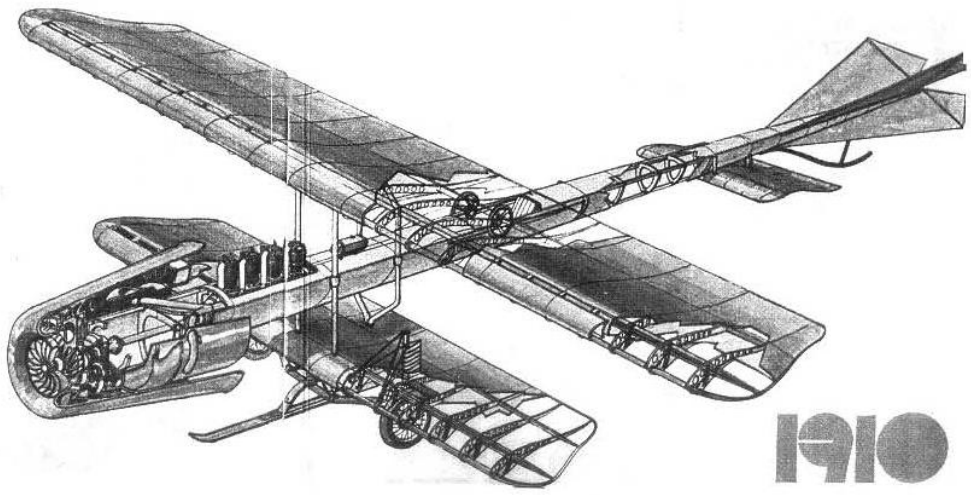
Review
of the Air Force Academy

The Scientific Informative Review, Vol. XV, No.2 (34)/2017
DOI: 10.19062/1842-9238.2017.15.2



Selection of papers presented within the Romanian Society of Probability and Statistics Conference, SPSR 2017, 20th edition.
Brasov, 28-29 of April, 2017

© September, 2017

Visa 0574-09/2017

I.S.S.N. 1842-9238

"Henri Coandă" Air Force Academy Publishing House, 160, Mihai Viteazul St., Braşov, ROMÂNIA
Phone: +40 268 423421, e-mail: editura@afahc.ro

C O N T E N T S

ON A PARTICULAR LIFETIME DISTRIBUTION

Ion VĂDUVA

University of Bucharest, Faculty of Mathematics and Computer Science, retired
(ivaduva@gmail.ro)

DOI: 10.19062/1842-9238.2017.15.2.1

Abstract: *The paper introduces a probability distribution as a mixture between a Gamma(0, $\lambda\eta$, υ) distribution and an exponential Exp(μ) distribution of η. The first and second order moments are calculated, together with the variance. Algorithms for simulation of the introduced distribution are presented. These include the inverse method and the rejection method. The last section discusses an application to reliability of a system with n components, with stochastic independent lifetimes, namely the distributions of maximum W and minimum V of lifetimes, when components have the introduced distribution. Simulation of W and V is also presented. The last part of the paper presents distributions of V and W when the number of components is (truncated) random with distributions: Poisson(λ), Geometric(p), or Binomial(n, p), n ≥ 1. Simulation of these distributions is also underlined.*

Keywords: *Gamma and Exponential distributions, Mixture distribution, Random variate simulation, Reliability.*

1. INTRODUCTION

In reliability theory, an important notion is the *lifetime*, i. e. a random variable *rv* which represents the running of a system until it *fails*. Let us denote L, a lifetime random variable which has the *cummulative distribution function (cdf)*, $F(x)=P(L<x)$ and the corresponding *probability density function (pdf)*, $f(x)=F'(x)$, assuming that the cdf is a continous (i.e. it is derivable). The *danger of system to fail* is given by the failure rate $r(x)$ defined as [7]

$$r(x)=\frac{f(x)}{\overline{F}(x)}\ ,\ \overline{F}(x)-1-F(x)+e^{-\int_0^x r(u)du} \tag{1}$$

where $F(x)$ is the *survival probability or reliability function.* The failure rate could be *increasing* i.e. the distribution of L is *IFR* (*increasing failure rate*) or is *DFR* (*decreasing failure rate*). As many real systems become *fatigue in time*, then many of reliability systems are IFR. (While, for instance, from reliability point of view, the lifetime of a computer program is DFR!, (see [7]). Examples of IFR (as well as DFR) cdf's are [7] the exponential distribution *Exp*(λ), of parameter $\lambda;\ \lambda>0;$ or a *Weibull*($0;\lambda;\ \upsilon$); $\lambda>0;\ \upsilon>1$ distribution, while when $0<\upsilon<1$, this distribution is DFR. Note that for any lifetime, the *pdf* $f(x)$, as well as *cdf* $F(x)$ is zero for $x\le 0$*:* Therefore, in the formulae like this, in the following we will specify the pdf $f(x)$ and the cdf $F(x)$, only for $x>0$.

Some time, *complex reliability systems* have a behavior which assumes at the beginning of their life an increasing failure rate and later on, a decreasing failure rate. In this situation are (see [7]) the *lognormal distribution LN*($\mu;\sigma$), $\mu>0;\ \sigma>0$ and the *Gamma*($0;\lambda;\upsilon$), $\lambda>0;\ \upsilon>1$ *distribution*, which has the pdf

$$f(x)=\frac{\lambda^{\nu}}{\Gamma(\nu)}x^{\nu-1}\ e^{-\lambda x}\ ,\ \ x>0\ ,\ \ where\ \ \Gamma(\nu)=\int_0^{\infty}x^{\nu-1}e^{-x}dx \tag{2}$$

Now, we assume the following situation inspired by [3,6]: the system was produced in a country (or in a *climate*) and it was stated that the lifetime L has a *Gamma*$(0;\lambda;\upsilon)$, $\lambda>0$; $\upsilon>1$ distribution. If the system is running (or used) in other conditions (i.e in another *climate*), then it is *assumed* that the initial life time distribution is altered, becoming L^*; such that this distribution becomes *Gamma*$(0;\lambda\eta;\upsilon)$, where η is a random variable *Exp*(μ): The problem is to determine the pdf of L^* which is a *mixture* or a *composition* from *Gamma*$(\lambda\eta;\upsilon)$ with respect to *Exp*(μ) distribution of η: (Note. In [2,3,6], L^* is a mixture between *Exp*$(\lambda\eta)$ and *Gamma*$(0;\lambda;\upsilon)$ of η which is a *Loomax distribution*).

2. THE PROBABILITY DISTRIBUTION OF L^*

Let us calculate the pdf of L^* as the **mixture** of a *Gamma*$(0;\lambda\eta;\upsilon)$ distribution with the *Exp*(μ) distribution of η: Using the pdf of η, then the pdf of the mixture is

$$f^*(x)=\int_0^{\infty}\frac{(\lambda\eta)^{\nu}}{\Gamma(\nu)}x^{\nu-1}e^{-\lambda\eta x}\mu e^{-\mu\eta}d\eta\ .$$

After some calculations we obtain

$$f^*(x)=\frac{\mu\lambda^{\nu}}{\Gamma(\nu)}x^{\nu-1}\int_0^{\infty}\eta^{\nu}e^{-\eta(\lambda x+\mu)}\mu e^{-\mu\eta}d\eta$$

which finally gives

$$f^*(x)=\frac{\mu\nu\lambda^{\nu}}{(\lambda x+\mu)^{\nu+1}}\ ,\ x>0 \tag{3}$$

or if we denote $\theta=\frac{\lambda}{\nu}$ the final form is

$$f^*(x)=\frac{\nu\theta^{\nu}x^{\nu-1}}{(\theta x+1)^{\nu+1}}\ ,\ x>0\ ,\ \theta>0\ ,\ \nu>0 \tag{3'}$$

The cdf of L^* is calculated as follows

$$F^*(x)=\int_0^{x}f^*(u)du=\int_0^{x}\frac{\nu\theta^{\nu}u^{\nu-1}}{(1+\theta u)^{\nu+1}}du$$

which is an integral of a bimome type, i.e.

$$I=\int_0^{x}u^m\left(a+bu^n\right)^p du\ ,$$

where a=1; b =θ, m= υ-1; n=1; $p=-\nu-1=\frac{\nu+1}{s}$ p=-υ-1, s=-1: According to Tchebycheff's conditions [1], this integral is calculated by using the transform

$$z^s=z^{-1}=1+\theta x\ ,\ hence\ \ u=\frac{1}{\theta}\cdot\frac{1-z}{z}\ ,\ du=-\frac{dz}{\theta z^2}$$

and because $u=0\rightarrow z=1$ *and* $u=x\rightarrow z=\frac{1}{1+\theta x}$ it results that

$$F^*(x) = -\int_1^{\frac{1}{1+\theta x}} \nu\theta^\nu \frac{(1-z)^{\nu-1}}{z^{\nu-1}} \frac{dz}{z^2} z^{\nu+1} = \int_{\frac{1}{1+\theta x}}^1 \nu(1-z)^{\nu-1} dz .$$

By simple calculations it results that

$$F^*(x) = \left(\frac{\theta x}{1+\theta x}\right)^\nu . \tag{4}$$

The moment of order *k* is calculated as follows

$$m_k = E\left(L^{*k}\right) = \nu\theta^\nu \int_0^\infty \frac{x^{\nu+k-1}}{(\theta x+1)^{\nu+1}} dx .$$

This is again a binome integral where $p = -\nu - 1 = \frac{\nu+1}{s}$, s=is an integer, $n = 1$, $m = \nu + k - 1$. According to mentioned Tchebytcheff's conditions [1], the integral can be also calculated by using transform $z^s = 1 + \theta x$, *i.e.* $z^{-1} = 1 + \theta x$.

