Review
of the Air Force Academy

© October, 2018

Visa 0574-10/2018

I.S.S.N. 1842-9238

C O N T E N T S

ON MIXING CONTINUOUS DISTRIBUTIONS WITH DISCRETE DISTRIBUTIONS USED IN RELIABILITY

Romică TRANDAFIR[*]**, Vasile PREDA**[**]**, Sorin DEMETRIU**[*]**,
Ion MIERLUŞ-MAZILU**[*]

[*]Technical University of Civil Engineering, Bucharest, Romania
(romica.trandafir@utcb.ro, demetriu@utcb.ro, mmi@utcb.ro), [**]University of Bucharest,
Romania (vasilepreda0@gmail.com)

Abstract: In this paper, new distributions with applications in the reliability of multi-component systems (and not only!) are obtained using the composition of two probability distributions. We consider the composition between: a) truncated binary distribution (Bin(n,p)) with Lindley distribution (Lindley(θ)), b) Kemp distribution (Kemp(α)) with exponential distribution, (Exp(λ)) and c) truncated Zipf distribution (Zipf(α,n)) with exponential distribution (Exp(λ)). Algorithms for numerical simulation of these probability distributions and some comparisons between their performances are presented.

Keywords: truncated binomial discrete distribution, Kemp distribution, Zipf truncated distribution, inverse method, composition method, lifetime variables

1. INTRODUCTION

Lindley (1958, 1965) [6], [7] introduced a new probability distribution that eventually triggered the interest of researchers. Known as the Lindley distribution, this new distribution was used in modelling system reliability [2]. Many researchers, including Faton Merovci [3] and a group of Romanian researchers coordinated by Professor Vasile Preda [8], introduced some generalizations of this distribution by gaining new divisions that proved appropriate in modelling practical situations. It is well known that the exponential distribution has wide applications in reliability. These distributions will be composed of truncated binomial discrete distribution, Kemp and truncated Zipf [4].

The Zipf (α, n) truncated distribution has applications in situations such as: in a statistical population a small number of individuals have a high frequency property, a large number of individuals occasionally have that property and a large number of individuals rarely have that property [12].

In this paper we consider that the lifetimes of the components of a system with n parallel-connected components are random and identically distributed variables (iid) with either Lindley distribution or exponential distribution. The number of independent components is also considered to be a random variable whose distribution is, in turn, Bin (n, p), Kemp (θ), respectively Zipf (α, n) [5], [11].

2. COMPOSITION OF PROBABILITY DISTRIBUTIONS

2.1. The *Bin-Lindley*(θ, *N, P*) distribution

The case when the lifetime variables are distributed *Lindley*(θ), ie they have the probability density function (PDF)

$$\varphi(x;\theta) = \frac{\theta^2}{1+\theta}(1+x)e^{-\theta x} \quad , \quad x > 0, \; \theta > 0 \tag{1}$$

and cumulative distribution function (CDF)

$$\Phi(x:\theta) = 1 - \frac{1+\theta+\theta x}{1+\theta}e^{-\theta x} \quad , \quad x > 0, \; \theta > 0 \tag{2}$$

Consider n random variables $L_1, ..., L_n \sim Lindley(\theta)$ iid and note

$$V = \min_{1 \le i \pi n} L_i \quad si \quad W = \max_{1 \le i \pi n} L_i \tag{3}$$

Then

a) The cumulative distribution function of the minimum random variable V is

$$\Phi_V(x;\theta,n) = P(V < x) = 1 - P(W \ge x) = 1 - (1 - \Phi(x;\theta))^n \quad , \; x > 0 \; , \; \theta > 0 \tag{4}$$

and the density probability function of random variable V is,

$$\varphi_V(x;\theta,n) = n\varphi(x;\theta)(1 - \Phi(x;\theta))^{n-1} \quad , \; x > 0 \; , \; \theta > 0 \tag{5}$$

b) The cumulative distribution function of the maximum random variable W is

$$\Phi_W(x;\theta,n) = P(W < x) = (\Phi(x;\theta))^n \quad , \; x > 0 \; , \; \theta > 0 \tag{6}$$

and the density probability function of random variable W is,

$$\varphi_W(x;\theta,n) = n\varphi(x;\theta)(\Phi(x;\theta))^{n-1} \quad , \; x > 0 \; , \; \theta > 0 \tag{7}$$

If random variable L represents the life of a component of a parallel-connected system with the same operating characteristics, variables V and W are used to determine the reliability of the multi-component system.

Consider now a random variable N whose distribution is truncated binomial [4], denoted $Bin(n, p)$, ie probability function

$$P(N = k) = \frac{C_n^k p^k q^{n-k}}{1 - q^n} \quad , \; k = 1, 2, ..., n \; , \; 0 < p < 1 \; , \; q = 1 - p \tag{8}$$

Suppose that n is a sample of the random variable $N \sim Bin(n, p)$. Then we will have:

i) the probability density function of random variable V

$$f_{V\_B\_L}(x;\theta,n,p) = \sum_{k=1}^{n} P(N = k)\varphi_v(x;\theta,n) = \sum_{k=1}^{n} \frac{C_n^k p^k q^{n-k}}{1 - q^n} k\,\varphi(x;\theta)(1 - \Phi(x;\theta))^{k-1} \tag{9}$$

After evaluation we obtain the expression

$$f_{V\_B\_L}(x;\theta,n,p) = \frac{np}{1-q^n}\varphi(x;\theta)(1 - p\Phi(x;\theta))^{n-1}$$
$$= \frac{np}{1-q^n}\frac{\theta^2}{1+\theta}(1+x)e^{-\theta x}\left(qp\frac{1+\theta+\theta x}{1+\theta}e^{-\theta x}\right)^{n-1} \quad , \; q = 1 - p \tag{10}$$
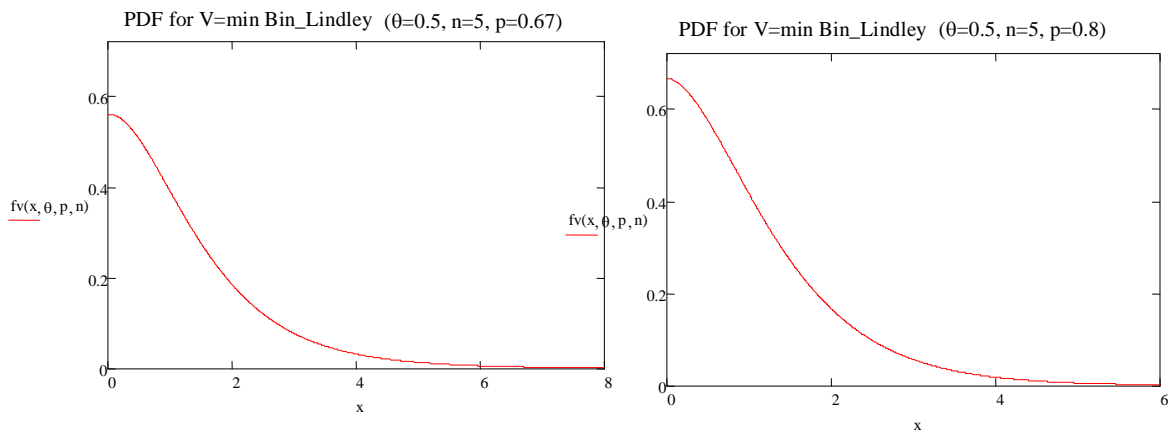
FIG. 1. PDF for random variable V with different parameters
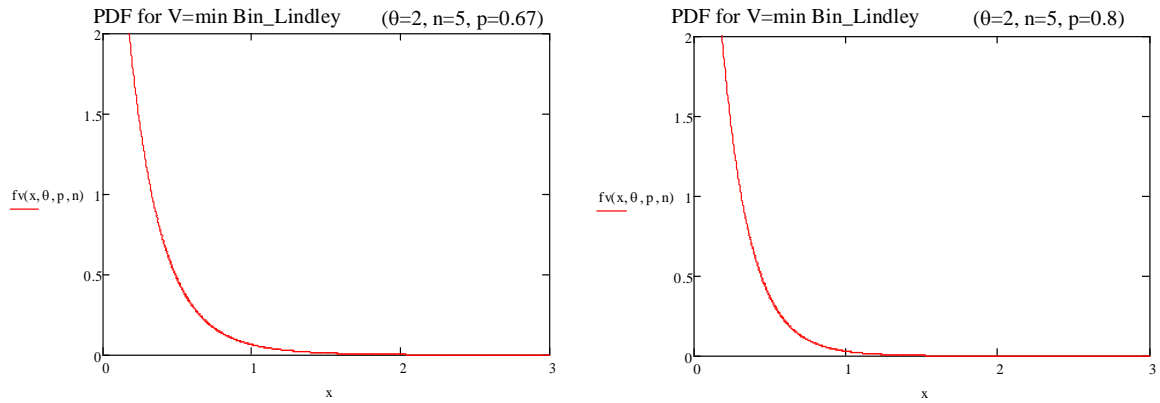
FIG. 2. PDF for random variable V with different parameters

and for the cumulative distribution function

$$F_{V\_B\_L}(x;\theta,n,p) = \int_0^x f_{V\_B\_L}(t;\theta,n,p)dt = \frac{1}{1-q^n}\left\{1 - \left(1 - p\frac{1+\theta+\theta x}{1+\theta}e^{-\theta x}\right)^n\right\} =$$

$$= \frac{1}{1-q^n}\left\{1 - \left(q + p\frac{1+\theta+\theta x}{1+\theta}e^{-\theta x}\right)^n\right\} \ , \ q = 1-p \tag{11}$$
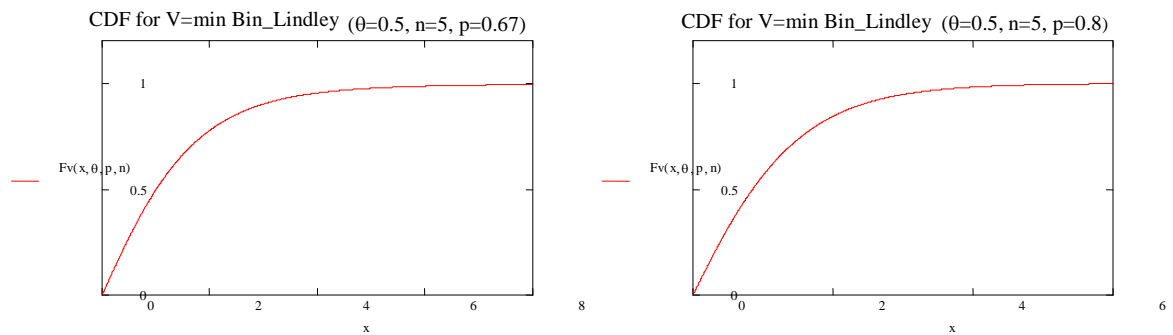
FIG. 3. CDF for random variable V with different parameters

FIG. 4. CDF for random variable V with different parameters

The analogue calculations lead to the following expressions for the probability density function and the distribution function of the random variable *W*

$$f_{W\_B\_L}(x;\theta,n,p) = \frac{np}{1-q^n}\varphi(x;\theta)(q + p\,\Phi(x;\theta))^{k-1} =$$

$$= \frac{np}{1-q^n}\frac{\theta^2}{1+\theta}(1+x)e^{-\theta x}\left(1 - p\frac{1+\theta+\theta x}{1+\theta}e^{-\theta x}\right)^{n-1} \tag{12}$$

7
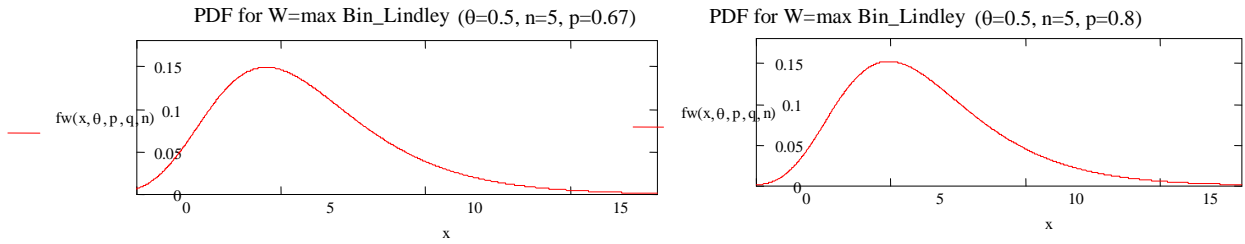
FIG. 5. PDF for random variable W with different parameters
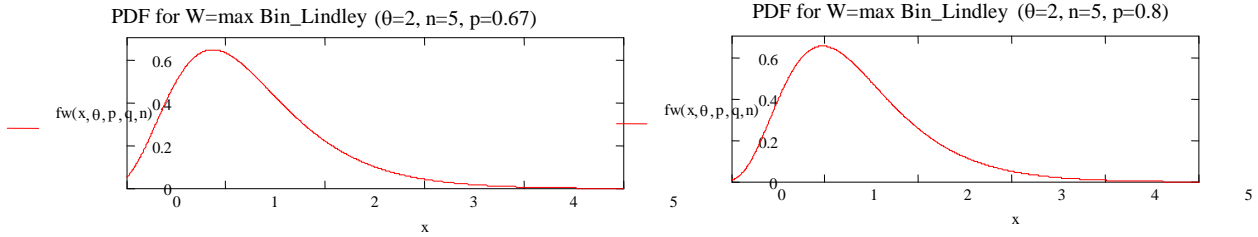
FIG. 6. PDF for random variable W with different parameters

Respectively

$$F_{W\_B\_L}(x;\theta,n,p) = \int_0^x f_{W\_B\_L}(t;\theta,n,p)\,dt = \frac{1}{1-q^n}\left\{\left(p\,\Phi(x;\theta)+q\right)^n - q^n\right\} =$$

$$= \frac{1}{1-q^n}\left\{\left(1 - p\,\frac{1+\theta+\theta x}{1+\theta}e^{-\theta x}\right)^n - q^n\right\}$$

$$(13)$$
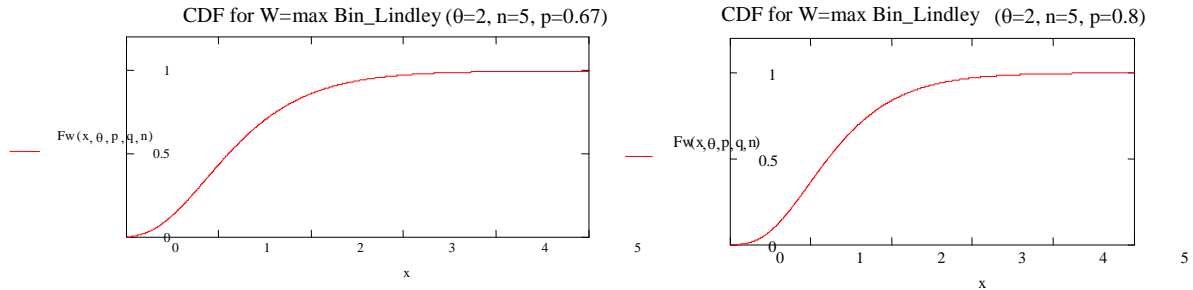
FIG. 7. CDF for random variable W with different parameters

2.2. Kemp- exponential distribution *Kemp_Exp*(θ, λ)

Consider the n variables $L_1,...,L_n \sim Exp(\lambda)$ iid having the exponential probability density function

$$\varphi(x;\lambda) = \lambda e^{-\lambda x} \ , \ \ x > 0 \ , \ \ \lambda > 0 \tag{14}$$

and the cumulative distribution function

$$\Phi(x;\lambda) = 1 - e^{-\lambda x} \ , \ \ x > 0 \ , \ \ \lambda > 0 \tag{15}$$

and N is a random variable having the *Kemp*(θ) distribution, i.e.

$$P(N=k) = -\frac{\theta^k}{k\ln(1-\theta)}, \ \ k = 1,2,..., \ \ 0 < \theta < 1 \tag{16}$$

With the notations above, relations (3), (4) - (7) become

a) The cumulative distribution function of the random variable *V* is

$$\Phi_V(x;\lambda) = P(V < x) = 1 - P(W \geq x) = 1 - (1 - \Phi(x;\lambda))^n =$$
$$= 1 - e^{-\lambda n x}, \; x > 0, \; \theta > 0 \tag{17}$$

and the probability density function of the random variable *V*,

$$\varphi_V(x;\lambda) = n\lambda e^{-\lambda x}(\lambda e^{-\lambda x})^{n-1} = n\lambda e^{-\lambda n x}, \; x > 0, \; \theta > 0 \tag{18}$$

b) The cumulative distribution function of the maximum random variable *W* is

$$\Phi_W(x;\lambda) = P(W < x) = (\Phi(x;\lambda))^n = (1 - e^{-\lambda x})^n, \; x > 0, \; \lambda > 0 \tag{19}$$

and the probability density function of the random variable *W*,

$$\varphi_W(x;\lambda) = n\lambda e^{-\lambda x}(1 - e^{-\lambda x})^{n-1}, \; x > 0, \; \lambda > 0 \tag{20}$$

Suppose now that *n* is a sample of the random variable $N \sim> Kemp(\theta)$. Let us consider the composition of the Kemp distribution with the distributions of the variables *V* and *W*. Then we will have:

i) The probability density function of random variable *V*

$$f_{V\_K\_E}(x;\lambda,\theta) = \sum_{k=1}^{\infty}\left(-\frac{\theta^k}{k\log(1-\theta)}\right)k\,\lambda\,e^{-\lambda a}(\lambda\,e^{-\lambda x})^{k-1} =$$
$$= -\frac{1}{\ln(1-\theta)}\sum_{k=1}^{\infty}\lambda e^{-\lambda x}\left(\lambda e^{-\lambda x}\right)^{k-1}\theta^k = \tag{21}$$

$$= -\frac{1}{\ln(1-\theta)}\sum_{k=1}^{\infty}\left(\theta\lambda\,e^{-\lambda x}\right)^k = -\frac{\theta\lambda}{\ln(1-\theta)}\left(\frac{e^{-\lambda x}}{1-\theta e^{-\lambda x}}\right), \; for \; (\theta e^{-\lambda x}) < 1$$

condition fulfilled because $0 < \theta < 1$, $\lambda > 0$, $x > 0$, $\theta < 1 < e^{\lambda x}$ and for the cumulative distribution function

$$F_{V\_K\_E}(x;\lambda,\theta) = \int_0^x f_{V\_K\_E}(t;\lambda,\theta)dt = \int_0^x\left(-\frac{\lambda\theta e^{-\lambda t}}{\ln(1-\theta)}\frac{1}{1-\theta e^{-\lambda t}}\right)dt =$$
$$= \left[-\frac{1}{\ln(1-\theta)}\ln\left(1-\theta e^{-\lambda t}\right)\right]\Big|_0^x = -\frac{1}{\ln(1-\theta)}\left[\ln\left(1-\theta e^{-\lambda x}\right) - \ln\left(1-\theta e^{-\lambda 0}\right)\right] = \tag{22}$$
$$= 1 - \frac{\ln\left(1-\theta e^{-\lambda x}\right)}{\ln(1-\theta)}$$
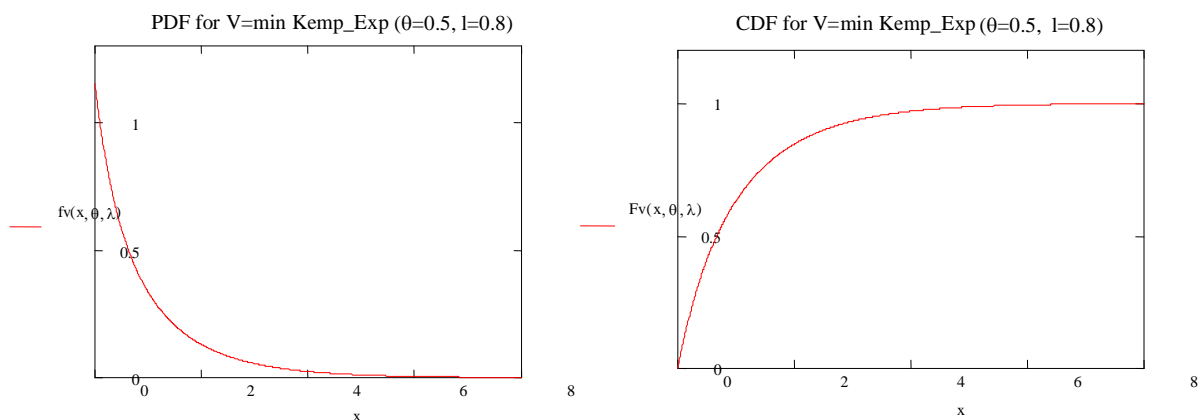
FIG. 8. PDF and CDF for random variable V with different parameters

ii) Similar calculations lead to the following expressions for the of the probability density function and the cumulative distribution function of the random variable *W*

$$f_{W\_K\_E}(x;\lambda,\theta) = \sum_{k=1}^{\infty} k\,\lambda\,e^{-\lambda x}\left(1-e^{-\lambda x}\right)^{k-1}\left(-\frac{\theta^k}{k\ln(1-\theta)}\right) =$$

$$= -\frac{\theta\,\lambda\,e^{-\lambda x}}{\ln(1-\theta)}\sum_{k=1}^{\infty}\left(\theta\left(1-e^{-\lambda x}\right)\right)^{k-1} = -\frac{\theta\,\lambda}{\ln(1-\theta)}\left(\frac{e^{-\lambda x}}{1-\theta\left(1-e^{-\lambda x}\right)}\right)$$

(23)

respectively

$$F_{W\_B\_L}(x;\lambda,\theta) = \int_0^x \frac{\theta\,\lambda\,e^{-\lambda t}}{\ln(1-\theta)}\frac{1}{1-\theta\left(1-e^{-\lambda x}\right)}\,dt = \frac{1}{\ln(1-\theta)}\ln\left(1-\theta\left(1-e^{-\lambda t}\right)\right)\Big|_0^x =$$

$$= \frac{1}{\ln(1-\theta)}\left[\ln\left(1-\theta\left(1-e^{-\lambda x}\right)\right) - \ln\left(1-\theta\left(1-e^{-\lambda 0}\right)\right)\right] = \frac{\ln\left(1-\theta\left(1-e^{-\lambda x}\right)\right)}{\ln(1-\theta)}$$

(24)
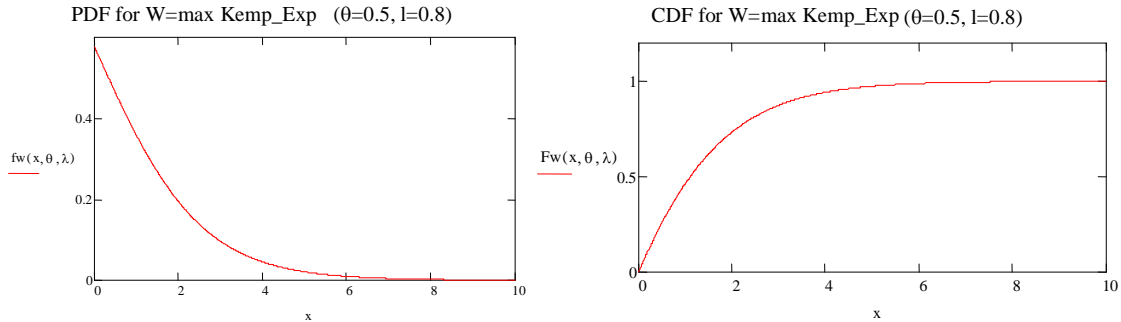
FIG. 9. PDF and CDF for random variable V with different parameters

Font: Times New Roman, 14, bold, centered, in Upper cases, spacing: before − two lines of 14 pt., after − two lines of 14 pt.

2.3. Zipf-exponential distribution $Zipf\_Exp(\alpha, N, \lambda)$

Consider, like in the previous case, n random variables $L_1,...,L_n \sim> Exp(\lambda)$ iid and N a random variable having the $Zipf(\alpha, n)$ distribution, ie its probability function is given by

$$P(N=k) = \frac{1}{k^\alpha \sum_{i=1}^{n}\frac{1}{i^\alpha}}, \quad k=1,2,...n, \quad \alpha \geq 0$$

(25)

If we note $H(\alpha;n) = \sum_{i=1}^{n}\frac{1}{i^\alpha}$ then the cumulative distribution function becomes

$$F(x) = P(N \leq x) = \frac{H_{x,\alpha}}{H_{n,\alpha}}, \quad x=1,2,...,n \,; \, \alpha \geq 0 \,,$$

(26)

Suppose N is a sample of the random variable distributed $Zipf(\alpha, n)$. Then, similar to relations (3), (4) - (7) we will have:

i) probability density function of random variable V

$$f_{V\_Z\_E}(x;\lambda,\alpha,n) = \sum_{k=1}^{n}\frac{1}{H_{\alpha,n}k^\alpha}k\lambda e^{-\lambda k x} = \frac{\lambda}{H_{\alpha,n}}\sum_{k=1}^{n}\frac{e^{-\lambda k x}}{k^{\alpha-1}}$$

(27)

and for the cumulative distribution function

$$F_{V\_Z\_E}(x;\lambda,\alpha,n) = \frac{1}{H_{\alpha,n}}\left(\sum_{k=1}^{n}\frac{1}{k^\alpha}\left(1-e^{-\lambda k x}\right)\right)$$

(28)

In the following figure are represented the graphs of the probability density function and the cumulative distribution function of the random variable $V \sim Zipf\_Exp(\alpha, n, \lambda)$ for the parameter values $\alpha = 3$, $n = 5$, $\lambda = 0.5$
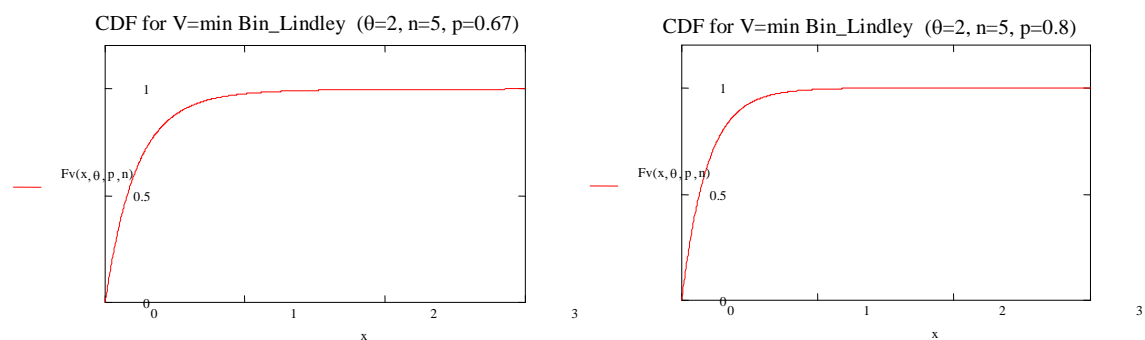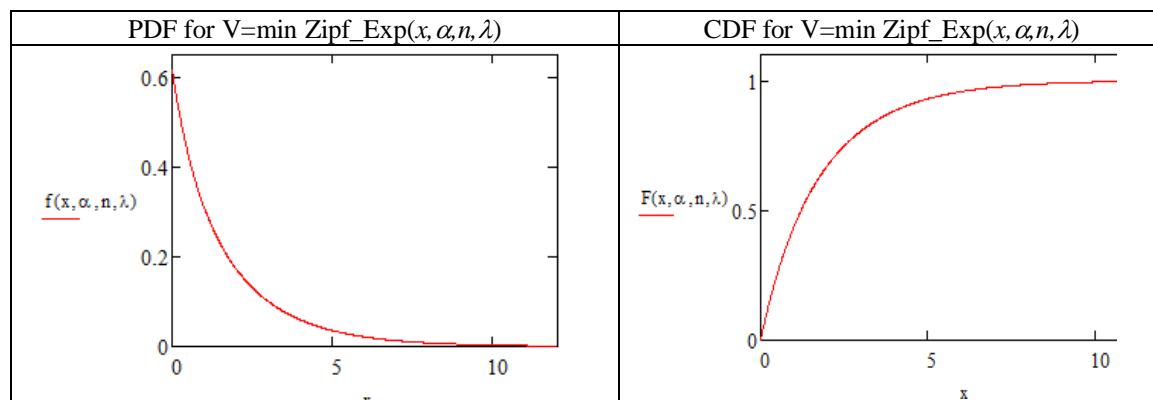

FIG. 10. PDF and CDF for random variable V with above parameters

Analogue calculations lead to the following expressions for the probability density function and the cumulative distribution function of the random variable W

$$f_{W\_Z\_E}(x; \lambda, \alpha, n) = \frac{1}{H_{n,\alpha}} \sum_{k=1}^{\infty} \frac{1}{k^{\alpha}} k \lambda \left(1 - e^{-\lambda x}\right)^{k-1} \tag{29}$$

respectively

$$F_{W\_Z\_E}(x; \lambda, \alpha, n) = \frac{1}{H_{\alpha,n}} \left(\sum_{k=1}^{n} \frac{1}{k^{\alpha}} \left(1 - e^{-\lambda x}\right)^{k} \right) \tag{30}$$

The following figures show the graphs of the probability density function and the cumulative distribution function of the random variable $W \sim Zipf\_Exp(\alpha, n, \lambda)$ for values of parameters $\alpha = 3$, $n = 5$, $\lambda = 0.5$
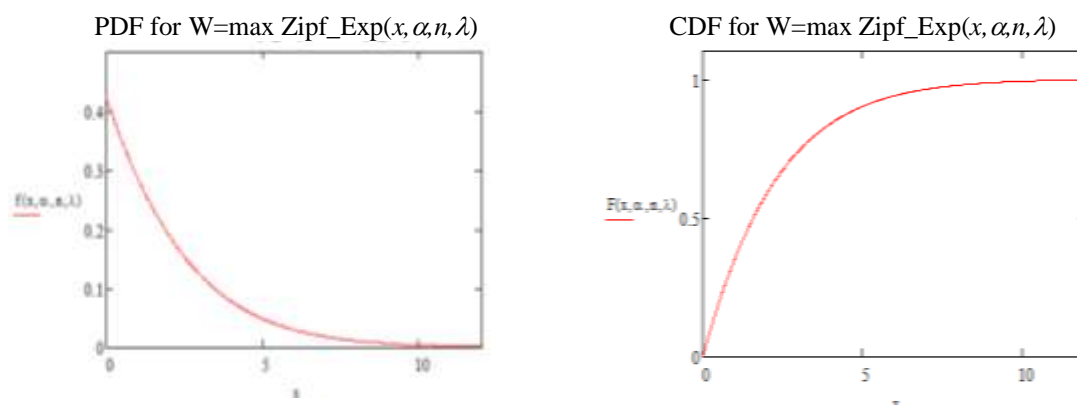
FIG. 11. PDF and CDF for random variable W with above parameters

3. NUMERICAL SIMULATION OF RANDOM VARIABLES *V / W* BY MEANS OF THE INVERSE METHOD

The following theorem is the basis of the inverse method.
Theorem. [9]
If $U \sim \mathcal{U}(0,1)$ then the random variable $\mathcal{F}^{-1}(U)$ has the cumulative distribution function $\mathcal{F}$.

Prove. $(\forall)\, u \in (0,\, 1)$ and $(\forall)\, x \in \mathcal{F}^{-1}([0,\, 1])$, the generalized inverse function checks $\mathcal{F}'(\mathcal{F}(x)) \leq x$ and $\mathcal{F}(\mathcal{F}^{-1}(u)) \geq u$. Then $\{(u,x)|\mathcal{F}^{-1}(u) \leq x\} = \{(u,x)|\mathcal{F}(x) \geq u\}$ and $P(\mathcal{F}^{-1}(U) \leq x) = P(U \leq \mathcal{F}(x)) = \mathcal{F}(x). \square$

3.1. The case of the Bin_Lindley(θ,N,P) distribution
i) Numerical simulation of the random variable V

Let $V \sim> F_{V\_B\_L}(x;\theta,n,p)$ and $U \sim> \mathcal{U}(0,1)$ be two independent random variable, then

the above theorem is obtained $F_{V\_B\_L}(x;\theta,n,p) = U$, ie $V = \dfrac{X}{\theta}$, where X is the solution

of the following equation

$$\frac{1+\theta+X}{e^X} - \frac{\left(\left(1-U(1-q^n)\right)^{\frac{1}{n}} - q\right)(1+\theta)}{p} = 0 \tag{31}$$

ii) Numerical simulation of the random variable W

Let $W \sim> F_{W\_B\_L}(x;\theta,n,p)$ and $U \sim> \mathcal{U}(0,1)$, then, from the above theorem is

obtained $F_{W\_B\_L}(x;\theta,n,p) = U$, that is $W = \dfrac{X}{\theta}$, where X is the solution of the equation

$$\frac{1+\theta+X}{e^X} - \frac{1-\left(U(1-q^n)+q^n\right)^{\frac{1}{n}}}{p}(1+\theta) = 0 \tag{32}$$

Table 1 summarizes the theoretical and empirical mean and variance obtained for 10000 simulated variables for parameter values.

Table 1. The theoretical and empirical mean and variance

Variable	Mean		Variance	
	Theoretical	Empirical	Theoretical	Empirical
V	0.2397	0.238	0.07411	0.07059
W	1.22824	1.22641	0.52662	0.5219

3.2. The case of the *Kemp_Exp*(θ, λ) distribution
i) Numerical simulation of the random variable V

Let $V \sim> F_{V\_K\_E}(x;\theta,\lambda)$ and $U \sim> \mathcal{U}(0,1)$, then, from the above theorem is obtained

$$F_{V\_K\_E}(x;\theta,\lambda) = U \text{ ie } V = \ln\left(\frac{\theta}{1-(1-\theta)^U}\right)^{\frac{1}{\lambda}}$$

ii) Numerical simulation of the random variable W

Let $W \sim> F_{W\_K\_E}(x;\theta,\lambda)$ and $U \sim> \mathcal{U}(0,1)$, then, from the above theorem is obtained

$$F_{W\_K\_E}(x;\theta,\lambda) = U \text{ ie } W = \ln\left(\frac{\theta}{\theta-1-(1-\theta)^U}\right)^{\frac{1}{\lambda}}$$

Table 2 shows the theoretical mean and variance and the empirical mean and variance obtained for 10000 simulated variables for parameter values.

Table 2. The theoretical and empirical mean and variance

Variable	Mean		Variance	
	Theoretical	Empirical	Theoretical	Empirical
V	1.04999	1.04248	1.3195	1.33747
W	1.48321	1.48257	1.86462	1.86015

3.3. The case of the *Zipf_Exp(α, N, λ)* distribution

The inverse simulation of the random variable $V \sim$ Zipf_Exp (α, n, λ) returns to solving by an approximate method of the equation

$$\frac{1}{H_{\alpha,n}}\left(\sum_{k=1}^{n}\frac{1}{k^{\alpha}}\left(1-e^{-\lambda k x}\right)\right)=U \qquad (33)$$

where $U \sim \mathcal{U}(0,1)$.

For simulation by the inverse method of the random variable $W \sim$ Zipf_Exp (α, n, λ), it is necessary to solve by an approximate method the equation

$$\frac{1}{H_{\alpha,n}}\left(\sum_{k=1}^{n}\frac{1}{k^{\alpha}}\left(1-e^{-\lambda x}\right)^{k}\right)=U \qquad (34)$$

where $U \sim \mathcal{U}(0,1)$.

