad (4)$$

где $q_{\nu}(\alpha)$ представляет собой комбинацию условных начальных моментов α_t , полученную из $q_{\nu}(x)$ заменой всех одночленов $x_1^{k_1}, \dots, x_r^{k_r}$ соответствующими начальными вероятностными моментами $\alpha_{k_1}, \dots, \alpha_{k_r}$.

Для биортогональной системы полиномов условие (3) выполняется только для $\mu \neq \nu$. Всякая биортогональная система приводится к биортонормальной путем деления полиномов p_{ν} и q_{ν} на множители ξ_{ν} и η_{ν} , произведение которых равно интегралу (3) при соответствующем ν и $\mu = \nu$. Очевидно, что при каждом ν один из множителей ξ_{ν}, η_{ν} может быть выбран произвольно.

Как известно [1, 2], существование всех вероятностных моментов для плотности необходимо и достаточно для существования интегралов (3).

Будем пользоваться векторной нумерацией полиномов $\{p_\nu, q_\nu\}$, так чтобы сумма координат $|\nu| = \nu_1 + \dots + \nu_r$ векторного индекса $\nu = [\nu_1 \dots \nu_r]^T$ была равна степени полиномов. Тогда число линейно независимых полиномов данной степени $\nu_* = |\nu|$ будет равно числу независимых одночленов степени ν_* , т. е. $C_{r+\nu_*-1}^{\nu_*}$.

Между начальными моментами $\alpha_{\nu t}$, семиинвариантами $\kappa_{\nu t}$ и коэффициентами ортогонального разложения $c_{\nu t}$ существуют следующие формулы связи [1, 2]:

$$\begin{aligned} \alpha_{l_1, \dots, l_r, t} &= \alpha_{l_1, \dots, l_r, t}^w + \sum_{l=1}^n \sum_{|\nu|=l} c_{\nu t} p_{\nu, l_1, \dots, l_r}(\alpha^w); \\ p_{\nu, l_1, \dots, l_r}(y_t) &= y_{1t}^{l_1} \dots y_{rt}^{l_r} p_\nu(y_t), \end{aligned} \quad (5)$$

где

$$\frac{c_{\nu t}}{\nu_1! \dots \nu_r!} = \sum_{h=1}^{[\nu|/3]} \frac{1}{h!} \frac{\kappa_{q_{1t}} \dots \kappa_{q_{ht}}}{q_{11}! \dots q_{hr}!} \quad (|\nu| = 6, 7, \dots).$$

Здесь $p_{\nu, l_1, \dots, l_r}(\alpha^w)$ получается из $p_{\nu, l_1, \dots, l_r}(y_t)$ так же, как $q_\nu(\alpha)$ и $q_\nu(y_t)$; $l_1, \dots, l_r = 0, 1, \dots, n$; $l_1 + \dots + l_r = 3, \dots, n$; $q_h = [q_{h1} \dots q_{hr}]^T$; $\nu = [\nu_1 \dots \nu_r]^T$.

Пользуясь известной обобщенной формулой Ито [1, 2, 13] для дифференциала нелинейной функции $\varphi(Y_t, \Theta, t)$:

$$\begin{aligned} d\varphi(Y_t, \Theta, t) &= \\ &= \left\{ \varphi_t(Y_t, \Theta, t) + \varphi_Y(Y_t, \Theta, t)^T a(Y_t, \Theta, t) + \frac{1}{2} \times \right. \\ &\times \text{tr} \left[\varphi_{YY}(Y_t, \Theta, t) b(Y_t, \Theta, t) \nu_0(\Theta, t) b(Y_t, \Theta, t)^T \right] \Big\} dt + \\ &+ \int_{R_0^q} \left[\varphi(Y_t + c(Y_t, \Theta, t, v)) - \varphi(Y_t, \Theta, t) - \right. \\ &- \varphi(Y_t, \Theta, t)^T c(Y_t, \Theta, t, v), t) - \varphi(Y_t, \Theta, t) - \\ &- \varphi_Y(Y_t, \Theta, t)^T c(Y_t, \Theta, t, v) \Big] \nu_P(\Theta, dt, dv) + \\ &+ \varphi_Y(Y_t, \Theta, t)^T b(Y_t, \Theta, t) dW_0(\Theta, t) + \\ &+ \int_{R_0^q} [\varphi(Y_t + c(Y_t, \Theta, t, v), \Theta, t) - \\ &- \varphi(Y_t, \Theta, t)] P_0(\Theta, dt, dv), \end{aligned}$$

получаем, что условные параметры МОР для МСтС (1) удовлетворяют следующим обыкновенным дифференциальным уравнениям:

$$\left. \begin{aligned} \dot{m}_t &= \varphi_{10}(m_t, K_t, \Theta, t) + \\ &+ \sum_{l=3}^n \sum_{|\nu|=l} \varphi_{1\nu}(m_t, K_t, \Theta, t) c_{\nu t}, \\ m(t_0) &= m_0; \\ \dot{K}_t &= \varphi_{20}(m_t, K_t, \Theta, t) + \\ &+ \sum_{l=3}^n \sum_{|\nu|=l} \varphi_{2\nu}(m_t, K_t, \Theta, t) c_{\nu t}, \\ K(t_0) &= K_0; \\ \dot{c}_{\kappa t} &= \varphi_{\kappa 0}(m_t, K_t, \Theta, t) + \\ &+ \psi_{\kappa 0}^{m_t}(m_t, K_t, \Theta, t)^T \dot{m}_t + \\ &+ \frac{1}{2} \text{tr} \left[\psi_{\kappa 0}^{K_t}(m_t, K_t, \Theta, t) \dot{K}_t \right] + \\ &+ \frac{1}{2} \text{tr} \left[\psi_{\kappa \nu}^{K_t}(m_t, K_t, \Theta, t) \dot{K}_t \right] + \\ &+ \sum_{l=3}^n \sum_{|\nu|=l} c_{\nu t} [\varphi_{\kappa \nu}(m_t, K_t, \Theta, t) + \\ &+ \psi_{\kappa \nu}^{m_t}(m_t, K_t, \Theta, t)^T \dot{m}_t], \quad c_{\kappa}(t_0) = c_{\kappa 0}. \end{aligned} \right\} \quad (6)$$

Здесь приняты следующие обозначения:

$$\left. \begin{aligned} \varphi_{10}(m_t, K_t, \Theta, t) &= M_{\Delta y}^w [a(Y_t, \Theta, t)]; \\ \varphi_{1\nu}(m_t, K_t, \Theta, t) &= M_{\Delta y}^{w p_\nu} [a(Y_t, \Theta, t)]; \end{aligned} \right\} \quad (7)$$

$$\left. \begin{aligned} \varphi_{20}(m_t, K_t, \Theta, t) &= \\ &= M_{\Delta y}^w \left[a(Y_t, \Theta, t) (Y_t - m_y)^T + \right. \\ &+ (Y_t - m_t) a(Y_t, \Theta, t)^T + \bar{\sigma}(Y_t, \Theta, t) \Big]; \\ \varphi_{2\nu}(m_t, K_t, \Theta, t) &= \\ &= M_{\Delta y}^{w p_\nu} \left[a(Y_t, \Theta, t) (Y_t - m_t)^T + \right. \\ &+ (Y_t - m_t) a(Y_t, \Theta, t)^T + \bar{\sigma}(Y_t, \Theta, t) \Big]; \end{aligned} \right\} \quad (8)$$

$$\left. \begin{aligned} \bar{\sigma}(Y_t, \Theta, t) &= \sigma(Y_t, \Theta, t) + \\ &+ \int_{R_0^q} c(Y_t, \Theta, t, v) c(Y_t, \Theta, t, v)^T \nu_P(t, \Theta) dv; \\ \sigma(Y_t, \Theta, t) &= b(Y_t, \Theta, t) \nu_0(\Theta, t) b(Y_t, \Theta, t)^T; \end{aligned} \right\} \quad (9)$$

$$\begin{aligned} \varphi_{\kappa \nu}(m_t, K_t, \Theta, t) &= \\ &= M_{\Delta y}^{w p_\nu} \left\{ \frac{\partial^T q_\kappa(Y_t)}{\partial Y_t} a(Y_t, \Theta, t) + \right. \\ &+ \frac{1}{2} \text{tr} \left[\frac{\partial}{\partial Y_t} \frac{\partial^T}{\partial Y_t} q_\kappa(Y_t) \sigma(Y_t, \Theta, t) \right] + \\ &+ \int_{R_0^q} \left[q_\kappa(Y_t + c(Y_t, \Theta, t, v)) - q_\kappa(Y_t) - \right. \\ &- \left. \frac{\partial q_\kappa(Y_t)}{\partial Y_t} c(Y_t, \Theta, t, v) \right] \nu_P(\Theta, t, dv) \Big\}; \end{aligned} \quad (10)$$

$$\left. \begin{aligned} \psi_{\kappa\nu}^{m_t}(m_t, K_t, \Theta, t) &= M_{\Delta^y}^{wp\nu} [q_{\kappa}^{m_t}(Y_t)]; \\ \psi_{\kappa\nu}^{K_t}(m_t, K_t, \Theta, t) &= M_{\Delta^y}^{wp\nu} [q_{\kappa}^{K_t}(Y_t)], \end{aligned} \right\} \quad (11)$$

а интегралы $\psi_{\kappa 0}^{m_t}(m_t, K_t, \Theta, t)$ и $\psi_{\kappa 0}^{K_t}(m_t, K_t, \Theta, t)$ выражаются согласно (11) при $p_{\nu}(y_t) = p_0(y_t) = 1$.

Таким образом, если существует одномерное распределение процесса Y_t в МСтС (1), то при фиксированном векторе параметров Θ и полиномиальных $\{p_{\nu}(y_t), q_{\nu}(y_t)\}$ уравнения (2), (6) при условиях (3), (4) и конечности интегралов (7)–(11) лежат в основе алгоритма МОР (теорема 1).

Уравнения теоремы 1 для \dot{m}_t и \dot{K}_t линейны относительно $c_{\kappa t}$, в то время как уравнения для $\dot{c}_{\kappa t}$ в силу (4) нелинейны.

Отметим, что для рассмотрения стационарных распределений с параметрами $\vartheta^{*y} = \{m^*, K^*, c_{\nu}^*\}$ достаточно правые части уравнений (5) приравнять к нулю.

Основной трудностью практического применения МОР для многомерных дифференциальных МСтС является быстрый рост числа уравнений для $c_{\nu t}$ с увеличением размерности p вектора состояния Y_t . В [1, 2] содержатся таблицы, отражающие эти закономерности. Например, для МСтС размерности $p = 10$, порядка учитываемого момента $N = 4$ и отрезка разложения одномерной плотности $n = 4$ количество уравнений составляет $Q^{\text{МОР}} = 85$.

Описанные в [1, 2] универсальные методы сокращения числа уравнений МОР, основанные на использовании метода С. В. Мальчикова и его обобщений, моментно-семиинвариантных соотношений, соответствующего ряда Эджуорта, а также их комбинации могут быть использованы и для МСтС (1). В тех случаях, когда известна аналитическая природа нелинейной задачи, применяют метод нормальных координат или непосредственно в уравнениях МСтС (1), или в уравнениях (6).

3 Метод квазимоментов

При аппроксимации одномерной плотности f отрезком разложения по многомерным полиномам Эрмита $\{H_{\nu}, G_{\nu}\}$ имеем согласно [1, 2]

$$p_{\nu}(y_t) = \frac{H_{\nu}(y_t - m_t)}{\nu_1! \dots \nu_p!}; \quad (12)$$

$$q_{\nu}(y_t) = G_{\nu}(y_t - m_t); \quad (13)$$

$$c_{\nu t} = q_{\nu}(\alpha) = G_{\nu}(\mu); \quad (14)$$

$$q_{\kappa}^{m_t} = 0; \quad q_{\kappa}^{K_t}(\alpha) = 0 \quad (|\kappa| = 3); \quad (14)$$

$$\left. \begin{aligned} q_{\kappa r}^{m_t} &= -\kappa_r c_{\kappa - e_r, t} \quad (r = \overline{1, p}; |\kappa| = 4, \dots, N); \\ q_{\kappa r s} &= -\kappa_r \kappa_s c_{\kappa - e_r - e_s, t} \\ &(r, s = \overline{1, p}; s > r; |\kappa| = 5, \dots, N). \end{aligned} \right\} \quad (15)$$

Таким образом, в основе МКМ для МСтС (1) лежат уравнения теоремы 1 при условиях (12)–(15) (теорема 2).

4 Оценка точности и чувствительности методов ортогональных разложений и квазимоментов

В задачах надежности и безопасности технических систем [14–16] для оценки точности МОР и МКМ можно, следуя [1, 2], применить метод сравнения вероятностей попадания на множества определенного класса или метод оценки вероятностных моментов четвертого порядка.

Метод теории чувствительности в инженерной практике широко применяется для приближенного анализа точности СтС со случайными параметрами Θ в предположении малых дисперсий этих параметров по сравнению с их математическими ожиданиями [17, 18]. Применяя описанный в разд. 2 МОР при фиксированных значениях Θ , найдем условные параметры $\vartheta^y = \{m_t; K_t; c_{\nu t}\}$ одномерной плотности. Уравнения функций чувствительности условных параметров ϑ^y для МОР получаются путем дифференцирования правых и левых частей уравнений теоремы 1 по Θ . В этом случае получаются следующие уравнения для $\nabla^{\Theta} m_t$ ($\partial/\partial\Theta$) и $\nabla^{\Theta} K_t = (\partial/\partial\Theta) K_t$:

$$\left. \begin{aligned} \nabla^{\Theta} \dot{m}_t &= \nabla^{\Theta} \varphi_{10} + \sum_{l=3}^n \sum_{|\nu|=l} (\nabla^{\Theta} \varphi_{1\nu} m_{\nu}^{c_t} + \\ &+ \varphi_{1\nu} \nabla^{\Theta} c_{\nu t}), \quad \nabla^{\Theta} m(t_0) = 0; \\ \nabla^{\Theta} \dot{K}_t &= \nabla^{\Theta} \varphi_{20} + \sum_{l=3}^n \sum_{|\nu|=l} (\nabla^{\Theta} \varphi_{2\nu} + \\ &+ \varphi_{2\nu} \nabla^{\Theta} c_{\nu t}), \quad \nabla^{\Theta} K(t_0) = 0; \\ \nabla^{\Theta} \dot{c}_{\kappa t} &= \nabla^{\Theta} \varphi_{\kappa 0} + \\ &+ \nabla^{\Theta} \left\{ \psi_{\kappa 0}^{m_t} \dot{m}_t + \frac{1}{2} \text{tr} \left(\psi_{\kappa 0}^{m_t} \dot{K}_t \right) + \right. \\ &+ \left. \frac{1}{2} \text{tr} \left(\psi_{\kappa \nu}^{K_t} \dot{K}_t \right) \right\} + \\ &+ \sum_{l=3}^n \sum_{|\nu|=l} \left\{ \nabla^{\Theta} c_{\nu t} \left[\varphi_{\kappa \nu} + \left(\psi_{\kappa \nu}^{m_t} \right)^T \dot{m}_t \right] + \right. \\ &+ \left. c_{\nu t} \nabla^{\Theta} \left[\left(\varphi_{\kappa \nu}^{m_t} \right)^T \dot{m}_t \right] \right\}, \quad \nabla^{\Theta} c_{\kappa}(t_0) = 0. \end{aligned} \right\} \quad (16)$$

При дифференцировании ϑ^y по Θ порядок уравнений возрастает пропорционально числу производных.

Аналогично выписываются уравнения для $\nabla^\Theta(\nabla^\Theta)^\top m_t$, $\nabla^\Theta(\nabla^\Theta)^\top K_t$ и $\nabla^\Theta(\nabla^\Theta)^\top c_{\nu t}$.

Таким образом, в условиях теоремы 1 уравнения алгоритма чувствительности МОР имеют вид (16) (теорема 3), а в условиях теоремы 2 уравнения алгоритма чувствительности МКМ имеют вид (16) при условиях (12)–(15) (теорема 4).

Наряду с частными функциями чувствительности часто используют обобщенные функции чувствительности [17, 18]. Допустим, что для оценки качества МСтС при гауссовских Θ выбрана условная функция потерь ρ , допускающая квадратичную аппроксимацию

$$\rho = \rho(\Theta) = \rho(m^\Theta) + \sum_{i=1}^{p^\Theta} \rho'_i(m^\Theta) \Theta_i^0 + \sum_{i,j=1}^{p^\Theta} \rho''_{ij}(m^\Theta) \Theta_i^0 \Theta_j^0.$$

Здесь $\rho(m^\Theta)$ — значение функции потерь при $\Theta = m^\Theta$; $\rho'_i = \rho'_i(m^\Theta)$ и $\rho''_{ij} = \rho''_{ij}(m^\Theta)$ — вектор и матрица первых и вторых функций чувствительности; p^Θ — число параметров Θ ; $\Theta_i^0 = \Theta_i - m_i^\Theta$ — случайные гауссовские отклонения параметров Θ с математическим ожиданием $m^\Theta = [m_1^\Theta \dots m_p^\Theta]^\top$ и ковариационной матрицей $K^\Theta = [K_{ij}^\Theta]$. Тогда применяют следующие оценки [17, 18]:

$$\varepsilon = \varepsilon_2^{1/4}, \quad \varepsilon_2 = M^N [\rho(\Theta)]^2 - \rho(m^\Theta)^2, \quad (17)$$

где

$$\begin{aligned} M^N [\rho(\Theta)] &> \rho(m^\Theta); \\ M^N [\rho(\Theta)^2] &= \rho(m^\Theta)^2 + \rho'(m^\Theta)^\top K^\Theta \rho'(m^\Theta) + \\ &+ 2\rho(m^\Theta) \text{tr} [\rho''(m^\Theta) K^\Theta] + \\ &+ \{ \text{tr} [\rho''(m^\Theta) K^\Theta] \}^2 + 2\text{tr} [\rho''(m^\Theta) K^\Theta]^2. \end{aligned} \quad (18)$$

Соотношение (18) показывает, что случайный разброс параметров Θ ухудшает качество МСтС, причем количественная оценка ухудшения качества проводится согласно (17).

При известной функции потерь ρ для вычисления функций чувствительности ρ'_i и ρ''_{ij} применяются известные численные методы аппроксимации в точке, в ряде точек и в области [17].

Для задач оценки надежности и безопасности типовые байесовские критерии ρ в виде условного и среднего риска приведены в [17, 18].

5 Пример

Рассмотрим СтС вида [1, 2]

$$\dot{Y}_t = -Y_t^3 + Y_t V(\Theta), \quad Y(t_0) = Y_0,$$

где $V = V(\Theta, t)$ — гауссовский белый шум, зависящий от случайного параметра Θ интенсивности

$\nu = \nu(\Theta)$. Уравнения МОР с учетом вероятностных моментов четвертого порядка $N = 4$ имеют следующий вид:

$$\left. \begin{aligned} \dot{m}_t &= -m_t (m_t^2 + 3D_t) - c_{3t}, \quad m(t_0) = m_0; \\ \dot{D}_t &= [\nu(\Theta) - 6D_t] (m_t^2 + D_t) - 6m_t c_{3t} - \\ &\quad - 2c_{yt}, \quad D(t_0) = D_0; \\ \dot{c}_{3t} &= 6[\nu(\Theta) - 3D_t] m_t D_t + \\ &\quad + 3[\nu(\Theta) - 3m_t^2 - 9D_t] c_{3t} - 9m_t c_{yt}, \\ &\quad c_3(t_0) = c_{30}; \\ \dot{c}_{4t} &= 6[\nu(\Theta) + 2m_t^2] D_t^2 + 36D_t^3 + 4c_{3t}^2 + \\ &\quad + 12[\nu(\Theta) - 6D_t] m_t c_{3t} + \\ &\quad + 6[\nu(\Theta) - 2m_t^2 - 8D_t] c_{4t}, \quad c_4(t_0) = c_4. \end{aligned} \right\} \quad (19)$$

Дифференцируя уравнения (19) по параметру $\nu = \nu(\Theta)$, получим систему уравнений для производных от m_t , D_t , c_{3t} и c_{4t} по ν при известных m_t^* , D_t^* , c_{3t}^* и c_{4t}^* :

$$\left. \begin{aligned} \nabla \dot{m}_t &= -3(m_t^{*2} + D_t^*) - 3m_t^* \nabla D_t - \nabla c_{3t}, \\ &\quad \nabla c_3(t_0) = 0; \\ \nabla \dot{D}_t &= m_t^{*2} + D_t^* + 2[m_t^* (\nu - 6D_t^*) - \\ &\quad - 3c_{3t}^*] \nabla m_t + [\nu - 6D_t^* - 6(m_t^{*2} + D_t^*)] \times \\ &\quad \times \nabla D_t - 6m_t^* \nabla c_{3t} - 2\nabla c_{4t}, \quad \nabla D(t_0) = 0; \\ \nabla \dot{c}_{3t} &= 3(2m_t^* D_t^* + c_{3t}^*) + \\ &\quad + [6(\nu - 3D_t^*) D_t^* - 18m_t^* c_{3t}^* - 9c_{4t}^*] \nabla m_t + \\ &\quad + 3[-6m_t^* D_t^* + 2(\nu - 3D_t^*) m_t^* - 9c_{3t}^*] \times \\ &\quad \times \nabla D_t + 3(\nu - 3m_t^{*2} - 9D_t^*) \nabla c_{3t} - \\ &\quad - 9m_t^* \nabla c_{4t}, \quad \nabla c_4(t_0) = 0; \\ \nabla \dot{c}_{4t} &= 6(D_t^{*2} + 2m_t^* c_{3t}^* + c_{4t}^*) + \\ &\quad + 12[m_t^* D_t^* + (\nu - 6D_t^*) c_{3t}^*] \nabla m_t + \\ &\quad + 4[3(\nu + 2m_t^{*2}) D_t^* + 27D_t^{*2} - 18m_t^* c_{3t}^* - \\ &\quad - 12c_{4t}^*] \nabla D_t + 4[2c_{3t}^* + 3(\nu - 6D_t^*) m_t^*] \times \\ &\quad \times \nabla c_{3t} + 6(\nu - 2m_t^{*2} - 8D_t^*) \nabla c_{4t}, \\ &\quad \nabla c_4(t_0) = 0. \end{aligned} \right\} \quad (20)$$

Анализ уравнений (20) показывает:

- (1) уравнения для $\nabla \dot{m}_t$ однородны и линейны относительно ∇m_t , ∇D_t и ∇c_{3t} ;
- (2) уравнения для $\nabla \dot{D}_t$ неоднородны и линейны относительно всех переменных;
- (3) уравнения для $\nabla \dot{c}_{3t}$ и $\nabla \dot{c}_{4t}$ неоднородны и линейны относительно всех переменных.

Для стационарного случая надо приравнять правые части (19) к нулю.

6 Заключение

Для нелинейных дифференциальных стохастических систем, в том числе и на многообразиях, понимаемых в смысле Ито, разработаны методы параметрического аналитического моделирования точности и чувствительности. Предполагается, что в качестве параметров одномерных распределений выбраны коэффициенты ортогональных разложений плотности. Особое внимание уделено методу квази моментов. Полученные результаты положены в основу разрабатываемого инструментального символьного программного обеспечения в среде MATLAB-MAPLE для линейных, линейных с мультипликативными шумами и нелинейных МСтС.

Для задач надежности и безопасности технических систем имеют важное значение вопросы сокращения числа уравнений точности и чувствительности МОР и МКМ путем использования соответствующих нормальных координат.

В качестве обобщений можно рассмотреть задачи оценки точности и чувствительности МОР и МКМ на основе многомерных распределений для различных моментов времени и критериев качества технических систем.

Литература

1. Пугачев В. С., Сеницын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — М.: Наука, 1990. 632 с. [Pugachev V. S., Sinitsyn I. N. Stochastic differential systems. Analysis and filtering. — Chichester, New York, NY, USA: John Wiley, 1987. 549 p.]
2. Pugachev, V. S., and I. N. Sinitsyn. 2001. *Stochastic systems. Theory and applications*. — Singapore: World Scientific. 908 p.
3. Сеницын И. Н. Стохастические информационные технологии для исследования нелинейных круговых стохастических систем // Информатика и её применения, 2011. Т. 5. Вып. 4. С. 78–89.
4. Sinitsyn I. N., Belousov V. V., Konashenkova T. D. Software tools for circular stochastic systems analysis // 29th Seminar (International) on Stability Problems for Stochastic Models: Abstracts. — Svetlogorsk, Russia, 2011. P. 86–87.
5. Сеницын И. Н. Математическое обеспечение для анализа нелинейных многоканальных круговых стохастических систем, основанное на параметризации распределений // Информатика и её применения, 2012. Т. 6. Вып. 1. С. 12–18.
6. Сеницын И. Н., Корепанов Э. Р., Белоусов В. В., Конашенкова Т. Д. Развитие математического обеспечения для анализа нелинейных многоканальных круговых стохастических систем // Системы и средства информатики, 2012. Вып. 22. № 1. С. 29–40.
7. Sinitsyn I. N., Belousov V. V., Konashenkova T. D. Software tools for spherical stochastic systems analysis and filtering // 30th Seminar (International) on Stability Problems for Stochastic Models and 6th Workshop (International) “Applied Problems in Theory of Probabilities and Mathematical Statistics Related to Modeling of Information Systems: Book of Abstracts / Eds. V. Yu. Korolev, S. Ya. Shorgin. — Moscow: IPI RAN, 2012. P. 91–93.
8. Сеницын И. Н. Параметрическое статистическое и аналитическое моделирование распределений в нелинейных стохастических системах на многообразиях // Информатика и её применения, 2013. Т. 7. Вып. 2. С. 4–16.
9. Сеницын И. Н., Сеницын В. И. Аналитическое моделирование нормальных процессов в стохастических системах со сложными нелинейностями // Информатика и её применения, 2014. Т. 8. Вып. 3. С. 2–4.
10. Сеницын И. Н., Сеницын В. И., Сергеев И. В., Белоусов В. В., Шоргин В. С. Математическое обеспечение аналитического моделирования стохастических систем со сложными нелинейностями // Системы и средства информатики, 2014. Т. 24. № 3. С. 4–29.
11. Сеницын И. Н., Сеницын В. И., Корепанов Э. Р. Моделирование нормальных процессов в стохастических системах со сложными иррациональными нелинейностями // Информатика и её применения, 2015. Т. 9. Вып. 1. С. 2–8.
12. Сеницын И. Н., Сеницын В. И., Сергеев И. В., Корепанов Э. Р., Белоусов В. В., Шоргин В. С. Математическое обеспечение моделирования нормальных процессов в стохастических системах со сложными иррациональными нелинейностями // Системы и средства информатики, 2015. Т. 25. № 2. С. 45–61.
13. Ватанабэ С., Икэда Н. Стохастические дифференциальные уравнения и диффузионные процессы. — М.: Наука, 1986. 474 с.
14. ГОСТ 23743-88. Изделия авиационной техники. Номенклатура показателей безопасности полета, надежности, контролепригодности, эксплуатационной и ремонтной технологичности.
15. Болотин В. В. Теория надежности машин // Машиностроение: Энциклопедия. Т. IV-3. Надежность машин. — М.: Машиностроение, 1998. 38 с.
16. Александровская Л. Н., Аронов И. З., Круглов В. И. и др. Безопасность и надежность технических систем. — М.: Университетская книга, Логос, 2008. 376 с.
17. Евланов А. Г., Константинов В. М. Системы со случайными параметрами. — М.: Наука, 1976. 568 с.
18. Справочник по теории автоматического управления / Под ред. А. А. Красовского. — М.: Наука, 1987. 712 с.

Поступила в редакцию 21.05.15

ANALYTICAL MODELING IN STOCHASTIC SYSTEMS ON MANIFOLDS BASED ON ORTHOGONAL EXPANSIONS

I. N. Sinitsyn

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: Problems of accuracy and sensitivity of one-dimensional distributions by parametrical analytical modeling algorithms on the basis of the orthogonal expansion method (OEM) and the quasi-moment method (QMM) in stochastic systems on manifolds (MStS) are considered. Stochastic system on manifolds is described by Ito linear, linear with multiplicative noises and nonlinear equations with Wiener and Poisson noises. The OEM and QMM equations are derived by generalized Ito formula. Methodological results are the basis of the original symbolic software tools for MATLAB-MAPLE. The problems of reduction of number of OEM and QMM equations are discussed, reliability and security algorithms are presented. Scalar nonlinear MStS with multiplicative white noise is investigated. Some possible generalizations are formulated.

Keywords: analytical modeling method (AMM); generalized Ito formula; Hermite polynomials; OEM and QMM accuracy equations; OEM and QMM sensitivity equations; orthogonal expansion method (OEM); quasi-moment method (QMM); stochastic system on manifold (MStS)

DOI: 10.14357/19922264150302

Acknowledgments

The work was financially supported by the Russian Foundation for Basic Research (project 15–07–02244).

References

1. Pugachev, V. S., and I. N. Sinitsyn. 1987. *Stochastic differential systems. Analysis and filtering*. Chichester, New York, NY: John Wiley. 549 p.
2. Pugachev, V. S., and I. N. Sinitsyn. 2001. *Stochastic systems. Theory and applications*. Singapore: World Scientific. 908 p.
3. Sinitsyn, I. N. 2011. Stokhasticheskie informatsionnye tekhnologii dlya issledovaniya nelineynykh krugovykh stokhasticheskikh sistem [Stochastic informational technologies for circular stochastic systems investigation]. *Informatika i ee Primeneniya — Inform Appl.* 5(4):78–89.
4. Sinitsyn, I. N., V. V. Belousov, and T. D. Konashenkova. 2011. Software tools for circular stochastic systems analysis. *29th Seminar (International) on Stability Problems for Stochastic Models: Abstracts*. Svetlogorsk, Russia. 86–87.
5. Sinitsyn, I. N. 2012. Matematicheskoe obespechenie dlya analiza nelineynykh mnogokanal'nykh krugovykh stokhasticheskikh sistem, osnovannoe na parametrizatsii raspredeleniy [Mathematical software for analysis of nonlinear multichannel circular stochastic systems based on distributions parametrization]. *Informatika i ee Primeneniya — Inform Appl.* 6(1):12–18.
6. Sinitsyn, I. N., E. R. Korepanov, V. V. Belousov, and T. D. Konashenkova. 2012. Razvitiye matematicheskogo obespecheniya dlya analiza nelineynykh mnogokanal'nykh krugovykh stokhasticheskikh sistem [Development of mathematical software for analysis of nonlinear multichannel circular stochastic systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 22(1):29–40.
7. Sinitsyn, I. N., V. V. Belousov, and T. D. Konashenkova. 2012. Software tools for spherical stochastic systems analysis and filtering. *30th Seminar (International) on Stability Problems for Stochastic Models and 6th Workshop (International) “Applied Problems in Theory of Probabilities and Mathematical Statistics Related to Modeling of Information Systems: Book of Abstracts*. Eds. V. Yu. Korolev and S. Ya. Shorgin. Moscow: IPI RAN. 91–93.
8. Sinitsyn, I. N. 2013. Parametricheskoe statisticheskoe i analiticheskoe modelirovanie raspredeleniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraznykh [Parametric statistical and analytical modeling of distributions in nonlinear stochastic systems on manifolds]. *Informatika i ee Primeneniya — Inform. Appl.* 7(2):4–16.
9. Sinitsyn, I. N., and V. I. Sinitsyn. 2014. Analiticheskoe modelirovanie normal'nykh protsessov v stokhasticheskikh sistemakh so slozhnymi nelineynostyami [Analytical modeling of normal processes stochastic systems with complex nonlinearities]. *Informatika i ee Primeneniya — Inform. Appl.* 8(3):2–4.
10. Sinitsyn, I. N., V. I. Sinitsyn, I. V. Sergeev, V. V. Belousov, and V. S. Shorgin. 2014. Matematicheskoe obespechenie analiticheskogo modelirovaniya stokhasticheskikh sistem so slozhnymi nelineynostyami [Mathematical software for analytical modeling of stochastic systems with complex nonlinearities]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 24(3):4–29.

11. Sinitsyn, I. N., V. I. Sinitsyn, and E. R. Korepanov. 2015. Modelirovanie normal'nykh protsessov v stokhasticheskikh sistemakh so slozhnymi irratsional'nymi nelineynostyami [Modeling of normal processes in stochastic systems with complex irrational nonlinearities]. *Informatika i ee Primeneniya — Inform. Appl.* 9(1):2–8.
12. Sinitsyn, I. N., V. I. Sinitsyn, I. V. Sergeev, E. R. Korepanov, V. V. Belousov, and V. S. Shorgin. 2015. Matematicheskoe obespechenie modelirovaniya normal'nykh protsessov v stokhasticheskikh sistemakh so slozhnymi irratsional'nymi nelineynostyami [Mathematical software for modeling of normal processes in stochastic systems with complex irrational nonlinearities]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(2):45–61.
13. Vatanabe, S., and N. Ikeda. 1986. *Stokhasticheskie differentsial'nye uravneniya i diffuzionnye protsessy* [Stochastic differential equations and diffusion processes]. Moscow: Nauka. 474 p.
14. Aleksandrovskaia, L. N., I. Z. Aronov, V. I. Kruglov, et al. 2008. *Bezopasnost' i nadezhnost' tekhnicheskikh sistem* [Security and reliability of technical systems]. Moscow: Universitetskaya kniga, Logos. 376 p.
15. Bolotin, V. V. 1998. Teoriya nadezhnosti mashin [Theory of machine reliability]. *Mashinostroenie. Entsiklopediya* [Machinebuilding. Encyclopedia]. Vol. IV-3. Nadezhnost' mashin [Reliability of machines]. Moscow: Mashinostroenie. 38 p.
16. GOST 23743–88. Izdeliya aviatsionnoy tekhniki. Nomenklatura pokazateley bezopasnosti poleta, nadezhnosti, kontroleprigodnosti, ekspluatatsionnoy i remontnoy tekhnologichnosti [Aircraft products. Nomenclature of flight safety, reliability, testability, operational and repair manufacturability].
17. Evlanov, A. G., and V. M. Konstantinov. 1976. *Sistemy so slozhnymi parametrami* [Systems with random parameters]. Moscow: Nauka. 568 p.
18. Krasovskiy, A. A., ed. 1987. *Spravochnik po teorii avtomaticheskogo upravleniya* [Handbook for automatic control]. Moscow: Nauka. 712 p.

Received May 21, 2015

Contributor

Sinitsyn Igor N. (b. 1940) — Doctor of Science in technology, professor, Honored scientist of RF, Head of Department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sinitsin@dol.ru

СОВМЕСТНОЕ СТАЦИОНАРНОЕ РАСПРЕДЕЛЕНИЕ ЧИСЛА ЗАЯВОК В m ОЧЕРЕДЯХ В N -КАНАЛЬНОЙ СИСТЕМЕ ОБСЛУЖИВАНИЯ С ПЕРЕУПОРЯДОЧЕНИЕМ ЗАЯВОК*

А. В. Печинкин¹, Р. В. Разумчик¹

Аннотация: Рассматривается функционирующая в непрерывном времени N -канальная система обслуживания с накопителем бесконечной емкости и переупорядочением заявок. В систему поступает пуассоновский поток заявок, время обслуживания каждым прибором распределено по экспоненциальному закону с одним и тем же параметром. При поступлении в систему всем заявкам присваивается порядковый номер. На выходе из системы сохраняется порядок между заявками, установленный при входе в нее. Заявки, завершившие обслуживание и нарушившие установленный порядок, формируют в бункере переупорядочения (БП) (неограниченной емкости) разные очереди. Если на приборах находится n ($n = \overline{1, N}$) заявок, то заявкой 1-го уровня будем называть ту из них, которая в систему поступила последней, 2-го уровня — предпоследней, ..., n -го уровня — первой. Находящиеся в БП заявки, поступившие между заявками 2-го и 1-го уровней, будем называть заявками 1-й очереди, заявки, поступившие между заявками 3-го и 2-го уровней, — заявками 2-й очереди, ..., заявки, поступившие между заявками N -го и $(N - 1)$ -го уровней, — заявками $(N - 1)$ -й очереди. Найдены математические соотношения в терминах производящих функций (ПФ), позволяющие алгоритмически вычислять совместное стационарное распределение числа заявок в накопителе и на приборах, а также в 1-й, 2-й, ..., m -й очередях в БП ($m = \overline{1, N - 1}$).

Ключевые слова: многолинейная система массового обслуживания; переупорядочение; отдельные очереди; совместное стационарное распределение

DOI: 10.14357/19922264150303

1 Введение

Производительность многосерверных систем с возможностью параллельной обработки информации существенным образом может зависеть от необходимости сохранения порядка в потоке передаваемых сообщений. Нарушение порядка может быть вызвано как внешними, так и внутренними причинами, и в результате складывается ситуация, когда задержка сообщений, передаваемых через систему, увеличивается, поскольку прежде, чем покинуть систему, они должны ожидать друг друга для восстановления порядка следования (т. е. для переупорядочения).

К настоящему времени для изучения влияния переупорядочения на характеристики производительности систем уже предложено много аналитических моделей. Некоторый обзор моделей и теоретических результатов, полученных до 1997 г., можно найти в [1, 2]. Один из подходов к изучению систем с переупорядочением, опирающийся на методы теории массового обслуживания, предполагает представление изучаемой системы в виде систе-

мы/сети массового обслуживания. На выходе из системы/сети имеется очередь (БП), в которой хранятся заявки, нарушившие порядок следования, установленный при входе. С использованием данного подхода было получено много результатов, касающихся распределения числа заявок в системе и БП при различных предположениях о входящем потоке и процессе обслуживания, распределения времен пребывания заявок в системе и БП и др. Среди наиболее интересных работ можно отметить [3–8].

Настоящее исследование является продолжением работ [9, 10], в которых рассматривалась система массового обслуживания (СМО) с переупорядочением в виде марковской многоканальной системы обслуживания неограниченной емкости и БП, также имеющим неограниченную емкость. В этих работах была получена система уравнений равновесия для совместного стационарного распределения числа заявок в системе и (суммарного числа заявок) в БП, допускающая рекуррентное решение, и приведены некоторые результаты численных расчетов. В недавней работе [11] было замечено,

* Работа выполнена при поддержке РФФИ (проект 13-07-00223).

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Российский университет дружбы народов, rrazumchik@gmail.com

что заявки, ожидающие в БП в рассматриваемой СМО, можно разделить на несколько групп в зависимости от того, сколько ⟨обслуживаний⟩ осталось прождать заявке, прежде чем она сможет покинуть БП (т.е. систему). Поскольку максимальное число ⟨обслуживаний⟩, которое может прождать заявка в БП, равно $(N - 1)$, а минимальное равно 1, то очередь в БП можно представить состоящей из нескольких отдельных очередей, общее число которых зависит от числа занятых приборов. Если занят только один прибор, то в БП не может ожидать ни одна заявка; при двух занятых приборах в БП может быть только одна очередь (из заявок, ожидающих окончания обслуживания на приборе, который был занят самым первым). Легко видеть, что при N занятых приборах в БП имеется $(N - 1)$ очередь. Общее число заявок в БП есть суммарное число заявок в отдельных очередях. Несомненный интерес представляет нахождение совместного стационарного распределения числа заявок в накопителе и на приборах, а также числа заявок в отдельных очередях в БП. В [11] приводится система уравнений равновесия для соответствующего совместного стационарного распределения и упоминается, что она допускает рекуррентное решение. В данной работе покажем, что традиционный подход, предполагающий нахождение стационарных вероятностей состояний в терминах ПФ, применим также и в данном случае.

Статья организована следующим образом. В разд. 2 дается описание системы. В разд. 3 приводится система уравнений равновесия для совместного стационарного распределения. Раздел 4 посвящен анализу системы равновесия в терминах ПФ. Здесь показано, что ПФ можно также находить рекуррентным образом. В заключении сформулированы основные результаты работы.

2 Описание системы

Рассмотрим функционирующую в непрерывном времени N -линейную ($N \geq 2$) СМО с накопителем неограниченной емкости, входящим пуассоновским потоком заявок интенсивности λ и экспоненциальным распределением времени обслуживания заявки каждым прибором с параметром μ .

При поступлении в систему всем заявкам присваивается порядковый номер. На выходе из СМО сохраняется порядок между заявками, установленный при входе в нее. Заявки, завершившие обслуживание и нарушившие установленный порядок, накапливаются на выходе системы в БП и покидают СМО только после того, как закончится обслу-

живание всех заявок с меньшими номерами. Такая СМО носит название системы с переупорядочением заявок.

Предполагается также выполнение необходимого и достаточного условия $\tilde{\rho} = \rho/N < 1$ существования стационарного режима функционирования СМО, где $\rho = \lambda/\mu$.

3 Система уравнений равновесия для стационарных вероятностей состояний

Пусть на приборах находится n ($n = \overline{1, N}$) заявок. Тогда заявкой 1-го уровня будем называть ту из них, которая в систему поступила последней, второго уровня — предпоследней, ..., n -го уровня — первой. При этом если $n = N$ (все приборы заняты), то находящиеся в БП заявки, поступившие между заявками 2-го и 1-го уровней, будем называть заявками 1-й очереди, заявки, поступившие между заявками 3-го и 2-го уровней, — заявками 2-й очереди, ..., заявки, поступившие между заявками N -го и $(N - 1)$ -го уровней, — заявками $(N - 1)$ -й очереди. Если же $n < N$, то заявками 1-й очереди будем называть заявки из БП, поступившие после заявки 1-го уровня, заявками 2-й очереди — заявки, поступившие между заявками 2-го и 3-го уровней, и т.д.

При $n \geq N$ обозначим через $p_{n; i_1, \dots, i_m}$, $m = \overline{1, N - 1}$, $i_1, \dots, i_m \geq 0$, стационарную вероятность того, что в системе на приборах и в накопителе находится n заявок, а в БП имеется i_1 заявок 1-й очереди, i_2 заявок 2-й очереди, ..., i_m заявок m -й очереди. Через $p_{n; i_1, \dots, i_m}$, $m = \overline{1, n}$, $i_1, \dots, i_m \geq 0$, обозначим аналогичную стационарную вероятность при $n = \overline{1, N - 1}$. Наконец, через p_0 обозначим стационарную вероятность отсутствия заявок в системе.

Через p_n , $n \geq 0$, обозначим стационарную вероятность того, что в системе на приборах и в накопителе (без учета числа заявок в БП) находится n заявок. Очевидно, что стационарные вероятности p_n определяются теми же самыми формулами, что и в обычной марковской СМО $M/M/N/\infty$ (см., например, [12]):

$$p_0 = \left(\sum_{i=0}^{N-1} \frac{\rho^i}{i!} + \frac{\rho^N}{(N-1)!(N-\rho)} \right)^{-1};$$

$$p_i = \frac{\rho^i}{i!} p_0, \quad i = \overline{1, N};$$

$$p_i = \frac{\rho^i}{N! N^{i-N}} p_0 = \tilde{\rho}^{i-N} p_N, \quad i \geq N + 1. \quad (1)$$

Поскольку целью настоящей работы является нахождение вероятностей в терминах ПФ, то не будем подробно останавливаться на системе уравнений равновесия и лишь заметим следующее. Для того чтобы ее выписать, сначала рассматриваются уравнения глобального баланса для вероятностей $p_{n;i_1}$, $n \geq 1$, $i_1 \geq 0$, затем для $p_{n;i_1,i_2}$, $n \geq 2$, $i_1, i_2 \geq 0$, и т.д. В последнюю очередь составляются уравнения для вероятностей $p_{n;i_1,\dots,i_{N-1}}$, $n \geq N-1$, $i_1, \dots, i_{N-1} \geq 0$. Произведя данные действия, можно получить следующий набор уравнений, которые приводим без пояснений.

Для вероятностей $p_{n;i}$, $n \geq N$, $i \geq 0$, справедливы уравнения:

$$p_{n;0}(\lambda + N\mu) = p_{n-1;0}\lambda + p_{n+1}(N-1)\mu, \quad n \geq N; \quad (2)$$

$$p_{n;i}(\lambda + N\mu) = p_{n-1;i}\lambda + p_{n+1;i-1}\mu, \quad n \geq N, \quad i \geq 1. \quad (3)$$

Соотношения для вероятностей $p_{N-1;i}$, $i \geq 0$, имеют вид:

$$p_{N-1;0}[\lambda + (N-1)\mu] = p_{N-2}\lambda + p_N(N-1)\mu; \quad (4)$$

$$p_{N-1;i}[\lambda + (N-1)\mu] = p_{N;i-1}\mu, \quad i \geq 1. \quad (5)$$

Наконец, для вероятностей $p_{n;i}$, $n = \overline{1, N-2}$, $i \geq 0$, имеют место следующие уравнения:

$$p_{n;0}(\lambda + n\mu) = p_{n-1}\lambda + p_{n+1;0}n\mu, \quad n = \overline{1, N-2}; \quad (6)$$

$$p_{n;i}(\lambda + n\mu) = p_{n+1;i}n\mu + \sum_{j=0}^{i-1} p_{n+1;i-j-1,j}\mu, \quad n = \overline{1, N-2}, \quad i \geq 1. \quad (7)$$

Начиная с вероятностей $p_{n;i_1,i_2}$, $n \geq 2$, $i_1, i_2 \geq 0$, и до вероятностей $p_{n;i_1,\dots,i_{N-1}}$, $n \geq N-1$, $i_1, \dots, i_{N-1} \geq 0$, уравнения получаются практически одинаковыми. Поэтому приведем их сразу в общем виде. Далее для сокращения записи примем соглашение, что $\sum_{i=0}^{-1} a_i = 0$. Тогда уравнения для вероятностей $p_{n;i_1,\dots,i_m}$, $m = \overline{2, N-1}$, $n \geq m$, $i_1, \dots, i_{N-1} \geq 0$, можно записать в следующем виде:

$$p_{n;0,i_2,\dots,i_m}(\lambda + N\mu) = p_{n-1;0,i_2,\dots,i_m}\lambda + p_{n+1;i_2,\dots,i_m}(N-m)\mu + \sum_{j=0}^{i_2-1} p_{n+1;j,i_2-j-1,i_3,\dots,i_m}\mu + \dots \dots + \sum_{j=0}^{i_m-1} p_{n+1;i_2,\dots,i_{m-1},j,i_m-j-1}\mu, \quad n \geq N, \quad i_2, \dots, i_m \geq 0; \quad (8)$$

$$p_{n;i_1,\dots,i_m}(\lambda + N\mu) = p_{n-1;i_1,\dots,i_m}\lambda + p_{n+1;i_1-1,i_2,\dots,i_m}\mu, \quad n \geq N, \quad i_1 \geq 1, \quad i_2, \dots, i_m \geq 0; \quad (9)$$

$$p_{N-1;0,i_2,\dots,i_m}[\lambda + (N-1)\mu] = p_{N-2;i_2,\dots,i_m}\lambda + p_{N;i_2,\dots,i_m}(N-m)\mu + \sum_{j=0}^{i_2-1} p_{N;j,i_2-j-1,i_3,\dots,i_m}\mu + \dots \dots + \sum_{j=0}^{i_m-1} p_{N;i_2,\dots,i_{m-1},j,i_m-j-1}\mu, \quad i_2, \dots, i_m \geq 0; \quad (10)$$

$$p_{N-1;i_1,\dots,i_m}[\lambda + (N-1)\mu] = p_{N;i_1-1,i_2,\dots,i_m}\mu, \quad i_1 \geq 1, \quad i_2, \dots, i_m \geq 0, \quad (11)$$

$$p_{n;0,i_2,\dots,i_m}(\lambda + n\mu) = p_{n-1;i_2,\dots,i_m}\lambda + p_{n+1;0,i_2,\dots,i_m}(n-m+1)\mu + \sum_{j=0}^{i_2-1} p_{n+1;0;j,i_2-j-1,i_3,\dots,i_m}\mu + \dots \dots + \sum_{j=0}^{i_m-1} p_{n+1;0,i_2,\dots,i_{m-1},j,i_m-j-1}\mu, \quad m \neq N-1, \quad n = \overline{m, N-2}, \quad i_2, \dots, i_m \geq 0; \quad (12)$$

$$p_{n;i_1,\dots,i_m}(\lambda + n\mu) = p_{n+1;i_1,\dots,i_m}(n-m+1)\mu + \sum_{j=0}^{i_1-1} p_{n+1;j,i_1-j-1,i_2,\dots,i_m}\mu + \dots \dots + \sum_{j=0}^{i_m-1} p_{n+1;i_1,\dots,i_{m-1},j,i_m-j-1}\mu, \quad m \neq N-1, \quad n = \overline{m, N-2}, \quad i_1 \geq 1, \quad i_2, \dots, i_m \geq 0. \quad (13)$$

Как указано в [11], данную систему можно решить рекуррентным образом. Однако традиционный подход, предполагающий нахождение стационарных вероятностей состояний в терминах ПФ, применим также и для рассматриваемой системы, что и будет показано в следующем разделе. Забегая вперед, отметим, что решение получается алгоритмическим.

4 Уравнения для производящих функций

Введем ПФ

$$p_n(z_1, \dots, z_m) = \sum_{i_1, \dots, i_m=0}^{\infty} z_1^{i_1} \dots z_m^{i_m} p_{n;i_1,\dots,i_m}, \quad n \geq 1, \quad m = \overline{1, \min(n, N-1)}, \quad 0 < z_k \leq 1, \quad k = \overline{1, m};$$

$$P(u, z_1, \dots, z_m) = \sum_{n=N}^{\infty} u^{n-N} p_n(z_1, \dots, z_m),$$

$$m = \overline{1, \min(n, N-1)}, \quad 0 < u \leq 1.$$

Учитывая (1), положим также

$$P(u) = \sum_{n=N}^{\infty} u^{n-N} p_n = \frac{1}{1 - \tilde{\rho}u} p_N, \quad 0 < u \leq 1.$$

Умножая теперь (2) и (3) на z^i и суммируя по всем возможным значениям i , получаем

$$(\lambda + N\mu)p_n(z) = \lambda p_{n-1}(z) + (N-1)\mu p_{n+1} + \mu z p_{n+1}(z), \quad n \geq N.$$

Умножая последнее выражение на u^{N-n} и суммируя по n от N до бесконечности, приходим к уравнению

$$(\lambda + N\mu)P(u, z) = \lambda p_{N-1}(z) + \lambda u P(u, z) + \frac{(N-1)\mu}{u} [P(u) - p_N] + \frac{\mu z}{u} [P(u, z) - p_N(z)]. \quad (14)$$

Обратимся к уравнениям (4) и (5). Умножая их на z^i и производя традиционные преобразования, получаем уравнение

$$[\lambda + (N-1)\mu]p_{N-1}(z) = \lambda p_{N-2} + (N-1)\mu p_N + \mu z p_N(z). \quad (15)$$

Наконец, из уравнений (6) и (7) после домножения на z^i и суммирования получаем соотношение

$$(\lambda + n\mu)p_n(z) = \lambda p_{n-1} + n\mu p_{n+1}(z) + \mu z p_{n+1}(z, z), \quad n = \overline{1, N-2}. \quad (16)$$

Отметим, что нахождение уравнений, которым удовлетворяют ПФ $p_n(z_1, \dots, z_m)$ и $P(u, z_1, \dots, z_m)$ при фиксированном m , осуществляется в 4 шага:

- (1) находится ПФ $p_n(z_1, \dots, z_m)$ при $n \geq N$;
- (2) находится выражение для двойной $P(u, z_1, \dots, z_m)$;
- (3) находится ПФ $p_{N-1}(z_1, \dots, z_m)$;
- (4) находится ПФ $p_n(z_1, \dots, z_m)$, $n = \overline{m, N-2}$.

Действуя подобным образом, уравнения (8)–(13) в терминах введенных ПФ можно привести к виду:

$$(\lambda + N\mu)p_n(z_1, \dots, z_m) = \lambda p_{n-1}(z_1, \dots, z_m) + (N-m)\mu p_{n+1}(z_2, \dots, z_m) + \mu z_1 p_{n+1}(z_1, \dots, z_m) + \mu z_2 p_{n+1}(z_2, z_2, \dots, z_m) + \mu z_3 p_{n+1}(z_2, z_3, z_3, \dots, z_m) + \dots + \mu z_m p_{n+1}(z_2, \dots, z_m, z_m), \quad n \geq N;$$

$$(\lambda + N\mu)P(u, z_1, \dots, z_m) = \lambda p_{N-1}(z_1, \dots, z_m) + \lambda u P(u, z_1, \dots, z_m) + \frac{(N-m)\mu}{u} [P(u, z_2, \dots, z_m) - p_N(z_2, \dots, z_m)] + \frac{\mu z_1}{u} [P(u, z_1, \dots, z_m) - p_N(z_1, \dots, z_m)] + \frac{\mu z_2}{u} [P(u, z_2, z_2, \dots, z_m) - p_N(z_2, z_2, \dots, z_m)] + \frac{\mu z_3}{u} [P(u, z_2, z_3, z_3, \dots, z_m) - p_N(z_2, z_3, z_3, \dots, z_m)] + \dots - p_N(z_2, z_3, z_3, \dots, z_m)] + \dots + \frac{\mu z_m}{u} [P(u, z_2, \dots, z_m, z_m) - p_N(z_2, \dots, z_m, z_m)], \quad (17)$$

$$[\lambda + (N-1)\mu]p_{N-1}(z_1, \dots, z_m) = \lambda p_{N-2}(z_2, \dots, z_m) + (N-m)\mu p_N(z_2, \dots, z_m) + \mu z_1 p_N(z_1, \dots, z_m) + \mu z_2 p_N(z_2, z_2, \dots, z_m) + \mu z_3 p_N(z_2, z_3, z_3, \dots, z_m) + \dots + \mu z_m p_N(z_2, \dots, z_m, z_m); \quad (18)$$

$$(\lambda + n\mu)p_n(z_1, \dots, z_m) = \lambda p_{n-1}(z_2, \dots, z_m) + (n-m+1)\mu p_{n+1}(z_1, \dots, z_m) + \mu z_1 p_{n+1}(z_1, z_1, \dots, z_m) + \mu z_2 p_{n+1}(z_1, z_2, z_2, \dots, z_m) + \dots + \mu z_m p_{n+1}(z_1, \dots, z_m, z_m), \quad m \neq N-1, \quad n = \overline{m, N-2}. \quad (19)$$

Нахождение выражений для ПФ $P(u, z)$, $p_{N-1}(z)$, $p_n(z)$, $P(u, z_1, \dots, z_m)$, $p_{N-1}(z_1, \dots, z_m)$, $m = \overline{2, N-1}$, $p_n(z_1, \dots, z_m)$, $m = \overline{2, N-2}$, $n = \overline{m, N-2}$, можно свести к последовательному решению систем линейных уравнений.

Сначала находятся ПФ числа заявок в накопителе и суммарного числа заявок в бункере. Если положить $z_1 = \dots = z_m = z$ в ПФ $P(u, z_1, \dots, z_m)$, то функция $P(u, z, \dots, z)$ будет представлять собой не что иное, как ПФ суммарного числа заявок в 1-й, 2-й, ..., m -й очередях и вероятность того, что заняты все приборы (в системе на приборах и в очереди не менее N заявок). Аналогично $p_n(u, z, \dots, z)$ — ПФ суммарного числа заявок в 1-й, 2-й, ..., m -й очередях и вероятность того, что занято ровно n приборов. Как показано в [9], при $z_1 = \dots = z_m = z$ система уравнений (14)–(16) и (17)–(19) решается рекуррентным образом.

Алгоритм нахождения $P(u, z_1, \dots, z_m)$, $m = \overline{m, N-1}$, $p_{N-1}(z_1, \dots, z_m)$, $n = \overline{m, N-1}$, $p_n(z_1, \dots, z_m)$, $m = \overline{2, N-2}$, $n = \overline{m, N-2}$, при фиксированном N состоит из $(N-1)$ шага. На

2. Dimitrov B. D., Green Jr. D., Rykov V. V., Stanchev P. L. On performance evaluation and optimization problems in queues with resequencing // *Advances in stochastic modelling* / Eds. J. R. Artalejo, A. Krishnamoorthy. — NJ, USA: Notable Publications Inc., 2002. P. 55–72.
3. Gogate N. R., Panwar S. S. Assigning customers to two parallel servers with resequencing // *IEEE Commun. Lett.*, 1999. Vol. 3. No. 4. P. 119–121.
4. Huisman T., Boucherie R. J. The sojourn time distribution in an infinite server resequencing queue with dependent interarrival and service times // *J. Appl. Probab.*, 2002. Vol. 39. No. 3. P. 590–603.
5. Lelarge M. Packet reordering in networks with heavy-tailed delays // *Math. Method. Oper. Res.*, 2008. Vol. 67. P. 341–371.
6. Xia Y., Tse D. N. C. On the large deviations of resequencing queue size: $2-M/M/1$ case // *IEEE Trans. Inform. Theory*, 2008. Vol. 54. No. 9. P. 4107–4118.
7. Leung K., Li V. O. K. A resequencing model for high-speed packet-switching networks // *Comput. Commun.*, 2010. Vol. 33. Iss. 4. P. 443–453.
8. Матюшенко С. И. Стационарные характеристики двухканальной системы обслуживания с переупорядочиванием заявок и распределениями фазового типа // *Информатика и её применения*, 2010. Т. 4. Вып. 4. С. 67–71.
9. Печинкин А. В., Разумчик Р. В. Совместное стационарное распределение числа заявок в накопителе и в очереди переупорядочения в многоканальной системе обслуживания с переупорядочением заявок // *Информатика и её применения*, 2014. Т. 8. Вып. 4. С. 3–10.
10. Pechinkin A. V., Caraccio I., Razumchik R. V. On joint stationary distribution in exponential multiserver reordering queue // *12th Conference (International) on Numerical Analysis and Applied Mathematics Proceedings*, 2015. Vol. 1648. 250003. 4 p.
11. Pechinkin A. V., Razumchik R. V. Joint stationary distribution of queues in multiserver resequencing queue // *18th Conference (International) on Distributed Computer and Communication Networks: Control, Computation, Communications Proceedings*, 2015 (in print).
12. Riordan J. *Stochastic service systems*. — New York, NY, USA: Wiley, 1962. 139 p.

Поступила в редакцию 1.07.15

JOINT STATIONARY DISTRIBUTION OF m QUEUES IN THE N -SERVER QUEUEING SYSTEM WITH REORDERING

A. V. Pechinkin and R. V. Razumchik^{1,2}

¹Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Peoples’ Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

Abstract: The paper considers a continuous-time N -server queueing system with a buffer of infinite capacity and customer reordering. The Poisson flow of customers arrives at the system. Service times of customers at each server are exponentially distributed with the same parameter. Each customer obtains a sequential number upon arrival. The order of customers upon arrival should be preserved upon departure from the system. Customers which violated the order form different queues in the reordering buffer which has infinite capacity. If there are n , $n = \overline{1, N}$, customers in servers, then the latest customer to occupy a server is called the 1st level customer, the last but one — the 2nd level customer, . . . , the first — the n th level customer. Customers in the reordering buffer that arrived between the 1st level and the 2nd level customers, form the queue number one. Customers, which arrived between the 2nd level and the 3rd level customers, form the queue number two, etc. Customers, which arrived between the N th level and the $(N - 1)$ th level customers, form the queue number $(N - 1)$ in the reordering buffer. Mathematical relations in terms of Z-transform, which make it possible to calculate the joint stationary distribution of the number of customers in the buffer, servers, and in the 1st, 2nd, . . . , m th queues ($m = \overline{1, N - 1}$) in the reordering buffer, are obtained.

Keywords: multiserver queueing system; reordering; separate queues; joint stationary distribution

DOI: 10.14357/19922264150303

Acknowledgments

The work was supported by the Russian Foundation for Basic Research (project 13-07-00223).

References

1. Boxma, O., G. Koole, and Z. Liu. 1994. Queueing-theoretic solution methods for models of parallel and distributed systems. *Performance evaluation of parallel and distributed systems: Solution methods*. CWI tract ser. Torino, Italy: CWI. 105-106:1–24. Available at: <http://oai.cwi.nl/oai/asset/1479/1479a.pdf> (accessed August 7, 2015).
2. Dimitrov, B. D., D. Green, Jr., V. V. Rykov, and P. L. Stanchev. 2002. On performance evaluation and optimization problems in queues with resequencing. *Advances in stochastic modelling*. Eds. J. R. Artalejo and A. Krishnamoorthy. NJ: Notable Publications Inc. 55–72.
3. Gogate, N. R., and S. S. Panwar. 1999. Assigning customers to two parallel servers with resequencing. *IEEE Commun. Lett.* 3(4):119–121.
4. Huisman, T., and R. J. Boucherie. 2002. The sojourn time distribution in an infinite server resequencing queue with dependent interarrival and service times. *J. Appl. Probab.* 39(3):590–603.
5. Lelarge, M. 2008. Packet reordering in networks with heavy-tailed delays. *Math. Method. Oper. Res.* (67):341–371.
6. Xia, Y., and D. N. C. Tse. 2008. On the large deviations of resequencing queue size: $2-M/M/1$ case. *IEEE Trans. Information Theory* 54(9):4107–4118.
7. Leung, K., and V. O. K. Li. 2010. A resequencing model for high-speed packet-switching networks. *Comput. Commun.* 33(4):443–453.
8. Matyushenko, S. I. 2010. Stacionarnye kharakteristiki dvukhkanal'noy sistemy obsluzhivaniya s pereporyadochivaniem zayavok i raspredeleniyami fazovogo tipa [Stationary characteristics of the two-channel queueing system with reordering customers and distributions of phase type]. *Informatika i ee Primeneniya — Inform. Appl.* 4(4):67–71.
9. Pechinkin, A. V., and R. V. Razumchik. 2014. Sovmestnoe stacionarnoe raspredelenie chisla zayavok v nakopitele i v bunkere pereporyadocheniya v mnogokanal'noy sisteme obsluzhivaniya s pereporyadocheniem zayavok [Joint stationary distribution of the number of customers in the system and reordering buffer in the multiserver reordering queue]. *Informatika i ee Primeneniya — Inform. Appl.* 8(4):3–10.
10. Pechinkin, A. V., I. Caraccio, and R. V. Razumchik. 2015. On joint stationary distribution in exponential multiserver reordering queue. *12th Conference (International) on Numerical Analysis and Applied Mathematics Proceedings*. 1648:250003. 4 p.
11. Pechinkin, A. V., and R. V. Razumchik. 2015 (in print). Joint stationary distribution of queues in multiserver resequencing queue. *18th Conference (International) on Distributed Computer and Communication Networks: Control, Computation, Communications Proceedings*. Moscow.
12. Riordan, J. 1962. *Stochastic service systems*. New York, NY: Wiley. 139 p.

Received July 1, 2015

Contributors

Pechinkin Alexander V. (1946–2014) — professor, Doctor of Science in physics and mathematics

Razumchik Rostislav V. (b. 1984) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; associate professor, Peoples’ Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; rrazumchik@iee.org

АППРОКСИМАЦИЯ ВРЕМЕНИ ОТКЛИКА СИСТЕМЫ ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ*

А. В. Горбунова¹, И. С. Зарядов², С. И. Матюшенко³, К. Е. Самуйлов⁴, С. Я. Шоргин⁵

Аннотация: Облачные вычисления представляют собой эволюционную технологию, позволяющую удаленному пользователю получать доступ к различным ресурсам посредством интернет-сервисов. В статье рассматривается система облачных вычислений, в которую поступают сложные пользовательские запросы, состоящие из нескольких подзапросов. Для обработки каждого подзапроса выделяется одна единица ресурса. Точная оценка эффективности облачных вычислительных ресурсов является необходимым условием обеспечения требуемого качества обслуживания. Время отклика системы, служащее одним из важнейших показателей этого качества, требует адекватного моделирования. Для его анализа построена математическая модель функционирования облачной системы в виде системы массового обслуживания (СМО), состоящей из нескольких подсистем, на которую поступает единый пуассоновский поток запросов. Каждый из поступивших запросов в момент поступления делится на несколько частей, обслуживаемых своей подсистемой массового обслуживания с накопителем неограниченной емкости и экспоненциальным распределением времени обслуживания на приборе. С помощью данной математической модели получена аналитическая формула для нахождения приближения среднего значения и дисперсии времени отклика системы — максимального времени обработки составляющих запрос подзапросов. Для оценки точности приближения выполнено имитационное моделирование средствами GPSS (General Purpose Simulation System).

Ключевые слова: система облачных вычислений; время отклика; обработка сложных запросов; система массового обслуживания; аппроксимация; имитационное моделирование

DOI: 10.14357/19922264150304

1 Введение

Облачные вычисления (cloud computing) представляют собой технологию, позволяющую удаленному пользователю в режиме реального времени по требованию получать доступ к вычислительным ресурсам (программным приложениям, серверам, устройствам хранения данных, сервисам и др.) через Интернет в рамках согласованного качества обслуживания для выбранной ценовой категории. В задачи поставщика таких услуг входит не только обеспечение требуемого уровня оказываемых услуг, но и, как следствие, избежание перегрузки ресурсов при обработке пользовательских требований [1–3].

Рассмотрим облачный центр, состоящий из нескольких физических машин, которые резервируются пользователями в порядке поступления запросов и могут использоваться совместно для их обработки посредством метода виртуализации. Разделение доступа к данным, виртуальным машинам и контроль их использования представляют собой

основные задачи поставщика облачных услуг. Если же говорить о качестве обслуживания, то его показателем в большинстве случаев считается время отклика системы, т. е. время обработки запроса пользователя [4].

В статье рассматриваются сложные запросы пользователей, содержащие несколько задач, для обработки каждой из которых требуется одна виртуальная машина: посылаемый запрос разбивается на K подзапросов и для его обработки выделяется K виртуальных машин. Запрос считается выполненным, т. е. пользователь получает отклик системы, после обработки всех его подзапросов.

Таким образом, время отклика системы облачных вычислений определяется как максимальное время обработки составляющих его заданий или подзапросов [5, 6].

Следует отметить, что основой для построения описываемой аналитической модели послужила так называемая fork-join система (система с расщеплением), которая, в свою очередь, является есте-

* Работа выполнена при частичной финансовой поддержке РФФИ (проект 15-07-03051).

¹ Российский университет дружбы народов, avgorbunova@rambler.ru

² Российский университет дружбы народов, izaryadov@sci.pfu.edu.ru

³ Российский университет дружбы народов, matushenko@list.ru

⁴ Российский университет дружбы народов, ksam@sci.pfu.edu.ru

⁵ Федеральный исследовательский центр «Информатика и управление» Российской академии наук, sshoragin@ipiran.ru

ственной моделью для многих реальных систем. В качестве примеров можно привести сферу обслуживания, медицинские приложения, производственные системы и др. [7]. Если же говорить об информационных технологиях, то прежде всего стоит упомянуть распределенные вычисления, технология которых оказала значительное влияние на концепцию облачных вычислений, а также параллельные вычисления, например обработку пакетов данных, телефонных вызовов и др. [8].

