Mathematics

Original article
UDC 51-77:004.89+311.2
DOI: https://doi.org/10.18721/JPM.14415

COX REGRESSION IN THE PROBLEM OF RISKY BEHAVIOR PARAMETER ESTIMATION BASED ON THE LAST EPISODES' DATA

V. F. Stoliarova [1] ✉, A. L. Tulupyev [2,1]

[1] St. Petersburg Federal Research Center of the RAS, St. Petersburg, Russia;
[2] St. Petersburg State University, St. Petersburg, Russia

✉ vfs@dscs.pro

Abstract: The article suggests an extension and refinement of the existing model for incomplete estimation of the person's behavior characteristics using several last episodes of his risky behavior and including individual characteristics and other factors, in particular, the sex, age, etc. This extension has been made with a novel approach to formalization of the gamma Poisson model of behavior based on the random point process theory. The general form of the likelihood function for the estimates was derived. This formalization allowed to use the Cox regression model for the estimation of the process intensity parameters on the condition of only three last episodes taking into account different characteristics of the domain. The possibilities of the chosen approach were shown using the data on publication of posts in the online social media.

Keywords: gamma Poisson model of behavior, mixed Poisson process, Cox regression, online media user profiling

Funding: The research was carried out in the framework of the project on state assignment SPC RAS No. 0073-2019-0003 (The formalization of the gamma Poisson model of behavior with the process intensity function). The research was supported by St. Petersburg State University, project No. 73555239 (The likelihood function for the estimation of the intensity function parameters with limited data on several last episodes taking into account external factors). RFBR project No. 20-07-00839 (The Cox regression model for the public posting data with respect to the information from the user's profile).

For citation: Stoliarova V. F., Tulupyev A. L., Cox regression in the problem of risky behavior parameter estimation based on the last episodes' data, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 14 (4) (2021) 202–217. DOI: https://doi.org/10.18721/JPM.14415

Научная статья
УДК 51-77:004.89+311.2
DOI: https://doi.org/10.18721/JPM.14415

РЕГРЕССИЯ КОКСА В ЗАДАЧЕ ОЦЕНКИ ПАРАМЕТРОВ РИСКООБРАЗУЮЩЕГО ПОВЕДЕНИЯ ИНДИВИДА ПО ДАННЫМ О ПОСЛЕДНИХ ЭПИЗОДАХ

В. Ф. Столярова [1] ✉, А. Л. Тулупьев [2,1]

[1] Санкт-Петербургский Федеральный исследовательский центр РАН, Санкт-Петербург, Россия;

[2] Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

✉ vfs@dscs.pro

Аннотация. В работе предлагается расширение и уточнение существующей модели оценки характеристик поведения индивида по неполным данным о нескольких последних эпизодах его рискованного поведения за счет учета ряда особенностей предметной области и иных факторов, в частности пола, возраста и т. п. Представлена формализация гамма-пуассоновской модели поведения в терминах случайных процессов. Выведен общий вид функции правдоподобия для построения искомых оценок. Привлекается модель регрессии Кокса для построения оценок максимального правдоподобия параметров интенсивности процесса при условии наблюдения лишь трех последних эпизодов с учетом индивидуальных особенностей и внешних факторов. Возможности этого подхода продемонстрированы на данных о публикации постов в социальной сети.

Ключевые слова: гамма-пуассоновская модель поведения, смешанный процесс Пуассона, регрессия Кокса, профиль пользователя онлайн социальной сети

Финансирование: Государственное задание Санкт-Петербургского Федерального исследовательского центра Российской академии наук СПИИРАН № 0073-2019-0003, поддержано Санкт-Петербургским государственным университетом, проект № 73555239; РФФИ, проект № 20-07-00839.

Для цитирования: Столярова В. Ф., Тулупьев А. Л. Регрессия Кокса в задаче оценки параметров рискообразующего поведения индивида по данным о последних эпизодах // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2021. Т. 14. № 4. С. 202–217. DOI: https://doi.org/10.18721/JPM.14415

Introduction

The problem of individual risk assessment, related to a person's behavior, arises in several areas.

Some problems in epidemiology and public health concern the interpersonal transmission of incurable diseases, like HIV. Transmission of such diseases is possible only via a person's participation in certain types of episodic risky behaviors [1, 2]. The Bell – Trevino model [3] combines the number of episodes for each type of such hazardous behavior with the cumulative risk of disease transmission in the population, which is used for the decision making. We note also that the main difficulty of using such models as the Bell – Trevino one is collecting the behavioral data, since deviant behaviors cannot be observed directly [2] and the interviews are highly prone to different types of cognitive biases [4].

Estimation of the behavior characteristics with self-reported data on episodes is relevant not only for behavioral risk assessment, but also for modelling health risks related to common behavior. There exists the concept of health-related behavior patterns [6] in the area of behavioral epidemiology [5]. For example, the patterns of alcohol and food consumption are related to depression [7]. Physical activity, screen time and eating habits are correlated with health-related quality of life in children [8] and adults [9, 10]. Some common behavioral patterns are related to the risk of cardiac attacks [10], diabetes mellitus [11]. The problem of behavioral data collection arises in this setting too. The common behaviors can be observed directly, but collecting such data takes some time, while the estimates may be needed in real time, e.g., during a routine visit to a primary care physician. For this reason, measures based on self-reporting are used. These measures allow to determine the behavior pattern and the related risk. Examples include the AUDIT score [12] and the Facebook Use Intensity Scale [13]. We note here that the frequency of behavior is estimated in such scales in any case, but questions like "How often did you do that?" are highly prone to cognitive biases [14].

The problem of behavioral simulation arises in the area of cybersecurity. An Internet user's behavioral pattern [15] includes several characteristics like the frequency of password change, the frequency of clicking on external links, the frequency of providing access to the user's account to third parties [16]. These peculiarities can be exploited by intruders to gain access the critical documents of an organization [17]. Therefore, the problem of user profiling is an important component of the organizational cybersecurity program. The problem with collecting data on episodic behavior also arises in this setting, as episodes of some behaviors are difficult to observe directly (e.g., sharing a password with third parties), and the estimates of the potential employee trustworthiness are required in the hiring process being limited in time.

Thus, several research areas deal with the problem of estimating episodic behavior characteristics, and the frequency (intensity) of the behavior is the main of those characteristics. Mathematical models of a person's behavior are proposed for assessing the risk associated with episodic behavior with limited and incomplete data.

The goal of the paper is formalizing the existing gamma Poisson mathematical model of personal behavior in terms of point stochastic processes [28] and deriving the maximal likelihood function for estimating the behavioral characteristics (the frequency of behavior) with data on several latest episodes of a person's behavior.