For the 10000 simulated values and for the parameters values $\alpha = 3$, $n = 5$, $\lambda = 0.5$ the theoretical and empirical mean and variance are presented in Table 3.

Table 3. The the theoretical and empirical mean and variance for α=3, n=5 , λ=0.5

Variable	Mean		Variance	
	Theoretical	Empirical	Theoretical	Empirical
V	1.82236	1.80352	3.67365	3.60838
W	2.20336	2.28037	4.44255	4.59143

4. NUMERICAL SIMULATION OF RANDOM VARIABLES V / W STARTING FROM THE DEFINITION OF THESE VARIABLES

4.1. The case of the *Bin_Lindley(θ, N, P)* distribution

We consider $U \sim \mathcal{U}(0,1)$ and $H(X,\theta,U)=e^{X}-\dfrac{1+\theta+X}{(1+\theta)U}$.

If X is the solution of the equation $H(X,\theta,U)=0$, taking $x=\dfrac{X}{\theta}$, $\theta \neq 0$, we have a sample of the random variable *Lindley*(θ).

We generate the random variable *Lindley*(θ) by inverse method with the following algorithm.

The Lindley(θ, N) *algorithm*

P0. Input θ - distribution parameter, N - sample volume;

P1. For $k := 1; N$

Generate $U \sim \mathcal{U}(0,1)$

$T:=(U>0)\wedge\left(H(0,\theta,U)\cdot H(30,\theta,U)<0)\right)$;

If $T = true$ then

If $x \neq 0$ then $L_k := \dfrac{X}{\theta}$;

P2. Returns L; Stop!

Algorithm for simulation of random variables V / W distributed Bin_Lindley(θ, n, p)
P0. Input: (θ, n, p), N-volume of the sample;
P1. For $i := 1; N$

   Generate $m \sim Bin(n, p)$

   while $m < 1$ Generate $U \sim \mathcal{U}(0,1)$, Generate $m \sim Bin(n, p)$;

   L: = Lindley (θ, m)

   $V_i := \min(L_1,...,L_m)$, $W_i := \max(L_1,...,L_m)$;

P3. Returns *V, W*; Stop!

Applying the algorithm for $n = 5$, $p = 0.67$, $\theta = 2$, for a sample of 10000 simulated values, the results from Table 4 are obtained.

Table 4. The theoretical and empirical mean and variance for $n = 5$, $p = 0.67$, $\theta = 2$

Variable	Mean		Variance	
	Theoretical	Empirical	Theoretical	Empirical
V	0.2397	0.24447	0.07411	0.07848
W	1.22824	1.2385	0.52662	0.52374

4.2. The case of the *Kemp_Exp(θ, λ)* distribution

To simulate the random variable *Kemp(θ)* we can use the composition method [1].

Let $\quad Y \sim F_Y(y) = \dfrac{\ln(1-y)}{\ln(1-\theta)}$, $0 < \theta < 1$, $0 < y < \theta$ $\quad$ and $\quad N \sim Kemp(\theta)$ $\quad$ then,

$P(N = k | Y = y) = (1-y)y^{k-1}$, $k = 1, 2,...$ that is, the distribution of the random variable N conditioned by $Y = y$ is a sample of the truncated geometric random variable *Geom(y)*.

Prove. The distribution density of the variable Y is

$$f_Y(y) = \frac{-1}{1-y}\frac{1}{\ln(1-\theta)} , \quad 0 < \theta < 1 , \quad 0 < y < \theta$$

Then

$$P(N = k) = \int_0^\theta P(N = k | Y = y) f_Y(y) dy = -\frac{1}{\ln(1-\theta)}\int_0^\theta y^{k-1} dy = -\frac{1}{\ln(1-\theta)}\frac{1}{k} y^k \Big|_0^\theta = -\frac{\theta^k}{k\ln(1-\theta)}$$

which means that N is a sample of the random variable *Kemp(θ)*.

To simulate the Geom (y) geometric random variable we use the inverse method as follows:

$P(N = k) = p q^{k-1}$, $k = 1, 2,...$

and $F(n) = P(N < n) = \sum_{k=1}^{n-1} p q^{k-1} = 1 - q^n$, $\quad F(n) = U$, $U \sim \mathcal{U}(0,1)$, ie $n = \left\lceil \dfrac{\ln U}{\ln q} \right\rceil$.

In this case $q = y$.

Algorithm for simulating the Kemp(θ) variable by the composition method
P0. Input: θ, N-volume of the sample
P1. For $i := 1; N$

   Generate $U \sim \mathcal{U}(0,1)$

   $y := 1 - (1 - \theta)^U$

   Generate $U \sim \mathcal{U}(0,1)$

      If $(y \cdot U \neq 0) \wedge (\theta \geq y)$ then $K_i := \left\lceil \dfrac{\ln U}{\ln y} + 0.5 \right\rceil$

P2. Returns *K*; Stop!

Algorithm for simulating V / W variables distributed Kemp_Exp(θ, λ)
P0. Input: (θ, n, p), *N*-volume of the sample
P1. For *i*: = 1; *N*

Generate $m \sim> Kemp(\theta)$

For *j*: = 1; *m*

Generate $U \sim> \mathcal{U}(0,1)$

If $U \neq 1$ calculate $L_j := -\frac{1}{\lambda}\ln(U)$

$V_i := \min(L_1,...,L_m)$, $W_i := \max(L_1,...,L_m)$;

P3. Returns *V, W*; Stop!

Applying the algorithm for θ = 0.8 and λ = 0.5, for a sample of 10000 simulated values, the results in Table 5.

Table 5. The theoretical and empirical mean and variance for θ = 0.8 and λ = 0.5

Variable	Mean		Variance	
	Theoretical	Empirical	Theoretical	Empirical
V	1.33561	0.08242	2.74246	0.00668
W	2.94505	7.44391	6.07521	6.20381

4.3. The case of the *Zipf_Exp(α, N, λ)* distribution

To simulate the *Zipf(α, n)* distribution we use a variant of the algorithm [10]
The Zipf_invers(α, n) algorithm
P0. Input: *α, n*-model parameters

j: = 0;

Generate $U \sim> \mathcal{U}(0,1)$;

P1. While $F(x) = \frac{H_{x,\alpha}}{H_{n,\alpha}} > U$ calculate *j*: = *j* + 1;

P2. Returns *N: = j;*Stop!

For the numerical simulation of the random variable *Zipf_Exp(α, n, λ)*, starting from the definition of variables *V* and *W*, we use the following algorithm.

The Zipf_Exp_Direct algorithm
P0. Input: (α, n, λ) - model parameters, *N* - volume of the sample;
P1. For *i*: = 1; *N* execute

m: = Zipf_invers(α, n);

For *j: = 1; m*

Generate $U \sim> \mathcal{U}(0,1)$;

$L_j := -\frac{1}{\lambda}\ln(U)$

$V_i := \min_{1 \leq j \leq m} L_j$, $W_i := \max_{1 \leq j \leq m} L_j$

P2. Calculate: the sample mean and variance of *V* and *W*: *M*[*V*], Var[*V*], *M* [*W*], Var[*W*];

Returns *M*[*V*], Var[*V*], *M*[*W*], Var[*W*];

Stop!

Table 6 shows the average and the theoretical dispersion for parameter values $\alpha = 3, n = 5, \lambda = 0.5$ for a 10,000 volume sample.

Table 6 The theoretical and empirical mean and variance for $\alpha=3$, $n=5$, $\lambda=0.5$

Variable	Mean		Variance	
	Theoretical	Empirical	Theoretical	Empirical
V	1.82236	1.80352	3.67365	3.60838
W	2.20336	2.28037	4.44255	4.59143

The numerical results of applying the two above algorithms are listed in Table 6.

5. CONCLUSIONS

In this paper we obtained three probability distributions with possible applications in the reliability of multi-component systems using the computation method (consisting of discrete distributions with continuous distributions). For these distributions we simulated 10,000 variables by the inverse method and using their definition for different parameter values and we compared the methods by considering the theoretical mean and the variance with the sampling mean and variance respectively. It can be concluded that the methods lead to good results as can be seen from the Tables 1-6.

REFERENCES

[1] Byron J.T. Morgan: *Elements of Simulation,* Chapman and Hall. London, New York, 1984

[2] Ghitany M.E., Atieh B. and Nadarajah S.: Lindley distribution and its application, *Mathematics and Computers in Simulation*,78 (2008) 49350

[3] Faton Merovci: Transmuted Lindley Distribution, *Int. J. Open Problems Compt. Math.*, Vol. 6, No. 2, June 2013, ISSN 1998-6262; Copyright ICSRS Publication, 2013, www.i-csrs.org

[4] Johnson, N. L., Kotz, S., Kemp, A. W.: *Univariate Discrete Distributions*, Second Edition, John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore, 1992

[5] Kus, C.: A new life time distribution, *Computational Statistics& data Analisys*, 51 (2007) 4497-4509

[6] Lindley, D.V.: Fiducial distributions and Bayes' theorem, *Journal of the Royal Statistical Society,* Series B 20 (1958) 102–107.

[7] Lindley, D.V.: *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Part II: Inference, Cambridge University Press, New York,1965.

[8] Preda, V., Bancescu, I., Mircea Dragulin, M., Tuculan (Diaconu) M-C : Compound Generalized Lindley distributions: Poisson, Binomial and Geometric type, *International Journal of Risk Theory*, Vol 6 (no.2), 2016, pag. 1-18.

[9] Robert, C. P. şi Casella, G.: *Monte Carlo Statistical Methods*, Springer, 1999.

[10] Vaduva, I.: *Simulation Models with Computers,* (Romanian), Ed. Tehnica, Bucuresti (1977).

[11] Vaduva, I.: Simulation of Some Mixed Lifetime Distributions, *Review of the Air Force Academy*, No.2 (34)/2017, pag. 5-16.

[12] www.math.wm.edu/~leemis/chart/UDR/UDR.html

ON THE SIMULATION OF SOME PARTICULAR DISCRETE DISTRIBUTIONS

Ion VĂDUVA, Mihăiţă DRĂGAN

University of Bucharest, Romania (vaduva@fmi.unibuc.ro, dragan_mihaita@yahoo.com)

Abstract: The paper summarizes the first several known discrete probability distributions which may describe the occurrence of a random number of events in various experiments, e.g in a reliability system, the occurrence of a random number of failures. Among these discrete distributions, the first mentioned are usual distributions such as: Poisson(X),\ > 0;Geometric(q),0 < q< l; Pascal(k,p),k £ N+, 0 < p < 1; Binomial(n,k,p),n,k £ N+,0 < p < 1. Then, some new discrete distributions are defined in terms of positive convergent series {an}, 1 < n < oo,an > 0. The paper presents methods of simulating the above mentioned distributions, which are either general, like the inverse method, or based on the rejection enveloping method. As enveloping distributions, either of the said distributions – i.e., Poisson and Geometric – or other less known distributions – such as the Zipf distribution or the Yule Distribution – are used. Comments related to testing these algorithms are finally presented.

Keywords: discrete probability distributions, Zipf distribution, Yule Distribution

1. INTRODUCTION

Any discrete distribution, in the form $p_n = P(N = n), n = 1,2,...$ can describe the occurence of a random number of events. In reliability, some usual distributions of this type are used, basically, the following distributions [2,3,5,9], truncated on [l,oo), (i.e. $n > 1$).

a. Geometric distribution *Geo(p)*, $O < p < 1$, definesd as

$$p_n = P(N = n) = q^n, n \text{ G } N^+, \tag{1.1}$$

b. Pascal distribution *Pas(p,k)*, $0 < p < 1$, $k \in N^+$, defined as

$$p_n = P(N = n) = C_{n+k-1}^{k-1} \frac{p^k q^n}{1-p^k}, n \in N^+, \tag{1.2}$$

c. *Poisson(λ)*, $\lambda > 0$ distribution defined as

$$p_n = P(N = n) = \frac{1}{e^\lambda - 1} \frac{\lambda^n}{n!} e^{-\lambda}, n \in N^+ \tag{1.3}$$

d. Binomial distribution *Binomial(n,p)*, $n \in N^+$, $0 < p < 1$ defined as

$$p_\alpha = P(N = \alpha) = \frac{1}{1-q^n} C_n^\alpha p^\alpha q^{n-\alpha}, q = 1 - p, \alpha \in N^+. \tag{1.4}$$

Methods for simulating these distributions are presented in various papers (see [1,5,9]).

Note that any convergent series of positive terms a_n, $n \geq 1$, could define *a discrete distribution*.

If

$$\beta = \sum_{n=1}^{\infty} a_n, 0 < \beta < \infty, \tag{1.5}$$

then the probabilities of the discrete distributions derived from such series are

$$p_n = \frac{a_n}{\beta}. \tag{1.5'}$$

Some other known discrete distributions, applied in different circumstances are the following.

e. *The distribution of Euler [5] defined as*

$$p_n = \left(\frac{1}{n} - \ln\left(1 + \frac{1}{n}\right)\right)\frac{1}{\beta}, n \ge 1, \tag{1.6}$$

where

$$\beta = \gamma = \lim_{n \to \infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \ln n\right) \tag{1.6'}$$

where γ is the *Euler's constant* $(0 < \gamma < 1)$ [8].

f. The *distribution of Kemp* defined as [4]

$$p_n = -\frac{\alpha^n}{n \log 1 - \alpha}, n \ge 1, 0 < \alpha < 1, \tag{1.7}$$

This distribution is also called *logarithmic series distribution of the parameter p*, $0 < p < 1$ and p_n is in equivalent form

$$p_n = \frac{a}{n} p^n, n = 1,2, \dots a = -\frac{1}{\log(1 - p)}. \tag{1.7'}$$

g. The *Zipf distribution* [1,5,6] of the parameter a, $a > 1$ defined as

$$p^n = \frac{1}{\zeta(a)n^a} n > 0, \tag{1.8}$$

where

$$\zeta(a) = \sum_{i=1}^{\infty} \frac{1}{i^a} \tag{1.8'}$$

is the *Riemann's function*. This distribution describes the occupied memory cells in the computer when the memory is dynamically allocated.

h. *The Yule(a)* distribution of the parameter a, $a > 1$, defined as [1,5,6]

$$p_n = P(N = n) = \frac{1}{c(a)} B(n, a), n \ge 1, c(a) = \sum_{n=1}^{\infty} B(n, a), \tag{1.9}$$

where $B(n,a)$ is *Beta function* defined as

$$B(n, a) = \int_0^1 (1 - u)^n u^a du, \tag{1.9'}$$

which is connected with function $\Gamma(p)$ by the formula

$$B(n, a) = \frac{\Gamma(n)\Gamma(a)}{\Gamma(n + a)}. \tag{1.9''}$$

The function $\Gamma(p)$ is defined as

$$\Gamma(p) = \int_0^\infty u^{p-1}e^{-u}du, p \in R^+. \tag{1.9'''}$$

Note that for $p \in N^+$, the function $\Gamma(p)$ is

$$\Gamma(p) = (p-1)!. \tag{1.9^{iv}}$$

The simulation of these distributions is presented in several books and papers (see [1,4,5,7]) and they will be briefly described as such in a following section of this paper.

In [8] **several positive convergent series** $\{a_n\}_{n \in N^+}$ are found which could define such discrete distributions, as

$$P(N = n) = p_n = \frac{a_n}{\beta}, \sum_{n=1}^\infty p_n = 1.$$

The following is a list of positive convergent series collected from [8]:

$$a_n = \frac{n}{a^n}, a > 1, \sum_{n=1}^\infty a_n = \frac{a}{(1-a)^2} = \beta, \tag{1.10}$$

$$a_n = \frac{1}{n(n+p)}, p \in N^+, \sum_{n=1}^\infty a_n = \frac{1}{p}\sum_{k=1}^p \frac{1}{k} = \beta \tag{1.11}$$

$$a_n = \frac{1}{(n-1)! + n!}, \sum_{n=1}^\infty a_n = 1 = \beta \tag{1.12}$$

$$a_n = \frac{n}{(2n+1)!}, \sum_{n=1}^\infty a_n = \frac{1}{2} = \beta \tag{1.13}$$

$$a_n = \frac{n}{(n+1)!}, \sum_{n=1}^\infty a_n = 1 = \beta \tag{1.14}$$

$$a_n = \frac{n!}{(n+k)!}, \sum_{n=1}^\infty a_n = \frac{1}{(k-1)k!} = \beta \tag{1.15}$$

$$a_n = \frac{1}{n^2}, \sum_{n=1}^\infty \frac{1}{2^n} = \varsigma(2) \tag{1.16}$$

$$a_n = \frac{n^2}{n!}, \sum_{n=1}^\infty a_n = 2 + 2e = \beta \tag{1.17}$$

$$a_n = \frac{n^3}{n!}, \sum_{n=1}^\infty a_n = 5e = \beta \tag{1.18}$$

$$a_n = \frac{(p+1)\dots(p+n)}{(q+1)\dots(q+n)}, q - p > 1, \beta = \frac{p+1}{q-p-1} \tag{1.19}$$

One aim of this paper is to present methods for simulating the distributions defined by (1.10)…(1.19).

2. THE INVERSE METHOD

Any probability distribution can be simulated by a general method, *the inverse method* [1,5,6,7,9]. If $F(x) = P(X < x)$ is the *cumulative distribution function (cdf)* of a random variable X, then a sampling value of X is simulated by the formula $X = F^{-1}(U)$, where U is a random number, uniformly distributed over (0,1). (See [1,6,7,9,10]). This induces the following algorithm:

Algorithm INV

begin

generate U an uniform random number over (0,1);

take $F^{-1}(U)$, (where F^{-1} is the inverse of function F(x)=P(X < x));

end.

The method can be used if there is an easy way to calculate the inverse function $F^{-1}(U)$. In the discrete case, where the function F is a "step" function, the jumps of the function are in the points 1,2,…. Thus, the distinct values of F(x) are F(i) defined as

$$F(i) = \begin{cases} 0, if\ x < 0 \\ \sum_{\alpha=0}^{i} p_\alpha, if\ i \le x < i+1, i = 1,2,... \end{cases} \tag{2.1}$$

To simulate a random sampling value i, we must calculate $F^{-1}(U)$ using F in the formula (2.1) as a step function. In other words, we must *search* the index i, such as $F(i) \le U < F(i+1)$. There are various possibilities to search i. One could be *binary search*. Here we use a simpler (but not faster!) procedure, based on dividing the interval (0,1) in *five* intervals, namely

$$(0,1) = (0,0.25) \cup [0.25,0.5) \cup [0.5,0.75) \cup [0.75,0.95) \cup [0.95, \infty).$$

The algorithm uses a table of distinct values of $F(i), 1 \le i \le i_0$, such as $F(I_0) = 0.95$ (i.e. $F(I_0)$ is large enough). The values F(i) are calculated as follows:

$$F(i) = \begin{cases} p_1\ if\ i = 1 \\ p_1 + p_2\ if\ i = 2 \\ ,... ... ... \\ p_1 + p_2 + \cdots + p_k\ if\ i = k \end{cases} \tag{2.1'}$$

(The means by which k can be determined is explained later; it is the I_0 index below).

Let us select the indexes I_1, I_2, I_3, I_0 as follows:

$$\begin{cases} F(I_1) \le 0.25 < F(I_1 + 1); \quad F(I_2 \le 0.5 \le F(I_2 + 1); \\ F(I3) \le 0.75 < F(I_3 + 1); \quad F(I_0) \le 0.95 < F(I_0 + 1). \end{cases} \tag{2.2}$$

The detailed algorithm **INV** is the following

Preparatory step; Calculate F(1),…F(I_0), determine I_1, I_2, I_3, I_0.

1. **generate** U **Uniform on** (0,1).

2. **if** U ≤ 0.25 **then begin**

i = I_1; **while** U $\le$ F(i) **do** i := i − 1 **end else if** U ≤ 0.5 **then**

begin i = I_2; **while** U $\le$ F(i) **do** i := i − 1 **end**

else if U ≤ 0.75 **then begin** i = I_3; **while** U $\le$ F(i) **do** i := i − 1 **end**

else if U $\le$ F(I_0) **then begin** i := I_0; **while** U $\le$ F(i) **do** i := i − 1 **end**

else begin i := I_0; **while** U > F(i) **do begin** i := i + 1; F(i) := F(i) + p_i

end; end;

deliver i.

(i.e. "i" is the generated sampling value). As $F(i), 1 \leq i \leq I_0$ are calculated only once and if I_0 is small, then the algorithm is fast for generating a sampling value *i*. But sometimes I_0 may not be small at all, and then the algorithm will be slow.

The detailed algorithm, described in steps **1.** and **2.** can be adapted and applied to each of the distributions (1.10-1.19).

3. THE ACCEPTANCE-REJECTION METHOD

There are various versions of this method (see [1,4,5,6,7,9,10]). Here, we will be using the rejection method based on **enveloping** the frequency function $f(n) = p_n$ of the distribution with another frequency function $h(n) = q_n$, which can be simulated. Let us assume that there is a constant α > 1 such as $\frac{f(n)}{h(n)} < \alpha, n = 1,2, ...$

The formal Theorem is the following: if X is a random variable with frequency function p_n (to be simulated) and if Y is another random variable (which can be simulated) whose frequency function is h_n, and if there is a constant $\alpha, 1 < \alpha < \infty$ such as

$$\frac{p_n}{h_n} \leq \alpha,$$

and if U is an uniform (0,1) random number independent of Y, then if

$$0 < U \leq \frac{p_n(Y)}{\alpha h_n(Y)},$$

the simulated value of X is X = Y.

The general simulation algorithm is:

> **Algorithm REJ**
> **repeat**
> *simulate a random variate U uniform (0,1);*
> *simulate j a random variate with the frequency function h(n);*
> **until** $U \leq \frac{f(j)}{\alpha h(j)}$;
>
> *deliver i = j;*

The value *i* is the simulated sampling value of *f(n)*.

The acceptance probability of the algorithm is $p_a = \frac{1}{\alpha}$ and if it is large, then the algorithm is fast. The function *h* is *the enveloping function*. To build up the algorithm **REJ**, it is important to find **a good** enveloping function *h* such a way as the **acceptance probability** which is large.

Discussions on simulating by rejection procedure REJ any of distributions (1.10)-(1.19) will be based on the following idea: the distribution *h(n)* could be a distribution which is decreasing with *n*, such as can happen in cases of convergent series with positive terms.

This suggests that sometimes (but not always), a candidate for *h(n)* could be the **geometric distribution** *Geom(p)*, $0 < p < 1, q = 1 - p$, for which

$$p_n = pq^n, 0 < p < 1, q = 1 - p, n = 0,1,2 \ldots \tag{3.1}$$

truncated to $n \geq 1$. Therefore, in this case, the distribution *h(n)* has probabilities

$$h_n = \frac{p}{1 - q}q^n = q^n, n = 1,2, \ldots \tag{3.2}$$

An enveloping candidate could be also *Poisson* or any other selected distribution.

First we have to specify how to simulate the truncated distribution *Geom(p)*, $n \geq 1$.

In [1,5,9] two procedures to simulate this distribution are presented. Note that this distribution is related to **Bernoulli triles**.

A Bernoulli trile is an experiment on an event with *constant probability p* which, when it occurs, we say that is a *success* and when it does not occur, we say that is a *failure*. The *number of failures N until a success occurs* is a random variable having distribution *Geom(p)*. Therefore, this can be simulated as:

Algorithm COUNT FAILURES

1. *Read p*, $0 < p < 1; j := 0$;

2. **Repeat**

Generate U uniform (0, 1); if $U \geq p$ then $j : +j + 1$

until $U < p$.

The value *j* is the simulated value of *Geom(p)*.

We can also use *the inverse method* to simulate *Geom(p)*. The cdf in this case is

$$F(n) = P(N < n) = \sum_{i=0}^{n-1} pq^i = p \frac{1-q^n}{1-q} = 1 - q^n. \tag{3.3}$$

The inverse method gives

$$j = \left[\frac{U}{q} \right] \tag{3.3'}$$

where [t] means the integer closest to real number t.

Since the values of *j* must be positive, we have to *reject* the value of j = 0, i.e. the algorithm is:

repeat

generate j from Geom(p)

until $j > 0$.

The value *j* is the simulated value of the **truncated** *Geom(p)*.

In order to build up the algorithm **REJ** for all discrete distributions in the form (1.10)-(1.19), it is enough, in each case, to specify the *possible envelope distribution and then to determine the constant α* in the algorithm **REJ**. For instance, to determine *q* of the enveloping *Geom*(p), we find first the maximum value of a_n and if this is $\beta = a_m$, then select *q* in the form $q = \beta$, β will be a normalizing number.

3.1 Simulation of the distribution defined by (1.10).

Method 1. Here is where we try to determine the *geometric distribution* as envelope.

The maximum of a_n is determined as the maximum of the function

$$f(x)_{max} = \max\left(\frac{1}{\beta} \frac{x}{a^x} \right), \qquad x \geq 0.$$

i.e. the maximum of

$$f(x) = xa^{-x}, \qquad a > 1.$$

After some calculations, it results that the maximum point of this function is

$$x_0 = \sqrt{\frac{a}{\ln(a)}} > 1, \qquad \ln(a) > 0,$$

and hence

$$(a_n)_{max} = \frac{(a-1)^2}{a} \frac{x_0}{a^{x_0}},$$ and this gives

$$q = \frac{(a-1)^2}{a} \frac{x_0}{a^{x_0}}. \tag{3.4}$$

To determine α, consider the ratio

$$r_n = \frac{a_n}{q^n}, \qquad i.e.\, r(x) = \frac{f(x)}{q^x}$$

which, in a similar manner has the same maximum point x_0, and after some calculation we finally obtain

$$\alpha = \frac{a^{x_0}+1}{a^{x_0}+1-(a-1)^2 x_0} > 1. \tag{3.4'}$$

Now, the construction of the algorithm **REJ** is terminated.

Method 2. An alternative method of simulating this distribution is to use *the inverse method* of the *equivalent* distribution directly:

$$f(n) = \frac{b^3}{(b-1)^2} n b^{n-1} = K^* n b^{n-1}, \qquad b = \frac{1}{a} < 1, K^* = \frac{b^3}{(b-1)^2}.$$

The cdf is therefore

$$F(n) = K^* \sum_{i=1}^{n} i b^{i-1} = K^* \left(\sum_{i=1}^{n} b^n\right)' = K^* \left(\frac{b^{n+1}-b}{b-1}\right)' = K^* \frac{(n+1)b^n}{b-1}$$

where *the derivative ()'* is calculated with respect to n. To apply the inverse method, we have to solve in n (numerically!) the equation

$$F(n) = U, i.e., K^* \frac{(n+1)b^n}{b-1} = U. \tag{3.4''}$$

where U is an uniform random number over (0,1). If n_0 is the solution of (3.4'') then the simulated value is n = int(n_0).

Method 3. Let us use as enveloping distribution the *Kemp distribution* i.e.

$$h_n = -\frac{1}{\log(1-p)} \frac{p^n}{n}, 0 < p < 1.$$

Then the ratio r_n becomes

$$r_n = -\frac{(a-1)^2 \log(1-p)}{a} \frac{n^2}{b^n}, b = ap.$$

If b>4, it is shown by induction that

$$\frac{n^2}{b^n} < 4$$

Therefore, when p is selected such as ap > 4 then $\alpha > 1$ and the algorithm **REJ** is obvious. With respect to p_a, it seems that **method 3** is preferable.

3.2 Simulation of the distribution defined by (1.11)

Method 1. Note that the sequence is

$$a_n = \frac{1}{\beta} \frac{1}{n(n+p)}, p_n = \frac{a_n}{\beta}, \beta = \frac{1}{p} \sum_{k=1}^{p} \frac{1}{k} = \frac{1}{p} K^*.$$

Let us choose this time as *enveloping distribution* a Zipf(2) distribution [1,5] defined as

$$q_n = \frac{1}{\zeta(2)n^2}, \qquad \zeta(a) = \sum_{n=1}^{\infty} \frac{1}{n^2}. \tag{3.5}$$

Consider the ratio

$$r(n) = \frac{p_n}{q_n} = \frac{\zeta(2)}{\beta} \frac{n^2}{n(n+p)}.$$

After some simple calculations we have

$$r_n \leq \frac{\zeta(2)}{\sum_{k=1}^{p} \frac{1}{k}} = \alpha > 1 \tag{3.5'}$$

and the construction of the algorithm **REJ** is finished. Simulation of the Zipf distribution is found in [1,5] and is presented in the last section. There is a version of this distribution [1] which is defined for n = 1,2,...$K^* < \infty$ (i.e. a finite series!), referring to a finite population of size K^*. Comments on this, will me made in the last section of the paper.

Method 2. Let us take as enveloping distribution that given by (1.10). Therefore we have

$$p_n = \frac{p}{\sum_{k=1}^{p} \frac{1}{k}} \frac{1}{n(n+p)} , h_n = \frac{(a-1)^2}{a} \frac{n}{a^n}.$$

then the ratio is

$$r_n = \frac{pa}{(a-1)^2 \sum_{k=1}^{p} \frac{1}{k}} \frac{a^n}{n^2(n+p)} = \frac{pa}{(a-1)^2 S} \frac{a^n}{n^2(n+p)} , \qquad S = \sum_{k=1}^{p} \frac{1}{k}$$

Note that ratio

$$R_n = \frac{a^n}{n^2(n+p)} < \frac{a}{p+1} , a > 1, p \geq 1,$$

proved by induction. Therefore

$$r_n < \alpha < \frac{pa}{(a-1)^2 S} \frac{a}{p+1},$$

Which if $a \geq p+1$ gives α > 1, and the algorithm **REJ** is obvious. To decide which of these method is preferable, it is necessary to numerically compare the p_a of the two methods.

3.3 Simulation of the distribution defined by (1.12)

Method 1. For the sequence a_n , we have

$$a_n = p_n = f(n) = \frac{1}{(n-1)!(n+1)} , \beta = 1,$$

we select the enveloping distribution h(n) as a *Poisson(1)* i.e.

$$q_n = \frac{1}{(n-1)!} , \sum_{i=1}^{\infty} q_n = e, h(n) = \frac{1}{e} \frac{1}{(n-1)!} \tag{3.6}$$

The ratio

$$r(n) = \frac{f(n)}{h(n)} = \frac{e(n-1)!}{(n-1)!(n+1)} \leq \frac{e}{2} = \alpha, \qquad \alpha > 1. \tag{3.6'}$$

Since elements f(n), h(n), α are specified, the algorithm **REJ** is obvious.

Method 2. If we select as enveloping distribution that given by (1.11), then we have

$$r_n = \frac{n(n+p)}{(n-1)!(n+1)} \frac{1}{\beta}$$

and

$$r_n = \frac{n^2(n+p)}{(n+1)n(n-1)!\,\beta} \le \frac{p+1}{\beta} = \alpha, \qquad \alpha = \frac{(p+1)\sum_{k=1}^{p}\frac{1}{k}}{p} > 1$$

therefore, the algorithm **REJ** is defined.

Method 3. Since $p_n \le \frac{1}{2}$, we can take as envelope the *Geo*(p), p $= \frac{1}{2}$ and

$$r_n = \frac{2.2^n}{(n-1)!\,(n+1)} \le \frac{4e}{2} = 2e^2 > 1$$

and again, the required algorithm is ready. Note that **method 1** is the best of the three methods, since in that case α is close to one.

3.4 Simulation of the distribution defined by (1.13)
Method 1. The sequence

$$a_n = \frac{n}{(2n+)!}, \beta = \frac{1}{2}, gives\; p_n = \frac{2n}{(2n+1)!}$$

If we select as enveloping distribution the *Poisson(1)* distribution in the form

$$h_n = \frac{1}{e}\frac{1}{(n-1)!}, n \ge 1, \tag{3.7}$$

we obtain

$$r_n = \frac{p_n}{h_n} = \frac{2en(n-1)!}{(2n+1)!} \le \frac{2e}{2} = \alpha = e. \tag{3.7'}$$

The algorithm **REJ** is obvious in this case.

Method 2. Let us take as envelope the distribution (1.12). In this case, we have

$$r_n = \frac{2n(n+1)(n-1)!}{(2n+1)!} = \frac{2(n+1)!}{(2n+1)!} \le \frac{2(2n+1)!}{(2n+1)!} = 2 = \alpha \tag{3.3''}$$

and algorithm **REJ** is obvious. This method is better than **method 1**, since p_a is larger.

3.5 Simulation of the distribution derived from (1.4)
Method 1. In this case

$$a_n = p_n = \frac{n}{(n+1)!}.$$

We select as an enveloping distribution the *Poisson*(1), i.e.

$$q_n = \frac{1}{(e-1)n!}. \tag{3.8}$$

The ratio r_n is

$$r_n = \frac{p_n}{q_n} = \frac{n(e-1)}{n+1} \le e-1 = \alpha \tag{3.8'}$$

and elements of the algorithm **REJ** are determined.

Method 2. Let us choose as enveloping distribution the *Kemp distribution*. Then the ratio becomes

$$r_n = -\log(1-p)\frac{n^2}{(n+1)!\,p^n} \le -\log(1-p)\frac{1}{2p} = \alpha, 0 < p < 1. \tag{3.8''}$$

If we choose p such as $e^{2p} > 1-p$, then α > 1 and algorithm **REJ** is defined. With respect to p_a, **method 1** is preferable.

3.6 Simulation of the distribution derived from (1.15)
In this case, note that

$$a_n = \frac{1}{\beta}\frac{n!}{(n+k)!}, \beta = \frac{1}{(k-1)k!}, a_n = \frac{n!\,k!\,(k-1)}{(n+k)!}$$

We choose as enveloping distribution *the Yule*(k) *distribution* in the form

$$h_n = \frac{B(n,k)}{c(k)} \, , c(k) = \sum_{n=1}^{\infty} B(n,k). \tag{3.9}$$

The ratio r_n is

$$r_n = \frac{n! \, k! \, (k-1)}{(n+k)!} \frac{(n+k-1)! \, c(k)}{(n-1)! \, (k-1)!} = \frac{nk(k-1)c(k)}{n+k} \leq \frac{k(k-1)c(k)}{k+1}.$$

Therefore

$$\alpha = \frac{k(k-1)c(k)}{k+1} > 1. \tag{3.9'}$$

All elements of the algorithm **REJ** are defined. The probability p_a can be calculated numerically.

3.7 Simulation of the distribution derived from (1.16)

Method 1 (known). This is the *Zipf(a) Distribution*, defined in its general form as

$$p_n = \frac{1}{\zeta(a)} \frac{1}{n^a} \, , a > 1, \zeta(a) = \sum_{i=1}^{\infty} \frac{1}{i^a} \, , \tag{3.10}$$

where the $\zeta(a)$ is the ζ *Riemann function*. The formula (3.10) shows that $\zeta(2) \leq 2$, (See [1,8]). An algorithm for simulation of the random variable X as Zipf(a) is presented in [1,4]. It uses as enveloping distribution the distribution of a random variable Y such as

$$q_i = P(Y = i) = \frac{1}{(i+1)^a} \left[(1 + \left(1 + \frac{1}{i}\right)^{a-1} - 1 \right], i \in N+, i > 1$$

For which the cdf is

$$H(i) = 1 - \frac{1}{i^{a-1}},$$

And the inverse method gives

$$Y = int(U^{\frac{-1}{a-1}}), \tag{3.10'}$$

(where *int* denotes "integer part"). Note that ratio

$$r_i = \frac{p_i}{q_i} < \frac{p_1}{q_1} = \frac{2^{a-1}}{\zeta(a)(2^{a-1}-1)} = \alpha > 1. \tag{3.10''}$$

Therefore, the algorithm is

Step1: Take b = 2^(a-1);
Step2: **repeat**
Generate U,V uniforms on (0,1), independent;
Take Y = int $(U^{-\frac{1}{a-1}})$, T=$(1+\frac{1}{Y})^{a-1}$;

Until VY $\frac{T-1}{b-1} \leq \frac{T}{b}$;

Deliver X = Y.