Но, несмотря на довольно большой перечень прикладных задач, которые могут быть решены с помощью такого рода модели, на сегодняшний день точные аналитические результаты для среднего времени отклика существуют только для случая расщепления на два подзапроса, причем с учетом некоторых дополнительных ограничений, а именно: условия равенства интенсивностей обслуживания [9]. Для случая же $K > 2$ большинство авторов в своих работах концентрируются непосредственно на различных методах аппроксимации искомой величины [7, 10, 11]. Это связано со сложностями, возникающими при поиске точного аналитического решения в рамках данной модели из-за существующей зависимости между очередями подзапросов вследствие общих моментов поступления. Отметим, что таких сложностей не возникает для случая, когда каждая подсистема состоит из бесконечного числа приборов [12, 13].

В большинстве статей из интуитивных соображений аппроксимация основывается на допущении о K параллельно функционирующих и независимых системах массового обслуживания, на вход которых поступают независимые потоки подзапросов. Подробный обзор существующих публикаций по данной тематике представлен в [14]. В данной работе с помощью анализа системы уравнений равновесия (СУР) получено логическое обоснование причин такого перехода и приведен достаточно подробный сравнительный анализ полученных результатов применительно к системе облачных вычислений.

Статья организована следующим образом: в разд. 2 описывается общая математическая модель функционирования рассматриваемой облачной системы, в разд. 3 проводится анализ ее времени отклика, в разд. 4 представлены численные результаты и, наконец, в разд. 5 кратко подведены итоги и сформулированы дальнейшие планы исследований.

2 Математическая модель

Итак, рассмотрим СМО с K независимыми параллельно работающими приборами, каждый из

которых имеет накопитель неограниченной емкости. В эту систему поступает пуассоновский поток заявок (запросов) с интенсивностью λ . При этом в момент поступления заявка делится на K подзапросов от 1-го до K -го типа и посылает их в каждую из K очередей. Заявка считается обслуженной в момент окончания обслуживания последнего из составляющих ее подзапросов. Времена обслуживания заявок-подзапросов являются независимыми случайными величинами, распределенными по экспоненциальному закону с параметром μ_k для подзапроса k -го типа, $k = \overline{1, K}$.

Пусть $x_k(t)$ — число подзапросов k -го типа, находящихся в системе в момент времени t . Обозначим $X(t) = (x_1(t), \dots, x_K(t))$. Случайный процесс $\{X(t), t \geq 0\}$, описывающий поведение системы во времени, является марковским (МП) с множеством состояний $X = \{\vec{n} := (n_1, \dots, n_k, \dots, n_K), n_1 \geq 0, \dots, n_k \geq 0, \dots, n_K \geq 0\}$.

В предположении о его существовании стационарное распределение должно удовлетворять следующей СУР:

$$\left(\lambda + \sum_{k=1}^K \mu_k u(n_k) \right) p(\vec{n}) = \lambda p(\vec{n} - \vec{1}) \prod_{k=1}^K u(n_k) + \sum_{k=1}^K \mu_k p(\vec{n} + \vec{e}_k), \quad \vec{n} \in X,$$

где \vec{e}_k — вектор, все элементы которого равны нулю, кроме k -го, который равен 1, $\vec{1}$ — единичный вектор и

$$u(n_k) = \begin{cases} 0, & \text{если } n_k \leq 0; \\ 1, & \text{если } n_k > 0. \end{cases}$$

Здесь не исследуется очевидное условие эргодичности рассматриваемой системы и считается далее, что $\rho = \max_{1 \leq k \leq K} \rho_k < 1$, где $\rho_k = \lambda / \mu_k$, $k = \overline{1, K}$.

3 Анализ времени отклика системы

Исследуя СУР, замечаем, что суммированием ее уравнений по всем индексам, кроме n_k -го, приходим к СУР, соответствующей СМО $M_\lambda / M_{\mu_k} / 1$. Следовательно, можем получить выражения для маргинальных вероятностей макросостояний МП $X(t)$:

$$p_{n_k} = (1 - \rho_k) \rho_k^{n_k}, \quad n_k \geq 0.$$

Далее выпишем функцию распределения времени отклика с учетом допущения о независимости времен пребывания подзапросов в подсистемах:

$$\begin{aligned} P(\eta < y) &= P(\max(\xi_1, \dots, \xi_K) < y) = \\ &= P(\xi_1 < y, \dots, \xi_K < y) = \\ &= P(\xi_1 < y) \cdots P(\xi_K < y) = \\ &= (1 - e^{-(\mu_1 - \lambda)y}) \cdots (1 - e^{-(\mu_K - \lambda)y}). \end{aligned}$$

Тогда среднее время отклика и его дисперсия будут иметь вид:

$$\begin{aligned} E\eta &= \sum_{l=1}^K \frac{1}{\mu_l - \lambda} - \sum_{l \neq m} \frac{1}{\mu_l + \mu_m - 2\lambda} + \\ &+ \sum_{l \neq m \neq n} \frac{1}{\mu_l + \mu_m + \mu_n - 3\lambda} + \dots \\ &\dots + (-1)^{2K-1} \frac{1}{\mu_1 + \mu_2 + \dots + \mu_K - K\lambda}; \end{aligned}$$

$$\begin{aligned} D\eta &= \sum_{l=1}^K \frac{2}{(\mu_l - \lambda)^2} - \sum_{l \neq m} \frac{2}{(\mu_l + \mu_m - 2\lambda)^2} + \\ &+ \sum_{l \neq m \neq n} \frac{2}{(\mu_l + \mu_m + \mu_n - 3\lambda)^2} + \dots \\ &\dots + (-1)^{2K-1} \frac{2}{(\mu_1 + \mu_2 + \dots + \mu_K - K\lambda)^2} - \\ &- \left(\sum_{l=1}^K \frac{1}{\mu_l - \lambda} - \sum_{l \neq m} \frac{1}{\mu_l + \mu_m - 2\lambda} + \dots \right) \end{aligned}$$

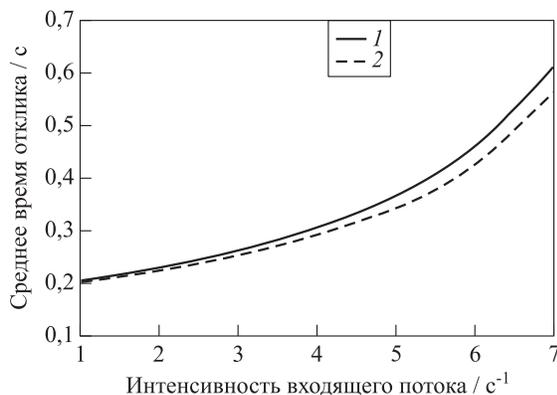


Рис. 1 Среднее время отклика: 1 — аналитическая модель; 2 — имитационная модель

$$\begin{aligned} &+ \sum_{l \neq m \neq n} \frac{1}{\mu_l + \mu_m + \mu_n - 3\lambda} + \dots \\ &\dots + (-1)^{2K-1} \frac{1}{\mu_1 + \mu_2 + \dots + \mu_K - K\lambda} \Big)^2. \end{aligned}$$

Кроме того, можно записать выражения и для других характеристик СМО, используя аналогичные результаты для системы типа $M/M/1$.

4 Численный пример

Для оценки полученной аппроксимации с помощью программных средств GPSS была построена имитационная модель. Поскольку интерес представляют характеристики системы в стационарном режиме ее работы, имитационное моделирование проводилось до тех пор, пока не стабилизировались параметры модели. Стабилизация наступила после прохождения через систему около 500 000 запросов.

На рис. 1 и 2 представлены графики зависимости среднего и среднеквадратического отклонения времени отклика системы облачных вычислений от интенсивности поступления запросов λ . Расчеты были выполнены для случая расщепления пользовательского запроса на 3 подзапроса ($K = 3$) для симметричного случая $\mu_1 = \mu_2 = 10 \text{ с}^{-1}$ при меняющейся интенсивности входящего потока λ . Видно, что полученная с помощью аналитической модели оценка среднего времени отклика является оценкой «сверху», а среднеквадратического отклонения — оценкой «снизу». Заметим, что для несимметричного случая качественных изменений численных результатов не наблюдается.

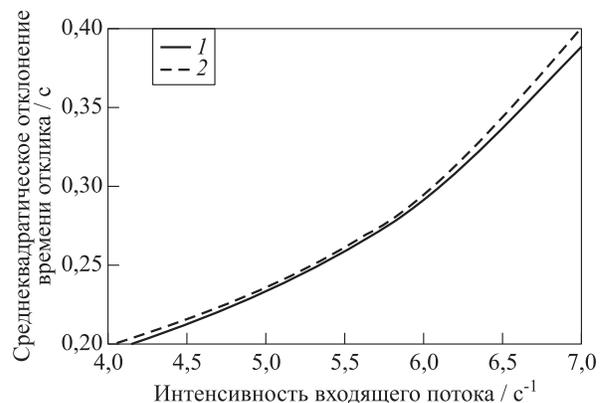


Рис. 2 Среднеквадратическое отклонение времени отклика: 1 — аналитическая модель; 2 — имитационная модель

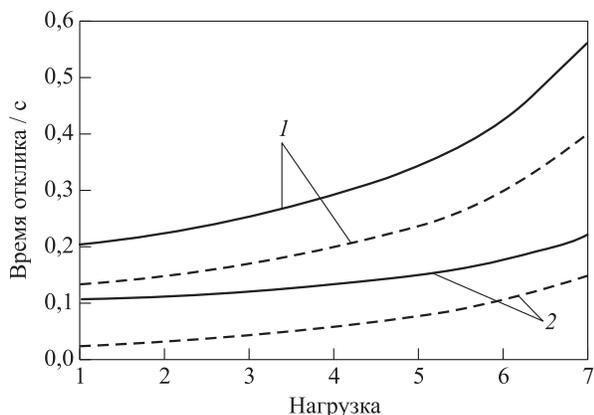


Рис. 3 Математическое ожидание (сплошные кривые) и среднее квадратическое отклонение (штриховые кривые) времени отклика: 1 — экспоненциальное обслуживание; 2 — детерминированное обслуживание

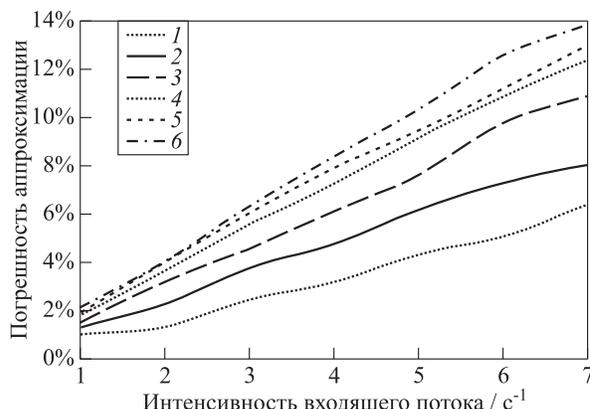


Рис. 4 Погрешность аппроксимации для среднего времени отклика: 1 — $K = 2$; 2 — 3; 3 — 4; 4 — 5; 5 — 6; 6 — $K = 7$

Квантили случайной величины времени отклика системы

Квантиль случайной величины времени отклика	Значения квантилей в зависимости от нагрузки ρ					
	0,10	0,30	0,50	0,70	0,80	0,90
Квантиль уровня 0,97	0,51	0,66	0,92	1,53	2,30	4,60
Квантиль уровня 0,98	0,56	0,71	1,00	1,67	2,50	5,00
Квантиль уровня 0,99	0,63	0,81	1,14	1,90	2,85	5,70
Квантиль уровня 0,999	0,89	1,14	1,60	2,67	4,00	8,01
Квантиль уровня 0,9999	1,15	1,47	2,06	3,44	5,15	10,31

Рисунок 3 иллюстрирует разницу в значениях времени отклика, полученных с помощью имитационного моделирования, для случаев экспоненциального и детерминированного обслуживания. С увеличением нагрузки математическое ожидание времени отклика при экспоненциальном обслуживании возрастает в среднем в два раза по сравнению с детерминированным обслуживанием. Заметим, что аналогичный результат аналитически следует из формулы Поллачека–Хинчина для СМО $M/M/1$ и $M/D/1$.

На рис. 4 изображены графики для погрешности аппроксимации математического ожидания времени отклика в зависимости от интенсивности входящего потока λ для различных значений $K = 2, 7$. Видно, что с увеличением числа подзапросов и интенсивности входящего потока увеличивается ошибка приближения.

В таблице представлены значения верхней границы для времени отклика с заданными вероятностями (квантили функции распределения времени отклика) для различных значений интенсивности нагрузки ρ в случае $K = 3$. Результаты, отраженные в таблице, подтверждают сделанное выше замечание об увеличении значений времени отклика с ростом нагрузки.

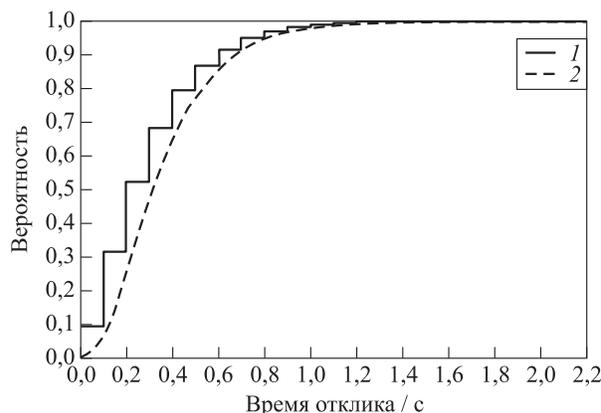


Рис. 5 Графики эмпирической (1) и аналитической (2) функций распределения

На рис. 5 приведены графики эмпирической функции распределения, построенной по данным имитационного моделирования, и аналитической функции распределения для времени отклика.

Из рисунка следует, что эмпирическая функция распределения аппроксимирует аналитическую функцию распределения сверху, что сочетается с результатами, иллюстрируемыми рис. 1.

5 Заключение

В статье рассмотрена и проанализирована модель системы облачных вычислений, в которую поступают сложные запросы. Получена формула для расчета приближения времени отклика системы, реализована имитационная модель и проведен сравнительный анализ.

Помимо получения изложенных в предыдущем разделе численных результатов с помощью имитационного моделирования исследовалась цепь Маркова, вложенная по моментам поступления запросов. Полученные значения стационарных вероятностей вложенной цепи Маркова оказались близкими к значениям стационарных вероятностей по МП, поэтому не следует ожидать, что полученная аппроксимация не точна из-за того, что вывод формул для характеристик времени отклика основывался именно на вероятностях по процессу. Также были вычислены попарные корреляции между очередями. Эксперимент проводился на ограниченном объеме данных, но можно увидеть, что корреляция положительна и составляет около 50%, т. е. очереди коррелированы и поэтому расчет времени отклика по маргинальным распределениям фактически означает расчет в предположении об их независимости. Но из полученных значений для корреляции следует, что очереди зависимы, и поэтому аналитически получается аппроксимация, которая дает положительный с точки зрения инженерных расчетов результат. Погрешность в параметрах, близких к реальным, составляет не более 14%.

В дальнейшем планируется улучшить аналитические и имитационные результаты, а также исследовать модели с другим типом входящего потока и иными законами распределения времени обслуживания, провести расчет их основных вероятностно-временных характеристик и показателей качества обслуживания.

Литература

1. Xiong K., Perros H. Service performance and analysis in cloud computing // 5th IEEE World Congress on Services (Services-I '09) Proceedings. — Los Angeles: IEEE, 2009. P. 693–700.
2. Buaya R., Broberg J., Goscinski A. M. Introduction to cloud computing // Cloud computing: Principles and paradigms. — John Wiley & Sons, 2011. P. 3–42.
3. Satyanarayana A., Suresh Varma P., Rama Sundari M. V., Sarada Varma P. Performance analysis of cloud computing under non homogeneous conditions // Int. J. Adv. Res. Comput. Sci. Softw. Eng., 2013. Vol. 3. P. 969–974.
4. Khazae H., Mistic J., Mistic V. B. A fine-grained performance model of cloud computing centers // IEEE Trans. Parallel. Distrib. Syst., 2012. Vol. 24. P. 2138–2147.
5. Мокров Е. В., Самуйлов К. Е. Модель системы облачных вычислений в виде системы массового обслуживания с несколькими очередями и с групповым поступлением заявок // Т-Comm — Телекоммуникации и транспорт, 2013. Т. 7. № 11. С. 139–141.
6. Basharin G. P., Gaidamaka Yu. V., Samouylov K. E. Mathematical theory of teletraffic and its application to the analysis of multiservice communication of next generation networks // Autom. Control Comp. Sci., 2013. Vol. 47. No. 2. P. 62–69.
7. Kemper B., Mandjes M. Mean sojourn time in two-queue fork-join systems: Bounds and approximations // OR Spectrum, 2012. Vol. 34. P. 723–742.
8. Ko S. S., Serfozo R. F. Response times in $M/M/s$ Fork-join networks // Adv. Appl. Probab., 2004. Vol. 36. No. 3. P. 854–871.
9. Nelson R., Tantawi A. N. Approximate analysis of fork/join synchronization in parallel queues // IEEE Trans. Comput., 1988. Vol. 37. P. 739–743.
10. Baccelli F., Makowski A. M., Shwartz A. The fork-join queue and related systems with synchronization constraints: Stochastic ordering and computable bounds // Adv. Appl. Probab., 1989. Vol. 21. No. 3. P. 629–660.
11. Kness C. On the diffusion approximation to a fork and join queueing model // SIAM J. Appl. Math., 1991. Vol. 51. P. 160–171.
12. Ивановская И. А., Моисеева С. П. Исследование математической модели параллельного обслуживания заявок смешанного типа // Известия Томского политехнического ун-та. Управление, вычислительная техника и информатика, 2010. Т. 317. № 5. С. 32–34.
13. Ивановская И. А., Моисеева С. П. Исследование модели параллельного обслуживания кратных заявок в нестационарном режиме // Вестник Томского гос. ун-та. Управление, вычислительная техника и информатика, 2010. № 3(12). С. 21–28.
14. Thomasian A. Analysis of fork-join and related queueing systems // ACM Comput. Surv. (CSUR), 2014. Vol. 47. No. 2. P. 1–71.

Поступила в редакцию 21.07.15

THE APPROXIMATION OF RESPONSE TIME OF A CLOUD COMPUTING SYSTEM

A. V. Gorbunova¹, I. S. Zaryadov¹, S. I. Matyushenko¹, K. E. Samouylov¹, and S. Ya. Shorgin²

¹Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

²Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: Cloud computing is an evolutionary technology that allows a remote user to gain access to resources through Internet services. The article discusses the cloud computing system, which receives complex user queries consisting of several subqueries. The allocation of one unit of the resource is required for processing each subquery. Accurate assessment of effectiveness of cloud computing resources is a prerequisite to ensure the required quality of service. Response time, i. e., the maximum of subqueries service times, is chosen as an indication of effectiveness of a cloud computing system. To analyze the characteristics of response time, a simplified mathematical model of a cloud system was constructed as a queuing system with single Poisson input flow of requests and several subsystems (a buffer and a server). Each request at the instant of arrival is divided into several parts, each of which is served by its queuing subsystem with unlimited storage capacity with exponentially distributed service time. The analytical formulas for approximation of mean response time and its variance are presented. To assess the accuracy of the approximation, a GPSS (General Purpose Simulation System) model was constructed.

Keywords: cloud computing system; response time; complex queries processing; queuing system; approximation; simulation

DOI: 10.14357/19922264150304

Acknowledgments

This work is partially supported by the Russian Foundation for Basic Research (grant No. 15-07-03051).

References

- Xiong, K., and H. Perros. 2009. Service performance and analysis in cloud computing. *5th IEEE World Congress on Services (Services-I'09) Proceedings*. Los Angeles. 693–700.
- Buyya, R., J. Broberg, and A. M. Goscinski. 2011. *Cloud computing: Principles and paradigms*. 1st ed. John Wiley & Sons. 664 p.
- Satyanarayana, A., P. Suresh Varma, M. V. Rama Sundari, and P. Sarada Varma. 2013. Performance analysis of cloud computing under non homogeneous conditions. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* 3:969–974.
- Khazae, H., J. Mistic, and V. B. Mistic. 2012. A fine-grained performance model of cloud computing centers. *IEEE Trans. Parall. Distr. Syst.* 24:2138–2147.
- Mokrov, E. V., and K. E. Samouylov. 2013. Model' sistemy oblachnykh vychisleniy v vide sistemy massovogo obsluzhivaniya s neskol'kimi ocheredyami i s gruppovym postupleniem zayavok [Modeling of cloud computing as a queuing system with batch arrival]. *T-Comm — Telecommunications and Transport* 7(11):139–141.
- Basharin, G. P., Yu. V. Gaidamaka, and K. E. Samouylov. 2013. Mathematical theory of teletraffic and its application to the analysis of multiservice communication of next generation networks. *Autom. Control Comp. Sci.* 47(2):62–69.
- Kemper, B., and M. Mandjes. 2012. Mean sojourn time in two-queue fork-join systems: Bounds and approximations. *OR Spectrum* 34:723–742.
- Ko, S. S., and R. F. Serfozo. 2004. Response times in $M/M/s$ fork-join networks. *Adv. Appl. Probab.* 36(3):854–871.
- Nelson, R., and A. N. Tantawi. 1988. Approximate analysis of fork/join synchronization in parallel queues. *IEEE Trans. Comput.* 37:739–743.
- Bacelli, F., A. M. Makowski, and A. Shwartz. 1989. The fork-join queue and related systems with synchronization constraints: Stochastic ordering and computable bounds. *Adv. Appl. Probab.* 21(3):629–660.
- Kness, C. 1991. On the diffusion approximation to a fork and join queueing model. *SIAM J. Appl. Math.* 51:160–171.
- Ivanovskaya, I. A., and S. P. Moiseeva. 2010. Issledovanie matematicheskoy modeli parallel'nogo obsluzhivaniya zayavok smeshannogo tipa [Studying of mathematical model of mixed type claim parallel service]. *Bulletin of the Tomsk Polytechnic University* 317(5):32–34.
- Ivanovskaya, I. A., and S. P. Moiseeva. 2010. Issledovanie modeli parallel'nogo obsluzhivaniya kratnykh zayavok v nestatsionarnom rezhime [Analysis of the parallel service model of multiple service requests operating in unsteady

- mode]. *Vestnik Tomskogo gos. un-ta. Upravlenie, vychislitel'naya tekhnika i informatika* [Tomsk State University J. Control and Computer Science] 3(12):21–28.
14. Thomasian, A. 2014. Analysis of fork-join and related queueing systems. *ACM Comput. Surv. (CSUR)* 47(2):1–71.

Received July 21, 2015

Contributors

Gorbunova Anastasiya V. (b. 1986) — PhD student, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; avgorbunova@rambler.ru

Zaryadov Ivan S. (b. 1981) — Candidate of Science (PhD) in physics and mathematics, associate professor, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; izaryadov@sci.pfu.edu.ru

Matyushenko Sergey I. (b. 1963) — Candidate of Science (PhD) in physics and mathematics, associate professor, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; matushenko@list.ru

Samuylov Konstantin E. (b. 1955) — Doctor of Science in technology, professor, Head of Department, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; ksam@sci.pfu.edu.ru

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics, professor, Deputy Director, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sshorgin@ipiran.ru

ЭКСТРЕМАЛЬНЫЕ ИНДЕКСЫ В СХЕМЕ СЕРИЙ И ИХ ПРИЛОЖЕНИЯ*

А. В. Лебедев¹

Аннотация: Проводится обобщение понятия экстремального индекса стационарной случайной последовательности на схему серий одинаково распределенных случайных величин со случайными длинами серий, стремящимися к бесконечности по вероятности. Введены новые экстремальные индексы с помощью двух определений, обобщающих основные свойства классического экстремального индекса. Доказан ряд полезных свойств новых экстремальных индексов. Показано, как поведение максимумов суммарных активностей на случайных графах (в моделях информационных сетей) и поведение максимумов признаков частиц в ветвящихся процессах (в моделях биологических популяций) могут быть описаны в терминах новых экстремальных индексов. Получены также новые результаты для моделей с копулами и пороговых моделей. Показано, что новые индексы могут принимать разные значения для одной системы, а также значения, большие единицы.

Ключевые слова: экстремальный индекс; схема серий; случайный граф; информационная сеть; ветвящийся процесс; копула

DOI: 10.14357/19922264150305

1 Введение

Экстремальный индекс стационарной (в узком смысле) случайной последовательности $\{\xi_n\}$ определяется следующим образом [1, §3.7].

Определение А.² Пусть $\xi_n, n \geq 1$, имеют распределение F и $M_n = \bigvee_{k=1}^n \xi_k$. Если для каждого $\tau > 0$ существует такая числовая последовательность $u_n(\tau)$, что $n\bar{F}(u_n(\tau)) \rightarrow \tau$ и $P(M_n \leq u_n(\tau)) \rightarrow e^{-\theta\tau}$, то θ называется экстремальным индексом.

При этом возможно любое значение $\theta \in [0, 1]$.

Заметим, что если взять максимумы \hat{M}_n последовательности независимых случайных величин с тем же распределением F , то

$$\lim_{n \rightarrow \infty} P(\hat{M}_n \leq u_n(\tau)) = e^{-\tau},$$

откуда следует

$$\lim_{n \rightarrow \infty} P(M_n \leq u_n(\tau)) = \left(\lim_{n \rightarrow \infty} P(\hat{M}_n \leq u_n(\tau)) \right)^\theta, \quad (1)$$

т. е. предельные функции распределения M_n и \hat{M}_n связаны степенной зависимостью,

$$\lim_{n \rightarrow \infty} P(M_n \leq u_n(\tau)) = \lim_{n \rightarrow \infty} P(\hat{M}_{[\theta n]} \leq u_n(\tau)), \quad \theta > 0, \quad (2)$$

т. е. M_n растет асимптотически как максимум $[\theta n]$ независимых случайных величин при $n \rightarrow \infty$ и

$$\lim_{n \rightarrow \infty} P(M_n \leq u_n(\tau)) \geq \lim_{n \rightarrow \infty} P(\hat{M}_n \leq u_n(\tau)), \quad (3)$$

т. е. M_n стохастически не превосходит максимума независимых случайных величин (в пределе).

Интерес к экстремальному индексу связан отчасти с тем, что его наличие сохраняет экстремальный тип предельного распределения максимумов. Напомним, что если для некоторых числовых последовательностей $a_n > 0, b_n, n \geq 1$, и невырожденного распределения G имеет место предел

$$\lim_{n \rightarrow \infty} P(\hat{M}_n \leq a_n x + b_n) = G(x), \quad \forall x \in \mathbb{R},$$

то G относится к одному из трех экстремальных типов, а именно: $G(x) = G_i(ax + b)$ для некоторых $a > 0$ и b , где $G_1(x) = \exp\{-e^{-x}\}$ (тип Гумбеля); $G_2(x) = \exp\{-x^{-\alpha}\}, x > 0, \alpha > 0$ (тип Фреше); $G_3(x) = \exp\{-(-x)^\alpha\}, x \leq 0, \alpha > 0$ (тип Вейбулла). Такие распределения G называют максимум-устойчивыми (max-stable) или распределениями экстремальных значений (extreme value distributions). Для любого $s > 0$ существуют такие $a(s) > 0, b(s)$, что $G^s(x) = G(a(s)x + b(s))$. Таким образом, возведение в степень $\theta > 0$ предельной функции распределения, возникающее в силу свойства (1), сохраняет экстремальный тип.

* Работа поддержана РФФИ (проект 14-01-00075).

¹ Московский государственный университет имени М. В. Ломоносова, механико-математический факультет, avlebed@yandex.ru

² Далее через \vee обозначается максимум, через \wedge — минимум, чертой сверху над функцией распределения — хвост: $\bar{F}(x) = 1 - F(x)$, через f^{-1} — обратная функция к f , а для функций распределения — обобщенная обратная: $F^{-1}(y) = \inf\{x : F(x) \geq y\}$, через $f(x)^n$ — n -я степень от $f(x)$.

Одна из интерпретаций экстремального индекса заключается в том, что превышения высокого уровня в последовательности происходят не по одиночке, а группами (кластерами) средней величины $1/\theta$. В приложениях это может означать природные катастрофы, отказы технических систем, потерю данных при передаче информации, финансовые потери и др. Понятно, что, когда такие события происходят несколько раз подряд, это гораздо опаснее, чем единичные случаи, и должно учитываться в управлении рисками. Более подробно об этом можно прочитать в [1–4].

С 1980-х гг. активные исследования в данной области ведутся в двух главных направлениях: вычисление экстремального индекса для различных случайных последовательностей и разработка статистических оценок экстремального индекса по наблюдениям.

С обзором результатов и библиографией можно ознакомиться, например, в [3, §8.1] и [4, §5.5]. Некоторым обобщениям классического понятия экстремального индекса и его статистическим оценкам посвящен раздел диссертации [5, §1.2]. В частности, можно дать следующее определение.

Определение Б. Пусть $\xi_n, n \geq 1$, имеют распределение F и $M_n = \bigvee_{k=1}^n \xi_k$. Если для каждой числовой последовательности $u_n, n \geq 1$, такой что

$$0 < \liminf_{n \rightarrow \infty} n\bar{F}(u_n) \leq \limsup_{n \rightarrow \infty} n\bar{F}(u_n) < \infty,$$

верно $P(M_n \leq u_n) - F(u_n)^{\theta n} \rightarrow 0, n \rightarrow \infty$, то θ называется экстремальным индексом.

Такое определение позволяет расширить понятие экстремального индекса на некоторые стационарные последовательности случайных величин с дискретным распределением (например, геометрическим), а для непрерывных эквивалентно определению А.

Исследованиями экстремумов и превышений высокого уровня в связи с моделями телекоммуникаций посвящены работы [6, 7], а в [8] изучались распределения и зависимость экстремумов в процессах выборочного исследования информационных сетей (network sampling processes), в том числе экстремальные индексы.

В диссертации [9] получены новые интересные результаты по экстремальным индексам последовательностей вида

$$X_n = A_n X_{n-1} + B_n,$$

где $(A_n, B_n), n \geq 1$, — независимые случайные пары со значениями в \mathbb{R}_+^2 . В некоторых случаях найдены в явном виде экстремальные индексы

и распределения размеров кластеров превышений высокого уровня, в более общем случае получены верхние и нижние границы для экстремального индекса. Доказана непрерывность экстремального индекса относительно некоторой сходимости распределений коэффициентов. Введены и изучены индексы многомерных последовательностей с тяжелыми хвостами. Часть полученных результатов представлена в [10, 11].

Однако на практике существует необходимость в изучении максимумов на более сложных структурах, чем множество натуральных чисел. Связанные с этим трудности обсуждались еще в [2, §3.9, §3.12]. Например, если речь идет о продолжительностях жизни компонент сложной системы (надежностной схемы), то непонятно, как пронумеровать их так, чтобы использовать модель стационарной последовательности, да и возможно ли это вообще. Несколько проще обстоит дело со случайными полями.

Экстремальный индекс естественным образом обобщается со случайных последовательностей на случайные поля на решетках \mathbb{N}^d [12]. Пусть, например, задано такое поле $\{\xi_{n_1, n_2}\}$ в \mathbb{N}^2 и $M_{n_1, n_2} = \bigvee_{k_1=1}^{n_1} \bigvee_{k_2=1}^{n_2} \xi_{k_1, k_2}$. Тогда если для каждого $\tau > 0$ существует такое $u_{n_1, n_2}(\tau)$, что $n_1 n_2 \bar{F}(u_{n_1, n_2}(\tau)) \rightarrow \tau$ и $P(M_{n_1, n_2} \leq u_{n_1, n_2}(\tau)) \rightarrow e^{-\theta\tau}$, то θ называется экстремальным индексом. Вычислению экстремального индекса случайного поля в \mathbb{N}^2 посвящена работа [13], а в [14] изучалось асимптотическое расположение максимума случайного поля с некоторым экстремальным индексом.

Поскольку вышеупомянутые результаты не имеют прямого отношения к теме статьи, не будем останавливаться на них более подробно, а подчеркнем лишь актуальность исследований экстремального индекса в различных моделях и приложениях.

Далее представлено новое обобщение экстремального индекса на схему серий со случайными длинами, которое позволит работать с более широким классом стохастических структур. Более того, дадим два различных определения.

Пусть задан набор случайных величин $\xi_{n, m}, n \geq 1, m \geq 1$, с распределениями F_n (здесь n — номер серии; m — номер случайной величины в серии), а также последовательность целочисленных случайных величин (длин серий) $\nu_n \xrightarrow{P} +\infty, n \rightarrow \infty$, и $M_n = \bigvee_{m=1}^{\nu_n} \xi_{n, m}$.

Определение 1.¹ Пусть для каждого $s \in (0, 1)$ существует такая последовательность $u_n(s)$, что $E F_n(u_n(s))^{\nu_n} \rightarrow s$, и $P(M_n \leq u_n(s)) \rightarrow \psi(s), n \rightarrow \infty$. Тогда ψ назовем экстремальной функцией.

¹По сравнению с определением А сделана замена $s = e^{-\tau}$ и соответственно переопределены функции $u_n, n \geq 1$.

Если $\psi(s) = s^\theta$, то θ назовем экстремальным индексом.

В общем случае можем определить частичные индексы

$$\theta^+ = \sup_{s \in (0,1)} \log_s \psi(s); \quad \theta^- = \inf_{s \in (0,1)} \log_s \psi(s),$$

тогда $\theta^+ \geq \theta^-$ и $s^{\theta^+} \leq \psi(s) \leq s^{\theta^-}$, $s \in (0, 1)$.

Смысл определения 1 заключается в сравнении предельных распределений M_n и максимумов \hat{M}_n независимых случайных величин в количестве ν_n (не зависящем от них) при одинаковой нормировке, определяемой условием $P(\hat{M}_n \leq u_n(s)) \rightarrow s$, $n \rightarrow \infty$. Тем самым обобщается свойство (1).

Понятно, что индексы, как и ранее, принимают неотрицательные значения, однако ограничение сверху единицей снимается, по крайней мере, для θ^+ , как будет показано далее (примеры 5.5, 6.1 и 6.2). Это связано с возможностью нарушения неравенства (3). Максимумы по сериям могут расти асимптотически быстрее, чем максимумы независимых случайных величин, взятых в тех же количествах, что соответствует случаю $\psi(s) < s$, $s \in (0, 1)$.

Определение 2. Пусть для каждого $s \in (0, 1)$ существует такая последовательность $u_n(s)$, что

$$EF_n(u_n(s))^{\nu_n} \rightarrow s; \\ P(M_n \leq u_n(s)) - EF_n(u_n(s))^{\theta \nu_n} \rightarrow 0, \quad n \rightarrow \infty,$$

тогда θ назовем экстремальным индексом.

Смысл определения 2 заключается в подборе такого значения θ , что предельные распределения M_n и максимумов независимых случайных величин в количестве $[\theta \nu_n]$ (не зависящем от них) совпадают при той же нормировке, что и в определении 1 (при $\theta > 0$). Тем самым обобщается свойство (2).

Существование экстремального индекса по определению 2 фактически означает, что экстремальная функция из определения 1 допускает представление

$$\psi(s) = \lim_{n \rightarrow \infty} EF_n(u_n(s))^{\theta \nu_n}.$$

Возникает вопрос, зачем давать два определения, нельзя ли обойтись каким-то одним. Действительно, во многих случаях оба индекса эквивалентны: они существуют и равны между собой (см. разд. 3, пример 5.1). Но бывает также, что обоих индексов не существует, а по определению 1 существует экстремальная функция и частичные индексы (примеры 5.2, 5.3, 6.1 и 6.2); бывает, что существует индекс по определению 2 и не существует индекса по определению 1, а экстремальная

функция и частичные индексы по-прежнему существуют (см. разд. 4); наконец, бывает удивительная ситуация, когда оба индекса существуют, но принимают *разные* значения (см. пример 5.4). Таким образом, это действительно две разные характеристики системы, не сводящиеся к одной.

Заметим, что ранее максимумы в схеме серий рассматривались в [15] для случайных величин, связанных ИТ-копулами (копулами преобразования независимости) и выводились условия, при которых максимумы растут асимптотически, как в случае независимых величин, т. е. в терминах данной статьи $\theta = 1$.

Для определенности терминологии далее будем говорить об экстремальных индексах *системы* (случайных величин), обозначаемой через $\{\xi_{n,m}; \nu_n\}$.

В разд. 2 доказаны основные свойства экстремальных индексов, в разд. 3 представлены их приложения к моделям информационных сетей, в разд. 4 — к моделям биологических популяций, в разд. 5 — к моделям с копулами, в разд. 6 — к пороговым моделям.

2 Основные свойства экстремальных индексов

Экстремальные индексы обладают следующими свойствами.

Свойство 1. Пусть η_n , $n \geq 1$, — стационарная последовательность с экстремальным индексом θ по определению А. Положим $\xi_{n,m} = \eta_m$, $m \geq 1$, и пусть задана целочисленная последовательность $l_n \rightarrow +\infty$, тогда система $\{\xi_{n,m}; l_n\}$ имеет экстремальный индекс θ по определениям 1 и 2.

Доказательство. Обозначим через $u_n^0(\tau)$, $n \geq 1$, последовательность, существующую по определению А, и положим $u_n(s) = u_{l_n}^0(-\ln s)$, тогда $F(u_n(s))^{l_n} \rightarrow s$, $P(M_n \leq u_n(s)) \rightarrow s^\theta$ и $F(u_n(s))^{\theta l_n} \rightarrow s^\theta$, что дает тот же экстремальный индекс по обоим определениям. \square

Свойство 2. Пусть система $\{\xi_{n,m}; \nu_n\}$ имеет экстремальный индекс по одному из определений 1 и 2 (или экстремальную функцию) и задана последовательность функций $g_n(x)$, $n \geq 1$, непрерывных и строго возрастающих на множестве точек роста F_n . Положим $\tilde{\xi}_{n,m} = g_n(\xi_{n,m})$, тогда система $\{\tilde{\xi}_{n,m}; \nu_n\}$ имеет тот же экстремальный индекс (экстремальную функцию).

Доказательство. Для новой системы $\tilde{F}_n(x) = F_n(g_n^{-1}(x))$. Пусть $\tilde{u}_n(s) = g_n(u_n(s))$, тогда $\tilde{F}_n(\tilde{u}_n(s)) = F_n(u_n(s))$ и $P(\tilde{M}_n \leq \tilde{u}_n(s)) = P(M_n \leq u_n(s))$, так что все пределы (в определениях 1 и 2) сохраняются. \square

Свойство 3. Пусть система $\{\xi_{n,m}; \nu_n\}$ имеет экстремальный индекс по одному из определений 1 и 2 и существует такая последовательность $c_n \rightarrow +\infty$, что $\nu_n/c_n \xrightarrow{P} 1$, $n \rightarrow \infty$, тогда система имеет тот же экстремальный индекс по другому определению.

Доказательство. В этом случае из $EF_n(u_n(s))^{\nu_n} = E(F_n(u_n(s))^{c_n})^{\nu_n/c_n} \rightarrow s \in (0, 1)$ следует $F_n(u_n(s))^{c_n} \rightarrow s$ и $EF_n(u_n(s))^{r\nu_n} \rightarrow s^r$, $r \geq 0$, $n \rightarrow \infty$. Таким образом, если θ — экстремальный индекс по определению 1, то из $P(M_n \leq u_n(s)) \rightarrow s^\theta$ следует $EF_n(u_n(s))^{\theta\nu_n} \rightarrow s^\theta$, а значит, θ — экстремальный индекс по определению 2. И наоборот, если θ — экстремальный индекс по определению 2, то из $EF_n(u_n(s))^{\theta\nu_n} \rightarrow s^\theta$ следует $P(M_n \leq u_n(s)) \rightarrow s^\theta$, а значит, θ — экстремальный индекс по определению 1. \square

Свойство 4. Пусть имеется система $\{\xi_{n,m}; \nu_n\}$ с экстремальным индексом $\theta > 0$ по определению 2, у которой:

- (а) $F_n \equiv F$;
- (б) для некоторого максимум-устойчивого закона G и функций $a(r) > 0$, $b(r)$, $r > 0$, верно

$$F^r(a(r)x + b(r)) \rightarrow G(x), \quad r \rightarrow \infty;$$

- (в) существует такая последовательность $c_n \rightarrow +\infty$, что $\nu_n/c_n \xrightarrow{d} \zeta > 0$, $n \rightarrow \infty$;
- (г) в определении 2 можно взять $u_n(s) = A_n H^{-1}(s) + B_n$, где $A_n = a(c_n)$; $B_n = b(c_n)$; $H(x)$ — непрерывная функция распределения.

Тогда¹

$$H(x) = EG(x)^\zeta;$$

$$P(M_n \leq A_n x + B_n) \rightarrow H(ax + b), \quad n \rightarrow \infty,$$

где $a > 0$ и b определяются из тождества $G(x)^\theta = G(ax + b)$. При этом экстремальная функция $\psi(s) = H(aH^{-1}(s) + b)$ по определению 1.

Доказательство. Согласно [1, следствие 1.3.2] для любого максимум-устойчивого закона существуют такие $a > 0$ и b , что $G(x)^\theta = G(ax + b)$. Пусть $x = H^{-1}(s)$. Поскольку

$$EF(A_n x + B_n)^{\nu_n} = E(F(a_{l_n} x + b_{l_n})^{l_n})^{\nu_n/l_n} \rightarrow EG(x)^\zeta, \quad n \rightarrow \infty,$$

то

$$H(x) = EG(x)^\zeta.$$

Тогда

$$EF(A_n x + B_n)^{\theta\nu_n} \rightarrow EG(x)^{\theta\zeta} = EG(ax + b)^\zeta = H(ax + b), \quad n \rightarrow \infty.$$

¹Если ζ вырождена и равна константе $c > 0$ почти наверное (п. н.), то очевидно $Eu^\zeta = u^c$.

По определению 2 отсюда следует $P(M_n \leq A_n x + B_n) \rightarrow H(ax + b)$, $n \rightarrow \infty$, а по определению 1 получаем $\psi(s) = H(aH^{-1}(s) + b)$. \square

Свойство 5. Пусть имеется система $\{\xi_{n,m}; \nu_n\}$, у которой:

- (а) $F_n \equiv F$;
- (б) существует такая последовательность $c_n \rightarrow +\infty$, что $\nu_n/c_n \xrightarrow{d} \zeta > 0$, $n \rightarrow \infty$;
- (в) для непрерывного распределения G и коэффициентов $A_n > 0$, B_n верно

$$F(A_n x + B_n)^{c_n} \rightarrow G(x);$$

$$P(M_n \leq A_n x + B_n) \rightarrow EG(x)^{\theta\zeta}, \quad n \rightarrow \infty.$$

Тогда θ является экстремальным индексом по определению 2.

Доказательство. Прежде всего получаем $EF(A_n x + B_n)^{\nu_n} \rightarrow EG(x)^\zeta$. Обозначим $H(x) = EG(x)^\zeta$ — это непрерывная функция, пробегающая все значения в $(0, 1)$. Полагаем $x = H^{-1}(s)$, $u_n(s) = A_n x + B_n$, тогда

$$EF(u_n(s))^{\nu_n} \rightarrow s; \quad EF(u_n(s))^{\theta\nu_n} \rightarrow EG(x)^{\theta\zeta}.$$

Поскольку по условию верно также $P(M_n \leq u_n(s)) \rightarrow EG(x)^{\theta\zeta}$, то

$$P(M_n \leq u_n(s)) - EF(u_n(s))^{\theta\nu_n} \rightarrow 0, \quad n \rightarrow \infty,$$

и θ является экстремальным индексом по определению 2. \square

Свойство 6. Пусть имеется система $\{\xi_{n,m}; l_n\}$ с экстремальным индексом θ по одному из определений, у которой:

- (а) $F_n \equiv F$ — непрерывное распределение;
- (б) l_n , $n \geq 1$, — целочисленная последовательность, $l_n \rightarrow +\infty$, $l_n \sim n^\alpha L(n)$, $n \rightarrow \infty$, $\alpha > 0$, $L(x)$ — медленно меняющаяся функция на \mathbb{R}_+ .

Пусть $\nu_n/l_n \xrightarrow{P} 1$, $n \rightarrow \infty$, тогда система $\{\xi_{n,m}; \nu_n\}$ имеет тот же экстремальный индекс по обоим определениям.

Доказательство. По свойству 3 каждая из систем имеет одинаковые экстремальные индексы по обоим определениям. Обозначим через M_n максимумы для системы $\{\xi_{n,m}; l_n\}$ и через \tilde{M}_n для системы $\{\xi_{n,m}; \nu_n\}$.

Для любого $\rho > 0$ верно $l_{[n\rho]} \sim \rho^\alpha l_n$, $n \rightarrow \infty$.

Поэтому из $\nu_n/l_n \xrightarrow{P} 1$, $n \rightarrow \infty$, следует

$$P(M_{[n(1-\varepsilon)]} \leq \tilde{M}_n \leq M_{[n(1+\varepsilon)]}) \rightarrow 1, \quad n \rightarrow \infty,$$

для любого $\varepsilon > 0$. Поскольку $F(u_n(s))^{l_n} \rightarrow s$, то

$$u_n(s) = F^{-1} \left(1 + \frac{(1 + o(1)) \ln s}{l_n} \right), \quad n \rightarrow \infty,$$

и для любого $\rho > 0$ в силу определения 1 верно

$$\begin{aligned} \mathbb{P}(M_{[n\rho]} \leq u_n(s)) &= \\ &= \mathbb{P} \left(M_{[n\rho]} \leq F^{-1} \left(1 + \frac{(1 + o(1)) \ln s}{l_n} \right) \right) = \\ &= \mathbb{P} \left(M_{[n\rho]} \leq F^{-1} \left(1 + \frac{(1 + o(1)) \ln s^{1/\rho^\alpha}}{l_{[n\rho]}} \right) \right) \rightarrow \\ &\rightarrow s^{\theta/\rho^\alpha}, \quad n \rightarrow \infty. \end{aligned}$$

Полагая $\rho = 1 \pm \varepsilon$, $\varepsilon > 0$, получаем

$$\begin{aligned} s^{\theta/(1+\varepsilon)^\alpha} &\leq \liminf_{n \rightarrow \infty} \mathbb{P} \left(\tilde{M}_n \leq u_n(s) \right) \leq \\ &\leq \limsup_{n \rightarrow \infty} \mathbb{P} \left(\tilde{M}_n \leq u_n(s) \right) \leq s^{\theta/(1-\varepsilon)^\alpha}. \end{aligned}$$

Переходя к пределу по $\varepsilon \rightarrow 0$, получаем $\lim_{n \rightarrow \infty} \mathbb{P}(\tilde{M}_n \leq u_n(s)) = s^\theta$; значит, θ является экстремальным индексом системы $\{\xi_{n,m}; \nu_n\}$ по определению 1. \square

Свойство 7. Пусть имеется система $\{\xi_{n,m}; \nu_n\}$ с экстремальным индексом θ по определению 1, у которой $F_n \equiv F$ — непрерывное распределение, $\nu_n/n \xrightarrow{P} c > 0$, $n \rightarrow \infty$, и пусть имеются целочисленная случайная последовательность η_n , не зависящая от $\{\xi_{n,m}; \nu_n\}$, и числовая последовательность $\mu_n \rightarrow +\infty$ такие, что $\eta_n/\mu_n \xrightarrow{d} \zeta > 0$, $n \rightarrow \infty$. Положим $\tilde{\xi}_{n,m} = \xi_{\eta_n, m}$, $\tilde{\nu}_n = \nu_{\eta_n}$, тогда система $\{\tilde{\xi}_{n,m}; \tilde{\nu}_n\}$ имеет экстремальный индекс θ по определению 2.

Доказательство. Для последовательности $u_n(s)$ из определения 1 для системы $\{\xi_{n,m}; \nu_n\}$ имеет место сходимость $EF(u_n(s))^{\nu_n} \rightarrow s$, откуда $F(u_n(s))^{cn} \rightarrow s$, $n \rightarrow \infty$, так что

$$u_n(s) = F^{-1} \left(1 + \frac{(1 + o(1)) \ln s}{cn} \right), \quad n \rightarrow \infty.$$

Для любого $x \in (0, 1)$ имеем

$$\begin{aligned} EF(u_{[\mu_n]}(x))^{\tilde{\nu}_n} &= EF(u_{[\mu_n]}(x))^{\tilde{\nu}_n} = \\ &= EF(u_{[\mu_n]}(x))^{(\nu_{\eta_n}/\eta_n)(\eta_n/\mu_n)\mu_n} \rightarrow Ex^\zeta, \quad n \rightarrow \infty. \end{aligned}$$

Обозначим

$$H(x) = Ex^\zeta; \quad x = H^{-1}(s); \quad \tilde{u}_n(s) = u_{[\mu_n]}(x),$$

тогда

$$EF(\tilde{u}_n(s))^{\tilde{\nu}_n} \rightarrow s; \quad EF(\tilde{u}_n(s))^{\theta \tilde{\nu}_n} \rightarrow Ex^{\theta \zeta}, \quad n \rightarrow \infty.$$

С другой стороны, по определению 1 для системы $\{\xi_{n,m}; \nu_n\}$ получаем

$$\begin{aligned} \mathbb{P}(\tilde{M}_n \leq \tilde{u}_n) &= \mathbb{P}(M_{\eta_n} \leq u_{[\mu_n]}(x)) = \\ &= \mathbb{P} \left(M_{\eta_n} \leq F^{-1} \left(1 + \frac{(1 + o(1)) \ln x}{c\mu_n} \right) \right) = \\ &= \mathbb{P} \left(M_{\eta_n} \leq F^{-1} \left(1 + \frac{(1 + o(1)) \ln x^{\eta_n/\mu_n}}{c\eta_n} \right) \right) \rightarrow \\ &\rightarrow Ex^{\theta \zeta}, \quad n \rightarrow \infty. \end{aligned}$$

Следовательно,

$$\mathbb{P}(\tilde{M}_n \leq \tilde{u}_n) - EF(\tilde{u}_n(s))^{\theta \tilde{\nu}_n} \rightarrow 0, \quad n \rightarrow \infty,$$

и θ является экстремальным индексом системы $\{\tilde{\xi}_{n,m}; \tilde{\nu}_n\}$ по определению 2. \square

Прокомментируем доказанные свойства.

Свойство 1 означает, что введенные индексы действительно являются обобщениями классического экстремального индекса (по определению А) и совпадают с ним, если в качестве серий брать растущие детерминированным образом куски стационарной последовательности.

Свойство 2 означает инвариантность экстремальных индексов относительно непрерывных строго возрастающих преобразований серий. Это означает, например, что в случае непрерывных случайных величин их все можно привести к равномерному распределению на $[0, 1]$ преобразованием с $g_n = F_n$. Аналогичное свойство имеет место для классического экстремального индекса (по определению А), когда речь идет об одном непрерывном строго возрастающем преобразовании ко всем членам последовательности.

Свойство 3 определяет ограничение на случайность длин серий, при котором оба новых индекса эквивалентны. Длины должны расти асимптотически эквивалентно неслучайной последовательности.

Свойство 4 обобщает известное утверждение для классического экстремального индекса [1, следствие 3.7.3]: предельное распределение максимумов стационарной последовательности имеет тот же экстремальный тип, что и предельное распределение максимумов независимых случайных величин с тем же частным распределением. В данном случае предельный закон уже не обязан быть максимум-устойчивым, однако его тип сохраняется. Максимум-устойчив он только при вырожденной случайной величине ζ (константе), т. е. при условии свойства 3.

Свойство 5 позволяет интерпретировать некоторый параметр предельного распределения максимумов зависимых случайных величин как экстремальный индекс по определению 2.

Свойство 6 дает достаточное условие на скорость роста длин серий, при котором индексы для случайных и неслучайных длин совпадают.

Свойство 7 показывает, что рандомизацией по случайно растущему номеру серии можно перейти от экстремального индекса по определению 1 к тому же индексу по определению 2.

3 Приложения к моделям информационных сетей

В работах автора [16–18] рассматривались максимумы суммарной активности в информационных сетях, описываемых степенными случайными графами (которые в отечественной литературе называют также интернет-графами или графами интернет-типа).

В качестве примеров недавних работ отечественных авторов о степенных графах можно указать [19, 20] и обзор [21]. Рекомендуем также зарубежный электронный учебник [22].

Рассмотрим следующий пример модели информационной сети [17].

Пусть каждый узел сети обладает индивидуальной случайной информационной активностью (интенсивностью производства информации). Предположим, что активности узлов независимы и одинаково распределены, причем их распределение A имеет тяжелый (правильно меняющийся) хвост, т. е. $\bar{A}(x) \sim x^{-a} L(x)$, $x \rightarrow \infty$, $a > 0$, где $L(x)$ — медленно меняющаяся функция. Активности и степени вершин (узлов) полагаются независимыми, и это существенное предположение.

Рассмотрим модель ориентированного случайного графа, где направления ребер соответствуют направлениям передачи информации. Пусть имеется n вершин и заданы независимые неотрицательные целочисленные случайные величины K_1, \dots, K_n , имеющие одинаковое распределение, заданное вероятностями $p_k \sim ck^{-\beta}$, $k \rightarrow \infty$, $\beta > 2$. Положим $D_i = \min\{K_i, n - 1\}$. Для i -й вершины выберем случайным образом (равновероятно и независимо от выбора для других вершин) D_i различных вершин из числа остальных (кроме i -й) и выпустим из них ребра, направленные в i -ю вершину. Полученный в результате граф можно отнести к степенным в том смысле, что число входящих ребер распределено асимптотически по степенному закону. Суммарной активностью в узле в данном случае будем считать сумму собственной активности узла и всех узлов, из которых в него поступает информация (его входящих соседей).

Подчеркнем, что активность никак не связана с понятиями «качества» или «веса» вершины, используемых в современных моделях формирования случайных графов. Здесь граф сформирован по описанному выше алгоритму, а индивидуальные активности представляют собой независимые дополнения к графу. Если говорить о социальных сетях, то бывает, что пользователь пишет много, но мало кого читает (или его мало кто читает), и наоборот, бывает, что он пишет мало, но читает многих (или его читают многие). Что касается суммарной активности, то она может оказаться большой за счет малого числа входящих соседей с большой индивидуальной активностью (или даже одного соседа), и наоборот, небольшой при большом числе входящих соседей с малой индивидуальной активностью. Как известно, для тяжелых хвостов характерно, что большие значения суммы принимаются за счет одного большого (максимального) слагаемого. Это свойство в данном случае обобщается и на суммы случайного числа слагаемых.

Обозначим через M_n максимум суммарных активностей. Пусть $v(r)$ — положительная неубывающая функция такая, что верно $r\bar{A}(v(r)) \rightarrow 1$, $r \rightarrow \infty$. Заметим, что $v(r)$ заведомо существует и является правильно меняющейся с показателем $1/a$ [23, §1.5].

Тогда при $a < \beta - 2$, если $2 < \beta < 3$, и при $a < (\beta - 1)/2$, если $\beta \geq 3$, имеет место предельный закон Фреше:

$$P\left(\frac{M_n}{v(n)} \leq x\right) = \exp\{-x^{-a}\}, \quad x > 0, \quad n \rightarrow \infty.$$

Отметим, что этот предельный закон обусловлен тем, что максимум суммарных активностей растет асимптотически эквивалентно (по вероятности) максимуму индивидуальных активностей по сети, что и доказано в [17, теорема 1].

С другой стороны, если число входящих соседей описывается случайной величиной K , не зависящей от активности, то предельное распределение F суммарной активности в каждом узле имеет хвост

$$\bar{F}(x) \sim (1 + EK)\bar{A}(x), \quad x \rightarrow \infty,$$

при условии (которое в данном случае выполняется)

$$EK^{1V(a+\varepsilon)} < \infty, \quad \varepsilon > 0,$$

согласно результатам [24] о распределении суммы случайного числа независимых случайных величин с тяжелыми хвостами. Поэтому

$$F(xv(n))^n \rightarrow \exp\{-(1 + EK)x^{-a}\}, \quad x > 0, \quad n \rightarrow \infty.$$

Обозначив

$$s = \exp\{-(1 + EK)x^{-a}\} \in (0, 1); \quad u_n(s) = xv(n),$$

приходим к выводу, что система суммарных активностей имеет экстремальный индекс $\theta = 1/(1 + EK)$ по определению 1 (а значит, и по определению 2 в силу свойства 3, поскольку $\nu_n = n$).

Для последовательностей значение $\theta \in (0, 1)$ означает, что превышения высокого уровня происходят не по одиночке, а группами (кластерами) средней величины $1/\theta$ [3, §8.1]. В рассматриваемом случае также можно предположить образование подобных кластеров.

Применительно к информационным сетям речь может идти о группах узлов с высокими суммарными активностями, вызванными высокой индивидуальной активностью одного узла, являющегося их общим входящим соседом.

Проверка наличия подобного эффекта в реальных сетях, разумеется, требует экспериментального исследования, выходящего за рамки данной работы, которая имеет теоретический характер.

Следует также признать, что выбор модели случайного графа в [17] был обусловлен не какими-то ее преимуществами в описании реальных сетей по сравнению с другими современными моделями (например, моделями преимущественного присоединения), а относительной простотой доказательства асимптотической эквивалентности роста максимумов суммарных и индивидуальных активностей методами работы автора [25]. При этом просто знания степенного закона для числа входящих вершин совершенно недостаточно и каждая модель случайного графа должна рассматриваться в индивидуальном порядке. Например, важна скорость роста максимальной степени вершины в графе. Если искусственно урезать степени вершин на растущем (с числом вершин) пороге, можно получить класс моделей с одинаковыми предельными распределениями степеней вершин, но разными скоростями роста максимальной степени, для которых будут возникать разные ограничения для a в зависимости от β . Играют также роль другие нюансы.

Недавно автором получены новые результаты для простых моделей с весами [18]. Похожие модели рассматривались в [22, гл. 6] как обобщенные случайные графы (generalized random graphs). Предполагается, что вершинам приписываются независимые веса w_i , $1 \leq i \leq n$, одинаково распределенные как неотрицательная случайная величина W , $EW^\beta < \infty$, $\beta \geq 1$.

В модели 1 полагаем $p_i = \varphi(w_i n^{-s/2})$, где $0 < s \leq 1$, и для φ на \mathbb{R}_+ верно $0 \leq \varphi(x) \leq \min\{1, x\}$, $\varphi(x) \sim x$, $x \rightarrow 0$. При известных w_i , $1 \leq i \leq n$, каждая пара вершин i и j соединяется ребром с вероятностью $p_i p_j$ независимо от других пар. Здесь граф полагается неориентированным, информация передается по ребру в обе стороны. Применительно

к социальным сетям веса могут отражать общительность пользователей.

В модели 2 при тех же предположениях о весах полагаем $p_i = \varphi(w_i n^{-s})$, $0 < s \leq 1$. При известных w_i , $1 \leq i \leq n$, в i -ю вершину входит ребро из любой другой вершины с вероятностью p_i независимо от других ребер. Здесь граф полагается ориентированным, информация передается по направлению ребра. Применительно к социальным сетям веса могут отражать любознательность пользователей.

В обоих случаях доказана асимптотическая эквивалентность роста максимумов суммарных и индивидуальных активностей при некоторых ограничениях на a в зависимости от β и s . При $s = 1$ существуют предельные распределения числа соседей (входящих соседей), а полученные результаты можно интерпретировать как наличие экстремальных индексов: $\theta = 1/(1 + (EW)^2)$ в модели 1 и $\theta = 1/(1 + EW)$ в модели 2 (по обоим определениям), аналогично предыдущему примеру.

Приложение разработанных методов к другим, более сложным и популярным моделям информационных сетей представляет собой дело будущего.

4 Приложение к моделям биологических популяций

В качестве моделей биологических популяций часто используются ветвящиеся процессы. При этом элементы популяции традиционно называют частицами. Частицы могут обладать какими-то случайными (количественными) признаками.

В случае живых организмов речь может идти о размерах, весе и других характеристиках, например удоях коров, яйценоскости кур, урожайности растений, чувствительности организмов к вредным и опасным факторам и др.

Ветвящимися процессами может также описываться распространение компьютерных вирусов. Полиморфные компьютерные вирусы способны не только размножаться, но и изменять свой код (подобно мутациям живых организмов). В качестве признаков могут рассматриваться какие-то характеристики кода вируса или его деятельности.

В [26, 27] изучались максимумы независимых случайных признаков частиц в ветвящихся процессах. При этом в качестве приложений в [26] речь шла о росте людей, а в [27] упоминалось разведение скаковых лошадей, с призовыми очками в качестве признака.

Приведем еще один пример. Если имеется колония вредных организмов с различными индивиду-

альными порогами чувствительности к какому-то фактору (яду, антибиотику или др.), то для уничтожения всей колонии нужна максимальная концентрация, иначе часть организмов колонии выживет и вновь размножится.

В ряде работ автора рассматривались максимумы случайных признаков частиц в надкритических ветвящихся процессах без вырождения (с конечными средним и дисперсией числа потомков). Так, в [28] рассмотрен процесс с непрерывным временем, а в [29, 30] — с дискретным. Однако при этом признаки разных частиц считались независимыми. В [31] впервые изучалась модель с зависимостью признаков частиц в поколении, обусловленной их общей наследственностью.

Сначала рассмотрен случай, когда признаки имеют стандартное нормальное распределение, а коэффициент корреляции признаков пары частиц мажорируется величиной r^k , $r \in (0, 1)$, если эти частицы имеют ближайшего общего предка k поколений назад. Показано, что максимумы по поколениям растут асимптотически так же, как в случае независимых признаков, что соответствует $\theta = 1$.

Далее рассмотрен случай, когда признаки имеют распределение с правильно меняющимся хвостом, а наследственность явно описывается процессом линейной авторегрессии первого порядка:

$$\xi_{n,m} = a\xi_{n-1,\kappa(n,m)} + b\xi_{n,m}^*, \quad a \in (0, 1), \quad b > 0, \quad (4)$$

где $\xi_{n,m}$ — признак m -й частицы в n -м поколении; $\kappa(n, m)$ — номер предка этой частицы в предыдущем поколении, а случайные величины $\xi_{n,m}^*$, $m \geq 1$, $n \geq 1$, независимы и имеют одинаковое распределение A , которое удовлетворяет условиям

$$\left. \begin{aligned} \bar{A}(x) \sim x^{-\gamma} L(x); \quad \frac{A(-x)}{\bar{A}(x)} \rightarrow p \geq 0; \\ x \rightarrow \infty; \quad \gamma > 0, \end{aligned} \right\} \quad (5)$$

где $L(x)$ — медленно меняющаяся функция.

В модели (4) существует и единственно стационарное распределение F . Предполагается, что все признаки частиц имеют это стационарное распределение. Для выявления роли наследственности «в чистом виде» желательно обеспечить независимость распределения признаков от коэффициентов авторегрессии (как это имело место в гауссовском случае). Здесь можно добиться этого только для строго устойчивых распределений с $0 < \gamma < 2$, полагая

$$a^\gamma + b^\gamma = 1. \quad (6)$$

Для произвольных A , удовлетворяющих (5), условие (6) обеспечивает асимптотическую эквивалентность хвостов: $\bar{F}(x) \sim \bar{A}(x)$, $x \rightarrow \infty$. Предполагается, что это условие выполнено.

В этом случае фактически показано, что

$$\theta = \frac{1 - a^\gamma}{1 - a^\gamma/\mu} \in (0, 1)$$

по определению 2 (по свойству 5), где $\mu > 1$ — среднее число потомков. Заметим, что θ стремится к 1 как при уменьшении параметра зависимости a , так и при уменьшении среднего числа потомков μ . Экстремального индекса по определению 1 в данном случае не существует.

Здесь также можно предположить образование кластеров. Очевидно, речь идет о родственных группах частиц, имеющих общего предка с аномально большим признаком и унаследовавшим эту мутацию. Этот вывод иллюстрируется компьютерным моделированием в [31].

5 Модели с копулами

Напомним понятия из теории копул [32, гл. 5, § 7.5; 33].

Копулой C называется функция многомерного распределения на $[0, 1]^d$, $d \geq 2$, если все частные распределения являются равномерными на $[0, 1]$. Согласно теореме Склера любая функция многомерного распределения в \mathbb{R}^d представима в виде:

$$G(x_1, \dots, x_d) = C(G_1(x_1), \dots, G_d(x_d)),$$

где G_i , $1 \leq i \leq d$, — функции частных распределений. Таким образом, всякому многомерному распределению можно поставить в соответствие его копулу. Если частные распределения непрерывны, то такое представление единственно.

Вектору с независимыми компонентами соответствует копула независимости

$$C(y_1, \dots, y_d) = y_1 \dots y_d.$$

В настоящее время математический аппарат копул активно используется в самых разных приложениях, в том числе проникает и в информатику. Отметим работу по рекуррентным нейронным сетям [34], где использовались копулы Стьюдента, Клейтона и Гумбеля. На их основе производилось успешное обучение человекоподобного робота.

В общем случае копулы могут описывать зависимость в поведении компонент сложных систем, обусловленную их взаимодействием или влиянием общих внешних факторов. В моделях предыдущих разделов зависимости суммарных активностей в сети и признаков частиц в поколении также описываются некими копулами, которые, однако, затруднительно выписать в явном виде и поэтому

лучше использовать иные методы. Отметим, что общение пользователей сети может приводить к зависимости их индивидуальных информационных активностей, что не учитывалось в [17]. В технических системах износ или выход из строя одних деталей может сказываться на других деталях, а на все вместе может влиять общий режим эксплуатации (температура, влажность и т. п.).

В финансах копулы используются для описания зависимости между колебаниями курсов различных акций и валют [32]. Эту зависимость следует учитывать как в финансовых расчетах, так и при программировании торговых (финансовых) роботов (black boxes).

Далее для простоты будем полагать, что $\nu_n = n$ (треугольная схема), $F_n(x) \equiv x$, $x \in [0, 1]$, а случайные величины $\xi_{n,m}$, $1 \leq m \leq n$, связаны n -мерной копулой C_n . Напомним, что к равномерному распределению можно перейти от любого непрерывного в силу свойства 2.

Пусть для любого $s \in (0, 1)$ последовательность $u_n(s)$ такова, что $u_n(s)^n \rightarrow s$, $n \rightarrow \infty$. Тогда

$$u_n(s) = 1 + \frac{(1 + o(1)) \ln s}{n}, \quad n \rightarrow \infty.$$

Пример 5.1. Копула Гумбеля–Хоугарда. Данная копула имеет вид:

$$C(y_1, \dots, y_d) = \exp \left\{ - \left(\sum_{i=1}^d (-\ln y_i)^\alpha \right)^{1/\alpha} \right\}, \quad \alpha \geq 1,$$

откуда следует

$$C(y, \dots, y) = y^{d^{1/\alpha}}.$$

Полагая C_n копулой Гумбеля–Хоугарда с $\alpha_n \geq 1$ и $(\alpha_n - 1) \ln n \rightarrow \gamma \geq 0$, получаем

$$P(M_n \leq u_n(s)) = u_n(s)^{n^{1/\alpha_n}} \rightarrow s^\theta, \quad \theta = e^{-\gamma} \in [0, 1].$$

Данная копула относится к классу копул экстремальных значений (или максимум-устойчивых). В общем случае для них имеет место представление Пикандса [32, с. 312, теорема 7.45]:

$$C(y_1, \dots, y_d) = \exp \left\{ B \left(\frac{\ln y_1}{\sum_{i=1}^d \ln y_i}, \dots, \frac{\ln y_d}{\sum_{i=1}^d \ln y_i} \right) \sum_{i=1}^d \ln y_i \right\}.$$

Здесь

$$B(w_1, \dots, w_d) = \int_{S^d} \left(\bigvee_{i=1}^d x_i w_i \right) dH(x),$$

где H — конечная мера на $S^d = \{x = (x_1, \dots, x_d) : x_i \geq 0, \sum_{i=1}^d x_i = 1\}$, причем эта мера должна быть нормирована так, что $\int_{S^d} x_i dH(x) = 1$ для всех $1 \leq i \leq d$ (о чем в [32] забыли упомянуть).