In the paper, we use the data on three last episodes for all respondents, but the proposed model can be adapted to any number of episodes. The problem setting lies within an analysis of recurrent event process [29] and a survival analysis [30]. The proposed formalization allows to use regression analysis to estimate behavioral characteristics, accounting for those affecting the behavior intensity. Application of regression analysis is illustrated for the problem on estimating the frequency of public posting in social media.

We note that although the gamma Poisson model was proposed earlier, the point process formalization with the possibility of including external factors is a novel approach in modelling personal behavior. This formalization extends and improves the existing Poisson process formalization, introducing a class of mixed Poisson processes. Practically speaking, the regression model approach accounting for external factors is novel.

Research methods

Estimation of episodic behavior characteristics with limited data (existing approach). Cheap and fast estimates of behavioral characteristics associated with a person's behavior are required in certain research areas. Sometimes the data on episodes of behavior is unavailable (e.g., in situations dealing with deviant behavior), sometimes the cheapest is the best (risk assessment during a check-up visit to a physician). The main characteristic of episodic behavior is its frequency (or intensity). This characteristic is often a component of existing risk scales, like the AUDIT score or Facebook Use Intensity Scale, and it can be directly related to a certain risk, like the frequency of physical activity. Due to cost and time limitations, it is preferable to rely on self-reports to gather the data on episodes of a person's behavior, subsequently using mathematical models to build desired estimates.

Various methods are used to extract a person's knowledge about their behavior. The Timeline Followback (TLFB) is one of the most widely used methods [18]. The individual is asked to outline the most memorable episodes in their behavior consulting a calendar. Holidays and events are marked too. This method is used to characterize different risky behaviors, like modelling consumption patterns of nicotine and alcohol [19], drugs [20], assessing HIV transmission patterns [21], and is considered to be reliable [22]. We note here that although the information on several memorable episodes is important in the context of behavioral pattern modelling, it has limited usage for estimating the frequency of behavior and is nevertheless prone to cognitive biases [23].

Refs. [24, 25] describe the method for behavior frequency estimation with limited data on several last consequent episodes and record intervals. The information extraction approach is similar to the TLFB method, and originally was implemented in the area of HIV transmission modelling to estimate the number of episodes of hazardous behavior and cumulative risk in the population [2]. The last episodes' methodology relies on answers to questions like "When was the last time (or earlier times) you participated in this behavior?" The answers to such questions are less prone to cognitive biases than answers to questions like "How many times did you participate in such behavior?". The information on several consequent episodes can be used to build the frequency estimates via the mathematical model of behavior.

The natural mathematical model for an episodic behavior is a point process, specifically, the Poisson one [26] that assumes that the sample is homogeneous with respect to the behavior in question. The gamma Poisson model was proposed to include additional interpersonal variability for heterogeneous samples [25]. In the latter case, Bayes Belief networks can be used to estimate the risky behavior characteristics [27]. We now proceed to formalization of the Poisson mathematical model of behavior.

Poisson model of person's behavior formalization. Let us assume that the episodes of a person's behavior occur in continuous time and only a finite number of episodes can occur in a finite interval of time. Several episodes cannot occur simultaneously. This property is natural when considering the person's episodic behavior.

The point process is defined via its intensity function [29]. Let us denote the starting point of the process as $t = 0$, moments of episode realizations as $0 \le T_1 < T_2 < \dots$. We exploit the property (1) here: two episodes cannot occur simultaneously; $N(s,t]$ is the number of episodes that occurred in the interval $(s,t]$, let $N(0,t) = N(t)$, and $H(t) = \{N(s), 0 \le s < t\}$ is the history of the process up to time $t > 0$.

The intensity function of a point process is the limit of the probability of episode realization in the infinitely small interval Δt:

$$\lambda(t \mid H(t)) = \lim_{\Delta t \to 0} \frac{\Pr\{\Delta N(t) = 1 \mid H(t)\}}{\Delta t}. \qquad (1)$$

This function is supposed to be bounded and continuous almost everywhere.

All characteristics of the process can be defined via its intensity function [29], including the distribution of the interval between two subsequent episodes, the distribution of the number of episodes $N(s, t]$ in the interval $(s, t]$, joint distribution of numbers $N(s_j, t_j)$ in the disjoint intervals $(s_j, t_j]$, $j = 1, 2, \dots, m$, the mean function $\mu(t) = E\{N(t)\}$, the dispersion function var $\{N(t)\}$, etc.

Now we will use write out the formalization of the Poisson model of the person's behavior in the terms of its intensity function. This model assumes that all individuals are equal in their behavior proneness and is useful when dealing with homogeneous samples.

If the realization of episodes for some type of behavior does not depend on the process history (see property (1)), then the Poisson point process is the basic model for such episodic behavior. The intensity function in that case is assumed to be a non-negative integrable deterministic function (1):

$$\lambda_{\text{Poiss}}(t \mid H(t)) = \rho(t), t > 0. \qquad (2)$$

Note. In the homogeneous case, the intensity function of the Poisson process $\lambda_{\text{Poiss}}(t \mid H(t)) = \rho$ is called 'intensity' and in the one-dimensional case it is 'rate'.

The Poisson model of behavior was a starting mathematical model for the problem of behavior characteristics estimation with limited data on several last episodes [1, 2, 24, 25]. Ref. [26] is dedicated to the maximum likelihood estimation of these characteristics.

The intensity function in the form (2) allows writing out different characteristics of the person's episodic behavior process in closed form [29]. Let us define the cumulative intensity function as

$$\mu(s,t)=\int_s^t \rho(v)dv,$$

where $\mu(s, t)$ is continuous and finite $t > 0$ (we denote $\mu(t) = \mu(0, t)$).

Then the following statements hold.

(*i*) The number of points $N(s,t]$ in the interval $(s,t]$ follows the Poisson distribution with the mean $\mu(s, t) = \mu(s) - \mu(t)$, $0 \leq s < t$:

$$\Pr(n \text{ episodes in } (s,t]) = \frac{\mu(s,t)^n}{n!}\exp\{-\mu(s,t)\};$$

(*ii*) For the Poisson model of the person's behavior, the mean number of episodes in the interval equals its variance:

$$\mathrm{E}\{N(t)\} = \mathrm{var}\{N(t)\} = \mu(t);$$

(*iii*) Numbers $N(s_1, t_1)$ and $N(s_2, t_2)$ in the disjoint intervals $[s_1,t_1)$, $[s_2,t_2)$ are independent of random variables;

(*iv*) For the homogeneous Poisson process with intensity ρ, the interval lengths between the times of subsequent episodes $W_j = T_j - T_{j-1}$, $j = 1, 2, ...$ are independent and identically exponentially distributed random variables with the same intensity ρ:

$$\Pr(W_j > w) = \exp(-\rho w), \;\; w > 0.$$

The Poisson model of the person's behavior can be used for estimating the behavioral characteristics in the following settings:

all individuals can be considered homogeneous in their behavior; they have identical intensity of the episode realization process;

as numbers of episodes in every two disjoint intervals are independent, the data on episodes can be collected in every moment.