X is the simulated value. The probability p_a can be easy approximated.

Method 2. Let us take as enveloping distribution, *the distribution derived from (1.10)* with the same a. Then we have

$$r_n = \frac{a}{(a-1)^2 \zeta(a)} \frac{a^n}{na^n} < \frac{a}{(a-1)^2 \zeta(a)} = \alpha, \alpha > 1 \tag{3.10'''}$$

and the algorithm **REJ** is terminated. To decide which of these methods is preferable, it is necessary to compare the p_a probabilities. The second method appears to be the best.

3.8 Simulation of the distribution derived from (1.17)

Method 1. The distribution is

$$p_n = \frac{1}{\beta}\frac{n^2}{n!}, \beta = 2 + 2e = 2(1+e).$$

We take as enveloping distribution the Zipf(2) distribution defined as

$$h_n = \frac{1}{\varsigma(2)n^2}, \varsigma(2) = \sum_{i=1}^{\infty}\frac{1}{i^2}, \tag{3.11}$$

where $\varsigma(2)$ is the Riemann function of the argument 2. Therefore

$$r_n = \frac{\varsigma(2)}{2(1+e)}\frac{n^4}{n!}.$$

It is shown by induction that

$$\frac{n^4}{n!} \leq 4^2.$$

Therefore

$$r_n \leq \frac{\varsigma(2)}{(1+e)} = \alpha > 1. \tag{3.11'}$$

The algorithm **REJ** is specified

Method 2. Let us take as enveloping distribution the Kemp distribution. Then, the ratio r_n becomes

$$r_n = -\frac{\log(1-p)}{2(1+e)}\frac{n^3}{n.n!p^n} \leq -\frac{\log(1-p)}{2p(1+e)} = \alpha. \tag{3.11''}$$

As parameter p is free, we can choose it as follows:

$$(1-p)e^{2p(1+e)} < 1, and\ then\ \alpha > 1. \tag{3.11'''}$$

Thus, algorithm **REJ** is terminated. To select the best method, the probabilities p_a must be estimated numerically.

3.9 Simulation of the distribution derived from (1.18)

Method 1. The distribution is

$$p_n = \frac{1}{5e}\frac{n^3}{n!}.$$

In this case we take again as enveloping distribution the *Zipf(2) distribution*, i.e.

$$h_n = \frac{1}{\varsigma(2)n^2}. \tag{3.12}$$

The ratio r_n in this case is $r_n = \frac{\varsigma(2)}{5e}\frac{n^5}{n!}.$

Here, again by induction, it is shown that

$$\frac{n^5}{n!} < 64,$$

and finally $\alpha = \frac{\varsigma(2).64}{5e} > 1.$ (3.12')

The algorithm **REJ** is obvious.

Method 2. Let us choose as enveloping distribution the Kemp distribution $0 < p < 1$,

i.e. $h_n = -\frac{1}{\log(1-p)}\frac{p^n}{n}.$ (3.12'')

The ratio r_n is

$$r_n = -\frac{\log(1-p)}{5e}\frac{n^4}{n!p^n} \leq -\frac{\log(1-p)}{p.5e} = \alpha = -\frac{\log(1-p)}{\log e^{5ep}}. \tag{3.12'''}$$

If we choose p as

$(1-p)e^{5ep} < 1,$ which can be done, then $\alpha > 1$, algorithm **REJ** is ready.

3.10 Simulation of the distribution derived from (1.19)

Method 1. In this case we have

$$p_n = \frac{q-p-1}{p+1} \frac{(p+1)...(p+n)}{(q+1)...(q+n)}, q-p > 1. \tag{3.13}$$

Let us take as enveloping distribution (1.10)

$$h_n = \frac{(a-1)^2}{a} \frac{n}{a^n}. \tag{3.13'}$$

Now the ratio $r_n = \frac{p_n}{h_n}$ is $r_n = \frac{(p+n)!}{(q+n)!} \frac{q!}{p!} \frac{a}{(a-1)^2} \frac{a^n}{n}$.

Since the function $\frac{a^x}{x} < 1$ for $x > int\left(\frac{1}{\log(a)+1}\right) = k$, we have

$$r_n \le \frac{(p+k)!}{(q+k)!} \frac{q!}{p!} \frac{a^2}{(a-1)^2} \frac{q-p-1}{p+1}.$$

If we now choose α as

$$\frac{(p+k)!}{(q+k)!(q-p-1)} \frac{a^2}{(a-1)^2} = \alpha, \tag{3.13''}$$

then the ratio becomes $r_n \le \alpha, \alpha \ge 1$, and the algorithm **REJ** is ready.

Method 2. Let us take as enveloping distribution the distribution of Kemp of the paramtere β, $0 < \beta < 1$, then we have

$$r_n = \frac{(p+n)!}{(q+n)!} \frac{q!}{p!} \frac{q-p-1}{p+q} \left[-\log(1-\beta)\frac{n}{\beta^n}\right].$$

Note that $\beta = e^{-\lambda}, \lambda > 0$, and then

$$\frac{n}{\beta^n} < \frac{n^e}{e^{-\lambda n}} < \frac{e^n}{e^{-\lambda n}}.$$

Finally, one obtain $r_n < \frac{p+1}{q+1} \frac{q-p-1}{p+q} [-\log(1-\beta)]e^{1+\lambda} = \alpha > 1.$ (3.13''')

i.e. the algorithm **REJ** is specified. Here again the probabilities p_a will show which method is preferable.

4. ADDENDA: SIMULATION OF USED DISTRIBUTIONS

The simulation of discrete distributions mentioned in the formulas (1.6)-(1.9) will be presented in the following.

4.1 Simulation of logarithmic series of the parameter p

This distribution is $p_n = -\log(1-p)\frac{p^n}{n}, n \ge 1, 0 < p < 1.$ (4.1)

Method 1. One method for simulating this distribution consists in the fact that the random variable X having this distribution is a mixture (see [5]) of the random variable Y with the cdf

$$F(y) = \frac{\log(1-y)}{\log(1-p)}, 0 \le y \le p. \tag{4.2}$$

with the *Geometric*(y) distribution. Therefore, the algorithm is *Generate a random variate y by the inverse method,* i.e solve the equation

$$\frac{\log(1-y)}{\log(1-\alpha)} = U; \tag{4.2'}$$

Generate X as Geometric(y).
Deliver X.
In [5], the inverse method for simulating X is presented, (i.e. the solution of (4.2')).

Method 2. Let us take as enveloping distribution the one deriving from (1.10).

Then,

$$r_n = -\log(1-p)\frac{p^n}{n}\frac{a}{(a-1)^2}\frac{a^n}{n} = -\log(1-p)\frac{a}{(a-1)^2}\frac{(pa)^n}{n^2} \leq -\log(1-p)\frac{a}{(a-1)^2}pa,$$

if $pa < 1$

In this case $r_n \leq -\log(1-p)\frac{a^2p}{(a-1)^2} = \alpha$,

which gives $\alpha > 1$ *if* $1 < a < 2$ and the algorithm **REJ** is defined. It seems difficult to compare these methods, if not by means of computer tests.

4.2 Simulation of $Zipf(a), a > 1$ distribution

Method 1. In this case

$$p_n = \frac{1}{\varsigma(a)n^a}, n \geq 1, \varsigma(a) = \sum_{n=1}^{\infty}\frac{1}{n^a}, \tag{4.3}$$

where $\varsigma(a)$ is the *Riemann function*. In [1,5] a *rejection method* which uses the enveloping distribution is presented:

$$q_n = p(Y = n) = \frac{1}{(n+1)^{a-1}}\left[\left(1+\frac{1}{n}\right)^{a-1} - 1\right], n \geq 1. \tag{4.4}$$

By calculating the ratio $r_n = \frac{p_n}{q_n}$, it results that $r_n = \frac{p_n}{q_n} \leq \frac{p_1}{q_1} = \frac{2^{a-1}}{\varsigma(a)(2^{a-1}-1)} = \alpha > 1.$ (4.5)

In [1] it is shown that

$$\alpha \leq \frac{12}{\pi^2} \; if \; a \geq 2 \; and \; \alpha \leq \frac{2}{\log(2)}, if \; 1 < a < 2. \tag{4.5'}$$

There are some remarks to be made regarding this distribution.

(1). If $1 \leq n \leq K^* < \infty$, it is used to represent random events, such as number of occupied cells of a computer memory of size $K *$, when memory is dynamically allocated;.

(2). For a finite $K *$ this distribution describe the random occurrence of words in a text of a given length (natural language).

Method 2, (New method). Let us select as enveloping function the h_n as the *Kemp* distribution. Then

$$h_n = -\frac{1}{\log(1-p)}\frac{p^n}{n}, 0 < p < 1 \; ,$$

and hence

$$r_n = \frac{p_n}{h_n} = -\frac{\log(1-p)}{\varsigma(a)}\frac{n}{n^a p^n} \leq -\frac{\log(1-p)}{p} = \alpha. \tag{4.5''}$$

We can choose the parameter p such as $\alpha > 1$ and the algorithm **REJ** is obvious. A more relevant comparison between these methods could be done by means of computer tests.

4.3 Simulation of Euler distribution

Method 1.(Known). This distribution is

$$p_n = \left[\frac{1}{n} - \log\left(1+\frac{1}{n}\right)\frac{1}{\gamma}\right], \gamma = \lim_{n\to\infty}(\sum_{k=1}^{n}\frac{1}{k} - \log(n)), \tag{4.6}$$

where γ is *the constant of Euler*. In this case, we are using a rejection method based on enveloping p_n, with the *distribution of Logarithmic series of parameter p, $0<p<1$*. The ratio is

$$r_n = -\frac{1}{\gamma \log(1-p)}(\frac{1}{n} - \log\left(1+\frac{1}{n}\right)\frac{n}{p^n} = -\frac{1}{\gamma \log(1-p)}$$

$$= 1 - \frac{\log\left(1+\frac{1}{n}\right)^n}{p^n} \leq -\frac{1}{\log(1-p)}\frac{1-\log(2)}{p} = \alpha. \tag{4.6'}$$

If we choose $p, 0 < p < 1$ such as

$$e > 2 + \left(\frac{1}{1-p}\right)^p$$

then $\alpha > 1$. The elements of the algorithm **REJ** are defined.

Method 2.(New) Let us take as enveloping distribution the *Geometric(q)*, where

$$q = \max\left(\frac{1}{n} - \log\left(1 + \frac{1}{n}\right) = \log\left(\frac{e}{2}\right) = q < 1.\right.$$

Since $= \log(2)$, we have

$$r_n \leq \frac{\log\left(\frac{e}{2}\right)}{\log(2)\,\gamma\left(\log\left(\frac{e}{2}\right)\right)} = \frac{1}{\log(2)\,\gamma} = \alpha > 1,$$

and **REJ** is defined. Here again, the comparison of methods could be done via computer tests.

4.4 Simulation of $Yule(a)$ distribution

The simulation is based on the following judgment: *The Yule(a) distribution is the mixture of the Geometric(p) distribution with*

$$p = e^{-\frac{Y}{a-1}}$$

and Exp(1) distribution of Y. This results in the following algorithm:

1. *Generate E and Exp(1) random variate. (i.e. Generate U uniform (0,1) and take E=-log(u)), U>0.*

2. *Generate $E^* \mapsto Exp(1)$, independent from E;*

3. *Calculate*

$$X = \text{int}\left\{\frac{E}{\log\left(1 - e^{-\frac{E^*}{a-1}}\right)}\right\} + 1.$$

The X variable is the required Yule(a) variable.

In [1], it is specified that the Yule distribution is a better approximation of word frequencies (in a natural language) than the Zipf distribution.

Comments. Computer tests were not performed yet. They could be performed following the hints in [10]. This could make a good exercise for an M.Sc. student. Such an exercise could be could be useful for comparing various methods of simulation for each distribution. The **inverse** algorithms must be first considered to assess the performance degree of these methods.

REFERENCES

[1] Luc Devroye, (1986).*Non Uniform Random Variate Generation*, Springer Verlag, New York, Berlin.

[2] I. B. Gerstbakh, (1989). *Statistical Reliability Theory*, Marcel Dekker, Inc., New York, London.

[3] N. L. Johnson and S.Kotz, (1972). Distributions in Statistics: Discrete univariate distributions,John Wiley and Sons, New York, London.

[4] A.W. Kemp, (1981). *Efficient generation of logarithmically distributed pseudo-random variables*, Appl. Statistics, 30(3), p.249-253.

[5] J.T. Byron Morgan (1984).*Elements of Simulation*, Chapman and Hall, London, New York.

[6] B. D. Ripley (1987). *Stochastic Simulation*, John Wiley and Sons, New York.

[7] Sheldom R.M. (1997). *Simulation*. Second Edition, Academic Press, San Diego, New York, London.

[8] Ghe. Siret (1985). *Calcul Diferential si Integral*, Vol.1 Editura Stiintica si Enciclopedica. Bucuresti.

[9] I. Vaduva, (1977). *Simulation Models with Computers,* (Romanian), Ed. Tehnica, Bucharest.

[10] I. Vaduva, (2011).*On Simulation of some Mixed Life Distributions*. Analele Univ. Bucuresti, Seria Informatica,Anul 2011, p.10-19.

THE RAYLEIGH'S FAMILY OF DISTRIBUTIONS

Bogdan Gheorghe MUNTEANU

"Henri Coandă" Air Force Academy, Braşov, Romania (munteanu.b@afahc.ro)

Abstract: Several general mathematical properties of the Rayleigh's family of distributions are examined in a consistent manner by using the power series distributions (PSD) class [1]. A new cumulative distribution function and probability density are obtained for the continuous type random variables which represent the maximum or the minimum in a sequence of independent, identically Rayleigh distributed random variables, in a random number by means of a power series distribution. An asymptotic result characterized by the Poisson Limit Theorem is formulated and analysed.

Keywords: power series distributions, Rayleigh distribution, distribution of the maximum and minimum, Poisson Limit Theorem

1. INTRODUCTION

In the paper [2] sets out to introduce and analyse the properties of the maximum and minimum distributions for a sample of power series distribution. This can serve as a mathematical model to describe the probabilistic behaviour of the signals used on a large scale in the field of radiolocation. In this paper, the distribution is presented as being the distribution of the maximum or minimum value from a sample of random volume Z from a Rayleigh distributed statistical population, where Z is a random value from the power series distribution class.

2. MIN RAYLEIGH AND MAX RAYLEIGH POWER SERIES DISTRIBUTIONS

It is know that a random variable admits a Rayleigh distribution with the parameter σ, and we note $X : Rayleigh(\sigma), \sigma > 0$, if the cumulative distribution function (cdf) is

$$F_{Ray}(x) = 1 - e^{-\frac{x^2}{2\sigma^2}}, x \geq 0,$$ while the corresponding probability density function (pdf)

$$f_{Ray}(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, x \geq 0.$$

We consider the random variables $U_{Ray} = \max\{X_1, X_2, ..., X_Z\}$ and $V_{Ray} = \min\{X_1, X_2, ..., X_Z\}$, where $(X_i)_{i \geq 1}$ are independent and identically distributed random variables, $X_i : Rayleigh(\sigma), \sigma > 0$ and $Z \in PSD$, that is $P(Z = z) = \frac{a_z \theta^z}{A(\theta)}, z = 1,2,...,$ where $a_1, a_2,...$ is a sequence of real, non-negative numbers. $\tau > 0$ the radius of convergence of the power series $A(\Theta) = \sum_{z \geq 1} a_z \Theta^z, \forall \Theta \in (0, \tau)$ and Θ the real parameter of the distribution.

We point out that the random variables $(X_i)_{i \geq 1}$ are independent of the random variable Z, the latter's distribution being part of the power series distributions class [1].

In accordance with the working methods in the paper [2,4], it can be stated that the random variables U_{Ray} follow the *Max Rayleigh power series distributions* of parameters σ and Θ (we note: U_{Ray} : *MaxRayleighPS* (σ, Θ)) and V_{Ray} follow the *Min Rayleigh power series distributions* of parameters σ and Θ (we note: V_{Ray} : *MinRayleighPS* (σ, Θ)) if the cumulative distribution functions (cdf) are characterized by the relation:

$$U_{Ray}(x) = \frac{A(\Theta F_{Ray}(x))}{A(\Theta)} = \frac{A\left[\Theta\left(1 - e^{-\frac{x^2}{2\sigma^2}}\right)\right]}{A(\Theta)}, \, x \geq 0 \tag{1}$$

and

$$V_{Ray}(x) = \frac{A(\Theta) - A\left[\Theta\left(1 - F_{Ray}(x)\right)\right]}{A(\Theta)} = \frac{A(\Theta) - A\left[\Theta e^{-\frac{x^2}{2\sigma^2}}\right]}{A(\Theta)}, \, x \geq 0. \tag{2}$$

The probability densities functions (pdf) are characterized by the relation:

$$u_{Ray}(x) = \frac{\Theta \cdot f_{Ray}(x) \cdot \frac{d}{dx}\left[A(\Theta F_{Ray}(x))\right]}{A(\Theta)} = \frac{\Theta x e^{-\frac{x^2}{2\sigma^2}} \frac{d}{dx}\left\{A\left[\Theta\left(1 - e^{-\frac{x^2}{2\sigma^2}}\right)\right]\right\}}{\sigma^2 A(\Theta)}, \, x \geq 0, \tag{3}$$

and,

$$v_{Ray}(x) = \frac{\Theta \cdot f_{Ray}(x) \cdot \frac{d}{dx}\left\{A\left[\Theta\left(1 - F_{Ray}(x)\right)\right]\right\}}{A(\Theta)} = \frac{\Theta \cdot x \cdot e^{-\frac{x^2}{2\sigma^2}} \cdot \frac{d}{dx}\left[A\left(\Theta e^{-\frac{x^2}{2\sigma^2}}\right)\right]}{\sigma^2 A(\Theta)}, \, x \geq 0. \tag{4}$$

Proposition 2.1. If $(X_i)_{i \geq 1}$ is a sequence of independent random variables, Rayleigh distributed with parameters $\sigma > 0$, while $U_{Ray} = \max\{X_1, X_2, ..., X_Z\}$ where $Z \in PSD$ with $P(Z = z) = \frac{a_z \theta^z}{A(\theta)}, z = 1,2,...,$ $A(\Theta) = \sum_{z \geq 1} a_z \Theta^z$, $\forall \Theta \in (0, \tau)$, then:

$$\lim_{\Theta \to 0^+} U_{Ray}(x) = \left[1 - e^{-\frac{x^2}{2\sigma^2}}\right]^k, \, x \geq 0, \quad \text{where } k = \min\{k \in N^*, a_k > 0\}.$$

Proposition 2.2. If $(X_i)_{i \geq 1}$ is a sequence of independent random variables, Rayleigh distributed with parameters $\sigma > 0$, while $V_{Ray} = \max\{X_1, X_2, ..., X_Z\}$ where $Z \in PSD$ with $P(Z = z) = \frac{a_z \theta^z}{A(\theta)}, z = 1,2,...,$ $A(\Theta) = \sum_{z \geq 1} a_z \Theta^z$, $\forall \Theta \in (0, \tau)$, then:

$$\lim_{\Theta \to 0^+} V_{Ray}(x) = 1 - e^{-\frac{x^2 l}{2\sigma^2}}, \, x \geq 0, \text{ where } l = \min\{l \in N^*, a_l > 0\}.$$

Corollary 2.1. The rth moments, $r \in N$, $r \geq 1$ of the random variables U_{Ray} : $MaxRayleighPS(\sigma, \Theta)$ and V_{Ray} : $MinRayleighPS(\sigma, \Theta)$ are given by:

$$EU_{Ray}^r = \sum_{z \geq 1} \frac{a_z \Theta^z}{A(\Theta)} E\left\{ \max\left[X_1, X_2, ..., X_z \right] \right\}^r$$

and

$$EV_{Ray}^r = \sum_{z \geq 1} \frac{a_z \Theta^z}{A(\Theta)} E\left\{ \min\left[X_1, X_2, ..., X_z \right] \right\}^r ,$$

where the pdfs of the random variables $\max\left[X_1, X_2, ..., X_z \right]$ and $\min\left[X_1, X_2, ..., X_z \right]$ are $f_{\max\left[X_1, X_2, ..., X_z \right]}(x) = z\, f_{Ray}(x)\left[F_{Ray}(x) \right]^{z-1}$ and $f_{\min\left[X_1, X_2, ..., X_z \right]}(x) = z\, f_{Ray}(x)\left[1 - F_{Ray}(x) \right]^{z-1}$.

3. SPECIAL CASES

3.1. The Max Rayleigh Binomial and Min Rayleigh Binomial distributions. The Max Rayleigh Binomial (MaxRayB) and Min Rayleigh Poisson (MinRayP) distributions are defined by the distribution functions presented in a general framework in [2], where $A(\Theta) = (\Theta + 1)^n - 1$, with $\Theta = \dfrac{p}{1-p}$, $p \in (0,1)$:

$$
U_{RayB}(x) = \frac{A\left(\Theta F_{Ray}(x) \right)}{A(\Theta)} = \frac{\left[\Theta\left(1 - e^{-\frac{x^2}{2\sigma^2}} \right) + 1 \right]^n - 1}{(\Theta+1)^n - 1}
$$

$$
= \frac{\left(1 - p \cdot e^{-\frac{x^2}{2\sigma^2}} \right)^n - (1-p)^n}{1 - (1-p)^n}, \; x \geq 0
$$

(5)

and

$$
V_{RayB}(x) = 1 - \frac{A\left[\Theta\left(1 - F_{Ray}(x) \right) \right]}{A(\Theta)} = \frac{(\Theta+1)^n - \left(\Theta e^{-\frac{x^2}{2\sigma^2}} + 1 \right)^n}{(\Theta+1)^n - 1}
$$

$$
= \frac{1 - \left(1 - p + p \cdot e^{-\frac{x^2}{2\sigma^2}} \right)^n}{1 - (1-p)^n}, \; x \geq 0.
$$

(6)

respectively.

3.2. The Max Rayleigh Poisson and Min Rayleigh Poisson distributions. The Max Rayleigh Poisson (MaxRayP) and Min Rayleigh Poisson (MinRayP) distributions are characterized by the cumulative distributions functions defined by the relations (1) and (2), where $A(\Theta^*) = e^{\Theta^*} - 1$, with $\Theta^* = \lambda$, $\lambda > 0$:

$$
U_{RayP}(x) = \frac{A\left(\Theta^* F_{Ray}(x) \right)}{A(\Theta^*)} = \frac{e^{\Theta^* F_{Ray}(x)} - 1}{e^{\Theta^*} - 1}
$$

(7)

$$
= \frac{e^{-\lambda e^{-\frac{x^2}{2\sigma^2}}} - e^{-\lambda}}{1 - e^{-\lambda}}, \; x \geq 0
$$

and

$$V_{RayP}(x) = 1 - \frac{A\left[\Theta^*\left(1 - F_{Ray}(x)\right)\right]}{A(\Theta^*)} = \frac{1 - e^{-\lambda\left(1 - e^{-\frac{x^2}{2\sigma^2}}\right)}}{1 - e^{-\lambda}}, \ x \geq 0. \tag{8}$$

3.3. On the Poisson limit theorem. The following theorems show that the MaxRayP and MinRayP distributions approximate the MaxRayB and MinRayB distributions depending on certain conditions.

Theorem 3.1. (Poisson limit theorem). *The MaxRayP and MinRayP distributions can be obtained as the limit of the MaxRayB, respectively MinRayB distributions with distribution functions given by (5) and (6) if $n \cdot \Theta \to \lambda$ when $n \to \infty$ and $\Theta \to 0^+$.*

In other words,

(i) $\lim\limits_{\substack{n\to\infty \\ p\to 0^+}} V_{RayB}(x) = V_{RayP}(x), \ \forall x \geq 0$ *, where $V_{RayB}(x)$ and $V_{RayP}(x), x \geq 0$ are the*

distribution functions of the random variables V_{RayB} : $MinRayleighB(\sigma, n, p)$ and

V_{RayP} : $MinRayleighPoi(\sigma, \lambda)$ *defined by (6) and (8);*

(ii) $\lim\limits_{\substack{n\to\infty \\ p\to 0^+}} U_{RayB}(x) = U_{RayP}(x), \ \forall x \geq 0,$ *where $U_{RayB}(x)$ and $U_{RayP}(x), x \geq 0$ are t*

distribution functions of the random variables U_{RayB} : $MaxRayleighB(\sigma, n, p)$ and

U_{RayP} : $MaxRayleighPoi(\sigma, \lambda)$ *defined by (5) and (7).*

Proof. We examine the convergence in terms of the maximum distributions $U_{RayB}(x)$ and $U_{RayP}(x)$, $x \geq 0$.

It is evident that:

$$\lim\limits_{\substack{n\to\infty \\ p\to 0^+}} (1-p)^n = \lim\limits_{\substack{n\to\infty \\ p\to 0^+}} \left[(1-p)^{-1/p}\right]^{-np} = e^{-\lambda};$$

$$\lim\limits_{\substack{n\to\infty \\ p\to 0^+}} (1 - p \cdot e^{-x^2/2\sigma^2})^n = \lim\limits_{\substack{n\to\infty \\ p\to 0^+}} \left[\left(1 - p \cdot e^{-x^2/2\sigma^2}\right)^{-1/p \cdot e^{-x^2/2\sigma^2}}\right]^{-np \cdot e^{-x^2/2\sigma^2}}$$

$$= e^{-\lim\limits_{\substack{n\to\infty \\ p\to 0^+}} np \cdot e^{-x^2/2\sigma^2}} = e^{-\lambda \cdot e^{-x^2/2\sigma^2}}.$$

$$\lim\limits_{\substack{n\to\infty \\ p\to 0^+}} U_{RayB}(x) = \lim\limits_{\substack{n\to\infty \\ p\to 0^+}} \frac{\left(1 - pe^{-x^2/2\sigma^2}\right)^n - (1-p)^n}{1 - (1-p)^n} = \frac{e^{-\lambda \cdot e^{-x^2/2\sigma^2}} - e^{-\lambda}}{1 - e^{-\lambda}} = U_{RayP}(x), \ x \geq 0.$$

Fig. 1 and 2 show the behaviour of the pdfs of *MinRayleighB*(σ, n, p), *MinRayleighPoi* (σ, λ), *MaxRayleighB*(σ, n, p) and *MaxRayleighPoi*(σ, λ) for some values of the parameters: $n = 40$, $p = \dfrac{1}{10}$, $\lambda = 4$, $\sigma = 5$.

Fig. 3 and 4 show the behaviour of the pdfs of *MinRayleighB*(σ, n, p), *MinRayleighPoi* (σ, λ), *MaxRayleighB*(σ, n, p) and *MaxRayleighPoi*(σ, λ) for some values of the parameters.
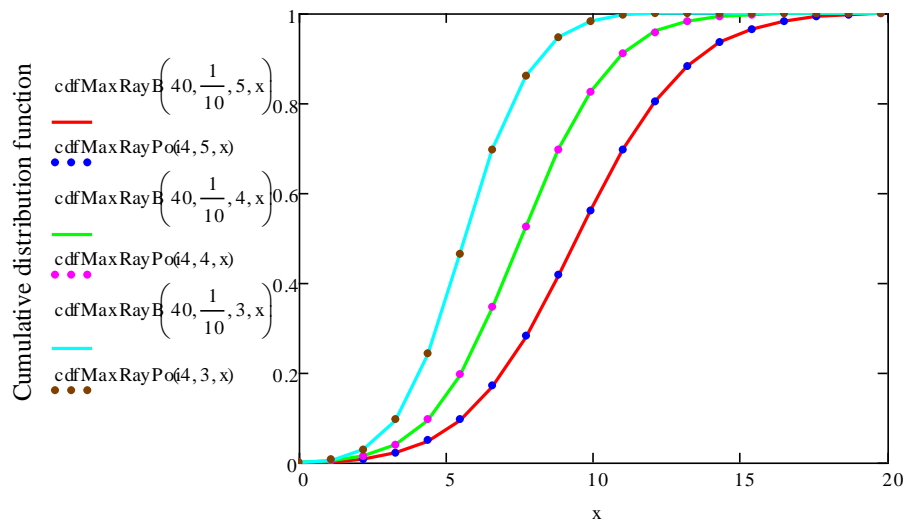
Fig. 1: Pdfs for the Max-Rayleigh-Binomial and Max-Rayleigh-Poisson distributions – graphical illustration of the Poisson Limit Theorem
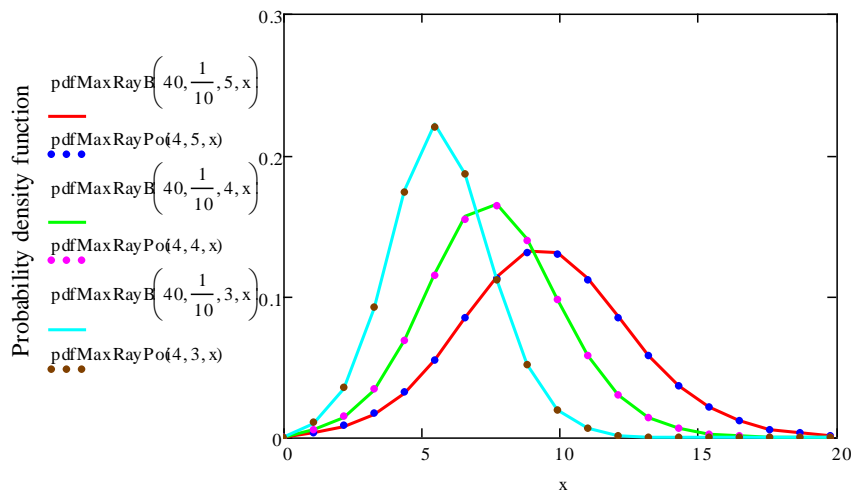
Fig. 2: Pdfs for the Max-Rayleigh-Binomial and Max-Rayleigh-Poisson distributions – graphical illustration of the Poisson Limit Theorem
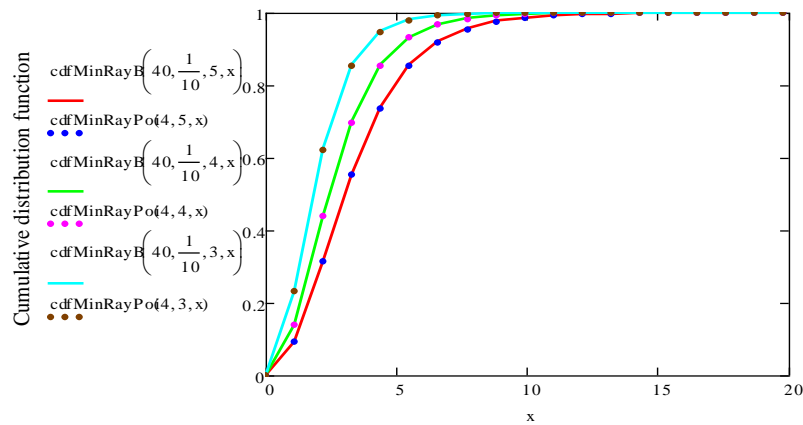
Fig. 3: Pdfs for the Min-Rayleigh-Binomial and Min-Rayleigh-Poisson distributions – graphical illustration of the Poisson Limit Theorem
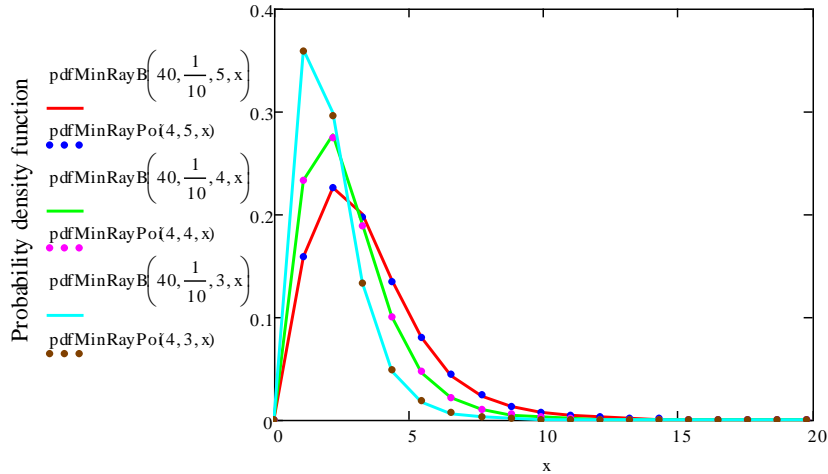
Fig. 4: Pdfs for the Min-Rayleigh-Binomial and Min-Rayleigh-Poisson distributions – graphical illustration of the Poisson Limit Theorem

CONCLUSIONS

The results formulated and examined in this paper are in connection with the study of the random variable distribution, which can be expressed as being the maximum or minimum of a sequence of independent random variables identically distributed in a random number. In practice, this translates as the emission and reception of some signals is a random number, signals whose amplitude is a random variable characterized by the Rayleigh distribution [3]. The signals that best record either a maximum or minimum amplitude are of special interest to us.

It has, thus, been presented in a consistent manner how to determine the maximum and minimum distribution of independent and identically distributed random variables, which form a random sequence.

The Poisson Limit Theorem has been formulated when the random variable number in a sequence has a zero truncated binomial distribution and the limit distribution of the minimum and maximum is of Poisson type.

REFERENCES

[1] N.L. Johnson , A. W. Kemp, S. Kotz, *Univariate Discrete Distribution*, New Jersey, 2005;

[2] A. Leahu, B. Gh. Munteanu, S. Cataranciuc, On the lifetime as the maximum or minimum of the sample with power series distributed size, *Romai J.*, vol. 9, no. 2, pp. 119-128, 2013;

[3] B. R. Mahafza, *Radar systems analysis and design using Matlab*, CRC Press Taylor & Francis Group, USA, 2013;

[4] B. Gh. Munteanu, The Min-Pareto power series distributions of lifetime, *Appl. Math. Inf. Sci.*, vol. 10, no. 5, pp. 1673-1679, 2016;

[5] B. Gh. Munteanu, Qualitative aspects of the Min Pareto binomial distribution, *Review of the Air Force Academy*, vol. XV, no.2(34), pp. 63-68, 2017.

THE GENETIC ALGORITHM FOR SOLVING THE NON-LINEAR TRANSPORTATION PROBLEM

Tatiana PAȘA

Moldova State University, Chişinău, Republic of Moldova
(pasa.tatiana@yahoo.co)

Abstract: The paper examines the transportation problem on a network with concave cost functions and presents a genetic algorithm implemented in Wolfram Language which is able to solves this problem. We experimentally prove that the algorithm is stable, effective and converges to an optimum solution. The results are compared to the ones obtained by means of the standard functions of Wolfram Language.