Заметим, что функция B однородна первого порядка. Таким образом, в общем случае верно

$$C(y, \dots, y) = y^{B(1, \dots, 1)}.$$

Обозначим $\beta_n = B_n(1, \dots, 1)$, тогда если $\beta_n/n \rightarrow \theta$, то θ — экстремальный индекс (по обоим определениям). Поскольку $0 \leq \beta_n \leq n$, то $\theta \in [0, 1]$.

Пример 5.2. Копула Клейтона. Данная копула имеет вид:

$$C(y_1, \dots, y_d) = \left(\sum_{i=1}^d y_i^{-\alpha} - d + 1 \right)^{-1/\alpha}, \quad \alpha \geq 0,$$

где вырожденный случай $\alpha = 0$ соответствует копуле независимости, возникающей в пределе при $\alpha \rightarrow 0$. Отсюда

$$C(y, \dots, y) = (d(y^{-\alpha} - 1) + 1)^{-1/\alpha}.$$

Пусть C_n — копула Клейтона с $\alpha_n \equiv \alpha > 0$, тогда

$$P(M_n \leq u_n(s)) = (n(u_n(s)^{-\alpha} - 1) + 1)^{-1/\alpha} \rightarrow (1 - \alpha \ln s)^{-1/\alpha} = \psi(s).$$

Здесь $\theta^- = 0$, $\theta^+ = 1$.

Пример 5.3. Копула Франка. Данная копула имеет вид:

$$C(y_1, \dots, y_d) = -\frac{1}{\alpha} \ln \left(1 - \frac{\prod_{i=1}^d (1 - e^{-\alpha y_i})}{(1 - e^{-\alpha})^{d-1}} \right), \quad \alpha \geq 0,$$

где вырожденный случай $\alpha = 0$ соответствует копуле независимости, возникающей в пределе при $\alpha \rightarrow 0$. Отсюда

$$C(y, \dots, y) = -\frac{1}{\alpha} \ln \left(1 - \frac{(1 - e^{-\alpha y})^d}{(1 - e^{-\alpha})^{d-1}} \right).$$

Пусть C_n — копула Франка с $\alpha_n \equiv \alpha > 0$, тогда, переходя к пределу, получаем:

$$P(M_n \leq u_n(s)) \rightarrow -\frac{1}{\alpha} \ln \left(1 - (1 - e^{-\alpha}) s^{\alpha/(e^\alpha - 1)} \right) = \psi(s).$$

В этом случае

$$\lim_{s \rightarrow 0} \log_s \psi(s) = \frac{\alpha}{e^\alpha - 1}; \quad \lim_{s \rightarrow 1} \log_s \psi(s) = 1,$$

а на интервале $(0, 1)$ функцией принимаются промежуточные значения. Поэтому $\theta^- = \alpha/(e^\alpha - 1) \in (0, 1)$, $\theta^+ = 1$.

Во всех трех примерах имеют место строго архимедовы копулы. Напомним, что строго архимедовой называется копула вида

$$C(y_1, \dots, y_d) = \varphi^{-1} \left(\sum_{i=1}^d \varphi(y_i) \right), \quad (7)$$

где φ — убывающая функция на $[0, 1]$, называемая генератором, $\varphi(0) = +\infty$, $\varphi(1) = 0$. При $d = 2$ достаточно, чтобы эта функция была выпуклой. Если потребовать, чтобы функция φ^{-1} была вполне монотонной на $(0, +\infty)$, то формула (7) определяет копулу при любом $d \geq 2$ [33, теорема 4.6.2]. Далее будем считать это условие на φ выполненным.

С другой стороны, функция f является преобразованием Лапласа–Стилтьеса некоторого распределения тогда и только тогда, когда f вполне монотонна и $f(0) = 1$ [35, гл. 13, §4, теорема 1]. Отсюда следует, что функция φ^{-1} должна быть преобразованием Лапласа–Стилтьеса некоторого распределения, причем в силу условия $\varphi(0) = +\infty$, а значит, и $\varphi^{-1}(+\infty) = 0$, это распределение не должно иметь атомов в нуле. Таким образом, существует некоторая случайная величина $\zeta > 0$ п. н. такая, что

$$\varphi^{-1}(u) = \mathbb{E}e^{-u\zeta}, \quad u \geq 0.$$

Введем обозначения:

$$x_0 = \inf\{x > 0 : \mathbb{P}(\zeta \leq x) > 0\}, \quad \mu = \mathbb{E}\zeta.$$

Будем для краткости обозначать $f(u) = \varphi^{-1}(u)$.

Теорема 5.1. Пусть $\mu < \infty$, тогда экстремальная функция $\psi(s) = f(-(\ln s)/\mu) = \mathbb{E}s^{\zeta/\mu}$, $\theta^+ = 1$, $\theta^- = x_0/\mu$.

Доказательство. Поскольку $1 - f(u) \sim \mu u$, $u \rightarrow 0 + 0$, то $\varphi(1 - t) \sim t/\mu$, $t \rightarrow 1 - 0$. Имеем:

$$\begin{aligned} \mathbb{P}(M_n \leq u_n(s)) &= f(n\varphi(u_n(s))) = \\ &= f\left(n\varphi\left(1 + \frac{(1 + o(1)) \ln s}{n}\right)\right) \rightarrow f\left(-\frac{\ln s}{\mu}\right), \\ & \qquad \qquad \qquad n \rightarrow \infty. \end{aligned}$$

Из неравенства Иенсена получаем:

$$\psi(s) = \mathbb{E}s^{\zeta/\mu} \geq s^{\mathbb{E}\zeta} = s.$$

С другой стороны, поскольку $\zeta > 0$ п. н., то $\psi(s) \leq s^{x_0/\mu}$. Следовательно, $\theta^+ \leq 1$ и $\theta^- \geq x_0/\mu$. Кроме того, получаем:

$$\lim_{s \rightarrow 0} \log_s \psi(s) = \frac{x_0}{\mu}; \quad \lim_{s \rightarrow 1} \log_s \psi(s) = 1,$$

так что эти оценки достигаются в пределе и верно $\theta^+ = 1$; $\theta^- = x_0/\mu$. \square

В случае копулы Клейтона генератор $\varphi(t) = t^{-\alpha} - 1$ и обратная функция $f(u) = 1/(1 + u)^{1/\alpha}$ соответствуют гамма-распределению с параметром формы $1/\alpha$, для которого $x_0 = 0$, так что $\theta^- = 0$; $\theta^+ = 1$.

В случае копулы Франка генератор $\varphi(t) = -\ln((1 - e^{-\alpha t})/(1 - e^{-\alpha}))$ и обратная функция $f(u) = -(1/\alpha)\ln(1 - (1 - e^{-\alpha})e^{-u})$ соответствуют дискретному распределению с вероятностями $\mathbb{P}(\zeta = k) = (1 - e^{-\alpha})^k/(\alpha k)$, $k \geq 1$. Тогда $x_0 = 1$ и $\mu = f'(0) = (e^\alpha - 1)/\alpha$, откуда $\theta^- = \alpha/(e^\alpha - 1)$; $\theta^+ = 1$.

В условия теоремы 5.1 из рассмотренных не вписывается только пример копулы Гумбеля–Хоугаарда, поскольку она имеет генератор $\varphi(t) = (-\ln t)^\alpha$ с обратной функцией $f(u) = \exp\{-u^{1/\alpha}\}$, $\alpha \geq 1$, что соответствует асимметричному $(1/\alpha)$ -устойчивому распределению на \mathbb{R}_+ , не имеющему конечного среднего.

Для изучения таких случаев применим следующую модификацию.

Заметим, что если $\varphi(t)$ — генератор с вполне монотонной обратной функцией, то $\varphi(t)^\beta$, $\beta \geq 1$, — также генератор с вполне монотонной обратной функцией [33, лемма 4.6.4].

Теорема 5.2. Пусть n -мерная копула C_n имеет генератор $\varphi_n(t) = \varphi(t)^{\beta_n}$, где $\beta_n \geq 1$, $(\beta_n - 1) \ln n \rightarrow \gamma \geq 0$, и для генератора $\varphi(t)$ верно $\mu < \infty$. Тогда

$$\psi(s) = f\left(-\frac{e^{-\gamma} \ln s}{\mu}\right); \quad \theta^- = \frac{x_0}{\mu} e^{-\gamma}; \quad \theta^+ = e^{-\gamma}.$$

Доказательство. Из $\varphi_n(t) = \varphi(t)^{\beta_n}$ следует $f_n(u) = f(u^{1/\beta_n})$. Имеем

$$\begin{aligned} \mathbb{P}(M_n \leq u_n(s)) &= f_n(n\varphi_n(u_n(s))) = \\ &= f\left(n^{1/\beta_n} \varphi\left(1 + \frac{(1 + o(1)) \ln s}{n}\right)\right) = \\ &= f\left(e^{((1-\beta_n) \ln n)/\beta_n} \left(-\frac{\ln s}{\mu}\right)\right) \rightarrow \\ & \qquad \qquad \qquad \rightarrow f\left(-\frac{e^{-\gamma} \ln s}{\mu}\right), \quad n \rightarrow \infty. \end{aligned}$$

Частичные индексы получаем из соотношения:

$$\begin{aligned} \log_s \psi(s) &= \frac{\ln f(-e^{-\gamma}(\ln s)/\mu)}{\ln s} = \\ &= e^{-\gamma} \frac{\ln f(-(\ln r)/\mu)}{\ln r}, \quad r = s^{e^{-\gamma}} \in (0, 1), \end{aligned}$$

где дробь в правой части представляет собой логарифм экстремальной функции из теоремы 5.1, принимающий значения от x_0/μ до 1. \square

В частности, результат примера 5.1 для копулы Гумбеля–Хоугаарда получаем при $\varphi_n(t) = (-\ln t)^{\alpha_n}$, где $\varphi(t) = -\ln t$ соответствует $\zeta = 1$ п. н. и копуле независимости.

Таким образом, видно, что в рассмотренных моделях с копулами могут быть любые $\theta \in [0, 1]$ и любые $0 \leq \theta^- < \theta^+ \leq 1$.

Рассмотрим теперь один поучительный пример модели с копулами и случайными длинами серий.

Пример 5.4. Пусть длины серий удовлетворяют условию $\nu_n/n \xrightarrow{d} \zeta$, $n \rightarrow \infty$, где ζ имеет устойчивое распределение с преобразованием Лапласа–Стилтьеса $Ee^{-u\zeta} = e^{-u^\beta}$, $0 < \beta < 1$, и в каждой серии случайные величины (не зависящие от ν_n) связаны копулой Гумбеля–Хоугаарда с $\alpha_n > 1$, $(\alpha_n - 1) \ln n \rightarrow \gamma > 0$, $n \rightarrow \infty$ (см. пример 5.1).

Предположим сначала, что

$$u_n(s)^n \rightarrow e^{-\tau}, \quad n \rightarrow \infty, \quad \tau > 0.$$

Тогда

$$Eu_n(s)^{\nu_n} = E(u_n(s)^n)^{\nu_n/n} \rightarrow Ee^{-\tau\zeta} = e^{-\tau^\beta}, \quad n \rightarrow \infty.$$

Возьмем $\tau = (-\ln s)^{1/\beta}$, тогда $Eu_n(s)^{\nu_n} \rightarrow s$, что и требовалось.

Далее имеем:

$$\begin{aligned} P(M_n \leq u_n(s)) &= u_n(s)^{\nu_n^{1/\alpha_n}} = \\ &= \left(u_n(s)^{n^{1/\alpha_n}}\right)^{(\nu_n/n)^{1/\alpha_n}} \rightarrow \\ &\rightarrow Ee^{-\tau e^{-\gamma}\zeta} = e^{-(e^{-\gamma}\tau)^\beta} = s^{e^{-\gamma\beta}}, \quad n \rightarrow \infty. \quad (8) \end{aligned}$$

Отсюда экстремальный индекс по определению 1 равен $e^{-\gamma\beta}$.

С другой стороны, для любого $\theta > 0$ верно

$$Eu_n(s)^{\theta\nu_n} \rightarrow Ee^{-\tau\theta\zeta} = e^{-(\theta\tau)^\beta} = s^{\theta^\beta}, \quad n \rightarrow \infty,$$

откуда и из (8) следует, что экстремальный индекс по определению 2 равен $e^{-\gamma}$.

Таким образом, система имеет два *разных* экстремальных индекса по двум разным определениям!

Во всех рассмотренных ранее моделях классическое свойство (3) сохраняло силу в форме $\psi(s) \geq s$ при всех $s \in [0, 1]$. В заключение рассмотрим пример, когда оно нарушается. При этом симметричную зависимость случайных величин в серии можно описать некоторой копулой, но проще сделать это конструктивно, построением.

Пример 5.5. Пусть $\eta_{n,m}$, $m \geq 1$, $n \geq 1$, независимы и имеют равномерное распределение на $[0, 1]$, $\nu_n = n$, κ_n принимают значения от 1 до n равномерно и не зависят от $\eta_{n,m}$, $1 \leq m \leq n$; $\gamma > 0$. Положим

$$\xi_{n,m} = \begin{cases} \eta_{n,m}^{1/(\gamma n)}, & m = \kappa_n; \\ \eta_{n,m}, & m \neq \kappa_n. \end{cases}$$

Тогда совместная функция распределения $\xi_{n,m}$, $1 \leq m \leq n$, имеет вид:

$$F^{(n)}(x_1, \dots, x_n) = \left(\prod_{m=1}^n x_m\right) \left(\frac{1}{n} \sum_{m=1}^n x_m^{\gamma n-1}\right),$$

откуда

$$\begin{aligned} F_n(x) &= x \left(1 + \frac{x^{\gamma n-1} - 1}{n}\right); \\ P(M_n \leq x) &= x^{(1+\gamma)n-1}. \end{aligned}$$

Полагая

$$u_n(s) = 1 - \frac{(1 + o(1))\tau}{n}, \quad \tau > 0,$$

получаем при $n \rightarrow \infty$:

$$\begin{aligned} F_n(u_n(s))^n &\rightarrow e^{-\tau} \exp\{e^{-\gamma\tau} - 1\} = s; \\ P(M_n \leq u_n(s)) &\rightarrow e^{-(1+\gamma)\tau} = \psi(s). \end{aligned}$$

Отсюда можно найти в явном виде обратную экстремальную функцию:

$$\psi^{-1}(u) = u^{1/(1+\gamma)} \exp\{u^{\gamma/(1+\gamma)} - 1\},$$

для которой верно $\psi^{-1}(u) > u$ при всех $u \in (0, 1)$, а значит, $\psi(s) < s$ при всех $s \in (0, 1)$. В этом случае можно показать, что $\theta^- = 1$, $\theta^+ = 1 + \gamma > 1$.

6 Пороговые модели

До сих пор рассматривались модели, в которых величина ν_n определялась внешними причинами по отношению к величинам $\{\xi_{n,m}\}$. Теперь введем модели, в которых ν_n представляет собой момент остановки относительно последовательности

$\{\xi_{n,m}, m \geq 1\}$, где $\xi_{n,m}, m \geq 1$, независимы и имеют равномерное распределение на $[0, 1]$, а остановка происходит в момент превышения очередной случайной величиной некоторого порога.

Подобные модели могут возникать, например, в задачах автоматизированного поиска объектов, обладающих некоторыми свойствами, простым перебором.

Заметим, что в [5, 36] рассматривалась модель максимумов случайных величин, в которой остановка происходила в момент превышения порога (а именно, времени t) не очередной величиной, а их накопленной суммой. Это обобщение классической задачи о наиболее длинной серии успехов в испытаниях Бернулли [3, §8.5]. Однако в этом случае момент остановки растет просто асимптотически пропорционально порогу и не происходит таких интересных эффектов, как в рассматриваемых моделях.

Пример 6.1. Пусть задана числовая последовательность $a_n \in (0, 1), n \geq 1$, и $a_n \rightarrow 1, n \rightarrow \infty$. Обозначим $\varepsilon_n = 1 - a_n > 0$, тогда $\varepsilon_n \rightarrow 0, n \rightarrow \infty$. Положим $\nu_n = \min\{m \geq 1 : \xi_{n,m} > a_n\}$. Тогда

$$\begin{aligned} P(M_n \leq u_n(s)) &= P(\xi_{n,\nu_n} \leq u_n(s)) = \\ &= P(\xi_{1,1} \leq u_n(s) | \xi_{1,1} > a_n) = \\ &= 0 \vee \frac{u_n(s) - a_n}{\varepsilon_n} = 0 \vee \left(1 - \frac{1 - u_n(s)}{\varepsilon_n}\right), \end{aligned} \quad (9)$$

где $u_n(s)$ определяются из условия:

$$Eu_n(s)^{\nu_n} = \frac{\varepsilon_n u_n(s)}{1 - (1 - \varepsilon_n)u_n(s)} \rightarrow s, \quad n \rightarrow \infty, \quad (10)$$

поскольку ν_n имеет геометрическое распределение (начиная от единицы) с параметром ε_n . Из (10) следует:

$$1 - u_n(s) \sim \varepsilon_n \frac{1 - s}{s}, \quad n \rightarrow \infty. \quad (11)$$

Подставляя это в (9) и переходя к пределу, получаем $\psi(s) = 0 \vee (2 - 1/s)$ по определению 1. В этом случае, как и в примере 5.5, $\psi(s) < s$ при всех $s \in (0, 1)$. Имеем $\psi(s) = 0$, а значит, $\log_s \psi(s) = +\infty$ при $s \in [0, 1/2]$, $\log_s \psi(s) > 1$ при $s \in (1/2, 1)$, а также $\log_s \psi(s) \rightarrow 1, s \rightarrow 1$. Отсюда $\theta^- = 1, \theta^+ = +\infty$.

Удивительно, что результат не зависит от выбора последовательности $a_n, n \geq 1$.

Что можно сказать в этом случае об экстремальном индексе по определению 2? Из (10) с учетом (11) получаем:

$$\begin{aligned} Eu_n(s)^{\theta \nu_n} &= \frac{\varepsilon_n u_n(s)^\theta}{1 - (1 - \varepsilon_n)u_n(s)^\theta} \rightarrow \\ &\rightarrow \frac{s}{\theta + (1 - \theta)s}, \quad n \rightarrow \infty, \end{aligned}$$

однако экстремальная функция не имеет такого вида; следовательно, экстремального индекса по определению 2 не существует.

Рассмотрим теперь модель со случайными порогом $\zeta_n, n \geq 1$. Пусть $0 < \zeta_n < 1$ п. н.; $\xi_{n,m}, m \geq 1$, не зависят от ζ_n и $\nu_n = \min\{m \geq 1 : \xi_{n,m} > \zeta_n\}$.

Теорема 6.1. Пусть $n(1 - \zeta_n) \xrightarrow{L_1} \zeta > 0, n \rightarrow \infty, E\zeta = 1$. Тогда $\psi(s) = g(f^{-1}(s))$, где $f(t) = E(\zeta/(t + \zeta))$ и $g(t) = E(\zeta - t)_+$.

Доказательство. При условии, что $\zeta_n = x \in (0, 1)$, длина серии ν_n имеет геометрическое распределение с параметром $1 - x$. Отсюда следует, что $(1 - \zeta_n)\nu_n \xrightarrow{d} \eta$ и $\nu_n/n \xrightarrow{d} \eta/\zeta, n \rightarrow \infty$, где η имеет стандартное показательное распределение и не зависит от ζ . Обозначим $f(t) = Ee^{-t\eta/\zeta}$, тогда $f(t) = E(\zeta/(t + \zeta))$.

Пусть $\tau > 0$, тогда

$$\begin{aligned} E\left(1 - \frac{\tau}{n}\right)^{\nu_n} &= E\left(\left(1 - \frac{\tau}{n}\right)^n\right)^{\nu_n/n} \rightarrow \\ &\rightarrow Ee^{-\tau\eta/\zeta} = f(\tau), \quad n \rightarrow \infty. \end{aligned}$$

Так что из $Eu_n(s)^{\nu_n} \rightarrow s$ следует

$$u_n(s) = 1 - \frac{(1 + o(1))f^{-1}(s)}{n}, \quad n \rightarrow \infty.$$

Получаем

$$\begin{aligned} P(M_n \leq u_n(s)) &= P(\xi_{1,1} \leq u_n(s) | \xi_{1,1} > \zeta_n) = \\ &= \frac{P(\zeta_n < \xi_{1,1} \leq u_n(s))}{\varepsilon_n} = \\ &= \frac{E(1 - (1 + o(1))f^{-1}(s)/n - \zeta_n)_+}{\varepsilon_n} \rightarrow \\ &\rightarrow E(\zeta - f^{-1}(s))_+, \quad n \rightarrow \infty. \end{aligned}$$

Напомним удобную для вычислений формулу

$$E(\zeta - t)_+ = \int_t^{+\infty} \bar{F}_\zeta(x) dx, \quad (12)$$

получаемую интегрированием по частям.

Пример 6.2. Пусть ζ равномерно принимает значения $1 - \delta$ и $1 + \delta, 0 < \delta < 1$ (случай $\delta = 0$ сводится к примеру 6.1). Тогда

$$\begin{aligned} f(t) &= \frac{1}{2} \left(\frac{1}{1 + t/(1 - \delta)} + \frac{1}{1 + t/(1 + \delta)} \right) = \\ &= \frac{(1 + t) - \delta^2}{(1 + t)^2 - \delta^2}, \end{aligned}$$

откуда

$$f^{-1}(s) = \frac{1 + \sqrt{1 - 4s(1-s)\delta^2}}{2s} - 1;$$

$$\psi(s) = \frac{1}{2} (1 - \delta - f^{-1}(s))_+ + \frac{1}{2} (1 + \delta - f^{-1}(s))_+.$$

Имеем $\psi(s) = 0$ при $0 < s < f(1 + \delta) = (2 - \delta)/4$, так что $\theta^+ = +\infty$.

В общем случае можно исследовать асимптотическое поведение экстремальной функции при $s \rightarrow 0$ и $s \rightarrow 1$. Обозначим:

$$\theta_0 = \lim_{s \rightarrow 0} \log_s \psi(s); \quad \theta_1 = \lim_{s \rightarrow 1} \log_s \psi(s).$$

Можно утверждать, что $\theta^- \leq \theta_0 \wedge \theta_1$ и $\theta^+ \geq \theta_0 \vee \theta_1$.

Следствие 6.1.

- (1) Если $\bar{F}_\zeta(x) \sim Cx^{-\alpha}$, $x \rightarrow \infty$, $C > 0$, $\alpha > 1$, то $\psi(s) \sim Cs^{\alpha-1}/(\alpha-1)$, $s \rightarrow 0$, и $\theta_0 = \alpha - 1$. Если $\bar{F}_\zeta(x)$ убывает быстрее любой степени, то $\theta_0 = +\infty$.
- (2) Если $E\zeta^{-1} < \infty$, то $\theta_1 = 1/E\zeta^{-1}$. Если $E\zeta^{-1} = \infty$, то $\theta_1 = 0$.

Доказательство.

- (1) Заметим, что

$$f(t) = E \frac{\zeta}{\zeta + t} \sim E \frac{\zeta}{t} = \frac{1}{t}, \quad t \rightarrow \infty.$$

Поэтому $f^{-1}(s) \sim 1/s$, $s \rightarrow 0$. Из $\bar{F}_\zeta(x) \sim Cx^{-\alpha}$, $x \rightarrow \infty$, по формуле (12) следует

$$g(t) \sim \frac{Ct^{-(\alpha-1)}}{\alpha-1}, \quad t \rightarrow \infty.$$

Следовательно,

$$\psi(s) = g(f^{-1}(s)) \sim \frac{Cs^{\alpha-1}}{\alpha-1}, \quad s \rightarrow 0,$$

и $\theta_0 = \alpha - 1$.

Если $\bar{F}_\zeta(x) = o(x^N)$, $x \rightarrow \infty$, то $g(t) = o(t^{N-1})$, $t \rightarrow \infty$, и $\psi(s) = o(s^{N-1})$, $s \rightarrow 0$, для любого $N > 0$, откуда $\theta_0 = +\infty$.

- (2) При $E\zeta^{-1} < \infty$ имеем

$$1 - f(t) = E \left(\frac{t}{\zeta + t} \right) \sim tE\zeta^{-1}, \quad t \rightarrow 0.$$

Отсюда

$$f^{-1}(s) \sim \frac{1-s}{E\zeta^{-1}}, \quad s \rightarrow 1.$$

Кроме того, $1 - g(t) \sim t$, $t \rightarrow 0$. Следовательно,

$$1 - \psi(s) \sim \frac{1-s}{E\zeta^{-1}}, \quad s \rightarrow 0,$$

откуда $\theta_1 = 1/E\zeta^{-1}$. Результат для $E\zeta^{-1} = \infty$ получаем предельным переходом.

По следствию 6.1 в примере 6.2 получаем $\theta_0 = +\infty$; $\theta_1 = 1 - \delta^2 \in (0, 1)$.

Понятно, что если из показателей θ_0 и θ_1 один больше единицы, а другой меньше, то график $\psi(s)$ неизбежно пересекает диагональ. При этом свойство (3) в форме $\psi(s) \geq s$ выполняется при некоторых $s \in (0, 1)$, а при некоторых — нарушается.

7 Заключение

В работе проведено обобщение понятия экстремального индекса стационарной случайной последовательности на схему серий со случайными длинами (двумя определениями). Изучены свойства новых экстремальных индексов. Рассмотрены их различные приложения для моделей информационных сетей и биологических популяций, моделей с копулами и пороговых моделей. Приведены примеры, когда существуют оба экстремальных индекса, только один из них или ни одного. В случаях, когда экстремальный индекс по первому определению не существует, найдены частичные индексы. Таким образом, сделан ряд важных шагов в построении нового математического аппарата, имеющего теоретическое и прикладное значение для описания экстремального поведения различных систем.

Разумеется, исследование новых экстремальных индексов, их свойств и приложений не может исчерпываться рамками одной статьи. Данная работа скорее призвана открыть цикл статей, а возможно, и целое научное направление, к которому могут присоединиться и другие авторы, подобно тому как это происходит в исследованиях классического экстремального индекса.

Литература

1. Лидбеттер М., Линдгрен Г., Ротсен Х. Экстремумы случайных последовательностей и процессов / Пер. с англ. — М.: Мир, 1989. 392 с. (Leadbetter M. R., Lindgren G., Rootzen H. Extremes and related properties of random sequences and processes. — Springer, 1986. 336 p.)
2. Галамбош Я. И. Асимптотическая теория экстремальных порядковых статистик / Пер. с англ. — М.: Наука, 1984. 304 с. (Galambos J. The asymptotic theory of extreme order statistics. — New York, NY, USA: Wiley, 1978. 352 p.)
3. Embrechts P., Klüppelberg C., Mikosch T. Modelling extreme events for insurance and finance. — Springer, 2003. 638 p.
4. De Haan L., Ferreira A. Extreme value theory. An introduction. — Springer, 2006. 420 p.

5. *Новак С. Ю.* Предельные теоремы и оценки скорости сходимости в теории экстремальных значений: Дис. . . . докт. физ.-мат. наук. — СПб.: ПОМИ РАН, 2014. 230 с.
6. *Markovich N. M.* Modeling clusters of extreme values // *Extremes*, 2013. Vol. 17. No. 1. P. 97–125.
7. *Markovich N. M.* Quality assessment of the packet transport of peer-to-peer video traffic in high-speed networks // *Perform. Evaluation*, 2013. Vol. 70. No. 1. P. 28–44.
8. *Avrachenkov K., Markovich N. M., Sreedharan J. K.* Distribution and dependence of extremes in network sampling processes. INRIA Research Report No. 8578, 2014. 25 p. <http://arxiv.org/abs/1408.2529>.
9. *Голдаева А. А.* Тяжелые хвосты, экстремумы и кластеры линейных стохастических рекуррентных последовательностей: Дис. . . . канд. физ.-мат. наук. — М.: МГУ, 2014. 94 с.
10. *Голдаева А. А.* Равномерная оценка экстремального индекса стохастических рекуррентных последовательностей // *Вестник Моск. ун-та. Сер. 1. Математика. Механика*, 2012. № 2. С. 51–55.
11. *Голдаева А. А.* Экстремальные индексы и кластеры в линейных стохастических рекуррентных последовательностях // *Теория вероятностей и ее применения*, 2013. Т. 58. № 4. С. 795–804.
12. *Choi H.* Central limit theory and extremes of random fields. — Chapel Hill: University of North Carolina at Chapel Hill, 2002. PhD Diss.
13. *Ferreira H., Pereira L.* How to compute the extremal index of stationary random fields // *Stat. Probabil. Lett.*, 2008. Vol. 78. P. 1301–1304.
14. *Pereira L.* The asymptotic location of the maximum of a stationary random field // *Stat. Probabil. Lett.*, 2009. Vol. 79. P. 2166–2169.
15. *Савинов Е. А.* Предельная теорема для максимума случайных величин, связанных ИТ-копулами t -распределения Стьюдента // *Теория вероятностей и ее применения*, 2014. Т. 59. № 3. С. 594–602.
16. *Лебедев А. В.* Максимумы активности в случайных сетях в случае тяжелых хвостов // *Проблемы передачи информации*, 2008. Т. 44. № 2. С. 96–100.
17. *Лебедев А. В.* Максимумы активности в безмасштабных случайных сетях с тяжелыми хвостами // *Информатика и её применения*, 2011. Т. 5. Вып. 4. С. 13–16.
18. *Лебедев А. В.* Максимумы активности в некоторых моделях информационных сетей со случайными весами и тяжелыми хвостами // *Проблемы передачи информации*, 2015. Т. 51. № 1. С. 72–81.
19. *Павлов Ю. Л.* О предельных распределениях степеней вершин в условных интернет-графах // *Дискретная математика*, 2009. Т. 21. № 3. С. 14–23.
20. *Лери М. М., Чеплюкова И. А.* Об одной статистической задаче для случайных графов интернет-типа // *Информатика и её применения*, 2011. Т. 5. Вып. 3. С. 34–40.
21. *Райгородский А. М.* Модели случайных графов и их применения // *Труды МФТИ*, 2010. Т. 2. № 4. С. 130–140.
22. *Van der Hofstad R.* Random graphs and complex networks. — Eindhoven University of Technology, 2014. Vol. 1. 328 p. <http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf>.
23. *Сенета Е.* Правильно меняющиеся функции / Пер. с англ. — М.: Наука, 1985. 144 с. (*Seneta E.* Regularly varying functions. — Springer, 1976. 116 p.)
24. *Stam A. J.* Regular variation of the tail of a subordinated probability distribution // *Adv. Appl. Probab.*, 1973. Vol. 5. P. 308–327.
25. *Лебедев А. В.* Общая схема максимумов сумм независимых случайных величин и ее приложения // *Математические заметки*, 2005. Т. 77. № 4. С. 544–550.
26. *Arnold B. C., Villasenor J. A.* The tallest man in the world // *Statistical theory and applications: Papers in honor of Herbert A. David* / Eds. H. N. Nagaraja, P. K. Sen, D. F. Morrison. — New York, NY, USA: Springer-Verlag New York, 1996. P. 81–88.
27. *Pakes A. G.* Extreme order statistics on Galton–Watson trees // *Metrika*, 1998. Vol. 47. No. 1. P. 95–117.
28. *Lebedev A. V.* Maxima of random particles scores in Markov branching processes with continuous time // *Extremes*, 2008. Vol. 11. No. 2. P. 203–216.
29. *Лебедев А. В.* Максимумы случайных признаков частиц в надкритических ветвящихся процессах // *Вестник Моск. ун-та. Сер. 1. Математика. Механика*, 2008. № 5. С. 3–6.
30. *Лебедев А. В.* Многомерные экстремумы случайных признаков частиц в надкритических ветвящихся процессах // *Теория вероятностей и ее применения*, 2012. Т. 57. № 4. С. 788–794.
31. *Лебедев А. В.* Асимптотическое поведение экстремумов случайных признаков частиц в ветвящихся процессах с наследственностью // *Ярославский педагогический вестник. Сер. Физико-математические и естественные науки*, 2010. № 1. С. 7–14.
32. *McNeil A. J., Frey R., Embrechts P.* Quantitative risk management. — Princeton University Press, 2005. 538 p.
33. *Nelsen R.* An introduction to copulas. — Springer, 2006. 276 p.
34. *Chatzis S. P., Demiris Y.* The copula echo state network // *Pattern Recogn.*, 2012. Vol. 45. P. 570–577.
35. *Феллер В.* Введение в теорию вероятностей и ее приложения / Пер. с англ. — М.: Мир, 1984. Т. 2. 752 с. (*Feller W.* An introduction to probability and its applications. — New York, NY, USA: Wiley, 1971. Vol. 2. 668 p.)
36. *Новак С. Ю.* О распределении максимума случайного числа случайных величин // *Теория вероятностей и ее применения*, 1991. Т. 36. № 4. С. 675–681.

Поступила в редакцию 13.02.15

EXTREMAL INDICES IN A SERIES SCHEME AND THEIR APPLICATIONS

A. V. Lebedev

Faculty of Mechanics and Mathematics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation

Abstract: The concept of an extremal index of a stationary random sequence is generalized to a series scheme of identically distributed random variables with random series sizes tending to infinity in probability. The new extremal indices are introduced through two definitions generalizing the basic properties of the classical extremal index. Some useful properties of the new extremal indices are proved. The paper shows how the behavior of aggregate activity maxima on random graphs (in information network models) and the behavior of maxima of random particles scores in branching processes (in biological populations models) can be described in terms of the new extremal indices. New results on models with copulas and threshold models are obtained. The paper shows that the new indices can take different values for one system and the values greater than one.

Keywords: extremal index; series scheme; random graph; information network; branching process; copula

DOI: 10.14357/19922264150305

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (project 14-01-00075).

References

- Leadbetter, M. R., G. Lindgren, and H. Rootzen. 1986. *Extremes and related properties of random sequences and processes*. Springer. 336 p.
- Galambos, J. 1978. *The asymptotic theory of extreme order statistics*. New York, NY: Wiley. 352 p.
- Embrechts, P., C. Klüppelberg, and T. Mikosh. 2003. *Modelling extremal events for insurance and finance*. Springer. 638 p.
- De Haan, L., and A. Ferreira. 2006. *Extreme value theory. An introduction*. Springer. 420 p.
- Novak, S. Yu. 2014. Predelnye teoremy i otsenki skorosti skhodimosti v teorii ekstremal'nykh znacheniy [Limit theorems and convergence rate estimation in the extreme value theory]. St. Petersburg. D.Sc. Diss. 230 p.
- Markovich, N. M. 2013. Modeling clusters of extreme values. *Extremes* 17(1):97–125.
- Markovich, N. M. 2013. Quality assessment of the packet transport of peer-to-peer video traffic in high-speed networks. *Perform. Evaluation* 70(1):28–44.
- Avrachenkov, K., N. M. Markovich, and J. K. Sreedharan. 2014. Distribution and dependence of extremes in network sampling processes. INRIA Research Report No. 8578. 25 p. Available at: <http://arxiv.org/abs/1408.2529> (accessed August 7, 2015).
- Goldaeva, A. A. 2014. Tyazhelye khvosty, ekstremumy i klastery lineynykh stokhasicheskikh rekurrentnykh posledovatel'nostey [Heavy tails, extremes, and clusters of linear stochastic recursive sequences]. Moscow: MGU. PhD Diss. 94 p.
- Goldaeva, A. A. 2012. Uniform estimator of the extremal index of stochastic recurrent sequences. *Moscow Univ. Math. Bull.* 67(2):82–85.
- Goldaeva, A. A. 2013. Extremal indices and clusters in the linear recursive stochastic sequences. *Teoriya Veroyatnostey i ee Primeneniya* [Theory Probab. Appl.] 58(4):689–698.
- Choi, H. 2002. Central limit theory and extremes of random fields. Chapel Hill: University of North Carolina at Chapel Hill. PhD Diss.
- Ferreira, H., and L. Pereira. 2008. How to compute the extremal index of stationary random fields. *Stat. Probabil. Lett.* 78:1301–1304.
- Pereira, L. 2009. The asymptotic location of the maximum of a stationary random field. *Stat. Probabil. Lett.* 79:2166–2169.
- Savinov, E. A. 2014. Predelnaya teorema dlya maksimuma sluchaynykh velichin, svyazannykh IT-kopulami t -raspredeleniya St'yudenta [Limit theorem for the maximum of random variables connected by IT-copulas of Student's distribution] *Teoriya Veroyatnostey i ee Primeneniya* [Theory Probab. Appl.] 59(3):594–602.
- Lebedev, A. V. 2008. Activity maxima in random networks in the heavy tail case. *Problems Information Transmission* 44(2):156–160.
- Lebedev, A. V. 2011. Maksimumy aktivnosti v bezmasshtabnykh sluchaynykh setyakh s tyazhelymi khvostami [Activity maxima in free-scale random networks with heavy tails]. *Informatika i ee Primeneniya — Inform. Appl.* 5(4):13–16.

18. Lebedev, A. V. 2015. Activity maxima in some models of information networks with random weights and heavy tails. *Problems Information Transmission* 51(1):66–74.
19. Pavlov, Yu. L. 2009. On the limit distributions of the vertex degrees of conditional Internet graphs. *Discrete Math. Appl.* 19(4):349–359.
20. Leri, M. M. 2011. Ob odnoy statisticheskoy zadache dlya sluchaynykh grafov internet-tipa [On a statistical problem for random Internet-type graphs]. *Informatika i ee Primeneniya — Inform. Appl.* 5(3):34–40.
21. Raigorodskii, A. M. 2010. Modeli sluchaynykh grafov i ikh primeneniya [Models of random graphs and their applications]. *Trudy MFTI* [MIPT Proceedings] 2(4):130–140.
22. Van der Hofstad, R. 2014. *Random graphs and complex networks*. Eindhoven University of Technology. Vol. 1. 328 p. Available at: <http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf> (accessed August 7, 2015).
23. Seneta, E. 1976. *Regularly varying functions*. Springer. 116 p.
24. Stam, A. J. 1973. Regular variation of the tail of a subordinated probability distribution. *Adv. Appl. Probab.* 5:308–327.
25. Lebedev, A. V. 2005. General scheme of maxima of sums of independent random variables and its applications. *Mathematical Notes* 77(4):503–509.
26. Arnold, B. C., and J. A. Villasenor. 1996. The tallest man in the world. *Statistical theory and applications. Papers in honor of Herbert A. David*. Eds. H. N. Nagaraja, P. K. Sen, and D. F. Morrison. New York, NY: Springer-Verlag New York. 81–88.
27. Pakes, A. G. 1998. Extreme order statistics on Galton–Watson trees. *Metrika* 47(1):95–117.
28. Lebedev, A. V. 2008. Maxima of random particles scores in Markov branching processes with continuous time. *Extremes* 11(2):203–216.
29. Lebedev, A. V. 2008. Maxima of random properties of particles in supercritical branching processes. *Moscow Univ. Math. Bull.* 63(5):175–178.
30. Lebedev, A. V. 2013. Multivariate extremes of random properties of particles in supercritical branching processes. *Teoriya Veroyatnostey i ee Primeneniya* [Theory Probab. Appl.] 57(4):678–683.
31. Lebedev, A. V. 2010. Asimptoticheskoe povedenie ekstremumov sluchaynykh priznakov chastits v vetyashchikh-sya protsessakh s nasledstvennost'yu [The asymptotic behavior of extremes of random particles scores in branching processes with a heredity] *Yaroslavskiy pedagogicheskiy vestnik. Ser. Fiziko-matematicheskie i estestvennyye nauki* [Yaroslavl Pedagogical Bull. Physics and mathematics and natural sciences] 1:7–14.
32. McNeil, A. J., R. Frey, and P. Embrechts. 2005. *Quantitative risk management*. Princeton University Press. 538 p.
33. Nelsen, R. 2006. *An introduction to copulas*. Springer. 276 p.
34. Chatzis, S. P., and Y. Demiris. 2012. The copula echo state network. *Pattern Recogn.* 45:570–577.
35. Feller, W. 1971. *An introduction to probability and its applications*. New York, NY: Wiley. Vol. 2. 668 p.
36. Novak, S. Yu. 1992. On the distribution of the maximum of a random number of random variables. *Teoriya Veroyatnostey i ee Primeneniya* [Theory Probab. Appl.] 36(4):714–721.

Received February 13, 2015

Contributor

Lebedev Alexey V. (b. 1971) — Candidate of Science (PhD) in physics and mathematics, associate professor, Faculty of Mechanics and Mathematics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; avlebed@yandex.ru

ПОРОГОВАЯ СТРАТЕГИЯ ОГРАНИЧЕНИЯ ДОСТУПА К РЕСУРСАМ В СИСТЕМЕ МАССОВОГО ОБСЛУЖИВАНИЯ $M/D/1$ С ФУНКЦИЕЙ ШТРАФОВ ЗА НЕСВОЕВРЕМЕННОЕ ОБСЛУЖИВАНИЕ ЗАЯВОК*

Я. М. Агаларов¹

Аннотация: Рассматривается задача оптимизации управления доступом на множестве динамических пороговых стратегий в системе $M/D/1$, в которой прием заявок в систему прекращается, если число одновременно находящихся в ней заявок превышает пороговое значение, и прием заявок возобновляется, если число заявок станет равно или меньше этого значения. В качестве целевой функции используется среднее значение предельного дохода системы, получаемого в единицу времени в стационарном режиме, при условии, что от каждой принятой на обслуживание заявки получает фиксированную плату, за несвоевременно обслуженную заявку система платит фиксированный штраф, а при отклонении заявки система не получает плату и не платит штраф. Получены оценки для оптимального значения целевой функции и оптимального порогового значения.

Ключевые слова: система массового обслуживания; пороговая стратегия; оптимизация

DOI: 10.14357/19922264150306

1 Введение

Одним из механизмов исключения перегрузок в системах, предназначенных для обработки стохастических потоков заданий, является контроль за перегрузкой и ограничение доступа для входной нагрузки на ресурсы системы. В большинстве существующих алгоритмов ограничения доступа используются пороговые стратегии управления доступом, суть которых заключается в том, что для некоторых параметров, определяющих качество работы сети, устанавливаются пороговые значения, при переходе которых соответствующими параметрами в большую или меньшую сторону по специальному алгоритму осуществляется ограничение или расширение доступа для входной нагрузки к ресурсам [1].

Основные результаты по оценке эффективности и оптимизации пороговых стратегий получены методами теории очередей в виде явных и неявных выражений для расчета стационарных вероятностно-временных характеристик (средней задержки, стационарного распределения вероятностей, интенсивности выходной нагрузки и т.д.) систем массового обслуживания (СМО) различного типа ($M/G/1$, $M_2/G/1$, $MMPP/G/1$, $M^x/G/1$, $G/M/m/K$, $MQ|M|m$ и др.) [2–7]. Задача оптимизации пороговых стратегий для СМО, за исключением, может быть, самых простых, аналитически

не решена, предлагаются либо эвристические решения, либо решение численными методами. В качестве целевой функции, как правило, используются усредненные характеристики системы: средняя задержка заявок в системе, среднее число заявок в системе, интенсивность потерь заявок.

В данной работе исследована задача выбора оптимального порога для системы $M/D/1$, в которой прием заявок в систему прекращается, когда число одновременно находящихся в ней заявок превышает порог, и прием заявок возобновляется, если число заявок опустится до порога или ниже. В качестве целевой функции используется среднее значение предельного дохода системы, получаемого в единицу времени в стационарном режиме, отнесенного к одной поступившей заявке, при условии, что от каждой принятой на обслуживание заявки система получает фиксированную плату, за несвоевременно обслуженную заявку платит фиксированный штраф, а при отклонении заявки система не получает плату и не платит штраф. Аналогичная задача с такой же целевой функцией рассмотрена в работе [6] для системы $G/M/m$. В работе [7] исследована оптимальность пороговой стратегии для системы $M/G/1$ в случае, когда система платит фиксированный штраф в каждую единицу времени ожидания в очереди независимо от длительности пребывания.

* Работа выполнена при поддержке РФФИ (проекты 15-07-03406 и 13-07-00223).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, agglar@yandex.ru

2 Описание системы и постановка задачи

Рассматривается одноканальная СМО типа $M/D/1$, на которую поступает пуассоновский поток однородных заявок с детерминированным временем обслуживания. Вновь поступившая заявка занимает свободный прибор, если таковой есть, иначе становится в очередь в порядке поступления согласно заданной стратегии управления доступом f . Система получает плату $C_0 > 0$, если время выполнения заявки (время пребывания заявки в системе) не превышает заданную величину («дедлайн») t_0 , и платит штраф $C_1 > 0$, если время выполнения заявки превышает t_0 . Если поступившая заявка согласно стратегии управления отклоняется, то система не получает плату и не платит штраф. Величину платы, получаемой системой при обслуживании заявки, за вычетом штрафа будем называть доходом от обслуживания заявки.

Среднее значение дохода от поступления заявки зависит от принятой в системе стратегии управления доступом (f), и его предельное значение определяется в виде предела

$$Q^f = \lim_{T \rightarrow \infty} \sum_{n=1}^{N_{\text{вх}}} \frac{d_n^f}{N_{\text{вх}}(T)}, \quad (1)$$

где d_n^f — доход, получаемый системой при обслуживании n -й в порядке поступления заявки при стратегии f ; $N_{\text{вх}}(T)$ — число заявок, поступивших на вход системы за отрезок времени $[0, T]$.

Пороговые стратегии управления доступом, рассматриваемые в данной статье, составляют класс динамических стационарных стратегий $F_d = \{f^{\bar{k}}\}$, определяемые следующим образом: $f^{\bar{k}} = \{f_i^{k_i}, i \geq 0\}$, где i — состояние системы (число заявок в системе, остающихся в момент завершения обслуживания заявки), $f_i^{k_i} = 1$ при $0 \leq i \leq k_i - 1$, $f_i^{k_i} = 0$ при $k_i \leq i$, $\infty > K > k_i > 0$, — пороговое значение (натуральное число), назначаемое в момент завершения обслуживания заявки. При этом если в момент поступления заявки в систему $f_i^{k_i} = 1$, то заявка ставится в очередь или сразу начинает обслуживаться (если прибор свободен), иначе она отклоняется и покидает систему. В дальнейшем для упрощения обозначений всюду, за исключением некоторых случаев, вместо слов «стратегия $f^{\bar{k}}$ » («политика $f_i^{k_i}$ ») будем писать «стратегия \bar{k} » («политика k_i »), вместо обозначений $f^{\bar{k}}$ и $f_i^{k_i}$ будем использовать соответственно \bar{k} и k_i . Чтобы исключить тривиальный случай, предположим, что $k_0 > 1$.

Как известно (см., например, [8]), процесс обслуживания заявок в данной системе описывается цепью Маркова, где переходы цепи определяются моментами окончания обслуживания, и состояние системы есть число заявок, остающихся в системе в момент ухода обслуженной заявки с прибора. Предположим, что цепь Маркова при любой стратегии $f \in F_d$ является цепью с одним положительным эргодическим классом [8]. Заметим, что цепь Маркова отвечает этому предположению, если, например, для класса стратегий F_d выполняется условие: для любого состояния цепи i существует состояние j такое, что $i \leq k_j - 1$.

Введем обозначения:

$\pi_i^{\bar{k}}, I - 1 \geq i \geq 0$, — стационарное распределение вероятностей цепи при стратегии \bar{k} ($\pi_i^{\bar{k}}$ — вероятность того, что цепь находится в состоянии i), $I = \max\{k_i\}$;

$g^{\bar{k}}$ — средний предельный доход, получаемый за один шаг марковской цепи (за интервал времени между двумя соседними моментами окончания обслуживания заявок) при стратегии \bar{k} ;

$q_i^{\bar{k}}$ — средний доход, получаемый системой в состоянии i (за время пребывания в состоянии i) при стратегии \bar{k} , $i \geq 0$.

Соотношение (1) можно написать в виде:

$$\begin{aligned} Q^f &= \lim_{T \rightarrow \infty} \sum_{n=1}^{N_{\text{вх}}(T)} \frac{d_n^{\bar{k}}}{N_{\text{вх}}(T)} = \\ &= \lim_{T \rightarrow \infty} \frac{N_{\text{вых}}(T)}{N_{\text{вх}}(T)} \sum_{n=1}^{N_{\text{вх}}(T)} \frac{d_n^{\bar{k}}}{N_{\text{вых}}(T)} = \\ &= \lim_{T \rightarrow \infty} \frac{N_{\text{вых}}(T)}{N_{\text{вх}}(T)} \lim_{T \rightarrow \infty} \sum_{n=1}^{N_{\text{вх}}(T)} \frac{d_n^{\bar{k}}}{N_{\text{вых}}(T)}, \end{aligned}$$

где $N_{\text{вых}}(T)$ — число заявок, обслуженных за отрезок времени $[0, T]$. Как следует из теории управляемых марковских процессов [8], предел

$\lim_{N_{\text{вх}} \rightarrow \infty} \sum_{n=1}^{N_{\text{вх}}} d_n^{\bar{k}} / N_{\text{вых}}$ равен среднему доходу за один шаг рассматриваемой цепи Маркова и

$$\lim_{N_{\text{вх}} \rightarrow \infty} \sum_{n=1}^{N_{\text{вх}}} \frac{d_n^{\bar{k}}}{N_{\text{вых}}} = \sum_{i=0}^{K-1} \pi_i^{\bar{k}} q_i^{\bar{k}}.$$

Ставится задача: найти оптимальную стратегию $\bar{k}^* \in F_d$ такую, что

$$\max_{\bar{k} \in F_d} Q^{\bar{k}} = Q^{\bar{k}^*}. \quad (2)$$

Здесь $Q^{\bar{k}} = (1 - \theta^{\bar{k}}) \sum_{i=0}^{I-1} \pi_i^{\bar{k}} q_i^{\bar{k}}$, где $\theta^{\bar{k}}$ — стационарная вероятность того, что система находится в состоянии i таком, что $f_i = 1$.

3 Метод решения

Получим выражения для расчета величины дохода q_i^k . Введем обозначения:

l -заявка — заявка, которая в данный момент имеет номер l в порядке поступления (будем считать, что заявки, находящиеся одновременно в системе, в любой момент пронумерованы числами $1, 2, \dots$ в порядке поступления, начиная с номера 1);

$t_0 \geq \nu > 0$ — время обслуживания заявки;

τ_l — случайная величина длины интервала времени между моментами поступления l -заявки и $(l-1)$ -заявки;

$\infty > \lambda > 0$ — интенсивность поступления заявок в систему;

$\rho = \lambda\nu$ — величина поступающей нагрузки;

$p_{ij}^{k_i}$ — вероятность перехода цепи из состояния i в состояние j при пороге k_i ;

$U_l^i(m, \nu)$ — случайная величина времени пребывания в системе l -заявки (интервал времени с момента поступления до момента окончания обслуживания), поступившей за время нахождения системы в состоянии i , при условии, что за время ν в систему поступило ровно m заявок.

Для $U_l^i(m, \nu) < l \leq m$ справедливо соотношение:

$$U_l^i(m, \nu) = (l+i)\nu - \sum_{j=1}^l \tau_j. \quad (3)$$

Из (3) получаем:

$$U_{l+1}^i(m, \nu) = U_l^i(m, \nu) + \nu - \tau_{l+1}. \quad (4)$$

Обозначим: B_m — событие, состоящее в том, что за время ν поступило ровно m заявок, $m > 0$; $A_{l,m}$ — событие, состоящее в том, что l -заявка обслужится неуспешно при условии B_m , $l = 1, \dots, \min\{m, k_i - i\}$; $E_{j,m}$ — событие, состоящее в том, что за время ν произошло ровно j неуспешных обслуживаний заявок при условии B_m , $j = 0, \dots, \min\{m, k_i - i\}$.

Обозначим $l_i^{\max} = \min\{m, k_i - i\}$. Тогда из (3) и (4) следует

$$E_{l_i^{\max}-l, m} = A_{l+1, m} \bar{A}_{l, m}, \quad m > 0, \quad l_i^{\max} \geq l \geq 0, \quad (5)$$

где $\bar{A}_{0, m} = \Omega$; $A_{m+1, m} = \Omega$;

$$E_{l_i^{\max}, m} = \Omega, \quad E_{j, m} = \bar{\Omega},$$

если $i > t_0/\nu$, $m > 0$, $l_i^{\max} > j \geq 0$.

Величина дохода, получаемого системой при пороге $k_i > 1$ за время нахождения в состоянии $i > 0$ (т. е. за один шаг соответствующей цепи Маркова) при условии B_m определяется как сумма доходов, получаемых при обслуживании допущенных в систему за время ν новых заявок и выражается формулами:

$$q_i^{k_i}(\nu, m; t_0) = \begin{cases} mC_0 - C_1 \sum_{l=0}^m (m-l)P(E_{m-l, m}), & m \leq k_i - i; \\ (k_i - i)C_0 - C_1 \sum_{l=0}^{k_i - i} (k_i - i - l)P(E_{k_i - i - l, m}), & m \geq k_i - i, \\ 0, & 0 < i \leq k_i - 1; \\ & i \geq k_i. \end{cases} \quad (6)$$

Величина дохода, получаемого системой при пороге $k_i = 1$ в состоянии $i = 0$, равна $g^k = q_0^{k_0} = C_0$; величина дохода, получаемого системой при пороге $k_i > 1$ в состоянии $i = 0$, составляет $q_0^{k_0} = q_1^{k_1} + C_0$.

Рассмотрим случай $0 < i \leq t_0/\nu$, $k_i > 1$. Как следует из (3)–(5), выполнение события $E_{l_i^{\max}-l, m}$ эквивалентно совместному выполнению неравенств:

$$\left. \begin{aligned} \sum_{j=1}^l \tau_j &\geq (i+l)\nu - t_0; \\ \sum_{j=1}^{l+1} \tau_j &< (i+l+1)\nu - t_0, \end{aligned} \right\} \quad (7)$$

где $0 \leq \sum_{j=1}^m \tau_j \leq \nu$. Из (5) и (7) следует:

$P(E_{l_i^{\max}-l, m}) = 0$ при $(i+l) > t_0/\nu + 1$ или $(i+l) \leq t_0/\nu - 1$. Отсюда вытекает, что для $0 < i \leq t_0/\nu$, $k_i > 0$ выражения (6) эквивалентны выражениям:

$$q_i^{k_i}(\nu, m; t_0) = \begin{cases} mC_0 - C_1 \sum_{l=[t_0/\nu]^+ - i - 1}^{\min\{m, [t_0/\nu] - i + 1\}} (m-l)P(E_{m-l, m}), & m \leq k_i - i; \\ (k_i - i)C_0 - C_1 \sum_{l=[t_0/\nu]^+ - i - 1}^{\min\{k_i - i, [t_0/\nu] - i + 1\}} (k_i - i - l)P(E_{k_i - i - l, m}), & m \geq k_i - i, \end{cases} \quad (8)$$

где $[t_0/\nu]$ — целая часть числа t_0/ν ;

$$\left[\frac{t_0}{\nu} \right]^+ = \begin{cases} \frac{t_0}{\nu}, & \text{если } \left[\frac{t_0}{\nu} \right] = \frac{t_0}{\nu}; \\ \left[\frac{t_0}{\nu} \right] + 1, & \text{если } \left[\frac{t_0}{\nu} \right] < \frac{t_0}{\nu}. \end{cases}$$

Как следует из (5), при $i > t_0/\nu$ для любого $m > 0$ имеют место равенства

$$P(E_{l_i^{\max}, m}) = 1;$$

$$q_i^{k_i}(\nu, m; t_0) = \begin{cases} m(C_0 - C_1), & 0 < m \leq k_i - i; \\ (k_i - i)(C_0 - C_1), & m \geq k_i - i. \end{cases}$$

Величина полного дохода, получаемого системой в состоянии i при пороге $k_i > 1$ и $i \leq t_0/\nu$, определяется с учетом (8) по формуле полной вероятности:

$$\begin{aligned} q_i^{k_i} &= \sum_{m=0}^{\infty} \frac{\rho^m}{m!} e^{-\rho} q_i^{k_i}(\nu, m; t_0) = \\ &= \sum_{m=1}^{k_i-i-1} \frac{\rho^m}{m!} e^{-\rho} \left[mC_0 - \right. \\ &\quad \left. - C_1 \sum_{l=m_i^+-1}^{\min\{m, m_i+1\}} (m-l)P(E_{m-l, m}) \right] + \\ &\quad + \sum_{m=k_i-i}^{\infty} \frac{\rho^m}{m!} e^{-\rho} \left[(k_i-i)C_0 - \right. \\ &\quad \left. - C_1 \sum_{l=m_i^+-1}^{\min\{k_i-i, m_i+1\}} (k_i-i-l)P(E_{k_i-i-l, m}) \right] = \\ &= (k_i-i)C_0 + C_0 e^{-\rho} \sum_{m=1}^{k_i-i-1} (m-k_i+i) \frac{\rho^m}{m!} - \\ &\quad - C_1 \sum_{m=1}^{k_i-i-1} \frac{\rho^m}{m!} e^{-\rho} \sum_{l=m_i^+-1}^{\min\{m, m_i+1\}} (m-l)P(E_{m-l, m}) - \\ &\quad - C_1 e^{-\rho} \sum_{l=m_i^+-1}^{\min\{k_i-i, m_i+1\}} (k_i-i-l) \sum_{m=k-i}^{\infty} \frac{\rho^m}{m!} P(E_{k_i-i-l, m}), \quad (9) \end{aligned}$$

где $m_i^+ = [t_0/\nu]^+ - i$; $m_i = [t_0/\nu] - i$.

Для краткости изложения в дальнейшем обозначим $\nu_{i,l}^+ = \min\{1, (i+l+1) - t_0/\nu\}$; $\nu_{i,l}^- = \max\{0, (i+l) - t_0/\nu\}$. Отметим еще раз (см. (5)), что для i, l таких, что $\nu_{i,l}^+ \leq 0$, $P(E_{j,m}) = 0$, $j \leq \min\{k_i - i, m\}$.

Вероятность $P(E_{m-l, m})$ для случая $i \leq t_0/\nu$, $k > 1$ вычисляется по формуле

$$\begin{aligned} P(E_{m-l, m}) &= \\ &= \frac{1}{\nu^m} \int_{\max\{0, (i+l)\nu - t_0\}}^{\min\{\nu, (i+l+1)\nu - t_0\}} \int_{x_1}^{\min\{\nu, (i+l+1)\nu - t_0\}} \frac{m!}{(l-1)!} \times \\ &\quad \times \frac{x_1^{l-1}(\nu - x_2)^{m-l-1}}{(m-l-1)!} dx_2 dx_1 = \\ &= \int_{\nu_{i,l}^-}^{\nu_{i,l}^+} \int_x^{\nu_{i,l}^+} \frac{m!}{(l-1)!} \frac{x^{l-1}(1-y)^{m-l-1}}{(m-l-1)!} dy dx \quad (10) \end{aligned}$$

при $k_i - i \geq m \geq l > 0$,

$$\begin{aligned} P(E_{m, m}) &= \frac{m}{\nu^m} \int_{\max\{0, i\nu - t_0\}}^{\min\{\nu, (i+1)\nu - t_0\}} (\nu - x_1)^{m-1} dx_1 = \\ &= m \int_0^{\nu_{i,0}^+} (1-x)^{m-1} dx = 1 - (1 - \nu_{i,0}^+)^m \end{aligned}$$

при $l = 0$, $k_i - i \geq m$. Формулы (10) следуют из (7) и из того, что распределение величин $S_l = \sum_{j=1}^l \tau_j$, $l = 1, \dots, m$, совпадает с распределением порядковых статистик из выборки n , взятой из равномерного распределения на $[0, \nu]$ [3].

Взяв в (10) интеграл по y и подставив формулу

$$\begin{aligned} x^{l-1}(1-x)^{m-l} &= \\ &= \sum_{j=0}^{m-l} (-1)^j \frac{(m-l)!}{j!(m-l-j)!} x^{l+j-1}, \quad (11) \end{aligned}$$

получим

$$\begin{aligned} P(E_{m-l, m}) &= \frac{m!}{(l-1)!(m-l-1)!} \times \\ &\quad \times \int_{\nu_{i,l}^-}^{\nu_{i,l}^+} x^{l-1} \frac{(1-x)^{m-l} - (1-\nu_{i,l}^+)^{m-l}}{m-l} dx = \\ &= \frac{m!}{(l-1)!(m-l)!} \left[\sum_{j=0}^{m-l} (-1)^j \frac{(m-l)!}{j!(m-l-j)!(l+j)} \times \right. \\ &\quad \left. \times \Delta_{i,l}(l+j) - \frac{\Delta_{i,l}(l)}{l} \left(\frac{1-\nu_{i,l}^+}{l} \right)^{m-l} \right], \quad (12) \end{aligned}$$

где $\Delta_{i,l}(j) = (\nu_{i,l}^+)^j - (\nu_{i,l}^-)^j$, $0 < l \leq m \leq k_i - i$. Подставив (10) в сумму $\sum_{m=k_i-i}^{\infty} (\rho^m/m!)P(E_{k_i-i-l,m})$, получим

$$\begin{aligned} & \sum_{m=k_i-i}^{\infty} \frac{\rho^m}{m!} \int_{\nu_{i,l}^-}^{\nu_{i,l}^+} \frac{x^{l-1}}{(l-1)!} \int_x^{\nu_{i,l}^+} \frac{m!(1-y)^{m-l-1}}{(m-l-1)!} dy dx = \int_{\nu_{i,l}^-}^{\nu_{i,l}^+} \frac{\rho^{l+1}x^{l-1}}{(l-1)!} \int_x^{\nu_{i,l}^+} \left[\sum_{m=l+1}^{\infty} \frac{\rho^{m-l-1}(1-y)^{m-l-1}}{(m-l-1)!} - \right. \\ & \left. - \sum_{m=l+1}^{k_i-i-1} \frac{\rho^{m-l-1}(1-y)^{m-l-1}}{(m-l-1)!} \right] dy dx = \int_{\nu_{i,l}^-}^{\nu_{i,l}^+} \frac{\rho^{l+1}x^{l-1}}{(l-1)!} \int_x^{\nu_{i,l}^+} \left[e^{\rho(1-y)} - \sum_{m=l+1}^{k_i-i-1} \frac{\rho^{m-l-1}(1-y)^{m-l-1}}{(m-l-1)!} \right] dy dx = \\ & = \int_{\nu_{i,l}^-}^{\nu_{i,l}^+} \frac{\rho^l x^{l-1}}{(l-1)!} \left[e^{\rho(1-x)} - e^{\rho(1-\nu_{i,l}^+)} \right] dx - \int_{\nu_{i,l}^-}^{\nu_{i,l}^+} \sum_{m=l+1}^{k_i-i} \frac{\rho^m x^{l-1}}{(l-1)!(m-l)!} \left[(1-x)^{m-l} - (1-\nu_{i,l}^+)^{m-l} \right] dx = \\ & = -e^{\rho(1-\nu_{i,l}^+)} \frac{\rho^l}{(l-1)!} \int_{\nu_{i,l}^-}^{\nu_{i,l}^+} x^{l-1} dx + e^{\rho} \frac{\rho^l}{(l-1)!} \int_{\nu_{i,l}^-}^{\nu_{i,l}^+} x^{l-1} e^{-\rho x} dx + \\ & + \sum_{m=l+1}^{k_i-i} (1-\nu_{i,l}^+)^{m-l} \frac{\rho^m}{(m-l)!(l-1)!} \int_{\nu_{i,l}^-}^{\nu_{i,l}^+} x^{l-1} dx - \frac{\rho^m}{(m-l)!(l-1)!} \sum_{m=l+1}^{k_i-i} \int_{\nu_{i,l}^-}^{\nu_{i,l}^+} x^{l-1} (1-x)^{m-l} dx. \quad (13) \end{aligned}$$

Используя в (13) формулы (11) и

$$\int x^{l-1} e^{-\rho x} dx = -\frac{e^{-\rho x}}{\rho} \left[x^{l-1} + (l-1)! \sum_{j=1}^{l-2} \frac{x^{l-1-j}}{j! \rho^j} + \frac{(l-1)!}{\rho^{l-1}} \right]$$

и обозначения

$$\left. \begin{aligned} \Delta_{i,l}(j) &= (\nu_{i,l}^+)^j - (\nu_{i,l}^-)^j; \\ D_{i,l}(j) &= e^{-\rho \nu_{i,l}^+} (\nu_{i,l}^+)^j - e^{-\rho \nu_{i,l}^-} (\nu_{i,l}^-)^j, \end{aligned} \right\} \quad (14)$$

после интегрирования получим:

$$\begin{aligned} \sum_{m=k_i-i}^{\infty} \frac{\rho^m}{m!} [1 - (1 - \nu_{i,0}^+)^m] &= \sum_{m=0}^{\infty} \frac{\rho^m}{m!} [1 - (1 - \nu_{i,0}^+)^m] - \\ &- \sum_{m=0}^{k_i-i-1} \frac{\rho^m}{m!} [1 - (1 - \nu_{i,0}^+)^m] = e^{\rho} (1 - e^{-\rho \nu_{i,0}^+}) + \sum_{m=0}^{k_i-i-1} \frac{\rho^m}{m!} [(1 - \nu_{i,0}^+)^m - 1] \end{aligned}$$

при $l = 0$;

$$\begin{aligned} \sum_{m=k_i-i}^{\infty} \frac{\rho^m}{m!} P(E_{m-l,m}) &= -e^{\rho(1-\nu_{i,l}^+)} \frac{\rho^l}{l!} \Delta_{i,l}(l) - e^{\rho} \rho^{l-1} \sum_{j=0}^{l-2} \frac{D_{i,l}(l-1-j)}{(l-j)! \rho^j} - e^{\rho} D_{i,l}(0) + \\ &+ \frac{\rho^m}{l!} \sum_{m=l+1}^{k_i-i} (1-\nu_{i,l}^+)^{m-l} \frac{\Delta_{i,l}(l)}{(m-l)!} - \frac{\rho^m}{(l-1)!} \sum_{m=l+1}^{k_i-i} \sum_{j=0}^{m-l} (-1)^j \frac{\Delta_{i,l}(l+j)}{j!(m-l-j)!(l+j)} \quad (15) \end{aligned}$$

при $0 < l \leq \min\{k_i - i, m\}$.

