Journal of Mathematical Extension
Vol. 16, No. 10, (2022) (10)1-16
URL: https://doi.org/10.30495/JME.2022.2173
ISSN: 1735-8299
Original Research Paper

A Continuous Analog of the Generalized Hypergeometric Distribution Generated by Dediscretization Method

D. Farbod*
Quchan University of Technology

A. Iranmanesh
Mashhad Branch, Islamic Azad University

M. Basirat
Mashhad Branch, Islamic Azad University

Abstract. In this paper, we shall attempt to introduce a new continuous analog of the regularly varying generalized hypergeometric distribution by the dediscretization method. The density, cumulative distribution, survival and hazard rate functions are obtained for the model. The density, cumulative distribution and hazard rate functions are illustrated by some figures. With the help of Monte Carlo method, simulation studies are done to obtain biases and mean square errors of the maximum likelihood estimations for the model's unknown parameters. We see that our simulation method has satisfactory results.

AMS Subject Classification: 62E10; 60E05; 62F10; 62E15
Keywords and Phrases: Continuous analog of the generalized hypergeometric distribution, Monte Carlo method, Dediscretization.

Received: September 2021; Accepted: January 2022
*Corresponding Author

1 Introduction and Preliminaries

The mechanism of large-scale biomolecular systems often is stated using Birth-Death Process (BDP) with various constraints on its coefficients. The stationary distributions of the process, which have skewness to the right, are considered as frequency distributions of different events arising in molecular evolution, biological networks and biosystems [3, 22]. Based on BDP and establishment the *statistical facts* such as regular variation at infinity, upward/downward convexity, convexity, unimodality, and etc several frequency distributions have been proposed (see [3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]). One of the important and useful frequency distributions is the three-parameter generalized hypergeometric distribution which was introduced by Danielian and Astola [7] (see also [3, 11]).

We note that all of distributions mentioned above are in discrete cases. In continuous cases, only Astola and Danielian [2] (see also [3]) considered continuous analog of the Waring distribution by the dediscretization method and then obtain new frequency distributions by discretization for the needs of bioinformatics systems. In this paper, by dediscretization method, we introduce a *Continuous Analog of Three-Parameter Regularly Varying Generalized Hypergeometric Distribution (in short we call CGHD)* as a new distribution and then consider some theoretical and numerical statistical inferences including simulation studies for this continuous distribution.

Astola and Danielian [2, 3] proposed some advantages and motivations of dediscretization idea, but after these studies in the year 2007, the dediscretization approach has not been considered more in researches and has been remained open problems in this subject. Hence, our motivation for this study is to:

a) introduce (by dediscretization method) a model CGHD as a new distribution;

b) obtain some statistical inferences such as cumulative distribution function (cdf), probability density function (pdf), hazard rate function (hrf) and survival function in practical forms for this continuous model;

c) do simulation studies by Monte Carlo method to obtain maximum likelihood estimations (MLEs) and their biases and mean square errors (MSEs) of the unknown parameters.

A three-parameter regularly varying generalized hypergeometric distribution generated by stochastic BDPs was introduced by Danielian and Astola [7] (see also [3, 11]). A revised form of the proposed frequency model by Danielian and Astola [7] has the following probability mass function [20]

$$\begin{cases} p_x(\alpha) = p_0(\alpha) \cdot \prod_{t=1}^{x} \frac{(t+r_1-1)(t+r_2-1)}{t(t+q)} \\ \\ p_0(\alpha) = \left[1 + \sum_{x=1}^{\infty} \prod_{t=1}^{x} \frac{(t+r_1-1)(t+r_2-1)}{t(t+q)}\right]^{-1} \end{cases} \tag{1}$$

where $p_\alpha(0)$ is the normalization factor, $x = 0, 1, 2, ...$; $\alpha = (r_1, r_2, q)$ is the unknown parameter such that $r_1 > 0$, $r_2 > 0$, $q > 0$. r_1 and r_2 are called numerator parameters and q denominator parameter. Also, $\rho = q + 2 - r_1 - r_2 > 1$ and $(-\rho)$ is called the exponent of regular variation of $p_x(\alpha)$. The discrete model (1) was introduced for discreption of phenomena arising in large-scale biomolecular systems. The model (1) so-called a three-parameter regularly varying generalized hypergeometric distribution generated by BDP.

