VOLUME/CİLT 14 NUMBER/SAYI 1 JULY/TEMMUZ 2022

İSTATİSTİK

JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION

TÜRK İSTATİSTİK DERNEĞİ DERGİSİ

ISSN: 1300-4077

SUPPORTED AND PRINTED BY

İZMİR UNIVERSITY OF ECONOMICS

İSTATİSTİK
JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION

ISSN: 1300-4077
https://dergipark.org.tr/tr/pub/ijtsa

SUPPORTED AND PRINTED BY

İZMİR UNIVERSITY OF ECONOMICS

Izmir University of Economics Sakarya Cad. No:156 Balcova - Izmir/TURKEY

Phone: +90 (232) 488-8139 Fax: + 90 (232) 279-2626 www.ieu.edu.tr

AIM AND SCOPE

İSTATİSTİK, Journal of the Turkish Statistical Association is a refereed journal which publishes papers containing original contributions to probability, statistics and the interface of them with other disciplines where prosperity of scientific thought emerges through innovation, vitality and communication on interdisciplinary grounds. The Journal is published four-monthly. The media of Journal is English. The main areas of publication are probability and stochastic processes, theory of statistics, applied statistics, statistical computing and simulation, and interdisciplinary applications in social, demographic, physical, medical, biological, agricultural studies, engineering, computer science, management science, econometrics, etc.

In addition, the journal contains original research reports, authoritative review papers, discussed papers and occasional special issues or relevant conference proceedings.

ABSTRACTED/INDEXED: İSTATİSTİK, Journal of the Turkish Statistical Association is indexed in MathSciNet, Zentralblatt MATH and EBSCO.

SUBMISSION OF MANUSCRIPTS:

Authors should submit their papers electronically by using online submission system (https://dergipark.org.tr/tr/pub/ijtsa).

As part of the submission process, authors are required to check their submission's compliance with all of the following items. Submissions that do not adhere to these guidelines may be returned to authors.

1. The submission has not been previously published, or is being considered for publication in another journal.

2. The submission file is in Portable Document Format (PDF) prepared in LaTeX (TEX) document file format.

3. Where available, URLs for the references have been provided.

4. The text is single-spaced; uses a 12-point font; employs italics, rather than underlining (except with URL addresses); and all illustrations, figures, and tables are placed within the text at the appropriate points, rather than at the end.

5. References should be listed in alphabetical order and should be in the following formats:

Cai, T. and Low, M. (2005). Non-quadratic estimators of a quadratic functional. *The Annals of Statistics*, 33, 2930-2956.

Meyer, Y. (1992). *Wavelets and Operators.* Cambridge University Press, Cambridge.

Cox, D. (1969). Some sampling problems in technology. *In New Developments in Survey Sampling* (N.L. Johnson and H. Smith, Jr., eds.). Wiley, New York, 506-527.

İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION
Vol. 14, No. 1, July 2022, pp. 1–10
ISSN 1300-4077 | 22 | 1 | 1 | 10

İSTATİSTİK

HYPOGEOMETRIC DISTRIBUTION AND RELATED DISCRETE TIME POINT PROCESS

Stefanka Chukova*
School of Mathematics and Statistics,
Victoria University of Wellington,
Wellington, NZ

Leda Minkova and Silvana Paralloi
Faculty of Mathematics and Informatics,
Sofia University,
Sofia, Bulgaria

Abstract: In this paper we propose and study a new distribution, called the hypogeometric distribution, which is a sum of independent geometrically distributed variables with different parameters. Also, we propose and study a discrete time point process based on this distribution. As an example, we focus on a particular form of this process. Also, we show that this type of processes could be used as an appropriate tool to model arrivals with increasing or decreasing time trends. Some possible extensions of this work are also included in the paper.

Key words: Geometric distribution, Hypogeometric distribution, Waiting time, Counting process.

1. Introduction

Most theoretical risk models are formulated for continuous time and results of interest for the particular study are derived. On the other hand, the practical world is discrete, and the continuous time models have to be modified and adjusted to the discrete time scenario. The results for discrete time risk models can provide a good background for better understanding the ideas of the continuous-time scenario and their results can be used as approximations or bounds for the corresponding results in the continuous case, see [3] and [2] for the approximating procedures. The discrete-time risk models have their special features and require specific set of ideas and apparatus to analyze. Also, they are of independent interest since formulas for discrete-time models are mostly recursive and hence suitable for computing the quantities of interest in practice while still reproducing, in limit, the corresponding continuous time results.

It is well-known that if the counting process in the discrete time risk model is the binomial process, the interarrival times are independent, identically distributed geometric random variables, see for example [5]. In this paper we consider a point process, with interarrival times that are independent, geometrically distributed with different parameters random variables. The geometric summands with different parameters are used in [1] and [6] for representing the number of shocks in an engineering system.

The main goal of this paper is to introduce the discrete-time hypogeometric process (HPGP), which has similar structure as the Binomial process, but the interarrival times are not identically distributed. In order to define this process, we firstly introduce the hypogeometric distribution, which is an analogue to the hypoexponential distribution given in [7], and use it to propose and study the HPGP.

In the next Section 2, we introduce the hypogeometric distribution. The corresponding discrete time pure birth process with some properties is introduced in Section 3. In Section 4 we illustrate

* Corresponding author. E-mail address:leda@fmi.uni-sofia.bg

our ideas on an example. The discussion in Section 5 provides some pictorial illustrations of the defined process and discuss some of its properties. Some conluding remarks for this study are given in Section 6.

2. Hypogeometric distribution

Let us consider the sequence

$$\{X_j,\ j=1,2,\ldots\} \tag{2.1}$$

of mutually independent, geometrically distributed random variables with different parameters $\pi_j \in (0,1)$ with $\pi_i \neq \pi_j,\ i \neq j$. The corresponding probability mass function (PMF), cumulative distribution function (CDF) and probability generating function (PGF) of X_j, $j=1,2,\ldots$ in (2.1) are given by

$$P(X_j=n)=(1-\pi_j)\pi_j^{n-1},\ n=1,2,\ldots, \tag{2.2}$$

$$F_{X_j}(n)=P(X_j\leq n)=1-\pi_j^n,\ n=1,2,\ldots, \tag{2.3}$$

and

$$\Psi_{X_j}(s)=\frac{(1-\pi_j)s}{1-\pi_j s},\ j=1,2,\ldots$$

Let $\tau_k=X_1+\ldots+X_k,\ k=1,2,\ldots$ be the sum of k of these random variables. Then, the following lemma holds:

Lemma 1. *The PMF of* $\tau_k,\ k=1,2,\ldots$ *is given by*

$$P(\tau_k=n)=\begin{cases}0, & n<k\\ \sum_{j=1}^{k} w(k,j)P(X_j=n), & n=k,k+1,\ldots,\end{cases} \tag{2.4}$$

where

$$w(k,j)=\Pi_{i=1,i\neq j}^{k}\frac{1-\pi_i}{\pi_j-\pi_i},\ j=1,2,\ldots,k, \tag{2.5}$$

and $w(1,1)=1$.

Proof. According to the definition, the PGF of τ_k is given by

$$\Psi_{\tau_k}(s)=s^k\ \Pi_{j=1}^{k}\frac{1-\pi_j}{1-\pi_j s}=s^k\,\Phi(s),$$

where

$$\Phi(s)=\Pi_{j=1}^{k}\frac{1-\pi_j}{1-\pi_j s}=\frac{\Pi_{j=1}^{k}(\frac{1}{\pi_j}-1)}{\Pi_{j=1}^{k}(\frac{1}{\pi_j}-s)}=\frac{\tilde{H}(s)}{D(s)}$$

is the PGF of some random variable Y with distribution $P(Y=n)=p_n,\ n=0,1,\ldots$. Then, due to the properties of PGF, we have the following representation for τ_k

$$\tau_k=Y+k,$$

i.e., $P(\tau_k=n)=P(Y=n-k)=p_{n-k},\ n\geq k$. Therefore, to find the distribution of τ_k it suffices to find the distribution of Y. We invert $\Phi(s)$ by using the partial-fraction expansion method, given in [4], p.220.

The roots of the denominator $D(s)=\Pi_{j=1}^{k}(\frac{1}{\pi_j}-s)$ are $s_j=\frac{1}{\pi_j},\ j=1,2,\ldots,k$ and its derivative is given by $D'(s)=-\sum_{i=1}^{k}\Pi_{j=1,j\neq i}^{k}(\frac{1}{\pi_j}-s)$. Then

$$D'(s_i)=D'\left(\frac{1}{\pi_i}\right)=-\Pi_{j=1,j\neq i}^{k}\left(\frac{1}{\pi_j}-\frac{1}{\pi_i}\right).$$

Then, for the coefficients f_i, we have

$$f_i = -\frac{\tilde{H}(s_i)}{D'(s_i)} = \frac{\Pi_{j=1}^{k}(\frac{1}{\pi_j}-1)}{\Pi_{j=1,j\neq i}^{k}(\frac{1}{\pi_j}-\frac{1}{\pi_i})},\ i=1,2,\dots k.$$

Hence, according to the inversion method, the PMF of the random variable Y is as follows

$$\begin{aligned} P(Y=n) &= \sum_{i=1}^{k} \frac{\Pi_{j=1}^{k}(\frac{1}{\pi_j}-1)}{\Pi_{j=1,j\neq i}^{k}(\frac{1}{\pi_j}-\frac{1}{\pi_i})}\,\pi_i^{n+1} \\ &= \sum_{i=1}^{k} \frac{\Pi_{j=1}^{k}(1-\pi_j)}{\Pi_{j=1,j\neq i}^{k}(1-\frac{\pi_j}{\pi_i})}\,\pi_i^{n} \\ &= \sum_{i=1}^{k} \Pi_{j=1,j\neq i}^{k}\left(\frac{1-\pi_j}{\pi_i-\pi_j}\right)(1-\pi_i)\pi_i^{n+k-1} \\ &= \sum_{i=1}^{k} w(k,i)P(X_i=n+k),\ n=0,1,\dots \end{aligned} \tag{2.6}$$

Now, using the distribution of Y in (2.6), we find that the PMF of τ_k is given by

$$P(\tau_k=n) = P(Y=n-k) = \begin{cases} 0, & n<k \\ \sum_{j=1}^{k} w(k,j)P(X_j=n), & n=k,k+1,\dots \end{cases}$$

□

Remark 1. Due to (2.4), the following identity is true

$$P(\tau_{k+1}=n) = \sum_{j=1}^{k+1} w(k+1,j)P(X_j=n) = 0,\ n=1,2,\dots,k. \tag{2.7}$$

Remark 2. The following identities are true

$$w(k,j) = \frac{\pi_j-\pi_{k+1}}{1-\pi_{k+1}}\,w(k+1,j),\ j=1,2,\dots k+1. \tag{2.8}$$

and

$$\sum_{j=1}^{k} w(k,j)\pi_j^{k-1} = 1,\ k=1,2,\dots \tag{2.9}$$

The equation (2.9) is equivalent to the fact that $\sum_{n=k}^{\infty} P(\tau_k=n)=1,\ k=1,2,\dots$

Definition 1. The distribution of τ_k, given in (2.4) is called a hypogeometric distribution with parameters $\pi_1,\pi_2,\dots,\pi_k$, $\pi_1\neq\pi_2\neq\dots\neq\pi_k$, and it is denoted by $HPG(\pi_1,\pi_2,\dots,\pi_k)$.

Lemma 2. *The cumulative distribution function of* $\tau_k,\ k=1,2,\dots$ *is given by*

$$P(\tau_k\leq n) = \sum_{j=1}^{k} w(k,j)\pi_j^{k-1}(1-\pi_j^{n-k+1}),\ k\leq n. \tag{2.10}$$

PROOF. According to the definition, we have the following

$$\begin{aligned} P(\tau_k \le n) &= \sum_{i=k}^{n} \sum_{j=1}^{k} w(k,j) P(X_j = i) \\ &= \sum_{j=1}^{k} w(k,j) \sum_{i=k}^{n} P(X_j = i) \\ &= \sum_{j=1}^{k} w(k,j)(1-\pi_j) \sum_{i=k}^{n} \pi_j^{i-1}, \end{aligned}$$

which leads to (2.10). □

3. Point process with $HPG(\pi_1, \pi_2, \ldots, \pi_k)$ distributed k^{th} waiting time

Let us consider a point process, whose waiting times are $HPG(\pi_1, \pi_2, \ldots, \pi_k)$. For this process, we will use the following notation $HPGP(\pi_1, \pi_2, \ldots, \pi_k)$. Then, it is well-known that the expected value and the variance of the j^{th} interevent time are equal to

$$E(X_j) = \frac{1}{1-\pi_j} \quad \text{and} \quad V(X_j) = \frac{\pi_j}{(1-\pi_j)^2}. \tag{3.1}$$

For this process, let us denote by $N(n)$ the number of events up to and including time n, with $N(0) = 0$. Then the following theorem holds:

THEOREM 1. *The PMF of $N(n)$, $n = 1, 2, \ldots$ is given by*

$$P(N(n) = k) = \begin{cases} \pi_1^n, & k = 0, \\ \frac{1}{1-\pi_{k+1}} \sum_{j=1}^{k+1} w(k+1,j) P(X_j = n+1), & k = 1, 2, \ldots, n. \end{cases} \tag{3.2}$$

PROOF. Firstly, let $k = 0$. The events $\{N(n) = 0\} \equiv \{X_1 > n\}$ are equivalent, i.e., $P(N(n) = 0) = P(X_1 > n)$. But $P(X_1 > n) = \pi_1^n$, therefore (3.2) is true for $k = 0$.