Therefore a k-st iteration (with respect to *k*) of the integral

$$I_k = \int_0^\infty \frac{x^{\nu+k-1}}{(\theta x+1)^{\nu+1}} dx$$

using the specified transform, gives

$$I_k = \frac{1}{\theta^{\nu+k}} \int_0^1 (1-z)^{\nu+k-1} z^k dz = -\frac{1}{\theta^{\nu+k}} \frac{1}{\nu+k} (1-z)^{\nu+k} z^k \Big|_0^1 + \frac{k}{\theta^{\nu+k}(\nu+k)} \int_0^1 (1-z)^{\nu+k} z^k dz \text{ i.e.}$$

$$I_k = \frac{k}{(\nu+k)\theta^{\nu+k}} \int_0^\infty (1-z)^{\nu+k} z^{k-1} dz .$$

For *k* = 1 one obtain

$$I_1 = \frac{1}{(\nu+1)\theta^{\nu+1}} \int_0^\infty (1-z)^{\nu+1} dz = \frac{1}{\theta^{\nu+1}(\nu+1)(\nu+2)} .$$

Therefore

$$m_1 = \frac{\nu}{\theta(\nu+1)(\nu+2)} .$$

(5)

For *k* = 2 one obtain

$$I_2 = \frac{2}{(\nu+2)\theta^{\nu+2}} \int_0^\infty (1-z)^{\nu+2} z dz = \frac{2}{\nu+2} \frac{1}{\theta^{\nu+2}} \left[-\frac{(1-z)^{\nu+3} z}{\nu+3} \Big|_0^1 + \frac{2}{(\nu+3)(\nu+2)\theta^{\nu+2}} \int_0^\infty (1-z)^{\nu+3} dz \right] =$$

$$= \frac{2}{(\nu+2)(\nu+3)(\nu+4)\theta^{\nu+2}}$$

Hence

$$m_2 = \frac{2\nu}{\theta^2(\nu+2)(\nu+3)(\nu+4)} . \tag{6}$$

Now the variance $\sigma^2 = Var\left(L^*\right)$ is calculated as

$$\sigma^2 = m_2 - m_1^2 = \frac{\nu}{(\nu+2)\theta^2} \left[\frac{2}{(\nu+3)(\nu+4)} - \frac{\nu}{(\nu+1)^2(\nu+2)} \right]$$

which is finally

$$\sigma^2 = \frac{\nu}{(\nu+2)\theta^2}\left[\nu^3 + \nu^{2-4\nu+2}\right] > 0\,. \tag{7}$$

In simulating reliability models which involve this distribution, it is interesting to built up algorithms (see [5,8]) for simulating it, i.e. algorithms for producing sampling values of L^*.

3. SIMULATION OF THE DISTRIBUTION

Such an algorithm is designed to produce a sampling value of L^* and when repeating it *n* times, to obtain a sample L_1^* , L_2^* , … , L_n^* .

In the following, we present simulation methods for L^*.

3.1 The inverse method

The Chintchin's lemma says (see [5,8]):

Lemma. *If a random variable X has the cdf F*(*x*) *and U is a rv uniformly distributed over* (0*;* 1), *then the cdf of* $F^{-1}(U)$ *is F*(*x*)*: (Note that equivalent relation* $U = F^{-1}(X)$ *is valid).*

Note that each computer (i.e. any language) has an algorithm (*generator*) to produce (when is *called*), an uniform random number *U*, and, when calling it next time, it produces *another* uniform random number *U*.

(In other words, successive calls of the generator, produce a sequence of *U's* independent and uniformly distributed). Note also (see [5,8]) that if in an algorithm appears operation 1-*U*, we can use instead *U*, because when *U* is a random number, 1-*U* is also a random number.

From the *Lemma* it results that in other words, to simulate a sampling value of *X;* the following algorithm is derived:

Alhorithm 1.

begin

-*Simulate a random number U uniformly distributed over* (0, 1);

-*Take* $X = F^{-1}(U)$ *;*

end

Therefore, the algorithm applied to L^* is:

The inverse algorithm for simulatin L^*

- *Simulate an uniform random number U*;

- *Calculate* $L^* = F^{-1}(U) = \frac{1}{\theta}\frac{U^{\frac{1}{\nu}}}{1-U^{\frac{1}{\nu}}}$

Note that if *U* is close to 1 (which might happen!), the algorithm fails, therefore we must *reject* a value of *U* for which $1-U^{\frac{1}{\nu}}$ is close *to the zero of the computer involved* and use the *next* random number *U:*

To simplify the writing, in the following when we use the *U,* it is assumed that it is a random number uniformly distributed over (0, 1)*:*

For simulating L^* we will use also the *acceptance-rejection method* which will be called in the following **the rejection method** (see [5,6,8]).

3.2 The rejection method

This method assumes [4,5,8] that we can simulate some *simpler* random variables S_1, S_2*; :::* until they satisfy *some condition*; the required random variable X is calculated in terms of random variables S_i , $i \in N$ which satisfy the condition.

There are several theorems which lead to rejection methods. We will use the following two theorems.

Theorem 1 [4,5,8,9]. *Assume that random variable X to be simulated, has a pdf* $f(x) \neq 0$, $x \in DD$ *and there is another random variable Y with pdf* $f(x) \neq 0$, $x \in D \subset R$ *which can be simulated, such as* $(\forall)x \in D$, $\frac{f(x)}{h(x)} \leq \alpha$, $\alpha = const$.

If U is uniform (0*;* 1) *stochasticaly independent from Y , and they satisfy the condition*

$$U \leq \frac{f(x)}{\alpha h(x)} \ , \tag{8}$$

then Y has the pdf f(*x*)*:*

Therefore the algorithm for simulating X is:

Algorithm 2.

repeat

- *generate U uniform* (0*;* 1);
- *generate Y having pdf h*(*x*)

until $U \leq \frac{f(x)}{\alpha h(x)}$

Take X = Y:

The performance of the algorithm is given by the *accepting probability*

$$p_a = \text{Prob}\left(U \leq \frac{f(Y)}{\alpha\, h(Y)} \right) = \frac{1}{\alpha} \tag{8'}$$

therefore it is necessary that $\alpha > 1$. (The value of p_a results from the proof of the theorem). The algorithm is fast if $p_a < 1$, is close to 1.

A rejection algorithm based on this theorem for simulating L^* is obtained if we take as enveloping pdf

$$h(x) = \frac{\theta}{(1+\theta x)^2}$$

In this case we have the ratio

$$r(x) = \frac{f^*(x)}{h(x)} = \frac{\nu\, \theta^{\nu-1} x^{\nu-1} (1+\theta x)^2}{1+\theta x^{\nu+1}}$$

Since $r(x)$ is an increasing function and positive, it results that

$\lim_{x\to} r(x) = \nu\, \theta^{\nu-1} = \alpha$.

Hence

$$p_a = \frac{1}{\nu\, \theta^{\nu-1}} \tag{8''}$$

which makes sense if $\nu\,\theta^{\nu-1} > 1$. The cdf $H(x) = \int_0^x f^*(u)du$ has the inverse

$$Y = \frac{1}{\theta}\left(1 - \frac{1}{\sqrt{U}}\right)$$

and therefore the rejection algorithm is obvious.

The above theorem is also called the *enveloping theorem* because the main assumption of the theorem says that $f(x) \le \alpha\, h(x)$, i.e. there is a α such as the graph of *f*(*x*) is *enveloped* by the graph of $\alpha\, h(x)$, $x \in D$.

Theorem 2 [4,5,8]. *Assume that the pdf f*(*x*) *of the r.v. X to be simulated is in the form*

$$f(x) = c\, r(x)\big(1 - Q(x)\big) \tag{9}$$

where c = const, r(*x*) *is the pdf of a random variable Y and Q*(*x*) *is the cdf of a r.v. Z: Then, the rv Y, satisfying condition* $Z \ge Y$ *with Z and Y independent random variables, has the pdf f*(*x*)*:*

Hence the theorem says that the sampling value *X* is the *accepted Y.*

The resulting rejection algorithm is

Algorithm 3.
repeat
- *simulate Z;*
- *simulate Y independent of Z*;
until $Z \ge Y$;
Take X = Y.

The performance of the algorithm is given by

$$p_a = \text{Prob}\big(Z \ge Y\big) = \frac{1}{c}$$

hence it is necessary that *c*>1. (The value p_a results from the proof of the theorem).

There is another form (a kind of **dual** of the theorem 2), let us call it
Theorem 2' in which the pdf is in the form

$$f(x) = c\, r(x) Q(x) \tag{9′}$$

and the condition becomes $Z \le Y$.

Therefore the algorithm deriving from **Theorem 2'** is

Algorithm 3'
repeat
- *simulate Z*;
- *simulate Y independent from Z*;
until $Z \le Y$;
Take X = Y.

We apply this theorem in the form (9') , i.e.

$$f^*(x) = \frac{\nu\,\theta^{\nu} x^{\nu-1}}{1 + \theta\, x^{\nu+1}} = c\, r(x) Q(x)$$

where

$$c = \nu\ ,\ Q(x) = \left(\frac{\theta\, x}{1 + \theta\, x}\right)^{\nu-1} \qquad r(x) = \frac{\theta}{(1 + \theta\, x)^2} \tag{9”}$$

and $p_a = \frac{1}{\nu}$ (i.e $\nu > 1$is required). The algorithm is obvious and random variables *Z* and *Y* are simulated by the inverse method according to formulae

$$Z = \frac{1}{\theta}\frac{U^{\frac{1}{q}}}{1-U^{\frac{1}{q}}}\ ,\ Y = \frac{1}{\theta}\frac{1-U}{U}\ .$$

The accepting probability is $p_a = \frac{1}{\nu}$ which works if $\nu > 1$, but not very large.