Из (9), подставив (15), получим для $i \leq t_0/\nu$, $k_i > 1$:

$$q_i^{k_i} = \begin{cases} \left((k_i - i)C_0 + C_0 e^{-\lambda\nu} \sum_{m=0}^{k_i-i-1} (m - k_i + i) \frac{\rho^m}{m!} - C_1 \sum_{m=0}^{k_i-i-1} \frac{\rho^m}{m!} e^{-\lambda\nu} \sum_{l=m_i^+-1}^{\min\{m, m_i+1\}} (m-l) P(E_{m-l, m}) - \right. \\ \left. - C_1 \sum_{l=m_i^+-1}^{\min\{k_i-i, m_i+1\}} (k_i - l + i) \left[e^{-\rho\nu_{i,l}^+} \frac{\rho^{l+1}}{l!} \Delta_{i,l}(l) - \frac{\rho^l}{(l-1)!} D_{i,l}(l-1) - \right. \right. \\ \left. - \rho^{l-1} \sum_{j=1}^{l-2} \frac{D_{i,l}(l-1-j)}{j! \rho^j (l-j)} - \rho D_{i,l}(0) - e^{-\rho} (1 - \nu_{i,l}^+) \frac{1 - (1 - \nu_{i,l}^+)^{k_i-i-l-1}}{\nu_{i,l}^+} \frac{\rho^{l+1}}{(l-1)!} \Delta_{i,l}(l) + \right. \\ \left. \left. + e^{-\rho} \frac{\rho^{l+1}}{(l-1)!} \sum_{m=l+1}^{k_i-i} \sum_{j=0}^{m-l} (-1)^j \frac{(m-l)!}{j!(m-l-j)!(m-j)} \Delta_{i,l}(m-j) \right] \right) \text{ при } 0 < i \leq k_i - 1, \\ 0 \text{ при } i \geq k_i; \end{cases} \quad (16)$$

$$q_0^{k_0} = C_0 + q_1^{k_1}, \quad (17)$$

где $P(E_{m-l, m})$ вычисляется по формуле (12); $\Delta_{i,l}(j)$ и $D_{i,l}(j)$ определены в (14).

Для $i > t_0/\nu$, преобразовав (9), получим

$$q_i^{k_i} = \begin{cases} \sum_{m=0}^{\infty} \frac{\rho^m}{m!} e^{-\rho} q_i^{k_i}(\nu, m; t_0) = (C_0 - C_1) \sum_{m=0}^{k_i-i-1} m \frac{\rho^m}{m!} e^{-\rho} + \\ + (k_i - i) (C_0 - C_1) \sum_{m=k_i-i}^{\infty} \frac{\rho^m}{m!} e^{-\rho} = (C_0 - C_1) \left[k_i - i - \sum_{m=0}^{k_i-i-1} (k_i - i - m) \frac{\rho^m}{m!} e^{-\rho} \right] \text{ при } i \leq k_i - 1; \\ 0 \text{ при } i \geq k_i. \end{cases}$$

Перейдем к рассмотрению параметров $\bar{\pi}_j^{\bar{k}}$, $j = 0, \dots, I-1$, $\bar{k} \geq 1$, $k_0 > 1$, и $\theta^{\bar{k}}$.

Для вероятностей переходов $p_{ij}^{\bar{k}}$ справедливы формулы:

$$\left. \begin{aligned} p_{ij}^{\bar{k}} &= \begin{cases} r_{j-i+1}, & i \leq j \leq k_i - 2; \\ 1 - \sum_{l=0}^{k_i-i-1} r_l, & j = k_i - 1, \end{cases} \text{ если } 1 \leq i \leq k_i - 1; \\ p_{0j}^{\bar{k}} &= \begin{cases} r_j, & i \leq j \leq k_0 - 2; \\ 1 - \sum_{l=0}^{k_0-2} r_l, & j = k_0 - 1; \end{cases} \\ p_{i, i-1}^{\bar{k}} &= \begin{cases} 1, & k_i \leq i, i \geq 1; \\ r_0, & k_i > i \geq 1; \end{cases} \\ p_{ij}^{\bar{k}} &= 0 \text{ при } i > k_i - 1, j > k_i - 1, j < i - 1, \end{aligned} \right\} \quad (18)$$

где

$$r_l = \frac{\rho^l}{l!} e^{-\rho} \text{ при } 0 \leq l \leq k_i - 2.$$

Для рассматриваемой цепи Маркова при стратегии \bar{k} стационарное распределение вероятностей является единственным решением системы уравнений [8]:

$$\left. \begin{aligned} \bar{\pi}_j^{\bar{k}} &= \sum_{\substack{i: 0 \leq i \leq j, \\ k_i > j}} \bar{\pi}_i^{\bar{k}} p_{ij}^{\bar{k}} + \bar{\pi}_{j+1}^{\bar{k}} p_{j+1, j}^{\bar{k}}, \quad j = 0, \dots, I-2; \\ \sum_{j=0}^{I-1} \bar{\pi}_j^{\bar{k}} &= 1. \end{aligned} \right\} \quad (19)$$

Из (18) и (19) следуют рекуррентные формулы для расчета стационарного распределения при стратегии \bar{k} :

$$\pi_j^{\bar{k}} = \pi_0^{\bar{k}} R_j, \quad j = 1, \dots, I-1, \quad (20)$$

где

$$\pi_0^{\bar{k}} = \left(\sum_{i=0}^{I-1} R_i \right); \quad R_0 = 1; \quad R_1 = \frac{1-r_0}{r_0};$$

$$R_{i+1} = \frac{1}{p_{i+1,i}^{\bar{k}}} \left(R_i - r_i - \sum_{\substack{0 \leq j \leq i, \\ k_j > i}} R_j r_{i-j+1} \right),$$

$$i = 1, \dots, I-2.$$

Для вычисления $\theta^{\bar{k}}$ рассмотрим вложенную цепь Маркова, которую образует последовательность $\{v_n, n \geq 0\}$, где v_n — число заявок в системе в момент $t_n, t_{n+1} = t_n + v, t_0 = 0$ [9]. Для состояний $i, j = 0, \dots, I-2$ вероятности переходов совпадают с $p_{ij}^{\bar{k}}$ в (18), т. е. уравнения равновесия для состояний $i = 0, \dots, I-2$ (первые $I-1$ уравнений) совпадают с соответствующими уравнениями в (19). Обозначим стационарные вероятности состояний новой вложенной цепи через $\theta_j^{\bar{k}}, j = 0, \dots, I$. Добавив уравнения $\sum_{j=0}^I \theta_j^{\bar{k}} = 1$ (уравнение полной вероятности) и $\rho = 1 - \theta_0^{\bar{k}} + \rho \theta^{\bar{k}}$ (уравнение баланса нагрузки), где $\theta^{\bar{k}} = \sum_{j=k_j} \theta_j^{\bar{k}}$, получим $I+1$ независимых уравнений равновесия для вычисления $I+1$ неизвестных переменных $\theta_j^{\bar{k}}, j = 0, \dots, I$. Аналогично (20) получим:

$$\theta_0^{\bar{k}} = \frac{1}{\rho \sum_{j: j \neq k_j} R_j + 1};$$

$$\theta_j^{\bar{k}} = \theta_0^{\bar{k}} R_j, \quad j = 1, \dots, I-1;$$

$$\theta_I^{\bar{k}} = 1 - \frac{1 - \theta_0^{\bar{k}}}{\rho} - \sum_{j: j = k_j, j \neq I} \theta_j^{\bar{k}}, \quad \theta_j^{\bar{k}} = 1 - \frac{1 - \theta_0^{\bar{k}}}{\rho}.$$

Упростим задачу (2), сократив класс допустимых пороговых стратегий, а именно: будем ниже рассматривать множество F_c статических пороговых стратегий $\bar{k} = \{k_i, i = 0, \dots, I-1\}, k_i = k > 1, i = 0, \dots, I-1, I = k$, и вместо обозначения \bar{k} будем писать k . Для исследования решения задачи (2) воспользуемся результатами теории управления марковскими процессами [8]. Пусть набор $V_j^k, j \geq 0$, — решение следующей системы уравнений:

$$\left. \begin{aligned} q_i^k + \sum_{j=i-1}^{k-1} r_{j-i+1} V_j^k &= V_i^k + g^k, \\ k-1 \geq i \geq 1; \\ q_0^k + \sum_{j=0}^{k-1} r_j V_j^k &= V_0^k + g^k; \\ V_{i-1}^k &= V_i^k + g^k, \quad i > k-1. \end{aligned} \right\} \quad (21)$$

Отметим, что каждое уравнение в (21) является линейной комбинацией остальных $k-1$ уравнений, и поэтому один из параметров V_j^k является произвольным. Действительно, если V_j^k — решение системы, то $V_j^k + const$ также являются решением. Если, например, каждое i -е уравнение умножить на π_i для $i = 1, \dots, k-1$ и просуммировать, то получим уравнение для $i = 0$.

Утверждение 1. Пусть для порогов $k > 0$ и $k+1 > 0$ справедлива система неравенств

$$q_i^k - q_i^{k+1} \leq -g^k \left(1 - \sum_{j=i-1}^{k-1} r_{j-i+1} \right),$$

$$0 \leq i \leq k. \quad (22)$$

Тогда если хотя бы одно неравенство в (22) является строгим, то справедливо неравенство $g^k < g^{k+1}$, иначе $g^k = g^{k+1}$.

Доказательство. Фиксируем порог k и решим систему уравнений (21). Из (21), проведя преобразования, получим

$$q_i^k + \sum_{j=i-1}^{k-2} r_{j-i+1} V_j^k + \left(1 - \sum_{j=i-1}^{k-2} r_{j-i+1} \right) V_{k-1}^k =$$

$$= V_i^k + g^k, \quad k-1 \geq i \geq 1;$$

$$g^k - q_i^k + \sum_{j=i-1}^{k-2} r_{j-i+1} (V_{k-1}^k - V_j^k) + (V_i^k - V_{k-1}^k) = 0.$$

Положив $V_{k-1}^k = 0$, получим

$$V_{k-2}^k = \frac{g^k - q_{k-1}^k}{r_0};$$

$$V_{i-1}^k = \frac{1}{r_0} \left(g^k - q_{i-1}^k + (1-r_1)V_i^k - \sum_{j=i+1}^{k-2} r_{j-i+1} V_j^k \right), \quad 1 \leq i \leq k-2.$$

Из (21) также получим $V_{i-1}^k = V_i^k + g^k$, т. е.

$$V_i^k = V_k^k = -g^k, \quad i \leq k \quad (23)$$

(так как $V_{k-1}^k = 0$). Далее воспользуемся теоремой 2.2 в [10]. Согласно этой теореме, если для стратегий $f > 0$ и $f' > 0$ выполняются неравенства

$$q_i^f + \sum_j p_{ij}^f V_j^f \leq q_i^{f'} + \sum_j p_{ij}^{f'} V_j^{f'} \quad (24)$$

и хотя бы одно неравенство в (24) является строгим, то справедливо неравенство $g^f < g^{f'}$, иначе $g^f = g^{f'}$.

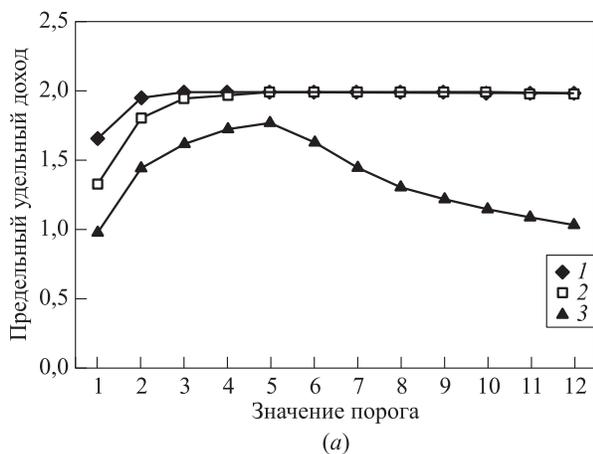
Рассмотрим неравенства (24) для порогов k и $k + 1$:

$$\begin{aligned} V_i^k + g^k &= q_i^k + \sum_{j=i-1}^{k-2} r_{j-i+1} V_j^k + \\ &+ \left(1 - \sum_{j=i-1}^{k-2} r_{j-i+1}\right) V_{k-1}^k \leq q_i^{k+1} + \sum_{j=i-1}^{k-1} r_{j-i+1} V_j^k + \\ &+ \left(1 - \sum_{j=i-1}^{k-1} r_{j-i+1}\right) V_k^k, \quad 0 \leq i \leq k. \end{aligned}$$

Сократив одинаковые слагаемые в левой и правой части неравенств и заменив $V_{k-1}^k = 0$ и $V_k^k = -g^k$, получим, что неравенства (24) для порогов k и $k + 1$ эквивалентны неравенствам (22).

Пусть $1 \leq k \leq [t_0/\nu]$. При $k \leq [t_0/\nu]$ выполняются равенства $P(E_{i,l,m}) = 1$, $l = 1, \dots, k - i$ (так как $i + l \leq k$). Тогда при $k \leq [t_0/\nu]$ из (10) (или из (16) и (17)) следует:

$$q_i^k = \begin{cases} C_0 \left[k - i - \sum_{m=0}^{k-i-1} (k-i-m) \frac{\rho^m}{m!} e^{-\rho} \right] & \text{при } i = 1, \dots, k - 1, k > 1; \\ 0 & \text{при } i \geq k; \end{cases}$$



$$q_0^k = \begin{cases} q_1^k + C_0 & \text{при } k > 1; \\ C_0 & \text{при } k = 1. \end{cases}$$

Используя (24), получим:

$$\begin{aligned} q_i^k - q_i^{k+1} &= C_0 \left(-1 + \sum_{j=0}^{k-i} r_j \right), \\ 0 \leq i \leq k &\leq \left[\frac{t_0}{\nu} \right] - 1. \end{aligned}$$

Подставив правую часть последнего равенства в (23), получим неравенства

$$C_0 \left(-1 + \sum_{j=0}^{k-i} r_j \right) \leq -g^k \left(1 - \sum_{j=0}^{k-i} r_j \right), \quad 0 \leq i \leq k,$$

которые эквивалентны неравенству $C_0 \geq g^k$, справедливому для любых $k \geq 1$. Так как $g^1 = C_0$, то согласно утверждению 1 для всех $1 \leq k \leq [t_0/\nu] - 1$ выполняются равенства $g^k = g^{k+1} = C_0$. Так как $\theta_k^k > \theta_{k+1}^{k+1}$, то из последних неравенств и (3) следует, что для оптимального порога k' выполняется неравенство $k' \geq [t_0/\nu]$. Для максимального значения целевой функции получаем оценку снизу:

$$\max_{k \in F_c} Q^k \geq C_0 (1 - \theta_{k_0}^{k_0}),$$

где $k_0 = [t_0/\nu]$, $\theta_{k_0}^{k_0} = 1 - (1 - \theta_0^{k_0})/\rho$, $\theta_0^{k_0} = 1 / (\rho \sum_{j=0}^{k_0-1} R_j + 1)$, $R_j, j = 0, \dots, k_0 - 1$, вычисляются по формулам (20).

Для поиска оптимального порогового значения на F_c предлагается метод простого перебора: последовательно для $k = k_0, k_0 + 1, \dots$ вычислить значение целевой функции Q^k , пока не выполнится условие $Q^k > Q^{k+1}$.

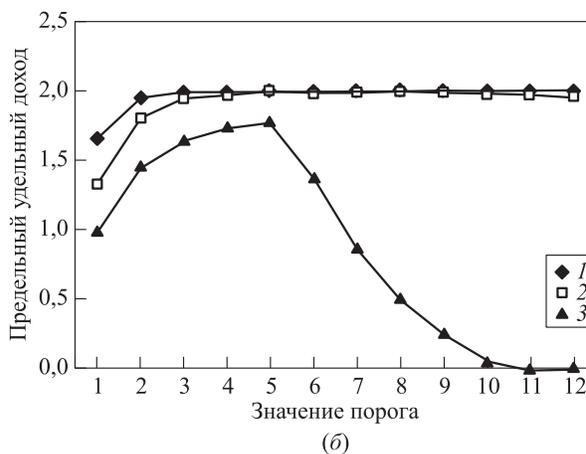


Рис. 1 Зависимость предельного дохода от порогового значения ($\nu = 1; t_0 = 5,1; C_0 = 2$) при $C_1 = 1,2$ (а) и 3 (б): 1 — $\lambda = 0,2$; 2 — $0,5$; 3 — $\lambda = 1$

4 Примеры

В качестве примеров рассмотрены задачи оптимизации статической пороговой стратегии управления доступом в СМО типа $M/D/1$ со следующими вариантами параметров ($\nu = 1; t_0 = 5,1; C_0 = 2$): (1) $\lambda = 0,2; C_1 = 1,2$ (рис. 1, *a*, кривая 1); (2) $\lambda = 0,5; C_1 = 1,2$ (см. рис. 1, *a*, кривая 2); (3) $\lambda = 1; C_1 = 1,2$ (см. рис. 1, *a*, кривая 3); (4) $\lambda = 0,2; C_1 = 3$ (см. рис. 1, *b*, кривая 1); (5) $\lambda = 0,5; C_1 = 3$ (см. рис. 1, *b*, кривая 2); (6) $\lambda = 1; C_1 = 3$ (см. рис. 1, *b*, кривая 3).

На рис. 1 приведены графики зависимости предельного дохода относительно одной поступившей заявки без учета длины интервала времени между соседними моментами поступления.

На рис. 2 приведены графики зависимости предельного дохода в единицу времени с учетом длины интервала времени между моментами поступления для СМО типа $M/D/1$ с параметрами ($\nu = 1; t_0 = 5,1; C_0 = 2; C_1 = 3$): (1) $\lambda = 0,2$ (кривая 1); (2) $\lambda = 0,5$ (кривая 2); (3) $\lambda = 1$ (кривая 3).

Во всех рассмотренных выше примерах при $\lambda = 1$ значение оптимального порога составляет $k_0 = 5$, в остальных случаях $k_0 = 6$.

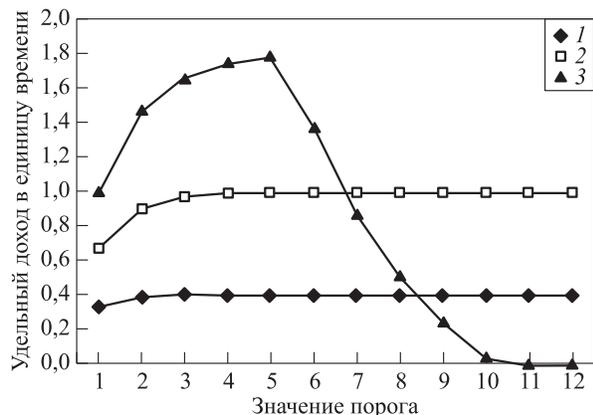


Рис. 2 Зависимость предельного дохода в единицу времени от значения порога ($\nu = 1; t_0 = 5,1; C_0 = 2; C_1 = 3$): 1 — $\lambda = 0,2$; 2 — $0,5$; 3 — $\lambda = 1$

5 Заключение

Исследование задачи оптимизации динамической пороговой стратегии ограничения доступа в СМО $M/D/1$, проведенное в данной работе, приводит к следующим выводам:

1. Рассмотренная выше задача оптимизации (2) является типичной задачей управления цепями Маркова.

2. Доказательство утверждения о том, что оптимальное решение задачи (2) единственное и принадлежит F_c , остается открытым (в отличие от постановки задачи в [7]).
3. Оптимальное решение \bar{k}' задачи (2) удовлетворяет условию $k'_i \geq [t_0/\nu], i = 1, \dots, I - 1$.
4. Численные эксперименты подтверждают гипотезу [6] о том, что целевая функция в задаче (2) является унимодальной выпуклой функцией.

Результаты работы могут быть применены при исследовании систем, в качестве модели которых может быть использована СМО $M/D/1$ (например, фрагментов компьютерных сетей) с целью повышения их эффективности.

Литература

1. Floyd S., Jacobson V. Random early detection gateways for congestion avoidance // IEEE/ACM Trans. Network., 1993. Vol. 1. No. 4. P. 397–413.
2. Welzl M. Network congestion control. — New York, NY, USA: Wiley, 2005. 282 p.
3. Агаларов Я. М., Соколов И. А. Динамическая стратегия распределения буферной памяти АТМ коммутатора // Информационные технологии и вычислительные системы, 2008. № 3, С. 14–21.
4. Hong Y., Huang C., Yan J. A comparative study of SIP overload control algorithms network and traffic engineering in emerging distributed computing applications // IGI Global, 2012. 20 p. <http://arxiv.org/ftp/arxiv/papers/1210/1210.1505.pdf>.
5. Печинкин А. В., Разумчик П. В. Стационарные характеристики системы $M2|G|1|r$ с гистерезисной политикой управления интенсивностью входящего потока // Информационные процессы, 2013. Т. 3. № 3. С. 125–140.
6. Коновалов М. Г. Об одной задаче оптимального управления нагрузкой на сервер // Информатика и её применения, 2013. Т. 7. Вып. 4. С. 34–43.
7. Гришунина Ю. Б. Оптимальное управление очередью в системе $M|G|1|\infty$ с возможностью ограничения приема заявок // Автоматика и телемеханика, 2015. № 3. С. 79–93.
8. Карлин С. Основы теории случайных процессов / Пер. с англ. — М.: Мир, 1971. 536 с. (Karlin S. A first course in stochastic processes. — New York and London: Academic Press, 1968. 502 p.)
9. Бочаров П. П., Печинкин А. В. Теория массового обслуживания. — М.: РУДН, 1995. 529 с.
10. Майн Х., Осаки С. Марковские процессы принятия решений / Пер. с англ. — М.: Наука, 1977. 176 с. (Mine H., Osaki S. Markovian decision processes. — New York, NY, USA: American Elsevier Publishing Co., 1970. 142 p.)

Поступила в редакцию 21.05.15

THE THRESHOLD STRATEGY FOR RESTRICTING ACCESS IN THE $M/D/1$ QUEUEING SYSTEM WITH PENALTY FUNCTION FOR LATE SERVICE

Ya. M. Agalarov

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The paper considers the problem of optimizing the access control on a set of dynamic threshold strategies in an $M/D/1$ system. If the number of concurrent requests in a system is more than the threshold, then the system stops accepting requests. If the number of requests is less or equal to this value, then the system resumes accepting requests. As a target function, the average value of the marginal revenue obtained per time unit in the stationary mode is used. It is assumed that the system receives a fixed fee for each accepted request and pays a fixed penalty for each overdue service of a request. The system does not receive a fee and does not pay a penalty for each rejected request. Estimates of the optimal value of the target function and the optimal threshold value are obtained.

Keywords: queueing system; threshold strategy; optimization

DOI: 10.14357/19922264150306

Acknowledgments

The work was supported by the Russian Foundation for Basic Research (projects 15-07-03406 and 13-07-00223).

References

1. Floyd, S., and V. Jacobson. 1993. Random early detection gateways for congestion avoidance. *IEEE ACM Trans. Network.* 1(4):397–413.
2. Welzl, M. 2005. *Network congestion control*. New York, NY: Wiley. 282 p.
3. Agalarov, Ya. M., and I. A. Sokolov. 2008. Dinamicheskaya strategiya raspredeleniya bufernoy pamyati ATM kommutatora [Dynamic allocation strategy of buffer memory in ATM commutator]. *Informatsionnye Tekhnologii i Vychislitel'nye Sistemy* [Information Technology and Computer Systems] 3:14–21.
4. Hong, Y., C. Huang, and J. Yan. 2012. A comparative study of SIP overload control algorithms network and traffic engineering in emerging distributed computing applications. *IGI Global*. 20 p. Available at: <http://arxiv.org/ftp/arxiv/papers/1210/1210.1505.pdf> (accessed June 24, 2015).
5. Pechinkin, A. V., and R. V. Razumchik. 2013. Statsionarnye kharakteristiki sistemy $M2|G|1|r$ s gisterezisnoy politikoy upravleniya intensivnost'yu vkhodyashchego potoka [Stationary characteristics of $M2|G|1|r$ queue with hysteric control policy of arrival rate]. *Informatsionnye Protsessy* [Information Processes] 3(3):125–140.
6. Konovalov, M. G. 2013. Ob odnoy zadache optimal'nogo upravleniya nagruzkoy na server [About one task of overload control]. *Informatika i ee Primeneniya — Inform. Appl.* 7(4):34–43.
7. Grishunina, Ju. B. 2015. Optimal'noe upravlenie ochered'yu v sisteme $M|G|1|\infty$ s vozmozhnost'yu ogranicheniya priema zayavok [Optimal control of queue in the $M|G|1|\infty$ system with possibility of customer admission restriction]. *Avtomatika i Telemekhanika* [Automation and Remote Control] 3:79–93.
8. Karlin, S. 1968. *A first course in stochastic processes*. New York and London: Academic Press. 502 p.
9. Bocharov, P. P., and A. V. Pechinkin. 1995. *Teoriya massovogo obsluzhivaniya* [Queueing theory]. Moscow: RUDN. 529 p.
10. Mine, H., and S. Osaki. 1970. *Markovian decision processes*. New York, NY: American Elsevier Publishing Co. 142 p.

Received May 21, 2015

Contributor

Agalarov Yaver M. (b. 1952) — Candidate of Science (PhD) in technology, associate professor; leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; agglar@yandex.ru

ПОЖАР НА КОНФИГУРАЦИОННОМ ГРАФЕ СО СЛУЧАЙНЫМИ ПЕРЕХОДАМИ ОГНЯ ПО РЕБРАМ*

М. М. Лери¹

Аннотация: Рассматривается случайный процесс распространения огня по ребрам двух видов конфигурационных графов со случайными степенями вершин. Степени вершин имеют либо степенное распределение, либо распределение Пуассона. Процесс распространения огня происходит в случайной среде, в которой вероятности перехода огня по ребрам графа имеют стандартное равномерное распределение. Оценены оптимальные значения параметров распределений степеней вершин, обеспечивающие максимальную сохранность вершин графа при двух типах возгорания: целенаправленном поджоге, когда пожар начинается с вершины, имеющей наибольшую степень, и случайном возгорании равновероятно выбранной вершины. В условиях одинаковых объемов рассматриваемых графов проведен сравнительный анализ степенной модели и модели с пуассоновским распределением с точки зрения числа уцелевших в пожаре вершин при разных условиях возгорания.

Ключевые слова: конфигурационные графы; степенное распределение; распределение Пуассона; устойчивость; модель лесного пожара

DOI: 10.14357/19922264150307

1 Введение

Одним из актуальных направлений в теории случайных графов в последние годы стали исследования их устойчивости к разрушающим воздействиям, которые передаются по ребрам графов [1–4]. Такое воздействие может рассматриваться как распространение огня во время лесного пожара (см., например, [5–7]), а также при изучении банковских кризисов [8]. В [9–11] были построены модели на конфигурационных графах, впервые предложенных в [12] и широко используемых для изучения структуры и динамики сложных сетей коммуникаций, например сети Интернет и систем мобильной связи [3, 4]. Наблюдения за реальными сетями (см., например, [13, 14]) показали, что в таких моделях степени вершин графов можно считать независимыми одинаково распределенными случайными величинами. Кроме того, число вершин степени k при больших значениях k пропорционально $k^{-(\tau+1)}$, где $\tau > 0$. Это означает, что распределение случайной величины ξ , равной степени произвольной вершины, можно задать следующим образом:

$$\mathbb{P}\{\xi \geq k\} = h(k)k^{-\tau}, \quad (1)$$

где $k = 1, 2, \dots$, а $h(k)$ — медленно меняющаяся функция. Такое распределение означает, что граф не имеет изолированных вершин. Изучение свойств конфигурационных графов с распределением степеней вершин (1) показало [14], что при

$k \rightarrow \infty$ предельное поведение важнейших числовых характеристик не зависит от выбора функции $h(k)$. Поэтому при исследовании динамики случайных графов с растущим числом вершин можно ограничиться простейшим случаем $h(k) \equiv const$.

2 Конфигурационные графы со случайными степенями вершин

Рассмотрим конфигурационные графы с N вершинами, степени которых $\xi_1, \xi_2, \dots, \xi_N$ являются независимыми одинаково распределенными случайными величинами с общим распределением. В настоящей работе рассматриваются два варианта распределений степеней вершин графа. Во-первых, это степенное распределение, заданное равенством

$$\mathbb{P}\{\xi \geq k\} = k^{-\tau}, \quad k = 1, 2, \dots, \tau > 1; \quad (2)$$

во-вторых, распределение Пуассона с единичным сдвигом:

$$\mathbb{P}\{\xi = k + 1\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots, \lambda > 0. \quad (3)$$

При построении графа сначала определяются степени вершин $1, 2, \dots, N$ как реализации случайной величины ξ , имеющей одно из двух рассматриваемых распределений (2) или (3). Будем считать

* Работа выполнена при поддержке РФФИ (проект 13-01-00009).

¹ Институт прикладных математических исследований КарНЦ РАН, leri@krc.karelia.ru

(следуя [14]), что из каждой вершины $i, i = 1, \dots, N$, выходит ξ_i так называемых «полуребер», т. е. ребер графа, для которых смежные вершины еще не определены. Все полуребра считаются различными (пронумерованными). В любом графе сумма степеней вершин должна быть четной, поэтому, если величина $\eta_N = \xi_1 + \dots + \xi_N$ оказывается нечетной, к равновероятно выбранной вершине добавляется еще одно полуребро. В [14] отмечается, что введение такого вспомогательного полуребра не влияет на поведение графа при больших значениях N . Далее попарными равновероятными соединениями полуребер образуются ребра графа, завершая его построение. Нетрудно заметить, что такие графы могут содержать кратные ребра и петли.

Структура описанной конструкции конфигурационного графа и ее асимптотическое поведение в случае, когда распределение степеней вершин имеет вид (2), подробно изучены в [14], где считалось, что параметр τ распределения (2) принадлежит интервалу (1, 2). Этот выбор интервала объясняется тем, что именно такие значения τ типичны для сложных сетей коммуникаций, к числу которых относится сеть Интернет.

3 Исследование устойчивости: модель лесного пожара

Моделирование «лесного пожара» в настоящей работе производилось следующим образом. Будем рассматривать вершины конфигурационного графа как деревья, растущие на некоторой ограниченной территории. Ребра графа означают возможные пути перехода «огня» от одной вершины к другой. В [9–11] для графов со степенным распределением

степеней вершин (2) была решена задача нахождения оптимальных значений параметра τ , зависящих от объема графа N и от заданных вероятностей перехода огня по ребру p , где $0 < p \leq 1$, при которых число уцелевших при пожаре вершин максимально. С целью согласования топологии конфигурационного графа с моделью «лесного пожара» введем вспомогательные графы следующего вида. Будем считать, что вершины графа находятся в узлах квадратной целочисленной решетки, как показано на рис. 1 для некоторых значений средней степени внутренней вершины (обозначим ее m).

В настоящей работе рассматривается решетка размера 100×100 , чтобы число вершин было достаточно большим. Таким образом, общее число «деревьев» на некоторой ограниченной «территории» не превосходит 10 000. Ребра графа соединяют только ближайшие вершины, т. е. находящиеся на расстоянии 1 друг от друга по горизонтали и вертикали и на расстоянии $\sqrt{2}$ по диагонали. Следовательно, степень любой вершины, находящейся внутри решетки, будет не более 8. Использование вспомогательных графов обусловлено необходимостью выявления зависимости между параметрами распределений степеней вершин (τ для (2) и λ для (3)) и числом исходных вершин конфигурационного графа N . Для установления соответствия вспомогательного решетчатого графа описанному ранее конфигурационному графу были вычислены средние значения $m = m(N)$ степеней вершин при различных значениях $N \leq 10\,000$ и различных вариантах размещения N вершин на решетке [11]. Число вершин конфигурационного графа N в данном случае равно числу вершин у соответствующего ему графа на решетке. Нетрудно проверить, что математическое ожидание распределения (2) равно

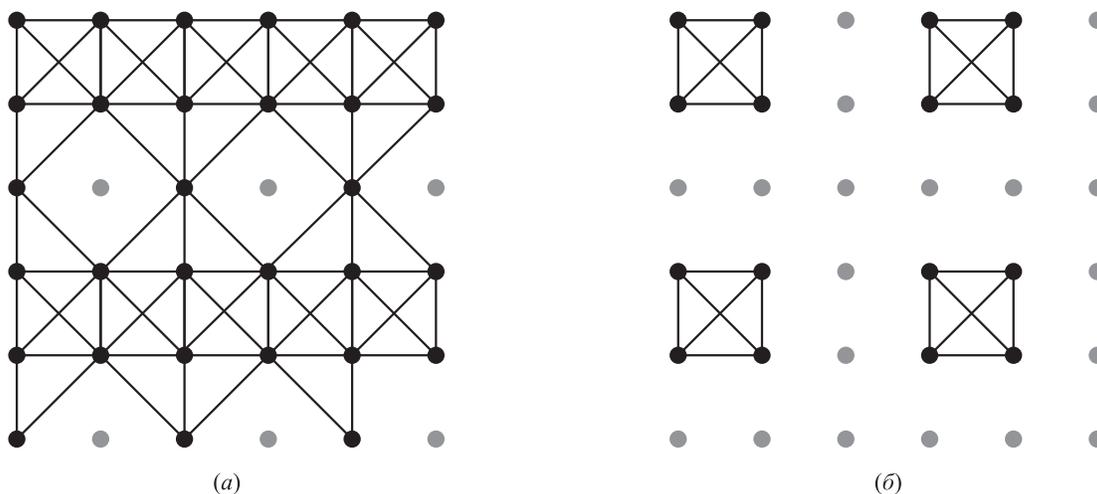


Рис. 1 Фрагменты топологий вспомогательного графа при $m = 6,53$ (а) и 3 (б)

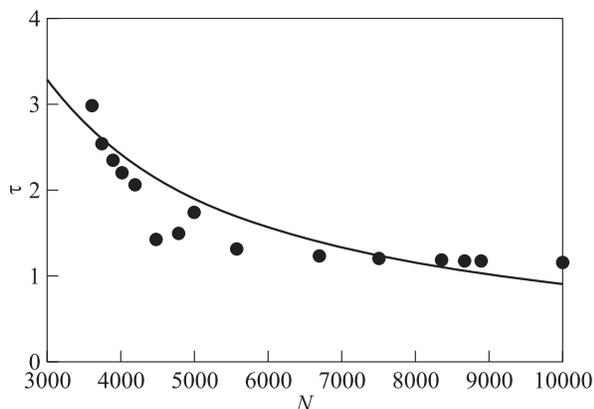


Рис. 2 Регрессионная зависимость τ от N

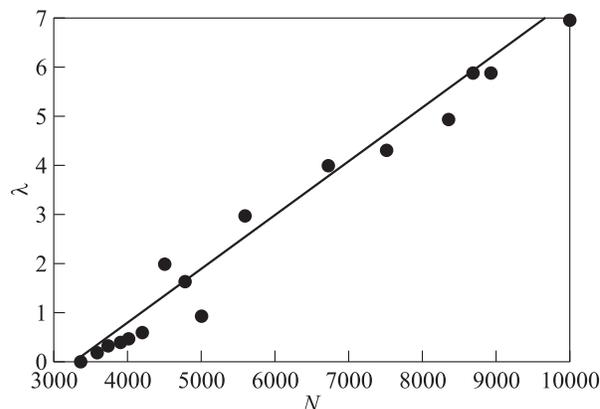


Рис. 3 Регрессионная зависимость λ от N

$\zeta(\tau)$, где $\zeta(\tau)$ — дзета-функция Римана. Принимая во внимание тот факт, что равенство $m(N) = \zeta(\tau)$ должно выполняться хотя бы приблизительно, была найдена регрессионная зависимость τ от N (рис. 2) следующего вида (с коэффициентом детерминации $R^2 = 0,87$):

$$\tau = 16\,813N^{-1,067}. \quad (4)$$

Кроме того, поскольку математическое ожидание распределения (3) равно $\lambda + 1$, на основе равенства $m(N) = \lambda + 1$ (рис. 3) была найдена зависимость параметра λ от N :

$$\lambda = 0,0011N - 3,61, \quad (5)$$

где $N \geq 3350$ и $R^2 = 0,98$.

Уравнения (4) и (5) позволяют находить значения параметров τ и λ распределений степеней вершин конфигурационного графа, наиболее близко соответствующие распределению N вершин на квадратной решетке. Отличие конфигурационного графа от решетчатого состоит еще и в том, что как распределение (2), так и распределение (3) допускают сколь угодно большие степени вершин, тогда как на решетке степень любой вершины не превосходит 8. Однако это отличие можно считать не очень существенным по двум причинам. Во-первых, вероятности больших степеней чрезвычайно малы и такие вершины встречаются довольно редко. Во-вторых, кратные ребра, соединяющие две вершины конфигурационного графа, существенно увеличивают степени этих вершин, в то время как в решетчатом графе между двумя соседними вершинами существует только одно ребро, означающее возможность перехода огня непосредственно от одной вершины к другой. Все вышесказанное позволяет предположить, что соответствие между двумя видами графов можно использовать для оценки значений параметров τ и λ распределений (2) и (3) с помощью полученных регрессионных

уравнений (4) и (5) соответственно в зависимости от исходного числа вершин N . В этом случае параметр распределения степеней вершин определяет не только точный вид распределения, но и в какой-то мере характеризует топологию рассматриваемого конфигурационного графа.

Установленная связь между параметрами τ и λ распределений (2) и (3) и числом вершин на решетке позволяет найти зависимости числа N исходных вершин конфигурационных графов от параметров распределений степеней вершин. Для распределения (2) была получена следующая модель [9]:

$$N = [9256\tau^{-1,05}], \quad (6)$$

а для распределения (3):

$$N = [907,5\lambda + 2509,4]. \quad (7)$$

Коэффициенты детерминации уравнений (6) и (7) равны 0,97 и 0,98 соответственно.

Далее рассматривался процесс распространения «огня» по ребрам конфигурационных графов двух типов: степенного и Пуассона. Предполагались две возможности «начала пожара» — случайное возгорание, при котором первой «загорается» равновероятно выбранная вершина графа, и целенаправленный поджог вершины, имеющей максимальную степень. При возгорании первой вершины огонь переходит по ребрам на смежные вершины с вероятностью $0 < p \leq 1$, т. е. с вероятностью $1 - p$ ребро становится негоряемым и огонь на смежную вершину не переходит. В настоящей работе, в отличие от предшествующих [9–11], процесс распространения огня происходит в так называемой случайной среде. Это означает, что вероятность p не является фиксированной величиной, она равномерно распределена на интервале $(0, 1]$ и своя для каждого ребра. Цель работы состояла в том, чтобы для

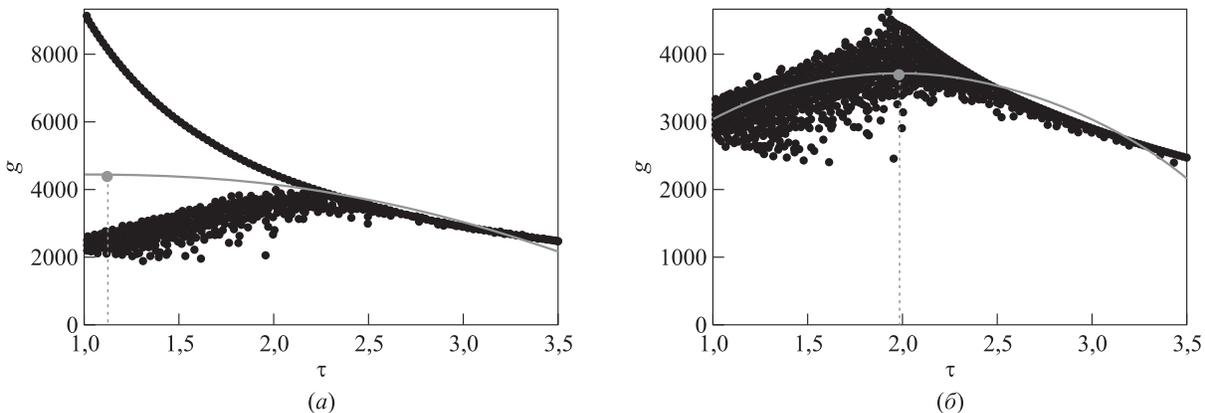


Рис. 4 Зависимость числа несгоревших вершин g от параметра τ : (а) при случайном возгорании, $0,2 \leq p \leq 1$; (б) при направленном поджоге, $0 < p \leq 1$

каждого типа графа (степенного и Пуассона) выяснить, какой из двух вариантов возгорания (случайное или целенаправленное) и при каких значениях параметров τ и λ обеспечивает наибольшее число уцелевших при пожаре вершин.

Имитационные эксперименты проводились на графах, задаваемых параметрами распределений степеней вершин $1 < \tau \leq 3,5$ и $0,3 \leq \lambda \leq 3$, объемы которых вычислялись с помощью уравнений (6) и (7). Для построения графов был использован алгоритм, описанный в [15], и датчик псевдослучайных чисел Mersenne twister [16]. Были построены уравнения зависимости числа сохранившихся после пожара вершин g от параметра распределения степеней вершин τ .

В случае случайного возгорания была получена следующая модель ($R^2 = 0,99$):

$$g = [2705,42\tau - 679,52\tau^2 + 1040,43] , \quad (8)$$

а при целенаправленном поджоге ($R^2 = 0,92$):

$$g = [917\tau - 408,43\tau^2 + 3955,53] . \quad (9)$$

Рисунок 4 отражает зависимости числа сохранившихся после пожара вершин g от параметра распределения степеней вершин τ .

На рис. 4, а видно, что при случайном возгорании облако регрессии распадается на 2 части. Это происходит вследствие того, что при одних и тех же значениях параметра распределения степеней вершин τ существует некоторая доля графов, которые почти не сгорают, т.е. после пожара в них остается более 98% вершин. Такая ситуация возникает, когда возгорание начинается с вершины, являющейся частью очень маленькой компоненты связности (не более 5 вершин), и, следовательно, огонь не достигнет остальных вершин графа.

Построенные уравнения (8) и (9) позволяют найти оптимальные значения параметра τ , обес-

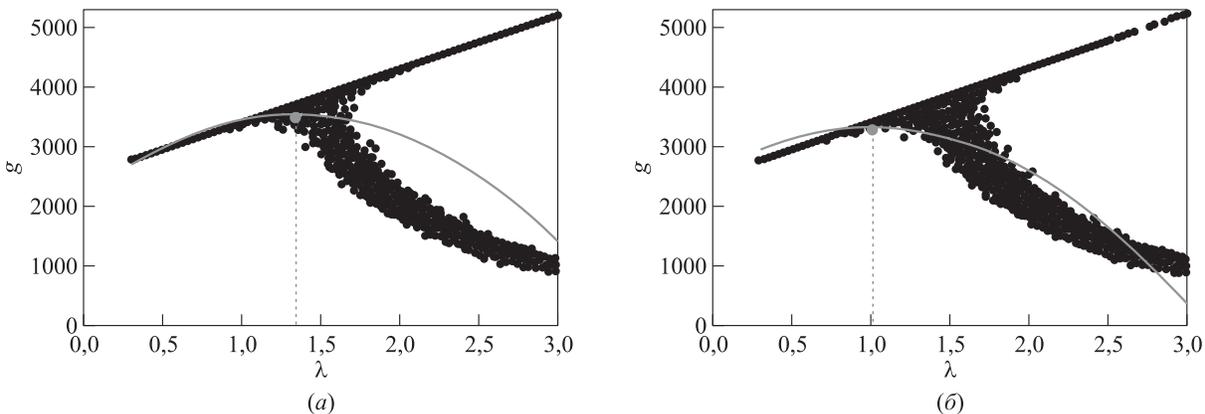


Рис. 5 Зависимость числа несгоревших вершин g от параметра λ , $0 < p \leq 1$: (а) при случайном возгорании; (б) при направленном поджоге

печивающие максимальную сохранность вершин. Так, в случае направленного поджога оно оказалось равным 1,99, а при случайном возгорании — 1,12. Следует отметить тот факт, что было введено ограничение снизу вероятности перехода огня по ребру значением 0,2, которое было обусловлено тем, что при меньших значениях распад облака регрессии увеличивается и становится невозможным нахождение оптимального значения параметра τ в пределах интервала его изменения.

Аналогичные эксперименты были проведены на конфигурационных графах с пуассоновским распределением степеней вершин (5). Были получены следующие модели зависимости числа уцелевших вершин g от параметра λ : при случайном возгорании (рис. 5, а)

$$g = [1510,75\lambda - 747,81\lambda^2 + 2552,62] , \quad (10)$$

где $R^2 = 0,97$, и при целенаправленном поджоге (рис. 5, б)

$$g = [2089,81\lambda - 779,04\lambda^2 + 2150] , \quad (11)$$

где $R^2 = 0,92$.

Из (10) и (11) нетрудно получить, что оптимальное значение параметра λ , обеспечивающее максимальную сохранность вершин графа при пожаре в случае направленного поджога равно 1,01, а при случайном возгорании — 1,34.

4 Исследование устойчивости: сравнение топологий

Далее рассматривалась задача сравнения двух моделей конфигурационных графов: со степенным распределением (2) и с распределением Пуассона (3). Имитационные эксперименты проводились

на графах одинаковых объемов $3000 \leq N \leq 10\,000$ с разными топологиями взаимного расположения вершин и ребер, зависящими от заданного распределения степеней вершин. Параметры τ и λ распределений (2) и (3) находились из уравнений (4) и (5). Шаг изменения объема графа N был равен 100. С целью сравнения топологий находились величины $V = V(N)$, равные разности между числом оставшихся после пожара вершин в модели (2) и в модели (3). На рис. 6, а приведены значения V в зависимости от N при случайном возгорании, а на рис. 6, б — при целенаправленном поджоге.

Проведенное исследование показало, что как при случайном возгорании, так и в случае направленного поджога степенной закон распределения степеней вершин (2) обеспечивает лучшую сохранность вершин во время пожара по сравнению с законом распределения Пуассона (3). Полученные в настоящей работе результаты для целенаправленного поджога отличаются от результатов сравнения топологий, приведенных в [17], где рассматривалась модель «лесного пожара» при заданной фиксированной вероятности p перехода огня по ребрам. В [17] было показано, что при некоторых значениях p и N пуассоновский закон распределения степеней вершин графа обеспечит их лучшую сохранность во время пожара по сравнению со степенным законом.

Автор выражает благодарность профессору Ю. Л. Павлову за помощь в постановке задачи и обсуждение полученных результатов.

Литература

1. Cohen R., Erez K., Ben-Avraham D., Havlin S. Resilience of the Internet to random breakdowns // Phys. Rev. Lett., 2000. Vol. 85. P. 4626–4628.

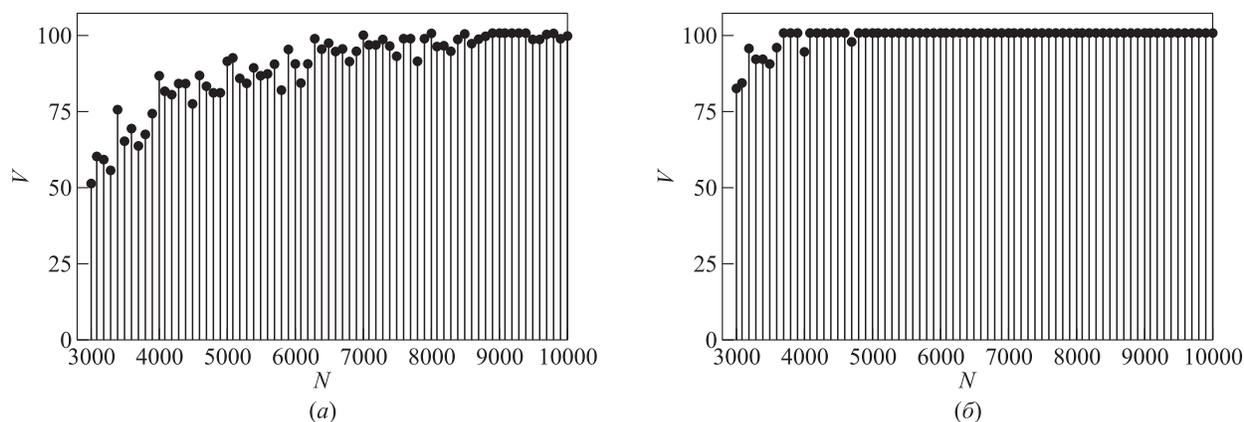


Рис. 6 Зависимость величины V от объема графа N , $0 < p \leq 1$: (а) при случайном возгорании; (б) при направленном поджоге

2. *Bollobas B., Riordan O.* Robustness and vulnerability of scale-free random graphs // *Internet Math.*, 2004. Vol. 1. No. 1. P. 1–35.
3. *Durrett R.* *Random graph dynamics.* — Cambridge: Cambridge University Press, 2007. 221 p.
4. *Hofstad R.* *Random graphs and complex networks.* — Eindhoven: Eindhoven University of Technology, 2011. 363 p.
5. *Drossel B., Schwabl F.* Self-organized critical forest-fire model // *Phys. Rev. Lett.*, 1992. Vol. 69. P. 1629–1632.
6. *Bertoin J.* Burning cars in a parking lot // *Commun. Math. Phys.*, 2011. Vol. 306. P. 261–290.
7. *Bertoin J.* Fires on trees // *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, 2012. Vol. 48. No. 4. P. 909–921.
8. *Arinaminparty N., Kapadia S., May R.* Size and complexity model financial systems // *Proc. Natl. Acad. Sci. USA*, 2012. Vol. 109. P. 18338–18343.
9. *Лери М. М., Павлов Ю. Л.* Лесной пожар на случайном графе со сгораемыми ребрами // *Ученые записки ПетрГУ. Сер. Естественные и технические науки*, 2013. № 2(131). С. 96–99.
10. *Leri M., Pavlov Yu.* Power-law graphs robustness and forest fires // *Computer Data Analysis and Modeling: Theoretical and Applied Stochastics: 10th Conference (International) Proceedings.* — Minsk, 2013. Vol. 1. P. 74–77.
11. *Leri M., Pavlov Yu.* Power-law random graphs' robustness: Link saving and forest fire model // *Austrian J. Stat.*, 2014. Vol. 43. No. 4. P. 229–236.
12. *Bollobas B. A.* A probabilistic proof of an asymptotic formula for the number of labelled regular graphs // *Eur. J. Combust.*, 1980. Vol. 1. P. 311–316.
13. *Faloutsos C., Faloutsos P., Faloutsos M.* On power-law relationships of the Internet topology // *Comp. Comm. Rev.*, 1999. Vol. 29. P. 251–262.
14. *Reittu H., Norros I.* On the power-law random graph model of massive data networks // *Perform. Evaluation*, 2004. Vol. 55. P. 3–23.
15. *Tangmunarunkit H., Govindan R., Jamin S., Shenker S., Willinger W.* Network topology generators: Degree-based vs. structural // *SIGCOMM'02 Proceedings.* — Pittsburgh, USA, 2002. P. 147–159.
16. *Matsumoto M., Nishimura T.* Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator // *ACM Trans. Model. Comput. Simul.*, 1998. Vol. 8. No. 1. P. 3–30.
17. *Leri M., Pavlov Yu.* Forest fire models on configuration random graphs // *3rd Russian–Finnish Symposium on Discrete Mathematics: Extended Abstracts.* — Petrozavodsk: Karelian Research Centre RAS, 2014. P. 68–70.

Поступила в редакцию 30.03.15

FOREST FIRE ON A CONFIGURATION GRAPH WITH RANDOM FIRE PROPAGATION

M. M. Leri

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences, 11 Pushkin-kaya Str., Petrozavodsk 185910, Russian Federation

Abstract: The paper considers a random process of fire propagation over links of two types of configuration graphs with random node degrees. Node degrees follow either a power law or the Poisson distribution. The process takes place in a random environment where the probabilities of fire propagation follow the standard uniform distribution. The optimal values of the node degree distribution parameters that ensure maximum node survival in case of a fire were estimated. The results were obtained for two cases of fire start: targeted start — when a fire starts from the node with the highest degree and random ignition — when a fire starts from an equiprobably chosen node. A comparative analysis of two graph models (power law and Poisson) in terms of the number of nodes remained after the fire was performed.

Keywords: configuration graphs; power-law distribution; Poisson distribution; robustness; forest fire model

DOI: 10.14357/19922264150307

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (project 13-01-00009).

References

1. Cohen, R., K. Erez, D. Ben-Avraham, and S. Havlin. 2000. Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.* 85:4626–4628.
2. Bollobas, B., and O. Riordan. 2004. Robustness and vulnerability of scale-free random graphs. *Internet Math.* 1(1):1–35.
3. Durrett, R. 2007. *Random graph dynamics*. Cambridge: Cambridge University Press. 221 p.
4. Hofstad, R. 2011. *Random graphs and complex networks*. Eindhoven: Eindhoven University of Technology. 363 p.
5. Drossel, B., and F. Schwabl. 1992. Self-organized critical forest-fire model. *Phys. Rev. Lett.* 69:1629–1632.
6. Bertoin, J. 2011. Burning cars in a parking lot. *Commun. Math. Phys.* 306:261–290.
7. Bertoin, J. 2012. Fires on trees. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques* 48(4):909–921.
8. Arinaminparty, N., S. Kapadia, and R. May. 2012. Size and complexity model financial systems. *Proc. Natl. Acad. Sci. USA* 109:18338–18343.
9. Leri, M. M., and Yu. L. Pavlov. 2013. Lesnoy pozhar na sluchaynom grafe so sgoraemymi rebrami [Forest fire on random graph with inflammable links]. *Uchenye Zapiski PetrGU. Estestvennye i tekhnicheskie nauki ser.* [Proceedings of Petrozavodsk State University. Natural and engineering sciences ser.] 2(131):96–99.
10. Leri, M., and Yu. Pavlov. 2013. Power-law graphs robustness and forest fires. *10th Conference (International) "Computer Data Analysis and Modeling: Theoretical and Applied Stochastics" Proceedings*, Minsk. 1:74–77.
11. Leri, M., and Yu. Pavlov. 2014. Power-law random graphs' robustness: Link saving and forest fire model. *Austrian J. Stat.* 43(4):229–236.
12. Bollobas, B. A. 1980. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Combust.* 1:311–316.
13. Faloutsos, C., P. Faloutsos, and M. Faloutsos. 1999. On power-law relationships of the internet topology. *Comp. Comm. Rev.* 29:251–262.
14. Reittu, H., and I. Norros. 2004. On the power-law random graph model of massive data networks. *Perform. Evaluation* 55:3–23.
15. Tangmunarunkit, H., R. Govindan, S. Jamin, S. Shenker, and W. Willinger. 2002. Network topology generators: Degree-based vs. structural. *SIGCOMM'02 Proceedings*. Pittsburgh, USA. 147–159.
16. Matsumoto, M., and T. Nishimura. 1998. Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Trans. Modeling Comput. Simul.* 8(1):3–30.
17. Leri, M., and Yu. Pavlov. 2014. Forest fire models on configuration random graphs. *3rd Russian Finnish Symposium on Discrete Mathematics: Extended Abstracts*. Petrozavodsk. 68–70.

Received March 30, 2015

Contributor

Leri Marina M. (b. 1969) — Candidate of Science (PhD) in technology, scientist, Institute of Applied Mathematical Research of Karelian Research Centre, Russian Academy of Sciences, 11 Pushkinskaya Str., Petrozavodsk 185910, Russian Federation; leri@krc.karelia.ru

АСИМПТОТИЧЕСКИЕ РАЗЛОЖЕНИЯ ВЫСОКОГО ПОРЯДКА ДЛЯ НЕСМЕЩЕННЫХ ОЦЕНОК И ИХ ДИСПЕРСИЙ В МОДЕЛИ ОДНОПАРАМЕТРИЧЕСКОГО ЭКСПОНЕНЦИАЛЬНОГО СЕМЕЙСТВА*

В. В. Чичагов¹

Аннотация: Рассмотрена модель повторной выборки объема n из распределения, принадлежащего естественному однопараметрическому экспоненциальному семейству. При неограниченном возрастании объема выборки изучено предельное поведение несмещенной оценки с равномерно минимальной дисперсией (НОРМД) заданной параметрической функции и НОРМД дисперсии этой оценки. Получены асимптотические разложения высокого порядка как для функций, определяющих несмещенные оценки, так и для дисперсий этих оценок. Результаты представлены как в канонической, так и в теап-параметризации.

Ключевые слова: естественное экспоненциальное семейство; несмещенная оценка; асимптотическое разложение

DOI: 10.14357/19922264150308

1 Введение

Несмещенные оценки широко применяются при решении задач теории надежности и статистического контроля качества, а также при построении моделей массового обслуживания. Проблема построения точечных несмещенных оценок в значительной степени себя исчерпала. Воиновым и Никулиным в [1] описаны основные методы построения несмещенных оценок, а также представлены в виде таблиц несмещенные оценки и функции от них для многих вероятностных законов.

Однако изучение предельного поведения несмещенных оценок и их характеристик применительно к естественным экспоненциальным семействам распределений далеко не закончено. Portnoy в [2] установил асимптотическую нормальность и асимптотическую эффективность несмещенной оценки заданной параметрической функции $g[\theta]$ с каноническим параметром $\theta \in \mathbb{R}^p$ при $g'[\theta] \neq 0$.

Случай $g'[\theta] = 0$ рассмотрели Lopez-Blazquez и Salamanca-Mino в [3] для скалярного параметра θ , а Blazquez и Rubio в [4] — случай многомерного параметра θ . Решение проблемы получено с использованием счетной системы ассоциированных с экспоненциальным семейством ортогональных полиномов, которые предложил Morris [5]. Задачу сравнения несмещенной оценки и оценки максимального правдоподобия с помощью асимптотиче-

ских разложений их среднеквадратических ошибок изучали Hwang и Hu [6].

Совершенно иной подход к изучению предельного поведения несмещенных оценок предложен в работах автора [7, 8]. Он основывается на применении локальных предельных теорем и разложения Эджворта [9, 10] к плотности условного распределения, определяющей несмещенную оценку плотности заданного распределения [1, 11]. На основе этого подхода в [12] получены стохастические разложения как для НОРМД заданной однопараметрической функции, так и для НОРМД дисперсии этой оценки. Другие возможности изложенного подхода представлены в работах [13, 14].

В данной работе, в отличие от [12], для функций, определяющих НОРМД плотности распределения и заданной параметрической функции, найдены разложения более высокого порядка. Это позволило получить асимптотические разложения дисперсий несмещенных оценок для более широкого, чем в [6], класса оценок, не требуя при этом существования разложений несмещенных оценок в сходящиеся ряды Тейлора.

2 Основная модель наблюдений и ее параметризации

Имеется $\vec{X} = (X_1, \dots, X_n)$ — повторная выборка, элементы которой имеют то же распределение,

* Работа выполнена при финансовой поддержке Министерства образования и науки РФ, проект № 2096.

¹Пермский государственный национальный исследовательский университет, chichagov@psu.ru

что и наблюдаемая случайная величина ξ . Распределение вероятностей случайной величины ξ принадлежит естественному экспоненциальному семейству распределений \mathcal{M} (см., например, [15]), которое параметризуется либо с помощью канонического параметра, либо с помощью параметра «среднее». В зависимости от типа параметра распределения ξ будем различать 2 формы параметризации семейства \mathcal{M} .

Каноническая параметризация. Естественное однопараметрическое экспоненциальное семейство \mathcal{M} с каноническим параметром $\theta \in \Theta = \text{Int}[\tilde{\Theta}] \subseteq \mathbb{R}$ определяется плотностью распределения

$$f[x; \theta] = \exp \{ \theta T[x] - \kappa[\theta] + d[x] \}, \quad x \in \mathbb{X}_G \subseteq \mathbb{R}, \quad (1)$$

относительно меры $\mu[x]$. Здесь \mathbb{X}_G — носитель распределения (1); $d[x]$ и $T[x]$ — известные борелевские функции; $\kappa[\theta]$ — кумулянтное преобразование данного распределения (см., например, [16, с. 13]); $\mu[x]$ — мера Лебега, если случайная величина ξ имеет абсолютно непрерывное распределение, или считающая мера, если ξ имеет решетчатое распределение. Функция $\kappa[\theta]$ в соответствии с теоремой 2.2 из [16] имеет производную любого порядка.

Как обычно, Θ — множество значений θ , при которых

$$\int \exp \{ \theta T[x] + d[x] \} d\mu[x] < \infty,$$

называемое естественным параметрическим множеством семейства \mathcal{M} .

Мean-параметризация. Величину $a = ET[\xi] = \kappa'[\theta] : \Theta \rightarrow \mathbb{A} \in \mathbb{R}$ называют параметром «среднего значения» [15, 16] семейства \mathcal{M} . В случае естественного экспоненциального семейства дисперсия $VT[\xi] = \kappa''[\theta] > 0$ при любом $\theta \in \Theta$, а потому между параметрами θ и a существует взаимно однозначное соответствие. Пусть $\theta = \Phi_1[a]$ — функция, обратная к $a = \kappa'[\theta]$. Обе функции $\theta = \Phi_1[a]$ и $a = \kappa'[\theta]$ бесконечное число раз дифференцируемы, при этом $\Phi_1'[a] > 0$ для любого $a \in \mathbb{A}$.

Функция плотности (1) при mean-параметризации принимает вид:

$$f[x; \Phi_1[a]] = \exp \{ \Phi_1[a]T[x] - \kappa[\Phi_1[a]] + d[x] \}, \quad x \in \mathbb{X}_G \subseteq \mathbb{R}. \quad (2)$$

3 Основные предположения и обозначения

(A₁) Распределение случайной величины ξ принадлежит естественному экспоненциальному

семейству, определяемому выражениями (1) и (2).

(A₂) Если $\mu[x]$ — мера Лебега, то для каждого $\theta \in \Theta$ существует $n_0 \in \mathbb{N}$, при котором случайная величина

$$Z_n = \frac{S_n - na}{b\sqrt{n}}$$

где

$$S_n = \sum_{i=1}^n T[X_i]; \quad a = ET[\xi]; \quad b = \sqrt{VT[\xi]},$$

имеет ограниченную плотность $f_{Z_n}[x; \theta]$. Если $\mu[x]$ — считающая мера, то носитель распределения $\mathbb{X}_G \subset \mathbb{Z}$, но не содержится ни в какой подрешетке решетки \mathbb{Z} .

Введем обозначения:

$g(\theta)$, $\theta \in \Theta$ ($G[a] = g[\Phi_1[a]]$, $a \in \mathbb{A}$, при mean-параметризации) — заданная параметрическая функция;

$\hat{g}[\theta] = \hat{g}_n[\theta|S_n]$ ($\hat{G}[a] = \hat{G}_n[a|S_n]$) — несмещенная оценка функции $g(\theta)$ ($G[a]$) по достаточной статистике S_n (отметим, что $\hat{g}[\theta] \equiv \hat{G}[a]$);

$\hat{g}_n[\theta; Z_n]$ ($\hat{G}_n[a; Z_n]$) — несмещенная оценка $\hat{g}_n[\theta|S_n]$ ($\hat{G}_n[a|S_n]$) как функция от нормированной суммы Z_n ;

$h'[z]$, $h''[z]$, $h^{(j)}[z]$ — соответственно производные 1-го, 2-го и j -го порядков функции $h[z]$;

$\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ — расширенное множество натуральных чисел;

$\kappa_j = \kappa^{(j)}[\theta] = \kappa^{(j)}[\Phi_1[a]]$ — кумулянт j -го порядка распределения $T[\xi]$;

$\rho_j = \kappa_j/b^j$ — нормированный кумулянт j -го порядка распределения $T[\xi]$;

$\varphi[z] = (1/\sqrt{2\pi})e^{-z^2/2}$ — плотность стандартного нормального распределения;

$H_j[z]$ — полином Чебышёва–Эрмита j -го порядка;

$$H_2[z] = z^2 - 1;$$

$$H_3[z] = z^3 - 3z;$$

$I[A]$ — индикатор события A .

4 Интегральное представление несмещенных оценок и несмещенно оцениваемых параметрических функций

В условиях (A₁) и (A₂) сумма S_n является полной достаточной статистикой для параметра θ ∈ Θ (a ∈ A). При этом если несмещенная оценка заданной параметрической функции является функцией от полной статистики, то она имеет равномерно минимальную дисперсию. В дальнейшем речь будет идти только о несмещенных оценках с равномерно минимальной дисперсией. Отметим также, что носитель распределения S_n не зависит от параметра θ (см. упражнение 2.11 в [15]).

Пусть несмещенная оценка ĝ_n[θ|S_n] функции g(θ), θ ∈ Θ, при n ≥ L₀ удовлетворяет уравнению несмещенности

$$g[\theta] = \int_{\mathbb{R}} \hat{g}_n[\theta|s] f_{S_n}[s; \theta] d\mu[s]. \quad (3)$$

Здесь и далее L₀ — минимальное значение n, при котором выполняется (3), f_{S_n}[s; θ] — плотность распределения статистики S_n относительно меры μ[·].

В условиях (A₁) и (A₂) функция f_{S_n}[s; θ] имеет вид:

$$f_{S_n}[s; \theta] = \exp \{ \theta s - n\kappa[\theta] + d_n[s] \},$$

где d_n[s] — некоторая борелевская функция.

Если использовать ĝ_L[θ|S_L] — несмещенную оценку функции g(θ) по первым L элементам выборки X̄, то уравнение несмещенности (3) примет вид:

$$g[\theta] = \int_{\mathbb{R}} \hat{g}_L[\theta|t] f_{S_L}[t; \theta] d\mu[t], \quad L_0 \leq L < n. \quad (4)$$

При сделанных предположениях НОРМД функции g(θ) определяется следующими выражениями [1, 4]:

$$\begin{aligned} \hat{g}_n[\theta|S_n] &= \int_{\mathbb{R}} \hat{g}_L[\theta|t] \hat{f}_{S_L}[t; \theta|S_n] d\mu[t]; \\ \hat{f}_{S_L}[t|S_n] &= \frac{f_{S_L}[t; \theta_0] f_{S_{n-L}}[S_n - t; \theta_0]}{f_{S_n}[S_n; \theta_0]}, \end{aligned} \quad (5)$$

где f̂_{S_L}[t; θ|S_n] — несмещенная оценка плотности f_{S_L}[t; θ]; θ₀ — произвольное значение параметра θ из области Θ (будем полагать θ₀ = θ).

Отметим, что значение оценки (5) в силу полноты статистики S_n не зависит от выбора значения L.