Keywords: transportation problem, concave cost function, genetic algorithm, population, crossover operation, mutation operation.

1. INTRODUCTION

The transport problem appeared when the need to solve a group of economic problems arose. These problems consist in finding the smallest transportation costs for the products as they travel towards their destination. A classical variant of the transportation problem is a linear programming problem, and a solution was proposed in 1947 by Dj. Dantzing [10] that solves it through the simplex method. It consists in determining the optimum transport plan for the required amount of a single product from a source to the destination, that minimizes the transportation cost. This model of the problem is used to optimize the supply of businesses with raw materials, the supply of stores with products from wholesalers and the design of telecommunication, water, gas or petrol networks. Several authors, D. R. Fulkerson [19], R. G. Busacker şi P. J. Goven [3], J. Edmonds and R. M. Karp [14], N. Tomizawa [40], M. Klein [27], D. D. Sleator and R. E. Tarjan [37], A. V. Goldberg and R. E. Tarjan [20, 21], P. T. Sokkalingam, R. K. Ahuja and J. B. Orlin [38], I-L.Wang, S-J. Lin [41], L. Ciupală [8], J. M. Davis and D. P. Williamson [11], P. Kovacs [28], A. Sifaleras [36], S. Ding [13], M. Dawuni and K. F. Darkwah [12], N. A. El-Sherbeny [15], M. B. Cohen, A. Madry, P. Sankowski, and A. Vladu [9], A. M. P. Chandrasiri and D. M. Samarathunge [4], R. A. Maher and F. A. Abdula [3], J. Erickson, K. Fox and L. Lkamsuren [16], S. Abdi, F. Baroughi and B. Alizadeh [1] have proposed solutions to the network transportation problem with linear cost functions, and also with uncertain cost and/or capacities for each edge in the network. They have also presented a theoretical and practical analysis of the algorithms, fitted with conclusions and recommendations.

When the network has non-linear cost functions, the problem becomes more complicated. In this case, there are several concepts that are widely used to solve it: the optimality conditions Kuhn-Tucker and the Lagrange multipliers, the first order derivatives (gradients), the second order derivatives (Hessian matrix), and also the penalty functions [39], [30].

The usage of these techniques is much more difficult in networks with concave cost functions, because there are multiple local minima and it can lead to only getting a single local optima. This type of problems have been studied in detail by R. Horst and P. M. Pardalos [24], R. Horst and H. Tuy [25], Q. He, A. Shabbir and G. L. Nemhauser [22].

We are especially interested in problems that describe real life situations which are typical for a modern economy. As this kind of models are extremely complex, our preoccupation resulted in genetic algorithms proposed to solve them. They have been described for the first time under the leadership of J. Holland [23] at the University of Michigan. The name "genetic" derives from the notions surrounding them, e.g. population, chromosomes, genes, selection, mutation, crossover. These are stochastic and heuristic algorithms, which mean that the obtained solutions is are not always optimal, but they come close to the optima. In general, these algorithms are polynomial and are used to solve complex problems. A comparative analysis of the genetic algorithms applied to optimization problems can be found in [33] and an overview of the algorithms is given in [29].

These algorithms are recommended because: one doesn't have to know gradients and Hessian matrices; they can't get stuck on local optima and still work very well on big problems with a many variables. Several authors have used genetic algorithms to solve transport network problems: D.B.M.M. Fontes and J.F. Gonçalves [18], A. Sadegheih and P. R. Drake [35]. The implementation of the algorithm is often slowed down by the need to use auxiliary variables and the computation of the fitness function many times.

Genetic algorithms have been proposed to solve micro circulation problems. Some of them are: the usage of adjacent roads to reduce the traffic from main roads [6], the minimization of the time spent before a plane can land [26] and the coordination of several urban bus routes [42].

The genetic algorithm described in [5] can plan efficiently the high-speed train stations. In [44] is present an algorithm that minimizes costs and streamlines the metro activity and [7] describes an algorithm that improves significantly the security and efficiency of corridors with multiple-rail grade crossings. Another genetic algorithm proposed in [2] optimizes the routes of the airport buses by minimizing the transfer time of the passengers. The problem of planning the deposit spaces in ports is studied in [43] with the purpose of minimizing taxes.

2. PROBLEM FORMULATION

We consider the transport network problem described by the convex graph $G = (V, E)$, $|V| = n$, $|E| = m$. The real production and consumption function $q = V \rightarrow R$ is defined on the finite set of vertices V. The concave non-decreasing functions of cost $\varphi_e(x_e)$ are defined on the edges E. To solve this non-linear optimization problem we must find a flow x^* that minimizes the function $F(x) = \sum_{e \in E} \varphi_e(x_e)$, thus satisfying the conditions:

$$F(x^*) = \min_{x \in X} F(x)$$

$$\sum_{e \in E^+(v)} x(e) - \sum_{e \in E^-(v)} x(e) = \begin{cases} -p(v), & v = v_0 \\ 0, & V/\{v_0, v_t\} \\ p(v), & v = v_t \end{cases}$$

$$x(e) \geq 0, for\ every\ e \in E.$$

X is the set of possible solutions, that satisfies the system of equations and the positivity restrictions, $E^-(v) = \{(v, u) | (v, u) \in E\}$, $E^+(v) = \{(u, v) | (u, v) \in E\}$.

We will consider the problem in which any quantity can be transported through an edge, the costs being described by a concave function. We also consider that in no intermediary point can the quantity of goods increase or decrease, thus the flow conservation condition being satisfied. Which means that no intermediary points consume or produce flow. Then, the problem requires the minimization of transportation costs from the source to intermediary centers and then to the destination.

In such situations, in which we have to check all admissible solutions, which can only be computed in a long time, the genetic algorithm is a solution that can provide the answer in a reasonable amount of time. Genetic algorithms are based on the theorem of templates [34]. A template H is defined as a pattern that describes a subset of chromosomes with similar genetic sections. Schemes have two properties:

• the degree of a template H, denoted by $o(H)$ – the number of fixed positions in a template;

• the definition length of a template H, denoted by $\delta(H)$ – the distance between the first and last position of the string of genes.

The first step in using a genetic algorithm is deciding how to encode the problem, how to describe the chromosomes as admissible solutions. The most often used is binary encoding, but it can also be numerical, symbolic and character-based depending on the problem. A population consists of chromosomes, which is actually a set of admissible solutions.

When creating a population, we must keep in mind that:

– the chromosomes will have a constant length;

– the number of chromosomes in a population is constant;

– every population $P(i + 1)$ is created from only the offspring of the population $P(i)$ or parents and offspring.

We will consider the elements of a population of size $4n$ (or $2n$) represented by numerical strings of length n constructed on the numerical alphabet $\mathcal{A} = \{0, 1, 2, \dots, n - 1\}$. The population will evolve to better solutions using *selection*, *mutation* and *crossover*. The value of the fitness function will have smaller and smaller values which means that the chromosomes will be closer to the minimum solution. The elitist selection is preferred in a genetic algorithm, because it guarantees that promising chromosomes will not be lost.

There are several ways in which we can make sure that we keep the best solution in a population. When applying *selection* at a step k, we will decide which chromosomes will participate in creating a new population. There are several methods of selection:

- the probability of choosing a particular chromosome depends on the value of –its fitness function;

- chromosomes are sorted in ascending order based on their fitness function and the probability depends on their position in the sorted list;

- for each 2 randomly chosen chromosomes, we take the one with a smaller fitness function.

For our problem, we will sort the chromosomes in increasing order and take the first half, thus we will never lose a good solution, even if it appears in the first population. These chromosomes will form the first half of the next population and will be parents to the offspring. Even if we aren't guaranteed to get better solutions this way, the chance is better than when choosing random parents and the offspring will be at least as good as the parents.

The other half of the population will be obtained through the **crossover** of the previously selected chromosomes. A pair of offspring will be created by taking two adjacent chromosomes and making a random cut. The first offspring will be formed from the left part of the mother chromosome and the right part of the father chromosome; the second one will be formed from the left part of the father chromosome and the right part of the father chromosome.

So if we have:

mother $(a_1, a_2, ..., a_n)$
father $(b_1, b_2, ..., b_n)$

then the offspring will be:

offspring1: $(a_1, a_2, ..., a_k, b_{k+1}, ..., b_n)$
offspring2: $(b_1, b_2, ..., b_k, a_{k+1}, ..., a_n)$

We can also use more cuts or create offspring from several chromosomes.

The random modification of a̶ the gene of a chromosome, also called **mutation**, has a mixed effect, it can improve or worsen the solution. The mutation will take place at a small rate of ε, e.g. $\varepsilon = 0.001$, so that we may avoid losing n̶o̶t̶ ̶l̶o̶s̶e̶ good solutions, but enough to produce new solutions in order to elude a local minima. Mutation will generate a random value to a randomly chosen gene.

We must also define the stop condition. Usually, we run the algorithm until k populations are created and return the best solution in it. But in this case we may get to a point where subsequent populations are the same. To avoid such a situation, we can stop breeding new populations when the condition $|f(i-1) - f(i)| \leq \varepsilon$ is satisfied for the best solutions in two consecutive populations $P(i-1)$ and $P(i)$.

Based on what we described, a genetic algorithm completes several steps and at the end an optimum solution, it results in [32], [17]:

Step 1. Generation of the initial population;
Step 2. Evaluation of the fitness function for each chromosome of the population;
Step 3. Selection of the chromosomes so that we don't lose any good solutions;
Step 4. Crossover of the selected chromosomes to create offspring with a fitness function at least as good as that of the parents;
Step 5. Mutation of a gene of a chromosome at a rate of ε;
Step 6. Test of the stop condition; if it is satisfied, then we STOP; if not, we go back to Step 2.

3. DESCRIPTION OF THE ALGORITHM

The genetic algorithm starts with a random population of chromosomes, each individual chromosome being an admissible solution to the transport network problem. Every chromosome will have length n. Using *selection*, *crossover* and *mutation* we will improve the population and obtain a smaller value of the fitness function.

The algorithm generates in each population chromosomes at least as good as in the previous population, because at every step we select only the chromosomes with the lowest objective function value and the rest of the population is filled with their offspring.

The genetic algorithm P1 proposed to solve the transport problem with non-linear concave functions consists of the following steps:

1. *Initialization.* The initial population is generated in the following way: a string of n random natural numbers will be generated, such that the first position will have a number nr_i between 1 and $n-1$ (from the first vertex called *source* there is at least one outgoing edge and at most $n-1$ edges), on every position $i = \overline{2, n-1}$ a number nr_i between 0 and $n-i$ will be generated, and the last position will have the number $nr_i = 0$ (there are no outgoing edges from the last vertex, the *destination*). This string will be one of the chromosomes in the population.

The population will have $4n$ chromosomes. We will also randomly generate a matrix $R = \{r_{ij} | i = \overline{1,n}, j = \overline{1,n}\}$ that shows how much of the flow from vertex i will go through edge (i, j). For this matrix $r_{ij} = 0$ only if the edge (i, j) doesn't exist and $\sum_{i=1}^{n} \sum_{j=1}^{n} r_{ij} = 1$.

2. *Evaluation* of the chromosomes from the current population means evaluating the objective function of each individual.

3. *Selection* of the parent chromosomes that will participate in the crossover is done so that their objective functions are the smallest possible. The chromosomes will be sorted in the ascending order of the objective function value. The first half of the new population will be formed from these chromosomes.

4. *Crossover* of the chromosomes is realized between the previously selected chromosomes to form the second half of the population. We will cut randomly each parent in the same place, and combine these parts as described earlier to create two new offspring. This way, each pair will have two offspring and the size of the population will be constant.

5. *Mutation* of a single gene of a chromosome will be done at a rate of $\varepsilon = 0,01 - 0,0001$, by generating a new random value for a gene ~~of~~ in a chromosome.

6. *Testing the stop condition* can be done in a few ways:
a) after creating k populations;
b) a time restriction for very serious problems;
c) stopping the algorithm when the condition $|f(i-1) - f(i)| \leq \varepsilon$ is satisfied, for the best solutions in two consecutive populations $P(i-1)$ and $P(i)$.

In Step 2 we evaluate the chromosomes of a population. As we said earlier, each chromosome is a string of numbers in which position i represent the number nr_i of outgoing edges in the subgraph that contain only edges through which the flow passes. To evaluate a chromosome, we must first obtain the solution encoded in it, and then evaluate the objective function.

To decode the solution from a chromosome we will do the following:

I. Let Nr_i be the number of outgoing edges from a vertex i in the graph G that describes the transport network. Then we have the following two cases:
- $nr_i \geq Nr_i$, the graph that describes the admissible solution contains all outgoing edges from the vertex i of graph G;
- $nr_i < Nr_i$, the graph that describes the admissible solution has only a subset of size nr_i of outgoing Nr_i edges from vertex i of graph G. These edges will be selected randomly.

II. We know that we need to transport a quantity of flow $p(v)$ through the network from the description of the problem. This flow will be assigned to the edges based on the matrix $R = \{r_{ij} | i = \overline{1,n}, j = \overline{1,n}\}$ generated in step 1, that shows how much of the flow from vertex i will go through edge (i, j). The obtained value is the flow x_k assigned to the edge (i, j) whose value will be placed in position k of the admissible solution of the form $(x_1, x_2, ..., x_m)$.

After constructing the admissible solution, we will compute the objective function. If the chromosome is deemed fit to go into the next population, this value will be stored so that in the next populations only the offspring will be evaluated. This allows us to decrease the execution time of the algorithm, which is very important, especially for big problems.

Theorem1: *The genetic algorithm P1 uses* $O(n^2)$ *memory.*

Proof: The transport network is described by an adjacency list of size $m < n^2$, the matrix $R = \{r_{ij} | i = \overline{1,n}, j = \overline{1,n}\}$ is of size n^2 and a population of chromosomes is of size $4n * n = 4n^2$, and every new population will be changed in place, so no other memory is necessary. As a result, the algorithm needs $O(n^2)$ memory. □

Theorem2: *The genetic algorithm P1 is polynomial and has a* $O(n^3)$ *complexity.*

Proof: To create the adjacency list that describes the graph $O(m)$ operations are needed. To create a population of size $4n$ with every chromosome of length n, $O(n^2)$ operations are needed. To evaluate all chromosomes in a population, we need $4n * m = 4nm$ operations, which creates a $O(nm)$ complexity.

The crossover has a $O(n^2)$ complexity, and the mutation $- O(n)$. As a result, the complexity of the algorithm is $O(n^3)$. $\square$

Remark1: For sparse graphs, the complexity is $O(nm)$.

Remark2: *The genetic algorithm P1* is convergent and always converges to a good solution. If the algorithm is run several times (for example in parallel or sequential) and we choose the best solution, we can get one very close to the global optima.

4. PRACTICAL APPLICATION

The algorithm described above was implemented in the Wolfram Language and tested on several random examples of different sizes. The tests were done using two stop conditions:

1. the algorithm was stopped when the condition $|f(i-1) - f(i)| \leq \varepsilon$ was satisfied, for the best solutions in two consecutive populations $P(i-1)$ and $P(i)$;

2. the algorithm was stopped after k populations were made.

As we can see in the following table (*Table 1.*), the execution time increases much slower for condition 1 compared to condition 2.

This happens because after a number of steps, even of a optima is found it will be sent to the next population because the condition 2 will not be satisfied. From this data we can recommend the first stop condition for the algorithm.

Table 1. Execution time of GA (seconds)

Nr. of vertices	t_ε	t_k	Nr. of vertices	t_ε	t_k
10	0,0262	0,0950	**60**	6,3544	46,7701
15	0,0809	0,3066	**65**	4,9332	83,5933
20	0,2421	0,7972	**70**	3,5041	102,1030
25	0,3294	1,6901	**75**	7,4183	195,1190
30	0,5540	3,700	**80**	18,0235	193,3200
35	1,1752	7,1764	**85**	12,8842	323,9760
40	0,9351	9,0325	**90**	10,5738	311,3430
45	1,1236	18,3405	**95**	45,6272	511,6740
50	3,5589	24,2091	**100**	55,9386	529,4140
55	2,2681	46,7701	**120**	117,9100	1621,2100

Using the standard Wolfram Language function $Minimize[\{f, cons\}, \{x, y, \dots\}]$ we can obtain the global minima for our problem.

Table 2. Execution time for *Minimiz* (seconds)

Nr. vertices	Minimize
4	0.17
6	3.08
8	345.61

The execution time for the function *Minimize* given in *Table 2.* increases starting from graphs with 8 vertices, i.e. the execution time on a graph with 8 vertices is 3 times longer than the genetic algorithm on a graph with 100 vertices.

By conducting these tests we could experimentally prove that the algorithm converges by computing the total objective function of a population and observing that it is always decreasing.

CONCLUSIONS

In this paper, we have discussed the transportation network problem with concave cost functions. To solve this problem, we proposed a genetic algorithm, which uses elements of graph theory, to transform a chromosome into an admissible solution.

1. The experimental results prove the correctness of the described algorithm, because we always get a good solution, and we can even get the global optima if we run it several times.

2. The algorithm is convergent, because the total fitness function of a population is always decreasing.

3. The execution time is much better comparing to standard Wolfram Language functions, which means that the algorithm is fast even for bigger networks.

REFERENCES

[1] Abdi, S., Baroughi, F., & Alizadeh, B., *The Minimum Cost Flow Problem of Uncertain Random Network,* Asia-Pacific Journal of Operational Research, vol. 35, no. 03, 2018.

[2] Bao, D., Gu, J., Di, Z., & Zhang, T., *Optimization of Airport Shuttle Bus Routes Based on Travel Time Reliability,* Mathematical Problems in Engineering, Article ID 2369350, https://doi.org/10.1155/2018/2369350 (vis. August 2018), 12 pages, 2018.

[3] Busacher, R. G., & Gowen, P. J., *A procedure for determining a family of minimum-cost network flow patterns.* Bethesda, MD: Technical Report ORO-TP-15, Operations Research Office, The John Hopkins University, 1960.

[4] Chandrasiri, A. M., & Samarathunge, D. M., *Application of Minimum Cost Flow Problem: A Case Study of Crown Distributors in Kegalle, Sri Lanka*, International Journal of Scientific & Engineering Research, vol. 8, no. 1, pp. 1850-1853, 2017.

[5] Chen, D., Ni, S., Xu, C., Lv, H., & Wang, S., *High-Speed Train Stop-Schedule Optimization Based on Passenger Travel Convenience,* Mathematical Problems in Engineering, Article ID 8763589, http://dx.doi.org/10.1155/2016/8763589, (vis. August 2018), 10 pages, 2016.

[6] Chen, Q., & Shi, F., *Model for Microcirculation Transportation Network Design.* Mathematical Problems in Engineering, Article ID 379867, doi:10.1155/2012/379867 (vis. August 2018), 11 pages, 2012.

[7] Chen, Y., & Rilett, L. R., *Signal Timing Optimization for Corridors with Multiple Highway-Rail Grade Crossings Using Genetic Algorithm,* Journal of Advanced Transportation, Article ID 9610430, https://doi.org/10.1155/2018/9610430 (vis. August 2018), 14 pages, 2018.

[8] Ciupală, L., *The minimum cost flow problem with surplus,* Bulletin of the Transylvania University of Braşov, Series III: Mathematics, Informatics, Physics , vol. 3, no. 52, pp. 177-182, 2010.

[9] Cohen, M. B., Madry, A., Sankowski, P., & Vladu, A., *Negative-weight shortest paths and unit capacity minimum cost flow in O(m[superscript 10/7]log W) time,* 28th Annual ACM-SIAM Symposium on Discrete Algorithm, 16-19 January 2017 (pp. 752-771). Barcelona, Spain: Association for Computing Machinery, http://hdl.handle.net/1721.1/113883, 2017.

[10] Dantzig, G. B., *Application of the Simplex Method to a Transportation Problem,* In T. C. Koopmans, Activity of Production and Allocation, New York: John Wiley and Sons, pp. 359-373, 1951.

[11] Davis, J. M., & Williamson, D. P., *A dual-fitting 3/2-aproximation algorithm for some minimum-cost graph problems,* 20th Annual European Symposium, no. 7501 in Lecture Notes in Computer Science, Springer, pp. 373-382, 2012.

[12] Dawuni, M., & Darkwah, K. F., *Maximum flow-minimum cost algorithm of a distribution company in Ghana: Case of 'NAAZO' Bottling Company,* Tamale, Tamale Metropolis, African Journal of Mathematics and Computer Science Research, vol. 8, no. 2, pp. 23-30, 2015.

[13] Ding, S., *Uncertain minimum cost flow problem,* Soft. Comput., vol. 18, pp. 2201-22017, 2014.

[14] Edmonds, J., & Karp, R., *Theoretical improvements in algorithmic efficiency for network flow problems,* Journal of the ACM, no. 19, pp. 248-264, 1972.

[15] El-Sherbeny, N. A., *Minimum Cost Flow Time-Windows Problem with Interval Bounds and Flows,* Theoretical Mathematics & Applications, vol. 6, no. 3, 2016.

[16] Erickson, J., Fox, K., & Lkhamsuren, L., *Holiest Minimum-Cost Paths and Flows in Surface Graphs,* Cornell University Library, arXiv:1804.01045, 2018.

[17] Eroglu, E., & Adiguzel, B., *A genetic algorithm based approach to the workload balancing problem,* 4th International Logistics and Supply Chain Management Congress, 2006.

[18] Fontes, D., & Goncalves, J. F., *Heuristic solutions for general concave minimum cost network flow problems,* Networks, An International Journal, https://doi.org/10.1002/net.20167 (vis.: August 2018), vol. 50, no. 1, pp. 67-76, 2007.

[19] Fulkerson, D. R., *An Out-of-Kilter Method for Minimal-Cost Flow Problems,* Journal of the Society for Industrial and Applied Mathematics, vol. 9, no. 1, pp. 18-27, 1961.

[20] Goldberg, A. V., & Tarjan, R. E., *Finding Minimum-Cost Circulations by Canceling Negative Cycles,* Journal of the Association for Computing Machinery, vol. 36, no. 4, pp. 873-886, 1989.

[21] Goldberg, A. V., & Tarjan, R. E., *Solving minimum-cost flow problems by successive approximations,* 19th ACM Symposium on Theory of Computing, STOC 87, New York: ACM Press, pp. 7-18, 1987.

[22] He, Q., Shabbir, A., & Nemhauser, G. L. *Minimum Concave Cost Flow Over a Grid Network*, Mathematical Programming, vol. 150, no. 1, pp. 79-98, 2015.

[23] Holland, J. H., *Genetic algorithms,* Scientific American , vol. 267, pp. 66-72, 1992.

[24] Horst, R., & Pardalos, P. M. *Handbook of Global Optimization*, Springer - Science + Business Media, 1st edition, 1995.

[25] Horst, R., & Tuy, H., *Global optimization, Deterministic Approaches.* New-York: Springer-Verlag, 3rd edition 1996.

[26] Jiang, Y., Xu, X., Zhang, H., & Luo, Y., *Taxiing Route Scheduling between Taxiway and Runway in Hub Airport.* Mathematical Problem in Engineering, Article ID 925139, http://dx.doi.org/10.1155/2015/925139 (vis. August 2018), 14 pages, 2015.

[27] Klein, M., *A primal method for minimal cost flows with applications to the assignment and transportation problems,* Management science, vol. 14, no. 3, pp. 205-220 1967.

[28] Kovacs, P., *Minimum-cost flow algorithms: An experimental evaluation,* Egervary Research Group, Technical reports, http://bolyai.cs.elte.hu/egres/ (vis. August 2018), 2013.

[29] Kudjo, P. K., & Ocquaye, E., *Review of Genetic Algorithm and Application in Software Testing,* International Journal of Computer Applications, vol. 160, no. 2, pp. 1-6, 2017.

[30] Luenberger, D. G., & Ye, Y., *Linear and nonlinear programming,* International Series in Operations Research and management science. Stanford: Springer, 2008.

[31] Maher, R. A., & Abdula, F. A., *An Algorithm for Cost-Minimizing in Transportation via Road Networks Problem,* International Journal of Mathematical and Computational Methods, http://www.iaras.org/iaras/journals/ijmcm, vol. 2, pp. 292-299, 2017.

[32] Moanţă, D., *Principii privind algoritmii genetici pentru soluţionarea unei probleme tridimensionale de transport,* Revista Informatica economică, no. 6, pp. 47-51, 1988.

[33] Osaba, E., Carballedo, R., Diaz, F., Onieva, E., Iglesia, I., & Perallos, A., *Crossover versus Mutation: A Comparative Analysis of the Evolutionary Strategy of Genetic Algorithms Applied to Combinatorial Optimization Problem,* The Scientific World Journal, Article ID 154676, http://dx.doi.org/10.1155/2014/154676, (vis. August 2018), 22 pages, 2014.

[34] Paiu, O. I., *Algoritmi genetici seriali şi paraleli,* Informatică economică, no. 6, pp.52-58, 1998.

[35] Sadegheih, A., & Drake, P. R., *A novel experimental analysis of the minimum cost flow problem,* IJE Transactiona A: Basics, vol. 22, no. 3, pp. 251-268, 2009.

[36] Sifaleras, A., *Minimum cost network flows: problems, algorithms, and software,* Yugoslav Journal of Operations Research , vol. 23, no. 1, pp. 3-17, 2013.

[37] Sleator, D. D., & Tarjan, R. E., *A data structure for dynamic trees.* Journal of Computer and System Sciences , vol. 26 no. 3, pp. 362-391, 1983.

[38] Sokkalingam, P. T., Ahuja, R. K., & Orlin, J. B., *New polynomial-time cycle-canceling for minimum cost flows.* Networks , vol. 36, pp. 53-63, 2000.

[39] Sun, W., & Yuan, Y.-X., *Optimization theory and methods, Nonlinear Programming*, Springer Science + Business Media, LLC, vol. 1, 2006.

[40] Tomizawa, N., *On some techniques useful for solution of transportation network problems.* Networks , vol. 1, pp. 173-194, 1971,

[41] Wang, I.-L., & Lin, S.-J., *A network simplex algorithm for solving the minimum distribution cost problem.* Journal of industrial and management optimization , vol. 5 no. 4, pp. 929-950, 2009.

[42] Yang, Z., Wang, W., Chen, S., Ding, H., & Li, X., *Genetic Algorithm for Multiple Bus Line Coordination on Urban Arterial,* Computational Intelligence and Neuroscience, Article ID 868521, http://dx.doi.org/10.1155/2015/868521 (vis. August 2018), 7 pages, 2015.

[43] Zhang, E., Mei, Q., Liu, M., & Zheng, F., *Stowage Planning in Multiple Ports with Shifting Fee Minimization.* Scientific Programming, Article ID 3450726, https://doi.org/10.1155/2018/3450726 (vis. August 2018), 9 pages, 2018.

[44] Zhang, P., Sun, Z., & Liu, X., *Optimized Skip-Stop Metro Line Operation Using Smart Card Data.* Journal of Advanced Transportation, Article ID 3097681, https://doi.org/10.1155/2017/3097681, (vis. August 2018), 17 pages, 2017.

REPRESENTING FUZZY SYSTEMS UNIVERSAL APPROXIMATORS

Iuliana IATAN[*], Stefan GIEBEL[**]

[*]Technical University of Civil Engineering, Bucharest, Romania (iuliafi@yahoo.com)
[**]Technical University of Braunschweig, Germany

Abstract: The described representation of a fuzzy system enables an approximate functional characterization of the inferred output of the fuzzy system. With polynomial subsystem inferences, the approximating function is a sum of polynomial terms of orders depending on the numbers of input membership functions. The constant, linear, and nonlinear parts of the fuzzy inference can hence be identified. The present work also includes two applications which show that the procedure can very well approximate the differential equations. In the case of no analytical solution, the procedure is a good alternative.

Keywords: fuzzy systems, approximating function, universal approximators, inferred output

1. INTRODUCTION

The aim of the paper is to prove that fuzzy systems are also universal approximators to continuous functions on compact domain in the case of the described subsystem inference representation corresponding to the fuzzy systems, as in the work of Kosko [11], Wang [16] and later Alci [1], Kim [10]. The paper is organized as follows: the first section provides a brief review on product sum fuzzy inference and introduces the concepts of additive and multiplicative decomposable systems; the second section presents a subsystem inference representation; the next sections discuss the cases of polynomial, sinusoidal, orthonormal and other designs of subsystem inferences; the last section presents some conclusions on the matter.

2. A FUZZY SYSTEM WITH TWO INPUT VARIABLES

A fuzzy system of n input variables a, b, K, y, z, with input membership functions $A_i, i = \overline{1, m_a}$, $B_j, j = \overline{1, m_b}, \mathrm{K}, Y_h, h = \overline{1, m_h}, Z_r, r = \overline{1, m_r}$ is expressible as an additive sum of $m_a \times m_b \times \mathrm{K} \times m_y \times m_z$ systems, each of which is multiplicative, and thus decomposable into n single variable subsystems.

Consider a fuzzy system with two input variables a and b with rule consequents embedded in the $m_a \times m_b$ matrix U_{m_a, m_b} from [7,8,15].

The inferred output is [7,8,15]:

$$\mathfrak{I}_{a,b}\left(U_{m_a, m_b}\right) = \sum_{q=1}^{m_a} \sum_{l=1}^{m_b} \gamma_{q,l} \cdot \mathfrak{I}_a\left(f_A^q\right) \cdot \mathfrak{I}_b\left(f_B^l\right),$$

(1)

where:

- $\gamma_{q,l}$ are the elements of the matrix ν_{m_a,m_b}, defined in [7,8,15];
- $f_A^q = \left(f_A^q(1) \quad f_A^q(2) \quad \mathrm{K} \quad f_A^q(m_a) \right), q = \overline{1,m_a}$ means a set of linear independent m_a

by one column vectors, selected for variable a and is associated to a subsystem $A^{(q)}$;

- $\Im_a\left(f_A^q\right)$ represents the inferred output of subsystem $A^{(q)}$;
- $f_B^l = \left(f_B^l(1) \quad f_B^l(2) \quad \mathrm{K} \quad f_B^l(m_b) \right), l = \overline{1,m_b}$ means a set of linear independent m_b by

one column vectors, selected for variable b;

- $\Im_b\left(f_B^l\right)$ represents the inferred output of subsystem $B^{(l)}$.

The selection of the vectors f_A^q and f_B^l should depend on the kind of approximation function one desires to use for the problem at hand, be it polynomial, sinusoidal, or other designs.

3. POLYNOMIAL SUBSYSTEM INFERENCE

The vectors $f_A^q, q = \overline{1,m_a}$ can be selected to emulate polynomial functions (they are termed polynomial subsystem vectors). The resulting subsystem inference $\Im_a\left(f_A^q\right)$ represents the polynomial subsystem inferences. In the case of a system with n fuzzy variables $a, b, \mathrm{K}, y, z,$ having $m_a, m_b, \mathrm{K}, m_y, m_z$ input membership functions, the inferred output is an approximation to the polynomial function, which contains polynomial terms up to orders of $m_a - 1, m_b - 1, \mathrm{K}, m_y - 1$ and $m_z - 1$ in a, b, K, y and z. Conversely, the polynomial function can be considered as an approximate output of the fuzzy system. The subsystem inference representation contributes to an approximate functional characterization of the inferred output in the sense that as $m_a, m_b, \mathrm{K}, m_y$ and m_z tend to large values, the polynomial inferences $\Im_a\left(f_A^i\right), \Im_b\left(f_B^j\right), \mathrm{K}, \Im_y\left(f_Y^h\right)$ and $\Im_z\left(f_Z^r\right)$ converge uniformly to the polynomial terms $a^{i-1}, b^{j-1}, \mathrm{K}, y^{h-1}$ and z^{r-1}.

4. SINUSOIDAL AND EXPONENTIAL SUBSYSTEM INFERENCES

Same as before for polynomial inferences, in the example of sinusoidal subsystem inferences, the fuzzy inferred output constitutes a piecewise linear approximation of a sinusoidal/ cosinusoidal function. Using the sinusoidal inferences, the approximating function is comprised of sine/ cosine, and cross product terms. With appropriate designs [7,8,15], sinusoidal inferences can be further manipulated into an orthonormal set.

In the example of exponential subsystem inferences, the inference of the fuzzy system constitutes [7,8,15] a piecewise linear approximation to an exponential term.

5. APPLICATIONS

In the case of the first application, we shall use the polynomial and exponential inferences together for the fuzzy approximation of the differential equation solution

$$\begin{cases} y''' = x\,e^{-x} \\ y(0) = 0 \\ y'(0) = 2 \\ y''(0) = 2 \end{cases} \tag{2}$$

We shall consider the fuzzy systems with two fuzzy variables: a is for approximating the polynomial term of x, and variables b is for realizing the exponential term e^{-x}.