Let $k = 1, 2, \ldots, n$. According to the well-known relation

$$P(N(n) = k) = P(\tau_k \le n) - P(\tau_{k+1} \le n),$$

we have that

$$P(N(n) = k) = \sum_{j=1}^{k} w(k,j) \pi_j^{k-1} (1 - \pi_j^{n-k+1}) - \sum_{j=1}^{k+1} w(k+1,j) \pi_j^k (1 - \pi_j^{n-k}).$$

Then, using the identity (2.8), we have

$$\begin{aligned} P(N(n) = k) &= \sum_{j=1}^{k} w(k+1,j) \tfrac{\pi_j - \pi_{k+1}}{1-\pi_{k+1}} (\pi_j^{k-1} - \pi_j^n) - \sum_{j=1}^{k} w(k+1,j) \pi_j^k (1 - \pi_j^{n-k}) \\ &\quad - w(k+1,k+1) \pi_{k+1}^k (1 - \pi_{k+1}^{n-k}) \\ &= \sum_{j=1}^{k} w(k+1,j) \frac{1-\pi_j}{1-\pi_{k+1}} (\pi_j^n - \pi_{k+1} \pi_j^{k-1}) - w(k+1,k+1)(\pi_{k+1}^k - \pi_{k+1}^n) \\ &= \sum_{j=1}^{k+1} w(k+1,j) \frac{1-\pi_j}{1-\pi_{k+1}} (\pi_j^n - \pi_{k+1} \pi_j^{k-1}). \end{aligned}$$

According to the identity (2.7), the second part of this expression is zero, which leads to (3.2). □

REMARK 3. The PMF of $N(n)$ for $k=1,2,\ldots,n$ given in (3.2), has the following equivalent representation:

$$P(N(n)=k)=\Pi_{m=1}^{k}(1-\pi_m)\sum_{j=1}^{k+1}\frac{\pi_j^n}{\Pi_{m=1,m\neq j}^{k+1}(\pi_j-\pi_m)}. \tag{3.3}$$

It follows from the definition of $w(k,j)$ in (2.5).

REMARK 4. Due to (3.2), the mean of $N(n)$ is given by

$$E(N(n))=\sum_{k=1}^{n}\frac{k}{1-\pi_{k+1}}\sum_{j=1}^{k+1}w(k+1,j)P(X_j=n+1),\ n=1,2,\ldots$$

Next, let us assume that the state transition probabilities of a counting process $N^*(n)$, with $N^*(0)=0$, are governed by the following assumptions:

$$P(N^*(1)=k)=\begin{cases}\pi_1, & k=0,\\ 1-\pi_1, & k=1,\\ 0, & k\geq 2,\end{cases} \tag{3.4}$$

and for every $k=0,1,\ldots,$ and $n=1,2,\ldots$

$$P(N^*(n+1)=k+j\mid N^*(n)=k)=\begin{cases}\pi_{k+1}, & j=0,\\ 1-\pi_{k+1}, & j=1,\\ 0, & j\geq 2,\end{cases} \tag{3.5}$$

which defines $N^*(n)$ as a discrete pure birth process. Next, we show that the following theorem holds.

THEOREM 2. *The process $N^*(n)$ defined by the assumptions (3.4) and (3.5) coincides with the counting process $N(n)$ whose interevent times are given by the sequence (2.1) with (2.2).*

PROOF. It suffices to show that the distribution of $N^*(n)$ coincides with the distribution of $N(n)$ given in Theorem 1. We use mathematical induction to prove this coincidence.

For $n=0$, we have $P(N(0)=0)=P(N^*(0)=0)=1$ by definition. For $n=1$, the distribution of $N^*(1)$ is given by (3.4). The distribution of $N(1)$, using (3.2) for $n=1$ and $\{k=0,1\}$, we obtain (3.4).

For $n=2$, the distribution of $N^*(2)$, using the probability rules and the total probability rule we get

- $k=0$
$$P(N^*(2)=0)=P(N^*(2)=0|P(N^*(1)=0)P(N^*(1)=0)=\pi_1\pi_1=\pi_1^2.$$
- $k=1$
$$\begin{aligned}P(N^*(2)=1)=\ & P(N^*(2)=1|P(N^*(1)=0)P(N^*(1)=0)\\ & +P(N^*(2)=1|P(N^*(1)=1)P(N^*(1)=1)\\ & =\pi_1(1-\pi_1)+(1-\pi_1)\pi_2=(1-\pi_1)(\pi_1+\pi_2).\end{aligned}$$
- $k=2$
$$P(N^*(2)=2)=P(N^*(2)=2|P(N^*(1)=1)P(N^*(1)=1)=(1-\pi_1)(1-\pi_2).$$

For the distribution of $N(2)$, using (3.2) for $n=2$ and $\{k=0,1,2\}$, we obtain the same distribution as for $N^*(2)$.

Now, assume that the two distributions coincide for some $n=i>2$, i.e., $P(N(i)=k)=P(N^*(i)=k)$ for $k=0,1,2,\ldots,i$. Next, we show that they coincide for $n=i+1$. The distribution of $N^*(i+1)$ is as follows

- $k=0$

$$P(N^*(i+1)=0)=P(N^*(i+1)=0|P(N^*(i)=0)P(N^*(i)=0)=\pi_1\pi_1^i=\pi_1^{i+1}.$$

- For $k=1,2,\ldots,i+1$, we have

$$\begin{aligned}P(N^*(i+1)=k) &= P(N^*(i+1)=k|P(N^*(i)=k-1)P(N^*(i)=k-1)\\ &\quad +P(N^*(i+1)=k|P(N^*(i)=k)P(N^*(i)=k)\\ &= \Pi_{m=1}^{k-1}(1-\pi_m)\sum_{j=1}^{k}\frac{\pi_j^i}{\Pi_{m=1,m\neq j}^{k}(\pi_j-\pi_m)}(1-\pi_k)\\ &\quad +\Pi_{m=1}^{k}(1-\pi_m)\sum_{j=1}^{k+1}\frac{\pi_j^i}{\Pi_{m=1,m\neq j}^{k+1}(\pi_j-\pi_m)}\pi_{k+1}\\ &= \Pi_{m=1}^{k}(1-\pi_m)\sum_{j=1}^{k}\frac{\pi_j^i}{\Pi_{m=1,m\neq j}^{k}(\pi_j-\pi_m)}\left(\frac{\pi_{k+1}}{\pi_j-\pi_{k+1}}+1\right)\\ &\quad +\frac{\Pi_{m=1}^{k}(1-\pi_m)}{\Pi_{m=1}^{k}(\pi_{k+1}-\pi_m)}\pi_{k+1}^{i+1}\\ &= \Pi_{m=1}^{k}(1-\pi_m)\sum_{j=1}^{k+1}\frac{\pi_j^{i+1}}{\Pi_{m=1,m\neq j}^{k+1}(\pi_j-\pi_m)}.\end{aligned}$$

Using (3.3) for $n=i+1$ and $\{k=1,2,\ldots,i+1\}$, we obtain that the distribution of $N(i+1)$ is the same as the distribution of $N^*(i+1)$.

□

THEOREM 3. *The counting process $N(n)$ satisfies the following recursion formula*

$$P(N(n)=k)=\begin{cases}\pi_1^n, & k=0,\\ \pi_{k+1}P(N(n-1)=k)+(1-\pi_k)P(N(n-1)=k-1), & k=1,2,\ldots,n-1\\ (1-\pi_1)\ldots(1-\pi_n), & k=n.\end{cases}$$

PROOF. The recursion follows from the assumptions (3.4) and (3.5). □

4. An example: $HPGP(\pi_1,\pi_1^a,\ldots,\pi_1^{a^{k-1}})$

According to the definition of the $HPGP(\pi_1,\pi_1^a,\ldots,\pi_1^{a^{k-1}})$, the PMF of X_j, $j=1,2,3,\ldots$ is given by

$$P(X_j=n)=(1-\pi_1^{a^{j-1}})\pi_1^{a^{j-1}(n-1)},\ n=1,2,\ldots \tag{4.1}$$

The mean and the variance of X_j, are given by

$$E(X_j)=\frac{1}{1-\pi_1^{a^{j-1}}}\quad\text{and}\quad V(X_j)=\frac{\pi_1^{a^{j-1}}}{(1-\pi_1^{a^{j-1}})^2}. \tag{4.2}$$

In this case, there are only two parameters that affect the behaviour of the process, that is why we denote it by $HPGP(a,\pi_1)$.

If $a>1$, it is easy to verify that $\pi_1^{a^{j-1}}\to 0$ and then $E(X_j)\to 1$ and $V(X_j)\to 0$, as $j\to\infty$. If $a<1$, $\pi_1^{a^{j-1}}\to 1$ and $E(X_j)\to\infty$ and $V(X_j)\to\infty$. It is always $a\neq 1$, due to the assumption $\pi_i\neq\pi_j$, $i\neq j$ in the initial settings of the process. If $a=1$, the random variables $X_1,X_2,\ldots$ are i.i.d., geometrically distributed as X_1, therefore $HPGP(a,\pi_1)$ is a discrete time renewal process and the corresponding counting process is the binomial process.

4.1. The k^{th} waiting time for $HPGP(a, \pi_1)$

Let τ_k is the waiting time until the occurrence of the k^{th} event in a HPGP(a, π_1), i.e. $\tau_k = X_1 + \ldots + X_k$. Therefore, using Lemma 1, we obtain the following theorem:

THEOREM 4. *The distribution of the waiting time τ_k until the k^{th} event for a $HPGP(a, \pi_1)$ is given by*

$$P(\tau_k = n) = \begin{cases} 0, & n < k \\ \sum_{j=1}^{k} w(k,j) P(X_j = n), & n = k, k+1, \ldots, \end{cases} \tag{4.3}$$

where

$$w(k,j) = \Pi_{l=1, l\neq j}^{k} \frac{1 - \pi_1^{a^{l-1}}}{\pi_1^{a^{j-1}} - \pi_1^{a^{l-1}}}, \ j = 1, 2, \ldots, k.$$

PROOF. It follows from Lemma 1 with $\pi_j = \pi_1^{a^{j-1}}$. □

Also, we get that

$$E(\tau_k) = \sum_{j=1}^{k} \frac{1}{1 - \pi_1^{a^{j-1}}} \quad \text{and} \quad V(\tau_k) = \sum_{j=1}^{k} \frac{\pi_1^{a^{j-1}}}{(1 - \pi_1^{a^{j-1}})^2}.$$

4.2. The $HPGP(a, \pi_1)$ counting process

Denote by $N(n)$ the counting process, representing the number of events in a HPGP(a, π_1) up to and including time $n \geq 0$, i.e., $N(n) = \max\{k, \tau_k \leq n\}$. The state space of $N(n)$ is $\mathcal{N}$, the set of the non-negative integers. Then we have the following result:

THEOREM 5. *The probability mass function of $N(n)$ in given by*

$$P(N(n) = k) = \begin{cases} \pi_1^n, & k = 0, \\ \frac{1}{1 - \pi_1^{a^k}} \sum_{j=1}^{k+1} w(k+1, j) P(X_j = n+1), & k = 1, 2, \ldots, n. \end{cases} \tag{4.4}$$

PROOF. The proof follows directly from Theorem 1 with $\pi_j = \pi_1^{a^{j-1}}$. □

According to Theorem 2, the assumption (3.5) is given by

$$P(N(n+1) = k + j \mid N(n) = k) = \begin{cases} \pi_1^{a^k}, & j = 0, \\ 1 - \pi_1^{a^k}, & j = 1, \\ 0, & j \geq 2, \end{cases} \tag{4.5}$$

for every $k = 0, 1, \ldots,$ and $n = 1, 2, \ldots,$ with the initial distribution given by

$$P(N(1) = k) = \begin{cases} \pi_1, & k = 0, \\ 1 - \pi_1, & k = 1, \\ 0, & k \geq 2, \end{cases} \tag{4.6}$$

which leads to the following equivalent definition of the HPGP:

DEFINITION 2. The counting process $N(n)$, $n = 1, 2, \ldots,$ defined by the assumptions (4.5) and (4.6) is a $HPGP(a, \pi_1)$.

4.3. The expected waiting time

Let us define

$$\tau_{N(n)} = \begin{cases} 0, & N(n) = 0, \\ X_1 + X_2 + \ldots + X_{N(n)}, & N(n) = 1, 2, \ldots, n. \end{cases}$$

THEOREM 6. *The expected waiting time of $\tau_{N(n)}$ is given by*

$$E(\tau_{N(n)}) = \sum_{j=1}^{n} E(X_j) P(N(n) \geq j). \tag{4.7}$$

PROOF. For the mean of $\tau_{N(n)}$ we have

$$\begin{aligned} E(\tau_{N(n)}) &= E[E(\textstyle\sum_{i=1}^{N(n)} X_i | N(n))] \\ &= \textstyle\sum_{k=1}^{n} E(\tau_k) P(N(n) = k) \\ &= \textstyle\sum_{k=1}^{n} \sum_{j=1}^{k} \frac{1}{1-\pi_1^{a^{j-1}}} P(N(n) = k) \\ &= \textstyle\sum_{j=1}^{n} \frac{1}{1-\pi_1^{a^{j-1}}} \sum_{k=j}^{n} P(N(n) = k), \end{aligned} \tag{4.8}$$

and then (4.7).

