ON A NEW REGULARLY VARYING GENERALIZED HYPERGEOMETRIC DISTRIBUTION OF THE SECOND TYPE

EDUARD DANIELIAN, ROBERT CHITCHYAN and DAVOOD FARBOD

Communicated by Dan Crişan

In this paper, we shall attempt to propose a *new* multi-parametric family of stationary distributions of *standard Birth-Death processes* with a special form of coefficients, so-called *Generalized Hypergeometric Distribution of the Second Type* (*GHS*). A subfamily of the *GHS* that varies regularly at infinity, exhibits asymptotically constant slowly varying component, decreases, is log-downward convex and unimodal are obtained for the needs of biomolecular systems. Moreover, as examples, we will examine such regularly varying frequency distribution with three real data sets in bioinformatics.

AMS 2010 Subject Classification: 60E05, 60J80, 60J28, 62P10.

Key words: asymptotic expansion, biomolecular systems, standard Birth-Death process, Generalized Hypergeometric Distribution of the Second Type (GHS), Kolmogorov-Smirnov (K-S) test, Regularly varying GHS (Regular GHS).

1. INTRODUCTION AND PRELIMINARIES

Based on data sets for various large-scale biomolecular systems some common *statistical properties* have been discovered. From the mathematical point of view these are: *skewness to the right*, *regular variation at infinity*, *unimodality*, *continuous dependence on the parameters (stability)*, *convexity*, etc. of frequency distributions. Any distribution satisfying the *statistical properties* above has a chance to be approved by biologists in order to be used, at least, in one among great variety of evolutionary biomolecular systems (see [2, p. 1]).

The mechanism of the dynamic of large-scale biomolecular systems are often modeled with the help of *standard Birth-Death processes* with various types of coefficients. The stationary distributions of the process, which have moderate growth and a skew to the right, may be applied as frequency distributions of different phenomena taking place in large-scale biomolecular systems (see [1, 2]). On the basis of the *standard Birth-Death* models several frequency distributions have been introduced. We refer the readers, for example, to the works of Glanzel and Schubert [8], Bornholdt and Ebel [4], Kuznetsov [12, 13], Kuznetsov *et al.* [14], Astola and Danielian [1], Danielian and Astola [5, 6]. But, the great variety and diversity of such systems do not allow to suggest a universal mathematical model which explains the mechanism of their dynamics (see [2]). In this paper, we continue to discover new possibilities of *standard Birth-Death processes* for biomolecular systems.

1.1. STANDARD BIRTH-DEATH PROCESS

We describe the *standard Birth-Death process*, say

$$\{\xi(t), t \geq 0\} \tag{1}$$

where t denotes the time. The transition probabilities (see [2, p. 149])

$$P_{m,g} = P\Big(\xi(t+h) = g \mid \xi(h) = m\Big) = P_{m,g}(h,t)$$

of the process (1) for any numbers $h, t \in R^+$ and for any integers $m = 0, 1, 2, ...,$ $g = 0, 1, 2, ...,$ do not depend on h. Meanwhile, when $t \longrightarrow 0$, we assume

$$P_{m,g}(t) = o(t) \quad for\ 1 < |m-g| < \infty, \quad P_{m,m+1}(t) = \lambda_m t + o(t),$$
$$P_{m+1,m}(t) = \mu_{m+1} t + o(t).$$

The assumptions conclude $P_{m,m}(t) = 1 - (\lambda_m + \mu_m)t + o(t)$ when $t \longrightarrow 0$.

According to the assumptions, the state probabilities, say $p_x(t) = P(\xi(t) = x)$ with $x = 0, 1, 2, ...$ at moment $t \geq 0$, satisfy the following differential equations (see [2, p. 149])

$$\begin{cases} \frac{dp_n(t)}{dt} = -(\lambda_n + \mu_n)p_n(t) + \lambda_{n-1}\, p_{n-1}(t) + \mu_{n+1}\, p_{n+1}(t), & n = 1, 2, ..., \\ \frac{dp_0(t)}{dt} = -\lambda_0 p_0(t) + \mu_1\, p_1(t), \end{cases} \tag{2}$$

with arbitrary initial conditions $p_m(0) \geq 0$, $m = 0, 1, 2, ...,$ $\sum_{m\geq 0} p_m(0) = 1$. Without loss of generality, let us have $p_0(0) = 1$ and $p_x(0) = 0$, $x = 1, 2, ...$.

Note 1. We notice that (1) ia a Markovian process with continuous time and countable numbers of states (see [2, p. 150]).

Note 2. It is well-known that the necessary and sufficient condition for the existence of the stationary solution of the system of differential equations (2) is

$$\sum_{n=1}^{\infty} \prod_{m=1}^{n} \frac{\lambda_{m-1}}{\mu_m} < \infty, \tag{3}$$

where $\{\frac{\lambda_{m-1}}{\mu_m}\}_1^\infty$ presents a sequence of ratios of *birth* and *death* coefficients. Also, $p_m = \lim_{t\longrightarrow\infty} p_m(t)$, $m = 0, 1, 2, \ldots$. See Astola and Danielian [2] for details.

If (3) holds, then the probabilities of the stationary states form a distribution of type $\{p_n\}$ (see, for example, [5, p. 406]):

$$(4)\qquad \begin{cases} p_x = p_0 \cdot \prod_{m=1}^{x} \frac{\lambda_{m-1}}{\mu_m}, & x = 1, 2, \ldots, \\ p_0 = \left(1 + \sum_{n=1}^{\infty} \prod_{m=1}^{n} \frac{\lambda_{m-1}}{\mu_m}\right)^{-1}. \end{cases}$$

Danielian and Astola [6], based on (1)–(4), built a three-parametric regular *Generalized Hypergeometric Distribution.* In this paper, again using (1)–(4), we construct a *new* multi-parametric family of *Generalized Hypergeometric Distributions*, say, of the *Second Type*, which presents a family of stationary distributions of the *Birth-Death processes* with a special form of coefficients. We call the *Second Type*, because this *new* family is created as a generalization of the regular *Generalized Hypergeometric Distribution* proposed by Danielian and Astola [6] as the First Type.

We extract from the introduced parametric family a subfamily of regularly varying distributions for the needs of biomolecular applications. In summary, the aim of this paper is:

1. to show that *GHS* is generated by the *standard Birth-Death process*;

2. to extract a subfamily of regularly varying *GHS* which we call *Regular GHS*;

3. to show that such regular subfamily satisfies in some important *statistical properties* such as *unimodality*, *convexity*, etc., in order to suggest as a *new* frequency distribution for biomolecular applications;

4. to fit some real data sets arising in biomolecular systems with such regular subfamily.