Trapezoidal input membership functions in Fig. 1 and Fig. 2 are assumed [7] for all variables.
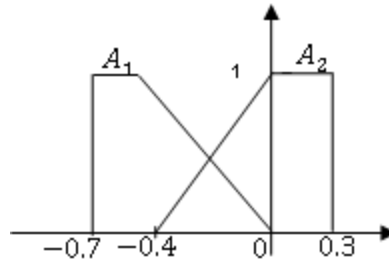
FIG. 1. Trapezoidal input membership functions for the fuzzy variable a

Let the domains of interest for a and b be [-0.7; 0.3] and [0.2; 0.8].
From the Fig. 1 we can notice that:

$$\mu_{A_1}(a) = \begin{cases} 1, & a \in \left[-0.7, -0.4\right] \\ 1 - \dfrac{a + 0.4}{0.4}, & a \in \left(-0.4, 0\right] \end{cases} \tag{3}$$

and

$$\mu_{A_2}(a) = \begin{cases} \dfrac{a + 0.4}{0.4}, & a \in \left[-0.4, 0\right] \\ 1, & a \in \left(0, 0.3\right] \end{cases} \tag{4}$$

For the variables a and b, one sets:

$$\begin{cases} m_a = m_b = 2 \\ a_1 = -0.4,\ a_2 = 0 \\ b_1 = 0.3,\ b_2 = 0.7 \\ f_A^{(1)} = f_B^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ f_A^{(2)} = \begin{pmatrix} 0 \\ a_2 - a_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0.4 \end{pmatrix} \\ f_B^{(2)} = \begin{pmatrix} 1 \\ e^{b_1 - b_2} \end{pmatrix} = \begin{pmatrix} 0 \\ e^{-0.4} \end{pmatrix}. \end{cases} \tag{5}$$

The inferred output of subsystem $A^{(2)}$ will be [7,8,15]:

$$\Im_a\left(f_A^2\right) = f_A^2(1)\cdot\mu_{A_1}(a) + f_A^2(2)\cdot\mu_{A_2}(a), \tag{6}$$

namely

$$\Im_a\left(f_A^2\right) = \begin{cases} 0.4\cdot\dfrac{a+0.4}{0.4}, & a\in[-0.4,0] \\ 0.4, & a\in(0,0.3] \end{cases} \tag{7}$$
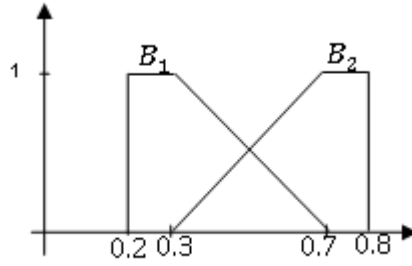
FIG. 2. Trapezoidal input membership functions for the fuzzy variable b

By observing Fig. 2, we can notice that:

$$\mu_{B_1}(b) = \begin{cases} 1, & b\in[0.2,0.3) \\ 1-\dfrac{b-0.3}{0.4}, & b\in[0.3,0.7] \end{cases} \tag{8}$$

and

$$\mu_{B_2}(b) = \begin{cases} \dfrac{b-0.3}{0.4}, & b\in[0.3,0.7] \\ 1, & b\in(0.7,0.8] \end{cases} \tag{9}$$

The inferred output of subsystem $B^{(2)}$ is [7]:

$$\Im_b\left(f_B^2\right) = f_B^2(1)\cdot\mu_{B_1}(b) + f_B^2(2)\cdot\mu_{B_2}(b), \tag{10}$$

namely

$$\Im_b\left(f_B^2\right) = \begin{cases} 1, & b\in[0.2,0.3) \\ 1+\left(1-e^{-0.4}\right)\cdot\dfrac{b-0.3}{0.4}, & b\in[0.3,0.7] \\ e^{-0.4}, & b\in(0.7,\,0.8] \end{cases} \tag{11}$$

$$\Im_b\left(f_B^2\right) = \begin{cases} e^{-0.4}, & x \in [0.2231, 0.3567) \\ 1 + \left(1 - e^{-0.4}\right) \cdot \dfrac{x - 0.3}{0.4}, & x \in [0.3567, 1.204] \\ 1, & x \in (1.204, 1.6092]. \end{cases} \tag{12}$$

As

$$\Im_{a,b}\left(U_{2,2}\right) = \sum_{q=1}^{2}\sum_{l=1}^{2} v_{q,l} \cdot \Im_a\left(f_A^q\right) \cdot \Im_b\left(f_B^l\right) = v_{2,2} \cdot \Im_a\left(f_A^2\right) \cdot \Im_b\left(f_B^2\right), \tag{13}$$

a fuzzy system to achieve (2) is [7]:

$$y''' = \Im_{a,b}\left(U_{2,2}\right) = \begin{cases} 0.4 \cdot e^{-0.4}, & x \in [0.2231, 0.3] \\ 0, & x \in [0.4, 0] \cup (0.3, 1.6092]. \end{cases} \tag{14}$$

where:

$$\begin{cases} y'(0) = 2 \\ y''(0) = 2 \end{cases}$$

and $U_{2,2}$ is given by [7].

Equation (14) will be analytically solved in Matlab 7.0, using the function **dsolve**; let $y_1(x)$ be the analytical solution of the equation (2), $y_2(x)$ and $y_3(x)$ the analytical solutions of the equation (14) for $x \in [0.4, 0] \cup (0.3, 1.6092]$ and respectively for $x \in [0.2231, 0.3]$.
We shall obtain:

$$\left|y_1(-0.1) - y_2(-0.1)\right| = 4.3376e\text{-}006$$
$$\left|y_1(0.4) - y_2(0.4)\right| = 9.118e\text{-}004$$
$$\left|y_1(0.7) - y_2(0.7)\right| = 0.0076$$
$$\left|y_1(1.25) - y_2(1.25)\right| = 0.0636$$
$$\left|y_1(0.25) - y_2(0.25)\right| = 5.5080e\text{-}004$$
$$\left|y_1(0.3) - y_2(0.3)\right| = 9.0670e\text{-}004.$$

Especially in the case of having no analytical solution at hand, the new procedure is interesting. The Lotka-Volterra equations are also called the predator-prey equations. The equations are a pair of first-order, non-linear, differential equations. They are needed to describe [8] the dynamics of biological systems in which two species interact with each other.

One is the predator and the other, its prey. If there are not enough preys, the population of predators will decrease. And if the population of preys increases, the predator population will also increase.

Furthermore, the Lotka-Volterra equations [6] are used in economics. Similar relations are established between different kinds of industries, as an example between engine construction and mining. Furthermore, the economic cycle in general can be simulated.

They develop in time according to the pair of equations [8]:

$$\begin{cases} \dfrac{\mathrm{d}\,x}{\mathrm{d}\,t} = x(\alpha - \beta y) \\[2mm] \dfrac{\mathrm{d}\,y}{\mathrm{d}\,t} = -y(\gamma - \delta x) \end{cases} \qquad (15)$$

where:

- $y = 10$ is the number of predators (for example, lions);
- $x = 800$ is the number of preys (for example, zebras);
- $\dfrac{\mathrm{d}\,x}{\mathrm{d}\,t}$ and $\dfrac{\mathrm{d}\,y}{\mathrm{d}\,t}$ represent the growth of the two populations against time;
- t represents the time;
- $\alpha = 3$, $\beta = 0.1$, $\gamma = 0.8$ and $\delta = 0.002$ are parameters representing the interaction of the two species.

The development of each species during a certain time interval can also be described by the upper procedure in the form of a polynomial. The values for t should be adapted to the procedure.

Fig. 3 shows that the method is useful if we seek to approximate (green) the function of one population(red).
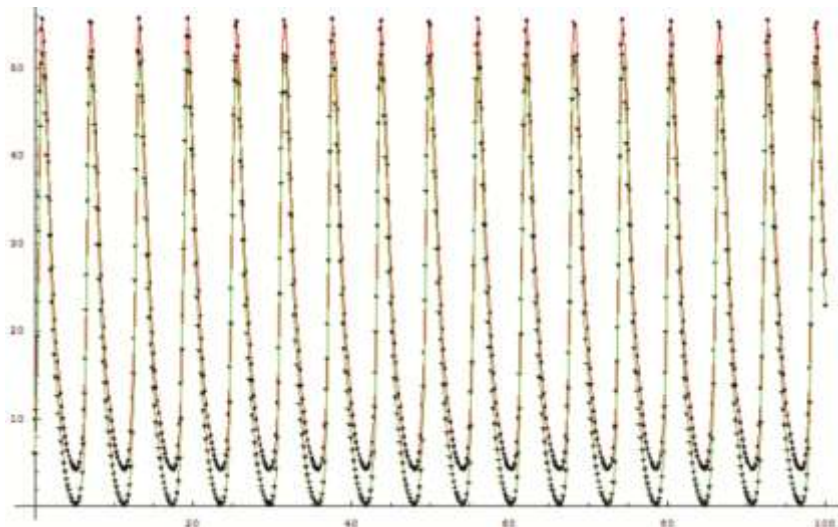

FIG. 3. The approximative solution of the population

The number of preys can be approximated easily by using the procedure above.

CONCLUSIONS

A new representation for fuzzy systems in terms of additive and multiplicative subsystem inferences of single variables is presented to prove that fuzzy systems are universal approximators to continuous functions on compact domain.

This representation enables an approximate functional characterization of the inferred output. The form of the approximating function depends on the choice of polynomial, sinusoidal, or other designs of subsystem inferences.

With polynomial subsystem inferences, the approximating function is a sum of polynomial terms of orders depending on the numbers of input membership functions.

Since polynomials are universal approximators [7,8,15], the same can be concluded regarding fuzzy systems.

With proper scaling, the sinusoidal inferences produce a set of orthonormal inferences.

The present work also includes two applications about constructing a fuzzy approximator for a function expressible in terms of sums and products of functions of a single variable. In this case, subsystem inferences that emulate the various single variable functions are adopted.

The second application [8] shows that the presented procedure can very well approximate a differential equation. In the case of no analytical solution the procedure, this is a good alternative.

REFERENCES

[1] M. Alci, Fuzzy rule-base driven orthogonal approximation, *Neural Computing & Applications*, vol. 17, no. 5-6, pp. 501-507, 2008;

[2] Y. Chen, B. Yang, A. Abraham and L. Peng, Automatic Design of Hierarchical Takagi-Sugeno Type Fuzzy Systems Using Evolutionary Algorithms, *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 3, pp. 385-397, 2007;

[3] M.T. Gan, M. Hanmandlu and A. H. Tan, From a Gaussian Mixture Model to Additive Fuzzy Systems, *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 3, pp. 303-315, 2005;

[4] G. Gandolfo. Giuseppe Palomba and the Lotka-Volterra equations. *Rendiconti Lincei*, vol. 19, no. 4, pp. 347-357, 2008;

[5] U. Grenander and M. Miller, *Pattern Theory from Representation to Inference*, Oxford University Press, New York, 2007;

[6] L. Hongxing, *From Probability Representations of Fuzzy Systems*, Science China Information Sciences, vol. 49, no. 3, pp. 339-363, 2006;

[7] I. Iatan, *A New Representation for The Fuzzy Systems in Terms of Some Additive and Multiplicative Subsystem Inferences*, Journal of Computational Analysis and Applications, vol. 14, no. 1, pp. 143-164, 2012;

[8] I. Iatan and S. Giebel, *A New Representation for The Fuzzy Systems*, in B. Hu et al. (eds.), *Operations Research Proceedings*, pp. 207-212, Selected Papers of the Annual International Conference of the German Operations Research Society (GOR) at Universität der Bundeswehr München, September 1–3, 2010, Springer, 2010;

[9] Y. Jin, Advanced Fuzzy Systems Design and Applications, Physica-Verlag Heidelberg, New York, 2003;

[10] J. Kim and J. S. Lee, *Single-Input Single-Output Convex Fuzzy Systems as Universal Approximators for Single-Input Single-Output Convex Functions,* pp. 350-355, Proceedings of the 18th International Conference on Fuzzy Systems, 2009;

[11] B. Kosko, Fuzzy Systems as Universal Approximators, *IEEE Transactions on Computers*, vol. 43, no. 11, pp. 1329-1333, 1994;

[12] Y. Liu, G. Chen and M. Ying, *Fuzzy Logic, Soft Computing and Computational Intelligence*, vol. 1, Tsinghua University Press and Springer, Beijing, 2005;

[13] Z. Miao, X. You and H. Zhen, A New Method for Fuzzy System Representation and Implementation, *Fuzzy Systems and Knowledge Discovery*, vol. 1, pp. 142-146, 2008;

[14] Y. Shi, Fuzzy Inference Modeling Based on Fuzzy Singleton-Type Reasoning, *International Journal of Innovative Computing*, *Information and Control*, vol. 3, no. 1, pp. 13-20, 2007;

[15] Y. Yam, Subsystem Inference Representation for Fuzzy Systems Based Upon Product-Sum-Gravity Rule, *IEEE Transactions on Fuzzy Systems*, vol. 5, no. 1, pp. 90-107, 1997;

[16] L. X. Wang, J. M. Mendel L. X. Wang and J. M. Mendel, Fuzzy Basis Functions, Universal Approximation, and Orthogonal Least Square Learning, *IEEE Trans. Neural Networks*, vol. 3, pp. 807-814, 1992.

STRICT STATIONARY TIME SERIES AND AUTOCOPULA

Daniel CIUIU

Technical University of Civil Engineering, Bucharest;
Romanian Institute for Economic Forecasting (dciuiu@yahoo.com)

Abstract: In this paper we consider not only the classical weakly stationarity of time series (same expectation, same variance and same correlations). We also aim to consider the strict stationarity of a time series. Therefore, each observation X_i from the time series $X_1,..., X_n$ has the same cumulative distribution function.

We consider that the common cdf F *is the common marginal distribution of* $X_1,..., X_n$*, and the dependence is expressed by a copula* C *of order* n*. An Archimedean copula is used, and the parameters (of the marginal cdf and the copula parameter θ) are estimated using the maximum likelihood method.*

Keywords: Time series, stationarity, smoothing.

1. INTRODUCTION

The definition of the term "copula" can be easily found in literature [2,15,10]. For time series, we use the theorem of Sklar, which establishes that every multivariate cumulative distribution function H (in our case the multivariate cdf of $\left(X_1,..., X_n \right)$) can be written

$$H \left(x_1,..., x_n \right) = C \left(F_1 \left(x_1 \right),..., F_n \left(x_n \right) \right), \tag{1}$$

where F_i is the marginal distribution of X_i.

For the copula C we use in this paper Archimedean copula, for which it is proved [2,8,11] that there exists $\varphi : [0,1] \to R$ decreasing and convex with $\phi(1) = 0$ having the pseudo-inverse g ($g(y) = x$ if x exists in such a way that $\phi(x) = y$, otherwise $g(y) = 0$) so that

$$C \left(u_1,..., u_n \right) = g \left(\sum_{i=1}^{n} \phi \left(u_i \right) \right). \tag{2}$$

For the Clayton copula, we have

$$\begin{cases} \phi(u) = \frac{u^{-\theta} - 1}{\theta} \\ g(x) = \left(\theta \cdot x + 1 \right)^{-\frac{1}{\theta}} \end{cases}, \text{ with } \theta > 0. \tag{3}$$

For the Frank family of copulas we have [2,15]

$$\begin{cases} \phi(u) = \ln \left(\frac{1 - e^{-\theta}}{1 - e^{-\theta u}} \right) \\ g(x) = -\frac{1}{\theta} \ln \left(\gamma e^{-x} + 1 \right), \text{ where } \gamma = e^{-\theta} - 1 \end{cases}, \text{ with } \theta \in R. \tag{4}$$

For the Gumbel-Hougaard copula we have [10,11]

$$\begin{cases} \phi(u) = (-\ln u)^{\theta} \\ g(x) = e^{-x^{\frac{1}{\theta}}} \end{cases}, \ with \ \theta \geq 1. \tag{5}$$

For the Gumbel-Barnet copula we have [15,11]

$$\begin{cases} \phi(u) = \frac{\ln(1-\theta\ln u)}{\theta} \\ g(x) = e^{\frac{1-e^{\theta x}}{\theta}} \end{cases}, \ with \ 0 < \theta \leq 1. \tag{6}$$

For the Ali-Mikhail-Haq copula we have [15,2]

$$\begin{cases} \phi(u) = \frac{1}{1-\theta} \cdot \ln\left(\theta + \frac{1-\theta}{u}\right) \\ g(x) = \frac{1-\theta}{e^{(1-\theta)x} - \theta} \end{cases}, \ with \ -1 \leq \theta \leq 1. \tag{7}$$

The Frank, Gumbel-Hougaard and Ali-Mikhail-Haq copulas contain the product copula (independence case) for $\theta = 0$, $\theta = 1$, $\theta = 0$, respectively. For the other copula families (Clayton and Gumbel-Barnett), the product copula is in both cases the limit case $\theta \to 0$.

Some copulas have been simulated in [15], and methods for simulating random variables and Monte Carlo methods can be found in [14].

When we consider the Gaussian time series, we test the weak stationarity using the Dickey-Fuller unit root test [7], and we stationarize the time series. Next, we find the ARMA model using the Box-Jenkins methodology [4,9,12]. If after we have found the ARMA model we have non-significant autocorrelations and partial autocorrelations, it means that in the Gaussian approach the model is correct. But, if we apply the BDS (Brock, Deckert and Skheinkman) test [3] and we obtain that the errors are not mutually independent and with the same distribution, and in the Gaussian approach the model is correct, it means that we must use a non-Gaussian approach. This test is as follows. First, we compute the probability

$$P_{k;\varepsilon} = P\left(\left\|(X_i,...,X_{i+k-1}) - (X_j,...,X_{j+k-1})\right\| < \varepsilon\right) \tag{8}$$

for a given ε empirically, where the above norm is the infinite norm (the maximum absolute value). If the values are independent with the same distribution, we have

$$P_{k;\varepsilon} = P_{1;\varepsilon}^k. \tag{8'}$$

Therefore we compute

$$Z_{k;\varepsilon} = \frac{P_{k;\varepsilon} - P_{1;\varepsilon}^k}{\sigma_{k;\varepsilon}}, \tag{8''}$$

where $\sigma_{k;\varepsilon}^2$ is the variance of the above numerator. The computation of $\sigma_{k;\varepsilon}^2$ is presented in [3], where it is mentioned that $Z_{k;\varepsilon}$ is asymptotically normal. Therefore, we have to compare these values to the cuantiles of the standard normal distribution, for $k = \overline{2, k_{max}}$.

2. THE MODEL

In order to apply the copula for strict stationary time series, we first have to take into account that strict stationarity yields to the same cdf for X_i. Therefore, in (1) we have $F_i = F$ (the same marginal cdf). The copula C models the dependence between the time series values at different moments. We call this copula autocopula, by analogy to the classical use of autocorrelations for classical ARMA models.

For estimating the parameters of the model, we take into account [15,2] that the multivariate pdf $h(x_1,...,x_n) = \frac{\partial^n H}{\partial x_1...\partial x_n}$ can be written

$$h(x_1,...,x_n) = \left| g^{(n)}\left(\sum_{i=1}^{n} \phi(F_i(x_i)) \right) \right| \cdot \prod_{i=1}^{n} |\phi'(F_i(x_i))| \cdot \prod_{i=1}^{n} f_i(x_i), \tag{9}$$

where f_i are the marginal pdfs.

Denote now by α the vector of parameters for the common marginal distribution having the cdf *F*, and by $V\!\!\!/\!o = \prod_{i=1}^{n} f_i(x_i;\alpha)$ the likelihood in the independence case (when $X_1,..., X_n$ are independent identical distributed, with the pdf *f* and cdf *F*). In the time series case, the above common pdf *h* is in fact the likelihood *V*, which must have a maximal value of 5 (we apply the maximum likelihood method). By computation, we obtain

$$\ln V = \ln\left(\left| g^{(n)}\left(\sum_{i=1}^{n} \phi(F(X_i;\alpha)) \right) \right| \right) + \sum_{i=1}^{n} \ln(-\phi'(F(X_i;\alpha))) + \ln V\!\!\!/\!o. \tag{10}$$

For solving the system of equations given by derivative on α components (α can be multiple, as in the normal case, when we have two parameters – the expectation and the variance), $\frac{\partial \ln V}{\partial \alpha_k} = 0$ and the derivative on θ, $\frac{\partial \ln V}{\partial \theta} = 0$, we try to express first the log-likelihood, because it is possible to separate α and θ. For instance, in the case of Clayton family we obtain

$$\ln V = \sum_{i=1}^{n-1} \ln(i \cdot \theta + 1) - \left(\tfrac{1}{\theta} + n\right)\ln\left(\sum_{i=1}^{n} F^{-\theta}(X_i;\alpha) - n + 1 \right) - (\theta+1)\left(\sum_{i=1}^{n} \ln F(X_i;\alpha) \right) + \ln V\!\!\!/\!o. \tag{11}$$

We notice that the last term, $\ln V\!\!\!/\!o$ does not depend on θ, the first sum does not depend on α, and the sum of logarithms multiplied by $\theta+1$ does not depend on θ. These make the computations easier, and we obtain for $\theta > 0$

$$\begin{cases} \dfrac{\partial \ln V}{\partial \alpha_k} = (n\theta+1)\dfrac{\sum\limits_{i=1}^{n} F^{-\theta-1}(X_i;\alpha)\cdot\frac{\partial F}{\partial \alpha_k}(X_i;\alpha)}{\sum\limits_{i=1}^{n} F^{-\theta}(X_i;\alpha)-n+1} - (\theta+1)\sum\limits_{i=1}^{n}\dfrac{\frac{\partial F}{\partial \alpha_k}(X_i;\alpha)}{F(X_i;\alpha)} + \dfrac{\partial \ln V\!\!\!/\!o}{\partial \alpha_k} \\[4mm] \dfrac{\partial \ln V}{\partial \theta} = \sum\limits_{i=1}^{n-1}\dfrac{i}{i\theta+1} + n\sum\limits_{i=1}^{n} F^{-\theta}(X_i;\alpha)\ln F(X_i;\alpha) - \sum\limits_{i=1}^{n}\ln F(X_i;\alpha) + \dfrac{\theta\left(\sum\limits_{i=1}^{n} F^{-\theta}(X_i;\alpha)\ln F(X_i;\alpha)\right)+\left(\sum\limits_{i=1}^{n} F^{-\theta}(X_i;\alpha)-n+1\right)\ln\left(\sum\limits_{i=1}^{n} F^{-\theta}(X_i;\alpha)-n+1\right)}{\theta^2\left(\sum\limits_{i=1}^{n} F^{-\theta}(X_i;\alpha)-n+1\right)} \end{cases}$$

$$\tag{12}$$

For the limit case, $\theta \to 0$ we obtain, using l'Hôpital

$$\begin{cases} \dfrac{\partial \ln V}{\partial \alpha_k} = \dfrac{\partial \ln V\!\!\!/\!o}{\partial \alpha_k} \\[4mm] \dfrac{\partial \ln V}{\partial \theta} = C_n^2 + (n-1)\left(\sum\limits_{i=1}^{n} \ln F(X_i;\alpha) \right) + \dfrac{\left(\sum\limits_{i=1}^{n}\ln F(X_i;\alpha)\right)^2}{2} - \dfrac{\sum\limits_{i=1}^{n}\ln^2 F(X_i;\alpha)}{2} \, . \end{cases} \tag{12'}$$

We can also prove that the Hessian is in the general case $\theta \neq 0$

$$
\begin{cases}
\dfrac{\partial^2 \ln V}{\partial \alpha_j \partial \alpha_k} = \dfrac{n\theta+1}{\left(\sum\limits_{i=1}^{n} F_i^{-\theta}-n+1\right)^2} \cdot \left(-(\theta+1)\left(\sum\limits_{i=1}^{n} F_i^{-\theta-2}\dfrac{\partial F_i}{\partial \alpha_j}\dfrac{\partial F_i}{\partial \alpha_k}\right)\left(\sum\limits_{i=1}^{n} F_i^{-\theta}-n+1\right) + \left(\sum\limits_{i=1}^{n} F_i^{-\theta-1}\dfrac{\partial^2 F_i}{\partial \alpha_j \partial \alpha_k}\right)\left(\sum\limits_{i=1}^{n} F_i^{-\theta}-n\right.\right. \\[4mm]
\left.\left.\left(\sum\limits_{i=1}^{k} F_i^{-\theta}\dfrac{\partial F_i}{\partial \alpha_j}\right)\left(\sum\limits_{i=1}^{k} F_i^{-\theta}\dfrac{\partial F_i}{\partial \alpha_k}\right)\right) -(\theta+1)\left(\sum\limits_{i=1}^{n}\dfrac{\frac{\partial^2 F_i}{\partial \alpha_j \partial \alpha_k}-\frac{\partial F_i}{\partial \alpha_j}\frac{\partial F_i}{\partial \alpha_k}}{F_i^2}\right) + \dfrac{\partial^2 \ln V\%}{\partial \alpha_j \partial \alpha_k} \\[4mm]
\dfrac{\partial^2 \ln V}{\partial \theta \partial \alpha_k} = \dfrac{n\left(\sum\limits_{i=1}^{n} F_i^{-\theta}-n+1\right)-(n\theta+1)\left(\sum\limits_{i=1}^{n} F_i^{-\theta}\ln F_i\right)}{\left(\sum\limits_{i=1}^{n} F_i^{-\theta}-n+1\right)^2}\cdot\left(\sum\limits_{i=1}^{n} F_i^{-\theta-1}\dfrac{\partial F_i}{\partial \alpha_k}\right) - \dfrac{n\theta+1}{\sum\limits_{i=1}^{n} F_i^{-\theta}-n+1}\cdot\left(\sum\limits_{i=1}^{n} F_i^{-\theta-1}\dfrac{\partial F_i}{\partial \alpha_k}\right) - \sum\limits_{i=1}^{n}\dfrac{\frac{\partial F_i}{\partial \alpha_k}}{F_i} \\[4mm]
\dfrac{\partial^2 \ln V}{\partial \theta^2} = -S_1 - nS_2 - \dfrac{\theta^2 S_2(S_0-n+1)+2\theta S_3(S_0-n+1)-\theta^2 S_3^2+2(S_0-n+1)\ln(S_0-n+1)}{\theta^3(S_0-n+1)^2}
\end{cases} \tag{13}
$$

, where

$$
\begin{cases}
S_0 = \sum\limits_{i=1}^{n} F^{-\theta}\left(X_i;\alpha,\theta\right) \\[3mm]
S_1 = \sum\limits_{i=1}^{n-1} \dfrac{i^2}{(i\theta+1)^2} \\[3mm]
S_2 = \sum\limits_{i=1}^{n} F^{-\theta}\left(X_i;\alpha,\theta\right)\ln^2 F\left(X_i;\alpha,\theta\right) \\[3mm]
S_3 = \sum\limits_{i=1}^{n} F^{-\theta}\left(X_i;\alpha,\theta\right)\ln F\left(X_i;\alpha,\theta\right)
\end{cases} . \tag{13'}
$$

In the independence case, $\theta = 0$, in the first equation, all the terms except $\frac{\partial^2 \ln V\%}{\partial \alpha_j \partial \alpha_k}$ vanish, and we can say the same about the right side of the second equation: $\lim\limits_{\theta \to 0}\dfrac{\partial^2 \ln V}{\partial \alpha_k \partial \theta} = 0$. In the case of the Clayton copula, the Hessian is negatively defined. For this reason, the non-linear system

$$
\begin{cases}
\dfrac{\partial \ln V}{\partial \alpha_k} = 0 \\[3mm]
\dfrac{\partial \ln V}{\partial \theta} = 0
\end{cases} \tag{14}
$$

is solved by means of the Newton-Raphson method, solving the involved linear system having the matrix of this system the Hessian (the Jacobean in the general case of non-linear systems) and the right sides given by the actual values of the left sides in (14) with the inverse sign by the Cholesky method. Before applying the Cholesky method, we multiply first the linear system by (-1) in order to obtain a positively defined matrix of the linear system.

3. APPLICATIONS

Consider the ROBOR rate between January 1, 2017 and April 3, 2018 (313 data points, observed daily, five days/ week).

First, we apply the classical Box-Jenkins approach. If we apply the Dickey-Fuller test, model III, we obtain the coefficients for β, Φ and γ 0.00545, -0.00635 and $5.92 \cdot 10^{-5}$, with the Student statistics 1.4531, -1.64969, respectively 2.5674. The time series is not stationary. After removing the moving average [9] of order $q = 2$, we obtain the new coefficients for β, Φ and γ -0.08792, -0.88554 and 0.000494, with the Student statistics -0.08976, -15.6655, respectively 0.091049. In fact, due to small values, the residues after removing the moving average is multiplied by the constant number 1000.

The SARMA stationary model is

$$X_t = 0.55063X_{t-1} - 0.33823X_{t-2} - 0.2505X_{t-5} + a_t + 0.5a_{t-1}$$

The maximum correlation in absolute value for the white noise a_t among the first 36 is $\rho_{33} = 0.181$, and the maximum partial correlation is $\hat{\rho}_{28} = 0.159$. For the first ten, the maximum autocorrelation and partial autocorrelation in absolute value are for lag 7: -0.081 and -0.073. This means that we have obtained the correct Gaussian model.

But, if we apply to the obtained white noise the BDS (Brock, Deckert and Sheinkman) test, we obtain the Z statistics for maximum dimension 6 and $\varepsilon = 0.7$ between 5.8352 (dimension=2) and 9.23336 (dimension=6). Therefore, we reject the null hypothesis of independence and same distribution, with the 1% threshold.

In the non-Gaussian case we first apply the Mann-Kendal test for initial data, and we obtain the statistics $Z = 17.19643$, which means that ROBOR data follows an increasing trend. After that, we apply the same linear transformations as in the Gaussian case, the Z statistics of Mann-Kendall test becomes $Z = -0.31851$, which means a non-significant decreasing trend (in fact we accept the null hypothesis of lack of trend).

Consider now the Clayton copula and the exponential marginal distribution. We obtain the following results, after 10 iterations using Newton-Raphson method.

Table 1. Results if we use the Clayton copula and the exponential marginal distribution

Value	Initial	Final
λ	0.0101	0.0206
θ	0	10.33141
ln V	-1751.25231	-89.08146
$\begin{pmatrix} \dfrac{\partial \ln V}{\partial \lambda} \\ \dfrac{\partial \ln V}{\partial \theta} \end{pmatrix}$	$\begin{pmatrix} 0 \\ -124723.4167 \end{pmatrix}$	$\begin{pmatrix} 0.00927 \\ -0.03483 \end{pmatrix}$
Hessian	$\begin{pmatrix} -3067317.013 & 0 \\ 0 & -10172519.95 \end{pmatrix}$	$\begin{pmatrix} -173360.23 & -20655.083 \\ -20655.083 & -4959.455 \end{pmatrix}$

CONCLUSIONS

In [5] we have presented a heavy tail smooth for non-stationary long memory time series. In this paper, we use auto-copula for stationary long memory time series: an Archimedean copula is the same dependence between X_n and X_{n-1}, and between X_n and X_1.

Before applying a model for a stationary time series, we have to test first the stationarity. In the Gaussian case, we apply the Dickey-Fuller unit root test, as we have mentioned before. But in the non-Gaussian case of our paper, we have to use other tests. For instance, we apply the Mann-Kendall test for lack of trend, used in [13] for discharges of Danube River. After we have found an increasing trend for the initial data, we have made the same transformations as in the Gaussian case to obtain stationary time series. Nevertheless, the Mann-Kendall test confirms the stationary of the transformed time series. An open problem is to check other transformations for non-normal distributions, for instance the ratio in exponential case. Such transformation avoids negative values, and we do not need to subtract the minimum value for obtaining positive values of last time series.

Another open problem is to solve the non-linear system (14) for other types of copula for which we can not separate the parameter θ from the marginal parameters, as in the Clayton case.

For instance, we can use the recurrence formulae obtained in [6] for the Frank copula, Gumbel-Hougaard copula, Gumbel-Barnet copula and Ali-Mikhail-Haq copula. For this, we need an analogue recurrence formula, but for mixed derivatives including u, and θ.

In the exponential case discussed in our paper, the non-linear system (14) has two variables: λ and θ. Therefore, another open problem is to consider other marginal distributions as well. The difficulty is not the number of variables, but the computation of the involved marginal cdf in (10) - (13), (12') and (13').

REFERENCES

[1] *** Piaţa Monetară Interbancară, ROBID - ROBOR: serii zilnice, *The Interactive Database of the National Bank of Romania*, www.bnr.ro, accessed on 3 Apr. 2018;

[2] G. Dall' Aglio, *Fréchet classes: the beginning*, in G. Dall' Aglio, S. Kotz and G. Salinetti (Eds) *Advances in Probability Distributions with Given Marginals. Beyond the Copulas*, pp. 1-12, Kluwer Academic Publishers, 1991;

[3] W.A. Brock, W.P. Deckert and J.A. Sheinkman, A Test for Independence Based on the Correlation Dimension, *Department of Economics, University of Wisconsin at Madison, University of Houston and University of Chicago, grants no. SES-8420872, 144-AH01 and SES-8420930*, 1987;

[4] P.J. Brockwell and R.A. Davis, *Springer Texts in Statistics. Introduction to Time Series and Forecasting*, Springer-Verlag, 2002;

[5] D. Ciuiu, *Heavy Tail Smooth and Application to Long Memory Time Series*, in L. Chivu, C. Ciutacu, V. Ioan-Franc and J.-V. Andrei (Eds) *Economic Dynamics and Sustainable Development - Resources, Factors, Structures and Policies*, pp. 157-164, Proceedings of The 3rd International Conference "Economic Scientific Research - Theoretical, Empirical and Practical Approaches ESPERA 2015", December 3-4 2015, Bucharest, Peter Lang, 2016;

[6] D. Ciuiu, Simulation of Queueing Systems with Many Stations and of Queueing Networks Using Copulas, *Mathematical Modeling in Civil Engineering*, no. 3, pp. 72-87, 2010;

[7] D.A. Dickey and W.A. Fuller, Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, *Econometrica* vol. 49, no. 4, pp. 1057-1072, 1981;

[8] C. Genest, Statistical Inference Procedures for Bivariate Archimedean Copulas, *Journal of American Statistical Association*, vol. 19, no. 1, pp. 1034-1043, 1993;

[9] D. Jula and N.M. Jula, *Prognoza economică*, Ed. Mustang, Bucharest, 2015;

[10] S. Kotz and J.P. Seeger, *A new approach to dependence in multivariate distributions*, in G. Dall' Aglio, S. Kotz and G. Salinetti (Eds) *Advances in Probability Distributions with Given Marginals. Beyond the Copulas*, pp. 113-127, Kluwer Academic Publishers, 1991;

[11] R. Nelsen, *Copulas and association*, in G. Dall' Aglio, S. Kotz and G. Salinetti (Eds) *Advances in Probability Distributions with Given Marginals. Beyond the Copulas*, pp. 51-74, Kluwer Academic Publishers, 1991;

[12] Th. Popescu, *Serii de timp. Aplicaţii în analiza sistemelor*, Technical Publishing House, Bucharest, 2000;

[13] R. Trandafir, D. Ciuiu, R. Drobot, *Testing Some Hypotheses for the Discharges of the Danube River*, in G. Păltineanu, P. Matei and G. Groza (Eds) *Proceedings of the 11th Workshop of the Department of Mathematics and Computer Science, Technical University of Civil Engineering Bucharest, May 27 2011, TUCB*, pp. 95-98, Matrix Rom, Bucharest, 2011;

[14] I. Văduva, *Modele de simulare*, Bucharest University Printing House, 2004;

[15] I. Văduva, Simulation of some multivariate distributions, *Analele Universităţii Bucureşti*, no. 1, pp. 127-140, 2003.

RANDOM WALK HYPOTHESIS ON BUCHAREST STOCK EXCHANGE

Sorina GRAMATOVICI, Corina-Mihaela MORTICI

Bucharest University of Economic Studies, Romania (sorina.gramatovici@csie.ase.ro, corinamortici@gmail.com)

Abstract: The aim of this paper is to examine the random walk in two of the stock indexes of , the Bucharest Stock Exchange (BET and BET Plus). Random walk hypothesis states that stock prices move randomly; as a result, the expected profit for the speculator is zero. Many economists believe that random walk can be applied to test the efficient market hypothesis in the weak level. Early literature used stochastic processes to test whether prices precluded everyone from easy profit and whether prices were following those processes or not. When stock prices do not fluctuate randomly, some investors can use past stock prices to gain an abnormal return. Assuming rationality and risk neutrality, a version "of the efficient market hypothesis states that information observable to the market prior to week t should not help to predict the return during week t". In other words, stock returns are not correlated to one another; consequently, the statistical model of the efficient market hypothesis holds and changes in returns are independent We employ several tests, such as econometric tests, Monte Carlo simulation using AI methods: Naive Bayes' Classifier, K Nearest Neighbors, Support Vector Machines. Daily data on returns covered the period February 2016 – November 2017. These tests support the common results that the random walk theory is valid for the two indexes therefore the Bucharest Stock Market is weak-form efficient.

Keywords: Random Walk, Stock Market, Efficient Market Hypothesis, Weak form efficiency, K Nearest Neighbors, Support Vector Machines

1. INTRODUCTION AND MOTIVATION

The Efficient Market Hypothesis (EMH), also known as Random Walk Theory, refers to the efficiency of information on stock markets. In economic literature, the term efficient market is used to explain the dependence between available information and stock price. These concepts were introduced and defined by Eugene Fama [3] in 1970, whose perspective has been that financial market efficiency is being driven by the speed of response and the adjustment of prices to new information in the market.