5. Discussion on $HPGP(\pi_1, \pi_2, \ldots, \pi_k)$

In what follows, we provide some insight on the behaviour of $HPGP(\pi_1, \pi_2, \ldots, \pi_k)$ and its particular version $HPGP(a, \pi_1)$. Recall that the consecutive interarrival times X_j, $j = 1, 2, 3 \ldots$ of the $HPGP(\pi_1, \pi_2, \ldots, \pi_k)$ are geometrically distributed with parameter $(1 - \pi_j)$. Also, see (3.1), we have

$$E(X_j) = \frac{1}{1-\pi_j} \quad \text{and} \quad V(X_j) = \frac{\pi_j}{(1-\pi_j)^2}.$$

Let us consider the following sequence of the consecutive parameters of $HPG(\pi_1, \pi_2, \ldots, \pi_k)$

$$\begin{gathered} \pi_1 < \pi_2 < \ldots < \pi_k, \text{ k} = 2, 3, 4, \ldots, \text{ i.e.,} \\ 1 - \pi_1 > 1 - \pi_2 > \ldots > 1 - \pi_k, \text{ k} = 2, 3, 4, \ldots \end{gathered} \tag{5.1}$$

Then, using formula (2.3) and the definition of usual stochastic order, denoted by "$\prec_{st}$", it is easy to see that

$$P(X_j > n) = \pi_j^n \leq \pi_{j+1}^n = P(X_{j+1} > n), \text{ then } X_j \prec_{st} X_{j+1}.$$

For details on stochastic orderings see [8]. Then, the consecutive interarrival times $\{X_j\}_1^\infty$ of the $HPGP(\pi_1, \pi_2, \ldots, \pi_k)$ form a stochastically increasing sequence. Also, it is easy to see that $E(X_1) < E(X_2) < \ldots < E(X_k) < \ldots$. Therefore, $HPGP(\pi_1, \pi_2, \ldots, \pi_k)$ can be used as a tool to model increasing trends over time.

Similarly, if

$$\begin{gathered} \pi_1 > \pi_2 > \ldots > \pi_k, \text{ k} = 2, 3, 4, \ldots, \text{ i.e.,} \\ 1 - \pi_1 < 1 - \pi_2 < \ldots < 1 - \pi_k, \text{ k} = 2, 3, 4, \ldots, \end{gathered} \tag{5.2}$$

then the consecutive interarrival times of the process form a stochastically decreasing sequence and $E(X_1) > E(X_2) > \ldots > E(X_k) > \ldots$.

Analogously, if $a < 1$,

$$\pi_1 < \pi_1^a < \pi_1^{a^2} < \ldots$$

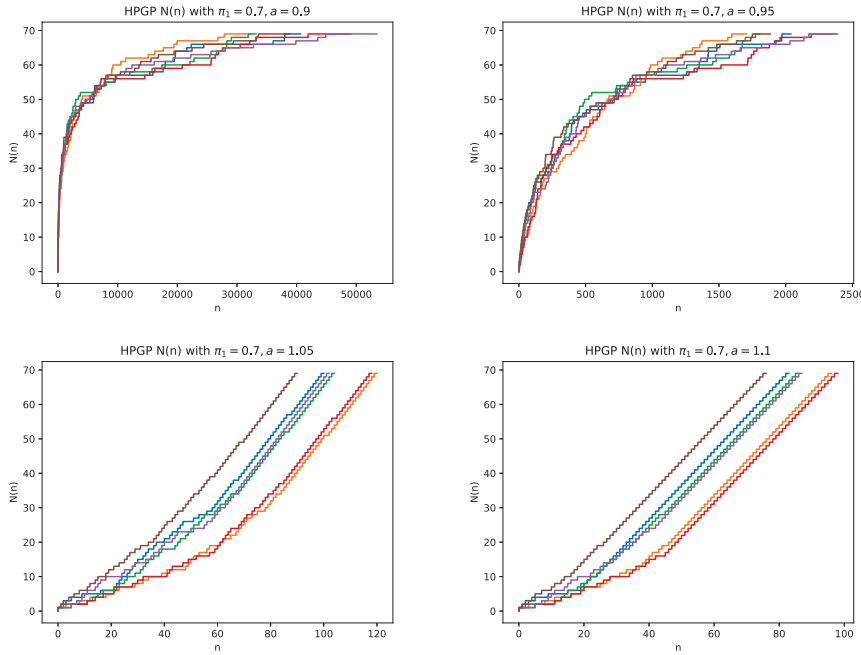


FIGURE 1. $N(n)$ for $HPGP(a, \pi_1)$ with different parameter values

and, using (5.1), we conclude that the $HPGP(a, \pi_1)$ is a stochastically increasing, whereas for $a > 1$ it is stochastically decreasing process, which is illustrated in Figure 1. In each of the plots included in Figure 1, we have shown six sample paths of the counting process of the $HPGP$ with the parameters as given in the plots' label. Comparing the two plots on the first column of Figure 1, we see that it takes much longer (it takes many more discrete time steps) for $HPGP(0.9, 0.7)$ to reach level 70 than for $HPGP(1.05, 0.7)$ to reach the same level. Similar comparison is in place for the second column plots of Figure 1.

Next, we provide some insight on the behaviour of $E(N(n))$ of $HPGP(a, \pi_1)$ depending on the values of its parameters. We use Remark 4 (and also simulation) to compute $E(N(n))$. The plots on Figure 2 agree with our intuition regarding the behaviour of $E(N(n))$. Indeed, $HPGP(0.8, 0.7)$ is a stochastically increasing process, therefore its expected number of events at time 30 should be less than corresponding number of events of $HPGP(1.10, 0.7)$, which forms a decreasing process, at the same time. So, Figure 2 depicts the corresponding $E(N(n))$'s for $HPGP$, with parameters given in the legend, having a relationship as we have expected.

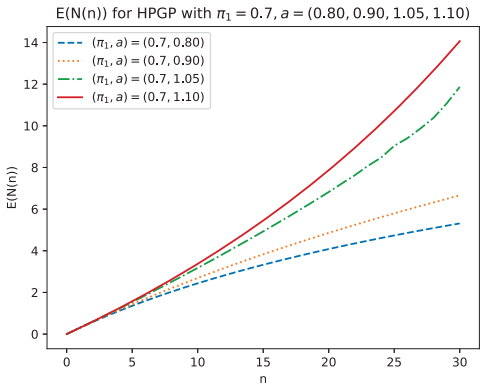


FIGURE 2. $E(N(n))$ for $HPGP(a, \pi_1)$ with different parameter values

6. Conclusion

In this study, analoguesly to the ideas of the hypoexponential distribution in [7], we proposed a new discrete distribution, called hypogeometric distribution, which is a sum of independent geometrically distributed random variables with different parameters. Also, we studied a point process with hypogeometrically distributed waiting times and derived some of its basic properties. An example of this type of process, with a particular hypogeometric distribution for its waiting times, is also included in the paper. In addition, a discussion on some useful properties of these type of processes to model time trends is included.

There are many open research questions related to the newly introduced hypogeometric distribution and related discrete-time point process. For example, questions related to the statistical inference for the distribution/process parameters as well as fitting these to real datasets. Also, how to introduce a compound $HPGP(\pi_1, \pi_2, \ldots, \pi_k)$ and $HPGP(a, \pi_1)$? What are the possible applications of these processes in risk theory?

Acknowledgment The research of the second author was partially supported by Grant DN12/11/20.dec.2017 of the Ministry of Education and Science of Bulgaria.

References

[1] Belzunce, F., Ortega E-M. and Ruiz, J.M. (2009). Aging properties of a discrete-time failure and repair model. *IEEE Transactions on Reliability*, 58, 161-171.

[2] Cossette, H., Landriault, D. and Marceau, É. (2004). Exact expressions and upper bound for ruin probabilities in the compound Markov binomial model. *Insurance: Mathematics and Economics*, 34, 449-466.

[3] Dickson, D.C.M., Egidio dos Reis, A. and Waters, H.R. (1995). Some stable algorithms in ruin theory and their applications. *ASTIN Bulletin*, 25, 153-175.

[4] Kobayashi, H., Mark, B.I. and Turin, W. (2012). *Probability, Random Processes, and Statistical Analysis*. Cambridge University Press, New York.

[5] Li, Sh. and Sendova, K.P. (2013). The finite-time ruin probability under the compound binomial risk model. *European Actuarial Journal*, 3, 249-271.

[6] Ortega, E-M., Alonso, J. and Ortega-Quiles, E. (2016). Comparison of multistate models with discrete-time pure-birth process for recurrent events and uncertain parameters. *Communications in Statistics - Theory and Methods*, 45, 1733-1746.

[7] Ross, S.M. (2007). *Introduction to Probability Models*. Ninth edition, Elsevier Inc., USA.

[8] Shaked, M. and Shanthikumar, J.G. (1994). *Stochastic Orders and Their Applications*. Academic Press, New York.

İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION
Vol. 14, No. 1, July 2022, pp. 11–16
ISSN 1300-4077 | 22 | 1 | 11 | 16

İSTATİSTİK

A COMPOUND POSITIVELY DEPENDENT FARLIE-GUMBEL-MORGENSTERN BIVARIATE COPULA

Selim Orhun Susam*
Department of Economics,
Munzur University,
62000, Tunceli, Turkey

Abstract: In this study, we propose a two parameter Farlie-Gumbel-Morgenstern (FGM) copula that maintains membership in the family in a way while adding the extra dependence parameter to the model by using the compound method. Also, we assess the performance of the new compound FGM copula among all the most used two-parameter families of FGM copulas. The new copula performs best for the real data having moderate dependence structure.

Key words: Copula, Compound disribution, FGM copula.

1. Introduction

Let X and Y be random variables having joint cumulative distribution function H and margins F, G, respectively. Sklar [15] defined copula representation of H as given by $H(x,y) = C(F(x),G(y))$, where C is a unique cumulative distribution function having uniform margins on unit interval. Copula must satisfy the following properties:

DEFINITION 1. A bivariate copula is a function with following properties:

1. C is 2-increasing function for all $x_1 \leq x_2$, $y_1 \leq y_2 \in [0,1]$ such that

$$C(x_2,y_2) - C(x_2,y_1) - C(x_1,y_2) + C(x_1,y_1) \geq 0,$$

2. C is grounded such that $C(x,0) = C(0,y) = 0$ for all $x,y \in [0,1]$,
3. C has uniform margins such that $C(x,1) = x$ and $C(1,y) = y$ for all $x,y \in [0,1]$,
4. $\dfrac{\partial^2 C(u,v)}{\partial u \, \partial v} \geq 0.$

We note that first condition is an equivalent condition for fourth condition in Definition 1 if $C(u,v)$ is twice differentiable. See Lu and Ghosh [10].

Many nonparametric measures of dependence can be viewed as functions of copula C. For bivarate case, Kendall's tau τ and Spearman's rho ρ can be defined, respectively, as

$$\tau = 4 \int_{[0,1]^2} C(u,v) dC(u,v) - 1,$$

$$\rho = 12 \int_{[0,1]^2} uv dC(u,v) - 3.$$

One of the most used bivariate family of the copula, because of its simple form, is the FGM family defined as

$$C(u,v) = uv + \theta u(1-u)v(1-v);\ \theta \in [-1,1],$$

* E-mail address: orhunsusam@munzur.edu.tr

and studied in Farlie [4], Gumbel [5] and Morgenstern [11]. Kendall's tau and Spearman's rho for FGM copula can given as

$$\tau = 2\frac{\theta}{9},\ \rho = \frac{\theta}{3}.$$

It is clear that, FGM copula has the positive dependence struructure for the dependence parameter $\theta \in [0,1]$. Essepecially, in this paper, we mainly interested for the FGM copula having positive dependence structures.

FGM copula have been stuied in different fields, such as finance (Cossette et al. [3]), economics (Patton [14]), and reliability engineering (Navarro et al. [13]; Navarro and Durante [12]). Since the FGM copula has only one parameter, many FGM generated copulas have been introduced with the aim of adding extra parameters to the model. Huang and Kotz [6] proposed the following copula:

$$C_1(u,v) = uv\Big(1+\theta(1-u)(1-v)+\beta uv(1-u)(1-v)\Big);$$

where

$$\theta \in [-1,1],\ \beta \le \frac{3-\theta+\sqrt{9-6\theta-3\theta^2}}{2}.$$

Huang and Kotz [7] proposed the following two copulas:

$$C_2(u,v) = uv\Big(1+\theta(1-u^\beta)(1-v^\beta)\Big),\ \beta>0, -\min(1,\frac{1}{\beta^2}) < \theta < \frac{1}{\beta},$$

$$C_3(u,v) = uv\Big(1+\theta(1-u)^\beta(1-v)^\beta\Big),\ \beta>1, -1<\theta<(\frac{\beta+1}{\beta-1})^{\beta-1}.$$

In addition, Lai and Xie [9], Bairamov and Bairamov [1] gave a generalization of the FGM copula family in their works.

In this study we prefer to use a different approach for enriching the FGM copula inspired by the Kelner et al. [8]. We attempt to replace θ, the generator parameter θ by new two parameters α and β, by using the compound tecnique in order to create new FGM copula while preserving its membership in FGM type copula. The paper is organized as follows: In Section 2, the new compound FGM copula is proposed and also, we investigate the its dependence structure. In Section 3, the performance of the proposed copula is investigated for the real data examples according to its goodness of fit results. Finally, last section is devoted for the conclusion.

2. Compound FGM copula

In this section, we present a tool for generating new compound positively dependent FGM copula having two dependence parameter α and β. This is achieved by using a compound of an existing positively dependent FGM copula cdf with respect to $f_{\alpha,\beta}(\theta)$ which is probability density of the dependence parameter θ,

$$C_T(u,v) = \int_0^1 C_\theta(u,v) f_{\alpha,\beta}(\theta) d\theta, \tag{2.1}$$

where $f_{\alpha,\beta}$ is the probabilty distribution function of beta distribution defined as

$$f_{\alpha,\beta}(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{Beta(\alpha,\beta)};\ \alpha,\beta>0.$$

Then the compuond FGM copula can be derived as following:

$$\begin{aligned} C_T(u,v) &= \int_0^1 \Big(uv+\theta u(1-u)v(1-v)\Big)\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{Beta(\alpha,\beta)}d\theta \\ &= uv + \frac{\alpha}{\alpha+\beta}u(1-u)v(1-v);\ \alpha,\beta>0, \end{aligned}$$

where $Beta$ is the Beta function defined as as $Beta(x_1, x_2) = \int_0^1 t^{x_1-1}(1-t)^{x_2-1}dt$ for x_1, x_2 positive integers. We now show that the new compound FGM copula satisfies the all properties of copula function.