The remainder of the paper is organized as follows. The main results of the paper are proposed in Sections 2, 3, 4, 5 and 6. Conclusion is given in Section 7.

2. **GHS**

Consider the following series

$$(5)\qquad T(\theta) \stackrel{def}{=} 1 + \sum_{n=1}^{\infty} \theta^n \cdot \frac{\left(\prod_{i=1}^{k} \Gamma(\alpha_i + n\nu_i)\right)}{\left(\prod_{j=1}^{r} \Gamma(\beta_j + n\omega_j)\right)},$$

where $\Gamma(\cdot)$ is Euler's Gamma Function, *i.e.* $\Gamma(z)=\int_0^{\infty} e^{-\varrho}\varrho^{z-1}d\varrho$.
Here: θ; $\alpha_1,\alpha_2,...,\alpha_k$; $\nu_1,\nu_2,...,\nu_k$; $\beta_1,\beta_2,...,\beta_r$; $\omega_1,\omega_2,...,\omega_r$ are *positive distinct* parameters.

LEMMA 1. *We have:*

1. The series (5) converges if $\sum_{i=1}^{k}\nu_i-\sum_{j=1}^{r}\omega_j<0$, *and diverges if* $\sum_{i=1}^{k}\nu_i-\sum_{j=1}^{r}\omega_j>0$.

2. Let

$$\sum_{i=1}^{k}\nu_i-\sum_{j=1}^{r}\omega_j=0, \tag{6}$$

Then the series (5) converges if $\sum_{i=1}^{k}\nu_i\ln\nu_i-\sum_{j=1}^{r}\omega_j\ln\omega_j<\ln\theta$, *and diverges if* $\sum_{i=1}^{k}\nu_i\ln\nu_i-\sum_{j=1}^{r}\omega_j\ln\omega_j>\ln\theta$.

3. Let (6) holds and

$$\sum_{i=1}^{k}\nu_i\ln\nu_i-\sum_{j=1}^{r}\omega_j\ln\omega_j=\ln\theta, \tag{7}$$

Then the series (5) converges if $\sum_{i=1}^{k}\alpha_i-\sum_{j=1}^{r}\beta_j<\frac{k-r-2}{2}$.

The proof of *Lemma 1* can be found in Appendix.

Now, let us consider the *standard Birth-Death process* with the following coefficients $(n=0,1,2,...,)$:

$$\lambda_n=\theta\cdot\prod_{i=1}^{k}\frac{\Gamma(\alpha_i)\cdot\Gamma(\alpha_i+(n+1)\nu_i)}{\Gamma(\alpha_i+n\nu_i)},\qquad \mu_{n+1}=\prod_{j=1}^{r}\frac{\Gamma(\beta_j)\cdot\Gamma(\beta_j+(n+1)\omega_j)}{\Gamma(\beta_j+n\omega_j)}. \tag{8}$$

The well-known condition $\sum_{n=1}^{\infty}\prod_{m=1}^{n}\frac{\lambda_{m-1}}{\mu_m}<+\infty$ of the existence of stationary distribution of the *standard Birth-Death process* in our case takes the form

$$+\infty>\sum_{n=1}^{\infty}\theta^n\cdot\prod_{s=1}^{n}\cdot\frac{\prod\limits_{i=1}^{k}\frac{\Gamma(\alpha_i)\Gamma(\alpha_i+s\nu_i)}{\Gamma(\alpha_i+(s-1)\nu_i)}}{\prod\limits_{j=1}^{r}\frac{\Gamma(\beta_j)\Gamma(\beta_j+s\omega_j)}{\Gamma(\beta_j+(s-1)\omega)}}=\sum_{n=1}^{\infty}\theta^n\cdot\frac{\prod\limits_{i=1}^{k}\frac{\Gamma(\alpha_i)\Gamma(\alpha_i+\nu_i)\cdots\Gamma(\alpha_i+n\nu_i)}{\Gamma(\alpha_i)\Gamma(\alpha_i+\nu_i)\cdots\Gamma(\alpha_i+(n-1)\nu_i)}}{\prod\limits_{j=1}^{r}\frac{\Gamma(\beta_j)\Gamma(\beta_j+\omega_j)\cdots\Gamma(\beta_j+n\omega_j)}{\Gamma(\beta_j)\Gamma(\beta_j+\omega_j)\cdots\Gamma(\beta_j+(n-1)\omega_j)}}$$

$$=\sum_{n=1}^{\infty}\theta^n\cdot\frac{\prod\limits_{i=1}^{k}\Gamma(\alpha_i+n\nu_i)}{\prod\limits_{j=1}^{r}\Gamma(\beta_j+n\omega_j)},$$

which is equivalent to the condition $T(\theta)<+\infty$, where $T(\theta)$ is given by formula (5). Then, in this case, the stationary distribution $\{p_n\}$ where

$$p_n=p_n(\theta,\alpha_i,\nu_i,\beta_j,\omega_j),\quad n=0,1,2,...,$$

takes the form

$$(9)\qquad \begin{cases} p_x = p_0 \cdot \theta^x \cdot \dfrac{\prod\limits_{i=1}^{k} \Gamma(\alpha_i + x\nu_i)}{\prod\limits_{j=1}^{r} \Gamma(\beta_j + x\omega_j)}, & x = 1, 2 \ldots, \\ p_0 = \left(1 + \sum\limits_{n=1}^{\infty} \theta^n \cdot \dfrac{\prod\limits_{i=1}^{k} \Gamma(\alpha_i + n\nu_i)}{\prod\limits_{j=1}^{r} \Gamma(\beta_j + n\omega_j)}\right)^{-1}, & \end{cases}$$

under the fulfilment of following conditions (see *Lemma 1*):

$$(a) \quad \sum_{i=1}^{k} \nu_i - \sum_{j=1}^{r} \omega_j < 0;$$

$$(b) \quad \sum_{i=1}^{k} \nu_i - \sum_{j=1}^{r} \omega_j = 0 \quad \text{and} \quad \sum_{i=1}^{k} \nu_i \ln \nu_i - \sum_{j=1}^{r} \omega_j \ln \omega_j < \ln \theta;$$

$$(10) \qquad (c) \begin{cases} \sum\limits_{i=1}^{k} \nu_i - \sum\limits_{j=1}^{r} \omega_j = 0, \quad \sum\limits_{i=1}^{k} \nu_i \ln \nu_i - \sum\limits_{j=1}^{r} \omega_j \ln \omega_j = \ln \theta, \\ \sum\limits_{i=1}^{k} \alpha_i - \sum\limits_{j=1}^{r} \beta_j < \frac{k-r-2}{2}. \end{cases}$$

COROLLARY 1. *The family of GHS, i.e. (9), has been constructed by (4) with coefficients of form (8).*

3. EXTRACTING REGULARLY VARYING DISTRIBUTION

Now, we are able to extract regularly varying frequency distribution from the family of type (9)–(10) as $n \to \infty$. Before that, let us remaind the following definitions (see [2, p. 11]).