More specifically, in the context of an efficient market, the prices of the current period's shares should fully reflect the relevant information in order to be able to forecast future prices so that there is no possibility of generating further profits using this information. Therefore, the main criteria in terms of efficiency with respect to what was described are: the extent to which the information is absorbed, the time it takes for it to accumulate, and the type of information so incorporated.

The price of an asset reflects the current value of the revenue it is speculated to generate in the forthcoming period. The expected revenue is influenced by determinants such as risks of volatility, liquidity, or bankruptcy.

While prices are determined and set reasonably, price changes are expected to be random and unpredictable, because new information is unpredictable by its nature. Thus, capital market prices are considered to follow a random walk process.

As examples of random walks, we mention throwing a coin or selecting a sequence of numbers from a random number table. Looking back to the financial markets, the current price is independent and uncorrelated with other evolution patterns of the past price.

Let X be a stochastic variable which follows a random process defined by the following equation

$$X(t + 1) = \delta + X(t) + \varepsilon_{t+1} \tag{1}$$

where δ is the drift parameter, ε_{t+1} identically and independently distributed prediction error.

The test of efficiency in its weak form has been widely studied in financial literature. Ayadi and Pyun investigated in [1] the prices of stock traded on Korean stock market between January 1984 and December 1988 and proved that The Korean stock market is a random walk. Kim and Shamsuddin [4] report the existence of a random walk for Hong Kong, Japon, Korea and Taiwan and rejected the random walk hypothesis for Indonesia, Malaysia and Phillipines. Lim et al. tested in [5] the efficiency of Shangai and Shenzhen stock markets and concluded that China's stock market has a week form efficiency. More recently, Chaibi [2] tested the weak form efficiency according to two indices of the Hong Kong stock exchange between July 1997 and December 2012 and rejected the random walk hypothesis pursuant to both of these indices. Mishra et al. [7] tested the random walk hypothesis for the Indian stock market using 19 years data on six indexes from National Stock Exchange and Bombay Stock Exchange. They used a unit root test that simultaneously accounts for heteroskedasticity and structural breaks and proved that Indian stock indexes are mean reverting.

The aim of the present paper is to analyze the validity and correspondence with the real markets of the theoretical concept of random walk process. In this respect, the practical illustration of the Romanian capital market case was considered relevant, namely by studying the evolution of two of the most important indices evaluated by the Bucharest Stock Exchange, i.e. BET and BET Plus using daily data throughout a period of two years. The random nature of these time series strictly corresponds to the real case if and only if the market is efficient in the weak form. Through the current paper, we intend to study the efficiency of the Romanian capital market, contributing to the results obtained in the existing literature.

Evaluating shares and stock indices is a crucial function of the financial markets, as it leads to the possibility of players' making investment strategies. Evaluating the value of shares is of particular importance to determine the behavior of markets, its behaviour being possible if and only if the type of efficiency is known. Remark that the available information is crucial because it can lead to arbitrage. Acquisition and sale of similar share simultaneously on two different markets as a result of its differences in price conceptualizes the idea of arbitrage. The effect of arbitration plays an essential role in the efficiency of a market because this phenomenon brings prices back to their intrinsic value.

If markets are efficient in a weak sense, it is not possible for players to buy a share whose price is underestimated and to sell them on other markets, where they are fairly valued or overestimated. The very event described makes it impossible for the players involved to „beat the market".

In this respect, the obvious question is still: why do investors analyze the market with sophisticated and time-consuming tools if their efforts are futile?

This seems to be the main concern of the current paper, since totally rational investors would not play or invest if they did not have the chance to beat the market. As Lo, Mamaysky and Wang [6] asserted, "With the help of sophisticated non-parametric techniques ... [analysts] would only enjoy a modest prediction power", often insufficient to play based on the fundamentals of these overwhelming strategies.

This paper is structured as follows. Section 2, which is preceded by the present introduction, covers the theoretical and methodological aspects that are the solid foundation of the case study undertaken in Section 3. Both sections deal in a multidisciplinary manner with the characteristics of the time series studied in order to test the existence of the "random walk" phenomenon, then to make predictions using econometric methods, but also using artificial intelligence. The final section summarizes the conclusions of the analysis – its significance and validity, and the author's contribution to the literature.

The period of February 15, 2016 - November 2, 2017 has been analyzed, namely the data on BET and BET Plus stock exchange index values from the Bucharest Stock Exchange. The data was collected from the official website of the Bucharest Stock Exchange (www.bvb.ro) and processed by the authors to perform the relevant tests in order to achieve the above mentioned objective.

This paper is outlined by the analysis and tests performed on the BET and BET Plus time series assuming rationality and risk neutrality, and not taking into account transaction costs and other costs that may be charged for players on the capital market. The analyzed period (435 observations) was the subjective decision of the author, and no attempt was made in order to identify an optimal dimension of time to observe indexes' patterns.

2. METHODOLOGY

2.1 The phenomenon of "ARCH Effect". Financial time series are frequently characterized by volatility, a phenomenon that is modeled by processes such as ARCH. The ARCH effect defines the hypothesis of financial market's speculators estimating the variance over a certain period of time with the information that has appeared in the previous period, and included in the model by the term ARCH. This context describes the well-known concept of "volatility clustering", which means that periods of high magnitude change are followed by periods of small fluctuations. In other words, significant changes in financial time series tend to cluster together, and low magnitude changes are of the same behavior. If the data series is affected by the ARCH, term, they might be predictable to a certain extent and respond to market speculation (a well-known example would be the weekend effect).

By testing the presence of this phenomenon with the ARCH LM test, we introduce the null test hypothesis

H_0: *There is no arch effect*

and the alternative

H_1: *There is an arch effect*

The regression to be estimated is given by the following equation.

$$u_t^2 = \propto_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_p u_{t-p}^2 + v_t \tag{2}$$

where u_t represents the residue of the initial regression estimated by the ordinary least squares method.

The statistic test is $T * R^2$, where T is the number of observations included in the analysis, and R^2 is the coefficient of determination of the initial regression. It follows a Chi-square distribution with p degrees of freedom.

2.2 Naïve Bayes Classifier. The naive Bayesian algorithm is a classification technique based on Bayes' theorem, based on the hypothesis of independence of predictors. In other words, the naive Bayes Classifier assumes that the presence of a certain characteristic in a class is uncorrelated with the presence of any other characteristic. For example, a fruit is identified as an apple if it is red, round, and has a diameter of about 5 cm. Although these characteristics depend on one another or, more clearly, the existence of each characteristic is dependent on the existence of the other, all these properties contribute, in an independent manner, to the likelihood that this fruit being called apple - and that is why this classifier is called "naive".

The naive Bayesian classifier is useful for large data sets, and, despite the simplicity of the assumptions on which it is based, it is known that it provides better performance than other complex classification techniques, and therefore has been included in this study.

Bayes' theorem represents a method to calculate the posterior probability P (c/x) based on the posterior probability of the class P(c), the posterior probability of the predictor P(x) and the probability of a predictor in a given class P(x/c), as follows.

$$P(c|x) = \frac{P(x|c) * P(c)}{P(x)} \tag{3}$$

Where

$$P(c|X) = P(x_1|c) * P(x_2|c) * ... * P(x_n|c) * P(c) \tag{4}$$

2.3 K Nearest Neighbors. The KNN method is a classification algorithm that includes all the observations and classifies the new observations generally based on a measure of similarity, and in most cases, this is a function of distance. The KNN has been used since the 1970s in statistical estimations and in the pattern recognition as a non-parametric technique.

The algorithm assumes that a new observation is classified based on the vote of its nearest K neighbors. A new instance is assigned to the most common class among its neighbors based on the distance function. The most widely used distance functions are: Euclidean distance, Manhattan distance and Minkowski distance, for continuous variables, and Hamming distance for discrete variables.

The decision to determine the optimum value for K is to be taken by inspecting the data series. Generally, a high K value leads to more accurate results, as it reduces noise. The cross validation procedure is another method by which the value of K is computed.

Based on the empirical studies already performed, for most data sets, a value of K between 3 and 10 produces satisfactory results and, moreover, more efficient results than the 1NN method.

2.4 Monte-Carlo Simulation. The Monte-Carlo simulation method is used in many branches of science in order to quantify the expectations of the evolution of a variable of an index whose behavior is similar to a random walk process.

In this case, however, no analytical function can describe its evolution, and the optimal decision is to generate random samples describing these variables course of action. The accuracy of the estimates obtained by the Monte-Carlo simulation method is inversely proportional to the number of extractions.

In the current study, the type of simulation used is called "time-driven", in the sense that, for this given period, we have built different scenarios that can lead to the fluctuations in the analysis.

If a stock market player uses the Monte-Carlo simulation for a past period, for which the evolution of the index in the analysis is already known, they will be aware of the various trajectories that this index could have followed, and thus the magnitude of the risk assumed it by choosing a discreet strategy among all possible. Computing a strategy, they certainly take need to take into account both the risk prize and the magnitude of a potential loss. Loss has to be taken into account, and it plays a crucial role in the computation of the expected profit, because by visualizing the many trajectories that an index may follow, it the high probability of the monetary loss involved can obviously be deduced. However, the more players are involved in capital market's game, the greater the risk, and so, the importance of using the Monte-Carlo method is obvious. The Monte-Carlo simulation is even superior to "What If" analysis, because in many cases, it is very difficult to identify or test the determinants of fluctuations in an index on the stock market. Moreover, when making a decision, it is especially important to include a graphical view of the various scenarios, so that the decision maker may becomes aware of the probability associated with the occurrence of each state of nature.

2.5 Support Vector Machine. Support vector machine is a supervised learning algorithm that can be used in regressions, but is especially suitable for classifications. In this method, each observation is represented in an n-dimensional space (where n is the number of states of the variable), the value of each state being represented by a coordinate. Classification is carried out by determining the hyperplane that segregates these classes, and in our investigation, the separation of the two classes.
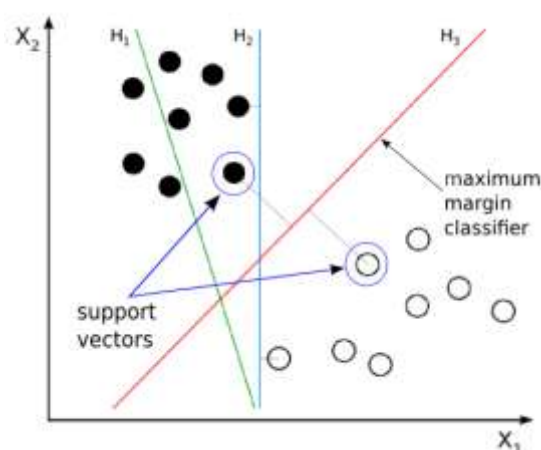

FIG. 1. Support Vector Machine representation [10]

Support vectors are simply the coordinates of an individual observation, and the support vector machine is the hyperplan, the border that optimally segregates the two classes.

In order to correctly determine this hyperplane, the shape of the nucleus must be computed. For most financial data sets, it is linear. Also, the gamma coefficient is associated to the kernel; the higher the value of the kernel, the more the algorithm will try to classify a new observation only if it has an almost perfect identification with a particular class.

However, the disadvantage of a too high gamma coefficient is that some of the observations will not be classified.

3. EMPIRICAL RESULTS

3.1 ARIMA-Modelling and Seasonality. ARIMA-modelling the evolution of stock indices is appropriate, as, mainly, in the context of financial time series there are trends, seasonality, errors, shocks, all these factors being taken into account in this type of model. To perform the procedure, we used the stationarized data series. For the analyzed period, we included monthly BET and BET Plus indices in order to capture the phenomenon of seasonality along with the influence on the evolution of time series. Moreover, we considered it important to add dummy variables corresponding to each month of the year in the analysis - to test whether they are statistically significant or not. Although, generally, we have to include 12 seasonal factors, we chose to exclude December, because, including the intercept, the dummy trap would have occurred.

The ultimate goal of the analysis is to predict the average values of the indices over the next two months based on the chosen model.
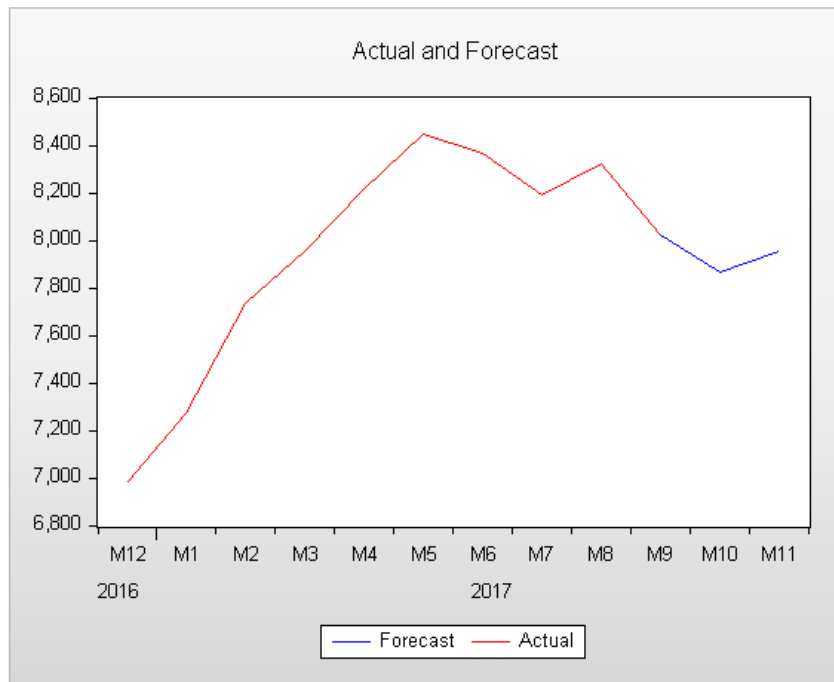

FIG.2. ARIMA Modelling

The most appropriate model was chosen using the Akaike informational criterion, out of 100 estimated candidate models. The minimum value (13.6714) corresponds to the most effective model. Proceeding with this analysis, the current values and the predicted values are presented in the table below.

Table 1 - Real values vs. predicted values

Month	Average real value	Average predicted value
October	7946.01	7868.38
November	7783.36	7955.372

As these values differ significantly from the real ones, and moreover, if the BET index actually shows, on average, a decrease from October to November, and using ARIMA modelling, it was forecasted to increase by almost 100 we can safely conclude that, even if the presence of the seasonal factors in the analysis is taken into account, the evolution of the Bucharest Stock Exchange still cannot be predicted, and therefore, there is a random walk.

3.2 Correlation Between BSE and international stock markets. Furthermore, we consider crucial to analyze the correlation between the evolution of the main indices of the Bucharest Stock Exchange and the evolution of the international stock markets. In this respect, it is relevant to observe the fluctuations of the US S&P500 index, for the same period.

By standardizing the data, we achieved a very strong correlation of 91.67% with the BET index and 92.29% with BET Plus.

Initially, not considering the seasonal factors, we attempted to predict the BET and BET Plus indices based on their own previous day value (taking into account the short-term dependency, the concept called volatility clustering) and the value of the previous day of the S&P500 index. The hypothesis is that, including a small lag over time, the indices studied follow, however, faithfully, the evolution of the main stock exchanges on the international level.

We present below the estimated models along with their most important properties.

$$DBET_t = 3.08 + 0.26 * DSP_{t-1} + 0.99 * DBET_{t-1} + \varepsilon_t$$

$$DBETPLUS_t = 1.27 + 0.26 * DSP_{t-1} + 0.99 * DBET_{t-1} + \varepsilon_t$$

Table 2 - Statistical properties of models

Index	Root Mean Square Error	F-statistic	Probability of Fischer-Snedecor test
BET	0.107	29463.05	0.000
BET Plus	0.108	32068.32	0.000

Although the estimated models are statistically significant, the forecast of the two indices based on them was not performing well. It is obvious that, using this type of technical analysis of the market, one cannot beat the market. It is to be noticed that the predicted values correspond to the actual values in very few days of the period considered. Moreover, for both indices, the upward trend of their evolution from the beginning of the period studied until the beginning of the third quarter of 2017 was validated. Otherwise, neither the magnitude nor the sign of the fluctuations were correctly predicted.

However, in order to improve this analysis, we explicitly admitted the existence of trends and included in the two models the effect of the seasonal factors in the attempt to explain precisely the fluctuations of the data series. We excluded the December analysis as a seasonality factor to avoid the dummy trap. The models obtained are as follows.

$$DBET_t = 7.47 + 0.06 * DSP_{t-1} + 0.01 * TR_t + 0.02 * mth1 + 0.06 * mth2 + 0.09$$
$$* mth3 + 0.09 * mth4 + 0.07 * mth5 + 0.05 * mth6 + 0.03 * mth7$$
$$+ 0.04 * mth8 + 0.02 * mth9 - 0.01 * mth10 - 0.01 * mth11 + \varepsilon_t$$

$$DBETPLUS_t = 3.49 + 0.04 * DSP_{t-1} + 0.08 * TR_t + 0.02 * mth1 + 0.01 * mth2$$
$$+ 0.05 * mth3 + 0.09 * mth4 + 0.06 * mth5 + 0.01 * mth6 + 0.03$$
$$* mth7 + 0.02 * mth8 + 0.02 * mth9 - 0.01 * mth10 - 0.01 * mth11$$
$$+ \varepsilon_t$$

Both models are statistically significant, but what is relevant in the analysis is that the estimation confirmed the significant effect of the seasonal factors. However, based on the probability associated with the t-Student test, it is to be noted that the seasonal effects of October and November are not significant. Based on these models, the results of the forecast are graphically represented below.
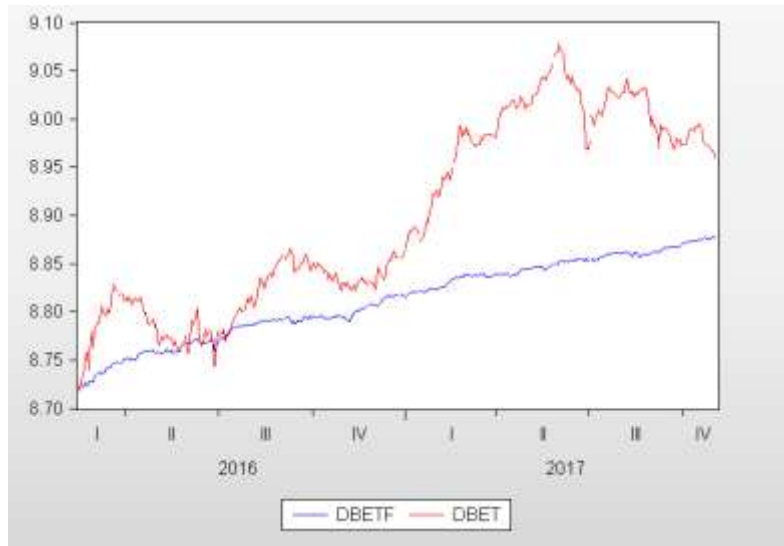

FIG. 3. Prediction of BET index based on short-past evolution of S&P500

FIG. 4. Prediction of BET Plus index based on short past evolution of S&P500

The graphs presented certainly point out to a more qualitative forecast, certainly due to the inclusion of seasonal factors, but also to the fact that the chosen method is static, which means that the predicted value at a certain iteration is based only on past real values of the index, and not the previously predicted values. Thus, in quantitative terms, we conclude that the predicted values of the BET index in the analyzed period correspond to 96.64% with the real values, and the predicted values of the BET Plus index are correlated with the real values in a higher proportion of 96.91%.
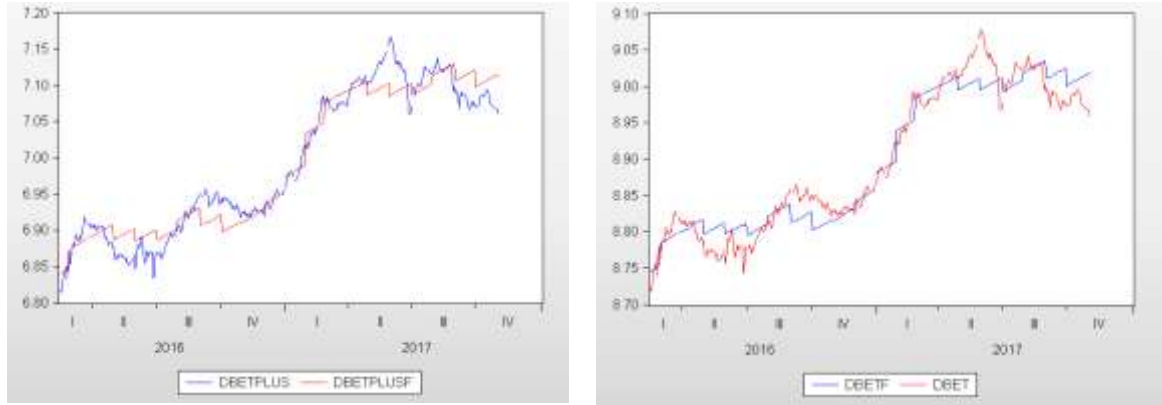
FIG. 5. BET Plus and BET index prediction based on seasonality and short term evolution of NYSE index

Based on these results, the international players' evolution is certainly an important determinant of the evolution of the Bucharest Stock Exchange and we naturally want in a more detailed analysis to test whether the BET and BET Plus indices are influenced by the recent-short or long-term evolution of the S&P500 US market index. In this attempt, we have improved the model by taking into account the values for the last 5 days of the S&P500 index, thus obtaining the following estimation.

$$DBET_t = 8.92 + 0.04 * DSP_{t-1} + 0.05 * DSP_{t-2} + 0.04 * DSP_{t-3} - 0.02 * DSP_{t-4}$$
$$+ 0.12 * DSP_{t-5} + 0.01 * TR_t + 0.02 * mth1 + 0.06 * mth2 + 0.09$$
$$* mth3 + 0.09 * mth4 + 0.07 * mth5 + 0.05 * mth6 + 0.03 * mth7$$
$$+ 0.04 * mth8 + 0.02 * mth9 - 0.01 * mth10 - 0.01 * mth11 + \varepsilon_t$$

$$DBETPLUS_t = 7.18 + 0.03 * DSP_{t-1} + 0.04 * DSP_{t-2} + 0.04 * DSP_{t-3} - 0.02$$
$$* DSP_{t-4} + 0.11 * DSP_{t-5} + 0.01 * TR_t + 0.02 * mth1 + 0.06 * mth2$$
$$+ 0.09 * mth3 + 0.09 * mth4 + 0.07 * mth5 + 0.05 * mth6 + 0.03$$
$$* mth7 + 0.04 * mth8 + 0.02 * mth9 - 0.01 * mth10 - 0.01 * mth11$$
$$+ \varepsilon_t$$

In this approach, we note that, based on the Wald test, these newly introduced coefficients are not significantly different from 0 for both of the time series. Actual values and predicted values are shown below.

FIG. 6. Predictions on BET and BET Plus index based on past 5-days evolution of NYSE index

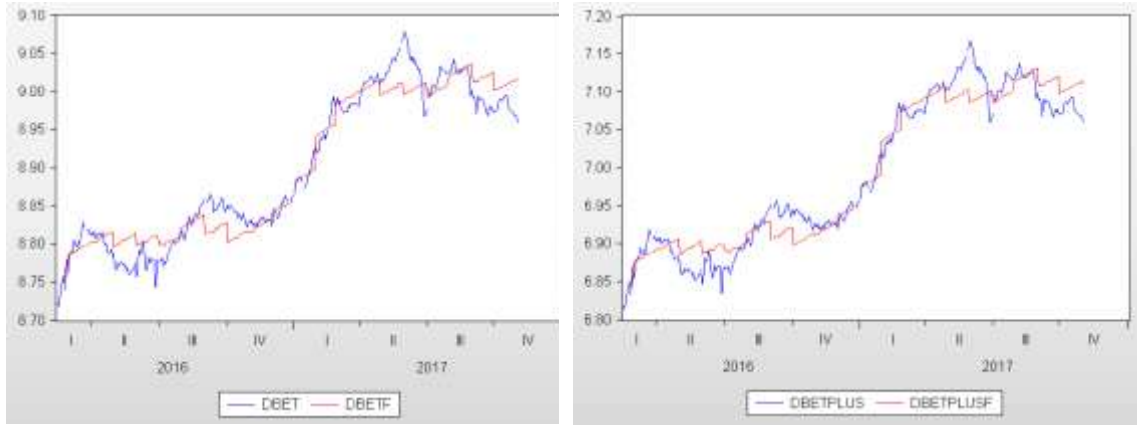For the BET index, the correlation coefficient between the actual values and the predicted values is 96.57% and for the BET Plus index it is 96.85%. In both cases, predictions based on the recent past value are better than considering a longer period. This behavior is probably due to the dynamic evolution of stock market indices; so, even if some shocks emerge, they quickly disappear and it would not be effective to include more past days in predicting the indexes in the analysis.

Finally, the most important determinant in the forecast being clearly the inclusion of the seasonality phenomenon, we added to the model that includes the seasonal factors a moving average term, an autoregressive term, and finally an AR and a MA term simultaneously. Based on the initial model and the three additional models, we have computed the average predicted values of the BET index evolution.

The models obtained are statistically significant, considering the validity of the Fischer-Snedecor test, and they take the following form.

$$\begin{aligned}
DBET_t = 3.31 &+ 0.98 * DBET_{t-1} + 0.25 * DSP_{t-1} + 0.01 * TR_t + 0.01 * mth1 + 0.01 \\
&* mth2 + 0.01 * mth3 + 0.01 * mth4 + 0.01 * mth5 + 0.01 * mth6 \\
&+ 0.01 * mth7 + 0.01 * mth8 + 0.01 * mth9 + 0.01 * mth10 + 0.01 \\
&* mth11 + \varepsilon_t
\end{aligned}$$

$$\begin{aligned}
DBET_t = 6.43 &+ 0.85 * \varepsilon_{t-1} + 0.11 * DSP_{t-1} + 0.01 * TR_t + 0.01 * mth1 + 0.01 \\
&* mth2 + 0.01 * mth3 + 0.01 * mth4 + 0.01 * mth5 + 0.01 * mth6 \\
&+ 0.01 * mth7 + 0.01 * mth8 + 0.01 * mth9 - 0.01 * mth10 - 0.01 \\
&* mth11 + \varepsilon_t
\end{aligned}$$

$$\begin{aligned}
DBET_t = 3.45 &+ 0.98 * DBET_{t-1} + 0.01 * \varepsilon_{t-1} + 0.24 * DSP_{t-1} + 0.01 * TR_t + 0.01 \\
&* mth1 + 0.01 * mth2 + 0.01 * mth3 + 0.01 * mth4 + 0.01 * mth5 \\
&+ 0.01 * mth6 + 0.01 * mth7 + 0.01 * mth8 + 0.01 * mth9 + 0.01 \\
&* mth10 + 0.01 * mth11 + \varepsilon_t
\end{aligned}$$

The overall representation of these time series together with the predictions obtained are presented below.

FIG. 7. Comparison between methods

In an attempt to achieve higher performance, predictions were computed by means of the static method, so only real values were taken into account. The average of the predictions is thus the most elaborate methodology among those tested in this study to predict the BET and BET Plus indices. The average of predictions along with real values are shown in FIG. 8.
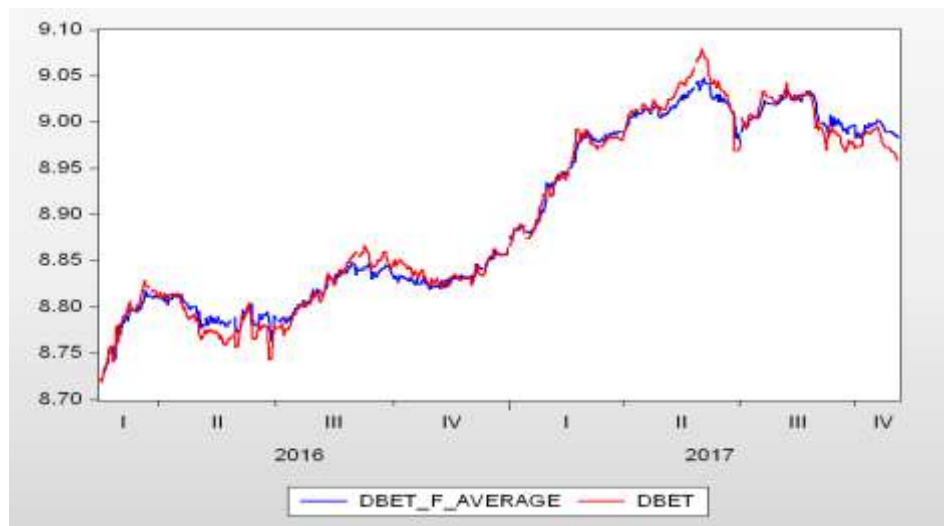
FIG. 8. Actual vs predicted values based on Forecasting Average best method

Although the Forecasting Average model based on the four sub-models is the most qualitative, the differences between the actual and the predicted values are still significant, with no significant profitability to be obtained if one uses analytical forecasting techniques. The predicted values increase when the real ones increase as well, and they decrease when the real ones show a negative evolution, but the magnitude of the fluctuation differs, sometimes significantly.

Moreover, as we have previously proved, it is not more significant to consider the last 5 days of the S&P500 than the previous day exclusively; so, any forecast obtained using analytical methods is anyway valid in the short term, but the evolution will instantly change significantly as compared to the one expected by speculators.

Once again, we have demonstrated that there are no profitable opportunities for analysts. Not even including correlation to international stock markets, we cannot predict the evolution of stock indices in order to gain surprofit.

As the quality of the forecast has increased considerably when considering the seasonal effect, we conclude that, it is important to thoroughly study and take into account all the determinants that exist in capital market. Even in this ideal case, shocks will remain unpredictable. However, the costs of these analyzes, as well as the transaction costs and other costs that appear on the real financial market, make these analyzes meaningless when players want a significant profit.

3.3 Predictions based on learning techniques – Naïve Bayes' Classifier and KNN.

The naive Bayesian classifier is a prediction technique whereby the data that compose the training set (70% of the data set) is converted into a frequency table. The probability associated with each event (selling or keeping the portfolio in the same structure – hold – as considered optimal strategy) is calculated and then computed the probability table based.

Using the Bayes' posterior probability formula, these probabilities are computed for all classes – the posterior probability for the sell strategy and the posterior probability for the hold strategy. Being the first index of Bucharest Stock Exchange, we considered relevant to include the BET index in the analysis.

A simple strategy is as follows: if there are positive fluctuations, the player will choose to choose the Buy event; otherwise, one will choose the "Hold" event. In fact, considering the Naive Bayes classifier, at each iteration, the class with the highest posterior probability will be chosen. Based on these assumptions, the observations in the test set (30% of the dataset) predicted the correct strategy in 62% of the cases.

```
Confusion Matrix and Statistics


preds  Buy Hold
  Buy   80   50
  Hold   0    1


                Accuracy : 0.6183
                  95% CI : (0.5294, 0.7018)
     No Information Rate : 0.6107
     P-Value [Acc > NIR] : 0.4669

                   Kappa : 0.0238
 Mcnemar's Test P-Value : 4.219e-12

             Sensitivity : 1.00000
             Specificity : 0.01961
          Pos Pred Value : 0.61538
          Neg Pred Value : 1.00000
              Prevalence : 0.61069
          Detection Rate : 0.61069
    Detection Prevalence : 0.99237
       Balanced Accuracy : 0.50980

        'Positive' Class : Buy
```
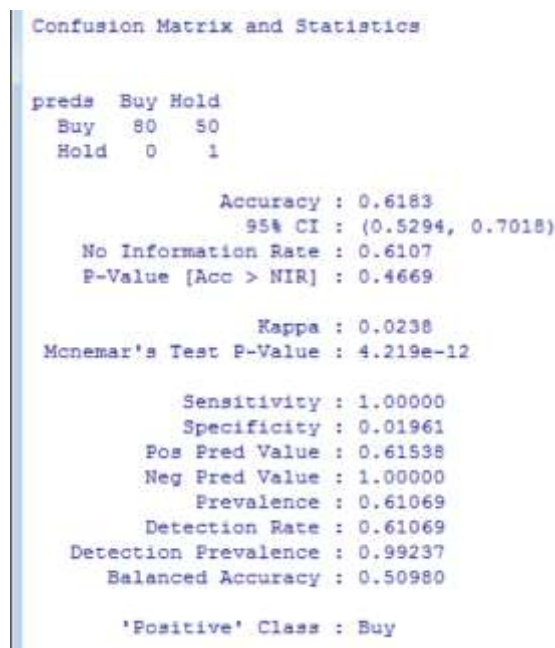
FIG. 9. The performance of Naïve Bayes classification algorithm

Furthermore, based on the confusion matrix in FIG. 9., 80 observations were predicted in the correct "Buy" class, and only one observation was predicted correctly in the "Hold" class. However, Cohen's coefficient shows a very low value of 2%. This shortcoming can be corrected by trying, for instance, to redefine the category variable (i.e., the strategy), adding more variables in the analysis, or even selectively choosing the determinants that certainly the strategy decision.

As for the k nearest neighbors method, we used the same test set and training set, and for both methodologies, we set the same seed to ensure the possibility of comparing their performance. The KNN method is one of the pattern recognition procedures, therefore, based on observations from the training set, for which we consider the real class known, we calculated the Euclidean distance between the observations in the test set and those in the training set. By initially setting k = 3, a new observation is classified based on the simple majority of the classes of the nearest three neighbors. For k = 10, we proceeded analogously and the results are provided below.
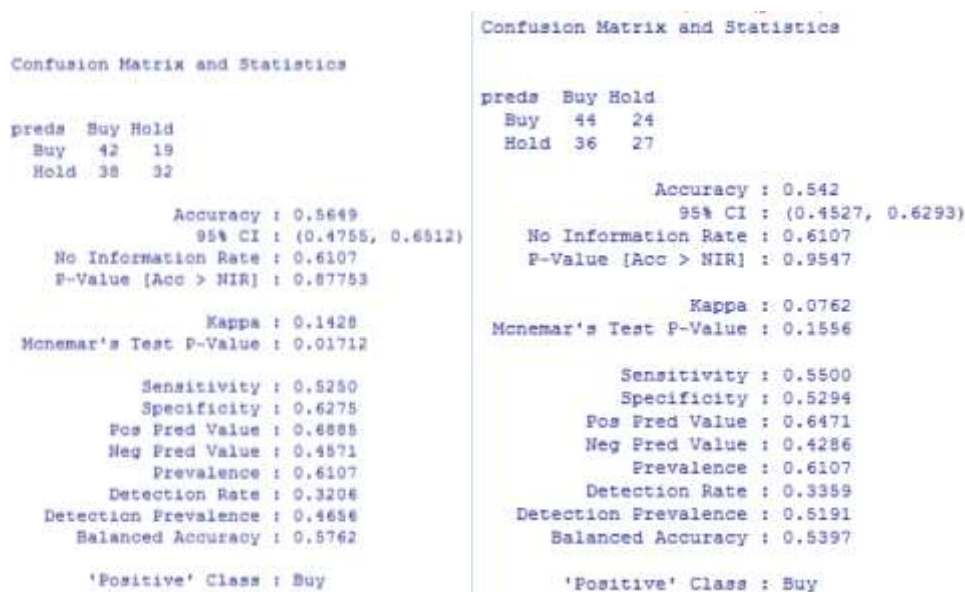

FIG. 10. Performance of the KNN method (K=3 and K=10)

Thus, we conclude that the KNN algorithm is more efficient than the naive Bayes classifier. Although accuracy is lower, it is important that the value of Cohen's coefficient increases considerably, so that for every class the ratio between the correctly predicted values and all values belonging to that class is balanced.