5 Вспомогательные результаты

Лемма 5.1. Пусть выполнены условия (A₁) и (A₂). Тогда центральные моменты μ_ℓ = μ_ℓ[a] = E(T[ξ] - a)^ℓ, ℓ = 0, 1, ..., 5, определяются выражениями:

$$\begin{aligned} \mu_0[a] &= 1; \quad \mu_1[a] = 0; \quad \mu_2[a] = \frac{1}{\Phi'_1[a]}; \\ \mu_3[a] &= -\frac{\Phi''_1[a]}{(\Phi'_1[a])^3}; \\ \mu_4[a] &= \frac{3(\Phi'_1[a])^3 + 3(\Phi''_1[a])^2 - \Phi'_1[a]\Phi_1^{(3)}[a]}{(\Phi'_1[a])^5}; \\ \mu_5[a] &= -\left(10(\Phi'_1[a])^3\Phi''_1[a] + 15(\Phi''_1[a])^3 - 10\Phi'_1[a]\Phi''_1[a]\Phi_1^{(3)}[a] + (\Phi'_1[a])^2\Phi_1^{(4)}[a]\right) / (\Phi'_1[a])^7. \end{aligned}$$

Для доказательства леммы 5.1 достаточно воспользоваться полученным в работе [6] рекуррентным соотношением

$$\mu_{\ell+1}[a] = \frac{\ell\mu_{\ell-1}[a] + \mu'_\ell[a]}{\Phi'_1[a]}.$$

Лемма 5.2. Первые 5 кумулянтов и моментов связаны соотношениями:

$$\begin{aligned} \kappa_1 &= a; \quad \kappa_2 = b^2; \quad \kappa_3 = \mu_3; \\ \kappa_4 &= \mu_4 - 3\mu_2^2; \quad \kappa_5 = \mu_5 - 10\mu_3\mu_2. \end{aligned}$$

Лемма 5.3. Пусть выполнены условия (A₁) и (A₂). Тогда нормированные кумулянты ρ_ℓ, ℓ = 3, 4, 5, имеют вид

$$\begin{aligned} \rho_3 &= -\frac{\Phi''_1[a]}{(\Phi'_1[a])^{3/2}}; \quad \rho_4 = \frac{3(\Phi''_1[a])^2 - \Phi'_1[a]\Phi_1^{(3)}[a]}{(\Phi'_1[a])^3}; \\ \rho_5 &= -\frac{15(\Phi''_1[a])^3}{(\Phi'_1[a])^{9/2}} + \frac{10\Phi''_1[a]\Phi_1^{(3)}[a]}{(\Phi'_1[a])^{7/2}} - \frac{\Phi_1^{(4)}[a]}{(\Phi'_1[a])^{5/2}}. \end{aligned}$$

Определение. Назовем взвешенным моментом j-го порядка центрированной суммы S_L величину

$$\nu_j[\theta; L] = E\left((S_L - La)^j \hat{g}_L[\theta|S_L]\right), \quad j \in \mathbb{N}_0. \quad (6)$$

Теорема 5.1. Пусть выполнены условия (A₁) и (A₂), Vĝ_L[θ|S_L] < ∞, L₀ ≤ L < n. Тогда последовательность взвешенных моментов {ν_j[θ; L], j ∈ N₀} удовлетворяет соотношениям:

$$\left. \begin{aligned} \nu_0[\theta; L] &= g[\theta]; \quad \nu_1[\theta; L] = g'[\theta]; \\ \nu_{j+1}[\theta; L] &= \nu'_j[\theta; L] + jLb^2\nu_{j-1}[\theta; L], \\ & \quad j = 1, 2, \dots, \end{aligned} \right\} \quad (7)$$

при этом ν_j[θ; L] < ∞ для любого j ∈ N₀.

Доказательство. Согласно [16] при сделанных предположениях существуют моменты любого порядка случайной величины $T[\xi]$. Поэтому $E|S_L - La|^j < \infty$ при любом $j \in \mathbb{N}_0$ и в силу неравенства Коши–Буняковского и ограниченности дисперсии $V\widehat{g}_L[\theta|S_L]$ имеем для любого $\theta \in \Theta$

$$\begin{aligned} & \left| E \left((S_L - La)^j \widehat{g}_L[\theta|S_L] \right) \right| \leq \\ & \leq \sqrt{E(S_L - La)^{2j} E(\widehat{g}_L[\theta|S_L])^2} < \infty. \end{aligned}$$

А значит, при каждом $j \in \mathbb{N}_0$ соотношение

$$\begin{aligned} \nu_j[\theta; L] &= \\ &= \int_{-\infty}^{\infty} (t - La)^j \widehat{g}_L[\theta|t] \exp\{\theta t - L\kappa[\theta] + d_L[t]\} d\mu[t] \end{aligned}$$

можно почленно дифференцировать по θ любое число раз.

Дифференцируя это соотношение и учитывая, что $a' = a'[\theta] = \kappa''[\theta]$, получим (7).

Следствие 5.1. В условиях теоремы 5.1

$$\begin{aligned} \nu_1[\theta; L] &= g'[\theta]; \\ \nu_2[\theta; L] &= b^2 L g[\theta] + g''[\theta]; \\ \nu_3[\theta; L] &= b^3 L \rho_3 g[\theta] + 3b^2 L g'[\theta] + g^{(3)}[\theta]; \\ \nu_4[\theta; L] &= 3b^4 L^2 g[\theta] + b^4 L \rho_4 g[\theta] + 4b^3 L \rho_3 g'[\theta] + \\ & \quad + 6b^2 L g''[\theta] + g^{(4)}[\theta]; \\ \nu_5[\theta; L] &= 10b^5 L^2 \rho_3 g[\theta] + b^5 L \rho_5 g[\theta] + 15b^4 L^2 g'[\theta] + \\ & \quad + 5b^4 L \rho_4 g'[\theta] + 10b^3 L \rho_3 g''[\theta] + 10b^2 L g^{(3)}[\theta] + g^{(5)}[\theta]. \end{aligned}$$

Следствие 5.2. Пусть $\gamma_j = E(Z_n^j)$, $j = 1, 2, \dots, 6$. Тогда в условиях теоремы 5.1 справедливы выражения:

$$\begin{aligned} \gamma_1 &= 0; \quad \gamma_2 = 1; \quad \gamma_3 = \frac{\rho_3}{n^{1/2}}; \quad \gamma_4 = 3 + \frac{\rho_4}{n}; \\ \gamma_5 &= \frac{10\rho_3}{n^{1/2}} + \frac{\rho_5}{n^{3/2}}; \quad \gamma_6 = 15 + \frac{10\rho_3^2 + 15\rho_4}{n} + \frac{\rho_6}{n^2}. \end{aligned}$$

Утверждение следствия 5.2 легко получить, воспользовавшись следствием 5.1, имея в виду, что $\gamma_j = \nu_j[\theta; L]/(b\sqrt{n})^j$ при $g[\theta] \equiv 1$, $L = n$.

Следствие 5.3. В условиях теоремы 5.1 любая несмещенно оцениваемая функция $g[\theta]$ имеет производные любого порядка.

Далее $p_n^{[k]}[z; \theta]$ — приближение Эджворта порядка k для плотности распределения нормированной суммы Z_n , определяемое выражением:

$$p_n^{[k]}[z; \theta] = \sum_{j=0}^k \frac{q_j[z]}{n^{j/2}}. \quad (8)$$

Здесь

$$q_j[z] = \varphi[z] \varrho_j[z], \quad \varrho_0[z] \equiv 1,$$

где

$$\varrho_j[z] = \sum H_{j+2s}[z] \prod_{m=1}^j \frac{1}{k_m!} \left(\frac{\rho_{m+2}}{(m+2)!} \right)^{k_m}, \quad j = \overline{1, k}. \quad (9)$$

Суммирование в (9) производится по всем целым неотрицательным решениям (k_1, \dots, k_j) уравнения $k_1 + 2k_2 + \dots + jk_j = j$ при $s = k_1 + k_2 + \dots + k_j$.

Основываясь на хорошо известном соотношении

$$(\varphi[z] H_k[z])^{(\ell)} = (-1)^\ell \varphi[z] H_{k+\ell}[z], \quad \ell \in \mathbb{N}_0,$$

и используя (8), (9), нетрудно получить следующее утверждение.

Лемма 5.4. Для любых $z \in \mathbb{R}$, $m \in \mathbb{N}_0$

$$\begin{aligned} q_0^{(m)}[z] &= (-1)^m \varphi[z] H_m[z]; \\ q_1^{(m)}[z] &= (-1)^m \varphi[z] \frac{\rho_3 H_{3+m}[z]}{6}; \\ q_2^{(m)}[z] &= (-1)^m \varphi[z] \left\{ \frac{\rho_4 H_{4+m}[z]}{24} + \frac{\rho_3^2 H_{6+m}[z]}{72} \right\}; \\ q_3^{(m)}[z] &= (-1)^m \varphi[z] \left\{ \frac{\rho_5 H_{5+m}[z]}{5!} + \right. \\ & \quad \left. + \frac{\rho_3 \rho_4 H_{7+m}[z]}{3!4!} + \frac{\rho_3^3 H_{9+m}[z]}{(3!)^4} \right\}; \\ q_4^{(m)}[z] &= (-1)^m \varphi[z] \left\{ \frac{\rho_6 H_{6+m}[z]}{6!} + \right. \\ & \quad \left. + \left(\frac{\rho_3 \rho_5}{3!5!} + \frac{\rho_4^2}{(4!)^2 2!} \right) H_{8+m}[z] + \right. \\ & \quad \left. + \frac{\rho_3^2 \rho_4 H_{10+m}[z]}{(3!)^2 2! 4!} + \frac{\rho_3^4 H_{12+m}[z]}{(3!)^4 4!} \right\}. \end{aligned}$$

Обозначим

$$Q_k[z; n] = \sum_{j=0}^k \frac{q_j[z]}{n^{j/2}} + \mathbf{O}\left(n^{-(k+1)/2}\right). \quad (10)$$

Лемма 5.5. Если A_0, \dots, A_k — некоторые постоянные, то для любых $\alpha > 0$, $0 < \delta < 0,5$ при $|z| \leq \sqrt{2\alpha \ln n}$, $(k+1)/2 - \alpha \geq 5/2$, $m = \overline{0, 4}$, $n \rightarrow \infty$ справедливы следующие соотношения:

$$\begin{aligned} (Q_k[z; n])^{-1} \sum_{j=0}^k \frac{A_j q_j^{(m)}[z]}{n^{j/2}} &= \\ &= \sum_{j=0}^{4-m} \frac{a_j[z]}{n^{j/2}} + \mathbf{O}\left(n^{-5/2+m/2+\delta}\right) = \mathbf{O}\left(\ln^{m/2} n\right), \end{aligned}$$

где

$$a_0[z] = \frac{A_0 q_0^{(m)}[z]}{\varphi[z]};$$

$$a_j[z] = \frac{A_j q_j^{(m)}[z]}{\varphi[z]} - \sum_{i=0}^{j-1} \frac{a_i[z] q_{j-i}[z]}{\varphi[z]}, \quad j = 1, \dots, k.$$

При этом для любого $k \in \mathbb{N}_0$

$$(Q_k[z; n])^{-1} = \mathbf{O}(n^\alpha). \quad (11)$$

Доказательство. Заметим, что $\varphi[\sqrt{2\alpha \ln n}] = \varphi[0]n^{-\alpha}$, $\{a_j[z], j = \overline{0, k}\}$ — многочлены от z , а функции $\{q_j^{(m)}[z], j = \overline{0, k}, m = \overline{0, 4}\}$ равномерно ограничены на числовой прямой. Поэтому для любого $j = \overline{0, k}$ при $|z| \leq \sqrt{2\alpha \ln n}$ верна оценка $a_j[z] = \mathbf{o}(n^\delta)$. При этом справедливы соотношения:

$$(Q_k[z; n])^{-1} = \frac{1}{\varphi[z](1 + \mathbf{o}(1)) + \mathbf{O}(n^{-(k+1)/2})} = \mathbf{O}(n^\alpha);$$

$$\sum_{j_1, j_2=0}^k \frac{q_{j_1}[z] a_{j_2}[z]}{n^{(j_1+j_2)/2}} = \sum_{j=0}^k \frac{A_j q_j^{(m)}[z]}{n^{j/2}} + \mathbf{O}(n^{-(k+1)/2+\delta}),$$

$$(Q_k[z; n])^{-1} \left\{ \sum_{j=0}^k \frac{A_j q_j^{(m)}[z]}{n^{j/2}} - \sum_{j_1, j_2=0}^k \frac{q_{j_1}[z] a_{j_2}[z]}{n^{(j_1+j_2)/2}} + \mathbf{O}(n^{-(k+1)/2+\delta}) \right\} = \mathbf{O}(n^{-(k+1)/2+\alpha+\delta}) = \mathbf{O}(n^{-5/2+\delta}).$$

Осталось заметить, что

$$\sum_{j=0}^4 \frac{a_j[z]}{n^{j/2}} + \mathbf{O}(n^{-5/2+\delta}) - \sum_{j=0}^{4-m} \frac{a_j[z]}{n^{j/2}} = \mathbf{O}(n^{-5/2+m/2+\delta}), \quad |a_0[z]| = \mathbf{O}(\ln^{m/2} n).$$

Лемма 5.6. Для любых $\alpha > 0$, $0 < \delta < 0,5$ при $|z| \leq \sqrt{2\alpha \ln n}$, $n \rightarrow \infty$, $(k+1)/2 - \alpha \geq 5/2$ справедливо разложение:

$$(Q_k[z; n])^{-1} \sum_{j=0}^k \frac{(j+1)(j+3)q_j^{(0)}[z]}{n^{j/2}} = 3 + \mathbf{O}(n^{-1/2+\delta}).$$

Утверждение леммы 5.6 нетрудно получить, воспользовавшись леммами 5.4 и 5.5 при $A_j = (j+1)(j+3)$, $a_0[z] = 3$, $m = 0$.

Лемма 5.7. Для любых $\alpha > 0$, $0 < \delta < 0,5$ при $|z| \leq \sqrt{2\alpha \ln n}$, $n \rightarrow \infty$, $(k+1)/2 - \alpha \geq 5/2$, $m = \overline{0, 2}$ справедливо разложение:

$$(Q_k[z; n])^{-1} \sum_{j=0}^k \frac{(j+1)q_j^{(m)}[z]}{n^{j/2}} = \sum_{j=0}^{2-m} \frac{b_{m,j}[z]}{n^{j/2}} + \mathbf{O}(n^{-3/2+m/2+\delta}),$$

где

$$b_{0,0}[z] = 1; \quad b_{0,1}[z] = \frac{\rho_3 H_3[z]}{6};$$

$$b_{0,2}[z] = \frac{1}{12} \{ \rho_3^2 (-3z^4 + 12z^2 - 5) + \rho_4 (z^4 - 6z^2 + 3) \};$$

$$b_{1,0}[z] = -z; \quad b_{1,1}[z] = -\frac{\rho_3}{6} (z^4 - 9z^2 + 6);$$

$$b_{2,0}[z] = H_2[z].$$

Утверждение леммы 5.7 нетрудно получить, воспользовавшись леммами 5.4 и 5.5 при $A_j = j+1$, $a_j[z] = b_{m,j}[z]$, $m = 0, 1, 2$.

Лемма 5.8. Для любых $\alpha > 0$, $0 < \delta < 0,5$ при $|z| \leq \sqrt{2\alpha \ln n}$, $n \rightarrow \infty$, $(k+1)/2 - \alpha \geq 5/2$, $m = \overline{0, 4}$ справедливо разложение:

$$(Q_k[z; n])^{-1} \sum_{j=0}^k \frac{q_j^{(m)}[z]}{n^{j/2}} = \sum_{j=0}^{4-m} \frac{c_{m,j}[z]}{n^{j/2}} + \mathbf{O}(n^{-5/2+m/2+\delta}),$$

где

$$c_{0,0}[z] = 1; \quad c_{0,j}[z] = 0, \quad j = \overline{1, 4};$$

$$c_{1,0}[z] = -z; \quad c_{1,1}[z] = \frac{\rho_3 H_2[z]}{2};$$

$$c_{1,2}[z] = \frac{z}{6} \{ -3(z^2 - 2)\rho_3^2 + (z^2 - 3)\rho_4 \};$$

$$c_{1,3}[z] = \frac{1}{24} \{ 3(5z^4 - 16z^2 + 5)\rho_3^3 - 2(5z^4 - 21z^2 + 8)\rho_3\rho_4 + \rho_5(z^4 - 6z^2 + 3) \};$$

$$c_{2,0}[z] = H_2[z]; \quad c_{2,1}[z] = -z\rho_3(z^2 - 2);$$

$$c_{2,2}[z] = \frac{1}{12} \{ 3\rho_3^2(5z^4 - 16z^2 + 5) - 2\rho_4(2z^4 - 9z^2 + 3) \};$$

$$c_{3,0}[z] = -H_3[z]; \quad c_{3,1}[z] = \frac{\rho_3}{2}(3z^4 - 12z^2 + 5);$$

$$c_{4,0}[z] = z^4 - 6z^2 + 3.$$

Утверждение леммы 5.8 нетрудно получить, воспользовавшись леммами 5.4 и 5.5 при $A_j \equiv 1$, $a_j[z] = c_{m,j}[z]$, $m = \overline{0,4}$.

Лемма 5.9. При $u = \sqrt{n/(n-L)}(z - \Delta/\sqrt{n})$, $L \in \mathbb{N}$, $\Delta = \mathbf{O}(n^\nu)$, $0 < \nu < 0,5$, $|z| \leq \sqrt{2\alpha \ln n}$, $n \rightarrow \infty$, для любых $m = \overline{0,4}$, $\alpha > 0$ справедливо разложение:

$$(u - z)^m = A_m[z, \Delta, n, L] + P_{m,1}[\Delta] \mathbf{O}\left(\frac{|z|^m}{n^{5/2}}\right) = P_{m,2}[\Delta] \mathbf{O}\left(\frac{1}{n^{m/2}}\right), \quad (12)$$

где $\{P_{m,j}[\Delta], j = 1, 2\}$ — некоторые многочлены от Δ степени m ,

$$A_1[z, \Delta, n, L] = \frac{3L^2 z}{8n^2} + \frac{Lz}{2n} - \left(\frac{L}{2n^{3/2}} + \frac{1}{\sqrt{n}}\right) \Delta;$$

$$A_2[z, \Delta, n, L] = \frac{L^2 z^2}{4n^2} - \frac{Lz}{n^{3/2}} \Delta + \left(\frac{L}{n^2} + \frac{1}{n}\right) \Delta^2;$$

$$A_3[z, \Delta, n, L] = \frac{3Lz}{2n^2} \Delta^2 - \frac{1}{n^{3/2}} \Delta^3;$$

$$A_4[z, \Delta, n, L] = \frac{\Delta^4}{n^2}, \quad A_5[z, \Delta, n, L] \equiv 0.$$

При этом для любого $z_* \in [\min(u, z), \max(u, z)]$ верна оценка:

$$\frac{\varphi[z_*]}{\varphi[z]} = \mathbf{O}(1).$$

Доказательство. Соотношение (12) легко получить с помощью элементарных преобразований из разложения

$$\begin{aligned} u - z &= \\ &= \left(1 + \frac{L}{2n} + \frac{3L^2}{8n^2} + \mathbf{O}(n^{-3})\right) \left(z - \frac{\Delta}{\sqrt{n}}\right) - z = \\ &= \frac{Lz}{2n} + \frac{3L^2 z}{8n^2} - \Delta \left(\frac{1}{\sqrt{n}} + \frac{L}{2n^{3/2}}\right) + \Delta \mathbf{O}\left(\frac{1}{n^{5/2}}\right). \end{aligned}$$

Теперь заметим, что при

$$z_* \in (\min(u, z), \max(u, z))$$

верны неравенства:

$$|z_* - z| \leq |u - z| = \mathbf{O}\left(\frac{\Delta}{\sqrt{n}}\right);$$

$$|z_* + z| \leq |z_* - z| + 2|z| = \mathbf{O}(\sqrt{\ln n}).$$

Отсюда

$$\begin{aligned} \frac{\varphi[z_*]}{\varphi[z]} &= \exp\left\{-\frac{(z_* - z)(z_* + z)}{2}\right\} = \\ &= \exp\left\{\mathbf{O}\left(\frac{z\Delta}{\sqrt{n}}\right)\right\} = \mathbf{O}(1), \end{aligned}$$

что завершает доказательство леммы 5.9.

Лемма 5.10. Пусть выполнены условия (\mathbf{A}_1) и (\mathbf{A}_2) . Тогда для любых $\alpha > 0$, $\beta > 0$ справедливы следующие оценки:

$$\mathbf{E}|Z_n|^\beta \leq c_\beta; \quad (13)$$

$$\mathbf{P}\left(|Z_n| \geq \sqrt{2\alpha \ln n}\right) = \mathbf{o}(n^{-\alpha}) \text{ при } n \rightarrow \infty; \quad (14)$$

$$\mathbf{E}\left(|Z_n|^\beta I\left[|Z_n| \geq \sqrt{2\alpha \ln n}\right]\right) = \mathbf{o}(n^{-\alpha/2}) \text{ при } n \rightarrow \infty,$$

где c_β — положительная постоянная, зависящая только от β .

Доказательство. Оценку (13) получим, воспользовавшись теоремой 20 из [17, с. 89] при $\beta \geq 2$ и неравенством Ляпунова при $0 < \beta < 2$:

$$\begin{aligned} \mathbf{E}|Z_n|^\beta &= \frac{\mathbf{E}|S_n - na|^\beta}{(b\sqrt{n})^\beta} \leq \\ &\leq \frac{C[\beta]n^{\beta/2-1}n\mathbf{E}|T[X_1] - a|^\beta}{(b\sqrt{n})^\beta}; \end{aligned}$$

$$\mathbf{E}|Z_n|^\beta \leq (\mathbf{E}|Z_n|^2)^{\beta/2} = 1,$$

где $C[\beta]$ — некоторая постоянная, зависящая только от β .

Применив теорему Амосовой [9, с. 309] при $c = \sqrt{2\alpha}$, оценим

$$\begin{aligned} \mathbf{P}\left(|Z_n| \geq c\sqrt{\ln n}\right) &= \\ &= \frac{2}{cn^{c^2/2}\sqrt{2\pi \ln n}} \left(1 + \mathbf{O}\left(\frac{1}{\ln n}\right)\right) = \\ &= \frac{1}{n^\alpha \sqrt{\alpha\pi \ln n}} \left(1 + \mathbf{O}\left(\frac{1}{\ln n}\right)\right) = \mathbf{o}(n^{-\alpha}). \end{aligned}$$

Используя неравенство Коши–Буняковского и (13)–(14), получим:

$$\begin{aligned} \mathbf{E}\left(|Z_n|^\beta I\left[|Z_n| \geq \sqrt{2\alpha \ln n}\right]\right) &\leq \\ &\leq \sqrt{\mathbf{E}|Z_n|^{2\beta} \mathbf{P}\left(|Z_n| \geq \sqrt{2\alpha \ln n}\right)} = \mathbf{o}(n^{-\alpha/2}). \end{aligned}$$

Лемма 5.11. Пусть выполнены условия (\mathbf{A}_1) и (\mathbf{A}_2) , $L_0 \leq L < n$.

1. Если

$$\delta_0 > 0; \beta > 0; \mathbf{E}|\widehat{g}_L[\theta|S_L] - g[\theta]|^{\max\{1, \beta + \delta_0\}} < \infty,$$

то

$$\begin{aligned} \mathbf{E}\left(|\widehat{g}_n[\theta|S_n] - g[\theta]|^\beta I\left[|Z_n| \geq \sqrt{2\alpha \ln n}\right]\right) &= \\ &= \mathbf{o}\left(n^{-\alpha\delta_0/(\beta + \delta_0)}\right). \end{aligned} \quad (15)$$

2. Если $\beta \geq 2$, $E|\widehat{g}_L[\theta|S_L] - g[\theta]|^\beta < \infty$, n_L — целая часть числа n/L , то

$$E|\widehat{g}_n[\theta|S_n] - g[\theta]|^\beta \leq d_\beta n_L^{-\beta/2} E|\widehat{g}_L[\theta|S_L] - g[\theta]|^\beta, \quad (16)$$

где d_β — положительная постоянная, зависящая только от β , и

$$E|\widehat{g}_n[\theta|S_n]|^\beta = |g[\theta]|^\beta + \mathbf{O}\left(\frac{1}{\sqrt{n}}\right). \quad (17)$$

Доказательство. Если $r = \beta + \delta_0 < 1$, то, используя неравенство Ляпунова и теорему Рао–Блэквэлла–Колмогорова, получим:

$$E|\widehat{g}_n[\theta|S_n] - g[\theta]|^r \leq (E|\widehat{g}_n[\theta|S_n] - g[\theta]|)^r \leq (E|\widehat{g}_L[\theta|S_L] - g[\theta]|)^r < \infty.$$

При $r \geq 1$, используя теорему Рао–Блэквэлла–Колмогорова, имеем:

$$E|\widehat{g}_n[\theta|S_n] - g[\theta]|^r \leq E|\widehat{g}_L[\theta|S_L] - g[\theta]|^r < \infty.$$

Докажем (15), воспользовавшись неравенством Гельдера с $p = 1 + \delta_0/\beta$ и (14):

$$\begin{aligned} E\left(|\widehat{g}_n[\theta|S_n] - g[\theta]|^\beta I\left[|Z_n| \geq \sqrt{2\alpha \ln n}\right]\right) &\leq \\ &\leq \left\{E|\widehat{g}_n[\theta|S_n] - g[\theta]|^{\beta+\delta_0}\right\}^{1/p} \times \\ &\times \left\{EI\left[|Z_n| \geq \sqrt{2\alpha \ln n}\right]\right\}^{(p-1)/p} = \\ &= \{o(n^{-\alpha})\}^{(p-1)/p} = o\left(n^{-\alpha\delta_0/(\beta+\delta_0)}\right). \end{aligned}$$

Теперь по каждой подвыборке $(X_{1+(j-1)L}, X_{2+(j-1)L}, \dots, X_{L+(j-1)L})$ построим несмещенную оценку с минимальной дисперсией $\widehat{g}_{L,j}[\theta|S_{L,j}]$, $j = 1, \dots, n_L$, где $S_{L,j} = \sum_{i=1}^L X_{i+(j-1)L}$. Используя теорему Рао–Блэквэлла–Колмогорова и теорему 20 [17, с. 89], убедимся в справедливости (16) при $\beta \geq 2$:

$$\begin{aligned} E|\widehat{g}_n[\theta|S_n] - g[\theta]|^\beta &\leq \\ &\leq E\left|\frac{1}{n_L} \sum_{j=1}^{n_L} \widehat{g}_{L,j}[\theta|S_{L,j}] - g[\theta]\right|^\beta \leq \\ &\leq \frac{d_\beta}{n_L^{\beta/2}} E|\widehat{g}_L[\theta|S_L] - g[\theta]|^\beta. \end{aligned}$$

Справедливость (17) при $\beta \geq 2$ следует из оценок:

$$\begin{aligned} \left(E|\widehat{g}_n[\theta|S_n]|^\beta\right)^{1/\beta} &\leq \\ &\leq \left(E|\widehat{g}_n[\theta|S_n] - g[\theta]|^\beta\right)^{1/\beta} + \left(E|g[\theta]|^\beta\right)^{1/\beta}; \end{aligned}$$

$$\begin{aligned} \left(E|g[\theta]|^\beta\right)^{1/\beta} - \left(E|\widehat{g}_n[\theta|S_n]|^\beta\right)^{1/\beta} &\leq \\ &\leq \left(E|\widehat{g}_n[\theta|S_n] - g[\theta]|^\beta\right)^{1/\beta}, \end{aligned}$$

которые получены с помощью неравенства Минковского.

Лемма 5.12. Пусть выполнены условия (\mathbf{A}_1) и (\mathbf{A}_2) , $\sqrt{V\widehat{g}_L[\theta|S_L]} < \infty$ при некотором $L_0 \leq L < n$, $B = \{t : |t - La| \leq n^\nu\}$ при $0 < \nu < 0,5$, $\Delta[t] = (t - La)/b$. Тогда для любых $\{\zeta_i > 0, i = 1, 2, 3\}, j \in \mathbb{N}_0$, при $n \rightarrow \infty$ справедливы следующие оценки:

$$E\{I[S_L \notin B]\} = P(|S_L - La| > n^\nu) = \mathbf{O}(n^{-\zeta_1}); \quad (18)$$

$$E\{I[S_L \notin B] \cdot |\widehat{g}_L[\theta|S_L] - g[\theta]|\} = \mathbf{O}(n^{-\zeta_2}); \quad (19)$$

$$E\{I[S_L \notin B] \cdot |\Delta^j[S_L] \widehat{g}_L[\theta|S_L]|\} = \mathbf{O}(n^{-\zeta_3}). \quad (20)$$

Доказательство. Поскольку существует момент любого порядка случайной величины S_L , то для любого $\beta > 0$ при $n \rightarrow \infty$ с учетом неравенства Маркова имеем:

$$\begin{aligned} P(|S_L - La| > n^\nu) &\leq \frac{E|S_L - La|^\beta}{n^{\nu\beta}} = \\ &= \frac{(b\sqrt{L})^\beta E|Z_L|^\beta}{n^{\beta\nu}} = \mathbf{O}(n^{-\beta\nu}). \end{aligned}$$

Отсюда в силу неравенства Коши–Буняковского и теоремы 5.1

$$\begin{aligned} E\{I[S_L \notin B] \cdot |\widehat{g}_L[\theta|S_L] - g[\theta]|\} &\leq \\ &\leq \sqrt{V\widehat{g}_L[\theta|S_L]} EI[S_L \notin B] = \mathbf{O}(n^{-\beta\nu/2}); \end{aligned}$$

$$\begin{aligned} E\{I[S_L \notin B] \cdot |\Delta^j[S_L] \widehat{g}_L[\theta|S_L]|\} &= \\ &= \mathbf{O}(n^{-\beta\nu/4}), \quad j \in \mathbb{N}_0. \end{aligned}$$

Произвольность выбора β в таком случае означает справедливость (18)–(20).

6 Асимптотические разложения функций, определяющих несмещенные оценки

Теорема 6.1. Пусть выполнены условия (\mathbf{A}_1) и (\mathbf{A}_2) , $L_0 \leq L < n$, $\Delta = (t - La)/b = \mathbf{O}(n^\nu)$ при $0 < \nu < 0,5$. Тогда для любых $\alpha > 0, 0 < \delta < 0,5$ при $|z| \leq \sqrt{2\alpha \ln n}, s = na + zb\sqrt{n}, n \rightarrow \infty$ справедливо асимптотическое разложение:

$$\frac{f_{S_{n-L}}[s-t; \theta]}{f_{S_n}[s; \theta]} = 1 + \sum_{j=1}^4 \frac{c_{jf}[z, L, \Delta]}{n^{j/2}} + P_5[\Delta] \mathbf{O}\left(n^{-5/2+\delta}\right), \quad (21)$$

где $P_5[\Delta]$ — некоторый многочлен от Δ степени 5,

$$\begin{aligned} c_{1f}[z, L, \Delta] &= z\Delta; \\ c_{2f}[z, L, \Delta] &= \frac{H_2[z]}{2} (\Delta^2 - \rho_3\Delta - L); \\ c_{3f}[z, L, \Delta] &= L\rho_3 \left(\frac{z^3}{3} - \frac{z}{2} \right) + \\ &+ \Delta \left\{ H_3[z] \left(\frac{\rho_3^2 - L}{2} - \frac{\rho_4}{6} \right) + \frac{\rho_3^2 z}{2} \right\} + \\ &+ \Delta^2 \rho_3 \left(z - \frac{z^3}{2} \right) + \Delta^3 \frac{H_3[z]}{6}; \\ c_{4f}[z, L, \Delta] &= L^2 \left(\frac{3}{8} - \frac{3z^2}{4} + \frac{z^4}{8} \right) - \\ &- L\rho_3^2 \left(\frac{5}{24} - z^2 + \frac{3z^4}{8} \right) + L\rho_4 \left(\frac{1}{8} - \frac{z^2}{2} + \frac{z^4}{8} \right) + \\ &+ \Delta \left\{ L\rho_3 \left(\frac{3}{4} - 2z^2 + \frac{7z^4}{12} \right) - \rho_3^3 \left(\frac{5}{8} - 2z^2 + \frac{5z^4}{8} \right) + \right. \\ &+ \rho_3\rho_4 \left(\frac{2}{3} - \frac{7z^2}{4} + \frac{5z^4}{12} \right) - \rho_5 \left(\frac{1}{8} - \frac{z^2}{4} + \frac{z^4}{24} \right) \left. \right\} + \\ &+ \Delta^2 \left\{ \rho_3^2 \left(\frac{5}{8} - 2z^2 + \frac{5z^4}{8} \right) - L \left(\frac{3}{4} - \frac{3z^2}{2} + \frac{z^4}{4} \right) - \right. \\ &- \rho_4 \left(\frac{1}{4} - \frac{3z^2}{4} + \frac{z^4}{6} \right) \left. \right\} - \Delta^3 \rho_3 \left(\frac{5}{12} - z^2 + \frac{z^4}{4} \right) + \\ &+ \Delta^4 \left(\frac{1}{8} - \frac{z^2}{4} + \frac{z^4}{24} \right). \end{aligned}$$

При этом для любого $t \in \mathbb{R}$

$$\frac{f_{S_{n-L}}[s-t; \theta]}{f_{S_n}[s; \theta]} = \mathbf{O}(n^\alpha). \quad (22)$$

Доказательство. Положим

$$u = \left(\frac{n}{n-L} \right)^{1/2} \left(z - \frac{\Delta}{\sqrt{n}} \right).$$

Используя уточнения локальной предельной теоремы (теоремы 13 и 15 из [9]) и приближение Эджворта (8), левую часть (21) при любом $k \in \mathbb{N}$ можно представить в виде:

$$\begin{aligned} \Psi_n &= \frac{f_{S_{n-L}}[s-t; \theta]}{f_{S_n}[s; \theta]} = \\ &= \sqrt{\frac{n}{n-L}} \frac{p_{n-L}^{[k]}[u; \theta] + \mathbf{O}(n^{-(k+1)/2})}{p_n^{[k]}[z; \theta] + \mathbf{O}(n^{-(k+1)/2})}, \quad (23) \end{aligned}$$

поскольку в случае абсолютно непрерывного распределения ξ

$$\begin{aligned} f_{S_n}[s; \theta] &= (b\sqrt{n})^{-1} f_{Z_n}[z; \theta] = \\ &= (b\sqrt{n})^{-1} \left(p_n^{[k]}[z; \theta] + \mathbf{O}(n^{-(k+1)/2}) \right), \end{aligned}$$

а в случае решетчатого распределения ξ верно соотношение:

$$b\sqrt{n} f_{S_n}[s; \theta] = p_n^{[k]}[z; \theta] + \mathbf{O}(n^{-(k+1)/2}).$$

Выберем k таким, что $(k+1)/2 - \alpha \geq 5/2$, и осуществим разложение в точке z следующего выражения:

$$\begin{aligned} \sqrt{\frac{n}{n-L}} p_{n-L}^{[k]}[u; \theta] &= \sum_{j=0}^k \frac{q_j[u]}{n^{j/2}} \left(\frac{n}{n-L} \right)^{(j+1)/2} = \\ &= \sum_{m=0}^4 \frac{(u-z)^m}{m!} \sum_{j=0}^k \frac{q_j^{(m)}[z]}{n^{j/2}} \left(\frac{n}{n-L} \right)^{(j+1)/2} + \\ &+ \frac{(u-z)^5}{5!} \varphi[z_*] \mathbf{O}(1), \end{aligned}$$

где $z_* \in [\min(u, z), \max(u, z)]$.

Используя (10) и разложение

$$\left(\frac{n}{n-L} \right)^{j/2} = 1 + \frac{jL}{2n} + \frac{j(j+2)L^2}{8n^2} + \mathbf{O}(n^{-3}),$$

представим левую часть (23) в виде суммы $\Psi_n = \sum_{i=1}^6 J_i$, где

$$J_1 = (Q_k[z; n])^{-1} \sum_{m=0}^4 \frac{(u-z)^m}{m!} \sum_{j=0}^k \frac{q_j^{(m)}[z]}{n^{j/2}};$$

$$\begin{aligned} J_2 &= (Q_k[z; n])^{-1} \frac{L}{2n} \sum_{m=0}^4 \frac{(u-z)^m}{m!} \times \\ &\times \sum_{j=0}^k \frac{(j+1)q_j^{(m)}[z]}{n^{j/2}}; \end{aligned}$$

$$\begin{aligned} J_3 &= (Q_k[z; n])^{-1} \frac{L^2}{8n^2} \sum_{m=0}^4 \frac{(u-z)^m}{m!} \times \\ &\times \sum_{j=0}^k \frac{(j+1)(j+3)q_j^{(m)}[z]}{n^{j/2}}; \end{aligned}$$

$$J_4 = (Q_k[z; n])^{-1} \mathbf{O}(n^{-3}) \sum_{m=0}^4 \frac{(u-z)^m}{m!} \sum_{j=0}^k \frac{q_j^{(m)}[z]}{n^{j/2}};$$

$$J_5 = (Q_k[z; n])^{-1} \frac{(u-z)^5}{5!} \varphi[z_*] \mathbf{O}(1);$$

$$J_6 = (Q_k[z; n])^{-1} \mathbf{O}(n^{-(k+1)/2}).$$

Применяя леммы 5.5–5.9, получим при $0 < \delta_i < 0,5$, $i = 1, 2, 3$, следующие оценки слагаемых J_1, \dots, J_6 :

$$J_1 = \sum_{m=0}^4 \frac{A_m[z, \Delta, n, L]}{m!} \sum_{j=0}^{4-m} \frac{c_{m,j}[z]}{n^{j/2}} + P_{4,1}[\Delta] \mathbf{O} \left(n^{-5/2+\delta_1} \right); \quad (24)$$

$$J_2 = \frac{L}{2n} \sum_{m=0}^2 \frac{A_m[z, \Delta, n, L]}{m!} \sum_{j=0}^{2-m} \frac{b_{m,j}[z]}{n^{j/2}} + P_{4,2}[\Delta] \mathbf{O} \left(n^{-5/2+\delta_2} \right); \quad (25)$$

$$J_3 = \frac{3L^2}{8n^2} + P_{4,3}[\Delta] \mathbf{O} \left(n^{-5/2+\delta_3} \right); \quad (26)$$

$$J_4 = \mathbf{O} \left(n^{-3} \right);$$

$$J_5 = P_{5,2}[\Delta] \mathbf{O} \left(n^{-5/2} \right);$$

$$J_6 = \mathbf{O} \left(n^{-5/2} \right),$$

где $A_0[z, \Delta, n, L] \equiv 1$, $P_{i,j}[\Delta]$ — некоторый многочлен порядка i от Δ .

Подставляя в (24)–(26) выражения коэффициентов $c_{m,j}[z]$ из леммы 5.8, $A_m[z, \Delta, n, L]$ — из леммы 5.9, $b_{m,j}[z]$ — из леммы 5.7 и используя лемму 5.6, установим справедливость разложения (21).

Справедливость (22) следует из (11), (23) и ограниченности на числовой прямой функций $\{q_j[z], j = \overline{0, k}\}$.

Следствие 6.1. В условиях теоремы 6.1 верно следующее разложение функции, определяющей несмещенную оценку плотности распределения $f_{S_L}[t; \theta]$:

$$\widehat{f}_{S_L}[t; \theta|s] = f_{S_L}[t; \theta] \left\{ 1 + \sum_{j=1}^4 \frac{c_{jf}[z, L, \Delta]}{n^{j/2}} + P_5[\Delta] \mathbf{O} \left(n^{-5/2+\delta} \right) \right\}.$$

Теорема 6.2. Пусть выполнены условия (\mathbf{A}_1) и (\mathbf{A}_2) , $V\widehat{g}_L[\theta|S_L] < \infty$ при некотором $L_0 \leq L < n$. Тогда для любых $\alpha > 0$ и $0 < \delta < 0,5$ при $|z| \leq \sqrt{2\alpha \ln n}$, $n \rightarrow \infty$ справедливо следующее асимптотическое разложение функции, определяющей несмещенную оценку $\widehat{g}_n[\theta; Z_n]$ функции $g[\theta]$:

$$\widehat{g}_n[\theta; z] = g[\theta] + \sum_{j=1}^4 \frac{c_{jg}[z]}{n^{j/2}} + \mathbf{O} \left(n^{-5/2+\delta} \right), \quad (27)$$

где

$$c_{1g}[z] = \frac{zg'[\theta]}{b}; \quad c_{2g}[z] = \frac{H_2[z](g''[\theta] - b\rho_3g'[\theta])}{2b^2};$$

$$c_{3g}[z] = (z^3 - 2z) \left(\frac{\rho_3^2 g'[\theta]}{2b} - \frac{\rho_3 g''[\theta]}{2b^2} \right) + H_3[z] \left(\frac{g^{(3)}[\theta]}{6b^3} - \frac{\rho_4 g'[\theta]}{6b} \right);$$

$$c_{4g}[z] = (5 - 16z^2 + 5z^4) \left(\frac{\rho_3^2 g''[\theta]}{8b^2} - \frac{\rho_3^3 g'[\theta]}{8b} \right) + (3 - 6z^2 + z^4) \left(\frac{g^{(4)}[\theta]}{24b^4} - \frac{\rho_5 g'[\theta]}{24b} \right) + \frac{(-5 + 12z^2 - 3z^4) \rho_3 g^{(3)}[\theta]}{12b^3} + \frac{(-3 + 9z^2 - 2z^4) \rho_4 g''[\theta]}{12b^2} + \frac{(8 - 21z^2 + 5z^4) \rho_3 \rho_4 g'[\theta]}{12b}.$$

Доказательство. Положим

$$B = \{t : |t - La| \leq n^\nu\} \text{ при } 0 < \nu < 0,5;$$

$$\Delta[t] = \frac{t - La}{b}; \quad s = na + zb\sqrt{n}; \quad h[S_L] = \widehat{g}_L[\theta|S_L].$$

Тогда, воспользовавшись леммой 5.12, соотношениями (22) и (19) при $\zeta_2 \geq -5/2 - \alpha$, получим оценку:

$$\mathbf{E} \left| I[S_L \notin B] h[S_L] \left(\widehat{f}_{S_L}[S_L; \theta|s] - f_{S_L}[S_L; \theta] \right) \right| \leq \mathbf{O} \left(n^\alpha \right) \mathbf{E} \{ I[S_L \notin B] |h[S_L]| \} = \mathbf{O} \left(n^{-5/2} \right).$$

Отсюда с учетом следствия 6.1, леммы 5.12, формул (4) и (5) найдем:

$$\begin{aligned} \widehat{g}_n[\theta; z] - g[\theta] &= \mathbf{E} \left\{ I[S_L \in B] h[S_L] \left(\widehat{f}_{S_L}[S_L; \theta|s] - f_{S_L}[S_L; \theta] \right) \right\} + \mathbf{O} \left(n^{-5/2} \right) = \\ &= \sum_{j=1}^4 \frac{1}{n^{j/2}} \mathbf{E} \{ \widehat{g}_L[\theta|S_L] c_{jf}[z, L, \Delta[S_L]] \} + \mathbf{O} \left(n^{-5/2+\delta} \right). \quad (28) \end{aligned}$$

Имея в виду, что для любого $j \in \mathbb{N}_0$ согласно (6)

$$\mathbf{E} \{ \widehat{g}_L[\theta|S_L] \Delta^j[S_L] \} = \frac{\nu_j[\theta; L]}{b^j},$$

и подставляя в (28) выражения коэффициентов $\{c_{jf}[\cdot], j = \overline{1, 4}\}$ из теоремы 6.1, после элементарных преобразований получим (27).

Следствие 6.2. В условиях теоремы 6.2 справедливо следующее асимптотическое разложение функции $\widehat{G}_n[a; Z_n]$, определяющей несмещенную оценку функции $G[a] = g[\Phi_1[a]]$:

$$\widehat{G}_n[a; z] = G[a] + \frac{zG'[a]}{\sqrt{n\Phi_1'[a]}} + \frac{H_2[z]G''[a]}{2n\Phi_1'[a]} + \sum_{i=3}^4 \frac{c_{iG}[z]}{n^{i/2}} + \mathbf{O}\left(n^{-5/2+\delta}\right),$$

где

$$c_{3G}[z] = \frac{zG''[a]\Phi_1''[a]}{2(\Phi_1'[a])^{5/2}} + \frac{H_3[z]G^{(3)}[a]}{6(\Phi_1'[a])^{3/2}};$$

$$c_{4G}[z] = -\frac{H_2[z]G''[a](\Phi_1''[a])^2}{2(\Phi_1'[a])^4} + \frac{(3z^2 - 2)\Phi_1''[a]G^{(3)}[a]}{6(\Phi_1'[a])^3} + \frac{H_2[z]G''[a]\Phi_1^{(3)}[a]}{4(\Phi_1'[a])^3} + \frac{(z^4 - 6z^2 + 3)G^{(4)}[a]}{24(\Phi_1'[a])^2}.$$

Далее $\widehat{g}_n^2[\theta|S_n] = \widehat{g}_n^2[\theta; Z_n]$ и $\widehat{V}\widehat{g}_n[\theta|S_n] = \widehat{V}\widehat{g}_n[\theta; Z_n] - \text{НОРМД}$ соответственно функции $g^2[\theta]$ и дисперсии несмещенной оценки $V\widehat{g}_n[\theta|S_n]$ по выборке объема n . Отметим, что оценка $\widehat{g}_n^2[\theta|S_n]$ заведомо существует при $n > 2L \geq L_0$, если существует дисперсия $V\widehat{g}_n[\theta|S_n]$, поскольку

$$\widehat{g}_n^2[\theta|S_n] = \mathbf{E}\{\widehat{g}_{L,1}[\theta|S_{L,1}]\widehat{g}_{L,2}[\theta|S_{L,2}]|S_n\},$$

где $S_{L,1} = \sum_{i=1}^L T[X_i]$; $S_{L,2} = \sum_{i=L+1}^{2L} T[X_i]$.

Через $\widehat{G}_n^2[a|S_n] = \widehat{G}_n^2[a; Z_n]$ и $\widehat{V}\widehat{G}_n[a|S_n] = \widehat{V}\widehat{G}_n[a; Z_n]$ обозначим НОРМД соответственно функции $G^2[a]$ и дисперсии несмещенной оценки $V\widehat{G}_n[a|S_n]$.

Теорема 6.3. Пусть выполнены условия (\mathbf{A}_1) и (\mathbf{A}_2) , $V\widehat{g}_L[\theta|S_L] < \infty$ при некотором $L \in \mathbb{N}$, $V\widehat{g}_M^2[\theta|S_M] < \infty$ при некотором $M \in \mathbb{N}$. Тогда для любых $\alpha > 0$, $|z| \leq \sqrt{2\alpha \ln n}$ и $0 < \delta < 0,5$ при $n \rightarrow \infty$ справедливо следующее асимптотическое разложение функции, определяющей несмещенную оценку дисперсии несмещенной оценки $\widehat{g}_n[\theta; Z_n]$:

$$\widehat{V}\widehat{g}_n[\theta; z] = \sum_{i=2}^4 \frac{c_{i\widehat{V}\widehat{g}}[z]}{n^{i/2}} + \mathbf{O}\left(n^{-5/2+\delta}\right), \quad (29)$$

где

$$c_{2\widehat{V}\widehat{g}}[z] = \frac{(g'[\theta])^2}{b^2};$$

$$c_{3\widehat{V}\widehat{g}}[z] = \frac{2zg'[\theta]g''[\theta]}{b^3} - \frac{z\rho_3(g'[\theta])^2}{b^2};$$

$$c_{4\widehat{V}\widehat{g}}[z] = \frac{(g'[\theta])^2(3z^2\rho_3^2 - 2\rho_3^2 + \rho_4 - z^2\rho_4)}{2b^2} + \frac{g'[\theta]g''[\theta](2\rho_3 - 3z^2\rho_3)}{b^3} + \frac{(g''[\theta])^2(2z^2 - 1)}{2b^4} + \frac{H_2[z]g'[\theta]g^{(3)}[\theta]}{b^4}.$$

Доказательство. Чтобы получить (29), достаточно заменить несмещенные оценки $\widehat{g}_n[\theta|S_n]$ и $\widehat{g}_n^2[\theta|S_n]$ в хорошо известном представлении несмещенной оценки дисперсии

$$\widehat{V}\widehat{g}_n[\theta|S_n] = (\widehat{g}_n[\theta|S_n])^2 - \widehat{g}_n^2[\theta|S_n]$$

на их асимптотические разложения. Разложение оценки $\widehat{g}_n[\theta|S_n]$ определяется формулой (27), разложение оценки $\widehat{g}_n^2[\theta|S_n]$ может быть получено после замены $g[\theta]$ на $g^2[\theta]$ в (27).

Следствие 6.3. Пусть выполнены условия (\mathbf{A}_1) и (\mathbf{A}_2) , $V\widehat{G}_L[a|S_L] < \infty$ при некотором $L \in \mathbb{N}$, $V\widehat{G}_M^2[a|S_M] < \infty$ при некотором $M \in \mathbb{N}$. Тогда для любых $\alpha > 0$, $|z| \leq \sqrt{2\alpha \ln n}$ и $0 < \delta < 0,5$ при $n \rightarrow \infty$ справедливо асимптотическое разложение функции, определяющей несмещенную оценку дисперсии несмещенной оценки $\widehat{V}\widehat{G}_n[a; Z_n]$,

$$\widehat{V}\widehat{G}_n[a; z] = \sum_{i=2}^4 \frac{c_{i\widehat{V}\widehat{G}}[z]}{n^{i/2}} + \mathbf{O}\left(n^{-5/2+\delta}\right),$$

где

$$c_{2\widehat{V}\widehat{G}}[z] = \frac{(G'[a])^2}{\Phi_1'[a]};$$

$$c_{3\widehat{V}\widehat{G}}[z] = \frac{2zG'[a]G''[a]}{(\Phi_1'[a])^{3/2}} - \frac{z(G'[a])^2\Phi_1''[a]}{(\Phi_1'[a])^{5/2}};$$

$$c_{4\widehat{V}\widehat{G}}[z] = \frac{(2z^2 - 1)(G''[a])^2 + 2H_2[z]G'[a]G^{(3)}[a]}{2(\Phi_1'[a])^2} - \frac{H_2[z](4G'[a]G''[a]\Phi_1''[a] + (G'[a])^2\Phi_1^{(3)}[a])}{2(\Phi_1'[a])^3} + \frac{H_2[z](G'[a])^2(\Phi_1''[a])^2}{(\Phi_1'[a])^4}.$$

7 Асимптотические разложения дисперсий несмещенных оценок

Теорема 7.1. Пусть выполнены условия (\mathbf{A}_1) и (\mathbf{A}_2) , а также существуют $L \in \mathbb{N}$ и $\delta_0 > 0$ такие, что

$C_L = E |\widehat{g}_L[\theta|S_L] - g[\theta]|^{2+\delta_0} < \infty$. Тогда для любого $0 < \delta < 0,5$ при $n \rightarrow \infty$ справедливо асимптотическое разложение дисперсии несмещенной оценки:

$$\begin{aligned} V\widehat{g}_n[\theta|S_n] &= \frac{(g'[\theta])^2}{nb^2} + \\ &+ \frac{(b\rho_3g'[\theta] - g''[\theta])^2}{2n^2b^4} + \mathbf{O}(n^{-5/2+\delta}). \end{aligned}$$

Доказательство. Представим при произвольном $\alpha > 0$ дисперсию несмещенной оценки в виде суммы

$$V\widehat{g}_n[\theta|S_n] = J_1 + J_2, \quad (30)$$

где

$$\begin{aligned} J_1 &= E \left(|\widehat{g}_n[\theta; Z_n] - g[\theta]|^2 I \left[|Z_n| \leq \sqrt{2\alpha \ln n} \right] \right); \\ J_2 &= E \left(|\widehat{g}_n[\theta; Z_n] - g[\theta]|^2 I \left[|Z_n| > \sqrt{2\alpha \ln n} \right] \right). \end{aligned}$$

Используя неравенство Гельдера при $p = 1 + \delta_0/2$, леммы 5.10–5.11 и теорему Рао–Блэквелла–Колмогорова, оценим

$$\begin{aligned} J_2 &\leq \left\{ E |\widehat{g}_n[\theta; Z_n] - g[\theta]|^{2+\delta_0} \right\}^{1/p} \times \\ &\quad \times \left\{ EI \left[|Z_n| > \sqrt{2\alpha \ln n} \right] \right\}^{(p-1)/p} \leq \\ &\leq C_L^{1/p} \mathbf{O} \left(n^{-\alpha\delta_0/(2+\delta_0)} \right) = \mathbf{O} \left(n^{-\alpha\delta_0/(2+\delta_0)} \right). \quad (31) \end{aligned}$$

Используя леммы 5.10–5.11 и теорему 6.2, представим

$$\begin{aligned} J_1 &= E \left\{ \left(\sum_{i=1}^4 \frac{c_{ig}[Z_n]}{n^{i/2}} + \mathbf{O} \left(n^{-5/2+\delta} \right) \right)^2 \times \right. \\ &\quad \left. \times I \left[|Z_n| \leq \sqrt{2\alpha \ln n} \right] \right\} = \\ &= EP_4[Z_n] + \mathbf{O} \left(n^{-5/2+\delta} \right) + \mathbf{O} \left(n^{-\alpha} \right), \quad (32) \end{aligned}$$

где

$$\begin{aligned} P_4[z] &= \frac{z^2(g'[\theta])^2}{nb^2} + \\ &+ \frac{(z^3 - z)(g'[\theta]g''[\theta] - b\rho_3(g'[\theta])^2)}{b^3n^{3/2}} + \\ &+ \frac{1}{12n^2b^4} \left\{ (g'[\theta])^2b^2(3\rho_3^2 - 30z^2\rho_3^2 + 15z^4\rho_3^2 + \right. \\ &\quad \left. + 12z^2\rho_4 - 4z^4\rho_4) + \right. \\ &\quad \left. + g'[\theta]g''[\theta]b(36z^2\rho_3 - 6\rho_3 - 18z^4\rho_3) + \right. \\ &\quad \left. + (g''[\theta])^2(3 - 6z^2 + 3z^4) + \right. \\ &\quad \left. + g'[\theta]g^{(3)}[\theta](4z^4 - 12z^2) \right\}. \end{aligned}$$

Применяя следствие 5.2, вычислим

$$\begin{aligned} EP_4[Z_n] &= \frac{(g'[\theta])^2}{nb^2} + \\ &+ \frac{(b\rho_3g'[\theta] - g''[\theta])^2}{2n^2b^4} + \mathbf{O} \left(n^{-5/2+\delta} \right). \quad (33) \end{aligned}$$

Подставляя (33) в (30)–(32) и выбирая $\alpha > 0$ достаточно большим, завершим доказательство теоремы 7.1.

Следствие 7.1. В условиях теоремы 7.1 дисперсия несмещенной оценки $V\widehat{G}_n[a|S_n]$ имеет следующее разложение:

$$V\widehat{G}_n[a|S_n] = \frac{(G'[a])^2}{n\Phi_1'[a]} + \frac{(G''[a])^2}{2n^2(\Phi_1'[a])^2} + \mathbf{O} \left(n^{-5/2+\delta} \right).$$

Следуя ходу доказательства теоремы 7.1, но используя разложения более высокого порядка, можно получить следующее утверждение.

Теорема 7.2. Пусть выполнены условия (A_1) и (A_2) , а также существуют $L \in \mathbb{N}$ и $\delta_0 > 0$ такие, что

$$C_L = E \left| \widehat{V}\widehat{G}_L[a|S_L] - V\widehat{G}_L[a|S_L] \right|^{2+\delta_0} < \infty.$$

Тогда для любых $\alpha > 0$ и $0 < \delta < 0,5$ при $|z| \leq \sqrt{2\alpha \ln n}$, $n \rightarrow \infty$ справедливо асимптотическое разложение дисперсии несмещенной оценки дисперсии:

$$E \left(\widehat{V}\widehat{G}_n[a|S_n] - V\widehat{G}_n[a|S_n] \right)^2 = \frac{c_1}{n^3} + \frac{c_2}{n^4} + \mathbf{O} \left(\frac{1}{n^4} \right),$$

где

$$\begin{aligned} c_1 &= \frac{4(G'[a]G''[a])^2}{(\Phi_1'[a])^3} - \frac{4(G'[a])^3G''[a]\Phi_1''[a]}{(\Phi_1'[a])^4} + \\ &\quad + \frac{(G'[a])^4(\Phi_1''[a])^2}{(\Phi_1'[a])^5}; \\ c_2 &= \frac{2(G''[a])^4}{(\Phi_1'[a])^4} + \frac{2(G'[a]G^{(3)}[a])^2}{(\Phi_1'[a])^4} - \\ &\quad - \frac{2(G'[a])^3G^{(3)}[a]\Phi_1^{(3)}[a]}{(\Phi_1'[a])^5} + \\ &\quad + \frac{4G'[a](G''[a])^2(-3G''[a]\Phi_1''[a] + 2\Phi_1''[a]G^{(3)}[a])}{(\Phi_1'[a])^5} - \\ &\quad - \frac{2(G'[a])^2G''[a](5\Phi_1''[a]G^{(3)}[a] + G''[a]\Phi_1^{(3)}[a])}{(\Phi_1'[a])^5} + \end{aligned}$$

$$\begin{aligned}
& + \frac{14(G'[a]G''[a]\Phi_1''[a])^2}{(\Phi_1'[a])^6} + \frac{(G'[a])^4(\Phi_1^{(3)}[a])^2}{2(\Phi_1'[a])^6} + \\
& + \frac{4(G'[a])^3\Phi_1''[a](\Phi_1''[a]G^{(3)}[a] + G''[a]\Phi_1^{(3)}[a])}{(\Phi_1'[a])^6} - \\
& - \frac{8(G'[a])^3G''[a](\Phi_1''[a])^3}{(\Phi_1'[a])^7} - \\
& - \frac{2(G'[a])^4(\Phi_1''[a])^2\Phi_1^{(3)}[a]}{(\Phi_1'[a])^7} + \frac{2(G'[a]\Phi_1''[a])^4}{(\Phi_1'[a])^8}.
\end{aligned}$$

В заключение отметим, что предложенный в работе подход позволяет получать для несмещенных оценок и их характеристик асимптотические разложения любого порядка. Возникающая при этом проблема проведения громоздких преобразований может быть решена с помощью одной из систем аналитических вычислений.

Литература

1. *Воинов В. Г., Никулин М. С.* Несмещенные оценки и их применения. — М.: Наука, 1989. 440 с.
2. *Portnoy S.* Asymptotic efficiency of minimum variance unbiased estimators // *Ann. Stat.*, 1977. Vol. 5. No. 3. P. 522–529.
3. *Lopez-Blazquez F., Salamanca-Mino B.* Limit distribution of unbiased estimators in natural exponential families // *Statistics*, 2002. Vol. 14. No. 4. P. 329–338.
4. *Blazquez F. L., Rubio D. G.* Unbiased estimation in the multivariate natural exponential family with simple quadratic variance function // *J. Multivariate Anal.*, 2003. Vol. 86. P. 1–13.
5. *Morris C. N.* Natural exponential families with quadratic variance functions // *Ann. Stat.*, 1982. Vol. 10. No. 1. P. 65–80.
6. *Hwang T.-Y., Hu C.-Y.* More comparisons of MLE with UMVUE for exponential families // *Ann. Inst. Statist. Math.*, 1990. Vol. 42. P. 65–75.
7. *Чичагов В. В.* Об асимптотическом поведении несмещенных оценок вероятностей для решетчатых распределений, достаточной статистикой которых является среднее // *Статистические методы оценивания и проверки гипотез: Межвуз. сб. науч. тр.* — Пермь: ПГУ, 2002. С. 106–120.
8. *Chichagov V. V.* Concerning asymptotic normality of a class of unbiased estimators in the case of absolutely continuous distributions. *Statistical methods of estimation and testing of hypotheses* // *J. Math. Sci.*, 2004. Vol. 119. No. 3. P. 336–341.
9. *Петров В. В.* Суммы независимых случайных величин. — М.: Наука, 1972. 416 с.
10. *Барндорф-Нильсен О., Коке Д.* Асимптотические методы в математической статистике / Пер. с англ. — М.: Мир, 1999. 255 с. (*Barndorff-Nielsen O. E., Cox D. R.* Asymptotic techniques for use in statistics. — London: Chapman and Hall, 1989. 252 p.)
11. *Лумельский Я. П., Сапожников П. Н.* Несмещенные оценки для плотностей распределений // *Теория вероятностей и ее применение*, 1969. Т. 14. № 2. С. 372–380.
12. *Чичагов В. В.* Стохастические разложения несмещенных оценок в случае однопараметрического экспоненциального семейства // *Информатика и её применения*, 2008. Т. 2. Вып. 2. С. 62–70.
13. *Чичагов В. В.* О несмещенной оценке вероятности $P(X < Y)$ в модели нагрузка–прочность // *Теория вероятностей и ее приложения: Тезисы докл. Междунар. конф., посвященной 100-летию со дня рождения Б. В. Гнеденко.* — М.: ЛЕНАНД, 2012. С. 264.
14. *Chichagov V.* Asymptotic of the mean absolute error of UNVUE and MLE in the case of one-parameter exponential family lattice distributions // *31st Seminar (International) on Stability Problems for Stochastic Models: Book of abstracts.* — Moscow: IPI RAN, 2013. P. 13–15.
15. *Jørgensen B.* The theory of dispersion models. — London: Chapman & Hall, 1997. 256 p.
16. *Brown L. D.* Fundamentals of statistical exponential families with applications in statistical decision theory // *Lecture notes — monograph ser.* — Hayward, CA, USA: Institute of Mathematical Statistics, 1986. Vol. 9. 284 p.
17. *Петров В. В.* Предельные теоремы для сумм независимых случайных величин. — М.: Наука, 1987. 320 с.

Поступила в редакцию 10.04.14

HIGHER-ORDER ASYMPTOTIC EXPANSIONS OF UNBIASED ESTIMATORS AND THEIR VARIANCES ON THE ONE-PARAMETER EXPONENTIAL FAMILY MODEL

V. V. Chichagov

Perm State University, 15 Bukireva Str., Perm 614990, Russian Federation

Abstract: The paper considers a model of duplicate sampling with the fixed size n from a distribution belonging to the natural one-parameter exponential family. A limiting behavior of the uniformly minimum variance unbiased

estimator (UMVUE) of the given parametric function and the UMVUE variance of this estimator is studied in the case of infinite size of the sample. Higher-order asymptotic expansions are obtained for functions defining unbiased estimators and variances of these estimators. The results are presented for both the canonical parameterization and the mean parameterization.

Keywords: natural exponential family; unbiased estimate; asymptotic expansion

DOI: 10.14357/19922264150308

Acknowledgments

The research was financially supported by the Russian Ministry of Education and Science (project No. 2096).

References

- Voinov, V. G., and M. S. Nikulin. 1989. *Nesmeshchennye otsenki i ikh primeneniya* [Unbiased estimators and their applications]. Moscow: Nauka. 440 p.
- Portnoy, S. 1977. Asymptotic efficiency of minimum variance unbiased estimators. *Ann. Stat.* 5(3):522–529.
- Lopez-Blazquez, F., and B. Salamanca-Mino. 2002. Limit distribution of unbiased estimators in natural exponential families. *Statistics* 14(4):329–338.
- Blazquez, F. L., and D. G. Rubio. 2003. Unbiased estimation in the multivariate natural exponential family with simple quadratic variance function. *J. Multivariate Anal.* 86:1–13.
- Morris, C. N. 1982. Natural exponential families with quadratic variance functions. *Ann. Stat.* 10(1):65–80.
- Hwang, T.-Y., and C.-Y. Hu. 1990. More comparisons of MLE with UMVUE for exponential families. *Ann. Inst. Statist. Math.* 42(1):65–75.
- Chichagov, V. V. 2002. Ob asimptoticheskom povedenii nesmeshchennykh otsenok veroyatnostey dlya reshchatykh raspredeleniy, dostatochnoy statistikoy kotorykh yavlyatsya srednee [On asymptotic behavior of unbiased probability estimators for lattice distributions with the mean as a sufficient statistic]. *Statisticheskie Metody Otsenivaniya i Proverki Gipotez: Mezhdunarodnyy Sbornik Nauchnykh Trudov* [Statistical methods for estimating and hypothesis testing: Interuniversity Collection of Research Papers]. Perm. 16:106–120.
- Chichagov, V. V. 2004. Concerning asymptotic normality of a class of unbiased estimators in the case of absolutely continuous distributions. Statistical methods of estimation and testing of hypotheses. *J. Math. Sci.* 119(3):336–341.
- Petrov, V. V. 1972. *Summy nezavisimyykh sluchaynykh velichin* [Sums of independent random variables]. Moscow: Nauka. 416 p.
- Barndorff-Nielsen, O. E., and D. R. Cox. 1989. *Asymptotic techniques for use in statistics*. London: Chapman and Hall. 252 p.
- Lumel'skiy, Ya. P., and P. N. Sapozhnikov. 1969. Nesmeshchennye otsenki dlya plotnostey raspredeleniy [Unbiased estimators for distribution densities]. *Teoriya veroyatnostey i ee primeneniya* [Theory Probab. Appl.] 14(2):372–380.
- Chichagov, V. V. 2008. Stokhasticheskie razlozheniya nesmeshchennykh otsenok v sluchae odnoparametricheskogo eksponentsial'nogo semeystva [Stochastic expansions of unbiased estimators for the case of one-parameter exponential family]. *Informatika i ee primeneniya — Inform. Appl.* 2(2):62–70.
- Chichagov, V. V. 2012. O nesmeshchennoy otsenke veroyatnosti $P(X < Y)$ v modeli nagruzka–prochnost' [An unbiased estimator of the probability $P(X < Y)$ in stress–strength model]. *Teoriya veroyatnostey i ee prilozheniya: Tezisy dokl. Mezhdunar. konf., posvyashchennoy 100-letiyu so dnya rozhdeniya B. V. Gnedenko* [Conference (International) “Probability Theory and Its Applications” in Commemoration of the Centennial of B. V. Gnedenko]. Moscow: LENAND. 264.
- Chichagov, V. 2013. Asymptotic of the mean absolute error of UNVUE and MLE in the case of one-parameter exponential family lattice distributions. *31st Seminar (International) on Stability Problems for Stochastic Models: Book of Abstracts*. Moscow. 13–15.
- Jørgensen, B. 1997. *The theory of dispersion models*. London: Chapman and Hall/CRC. 256 p.
- Brown, L. D. 1986. *Fundamentals of statistical exponential families with applications in statistical decision theory*. Lecture notes — monograph ser. Hayward, CA: Institute of Mathematical Statistics. Vol. 9. 284 p.
- Petrov, V. V. 1987. *Predel'nye teoremy dlya summ nezavisimyykh sluchaynykh velichin* [Limit theorems for sums of independent random variables]. Moscow: Nauka. 320 p.

Received April 10, 2014

Contributor

Chichagov Vladimir V. (b. 1955) — Candidate of Science (PhD) in physics and mathematics, associate professor, Perm State University, 15 Bukireva Str., Perm 614990, Russian Federation; chichagov@psu.ru

РЕАЛИЗУЕМОСТЬ ВЕРОЯТНОСТНЫХ РЕАКЦИЙ КОНЕЧНЫМИ ВЕРОЯТНОСТНЫМИ АВТОМАТАМИ*

А. М. Миронов¹

Аннотация: Формулируется и доказывается критерий реализуемости вероятностной реакции (ВР) конечным вероятностным автоматом (ВА). Этот критерий усиливает соответствующий критерий Р. Г. Бухараева и Х. Хомута, который имеет следующий вид: ВР реализуется в конечном ВА тогда и только тогда, когда существует конечное опорное множество для множества состояний этой ВР, выпуклое относительно полугруппы всех вращений. Сформулированный и доказанный в настоящей работе критерий реализуемости ВР в конечном ВА не связан с построением множества состояний для ВР и имеет следующий вид: ВР реализуема конечным ВА тогда и только тогда, когда она принадлежит конусу над некоторым конечным множеством ВР, устойчивому относительно сдвигов. Доказательство этого критерия имеет существенно более простой вид, чем доказательство критерия Бухараева–Хомута.