Definition 1. The frequency distribution $\{p_n\}$ varies regularly at infinity with exponent $(-\rho)$ if it may be presented in the form

$$(11) \qquad p_n = n^{-\rho} \cdot L(n)(1 + o(1)), \quad n \to +\infty,$$

where $L(n) > 0$ for $n = 1, 2, ...$, and for $s = 2, 3, ...$, $\lim_{n\to\infty} \frac{L(sn)}{L(n)} = s^{-\rho}$.

Definition 2. If for $s = 2, 3, ...$, the limit exists

$$(12) \qquad \lim_{n\to\infty} \frac{L(sn)}{L(n)} = 1,$$

then we say that $\{p_n\}$ exhibits the asymptotically constant slowly varying component if in representation (12) we have

$$\lim_{n\to+\infty} L(n) = L \in R^+ = (0, +\infty).$$

Let us propose the following theorem.

Theorem 1. *The GHS $\{p_n\}$ given by (9) varies regularly at infinity if the condition (c) (i.e. (10)) holds. Then, the exponent of the regular variation of $\{p_n\}$ equals to*

$$-\rho = \left(\sum_{i=1}^{k} \alpha_i - \sum_{j=1}^{r} \beta_j - \frac{k-r}{2} \right) < -1 \tag{13}$$

Moreover, $\{p_n\}$ exhibits the asymptotically constant slowly varying component

$$L = p_0 \cdot \left((2\pi)^{\frac{k-r}{2}} \cdot \frac{\prod_{i=1}^{k} \nu_i^{\alpha_i-(1/2)}}{\prod_{j=1}^{r} \omega_j^{\beta_j-(1/2)}} \right) \in R^+. \tag{14}$$

Proof. From (9) and (20), (21), (22) (in Appendix) it follows that only under the condition (c) the ratio (p_n/p_0) has the asymptotic representation (11)–(12). In other words, we have

$$\begin{aligned} p_n &= p_0 \cdot \left((2\pi)^{\frac{k-r}{2}} \cdot \frac{\prod_{i=1}^{k} \nu_i^{\alpha_i-(1/2)}}{\prod_{j=1}^{r} \omega_j^{\beta_j-(1/2)}} \right) \cdot n^{\sum_{i=1}^{k} \alpha_i - \sum_{j=1}^{r} \beta_j - \frac{k-r}{2}} (1+o(1)) \\ &= L \cdot n^{-\rho} \cdot (1+o(1)), \quad n \to \infty, \end{aligned} \tag{15}$$

where $-\rho$ and L are given by formulas (13) and (14) respectively. So, $-\rho$ is the exponent of $\{p_n\}$'s regular variation. At the same time, remaind that the asymptotically constant slowly varying component (if exists) is defined by a limit $L = \lim_{n\to+\infty} L(n) \in R^+$ or $L = \lim_{n\to+\infty} n^\rho p_n \in R^+$. Comparing it with (15) we conclude that L given by (14) is an asymptotically constant slowly varying component for $\{p_n\}$. The proof of *Theorem 1* is complete. □

4. ASYMPTOTIC EXPANSION WITH TWO TERMS

We give the following theorem.

Theorem 2. *Let for the GHS (9) the condition (c), i.e. (10), holds. Then it exhibits the asymptotic expansion*

$$p_n = \frac{L}{n^\rho} + \frac{M}{n^{\rho+1}} + o\left(\frac{1}{n^{\rho+1}}\right), \; n \to +\infty,$$

with ρ, L given by (13),(14) respectively, and

$$M = \frac{L}{2}\left\{\sum_{i=1}^{k}\frac{1}{\nu_i}\left(\alpha_i(\alpha_i-1)+\frac{1}{6}\right)-\sum_{j=1}^{r}\frac{1}{\omega_j}\left(\beta_j(\beta_j-1)+\frac{1}{6}\right)\right\}.$$

See Appendix for the proof of *Theorem 2.*

5. UNIMODALITY AND CONVEXITY

Unimodality and convexity are features of interest for distributions arising in bioinformatics. For more details see, for example, [2, 3]. In this section, we would like to prove such features.

An interesting particular case arises when in (9) we put $\nu_1 = \nu_2 = \cdots = \nu_k = 1$ and $\omega_1 = \omega_2 = \cdots = \omega_r = 1$. Considering the regularly varying distribution in this case we have to take condition (c) given by (10). The condition (c) implies

$$k = r, \quad \theta = 1, \quad \sum_{i=1}^{k}(\alpha_i - \beta_i) < -1. \tag{16}$$

From *Theorem 2*, we conclude (see, (13) and (14))

$$\rho = -\left(\sum_{i=1}^{k}(\alpha_i-\beta_i)\right) > 1, \quad L = p_0 \in R^+, M = \frac{p_0}{2}\left\{\sum_{i=1}^{k}(\alpha_i^2-\beta_i^2)+\rho\right\}.$$

From the model (9) we are able to extract a subfamily of regular varying distribution $\{\widehat{p}_n\}$ where $\widehat{p}_n = \widehat{p}_n(\alpha_i, \beta_i)$, $n = 0, 1, 2, ...$, are of the form

$$\begin{cases} \widehat{p}_x = \widehat{p}_0 \cdot \prod_{i=1}^{k}\frac{\Gamma(\alpha_i+x)}{\Gamma(\beta_i+x)}, & x = 1, 2, ..., \\ \widehat{p}_0 = \left(1+\sum_{n=1}^{\infty}\prod_{i=1}^{k}\frac{\Gamma(\alpha_i+n)}{\Gamma(\beta_i+n)}\right)^{-1}. \end{cases} \tag{17}$$

$\alpha_i \in (0,\infty)$ and $\beta_i \in (0,\infty)$, $i = 1, 2, ..., k$, are the parameters of the model.