Astola and Danielian [2, 3] suggested a method of replacement of sums in (1) by integrals (we present $\prod$ with the help of $\sum$), which does not change the behavior of distributions. It simplifies the obtained formulas and allows to suggestion of new distributions for bioinformatics. This method so-called *dediscretization*. Thereafter, it can be proposed new discrete distributions with the same properties as before for bioinofrmatics.

The remainder of the paper is formed as follows. In section 2 we introduce the model CGHD by the dediscretization method and then obtain cdf, pdf, hrf and survival function for this model in practical forms. Also, some figures for the cdf, pdf and hrf of the CGHD model are plotted. In section 3, using Monte Carlo method we do simulation study to obtain biases and MSEs for the MLEs of the unknown parameters. Section 4 concludes.

2 On the CGHD

For the formula (1), in dediscretization method we use equality

$$\prod_{t=1}^{x} \frac{(t+r_1-1)(t+r_2-1)}{t(t+q)} = \exp\Big(\sum_{t=1}^{x} \ln\big(\frac{(t+r_1-1)(t+r_2-1)}{t(t+q)}\big)\Big)$$

Let us have definition of dediscretization. Compared to Astola and Danielian [3] we present the following definition for the model (1).

Definition 1. *The function $F_x(\alpha)$ defined on $x \in (0,\infty)$*

$$F_x(\alpha) = \frac{\int_0^x \exp\big(\int_0^t \ln\big(\frac{(u+r_1-1)(u+r_2-1)}{u(u+q)}\big)du\big)dt}{\int_0^\infty \exp\big(\int_0^t \ln\big(\frac{(u+r_1-1)(u+r_2-1)}{u(u+q)}\big)du\big)dt} \tag{2}$$

is a dediscretization of $\{p_x(\alpha)\}$.

Let us explain the dediscretization for some class of distributions $\{p_x(\alpha)\}$. The dediscretization is a procedure on a class of distributions which leads to the concrete construction of corresponding and in some sense close class of smooth enough (for example, infinite differentiable) distribution functions. The constructed class of distribution function gives a continuous form and it has to satisfy definite restrictions. In the other words, it must conserve the main properties of distributions of the original class such as regular variation at infinity with the same exponent, upward/downward convexity, unimodality, convexity, existence of moments, etc (for more details see [2, 3]).

Now, we obtain cdf, pdf, hrf and survival function of the class of CGHD. At first and using (2) we give the following lemma.

Lemma 1. *The cdf $F_x(\alpha) = P(X \le x, \alpha)$ of CGHD is given for $x \in (0,\infty)$ as*

$$F_x(\alpha) = \frac{1}{c(\alpha)} \times \int_0^x \frac{(t+r_1-1)^{t+r_1-1}(t+r_2-1)^{t+r_2-1}}{t^t(q+t)^{q+t}}dt \tag{3}$$

where $c(\alpha)$ is the normalization factor as

$$c(\alpha) = \int_0^{\infty} \frac{(t+r_1-1)^{t+r_1-1}(t+r_2-1)^{t+r_2-1}}{t^t(q+t)^{q+t}}dt. \tag{4}$$

Proof. We define the cdf of discrete model generalized hypergeometric distribution (1) as follows for $x \in (0,\infty)$

$$\widehat{F}_x(\alpha) = \widehat{F}_0(\alpha) \times \sum_{n=0}^{[x]} \exp\Big(\sum_{m=1}^{n} \ln \frac{(m+r_1-1)(m+r_2-1)}{m(m+q)}\Big) \tag{5}$$