Finally, we note that for simulating L^* the **algoritnms 1 and 2** are prefered. (They are faster!).

4. AN APPLICATION TO RELIABILITY

Assume that a system consists in *n* components having the lifetimes L_1^* , L_2^* , … , L_n^* *n* independent and identicaly distributed. In reliability is interesting to consider random variables

$$V = \min_{1\le i\le n} L_i^*\ ,\ W = \max_{1\le i\le n} L_i^* \tag{10}$$

(The lifetime *V* is applied when *all components fail* and the lifetime *W* is applied when *all components run*). It is obvious that cdf's of these rv's are respectively

$$F_v^*(x) = 1-\left(1-F^*(x)\right)^n\ ,\ F_w^*(x) = \left(F^*(x)\right)^n \tag{11}$$

and the corresponding pdf's are [10]

$$f_v^*(x) = nf^*(x)\left(1-F^*(x)\right)^{n-1}\ ,\ f_w^*(x) = f^*(x)\left(F^*(x)\right)^{n-1} \tag{11'}$$

i.e for our rv L^* we have

$$f_v^*(x) = \frac{n\nu\theta^\nu x^{\nu-1}}{(1+\theta x)^{\nu+1}}\left(1-\left(\frac{\theta x}{(1+\theta x)}\right)^\nu\right)^{n-1}\ ,\ f_w^*(x) = \frac{n\nu\theta^\nu x^{\nu-1}}{(1+\theta x)^{\nu+1}}\left(\frac{\theta x}{(1+\theta x)}\right)^{\nu(n-1)} \tag{11"}$$

Simulation of *V* and *W* can be done using directly formula (10) (i.e. calculating *min* and *max* of the simulated sample L_1^* , L_2^* , … , L_n^*).

Taking into consideration formulae (11'),(11") and theorems 1 and 2, the following concluding theorem is valid

Theorem 3. *The simulation of V and W can be done by using theorem 1 with enveloping density $f^*(x)$ or using theorems 2 and 2', noticing that pdf 's $f_v^*(x) =$, $f_w^*(x)$ are in the forms (9) or (9') taking into account froms (11").*

The proof is obvious, existing $0 < \alpha < \infty$ (in theorem 1), $r(x)$ is $f^*(x)$ (from (11') or (11")) and $Q(x)$ in theorems 2 or 2' are either $\left(\frac{\theta x}{(1+\theta x)}\right)^{\nu(n-1}$, *or* $\left(1-\left(\frac{\theta x}{1+\theta x}\right)^\nu\right)^{n-1}$.

4.1. The case when n random

In some practical situations, the number of components of the system is a random variable, say N^* , $N^* > 0$ [10]. Possible discrete distributions [10] are *Poisson*(λ)*; λ>0,* or *Geometric*(*p*)*; 0<p<*1*;* and these distributions are *truncated* on $[1, \infty)$.

In this case the distribution of L^* is called [10] *target distribution.* (In [10] there are used target distributions Weibull and Loomax). Here we will consider also the *new* case when N^* is *Binomial*(*n; p*); $n \in N$, <*p*<1.

Distributions of *V* and *W* when N^* is *Geometric*(*p*). The frequency of *Geometric*(*p*) distribution, truncated on $[1, \infty)$is

$$P(N^* = k) = p\,q^{k-1}\ ,\ k = 1, 2, \ldots \tag{12}$$

The cdf of *V* is the *mixture*

$$\Phi_v(x) = \sum_{k=1}^{\infty} p\,q^{k-1}\left(1-\left(1-\left(\frac{\theta x}{1+\theta x}\right)^{\nu}\right)^{k}\right) \tag{13}$$

which gives

$$\Phi_v(x) = \frac{p}{1-q} - p\sum_{k=1}^{\infty} q^{k-1}\left(1-\left(\frac{\theta x}{1+\theta x}\right)^{\nu}\right)^{k} = 1 - p\left(1-\left(\frac{\theta x}{1+\theta x}\right)^{\nu}\right)\sum_{k=1}^{\infty} q^{k}\left(1-\left(\frac{\theta x}{1+\theta x}\right)^{\nu}\right)^{k}$$

which finally gives

$$\Phi_v(x) = 1 - p\left(1-\left(\frac{\theta x}{1+\theta x}\right)^{\nu}\right)\frac{1}{1-q\left(1-\left(\frac{\theta x}{1+\theta x}\right)^{\nu}\right)} \tag{14}$$

The rv *V* from (10) can be simulated by the inverse method (based on the *inverse* $\Phi_v^{-1}(U)$ of (14)).

Using the truncated distribution (12), the cdf of *W* is

$$\Phi_w(x) = \sum_{k=1}^{\infty} pq^{k-1}\left(\left(\frac{\theta x}{1+\theta x}\right)^{\nu}\right)^{k} = p\left(\frac{\theta x}{1+\theta x}\right)^{\nu}\sum_{k=1}^{\infty} q^{k-1}\left(\left(\frac{\theta x}{1+\theta x}\right)^{\nu}\right)^{k-1}$$

which gives

$$\Phi_w(x) = p\left(\frac{\theta x}{1+\theta x}\right)^{\nu}\frac{1}{1-q\left(\frac{\theta x}{1+\theta x}\right)^{\nu}} \tag{15}$$

The rv *W* can be simulated also by the inverse method (use the *inverse* $\Phi_w^{-1}(U)$ of (15)).

Distributions of *V* and *W* when *N_* is *Poisson*(λ) .

The *truncated Poisson*(λ) distribution is

$$P(N^* = k) = \frac{1}{e^{\lambda}-1}\frac{\lambda^k}{k!}e^{-\lambda}\ ,\ k \geq 1\ . \tag{16}$$

The cdf of *V* is the mixture

$$\psi_v(x) = \frac{e^{-\lambda}}{e^{\lambda}-1}\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}\left[1-\left(1-\frac{\theta x}{1+\theta x}\right)^{\nu}\right]^{k},$$

which finaly is

$$\psi_v(x)=\frac{e^{-\lambda}}{e^{\lambda}-1}\left[e^{\lambda}-e^{\lambda\left(1-\left(\frac{\theta x}{1+\theta x}\right)^{\nu}\right)}\right]. \tag{17}$$

Therefore, the rv *V* can be simulated by the inverse method.

The cdf of *W* is the mixture

$$\psi_w(x)=\frac{e^{-\lambda}}{e^{\lambda}-1}\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}\left(\frac{\theta x}{1+\theta x}\right)^{k\nu}=\frac{e^{-\lambda}}{e^{\lambda}-1}\sum_{k=1}^{\infty}\frac{1}{k!}\left[\lambda\left(\frac{\theta x}{1+\theta x}\right)^{\nu}\right]^k$$

which finaly gives

$$\psi_w(x)=\frac{e^{-\lambda}}{e^{\lambda}-1}\left(e^{\lambda\left(\frac{\theta x}{1+\theta x}\right)^{\nu}}-1\right). \tag{18}$$

In this case, the rv *W* is also easy simulated by the inverse method.

Simulation of *V* and *W* when *N_* is *Binomial*(*n; p*).

The truncated *Binomial*(*n; p*), $0<p<1$, distribution for $N^*=k$, $k=1,2,\ldots,n$ has the *frequency* function

$$P\left(N^*=k\right)=p\frac{1}{1-q^n}C_n^k p^k q^{n-k},\ q=1-p,\ k=1,2,\ldots \tag{19}$$

The mixture cdf of *V* in this case is

$$\Phi_v(x)=\sum_{k=1}^{\infty}C_n^k p^k q^{n-k}\left(1-\left(1-\frac{\theta x}{1+\theta x}\right)^{\nu}\right)^k=\frac{1}{1-q^n}\sum_{k=1}^{\infty}C_n^k p^k q^{n-k}\left(1-\left(1-\frac{\theta x}{1+\theta x}\right)^{\nu}\right)^k$$

which finaly gives

$$\Phi_v(x)=1-\frac{1}{1-q^n}\left\{\left[p\left(1-\left(\frac{\theta x}{1+\theta x}\right)^{\nu}+q\right)\right]^n\right\}. \tag{20}$$

The mixture of *W* with truncated *Binomial*(*n; p*) is

$$\Phi_w(x)=\frac{1}{1-q^n}\sum_{k=1}^{\infty}C_n^k p^k q^{n-k}\left(\frac{\theta x}{1+\theta x}\right)^{\nu k}$$

which finaly gives

$$\Phi_w(x)=\frac{1}{1-q^n}\left\{\left[p\left(\frac{\theta x}{\theta x+1}\right)^{\nu}+q\right]^n-q^n\right\} \tag{21}$$

Since cdf's (20) and (21) can be easily inversed, the inverse method for simulating *V* and *W* can be applied. Finally, note that the *hypothesis that N_ is binomial is more realistic* for a sistem with *n* components (*n*= fixed), which might have *only* α components runing ($1<\alpha<n$) at a given time.