We are going to find conditions under which in our case $\{\widehat{p}_n\}$ *decreases* and is *log-downward convex.* It means that $\widehat{p}_n > \widehat{p}_{n+1}$ and $\frac{\widehat{p}_n}{\widehat{p}_{n+1}} > \frac{\widehat{p}_{n+1}}{\widehat{p}_{n+2}}$ for all n.

Notice that the Gamma Function $\Gamma(x)$ increases when its positive argument x increases. Because of the symmetry of multipliers under the product at the right-hand side of formula for $\widehat{p}_n$ we may propose the following assertion.

LEMMA 2. *If there is a finite sequence of natural numbers $i_1, i_2, ..., i_k$ with $\{i_1, i_2, ..., i_k\} = \{1, 2, ..., k\}$ such that*

$$\alpha_{i_1} < \beta_1, \quad \alpha_{i_2} < \beta_2, \quad \ldots, \quad \alpha_{i_k} < \beta_k, \tag{18}$$

then the frequency distribution (17) satisfying condition (16) decreases and is log-downward convex.

Proof. Without loss of generality, we may assume that

$$\alpha_1 < \beta_1, \quad \alpha_2 < \beta_2, \quad \ldots, \quad \alpha_k < \beta_k. \tag{19}$$

From (17) and based on the identity $\Gamma(x+1) = x \cdot \Gamma(x)$ for $x \in R^+$, it follows that

$$\frac{\widehat{p}_{n+1}}{\widehat{p}_n} = \prod_{i=1}^{k} \frac{\Gamma(\alpha_i + n + 1)}{\Gamma(\alpha_i + n)} \cdot \frac{\Gamma(\beta_i + n)}{\Gamma(\beta_i + n + 1)} = \prod_{i=1}^{k} \frac{\alpha_i + n}{\beta_i + n}, \qquad n = 0, 1, 2, \ldots.$$

It at once implies that if the condition (19) holds, then the proof of *Lemma 2* is complete. □

Corollary 2. *The subfamily of Regular GHS (17) satisfies some observed statistical properties (mentioned in the section 1) and hence it can be considered, under conditions, as a model for biomolecular needs.*

6. FITTING OF THE REGULAR GHS

In this section, we shall examine the model introduced for three real data sets. That is why, let ξ be a random variable with probability distribution (17). In order to apply the probability function (17) to the data (compare to [14, p. 399; 13, p. 378; 7, p. 215]) we consider the random variable ξ as *doubly-truncated*. Namely, random variable ξ is restricted from 1 to the maximum observed in each data set. In addition, some plots of the distribution (17) for different values of the parameters are presented. In order to numerical studies, in this section, let us assume $k = r = 4$, *i.e.* $i = 1, 2, 3, 4$.

Example 1. We consider the number of amino acids in the protein chain (see [10]) as a real data set in the following Table:

Table 1

36	153	146	97	83	46	150	43
29	30	71	58	26	40	70	138

The *p-value* of the *K-S Test* is *0.6834*, which does not reject the adequacy of the *Regular GHS* (17) for the number of amino acids. Comparing to Farbod and Gasparian [7], we plot the *empirical cumulative distribution function* (ecdf) and fitted *cumulative distribution function* (cdf) for the number of amino acids data in Fig. 1.

Example 2. Let us have the number of residues in globular proteins (see [11]) as a real data set in the Table 2.

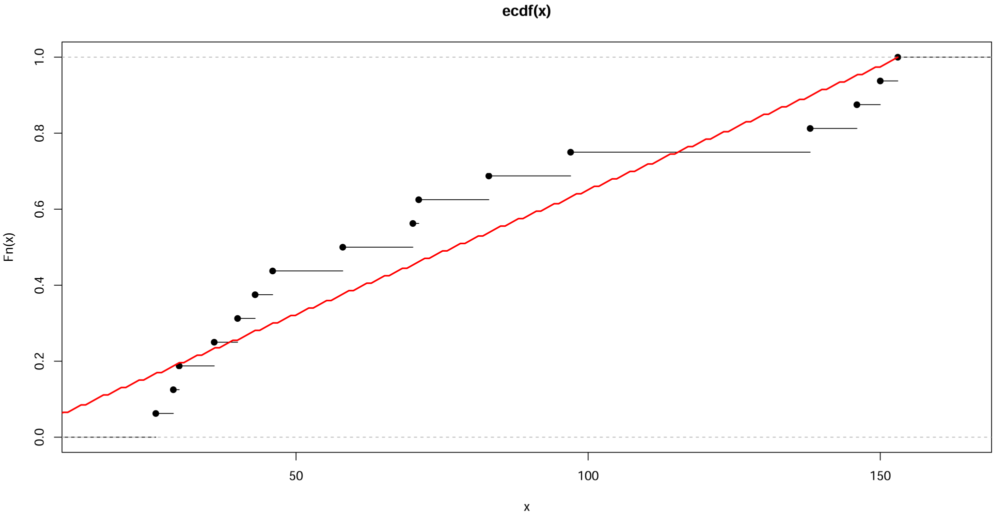

Fig. 1 – Fitting of the doubly-truncated *Regular GHS* to the data of Table 1. The dashed line is the ecdf of data and the solid line is the fitted cdf.

Table 2

85	103	103	112	134	82	54	98	138
54	125	99	36	29	51	71	26	62

Using *K-S Test* the *p-value* is *0.449*, which does not reject the adequacy of the *Regular GHS* (17) for the number of residues. Again, by comparing to Farbod and Gasparian [7], we plot the ecdf and fitted cdf of the number of residues data in Fig. 2.

Example 3. As a real data set, let us have the number of exons in human genes (see [15]) in the Table 3:

Table 3

2	3	3	8	14	118	29	26	26	27	79

The *p-value*, with the help of *K-S Test*, is *0.8739* which does not reject the adequacy of the *Regular GHS* (17) for the number of exons in human genes. Fig. 3 indicates plot the ecdf and fitted cdf of the number of exons in human genes.

6.1. FIGURES OF THE MODEL

We present some plots of the doubly-truncated *Regular GHS* (17) for different values of the parameters in Fig. 4. It is readily seen that the Plots have right skewness.