However, the Poisson model of behavior has limited use in real-life applications. In particular, the property (*ii*) is often violated. Gamma Poisson model of behavior was formulated to incorporate this overdispersion with the random intensity function that can model inter-individual variability of the intensity function.

Results obtained

Formalization of gamma Poisson model of the person's behavior. This mathematical model describes the person's episodic behavior assuming that the episodes occur for every individual according to the Poisson point process and the intensity parameter varies between individuals in the population. This parameter is modelled with the gamma distributed variable. The assumptions of the gamma Poisson model are listed below.

1. Episodes of the person's behavior occur in continuous time and only a finite number of episodes can occur in a finite interval of time. Several episodes cannot occur simultaneously.

2. The time of realization of the episode for every individual does not depend on the times of realization of the previous episodes.

3. All individuals vary in their behavior proneness that remains constant in time.

The second assumption may be verified with statistical data (e.g., with methods of a dependency analysis), but as we consider the situations with lack of such data, we have to state it as the assumption. The third assumption describes the natural heterogeneity of the sample, and the requirement that behavior proneness be constant in time may be relaxed with appropriate mathematical considerations (not covered in Ref. [29]).

It is assumed in the gamma Poisson model that the behavior intensity varies between individuals [25], and it cannot be observed directly. The mixed Poisson process is a background for this model. The episodes of the process for a particular individual occur according to the standard Poisson process, but the intensity of this process varies between individuals in the population and is modelled with the random variable [31].

Historically, the mixed Poisson distribution emerged in a similar setting in insurance studies [32, 33]. The authors proposed to take into account an individual accident proneness when modelling the number of insurance claims in some period of time. This accident proneness is assumed to have a gamma distribution.

Now we proceed to the gamma Poisson model of behavior formalization. We introduce unobserved random covariates u_i, i = 1, 2, ..., m to model (1) to take into account inter-individual differences in the behavior. Conditioned on u_i, the intensity function of the episode realization process $N_i(t),\ t \geq 0$ has the form

$$\lambda_i(t \mid u_i = \lambda_i, H(t)) = \lambda_i \rho_i(t), \tag{3}$$

where u_i is a random variable with the distribution function $G_i(u)$ and finite expectation, λ_i is the behavior proneness for an individual i and is the realization of u_i for a current individual.

We assume that all u_i are independent and identically distributed random variables with the probability density function (PDF) $G(u)$. For simplicity of calculations, in this paper we assume that $G(u)$ has the gamma distribution

$$g(u;k,\phi) = \frac{u^{k-1}\exp(-u/\phi)}{\phi^k \Gamma(k)},\ u > 0, \tag{4}$$

where $k > 0$ is the shape parameter, $\phi > 0$ is the scale one.

The gamma distribution is frequently chosen for modelling different real-life values that are skewed and positive, e.g., the inter-episode time intervals. The usefulness of the gamma distribution for the purpose of this study is the property of the gamma mixture of the Poisson distribution: it has a closed form. Moreover, it is widely used as the conjugate prior in Bayesian statistics.

Now we write out some properties of the mixed gamma Poisson process that is the foundation for the gamma Poisson model of the person's behavior. In the following, we denote $\mu_i(s,t) = \int_s^t \rho_i(v)dv$.

GP. 1. Conditioned on u_i, the number of episodes in some interval has the Poisson discrete distribution

$$\mathrm{P}(N_i(s,t) = n \mid u_i = \lambda_i) = \frac{(\lambda_i \mu_i(s,t))^n}{n!}\exp(-\lambda_i \mu_i(s,t)),$$

and without this condition it has a negative binomial distribution:

$$\mathrm{P}(N_i(s,t) = n) = \int_0^\infty \frac{[u\mu_i(s,t)]^n}{n!} \cdot \exp(-u\mu_i(s,t)) g(u;k,\phi)du =$$

$$= \int_0^\infty \frac{[u\mu_i(s,t)]^n}{n!} \exp(-u\mu_i(s,t)) \cdot \frac{u^{k-1}\exp(-u/\phi)}{\phi^k\Gamma(k)} du =$$

$$= \frac{\mu_i(s,t)^n}{n!\phi^k\Gamma(k)} \cdot \int_0^\infty u^{n+k-1}\exp(-u(\mu_i(s,t) + 1/\phi))du =$$

$$= \frac{\mu_i(s,t)^n}{n!\phi^k\Gamma(k)} \frac{\phi^{n+k}}{[\phi\mu_i(s,t)+1]^{n+k}} \Gamma(n+k) = \frac{\Gamma(n+k)}{n!\Gamma(k)} \frac{[\phi\mu_i(s,t)]^n}{[1+\phi\mu_i(s,t)]^{n+k}}, n = 0,\ 1,\ \ldots\ .$$

The negative binomial distribution has an additional parameter compared to binomial, incorporating the overdispersion emerging due to inter-individual variability.

The negative binomial distribution has an additional parameter compared to binomial, incorporating the overdispersion emerging due to inter-individual variability.

GP. 2. Using the fact that the expectation of the gamma distributed random variable (4) is $\mathrm{E}\{u_i\} = k\phi$ and the formulae for the complete expectation and complete variance, we derive that the expectation of the number of episodes in some interval for the gamma Poisson model of behavior is

$$\mathrm{E}\{N_i(s,t)\} = \mathrm{E}\{\mathrm{E}\{N_i(s,t)\,|\,u_i\}\} = \mu_i(s,t)k\phi,$$

and the variance is:

$$\mathrm{var}\{N_i(s,t)\} = \mathrm{E}(\mathrm{var}[N_i(s,t)\,|\,u_i]) + \mathrm{var}(\mathrm{E}[N_i(s,t)\,|\,u_i]) =$$
$$= \mathrm{E}(u_i\mu_i(s,t)) + \mathrm{var}(u_i\mu_i(s,t)) = \mu_i(s,t)k\phi + k(\mu_i(s,t)\phi)^2.$$