REFERENCES

[1] Demidovitch, B. *Recueil d'Exercices et de Problemes d'Analyse Mathematique*, Edition MIR Moscoou, 1971.
[2] Johnson, N.L., Kotz, S. *Distributions in statistics:continuous multivariate distributions.* John Wiley and Sons, New York, London, 1972.
[3] Tapan Kunar Nayak. "Multivariate Loomax distribution:properties and usefulnees in reliability". *J.Appl.Probab., p. 170-177, 1987.*
[4] Vaduva,I. "Computer generation of gamma random variables by rejection and composition procedures".*Math.Oper.Forsch.u.Statist., Ser.Statistics,Vol.8,Nr.4, pp546-576, 1977.*
[5] Vaduva,I. *Modele de simulare cu calculatorul,* Editura Tehnica, Bucuresti 1977, 358 p.
[6] Vaduva,I. *Fast algorithms for computer generation of random vectors used in reliability and applications.* Preprint Nr.1603, TH Darmstadt, FB Mathematik, Januar, 1994, 36p.
[7] Vaduva, I. *Fiabilitatea programelor.* Ed. Universit_at_ii Bucuresti, 2003, 160 p.
[8] Vaduva,I. *Modele de simulare.* Editura Universit_at_ii din Bucuresti, 2004, 190p.
[9] Vaduva,I. "Simulation of some mixed lifetime distributions". *Analele Univ.Bucuresti, Seria Informatica, 2011, p.10-19.*
[10] Vaduva,I. "On simulation of some life data distributions".*Analele Univ. "Spiru Haret", Vol. IX, Nr. 1, 2013, p.5-16.*

DEALING THE NONLINEARITY ASSOCIATED WITH THE DATA USING ARTIFICIAL NEURAL NETWORKS

Iuliana IATAN

Technical University of Civil Engineering Bucharest, Romania (iuliafi@yahoo.com)

DOI: 10.19062/1842-9238.2017.15.2.2

Abstract: *The Artificial Neural Networks (ANNs) are well-suited for a very broad class of nonlinear approximations and mappings. The ANN with nonlinear activation functions are more effective than linear regression models in dealing with nonlinear relationships. We are trying to find out how relevant is to use a Fuzzy Neural Network for prediction because it handles well the nonlinearity associated with the data.*

Keywords: *nonlinear, prediction, fuzzy neural network, personality*

1. INTRODUCTION

In this work we propose a specific neural network for predicting personality, by special type of Fuzzy Gaussian Neural Network (FGNN) understanding that it has so special the connections (between the second and third layers) and the operations with the nodes, too.

We shall propose to apply the FGNN for predicting a users' Big Five personality traits (the five factor model of personality) from the public information they share on Facebook. The Big Five traits are characterized by the following: *Neuroticism*, *Extraversion*, *Openness*, *Agreeableness*, and *Conscientiousness*.

To emphasize the performances of our proposed approach for predicting personality we have compared it both with a neural method of regression (like Multilayer Perceptron=MP) and with a non-neural approach Multiple Linear Regression Model (MLRM). The comparison of FGNN and respectively MP versus MLRM marks both the competition nonlinear over linear and of neural over statistical, too.

To test the performance of the neuro- fuzzy prediction achieved based on FGNN we shall use the Normalized Root Mean Square Error (NRMSE). According with the NRMSE criterion, we have achieved that the prediction with FGNN is better than with others two methods both over the training lot and over the test lot, too.

2. RELATED WORK

One distinguish some methods used in prediction with Social Media [12]: Bayes classifier, *K*-nearest neighbor classifier, Artificial Neural Networks, decision trees, model based prediction.

In [3], Golbeck made a Pearson correlation analysis between subjects' personality scores and each of the features obtained from analyzing their tweets and public account data. There are a number of significant correlations here, however none of them are strong enough to directly predict any personality trait. He described later in [4] the results of predicting personality traits through MLRM.

In the case of the MLRM applied in [4] for predicting personality, the optimal parameters were computed using the correlations between each profile feature and personality factor.

More recently, [10] studied the relationship between sociometric popularity (number of Facebook contacts) and personality traits on a far larger number of subjects.

The paper from [9] develops a fuzzy neural network approach to financial engineering; this model was successfully applied to the prediction of daily exchange rates (US Dollar-Romanian Lei). In this work we extend the application domain of fuzzy neural networks, viz. in the field of text mining, to predict personality traits. This FGNN having *M* output neurons is unlike the exchange rate FGNN [9], which uses a single neuron in the last layer to estimate the current exchange rate based on the previous m daily exchange rates.

We are trying to find out how relevant is to use the FGNN for predicting personality because it handles well the nonlinearity associated with the data.

3. BASELINES

The Artificial Neural Networks (ANNs) are well-suited for a very broad class of nonlinear approximations and mappings. The ANN with nonlinear activation functions are more effective than linear regression models in dealing with nonlinear relationships.

A feed-forward neural network is a nonparametric statistical model for extracting nonlinear relations in the data, namely it is a useful statistical tool for nonparametric regression.

A feed-forward neural network with an specific activation function is identical to a linear regression model:

- the input neurons are equivalent to independent variables or regressors;
- the output neuron is the dependent variable;
- the various weights of the network are equivalent to the estimated coefficients of a regression model.

Some advanced neural network techniques are related to more complex statistical methods such as:

1) kernel discriminant analysis
2) *k*-means cluster analysis
3) Principal Component Analysis(PCA).

Some neural networks do not have any close parallel in statistics, such as:

1) Kohonen’s self-organizing maps
2) Fuzzy Gaussian Neural Network.

The regression and correlation are related as the both of them are designed to extract relations between some variables.

In the case of a linear regression model, of the first order, "the slope of the regression line is the correlation coefficient times the ratio of the standard deviation of *y* to that of *x*."

MLRM is a method used to model multiple linear relationship between a dependent variable and more independent variables.

A major problem with multiple regression consists in the large number of predictors that are available, although only a few of them are actually significant.

The advantage of MLRM is that it can be implemented very easy.

Example 1. We are interested [6] in exploring for a sample of 32 vehicles the relationship between: the number of gears of a vehicle, the overall length (in inches) and its fuel efficiency (measured in miles per gallon).

```
>> x1=[3 3 3 3 3 3 3 3 4 5 3 4 4 3 4 3 3 3 3 3 3 3 3 3 5 4 3 5 3 3 3 3 3];
>> x2=[200.3 199.6 196.7 199.9 194.1 184.5 179.3 179.3 155.7 165.2 195.4 ...
160.6 170.4 171.5 168.8 199.9 224.1 231 196.7 197.6 179.3 214.2 196 165.2 ...
176.4 228 171.5 215.3 215.5 216.1 209.3 185.2];
>> y=[18.9 17 20 18.25 20.07 11.2 22.12 21.47 34.7 30.4 16.5 36.5 21.5 19.7...
20.3 17.8 14.39 14.89 17.8 16.41 23.54 21.47 16.59 31.9 29.4 13.27 23.9 ...
19.73 13.9 13.27 13.77 16.5];
>> X=[ ones(length(y),1) x1' x2'];
>> u=(X'*X)^-1*X'*y'

u =

   36.4857
    3.8272
   -0.1514

>> [xx1,yy1]=meshgrid(1:0.1:7,150:0.5:250);
>> zz=u(1)+u(2)*xx1+u(3)*yy1;
>> plot3(xx1,yy1,zz,x1,x2,y,'om')
```

FIG. 1. Matlab code

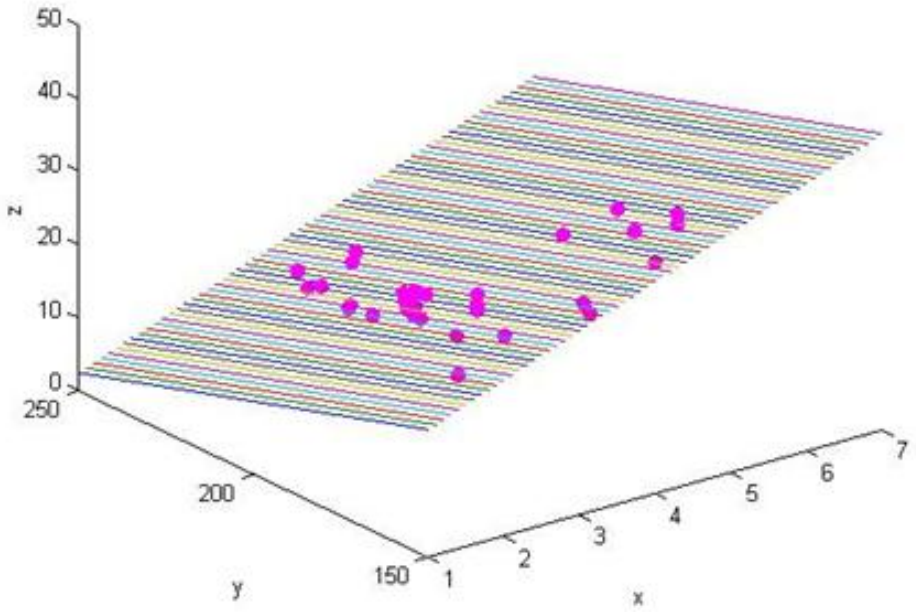


FIG. 2. Multiple linear regression through a scatter plot in space to which a plane of the form z = 36.4857 + 3.8272x + 0.1514y

Our FGNN represents [2] a modified version of Chen and Teng fuzzy neural network, by transforming the function of approximation into a function of classification. The change affects:

- the number of the classes (the number of the neurons belonging to the last layer);
- the equations of the fourth layer, but the structure diagram is similar.

4. FGNN ARCHITECTURE

The four-layer structure of the Fuzzy Gaussian Neural Network (FGNN) is shown in the Fig. 3.