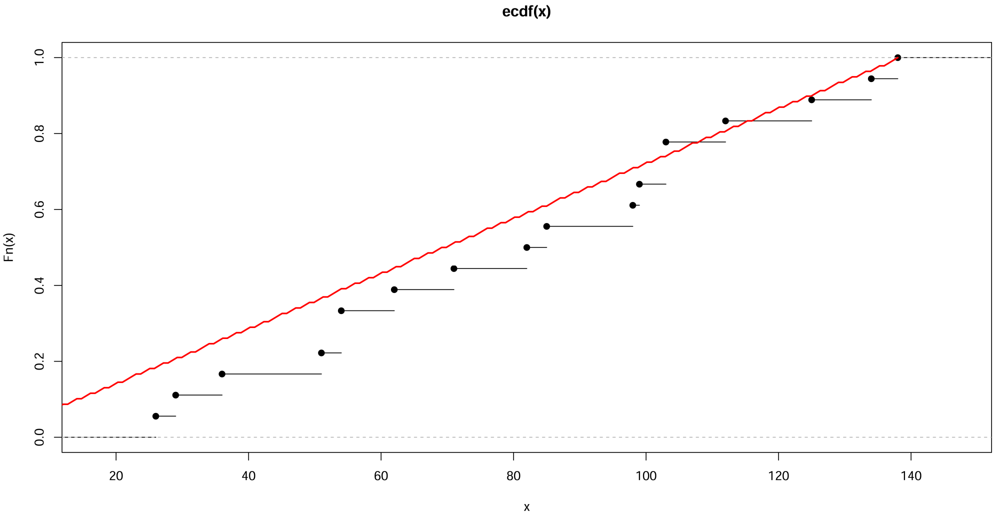

Fig. 2 – Fitting of the doubly-truncated *Regular GHS* to the data of Table 2. The dashed line is the ecdf of data and the solid line is the fitted cdf.

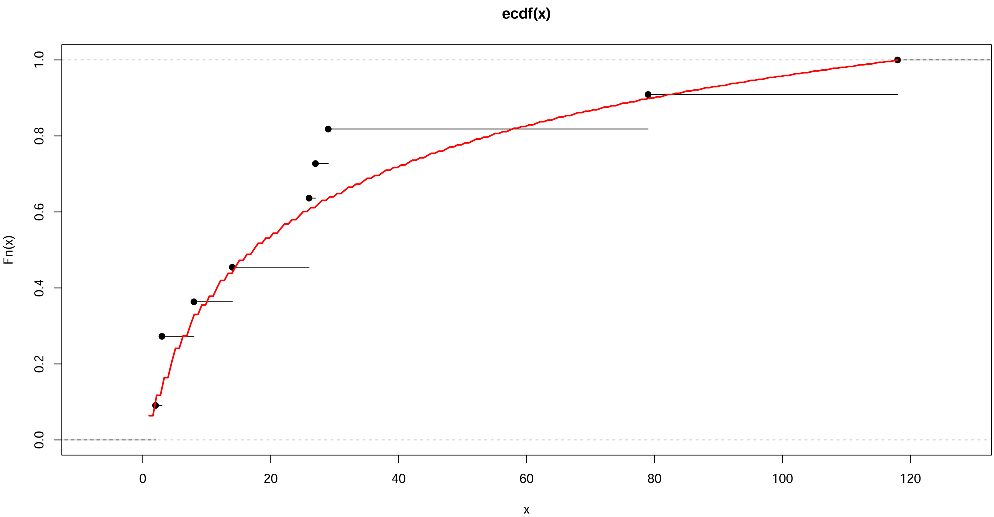

Fig. 3 – Fitting of the doubly-truncated *Regular GHS* to the data of Table 3. The dashed line is the ecdf of data and the solid line is the fitted cdf.

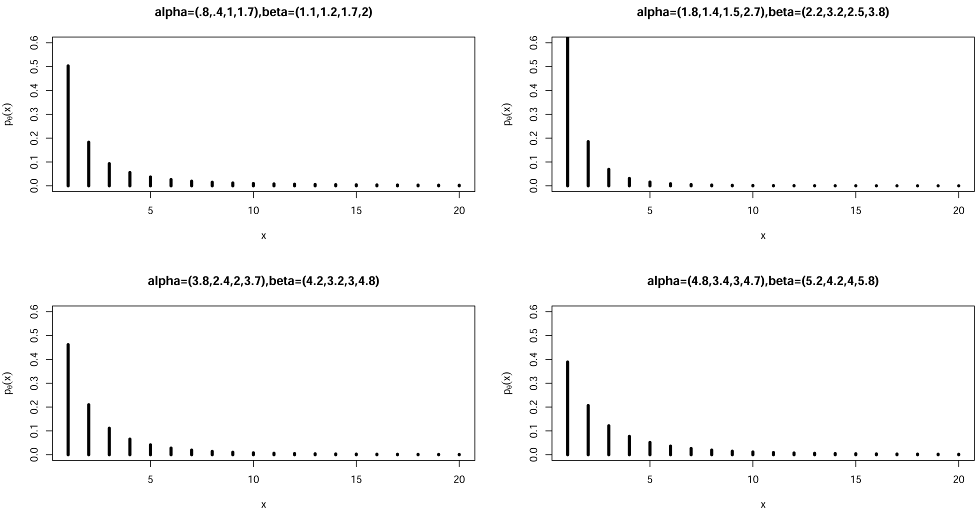

Fig. 4. – Illustrations of the doubly-truncated *Regular GHS* for different values of the parameters α_i and β_i, when $i = 1, 2, 3, 4$.

7. CONCLUSION

In this study, a *new* multi-parametric family of stationary distribution of *standard Birth-Death process*, say *GHS*, has been proposed. A subfamily (*Regular GHS*) of such distribution that varies regularly at infinity has been extracted and the properties of this subfamily have been investigated from the point of view of biomolecular systems. Meanwhile, as examples, three real data sets on the number of amino acids in the protein chain, the number of residues in globular protein and the number of exons in human genes have been fitted with the *Regular GHS* (17). As we saw from the Examples 1–3, the *p-values* are 0.6834, 0.449 and 0.8739, respectively. It indicates that the *Regular GHS* (17) is a suitable candidate model to fit such discrete data.