GP. 3. The number of episodes in the disjoint intervals for this model is not independent. Let $s_1 < t_1 < s_2 < t_2$, then we have:

$$\mathrm{cov}\left\{N_i\left(s_1,\ t_1\right),\ N_i\left(s_2,\ t_2\right)\right\} = \mathrm{E}(N_i(s_1,t_1)\cdot N_i(s_2,t_2)) - \mathrm{E}N_i(s_1,t_1)\mathrm{E}N_i(s_2,t_2) =$$
$$= \mathrm{E}[\mathrm{E}(N_i(s_1,t_1)\cdot N_i(s_2,t_2)\,|\,u_i)] - (k\phi)^2\mu_i(s_1,t_1)\mu_i(s_2,t_2) =$$
$$= \mathrm{E}(u_i^2)\mu_i(s_1,t_1)\mu_i(s_2,t_2) - (k\phi)^2\mu_i(s_1,t_1)\mu_i(s_2,t_2) = k\phi^2\mu_i(s_1,t_1)\mu_i(s_2,t_2).$$

GP. 4. (gamma Poisson Unconditioned Intensity function). The unconditioned intensity function for the gamma Poisson process has the form:

$$\lambda_i(t\,|\,H_i(t)) = \left\{\frac{N_i(t^-)+k}{1+\mu_i(t)\phi}\right\}\phi\rho_i(t),\ \ t>0. \tag{5}$$

Theorem. *The unconditioned intensity function for the gamma Poisson process has the following form in the assumptions of the gamma Poisson model:*

$$\lambda_i(t\,|\,H_i(t)) = \left\{\frac{N_i(t^-)+k}{1+\mu_i(t)\phi}\right\}\phi\rho_i(t),\ \ t>0.$$

P r o o f. Now we proceed to the derivation of Eq. (5). Let us consider an individual episode realization process with conditioned (for every individual) intensity:

$$\lambda(t\,|\,H(t),u) = u\rho(t),$$

where u has gamma distribution (4) with the cumulative distribution function (cdf) $g(u;\,k,\,\phi)$. Based on the full expectation rule, following definition (1) and bearing in mind that it is impossible for episodes to occur simultaneously (the property according to gamma Poisson assumption (1)), in that case we have

$$P\{N(t,t+\Delta t)=1\,|\,H(t),u\} = \lambda(t\,|\,H(t),u)\Delta t + o(\Delta t),$$

we derive

$$P\{N(t,t+\Delta t)=1\,|\,H(t)\} = \frac{P\{N(t,t+\Delta t)=1;H(t)\}}{P(H(t))} =$$

$$= \frac{\int_0^\infty P\{N(t,t+\Delta t)=1; H(t)\,|u\} g(u)du}{\int_0^\infty P(H(t)\,|u) g(u)du} =$$

$$= \frac{\int_0^\infty P\{N(t,t+\Delta t)=1\,|u\} P\{H(t)\,|u\} g(u)du}{\int_0^\infty P(H(t)\,|u) g(u)du} = \frac{\int_0^\infty [u\rho(t)\Delta t + o(\Delta t)]\, P\{H(t)\,|u\} g\left(u\right)du}{\int_0^\infty P(H(t)\,|\,u) g(u)du}.$$

Further we note that $N(t)|u$ is a standard Poisson process with the intensity $u\rho(t)$, and, therefore,

$$P(H(t)\,|\,u) = P(N(t^-)\,|\,u) \sim \text{ Poisson}(u\rho(t)).$$

As $g(u)$ follows a gamma distribution, $\int_0^\infty P(H(t)\,|\,u)\, g(u)du$ is the probability of the variable

with a negative binomial distribution [31] equal to $N(t^-)$:

$$\int_0^\infty P(H(t)\,|\,u)\, g(u)du = P_{\text{NB}}(N(t^-)),$$

where P_{NB} denotes the probability mass function of the negative binomial distribution.

Then we write:

$$\int_0^\infty [u\rho(t)\Delta t + o(\Delta t)] P\{H(t)\,|\,u\} g(u)du = \frac{N(t^-)+1}{\mu(t)} P_{\text{NB}}(N(t^-)+1).$$

Therefore, the intensity function has the form

$$\lambda\left(t\,|\,H(t)\right) = \lim_{\Delta t\to 0} \frac{P\left\{N(t,t+\Delta t)=1\,|\,H(t)\right\}}{\Delta t} = \frac{N\left(t^-\right)+1}{\mu(t)} \frac{P_{\text{NB}}\left(N(t^-)+1\right)}{P_{\text{NB}}\left(N(t^-)\right)} \rho(t) =$$

$$= \left(\frac{N(t^-)+1}{\mu(t)}\right) \cdot \frac{C_{k-1}^{N(t^-)+k} \left(\frac{\mu(t)\phi}{1+\mu(t)\phi}\right)^{N(t-)+1} \left(\frac{1}{1+\mu(t)\phi}\right)^k}{C_{k-1}^{N(t^-)+k-1} \left(\frac{\mu(t)\phi}{1+\mu(t)\phi}\right)^{N(t-)} \left(\frac{1}{1+\mu(t)\phi}\right)^k} \rho(t) = \left\{\frac{N(t^-)+k}{1+\mu(t)\phi}\right\} \phi\rho(t).$$

Q. E. D.

If the behavior proneness u has the gamma distribution with the expectation 1 (one) (in that case $\mathrm{E}(u) = k\phi = 1$), then the intensity function has the form:

$$\lambda(t\,|\,H(t)) = \left\{\frac{1+\phi N(t^-)}{1+\phi\mu(t)}\right\} \rho(t), \;\; t > 0. \tag{6}$$

Therefore, the gamma Poisson model of the person's behavior reflects the following properties:

a) all individuals in the population vary in their behavior proneness and the intensity of the episode's realization;

b) gamma Poisson model takes into account the overdispersion in the number of episodes that occurs due to inter-individual variability;

c) as $k,\phi > 0$, it follows from (GP. 3) that the numbers of episodes in the disjoint intervals are clustered; this means that the probability for $N(s_1,t_1)$ and $N(s_2,t_2)$ to take simultaneously large (or small) values is greater than in the Poisson case;

d) in the homogeneous case, according to (*ii*) and Eq. (2), the expected intensity of episode realization in the population is

$$\mathrm{E}(\lambda\,|\,N(t)=n) = \frac{k+n}{1/l\phi + t}$$

(this follows from GP. 4), which is close to the classical notion of the rate as the number of episodes per unit time.

Note on the concept of behavior intensity. The notion of behavior intensity is introduced to deal with the characteristics of episodic types of behavior. This notion is sometimes overloaded, and the proposed formalization of the gamma Poisson model forms the following system of notions.

The intensity function $\lambda(t \mid H(t))$ is the probability limit of the realization for one episode in the infinitely small unit of time Δt. If the intensity function is constant (the case of standard homogeneous Poisson process), it is called 'intensity'; the notion of 'rate' can be used in the one-dimensional case.

Gamma Poisson model of the person's behavior: the likelihood function for estimating the behavior characteristics with the data on several last episodes

Therefore, the proposed formalization allows interpreting the estimation of episode frequency as a problem on intensity function restoration and estimation of the parameters of gamma distribution that models individual behavior proneness. The data on several last consequent episodes are less prone to cognitive biases and are easy to extract [2].