Ключевые слова: вероятностные автоматы; вероятностные реакции; случайные функции

DOI: 10.14357/19922264150309

1 Введение

1.1 Понятие вероятностного автомата

Понятие **вероятностного автомата** впервые было сформулировано в 1963 г. в основополагающей работе М. Рабина [1]. Данное понятие возникло как синтез понятий конечного детерминированного автомата [2] и цепи Маркова [3] и было предназначено для построения математических моделей динамических систем, в которых присутствует неопределенность, описываемая статистическими закономерностями. Эта неопределенность связана:

- с неточностью знаний о состояниях, в которых моделируемые системы находятся в процессе своего функционирования;
- с недетерминированностью правил изменения этих состояний.

Неопределенность в ВА может быть вызвана различными причинами, которые подразделяются на два класса.

1. Причины из первого класса связаны с природой системы, моделируемой ВА. К ним относятся:

- влияние случайных факторов на функционирование системы, например: случайные сбои компонентов системы или отказы в их работе, случайное изменение условий

функционирования анализируемой системы, случайность потока заявок в системе массового обслуживания и т. п.;

- несовершенство (или невозможность) точного измерения состояний этой системы.

2. Второй класс причин связан с преднамеренным внесением неточности и неопределенности в математические модели анализируемых систем. Это делается в тех случаях, когда точные модели анализируемых систем имеют неприемлемо высокую сложность и проведение анализа поведения таких систем возможно только с использованием их упрощенных моделей, в которых некоторые компоненты состояний этих систем игнорируются. В частности, анализ поведения сложной программной системы (например, операционной системы компьютера) в большинстве случаев возможен только с использованием таких упрощенных математических моделей этих систем, в которых принимаются во внимание значения лишь некоторых программных переменных, от которых существенно зависит поведение анализируемой программной системы.

Как правило, моделирование систем при помощи ВА производится:

- либо с целью анализа свойств этих систем (к числу которых относятся, например, корректность, безопасность, надежность, устой-

* Работа выполнена при частичной поддержке РФФИ (проект 12-07-00109).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, amironov66@gmail.com

- чивость функционирования в непредусмотренных ситуациях и др.),
- либо с целью вычисления различных количественных характеристик анализируемых систем, среди которых могут быть, например, следующие:
 - частота выполнения тех или иных действий или переходов в анализируемых системах;
 - вероятность отказа компонентов анализируемых систем;
 - вероятность вторжения злоумышленника в компьютерную сеть;
 - математическое ожидание времени отклика веб-сервиса.

1.2 Исторический обзор

Первоначальное понятие ВА, введенное в работе М. Рабина [1], было предназначено главным образом для изучения вопросов представимости регулярных языков вероятностными автоматами. Затем оно было обобщено до такого понятия, которое позволило моделировать вероятностные преобразователи информации. Определение ВА в общей форме было введено независимо в работах Дж. Карлайла [4], Р. Г. Бухараева [5] и П. Штарке [6].

С начала возникновения понятия ВА исследовательская деятельность в этой области отличалась высокой активностью. Результаты первых лет исследований в области ВА были систематизированы в книге [7]. Подробный список (около 500) ссылок на работы с наиболее существенными теоретическими и практическими результатами по ВА, полученными до 1985 г., можно найти в фундаментальной монографии Р. Г. Бухараева [8], которую можно рассматривать как итог первого периода развития теории ВА, продолжавшегося более двух десятилетий.

В последующие годы произошло некоторое снижение активности исследований в этой области, но в настоящее время теория ВА вновь находится в состоянии подъема. Возрождение исследовательской активности в области ВА в значительной степени связано с тем, что в связи с бурным развитием современных информационных технологий возник широкий круг новых задач, в решении которых ВА могут служить эффективным инструментом. К числу таких задач относятся задачи в следующих областях:

- верификация программ и протоколов передачи данных в компьютерных сетях;
- информационный поиск в Интернете;

- финансово-экономический анализ;
- обработка и извлечение знаний из больших массивов данных (data mining и process mining); в частности в задачах анализа бизнес-процессов, биоинженерии и биоинформатики;
- извлечение смысла из текстов на естественных языках;
- машинное зрение и обработка изображений и др.

Началом современного этапа развития теории ВА можно считать работу [9], в которой рассмотрены ВА, возникающие при моделировании параллельных вычислительных систем с асинхронным взаимодействием. В качестве вводных текстов в современную теорию ВА можно назвать работы [10, 11].

Главное отличие нового понятия ВА от того, которое изучалось в предшествующий период, заключается в том, что в новом понимании ВА определяется как **система переходов** (transition system), с которой связано некоторое множество переменных. Вероятностный автомат функционирует путем выполнения переходов, после каждого из которых происходит обновление значений переменных этого ВА. Можно доказать, что если множество переменных ВА конечно и множества значений этих переменных тоже конечны, то новое и старое понятия ВА будут эквивалентны.

Наряду с упомянутыми выше понятиями ВА существуют и другие модели динамических систем со случайным поведением, например скрытые марковские модели (hidden Markov models) [12], байесовские сети (Bayesian networks) [13], вероятностные графические модели [14], марковские решающие процессы (Markov decision processes) [15], вероятностные I/O автоматы (probabilistic I/O automata) [16]. Все эти модели являются частными случаями исходного понятия ВА общего вида [8].

Наряду с перечисленными выше моделями в последние годы изучаются модели динамических систем со случайным поведением, переходы в которых могут быть ассоциированы не только с вероятностями их выполнения, но и с модальностями must и may, которые позволяют существенно усилить выразительные возможности этих моделей по сравнению с другими упомянутыми выше моделями. Основные концепции и методы, относящиеся к таким моделям, содержатся в статье [17].

Также изучаются и другие обобщения понятия ВА, в частности вероятностные сети Петри [18, 19], ВА с непрерывным временем [20], вероятностные процессные алгебры [21].

1.3 Содержание статьи

Главный результат настоящей работы (теорема 4) представляет собой усиление критерия реализуемости многотактного канала в конечном ВА, полученного независимо Х. Хомутом [22] и Р. Г. Бухараевым [23].

Понятие **многотактного канала** (МК) было введено независимо Р. Г. Бухараевым [5] и П. Штарке [6] соответственно. Это понятие служит характеристикой внешнего поведения ВА и представляет собой вероятностную зависимость между входными сигналами, поступающими в ВА, и выходными сигналами, выдаваемыми этим ВА. Бухараев [5] и П. Штарке [6] ввели условие на МК, которое они называют условием автоматности. Автор настоящей работы счел целесообразным ввести новое понятие — **вероятностной реакции**. Можно доказать эквивалентность понятий ВР и МК, удовлетворяющего условию автоматности. Основным результатом настоящей работы формулируется в терминах ВР.

Суть критерия Бухараева—Хомута реализуемости МК в конечном ВА [8, теорема 2.2.1, с. 50] заключается в следующем. Вводятся понятия множества S_f состояний МК f и опорного множества для S_f . Критерий имеет следующий вид: МК f , удовлетворяющий условию автоматности, реализуется в конечном ВА тогда и только тогда, когда существует конечное опорное множество для S_f , выпуклое относительно полугруппы всех вращений. Проверка данного критерия представляется затруднительной по той причине, что для этого необходимо вычислить все множество S_f , что сложно сделать в том случае, когда данное множество бесконечно.

Доказательство этого критерия основано на представлении МК как точек счетномерного линейного пространства с индексацией координат парами элементов свободных полугрупп над входным и выходным алфавитами. В доказательстве используются бесконечномерные матрицы, индексированные парами, компонентами которых являются пары элементов этих свободных полугрупп.

Сформулированный и доказанный в настоящей работе критерий реализуемости ВР в конечном ВА не связан с построением бесконечного множества состояний для ВР и имеет следующий вид: ВР реализуема конечным ВА тогда и только тогда, когда она принадлежит конусу над некоторым конечным множеством ВР, устойчивому относительно сдвигов. Доказательство этого критерия имеет существенно более простой вид, чем доказательство критерия Бухараева—Хомута, и не использует бесконечномерных матриц.

1.4 Мотивация и актуальность работы

Проблема реализуемости ВР в конечных ВА является вероятностным обобщением известной проблемы реализуемости отображений реакции в конечных детерминированных автоматах, т. е. проблемы построения автоматов по информации о выходных сигналах, которые они выдают в ответ на входные воздействия. Данная проблема была самой первой проблемой, относящейся к теории автоматов (первой работой, посвященной данной проблеме, является [24]). На всем протяжении развития теории автоматов специализация данной проблемы на различные классы автоматов играла ключевую роль в развитии теории автоматов.

В настоящее время важность проблемы реализуемости ВР в классе ВА обосновывается значительным числом актуальных прикладных задач, приводящих к данной проблеме. Например, большой класс задач распознавания образов сводится к задаче построения распознающей системы (в качестве которой может выступать, например, ВА), реакция которой на заданный класс входных воздействий должна удовлетворять некоторым условиям (например, система должна правильно реагировать на некоторые входные воздействия с заданной вероятностью ошибки). Математической формализацией данной задачи является проблема построения ВА, реакция которого обладает заданными свойствами. Иногда не существует ни одного ВА, реакция которого обладает заданными свойствами. Поэтому еще до синтеза ВА, соответствующего некоторой ВР, необходимо обосновать существование такого ВА. Доказываемая в настоящей работе теорема 4 связана именно с этой задачей. Теорема 4 сводит проблему обоснования существования ВА, реакция которого совпадает с заданной ВР f , к проблеме нахождения конечного множества Γ_f ВР, обладающего свойствами, упомянутыми в конце п. 1.3. Из этого результата не вытекает существования алгоритма построения множества Γ_f в том случае, когда это множество существует. Не исключено, что задача нахождения множества Γ_f является алгоритмически неразрешимой. Однако достоинством теоремы 4 является то, что она сужает направление поисков в решении задачи распознавания реализуемости заданной ВР f .

Отметим, что с момента появления в начале 1970-х гг. критерия Бухараева—Хомута реализуемости ВР не появилось никаких результатов, связанных с усилением данного критерия. В недавней работе Бухараева [25] этот критерий упомянут в той же самой формулировке, в которой он был приведен в [23, 22].

2 Вспомогательные понятия

2.1 Случайные функции и распределения

Пусть задана пара множеств X, Y . **Случайной функцией (СФ)** из X в Y называется произвольная функция f вида

$$f : X \times Y \rightarrow [0, 1],$$

такая что $\forall x \in X$ множество $\{y \in Y \mid f(x, y) > 0\}$ конечно или счетно и $\forall x \in X \sum_{y \in Y} f(x, y) = 1$.

Для любых $x \in X$ и $y \in Y$ значение $f(x, y)$ можно интерпретировать как вероятность того, что СФ f отображает x в y .

Если f — СФ из X в Y , то будем обозначать этот факт записью $f : X \xrightarrow{r} Y$. Будем называть X **областью определения** СФ f , а Y — **областью значений** СФ f .

Случайная функция называется **конечной (КСФ)**, если ее область определения и область значений являются конечными множествами.

Пусть задана КСФ $f : X \xrightarrow{r} Y$ и на X и Y заданы упорядочения их элементов, которые имеют вид (x_1, \dots, x_m) и (y_1, \dots, y_n) соответственно. Тогда f можно представить в виде матрицы (обозначаемой тем же символом f):

$$f = \begin{pmatrix} f(x_1, y_1) & \cdots & f(x_1, y_n) \\ \vdots & & \vdots \\ f(x_m, y_1) & \cdots & f(x_m, y_n) \end{pmatrix}. \quad (1)$$

Ниже будем отождествлять каждую КСФ f с соответствующей ей матрицей (1). Будем предполагать, что для каждого множества X , являющегося областью определения или областью значений какой-либо из рассматриваемых КСФ, на X задано фиксированное упорядочение его элементов. Таким образом, для каждой рассматриваемой КСФ соответствующая ей матрица определена однозначно.

Вероятностным распределением (или просто **распределением**) на множестве X называется СФ вида $\xi : \mathbf{1} \xrightarrow{r} X$, где $\mathbf{1}$ — множество, состоящее из одного элемента, который будем обозначать символом e . Совокупность всех распределений на X будем обозначать записью X^Δ . Для каждого $x \in X$ и каждого $\xi \in X^\Delta$ значение $\xi(e, x)$ будем обозначать более коротко записью x^ξ . Для каждого $x \in X$ будем обозначать записью ξ_x распределение из X^Δ , определяемое следующим образом: $\forall y \in X \quad y^{\xi_x} \stackrel{\text{def}}{=} 1$, если $y = x$, и $y^{\xi_x} \stackrel{\text{def}}{=} 0$, если $y \neq x$.

2.2 Строки

Для каждого множества X будем обозначать записью X^* совокупность всех конечных строк, компонентами которых являются элементы X . Множество X^* содержит **пустую строку**, она обозначается символом ε . Для каждого $x \in X$ строка, состоящая из одного этого элемента, обозначается той же записью x . Для каждой строки $u \in X^*$ ее **длиной** называется число компонентов этой строки. Длина пустой строки равна нулю. Длина строки u обозначается записью $|u|$. Для каждой пары строк $u, v \in X^*$ их **конкатенацией** называется строка, обозначаемая записью uv и определяемая следующим образом: $u\varepsilon \stackrel{\text{def}}{=} \varepsilon u \stackrel{\text{def}}{=} u$, и если $u = x_1 \cdots x_n$ и $v = x'_1 \cdots x'_m$, то $uv \stackrel{\text{def}}{=} x_1 \cdots x_n x'_1 \cdots x'_m$.

3 Вероятностные автоматы

3.1 Понятие вероятностного автомата

Вероятностный автомат — это пятерка A вида

$$A = (X, Y, S, P, \xi^0), \quad (2)$$

компоненты которой имеют следующий смысл:

- (1) X, Y и S — конечные множества, элементы которых называются соответственно **входными сигналами, выходными сигналами и состояниями** ВА A ;
- (2) P — СФ вида $P : S \times X \xrightarrow{r} S \times Y$, называемая **поведением** ВА A . $\forall (s, x, s', y) \in S \times X \times S \times Y$ значение $P(s, x, s', y)$ понимается как вероятность того, что если в текущий момент времени (t) A находится в состоянии s и в этот момент времени на его вход поступил сигнал x , то в следующий момент времени ($t+1$) A будет находиться в состоянии s' и в момент времени t выходной сигнал A равен y ;
- (3) ξ^0 — распределение на S , называемое **начальным распределением** ВА A . $\forall s \in S$ значение s^{ξ^0} понимается как вероятность того, что в начальный момент времени ($t = 0$) ВА A находится в состоянии s .

Пусть A — ВА вида (2) и упорядочение множества S его состояний имеет вид (s_1, \dots, s_n) . Для любых $x \in X$ и $y \in Y$ будем обозначать записью A^{xy} матрицу порядка n

$$\begin{pmatrix} P(s_1, x, s_1, y) & \cdots & P(s_1, x, s_n, y) \\ \vdots & & \vdots \\ P(s_n, x, s_1, y) & \cdots & P(s_n, x, s_n, y) \end{pmatrix} \quad (3)$$

и для любой пары строк $u \in X^*$, $v \in Y^*$ будем обозначать записью $A^{u,v}$ (запятая в этой записи может опускаться) матрицу порядка n , определяемую следующим образом:

- $A^{\varepsilon,\varepsilon} = E$ (единичная матрица),
- если $|u| \neq |v|$, то $A^{u,v} = 0$ (нулевая матрица),
- если $u = x_1 \cdots x_k$ и $v = y_1 \cdots y_k$, то $A^{u,v} = A^{x_1 y_1} \cdots A^{x_k y_k}$.

Для любых $s, s' \in S$ будем обозначать записью $A_{s,s'}^{u,v}$ коэффициент матрицы $A^{u,v}$, находящийся в строке, соответствующей состоянию s , и в столбце, соответствующем состоянию s' .

Если строки $u \in X^*$ и $v \in Y^*$ имеют вид $x_0 \cdots x_k$ и $y_0 \cdots y_k$ соответственно, то $A_{s,s'}^{u,v}$ можно понимать как вероятность того, что если в текущий момент (t) A находился в состоянии s и начиная с этого момента на вход A последовательно поступали элементы строки u (т.е. в момент t поступил сигнал x_0 , в момент $t+1$ поступил сигнал x_1 и т.д.), то в моменты $t, t+1, \dots, t+k$ выходные сигналы A равны y_0, \dots, y_k соответственно и в момент $t+k+1$ A будет находиться в состоянии s' .

Для каждого ВА $A = (X, Y, S, P, \xi^0)$ **графовым представлением** ВА A называется граф Γ_A с множеством вершин S , каждая вершина s которого помечена числом s^{ξ^0} и ребра которого помечены тройками вида (x, y, a) , где $x \in X$, $y \in Y$, $a \in [0, 1]$; если $A_{s,s'}^{xy} = a > 0$, то Γ_A содержит ребро из s в s' с меткой (x, y, a) .

Данное графовое представление ВА полностью аналогично графовому представлению обычных автоматов. В частности, для ВА тоже можно ввести понятие **вычисления (computation)**, которое представляет собой путь в графе Γ_A . Если путь в графе Γ_A имеет вид:

$$s_0 \xrightarrow{(x_1, y_1, a_1)} s_1 \xrightarrow{(x_2, y_2, a_2)} \dots \xrightarrow{(x_n, y_n, a_n)} s_n,$$

то произведение $a_1 \cdots a_n$ можно интерпретировать как вероятность того, что при подаче на вход ВА A слова $x_1 \cdots x_n$ соответствующее выходное слово будет иметь вид $y_1 \cdots y_n$.

Для каждого $y \in Y$ **принимаящим вычислением (accepting computation)** относительно сигнала y является произвольное вычисление в графе Γ_A , имеющее вид:

$$s_0 \xrightarrow{(x_1, y_1, a_1)} s_1 \xrightarrow{(x_2, y_2, a_2)} \dots \xrightarrow{(x_n, y_n, a_n)} s_n \xrightarrow{(x, y, a)} s_{n+1}.$$

3.2 Реакция вероятностного автомата

Пусть задан ВА A вида (2) и распределение $\xi \in S^\Delta$. Будем говорить, что **ВА A в момент времени t**

имеет распределение ξ , если для каждого состояния $s \in S$ вероятность того, что A в момент времени t находится в состоянии s , равна s^ξ . **Реакцией ВА A в распределении ξ** называется функция

$$A^\xi : X^* \times Y^* \rightarrow \mathbb{R},$$

определяемая следующим образом:

$$\forall u \in X^*, \forall v \in Y^* \quad A^\xi(u, v) \stackrel{\text{def}}{=} \xi A^{u,v} I,$$

где запись I обозначает вектор-столбец порядка $|S|$, все компоненты которого равны 1.

Реакцией ВА A будем называть реакцию этого ВА в его начальном распределении. Будем обозначать реакцию ВА A записью f_A . Если строки $u \in X^*$ и $v \in Y^*$ имеют вид $x_0 \cdots x_k$ и $y_0 \cdots y_k$ соответственно, то $f_A(u, v)$ можно понимать как вероятность того, что если начиная с момента 0 на вход A последовательно поступали элементы строки u (т.е. в момент 0 поступил сигнал x_0 , в момент 1 поступил сигнал x_1 и т.д.), то в моменты $0, 1, \dots, k$ выходные сигналы A равны y_0, \dots, y_k соответственно.

Теорема 1. Если A — ВА вида (2) и $\xi \in S^\Delta$, то A^ξ — СФ.

Доказательство. Поскольку $\forall u \in X^*, \forall v \in Y^* \quad A^\xi(u, v) \geq 0$, то для доказательства теоремы достаточно доказать, что

$$\forall u \in X^* \quad \sum_{v \in Y^*} A^\xi(u, v) = 1,$$

т.е.

$$\forall u \in X^* \quad \sum_{v \in Y^*} \xi A^{u,v} I = 1. \quad (4)$$

Поскольку $A^{u,v} = 0$ при $|u| \neq |v|$, то (4) эквивалентно условию:

$$\forall k \geq 0, \forall u \in X^k \quad \sum_{v \in Y^k} \xi A^{u,v} I = 1. \quad (5)$$

Докажем (5) индукцией по k . Если $k = 0$, то (5) следует из того, что $A^{\varepsilon,\varepsilon} = E$ и $\xi E I = \xi I = 1$ (так как $\xi \in S^\Delta$).

Пусть (5) верно для некоторого k . Докажем, что

$$\forall u \in X^{k+1} \quad \sum_{v \in Y^{k+1}} \xi A^{u,v} I = 1. \quad (6)$$

Соотношение (6) эквивалентно соотношению

$$\forall u \in X^k, \forall x \in X \quad \sum_{v \in Y^k, y \in Y} \xi A^{u,xy} I = 1. \quad (7)$$

Так как $A^{u,x,vy} = A^{u,v}A^{xy}$, то (7) можно переписать в виде

$$\forall u \in X^k, \forall x \in X \quad \sum_{v \in Y^k} \xi A^{u,v} \left(\sum_{y \in Y} A^{xy} I \right) = 1. \quad (8)$$

Соотношение (8) следует из (5) и из равенства

$$\sum_{y \in Y} A^{xy} I = I, \quad (9)$$

которое верно потому, что если A^{xy} имеет вид (3), то $\forall i \in \{1, \dots, n\}$ элемент с индексом i столбца $\sum_{y \in Y} A^{xy} I$ равен сумме

$$\sum_{y \in Y, j=1, \dots, n} P(s_i, x, s_j, y),$$

которая равна 1, так как P есть СФ вида:

$$P : S \times X \xrightarrow{r} S \times Y. \quad \square$$

4 Вероятностные реакции

4.1 Понятие вероятностной реакции

Пусть X и Y — конечные множества.

Вероятностной реакцией из X в Y называется СФ $f : X^* \xrightarrow{r} Y^*$, удовлетворяющая условию: $\forall u \in X^*, \forall v \in Y^*$,

$$\left. \begin{array}{l} \text{если } |u| \neq |v|, \text{ то } f(u, v) = 0, \\ \forall x \in X \quad f(u, v) = \sum_{y \in Y} f(ux, vy). \end{array} \right\} \quad (10)$$

Запись $R(X, Y)$ обозначает совокупность всех ВР из X в Y .

Теорема 2. Множество $R(X, Y)$ замкнуто относительно выпуклых комбинаций, т. е. если $f_1, \dots, f_n \in R(X, Y)$ и $a_1, \dots, a_n \in [0, 1]$, где $\sum_{i=1}^n a_i = 1$, то

$$\sum_{i=1}^n a_i f_i \in R(X, Y).$$

Доказательство. $\sum_{i=1}^n a_i f_i$ — СФ, поэтому

$$\begin{aligned} & - \forall u \in X^*, \forall v \in Y^*, \text{ если } |u| \neq |v|, \text{ то} \\ & \quad \forall i = 1, \dots, n \quad f_i(u, v) = 0 \Rightarrow \sum_{i=1}^n a_i f_i(u, v) = 0; \\ & - \forall x \in X \quad \sum_{y \in Y} \sum_{i=1}^n a_i f_i(ux, vy) = \sum_{i=1}^n a_i \times \\ & \quad \times \sum_{y \in Y} f_i(ux, vy) = \sum_{i=1}^n a_i f_i(u, v) = f(u, v). \quad \square \end{aligned}$$

Теорема 3. Для каждого ВА $A = (X, Y, S, P, \xi^0)$ и каждого $\xi \in S^\Delta$

$$A^\xi \in R(X, Y).$$

Доказательство. Докажем, что $\forall u \in X^*, \forall v \in Y^*$ СФ $f \stackrel{\text{def}}{=} A^\xi$ удовлетворяет условию (10):

- если $|u| \neq |v|$, то $A^{u,v} = 0$; поэтому $A^\xi(u, v) = \xi A^{u,v} I = 0$,
- $\forall x \in X$

$$\begin{aligned} \sum_{y \in Y} A^\xi(ux, vy) &= \sum_{y \in Y} \xi A^{ux,vy} I = \\ &= \sum_{y \in Y} \xi A^{u,v} A^{xy} I = \xi A^{u,v} \left(\sum_{y \in Y} A^{xy} I \right) = \\ &= \xi A^{u,v} I = A^\xi(u, v) \quad (11) \end{aligned}$$

(в (11) используется равенство (9)). \square

4.2 Реализуемость вероятностных реакций

Вероятностная реакция f называется **реализуемой**, если \exists ВА $A : f_A = f$.

Пусть X и Y — конечные множества. Будем использовать следующие определения и обозначения:

- запись $[0, 1]^{X^* \times Y^*}$ обозначает множество функций вида

$$f : X^* \times Y^* \rightarrow [0, 1];$$

- для каждого $\Gamma \subseteq [0, 1]^{X^* \times Y^*}$ **конусом** над Γ называется подмножество $C_0(\Gamma) \subseteq [0, 1]^{X^* \times Y^*}$, состоящее из функций вида $\sum_{i=1}^n a_i f_i$, где

$$\bullet \forall i = 1, \dots, n \quad a_i \in [0, 1], f_i \in \Gamma, \sum_{i=1}^n a_i \leq 1;$$

$$\bullet \forall (u, v) \in X^* \times Y^* \quad \left(\sum_{i=1}^n a_i f_i \right) (u, v) \stackrel{\text{def}}{=} \sum_{i=1}^n a_i f_i(u, v);$$

- $\forall x \in X, \forall y \in Y$ запись D^{xy} обозначает отображение вида

$$D^{xy} : [0, 1]^{X^* \times Y^*} \rightarrow [0, 1]^{X^* \times Y^*},$$

называемое **сдвигом** и сопоставляющее каждой функции f из $[0, 1]^{X^* \times Y^*}$ функцию, обозначаемую записью $f D^{xy}$, где

$$\forall u \in X^*, \forall v \in Y^* \quad (f D^{xy})(u, v) \stackrel{\text{def}}{=} f(xu, yv); \quad (12)$$

– подмножество $\Gamma \subseteq [0, 1]^{X^* \times Y^*}$ называется **устойчивым относительно сдвигов**, если

$$\forall f \in \Gamma, \forall x \in X, \forall y \in Y \quad fD^{xy} \in C_0(\Gamma).$$

Теорема 4. Пусть X и Y — конечные множества и $f \in R(X, Y)$. Следующие условия эквивалентны:

- f реализуема;
- существует конечное множество $\Gamma_f \subseteq R(X, Y)$, устойчивое относительно сдвигов и такое, что $f \in C_0(\Gamma_f)$.

Доказательство. Пусть f реализуема, т. е. \exists ВА $A = (X, Y, S, P, \xi^0)$:

$$\forall u \in X^*, \forall v \in Y^* \quad f(u, v) = \xi^0 A^{u,v} I.$$

$\forall s \in S$ обозначим записью A_s ВА (X, Y, S, P, ξ_s) . В качестве искомого Γ_f можно взять множество $\{f_{A_s} \mid s \in S\}$.

$f \in C_0(\Gamma_f)$, так как $f = \sum_{s \in S} s \xi^0 f_{A_s}$ и $\Gamma_f \subseteq \subseteq R(X, Y)$ (по теореме 3).

Докажем, что Γ_f устойчиво относительно сдвигов, т. е. $\forall s \in S, \forall x \in X, \forall y \in Y \quad f_{A_s} D^{xy} \in C_0(\Gamma_f)$. Согласно (12),

$$\begin{aligned} \forall u \in X^*, \forall v \in Y^* \quad (f_{A_s} D^{xy})(u, v) &= \\ &= f_{A_s}(xu, yv) = \xi_s A^{xu, yv} I = \xi_s A^{xy} A^{u,v} I. \end{aligned}$$

Нетрудно видеть, что

$$\xi_s A^{xy} A^{u,v} I = \sum_{s' \in S} a_{s'} f_{A_{s'}}(u, v),$$

где $\forall s' \in S \quad a_{s'}$ — компонента вектор-строки $\xi_s A^{xy}$, соответствующая состоянию s' (т. е. элемент матрицы A^{xy} , находящийся в строке s и столбце s'). Свойства $\forall s' \in S \quad a_{s'} \in [0, 1]$ и $\sum_{s' \in S} a_{s'} \leq 1$ являются следствием соответствующих свойств матрицы A^{xy} .

Обратно, пусть $f \in C_0(\Gamma_f)$, где $\Gamma_f = \{f_1, \dots, f_n\} \subseteq R(X, Y)$ и Γ_f устойчиво относительно сдвигов. Определим A как ВА

$$A \stackrel{\text{def}}{=} (X, Y, S, P, \xi^0), \quad (13)$$

компоненты которого имеют следующий вид:

- $S \stackrel{\text{def}}{=} \{1, \dots, n\}$;
- $\xi^0 = (a_1, \dots, a_n)$, где a_1, \dots, a_n — коэффициенты представления f в виде суммы

$$f = \sum_{i=1}^n a_i f_i, \quad (14)$$

$$\text{где } \forall i = 1, \dots, n \quad a_i \geq 0, \sum_{i=1}^n a_i \leq 1.$$

По предположению, $f \in R(X, Y)$, в частности $f(\varepsilon, \varepsilon) = 1$, откуда следует равенство $\sum_{i=1}^n a_i = 1$, поэтому $\xi^0 \in S^\Delta$;

- поведение $P : S \times X \times S \times Y \rightarrow [0, 1]$ ВА (13) определяется матрицами A^{xy} порядка n ($x \in X, y \in Y$):

$$P(i, x, j, y) \stackrel{\text{def}}{=} A_{ij}^{xy},$$

где $\forall x \in X, \forall y \in Y, \forall i = 1, \dots, n$ строка i матрицы A^{xy} состоит из коэффициентов a_{i1}, \dots, a_{in} представления функции $f_i D^{xy}$ в виде суммы $\sum_{j=1}^n a_{ij} f_j$ ($\forall i, j \quad a_{ij} \geq 0, \sum_{j=1}^n a_{ij} \leq 1$).

Докажем, что P является СФ вида $S \times X \times S \times Y$. Данное утверждение эквивалентно соотношению $\left(\sum_{y \in Y} A^{xy} \right) I = I$.

$\forall i = 1, \dots, n$ из

$$f_i D^{xy} = \sum_{j=1}^n A_{ij}^{xy} f_j \quad (15)$$

следует, что

$$(f_i D^{xy})(\varepsilon, \varepsilon) = \sum_{j=1}^n A_{ij}^{xy} f_j(\varepsilon, \varepsilon). \quad (16)$$

Так как $\forall i = 1, \dots, n \quad f_j \in R(X, Y)$, то $f_j(\varepsilon, \varepsilon) = 1$. Кроме того, левая часть (16) равна $f_i(x, y)$. Поэтому (16) можно переписать в виде $f_i(x, y) = \sum_{j=1}^n A_{ij}^{xy}$, откуда следует соотношение

$$\sum_{y \in Y} f_i(x, y) = \sum_{y \in Y} \sum_{j=1}^n A_{ij}^{xy}. \quad (17)$$

Так как $f_i \in R(X, Y)$, то согласно второму соотношению в (10) левая часть (17) равна $f_i(\varepsilon, \varepsilon)$, т. е. равна 1. Учитывая это и меняя порядок суммирования в правой части (17), получаем соотношение

$$\sum_{j=1}^n \sum_{y \in Y} A_{ij}^{xy} = 1. \quad (18)$$

Нетрудно видеть, что истинность (18) $\forall i = 1, \dots, n$ эквивалентна доказываемому равенству $\left(\sum_{y \in Y} A^{xy} \right) I = I$.

Докажем, что реакция ВА (13) совпадает с f , т. е.

$$\forall u \in X^*, \forall v \in Y^* \quad \xi^0 A^{u,v} I = f(u, v). \quad (19)$$

Если $|u| \neq |v|$, то левая часть равенства в (19) равна 0 по определению матриц вида $A^{u,v}$ и правая часть

равенства в (19) равна 0 согласно предположению $f \in R(X, Y)$ и первому соотношению в (10).

Пусть $|u| = |v|$. Докажем (индукцией по $|u|$), что

$$A^{u,v}I = \begin{pmatrix} f_1(u, v) \\ \vdots \\ f_n(u, v) \end{pmatrix}. \quad (20)$$

Если $u = v = \varepsilon$, то обе части (20) равны I .

Если $u = xu'$ и $v = yv'$, то, предполагая верным равенство (20), в котором u и v заменены на u' и v' , имеем

$$\begin{aligned} A^{u,v}I &= A^{xu',yv'}I = A^{xy}A^{u',v'}I = \\ &= A^{xy} \begin{pmatrix} f_1(u', v') \\ \vdots \\ f_n(u', v') \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n A_{1j}^{xy} f_j(u', v') \\ \vdots \\ \sum_{i=1}^n A_{nj}^{xy} f_j(u', v') \end{pmatrix}. \end{aligned} \quad (21)$$

Из (15) следует, что правую часть в (21) можно переписать в виде

$$\begin{pmatrix} (f_1 D^{xy})(u', v') \\ \vdots \\ (f_n D^{xy})(u', v') \end{pmatrix} \quad (22)$$

Согласно определению (12) функций вида $f D^{xy}$, столбец (22) совпадает с правой частью доказываемого равенства (20).

Таким образом, равенство (20) доказано. Согласно этому равенству левая часть доказываемого равенства (19) равна

$$\xi^0 \begin{pmatrix} f_1(u, v) \\ \vdots \\ f_n(u, v) \end{pmatrix} = \sum_{i=1}^n \xi_i^0 f_i(u, v). \quad (23)$$

По определению ξ^0 (см. (14)) правая часть (23) равна $f(u, v)$, т. е. правой части доказываемого равенства (19). \square

5 Примеры реализуемых и нереализуемых вероятностных реакций

5.1 Пример реализуемой вероятностной реакции

В этом пункте приводится пример ВР, реализуемость которой обосновывается теоремой 4.

Пусть $X = \{x\}$, $Y = \{y_1, y_2\}$ и реакция $f : X^* \times Y^* \rightarrow [0, 1]$ имеет следующий вид:

$$f(\varepsilon, \varepsilon) \stackrel{\text{def}}{=} 1$$

и $\forall u \in X^*, \forall v \in Y^*$,

- если $|u| \neq |v|$, то $f(u, v) = 0$;
- если $|u| = |v| > 0$, то v можно представить в виде одной из следующих конкатенаций:

$$\left. \begin{aligned} &y_1^{p_1} \quad (p_1 > 0); \\ &y_2^{q_1} \quad (q_1 > 0); \\ &y_1^{p_1} y_2^{q_1} \quad (p_1 > 0, q_1 > 0); \\ &y_1^{p_1} y_2^{q_1} \cdots y_k^{p_k} y_2^{q_k} \quad \left(\begin{array}{l} k > 1, p_1 \geq 0, q_k \geq 0, \\ \text{остальные } p_i, q_i > 0 \end{array} \right), \end{aligned} \right\} \quad (24)$$

где $\forall n \geq 0 y_i^n$ является строкой из n символов y .

Обозначим $p \stackrel{\text{def}}{=} \sum_{i=1}^k p_i$, $q \stackrel{\text{def}}{=} \sum_{i=1}^k q_i$ и определим

$$f(u, v) \stackrel{\text{def}}{=} c \left(\frac{1}{5} \right)^p \left(\frac{4}{5} \right)^q, \quad (25)$$

где число c определяется в соответствии с видом v из перечисления (24), т. е. c равно соответственно

$$\left. \begin{aligned} &1 + \frac{4p_1}{3}; \\ &\frac{2}{3}; \\ &\frac{2}{3} + \frac{4p_1}{3}; \\ &\text{если } p_1 = 0, \text{ то } \frac{2}{3}, \text{ иначе } \frac{2}{3} + \frac{4p_1}{3} \end{aligned} \right\} \quad (26)$$

(строки в (26) соответствуют строкам в (24)).

Определим $\Gamma_f \stackrel{\text{def}}{=} \{f_1, f_2\}$, где $f_i(\varepsilon, \varepsilon) = 1$ ($i = 1, 2$) и

- f_1 имеет такой же вид, как f в (25), но число c определяется иначе и равно соответственно

$$\begin{aligned} &1 + 4p_1; \\ &0; \\ &4p_1; \\ &\text{если } p_1 = 0, \text{ то } 0, \text{ иначе } 4p_1; \end{aligned}$$

- f_2 имеет такой же вид, как f в (25), где $c \stackrel{\text{def}}{=} 1$.

Докажем, что определенное выше множество Γ_f удовлетворяет условиям теоремы 4.

1. Докажем, что Γ_f устойчиво относительно сдвигов. Нетрудно видеть, что если $|u| = |v|$, то

$$\begin{aligned}
 - (f_1 D^{xy_1})(u, v) &= f_1(xu, y_1v) = c \left(\frac{1}{5}\right)^{p+1} \times \\
 &\times \left(\frac{4}{5}\right)^q, \\
 \text{где } p \text{ и } q &\text{ — количество вхождений } y_1 \text{ и } y_2 \\
 &\text{ в } v, \text{ и}
 \end{aligned}$$

$$c \stackrel{\text{def}}{=} \begin{cases} 1 + 4(p+1), & \text{если } v = y_1^p \ (p \geq 0); \\ 4(p+1), & \text{если } v = y_1^p y_2 w \\ & (p > 0, w \in Y^*); \\ 4, & \text{если } v = y_2 w \ (w \in Y^*); \end{cases}$$

$$- (f_1 D^{xy_2})(u, v) = f_1(xu, y_2v) = 0;$$

$$\begin{aligned}
 - (f_2 D^{xy_1})(u, v) &= f_2(xu, y_1v) = \left(\frac{1}{5}\right)^{p+1} \times \\
 &\times \left(\frac{4}{5}\right)^q, \\
 \text{где } p \text{ и } q &\text{ — количество вхождений } y_1 \text{ и } y_2 \text{ в } v;
 \end{aligned}$$

$$\begin{aligned}
 - (f_2 D^{xy_2})(u, v) &= f_2(xu, y_2v) = \left(\frac{1}{5}\right)^p \times \\
 &\times \left(\frac{4}{5}\right)^{q+1}.
 \end{aligned}$$

Из вышеприведенных соотношений следуют равенства:

$$(f_1 D^{xy_1})(u, v) = \frac{1}{5} f_1(u, v) + \frac{4}{5} f_2(u, v);$$

$$(f_1 D^{xy_2})(u, v) = 0;$$

$$(f_2 D^{xy_1})(u, v) = \frac{1}{5} f_2(u, v);$$

$$(f_2 D^{xy_2})(u, v) = \frac{4}{5} f_2(u, v),$$

которые означают, что Γ_f устойчиво относительно сдвигов.

- Докажем, что $f \in C_0(\Gamma_f)$. Данное свойство является непосредственным следствием легко проверяемого равенства

$$f(u, v) = \frac{1}{3} f_1(u, v) + \frac{2}{3} f_2(u, v).$$

Таким образом, на основании теоремы можно заключить, что определенная выше функция 4 является реализуемой. В доказательстве теоремы 4 содержится алгоритм построения ВА, реакция которого совпадает с f . \square

Отметим, что для применения к данной функции f критерия Бухараева–Хомута необходимо построить множество S_f состояний для МК f , которое, как нетрудно доказать, в данном случае будет бесконечным.

5.2 Пример нереализуемой вероятностной реакции

Пусть множество X имеет вид $\{a, b\}$, а множество Y имеет вид $\{0, 1\}$. Определим функцию $f : X^* \times Y^* \rightarrow [0, 1]$ следующим образом: $f(\varepsilon, \varepsilon) \stackrel{\text{def}}{=} 1$ и $\forall u \in X^*, \forall v \in Y^*$

$$- \text{если } |u| \neq |v|, \text{ то } f(u, v) \stackrel{\text{def}}{=} 0;$$

- если строки u и v имеют вид $x_1 \cdots x_k$ и $y_1 \cdots y_k$ соответственно, где $k \geq 1$ и $\forall i = 1, \dots, k$ $x_i \in X$ и $y_i \in Y$, то $\forall i = 1, \dots, k$ компонента y_i строки v имеет следующий вид:

- если число вхождений символа a в префикс $x_1 \cdots x_i$ строки u совпадает с числом вхождений символа b в этот префикс, то $y_i = 1$,
- иначе $y_i = 0$.

Нетрудно видеть, что определенная выше функция f является

- СФ, так как $\forall u \in X^*, \exists v \in Y^* : f(u, v) = 1$, $\forall v' \in Y^* \setminus \{v\} f(u, v') = 0$;

- ВР, так как $\forall u \in X^*, \forall v \in Y^*$ первое соотношение в условии (10) из определения понятия ВР верно по определению функции f , а второе соотношение

$$\forall x \in X \quad f(u, v) = \sum_{y \in Y} f(ux, vy) \quad (27)$$

верно потому, что

- если $f(u, v) = 0$, то $\forall x \in X, \forall y \in Y$ $f(ux, vy) = 0$;
- если $f(u, v) = 1$, то $\forall x \in X \exists y \in Y$:

$$f(ux, vy) = 1, \quad \forall y' \in Y \setminus \{y\} \quad f(ux, vy') = 0,$$

поэтому в данном случае соотношение (27) также верно.

Можно доказать, что не существует ВА, реакция которого совпадает с ВР f .

6 Заключение

Критерий реализуемости вероятностных реакций конечными вероятностными автоматами, изложенный в настоящей работе, является более простым, чем соответствующий критерий Бухараева–Хомута. Однако проверка этого критерия для заданной ВР f может представлять некоторые трудности, поскольку для доказательства реализуемости f необходимо построить конечное мно-

жество ВР Γ_f , удовлетворяющее условию теоремы 4. Одним из направлений развития изложенного в настоящей работе результата может быть нахождение стратегий построения для заданной ВР f соответствующего множества ВР Γ_f .

Кроме того, поскольку множество Γ_f для заданной ВР f можно рассматривать как множество состояний одного из ВА, реакция которого совпадает с f , то, следовательно, к проблеме построения для заданной ВР f соответствующего множества Γ_f с наименьшим возможным числом элементов сводится проблема построения для заданного ВА A такого ВА, реакция которого совпадает с реакцией ВА A и который содержит наименьшее возможное число состояний (поскольку в качестве исходной ВР f можно рассматривать реакцию ВА A). Данная проблема известна в литературе по теории автоматов как проблема минимизации автоматов и является одним из наиболее популярных предметов исследований в области теории ВА. Среди последних результатов, относящихся к решению данной проблемы, отметим работы [26–28]. Одним из путей развития данных результатов может быть разработка на их основе методов построения для заданной ВР f соответствующего множества Γ_f , которое содержит как можно меньшее число элементов.

Литература

1. Rabin M. O. Probabilistic automata // Information Control, 1963. Vol. 6. No. 3. P. 230–245.
2. Хопкрофт Д., Мотвани Р., Ульман Дж. Введение в теорию автоматов, языков и вычислений / Пер. с англ. — М.: Вильямс, 2002. 528 с. (Hopcroft J. E., Motwani R., Ullman J. D. 2006. Introduction to automata theory, languages, and computation. — 3rd ed. — Pearson. 750 p.)
3. Кемени Дж., Снелл Дж. Конечные цепи Маркова / Пер. с англ. — М.: Наука, 1970. 274 с. (Kemeny J. G., Snell J. L. 1976. Finite Markov chains. — New York — Berlin — Heidelberg — Tokyo: Springer-Verlag. 225 p.)
4. Carlyle J. W. Reduced forms for stochastic sequential machines // J. Math. Anal. Appl., 1963. Vol. 7. No. 2. P. 167–175.
5. Бухараев Р. Г. Некоторые эквивалентности в теории вероятностных автоматов // Ученые записки Казанского университета, 1964. Т. 124. № 2. С. 45–65.
6. Starke P. H. Theorie stochastischen Automaten // Elektronische Informationsverarbeitung und Kybernetik, 1965. Vol. 1. No. 2. P. 5–32.
7. Paz A. Introduction to probabilistic automata. — New York, NY, USA: Academic Press, 1971. 228 p.
8. Бухараев Р. Г. Основы теории вероятностных автоматов. — М.: Наука, 1985. 288 с.
9. Segala R., Lynch N. A. Probabilistic simulations for probabilistic processes // Nordic J. Computing, 1995. Vol. 2. No. 2. P. 250–273.
10. Stoelinga M. An introduction to probabilistic automata // Bull. Eur. Assoc. Theor. Comput. Sci., 2002. Vol. 78. P. 176–198.
11. Sokolova A., de Vink E. P. Probabilistic automata: System types, parallel composition and comparison // Validation of stochastic systems — a guide to current research / Eds. Ch. Baier, B. R. Haverkort, H. Hermanns, J.-P. Katoen, M. Siegle. — Lecture notes in computer science ser. — Springer, 2004. Vol. 2925. P. 1–43.
12. Rabiner L. R. A tutorial on hidden Markov models and selected applications in speech recognition // Proc. IEEE, 1989. Vol. 77. No. 2. P. 257–286.
13. Darwiche A. Modeling and reasoning with Bayesian networks. — Cambridge: Cambridge University Press, 2009. 562 p.
14. Koller D., Friedman N. Probabilistic graphical models. Principles and techniques. — Massachusetts: MIT Press, 2009. 1280 p.
15. Handbook of Markov decision processes / Eds. E. A. Feinberg, A. Shwartz. — Boston, MA, USA: Kluwer, 2002. 562 p.
16. Wu S.-H., Smolka S. A., Stark E. W. Composition and behaviors of probabilistic I/O automata // Theor. Comput. Sci., 1997. Vol. 176. P. 1–38.
17. Delahaye B., Katoen J.-P., Larsen K. G., Legay A., Pedersen M. L., Sher F., Wasowski A. Abstract probabilistic automata // Inform. Comput., 2013. Vol. 232. P. 66–116.
18. Kudlek M. Probability in Petri nets // Fund. Inform., 2005. Vol. 67. No. 1. P. 121–130.
19. Liu Y., Miao H., Zeng H., Li Z. Probabilistic Petri net and its logical semantics // 9th Conference (International) on Software Engineering Research, Management and Applications Proceedings. — Baltimore: IEEE Computer Society, 2011. P. 73–78.
20. Eisentraut C., Hermanns H., Zhang L. On probabilistic automata in continuous time // 25th Annual IEEE Symposium on Logic in Computer Science (LICS) Proceedings, 2010. P. 342–351.
21. Jonsson B., Larsen K. G., Yi W. Probabilistic extensions of process algebras // Handbook of process algebras. — North Holland: Elsevier, 2001. P. 685–710.
22. Homuth H. H. A type of stochastic automation applicable to the communication channel // Angewandte Informatik, 1971. No. 8. P. 362–372.
23. Бухараев Р. Г. Теория абстрактных вероятностных автоматов // Проблемы кибернетики, 1975. Вып. 30. С. 147–198.
24. Мур Э. Ф. Умозрительные эксперименты с последовательными машинами // Автоматы. — М.: ИЛ, 1956. С. 179–210.

25. Бухараев П. Г. Сети вероятностных процессов // Математические вопросы кибернетики, 2007. Вып. 16. С. 57–72.
26. Mateus P., Qiu D., Li L. On the complexity of minimizing probabilistic and quantum automata // Inform. Comput., 2012. Vol. 218. P. 36–53.
27. Миронов А. М., Френкель С. Л. Минимизация вероятностных моделей программ // Фундаментальная и прикладная математика, 2014. Т. 19. Вып. 1. С. 121–163.
28. Kiefer S., Wachter B. Stability and complexity of minimizing probabilistic automata // Automata, languages, and programming / Eds. J. Esparza, P. Fraigniaud, Th. Husfeldt, E. Koutsoupias. — Lecture notes in computer science ser. — Springer, 2014. Vol. 8573. P. 268–279.

Поступила в редакцию 05.05.15

REALIZABILITY OF PROBABILISTIC REACTIONS BY FINITE PROBABILISTIC AUTOMATA

A. M. Mironov

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The paper considers the problem of optimizing the access control on a set of dynamic threshold strategies in an $M/D/1$ system. If the number of concurrent requests in a system is more than the threshold then the system stops accepting requests. If the number of requests is less or equal to this value, then the system resumes accepting requests. As a target function, the average value of the marginal revenue obtained per time unit in the stationary mode is used. It is assumed that the system receives a fixed fee for each accepted request and pays a fixed penalty for each overdue service of a request. The system does not receive a fee and does not pay a penalty for each rejected request. Estimates of the optimal value of the target function and the optimal threshold value are obtained.

Keywords: probabilistic automata; probabilistic reaction; random functions

DOI: 10.14357/19922264150309

Acknowledgments

The work was partially supported by the Russian Foundation for Basic Research (project 12-07-00109).

References

- Rabin, M. O. 1963. Probabilistic automata. *Information Control* 6(3):230–245.
- Hopcroft, J. E., R. Motwani, and J. D. Ullman. 2006. *Introduction to automata theory, languages, and computation* 3rd ed. Pearson. 750 p.
- Kemeny, J. G., and J. L. Snell. 1976. *Finite Markov chains*. New York – Berlin – Heidelberg – Tokyo: Springer-Verlag. 225 p.
- Carlyle, J. W. 1963. Reduced forms for stochastic sequential machines. *J. Math. Anal. Appl.* 7(2):167–175.
- Bukharaev, R. G. 1964. Nekotorye ekvivalentnosti v teorii veroyatnostnykh avtomatov [Certain equivalences in the theory of probabilistic automata]. *Uchenye zapiski Kazanskogo Universiteta* [Lecture Notes of Kazan University] 124(2):45–65.
- Starke, P. H. 1965. Theorie stochastischen Automaten. *Elektronische Informationsverarbeitung und Kybernetik* 1(2):5–32.
- Paz, A. 1971. *Introduction to probabilistic automata*. New York, NY: Academic Press. 228 p.
- Bukharaev, R. G. 1985. *Osnovy teorii veroyatnostnykh avtomatov* [Foundations of probabilistic automata theory]. Moscow: Nauka. 288 p.
- Segala, R., and N. A. Lynch. 1995. Probabilistic simulations for probabilistic processes. *Nordic J. Computing* 2(2):250–273.
- Stoelinga, M. 2002. An introduction to probabilistic automata. *Bull. Eur. Assoc. Theor. Comput. Sci.* 78:176–198.
- Sokolova, A., and E. P. de Vink. 2004. Probabilistic automata: System types, parallel composition and comparison. *Validation of stochastic systems — a guide to current research*. Eds. Ch. Baier, B. R. Haverkort, H. Hermanns, J.-P. Katoen, and M. Siegle. Lecture notes in computer science ser. Springer. 2925:1–43.
- Rabiner, L. R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* 77(2):257–286.

13. Darwiche, A. 2009. *Modeling and reasoning with Bayesian networks*. Cambridge University Press. 562 p.
14. Koller, D., and N. Friedman. 2009. *Probabilistic graphical models. Principles and techniques*. Massachusetts: MIT Press. 1280 p.
15. Feinberg, E. A., and A. Schwartz, eds. 2002. *Handbook of Markov decision processes*. Boston, MA: Kluwer. 562 p.
16. Wu, S.-H., S. A. Smolka, and E. W. Stark. 1997. Composition and behaviors of probabilistic I/O automata. *Theor. Comput. Sci.* 176(1):1–38.
17. Delahaye, B., J.-P. Katoen, K. G. Larsen, A. Legay, M. L. Pedersen, F. Sher, and A. Wasowski. 2013. Abstract probabilistic automata. *Inform. Comput.* 232:66–116.
18. Kudlek, M. 2005. Probability in Petri nets. *Fund. Inform.* 67(1):121–130.
19. Liu, Y., H. Miao, H. Zeng, and Z. Li. 2011. Probabilistic Petri net and its logical semantics. *9th Conference (International) on Software Engineering Research, Management and Applications Proceedings*. 73–78.
20. Eisentraut, C., H. Hermanns, and L. Zhang. 2010. On probabilistic automata in continuous time. *25th Annual IEEE Symposium on Logic in Computer Science (LICS) Proceedings*. 342–351.
21. Jonsson, B., K. G. Larsen, and W. Yi. 2001. Probabilistic extensions of process algebras. *Handbook of process algebra*. North Holland: Elsevier. 685–710.
22. Homuth, H. H. 1971. A type of stochastic automation applicable to the communication channel. *Angewandte Informatik* 8:362–372.
23. Bukharaev, R. G. 1975. Teoriya abstraktnykh veroyatnostnykh avtomatov [A theory of abstract probabilistic automata]. *Problemy Kibernetiki* [Problems Cybernetics] 30:147–198.
24. Moore, E. F. 1956. Gedanken-experiments on sequential machines. *Automata Studies. Annals of Mathematical Studies*. Princeton University Press. 34:129–153.
25. Bukharaev, R. G. 2007. Seti veroyatnostnykh protsessorov [Networks of probabilistic processors]. *Matematicheskie Voprosy Kibernetiki* [Mathematical Issues of Cybernetics] 16:57–72.
26. Mateus, P., D. Qiu, and L. Li. 2012. On the complexity of minimizing probabilistic and quantum automata. *Inform. Comput.* 218:36–53.
27. Mironov, A. M., and S. L. Frenkel. 2014. Minimizatsiya veroyatnostnykh modeley programm [Minimization of probabilistic models of programs]. *Fundamental'naya i Prikladnaya Matematika* [Fundamental Applied Mathematics] 19(1):121–163.
28. Kiefer, S., and B. Wachter. 2014. Stability and complexity of minimising probabilistic automata. *Automata, languages, and programming*. Eds. J. Esparza, P. Fraignaud, Th. Husfeldt, and E. Koutsoupias. Lecture notes in computer science ser. Springer. 8573:268–279.

Received May 5, 2015

Contributor

Mironov Andrew M. (b. 1966) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilova Str., Moscow 119333, Russian Federation; amironov66@gmail.com

ССЫЛОЧНАЯ ЦЕЛОСТНОСТЬ ДАННЫХ В КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

В. С. Зыкин¹

Аннотация: Решается прикладная задача автоматизации построения избыточного множества ссылочных ограничений на данные. Эти ограничения позволяют регламентировать на предприятии бизнес-правила в использовании информации, которая хранится в реляционной базе данных (БД) и обслуживается системой управления БД (СУБД). Теоретической основой ограничений служат зависимости включения, которые в данной статье получили обобщение, позволяющее использовать неопределенные значения. Такое обобщение является следствием их практической значимости. Для корректного решения указанной проблемы вводится и исследуется понятие ациклических схем БД. Введена интерпретация ациклических схем БД в виде ассоциированных гиперграфов, доказана теорема о цикличности таких гиперграфов. Представлено автоматическое построение множества всевозможных ссылочных ограничений и предложен алгоритм автоматизации удаления избыточных ссылочных ограничений целостности.

Ключевые слова: ссылочная целостность; неопределенные значения; ациклические схемы

DOI: 10.14357/19922264150310

1 Введение

Целостность БД (database integrity) — соответствие имеющейся в БД информации ее внутренней логике, структуре и всем явно заданным правилам. Каждое правило, налагающее некоторое ограничение на возможное состояние БД, называется ограничением целостности (integrity constraint). Бизнес-процессы оперируют данными, которые могут быть изменены и обновлены в любое время. Данные, участвующие в бизнес-процессе, могут быть связаны с хранимыми данными и должны удовлетворять правилам, которые регулируют поведение компании [1]. Ограничения целостности оказывают существенное влияние на процесс интеграции схем БД при слиянии компаний или расширении сферы их деятельности [2].

Ссылочные ограничения целостности на данные (referential integrity) — один из основных видов ограничений в БД, которые позволяют сохранить структурную целостность БД. В большинстве существующих СУБД поддерживается такой вид ограничений и задаются эти ограничения в виде связей (relationship) на схеме БД. При «ручном» проектировании схемы БД ссылочные ограничения задаются проектировщиком «вручную» по мере формирования очередных отношений (relation) БД. Однако в настоящее время стали широко использоваться средства автоматизации проектирования схем БД. При этом связи между автоматически сформиро-

ванными отношениями на схеме БД по-прежнему необходимо формировать либо корректировать вручную, а проблема идентификации и удаления избыточных связей вообще не решается. Таким образом, становится актуальной проблема построения корректного и избыточного набора ссылочных ограничений целостности.

Ответственность за поддержание ограничений целостности в технологии БД возлагается на СУБД. Любая операция по изменению состояния БД, нарушающая какое-либо ограничение целостности, будет заблокирована СУБД.

Традиционно в технологии БД используются следующие виды ограничений целостности.

1.1 Ограничение домена

При определении структуры отношений БД могут быть заданы ограничения на допустимые значения в столбцах. Определение типа атрибута в отношении задает базовое ограничение, которое контролируется СУБД. В указанном столбце не может появиться значение, противоречащее выбранному типу, например символьная строка в столбце, для которого указан тип «Дата». Кроме того, на значения атрибутов в столбце отношения могут быть дополнены ограничения, например дата должна быть задана в определенном интервале. Попытка ввести значение этого атрибута, лежащее за пределами указанного интервала, будет заблокирована СУБД.

* Работа выполнена при поддержке РГНФ, грант № 14-16-70008 а(р).

¹ Омский государственный технический университет, vszykin@omgtu.ru

1.2 Целостность сущностей

В каждом отношении должен быть задан первичный ключ, который имеет уникальное непустое значение в каждой строке отношения. Основанием для определения первичного ключа служит множество функциональных зависимостей, формируемых проектировщиком БД. Определенному таким образом набору атрибутов отношения ставится в соответствие свойство PRIMARY KEY. В пределах отношения это свойство может указываться только один раз. Другим способом гарантировать уникальность значений подмножества атрибутов для альтернативных ключей отношения является присвоение этому подмножеству свойства UNIQUE.

1.3 Ссылочная целостность

Если в некотором отношении значения атрибутов в строке могут принимать только те значения, для которых есть совпадающие значения соответствующих атрибутов в другом отношении, то такое ограничение называется ссылочной целостностью. Для его реализации в СУБД имеется аппарат внешних ключей (FOREIGN KEY). При этом определяется главное отношение и подчиненное отношение: в подчиненном отношении не может быть строк, которым нет соответствующей строки в главном отношении. Система управления БД не позволит выполнить операцию дополнения строки в подчиненное отношение, если в главном отношении нет соответствующего кортежа, и операция удаления строки в главном отношении будет отвергнута, если в подчиненном отношении имеется соответствующая связанная строка (запрет висячих ссылок).

Необходимо отметить, что ограничения целостности могут взаимодействовать между собой и, более того, противоречить друг другу. В литературе (см., например, [3–5]) активно обсуждается проблема взаимодействия функциональных зависимостей и зависимостей включения и, как следствие, взаимодействия целостности сущностей и ссылочной целостности. Ограничения домена, в свою очередь, могут существенно усилить ссылочную целостность вплоть до блокировки ввода записей в подчиненное отношение. В данной работе предполагается отсутствие влияния на ссылочные ограничения других ограничений. Это не лишает актуальности исследования, поскольку используются типизированные зависимости включения, которые имеют огромное практическое значение, но при этом не взаимодействуют с функциональными зависимостями. Ограничения домена могут только сузить область допустимых значений, но это не

противоречит результатам данной статьи. Проблема взаимодействия ограничений может быть изучена в дальнейшем как расширение представленных результатов исследования.

В настоящее время известны различные варианты автоматизации построения первых двух видов ограничений целостности, а ссылочная целостность остается без должного внимания. Однако при определении ссылочной целостности могут быть использованы средства автоматизации. Это обусловлено тем, что внешние ключи являются следствием декомпозиции и синтеза схемы БД с использованием функциональных зависимостей.

При правильном (классическом) проектировании схемы БД [6, 7] в большинстве случаев внешние ключи устанавливаются на атрибутах первичного ключа главного отношения. Этот факт и будет использован для автоматизации построения связей на схеме БД.

Классический способ проектирования схемы БД [6, 7] основан на зависимостях: функциональных, многозначных, соединения и включения. При этом сущности (отношения) БД формируются в процессе проектирования, что позволяет реализовать принцип независимости данных и, как следствие, устойчивость проекта БД при последующей модернизации.

Альтернативным подходом к созданию проекта БД служит объектно-ориентированный подход. В этом случае сначала создаются сущности (объекты) БД. Единственным гарантом правильности их формирования является интуиция проектировщика. Зачастую объекты дублируют структуру документооборота предприятия, что нарушает принцип независимости данных и приводит к разрушению структуры БД при необходимости ее модернизации. Затем сразу создаются связи между объектами, также на основе интуиции проектировщика (ER-диаграмма — entity–relationship). Типичной ошибкой при таком подходе становится «нагрузка» связей семантикой приложения, т. е. в связях «прячутся» объекты БД. Несмотря на все недостатки объектно-ориентированного подхода, он получил достаточно широкое распространение на практике за счет двух своих преимуществ: наглядности проектирования и низкого уровня требований к квалификации проектировщика. В качестве примеров инструментария таких систем можно привести ERWin, BPWin и др.

При преобразовании ER-диаграммы в схему БД используется механизм автоматического формирования связей. Правила преобразования при этом являются эвристическими с множеством исключений и нереализуемых ситуаций, причина которых

в неоднозначной интерпретации сущностей и связей. Следовательно, говорить о корректности результатов этого процесса нет смысла.

Цель данной работы — исследование свойств ссылочных ограничений целостности и реализация программного обеспечения, позволяющего автоматизировать построение корректного и избыточного набора связей на схеме БД, реализующего ссылочные ограничения целостности. На основе разработки нового математического аппарата и на основе известного механизма использования связей на схеме БД разработаны алгоритмы автоматического формирования ссылочных ограничений целостности. Доказаны корректность и избыточность результатов построений.

2 Постановка задачи

Ссылочные ограничения целостности на данные — это один из основных видов ограничений в БД, которые поддерживаются большинством существующих СУБД. Для этого в СУБД используются связи, установленные на схеме БД. Пусть $U = \{A_1, A_2, \dots, A_n\}$ — множество атрибутов, определенных в БД; $[R_i]$ — множество атрибутов, на которых определено отношение R_i , $[R_i] \subseteq U$; $\mathfrak{R} = (R_1, R_2, \dots, R_k)$ — БД; $S = \{[R_1], [R_2], \dots, [R_k]\}$ — схема БД, $1 \leq i \leq k$. Значения атрибутов, по которым связаны отношения, заимствуются из главных отношений (справочников) в подчиненные отношения. Между ними устанавливается связь $1 : M$ либо $1 : 1$ по направлению от главного отношения к подчиненному. Формальным основанием для установления связей служат зависимости включения (inclusion dependencies) [3]:

Определение 1. Пусть $[R_i]$ и $[R_j]$ — схемы отношений (необязательно различные), $V \subseteq [R_i]$ и $W \subseteq [R_j]$, $|V| = |W|$, тогда соотношение $R_i[V] \subseteq R_j[W]$ называется зависимостью включения.

В этом определении $|V|$ — мощность множества V ; $R_i[V] = \pi_V(R_i)$ — проекция отношения R_i по атрибутам V .

Рассмотрим практические аспекты использования зависимостей включения. Будем предполагать, что условие $V = W$ является необходимым для установления связи. Такие зависимости включения называются типизированными (typed) [4, 5]. Это дополнительное ограничение вполне согласуется с общепринятым свойством связей на схеме БД: связи отражают количественное соотношение кортежей в отношениях и не обладают какой-либо семантикой. Необходимость связывания различных по смыслу атрибутов, скорее всего, является признаком потери какой-либо функциональной

зависимости для связываемых атрибутов либо, как в примере 1.1 из [5], следствием того, что оставшаяся зависимость включения, после удаления избыточных зависимостей, установлена для атрибутов-синонимов и им соответствует тривиальная функциональная зависимость, т. е. решение семантических проблем на предварительном этапе проектирования схемы БД в большинстве случаев позволяет избавиться от необходимости использования нетипизированных зависимостей включения.

В прикладных задачах часто приходится иметь дело с неопределенными данными, когда часть характеристик объектов, кроме идентификатора (первичного ключа), неизвестна. В этом случае условие $\pi_V(R_i) \subseteq \pi_W(R_j)$ в определении 1 может быть не выполнено, т. е. в отношении R_i могут быть кортежи с неопределенными значениями. Этим кортежам соответствуют кортежи в R_j , но с определенными значениями. В технологии БД такое соответствие называется расширенной связью. Далее определение 1 будет изменено в соответствии с этим дополнением.

Рассмотрим пример, когда между отношениями должна быть установлена связь $1 : 1$.

Пример 1. Пусть A_1 — «Табельный номер сотрудника», A_2 — «ФИО сотрудника», A_3 — «Дата увольнения сотрудника». Существуют функциональные зависимости: $A_1 \rightarrow A_2$ и $A_1 \rightarrow A_3$. Области определения этих зависимостей различны: первая зависимость определена для всего множества сотрудников, а вторая — только для уволенных. Это и должно служить основанием для построения декомпозиции: $R_i[A_1, A_2]$ и $R_j[A_1, A_3]$. Между R_i и R_j устанавливается связь $1 : 1$, где R_i — главное отношение, а R_j — подчиненное. Заметим, что такая декомпозиция решает проблему заведомо неопределенных значений: в объединенном отношении $R_i[A_1, A_2, A_3]$ значение атрибута A_3 будет иметь неопределенное значение для всех неуволенных сотрудников. Формальное определение области зависимости включения является трудоемкой задачей, требующей отдельного исследования, в данной статье ограничимся данным примером для понимания термина «область определения функциональной зависимости».

Введем обозначения: $PK(R_i)$, или просто $PK(i)$, — множество атрибутов, служащее первичным ключом в отношении R_i ; $L(i, j, X)$ — связь $1 : 1$ либо $1 : M$ от R_i к R_j , установленная по множеству атрибутов X , где R_i — главное отношение, а R_j — подчиненное отношение; $L_1(i, j, X)$ — связь $1 : 1$ от R_i к R_j ; $L_M(i, j, X)$ — связь $1 : M$ от R_i к R_j . Заметим, что в отношении R_i может существовать несколько альтернативных первичных

ключей и множество связей, в которых R_i будет главным или подчиненным.

Далее рассмотрим определение связей между отношениями БД, которые будем использовать в качестве расширенных ссылочных ограничений целостности, учитывающих наличие неопределенных значений.

Определение 2. Между отношениями R_i и R_j допустима связь $L_1(i, j, X)$, если $X = PK(R_i) = PK(R_j)$ и для любых реализаций R_i и R_j выполнено $\pi_X(R_j) \subseteq \pi_X(R_i)$.

В этом определении невозможно наличие неопределенных значений, поскольку атрибуты X в обоих отношениях являются компонентами первичного ключа.

Определение 3. Между отношениями R_i и R_j допустима связь $L_M(i, j, X)$, если $PK(R_i) \neq PK(R_j)$ и $PK(R_i) \subseteq [R_j]$.