8. APPENDIX

8.1. PROOF OF LEMMA 1

Using the asymptotic expansion (see [9])

$$\Gamma(z) = z^{z-(1/2)} \cdot e^{-z} \cdot \sqrt{2\pi} \cdot \left(1 + O\left(\frac{1}{z}\right)\right), \quad z \to \infty.$$

for the n_{th} term at the right-hand-side of (5) we obtain

$$\theta^n \cdot \frac{\prod_{i=1}^{k} \Gamma(\alpha_i + n\nu_i)}{\prod_{j=1}^{r} \Gamma(\beta_j + n\omega_j)} = \theta^n \cdot (2\pi)^{\frac{k-r}{2}} \cdot e^{-\left(\sum_{i=1}^{k} \alpha_i - \sum_{j=1}^{r} \beta_j\right)} \cdot e^{-n\left(\sum_{i=1}^{r} \nu_i - \sum_{j=1}^{r} \omega_j\right)}$$

$$\cdot \frac{\prod_{i=1}^{k} (\alpha_i + n\nu_i)^{\alpha_i + n\nu_i - \frac{1}{2}}}{\prod_{j=1}^{r} (\beta_j + n\omega_j)^{\beta_j + n\omega_j - \frac{1}{2}}} \times \left(1 + O\left(\frac{1}{n}\right)\right), \quad n \to +\infty.$$

Taking into account that

$$\frac{\prod_{i=1}^{k} (\alpha_i + n\nu_i)^{\alpha_i + n\nu_i - \frac{1}{2}}}{\prod_{j=1}^{k} (\beta_j + n\omega_j)^{\beta_j + n\omega_j - \frac{1}{2}}} = n^{\sum_{i=1}^{k} \alpha_i - \sum_{j=1}^{r} \beta_j - \frac{k-r}{2}} \cdot n^{n\left(\sum_{i=1}^{k} \nu_i - \sum_{j=1}^{r} \omega_j\right)} \cdot \frac{\prod_{i=1}^{k} \left(1 + \frac{\alpha_i}{n\nu_i}\right)^{n\nu_i}}{\prod_{j=1}^{r} \left(1 + \frac{\beta_j}{n\omega_j}\right)^{n\omega_j}}$$

$$\times \left(\frac{\prod_{i=1}^{k} \nu_i^{\alpha_i - \frac{1}{2}}}{\prod_{j=1}^{r} \omega_j^{\beta_j - \frac{1}{2}}} \right) \cdot \left(\frac{\prod_{i=1}^{k} \nu_i^{\nu_i}}{\prod_{j=1}^{r} \omega_j^{\omega_j}} \right)^n \cdot \left(1 + O\left(\tfrac{1}{n}\right)\right), \quad n \to +\infty,$$

we derive

(20)

$$\theta^n \cdot \frac{\prod_{i=1}^{k} \Gamma(\alpha_i + n\nu_i)}{\prod_{j=1}^{r} \Gamma(\beta_j + n\omega_j)} = \left[(2\pi)^{\frac{k-r}{2}} \cdot \frac{\prod_{i=1}^{k} \nu_i^{\alpha_i - \frac{1}{2}}}{\prod_{j=1}^{r} \omega_j^{\beta_j - \frac{1}{2}}} \right] \cdot \left(\theta \cdot \frac{\prod_{i=1}^{k} \nu_i^{\nu_i}}{\prod_{j=1}^{r} \omega_j^{\omega_j}} \right)^n \cdot n^{\sum_{i=1}^{k} \alpha_i - \sum_{j=1}^{r} \beta_j - \frac{k-r}{2}}$$

$$\times n^{n\left(\sum_{i=1}^{k} \nu_i - \sum_{j=1}^{r} \omega_j\right)} \cdot e^{-n\left(\sum_{i=1}^{k} \nu_i - \sum_{j=1}^{r} \omega_j\right)} \cdot \left(1 + O\left(\frac{1}{n}\right)\right), \quad n \to +\infty.$$

Here, the following limit equalities were used

$$\lim_{n \to +\infty} \prod_{i=1}^{k} \left(1 + \frac{\alpha_i}{n\nu_i}\right)^{n\nu_i} = e^{\alpha_1 + \alpha_2 + \cdots + \alpha_k},$$

$$\lim_{n \to +\infty} \prod_{j=1}^{r} \left(1 + \frac{\beta_j}{n\omega_j}\right)^{n\omega_j} = e^{\beta_1 + \beta_2 + \cdots + \beta_r}.$$

The *main* term in (20) equals to $n^{n\left(\sum_{i=1}^{k} \nu_i - \sum_{j=1}^{r} \omega_j\right)}$. That is why the tail of the series (5), namely,

$$\sum_{n \geq n_0} \theta^n \cdot \frac{\prod_{i=1}^{k} \Gamma(\alpha_i + n\nu_i)}{\prod_{j=1}^{r} \Gamma(\beta_i + n\omega_j)} \qquad (21)$$

for n large enough converges if $\sum_{i=1}^{k}\nu_i - \sum_{j=1}^{r}\omega_j < 0$, and diverges if $\sum_{i=1}^{k}\nu_i - \sum_{j=1}^{r}\omega_j > 0$.

Now, let (6) holds. Due to (20), we have

(22)
$$\theta^n \cdot \frac{\prod_{i=1}^{k}\Gamma(\alpha_i+n\nu_i)}{\prod_{j=1}^{r}\Gamma(\beta_j+n\omega_j)} = \left((2\pi)^{\frac{k-r}{2}}\cdot\frac{\prod_{i=1}^{k}\nu_i^{\alpha_i-\frac{1}{2}}}{\prod_{j=1}^{r}\omega_j^{\beta_j-\frac{1}{2}}}\right)\cdot\left(\theta\cdot\frac{\prod_{i=1}^{k}\nu_i^{\nu_i}}{\prod_{j=1}^{r}\omega_j^{\omega_j}}\right)^n \cdot n^{\sum_{i=1}^{k}\alpha_i-\sum_{j=1}^{r}\beta_j-\frac{k-r}{2}} \times\left(1+O\left(\frac{1}{n}\right)\right), \quad n\to+\infty.$$

The *main* term at the right-hand-side of (22) equals to $\left(\theta\cdot\frac{\prod_{i=1}^{k}\nu_i^{\nu_i}}{\prod_{j=1}^{r}\omega_j^{\omega_j}}\right)^n$.
That is why, in this case, the tail of the series (5) for n large enough converges if

$$\left(\theta\cdot\frac{\prod_{i=1}^{k}\nu_i^{\nu_i}}{\prod_{j=1}^{r}\omega_j^{\omega_j}}\right)^n < 1 \quad \textit{or} \quad \sum_{i=1}^{k}\nu_i\ln\nu_i - \sum_{j=1}^{r}\omega_j\ln\omega_j < \ln\theta,$$

and diverges if

$$\left(\theta\cdot\frac{\prod_{i=1}^{k}\nu_i^{\nu_i}}{\prod_{j=1}^{r}\omega_j^{\omega_j}}\right)^n > 1 \quad \textit{or} \quad \sum_{i=1}^{k}\nu_i\ln\nu_i - \sum_{j=1}^{r}\omega_j\ln\omega_j > \ln\theta.$$