By dediscretization method and from (5), we define the cdf for continuous analog of the model (1) in the following when $x \in (0,\infty)$

$$\begin{aligned} F_x(\alpha) = & \frac{1}{\int_0^{\infty} \exp\Big(\int_0^t \ln \frac{(u+r_1-1)(u+r_2-1)}{u(u+q)}du\Big)dt} \\ & \times \int_0^x \exp\Big(\int_0^t \ln \frac{(u+r_1-1)(u+r_2-1)}{u(u+q)}du\Big)dt. \end{aligned} \tag{6}$$

Suppose that $f(u) = \ln\Big(\frac{(u+r_1-1)(u+r_2-1)}{u(u+q)}\Big), \quad u \in (0,\infty)$. Then we have

$$\begin{aligned} \int_0^t f(u)du &= t\cdot f(t) - \int_0^t uf(u)du \\ &= t\ln(t+r_1-1) + t\ln(t+r_2-1) - t\ln t - t\ln(q+t) \\ &\quad - \int_0^t \frac{u}{u+r_1-1}du - \int_0^t \frac{u}{u+r_2-1}du + \int_0^t \frac{u}{u}du + \int_0^t \frac{u}{q+u}du \\ &= \ln\Big(\frac{(t+r_1-1)^{t+r_1-1}\ (t+r_2-1)^{t+r_2-1}\ q^q}{t^t\ (q+t)^{q+t}\ (r_1-1)^{r_1-1}\ (r_2-1)^{r_2-1}}\Big) \end{aligned} \tag{7}$$

By substituting (7) into (6) the proof is completed. □

Corollary 1. *From (3) in Lemma 1 and using $f_x(\alpha) = \frac{dF_x(\alpha)}{dx}$, we can obtain the pdf of the model. The pdf of the CGHD is given as follows for $x \in (0,\infty)$*

$$f_x(\alpha) = \frac{1}{c(\alpha)} \times \frac{(r_1+x-1)^{r_1+x-1}(r_2+x-1)^{r_2+x-1}}{x^x(q+x)^{q+x}} \tag{8}$$

where $c(\alpha)$ is the normalization factor as in (4).

Remark 1. *We call (3) as CGHD. Its pdf is written in the form (8).*

Corollary 2. *From (3) and (4) it is possible to propose the asymptotic as $x \longrightarrow \infty$ of the function $F_x(\alpha)$ tail. So, we have when $x \longrightarrow \infty$*

$$\begin{aligned}
1 - F_x(\alpha) &= \tfrac{1}{c(\alpha)} \times \int_x^{\infty} \tfrac{(t+r_1-1)^{t+r_1-1}\ (t+r_2-1)^{t+r_2-1}(1+\frac{r_1-1}{t})^t(1+\frac{r_2-1}{t})^t}{t^t\ (q+t)^{q+t}\ 1^t(1+\frac{q}{t})^t} dt \\
&\approx \tfrac{1}{c(\alpha)} \times \exp(-(q+2-r_1-r_2)) \times \int_x^{\infty} \tfrac{(t+r_1-1)^{t+r_1-1}\ (t+r_2-1)^{t+r_2-1}}{t^t\ (q+t)^{q+t}} dt \\
&\approx \tfrac{1}{c(\alpha)} \times \exp(-(q+2-r_1-r_2)) \times \int_x^{\infty} t^{-(q+2-r_1-r_2)} dt \\
&= \tfrac{1}{c(\alpha)} \times \exp(-(q+2-r_1-r_2)) \times \tfrac{1}{q+2-r_1-r_2} x^{-(q+2-r_1-r_2)}.
\end{aligned} \tag{9}$$

From (9) we conclude that the only tail $1 - F_x(\alpha)$, $x \in \mathbb{R}^+$ of distribution function $F_x(\alpha)$ varies regularly with exponent $(-\rho)$ such that

$$-\rho = -(q+2-r_1-r_2) < -1$$

and it turns out

$$1 - F_x(\alpha) = x^{-(q+2-r_1-r_2)} L(x) \tag{10}$$

of the tail the slowly varying component $L(x)$ holds condition

$$\lim_{x \longrightarrow \infty} L(x) = \frac{1}{c(\alpha)}\ \exp(-(q+2-r_1-r_2))\ \frac{1}{q+2-r_1-r_2}. \tag{11}$$

Now, based on (10) and (11) we see that $1 - F_x(\alpha)$ shows a constant slowly varying component.