However, it is remarkable that considering only three neighbors, the quality of the algorithm was higher than choosing ten neighbors. This may be, for instance, due to the fact that the choice of the class for a non-classified observation by the simple majority determines that choosing an odd number of closest neighbors is certainly a rational decision.

3.4 Monte-Carlo Simulation of the behavior of Bucharest Stock Exchange. In an attempt to establish a stock market optimal strategy, a player must consider all the trajectories that a stock index may follow in a given period of time. In this analysis, we chose to track the potential fluctuations of the BET index, being the first index of the Romanian stock exchange, which has highlighted the performance of the most liquid and active ten traded companies since the beginning of the BSE.

We considered it important not only to observe its past values but, above all, the vast trajectories that it could have followed in the context of different states of nature.

By illustrating them as shown in FIG. 11, a decision maker whose purpose is to obtain profit from stock market games, will make rational decisions, because by conducting this analysis, one can compute the probability of assuming that a particular event, of all possible events, will take place and will generate the expected profit.

In the context of such a dynamic system, this assumption is crucial, because if another event occurs, it will not bring the expected revenue or, in a pessimistic way of thinking, it can even produce significant losses.

In fact, the player should propose a strategy, quantifying the effects of a possible error in which the expected scenario will not occur.
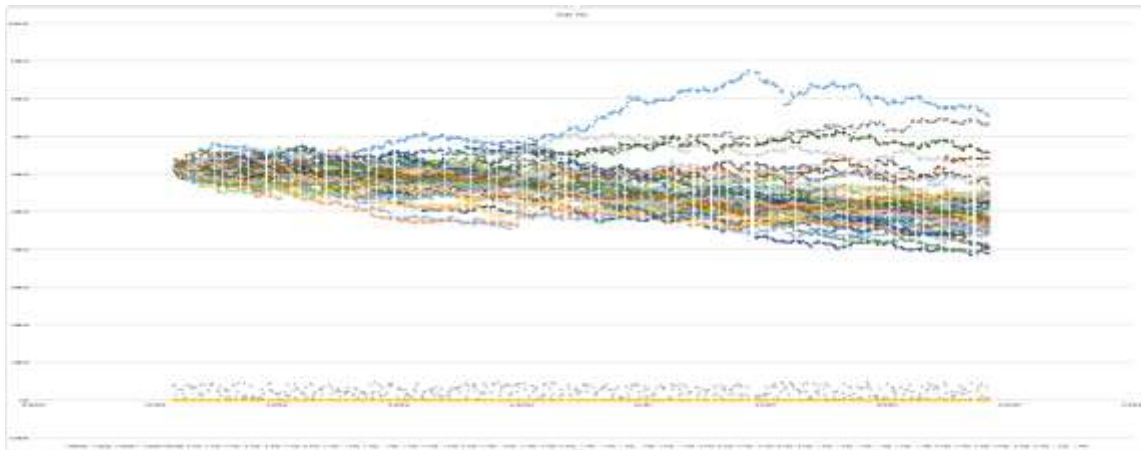

FIG. 11. Monte Carlo simulation of possible trajectories of BSE

3.5 Predictions based on Support Vector Machine learning technique. Predictions based on vector-based machines imply the division of the dataset into two subsets, as follows: one-third represents the test set, and two-thirds constitute the training set. To train the VSM, we considered the kernel to be linear, as in almost all cases, this type of kernel is suitable for the financial data series.

For the entire dataset, an UPDOWN categorial variable was included, so that when the BET index shows positive fluctuations, the variable takes the value Up, and when decreases occur, the variable takes the value Down.

Based on the training set, the support vector machine was trained to be able to predict the daily BET index fluctuations for the period included in the test set.

Subsequently, the results obtained based on the supervised learning algorithm were compared with the actual fluctuations.

```
[1] 145
> na.omit(settestare[,2])
  [1] Up    Up    Up    Up    Down Up    Down Down Up    Down Down Down Up    Up    Down
 [16] Up    Down Up    Down Up    Up    Up    Up    Down Up    Up    Down Down Up    Up
 [31] Down Up    Up    Up    Up    Down Down Up    Down Up    Down Down Up    Down Down
 [46] Up    Up    Down Up    Up    Up    Down Up    Down Down Down Up    Down Down Up
 [61] Up    Up    Down Down Up    Up    Up    Down Down Down Up    Down Down Down Down
 [76] Down Down Down Up    Up    Up    Down Up    Up    Down Up    Down Down Down Up
 [91] Up    Up    Down Up    Down Up    Down Down Up    Up    Up    Up    Up    Up    Up
[106] Down Down Up    Down Up    Up    Up    Up    Up    Up    Down Down Down Up    Down
[121] Down Up    Down Up    Up    Down Down Up    Down Up    Up    Down Down Up    Down
[136] Up    Up    Down Down Down Up    Up    Up    Up    Down
Levels: Down Up
> classAgreement(confuzie)
$diag
[1] 0.537931
```

FIG. 12. Performance of Support Vector Machine learning algorithm

The prediction in FIG. 12. coincide 53.79% with the actual results for the 145 observations included in the test set. Although the algorithm has made accurate predictions in more than half of all cases, it does not guarantee that there will be profitable opportunities.

However, the purpose of this analysis was to prove the possibility of predicting the sign of the Bet indexes, and not the magnitude of these fluctuations.

This distinction should be highlighted as a decision-maker cannot compute the expected monetary value of making a decision (choosing a particular strategy), even if assuming that they ~~he~~ know for sure whether or not the index will show a positive or negative trend the next day. Stock market players generally have a varied portfolio of shares that can include both shares whose evolution follows the market, and other that do the contrary, so the magnitude of a change in the evolution of the BET index is of considerable importance. We can therefore conclude that the accurate forecast of 53.79% of the BET index trajectory does not add value to the market players' strategies.

CONCLUSIONS

As proven in this paper, the challenge of the random walk theory on the Romanian stock market, to its players is outlined by the following context: if markets are efficient, then the prices of the shares at any time will represent the consistent estimation of their intrinsic value. In this respect, the fundamental analysis is useful if and only if the analyst has new information, which is not yet available on the market, so it was not considered in the formation of current prices. If the analyst does not have new information or contexts not yet exploited in the market, then the optimal decision should be choosing shares in a portfolio or transaction through a purely random procedure.

In essence, the tests performed were not able to reject the hypothesis of describing the evolution of prices on the Romanian stock market as a random walk process. Further work may be developed. The present study could be extended in any of the following directions.

- A random walk study can be carried out in the context of efficient market using data from the Sibiu Stock Exchange, in order to supplement the framework for the Romanian stock market.
- The same analysis can be carried out using other analysis tools to improve the results and conclusions obtained in this paper.
- The period under analysis can be extended in order to achieve more consistent results.

REFERENCES

[1] Ayadi, O.F., Pyun, C.S., „*An application of variance ratio test to the Korean securities market*", Journal of Banking and Finance, 18(4),pp.643-658

[2] Chaibi L.F., „ *Testing the random walk: The case of Hong Kong stock exchange*", Journal of Empirical Studies,2014, 1 (2), pp.54-61

[3] Fama E., „ *Efficient capital market: A reviw of theory and empirical work*", Journal of Finance, 25(2), 1970, pp.387-417

[4] Kim, J.H., Shamsuddin A., „*Are Asian stock market efficient? Evidence from new multiple variance ratio tests*", Journal of Empirical Finance,15(3), 2008, pp.518-532

[5] Lim, P.K., Habibullah, M.S., Hinich, M.J., „ *The weak form efficiency of Chinese stock markets: Thin trading, nonlinearity and episodic serial dependencies*", Journal of Emerging Market Finance, 8(1), 2008, pp.133-163

[6] Lo, A.W., Mamaysky,H., Wang, J., *„Foundations of Technical Analysis: Computational Algorithms, Statistical Inference and Empirical Implementation”*, The Journal of Finance, 55(4), Papers and Proceedings of the Sixtieth Annual Meeting of the American Finance Association, Boston, Massachusetts, January 7-9, 2000 (Aug., 2000), pp. 1705-1765

[7] Mishra, A., Mishra,V., Smyth R., *„ The random walk hypothesis on the Indian stock market”*, Journal of Emerging Markets Finance and Trade, 51 (5), 2015, pp. 879-892

[8] Wooldridge, J., *„Introductory Econometrics. A Modern Approach”,* Pearson, 2009

[9] www.bvb.ro.

[10] http://scikit-learn.org/stable/modules/svm.html.

VARMA-TSALLIS ENTROPY: PROPERTIES AND APPLICATIONS

Romică TRANDAFIR[*], **Vasile PREDA**[**], **Sorin DEMETRIU**[*],
Ion MIERLUȘ-MAZILU[*]

[*]Technical University of Civil Engineering, Bucharest, Romania
(romica.trandafir@utcb.ro, demetriu@utcb.ro, mmi@utcb.ro),
[**]University of Bucharest, Romania (vasilepreda0@gmail.com)

Abstract: Entropy represents a universal concept in science suitable for quantifying the uncertainty of a series of random events. In this paper we obtain a new type of entropy named Varma-Tsallis entropy starting from Tsallis entropy and Varma entropy. Some properties and applications of the proposed entropy in water engineering are presented.

Keywords: entropy, principle of maximum entropy, Lagrange multipliers, maximum annual discharge

1. INTRODUCTION

In this paper we present some types of entropy, the connections between them and we propose a new type of entropy starting from Tsallis entropy and Varma entropy, namely, the Varma-Tsallis entropy. For this proposed entropy we present some properties and a procedure that shows how it can be applied in practical applications. An application of the proposed entropy for the determination of the cumulative distribution function (cdf) for the recorded annual discharges of the Prut river and the Somes river.

2. TYPES OF ENTROPY

2.1. Boltzmann-gibbs-shannon entropy (referred to as the shannon entropy, 1948) [7]

o **Discrete case**

Let X be a random variable that takes on values x_i, $i = \overline{1, N}$, that occur with probabilities p_i, $0 \leq p_i \leq 1$, $i = \overline{1, N}$ and $\sum_{i=1}^{N} p_i = 1$. The information gain from the occurrence of any event x_i, is given by

$$\Delta H(x_i) = -\log_2(p_i) \tag{1}$$

i.e. the information gained is the logarithm of inverse of the probability of occurrence. For the all N events the average of information gain H_S can be expressed as

$$S = H_S = \sum_{i=1}^{N} p_i H(x_i) = -\sum_{i=1}^{N} p_i \log_2(p_i) \tag{2}$$

Equation (2) is the Shannon entropy, also called *informational entropy*.

o **Continuous case**

If the random variable is non-negative continuous with a probability density function (pdf) $f(x)$, the the Shannon entropy can be writen as:

$$S = H_S(X) = H_S(f) = -\int_0^\infty f(x)\log_2(f(x))dx \tag{2'}$$

2.2. Renyi entropy (1961) [6]

Renyi proposed a generalized entropy of order α as

o **Discrete case**

$$R = H_\alpha(X) = \frac{1}{1-\alpha}\log\left(\sum_{i=1}^N p_i^\alpha\right) , \ \alpha > 0 , \ \alpha \neq 1 \tag{3}$$

Remark. Renyi's entropy contains the Shannon entropy as a special case ($\lim_{\alpha \to 1} H_\alpha(X) = S$).

o **Continuous case**

$$R = H_\alpha(X) = \frac{1}{1-\alpha}\log\left(\int_0^\infty f_X(x)^\alpha dx\right) , \ \alpha > 0 , \ \alpha \neq 1 \tag{3'}$$

2.3. Varma entropy (1966) [12]

o **Discrete case**

$$V = H_\alpha^\beta(X) = \frac{1}{\beta-\alpha}\sum_{i=1}^N p_i^{\alpha+\beta-1} , \ \beta-1 < \alpha < \beta , \ \beta \geq 1 \tag{4}$$

o **Continuous case**

$$V = H_\alpha^\beta(X) = \frac{1}{\beta-\alpha}\ln\left(\int_0^\infty f_X(x)^{\alpha+\beta-1} dx\right) , \ \beta-1 < \alpha < \beta , \ \beta \geq 1 \tag{4'}$$

Remark:

The Varma entropy includes, as particular cases, the Renyi entropy.

$$\lim_{\beta \to 1} H_\alpha^\beta(X) = R_\alpha(X) = -\frac{1}{\alpha-1}\log\int_0^\infty (f(x))^\alpha \, dx \tag{5}$$

and the Shannon entropy

$$\lim_{\substack{\beta \to 1 \\ \alpha \to 1}} H_\alpha^\beta(X) = S = -\int_0^\infty f(x)\cdot\log(f(x))dx \tag{5'}$$

2.4. Kapur entropy (1967) [3]

o **Discrete case**

$$K = H_{\alpha,\beta}(X) = \frac{1}{\beta-\alpha}\frac{\sum_{i=1}^N p_i^{\alpha+\beta-1}}{\sum_{i=1}^N p_i^\beta} , \ \alpha \neq \beta , \ \alpha > 0, \ \beta > 0 \tag{6}$$

o **Continuous case**

$$K = H_{\alpha,\beta}(X) = \frac{1}{\beta-\alpha}\frac{\int_0^\infty (f_X(x))^\alpha \, dx}{\int_0^\infty (f_X(x))^\beta \, dx} , \ \alpha \neq \beta , \ \alpha > 0, \ \beta > 0 \tag{6'}$$

2.5. Tsallis entropy (1988) [10]
o **Discrete case**

$$T_m = H_m(X) = \frac{1}{m-1} \sum_{i=1}^{N} p_i \left(1 - p_i^{m-1}\right) \ , \ m \neq 1, \ m \in R \tag{7}$$

o **Continuous case**

$$T_m = H_m(X) = H_m(f) = \frac{1}{m-1} \int_0^\infty \left\{ f_X(x) - [f(x)]^m \right\} dx \ , \ m \neq 1 \ , \ m \in R \tag{7'}$$

Remarks:
1. For $m \to 1$ Tsallis entropy converges to Shannon entropy.
2. For $m < 0$ Tsallis entropy is concave and for $m > 0$ Tsallis entropy is convex.
3. For all m the Tsallis entropy decreases as m increases.

2.6. Varma-Tsallis entropy
If we denote
$$m = m + r - 1 \tag{8}$$

o **Discrete case**

$$VT_{m,r} = H_{m,r} = \frac{1}{m-r} \sum_{i=1}^{N} P_{i(1-p_i^{m+r-2})} = \frac{m-1}{m-r}\left[\frac{1}{m-1} \sum_{i=1}^{N} p_i\left(1 - p_i^{m-1}\right)\right], m \neq 1, r \neq 1 \tag{9}$$

$$VT_{m,r} = \frac{m-1}{m-r} T_m, m \neq 1, r \neq 1$$

o **Continuous case**

$$VT_{m,r} = H_{m,r}(X) = \frac{1}{m-r}\left(1 - \int_0^\infty f_x(x)^{m+r-1} dx\right), m, r \neq 1 \tag{9'}$$

The equation (9') becomes

$$VT_{m,r}(X) = \frac{m-1}{m-r}\left\{\frac{1}{m-1}\left(1 - \int_0^\infty f_x(x)^m dx\right)\right\}, m \neq 1, r \neq 1 \tag{9''}$$

Thus

$$VT_{m,r} = \frac{m-1}{m-r} T_m, m \neq 1, r \neq 1 \tag{10}$$

Let $\alpha = \dfrac{m-r}{m+r-2}$ be a parameter. Then

$$T_m = \alpha VT_{m,r}, m \neq 1, r \neq 1 \tag{10'}$$

3. PROPERTIES OF THE VARMA-TSALLIS ENTROPY

3.1. Concavity, convexity
It can be shown that for

$$0 < a < 1, \ P = \{p_i\}_{1 \leq i \leq N} \ , \ Q = \{q_i\}_{1 \leq i \leq N} \ , G = \{g_i = ap_i + (1-a)q_i\}_{1 \leq i \leq N} \tag{11}$$

then
i)
$$\alpha \cdot VT_{m,r}(G) \geq a \cdot \alpha \cdot VT_{m,r}(P) + (1-a)\alpha \cdot VT_{m,r}(Q) \tag{12}$$

for $m.\alpha > 0$,

ii)
$$\alpha \cdot VT_{m,r}(G) \leq a \cdot \alpha \cdot VT_{m,r}(P) + (1-a)\alpha \cdot VT_{m,r}(Q) \tag{13}$$
for $m > 0, \alpha < 0$,

iii)
$$\alpha \cdot VT_{m,r}(G) \leq a \cdot \alpha \cdot VT_{m,r}(P) + (1-a)\alpha \cdot VT_{m,r}(Q) \tag{14}$$
for $m < 0, \alpha > 0$,

iv)
$$\alpha \cdot VT_{m,r}(G) \geq a \cdot \alpha \cdot VT_{m,r}(P) + (1-a)\alpha \cdot VT_{m,r}(Q) \tag{15}$$
for $m < 0, \alpha < 0$.

$$VT_{m,r,extreme} = \frac{m+r-2}{m-r}\frac{N^{m-1}-1}{m-1}$$

3.2 maximum value

It is well known that the Tsallis entropy attains an extreme value for all values of m when all p_i, $i = \overline{1,N}$ are equal, i. e. $p_i = \dfrac{1}{N}$ and this extreme value is

$$T_{m,extreme} = \frac{N^{m-1}-1}{m-1} \tag{16}$$

For $m > 0$ this extreme value is a maximum value and for $m < 0$ this extreme value is a minimum value. Considering the equations (10), (10') the extreme value for Varma-Tsallis entropy will be given by

$$VT_{m,r,extreme} = \frac{m+r-2}{m-r}\frac{N^{m-1}-1}{m-1} \tag{17}$$

4. THE PRINCIPLE OF MAXIMUM ENTROPY

Considering the following principles of ancient wisdom:
- "speak truth and nothing but truth
- make use of all the given information you are given and scrupulously avoid using the information not given to you
- make use of all the given and be maximally uncommitted to the missing information or be maximally uncertain about it" [9], E. T. Jaynes (1957) [1,2] formulated the principle of maximum entropy (POME), which states that *"one should choose the distribution that has the highest entropy, subject to the given information"*.

The implication here is that POME considers all of the given information and, at the same time, avoids consideration of any information that is not given. This is consistent with Laplace's principle of insufficient reason (or principle of indifference), according to which all outcomes of an experiment should be considered equally likely unless there is information to the contrary.

Therefore, POME enables entropy theory to achieve the probability distribution of a given random variable [8].

To obtain the probability distribution of a given random variable by POME, it can be used the following procedure:
- fix the kind of entropy, in this case Varma-Tsallis entropy (9')
- give the constraints
- maximize the entropy by POME
- obtain the probability distribution according to constraints
- determine the Lagrange multipliers
- determine the maximum entropy.

4.1 SPECIFICATION OF CONSTRAINTS

Given a sample of random variable X, $\left(x_1, x_2, ..., x_N\right)$, a type of restriction can be given by the following equations

$$\int_0^\infty x^k f(x)dx = \overline{x^k} \ , \ k = 0,1,2,3... \tag{18}$$

where $\overline{x^k}$, $k = 0,1,2,3...$ are empirical moments of random variable X.

Remark: In water engineering, empirical moments $k=0,1,2,3$ are considered.

The constraints (18) are not sufficient to determine $f(x)$ uniquely, because there may be many, even infinity of probability distributions satisfying (18).

4.2. Entropy maximization using lagrange multipliers

To determine $f(x)$ we should maximize the Varma-Tsallis entropy (9') subject to (18) using the method of Lagrange multipliers.

There are two fortunate circumstances favoured the great success of the POME, since in all optimization problems the difficulties arise when we have to decide whether
- the extreme value found is a maximum or minimum
- the maximum obtained is local or global
- the non-negativity constraints are satisfied

namely: the Varma-Tsallis entropy function is a concave function and the pdf is always non-negative. The Lagrangian function L is given, in this case, by

$$L = \frac{1}{m-r}\left\{1 - \int_0^\infty \left(f(x)\right)^{m+r-1} dx\right\} - \lambda_0 \left\{\int_0^\infty f(x)dx - 1\right\} -$$
$$- \sum_{i=1}^k \lambda_i \left\{\int_0^\infty x^i f(x)dx - \overline{x^i}\right\} \ , \ k = 1,2,3... \tag{19}$$

where λ_i , $i = \overline{1,k}$ are the Lagrange multipliers.

Differentiating equation (19) with respect to $f(x)$ and equating the derivative to zero, we obtain:

$$\frac{\partial L}{\partial f(x)} = 0 = \frac{1}{m-r} - \frac{m+r-1}{m-r}\left[f(x)\right]^{m+r-2} - \sum_{i=0}^k \lambda_i x^i \ , \ k = 1,2,3... \tag{20}$$

Thus, the pdf of X is

$$f(x) = \left\{\frac{1}{m+r-1}\left[\lambda_0 - (m-r)\sum_{i=1}^k \lambda_i \cdot x^i\right]\right\}^{\frac{1}{m+r-2}} \ , \ k = 1,2,3... \tag{21}$$

Substituting equation (21) in equation (18) the result is, respectively:

$$\int_0^\infty \left\{\frac{1}{m+r-1}\left[\lambda_0 - (m-r)\sum_{i=0}^k \lambda_i \cdot x^i\right]\right\}^{\frac{1}{m+r-2}}dx = 1 \ , \ k = 1,2,3... \tag{22}$$

$$\int_0^\infty x^j \left\{\frac{1}{m+r-1}\left[\lambda_0 - (m-r)\sum_{i=0}^k \lambda_i \cdot x^i\right]\right\}^{\frac{1}{m+r-2}}dx = \overline{x^j}, \ j = \overline{1,k}, \ k = 1,2,3... \tag{23}$$

The system given by equations (22)-(23) do not have generally an analytical solution but can be solved with numerical methods.

Substitution of equation (21) in equation (9') leads to maximum Varma-Tsallis entropy

$$VT_{m,r} = \frac{1}{m-r}\left\{1 - \int_0^\infty \left[\frac{1}{m+r-1}\left(\lambda_0 - (m-r)\sum_{i=1}^k \lambda_i \cdot x^i\right)^{\frac{1}{m+r-2}}\right]^{m+r-1} dx\right\} , \ k = 1,2,3... \quad (24)$$

Last equation shows that the $VT_{m,r}$ of the distribution probability of X depends only on the constraints, since the Lagrange multipliers themselves depend on the same constraints.

5. APPLICATIONS

The design of the hydraulic structures like spillways, dykes or diversions is based on the maximum discharges corresponding to standard values of the annual probability of exceedance (usually in the range 1% - 0,1%). The length of the registered data rarely exceeds 50 years, which means that the empirical probabilities of exceedance of the maximum annual discharges are in the range 2-98%. The main problem is the real probability of exceedance of the outliers is not known, meaning that the values of the statistical parameters are influenced by the empirical probability which is assigned to the extreme values.

The method described above was used to determine the pdf for the maximum annual discharges of Prut River recorded at Radauti and the pdf of the data for the Somes River recorded at Satu Mare. The maximum annual discharges rates of the river Prut at Radauti gauge station between 1978 and 2015 (Fig.1) , X_i , $i = \overline{1,38}$, and of the river Somes at Satu Mare gauge station between 1928 and 1988 (Fig 2) , X_i , $i = \overline{1,64}$ are used to obtain the probability distributions of discharges in order to be able to make predictions of floods.

The measured discharge data are normalized

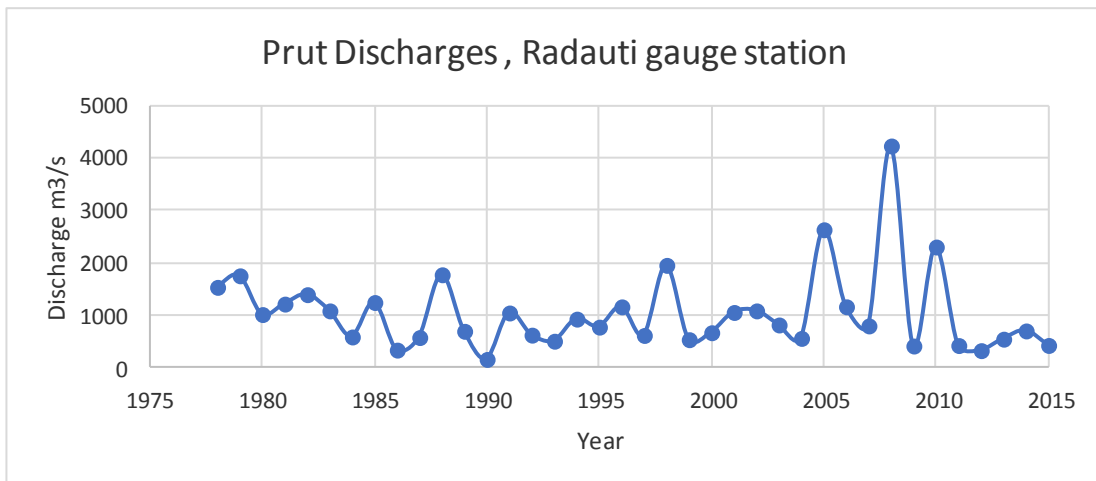$$x_i = \frac{X_i}{\max(X_i) - \min(X_i)} , \ i = \overline{1,38} , \ respectively \ i = \overline{1,64}$$

FIG.1. Maximum annual discharges of the Prut River recorded at Radauti station

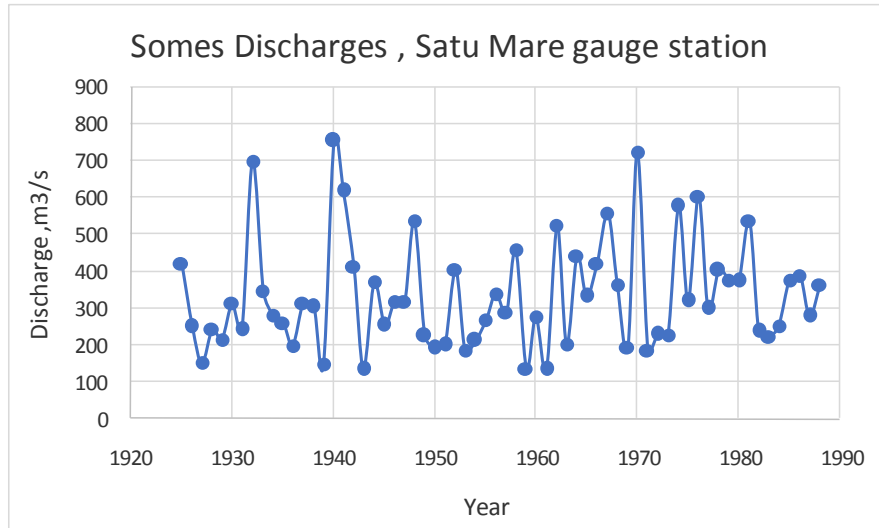$X\text{max}=4240$, $X\text{min}=163$, $w=X\text{max}-X\text{min}=4077$

FIG. 2. Maximum annual discharges of the Somes River recorded at Satu Mare station
Xmax=756 , Xmin=134 , $w=X$max$-X$min=622

The empirical moments of these normalized records make out the first data set:
$$\overline{x^1} = 0.25729 , \quad \overline{x^2} = 0.10104 , \quad \overline{x^3} = 0.05829$$
and for the second data set
$$\overline{x^1} = 0.53851 , \quad \overline{x^2} = 0.34619 , \quad \overline{x^3} = 0.26046$$

The non-linear equations systems for Lagrange's multiplier, considering $m = 2$, $r = 0.5$, $k = 2$ are solved for each case using a numerical method.

We obtain for the recorded discharges at Radauti
$$\lambda_0 = -2.579 , \quad \lambda_1 = -2.671 , \quad \lambda_2 = 1.104 .$$

and for the recorded discharges at Satu Mare
$$\lambda_0 = -2.002 , \quad \lambda_1 = -0.581 , \quad \lambda_2 = -0.251 .$$

Finally, the pdf for $\left(m = 2 , r = 0.5 , k = 2 \right)$ is given by

$$f(x, \lambda, m, r, k) = \frac{1}{m+r-1} \left[\lambda_0 - (m-r) \sum_{i=1}^{k} \left(\lambda_i \cdot x^i \right)^{\frac{1}{m+r-2}} \right]$$

and cumulative distribution function

$$F(x, \lambda, m, r, k) = \frac{1}{m+r-1} \int_{0}^{x} \left[\lambda_0 - (m-r) \sum_{i=1}^{k} \left(\lambda_i \cdot t^i \right)^{\frac{1}{m+r-2}} \right] dt$$
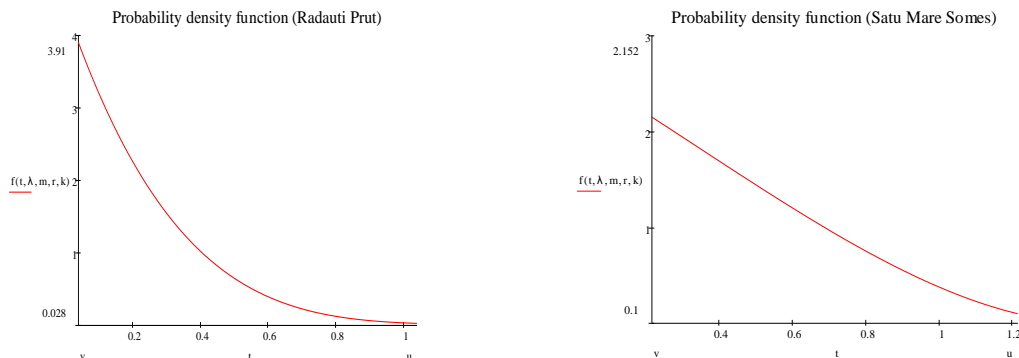

FIG. 3. Graphics of probability density functions for the two gauges

A comparison between the maximum annual discharge quantiles corresponding to different mean return intervals (periods) that are estimated using the obtained probability distributions and other reference probability distributions recommended in Statistical Hydrology is presented in Table 1.

Table 1. The maximum annual discharge quantiles corresponding to different mean return periods

Gauge station	Quantiles	Zα				
	Probabilities	α=0.90	α=0.95	α=0.98	α=0.99	α=0.995
	T (years)	**10**	**20**	**50**	**100**	**200**
Satu Mare, Somes	Varma-Tsallis	555	620	681	711	730
	Log-Pearson type III	532	627	756	857	963
	Lognormal 3-parameter	534	633	768	875	987
	Generalized Extreme Value	530	627	760	866	979
	Gumbel Max	529	612	720	801	882
Radauti, Prut	Varma-Tsallis	2138	2598	3127	3467	3742
	Log-Pearson type III	1953	2487	3272	3934	4662
	Lognormal 3-parameter	1924	2421	3135	3725	4361
	GEV (Generalized Extreme Value)	1900	2477	3415	4292	5348
	Gumbel Max	2055	2488	3048	3468	3886

5. CONCLUSIONS

In this paper we introduced a generalization of Tsallis entropy, called Varma-Tsallis entropy, highlighted some properties and showed how it can be used to determine a probability density function of a random variable. The method presented here was used for the recorded maximum annual discharges of the Prut River and the Somes River. The results we obtained show that the method is reliable.

REFERENCES

[1] Jaynes, E.T., *Information theory and statistical mechanics*, I. Phys. Rev. ,1957, 106, 620–630;

[2] Jaynes, E.T., *Information theory and statistical mechanics*, II. Phys. Rev. ,1957, 108, 171–190;

[3] Kapur, J.N., Generalized entropy of order α and type β . *The Math. Seminar*, 1967, 4, 78–94;.

[4] Madan Mohan Sati, Nitin Gupta, *On Partial Monotonic Behaviour of Varma Entropy and Its Application in Coding Theory*, Journal of the Indian Statistical Association, May 24, 2016;

[5] Papalexiou, S.M.; Koutsoyiannis, D., *Entropy based derivation of probability distributions: A case study to daily rainfall.* Adv. Water Resources, 2012, 45, 51–57;

[6] Renyi, A., *On measures of entropy and information.* In: Neyman, J. (Ed.), Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability,1961, Vol. I. University of California Press, Berkeley, CA, 547-56;

[7] Shannon, C.E., *A mathematical theory of communication*,1948, Bell System Technical Journal, 27, 379-423, 623-656;

[8] Singh V. P., Sivakumar, B., Cui H., *Tsallis Entropy Theory for Modeling in Water Engineering: A Review*, Entropy ,2017, 19, 641; doi:10.3390/e19120641, www.mdpi.com/journal/entropy;

[9] Singh, V. P., Fiorentino, M. (editors), *Entropy and Energy Dissipation in Water Resoources*, 3-20,1992, Kluwer Academic Publishers;

[10] Tsallis, C., *Possible generalization of Boltzmann-Gibbs statistics.* J. Stat. Phys. 1988, 52, 479–487;

[12] Varma, R.S., Generalization of Renyi's entropy of order α, *Journal of Mathematical Sciences*, 1966,1, 34-48.

Q-LOG EXPONENTIAL DISTRIBUTION IN URBAN AGGLOMERATION

Aurelia CĂȘARU[*], **Irina BĂNCESCU**[**], **Vasile PREDA**[**]

[*]PhD School of Mathematics, University of Bucharest, Romania
(aurelia.casaru@gmail.com)
[**]National Institute for Economic Research, Bucharest, Romania

Abstract: A purpose of urban theory is to describe how cities develop based on the number of inhabitants. Many statistical models and laws of growing of cities have been suggested, such as Zipf's, Mandelbrot-Zipf's and Gibrat's laws. This paper studies the urban agglomeration of Romania for the years between 2007-2017, using the q-log exponential distribution.

Keywords: q-log exponential distribution, Romania's cities distribution, Tsallis statistics.

1. INTRODUCTION

Cities develop in different ways all around the world, depending on many social-economic factors such as: economic growth of the area, economic activity, ethnic factors, infrastructure, and not only. Statisticians have studied the cities distribution for a long time, considering big cities, small cities or all together. Some laws of probability have Pareto tails for the lower and upper tails and different bodies: log-normal or Singh-Maddala [1,2]. Other statistical models used in urban theory are the q-exponential, Pareto, log-normal [3,4], and more recently, the q-log distribution family [5].

This paper studies the cities and municipalities agglomeration of Romania from 2007 to 2017 using the q-log exponential distribution. We apply the Kolmogorov-Smirnov test and graphically show how well this probability law models the data.

The paper is organized as follows. In Section 2, we present the q-log exponential distribution. Empirical analysis of Romania's cities population is performed in Section 3, while Section 4 concludes the paper.