Now, we assume that (6) and (7) hold. Due to (22), for $n\to+\infty$, we have

(23)
$$\theta^n \cdot \frac{\prod_{i=1}^{k}\Gamma(\alpha_i+n\nu_i)}{\prod_{j=1}^{r}\Gamma(\beta_j+n\omega_j)} = \left((2\pi)^{\frac{k-r}{2}}\cdot\frac{\prod_{i=1}^{k}\nu_i^{\alpha_i-\frac{1}{2}}}{\prod_{j=1}^{r}\omega_j^{\beta_j-\frac{1}{2}}}\right)\cdot n^{\sum_{i=1}^{k}\alpha_i-\sum_{j=1}^{r}\beta_j-\frac{k-r}{2}}\cdot\left(1+O\left(\frac{1}{n}\right)\right).$$

The formula (23) states that the series (5) converges if $\sum_{i=1}^{k}\alpha_i - \sum_{j=1}^{r}\beta_j <$

$\frac{k-r-2}{2}$, and diverges if $\sum\limits_{i=1}^{k}\alpha_i - \sum\limits_{j=1}^{r}\beta_j \geq \frac{k-r-2}{2}$. The proof of *Lemma 1* is completed. □

8.2. PROOF OF THEOREM 2

Let us repeat the calculations of the n_{th} term at the right-hand-side of (5) now using the asymptotic expansion (see [9]) with two terms

$$\Gamma(z) = z^{z-1/2}e^{-z}\sqrt{2\pi}\left(1+\frac{1}{12z}+O\left(\frac{1}{z^2}\right)\right), \quad z\to+\infty,$$

and also taking into account condition (c) (see, (10) and (13)). We have for $n\to+\infty$

$$(24)\quad \theta^n\cdot\frac{\prod\limits_{i=1}^{k}\Gamma(\alpha_i+n\nu_i)}{\prod\limits_{j=1}^{k}\Gamma(\beta_j+n\omega_j)} = \theta^n\cdot(2\pi)^{\frac{k-r}{2}}e^{\rho-\frac{k-r}{2}}$$

$$\cdot\frac{\prod\limits_{i=1}^{k}\left\{(\alpha_i+n\nu_i)^{\alpha_i+n\nu_i-\frac12}\cdot\left(1+\frac{1}{12\cdot(\alpha_i+n\nu_i)}+O\left(\frac{1}{n^2}\right)\right)\right\}}{\prod\limits_{j=1}^{r}\left\{(\beta_j+n\omega_j)^{\beta_j+n\omega_j-\frac12}\cdot\left(1+\frac{1}{12\cdot(\beta_j+n\omega_j)}+O\left(\frac{1}{n^2}\right)\right)\right\}}$$

$$=\theta^n\cdot(2\pi)^{\frac{k-r}{2}}e^{\rho-\frac{k-r}{2}}\cdot\left(\frac{\prod\limits_{i=1}^{k}(\alpha_i+n\nu_i)^{\alpha_i+n\nu_i-\frac12}}{\prod\limits_{j=1}^{r}(\beta_j+n\omega_j)^{\beta_j+n\omega_j-\frac12}}\right)$$

$$\times\left(1+\frac{1}{12n}\left(\sum_{i=1}^{k}\frac{1}{\nu_i}-\sum_{j=1}^{r}\frac{1}{\omega_j}\right)+O\left(\frac{1}{n^2}\right)\right).$$

Here

$$\theta^n\cdot\frac{\prod\limits_{i=1}^{k}(\alpha_i+n\nu_i)^{\alpha_i+n\nu_i-\frac12}}{\prod\limits_{j=1}^{r}(\beta_j+n\omega_j)^{\beta_j+n\omega_j-\frac12}} = n^{-\rho}\cdot\left(\frac{\prod\limits_{i=1}^{k}\left(1+\frac{\alpha_i}{n\nu_i}\right)^{n\nu_i}}{\prod\limits_{j=1}^{r}\left(1+\frac{\beta_j}{n\omega_j}\right)^{n\omega_j}}\right)\cdot\left(\frac{\prod\limits_{i=1}^{k}\nu_i^{\alpha_i-\frac12}}{\prod\limits_{j=1}^{r}\omega_j^{\beta_j-\frac12}}\right)$$

$$\cdot\left(\frac{\prod\limits_{i=1}^{k}\left(1+\frac{\alpha_i}{n\nu_i}\right)^{\alpha_i-(1/2)}}{\prod\limits_{j=1}^{r}\left(1+\frac{\beta_j}{n\omega_j}\right)^{\beta_j-(1/2)}}\right),$$

or, according to the asymptotic expansion $(1+x)^\alpha = 1+\alpha x+O(x^2), x\to 0, \alpha>0$, we get

$$(25)\quad \theta^n \cdot \frac{\prod_{i=1}^{k} (\alpha_i + n\nu_i)^{\alpha_i + n\nu_i - \frac{1}{2}}}{\prod_{j=1}^{r} (\beta_j + n\omega_j)^{\beta_j + n\omega_j - \frac{1}{2}}} = n^{-\rho} \cdot \left(\frac{\prod_{i=1}^{k} \left(1 + \frac{\alpha_i}{n\nu_i}\right)^{n\nu_i}}{\prod_{j=1}^{r} \left(1 + \frac{\beta_j}{n\omega_j}\right)^{n\omega_j}} \right) \cdot \left(\frac{\prod_{i=1}^{k} \nu_i^{\alpha_i - \frac{1}{2}}}{\prod_{j=1}^{r} \omega_j^{\beta_j - \frac{1}{2}}} \right)$$

$$\times \left(1 + \frac{1}{n} \left(\sum_{i=1}^{k} \frac{\alpha_i(\alpha_i - \frac{1}{2})}{\nu_i} - \sum_{j=1}^{r} \frac{\beta_j(\beta_j - \frac{1}{2})}{\omega_j} \right) + O\left(\frac{1}{n^2}\right) \right), \quad n \to +\infty.$$