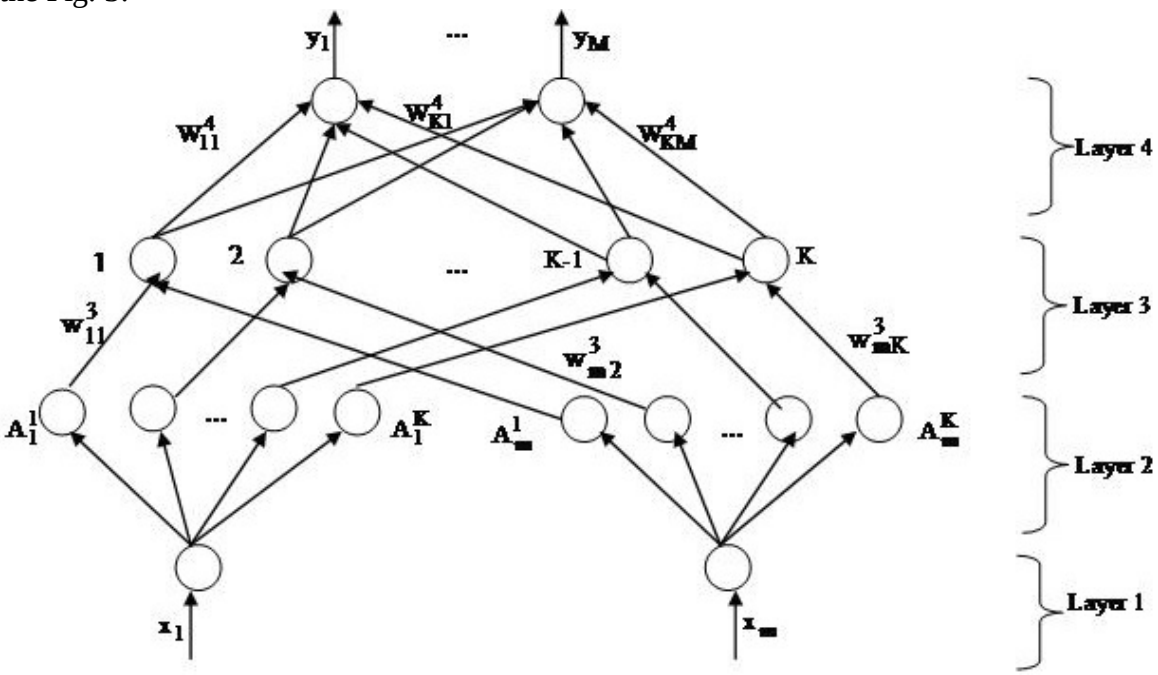


FIG. 3. Structure of FGNN

- *m* means the number of the neurons corresponding to the input layer;
- $X = (x_1, \ldots, x_m)$ represents the vector which one applies to the FGNN input;
- $\{W_{ij}^3\}_{i=\overline{1,m},\ j=\overline{1,K}}$ is the weight between the $(i-1)K + j$ -th neuron of the second layer and the neuron *j* of the third layer, where *K* is the number of the neurons from the third layer;
- $\{W_{ij}^4\}_{i=K,\ j=\overline{1,m}}$ is the connection from the neuron *i* from the third layer and the neuron *j* from the last layer of the FGNN;
- *M* represents the number of the classes;
- $Y = (y_1, \ldots, y_M)$ is the output of the FGNN.

The construction of FGNN is based on fuzzy rules of the form:

$\Re_j$: *If* x_1 *is* A_1^j *and* x_2 *is* A_2^j *... and* x_m *is* A_m^j, *then* y_1 *is* β_1^j, …, y_M *is* β_M^j,

where:

- *m* is the dimension of the input vectors (number of the retained features);
- $j,\ j = \overline{1,K}$ is the rule index;
- *M* is the number of the output neurons (it corresponds to the number of classes);
- $X = (x_1, \ldots, x_m)$ is the input vector, corresponding to the rule $\Re_j$;
- $A_i^j,\ i = \overline{1,m}$ are some fuzzy sets corresponding to the input vector;
- $Y = (y_1, \ldots, y_M)$ is the vector of the real outputs, corresponding to the rule $\Re_j$;
- $\beta_i^j,\ i = \overline{1,M}$ are some fuzzy sets corresponding to the output vector.

The *j*-th fuzzy rule is illustrated in Fig. 4.

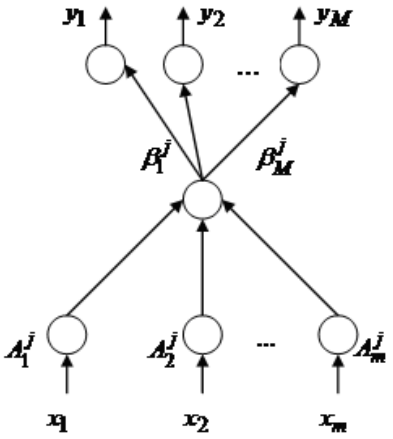


FIG. 4. The *j*-th component of FGNN

As in the case of the other neural networks, the FGNN input layer is a transparent layer, without a role in the data processing; the neurons of the first level (*input level*) do not process the signals; they only transmit the information to the next level.

The neurons of the second layer linguistic term layer (*level 2*) of the FGNN are membership neurons, resulting by the fuzzification of the first layer neurons. Each neuron of this level performs a Gaussian membership function [2], [6], [9]. The FGNN parameters have a physical significance, in the meaning that:

- m_{ij}, $i=\overline{1,m},\ j=\overline{1,K}$ represents the average;
- σ_{ij}, $i=\overline{1,m},\ j=\overline{1,K}$ is the variance

of the membership functions corresponding to some fuzzy sets, *m* being the number of the neurons from the input layer of the FGNN and *K* representing the number of the fuzzy considered rules.

The number of neurons characterizing this level is mK. Each input x_{ki}^2 is transformed by this layer into a fuzzy membership degree.

The third layer of the FGNN is called the *rule layer*. The connections between the membership neurons of the second layer and the rule neurons that characterize the third layer of the FGNN indicate the premise of the fuzzy rules. This layer computes the antecedent matching by the product operation [2], [6], [9].

The last layer of the FGNN is the output layer, which contains the output neurons. The conclusion (the consequence) of the rules is evidenced by the connections between the neurons of the third layer and the neurons of the output layer. This level performs the defuzzification of its inputs, providing *M* non-fuzzy outputs.

The FGNN parameters one initialize according to the *on- line initialization algorithm* [2], [6], [9] and they will be refined during the training algorithm [2], [6- 9].

5. TRAINING ALGORITHM

The training algorithm is of type back- propagation (BP), in order to minimize the error function:

$$E=\frac{1}{K}\cdot\sum_{k=1}^{K}E_k, \tag{1}$$

where:

- ✓ E_k represents the error for the rule *k*:

$$E_k = \frac{1}{2} \cdot \sum_{i=1}^{M} (d_{ki} - y_{ki})^2, \; k = \overline{1, K}, \tag{2}$$

- ✓ $d_k = (d_{k1}, \ldots, d_{kM})$ is the ideal output vector of the FGNN when at its input is applied the vector having the index *k*;
- ✓ $y_k = (y_{k1}, \ldots, y_{kM})$ is the corresponding real output vector of the FGNN ($k = \overline{1, K}$).

The training of this neural network is supervised, namely for of the K vectors from the training lot, we know the set of the ideal outputs. The refining of the FGNN parameters can be divided into two phases, depending on the parameters of premises and respective of conclusions of the rules, as follows:

A) in the part of the premise of the rules, the means and variances of the Gaussian functions one refine.
B) in the conclusions of the rules, the weights relating to the latest layer of FGNN must to be refined, the others being equal to 1.

7. EXPERIMENTAL EVALUATION

We are trying to find out how relevant is to use the Fuzzy Gaussian Neural Network for predicting personality because it handles well the nonlinearity associated with the data. We are also asking if the FGNN proves very good prediction performances over a statistical approach of prediction like MLRM and over a neural network as MP, too.

We use a data set made available by [3], [11]. The personality test called "The Big Five" (the five factor model of personality) represents the most comprehensive, reliable and useful test of personality concepts. It has emerged as one of the most well-researched and well-regarded measures of personality structure in recent years.

The Big Five traits are characterized [3] by the following:

- ➢ *Openness*: curious, intelligent, imaginative. High scorers tend to be artistic and sophisticated in taste and appreciate diverse views, ideas, and experiences;
- ➢ *Conscientiousness*: responsible, organized, persevering. Conscientious individuals are extremely reliable and tend to be high achievers, hard workers, and planners;
- ➢ *Extroversion*: outgoing, amicable, assertive. Friendly and energetic, extroverts draw inspiration from social situations;
- ➢ *Agreeableness*: cooperative, helpful, nurturing. People who score high in agreeableness are peace-keepers who are generally optimistic and trusting of others.
- ➢ *Neuroticism*: anxious, insecure, sensitive. Neurotics are moody, tense, and easily tipped into experiencing negative emotions.

The data is preprocessed in the following manner: we shall build a data set of 300 vectors, a half of them representing the training lot and the other half being the test lot.

These vectors have 20 components, each of them characterizing a personality trait. Each component means a correlation between the Big Five and individual words. For example [11]:

- *Neuroticism* correlates positively with negative emotion words (e.g. awful (0.26), though (0.24), lazy (0.24), worse (0.21), depressing (0.21), irony (0.21), terrible (0.2), road (-0.2), Southern (-0.2), stressful (0.19), horrible (0.19), sort (0.19), visited (-0.19), annoying (0.19), ashamed (0.19), ground (-0.19), ban (0.18), oldest (-0.18), invited (-0.18), completed (-0.18));
- *Extraversion* correlates positively with words reflecting social settings or experiences (e.g. Bar (0.23), other (-0.22), drinks (0.21), restaurant (0.21), dancing (0.2), restaurants (0.2), cats (-0.2), grandfather (0.2), Miami (0.2), countless (0.2), drinking (0.19), shots (0.19), computer (-0.19), girls (0.19), glorious (0.19), minor (-0.19), pool (0.18), crowd (0.18), sang (0.18), grilled (0.18));
- *Openness* shows strong positive correlations with words associated with intellectual or cultural experience (e.g. folk (0.32), humans (0.31), of (0.29), poet (0.29), art (0.29), by (0.28), universe (0.28), poetry (0.28), narrative (0.28), culture (0.28), giveaway (-0.28), century (0.28), sexual (0.27), films (0.27), novel (0.27), decades (0.27), ink (0.27), passage (0.27), literature (0.27), blues (0.26));
- *Agreeableness* correlates with words like: wonderful (0.28), together (0.26), visiting (0.26), morning (0.26), spring (0.25), porn (-0.25), walked (0.23), beautiful (0.23), staying (0.23), felt (0.23), share (0.23), gray (0.22), joy (0.22), afternoon (0.22), day (0.22), cost (-0.23), moments (0.22), hug (0.22), glad (0.22), fuck (-0.22);
- *Conscientiousness* has strong positive correlations with words like: completed (0.25), adventure (0.22), stupid (-0.22), boring (-0.22), adventures (0.2), desperate (-0.2), enjoying (0.2), saying (-0.2), Hawaii (0.19), utter (-0.19), extreme (-0.19), it's (-0.19), deck (0.18).