LEMMA 1. *Let $C_T(u,v) = \int_0^1 C_\theta(u,v) f_{\alpha,\beta}(\theta)d\theta$ be a compound FGM copula based on density function $f_{\alpha,\beta}(\theta)$ of $\theta \in [0,1]$. Then for any $C_\theta(u,v)$ the compound FGM copula C_T is also an valid copula function.*

For the proof of Lemma 1, using the fact that for existing FGM copula complies properties defined in Definition 1 we get

$$C_T(0,v) = C_T(u,0) = \int_0^1 C_\theta(0,v) f_{\alpha,\beta}(\theta)d\theta = \int_0^1 C_\theta(u,0) f_{\alpha,\beta}(\theta)d\theta = 0,$$

$$C_T(1,v) = \int_0^1 C_\theta(1,v) f_{\alpha,\beta}(\theta)d\theta = v\int_0^1 f_{\alpha,\beta}(\theta)d\theta = v,$$

$$C_T(u,1) = \int_0^1 C_\theta(u,1) f_{\alpha,\beta}(\theta)d\theta = u\int_0^1 f_{\alpha,\beta}(\theta)d\theta = u.$$

Also we know that $\frac{\delta^2 C_\theta(u,v)}{\delta\, u\, \delta\, v} \geq 0$ then

$$\frac{\partial^2 C_T(u,v)}{\partial\, u\, \partial\, v} = \int_0^1 \frac{\delta^2 C_\theta(u,v)}{\delta\, u\, \delta\, v} f_{\alpha,\beta}(\theta)d\theta \geq 0,$$

then proof is completed.

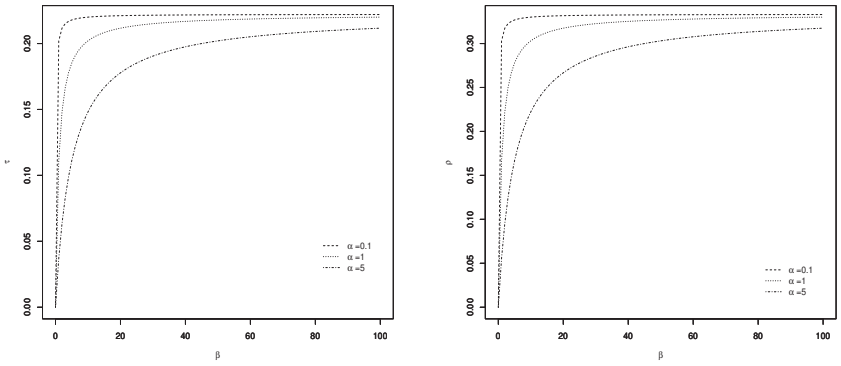
Proposed compound FGM copula has the following Kendall's tau, Spearman's rho, lower tail dependence and upper tail dependence coefficients given by respectively

$$\tau_T = \frac{11\,\alpha + 9\,\beta}{9(\alpha+\beta)} - 1,$$

$$\rho_T = \frac{10\,\alpha + 9\,\beta}{3(\alpha+\beta)} - 3,$$

$$\lambda_{T,L} = 0,\ \lambda_{T,U} = 0.$$

Figure 1 represents the values for the Kendall's tau and Spearman's rho of the compound FGM copula with different fixed values of α and varying values of β. From the this figure, it can be conclude that the compound FGM copula exhibits a varying dependence structures depending on the dependence parameters. Also, we can state that Kendall's tau and Spearman's rho are, respectively, limited to $(0, 0.22)$ and $(0, 0.33)$ as in FGM copula.



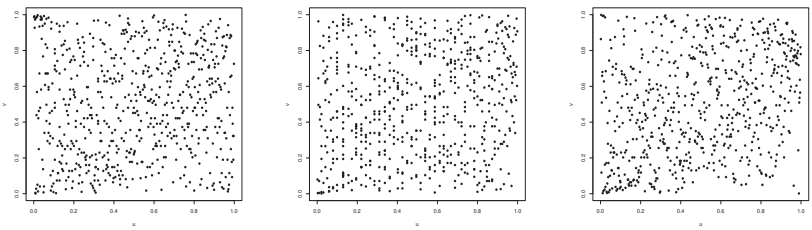
(a) τ values (b) ρ values

FIGURE 1. τ and ρ values for compound FGM copula

3. Case study

In this section, we put the new compound copula into the most used double parameter families of FGM copula as discussed in Section 1. Especially, in this case study, it is aimed to investigating the goodness of fit performance of proposed compund FGM copula under the different dependence structures. We use uranium dataset available in R package "copula". According the this package "These data consist of log concentrations of 7 chemical elements in 655 water samples collected near Grand Junction, CO (from the Montrose quad-rangle of Western Colorado). Concentrations were measured for the following elements: Uranium (U), Lithium (Li), Cobalt (Co), Potassium (K), Cesium (Cs), Scandium (Sc), And Titanium (Ti)." We prefer to modelling the pairs of variables U-Co, Li-Sc and K-Cs.

To avoid decision about marginal distributions, the observations were transformed to pseudo-observation (normalized ranked data) by their corresponding empirical distribution functions. Figure 2(a), 2(b) and 2(c) show the scatter plots of pseudo-observation for the pairs U-Co, Li-Sc and K-Cs, respectively. Looking at the graphs, strong positive dependence structure with $\tau = 0.2074$, mild positive dependence structure with $\tau = 0.0595$ and moderate dependence structure with $\tau = 0.1021$ can be observed for the pairs of K-Cs, U-Co and Li-Sc, respectively.



(a) U-Co ($\tau = 0.0595$) (b) Li-Sc ($\tau = 0.1021$) (c) K-Cs ($\tau = 0.2074$)

FIGURE 2. Scatter plots of real data sets

TABLE 1. Goodness-of-Fit results for real data sets

Copula	Data set	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\beta}$	CvM	P-Val
C_θ	U-Co	0.2681	—	—	0.0612	0.0241
C_1		0.3812	—	-0.4700	0.0565	0.501
C_2		0.7100	—	0.5142	0.0583	0.0487
C_3		0.4488	—	1.3734	0.0521	0.0547
C_T		—	0.0676	0.1920	0.0611	0.0251
C_θ	Li-Sc	0.3762	—	—	0.0618	0.0231
C_1		0.3081	—	-0.1990	0.0308	0.0905
C_2		0.5650	—	0.5873	0.0307	0.0912
C_3		0.3347	—	1.1980	0.0293	0.0847
C_T		—	0.0742	0.2121	0.0218	0.1027
C_θ	K-Cs	0.9333	—	—	0.0397	0.0784
C_1		0.9543	—	-0.3472	0.0340	0.0817
C_2		1.2087	—	0.7888	0.0326	0.0907
C_3		0.4255	—	0.6905	0.0991	0.0034
C_T		—	0.2682	0.0401	0.0367	0.0797

In order to asses goodness of fit we use Cramér-von Mises distance which measure the distance beetwen empirical copula and null hypothesis copula distribution functions are given by

$$CvM = \int_0^1 \int_0^1 n\Big(C_n(u,v) - C_\theta(u,v)\Big)^2 dC_n(u,v), \tag{3.1}$$

where empirical copula is defined by

$$C_n(u,v) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{I}(U_i \le u, V_i \le v).$$

Thus test statistic defined in Eq. (3.1) allows us to compare the distances among copulas (smaller is the better). Also, the parameters of the copulas are estimated by minimizing the Eq. (3.1) under the consideration of constraints (α, $\beta > 0$). Goodness of fit results and estimated parameters for the pairs of K-Cs, U-Co and Li-Sc are shown in Table 1. P-values of the test statistic is computed according to Berg [2]. According to the this table, C_3 is the best performing copula model for the pair U-Co since it possesses the greatest p-value (0.0547) and lowest CvM (0.0521) values. Similarly, from Table 1, the best fit among all possible copulas for the pairs of Li-Sc and K-Cs are C_T (P-val:0.1027) and C_2 (P-val:0.0907), respectively.

For the performance of the new compound FGM copula in real data examples, it has smaller CvM diastance and greates P-Value for the pair Li-Sc which has a modarate dependence coefficient. On the contrary, for the pairs of U-Co and K-Cs which have mild and high dependence respectively, there is no difference with the classical FGM copula in terms of real data performance.

4. Conclusions

We have introduced compound FGM copula, describing its Kendall's tau and Spearman's rho with closed-form. We create new copula using a compound distribution method with a Beta probability density function of its dependence parameter θ. The proposed compound FGM copula make us possible to work with powerful models that can provide a much better goodness-of-fit results for the data set which have moderate dependence structure.

References

[1] Bairamov, I. and Bairamov, K. (2013). From the Huang Kotz FGM distribution to Baker's bivariate distribution. *Journal of Multivariate Analysis*, 113, 106-115.

[2] Berg, D. (2009). Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance*, 15, 675-701.

[3] Cossette, H., Cote, M.P., Marceau, E. and Moutanabbir, K. (2013). Multivariate distribution defned with Farlie-Gumbel-Morgenstern copula and mixed Erlang marginals: Aggregation and capital allocation. *Insurance: Mathematics and Economics*, 52(3), 560-572.

[4] Farlie, D.J.G. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, 47 (3-4), 307-23.

[5] Gumbel, E. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292), 698-707.

[6] Huang, J.S. and Kotz, S. (1984). Correlation structure in iterated Farlie-Gumbel-Morgenstern distributions. *Biometrika*, 71(3), 633-636.

[7] Huang, J.S. and Kotz, S. (1999). Modifications of the Farlie-Gumbel-Morgenstern distributions. A tough hill to climb. *Metrika*, 49(2),135-45.

[8] Kelner, M., Landsman, Z. and Makov, U.E. (2021). Compound Archimedean copulas. *International Journal of Statistics and Probability*, 10(3), 126-126.

[9] Lai, C.D. and Xie, M. (2000). A new family of positive quadrant dependent bivariate distributions. *Statistics & Probability Letters*, 46, 359-634.

[10] Lu, L. and Ghosh, S.K. (2021). Nonparametric estimation and testing for positive quadrant dependent bivariate copula. *Journal of Business and Economic Statistics*, In Press.

[11] Morgenstern, D. (1956). Einfache beispiele zweidimensionaler verteilungen. *Mitteilingsblatt Feur Mathematische Statistik*, 8, 234-235.

[12] Navarro, J. and Durante, F. (2017). Copula-based representations for the reliability of the residual lifetimes of coherent systems with dependent components. *Journal of Multivariate Analysis*, 158, 87-102.

[13] Navarro, J., Ruiz, J.M. and Sandoval, C.J. (2007). Properties of coherent systems with dependent components. *Communications in Statistics-Theory and Methods*, 36, 175-191.

[14] Patton, A. J. (2006). Estimation of multivariate models for time series of possibly diferent lengths. *Journal of Applied Econometrics*, 21(2), 147-173.

[15] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8, 229-231.

İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION
Vol. 14, No. 1, July 2022, pp. 17–26
ISSN 1300-4077 | 22 | 1 | 17 | 26

İSTATİSTİK

APPLICATION OF TIME SERIES ANALYSIS TO CLINICAL DATA (HEART RATE (HR), SYSTOLIC BLOOD PRESSURE (SBP), AND DIASTOLIC BLOOD PRESSURE (DBP))

Kamber Kaşali*
Department of Biostatistics,
Ataturk University,
25240, Erzurum, Turkey

Ahmet Dirican
Department of Biostatistics,
Istanbul University,
34098, Istanbul, Turkey

Abstract: Multivariable analysis methods are frequently used in studies in the field of health carried out through the variables such as heart rate (HR), systolic blood pressure (SBP), and diastolic blood pressure (DBP), etc. In this respect, the basic purpose of this study is to demonstrate that it is more appropriate to analyze the clinical variables that change over time with time series analysis. Data used in the study were obtained from twenty-four-hour rhythm and blood pressure results of holter monitor worn by the patients who have consulted cardiology policlinic with the complaint of blood pressure and heart attack. Heart rate rates (HR), systolic blood pressure (SBP) and diastolic blood pressure (DBP) variables were obtained from the appropriate 250 files. According to the results, there is a causal relationship between HR with SBP and DBP for male and female patients. The p values are 0.0017 and 0.0084 for males and 0.0056 and 0.0001 for females, respectively. This result shows that SBP and DBP can be used to predict HR. According to the results of the time series analysis, it is shown that HR and SBP and DBP variables are correlated but correlations are immediate, and stabilized over time. In our study, it has been shown that applying time series analysis for the time-varying data will give more detailed results.

Key words: Time series, Cointegration, Granger causality.

1. Introduction

Time series is called as a dataset of consecutive observations of an event in a given time period (hours, days, weeks, months, years, etc.). Changes in observations in time series arise from trends, seasonal movements, cyclical movements, and irregular fluctuations [3]. Analysis approaches in time series depends on whether the series is stationary or not. Whether initially non-stationary series act in the same way in later period is examined by cointegration analysis [7]. The time series consist of four components [10]:

a) Trend (T).
b) Seasonal Variations (S).
c) Cyclic Variations (C).
d) Random or Irregular movements (R).

* Corresponding author. E-mail address:kmbrkasali76@gmail.com

A time series can contain one or a few of the above components. Between the actual observation values of the time series Y and the above components

$$Y = T + S + C + R, Y = T * S * C * R, \tag{1.1}$$

there is such a relationship given in Eq. (1.1). Many clinical variables in the field of health are likely to change over time (HR, SBP and DBP, etc.). Use of time series in examining these variables is believed to help reveal important findings of clinical facts [4].