Corollary 3. *The survival function of CGHD can be given in a practical form as the following integral*

$$S_x(\alpha) = 1 - F_x(\alpha) = \frac{1}{c(\alpha)} \times \int_x^{\infty} \frac{(t+r_1-1)^{t+r_1-1}(t+r_2-1)^{t+r_2-1}}{t^t(q+t)^{q+t}} dt$$

and hrf as

$$h_x(\alpha) = \frac{f_x(\alpha)}{S_x(\alpha)} = \left(\int_x^\infty \frac{(t+r_1-1)^{t+r_1-1}(t+r_2-1)^{t+r_2-1}}{t^t(q+t)^{q+t}} dt \right)^{-1} \times \frac{(x+r_1-1)^{x+r_1-1}(x+r_2-1)^{x+r_2-1}}{x^x(q+x)^{q+x}}.$$

Remark 2. *From (9) it can also be considered survival function and hrf in more suitable forms approximately as* $x \longrightarrow \infty$.

2.1 Figures

Here, we plot the cdf, hrf and pdf of CGHD for different possible values of parameters r_1, r_2 and q in Figures 1 and 2.

From Figure 1, we see that the hrf is very flexible. From Figure 2 and as we expected, we see that for the different parameter sets the density functions are unimodal. Moreover, the values of ρ corresponding to the selected three parameters are satisfied in the condition $\rho > 1$.

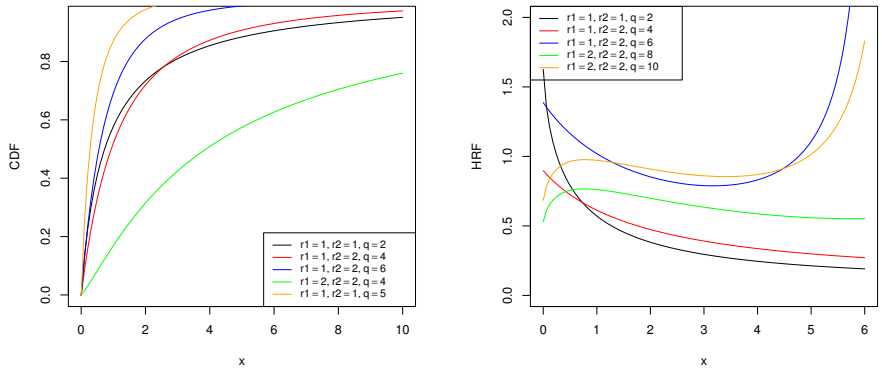

Figure 1: Plots of the cdf and hrf of CGHD for different values of parameters r_1, r_2 and q

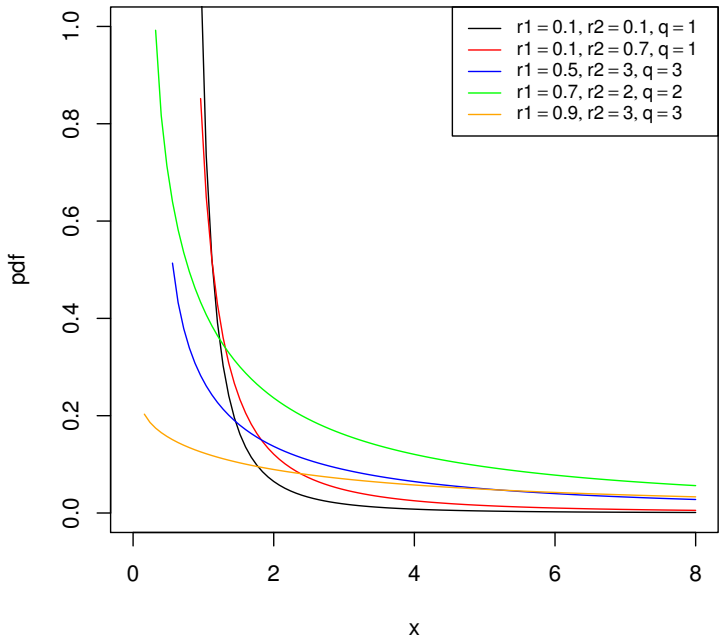

Figure 2: Plots of the pdf of CGHD for different values of parameters r_1, r_2 and q

3 Estimation

In this section, we are going to obtain MLEs of the unknown parameters of the model (8), but it is not possible to have the estimates analytically. Hence, numerical methods and simulation studies shall be used to evaluate the MLEs. To do simulation studies we use Monte Carlo method.