2. METHODOLOGY

The q-log-location-scale exponential model was first introduced in 2018, as a submodel of q-log-location-scale distributions [5]. This class of distributions has been obtained by applying a q-logarithm Tsallis transformation to a baseline location-scale distribution. The q-logarithm and q-exponential functions are defined by

$$\log_q^T(x) = \log(x), if\, x > 0, q = 1 \text{ and } \log_q^T(x) = \frac{x^{q-1} - 1}{q-1}, if\, x > 0, q \neq 1$$

while $e_q(x) = \exp(x), if\, q = 1$, $e_q(x) = [1 + (q-1)x]^{1/(q-1)}, if\, q \neq 1, 1 + (q-1)x > 0$, and $e_q(x) = 0^{1/(q-1)}, if\, q \neq 1, 1 + (q-1)x \leq 0$ where q is a real parameter.
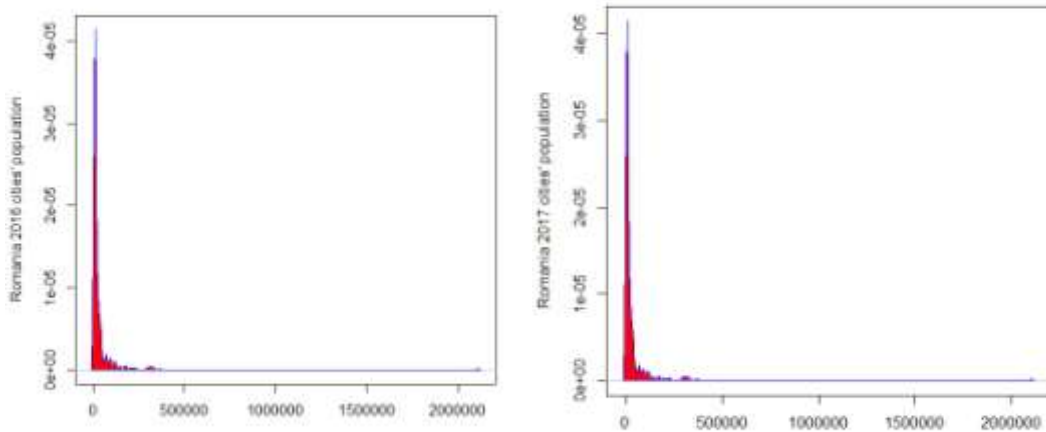
FIG. 1. Empirical density of city size "Romania2016" and "Romania2017" data

The q-log exponential distribution of parameters $\rho > 0$, $\theta > 0$, $q_1 \in (0,1)$, and $q_2 > 1$ is obtained considering as baseline distribution the q_1-exponential model. These two statistical models are presented next.

The q_1-exponential distribution is defined by the following cumulative distribution and density functions, respectively

$$G(x) = 1 - \left[e_{q_1}(-\rho x) \right]^{q_1}, \; g(x) = q_1 \rho e_{q_1}(-\rho x), q_1 > 0, \rho > 0, 0 < x < C_{q_1 \rho} \qquad (1)$$

Where

$$C_{q_1 \rho}(1) = \infty, if q_1 \in (0,1) and \, C_{q_1 \rho} = \frac{1}{\rho(q_1 - 1)}, if q_1 > 1.$$

The q-log exponential distribution (qLE) of parameters q_1, q_2, ρ, and θ is defined by the following distribution and density functions, respectively

$$F(x) = 1 - \left[e_{q_1}\left(-\rho \log_{q_2}^T \left(\frac{x}{\theta} \right) \right) \right]^{q_1}, \; f(x) = \frac{q_1 \rho}{\theta} \left(\frac{x}{\theta} \right)^{q_2 - 2} e_{q_1}\left(-\rho \log_{q_2}^T \left(\frac{x}{\theta} \right) \right), x > \theta$$

where $q_1 \in (0,1), q_2 > 1, \rho > 0$, while $\theta > 0$ is chosen as the minimum value of each dataset considered.

3. EMPIRICAL ANALYSIS

In this section, we discuss the analysis of cities' size distribution in Romania between 2007 and 2010. We perform the Kolmorogov-Smirnov test based on maximum likelihood estimation of parameters.

3.1 Data

Considering demographic data provided by INS we analyze the cities' population data using q-log exponential distribution. Some characteristics of the datasets considered such as maximum and minimum values, number of observations, measures of skewness and kurtosis, standard deviation, and mean are displayed in Table 1. It can be observed that for each datasets, the measure of kurtosis is extremely high, suggesting a heavy-tail distribution. Also, the skewness is high for these datasets.

Because all datasets have almost the same values for these measures, we have only displayed the empirical densities for years 2016 and 2017. The empirical densities of datasets "Romania2016", and "Romania2017" are displayed in Fig. 1.

Table 1. Descriptive statistics of Romania cities population

Year	Nr. of obs.	Mean	SD	Min	Max	Skewness	Kurtosis
2007	319	40,161.72	131,960.5	1,811	2,156,978	13.20	205.09
2008	319	40,051.41	131,970.5	1,784	2,158,816	13.23	205.78
2009	319	40,011.10	132,033.1	1,750	2,160,627	13.24	206.11
2010	319	39,960.93	132,077.1	1,732	2,162,037	13.26	206.41
2011	319	39,816.94	131,755.7	1,710	2,157,282	13.27	206.63
2012	319	39,671.51	131,410.2	1,704	2,151,758	13.27	206.70
2013	319	39,589.31	130,890.8	1,695	2,140,816	13.23	205.72
2014	319	39,428.26	129,371.9	1,674	2,110,752	13.14	203.55
2015	319	39,317.36	128,847.3	1,677	2,100,519	13.11	202.87
2016	320	39,243.34	129,026.9	1,684	2,107,399	13.15	203.86
2017	320	39,136.24	128,783.3	1,663	2,103,251	13.14	203.81

3.2 Parameter estimation and discussion

In order to assess if the q-log exponential model is appropriate to model the datasets considered, we utilize the Kolmogorov-Smirnov (KS) test. Also, we discuss the maximum likelihood estimation of the parameters. Distributions having few parameters are well fitted to data by maximum likelihood method. Hence, the maximum likelihood method applied for qLE model is described next.

Let $x_1, x_2, ..., x_n$ be a random sample of size n from $qLE(\rho, q_1, q_2, \theta)$ distribution of parameters $\rho > 0$, $\theta > 0$, $q_1 \in (0,1)$ and $q_2 > 1$. The log-likelihood function for the vector of parameters $\delta = (\rho, q_1, q_2, \theta)^T$ can be expressed as

$$l(\delta) = n\log(q_1) + n\log(\rho) - n\log(\theta) + \sum_{i=1}^{n} (q_2 - 2)[\log(x_i) - \log(\theta)] + \sum_{i=1}^{n} \log e_{q_1}\left(-\rho \log_{q_2}^T\left(\frac{x_i}{\theta}\right)\right)$$

The log-likelihood can be maximized by solving the nonlinear likelihood equations obtained by differentiating the equation above. However, the maximum likelihood estimator of θ is very simple

$$\theta = \min x_i$$

In other words, we choose the parameter θ to be equal to the smallest value of the dataset considered for analysis. The components of the score vector $U(\delta)$ are

$$U_\rho(\delta) = \frac{\partial l}{\partial \rho} = \frac{n}{\rho} - \sum_{i=1}^{n} \log_{q_2}^T\left(\frac{x_i}{\theta}\right)\left[e_{q_1}\left(-\rho \log_{q_2}^T\left(\frac{x_i}{\theta}\right)\right)\right]^{1-q_1}$$

$$U_{q_1}(\delta) = \frac{\partial l}{\partial q_1} = \frac{n}{q_1} - \sum_{i=1}^{n}\left\{\frac{\log\left[1 + (q_1-1)\left(-\rho \log_{q_2}^T\left(\frac{x_i}{\theta}\right)\right)\right]}{(q_1-1)^2} + \frac{\rho \log_{q_2}^T\left(\frac{x_i}{\theta}\right)}{(q_1-1)\left[1+(q_1-1)\left(-\rho \log_{q_2}^T\right)\left(\frac{x_i}{\theta}\right)\right]}\right\}$$

$$U_{q_2}(\delta) = \frac{\partial l}{\partial q_2} = \sum_{i=1}^{n} \log\left(\frac{x_i}{\theta}\right) - \sum_{i=1}^{n} \left[e_{q_1}\left(-\rho \log_{q_2}^{T}\left(\frac{x_i}{\theta}\right)\right)\right]^{1-q_1} \frac{\rho}{(q_2-1)^2} \left\{\left(\frac{x_i}{\theta}\right)^{q_2-1}\left[\log\left(\frac{x_i}{\theta}\right)(q_2-1)+1\right]+1\right\}$$

Solving the nonlinear likelihood equations above requires the use of numerical Methods that admit restrictions, such as an extended Nelder-Mead method [6].

Table 2 displays the maximum likelihood estimates of Romania's cities population. The standard errors were calculated by considering 500 bootstrapped samples, while the software used is R. All parameter estimates are highly significant as indicated by the low standard errors. The MLE of $\hat{\rho}$ ranges from 0.0284 to 0.0320, of $\hat{q}_1$ ranges from 0.20417 to 0.21728, while of $\hat{q}_2$ ranges from 4.58516 to 4.82080. The smallest city in Romania for year 2007 had a population of 1,811 inhabitants, while in 2017 this value decreased to 1,663 inhabitants. The estimate of parameter θ is taken as the minimum value of each dataset.

Table 2. Parameter estimates of q-log exponential distribution of Romania's cities population

Year	Parameter estimators (standard errors)			
	ρ	$\hat{q}_1$	$\hat{q}_2$	θ
2007	0.0320 (0.6836)	0.20417 (0.02952)	4.82080 (0.54629)	1,811
2008	0.0308 (0.65560)	0.20610 (0.03013)	4.79259(0.54740)	1,784
2009	0.0291 (0.68319)	0.20745 (0.03194)	4.76868 (0.57177)	1,750
2010	0.0288 (0.69591)	0.20941 (0.03049)	4.73258 (0.52923)	1732
2011	0.0284 (0.69357)	0.21164 (0.03059)	4.69677 (0.52232)	1,710
2012	0.0288 (0.68885)	0.21293 (0.03277)	4.67398 (0.54680)	1,704
2013	0.0292 (0.69707)	0.21453 (0.03077)	4.64519 (0.51047)	1,695
2014	0.0292 (0.67057)	0.21650 (0.03288)	4.60674 (0.51194)	1,674
2015	0.0304 (0.69024)	0.21687 (0.03106)	4.59454 (0.49957)	1,677
2016	0.0310 (0.67105)	0.21648 (0.03209)	4.60079 (0.54209)	1,684
2017	0.0304 (0.65839)	0.21728 (0.03227)	4.58516 (0.54138)	1,663

To predict city sizes $\hat{x}$, we substitute θ, and $\hat{q}_2$ into the q-log exponential CDF and solve for

$$\hat{x} = \hat{\theta} e_{\hat{q}_2}\left[\frac{-1}{\hat{\rho}} \log_{\hat{q}_1}^{T}\left[(1-F(.))^{1/\hat{q}_1}\right]\right]$$

The log of actual and predicted values of x can be plotted against the log rank to obtain the rank-size plot.

3.3 Graphical analysis

In this section, we graphically analyze the modelling of Romania's cities population. We perform the Kolmogorov- Smirnov test and display the rank-size plots of both data and predicted values. The Kolmogorov-Smirnov (KS) test considers the goodness-of-fit by analyzing the supremum of the difference between the theoretical and empirical CDF. Table 3 reports the KS test values of q-log exponential distribution based on the selected data. A p-value of the KS test close to 1 indicates extreme evidence for the data to have come from the distribution fitted.

Table 3. Kolmogorov-Smirnov test results

Year	KS	p-value
2007	0.02612	0.981
2008	0.02646	0.978
2009	0.02631	0.979
2010	0.02653	0.978
2011	0.02692	0.974
2012	0.02694	0.974
2013	0.02719	0.972
2014	0.02727	0.971
2015	0.02676	0.976
2016	0.02622	0.980
2017	0.02698	0.973

CONCLUSIONS

Romania's cities population can be very well modelled by means of q-log exponential distributions for each year. Since a large portion of the population of Romania (56.4%) is living in cities, the economic activities in this part of the country are important to the national economic growth. At present, in 2017, the capital, Bucharest, is the most developed city in the country, having more than 2 millions inhabitants, while the second largest city Iasi, has 368,866 inhabitants. This fact suggests a large gape in development between the capital and the rest of the cities.
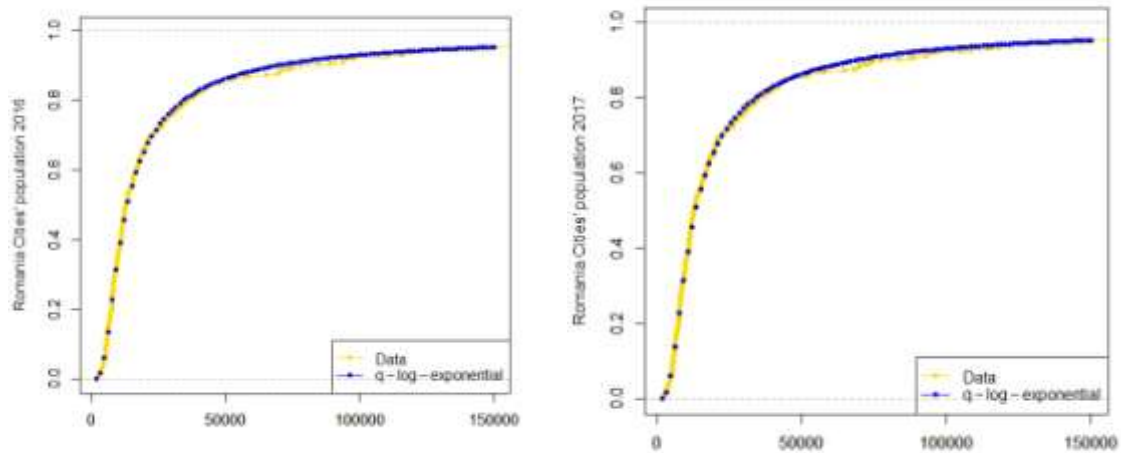

FIG. 2. Empirical and theoretical cumulative distribution of city size "Romania2016" and "Romania2017" data
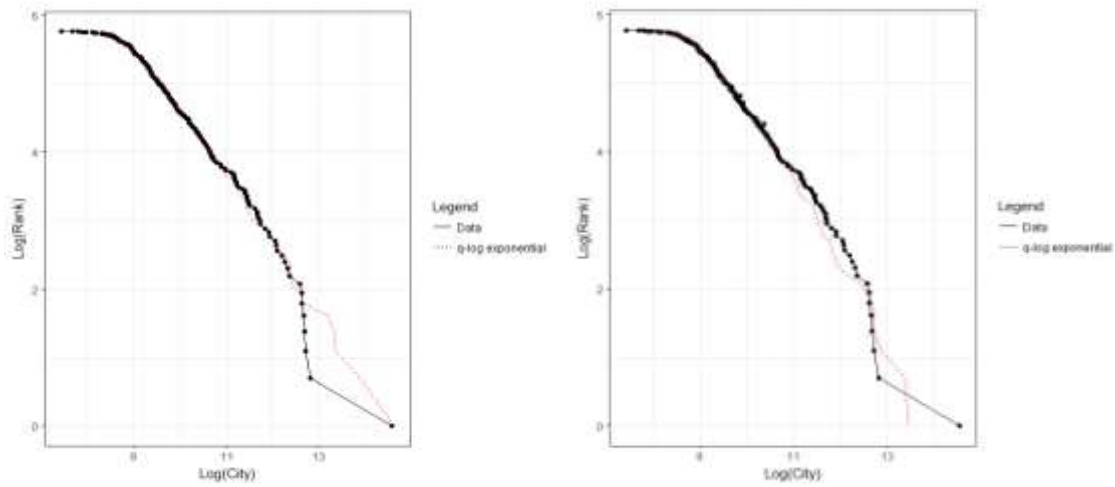
FIG. 3. Rank-size plots for 2016 and 2017

REFERENCES

[1] J. Luckstead, S. Devadoss, *Pareto tails and lognormal body of US cities size distribution*, *573-578*, Physica A: Statistical Mechanics and its Applications 465, 2017;

[2] M. Puente-Ajovın, A. Ramos, *On the parametric description of the French, German, Italian and Spanish city size distributions*, *489-509*, The Annals of Regional Science, 54(2), 2015;

[3] L.C. Malacarne, R.S. Mendes, E.K. Lenzi, *q-Exponential distribution in urban agglomeration*, Physical Review E 65, (1), 2001;

[4] P. Cirillo, *Are your data really Pareto distributed? 5947–5962,* Physica A: Statistical Mechanics and its Applications 392 (2013);

[5] I. Bancescu, *q-log-distributions. Log-concavity and Log-convexity, 1-16*, The European Physical Journal Plus, 133, 2018;

[6] Subrahmanyam M.B. *An Extension of the simplex method to constrained nonlinear optimization*, 311-319, Journal of Optimization Theory and Applications, 62(2), 1989.

A COMPARISON BETWEEN DETRENDING METHODS: HODRICK-PRESCOTT, BAXTER-KING, CHRISTIANO-FITZGERALD FILTERS. A SHORT SURVEY

Bogdan Corneliu BIOLAN

University of Bucharest, Romania (bbiolanc@gmail.com)

Abstract: This article examines the differences between the extracted cyclical components of some macro-economic time series using three detrending methods: HP (Hodrick-Prescott), BK (Baxter-King) and CF (Christiano-Fitzgerald). We use different approaches to compare the differences. A standard examination of the cyclical component is done. We also use a frequency domain approach and examine the sample spectra for each cycle. Furthermore, impulse responses and the correlation between the cyclical components extracted by each detrending method are studied. Our conclusion is that for quarterly data HP, BK and CF produce similar cycles. However, when we considered annual data the HP method gives us significant differences from the BK and CF methods.

Keywords: Hodrick-Prescott filter, BK (Baxter-King) filter, CF (Christiano-Fitzgerald) filter, DSP (Digital Signal Processing), macro-economy.

1. INTRODUCTION

Correctly estimating business cycles is important for macro-economic research. Several methods of extracting business cycles from some given time series were developed, but none of these lead to different results one to another. This is the reason why the authors from [22], investigate these three de-trending tools: HP (Hodrick-Prescott), BK (Baxter-King) and CF (Christiano-Fitzgerald). All these three are also considered to be filters, because their aim is to separate the trend component from the cyclical component. The analyses carried out in this article include a basic analysis of the cyclical component, a frequency domain examination of the spectral densities of each filter, also the correlation between the cyclical components found by the detrending methods and the impulse responses are also studied.

Also, the purpose of this research article is to determine, in future researches, what other types of signals, besides the time series presented here, would be suitable to adapt these filters (HP, BK or CF) to extract important observations from input data. By adapting these types of filters to other datasets we mean to modify not only the constants involved in their own definition, but also the mathematical definitions or equations that describe them.

From the conclusions obtained in [22], we are now sure that for one quarter of the data, the HP, BK and CF filters extract similar cycles and this is an very important result.

The second important result that could be used for other datasets concerns the behavior of the HP filter for annual data, which differs from the other two band pass filters (BK and CF).

2. MAIN RESULTS CONCERNING THE HODRICK-PRESOTT FILTER, THE BAXTER-KING FILTER, THE CHRISTIANO-FITZGERALD FILTER

2.1 The Hodrick-Prescott Filter

For this section we follow the line from [21].

The Hodrick-Prescott (HP) filter is a standard tool in macroeconomics for differentiating the long trend in a data series from short run fluctuations. It is also a smoothing method whose aim is to obtain a smooth component from the trend. Let's assume we have the following time series:

$y_t = \tau_t + c_t, t = 1,2, \dots, T$.

The HP filter smoothed the series $\hat{\tau}_T = (\hat{\tau}_{T1}, \hat{\tau}_{T2}, \dots, \hat{\tau}_{TT})$ as defined and described in economics by Hodrick and Prescott (1980, 1997) ([3] and [4]) results from minimizing, over all $\tau \in \mathbb{R}^T$,

$$\sum_{t=1}^{T} (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} (\tau_{t+1} - 2\tau_t + \tau_{t-1})^2, \tag{1}$$

where T denotes the sample size, λ is the nonnegative smoothing parameter so that for quarterly of the data is often chosen to be equal to 1600, and $y = (y_1, y_2, \dots, y_T)'$ is the data series to be smoothed.

In [5] a similar filtering technique was introduced [1]. Usually, $\hat{\tau}_{Tt}$ are referred to as "trend components", while $\hat{c}_{Tt} = y_t - \hat{\tau}_{Tt}$ it is named "cyclical component".

It was mentioned in [1] that there exists a minimizer, which is unique, to the minimization problem for equation (1), so that, for a known positive definite $(T \times T)$ matrix F_T, letting I_T denote the $(T \times T)$ identity matrix, $y = (\lambda F_T + I_T)\hat{\tau}_{Tt}$ and
$\hat{\tau}_{Tt} = (\lambda F_T + I_T)^{-1} y; \qquad (2)$.

So the trend component $\hat{\tau}_{Tt}$ and the cyclical component $\hat{c}_{Tt}$ are both weighted averages of y_t, so therefore we will re-write:

$$\hat{\tau}_{Tt} = \sum_{s=1}^{T} w_{Tts} y_s .$$

For notational convenience, the dependence of w_{Tts} and $\hat{\tau}_{Tt}$ on λ is suppressed. One of the purposes set in [1] is to find a new representation for w_{Tts} and see immediate consequences of this representation. This approach eliminates the inability to discover a simple expression for the elements of $(\lambda F_T + I_T)^{-1}$, which prevented other researchers from finding a simple expression for the weights that are implicit for the HP filter (for more, see [1]).

We notice first of all, that: $\hat{\tau}_{Tt}(y_1 + 1, y_2 + 1, \dots, y_T + 1) = \hat{\tau}_{Tt}(y_1, y_2, \dots, y_T) + 1$, so this means that $\sum_{s=1}^{T} w_{Tts} = 1$ for $t \in \{1,2, \dots, T\}$. Also, we have that: $\hat{\tau}_{Tt}(1,2, \dots, T) = t$, and therefore we have that: $\sum_{s=1}^{T} w_{Tts} s = t$, for $t \in \{1,2, \dots, T\}$.

Authors in [1] obtained that this way, a quadratic trend is not absorbed in $\hat{\tau}_{Tt}$.

They also mentioned that previous literature on the HP filter is only based on the observation that the first order condition for $\hat{\tau}_{Tt}, t \in \{3, \dots T - 2\}$ is:

$$-2(y_t - \hat{\tau}_{Tt}) - 4\lambda(\hat{\tau}_{T,t+1} - 2\hat{\tau}_{Tt} + \hat{\tau}_{T,t-1}) + 2\lambda(\hat{\tau}_{Tt} - 2\hat{\tau}_{T,t-1} + \hat{\tau}_{T,t-2})$$
$$+ 2\lambda(\hat{\tau}_{T,t+2} - 2\hat{\tau}_{T,t+1} + \hat{\tau}_{Tt}) = 0 \tag{3}$$

Let $\bar{B}$ denote the forward operator and B the backward operator, then according to [1], this simplifies to the following relation:

$$y_t = (\lambda \bar{B}^2 - 4\lambda \bar{B} + (1+6\lambda) - 4\lambda B + \lambda B^2)\hat{t}_{Tt}, \tag{4}$$

which can also be re-written as:

$$y_t = (\lambda|1-B|^4 + 1)\hat{t}_{Tt} \tag{5}$$

Papers that, according to [1], that analyze the HP filter based on the first order condition are for example [6], [7], [8], [9] and [10].

A high value of λ will give a more linear trend and will allow for increased variation in the cyclical component, [22]. For $\lambda = 0$, the trend component is, obviously, equivalent to the actual time series, y_t.

The determinant key for the minimization problem (1) is the value of λ.

2.2 The Baxter-King filter

For following we consider the results from [22].

BK filter is decomposing a time series, let's say y_t, into three different components: trend, cycle and the irregular component.

$$y_t = \tau_t + c_t + \varepsilon_t \tag{6}$$

where τ_t is the trend component, c_t is the cyclical component, ε_t is the irregular component, [15].

As a result, a new time series $y_t{}^*$ is obtained when we apply a finite symmetric moving average.

So, we define the following symmetric moving average: $y_t{}^* = \sum_{k=-K}^{K} \beta_k y_{t-k}$, $\qquad$ (7) where β_k are fixed constants or let's say weights and K the maximum lag length. In order to extract the cyclical components from the above time series, BK uses weights that will add up to zero, meaning: $\sum_{k=-K}^{K} \beta_k = 0$. That will be the trend elimination property, [15]. These weights that add up to zero and the moving average have some good elimination properties that generate a so called stationary time series.

This is an very important fact, because economic time series have the tendency to be non-stationary, [15].

Authors in [15] derive this filter through the frequency-domain perspective. The point from where to start with the BK filter is, according to [22], the Cramer representation theorem, which states the following:

$$y_t = \int_{-\pi}^{\pi} \xi(\omega)d\omega \tag{8}$$

under some suitable conditions. In this representation, the time series is now written as the integral of the random periodic components, $\xi(\omega)$, and where ω is expressed in radians. According to [22], if we apply Cramer Theorem to equation (7), we have the following:

$$y_t{}^* = \int_{-\pi}^{\pi} \beta(\omega)\xi(\omega)d\omega \tag{9}$$

where $\beta(\omega) = \sum_{k=-K}^{K} \beta_k \exp(i\omega k)$ it is the frequency response function for this filter.

Another interpretation, according to [22] would be: how much y_t^* responds to y_t at a given frequency ω with respect to the weight $\beta(\omega)$, also to the random and periodic component $\xi(\omega)$. An important observation for the BK filter would be that $\beta(\omega)$ has the value 0 at frequency 0, meaning: $\beta(0) = 0$.

The cyclical component is defined or extracted as follows:

$$c_t = \sum_{j=-K}^{K} b_j \, y_{t-j} \qquad (10)$$

where the weights b_j can found by applying the inverse Fourier transform to the frequency response function, [12].

$$b_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} \beta(\omega) \exp(i\omega j) d\omega \qquad (11)$$

Considering the definition from [22], ω has to be between $\frac{\pi}{16}$ and $\frac{\pi}{3}$. The construction of BK filter is made from two low pass filters and is having two frequency bands, $\frac{\pi}{16}$ and $\frac{\pi}{3}$. Because of this, we define two frequency response functions: $\beta_1(\omega) = 1$, for $|\omega| \le \omega_1 = \frac{\pi}{16}$ (zero otherwise), $\beta_2(\omega) = 1$, for $|\omega| \le \omega_2 = \frac{\pi}{3}$. To obtain the weights b_j, we get $\beta_1(\omega)$ from $\beta_2(\omega)$ and we will have the desired frequency response function for $\frac{\pi}{16} \le \omega \le \frac{\pi}{3}$ or $-\frac{\pi}{3} \le \omega \le -\frac{\pi}{16}$ and zero otherwise. The weights b_j can then be obtained from equation (11).

2.3 The Christiano and Fitzgerald filter

The Christiano and Fitzgerald filter (CF), like the BK filter, is the approximation of an ideal band pass filter. Assuming a symmetric moving average is not the case here [16]. The cyclical component, according to [22], is given by:

$$c_t = b_0 y_t + b_1 y_{t+1} + \cdots + b_{T-1-t} y_{T-1} +$$
$$\tilde{b}_{T-t} y_t + b_1 y_{t-1} + \cdots + \tilde{b}_{T-1} y_1 + b_{t-2} y_2 \qquad (12)$$

For t=3,4,…,T-2, where:

$$b_j = \frac{\sin(jc) - \sin(ja)}{\pi j}, j \ge 1$$

$$b_0 = \frac{c - a}{\pi}, a = \frac{2\pi}{p_h}, c = \frac{2\pi}{p_h}$$

$$\tilde{b}_k = -\frac{1}{2} b_0 - \sum_{j=1}^{k-1} b_j$$

and p_t and p_h are defined like in [16]. So, cycles that are longer than p_t but shorter than p_h are defined to be the actual cyclical component c_t. The CF filter is not symmetric, [17].

This is in contrast with BK filter, which is considered to be symmetric. CF is consistent, when compared to BK filter, because it converges to an ideal band pass filter when the sample size T is increased, [22], [17].

3. THE IMPULSE RESPONSE FUNCTION

The Impulse Response Function (IRF) is in general obtained from Variance Autoregressive model (VAR). Consider the following VAR (1) systems with these two equations:

$$x_t = a_1 + b^1_{1,1}x_{t-1} + b^1_{1,1}x_{t-1} + b^1_{1,2}z_{t-1} + \varepsilon_{t,x}$$
$$z_t = a_2 + b^1_{2,1}x_{t-1} + b^1_{1,1}x_{t-1} + b^1_{2,2}z_{t-1} + \varepsilon_{t,z}$$

where x_t and z_t appear as is [22]. This shows that if there is a shock in x_t, there will also be an effect in z_t.

Let's suppose that at a given time t there is a shock to z by one standard deviation, σ_z , and x is not shocked. Furthermore, suppose that for period: $t+1, t+2, \dots t+n$ we don't have any shock in either x or z [20]. This response to the shock will then be:

➔ Time t (when the shock occurs):

The effect on z is σ_z and on x we have no effect.

➔ Time $t+1$:

The effect on z is $\beta^1_{2,2}\sigma_z$ and the effect on x is $\beta^1_{1,2}\sigma_z$

➔ Time $t+2$:

The effect on z is $(\beta^1_{2,2})^2\sigma_z + \beta^1_{2,1}\beta^1_{1,2}\sigma_z$ and the effect on z is $\beta^1_{1,2}\beta^1_{2,2}\sigma_z + \beta^1_{1,1}\beta^1_{1,2}\sigma_z$

In future periods, $t+n$, the effect from the shock will be different from the values of $\beta^1_{i,j}$.

The coefficients $\beta^k_{i,j}$ are obtained from the following VAR(k) system, with two equations, [22]:

$$\tau_t = \alpha_1 + \beta^1_{1,1}\tau_{t-1} + \beta^1_{1,2}c_{t-1} + \cdots + \beta^k_{1,1}\tau_{t-k} + \beta^k_{1,2}c_{t-k} + \varepsilon_{1,t},$$
$$c_t = \alpha_2 + \beta^1_{2,1}\tau_{t-1} + \beta^1_{2,2}c_{t-1} + \cdots + \beta^k_{2,1}\tau_{t-k} + \beta^k_{2,2}c_{t-k} + \varepsilon_{2,t},$$

where: α_i is a constant $(i = 1,2)$, τ_t is obviously the original time series and c_t is the cyclical component for the time series. Above, k represents the number of lags and $\varepsilon_{i,t}$ is a stochastic error term $(i = 1,2)$.

In order to decide about the number of lags in this the model, we proceeded as in [22].

4. CONCLUSIONS

This survey article reveals the new mathematically rigorous results and properties of the HP filter, BK filter, CF filter obtained by the authors in [22] and also establishes new lines of research in this field, having set for the future clear objectives concerning the kind of results that are to be obtained.

Future work will include possible applications of these filters in DSP, by calibrating the constants in each filter after replacing the GDP, consumption, investment and inflation test datasets with digitized samples from complex valued signals, reconstruct the impulse response functions for each filter and after plot the spectral density for each filter in the detrended dataset.

REFERENCES

[1] R.M. de Jong and N. Sakarya, The Econometrics of the Hodrick-Prescott filter, *Review of Economics and Statistics*, vol. 98, no. 2, pp. 310-317, 2015;

[2] M.O. Ravn and H. Uhlig, On adjusting the Hodrick-Prescott filter for the frequency of observations, *Review of Economics and Statistics*, vol. 84, no. 2, pp. 371-376, 2002;

[3] R. Hodrick and E.C. Prescott, *Postwar US business cycles: an empirical investigation*, Carnegie Mellon University discussion paper, no. 451, 1980;

[4] R. Hodrick and E.C. Prescott, Postwar US business cycles: an empirical investigation, *Journal of Money, Credit, and Banking*, vol. 29, no. 1, pp. 1-16, 1997;

[5] E.T. Whittaker, On a New Method of Graduations, *Proceedings of the Edinburgh Mathematical Society*, vol. 41, pp. 63-75, 1923;

[6] R.G. King and S.T. Rebelo, Low frequency filtering and real business cycles, *Journal of Economic Dynamics and Control*, vol. 17, no. 1, pp. 207-231, 1993;

[7] T. Cogley and J.M. Nason, Effects of the Hodrick-Prescott filter on trend and difference stationary time series: Implications for business cycle research, *Journal of Economic Dynamics and Control* vol. 19, no. 1, pp. 253-278, 1995;

[8] P.C.B. Phillips and S. Jin, *Limit Behavior of the Whittaker (HP) Filter*, Working Paper, 2002;

[9] P.C.B. Phillips, Two New Zealand pioneer econometricians, *New Zeeland Economic Papers*, vol. 44, pp. 1-26, 2010;

[10] T. McElroy, Exact formulas for the Hodrick Prescott filter, *The Econometrics Journal*, vol. 11, no. 1, pp. 209-217, 2008;

[11] K.S. Miller, On the inverse of the sum of matrices, *Mathematics Magazine*, vol. 54, no. 2, pp. 67-72, 1981;

[12] Guay, A. & St-Amant, P., *Do the Hodrick-Prescott and Baxter-King Filters Provide a Good Approximation of Business Cycles?*, Cahiers de recherche CREFE / CREFE Working Papers 53, CREFE, Université du Québec à Montréal, 1997;

[13] Christiano, L.J. & Fitzgerald, T.J., *The Band Pass Filter*, International Economic Review, vol. 44, pp. 435-465, 2003;

[14] Burns, A. M., & W. C. Mitchell, *Measuring Business Cycles*. New York: National Bureau of Economic Research, 1946;

[15] Baxter, M., & King, R. G., *Measuring Business Cycles: Approximate Band-Pass Filters For Economic Time Series*. The Review of Economics and Statistics, MIT Press, vol. 81(4), pp. 575-593, 1999;

[16] Christiano, L.J. & Fitzgerald, T.J., *The Band Pass Filter*, International Economic Review, vol. 44, pp. 435-465, 2003;

[17] Haug A.A., & Dewald W.G., *Longer-term Effects of Monetary Growth on Real and Nominal Variables, Major Industrial Countries, 1880-2001*, Working Paper Series 382, European Central Bank, 2004;

[18] de Haan, J., Inklaar, R., & Jong-A-Pin, R., *Will Business Cycles In The Euro Area Converge? A Critical Survey of Empirical Research*. Journal of Economic Surveys vol. 22, pp. 234-273, 2008;

[19] Van Aarle, B., Kappler, M., Sachs, A., Seymen, A., & Weyerstrass, K., *Study on Economic Integration and Business Cycle Synchronization*. Mannheim & Vienna: Centre for European Economic Research (ZEW) and Institute for Advance Studies (IHS), 2008;

[20] Hill, R. Carter, Griffiths, William E. & Lim, Guay C. (2008). *Principles of econometrics*. 3. ed. Hoboken, N.J.: Wiley, 2008;

[21] Bogdan Corneliu Biolan, *On Hodrick-Prescott Filter. A short survey and applications, 113-118*, Review of the Air Force Academy, Vol.15, No. 2(34) / 2017, 2017;

[22] Gabrielle Larsson and Tamas Vasi, Department of statistics, Uppsala University, *Comparison of detrending methods*, 2012.