2. Material and methods

2.1. Methods

2.1.1. The stationarity in time series

In time series, if the first mean and variance of the series as well as the high-order moments do not show a change with respect to time, the series is expressed as a stationary time series if it is free from periodic fluctuations in time [9]. The conditions required for any Y_t series to be stationary can be listed as follows:

$$E(Y_t) = \mu \text{ (for all } t\text{'s, Constant average)}, \tag{2.1}$$

$$Var(Y_t) = E(Y_t - \mu) = \sigma^2 \text{ (for all } t\text{'s, Constant variance)}, \tag{2.2}$$

$$\gamma_t = E[(Y_t - \mu)(Y_{t+k}) - \mu \text{ (for all } t \text{ and all } k \neq 0\text{, Constant covariance based on delay distance)}. \tag{2.3}$$

Here k is the lag distance. γ_t is the covariance between two values with k period difference between them. In addition, if the joint and conditional probability distribution process does not change over time, the series is expressed as strongly stationary [13].

2.1.2. Model selection criteria

In time series analyzes, criteria such as R^2 are used to select the most suitable model

$$Y_t = \delta + \phi_1 Y_{t-1} + e_t; t = 1, 2, 3, ..T \tag{2.4}$$

R^2 value in above AR (1) model,

$$R^2 = 1 - \sigma^2/[\sigma^2/(1 - \phi_1^2)] = \phi_1^2. \tag{2.5}$$

The R^2 value depends on ϕ_1 and as the ϕ_1 value gets bigger, the R^2 value will also increase. For this reason, R^2 value in time series is not used much as a selection criterion. There are many selection criteria in time series models. The most commonly used of these are the information criteria put forward by [1] and [15].

$$AIC = ln(ESS/n) + 2k/n, \tag{2.6}$$

$$SIC = ln(ESS/n) + (k.lnn)/n. \tag{2.7}$$

Here n is the number of observations, k is the number of estimated parameters, the ESS is the sum of squares of error terms, and it is expressed as follows:

$$ESS = \sum (Y - Y')^2. \tag{2.8}$$

AIC and SIC criteria are also required to have small values. The delay order with the smallest values is accepted as the most appropriate delay order [7].

2.1.3. The stationarity tests

Many statistical methods are used to determine whether the series is stationary or not. These are generally: Graphical Analysis, Correlogram Analysis and Unit Root Analysis. The Unit Root Analysis: One of the most common methods used to determine stationarity in time series is the "Unit Root" analysis. This analysis is tested with different methods that take into account the breakage that may occur in the series. Dickey-Fuller (DF) Unit Root Test: Dickey and Fuller have revealed whether time series models have a unit [5]. If the following AR (1) process is considered;

$$Y_t = \alpha_1 Y_{t-1} + u_t,$$

in this process, there are 3 different situations for α_1.
1. $|\alpha_1| < 1$ if so, there is a stable root and the series is stationary.
2. $|\alpha_1| = 1$ if so, the series is not stationary, that is, it is unit rooted.
3. $|\alpha_1| > 1$ if so, it is unstable and there is no unit root.

Augmented Dickey-Fuller (ADF) Unit Root Test: Correlation can occur between variables in analysis in time series. In cases with such problem, the Augmented Dickey-Fuller (ADF) test, the extended version of Dickey-Fuller test, is used [8].

Phillips-Perron (PP) Unit Root Test: When the assumptions of DF and ADF tests were not followed, Phillips and Perron [11] assert the Phillips and Perron test. Phillips and Perron is a non-parametric test that predicts correcting error terms. The Phillips and Perron test's model is given below:

$$Y_t = \mu + \phi_1 Y_{t-1} + u_1, \tag{2.9}$$

and

$$(1 - \phi_1 L) Y_t = \mu + u_1. \tag{2.10}$$

Here t= 1,2,..,T and the unit root for this model are calculated with $1/\phi_1$. If $\phi_1 = 1$ is in the model, the serial is unit root.

Cointegration Analysis: In case the time series is not stationary, whether the series act together in the long term is investigated by cointegration analysis. Engle-Granger Cointegration Test: Engle-Granger [12] was the first to mention the cointegration relationship between series. With this method, the long-term balance relationship between two variables is investigated [2].

2.1.4. Vector error correction model (VECM)

If there is cointegration between time series variables, it is more appropriate to make the causality between variables with error correction (VECM) model. The VECM model is used to distinguish between the long-term balance of variables and short-run dynamics between variables. The VECM model is given in (2.11):

$$\Delta X_t = \alpha + \sum_{t=1}^{m} \beta_i \Delta X_{t-i} + \sum_{t=1}^{n} \gamma_i \Delta Y_{t-i} + \sum_{t=1}^{p} \psi_i \Delta Z_{t-i} + \lambda EC_{t-1} + e_t. \tag{2.11}$$

The x in the model is the error correction value that allows the variables to come to equilibrium in the long term.

2.1.5. Vector autoregression (VAR) model

If there is no cointegration between time series variables, it is expressed with the serial vector autoregressive (VAR) model. The VAR model is given below:

$$Y_t = \alpha_1 + \sum_{i=1}^{m} \alpha_2 Y_{t-i} + \sum_{j=1}^{n} \alpha_3 X_{t-j} + e_{1t}. \tag{2.12}$$

$$X_t = \beta_1 + \sum_{i=1}^{p} \beta_2 Y_{t-i} + \sum_{j=1}^{q} \beta_3 X_{t-j} + e_{2t}. \tag{2.13}$$

In the VAR model, there are dependent and independent variables. Sims [14] said that no distinction should be made between intrinsic and extrinsic variables in the VAR model. Sims proposed the VAR model.

2.1.6. Granger causality test

The testability of the causality of two AR model variables was demonstrated by Granger [6]. The applicability of the test depends on whether both variables are stationary and stochastic.

$$Y_t = \alpha + \sum_{i=1}^{r} b_i Y_{t-i} + \sum_{j=1}^{m} c_j X_{t-j} + e_t; t = 1, 2, ..T \tag{2.14}$$

In the model given in Eq. (2.14), α is constant; b_i, Y_t's previous period coefficient; c_j, X_t's previous period coefficient and is an error term with a white noise process [6].

2.2. Data

The data used in the study were collected from 24-hour rhythm and blood pressure results of holter monitors worn by patients who came to the cardiology outpatient clinic of Haseki Training and Research Hospital with the complaint of blood pressure and heart palpitations. 450 folders of patients were analyzed from hospital records and thus the data were gathered from 250 files as 125 men's and 125 women's files. Those with missing measurements in the data of 450 patients were excluded from the study. The results of 125 men and women were averaged and A single 24-hour data set was obtained for males and females. The variables such as heart rate (HR), systolic blood pressure (SBP) and diastolic blood pressure (DBP) were obtained from the files that met the criteria. The definitions for the variables used in the study are summarized below:

1. Heart rate is a fluctuation in the endpoints of the arteries when blood is pumped from the left ventricle to the major arteries,
2. Systolic blood pressure is the pressure in the vein wall of the blood that is excreted from the heart towards the veins when the heart contracts,
3. Diastolic blood pressure is the pressure that is still present in the vessel wall when the heart relaxes.

2.3. Statistical analysis

The results were presented as means, standard deviations, median, minimum and maximum, percentages, and frequencies. The normality distribution of continuous variables was investigated with the Shapiro Wilk test. If there is normality, we used independent samples t-test for two groups comparison. If not, we used Mann Whitney U test for two groups comparison. These analyses were conducted with a statistical analysis program, IBM SPSS 20. The stationarity of HR, SBP and DBP variables were checked using ADF and PP Analyses. Non-stationarity variables were made stationarity by taking their differences. Autocorelation of the model was investigated by the Lagrange Multiplier (LM) test and its heteroskedasticity (varying variance) was checked using White test. When the variables were determined to be an integrated series of the same degree, the cointegration test was conducted using Engle Granger and Johansen methods. Since cointegration was present between the variables, the vector autoregressive model was selected as a candidate model, and thus, the Granger causality test was applied. These analyses were carried out with the EViews 8 statistical analysis program and statistical significance was defined as $p < 0.05$.

3. Results

Findings of this study were summarized below.

Table 1. Descriptive statistics

		MALE (N:125 t:24)			FEMALE (N:125 t:24)		
		HR	SBP	DBP	HR	SBP	DBP
Mean		76.14	125.17	77.1	74.65	132.23	80.45
95% confidence interval	Upper limit	73.36	123.16	75.29	71.7	130.17	78,65
	Lower limit	78.93	127.19	78.91	77.6	134.3	82.25
Standard Deviation		6.59	4.77	4.29	6.98	4.9	4.27
Median		78.53	126.72	78.6	76.77	134.34	81.55
Minimum		65.86	116.15	68.5	63.7	122.87	73.16
Maksimum		85.7	131.7	83.81	83.45	137.66	87.76
Range		19.84	15.55	15.31	19.75	14.79	14.6
IQR (İnterquartile Range)		12.68	8.77	6.85	14.15	9.09	7.65
Skewness		-0.27	-0.69	-0.62	-0.38	-0.64	-0.4
Kurtosis		-1.41	-0.85	-0.76	-1.45	-1.18	-1.09

Table 2. Gender comparison

	Male (N:125 t:24)		Female (N:125 t:24)			
	Mean ± SD	Median ± IQR	Mean ± SD	Median ± IQR	Z	p
HR	76.14 ± 6.59	78.53 ± 12.68	74.65 ± 6.98	76.77 ± 14.15	-0.969	0.332
SBP	125.17 ± 4.77	126.72 ± 8.77	132.23 ± 4.9	134.34 ± 9.09	-3.794	<0.001
DBP	77.1 ± 4.29	78.6 ± 6.85	80.45 ± 4.27	81.55 ± 7.65	-2.608	0.009
Z: Statistical value for Mann Whitney U Test						

As shown in the Table 2, there was not a statistically significant difference between male and female patients in terms of HR (p=0.332), whereas the SBP (p<0.001) and DBP (p<0.001) variables were statistically significant in terms of gender. In this study, the data of men and women were analyzed as different layers in order to examine the trends of the sexes separately.

3.1. Results of time series analysis

The relationship between HR and SBP and DBP obtained from 125 male patients was examined using both Engle-Granger and Johansen cointegration tests. Before applying the cointegration tests, the series should have become stationary when the differences of the same degree are taken. Because the degree of integration of the series must be the same. ADF and PP unit root tests will be applied to the series to show whether this required condition is fulfilled. Therefore, the graphs will be first examined to see the properties of series. The graphics for the series dealt with are shown separately and together below (Figure 1, Figure 2).

When looking carefully at the graphs (Figure 1, Figure 2), it is seen that the series show non-stationary properties and move parallel together. Before moving on to the analysis, our expectation is that the series are unit-rooted at the level and are related in the long run. Unit root tests are required to indicate this condition. The unit root test results of the series are given below (Table 3).

When looking at the graphs of the series (Figure 1, Figure 2), it is understood that fixed and trending equations should be considered. Since there is a 24-hour-time-series, maximum 5 delays are given and the appropriate delay is determined according to Schwarz. The obtained results from ADF and PP tests are as follows:

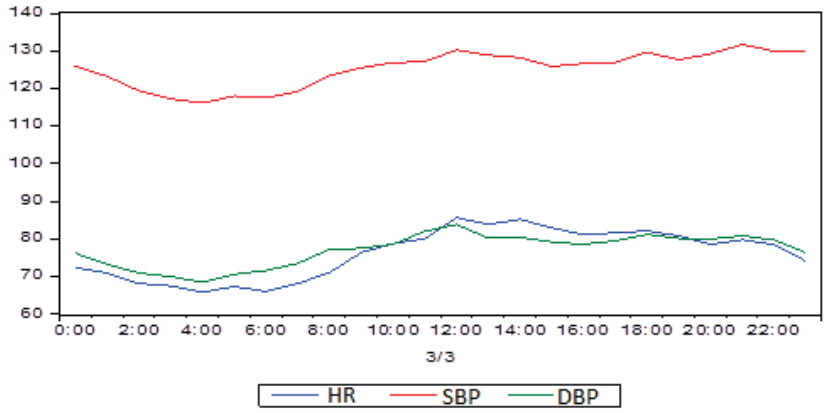


FIGURE 1. HR, SBP and DBP graph of male patients' average.

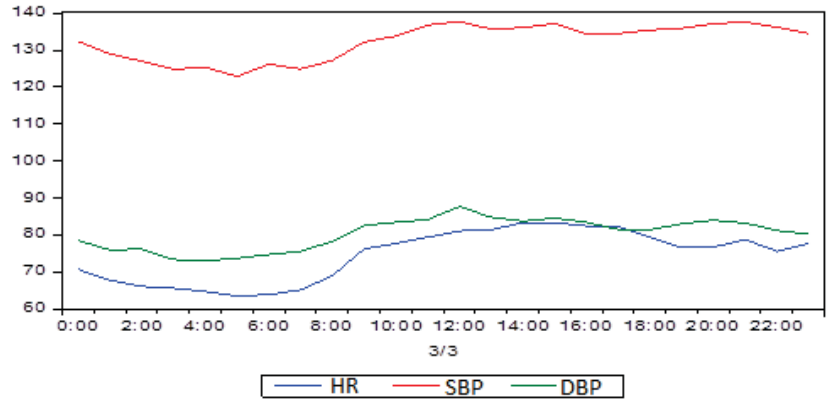


FIGURE 2. HR, SBP and DBP graph of female patients' average.

TABLE 3. ADF and PP tests results

	Male Patients				Female Patients			
	I(0)		I(1)		I(0)		I(1)	
	ADF	PP	ADF	PP	ADF	PP	ADF	PP
HR	0.755	0.677	0.042	0.041	0.794	0.714	0.047	0.047
SBP	0.799	0.702	0.013	0.013	0.746	0.662	0.006	0.006
DBP	0.704	0.61	0.037	0.043	0.706	0.628	0.012	0.013

Looking at the results in the tables, the ADF and PP unit root test results of the HR, SBP and DBP series can be seen. When $p>0.05$, the series are unit-rooted, not stationary. In the original I(0) state of the series, it is seen that they are not stationary ($p>0.05$). For this reason, the test was applied again by taking the first difference I(1) of the series. As the trend effect disappeared for the first difference series, it became stationary ($p<0.05$).