3.1 Simulation study

A simulation study is done to evaluate the performance of the maximum likelihood method. We write a new program in R statistical software by

using the Monte Carlo method. Samples of size $n = 20, 50, 100, 200, 500$ from (8) for selected values of parameters are generated. For each sample, the MLEs are derived and repeated in this process 1000 times. Then the average of biases and the MSEs are computed. The results are proposed in Tables 1 – 3. We consider different possible values for initial values of the unknown parameters and as we see from Tables 1 – 3, the estimated parameters have satisfactory values and by increasing sample sizes the biases and MSEs decrease.

Note 1. To the best of our knowledge, it has not been proposed any closed form for the cdf of the model (8) and hence we can not generate data based on the cdf. To overcome this difficulty, we use Monte Carlo method.

Note 2. In simulation studies, we see that when $q - r_1 - r_2$ is small, the results are better. On the other hand, a small value of $q - r_1 - r_2$ causes smaller biases and MSEs (see Tables 1–3).

Table 1: The biases and MSEs of the MLEs for the simulated data

n	r_1	r_2	q	bias (r_1)	bias (r_2)	bias (q)	mse (r_1)	mse (r_2)	mse (q)
20	0.5	1	1.7	0.0816	-0.2133	-0.0418	0.1946	0.2638	0.1694
50	0.5	1	1.7	0.0617	-0.1773	-0.0525	0.1628	0.2156	0.0892
100	0.5	1	1.7	0.0605	-0.1373	-0.0481	0.1405	0.1747	0.0524
200	0.5	1	1.7	0.0699	-0.0869	-0.0286	0.1214	0.1473	0.0294
500	0.5	1	1.7	0.0841	-0.0590	-0.0142	0.1023	0.1341	0.0144
20	0.5	1	2	0.1867	-0.0722	0.0026	0.3244	0.3352	0.2289
50	0.5	1	2	0.1468	-0.0543	-0.0218	0.2755	0.3038	0.1284
100	0.5	1	2	0.1334	-0.0050	-0.0184	0.2406	0.2670	0.0782
200	0.5	1	2	0.1175	0.0652	-0.0004	0.2055	0.2643	0.0468
500	0.5	1	2	0.1051	0.1114	0.0106	0.1726	0.2796	0.0247
20	0.5	2	2.7	0.3978	-0.7170	-0.0605	0.6400	1.0750	0.3316
50	0.5	2	2.7	0.3420	-0.6764	-0.0853	0.5383	0.9677	0.1898
100	0.5	2	2.7	0.3335	-0.6390	-0.0845	0.4917	0.8527	0.1235
200	0.5	2	2.7	0.3296	-0.5489	-0.0544	0.4314	0.7159	0.0727
500	0.5	2	2.7	0.3186	-0.5103	-0.0403	0.3619	0.6845	0.0424
20	0.5	2	3	0.5008	-0.5840	-0.0266	0.8751	1.0511	0.4114
50	0.5	2	3	0.4159	-0.5443	-0.0661	0.7265	0.9771	0.2464
100	0.5	2	3	0.4020	-0.5188	-0.0699	0.6624	0.8878	0.1644
200	0.5	2	3	0.3827	-0.4104	-0.0376	0.5794	0.7587	0.0992
500	0.5	2	3	0.3428	-0.3470	-0.0252	0.4648	0.7472	0.0588