We want to predict M components (M being the number of the neurons from the output layer of FGNN) for every vector in order to complete the behavior corresponding to a person.

For evaluation, we use the Normalized Root Mean Square Error (NRMSE) [1].

Following [13], the prediction is considered:

- **excellent** if NRMSE≤0.1;
- **good** if 0.1 < NRMSE≤0.2;
- **fair** if 0.2 < NRMSE≤0.3;
- **poor** if NRMSE > 0.3.

For significance testing we use the three models: Multiple Linear Regression Model (MLRM), Multilayer Perceptron (MP) and Fuzzy Gaussian Neural Network (FGNN).

The three models: MLRM, MP and FGNN have been evaluated using a corresponding test lot, having a number of vectors equal to that of the training lot. The performance of FGNN over MP is based on [6] the fuzzy properties of FGNN, while the MP is a crisp neural network. The comparison [6] of FGNN and respectively MP versus MLRM marks both the competition nonlinear over linear and of neural over statistical, too.

CONCLUSIONS

The ability to predict personality has implications in many areas:

- like other studies relating to personality and language we adopted the five factor model of personality, which describes the following traits on a continuous scale: *neuroticism*, *extraversion*, *openness*, *agreeableness* and *conscientiousness*;

- ✓ in justice as the **personality rights** are some non- patrimonial civil rights, being regulated in article 58 NCC(New Civil Code); the protection of human personality is regulated by the Constitution of the Romania and the NCC(see the article 252).

To emphasize the performances of our proposed approach for predicting personality we have compared it both with a neural method of regression (like MP) and with a nonneural approach (MLRM), too.

According with the NRMSE criterion, we have achieved that the prediction with FGNN is better than with others two methods both over the training lot and over the test lot, too.

The advantage of the FGNN consists in the fact that, for certain values of the overlapping parameters one achieve very good recognition rates of the test lot.

A major problem with multiple regression consists in the large number of predictors that are available, although only a few of them are actually significant.

REFERENCES

[1] M. Benaddy, M. Wakrim and S. Aljahdali, *Evolutionary prediction for cumulative failure modeling: A comparative study*, in 2011 Eighth International Conference on Information Technology: New Generations, pp. 41-47, 2011;

[2] C. Y. Chen and C.C. Teng, *Fuzzy Neural Network System in Model Reference Control Systems* in Fuzzy Logic and Expert Systems Applications, pp. 285-313, Academic Press, San Diego-Toronto, 1998;

[3] J. Golbeck, C. Robles, M. Edmondson and K. Turner, *Predicting personality from TWITTER*, in IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, pp. 149-156, 2011;

[4] J. Golbeck, C. Robles and K. Turner, *Predicting personality with Social Media*, in IEEE International Conference on Privacy, Security, Risk, and Trust, and Proceedings of the ACM Conference on Human Factors in Computing, pp. 253-262, 2012;

[5] I. Iatan, *Neuro- Fuzzy Systems for Pattern Recognition* (in Romanian), Ph.D. Thesis, Faculty of Electronics, Telecommunications and Information Technology- University Politehnica of Bucharest, PhD supervisor: Prof. dr. Victor Neagoe, 2003;

[6] I. Iatan, *Issues in the Use of Neural Networks in Information Retrieval*, Springer-Verlag Berlin Heidelberg, 2016;

[7] V. Neagoe, and I. Iatan, *Face recognition using a fuzzy-gaussian neural network*, in Proceedings of First IEEE International Conference on Cognitive Informatics, ICCI 2002, 19-20 August 2002, Calgary, Alberta, Canada, pp. 361-368, 2002.

[8] V. Neagoe, I. Iatan and S. Grunwald, *A neuro- fuzzy approach to ECG signal classification for ischemic heart disease diagnosis* in the American Medical Informatics Association Symposium (AMIA 2003), Nov. 8- 12 2003, Washington DC, pp. 494-498, 2003.

[9] V. Neagoe, R. Iatan and I. Iatan, *A nonlinear neuro-fuzzy model for prediction of daily exchange rates* in Proceedings of World Automation Congress WAC'04 Seville, Spain, 17, pp. 573-578, 2004.

[10] D. Quercia, R. Lambiotte, D. Stillwell, M. Kosinski and J. Crowcroft, *The personality of popular FACEBOOK users* in 2012 ACM Conference on Computer Supported Cooperative Work (CSCW 2012), Session: Social Network Analysis, pp. 955-964, 2012.

[11] T. Yarkoni, Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers, *Journal of Research in Personality*, vol. 44, pp. 363-373, 2010;

[12] S. Yu and S. Kak. *A survey of prediction using Social Media*, 2012. Available at http://arxiv.org/ftp/arxiv/papers/1203/1203.1647.pdf

[13] H. Zeng, L. Li, J. Hu, L. Liang, J. Li, B. Li and K. Zhang, Accuracy validation of TRMM multisatellite precipitation analysis daily precipitation products in the lancang river basin of China, *Theoretical and Applied Climatology*, pp. 1-13, 2012.

A FAST SELF-ADAPTIVE APPROACH TO RELIABILITY OPTIMIZATION PROBLEMS

George ANESCU, Paul ULMEANU

"Politehnica" University, Power Engineering Faculty, Bucharest, Romania
(george.anescu@gmail.com, paul.ulmeanu@upb.ro)

DOI: 10.19062/1842-9238.2017.15.2.3

Abstract: *The paper is investigating the suitability of the FSA-DE optimization method for solving reliability optimization problems by approaching a set of three case studies: a known RAP case study, a FTO case study and a ETO case study. For the RAP case study the numerical results obtained by FSA-DE are compared with the ones obtained by other known optimization methods.*

Keywords: *Mixed Integer Non-Linear Programming (MINLP), Global Optimization (GO), Fast Self-Adaptive Differential Evolution (FSA-DE), Reliability Optimization, Redundancy Allocation Problem (RAP), Fault Tree Optimization (FTO), Event Tree Optimization (ETO)*

1. INTRODUCTION

In order to become more competitive on the market, many manufacturers are investing resources for improving the reliability of the systems and components they produce. Two approaches are commonly used in order to reach a high reliability of a system.

In the first approach the system's reliability is increased during the design phase by increasing the number of redundant components in the various subsystems of the considered system. But by increasing the number of identical components there are also involved increases in the cost, the weight or the volume of the sub-systems, which impose additional constraints on the overall cost, weight or volume of the system. This first model is called the Redundancy Allocation Problem (*RAP*) and was first introduced by Fyffe et al. in [1]. There are many varieties of RAP problems in the field of reliability optimization, which were widely investigated by using many optimization methods, including the meta-heuristic ones. For an overview of *RAP* see Kuo and Prasad [2], and for surveys of the most recent research advances in *RAP* problems see Kuo [3] and Chambari et al. [4].

In the second approach the system reliability is increased by increasing the reliability of the components, and it can be applied to both the design and operational phases of the system. In order to determine the components which should be considered for reliability improvement and their optimal reliability values, taking into account that there are also some economic cost limitations, as opposed to the reliability (safety) requirements, some combinations of Fault Tree Analysis (*FTA*) (see [5]) techniques and mathematical optimization techniques are employed. This second model is called the Fault Tree Optimization (*FTO*) problem and it was investigated by applying mainly Genetic Algorithms (*GA*) optimization techniques (see [6], [7]).

The models and methodologies based on probabilistic risk analysis and optimization can be extended from optimizing the design and operation of systems and sub-systems to optimizing the design and operation of complex industrial systems, like nuclear power plants, or fossil power plants (see [8]).

The goal of such a methodology is to minimize the risk to have a nuclear accident or the economic risk to shut down the production for all the possible reasons.

In design the focus is on component quality and redundancy levels, while in maintenance and testing the focus is on scheduling tasks and human reliability. After modeling the systems and sub-systems by using the *FTA* methodology, the next step is to model the Accidental Sequences (*AS*) with Event Trees (*ET*) ([9]). In an *AS* several systems are performing their functions successfully or unsuccessfully. The combination of different systems performing their functions right or wrong drive the *AS* to different final Plant States (*PS*) which can be grouped, according to the degree of damage produced, as totally successful, partially successful, or unsuccessful. The plant states can be quantified and some constraints can be imposed on the unsuccessful states according to some permissible upper and lower risk limits. When the total investment and the operating budget is limited, the Event Tree Optimization (*ETO*) problem consists in how to optimally distribute the funds so that all the unsuccessful Plant States in *ET* are observing the imposed permissible risk limits.