3.2. Examination of autocorrelation between data of patients by LM test

When the LM test results (Table 4) are looked, it is seen that there is autocorrelation since P<0.05 is present even in 1 delayed state. The Newey-West Test was used to eliminate the problem of autocorrelation. The Newey-West test result is shown below:

Table 4. LM tests results

	Male Patients			Female Patients		
SBP, DBP with HR	LM(1)	LM(2)	LM(3)	LM(1)	LM(2)	LM(3)
P	0.001	0.006	0.017	0.009	0.016	0.027

3.3. Heteroskedasticity white correction test between the data of patients

According to the White test result (Table 5), there is no heteroskedasticity problem in our model. The Newey-west test is used to eliminate this problem. The Newey-west Test simultaneously eliminates both the problems of autocorelation and heteroskedasticity. The Newey-west correction version of our model is given below once again The equation obtained from this test will be used to find the cointegration relationship.

Table 5. Heteroskedasticity white correction test results

Heteroskedasticity Test: White					
Male Patients			Female Patients		
Statistical value		p	Statistical value		p
F-statistic	1.521.049	0.2327	F-statistic	1.181.623	0.3566
Obs*R-squared	7.128.456	0.2113	Obs*R-squared	5.930.823	0.3130
Scaled explained SS	4.044.196	0.5431	Scaled explained SS	3.721.332	0.5902

3.4. Newey-west correction test for autocorelation and heteroskedasticity between data of patients

The unit root tests of the residue series obtained from the above model were examined by applying ADF test and PP test (Table 6).

In the Enger-Granger method, the residue series obtained from the regression model is stable with respect to all levels of significance (Table 7). Results were obtained from ADF and PP tests. This result shows that a cointegration relationship exists between HR and SBP and DBP in males.In other words, it shows that these series have acted together in the long term. The series has been found to have no short-term relationship.

Table 6. Heteroskedasticity Newey-west correction test results

Dependent variable: D(HR)									
Male Patients					Female Patients				
Variable	Coefficient	Std. Error	T	p	Variable	Coefficient	Std. Error	t	p
D(SBP)	0.410275	0.33894	1.210.45	0.240	D(SBP)	0.514639	0.19845	2.593.28	0.017
D(DBP)	0.573361	0.31488	1.820.84	0.083	D(DBP)	0.269124	0.29050	0.92639	0.365
C	-0.002745	0.35370	-0.00776	0.993	C	0.235660	0.43854	0.53736	0.596

Table 7. ADF and PP test results

	Male Patients		Female Patients	
	ADF	PP	ADF	PP
Residue series	0,004	0,003	0,016	0,016

Table 8. Granger causality test results

VEC Granger Causality/Block Exogeneity Wald Tests					
Male Patients			Female Patients		
Dependent variable: D(HR)			Dependent variable: D(HR)		
	Chi-Square	P		Chi-Square	p
D(SBP)	12.78838	0.0017	D(SBP)	7.684185	0.0056
D(DBP)	9.555896	0.0084	D(DBP)	16.21758	0.0001
All	14.34794	0.0063	All	16.58381	0.0003
Dependent variable: D(SBP)			Dependent variable: D(SBP)		
	Chi-Square	P		Chi-Square	p
D(HR)	2.656929	0.2649	D(HR)	0.787467	0.3749
D(DBP)	1.194488	0.5503	D(DBP)	4.477577	0.0343
All	4.238053	0.3747	All	9.673283	0.0079
Dependent variable: D(DBP)			Dependent variable: D(DBP)		
	Chi-Square	P		Chi-Square	p
D(HR)	0.684889	0.7100	D(HR)	0.145798	0.7026
D(SBP)	4.131397	0.1267	D(SBP)	1.533346	0.2156
All	5.729465	0.2203	All	1.539706	0.4631

3.5. Granger causality test

According to the results in the Table 8, there is a causality between SBP and DBP with HR for male patients. Because the p values were 0.0017 and 0.0084 respectively. These p values lead to the rejection of the Ho hypothesis for male patients, which states that "the HR series of SBP and DBP series is not the cause of Granger". This result shows us that SBP and DBP are the causes of Granger of HR, that is, they can be used for estimation. Our results show that DBP is not the cause of Granger SBP (p=0.553) and that SBP is not the cause of Granger DBP (p=0.126). Likewise, it is observed that KAH was not seen to be the cause of Granger neither for the SKB nor for the DKB. (P=0.2649; p=0.7100). Consequently, our study shows that male patients have a one-way causality between SBP and DBP with HR. Likewise, the direction of this causality appears to be from SBP and DBP to HR. According to the results in the table, there is a causality between SBP and DBP with HR for female patients since the p values were 0.0056 and 0.0001 respectively. These p values lead to the rejection of the H_0 hypothesis for female patients which states that "the HR series of SBP and DBP series is not the cause of Granger". This result shows us that SBP and DBP are the cause of Granger of HR, that is, they can be used for estimation. Our results show that DBP is the cause of Granger of SBP (p=0.034) and that SBP is not the cause of Granger of DBP (p=0.215). Likewise, it is observed that HR was not seen to be the cause of Granger neither for SBP nor for DBP (P=0.375; p=0.703). As a result, Our study shows that female patients have a one-way causality between SBP and DBP with HR, and the direction of this causality appears to be from SBP and DBP to HR. It also shows that DBP is the cause of Granger of SBP and that the DBP variable can be used in estimating the SBP variable.

4. Discussion

It is important how the time series coheres with each other, or how they relate to each other. Whether these series act together in the short term or the long term is a question of curiosity. How the changing variables change over time provides information about the patients' condition. This information can be extremely important to save lives. While multiple comparisons are being made for time-varying variables in the field of biosatatistics, time series analysis is not used very often. Nevertheless Studies on time-varying data in the medical field have increased recently. Analyses of these data have been tried to be explained by using multiple comparison methods. These methods do not provide information about the change of the series over time. Under these circumstances, it leads misunderstanding in time series. HR, SBP and DBP variables were used in our study. The relationship between variables was examined by using the time series cointegration method. In 2017, Diego Giulliano Destro Christofaro et al. [4] in their study titled "Relationship between Resting heart Rate, Blood Pressure and Pulse Pressure in Adolescents" analysed the relationship between the same variables by using linear regression multiple comparison method. In the study published by Diego Giulliano Destro Christofaro et al., 24-hour data from 716 female and 515 male adolescent patients aged between 14-17 were collected. HR values were calculated (80.1±11.0 beats/minute) for women and (75.9±12.7 beats/minute) for men and were statistically significant ($p<0.001$). In the same study, Resting HR was associated with SBP in males (Beta=0.15 [0.04-0.26]) and female (Beta=0.24 [0.16-0.33]), with DBP in male (Beta=0.50 [0.37-0.64]) and female (Beta=0.41 [0.30-0.53]). Results were calculated and found to be statistically significant. The relationship between variables was revealed in the study, but no information was given about the change of variables over time and about the relationship between short and long term. In our study, when the results of the time series cointegration analysis were checked, the probability values between HR and SBP and DBP were calculated as 0.0017 and 0.0084 in males and 0.0056 and 0.0001 p in females respectively. These p values show us that the Ho hypothesis stating that "SBP and DBP are not the cause of Granger of HR" is rejected. This shows us that SBP and DBP are the causes of granger of HR, that is, they can be used for HR values estimation. In conclusion, both male and female patients have a one-way causality between SBP and DBP with HR. The direction of causality appears to be from SBP and DBP to HR.

5. Conclusions

Looking at the results, it is seen that time series analysis results put forward more detailed results than multiple comparison methods. According to the time series analysis results, it was shown that SBP and DBP with HR variables are related, relations are instantaneous relations, and they come to equilibrium over the long term.

References

[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723.

[2] Bozkurt, H. (2007). *Zaman Serileri Analizi*. Ekin Kitabevi, Bursa.

[3] Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*, Revised Edition, Holden Day, San Francisco.

[4] Christofaro, D.G.D., Casonatto, J., Vanderlei, L.C.M., Cucato, G.G., Dias, R.M.R. (2017). Relationship between resting heart rate, blood pressure and pulse pressure in adolescents. *Arquivos Brasileiros de Cardiologia*, 108(5), 405-410.

[5] Dickey, D.A. and Fuller, W.A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49, (4), 1057-1072.

[6] Granger, C.W.J. (1969). Investigating causal relations by econometrics models and cross spectral methods. *Econometrica*, 37, 3, 424-438.

[7] Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press, Princeton.

[8] Kadilar, C. (2000). *Uygulamalı Çok Değişkenli Zaman Serileri Analizi*. Bizim Büro Basımevi, Ankara.

[9] Kayim, H. (1985). *İstatistiksel Ön Tahmin Yöntemleri*. Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi, Ankara.

[10] Newbold, P., Carlson, W.L., Thorne, B.M. (2000). *Statistics for Business and Economics: Global Edition*. Pearson Education, New York.

[11] Phillips, P.C.B., Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika* , 75(2), 335-346.

[12] Engle, R.F., Granger, C.W.J. (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica* , 55, 251-276.

[13] Sevuktekin, M. and Nargelecekler, M. (2007). *Ekonometrik Zaman Serileri Analizi*. Nobel Akademik Yayıncılık, Ankara.

[14] Sims, C.A. (1980). Macroeconomics and reality. *Econometrica*, 48, 1-48.

[15] Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.

İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION
Vol. 14, No. 1, July 2022, pp. 27–37
ISSN 1300-4077 | 22 | 1 | 27 | 37

İSTATİSTİK

GOMPERTZ-EXPONENTIAL DISTRIBUTION: RECORD VALUE THEORY AND APPLICATIONS IN RELIABILITY

Shakila Bashir* and Ahmad Mahmood Qureshi
Department of Statistics,
Forman Christian College (A Chartered University),
Lahore, Pakistan

Abstract: The continuous probability distributions have wide applications in the field of transportation and reliability engineering. The continuous distributions are used to estimate how funds can be allocated to improve roads, railways, bridges, waterways, airports etc. and used to check the reliability/performance of a product. The Gompertz exponential (GoE) distribution is derived using Gompertz G generator. Some basic properties of the model have been derived. The parameters of the GoE distribution are estimated by maximum likelihood estimation method. The upper record values from the GoE distribution have also been introduced with various properties. Moreover, applications of the GoE distributions has been provided in the field of reliability to check the performance of some transportation related parts and the suggested model provides better fit than the existing well-known models. Finally, a simulation study is carried out. Random numbers of size 50 are generated 15 times for GoE distribution and upper records has been noted.

Key words: Gompertz family of distributions, Exponential, MLE, GoE, Reliability.

1. Introduction

In several real-life applications, the classical distributions do not appropriately work to some real data sets. Thus, researchers introduced many generators by introducing one or more parameters to generate new distributions. The new generated distributions are more flexible as compared to the classical distributions.

Some well-known generators are as follows: Marshall and Olkin [14] generated Marshal-Olkin family, Eugene et al [12] and Jones [13] introduced Beta G, Cordeiro and de Castro [10] developed Kumaraswamy G, Alexander et al [2] presented McDonald G, Zografos and Balakrishnan [22] introduced gamma G type 1, Risti´c and Balakrishnan [19] introduced gamma G type 2, Torabi and Hedesh [21] developed gamma G type 3, Amini et al [6] developed log-gamma G, Cordeiro et al [11] developed exponentiated generalized G, Alzaatreh et al [4] and Alzaghal et al [5] introduced transformed transformer T-X and exponentiated T-X respectively, Bourguignon et al [7] developed Weibull G, Cordeiro et al [9] generated exponentiated half logistic family. Morad et al [15] introduced another generator for continuous distributions called the Gompertz G generator and presented several mathematical properties of it.

In this article the Gompertz family of distribution is considered to develop a new model. Alizadeh et al [3], and Abdal-Hameed, et al [1] used this generator in their work. The cumulative distribution function (cdf) and probability density function (pdf) of the Gompertz family of distributions is

$$F(x) = 1 - e^{\frac{\alpha}{\beta}[1-[1-G(x)]^{-\beta}]}, \quad \alpha > 0, \beta > 0 \tag{1.1}$$

$$f(x) = \alpha g(x)[1-G(x)]^{-\beta-1} e^{\frac{\alpha}{\beta}[1-[1-G(x)]^{-\beta}]}, \quad \alpha > 0, \beta > 0 \tag{1.2}$$

* Corresponding author. E-mail address: shakilabashir@fccollege.edu.pk

where α and β are extra shape parameters and the cdf in Eq. (1.1) and the pdf in Eq. (1.2) was developed using the following transformation:

$$F(x) = \int_0^{-log[1-G(x)]} w(t)dt,$$

$w(t)$ is the probability density function (pdf) of the Gompertz distribution, where t is a random variable. $G(x)$ and $g(x)$ are the cdf and pdf of the baseline distribution. The probability density function (pdf) of the exponential distribution is

$$f(x) = \theta e^{-\theta x}, \theta > 0, x > 0, \tag{1.3}$$

where θ is scale parameter.

The cumulative distribution function (cdf) of the exponential distribution is

$$F(x) = 1 - e^{-\theta x}, \theta > 0, x > 0. \tag{1.4}$$

If an observation is greater than (or less than) of all the values in the experiment, then this value is called a record values. Records values, extreme and lower both have wide application in the fields of studies such as in the science of climates, sports, engineering, medical fields, traffic and transportation, and other industry. Basically, Chandler [8] developed the initials of the record value theory and then later many works has been done it. Further work done by many researchers on almost every continuous probability distribution. The pdf of the sequence of upper record values $[X_{U(n)}, n > 1]$ is

$$f_n(x) = \frac{[R(x)]^{n-1}}{\Gamma(n)} f(x), -\infty < x < \infty, \tag{1.5}$$

where $R(x) = -\ln[1 - F(x)]$.