We also note that it is important to take into account external factors that influence the person's behavior. Regression analysis allows to combine the mathematical gamma Poisson model of behavior and available additional data on external factors. To use regression analysis, we derive the maximum likelihood function for the realization of several last episodes.

We observe m individuals from the moment $t = \tau_{i0}$ up to τ_i. Next, we extract self-reported data on the times when the latest episodes of some type of behavior occurred (the available number of episodes can be 0 or 3 – 4 for different respondents). Therefore, we observe several epochs of the point stochastic process

$$\{N_i(t), 0 \le t\}, i = 1,\ 2,\ \ldots,\ m$$

in the interval $\tau_i - \tau_{i0}$.

Let us denote as $H_i(t) = \{N_i(s) : 0 \le s < t\}$ the history of a point process for an individual i. Such process has an intensity function of the form (3) in the gamma Poisson model.

Let us assume that p external covariates (e.g., age, sex, etc.) influence the process of episode realization:

$$\mathbf{x}_i(t) = (x_{i1}(t),\ x_{i2}(t),\ \ldots\ ,\ x_{ip}(t))',\ i = 1,\ 2,\ \ldots,\ m.$$

There exist several methods for incorporating these factors to the intensity function, one of them is the assumption of proportional hazards that is the foundation for the Cox proportional hazards model [36]. In that case, all external factors are included to the deterministic factor of the intensity function (3):

$$\rho_i(t) = \rho_0(t;\ \alpha)\exp(\mathbf{x}_i\ (t)\boldsymbol{\beta}), \tag{7}$$

where $\rho_0(t;\alpha)$ is the baseline intensity function for the individuals with $\mathbf{x}_i\ (t) = 0$; $\boldsymbol{\beta}$ is the vector of regression coefficients.

If there is no assumed form of the function then the regression estimation method is semiparametric. The semiparametric method for the recurrent events data is called the Andersen – Gill model [29, 30]. The semiparametric derivation allows to simplify the overall maximum likelihood function using gamma distribution (4) with 1 (one) expectation $\mathrm{E}(u_i) = k\phi = 1$:

$$g_1(s;\phi) = \frac{s^{\phi^{-1}-1}\exp(-s/\phi)}{\phi^{\phi^{-1}}\Gamma(1/\phi)},\ s > 0. \tag{8}$$

Therefore, in order to restore the intensity function (3) we should estimate the parameters $\theta = (\alpha, \beta, \phi)$ from data. The likelihood function is comprised from the episode realization likelihoods for every individual in the sample:

$$L(\theta) = \prod_{i=1}^{m} L_i(\theta).$$

If u_i were observed, then the individual likelihood function of the realization of data $(n_i, t_{i1}, \ldots, t_{in_i}, u_i)$ (see Ref. [30]) is

$$L_i^0(\theta) = P(N_i(\tau_{i0}, \tau_i) = n_i; u_i) = \prod_{j=1}^{n_i} u_i \rho_i(t_{ij}) \exp\left\{-\int_0^\infty u_i \rho_i(s) ds\right\}.$$

However, u_i represent an individual proneness and are therefore unobserved. The likelihood function of the data $(n_i, t_{i1}, \ldots, t_{in_i})$ is the gamma mixture of the individual functions $L_i^0(\theta)$:

$$L_i(\theta) = \int_0^\infty L_i^0(\theta) dG(u, \phi) = \int_0^\infty \left[\prod_{j=1}^{n_i} u_i \rho_i(t_{ij}) \exp\left\{-\int_0^\infty u_i \rho_i(s) ds\right\}\right] dG(u, \phi).$$

Using the formula for the realization probability for n episodes of mixed Poisson distribution, we derive

$$L_i(\theta, \phi) = \left\{\prod_{j=1}^{n_i} \frac{\rho_0(t_{ij})}{\mu_0(\tau_i)}\right\} \frac{\Gamma(n_i + \phi^{-1})}{\Gamma(\phi^{-1})} \cdot \frac{(\phi\mu_i(\tau_i))^{n_i}}{(1 + \phi\mu_i(\tau_i))^{n_i + \phi^{-1}}}.$$

The EM (Expectation-Maximization) algorithm is used to estimate the parameters [29]. The main assumption of this method is the proportional hazards assumption (7) that defines the multiplicative form of the dependency between intervals between subsequent episodes in the gamma Poisson model and external factors.

Program realization. The data were gathered for 1,500 random users of the ‘VKontakte’ social network who have written at least three public wall posts during the last year from their last visit. The data included information on the times when the posts were made, sex and age listed in the profile, and the number of friends. As the main purpose of the data analysis was illustrating how to apply the regression methodology to the problem of behavior intensity estimation, only full observations were included.

Table 1 demonstrates the survival data on behavioral episodes in regression analysis. The

Table 1

The survival data for public posts

The interval				Posting episode
Start	End	Start	End	
		(UNIX time format)		
Feb. 1, 2021 17:45:11	Feb. 21, 2021 14:55:11	1612190711	1613908511	0
Feb. 1, 2021 17:15:07	Feb. 1, 2021 17:45:11	1612188907	1612190711	1
May 31, 2020 16:02:00	Feb. 1, 2021 17:15:07	1590930120	1612188907	1
Feb. 21, 2020 14:55:11	May 31, 2020 16:02:00	1590930020	1590930120	1

Footnote: user ID = 22.

Table 2

The coefficients obtained for the fitted Cox regression

Variable	Coefficient exponentiated	Standard error	z-score with significance
Sex (male)	0.745	0.079	−3.71*)
Age	1.03	0.006	5.08*)
Number of friends	1.00	0.000	6.98*)

Footnote: *) denotes that p-value is less than 0.01.

intervals between public posts for user ID 22 are given: the last visit was at 1613908511 (2021-02-21 14:55:11 MSK), and the times of public posts were 1612190711 (2021-02-01 17:45:11 MSK), 1612188907 (2021-02-01 17:15:07 MSK), 1590930120 (2020-05-31 16:02:00 MSK).

The general regression model for estimating the behavior characteristics with the data on last episodes accounting for the user profile data is as follows:

$$\text{Surv. } (\textit{Start, End, Posting episode}) \sim$$
$$\text{Cluster (ID)} + \textit{Sex} + \textit{Age} + \textit{Number of friends}.$$

This model reflects the dependency between the interval, starting at Start, ending at End, and other characteristics from the user's profile. We used the emfrail [35] R package for semiparametric regression fitting. The fitted model is presented in Table 2.