В определении 3 отношение R_j может содержать неопределенные значения для атрибутов, не принадлежащих первичному ключу. Заметим, что определения 2 и 3 соответствуют типизированным зависимостям включения, которые поддерживаются СУБД за счет создания внешних ключей (foreign key). Ограничение целостности, задаваемое связью $L_M(i, j, X)$, не подразумевает выполнения условия $\pi_X(R_j) \subseteq \pi_X(R_i)$, где $X = R_i \cap R_j$, поскольку атрибуты X , не принадлежащие $PK(R_j)$, могут принимать неопределенные значения, тогда как в R_i им соответствуют определенные значения. Смысл ограничения в том, что неопределенное значение какого-либо атрибута в R_j может быть заменено только тем определенным значением, которое есть в R_i . Поиск связей, соответствующих определениям 2 и 3, достаточно просто алгоритмизируется, что позволяет выявлять большинство ссылочных ограничений целостности в автоматическом режиме.

В схемах БД [8] кроме рассмотренных используются следующие связи:

- (1) «one or more mandatory»;
- (2) «one and only one»;
- (3) «zero or one optional»;
- (4) «zero or many optional».

Поскольку связи на схеме БД являются парными, то связь 1 служит частным случаем связи $1 : M$ со стороны второго отношения, связь 2 — частным случаем связи $1 : 1$ со стороны обоих отношений, связь 3 совпадает со связью $1 : 1$ со стороны второго отношения, связь 4 совпадает со связью $1 : M$ со стороны второго отношения. Другие варианты связей, имеющие тип связи 3 со стороны

первого отношения, противоречат определению зависимости включения и не являются ограничением ссылочной целостности данных. Такие связи будем далее игнорировать. Особый случай представляет вариант связи $1 : M$, который допускает наличие кортежей с неопределенными значениями общих атрибутов во втором отношении, когда они не являются ключевыми. Поскольку неопределенное значение не равно определенному значению, то этот вариант связи противоречит зависимости включения, но остается ограничением целостности. Оставим такую связь в рассмотрении, назвав ее расширенной связью.

Таким образом, задача состоит в разработке алгоритмов для формирования ссылочных ограничений целостности в соответствии с определениями 2 и 3.

3 Ациклические схемы баз данных

В работе [4] условие ациклическости зависимостей включения формулируется следующим образом: множество зависимостей включения является ациклическим, если не существует объекта $R[X] \subseteq R[Y]$ и $X \neq Y$ и не существует отношений R_1, R_2, \dots, R_n ($n > 1$) таких, что $R_1[X_1] \subseteq R_2[Y_2], R_2[X_2] \subseteq R_3[Y_3], \dots, R_n[X_n] \subseteq R_1[Y_1]$. Тогда условие ациклическости зависимостей включения можно проверять с использованием полиномиального алгоритма построения замыкания множества отношений.

Рассмотрим простейший пример с типизированными зависимостями включения $R_1[X_1] \subseteq R_2[X_1], R_2[X_1] \subseteq R_3[X_1]$ и пусть атрибуты X_1 в отношении R_2 допускают неопределенные значения. Тогда замыкание $R_1[X_1] \subseteq R_3[X_1]$ не выполнено, поскольку не выполнена зависимость $R_2[X_1] \subseteq R_3[X_1]$. Это говорит о том, что учет расширенных связей не позволяет использовать аппарат выводимости и построения замыкания. В предлагаемом подходе условие $R_2[X_1] \subseteq R_3[X_1]$ выполнено, если X_1 является первичным ключом в R_2 , а в первичном ключе неопределенные значения не допускаются. Это несколько сужает область применимости, но позволяет формировать корректные построения с учетом расширенных связей.

Сформулируем условие ациклическости для совокупности отношений по аналогии с условием ациклическости зависимостей включения [4], но учитывающее расширенные связи.

Определение 4. Совокупность отношений \mathfrak{R} будем называть ациклической, если не существует упорядоченного подмножества отношений

$$\{R_{m(1)}, R_{m(2)}, \dots, R_{m(s)}\} \subseteq \mathfrak{R}$$

такого, что имеются связи

$$L(m(1), m(2), X_1), L(m(2), m(3), X_2), \dots, L(m(s), m(1), X_s), \quad (1)$$

$s > 1$ и $X_1 \subseteq X_2 \subseteq \dots \subseteq X_s$, в противном случае совокупность отношений \mathfrak{R} будем называть циклической. Последовательность (1) может содержать расширенные связи.

В работе [9] рассмотрено несколько различных условий ацикличности \mathfrak{R} и доказана их эквивалентность. Одно из условий, рассмотренное также в работе [7], устанавливает, что совокупность \mathfrak{R} является ациклической, если ацикличесок ассоциированный с ней гиперграф. Поскольку предполагается, что в \mathfrak{R} отсутствуют отношения, для которых $[R_i] \subseteq [R_j]$ при $i \neq j$, то гиперграф будет редуцированным.

Определение 5. В контексте БД гиперграфом называется пара $G = (U, S)$, где вершинами U служат атрибуты БД, а элементами S являются гиперребра $[R_i]$, которые объединяют вершины — атрибуты, входящие в $[R_i]$. Гиперграф будет циклическим, если существует атрибут A и последовательность $[R'_1], [R'_2], \dots, [R'_m]$ такая, что $A \in [R'_1]$ и $A \in [R'_m]$, и $[R'_i] \cap [R'_{i+1}] \neq \emptyset$ для $1 \leq i \leq m - 1$.

Теорема 1. Если совокупность \mathfrak{R} циклическая, то ассоциированный с ней гиперграф также будет циклическим.

Доказательство. Подмножество отношений $\{R_{m(1)}, R_{m(2)}, \dots, R_{m(s)}\} \subseteq \mathfrak{R}$ соответствует ребрам в ассоциированном гиперграфе. Ребра связывают совокупности атрибутов — узлы гиперграфа. Существование последовательности $L(m(1), m(2), X_1), L(m(2), m(3), X_2), \dots, L(m(s), m(1), X_s)$ гарантирует, что $R_{m(i)} \cap R_{m(i+1)} \neq \emptyset, i = 1, 2, \dots, s$, где $m(s+1) = m(1)$. Следовательно, подмножество $\{R_{m(1)}, R_{m(2)}, \dots, R_{m(s)}\} \subseteq \mathfrak{R}$ вместе со связываемыми узлами образует цикл в общем гиперграфе. А значит, общий гиперграф для \mathfrak{R} будет циклическим.

Исследование соотношения условий ацикличности в [9] с определением 4 продолжим рассмотрением примера 13.15 из [7]: $\mathfrak{R} = \{R_1, R_2, R_3\}$, где $R_1[ABC]; R_2[BD]; R_3[CD]$. Дополним этот пример функциональными зависимостями $\{BC \rightarrow A, B \rightarrow D, C \rightarrow D\}$, которые могли быть использованы при построении \mathfrak{R} . Ассоциированный с \mathfrak{R} гиперграф является циклическим, тогда как по определению 4 совокупность \mathfrak{R} является ациклической, поскольку для нее существуют только две связи: $L_M(2, 1, B)$ и $L_M(3, 1, C)$. Таким образом, класс ациклических схем БД по определению 4 шире, чем по условиям в [9].

4 Формирование избыточного набора связей

Далее будем предполагать, что совокупность \mathfrak{R} является ациклической в соответствии с определением 4.

Определение 6. Связь $L(i, j, X)$ задает множество допустимых кортежей T_i в R_j . Тогда связь $L(i, j, X)$ будет избыточной, если существуют связи $L(l, j, Y_l), l = 1, \dots, m$, где $\bigcap_{l=1}^m T_l \subseteq T_i$.

Заметим, что под определение 6 не попадают избыточные связи, являющиеся следствием взаимодействия функциональных зависимостей и зависимостей включения [3, 5], поскольку для типизированных зависимостей включения такое взаимодействие удаляется на этапе проектирования схем отношений БД. По этой же причине между парой отношений может быть установлена только одна связь (другая может быть только циклической). Далее будем рассматривать только такие множества \mathfrak{R} .

Для формирования избыточного набора связей на схеме БД сначала формируется множество отношений $\mathfrak{R} = \{R_1, R_2, \dots, R_q\}$ в соответствии с нормальной формой Бойса—Кодда. По определениям 2 и 3 между сформированными отношениями устанавливаются связи $L_M(i, j, X)$ и $L_1(i, j, X)$. Затем осуществляется удаление избыточных связей.

Связи $L_M(i, j, X)$ не могут быть установлены простым алгоритмом перебора по определению 3, поскольку наличие связи предполагает ссылочное ограничение целостности, а оно, в свою очередь, зависит от семантики прикладной области. Для демонстрации этого факта рассмотрим пример.

Пример 2. Пусть задано некоторое множество отношений — фрагмент схемы БД учебного заведения, где подчеркнуты ключевые атрибуты отношений:

$R_1 = \text{Студенты}$ (№ студента, № группы, ФИО студента);

$R_2 = \text{Список групп}$ (№ группы, Код группы, № специальности, № курса);

$R_3 = \text{Предметы}$ (№ предмета, Предмет);

$R_4 = \text{Оценки}$ (№ студента, № группы, № предмета, Оценка);

$R_5 = \text{Аттестация}$ (№ студента, № группы, № предмета, Вид аттестации, Балл).

На схеме должны быть установлены следующие связи:

$L_M(2, 1, \langle \text{№ группы} \rangle)$;

$L_M(2, 4, \langle \text{№ группы} \rangle)$;

$L_M(2, 5, \langle \mathbb{N} \text{ группы} \rangle)$;
 $L_M(1, 4, \langle \mathbb{N} \text{ студента, } \mathbb{N} \text{ группы} \rangle)$;
 $L_M(1, 5, \langle \mathbb{N} \text{ студента, } \mathbb{N} \text{ группы} \rangle)$;
 $L_M(3, 4, \langle \mathbb{N} \text{ предмета} \rangle)$;
 $L_M(3, 5, \langle \mathbb{N} \text{ предмета} \rangle)$.

Формально, по определению 2 должна быть установлена связь $L_M(4, 5, \langle \mathbb{N} \text{ студента, } \mathbb{N} \text{ группы, } \mathbb{N} \text{ предмета} \rangle)$, которая содержательно задает следующее ограничение целостности: аттестация студентов может осуществляться только по тем предметам, по которым проставлены оценки. Однако это противоречит прикладной области, и связь не должна присутствовать на схеме БД. Проектировщику БД в этом случае необходимо ответить на вопрос: нужна или нет данная связь?

Для установления связи $L_1(i, j, X)$ дополнительно надо определить: какое отношение в этой связи является главным, а какое — подчиненным. При пустой БД, на этапе проектирования, включение $\pi_V(R_j) \subseteq \pi_V(R_i)$, где V — первичный ключ отношений R_i и R_j , не может быть установлено алгоритмически. Поэтому при установлении связей необходимо учитывать семантику прикладной области, которую определяет проектировщик схемы БД. При определении включения между R_i и R_j проектировщику БД достаточно ответить на вопрос: «Как соотносятся области значений атрибутов $R_i \cap R_j$ в отношениях R_i и R_j ?», предоставив выбор одного из трех вариантов:

- «1. Значения в R_i являются подмножеством значений в R_j по множеству атрибутов V »;
- «2. Значения в R_j являются подмножеством значений в R_i по множеству атрибутов V »;
- «3. Не выполнены условия 1-го и 2-го вариантов».

При выборе первого варианта отношение R_j становится главным для R_i , второго — R_i становится главным для R_j , третьего — отсутствие связи между R_i и R_j . Однако для решения указанной проблемы достаточно спросить: нужна связь $L(i, j, V)$ либо связь $L(j, i, V)$. Такое решение будет пригодно для обоих видов связей. Допустим, что этот выбор реализуется функцией $\text{Ch}(R_i, R_j, V)$, имеющей значения $L(i, j, V)$ либо пусто (\emptyset).

В соответствии с рассмотренным обоснованием алгоритм формирования связей для множества \mathfrak{R} , где L — множество всех связей, может быть записан в следующем виде:

```

L = ∅
do i = 1 to k
  for each PK(Ri) in Ri
    do j = 1 to k
      if i ≠ j then

```

```

    if PK(Ri) ∈ Rj then
      L = L ∪ Ch(Ri, Rj, V)
    endif
  endif
enddo
endfor
enddo

```

Сложность данного алгоритма является полиномиальной по времени: $O(k^2)$.

В работе [4] без доказательства сформулировано условие избыточной зависимости включения. Докажем аналогичное условие для связей, в том числе и расширенных, опираясь на реализации отношений, а не на выводимость.

Теорема 2. *Связь является избыточной, если существуют связи*

$$L(i, m(1), X_0), L(m(1), m(2), X_1), \dots, L(m(p), j, X_p) \quad (2)$$

и

$$X \subseteq \text{PK}(i) \subseteq X_s \subseteq R_{m(s)}, \quad s = 2, 3, \dots, p, \quad (3)$$

где m — массив номеров отношений.

Доказательство. Заметим, что условия $\text{PK}(i) \subseteq R_j$ и $\text{PK}(i) \subseteq R_{m(1)}$ являются следствием наличия связей $L(i, j, X)$ и $L(i, m(1), X_0)$ соответственно.

1. Пусть выполнены условия (2) и (3). Поскольку множество атрибутов $\text{PK}(i)$ присутствует в каждом отношении цепочки

$$R_i, R_{m(1)}, R_{m(2)}, \dots, R_{m(p)}, R_j, \quad (4)$$

то по определению оно участвует во всех связях последовательности (2); следовательно,

$$\pi_{\text{PK}(i)}(R_i) \supseteq \pi_{\text{PK}(i)}(R_{m(p)}) \supseteq \dots \supseteq \pi_{\text{PK}(i)}(R_{m(1)}) \supseteq \pi_{\text{PK}(i)}(R_j).$$

Данная последовательность включений гарантирует, что в R_j не появится кортежа со значением $\text{PK}(i)$, не содержащимся в других отношениях цепочки (4). Это служит более сильным ограничением, чем задает связь $L(i, j, X)$ и, следовательно, она избыточна в соответствии с определением 6. Пусть связь $L(m(p), j, X_p)$ в последовательности (2) является расширенной, а связь $L(i, j, X)$ не является расширенной. Другие связи в последовательности (2) не могут быть расширенными, так как общие атрибуты во вторых отношениях должны быть ключевыми. В отношении R_j не появится кортеж с пустыми значениями всех атрибутов $\text{PK}(i)$, поскольку данное ограничение устанавливается внутри R_j . Поэтому связь $L(i, j, X)$ может быть удалена.

2. Предположим, что в цепочке (4) имеется отношение $R_{m(s)}$, которое не содержит полностью множество атрибутов $PK(i)$. Тогда при переходе от $R_{m(s-1)}$ к $R_{m(s)}$ снимаются ограничения на атрибуты $PK(i) - [R_{m(s)}]$, обусловленные последовательностью связей (2). Следовательно, эти атрибуты могут принимать произвольные значения, в том числе и не принадлежащие проекции $\pi_{PK(i)}(R_i)$. Поэтому последовательность (2) не является основанием для удаления связи $L(i, j, X)$.
3. Если выполнено условие (3), но отсутствует последовательность (2), то и ограничения на $PK(i)$ в R_j , обусловленные этой последовательностью, отсутствуют.

Следствие. Если существует (2) и $p = 1$, то связь $L(i, j, X)$ избыточна.

Замечание. Последовательность нескольких связей не может быть избыточна. Допустим, что существуют последовательности (2) и $L(i, v, X_i)$, $L(v, j, X_v)$, $v \neq m(s)$, $s = 1, 2, \dots, p$. Удаление связей $L(i, v, X_i)$ и $L(v, j, X_v)$, независимо от условия (3), означало бы снятие с отношения R_v ограничений, накладываемых связью $L(i, v, X_i)$.

Поиск избыточных связей с использованием соотношений (2) и (3) будет экспоненциальной задачей. Поскольку связи являются реализацией типизированных зависимостей включения, то к поиску избыточных связей применим алгоритм построения замыкания отношений [2], полиномиальный по времени. Рассмотрим адаптацию этого алгоритма для удаления избыточных связей без использования графа зависимостей включения [4]. Пусть $PK(R_i)$ — ключ отношения R_i , соответствующий связи $L(i, j, X)$. Тогда

```

for each  $L(i, j, X)$  in  $L$ 
   $l = 1$ 
   $m(l) = i$ 
   $iterations = true$ 
  do while  $iterations$ 
    for each  $L(v, w, X_l)$  in  $L$ 
      where  $L(i, j, X) \neq L(v, w, X_l)$ 
         $substitution = false$ 
        if  $v \in m[1, \dots, l]$  and  $PK(R_i) \subseteq R_w$  then
          if  $w = j$  then
             $L = L - L(i, j, X)$ 
          exit do
        else
          if  $w \notin m[1, \dots, l]$  then
             $l = l + 1$ 
             $m(l) = w$ 
             $substitution = true$ 
          endif
        endif
  endif
endif

```

```

endif
endifor
if not  $substitution$  then  $iterations = false$ 
enddo
endifor

```

Пусть $|L|$ — число связей в исходном множестве L . С учетом того, что на каждой итерации (цикл do while) к массиву m присоединяется не менее одного номера отношения (их k штук), рассмотренный алгоритм будет иметь полиномиальную сложность: $O(|L|^2 k^2)$. Дополнительная степень при k обусловлена выполнением операций $v \in m[1, \dots, l]$ и $w \notin m[1, \dots, l]$. Доказательство корректности алгоритма выполняется индукцией по числу итераций по аналогии с доказательством теоремы 5.2 в [6].

Если между отношениями установлены связи $L_1(m(1), m(2), X_1), \dots, L_1(m(p-1), m(p), X_{p-1})$, то эти последовательности могут иметь общие узлы (отношения). Если исходное отношение R_i в соответствии с определением 3 имело связь $L_M(i, j, X)$, то связи $L_M(m(l), j, X_l)$, $l = 1, \dots, p$, также будут иметь место. Алгоритм удаления избыточных связей оставит исходящие связи $L_M(m(p), j, X_p)$, т.е. исходящие связи из последних отношений последовательностей. Эти связи будут задавать более жесткие ограничения, чем отношение R_i , т.е. сократится количество допустимых значений. Следовательно, связи необходимо «передвинуть» к тем отношениям последовательностей, которые соответствуют реальным ограничениям. Число таких связей регламентируется следующим очевидным свойством, вытекающим из теоремы 2.

Свойство 1. Последовательность отношений $R_{m(1)}, R_{m(2)}, \dots, R_{m(p)}$, между которыми установлены связи $L_1(m(1), m(2), X_1), \dots, L_1(m(p-1), m(p), X_{p-1})$, и произвольное отношение R_j могут иметь не более одной связи $L_M(m(l), j, X_l)$.

Таким образом, число связей $L_M(m(l), j, X_l)$ с отношением R_i из указанных последовательностей связей может быть не больше, чем соответствующих последовательностей $R_{m(1)}, R_{m(2)}, \dots, R_{m(p)}$. Дополнительная проблема заключается в том, что все эти связи в совокупности задают конъюнктивное ограничение на значения атрибутов $PK(R_i)$ в отношении R_j . Однако может потребоваться дизъюнктивное ограничение для какой-либо группы связей. В БД не используются связи, устанавливаемые для трех и более отношений и позволяющие задавать ограничения в виде дизъюнктивной (ДНФ) или конъюнктивной (КНФ) нормальной форме. Схема БД в этом случае была бы гиперграфом (не надо путать с гиперграфом из [4, 9], поскольку связей там вообще нет). В рассматриваемом случае для упомянутой группы связей при-

дется создать фиктивное отношение с атрибутами РК(R_i), которое объединит эти связи и будет встроено во все соответствующие этим связям последовательности. Заметим, что, пользуясь определением выводимости зависимостей включения без учета неопределенных значений, невозможно получить аналогичное преобразование.

5 Заключение

Область применения предложенного подхода ограничивается определениями 2 и 3, что вполне достаточно для корректно сформированных отношений БД. Обоснование этого утверждения может быть сделано только на уровне анализа семантики конкретных БД. Если же проектировщик дополнит связи на схеме БД, не удовлетворяющие определениям 2 и 3, то тогда ему будет необходимо перейти к построению нормальной формы IDNF [3]. Если же проектировщик захочет усилить ограничения, используя связи [8], то для этого будет достаточно подправить (усилить) свойства существующих связей.

В работе рассмотрены простейшие виды парных ограничений целостности данных, заимствованные из объектно-ориентированных моделей данных. Однако на практике имеются структурные ограничения на данные, которые связывают не только пары, но и большее количество компонентов БД в одном ограничении, например для пятой нормальной формы, когда связывать необходимо три и более отношений, полученных после декомпозиции. Для решения указанной задачи потребуется развить формальную теорию зависимостей включения (теоретическую основу ссылочной целостности) на случай множества отношений: построение избыточного множества зависимо-

стей, ацикличность, разработка алгоритмов автоматического формирования программ (триггеров), обслуживающих ограничения, и т. д. В дальнейшем эти исследования позволят существенно расширить функциональные возможности приложений, работающих с БД.

Литература

1. *Gómez-López M. T., Gasca R. M., PérezÁlvarez J. M.* Compliance validation and diagnosis of business data constraints in business processes // *Inform. Syst.*, 2015. Vol. 48. P. 26–43.
2. *Visser J.* Coupled transformation of schemas, documents, queries, and constraints // *Electronic Notes in Theoretical Computer Sci.*, 2008. Vol. 200. Iss. 3. P. 3–23.
3. *Casanova M., Fagin R., Papadimitriou C.* Inclusion dependencies and their interaction with functional // *J. Comp. Syst. Sci.*, 1984. Vol. 28. Iss. 1. P. 29–59.
4. *Missaoui R., Godin R.* The implication problem for inclusion dependencies: A graph approach // *SIGMOD Record*, 1990. Vol. 19. Iss. 1. P. 36–40.
5. *Levene M., Vincent M. W.* Justification for inclusion dependency normal form // *IEEE Trans. Knowledge Data Eng.*, 2000. Vol. 12. Iss. 2. P. 281–291.
6. *Ульман Дж.* Основы систем баз данных. — М.: Финансы и статистика, 1983. 334 с.
7. *Ме́йер Д.* Теория реляционных баз данных / Пер. с англ. — М.: Мир, 1987. 608 с. (*Maier D.* The theory of relational databases. — Computer Science Press, 1983. 656 p.)
8. *Garmany J., Walker J., Clark T.* Logical database design principles. — New York, NY, USA: Auerbach Publications, 2005. 200 p.
9. *Beeri C., Fagin R., Maier D., Yannakakis M.* On the desirability of acyclic database schemes // *ACM*, 1983. Vol. 30. Iss. 3. P. 479–513.

Поступила в редакцию 05.01.15

REFERENTIAL INTEGRITY OF DATA IN CORPORATE INFORMATION SYSTEMS

V. S. Zykin

Omsk State Technical University, 11 Mira Av., Omsk 644050, Russian Federation

Abstract: The paper deals with the task of construction of a nonredundant set of referential constraints on data. This set of constraints allows regulating the business rules of using information on the enterprise, which is supported by a database management system and is located in a relation database. The inclusion dependences are the theoretical bases of the restrictions and they have obtained generalization in this paper; so, it is possible to use null values. This generalization is a consequence of their practical significance. The term “acyclic database schemes” is introduced

and investigated for correct solution of this problem. Some attention is given to interpretation of acyclic schemas with hypergraphs, the theorem of acyclic hypergraph is proved. It is proposed to construct the set of all referential constraints automatically by using the rule of decomposition of relations. An algorithm of removing redundant referential integrity constraints is presented.

Keywords: referential integrity; undefined values; acyclic schemes

DOI: 10.14357/19922264150310

References

1. Gómez-López, M. T., R. M. Gasca, and J. M. Pérez-Álvarez. 2015. Compliance validation and diagnosis of business data constraints in business processes. *Inform. Syst.* 48:26–43.
2. Visser, J. Coupled transformation of schemas, documents, queries, and constraints. *Electronic Notes Theoretical Computer Sci.* 200(3):3–23.
3. Casanova, M. 1984. Inclusion dependencies and their interaction with functional. *J. Comp. System Sci.* 28(1):29–59.
4. Missaoui, R. 1990. The implication problem for inclusion dependencies: A graph approach. *SIGMOD Record* 19(1):36–40.
5. Levene, M. 2000. Justification for inclusion dependency normal form. *IEEE Trans. Knowledge Data Eng.* 12(2):281–291.
6. Ul'man, J. 1983. *Osnovy sistem baz dannykh* [Fundamentals of database systems]. Moscow: Finance and Statistics. 334 p.
7. Maier, D. 1983. *The theory of relational databases*. Computer Science Press. 656 p.
8. Garmany, J., J. Walker, and T. Clark. 2005. *Logical database design principles*. Auerbach Publications. 200 p.
9. Beeri, C., R. Fagin, D. Maier, and M. Yannakakis. 1983. On the desirability of acyclic database schemes. *ACM* 30(3):479–513.

Received January 5, 2015

Contributor

Zykin Vladimir S. (b. 1992) — assistant professor, Omsk State Technical University, 11 Mira Av., Omsk 644050, Russian Federation; vszykin@omgtu.ru

ПРОЦЕССЫ ЦЕЛЕНАПРАВЛЕННОЙ ГЕНЕРАЦИИ И РАЗВИТИЯ КРОСС-ЯЗЫКОВЫХ ЭКСПЕРТНЫХ ЗНАНИЙ: СЕМИОТИЧЕСКИЕ ОСНОВАНИЯ МОДЕЛИРОВАНИЯ*

И. М. Зацман¹

Аннотация: Представлены результаты разработки семиотических оснований для создания моделей процессов целенаправленной генерации и развития новых экспертных знаний и разработки технологий, обеспечивающих эти процессы. Необходимость разработки таких технологий проявляется наиболее наглядно в ситуациях, когда имеющиеся системы экспертных знаний не удовлетворяют новым социально или технологически значимым целям, отражающим новые или изменившиеся потребности общества. В статье речь идет не о хорошо известных в области искусственного интеллекта методах и моделях представления знаний, процессах управления формами их представления, а о разработке новых моделей процессов целенаправленной генерации знаний, отражающих динамику их формирования. Рассматриваемый подход к моделированию этих процессов и разработке обеспечивающих их технологий ориентирован на те прикладные области, где знания генерируются экспертами в процессе анализа текстов или других объектов интерпретации, которые могут изменяться во времени, с последующим представлением экспертных знаний в надкорпусных базах данных (НБД). Отличительная черта предлагаемого подхода к моделированию заключается в явном описании отношений между новыми экспертными знаниями и теми объектами интерпретации, на основе анализа которых были сгенерированы элементы новых знаний. Другая отличительная черта заключается в явном описании изменяемых во времени элементов знаний, соответствующих объектам интерпретации. Реализуемость такого подхода демонстрируется на примере экспериментальной информационной технологии, которая поддерживает целенаправленную генерацию экспертами кросс-языковых знаний о переводах глагольных конструкций русского языка на французский. Эти кросс-языковые знания формируются в процессе анализа параллельных текстов на русском и французском языках, пары выровненных предложений которых являются объектами интерпретации.

Ключевые слова: кросс-языковые экспертные знания; компьютерное моделирование; генерация знаний; объекты интерпретации; семиотические основания; модели процессов генерации знаний; надкорпусные базы данных

DOI: 10.14357/19922264150311

1 Введение

Модели процессов генерации и развития новых знаний (далее — модели генерации) стали активно разрабатываться в последнем десятилетии прошлого века. Наиболее известную модель генерации, названную автором спиральной, предложил Икудзиро Нонака [1, 2]. В процессе ее построения Нонака рассматривал личностные знания человека и коллективные (согласованные) знания группы людей, которые были разделены на выражаемые (explicit) и невыражаемые знания (tacit). Таким образом, спиральная модель генерации включает в рассмотрение следующие четыре понятия и соответствующие им четыре множества знаний (рис. 1):

(1) личностные невыражаемые знания (individual tacit knowledge);

- (2) коллективные невыражаемые знания (group tacit knowledge);
- (3) личностные выражаемые знания (individual explicit knowledge);
- (4) коллективные выражаемые знания (group explicit knowledge).

Наряду с этими четырьмя понятиями были определены следующие четыре вида процессов:

- (1) социализация личностных невыражаемых знаний;
- (2) экстернализация коллективных невыражаемых знаний;
- (3) синтез личностных выражаемых знаний;
- (4) интернализация личностных выражаемых знаний.

* Работа выполнена при поддержке РФФИ (проекты 14-07-00785, 13-06-00403) и РГНФ (проект 15-04-00507).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, iz-ipi@al70.ipi.ac.ru

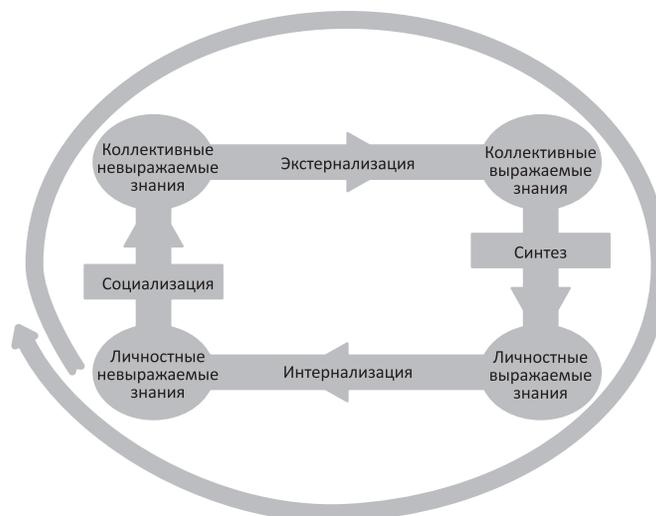


Рис. 1 Спиральная модель генерации знаний Икуджири Нонака [3, с. 69]

Используя эти четыре процесса, Нонака ввел метафорическое понятие спирали генерации знаний, каждый виток которой включает следующую последовательность: социализация → экстернализация → синтез → интернализация → социализация (как начало следующего витка спирали). Было показано на примерах, что эта спираль может служить качественной моделью итерационного процесса генерации новых знаний во время «мозгового штурма».

Обобщение и существенное развитие спиральной модели генерации знаний было предложено в работах Йошитеру Накамори и Анджея Вежбицкого в рамках создаваемой ими научной дисциплины, которую они называют «Наука о знаниях» [3–8]. В этих работах знания разделены на личностные знания человека, коллективные и конвенциональные знания. Это деление они называют социальным аспектом, или измерением, так как в результате обобщения было определено три уровня социализации знаний (от первого личностного уровня до третьего конвенционального). С учетом деления на выражаемые и невыражаемые знания в результате обобщения были определены еще два новых множества знаний, которых нет в спиральной модели: конвенциональные невыражаемые и конвенциональные выражаемые знания.

Вежбицкий и Накамори определили систему отношений между множествами знаний. Свою модель, включающую шесть множеств знаний и систему отношений между ними, они в совокуп-

ности назвали *креативным пространством* (далее в статье термины «креативное пространство» и «модель Вежбицкого–Накамори» будут использоваться как синонимы). Кроме шести множеств знаний ими были также определены множества эмоций (личностные, коллективные и конвенциональные), которые в статье не рассматриваются.

В предлагаемых далее в статье моделях не используются три множества невыражаемых знаний человека, которые по определению непосредственно не поддаются экспликации. Однако в процессе компьютерного моделирования эти три множества могут использоваться опосредованно. Например, в технологии генерации новых кросс-языковых знаний, рассматриваемой далее в статье, анализируются объекты интерпретации, которые сформированы в процессе перевода текстов на русском языке и являются результатом применения переводчиком одновременно как конвенциональных знаний, так и его личностных невыражаемых знаний¹. Этот анализ является примером извлечения и опосредованного использования послед-

Отметим, что в модели Вежбицкого–Накамори и спиральной модели, которые относятся к категории качественных, нет явно определенной оси времени. Это не дает возможности фиксировать моменты времени генерации каждого нового структурного элемента экспертных знаний. Кроме того, в этих моделях не рассматриваются объекты интерпретации, служащие источниками новых знаний.

¹ В статье по смысловому содержанию разделяются понятия невыражаемых (tacit) и подразумеваемых (implicit) знаний. Невыражаемые и незакрепленные в знаковой форме знания используются переводчиком часто неявно и могут быть известны только ему, т. е. такие знания могут являться личностными. Подразумеваемые знания косвенно выражаются в знаковой форме. Например, фраза «фирма закрыла отдел разработки прикладных программ» подразумевает, что раньше в этой фирме существовал отдел прикладного программирования. Такие знания могут быть коллективными или конвенциональными.

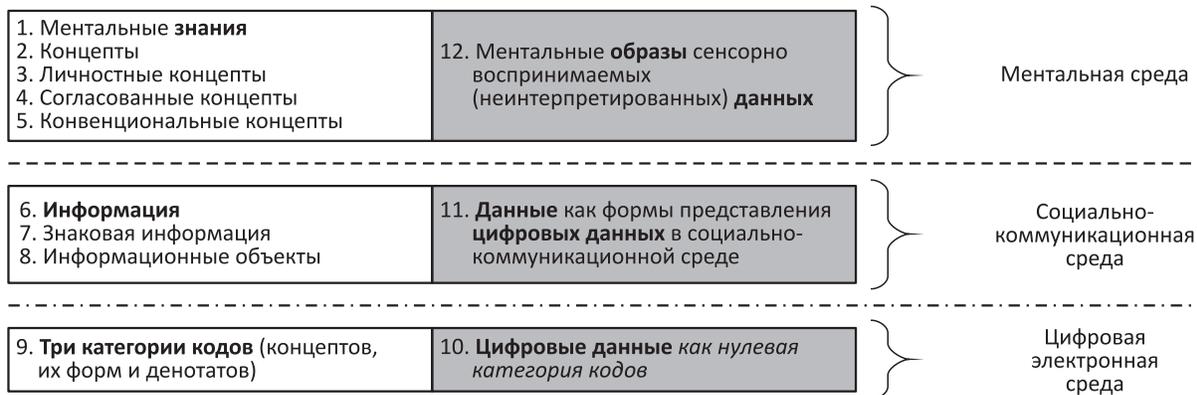


Рис. 2 Система терминов для описания объектов трех сред предметной области информатики и интерфейсов между ними [10, 11] (*денотат по определению является компонентом, отношением или свойством объекта интерпретации*¹)

Основная цель статьи заключается в описании разработанных семиотических оснований для создания количественных моделей процессов целенаправленной генерации и развития экспертных знаний, а также для разработки обеспечивающих их технологий и баз данных. С использованием этих оснований в статье предлагается развитие модели Вежбицкого–Накамори в следующих двух направлениях:

- (1) вводится ось времени, на которой фиксируются дискретные моменты времени, в которые порождаются новые элементы выражаемых экспертных знаний, а также фиксируются моменты времени их изменения;
- (2) в явном виде специфицируется связь каждого нового элемента знаний с тем объектом интерпретации, в результате анализа которого формируется этот элемент и дается формализованное его описание.

2 Семиотические основания

В процессе развития модели Вежбицкого–Накамори использовалась ранее разработанная система терминов [9, 10]. Она включает, в частности, термины и дефиниции для таких понятий, как концепты знаний человека (личностные, коллективные, конвенциональные), семантическая информация, сенсорно воспринимаемые данные, цифровая информация, цифровые данные, цифровые коды нескольких категорий и ряд других терминов, а также задает их распределение по трем средам предметной области информатики (ментальной, социально-коммуникационной и цифровой) и описание отношений между этими терми-

нами. Первая отличительная черта этой системы терминов заключается в том, что значения терминов «знания», «семантическая информация», «сенсорно воспринимаемые данные», «цифровая информация» и «цифровые данные» четко разграничены и они по определению не пересекаются по их смысловому содержанию (рис. 2). Вторая ее отличительная черта состоит в том, что она является «масштабируемой» по числу сред, так как определен принцип добавления новых сред в предметную область информатики как информационно-компьютерной науки, а также соответствующего расширения системы терминов за счет именованной объектов каждой новой среды и их интерфейсов с объектами уже существующих сред. Этот принцип получил название аксиомы герметичности сред предметной области информатики [11, 12].

Основная идея развития модели Вежбицкого–Накамори заключается в установлении связи каждого нового концепта с объектом интерпретации. При этом анализ объектов интерпретации выполняется в соответствии с явно определенными целями формирования новых знаний, так как статья посвящена именно целенаправленным процессам формирования новых выражаемых знаний на основе извлечения и экспликации невыражаемых.

Развитие модели Вежбицкого–Накамори было выполнено в несколько этапов. Сначала на первом этапе в модель были добавлены следующие шесть множеств:

- (1) формы представления личностных концептов (структурированные тексты, изображения или другие виды семантической информации);
- (2) формы представления коллективных концептов;

¹Отличие денотата от объекта интерпретации будет описано далее в примере генерации кросс-языковых знаний.

- (3) формы представления конвенциональных концептов;
- (4) цифровые коды личностных концептов и форм их представления;
- (5) цифровые коды коллективных концептов и форм их представления;
- (6) цифровые коды конвенциональных концептов и форм их представления.

В последних трех множествах разделяются коды концептов и их имен (как частного случая форм представления концептов), например коды для представления смыслового содержания слов и последовательностей их литер в цифровой среде относятся к разным категориям (см. рис. 2).

Затем для количественного описания динамики социализации, определяемой процессами согласования личностных концептов в группе экспертов, была введена ось с числами от нуля до бесконечности. Единица на оси социализации обозначает личностный концепт, $N > 1$ — коллективный, который согласован группой из N экспертов, а бесконечность — конвенциональный концепт. Необходимость в нуле на этой оси возникла в процессе проведения эксперимента для обозначения тех «бывших» личностных концептов, от которых со временем отказались их авторы. Ось социализации позволяет кодировать степень согласованности между экспертами результатов анализа динамически изменяемых объектов интерпретации и изменение степени согласованности во времени.

Таким образом, на первом этапе развития модели Вежбицкого—Накамори были определены шесть новых множеств знаний и введена ось социализации. Эта ось служит для обозначения уровня социализации не только выражаемых знаний, но также форм их представления и их цифровых кодов. Отметим, что система отношений между множествами знаний, определенная Вежбицким и Накамори, не охватывает шесть новых множеств знаний. Поэтому далее потребуется доопределить или построить новую систему отношений.

На втором этапе была добавлена ось времени, на которой фиксируются моменты порождения новых личностных, коллективных и конвенциональных концептов, а также моменты времени их изменений. Следствием этого этапа является то, что появляется возможность фиксировать на оси времени не только моменты порождения и изменения концептов, но также форм их представления и их цифровых кодов.

Третий этап заключается в добавлении множества объектов интерпретации, каждый из которых имеет уникальный цифровой код. Следствием этого этапа является то, что появляется потенциальная

возможность фиксировать на оси времени не только моменты порождения и изменения концептов, форм их представления и их цифровых кодов, но также динамику изменения соответствующих им объектов интерпретации.

Результаты трех перечисленных этапов развития модели Вежбицкого—Накамори позволяют фиксировать на оси времени:

- моменты изменения экспертами объектов интерпретации, являющихся источниками новых знаний;
- моменты порождения новых концептов в процессе интерпретации объектов и интроспекции результатов интерпретации;
- моменты изменения формируемых концептов, форм их представления (их имен) и их цифровых кодов.

Определим новую систему отношений между объектами интерпретации, концептами как элементами множеств знаний, их именами и цифровыми кодами, являющимися их идентификаторами. Она необходима потому, что система отношений в модели Вежбицкого—Накамори не охватывает шесть новых множеств знаний. Для описания отношений между объектами интерпретации, концептами и именами предлагается использовать треугольник Фреге [13–15], что представляет собой *первое семиотическое основание* моделирования процессов генерации знаний. Семиотический треугольник Фреге по определению связывает сам объект интерпретации (точнее, некоторый денотат, определяемый в процессе анализа объекта интерпретации; как правило, это его компонент, отношение или свойство), понимание денотата (его смысловое содержание), т. е. его концепт, а также некоторое имя как текстовую, или невербальную, форму обозначения денотата и его концепта. Для цифрового кодирования денотатов, концептов и имен предлагается использовать цифровой семиотический треугольник, что представляет собой *второе семиотическое основание* моделирования процессов генерации знаний. Его определение дано в работе [16], в которой рис. 6 иллюстрирует взаимосвязи этих двух треугольников.

Основная идея цифрового семиотического треугольника заключается в том, что для каждой из трех вершин треугольника Фреге используется своя категория цифровых кодов, в том числе и для концептов. Предлагаемое введение отдельной кодировки для концептов дает возможность строить взаимно однозначные отношения между концептами и цифровыми кодами, являющимися их идентификаторами. Построение таких отношений является основой компьютерного моделирования про-

цессов генерации знаний. Важно отметить, что цифровой семиотический треугольник дает возможность строить взаимно однозначные отношения, но оставляет открытым вопрос о конкретных методах приписывания цифровых кодов концептам. В ряде задач компьютерного моделирования процессов генерации знаний существует свобода выбора метода назначения кодов. Однако в задачах оценивания релевантности новых знаний явно определенным целям их генерации этот метод может быть во многом обусловлен заданными целями генерации. Примеры таких целей рассматриваются далее в статье.

Перечислим кратко основные результаты, которые были получены в процессе развития модели Вежбицкого—Накамори. С помощью цифр 1 и 2 в списке отмечены положения, взятые из модели Вежбицкого—Накамори (1) и полученные в результате развития этой модели (2):

- три множества выражаемых знаний (личностные, коллективные и конвенциональные) (1), которые состоят из концептов соответствующих категорий (2);
- три множества имен (форм представления концептов) (2);
- множество объектов интерпретации (2);
- три множества цифровых кодов концептов (2) (личностные, коллективные и конвенциональные);
- три множества цифровых кодов форм представления концептов (2);
- множество цифровых кодов, построенное на основе уникальных идентификаторов объектов интерпретации (2);
- ось времени для отражения динамики процессов генерации знаний (2);
- ось социализации выражаемых экспертных знаний (1 и 2, так как в модели Вежбицкого—Накамори нет детализации коллективных знаний в зависимости от числа экспертов в группе, а в результате развития этой модели добавлена их детализация);
- система отношений между объектами интерпретации, денотатами, концептами выражаемых знаний, именами и их цифровыми кодами (1 и 2, так как система отношений задана в модели Вежбицкого—Накамори только между знаниями с учетом трех уровней их социализации; в результате развития модели система отношений дополнена треугольником Фреге и цифровым семиотическим треугольником).

Для описания еще одного, третьего, семиотического основания необходимо вернуться к определению классического треугольника Фреге. В семиотике он определяется как треугольник с тремя вершинами (денотат, концепт как идеальная вершина треугольника, имя как форма представления концепта), находящимися в отношениях устойчивой связи, опосредованной сознанием, представляет собой устойчивое единство, которое посредством сенсорно воспринимаемой формы *конвенционально* репрезентирует концепт и денотат. Строго говоря, в новой системе отношений классический треугольник Фреге применим только для случая конвенциональной репрезентации концепта сенсорно воспринимаемой формой. Следовательно, дополнительно необходим некоторый способ для построения личностного и коллективного семиотических треугольников Фреге, аналогичных классическому треугольнику Фреге и применимых в случае генерации личностных и коллективных концептов. При этом необходимо учитывать то обстоятельство, что личностные и коллективные концепты понимают только их авторы, так как для них отсутствует конвенциональная репрезентация их формой. Следовательно, для экспертов, участвующих в процессе формирования новых знаний, но не являющихся авторами генерируемых новых концептов, они будут недоступны, если их авторы не эксплицируют свою личностную или коллективную репрезентацию в форме, доступной другим экспертам тем или иным способом.

Отметим, что в процессе личностной репрезентации кроме объекта интерпретации, денотата, концепта и имени «задействовано» персональное авторское сознание. В приведенном определении классического треугольника Фреге говорится об устойчивой связи, *опосредованной сознанием*, но ничего не говорится о том, как и где (в какой среде или средах) эта связь закреплена материально. Остановимся кратко на истории этого вопроса, чтобы затем предложить способ экспликации личностной репрезентации концепта в форме, доступной другим экспертам.

В 1988 г. была сформирована рабочая группа «FRamework of Information System COnccepts — FRISCO» в рамках Международной федерации по обработке информации (International Federation for Information Processing — IFIP). Основной целью этой группы было создание системы определений для базовых терминов, которую затем можно было бы предложить использовать как терминологическую основу разработки и описания информационных систем. Итоги ее работы опубликованы в виде отчета в 1998 г. [17]. В результате работы группы FRISCO, в частности, было определено понятие

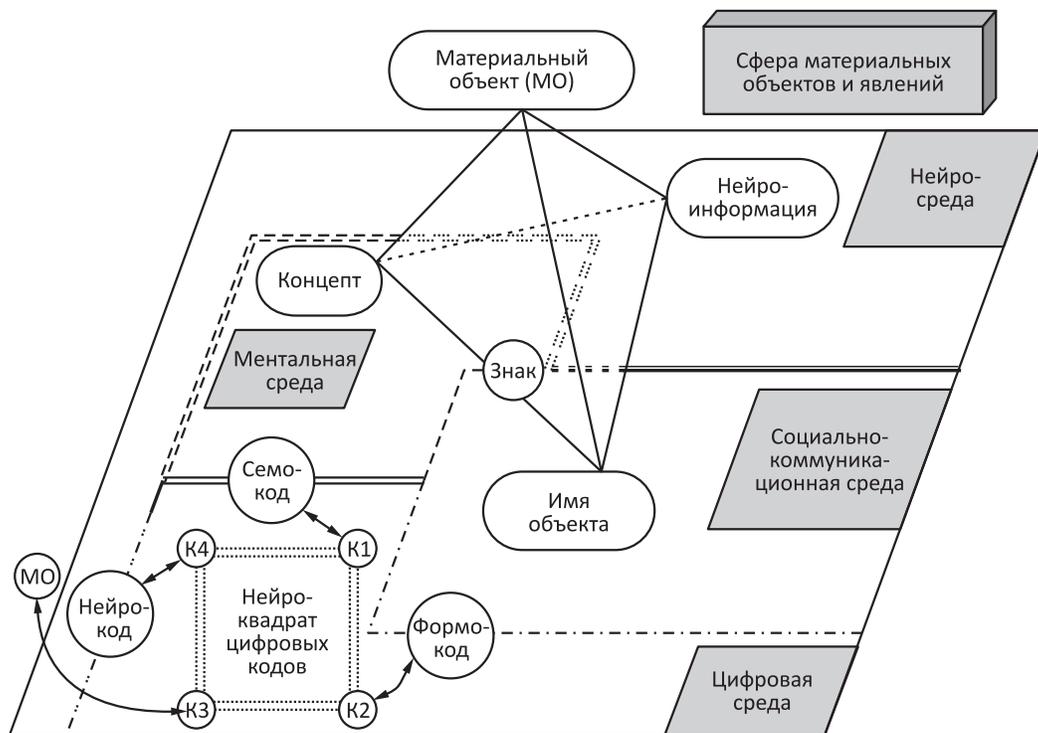


Рис. 3 Семиотический нейротетраэдр и нейроквадрат [11]

семиотического тетраэдра с вершинами «объект, концепт, имя объекта и интерпретатор». Отметим, что в определении этого понятия есть субъект, который *интерпретирует* объект, генерирует концепт и имя объекта, а также *устанавливает* связь между ними [18]. Идея субъекта-интерпретатора или интерпретанта¹, который *является носителем* этой связи, зародилась еще раньше в общей семиотике и в рамках этой научной дисциплины оказалась весьма продуктивной [19, 20].

Однако с точки зрения разработки и описания информационных систем и технологий включение субъектов в дефиниции терминов вместо объектов обладает рядом недостатков. Основной из них заключается в том, что в задачах когнитивной информатики и нейроинформатики, например при разработке нейрокоммуникаторов, необходимо различать объекты *ментальной, нейрофизиологической, социально-коммуникационной и цифровой* сред. В традиционных моделях с субъектами-интерпретаторами и интерпретантами это различие между объектами трудно провести, так как в них ментальная среда и нейрофизиологическая среда (далее — нейросреда), а также их объекты, как правило, не различаются. Поэтому в работе [11] эти две

среды было предложено рассматривать в информатике отдельно, а систему терминов дополнить следующими понятиями (рис. 3):

- «нейроинформация», в частности фиксирующая постоянные или временные связи между объектом интерпретации, концептом (личностным, коллективным или конвенциональным) и именем объекта;
- «нейросемиотический тетраэдр» с вершинами объект, концепт, имя и нейроинформация;
- «нейроквадрат» как четверка кодов следующих категорий:

- K1 — для концептов ментальных знаний человека;
- K2 — для слов как имен объектов и других знаковых форм для обозначения ментальных знаний;
- K3 — для кодирования предметов материальной сферы (в общем случае — для кодирования любых объектов, в результате семантической интерпретации которых человеком определяются денотаты и генерируются концепты);

¹Интерпретатор — это только человек, который анализирует предмет, генерирует концепт и его имя, а интерпретант — это не обязательно человек. В общей семиотике функция абстрактного интерпретанта заключается в интерпретации предметов, генерации концептов и присвоении им имен [19, с. 15, 16].

К4 — для кодирования в цифровой среде нейроинформации, с помощью которой фиксируются связи между объектом интерпретации, концептом и именем;

- «нейрокод» как аналог компьютерных таблиц кодировки символов для нейроинформации.

Определение нейросемиотического тетраэдра по своему содержанию во многом совпадает с термином «психосемиотический тетраэдр» Ф. Е. Василюка с вершинами «предмет, личностный концепт, имя предмета и чувственная ткань (которая связывает воедино первые три вершины)» [21]. Различие этих двух тетраэдров заключается в том, что первый описывает любые концепты (личностный, коллективный и конвенциональный), а второй был определен Ф. Е. Василюком только для личностных концептов. В нейросемиотическом тетраэдре три из четырех его вершин являются сущностями трех разных сред: ментальной, социально-коммуникационной и нейросреды, к которой принадлежит нейроинформация. Таким образом, в определении нейросемиотического тетраэдра использовался альтернативный подход по сравнению с традиционной идеей субъекта-интерпретатора или интерпретанта.

Сопоставление рис. 3 и рис. 6 из работы [16], в которой даны определения понятий «формокод» и «семокод», наглядно иллюстрирует то, что нейроквадрат является обобщением цифрового семиотического треугольника. Итак, нейротетраэдр и нейроквадрат являются *третьим семиотическим основанием* моделирования процессов генерации знаний.

3 Семиотические модели

В этом разделе три семиотических основания, рассмотренные в предыдущем разделе, будут использованы в процессе построения двух моделей генерации знаний: модели фиксированного состояния и нестационарной модели. Они являются результатом обобщения двух ранее разработанных моделей, основанных на треугольнике Фреге и цифровом семиотическом треугольнике [16, 22].

Кроме трех семиотических оснований исходными данными для обобщения является следующее описание этих двух моделей, разработанных для итерационных процессов генерации знаний экспертами, которые используют информационную систему для фиксации эволюции объектов интерпретации, определения денотатов, сгенерированных ими концептов, описания которых эксперты

формируют в процессе интроспекции, имен концептов и денотатов. По определению из работы [16] первая модель фиксирует *состояние процесса генерации в момент времени*, соответствующий окончанию некоторой итерации этого процесса, и состоит из:

- трех сред предметной области информатики: ментальной, социально-коммуникационной и цифровой;
- треугольника Фреге, включающего объект интерпретации, концепт, сгенерированный или измененный на этой итерации некоторым экспертом, и имя;
- цифрового семиотического треугольника, включающего коды объекта интерпретации, концепта и имени, сгенерированные в этот же момент времени информационной системой, обеспечивающей работу экспертов.

Согласно первой модели три вершины треугольника Фреге кодируются тремя цифровыми кодами разных категорий, которые генерируются в конце каждой итерации:

- семантическим кодом концепта (К1);
- информационным кодом его имени, если оно создано экспертом (в противном случае этот код равен нулю) (К2);
- объектным кодом объекта интерпретации (К3).

Таким образом, после завершения каждой итерации в информационной системе по некоторому заданному алгоритму генерируются три цифровых кода разных категорий. Если на некоторой итерации принял участие один эксперт, то создается только одна запись с результатами личностной семантической интерпретации рассматриваемого объекта и три кода, а если несколько экспертов, то для каждого из них — одна запись и три кода.

В процессе целенаправленной генерации знаний существует отдельный вид итераций для согласования личностных концептов и имен, которые могут меняться в пределах итераций, но не между ними. На этих итерациях эксперты ставят своей целью согласовать между собой свои личностные интерпретации и сформировать коллективные концепты и имена. Если это удастся сделать, то в информационной системе создается еще одна запись с результатами коллективной семантической интерпретации и еще три кода с указанием идентификаторов всех экспертов, которые приняли участие в процессе согласования и выработали единую позицию. В общем случае генерация и согласование концептов могут быть совмещены в рамках одной комплексной итерации. Отметим, что между двумя любыми итерациями объекты интерпретации могут меняться экспертами, но не в пределах итераций.

Регистрация в информационной системе изменений объектов интерпретации между итерациями, описаний новых концептов и результатов их согласования дает возможность восстановить ретроспективно все этапы процесса генерации знаний.

Вторая модель предназначена для описания динамики процесса генерации [22]. Концепты, имена и объекты интерпретации могут изменяться в широком диапазоне в процессе согласования экспертами их личностных концептов и имен. По определению вторая модель описывает динамику процесса генерации концептов одним экспертом и состоит из:

- трех сред предметной области информатики: ментальной, социально-коммуникационной и цифровой;
- треугольников Фреге, построенных экспертом в моменты времени окончания итераций $t_i, i = 1, 2, \dots$;
- цифровых семиотических треугольников, построенных информационной системой в эти же моменты времени t_i .

На основе этой модели динамики процесса генерации концептов в работе [22] было дано определение пространства Фреге как 4-мерного множества точек для трех кодов разных категорий $\{t_i, \text{семантический код } (t_i), \text{информационный код } (t_i), \text{объектный код } (t_i) \text{ при } i = 1, 2, \dots\}$, сгенерированных информационной системой в процессе работы одного эксперта. Аналогичную модель и соответствующее пространство Фреге можно определить для случая генерации знаний коллективом экспертов, пример которого рассматривается в работе [23].

Пространство Фреге имеет три оси координат цифровых кодов: семантическую, информационную и объектную, а также четвертую — ось времени, содержащую дискретный набор точек начала и окончания итераций генерации концептов, их согласования или комплексных итераций. Пространство Фреге дает возможность представить графически динамику процесса, используя последовательности значений семантических, информационных и объектных кодов, сгенерированные информационной системой в дискретные моменты времени окончания итераций.

Отметим, что вид дискретных траекторий точек будет определяться алгоритмами назначения семантических, информационных и объектных кодов. Можно ли задать некоторую метрику в пространстве Фреге? В настоящее время этот вопрос остается открытым. Пока этот вопрос находится в стадии изучения, термин «пространство Фреге» определен только как семиотическое понятие, но

не математическое. При этом если в информационной системе фиксируется содержательная эволюция объектов интерпретации, сгенерированных концептов, описания которых эксперты формируют в процессе интроспекции, и имен, то пространство Фреге служит для количественного описания процесса генерации концептов.

Прежде чем приступить к построению модели фиксированного состояния и нестационарной модели на основе обобщения двух ранее разработанных моделей, отметим, что в последних использовались только три среды предметной области информатики (ментальная, социально-коммуникационная и цифровая среды) и первые два семиотических основания для моделирования процессов генерации знаний (треугольник Фреге и цифровой семиотический треугольник). Перечислим исходные данные построения модели фиксированного состояния и нестационарной модели на основе обобщения двух ранее разработанных моделей:

- определение семиотического нейротетраэдра и нейроквадрата, которые служат третьим семиотическим основанием;
- первая исходная модель, которая фиксирует состояние процесса генерации концептов в момент времени, соответствующий окончанию некоторой итерации этого процесса;
- вторая исходная модель, которая описывает динамику процесса генерации концептов.

Для обобщения перечисленных двух моделей рассмотрим четыре среды предметной области информатики как информационно-компьютерной науки: ментальную, социально-коммуникационную, нейро- и цифровую среды [11]. С формальной точки зрения это обобщение представляет собой замену треугольника Фреге на семиотический тетраэдр, а цифрового семиотического треугольника — на нейроквадрат кодов (см. рис. 3). Сделав такую замену, получаем два следующих обобщения.

Для случая четырех сред модель фиксированного состояния процесса генерации знаний состоит из:

- ментальной, социально-коммуникационной, цифровой сред и нейросреды;
- семиотического тетраэдра, включающего объект интерпретации, концепт, сгенерированный или измененный на этой итерации некоторым экспертом, имя объекта интерпретации, которое одновременно является и именем концепта в этот момент времени, а также нейроинформацию о связях между объектом, концептом и их именем;

- нейроквadrата, включающего цифровые коды объекта, концепта, имени и связывающей их нейроинформации, сгенерированные в этот же момент времени информационной системой, обеспечивающей работу экспертов.

Нестационарная модель динамики процесса генерации знаний одним экспертом состоит из:

- тех же самых четырех сред предметной области информатики;
- семиотических тетраэдров, построенных экспертом в моменты времени окончания итераций $t_i, i = 1, 2, \dots$;
- нейроквadrатов, построенных в эти же моменты времени t_i .

На основе этой модели динамики процесса аналогично можно определить обобщенное пространство Фреге как 5-мерное множество точек для четырех кодов разных категорий $\{t_i, \text{семантический код } (t_i), \text{информационный код } (t_i), \text{объектный код } (t_i), \text{код нейроинформации } (t_i) \text{ при } i = 1, 2, \dots\}$, сгенерированных информационной системой на i -й итерации работы эксперта.

Главное содержание приведенного формального обобщения состоит в замене треугольника Фреге на семиотический тетраэдр, основное отличие которого от треугольника Фреге заключается в наличии нейроинформации. По определению она фиксирует связи между объектом, концептом и именем. Однако остается открытым вопрос о практических способах получения нейроинформации об этих связях в процессе решения прикладных задач. Раньше, когда такие задачи решались с использовани-

ем моделей, которые охватывали объекты только трех сред, объекты интерпретации были доступны экспертам для изменений и анализа, так как они представляли собой изменяемые во времени:

- компьютерные программы и данные, используемые для вычисления значений новых индикаторов [9, 23];
- фрагменты параллельных текстов на русском и французском языке (рис. 4), в результате контрастивного анализа которых определялись денотаты и формировались их кросс-языковые концепты в процессе анализа параллельных фрагментов [24, 25].

Новые концепты знаний, принадлежащие ментальной среде, описывались экспертами в результате субъективной интроспекции с последующим присвоением имен сформированным ими концептам. Иначе говоря, в моделях, которые охватывали объекты трех сред, эксперты сами анализировали объекты интерпретации, описывали концепты и давали имена. После расширения числа сред до четырех в обобщенных моделях появляется нейроинформация о связях между объектом интерпретации, концептом и именем, которая экспертам недоступна. Поэтому и возникает вопрос о способах получения нейроинформации в процессе решения практических задач.

Сегодня есть возможность отобразить в компьютерной форме уровень активности разных участков мозга экспертов в режиме реального времени, используя метод функциональной магнитно-резонансной томографии (functional Magnetic

Номер пары	Оригинальный текст	Перевод
9	Цвет лица у Ильи Ильича не был ни румяный, ни смуглый, ни положительно бледный, а безразличный или казался таким, может быть, потому, что Обломов как-то обрюзг не по летам: от недостатка ли движения или воздуха, а может быть, того и другого.	Le teint d'Ilya Ilitch n'était ni rose, ni hâlé, ni carrément pâle, mais indifférent ou, du moins, il le paraissait. Peut-être parce que la chair d'Oblomov était prématurément flasque: faute d'exercice ou manque d'air, peut-être l'un et l'autre.
18	Халат имел в глазах Обломова тьму неопценных достоинств: он мягок, гибок; тело не чувствует его на себе; он, как послушный раб, покоряется самомалейшему движению тела.	Aux yeux d'Oblomov cette robe de chambre avait une foule de qualités inappréciables: elle était douce, souple, ne pesait pas sur le corps; telle une esclave docile, elle se pliait au moindre mouvement.
21	Лежанье у Ильи Ильича не было ни необходимостью, как у больного или как у человека, который хочет спать , ни случайностью, как у того, кто устал, ни наслаждением, как у лентяя: это было его нормальным состоянием.	La position allongée n'était pour Ilya Ilitch ni nécessaire, comme pour un malade ou pour un homme qui veut dormir , ni accidentelle, comme pour une personne fatiguée, ni voluptueuse comme chez le fainéant; c'était son état normal.

Рис. 4 Три предложения параллельных текстов на русском языке и их переводы (полужирным шрифтом выделен контекст, используемый далее на рис. 6)

Resonance Imaging — fMRI) [26]. Этот метод позволяет использовать объективные индикаторы уровня активности, наблюдая количественные измерения мозговой деятельности, одновременно фиксируя и описывая концепты как результаты личностного анализа экспертами объектов интерпретации в процессе субъективной интроспекции.

Но и здесь возникают вполне закономерные вопросы. Можно ли использовать объективные индикаторы уровня активности в процессе генерации новых концептов для описания связей между объектом интерпретации, концептом и именем, а также для сопоставления с результатами личностного анализа? Можно ли их использовать для описания процесса согласования новых концептов между экспертами?

В настоящее время действительно есть возможность наблюдать одновременно и количественные данные измерения мозговой деятельности, и результаты субъективного мышления в процессе субъективной интроспекции, но из первых сегодня трудно получить именно ту нейроинформацию, которая соответствует личностным или согласованным концептам экспертов, чтобы провести ее сопоставление с результатами личностного субъективного анализа. Кроме того, есть гипотеза и подтверждающие ее экспериментальные данные, что у экспертов часть нейроинформации, соответствующей устоявшимся конвенциональным концептам знаний, носит структурный характер, скорее всего, на уровне связей между нейронами долговременной памяти, что не фиксируется fMRI и другими современными методами. Стремительное развитие когнитивной нейронауки и ее инструментальных средств позволяет надеяться, что в будущем станет возможным соотнести структурную нейроинформацию и количественные нейроданные измерений мозговой деятельности с устоявшимися и новыми концептами экспертных знаний [26–29].

Однако в настоящее время при разработке информационных технологий и решении практических задач, когда недоступна *объективная нейроинформация* о связях между объектом интерпретации, концептом и именем, предлагается по-прежнему использовать результаты *субъективной интроспекции*. Иначе говоря, в процессе итерационной генерации новых знаний эксперты в информационной системе должны описывать не только свои концепты и присваивать им имена, но также устанавливать и фиксировать их связи с объектами интерпретации, определенными ими денотатами и присвоенными именами.

Предлагаемый подход позволяет уже сегодня использовать обобщенные модели при разработке информационных технологий, поддерживающих ге-

нерацию новых знаний, и решении практических задач. Кроме того, использование цифровой среды информационной системы как носителя этих связей обеспечит доступ всех экспертов к описаниям личностных и коллективных концептов. Другими словами, анализировать и обсуждать описания таких концептов смогут все эксперты, а не только их авторы, если в процессе итерационной генерации новых знаний эксперты в информационной системе описывают свои концепты, устанавливают и фиксируют их связи с объектами интерпретации, определенными ими денотатами и присвоенными именами.

4 Технология, обеспечивающая генерацию знаний

Модели фиксированного состояния и динамики процесса генерации знаний были использованы при разработке информационной технологии, обеспечивающей целенаправленную генерацию и развитие кросс-языковых знаний коллективом экспертов. Необходимость разработки подобной технологии проявляется наиболее наглядно в ситуации, когда необходимо повысить качество машинного перевода и для этого требуется существенное развитие контрастивных грамматик на основе формирования новых кросс-языковых знаний. При этом направления развития контрастивных грамматик должны определяться явно эксплицированными целями, достижение которых и должно непосредственно способствовать повышению качества машинного перевода.

При таком подходе кроме моделей состояния и динамики процесса генерации знаний необходимо использовать некоторый способ описания целей. Как было уже отмечено, рассматриваемый подход к моделированию процесса генерации знаний при явном описании целей и разработке обеспечивающей технологии ориентирован на те прикладные области, где генерируемые экспертные знания являются результатом анализа объектов интерпретации. В рассматриваемом примере целенаправленного формирования кросс-языковых знаний объектами интерпретации являются предложения параллельных текстов на русском и французском языках, а денотатами — пары тех параллельных фрагментов, которые выделяются экспертами согласно рассматриваемому ими направлению развития контрастивной грамматики (выделенные полужирным шрифтом на рис. 4 параллельные фрагменты в паре № 21 станут далее одним из объектов анализа).

Отличительная черта предлагаемого подхода к моделированию заключается в явном описании отношений между новыми экспертными знаниями, объектами интерпретации и денотатами, на основе анализа которых могут быть сгенерированы элементы новых знаний (т. е. не каждый анализируемый объект и определенный в процессе анализа денотат всегда порождает новый концепт). Реализуемость такого подхода была продемонстрирована в процессе выполнения контрастных исследований, включающих задачи целенаправленной генерации кросс-языковых знаний:

- о переводах глагольных конструкций русского языка на французский;
- о возможных вариантах перевода лингвистических слов русского языка на французский.

При проведении этих контрастных исследований кросс-языковые знания формировались экспертами в процессе анализа параллельных текстов на русском и французском языках с использованием НДБ [30, 31]. Отметим, что переводной текст является результатом применения перевод-

чиком как конвенциональных знаний (в этом случае анализ соответствующих параллельных текстов не приводит к генерации новых концептов), так и его невыражаемых знаний, что может привести к генерации новых концептов. Невыражаемые знания могут использоваться переводчиками неявно, при этом быть новыми и неописанными в контрастных грамматиках в явной (эксплицитной) форме. В приведенных далее примерах рассматриваются оригинальные тексты на русском языке, при переводе которых на французский язык невыражаемые знания использовались переводчиками, что и нашло свое отражение в результатах перевода. Поэтому результаты сопоставления оригинальных текстов на русском языке и их переводов могут помочь сформировать и описать новые знания.