2. Development of the GoE Distribution

The cumulative distribution function of the GoE distribution is obtained by substituting Eq. (1.4) in Eq. (1.1),

$$F(x) = 1 - e^{\frac{\alpha}{\beta}[1-e^{\beta\theta x}]}; \quad \alpha, \beta, \theta > 0, x > 0, \tag{2.1}$$

$$f(x) = \alpha\theta e^{\beta\theta x} e^{\frac{\alpha}{\beta}[1-e^{\beta\theta x}]}; \quad \alpha, \beta, \theta > 0, x > 0, \tag{2.2}$$

where θ is scale parameter and α,β are shape parameters. The Gompertz Exponential distribution is graphically represented in Figures 1 and 2.
In Figure 1, the proposed pdf is positively skewed for various combinations of parameters.

2.1. Properties of the GoE Distribution

In this section, some GoE distribution properties have been derived. The graph for the reliability measures have also been presented and discussed.

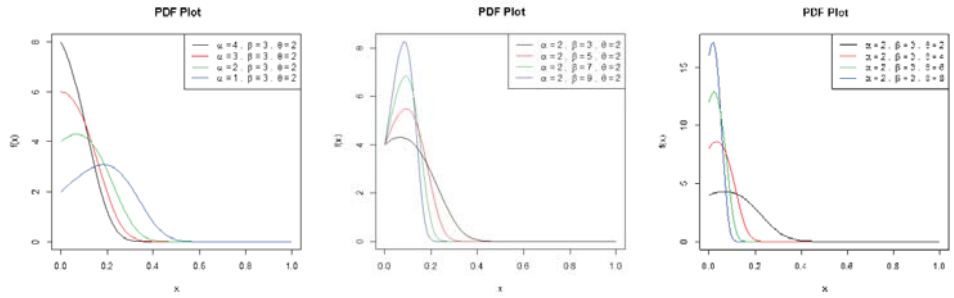
The mean of the GoE distribution is

$$E(X) = \alpha\theta \int_0^{\infty} x e^{\beta\theta x} e^{\frac{\alpha}{\beta}} (1 - e^{\beta\theta x}) dx. \tag{2.3}$$

The variance of the GoE distribution can be calculated by solving the integral in eq (2.3) and in eq (2.4)

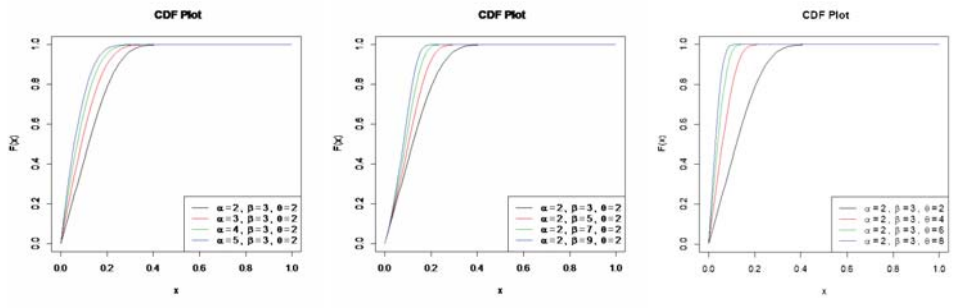
$$E(X^2) = \alpha\theta \int_0^{\infty} x^2 e^{\beta\theta x} e^{\frac{\alpha}{\beta}} (1 - e^{\beta\theta x}) dx. \tag{2.4}$$

The above integrals are unsolvable therefore the numerical values for the mean and variance have been calculated for different values of parameters and presented in Table 1.



(a) (b) (c)

FIGURE 1. Pdf plots for various parametric values



(a) (b) (c)

FIGURE 2. Cdf plots for various parametric values

TABLE 1. Mean and variance for the Gompertz exponential distribution

$\theta=2,\ \beta=3$	α	Mean	Variance	CV
	1	0.192801	0.01232	57.57
	2	0.129332	0.007197	65.59
	3	0.099391	0.004897	70.41
	4	0.081352	0.003604	73.79
	5	0.069123	0.002786	76.36
$\theta=2,\ \alpha=2$	β	Mean	Variance	CV
	4	0.115364	0.005150	62.21
	5	0.104783	0.003903	59.62
	7	0.089546	0.002502	55.86
	9	0.078908	0.001509	49.23
	10	0.074667	0.001509	52.03
$\alpha=2,\ \beta=2$	θ	Mean	Variance	CV
	1	0.258665	0.030330	67.32
	3	0.086222	0.003199	65.60
	4	0.064666	0.001798	65.57
	5	0.051733	0.001152	65.61
	6	0.043111	0.000799	65.57

From Table 1, mean and variance are decreasing as α increasing at fixing β and θ. Similarly mean and variance are decreasing while fixing α and θ with increasing β and same trend is with fixing α and β and increasing θ. While CV is increasing with increasing α, CV is decreasing with increasing β and CV is not varying much with increasing θ.

Reliability function, hazard rate function, reversed hazard rate function and odd function are as follows respectively,

$$R(x) = e^{\frac{\alpha}{\beta}[1-e^{\beta\theta x}],} \tag{2.5}$$

$$h(x) = \alpha\theta e^{\beta\theta x}. \tag{2.6}$$

In Eq. (2.6), for $\beta = 0$ the hazard rate function is constant as given below,

$$h(x) = \alpha\theta. \tag{2.7}$$

The hazard rate function given in Eq. (2.7) is constant due to putting $\beta = 0$, in Eq. (2.6). Therefore, we can say that if the shape parameter β is zero then GoE distribution showing the constant hazard rate.

$$r(x) = \frac{\alpha\theta e^{\beta\theta x} e^{\frac{\alpha}{\beta}[1-e^{\beta\theta x}]}}{1 - e^{\frac{\alpha}{\beta}[1-e^{\beta\theta x}],}} \tag{2.8}$$

$$O(x) = \frac{1 - e^{\frac{\alpha}{\beta}[1-e^{\beta\theta x}]}}{e^{\frac{\alpha}{\beta}[1-e^{\beta\theta x}].}} \tag{2.9}$$

The cumulative hazard function for GoE distribution is

$$H(x) = \frac{\alpha}{\beta}[e^{\beta\theta x} - 1]. \tag{2.10}$$

The Shannon entropy for GoE distribution is

$$S(x) = \frac{\theta(\alpha+\beta)}{\beta} - \ln(\alpha\theta) - \frac{\alpha^2\theta}{\beta^2} - \frac{\alpha e^{\frac{\alpha}{\beta}}}{\beta}\Psi_X(\frac{\alpha}{\beta}), \tag{2.11}$$

where $\Psi_X = \int_1^\infty \ln(x) e^{-\frac{\alpha}{\beta}x} dx$.

The graphs of the reliability function and hazard rate function of the GoE are presented in Figures 3 and 4. From Figure 3, the reliability function is monotonically decreasing with varying parametric values.

From Figure 4, we can see that the hazard rate function of the GoE distribution shows the increasing form of bath tub (IBT) shape or J-shaped. As the time passes on, or aging a person, there is greater chance of death and same concept with the products, as the life time of the product increases there more chances of failure that product.

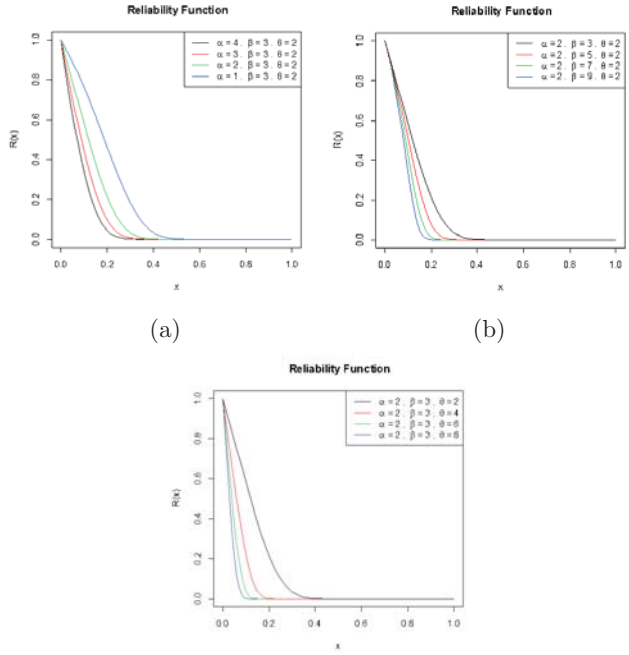


FIGURE 3. Reliability graphs for different parametric values

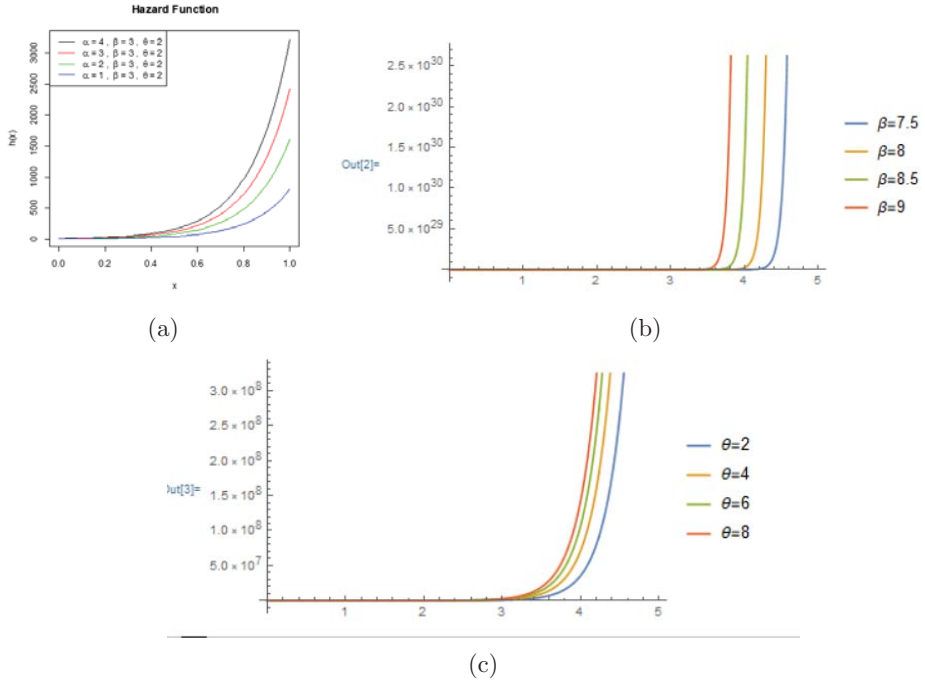


FIGURE 4. Hazard rate graphs for different parametric values

2.2. Quantile Function and Median

The quantile function of the GoE distribution is

$$Q(u) = \frac{ln[1 - \frac{\beta}{\alpha} ln[1-u]]}{\beta\theta}. \tag{2.12}$$

The median of the GoE distribution is

$$median = \frac{ln[1 - \frac{\beta}{\alpha} ln[0.5]]}{\beta\theta}. \tag{2.13}$$

Mode of the GoE distribution

$$mode = \frac{ln[\frac{\beta}{\alpha}]}{\beta\theta}. \tag{2.14}$$

2.3. Order Statistics of the Gompertz Exponential Distribution

The probability density function of the r^{th} order statistics from GoE distribution is

$$f_{r:n}(x) = \frac{n!}{(r-1)!(n-r)!} \alpha\theta e^{\beta\theta x} e^{\frac{(n-r+1)\alpha}{\beta}(1-e^{\beta\theta x})} [1 - e^{\frac{\alpha}{\beta}[1-e^{\beta\theta x}]}]^{r-1}; \alpha, \beta, \theta, x > 0. \tag{2.15}$$

The probability density function of smallest and largest order statistics from GoE distribution is

$$f_{1:n}(x) = n\alpha\theta e^{\beta\theta x} e^{\frac{n\alpha}{\beta}(1-e^{\beta\theta x})}, ; \alpha, \beta, \theta, x > 0, \tag{2.16}$$

$$f_{n:n}(x) = n\alpha\theta e^{\beta\theta x} e^{\frac{\alpha}{\beta}(1-e^{\beta\theta x})} [1 - e^{\frac{\alpha}{\beta}[1-e^{\beta\theta x}]}]^{n-1}; \alpha, \beta, \theta, x > 0. \tag{2.17}$$

3. Record Values from GoE distribution

Using Eq. (2.1) and Eq. (2.2) in Eq. (1.5) the pdf of the upper record values from Gompertz exponential distribution (UR-GoED) is

$$f_n(x) = \frac{\alpha^n \theta}{\beta^{n-1}\Gamma(n)} e^{\beta\theta x} (e^{\beta\theta x} - 1)^{n-1} e^{\frac{\alpha}{\beta}(1-e^{\beta\theta x})}; \alpha, \beta, \theta, x > 0. \tag{3.1}$$

The cumulative distribution function of the UR-GoED is

$$F_n(x) = \frac{1}{\Gamma(n)} \gamma(n, \frac{\acute{x}\alpha}{\beta}). \tag{3.2}$$

The reliability function of the UR-GoED is

$$R_n(x) = \frac{1}{\Gamma(n)}\Gamma(n, \frac{\acute{x}\alpha}{\beta}). \tag{3.3}$$

The hazard rate function of the UR-GoED is

$$h_n(x) = \frac{\alpha^n \theta e^{\beta\theta x}(e^{\beta\theta x}-1)^{n-1} e^{\frac{\alpha}{\beta}(1-e^{\beta\theta x})}}{\beta^{n-1}\Gamma(n, \frac{\acute{x}\alpha}{\beta})}. \tag{3.4}$$