The overall regression performance is good: the p-value for the likelihood ratio test is lower than 0.01, so that the model describes the data better than the saturated model, the Commenges – Andersen test for the frailty significance p-value is lower than 0.01. Regression fitting allows estimating the parameter of gamma density: 0.769 (95% confidence interval (0.691, 0.861)). Furthermore, all three characteristics (Sex, Age, Number of friends) of the user's profile are statistically significant in the regression. The negative regression coefficient for the sex variable can be interpreted to indicate that women are more likely to write public posts. Other variables have positive valued coefficients, meaning that an increase in those values results in an increase in the length of intervals between subsequent episodes.

Discussion of the results

The proposed methodology allows using the Poisson model of behavior in heterogeneous samples. The assumptions made are that the realization process of an individual episode follows the Poisson model, and the proportional hazards assumption holds. The former assumption does not hold for some types of behavior, e.g., planned behavior. The gamma Poisson model can be used in situations when episodes are more or less spontaneous, like deviations from the dietary plan. The latter assumption can also be violated, and there exist several other forms of dependency between behavioral episodes and other influencing factors [29].

Summary

The paper addresses the problem of estimating risky behavioral characteristics with data on several latest episodes accounting for external factors that influence how the episode unfolds. The novel formalization of the gamma Poisson model of behavior via the process intensity function is presented. The proposed formalization extends the existing one, developed for the Poisson model of behavior, allowing to include external factors (like age or sex) and take into account the interpersonal variability in the behavior proneness. The gamma Poisson model has some limitations as it was developed to model occasional (unplanned) behavior. The practical application of the proposed formalization includes the survival analysis regression model: the Cox proportional-hazards regression. Although the proportional-hazards assumption can be violated too, other approaches exist. The Cox regression usage was demonstrated on the public posting data.

The factors that influence how episodes evolve can have a complex dependency structure. This peculiarity can be addressed with probabilistic graphical networks that may also be used for decision making under uncertainty and incorporate different types of uncertainties in the domain.

REFERENCES

1. **Paschenko A. E., Tulupiev A. L., Tulupieva T. V., et al.,** Indirect assessment of the probability of HIV infection on the basis of data on last risky behavior episodes, Health Care of the Russian Federation. (2) (2010) 32–35 (in Russian).
2. **Tulupieva T. V., Paschenko A. E., Tulupiev A. L., et al.,** Modeli VICH-riskovannogo povedeniya v kontekste psihologicheskoy zaschity i drugih adaptivnyh stiley [Models of HIV risky behavior in the context of psychological protection and other responsive behaviors], Nauka, St. Petersburg, 2008 (in Russian).
3. **Bell D. C., Trevino R. A.,** Modeling HIV risk, J. Acq. Immun. Def. Syndrom. 22 (3) (1999) 280–287.
4. **Rao A., Tobin K., Davey-Rothwell M., Latkin C. A.,** Social desirability bias and prevalence of sexual HIV risk behaviors among people who use drugs in Baltimore, Maryland: implications for identifying individuals prone to underreporting sexual risk behaviors, AIDS and Behavior. 21 (7) (2017) 2207–2214.
5. **Kaplan R. M., Criqui M. H.** (Eds.), Behavioral epidemiology and disease prevention, Springer Science & Business Media, USA, 2012.
6. **Oftedal S., Kolt G. S., Holliday E. G., et al.,** Associations of health-behavior patterns, mental health and self-rated health, Prevent. Med. 118 (January) (2019) 295–303.
7. **Guertler D., Moehring A., Krause K., et al.,** Latent alcohol use patterns and their link to depressive symptomatology in medical care patients, Addiction. 116 (5) (2021) 1063–1073.
8. **Dumuid D., Olds T., Lewis L., et al.,** Health-related quality of life and lifestyle behavior clusters in school-aged children from 12 countries, J. Pediatrics. 183 (April) (2017) 178–183.
9. **Blom E. E., Aadland E., Skrove G. K., et al.,** Health-related quality of life, intensity-specific physical activity in high-risk adults attending a behavior change service within primary care, PloS One. 14 (12) (2019) e0226613.
10. **Hart L. M., Gordon A. R., Sarda V., et al.,** The association of disordered eating with health-related quality of life in US young adults and effect modification by gender, Qual. Life Res. 29 (5) (2020) 1203–1215.
11. **Bishwajit G., Tang S., Yaya S., et al.,** Lifestyle behaviors, subjective health, and quality of life among Chinese men living with type 2 diabetes, Am. J. Men's Health. 11 (2) (2017) 357–364.
12. **Neufeld M., Rehm J., Bunova A., et al.,** Validation of a screening test for alcohol use, the Russian Federation, Bull. World Health Organ. 99 (7) (2021) 496–505.
13. **Orosz G., Tóth-Király I., Bőthe B.,** Four facets of Facebook intensity – The development of the multidimensional Facebook intensity scale, Pers. Individ. Differ. 100 (October) (2016) 95–104.
14. **Kreitchmann R. S., Abad F. J., Ponsoda V., et al.,** Controlling for response biases in self-report scales: Forced-choice vs. psychometric modeling of Likert items, Front. Psychol. 10 (15 October) (2019) 2309.
15. **Eke C. I., Norman A. A., Shuib L., Nweke H. F.,** A survey of user profiling: State-of-the-art, challenges, and solutions, IEEE Access. 7 (17 October) (2019) 144907–144924.
16. **Tulupieva T. V., Abramov M. V., Tulupiev A. L.,** Tsifrovaya kultura: sotsialniye seti i sotsioinzhenerniye ataki [Digital culture: Social networks and social engineering attacks], In: Psychologicheskoye zdorovye i technologii zdorovyesberejeniya v sovremennoy obrazovatelnoy srede [Psychological health and health preservation technologies in the modern educational environment (monograph)], Ed. by V. M. Golyanich, NIC ART, St. Petersburg (2019) 322–345 (in Russian).
17. **Azarov A. A., Tulupieva T. V., Suvorova A. V., et al.,** Sotsioinzhenerniye ataki: problemy analisa [Social engineering attacks: the problems of analysis], Nauka, St. Petersburg, 2016 (in Russian).
18. **Sobell L. C., Sobell M. B.,** Timeline follow-back: A technique for assessing self-reported alcohol consumption, In: Measuring alcohol consumption: Psychological and biochemical methods. Litten R. Z., Allen J. P. (Eds.), Humana Press, Totowa, NJ, USA (1992) 41–72.
19. **Ray L. A., Du H., Grodin E., et al.,** Capturing habitualness of drinking and smoking behavior

in humans, Drug Alcohol Depend. 207 (1 February) (2020) 107738.

20. **Martin-Willett R., Helmuth T., Abraha M., et al.,** Validation of a multisubstance online Timeline Followback assessment, Brain Behav. 10 (1) (2020) e01486.