Разработанная технология [24, 32, 33], обеспечивающая генерацию и целенаправленное формирование кросс-языковых знаний, основана на методике, созданной Анной А. Зализняк [32–34], и включает следующие основные этапы (рис. 5):

- из корпуса параллельных текстов отбираются пары предложений как объекты интерпрета-

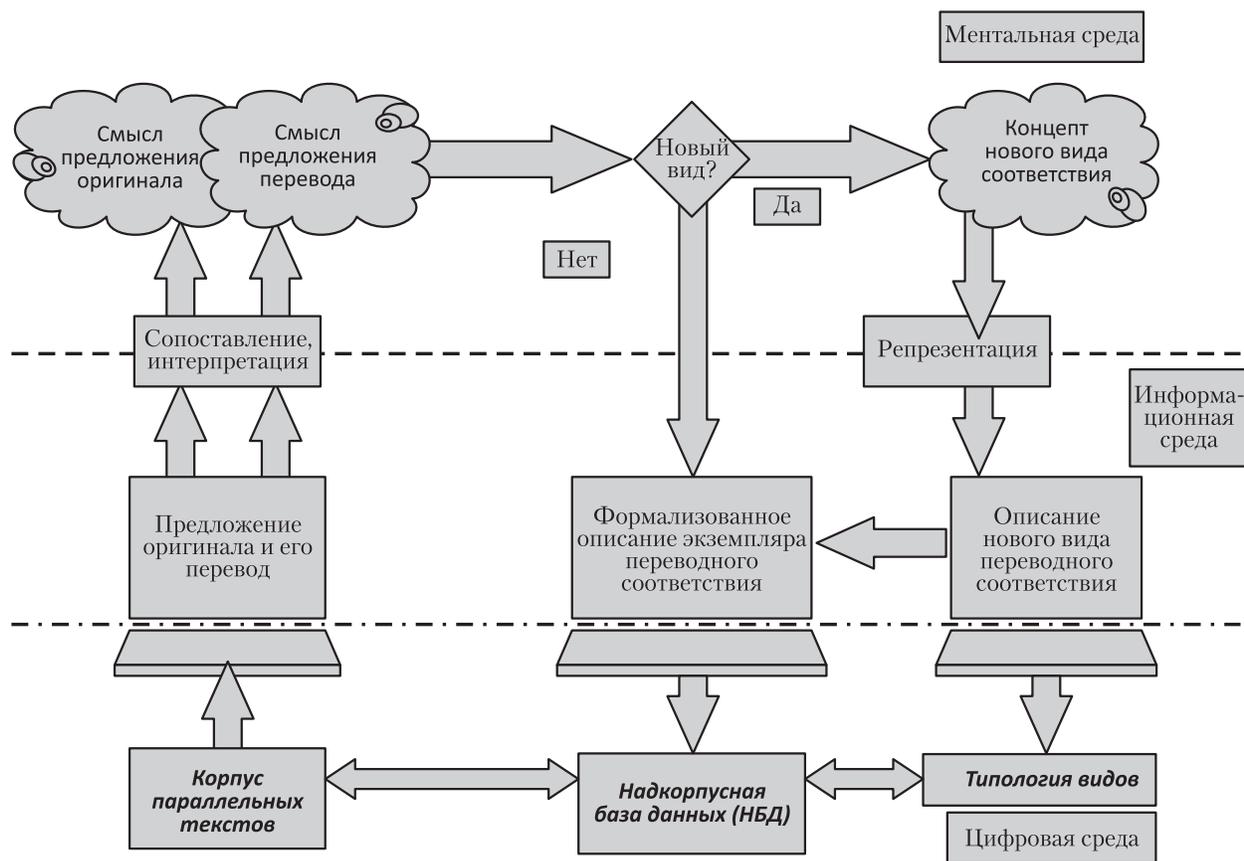


Рис. 5 Основные этапы технологии (нейросреда не показана; социально-коммуникационная среда для краткости обозначена как «Информационная среда»)

ции, содержащие исследуемые языковые объекты (см. пару выделенных фрагментов на рис. 4, которая является примером денотата)¹;

- в каждой из отобранных пар предложений эксперты анализируют перевод исследуемого языкового объекта на французский язык и определяют его функционально эквивалентный фрагмент (ФЭФ), в терминологии Д. О. Добровольского;
- языковой объект текста оригинала сопоставляется с его ФЭФ согласно заданному направлению развития контрастивной грамматики²;
- результат сопоставления описывается экспертами в формализованном виде с использованием методики Анны А. Зализняк;
- если вариант перевода исследуемого языкового объекта из текста оригинала уже включен в существующие контрастивные грамматики (это случай пары выделенных фрагментов на рис. 4), то формализованное описание анализируемого экземпляра переводного соответствия вводится в НБД без дополнения (развития) контрастивной грамматики;
- если вариант перевода исследуемого языкового объекта из текста оригинала экспертами считается новым, то формализованное описание этого экземпляра переводного соответствия вводится в НБД, а в типологию включается новый вид соответствия исследуемого языкового объекта и его ФЭФ;
- одновременно НБД генерирует четыре цифровых идентификатора (см. рис. 3) для обозначения:
 - объекта интерпретации с выделенным в нем денотатом, который представляет собой пару параллельных текстовых фрагментов;
 - концепта денотата как исследуемого языкового объекта или явления, который эксперты описывают в формализованном виде;
 - имени концепта, которое одновременно является и именем денотата;
 - связей между объектом интерпретации, включающим денотат, концептом и именем.

¹На рис. 4 приведена пара предложений № 21 с глаголом настоящего времени русского языка, который в приведенном далее примере станет исследуемым языковым объектом. На момент проведения эксперимента, описанного в статье, в русско-французском подкорпусе Национального корпуса русского языка было около 4 тыс. пар с глаголами настоящего времени (сейчас их около 20 тыс.).

²Например, целью развития контрастивной грамматики может быть формирование расширенного списка вариантов перевода глагольных конструкций, включая низкочастотные варианты, в том числе такие варианты, которые могут отсутствовать в существующих описаниях контрастивной грамматики [35, 36], но которые могут использоваться переводчиками. Тогда их можно извлекать в процессе сопоставления текстов оригинала и перевода.

Реализуемость разработанной технологии была проверена в процессе проведения эксперимента по анализу переводов глагольных конструкций русского языка на французский [25]. Целью анализа было развитие типологии видов соответствия исследуемых глагольных конструкций и их ФЭФ. Из русско-французского подкорпуса Национального корпуса русского языка (НКРЯ) были отобраны около 4000 пар предложений, содержащих глагольные конструкции настоящего времени русского языка и их переводы на французский. Согласно работам [35, 36] они могут быть переведены с помощью следующих 9 грамматических конструкций: *présent*, *imparfait*, *infinitif*, *passé composé*, *futur simple*, *subjonctif présent*, *gérondif*, *futur immédiat* и *impératif*.

Таким образом, до начала эксперимента типология видов включала девять записей для русского настоящего времени (НастВ). Первая запись имела вид (НастВ, *présent*), вторая — (НастВ, *imparfait*) и т.д. до (НастВ, *impératif*). Во время эксперимента эксперты сравнивали оригинальный и переведенный текст в отобранных парах предложений, выделяя глагольную конструкцию НастВ в тексте оригинала и ее ФЭФ в тексте перевода, которые в совокупности являются денотатом.

Если вариант перевода глагольной конструкции НастВ уже был изначально включен в типологию (видов соответствия), то формализованное описание этой конструкции и ее ФЭФ (рис. 6) добавляются в НБД без изменения типологии видов. Если вариант перевода глагольной конструкции НастВ эксперты считают новым, то формализованное описание этой конструкции и ее ФЭФ добавляются в НБД, а в типологию включается новый вид (см. рис. 5). Формализованное описание создается экспертами на основе смыслового содержания соответствия глагольной конструкции и ее ФЭФ. Это смысловое содержание вида соответствия и является тем концептом, который формируется в процессе анализа этого соответствия. Если концепт оказывается новым, то типология дополняется его именем в формате (вид глагольной конструкции, вид ее ФЭФ).

Данные эксперимента по извлечению и описанию новых концептов, полученные на первом и втором его этапах, приведены в табл. 1 и 2 соответственно.

Контекст глагольной конструкции	Вид глагольной конструкции и грамматические признаки ее контекста	Контекст ФЭФ	Вид конструкции ФЭФ и грамматические признаки его контекста
как [. . .] у человека, который хочет спать ,	НастВ < SubInf-IPF > < SubAttr >	comme [. . .] pour un homme qui veut dormir ,	Present < SubInf > < SubAttr >

Рис. 6 Формализованное описание глагольной конструкции и ее ФЭФ с известным типологическим видом соответствия (НастВ, présent) (контекст извлечен из пары № 21 на рис. 4)

Таблица 1 Четыре новых типологических вида

№ п/п	Типологический вид соответствия	Число экземпляров вида в НБД	Статус вида (известный до начала эксперимента или новый)
1	(НастВ, présent)	311	известный
2	(НастВ, imparfait)	53	известный
3	(НастВ, infinitif)	15	известный
4	(НастВ, passé composé)	8	известный
5	(НастВ, conditionnel présent)	7	новый
6	(НастВ, futur simple)	5	известный
7	(НастВ, subjonctif présent)	4	известный
8	(НастВ, participe passé)	2	новый
9	(НастВ, gérondif)	2	известный
10	(НастВ, subjonctif imparfait)	1	новый
11	(НастВ, plus-que-parfait)	1	новый
12	(НастВ, futur immédiat)	0	известный
13	(НастВ, impératif)	0	известный
Всего записей		409	

Таблица 2 Восемь новых типологических видов

№ п/п	Типологический вид соответствия	Число экземпляров вида в НБД	Статус вида (известный до начала эксперимента или новый)
1	(НастВ, présent)	1587	известный
2	(НастВ, imparfait)	328	известный
3	(НастВ, infinitif)	71	известный
4	(НастВ, passé composé)	30	известный
5	(НастВ, conditionnel présent)	23	новый
6	(НастВ, participe passé)	22	новый
7	(НастВ, subjonctif présent)	19	известный
8	(НастВ, futur simple)	19	известный
9	(НастВ, participe présent)	19	новый
10	(НастВ, gérondif)	15	известный
11	(НастВ, futur immédiat)	10	известный
12	(НастВ, passé simple)	10	новый
13	(НастВ, plus-que-parfait)	8	новый
14	(НастВ, subjonctif imparfait)	6	новый
15	(НастВ, impératif)	5	известный
16	(НастВ, infinitif passé)	3	новый
17	(НастВ, passé immédiat)	1	новый
Всего записей		2176	

Таблица 1 содержит результаты первого этапа анализа 409 пар предложений из 4000, т.е. приблизительно 10% от общего их числа. На этом этапе эксперты выявили и описали четыре новых типологических вида перевода русского настоящего времени, которым присвоили следующие имена: (НастВ, *conditionnel présent*), (НастВ, *participe passé*), (НастВ, *subjonctif imparfait*) и (НастВ, *plus-que-parfait*). В то же самое время они не нашли примеры вариантов перевода с глагольными конструкциями *futur immédiat* и *impératif*.

Таким образом, первый вариант описания цели развития этой типологии видов мог бы состоять в том, чтобы найти примеры для всех изначально известных девяти типологических видов и описать найденные новые виды (как минимум найти и описать один новый вид). В этом случае обработка 409 пар недостаточна, так как для достижения такой цели эксперты должны продолжать искать примеры французских переводов с глагольными конструкциями *futur immédiat* и *impératif* (см. табл. 1).

Таблица 2 содержит результаты второго этапа анализа 2176 пар предложений из 4000, т.е. приблизительно 54% от общего их числа. Эксперты выявили и описали еще четыре новых типологических вида перевода русского настоящего времени: (НастВ, *participe présent*), (НастВ, *passé simple*), (НастВ, *infinitif passé*) и (НастВ, *passé immédiat*) (см. табл. 2).

Одновременно они нашли французские переводы с глагольными конструкциями *futur immédiat* и *impératif*. В итоге проведенного эксперимента все восемь новых типологических видов были добавлены экспертами к девяти уже имеющимся в типологии. Дополненная типология включала 17 видов после обработки 2176 пар предложений из 4000 (см. табл. 2).

Второй вариант описания цели развития этой типологии видов мог бы состоять в том, чтобы найти примеры для всех изначально известных девяти видов, описать найденные новые типологические виды (как минимум найти и описать один новый вид) при условии, что эксперты должны обработать не менее чем 55% от всех пар предложений с глагольной конструкцией *НастВ*, имеющихся в русско-французском подкорпусе НКРЯ, т.е. 55% от 4000 на момент проведения эксперимента. В этом случае анализ 2176 пар был недостаточен и эксперты должны были бы продолжить свою работу, пока не будет обработано 2200 пар.

Кроме развития типологии видов соответствия глагольной конструкции и ее ФЭФ разработанная технология в настоящее время используется для формирования списков возможных вариантов перевода лингвоспецифичных слов русского языка на

французский язык [37], а также для формирования методологии контрастного корпусного исследования категории безличности в русском языке. Таким образом, компьютерная поддержка процессов генерации и целенаправленного развития экспертами кросс-языковых знаний была опробована в процессе развития контрастной русско-французской грамматики для глагольных конструкций и продолжает применяться для контрастных исследований лингвоспецифичных слов русского языка.

5 Заключение

Двухязычные параллельные корпуса, в которых каждому тексту на русском языке соответствует один или несколько его переводов на другой язык, являются потенциальным и неисчерпаемым источником генерации новых, но трудно извлекаемых кросс-языковых знаний. Являясь уникальным и постоянно пополняемым, он может быть использован для существенного повышения качества машинного перевода, актуализации моно- и двухязычных грамматик, а также для обновления широкого спектра образовательных курсов по лингвистике, теории и практике перевода.

Однако функциональность традиционных электронных корпусов не обеспечивает извлечения тех невыражаемых знаний переводчиков, которые менялись ими в процессе перевода. Эти знания могут быть личностными или коллективными и передаваться в процессе их социализации (см. рис. 1), например в процессе демонстрации образцов перевода в процессе обучения, но при этом они могут продолжать оставаться невыражаемыми и неэксплицированными. Наблюдается парадокс: с одной стороны, в электронных корпусах есть образцы переводов, полученные с применением невыражаемых знаний переводчиков; с другой стороны, традиционные параллельные корпуса не могут поддержать процессы извлечения и экспликации этих знаний. Поэтому понадобилось существенное дополнение функциональности традиционных корпусов за счет разработки новой информационной технологии целенаправленной генерации знаний.

Разработка этой технологии была связана с развитием семиотических оснований информатики как информационно-компьютерной науки и созданием новых моделей целенаправленной генерации и развития новых знаний с использованием четырех сред ее предметной области. К ментальной, социально-коммуникационной и цифровой средам была добавлена нейросреда. Ее добавление стало основой для определения нейросемиотического тетраэдра, что представляет собой качественно

новое развитие понятия семиотического тетраэдра, предложенного группой FRISCO в конце прошлого века.

Суть этого развития в том, что предложено разделять объекты ментальной среды и нейросреды в предметной области информатики. На практике такое деление уже используется в процессе разработки ряда когнитивных и нейрокоммуникационных технологий. Следовательно, это должно найти свое отражение и в теоретических основаниях информатики, а также в образовательных курсах по ее изучению в системе среднего и высшего профессионального образования. Отметим, что разделение объектов ментальной среды и нейросреды существенно увеличивает спектр интерфейсов, которые являются новыми объектами исследований в информатике [11].

Адаптация разработанной технологии для проведения контрастных исследований повлекла за собой необходимость в новой категории информационных лингвистических ресурсов, получивших название НБД, методы формирования которых разработаны М. Г. Кружковым [30, 31, 33, 34]. С прикладной точки зрения реализация в этих базах данных моделей и технологии целенаправленной генерации и развития новых знаний дала возможность существенно дополнить функциональность электронных корпусов текстов и тем самым обеспечить извлечение тех невыражаемых и труднодоступных знаний переводчиков, которые применялись ими, являясь неэксплицированными и новыми в контрастной лингвистике.

Литература

1. *Nonaka I.* The knowledge-creating company // *Harvard Bus. Rev.*, 1991. Vol. 69. No. 6. P. 96–104.
2. *Nonaka I.* A dynamic theory of organizational knowledge creation // *Organ. Sci.*, 1994. Vol. 5. No. 1. P. 14–37.
3. *Wierzbicki A. P., Nakamori Y.* Basic dimensions of creative space // *Creative space: Models of creative processes for knowledge civilization age* / Eds. A. P. Wierzbicki, Y. Nakamori. — Berlin–Heidelberg: Springer Verlag, 2006. P. 59–90.
4. *Wierzbicki A. P., Nakamori Y.* Knowledge sciences: Some new developments // *Zeitschrift für Betriebswirtschaft*, 2007. Vol. 77. No. 3. P. 271–295.
5. *Wierzbicki A. P., Nakamori Y.* The importance of multimedia principle and emergence principle for the knowledge civilisation age // *J. Syst. Sci. Syst. Eng.*, 2008. Vol. 17. No. 3. P. 297–318.
6. *Nakamori Y.* Methodology for knowledge synthesis // *Cutting-edge research topics on multiple criteria decision making* / Eds. Y. Shi, S. Wang, Y. Peng, J. Li, Y. Zeng. — Communications in computer and information science ser. — Berlin: Springer, 2009. Vol. 35. P. 311–317.
7. *Knowledge science — modeling the knowledge creation process* / Ed. Y. Nakamori. — London — New York: CRC Press, 2011. 177 p.
8. *Nakamori Y.* Knowledge and systems science — enabling systemic knowledge synthesis. — London — New York: CRC Press, 2013. 234 p.
9. *Zatsman I., Buntman P.* Theoretical framework and denotatum-based models of knowledge creation for monitoring and evaluating R&D program implementation // *Int. J. Softw. Sci. Comput. Intell.*, 2013. Vol. 5. No. 1. P. 15–31.
10. *Зацман И. М.* Построение системы терминов информационно-компьютерной науки: проблемно-ориентированный подход // *Теория и практика общественной научной информации*. — М.: ИНИОН РАН, 2013. Вып. 21. С. 120–159.
11. *Зацман И. М.* Таблица интерфейсов информатики как информационно-компьютерной науки // *Науч.-технич. информация. Сер. 1: Организация и методика информационной работы*, 2014. № 11. С. 1–15.
12. *Зацман И. М.* Информационно-компьютерная наука: технологические предпосылки становления // *Информационные технологии*, 2014. № 3. С. 3–12.
13. *Успенский В. А.* К публикации статьи Г. Фреге «Смысл и денотат» // *Семиотика и информатика*, 1997. Вып. 35. С. 351–352.
14. *Фреге Г.* Смысл и денотат // *Семиотика и информатика*, 1997. Вып. 35. С. 352–379.
15. *Фреге Г.* Понятие и вещь // *Семиотика и информатика*, 1997. Вып. 35. С. 380–396.
16. *Зацман И. М.* Семиотическая модель взаимосвязей концептов, информационных объектов и компьютерных кодов // *Информатика и её применения*, 2009. Т. 3. Вып. 2. С. 65–81.
17. A framework of information system concepts (Web edition): The FRISCO Report. — IFIP, 1998. <http://www.mathematik.uni-marburg.de/~hesse/papers/fri-full.pdf>.
18. *Hesse W., Verrijn-Stuart A. A.* Towards a theory of information systems: The FRISCO approach // *Information modelling and knowledge bases XII* / Eds. H. Kangassalo, H. Jaakkola, E. Kawaguchi. — Amsterdam: IOS Press, 2001. P. 81–91.
19. *Eco U.* A theory of semiotics. — Bloomington: Indiana University Press, 1976. 356 p.
20. *Пирс Ч.* Логические основания теории знаков / Пер. с англ. — СПб.: Алетейя, 2000. 352 с.
21. *Василюк Ф. Е.* Структура образа // *Вопросы психологии*, 1993. № 5. С. 5–19.
22. *Зацман И. М.* Нестационарная семиотическая модель компьютерного кодирования концептов, информационных объектов и денотатов // *Информатика и её применения*, 2009. Т. 3. Вып. 4. С. 87–101.
23. *Зацман И. М., Бунтман П. С.* Проектирование индикаторов мониторинга в сфере науки: теоретические

- основания и модели // Онтология проектирования, 2014. № 3(13). С. 32–51.
24. Zatsman I., Buntman N., Krushkov M., Nuriev V., Zalizniak Anna A. Conceptual framework for development of computer technology supporting cross-linguistic knowledge discovery // 15th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2014. Vol. 3. P. 1063–1071.
 25. Zatsman I., Buntman N. Outlining goals for discovering new knowledge and computerised tracing of emerging meanings discovery // 16th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2015. P. 851–860.
 26. Баарс Б., Гейдж Н. Мозг, познание, разум: введение в когнитивные нейронауки / Пер. с англ. — М.: БИНОМ. Лаборатория знаний, 2014. Ч. 1. 544 с.; Ч. 2. 464 с. (Baars B., Gage N. Cognition, brain, and consciousness: Introduction to cognitive neuroscience. — Burlington, MA, USA: Academic Press/Elsevier, 2010. 677 p.)
 27. Секерина И. А. Метод вызванных потенциалов мозга в экспериментальной психолингвистике // Вопросы языкознания, 2006. № 3. С. 22–45.
 28. De Charms R. C. Applications of real-time fMRI // Nat. Rev. Neurosci., 2008. Vol. 9. No. 9. P. 720–729.
 29. Kumaran D., Summereld J. J., Hassabis D., Maguire E. A. Tracking the emergence of conceptual knowledge during human decision-making // Neuron, 2009. Vol. 63. No. 6. P. 889–901.
 30. Зализняк А. А., Зацман И. М., Инькова О. Ю., Кружков М. Г. Надкорпусные базы данных как лингвистический ресурс // Корпусная лингвистика-2015: Тр. 7-й Междунар. конф. — СПб.: СПбГУ, 2015. С. 211–218.
 31. Кружков М. Г. Информационные ресурсы контрастных лингвистических исследований: электронные корпуса текстов // Системы и средства информатики, 2015. Т. 25. Вып. 2. С. 140–159.
 32. Loiseau S., Sitchinava D. V., Zalizniak Anna A., Zatsman I. M. Information technologies for creating the database of equivalent verbal forms in the Russian-French multivariant parallel corpus // Информатика и её применения, 2013. Т. 7. № 2. С. 100–109.
 33. Krushkov M. G., Buntman N. V., Loshchilova E. Ju., Sitchinava D. V., Zalizniak Anna A., Zatsman I. M. A database of Russian verbal forms and their French translation equivalents // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам ежегодной Междунар. конф. «Диалог». — М.: РГГУ, 2014. Вып. 13(20). С. 284–297.
 34. Бунтман Н. В., Зализняк Анна А., Зацман И. М., Кружков М. Г., Ложилова Е. Ю., Сичинава Д. В. Информационные технологии корпусных исследований: принципы построения кросслингвистических баз данных // Информатика и её применения, 2014. Т. 8. Вып. 2. С. 98–110.
 35. Гак В. Г. Русский язык в сопоставлении с французским. — М.: УРСС, 2006. 264 с.
 36. Kouznetsova I. N. Grammaire contrastive du francais et du russe. — Moscow: Nestor Academic Publs., 2009. 272 p.
 37. Зализняк Анна А. Лингвоспецифичные единицы русского языка в свете контрастного корпусного анализа // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам ежегодной Междунар. конф. «Диалог». — М.: РГГУ, 2015. Вып. 14(21). Т. 1. С. 683–695.

Поступила в редакцию 22.07.15

GOAL-ORIENTED PROCESSES OF CROSS-LINGUAL EXPERT KNOWLEDGE CREATION: SEMIOTIC FOUNDATIONS FOR MODELING

I. M. Zatsman

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The results of development of semiotic foundations for modeling goal-oriented processes of cross-lingual expert knowledge creation are described. The technology supporting these processes is outlined. The demand for such technologies is obvious in situations where present systems of expert knowledge do not answer to new socially or technologically significant purposes, corresponding to new or changed requirements of modern society. Instead of centering on the well-known artificial intelligence methods and models of information processing for knowledge representation, this paper focuses on development of new models of goal-oriented processes of expert knowledge creation reflecting dynamics of its formation. The suggested approach to modeling these processes and to development of technologies supporting them is focused on those applied areas where expert knowledge is elicited from domain experts. The experts analyze texts or other interpretation objects which can vary over time and enter

the results of analysis into supracorpus databases. The distinguishing feature of the semiotic approach to modeling is the explicit description of relations between new expert knowledge and those interpretation objects, from which parts of new knowledge were generated. Other important feature is the explicit description of parts of knowledge corresponding to interpretation objects that may vary over time. Feasibility of the approach is demonstrated on the example of information technology, which supports the processes of creation of cross-lingual expert knowledge based on French translations of Russian verbal constructions. Cross-lingual knowledge is generated in the course of analysis of parallel texts in Russian and French languages.

Keywords: cross-lingual expert knowledge; computer modeling; knowledge creation; interpretation objects; semiotic foundations; models of knowledge creation processes

DOI: 10.14357/19922264150311

Acknowledgments

The work was supported by the Russian Foundation for Basic Research (projects 14-07-00785, 13-06-00403) and the Russian Foundation for Humanities (project 15-04-00507).

References

1. Nonaka, I. 1991. The knowledge-creating company. *Harvard Bus. Rev.* 69(6):96–104.
2. Nonaka, I. 1994. A dynamic theory of organizational knowledge creation. *Organ. Sci.* 5(1):14–37.
3. Wierzbicki, A. P., and Y. Nakamori. 2006. Basic dimensions of creative space. *Creative space: Models of creative processes for knowledge civilization age*. Eds. A. P. Wierzbicki, and Y. Nakamori. Berlin–Heidelberg: Springer Verlag. 59–90.
4. Wierzbicki, A. P., and Y. Nakamori. 2007. Knowledge sciences: Some new developments. *Zeitschrift für Betriebswirtschaft* 77(3):271–295.
5. Wierzbicki, A. P., and Y. Nakamori. 2008. The importance of multimedia principle and emergence principle for the knowledge civilisation age. *J. Syst. Sci. Syst. Eng.* 17(3):297–318.
6. Nakamori, Y. 2009. Methodology for knowledge synthesis. *Cutting-edge research topics on multiple criteria decision making*. Eds. Y. Shi, S. Wang, Y. Peng, J. Li, and Y. Zeng. Communications in computer and information science ser. Berlin: Springer. 35:311–317.
7. Nakamori, Y., ed. 2011. *Knowledge science — modeling the knowledge creation process*. London — New York: CRC Press. 177 p.
8. Nakamori, Y. 2013. *Knowledge and systems science — enabling systemic knowledge synthesis*. London — New York: CRC Press. 234 p.
9. Zatsman, I., and P. Buntman. 2013. Theoretical framework and denotatum-based models of knowledge creation for monitoring and evaluating R&D program implementation. *Int. J. Softw. Sci. Comput. Intell.* 5(1):15–31.
10. Zatsman, I. 2013. Postroenie sistemy terminov informatsionno-komp'yuternoy nauki: problemno-orientirovanny podkhod [Construction of the system of terms of information and computer science: A problem-oriented approach]. *Teoriya i praktika obshchestvennoy nauchnoy informatsii* [Theory and practice of scientific information for social sciences]. Moscow: INION RAS. 120–159.
11. Zatsman, I. 2014. Tablitsa interfeysov informatiki kak informatsionno-komp'yuternoy nauki [A table of interfaces of informatics as computer and information science]. *Nauchno-tekhnicheskaya informatsiya. Ser. 1: Organizatsiya i metodika informatsionnoy raboty* [Scientific and Technical Information. Ser. 1: Management and methodology of information work] (11):1–15.
12. Zatsman, I. 2014. Informatsionno-komp'yuternaya nauka: Tekhnologicheskie predposylki stanovleniya [Information and computer science: Technological prerequisites of formation]. *Informatsionnye Tekhnologii* [Information Technologies] (3):3–12.
13. Uspenskiy, V. A. 1997. K publikatsii stat'i G. Frege "Smysl i denotat" [To the publication of the paper of G. Frege "Sense and reference"]. *Semiotika i Informatika* [Semiotics and Informatics] 35:351–352.
14. Frege, G. 1997. Smysl i denotat [Sense and reference]. *Semiotika i Informatika* [Semiotics and Informatics] 35:352–379.
15. Frege, G. 1997. Ponyatie i veshch' [Concept and thing]. *Semiotika i Informatika* [Semiotics and Informatics] 35:380–396.
16. Zatsman, I. 2009. Semioticheskaya model' vzaimosvyazey kontseptov, informatsionnykh ob'ektov i komp'yuternykh kodov [Semiotic model of relationships of concepts, information objects, and computer codes]. *Informatika i ee Primeneniya — Inform. Appl.* 3(2):65–81.
17. FRISCO. 1998. A framework of information system concepts. Report. Available at: <http://www.mathematik.uni-marburg.de/~hesse/papers/fri-full.pdf> (accessed July 29, 2015).
18. Hesse, W., and A. A. Verrijn-Stuart. 2001. Towards a theory of information systems: The FRISCO approach. *Information modelling and knowledge bases XII*. Eds. H. Kangasalo, H. Jaakkola, and E. Kawaguchi. Amsterdam: IOS Press. 81–91.
19. Eco, U. 1976. *A theory of semiotics*. Bloomington: Indiana University Press. 356 p.
20. Peirce, Ch. S. 1931–1958. *Collected papers of Charles S. Peirce*. Cambridge: Harvard University Press. 8 vols.

21. Vasilyuk, F. E. 1993. Struktura obraza [Structure of image]. *Voprosy Psikhologii* [Questions of Psychology] (5):5–19.
22. Zatsman, I. 2009. Nestatsionarnaya semioticheskaya model' komp'yuternogo kodirovaniya kontseptov, informatsionnykh ob"ektov i denotatov [Nonstationary semiotic model of computer coding of concepts, information objects and denotata]. *Informatika i ee Primeneniya — Inform. Appl.* 3(4):87–101.
23. Zatsman, I., and P. Buntman. 2014. Proektirovanie indikatorov monitoringa v sfere nauki: Teoreticheskie osnovaniya i modeli [Design of indicators for monitoring in science: Theoretical foundations and models]. *Ontologiya Proektirovaniya* [Ontology of Design] (3):32–51.
24. Zatsman, I., N. Buntman, M. Kruzhkov, V. Nuriev, and Anna A. Zalizniak. 2014. Conceptual framework for development of computer technology supporting cross-linguistic knowledge discovery. *15th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 3:1063–1071.
25. Zatsman, I., and N. Buntman. 2015. Outlining goals for discovering new knowledge and computerised tracing of emerging meanings discovery. *16th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 851–860.
26. Baars, B., and N. Gage. 2010. *Cognition, brain, and consciousness: Introduction to cognitive neuroscience*. Burlington, MA: Academic Press/Elsevier. 677 p.
27. Sekerina, I. 2006. Metod vyzvannykh potentsialov mozga v eksperimental'noy psikholingvistike [Method of evoked potentials of brain in experimental psycholinguistics]. *Voprosy Yazykoznaniiya* [Topics in the Study of Language] 3:22–45.
28. De Charms, R. C. 2008. Applications of real-time fMRI. *Nat. Rev. Neurosci.* 9(9):720–729.
29. Kumaran, D., J.J. Summereld, D. Hassabis, and E. A. Maguire. 2009. Tracking the emergence of conceptual knowledge during human decision-making. *Neuron* 63(6):889–901.
30. Zalizniak, Anna A., I. Zatsman, O. Inkova, and M. Kruzhkov. 2015. Nadkorpusnye bazy dannykh kak lingvisticheskiy resurs [Supracorpus database as linguistic resource]. *Tr. 7-y konf. po Korpusnoy Lingvistike* [7th Conference on Corpus Linguistics Proceedings]. St. Petersburg. 211–218.
31. Kruzhkov, M. 2015. Informatsionnye resursy kontrastivnykh lingvisticheskikh issledovaniy: Elektronnyye korpusa tekstov [Information resources for contrastive studies: Digital text corpora]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(2):140–159.
32. Loiseau, S., D. V. Sitchinava, Anna A. Zalizniak, and I. M. Zatsman. 2013. Information technologies for creating the database of equivalent verbal forms in the Russian-French multivariant parallel corpus. *Informatika i ee Primeneniya — Inform.s Appl.* 7(2):100–109.
33. Kruzhkov, M. G., N. V. Buntman, E. Ju. Loshchilova, D. V. Sitchinava, Anna A. Zalizniak, and I. M. Zatsman. 2014. A database of Russian verbal forms and their French translation equivalents. *Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii. Po mat-lam ezhegodnoy Mezhdunar. konf. "Dialog-2014"* [Computational Linguistics and Intellectual Technologies: Conference (International) "Dialog-2014" Proceedings]. Moscow. 13(20):284–297.
34. Buntman, N. V., Anna A. Zaliznyak, I. M. Zatsman, M. G. Kruzhkov, E. Yu. Loshchilova, and D. V. Sitchinava. 2014. Informatsionnye tekhnologii korpusnykh issledovaniy: Printsipy postroeniya krosslingvisticheskikh baz dannykh [Information technologies for corpus studies: Underpinnings for cross-linguistic database creation]. *Informatika i ee Primeneniya — Inform. Appl.* 8(2):98–110.
35. Gak, V. G. 2006. *Russkiy yazyk v sopostavlenii s frantsuzskim* [Russian in comparison to French]. Moscow: URSS. 264 p.
36. Kouznetsova, I. N. 2009. *Grammaire contrastive du francais et du russe*. Moscow: Nestor Academic Pubs. 272 p.
37. Zalizniak, Anna A. Lingvospetsifichnye edinitsy russkogo yazyka v svete kontrastivnogo korpusnogo analiza [Russian language-specific words in light of the contrastive corpus analysis]. *Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii. Po mat-lam ezhegodnoy Mezhdunar. konf. "Dialog-2015"* [Computational Linguistics and Intellectual Technologies: Conference (International) "Dialog-2015" Proceedings]. Moscow. 14(21):683–695.

Received July 22, 2015

Contributor

Zatsman Igor M. (b. 1952) — Doctor of Science in technology, Head of Department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; iz_ipi@a170.ipi.ac.ru

ВЛИЯНИЕ МОРФОЛОГИЧЕСКИХ ОПЕРАЦИЙ НА РАСПОЗНАВАНИЕ ФИГУРЫ ДВИЖУЩЕГОСЯ ЧЕЛОВЕКА ПО ПОТОКУ ИЗОБРАЖЕНИЙ*

М. О. Абрамов¹, М. Ю. Катаев²

Аннотация: Приведено описание методики выделения фигуры движущегося человека по потоку изображений с использованием морфологических операций. Поставленная задача определяется получением максимальной точности выделения фигуры человека на бинарном изображении. Приведены численные результаты, полученные на основе целого ряда методов. Сделаны выводы и даны рекомендации по распознаванию фигуры человека на изображении.

Ключевые слова: морфологические операции; одна камера; распознавание фигуры человека; цифровая обработка; бинарные изображения

DOI: 10.14357/19922264150312

1 Введение

Цифровая обработка изображений [1–6] находит широкое применение практически во всех областях человеческой деятельности. Одними из наиболее сложных в этом направлении являются задачи автоматического извлечения информации из изображений в различных условиях их получения. Особую уникальность приобретают методы обработки потоков изображений, так как они сложны для автоматического анализа из-за множества параметров, которые определяются условиями формирования динамического изображения, а также множества вариантов их компенсации или выделения. Каждый из методов обработки изображений может быть реализован с помощью различных алгоритмов, а их выбор и сочетание определенно влияют на быстродействие и точность.

Рассматриваемая в статье задача является развитием цикла работ [7–9], связанных с оценкой двигательной активности человека (на примере его походки). Параметры двигательной активности: скорость, энергия, длина и симметричность шагов и др. — восстанавливаются методом, связанным с оценкой центра масс по изображению фигуры. Точность метода существенно зависит от качества выделения фигуры.

Изменение свойств динамического фона от кадра к кадру затрудняет применение типизованных математических процедур обработки изображений,

что, например, при выделении фигуры человека приводит к ее изменению. Изменение фигуры человека проявляется в разрывах фигуры, уменьшении площади или появлении дополнительных артефактов, чаще всего связанных с тенью, и др.

Для получения оценок изменения центра масс в данной статье рассматриваются особенности, которые связаны с применением морфологических операций.

2 Постановка задачи

В последние годы ведутся активные работы в области приложения математических методов к обработке видеоизображений, что обусловлено быстрым развитием техники. Появляются новые способы использования уже известных методов, что требует проведения специальных исследований относительно поставленной задачи. Одной из важнейших является задача обнаружения и выделения человека на изображении, сложность которой повышается по разным причинам: высокой изменчивости освещенности различных участков изображения, появления тени, динамического фона, различного числа объектов фона, различных форм фигуры человека, его положения относительно камеры и пр. [10]. Однако при решении конкретных задач контроля, медицины, спорта необходимо иметь четкие и однозначные оценки направления

* Работа выполнена при поддержке РФНФ, грант № 14-16-70008 а(р).

¹Национальный исследовательский Томский государственный университет, maxim_amo@mail.ru

²Томский государственный университет систем управления и радиоэлектроники; Юргинский технологический институт (филиал) Национального исследовательского Томского политехнического университета, kataev.m@sibmail.com

движения человека, его антропометрических характеристик при различных условиях освещенности, типа фона, положения и наклона тела.

Данное исследование проводится в медицинских целях и посвящено разработке автоматизированной программной системы оценки двигательной способности постинсультных пациентов [7, 8]. От точности работы программной системы зависит оценка двигательной активности, которая определяет выбор тех или иных реабилитационных мероприятий. Медицинская направленность приложения накладывает жесткие условия на точность и скорость работы алгоритмов обработки изображений, которые должны работать в режиме, близком к реальному времени.

Оценка двигательной активности опирается на метод, связанный с изучением положения при движении центра масс [8]. Для точного определения центра масс фигура человека в процессе движения должна быть определена максимально близко к реальной. Авторами было проведено исследование влияния методов предобработки на качество выделения фигуры движущегося человека на изображениях [9]. В результате был получен диапазон данных по изменению восстановленной фигуры человека, отображающий влияние комбинаций тех или иных методов обработки на качество выделения. Оценка качества восстановления проводилась по критерию сравнения площади полученной фигуры и реальной фигуры, выделенной вручную максимально близко к оригиналу.

Очевидно, что условия измерений и сцена определяют тип и стабильность фона. Особую проблему для выделения фигуры человека создает динамический фон, когда наблюдаются изменения сцены, освещенности, возникающие тени и др. Эти особенности учитывались при разработке программно-алгоритмической части. Ранее авторами было исследовано влияние различных методов предобработки и обработки на конечный результат [9]. Анализ этой работы позволил выявить, что площадь полученной фигуры после обработки занижена по сравнению со своими реальными значениями. Одним из вариантов решения этой проблемы является применение морфологических операций. Дело в том, что в некоторых случаях фигура человека после выделения теряет свою целостность, т. е. некоторые части фигуры получают отделенными друг от друга. Нарушение целостности фигуры приводит к тому, что при фильтрации шумов на изображении отдельные части фигуры могут быть просто отброшены как случайные помехи. Морфологические операции позволяют объединить соседние части фигуры и заполнить получающиеся артефакты (пустоты) внутри фигуры.

3 Математическая морфология

Термин «морфология» [11] относится к описанию свойств формы и структуры каких-либо объектов. В контексте машинного зрения этот термин относится к описанию свойств формы областей на изображении. Операции математической морфологии изначально были определены как операции над множествами, в дальнейшем было определено, что они также полезны в задачах обработки множества точек в двумерном пространстве. Множествами в математической морфологии представляются объекты на изображении.

Так как в рассматриваемом случае математическая морфология используется применительно к задаче распознавания объектов на двумерном изображении, то в первую очередь представляет интерес бинарная морфология, описывающая действия над бинарными изображениями, представленными в виде упорядоченного множества точек — черных и белых, задаваемых как 1 и 0 соответственно. Под областью изображения обычно понимается подмножество единиц изображения. Каждая операция двоичной морфологии является некоторым преобразованием этого множества. В качестве исходных данных принимаются двоичное изображение S и некоторый структурный элемент (СЭ), или примитив, B . Результатом операции также является двоичное изображение.

Структурный элемент представляет собой некоторое двоичное изображение, которое имеет произвольную форму и размер и у которого определен один элемент, называемый ядром (началом). Как правило, используются симметричные СЭ в форме квадрата, креста или круга, а их ядром является точка, находящаяся посередине. Размер СЭ в пикселах примем за r .

Базовыми операциями морфологии являются расширение (дилатация) и сужение (эрозия). Расширение двоичного изображения S на СЭ B обозначается как

$$(S \oplus B)(u, v) = \max_{(i, j) \in B} \{S(u + i, v + j) + B(i, j)\}$$

и понимается как объединение всех копий СЭ, центрированного по каждому пикселу изображения S . Сужение двоичного изображения S на СЭ B обозначается как

$$(S \ominus B)(u, v) = \min_{(i, j) \in B} \{S(u + i, v + j) + B(i, j)\}.$$

Операции сужения и расширения редко используются по отдельности, так как они изменяют размеры областей изображения. Однако если обе

операции используются последовательно, то это позволяет добиться различных результатов в зависимости от порядка их следования. Тип морфологической операции, когда сначала производится расширение, а потом сужение, называется закрытием (замыканием). Обратному порядку соответствует открытие (размыкание). Открытие обозначается следующим образом:

$$S \circ B = (S! B) \oplus B, \tag{1}$$

объекты, которые меньше, чем СЭ B , исчезают, а большие объекты остаются.

Морфологическое закрытие:

$$S \bullet B = (S \oplus B)! B, \tag{2}$$

отверстия на области изображения, которые меньше B , будут заполнены [12].

4 Исследование морфологических операций

В рассматриваемой работе морфологическая обработка применяется для следующих целей: удаление шумов, объединение разделенных частей фигуры и заполнение пустот внутри фигуры. В результате работы морфологических операций на выходе должна получиться замкнутая фигура человека, которая максимально точно соответствует по форме и площади реальной фигуре.

Чтобы исследовать особенности использования морфологических операций применительно к данной задаче, был проведен ряд тестов, показывающих их основные достоинства и недостатки. На вход для обработки подавались изображения с различными комбинациями проблемных для обработки ситуаций, возникающих в реальных условиях (например, мелкие помехи, крупные помехи, слабая фрагментация фигуры человека, сильная фрагментация фигуры человека и т. д.).

Основными морфологическими методами были выбраны операции закрытия (2) и открытия (1) [13]. Исследовалось как их отдельное использование, так и последовательное, изменялся размер матрицы СЭ и ее форма. По результатам исследования определялись наиболее подходящие параметры для решения поставленной задачи (рис. 1 и 2).

Визуальная оценка результатов показывает, что наиболее подходящей комбинацией для решения поставленной задачи является последовательное выполнение над изображением операций открытия и закрытия (рис. 3). При использовании такого подхода удалось очистить изображение от мелких

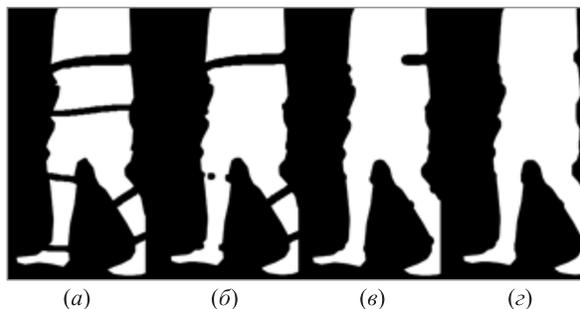


Рис. 1 Действие операции морфологического закрытия на изображение ((a) исходное изображение) фрагментированного силуэта человека при различном размере СЭ: (б) $r = 10$; (в) $r = 15$; (г) $r = 20$

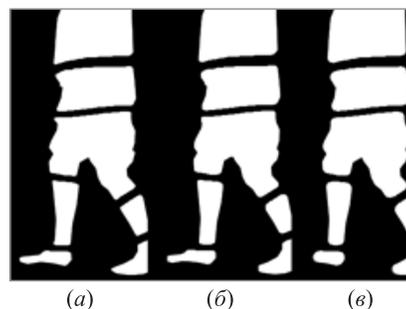


Рис. 2 Действие операции морфологического открытия на изображение фрагментированного силуэта человека при различном размере СЭ: (a) $r = 5$; (б) $r = 10$; (в) $r = 15$

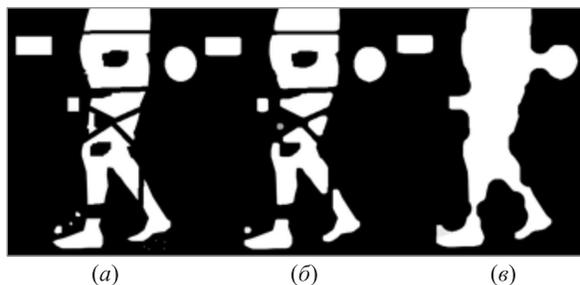


Рис. 3 Последовательное выполнение морфологического открытия и закрытия: (a) исходное изображение; (б) морфологическое открытие $r = 10$; (в) морфологическое закрытие $r = 30$

артефактов, размеры которых меньше размерности матрицы СЭ, не изменяя части фигуры человека. Также удалось добиться соединения разделенных участков фигуры на расстоянии, не превышающем размеров матрицы СЭ. Наиболее удачной формой СЭ был определен круг, так как контур фигуры человека состоит в основном из плавных изгибов и непрямолинейных линий.

Таким образом, для достижения максимальной эффективности морфологической обработки было принято решение использовать последовательно операции открытия и закрытия с различным размером матрицы ядра. Данный подход позволяет очистить изображение от мелкого шума без существенного влияния на фигуру, а также объединить фрагментированные участки фигуры человека и заполнить возникающие пустоты, тем самым улучшая качество всего процесса распознавания фигуры человека.

5 Результаты

Авторами был разработан программно-аппаратный комплекс, предназначенный для исследования эффективности различных комбинаций методов предобработки, обработки и постобработки с целью распознавания человека на потоке изображений и дальнейшего анализа. Все этапы обработки проходят последовательно над всеми изображениями из потока. На каждом этапе можно использовать до 4–5 разных вариантов обработки, таких как выбор цветового пространства, метод фильтрации, алгоритм распознавания, эквализация гистограммы и др. Суммарное число возможных комбинаций всех вариантов методов (добавленных в программно-аппаратный комплекс) равно 48. Для распознавания фигуры в данном исследовании используется модифицированный алгоритм на базе адаптивной гауссовской смешанной модели [14]. На вход подаются последовательно все кадры видеоряда, начиная с кадра, на котором нет иссле-

дуемого субъекта. Более подробно используемые алгоритмы и методы описаны в работе [9].

Все выбранные методы обработки применялись к базе видеопоследовательностей, собранной авторами статьи. На данный момент в нее входят изображения 30 человек. Съемки проводились врачами местных поликлиник [8] с разрешения пациентов с нарушениями походки и без, в разные периоды времени. Разрешение видео — 640 × 480 пикселей. Средняя продолжительность — 8–12 с с частотой 25 кадр/с.

Для исследования влияния морфологических операций на качество распознавания фигуры человека на изображении использовались алгоритмы на основе формул (1) и (2). Для численной оценки точности выделения фигуры было проведено сравнение результатов предыдущей работы [9] с результатами, полученными после проведения морфологических операций. Критерием точности служило отклонение найденной площади фигуры человека от эталонной, размер которой известен.

Морфологическая обработка применялась на финальной стадии работы программно-аппаратного комплекса. Использовались последовательно операции открытия с диаметром СЭ $r = 5$ и закрытия с $r = 20$ в указанном порядке.

В приведенной таблице в колонке $\Delta\bar{S}$ указано отличие в процентах полученной площади фигуры от эталонного значения [9]. Несмотря на сохранение тенденции к занижению результатов от реальной площади, наблюдается существенное улучшение по сравнению с циклом обработки без морфологических операций. Это улучшение отражено в колонке $\Delta\bar{S}'$ таблицы в процентах от ре-

Сравнение результатов, полученных с использованием морфологических операций и без них

№ метода	$\Delta\bar{S}$, %	$\Delta\bar{S}'$, %	№ метода	$\Delta\bar{S}$, %	$\Delta\bar{S}'$, %	№ метода	$\Delta\bar{S}$, %	$\Delta\bar{S}'$, %
RGB			HSV			YUV		
1	-29,61	6,03	17	-20,41	43,66	33	-51,55	5,23
2	-23,63	5,04	18	-40,88	4,84	34	-24,49	10,04
3	-29,54	6,05	19	-0,68	36,56	35	-46,94	5,45
4	-23,56	4,95	20	10,62	18,31	36	-25,25	9,64
5	-28,91	5,24	21	-18,83	36,73	37	-51,40	4,52
6	-24,16	3,70	22	-39,38	33,16	38	-26,30	6,97
7	-28,98	5,23	23	-1,94	26,48	39	-46,63	5,25
8	-24,07	4,44	24	15,56	-0,41	40	-26,09	7,09
9	-29,27	4,19	25	-23,37	18,47	41	-51,51	3,97
10	-25,03	3,32	26	-47,44	-39,70	42	-27,62	5,03
11	-29,22	4,26	27	-6,50	14,80	43	-47,26	5,05
12	-24,91	3,34	28	9,03	2,88	44	-28,09	4,45
13	-30,46	5,48	29	-20,17	51,14	45	-53,20	4,09
14	-22,53	4,90	30	-44,76	31,63	46	-25,24	9,04
15	-30,39	5,48	31	1,09	41,10	47	-47,30	5,33
16	-22,28	5,06	32	11,48	24,93	48	-24,86	9,46

зультатов, полученных без морфологических операций.

Для того чтобы оценить возможности различных цветовых пространств для выделения фигуры человека на изображении, проводились одинаковые расчеты для таких цветовых пространств, как RGB, HSV и IUV (см. таблицу).

6 Выводы

Среднее процентное отклонение (в абсолютном выражении) площади восстановленной фигуры при использовании морфологических операций от эталонной составляет 25,98%, что на 12,86% лучше, чем без их использования (38,84% [9]). Помимо значительного среднестатистического улучшения результата при использовании методов 19, 23 и 31 удалось добиться максимального приближения значения восстановленной фигуры к эталонной (отклонение менее 2%). Таким образом, именно эти методы являются наиболее подходящими для решения поставленной задачи. Обращаем внимание на то, что помимо цветового пространства HSV (методы 17–32, см. таблицу) все эти методы объединяет обработка методом коррекции баланса белого «Серый мир» и использование методов фильтрации для снижения уровня шума.

Стоит отметить, что в зависимости от используемого цветового пространства изменяются особенности полученной фигуры человека. При использовании пространства RGB (методы 1–16, см. таблицу) получают наиболее стабильные результаты, а при визуальной оценке фигура человека наиболее четко соответствует реальной. Однако данное цветовое пространство наименее восприимчиво к внешним факторам, таким как схожесть цветового тона элементов фона и исследуемого объекта. На всех полученных изображениях с использованием RGB те части тела человека, что были свободны от одежды, имели схожие тона с фоном (светлая кожа и светлый фон), поэтому после всех процедур обработки эти части относились алгоритмом к фону и обрезались. Данную проблему легко решить при наличии на человеке одежды, средне или сильно отличающейся от фона. Поэтому можно рекомендовать использование этого цветового пространства в том случае, если есть возможность влиять на объект или среду исследования. Пространство HSV в меньшей степени чувствительно к внешним факторам, оно позволило распознать все части исследуемого объекта. Однако, в отличие от RGB, силуэт полученного человека получается искаженным, нечеткие линии на краях и различного рода артефакты делают это пространство очень

чувствительным к используемым методам предобработки. Наилучшие результаты получились при его использовании, поэтому можно рекомендовать выбор этого цветового пространства в тех случаях, когда исследователь не может контролировать среду или объект записи. Пространство YUV (методы 33–48, см. таблицу) при визуальном анализе и согласно таблице показало наихудшие результаты. Несмотря на стабильность метода при использовании контрастирования на этапе предобработки, следует воздержаться от использования данного цветового пространства для распознавания движущихся объектов на изображении.

Все сказанное выше дает право заключить, что использование морфологических операций приводит к улучшению качества распознавания фигуры движущегося человека на изображении. В данной работе сделан упор на описание практического действия морфологических операций, поэтому в совокупности с предыдущей публикацией по предобработке [9] она может оказать содействие желающим в создании своей собственной системы распознавания или в изучении действия методов предобработки и морфологических операций на изображения.

Литература

1. Грузман И. С., Киричук В. С., Косых В. П., Перетягин Г. И., Спектор А. А. Цифровая обработка изображений в информационных системах. — Новосибирск: НГТУ, 2002. 352 с.
2. Форсайт А., Понс Дж. Компьютерное зрение. Современный подход. — М.: Вильямс, 2004. 928 с.
3. Яне Б. Цифровая обработка изображений / Пер. с англ. — М.: Техносфера, 2007. 584 с.
4. Гонсалес Р., Вудс Р., Эддинс С. Цифровая обработка изображений в среде MATLAB. — М.: Техносфера, 2006. 616 с.
(Jane B. Digital image processing. — 6th ed. — Berlin: Springer, 2005. 608 p.)
5. Красильников Н. Н. Цифровая обработка 2D- и 3D-изображений. — СПб.: БХВ-Петербург, 2011. 608 с.
6. Захаров Р. К. Методы повышения качества изображений в задачах распознавания // Современные научные исследования и инновации, 2012. № 8. С. 9–10.
7. Катаева Н. Г., Катаев М. Ю., Чистякова В. А. Автоматизированная оценка степени нарушения ходьбы после инсульта // Медицинская техника, 2012. № 1. С. 40–43.
8. Катаев М. Ю., Катаева Н. Г., Катаев С. Г., Абрамов М. О., Чистякова В. А. Определение и анализ двигательной активности постинсультного пациента по видеопотоку // Бюллетень сибирской медицины, 2014. Т. 13. № 5. С. 36–41.
9. Абрамов М. О., Катаев М. Ю. Влияние методов предобработки на восстановление фигуры движущегося

- человека из потока изображений // Докл. ТУСУР, 2014. № 4(33). С. 65–69.
10. *Бойко И. А., Гурьянов Р. А.* Распознавание объектов на основе видеосигнала, полученного с камеры, установленной на подвижной платформе // Молодой ученый, 2013. № 6. С. 34–36.
 11. *Огнев И. В., Сидорова Н. А.* Обработка изображений методами математической морфологии в ассоциативной осцилляторной среде // Известия вузов. Поволжский регион. Технические науки, 2007. № 4. С. 87–97.
 12. *Burger W., Burge M. J.* Digital image processing: An algorithmic introduction using Java. — New York, NY, USA: Springer Science & Business Media, 2007. 566 p.
 13. *Najman L.* Mathematical morphology: From theory to applications. — London: Wiley, 2010. 520 p.
 14. *Kaew Trakulpong P., Bowden R.* An improved adaptive background mixture model for real-time tracking with shadow detection // Computer Vision and Distributed Processing: 2nd European Workshop on Advanced Video-based Surveillance Systems, AVBS01, Proceedings. — Kingston: Kluwer Academic Publs., 2002. P. 135–144.

Поступила в редакцию 26.01.15

INFLUENCE OF MORPHOLOGICAL OPERATIONS ON RECOGNITION OF A MOVING HUMAN FIGURE FROM A SET OF IMAGES

M. O. Abramov¹ and M. Yu. Kataev^{2,3}

¹National Research Tomsk State University, 36 Lenin Av., Tomsk 634050, Russian Federation

²Tomsk State University of Control Systems and Radioelectronics, 40 Lenin Av., Tomsk 634050, Russian Federation

³Yurga Institute of Technology (Branch) of National Research Tomsk Polytechnic University, 26 Leningradskaya Str., Yurga, Kemerovo Region 652055, Russian Federation

Abstract: The article describes the methodology of moving human figure recognition from set of images with the help of morphological operations. Accuracy of recognition results is determined by the difference between the true and recovering figure squares from binary images. Comparison of the obtained accuracy for different morphological operations for some color spaces (RGB, HSV, and YUV) is investigated. The paper provides recommendations for improvement of human figure recognition from set of images.

Keywords: morphological operations; one camera; human figure recognition; digital processing; binary images

DOI: 10.14357/19922264150312

Acknowledgments

Work was performed with financial support of the Russian Foundation for Humanities, project No. 14-16-70008) a(r).

References

1. Gruzman, I. S., V. S. Kirichuk, V. P. Kosykh, G. I. Peretyagin, and A. A. Spektor. 2002. *Tsifrovaya obrabotka izobrazheniy v informatsionnykh sistemakh* [Digital image processing in information systems]. Novosibirsk: NGTU Publ. 352 p.
2. Forsayt, A. and J. Pons. 2004. *Komp'yuternoe zrenie. Sovremennyy podkhod* [Computer vision. Modern approach]. Moscow: Vil'yams. 928 p.
3. Jane, B. 2005. *Digital image processing*. 6th ed. Berlin: Springer. 608 p.
4. Gonsales, P., P. Vuds, and C. Eddins. 2006. *Tsifrovaya obrabotka izobrazheniy v srede MATLAB* [Digital image processing in MATLAB]. Moscow: Technosphere. 616 p.
5. Krasil'nikov, N. N. 2011. *Tsifrovaya obrabotka 2D i 3D izobrazheniy* [Digital 2D and 3D image processing]. St. Petersburg: BKhV-Peterburg. 608 p.
6. Zakharov, R. K. 2012. *Metody povysheniya kachestva izobrazheniy v zadachakh raspoznavaniya* [Methods of improvement quality of images in the recognition tasks]. *Sovremennyye Nauchnye Issledovaniya i Innovatsii* [Modern Scientific Researches and Innovations] 8:9–10.
7. Kataeva, N. G., M. Yu. Kataev, V. A. Chistyakova, and Ya. A. Khamaganov. 2012. *Avtomatizirovannaya otsenka stepeni narusheniya khod'by posle insul'ta* [Automated estimation of the severity of walking disorders in patients after stroke]. *Meditsinskaya Tekhnika* [Biomedical Engineering] 1:40–43.

8. Kataev, M. Yu., N. G. Kataeva, S. G. Kataev, M. O. Abramov, and V. A. Chistyakova. 2014. Opredelenie i analiz dvigatel'noy aktivnosti postinsul'tnogo patsienta po videopotoku [Definition and analysis of motion activity of afterstroke patient from the video stream]. *Byulleten' Sibirskoy Meditsiny* [Bulletin of Siberian Medicine] 5:36–41.
9. Abramov, M. O., and M. Yu. Kataev. 2014. Vliyaniye metodov predobrabotki na vosstanovleniye figury dvizhushchegosya cheloveka iz potoka izobrazheniy [The effect of preprocessing methods on figure of the moving human retrieving from an images sequence]. *Dokl. TUSURa* [Proc. TUSUR] 4:65–69.
10. Boyko, I. A., and R. A. Gur'yanov. 2013. Raspoznavaniye ob"ektov na osnove videosignala, poluchennogo s kamery, ustanovlennoy na podvizhnoy platforme [Object recognition based on the video signal obtained from a camera placed on mobile platform]. *Molodoy Uchenyy* [Young Scientist] 6:34–36.
11. Ognev, I. V., and N. A. Sidorova. 2007. Obrabotka izobrazheniy metodami matematicheskoy morfologii v assotsiativnoy ostsillyatornoy srede [Image processing by means of mathematical morphology in the associative oscillatory medium]. *Izvestiya VUZov. Povolzhskiy region. Tekhnicheskie Nauki* [University Proceedings. Volga Region. Technical Sciences] 4:87–97.
12. Burger, W., and M. J. Burge. 2007. *Digital image processing: An algorithmic introduction using Java*. New York, NY: Springer Science & Business Media. 566 p.
13. Najman, L. 2010. *Mathematical morphology: From theory to applications*. London: Wiley. 520 p.
14. Kaew Trakulpong, P., and R. Bowden. 2002. An improved adaptive background mixture model for real-time tracking with shadow detection. *Computer Vision and Distributed Processing: 2nd European Workshop on Advanced Video-based Surveillance Systems, AVBS01, Proceedings*. Kingston: Kluwer Academic Publs. 135–144.

Received January 26, 2015

Contributors

Abramov Maksim O. (b. 1990) — PhD student, National Research Tomsk State University, 36 Lenin Av., Tomsk 634050, Russian Federation; maxim_amo@mail.ru

Kataev Mikhail Yu. (b. 1961) — Doctor of Science in technology; professor, Tomsk State University of Control Systems and Radioelectronics, 40 Lenin Av., Tomsk 634050, Russian Federation; professor, Yurga Institute of Technology (Branch) of National Research Tomsk Polytechnic University, 26 Leningradskaya Str., Yurga, Kemerovo Region 652055, Russian Federation; kataev.m@sibmail.com

Абрамов Максим Олегович (р. 1990) — аспирант Национального исследовательского Томского государственного университета

Агаларов Явер Мирзабекович (р. 1952) — кандидат технических наук, доцент, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Беляев Юрий Константинович (р. 1932) — доктор физико-математических наук, профессор Университета Умеа, Швеция

Горбунова Анастасия Владимировна (р. 1986) — аспирант Российского университета дружбы народов

Зарядов Иван Сергеевич (р. 1981) — кандидат физико-математических наук, доцент Российского университета дружбы народов

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зыкин Владимир Сергеевич (р. 1992) — ассистент Омского государственного технического университета

Катаев Михаил Юрьевич (р. 1961) — доктор технических наук; профессор Томского государственного университета систем управления и радиоэлектроники; профессор Юргинского технологического института (филиала) Томского политехнического университета

Кристрём Бенгт (р. 1960) — доктор философии, директор Центра экономики природных ресурсов и окружающей среды Шведского университета сельскохозяйственных наук

Лебедев Алексей Викторович (р. 1971) — кандидат физико-математических наук, доцент механико-математического факультета Московского государственного университета им. М. В. Ломоносова

Лери Марина Муксумовна (р. 1969) — кандидат технических наук, научный сотрудник Института прикладных математических исследований Карельского научного центра Российской академии наук

Матюшенко Сергей Иванович (р. 1963) — кандидат физико-математических наук, доцент Российского университета дружбы народов

Миронов Андрей Михайлович (р. 1966) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Российской академии наук Федерального исследовательского центра «Информатика и управление» Российской академии наук

Печинкин Александр Владимирович (1946–2014) — доктор физико-математических наук, профессор

Разумчик Ростислав Валерьевич (р. 1984) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, доцент Российского университета дружбы народов

Самуйлов Константин Евгеньевич (р. 1955) — доктор технических наук, профессор, заведующий кафедрой Российского университета дружбы народов

Синицын Игорь Николаевич (р. 1940) — доктор технических наук, профессор, заслуженный деятель науки РФ, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Чичагов Владимир Витальевич (р. 1955) — кандидат физико-математических наук, доцент Пермского государственного национального исследовательского университета

Шоргин Сергей Яковлевич (р. 1952) — доктор физико-математических наук, профессор, заместитель директора Федерального исследовательского центра «Информатика и управление» Российской академии наук

Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация не должна нарушать закон об авторских правах.

Направляя рукопись в редакцию, авторы сохраняют все права собственников данной рукописи и при этом передают учредителям и редколлегии неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на ее распространение в России и за рубежом. Авторы должны представить в редакцию письмо в следующей форме:

Соглашение о передаче права на публикацию:

«Мы, нижеподписавшиеся, авторы рукописи «. . .», передаем учредителям и редколлегии журнала «Информатика и её применения» неисключительное право опубликовать данную рукопись статьи на русском языке как в печатной, так и в электронной версиях журнала. Мы подтверждаем, что данная публикация не нарушает авторского права других лиц или организаций, а также не содержит сведений, запрещенных к опубликованию в открытой печати.

Подписи авторов: (ф. и. о., дата, адрес)».

Это соглашение может быть представлено в бумажном виде или в виде отсканированной копии (с подписями авторов).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации представленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.

3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам.

Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.

5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.

6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или ЛАТЭХ (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.

7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 20 страниц указанного формата.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете: <http://www.ipiran.ru/journal/template.doc>

8. Статья должна содержать следующую информацию на *русском и английском языках*:

- название статьи;
- Ф.И.О. авторов, на английском можно только имя и фамилию;
- место работы, с указанием почтового адреса организации и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
http://www.ipiran.ru/journal/issues/2013_07_01_rus/authors.asp и
http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;
- аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
- ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами;
- источники финансирования работы (ссылки на гранты, проекты, поддерживающие организации и т. п.).

9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала;
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://translit.ru/> бесплатной программой транслитерации русского текста в латиницу, при этом в закладке «варианты. . . » следует выбрать опцию BGN.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Ниже приведены примеры ссылок на различные виды публикаций в списке “References”.

Описание статьи из журнала:

Zagurenko, A. G., V. A. Korotovskikh, A. A. Kolesnikov, A. V. Timonov, and D. V. Kardymon. 2008. Tekhniko-ekonomicheskaya optimizatsiya dizayna gidrorazryva plasta [Technical and economic optimization of the design of hydraulic fracturing]. *Neftyanoe hozyaystvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Описание статьи из продолжающегося издания (сборника трудов):

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal”–kompozit [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

Описание материалов конференций:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W. Vibration problems in engineering. — 4th ed. — N.Y.: Wiley, 1974. 521 p.*)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. N.Y.: Wiley. 521 p.

Описание неопубликованного документа:

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. *Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor* [Mathematical modeling of the plasma in the compact torus]. D.Sc. Diss. Moscow. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. *Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv* [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. *Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya* [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.
11. При отправке файлов по электронной почте просим придерживаться следующих правил:
 - указывать в поле subject (тема) название журнала и фамилию автора;
 - использовать attach (присоединение);
 - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Информатика и её применения»:
Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН
Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05
e-mail: rust@ipiran.ru (Сейфуль-Мулюков Рустем Бадриевич)
<http://www.ipiran.ru/journal/issues/>

Requirements for manuscripts submitted to Journal “Informatics and Applications”

Journal “Informatics and Applications” (Inform. Appl.) publishes theoretical, review, and discussion articles on the research and development in the field of informatics and its applications.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

The topics covered include the following areas:

- theoretical fundamentals of informatics;
 - mathematical methods for studying complex systems and processes;
 - information systems and networks;
 - information technologies; and
 - architecture and software of computational complexes and networks.
1. The Journal publishes original articles which have not been published before and are not intended for simultaneous publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

Agreement on the transfer of rights to publish:

“We, the undersigned authors of the manuscript “. . .”, pass to the Founder and the Editorial Board of the Journal “Informatics and Applications” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.

Author(s) signature(s): (name(s), address(es), date).

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If the authors receive their article for correction after reviewing, it does not mean that the article is approved for publication. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or \LaTeX (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.

7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 20 pages of the specified format.

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

All pages of the manuscript should be numbered.

The templates for the manuscript typesetting are presented on site: <http://www.ipiran.ru/journal/template.doc>.

8. The articles should enclose data both in **Russian and English:**

- title;
- author’s name and surname;
- affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
- data on authors according to the format: (see site)

http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp and

http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;

- abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae;
 - indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.
Important! Keywords must not be sentences;
 - Acknowledgments.
9. References. Russian references have to be presented both in English translation and Latin transliteration (refer <http://www.translit.ru>, option BGN).
Please take into account the following examples of Russian references appearance:
- Article in journal:**
Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.
- Journal article in electronic format:**
Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).
- Article from the continuing publication (collection of works, proceedings):**
Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental’noe issledovanie prochnosti soedineniy “stal’–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.
- Conference proceedings:**
Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma “Novye resursosberegayushchie tekhnologii nedropol’zovaniya i povysheniya neftegazootdachi” [6th Symposium (International) “New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact” Proceedings]*. Moscow. 267–272.
- Books and other monographs:**
Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem [Operation of turbine generators with direct cooling]*. Moscow: Energy Publs. 352 p.
- Dissertation and Thesis:**
Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.
- State standards and patents:**
GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch’yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publs. 10 p.
Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel’nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.
- References in Latin transcription are presented in the original language.
References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.
10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.
11. Submissions of files by e-mail must include:
- the journal title and author’s name in the “Subject” field;
 - an article and additional materials have to be attached using the “attach” function;
 - an electronic version of the article should contain the file with the text and a separate file with figures.
12. “Informatics and Applications” journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia
Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05
e-mail: rust@ipiran.ru (to Prof. Rustem Seyful-Mulyukov)
<http://www.ipiran.ru/english/journal.asp>