2. MIXED INTEGER NON-LINEAR PROGRAMMING (*MINLP*) PROBLEM

The most general form of the reliability optimization problems treated in this paper is the *MINLP* formulation where equality or inequality constraints can be applied to the objective function and some of the decision variables can take continuous real values in real intervals, while other decision variables are restricted to integer values in sets of consecutive integer values ([10]):

$$\begin{aligned} &\text{minimize} && f(\mathbf{x}) \\ &\text{subject to} && \mathbf{x} \in \mathrm{D} \end{aligned} \tag{1}$$

with:

$$\mathrm{D} = \left\{ \mathbf{x}: \mathbf{l} \leq \mathbf{x} \leq \mathbf{u};\ \text{and}\ g_i(\mathbf{x}) \leq 0, i = 1, \ldots, \mathrm{G};\ \text{and}\ h_j(\mathbf{x}) = 0, j = 1, \ldots, \mathrm{H} \right\} \tag{2}$$

where, $\mathbf{x} \in \mathbb{R}^n$ is a real n-dimensional vector of decision variables ($\mathbf{x} = (x_1, x_2, \ldots, x_n)$), there is a number $0 \leq n_C \leq n$ such that the last $n - n_C$ decision variables are restricted to integer values, $f: \mathbb{R}^n \to \mathbb{R}$ is the continuous objective function, $\mathrm{D} \subset \mathbb{R}^n$ is the non-empty set of feasible decisions (a proper subset of $\mathbb{R}^n$), $\mathbf{l}$ and $\mathbf{u}$ are explicit, finite (component-wise) lower and upper bounds of $\mathbf{x}$, $g_i: \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, \mathrm{G}$ is a finite collection of continuous inequality constraint functions, and $h_j: \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, \mathrm{H}$ is a finite collection of continuous equality constraint functions. No other additional suppositions are made on the *MINLP* problem and it is assumed that no additional knowledge about the objective function and constraint functions can be obtained, in this way treating the *MINLP* problem as a black box, i.e. for any point $\mathbf{x}$ in the boxed domain $\{\mathbf{x}: \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}\}$ it is assumed the ability to calculate the values of the functions $f(\mathbf{x})$, $g_i(\mathbf{x}), i = 1, \ldots, \mathrm{G}$, $h_j(\mathbf{x}), j = 1, \ldots, \mathrm{H}$, but nothing more. In order to efficiently handle the constraints in constrained optimization problems one of the best approaches is to apply the Deb's Rules (see [11]). For a detailed constraints handling methodology based on Deb's Rules see [12]. In the *MINLP* model another important issue is the handling of the integer constraints. In the population based meta-heuristic optimization methods the integer decision variables are treated like the continuous variables, but when the objective function $f(\mathbf{x})$ is evaluated the values rounded to the closest integer, $x'_j = round(x_j), j = n_C + 1, \ldots, n$, are used in the evaluation.

In [12] the *FSA-DE* variant of Differential Evolution (*DE)* was constructed by implementing and experimentally testing a set of four gradual and cumulative improvements to the initial *DE/rand/1/bin* scheme (originally introduced in [13]): 1) a randomization of the scaling control parameter in the real interval $[0,1)$, 2) a Random Greedy Selection method (*RGS*, see[14]);

3) the use of a normal (Gaussian) probability distribution for sampling the crossover probability, and 4) a resetting mechanism. *FSA-DE* proved better performance, while the dependence on method parameters was eliminated.

3. RAP CASE STUDY

We consider a known *RAP* case study, the 5-unit bridge structure shown in **FIG. 1** (see [15]):

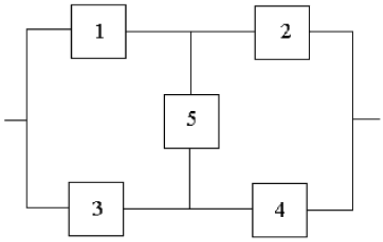


FIG. 1. The schematic diagram of 5-unit bridge system ([15])

In order to efficiently compute the reliability of this system we first eliminate the bridge components by applying the Bayes Total Probability Theorem:

$$R_S(\mathbf{r},\mathbf{n}) = R(s_5)R(S|s_5) + R(\bar{s}_5)R(S|\bar{s}_5) = \\ = R_5R(S_1) + (1 - R_5)R(S_2) \tag{3}$$

where we used the notations S for the original bridge system, s_1, s_2, …, s_5 for the boolean states of the sub-systems (*true* means perfectly working, while *false* means totally defective) and $R_i = R(s_i),\ i = 1,\dots,5$ are the reliabilities of the sub-systems. The simplified systems $R(S_1)$ and $R(S_2)$ are of serial-parallel type and their reliabilities can be easily calculated: $R(S_1) = (1-(1-R_1)(1-R_3))(1-(1-R_2)(1-R_4))$ and $R(S_2) = 1-(1-R_1R_2)(1-R_3R_4)$. The sub-systems are comprising identical redundant (parallel connected) components, and their reliabilities are simply calculated as $R_i = 1-(1-r_i)^{n_i},\ i = 1,\dots,5$, with n_i the number of components in subsystem i. The optimization problem for the 5-unit has the dimension 10, the decision variables being the numbers of components n_i and the component reliabilities r_i:

$$\begin{aligned} &\text{maximize} && R_S(\mathbf{r},\mathbf{n}) \\ &\text{subject to} && C_S(\mathbf{r},\mathbf{n}) \le C_{lim},\ W_S(\mathbf{r},\mathbf{n}) \le W_{lim},\ V_S(\mathbf{r},\mathbf{n}) \le V_{lim} \end{aligned} \tag{4}$$

where: $C_S(\mathbf{r},\mathbf{n}) = \sum_{i=1}^{5} \alpha_i \left(-\frac{T}{\ln(r_i)}\right)^{\beta_i} (n_i + e^{n_i/4})$, $W_S(\mathbf{r},\mathbf{n}) = \sum_{i=1}^{5} w_i n_i e^{n_i/4}$ and $V_S(\mathbf{r},\mathbf{n}) = \sum_{i=1}^{5} w_i v_i {n_i}^2$. The other constraints applied to the 5-unit bridge system are: $1 \le n_i \le 10, 0.0 \le r_i \le 1.0,\ i = 1,\dots,5$, with n_i taking integer values and r_i taking real values. The design data are $C_{lim} = 175.0$, $W_{lim} = 200.0$, $V_{lim} = 110.0$ and $T = 1000h$. The other design constants are the ones given in [15]. **Table 1** gives the best results obtained by *FSA-DE* method for 5-unit bridge system problem.

Table 1 – *RAP* Case Study, 5-Unit Bridge System, *FSA-DE* results

Objective	Stage	r_i	n_i	Attribute
$Max\ R_S$	1	0.828045581	3	$R_S = 0.99988964$
	2	0.857778608	3	$C_S = 74.9994$
	3	0.914351326	2	$W_S = 198.439534$
	4	0.648110393	4	$V_S = 105$
	5	0.704001133	1	

Table 2 gives some comparison results between various other recently employed optimization methods and *FSA-DE* method for the 5-unit bridge system case study, and it can be observed that *FSA-DE* method was able to achieve the best known maximum reliability.

Table 2 - *RAP* Case Study, 5-Unit Bridge System, comparative results

Parameter	*HS* [16]	*IPSO* [17]	*ABC* [18]	*ICS* [15]	*FSA-DE*
$R_S(\mathbf{r}, \mathbf{n})$	0.99988962	0.99988963	0.99988962	0.99988964	**0.99988964**
n_1	3	3	3	3	**3**
n_2	3	3	3	3	**3**
n_3	2	2	2	2	**2**
n_4	4	4	4	4	**4**
n_5	1	1	1	1	**1**
r_1	0.82883148	0.82868361	0.828087	0.828094038	**0.828045581**
r_2	0.85836789	0.85802567	0.857805	0.858004485	**0.857778608**
r_3	0.91334996	0.91364616	0.914240	0.914162924	**0.914351326**
r_4	0.64779451	0.64803407	0.648146	0.647907792	**0.648110393**
r_5	0.70178737	0.70227595	0.704163	0.704565982	**0.704001133**

6. FTO CASE STUDY

The first step in a *FTA* methodology involves the construction of the fault tree representation of the system. Usually a top-down approach is adopted, starting from the definition of the general failure condition of the system (the top event) and logically developing the fault tree structure, through *OR*, *AND* and *NOT* logical gates, from more general failure events to more specific failure events associated to the sub-systems and the components of a system. When the logic cannot be further developed, the last generated events are considered the basic events of the fault tree. Each basic event has an associated reliability r_i and the corresponding amount of investment c_i which is needed to achieve the reliability. For each basic event it is needed a reliability-cost curve, which is available from mathematical modeling and calculation, historical data, or can be provided by the manufacturer. A typical reliability-cost curve is presented in **FIG. 2**, where it can be observed that the reliability is increasing with the cost and it is asymptotically approaching the value of 1.0 with a very high cost. When the costs are known, once calculated the reliabilities of the basic events $r_i(c_i)$, the overall reliability of the system can be calculated using the fault tree logic. We can define the following optimization problem for Reliability Maximization (achieving the maximum possible system reliability within a given amount of investment):

maximize $R_{TOP}(\mathbf{c})$

subject to $\sum_{i=1}^{N_{be}} c_i \leq C_{tot}$ (5)

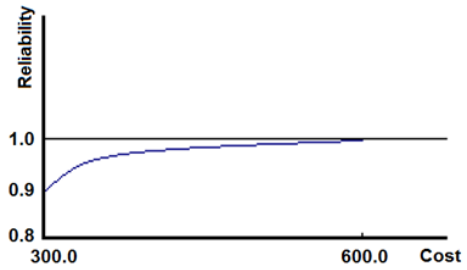


FIG. 2. A typical reliability-cost curve ([6])

where N_{be} is the number of basic events, $c_i, i = 1, \dots, N_{be}$ are the associated costs and C_{tot} is the total investment. We built a simple *FTO* case study starting from a simple fault tree with seven basic events (see [19]) presented in **FIG. 3**.