21. **Irwin T. W., Morgenstern J., Parsons J. T., et al.,** Alcohol and sexual HIV risk behavior among problem drinking men who have sex with men: An event level analysis of timeline followback data, AIDS and Behavior. 10 (3) (2006) 299–307.

22. **Santos G. M., Strathdee S. A., El-Bassel N., et al.,** Psychometric properties of measures of substance use a systematic review and meta-analysis of reliability, validity and diagnostic test accuracy, BMC Med. Res. Methodol. 20 (1) (2020) 1–22.

23. **Kaplan B. A., Koffarnus M. N.,** Timeline followback self-reports underestimate alcohol use prior to successful contingency management treatment, Alcohol Alcohol. 54 (3) (2019) 258–263.

24. **Paschenko A. E., Tulupiev A. L., Nikolenko S. I.,** HIV infection modeling based on the last episodes of risky behavior, J. Instr. Engin. 49 (11) (2006) 33–34 (in Russian).

25. **Paschenko A. E., Tulupiev A. L., Nikolenko S. I.,** HIV-acquisition risk statistical estimates based on the data about several last episodes of risky behavior, Trudy SPIIRAN (SPIIRAS Proceedings). 2 (3) (2006) 257–268 (in Russian).

26. **Stepanov D. V., Musina V. F., Suvorova A. V., et al.,** Risky behavior Poisson model identification: heterogeneous arguments in likelihood, Trudy SPIIRAN (SPIIRAS Proceedings). (4 (23)) (2012) 157–184 (in Russian).

27. **Suvorova A.,** Exploring Bayesian belief network for risky behavior modelling: discretization and latent variables, Proc. the II Internation. Sci. & Pract. Conf. “Fuzzy Technologies in the Industry–FTI”. Ulyanovsk, Russia, October 23 – 25 (2018) 63–70.

28. **Daley D. J., Vere-Jones D.,** An introduction to the theory of point processes, Vol. I: Elementary theory and methods, 2nd edition, Springer, New York, 2003.

29. **Cook R. J., Lawless J.,** The statistical analysis of recurrent events, Springer Science & Business Media, New York, 2007.

30. **Andersen P. K., Borgan O., Gill R. D., Keiding N.,** Statistical models based on counting processes, Springer Science & Business Media, New York, 2012.

31. **Grandell J.,** Mixed Poisson processes [Monographs on Statistics and Applied Probability, Vol. 77], Chapman and Hall / CRC Press, UK, 1997.

32. **Greenwood M., Yule G. U.,** An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents, J. Royal Stat. Soc. A. 83 (2) (1920) 255–279.

33. **Newbold E. M.,** A contribution to the study of the human factor in the causation of accidents, Industrial Health Research Board, London, 1926.

34. **Therneau T. M., Grambsch P. M.,** Modeling survival data: Extending the Cox model, Springer Science + Business Media, New York, 2000.

35. **Balan T. A., Putter H.,** frailtyEM: An R package for estimating semiparametric shared frailty models, J. Stat. Softw. 90 (7) (2019) 1–29.

СПИСОК ЛИТЕРАТУРЫ

1. **Пащенко А. Е., Тулупьев А. Л., Тулупьева Т. В., Красносельских Т. В., Соколовский Е. В.** Косвенная оценка вероятности заражения ВИЧ-инфекцией на основе данных о последних эпизодах рискованного поведения // Здравоохранение Российской Федерации. 2010. № 2. С. 32–35.

2. **Тулупьева Т. В., Пащенко А. Е., Тулупьев А. Л., Красносельских Т. В., Казакова О. С.** Модели ВИЧ-рискованного поведения в контексте психологической задачи и других адаптивных стилей. СПб.: Наука, 2008. 147 с.

3. **Bell D. C., Trevino R. A.** Modeling HIV risk // Journal of Acquired Immune Deficiency Syndromes. 1999. Vol. 22. No. 3. Pp. 280–287.

4. **Rao A., Tobin K., Davey-Rothwell M., Latkin C. A.** Social desirability bias and prevalence of sexual HIV risk behaviors among people who use drugs in Baltimore, Maryland: implications for identifying individuals prone to underreporting sexual risk behaviors // AIDS and Behavior. 2017. Vol. 21. No. 7. Pp. 2207–2214.

5. **Kaplan R. M., Criqui M. H.** (Eds.) Behavioral epidemiology and disease prevention. USA: Springer Science & Business Media, 2012. 299 p.

6. **Oftedal S., Kolt G. S., Holliday E. G., Stamatakis E., Vandelanotte C., Brown W. J., Duncan M. J.** Associations of health-behavior patterns, mental health and self-rated health // Preventive Medicine. 2019. Vol. 118. January. Pp. 295—303.

7. **Guertler D., Moehring A., Krause K., et al.** Latent alcohol use patterns and their link to depressive symptomatology in medical care patients // Addiction. 2021. Vol. 116. No. 5. Pp. 1063—1073.

8. **Dumuid D., Olds T., Lewis L., et al.** Health-related quality of life and lifestyle behavior clusters in school-aged children from 12 countries // The Journal of Pediatrics. 2017. Vol. 183. April. Pp. 178—183.

9. **Blom E. E., Aadland E., Skrove G. K., Solbraa A. K., Oldervoll L. M.** Health-related quality of life, intensity-specific physical activity in high-risk adults attending a behavior change service within primary care // PloS One. 2019. Vol. 14. No. 12. P. e0226613.

10. **Hart L. M., Gordon A. R., Sarda V., Calzo J. P., Sonneville K. R., Samnaliev M., Austin S. B.** The association of disordered eating with health-related quality of life in US young adults and effect modification by gender // Quality of Life Research. 2020. Vol. 29. No. 5. Pp. 1203—1215.

11. **Bishwajit G., Tang S., Yaya S., He Z., Feng Z.** Lifestyle behaviors, subjective health, and quality of life among Chinese men living with type 2 diabetes // American Journal of Men's Health. 2017. Vol. 11. No. 2. Pp. 357—364.

12. **Neufeld M., Rehm J., Bunova A., et al.** Валидация скринингового теста на употребление алкоголя, Российская Федерация // Bulletin of the World Health Organization. 2021. Vol. 99. No. 7. Pp. 496—505.

13. **Orosz G., Tóth-Király I., Bőthe B.** Four facets of Facebook intensity — The development of the multidimensional Facebook intensity scale // Personality and Individual Differences. 2016. Vol. 100. October. Pp. 95—104.

14. **Kreitchmann R. S., Abad F. J., Ponsoda V., Nieto M. D., Morillo D.** Controlling for response biases in self-report scales: Forced-choice vs. psychometric modeling of Likert items // Frontiers in Psychology. 2019. Vol. 10. 15 October. P. 2309.