Let us consider the following expression's behavior as $x \to +\infty$

$$\left(1 + \frac{c}{x}\right)^x = exp\left\{x \cdot ln\left(1 + \frac{c}{x}\right)\right\} = \exp\left\{x\left(\frac{c}{x} - \frac{c^2}{2x^2} + O\left(\frac{1}{x^3}\right)\right)\right\}$$

$$= (\exp c) \exp\left\{-\frac{c^2}{2x} + O\left(\frac{1}{x^2}\right)\right\} = (\exp c)\left(1 - \frac{c^2}{2x} + O\left(\frac{1}{x^2}\right)\right), \; x \to +\infty.$$

Applying the above asymptotic expansion, as $n \to +\infty$, we obtain

$$(26)\quad \frac{\prod_{i=1}^{k}\left(1+\frac{\alpha_i}{n\nu_i}\right)^{n\nu_i}}{\prod_{j=1}^{r}\left(1+\frac{\beta_j}{n\omega_j}\right)^{n\omega_j}} = e^{\sum_{i=1}^{k}\alpha_i - \sum_{j=1}^{r}\beta_j} \cdot \frac{\prod_{i=1}^{k}\left(1-\frac{\alpha_i^2}{2n\nu_i}+O\left(\frac{1}{n^2}\right)\right)}{\prod_{j=1}^{r}\left(1-\frac{\beta_j^2}{2n\omega_j}+O\left(\frac{1}{n^2}\right)\right)}$$

$$= e^{-\rho + \frac{k-r}{2}} \cdot \left(1 - \frac{1}{2n} \cdot \left(\sum_{i=1}^{k} \frac{\alpha_i^2}{\nu_i} - \sum_{j=1}^{r} \frac{\beta_j^2}{\omega_j}\right) + O\left(\frac{1}{n^2}\right)\right).$$

Substituting (26) into (25), and after that (25) into (24) we come to the following asymptotic expansion

$$\theta^n \cdot \frac{\prod_{i=1}^{k} \Gamma(\alpha_i + n\nu_i)}{\prod_{j=1}^{r} \Gamma(\beta_j + n\omega_j)} = (2\pi)^{\frac{k-r}{2}} \left(\frac{\prod_{i=1}^{k} \nu_i^{\alpha_i - \frac{1}{2}}}{\prod_{j=1}^{r} \omega_j^{\beta_j - \frac{1}{2}}} \right)$$

$$\times \left\{ 1 + \frac{1}{2n} \left(\sum_{i=1}^{k} \left(\frac{\alpha_i(\alpha_i - 1)}{\nu_i} + \frac{1}{6\nu_i} \right) - \sum_{j=1}^{r} \left(\frac{\beta_j(\beta_j - 1)}{\omega_j} - \frac{1}{6\omega_j} \right) + O\left(\frac{1}{n^2}\right) \right) \right\}.$$

Based on the last asymptotic expansion the proof is finished. □

Acknowledgments. The authors are thankful to the anonymous referees for valuable comments and suggestions to improve the paper.

REFERENCES

[1] J. Astola and E. Danielian, *On regularly varying distributions generated by birth-death processes.* Facta Univ. Ser. Elect. Energ. **19** (2006), *1*, 109–131.

[2] J. Astola and E. Danielian, *Frequency Distributions in Biomolecular Systems and Growing Networks.* Tampere International Center for Signal Processing (TICSP) **31**, Tampere, Finland, 2007.

[3] J. Astola, E. Danielian and S. Arzumanyan, *Frequency distributions in bioinformatics: the development.* A review, Proc. Yerevan State Univ.: Ser. Phys. Math. Sci. **223** (2010), *3*, 3–22.

[4] S. Bornholdt and H. Ebel, *Wolrd wide web scaling exponent from Simon's 1955 model*, Phys. Rev. E **64** (2001), 035104(4).

[5] E. Danielian and J. Astola, *On the steady state of birth-death process with coefficients of moderate growth*, Facta Univ. Ser.: Elec. Energ. **17** (2004), *3*, 405–419.

[6] E. Danielian and J. Astola, *On regularly varying hypergeometric distributions*, Proceedings International TICSP Workshop on Spectral Methods and Multirate Signal Processing, Florence, Italy, 2-3 Sept., (2006); In Astola *et al.* (eds.): TICSP Series **34**, Tampere, Finland, 127–132.

[7] D. Farbod and K. Gasparian, *On the maximum likelihood estimators for some generalized Pareto-like frequency distribution.* J. Iran. Stat. Soc. (JIRSS), **12** (2013), *2*, 211–233.

[8] W. Glanzel and A. Schubert, *Predictive aspects of stochastic model for citation processes.* Inform. Process. Manag. **31** (1995), 1, 69–80.

[9] I.S. Gransteyn and I.M. Ryznik, *Tables of Integrals, Series and Products.* Academic Press. Translated from Russian by S. Technica, edited by A. Jeffrey and D. Zwillinger, 2007.

[10] U. Hobohm, M. Schare, R. Schneider and Ch. Sander, *Selection of representative protein data sets.* Protein Sci. (1992), **1**, 409–417.

[11] W. Kabsch and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.* Biopolymers **22** (1983), *12*, 2577–2637.

[12] V.A. Kuznetsov, *Distributions associated with stochastic processes of gene expression in a single eukaryotic cell.* EURASIP J. Appl. Signal Processing, **4** (2001), 258–296.

[13] V.A. Kuznetsov, *Family of skewed distributions associated with the gene expression and proteome evolution.* Signal Processing **33** (2003), **4**, 889–910.

[14] V.A. Kuznetsov, V.A. Pickalov, O.V. Senko and G.D. Knott, *Analysis of the evolving proteomes: Predictions of the number for protein domains in nature and the number of genes in eukaryotic organisms.* J. Biol. Systems **10** (2002), *4*, 381–407.

[15] T. Strachan and A.P. Read, *Human Molecular Genetics.* 2nd edition, NewYork, 1999.

Received 19 June 2014

"Yerevan" State University,
Chair of Probability Theory
and Mathematical Statistics,
0025 Yerevan, Armenia

"Yerevan" State University
Chair of Actuarial Mathematics,
0025 Yerevan, Armenia
chitchyanrobert@gmail.com
rchitchyan@gmail.com

"Quchan" University of Advanced Technology,
Department of Mathematics,
P.O. Box 94771-67335,
Quchan, Iran
d.farbod@qiet.ac.ir
d.farbod@gmail.com