Table 2: The biases and MSEs of the MLEs for the simulated data

n	r_1	r_2	q	bias (r_1)	bias (r_2)	bias (q)	mse (r_1)	mse (r_2)	mse (q)
20	0.1	0.5	1	0.2354	-0.0562	0.0783	0.1597	0.1410	0.1254
50	0.1	0.5	1	0.1860	-0.0576	0.0425	0.1199	0.1162	0.0642
100	0.1	0.5	1	0.1670	-0.0470	0.0287	0.1078	0.1096	0.0380
200	0.1	0.5	1	0.1241	-0.0184	0.0229	0.0734	0.1048	0.0235
500	0.1	0.5	1	0.0730	0.0203	0.0146	0.0430	0.1108	0.0104
20	0.1	0.5	1.5	0.3896	0.1614	0.1827	0.4036	0.3484	0.2618
50	0.1	0.5	1.5	0.2660	0.1423	0.1079	0.2579	0.3450	0.1410
100	0.1	0.5	1.5	0.2219	0.1243	0.0757	0.2156	0.3247	0.0880
200	0.1	0.5	15	0.1576	0.1170	0.0545	0.1479	0.3016	0.0534
500	0.1	0.5	1.5	0.0859	0.1236	0.0327	0.0793	0.2868	0.0228
20	0.1	0.5	2	0.5727	0.3786	0.2908	0.8170	0.7155	0.4599
50	0.1	0.5	2	0.3897	0.3234	0.1764	0.5184	0.7231	0.2583
100	0.1	0.5	2	0.3083	0.2812	0.1242	0.4024	0.6755	0.1654
200	0.1	0.5	2	0.2209	0.2550	0.0907	0.2914	0.6333	0.1021
500	0.1	0.5	2	0.1308	0.2215	0.0538	0.1843	0.5714	0.0444
20	0.1	0.5	2.5	0.7626	0.6105	0.3949	1.3536	1.2642	0.7103
50	0.1	0.5	2.5	0.5487	0.4938	0.2442	0.9361	1.1953	0.4101
100	0.1	0.5	2.5	0.4461	0.4206	0.1771	0.7487	1.0822	0.2718
200	0.1	0.5	2.5	0.3114	0.4138	0.1327	0.5312	1.1200	0.1741
500	0.1	0.5	2.5	0.1877	0.3587	0.0818	0.3593	1.0368	0.0783

Table 3: The biases and MSEs of the MLEs for the simulated data

n	r_1	r_2	q	bias (r_1)	bias (r_2)	bias (q)	mse (r_1)	mse (r_2)	mse (q)
20	0.5	1	2.5	0.3653	0.1435	0.0699	0.6326	0.5844	0.3632
50	0.5	1	2.5	0.2875	0.1461	0.0249	0.5382	0.5458	0.2126
100	0.5	1	2.5	0.2678	0.1683	0.0186	0.4884	0.5233	0.1407
200	0.5	1	2.5	0.2456	0.2492	0.0373	0.4590	0.5890	0.0878
500	0.5	1	2.5	0.1726	0.3381	0.0408	0.3486	0.7103	0.0494
20	0.5	1	3	0.5544	0.3764	0.1402	1.0747	1.0035	0.5312
50	0.5	1	3	0.4310	0.3421	0.0673	0.8720	0.9255	0.3154
100	0.5	1	3	0.3935	0.3609	0.0520	0.7947	0.9039	0.2188
200	0.5	1	3	0.3537	0.4288	0.0653	0.7411	1.0226	0.1466
500	0.5	1	3	0.2614	0.5301	0.0662	0.5910	1.2683	0.0848
20	0.5	1	3	0.7520	0.5836	0.2087	1.6276	1.4618	0.7155
50	0.5	1	3	0.5867	0.5308	0.1062	1.3279	1.3921	0.4309
100	0.5	1	3	0.5574	0.5102	0.0782	1.2729	1.3445	0.3110
200	0.5	1	3	0.4995	0.6090	0.0957	1.1670	1.5675	0.2158
500	0.5	1	3	0.3683	0.7135	0.0885	0.9139	1.9538	0.1323
20	0.5	1	4	0.9575	0.7622	0.2654	2.2882	2.0179	0.9206
50	0.5	1	4	0.7530	0.7016	0.1398	1.8988	1.9271	0.5653
100	0.5	1	4	0.6897	0.6925	0.1016	1.7075	1.9231	0.4134
200	0.5	1	4	0.6582	0.7512	0.1172	1.6907	2.1209	0.2977
500	0.5	1	4	0.5201	0.8451	0.1088	1.4182	2.5770	0.1914