The mean of the UR-GoED is

$$E(X_{U(n)}) = \frac{1}{\theta\beta\Gamma(n)} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}\Gamma(k+n)}{k} \left(\frac{\beta}{\alpha}\right)^k. \tag{3.5}$$

The relation between pdf and cdf of the GoE distribution using Eq. (2.1) and Eq. (2.2), we get

$$f(x) = \alpha\theta e^{\beta\theta x}[1 - F(x)]. \tag{3.6}$$

4. Parameter Estimation

The Maximum Likelihood Estimation (MLE) methodology is used to estimate the parameters of the GoE distribution. Let $x_1, x_2, x_3, ..., x_n$ be the random samples distributed GoE distribution.

$$L(x_1, x_2, ..., x_n; \alpha, \beta, \theta) = \prod_{i=1}^{n} \alpha\theta e^{\beta\theta x_i} e^{\frac{\alpha}{\beta}(1-e^{\beta\theta x_i})},$$

$$\ln L(\alpha, \beta, \theta) = n\log\theta + n\log\alpha + \beta\theta \sum_{i=1}^{n} x_i + \frac{\alpha}{\beta}\sum_{i=1}^{n}(1 - e^{\beta\theta x_i}), \tag{4.1}$$

$$\frac{\partial L(\alpha, \beta, \theta)}{\partial\theta} = \frac{n}{\theta} + \beta\sum_{i=1}^{n} x_i - \alpha\sum_{i=1}^{n} x_i e^{\beta\theta x_i}, \tag{4.2}$$

$$\frac{\partial L(\alpha, \beta, \theta)}{\partial\alpha} = \frac{n}{\alpha} + \frac{1}{\beta}\sum_{i=1}^{n}(1 - e^{\beta\theta x_i}), \tag{4.3}$$

$$\frac{\partial L(\alpha, \beta, \theta)}{\partial\beta} = \theta\sum_{i=1}^{n} x_i - \frac{\alpha\theta}{\beta}\sum_{i=1}^{n} x_i e^{\beta\theta x_i} - \frac{\alpha}{\beta^2}\sum_{i=1}^{n}(1 - e^{\beta\theta x_i}). \tag{4.4}$$

5. Simulations

Random numbers of size 50 from GoE distribution are generated 15 times and the upper record are noted. That are considered the upper record values from GoE distribution. To simulate the random numbers, we use the quantile function given in Eq. (2.12). Simulations are done by using R-package. Then some descriptive measures are calculated from the upper records from GoE distribution.

The below table mentioned that how we can record the upper records (lower records) in real life situations and used them to forecast the results.

TABLE 2. Descriptive measures for UR-GoE distribution when $n = 15, \alpha = 0.05, \beta = 0.25, \theta = 0.3$

Mean	Median	G.M	H.M	Variance	S.D	M.D	C.V
41.4374	41.1112	41.2872	41.1407	12.7347	3.5686	2.9141	8.6120%

6. Model Validation and Application

In this section the proposed distribution GoE is applied on two real life data sets and compared with some well-known models. R software is used for the applications and the criterion used for model selection are AIC, CAIC, BIC, NLL and HQIC.

Firstly, we have used data of failure and service times for a particular windshield taken from Murthy et al. [16] Ramos et al. [18] also used this data. There are 147 observations in the data from which 84 are failed windshields, and 63 are service times of windshields that had not failed at the time of observation.

Secondly, this data is relating to the strengths of 1.5cm glass fibers which taken from Oguntunde et al. [17], and Bourguignon, Silva, and Cordeiro [7], Smith and Naylor [20] were also used this data.

TABLE 3. NLL and goodness of fit criterion for failure times of 84 Aircraft Windshield data

Models	Estimates	NLL	AIC	CAIC	BIC	HQIC
GoED (proposed)	$\hat{\alpha}$=0.0753 $\hat{\beta}$=0.7658 $\hat{\theta}$=1.0372	128.71	263.42	263.72	266.28	266.35
WL	$\hat{a}$=0.0118 $\hat{b}$=0.6462 $\hat{\alpha}$=5.9950 $\hat{\lambda}$=1.3790	127.837	263.675	264.175	273.446	267.605
EL	$\hat{\delta}$=0.0237 $\hat{\alpha}$=6.3074 $\hat{\lambda}$=3.7412	130.549	267.099	267.099	274.427	270.046
KGL	$\hat{a}$=2.9739 $\hat{b}$=19.877 $\hat{\alpha}$=2.2171 $\hat{\lambda}$=12.385	134.888	277.776	278.276	287.546	281.706
GL	$\hat{\alpha}$=5.7000 $\hat{\lambda}$=3.7700 $\hat{\beta}$=1.7e+5 $\hat{\sigma}$=5.1e+4	135.071	278.143	278.643	287.913	282.073
Exponentiated Lomax	$\hat{a}$=3.5417 $\hat{\alpha}$=11653 $\hat{\lambda}$=15521	141.405	288.811	289.107	296.139	291.758
Exponentiated LP	$\hat{\gamma}$=1e-10 $\hat{\sigma}$=3.5473 $\hat{\alpha}$=22063 $\hat{\lambda}$=29382	141.404	290.808	291.308	300.578	294.738
Lomax	$\hat{\alpha}$=30087 $\hat{\lambda}$=76941	164.989	333.978	334.124	338.863	335.943

*WL (Weibull Lomax), EL (exponential Lomax), KGL (Kumaraswamy-Generalized Lomax), GL (Gumbel-Lomax), Exponentiated LP (Lomax Poisson)

Table 4. NLL and goodness of fit criterion for service times of 63 Aircraft Windshield data

Models	Estimates	NLL	AIC	CAIC	BIC	HQIC
GoED (proposed)	$\hat{\alpha}$=0.25528 $\hat{\beta}$=0.59955 $\hat{\theta}$=0.81140	98.27665	202.5533	202.9601	204.8396	205.0820
WL	$\hat{a}$=0.1276 $\hat{b}$=0.9204 $\hat{\alpha}$=3.9136 $\hat{\beta}$=3.0067	98.11712	204.2342	204.9239	212.8068	207.6059
McL	$\hat{a}$=1.3230 $\hat{b}$=53.7712 $\hat{c}$=5.7144 $\hat{\alpha}$=7.4371 $\hat{\beta}$=42.8972	98.5883	207.1766	208.2292	217.8923	211.3911
KwL	$\hat{a}$=1.6691 $\hat{b}$=60.5673 $\hat{\alpha}$=2.5649 $\hat{\beta}$=65.0640	100.8676	209.7353	210.4249	218.3078	213.1069
GL	$\hat{a}$=1.9073 $\hat{\alpha}$=35842.4330 $\hat{\beta}$=39197.5715	102.8332	211.6663	212.0731	218.0958	214.1951
BL	$\hat{a}$=1.9218 $\hat{b}$=31.2594 $\hat{\alpha}$=4.9684 $\hat{\beta}$=169.5719	102.9611	213.9223	214.6119	222.4948	217.2939
EL	$\hat{a}$=1.9145 $\hat{\alpha}$=22971.1536 $\hat{\beta}$=32881.9966	103.5498	213.0995	213.5063	219.5289	215.6282
Lomax Distribution	$\hat{\alpha}$=99269.7800 $\hat{\beta}$=207019.3700	109.2988	222.5976	222.7976	226.8839	224.2834

*WL (Weibull Lomax), McL (McDonald Lomax), KwL (Kumaraswamy Lomax),
GL (Gumbel-Lomax), BL (Beta Lomax), EL (Exponential Lomax)

Table 5. NLL and goodness of fit criterion for the strength of 1.5cm glass fibers

Models	Estimates	NLL	AIC	CAIC	BIC	HQIC
GoED (proposed)	$\hat{\alpha}$=0.007459 $\hat{\beta}$=2.762845 $\hat{\theta}$=1.298306	14.8237	35.64743	36.0542	37.9337	38.1761
GoWei.	$\hat{\alpha}$=0.228488761 $\hat{\beta}$=0.009628097 $\hat{\theta}$=0.794918813 $\hat{\lambda}$=5.612111282	15.18847	38.37694	39.06659	46.94948	41.74856
GL	$\hat{\alpha}$=0.004592168 $\hat{\beta}$=8.179090955 $\hat{\theta}$=0.506999370 $\hat{\lambda}$=1.515829085	14.50274	37.00548	37.69513	45.57802	40.3771

*GoWei (Gompertz Weibull), GL (Gompertz Lomax)

7. Conclusions

The proposed model named Gompertz exponential (GoE) distribution is derived in this article and shape of the model can be seen from Figure 1, it is having longer right tail means positively sewed distribution. Generally, positively skewed distributions are mostly preferred in lifetime data sets. Some basic properties of the GoE model have been derived including some reliability measures. Order statistics and upper record values have also been introduced for the proposed model. Parameters of the GoE distribution are estimated by the method of maximum likelihood estimation. A simulation study is carried out for the GoE model by generating random numbers. From the simulated data the upper records have been noted and some descriptive measures are calculated. The record values are used to learn how we can get record from any known model and used them to get results. Finally, the model is applied on three lifetime data sets (Failure times of 84 Aircraft Windshield, Service times of 63 Aircraft Windshield and strengths of 1.5cm glass fibers) and compared with other models.It can be seen from Table 3, 4 and 5 GoE distribution is better fitted as compared to the other well-known distributions. GoE distribution is more flexible as compared to GoWei, Gompertz Lomax, WL, EL, KGL, GL, Exponentiated LP, Lomax, McL, BL, KwL. The continuous probability distributions have great importance in the field of transportations (they are used to estimate how funds can be allocated for to improve roads, railways, bridges, waterways, airports etc.), reliability engineering (to check the reliability of a product or even to check the reliability of a system, failure chances etc.). The newly derived distribution GoE, is applied here in this article on the data sets mentioned previously and it can be seen not only the newly derived model is providing better approach on these data sets with comparison of other well-known existing models also it is showing application in the theory of reliability and transportation. The record values derived from GoE is a new approach that if we find maximum or minimum records in these fields (reliability, transportation, or others) then the record values from GoE distribution can be the best option to apply.

References

[1] Abdal-Hameed, M. Kh., Khaleel, M.A., Abdullah, Z.M., Oguntunde, P.E., Adejumo, A.O. (2018). Parameter estimation and reliability, hazard functions of Gompertz Burr Type XII distribution. *Tikrit Journal for Administration and Economics Sciences*, 1, 381-400.

[2] Alexander, C., Cordeiro, G.M., Ortega, E.M.M. and Sarabia, J.M. (2012). Generalized beta- generated distributions. *Computational Statistics and Data Analysis*, 56, 1880-1897.

[3] Alizadeh, M., Cordeiro, G.M., Bastos Pinho, L.G. and Ghosh, I. (2017). The Gompertz-g family of distributions. *Journal of Statistical Theory and Practice*, 11(1), 179–207. doi:10.1080/15598608.2016.1267668.

[4] Alzaatreh, A., Lee, C. and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71, 63-79.

[5] Alzaghal, A., Famoye, F. and Lee, C. (2013). Exponentiated T-X family of distributions with some applications. *International Journal of Statistics and Probability*, 2, 1-31.

[6] Amini, M., MirMostafaee, S.M.T.K. and Ahmadi, J. (2012). Log-gamma-generated families of distributions. *Statistics*, 1, 1-20.

[7] Bourguignon, M., Silva, R.B. and Cordeiro, G.M. (2014). The Weibull-G family of probability distributions. *Journal of Data Science*, 12, 53-68.

[8] Chandler, K.N. (1952). The distribution and frequency of record values. *Journal of Royal Statistical Society*, B, 14, 220-228.

[9] Cordeiro, G.M., Alizadeh, M. and Ortega, E.M. (2014). The exponentiated half-logistic family of distributions: Properties and applications. *Journal of Probability and Statistics*, 2014. doi:10.1155/2014/864396.

[10] Cordeiro, G.M. and de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, 81, 883-898.

[11] Cordeiro, G.M., Ortega, E.M.M. and Silva, G.O. (2011). The exponentiated generalized gamma distribution with application to lifetime data. *Journal of Statistical Computation and Simulation*, 81, 827-842.

[12] Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics-Theory and Methods*, 31, 497-512.

[13] Jones, M.C. (2004). Families of distributions arising from distributions of order statistics. *Test*, 13, 1-43.

[14] Marshall, A.W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*, 84, 641-652.

[15] Morad A., Gauss M.C., Luis G.B.P. and Indranil, G. (2016). The Gompertz-G family of distributions. *Journal of Statistical Theory and Practice*, 11(1), 179-207.

[16] Murthy, D.N.P., Xie, M. and Jiang, R. (2004). *Weibull Models*. Wiley.

[17] Oguntunde, P.E., Adejumo, A.O., and Owoloko, E.A. (2017). Application of Kumaraswamy inverse exponential distribution to real lifetime data. *International Journal of Applied Mathematics and Statistics*, 56(5), 34-47.

[18] Ramos, M.W., Marinho, P.R.D., Silva, R.V. and Cordeiro, G.M. (2013). The exponentiated Lomax Poisson distribution with an application to lifetime data. *Advances and Applications in Statistics*, 34, 107-135, 2013.

[19] Risti c, M.M. and Balakrishnan, N. (2012). The gamma-exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, 82, 1191-1206.

[20] Smith, R.I. and Naylor, J.C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Applied Statistics*, 36, 258-369. doi:10.2307/2347795.

[21] Torabi, H. and Hedesh, N.M. (2012). The gamma-uniform distribution and its applications. *Kybernetik*a, 1, 16-30.

[22] Zografos, K. and Balakrishnan, N. (2009). On families of beta- and generalized gamma- generated distributions and associated inference. *Statistical Methodology*, 6, 344-362.

VOLUME/CİLT 14 NUMBER/SAYI 1 JULY/TEMMUZ 2022

İSTATİSTİK

JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION

TÜRK İSTATİSTİK DERNEĞİ DERGİSİ

ISSN: 1300-4077