15. **Eke C. I., Norman A. A., Shuib L., Nweke H. F.** A survey of user profiling: State-of-the-art, challenges, and solutions // IEEE Access. 2019. Vol. 7. 17 October. Pp. 144907—144924.

16. **Тулупьева Т. В., Абрамов М. В., Тулупьев А. Л.** Цифровая культура: социальные сети и социоинженерные атаки // Психологическое здоровье и технологии здоровьесбережения в современной образовательной среде (Монография). Ред. В. М. Голянич. СПб.: НИЦ АРТ, 2019. С. 322—345.

17. **Азаров А. А., Тулупьева Т. В., Суворова А.Л., Тулупьев А. Л., Абрамов М. В., Юсупов Р. М.** Социоинженерные атаки. Проблемы анализа. СПб.: Наука, 2016. 352 с.

18. **Sobell L. C., Sobell M. B.** Timeline follow-back: A technique for assessing self-reported alcohol consumption // Measuring alcohol consumption: Psychological and biochemical methods. Litten R. Z., Allen J. P. (Eds.) Totowa, NJ (USA): Humana Press, 1992. Pp. 41—72.

19. **Ray L. A., Du H., Grodin E., Bujarski S., Meredith L., Ho D., Nieto S., Wassum K.** Capturing habitualness of drinking and smoking behavior in humans // Drug and Alcohol Dependence. 2020. Vol. 207. 1 February. P. 107738.

20. **Martin-Willett R., Helmuth T., Abraha M., Bryan A. D., Hitchcock L., Lee K., Bidwell L. C.** Validation of a multisubstance online Timeline Followback assessment // Brain and Behavior. 2020. Vol. 10. No. 1. P. e01486.

21. **Irwin T. W., Morgenstern J., Parsons J. T., Wainberg M., Labouvie E.** Alcohol and sexual HIV risk behavior among problem drinking men who have sex with men: An event level analysis of timeline followback data // AIDS and Behavior. 2006. Vol. 10. No. 3. Pp. 299—307.

22. **Santos G. M., Strathdee S. A., El-Bassel N., et al.** Psychometric properties of measures of substance use a systematic review and meta-analysis of reliability, validity and diagnostic test accuracy // BMC Medical Research Methodology. 2020. Vol. 20. No. 1. Pp. 1—22.

23. **Kaplan B. A., Koffarnus M. N.** Timeline followback self-reports underestimate alcohol use prior to successful contingency management treatment // Alcohol and Alcoholism. 2019. Vol. 54. No. 3. Pp. 258—263.

24. **Пащенко А. Е., Тулупьев А. Л., Николенко С. И.** Моделирование заражения ВИЧ-инфекцией на основе данных о последних эпизодах рискованного поведения // Известия высших учебных заведений. Приборостроение. 2006. Т. 49. № 11. С. 33–34.

25. **Пащенко А. Е., Тулупьев А. Л., Николенко С. И.** Статистическая оценка вероятности заражения ВИЧ-инфекцией на основе данных о последних эпизодах рискованного поведения // Труды СПИИРАН. 2006. Т. 2. № 3. С. 257–268.

26. **Степанов Д. В., Мусина В. Ф., Суворова А. В., Тулупьев А. Л., Сироткин Т. В., Тулупьева Т. В.** Функция правдоподобия с гетерогенными аргументами в идентификации пуассоновской модели рискованного поведения // Труды СПИИРАН. 2012. № 4 (23). С. 157–184.

27. **Suvorova A.** Exploring Bayesian belief network for risky behavior modelling: discretization and latent variables // Proceedings of the II International Scientific and Practical Conference "Fuzzy Technologies in the Industry–FTI". Ulyanovsk, Russia, October 23 – 25, 2018. Pp. 63–70.

28. **Daley D. J., Vere-Jones D.** An introduction to the theory of point processes. Vol. I: Elementary theory and methods. 2nd Edition. New York: Springer, 2003. 491 p.

29. **Cook R. J., Lawless J.** The statistical analysis of recurrent events. New York: Springer Science & Business Media, 2007. 404 p.

30. **Andersen P. K., Borgan O., Gill R. D., Keiding N.** Statistical models based on counting processes. New York: Springer Science & Business Media, 2012. 784 p.

31. **Grandell J.** Mixed Poisson processes. [Monographs on Statistics and Applied Probability. Vol. 77]. UK: Chapman and Hall / CRC Press. 1997. 280 p.

32. **Greenwood M., Yule G. U.** An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents // Journal of Royal Statistical Society. A. 1920. Vol. 83. No. 2. Pp. 255–279.

33. **Newbold E. M.** A contribution to the study of the human factor in the causation of accidents. London: Industrial Health Research Board, 1926. 75 p.

34. **Therneau T. M., Grambsch P. M.** Modeling survival data: Extending the Cox model, New York: Springer Science + Business Media, 2000. 287 p.

35. **Balan T. A., Putter H.** frailtyEM: An R package for estimating semiparametric shared frailty models // Journal of Statistical Software. 2019. Vol. 90. No. 7. Pp. 1–29.

THE AUTHORS

STOLIAROVA Valeriia F.
St. Petersburg Federal Research Center of the Russian Academy of Sciences
39, 14-th Liniya V. I., St. Petersburg, 199178, Russia
vfs@dscs.pro
ORCID: 0000-0002-1666-2186

TULUPYEV Alexander L.
St. Petersburg State University
St. Petersburg Federal Research Center of the Russian Academy of Sciences
7–9, Universitetskaya Emb., St. Petersburg, 199034, Russia
alt@dscs.pro
ORCID: 0000-0003-1814-4646

СВЕДЕНИЯ ОБ АВТОРАХ

СТОЛЯРОВА Валерия Фуатовна – *младший научный сотрудник Санкт-Петербургского Федерального исследовательского центра Российской академии наук, Санкт-Петербург, Россия.*
199178, Россия, г. Санкт-Петербург, 14-я линия В. О., 39
vfs@dscs.pro
ORCID: 0000-0002-1666-2186

ТУЛУПЬЕВ Александр Львович — *профессор, доктор физико-математических наук, профессор кафедры информатики Санкт-Петербургского государственного университета, главный научный сотрудник Санкт-Петербургского Федерального исследовательского центра Российской академии наук, Санкт-Петербург, Россия.*

199034, Россия, г. Санкт-Петербург, Университетская наб., 7—9

alt@dscs.pro

ORCID: 0000-0003-1814-4646

Received 02.11.2021. Approved after reviewing 03.12.2021. Accepted 03.12.2021.

Статья поступила в редакцию 02.11.2021. Одобрена после рецензирования 03.12.2021. Принята 03.12.2021.