In Figure 3 we show the pdf of CGHD (8) for true values of parameters and estimated parameters. This Figure confirms the accuracy of the simulation calculations. Meanwhile, simulation observations are satisfied with the variation of the value of regular variation ρ, at which is (negative) exponent of the pdf.

4 Conclusion

In the present paper, by dediscretization method we have introduced a new three-parameter continuous distribution so-called CGHD. Some statistical inferences for this continuous distribution were considered.

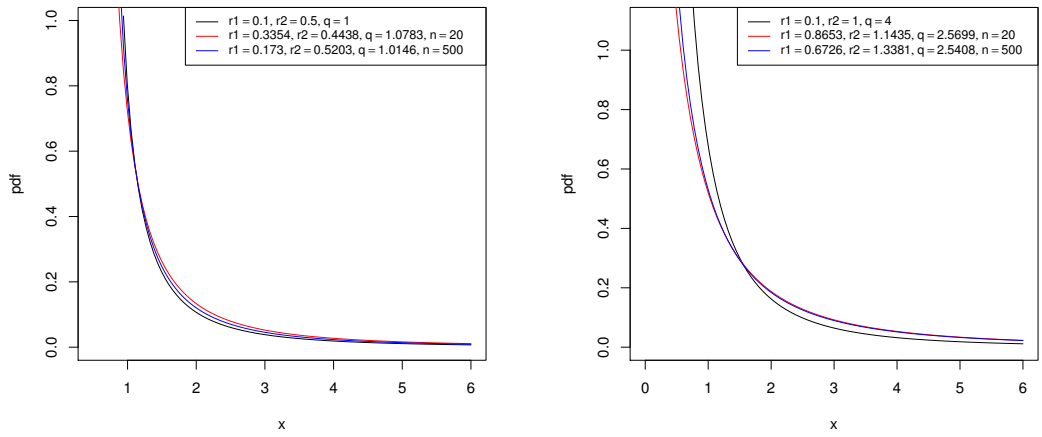

Figure 3: Plots of the pdf of CGHD (8) for true values of parameters (black) and estimated parameters (blue for $n = 20$ and red for $n = 500$)

Using the Monte Carlo method we have done simulation studies to obtain biases and MSEs of the unknown parameters. It was seen that our suggested method works well and the numerical results are acceptable. For the model CGHD, some figures of the pdf, cdf and hrf for different values of parameters have been presented. All computations have been done by R statistical software (version 4.0.3).

Future work. In this paper, using dediscretization method we introduced a new continuous distribution. As future studies, real applications of the proposed model (8) may be considered. Moreover, using the model (8) and discretization method (the reverse to dediscretization procedure introduced by Astola et al. [4, 5]), we are able to construct some new discrete distributions for the needs of biosystems.

Acknowledgement. The authors would like to thank the anonymous referees for their helpful comments.

References

[1] J. Astola and E. Danielian, On regularly varying distributions generated by birth-death processes, *Facta Universitatis: Series Electronics Energetics*, **19**(1) (2006), 109-131.

[2] J. Astola and E. Danielian, Dediscretization of distributions arising in macroevolution models, *Facta Universitatis, Series: Electronics and Energetics*, **20**(2) (2007), 119-146.

[3] J. Astola and E. Danielian, *Frequency Distributions in Biomolecular Systems and Growing Networks*, Tampere International Center for Signal Processing (TICSP), Series no. 31, Tampere, Finland, 2007.

[4] J. Astola, E. Danielian and A. H. Arakelyan, Frequency distributions in growing biomolecular networks based on stable densities, *Reports National Academy of Sciences of Armenia*, **107**(1) (2007), 26-36.

[5] J. Astola, E. Danielian and A. H. Arakelyan, On distance in variation for frequency distributions generated by stable laws, *Reports National Academy of Sciences of Armenia*, **108**(2) (2008), 99-109.

[6] J. Astola, E. Danielian and S. Arzumanyan, Frequency distributions in bioinformatics, a review, *Proceedings of the Yerevan State University: Physical Mathematical Sciences*, **223**(3) (2010), 3-22.

[7] 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 September; In Astola et al. (eds.): TICSP Series 34, Tampere, Finland, pp. 127-132, 2006.

[8] E. Danielian, R. Chitchyan and D. Farbod, On a new regularly varying generalized hypergeometric distribution of the second type, *Mathematical Reports*, **18(68)**(2) (2016), 217-232.

[9] D. Farbod, Some statistical inferences for two frequency distributions arising in bioinformatics, *Applied Mathematics E-Notes*, **14** (2014), 151-160.

[10] D. Farbod, Asymptotic properties of maximum likelihood estimators for a generalized Pareto-type distribution, *Journal of Contemporary Mathematical Analysis*, **50**(1) (2015), 44-51.

[11] D. Farbod, Estimation of parameters for a generalized hypergeometric distribution generated by birth-death process, *Journal of Mathematical Researches*, **6**(3) (2020), 417-428, (in Persian).

[12] D. Farbod and K. Gasparian, On the maximum likelihood estimators for some generalized Pareto-like frequency distribution, *Journal of the Iranian Statistical Society (JIRSS)*, **12**(2) (2013), 211-233.

[13] J. N. Kapur, Application of generalized hypergeometric functions to generalized birth and death processes, *Indian Journal of Pure and Applied Mathematics*, **9** (1978), 1059-1069.

[14] A. W. Kemp, A wide class of discrete distributions and the associated differential equations, *Indian Journal of Statistics, Series A*, **30** (1968), 401-410.

[15] V. A. Kuznetsov, *Statistics of the number of transcripts and protein sequence encoded in the genome*. In: Zhang W, Shmulevich I (eds), Computational and statistical approaches to genomics, 1st ed., 125-171, Springer, Boston, (2002).

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

[17] V. A. Kuznetsov, Hypergeometric stochastic model of evolution of conserved protein coding sequence in the archaeal, bacterial and eukaryotic proteomes, *Fluctuation and Noise Letters*, **3** (2003), 295-324.

[18] V. A. Kuznetsov, Stochastic model of evolution of conserved protein coding sequences, *AIP Conference Proceedings*, American Institute of Physics, **665** (2003), 369-380.

[19] V. A. Kuznetsov, *Mathematical modeling of avidity distribution and estimating general binding properties of transcription factors from*

genome-wide binding profiles, In: Tatarinova,T. V. and Nikolsky, Y. (eds), Biological Networks and Pathway Analysis, pp. 193-276, Springer, New York, 2017.

[20] V. A. Kuznetsov, A. Grageda and D. Farbod, Generalized hypergeometric distributions generated by birth-death process in bioinformatics, *Markov Processes and Related Fields*, (2022), accepted.

[21] 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, *Journal of Biological Systems*, **10**(4) (2002), 381-407.

[22] A. S. Novozhilov, G. P. Karev and E. V. Koonin, Biological applications of the theory of birth-and-death processes, a review, *Briefing in Bioinformatics*, **7**(1) (2006), 70-85.

Davood Farbod
Assistant Professor of Mathematical Statistics
Department of Mathematics
Quchan University of Technology
Quchan 9477177870, Iran
E-mail: d.farbod@qiet.ac.ir and d.farbod@gmail.com

Anis Iranmanesh
Assistant Professor of Statistics
Department of Mathematics and Statistics
Mashhad Branch, Islamic Azad University
Mashhhad, Iran
E-mail: anisiranmanesh@yahoo.com

Maryam Basirat
Assistant Professor of Statistics
Department of Mathematics and Statistics
Mashhad Branch, Islamic Azad University
Mashhhad, Iran
E-mail: basirat_mb@yahoo.com