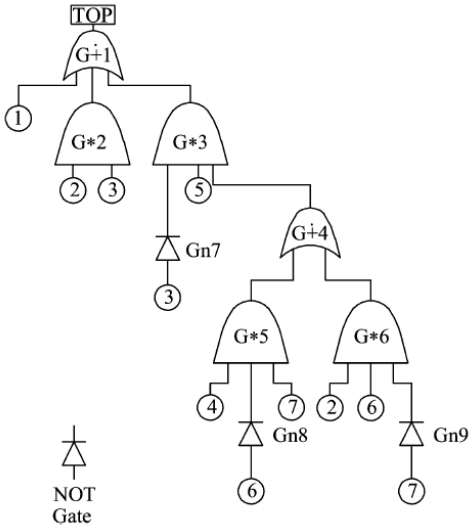


FIG. 3. A fault tree example ([19])

We can logically evaluate the *TOP* event by applying a top-down approach: $TOP = e_1 + G_2 + G_3 = e_1 + e_2 e_3 + \bar{e}_3 e_5 G_4$, with $G_4 = G_4 + G_4 = e_4 \bar{e}_6 e_7 + e_2 e_6 \bar{e}_7$. In order to evaluate the system reliability we apply recursively the Bayes Total Probability Theorem by first eliminating the basic events which appear multiple times, and after that evaluating the remaining tree structures by applying a simple bottom up approach. We have:

$$R_{TOP} = R(TOP) = 1 - Q(TOP) \tag{6}$$

$$\begin{aligned} Q(TOP) &= q_3 Q(TOP|e_3) + r_3 Q(TOP|\bar{e}_3) = \\ &= q_3(1 - r_1 r_2) + r_3\left(1 - r_1\left(1 - q_5 Q(G_4)\right)\right) \end{aligned} \tag{7}$$

with

$$Q(G_4) = q_6 Q(G_4|e_6) + r_6\, Q(G_4|\bar{e}_6) = q_6 q_2 r_7 + r_6 q_4 q_7 \tag{8}$$

We modeled the reliability-cost curve by simply using the hyperbolic tangent function:

$$r_i(c) = \frac{e^{\alpha_i c} - 1}{e^{\alpha_i c} + 1} \tag{9}$$

with the constants $\alpha_i > 0,\ i = 1, \dots, 7$ (which control the slope of the curve) given in **Table 3**.

We assumed that we started from an initial investment of 7000 units which is distributed among the basic events, so that with the minimal costs $c_{min,i}, i = 1, ..., 7$ given in **Table 3**, the unreliabilities $q0_i, i = 1, ..., 7$ are the same as given in [19].

Table 3 - Data for *FTO* case study

i	α_i	$q0_i$	$c_{min,i}$
1	0.003701	0.01614	1300.0
2	0.004292	0.0625	800.0
3	0.004215	0.3125	400.0
4	0.003701	0.01614	1300.0
5	0.002837	0.00125	2600.0
6	0.003662	0.5	300.0
7	0.003662	0.5	300.0

The available investment was $C_{tot} = 12000$ units, with an additional investment of 5000 units. **Table 4** gives the best result obtained by *FSA-DE* method. It can be observed that with an additional investment of 5000 units distributed among the basic events e_1, e_2 and e_4 the system's reliability was increased from 0.96461053 (which is obtained with the minimal investment of 7000 units) to 0.99999577.

Table 4 - *FTO* case study, *FSA-DE* results

c_1	c_2	c_3	c_4	c_5	c_6	c_7	$R_{TOP}(\mathbf{c})$
3815.276	3053.641	400.0	1531.068	2600.0	300.0	300.0	0.99999577

7. ETO CASE STUDY

The *ETO* case study considered in this section (see **FIG. 4**.) is based on a simplified Event Tree obtained from the original Event Tree built for CAREM 25 Project ([20]). CAREM 25 is a CNEA (Comisión Nacional de Energía Atómica) project from Argentina aiming to develop, design and construct a small nuclear power plant with an electrical output of about 27 MW. According to [20], the Accidental Sequences (*AS*) were built simplifying the headers to show only the human error intervention. Five models were taken into account for the representation of the human behavior: 1) Technician, 2) Technician and supervision, 3) Technician and supervision with written procedures, 4) Technician and administrative control, 5) Technician, supervision and administrative control.

To each human behavior model a human error probability (or error frequency) $\boldsymbol{q_j, j = 1, ..., 5}$, was associated, as determined in [20] by applying Human Event Tree (*HEP*) modeling. Also a cost, $\boldsymbol{c_j, j = 1, ..., 5}$, was associated, the data being presented in **Table 5**. The *AS*s in an *ET* are initiated at the left by an undesired event ue, which in the considered case study also came from a human error. The ASs are simplified in order to show in the headers only the human interventions : $\boldsymbol{he_1, he_2, he_3, he_4}$. An up branch is representing a successful intervention and an down branch is representing a wrong intervention. On the right hand side the final Plant States (*PS*) are represented.

The successful *PS*s and the successful human interventions are represented underlined. For an unsuccessful *PS* a frequency between 10^{-7} and 10^{-9} is considered a reasonable value. Any value lower than 10^{-9} represents a too high cost and any value higher than 10^{-7} represents a very high risk.

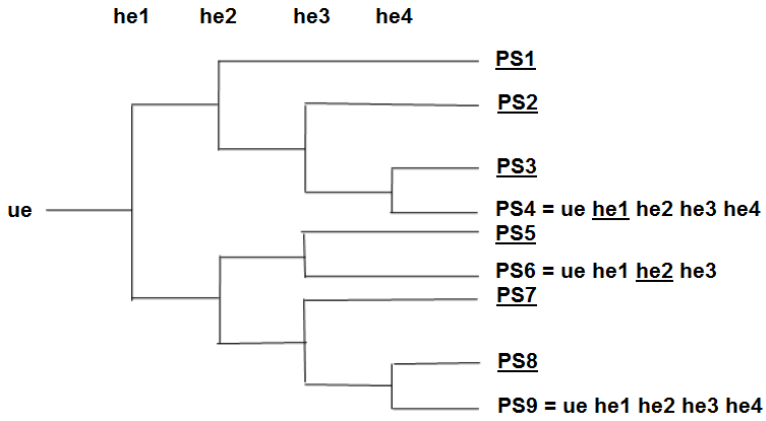


FIG. 4. The simplified CAREM 25 Event Tree ([20])

Table 5 - Data for *ETO* case study

Model	1	2	3	4	5
error freq.	0.505	0.1009	0.001105	0.0209	0.0011049
Cost	240.0	520.0	1530.0	950.0	1670.0

We are required to solve here a combinatorial type of optimization problem where we need to associate to each human intervention $he_i, i = 1, \dots, 4$ in the *ET* an appropriate model $m_i, i = 1, \dots, 4$, from the five given models, so that the total cost associated to the human intervention is minimal, while the frequencies of the unsuccessful *PS*s are observing the required constraints. This problem can be easily modeled as an integer programming problem by taking as decision variables the indices of the selected human behavior models:

minimize $C_{tot} = \sum_{i=1}^{4} C(m_i)$

subject to $10^{-9} \leq Q(PS_4), Q(PS_6), Q(PS_9) \leq 10^{-7}$ (10)

where $Q(PS_4) = Q(ue\ \overline{he_1} he_2 he_3 he_4) = Q(ue)(1 - Q(m_1))Q(m_2)Q(m_3)Q(m_4)$, $Q(PS_6) = Q(ue\ he_1 \overline{he_2} he_3) = Q(ue)Q(m_1)(1 - Q(m_2))Q(m_3)$ and $Q(PS_9) = Q(ue\ he_1 he_2 he_3 he_4) = Q(ue)Q(m_1)Q(m_2)Q(m_3)Q(m_4)$. The frequency associated to the undesired event was $Q(ue) = 1.1 \times 10^{-7}$. **Table 6** gives the best result obtained by *FSA-DE* method for *ETO* case study.

Table 6 - *ETO* case study, *FSA-DE* results

m_1	m_1	m_1	m_1	C_{tot}
2	1	3	2	2810.0

CONCLUSIONS

The paper investigated the suitability of the *FSA-DE* optimization method for solving reliability optimization problems. *FSA-DE* is advantageous over other optimization methods since it is an almost parameter free method. First a known *RAP* case study was investigated by applying the *FSA-DE* optimization method and the obtained results were compared with other results published in the literature.

Finally, for illustrative purposes, two new optimization case studies were built inspired from published information: a *FTO* case study and an *ETO* case study, and the numerical optimization results obtained by applying the *FSA-DE* method were presented. The study proved that *FSA-DE* is a competitive optimization method for solving reliability optimization problems.

REFERENCES

[1] D.E. Fyffe, W.W. Hines and N.K. Lee, System reliability allocation and a computational algorithm, *IEEE Transaction on Reliability*, vol. 17, no. 2, pp. 64-69, 1968;

[2] W. Kuo and V.R. Prasad, An annotated overview of system reliability optimization, *IEEE Transaction on Reliability*, vol. 49, no. 2, pp. 176-187, 2000;

[3] W. Kuo, Recent advances in optimal reliability allocation, *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 37, no. 2, pp. 143-156, 2007;

[4] A. Chambari, S.H.A. Rahmati, A.A. Najafi and A. Karimi, A bi-objective model to optimize reliability and cost of system with a choice of redundancy strategies, *Computers and Industrial Engineering*, vol. 63, no. 1, pp. 109-119, 2012;

[5] W. Vesely et al., *Fault Tree Handbook*, NUREG-0492, Ed. NRC, Washington, 1981;

[6] A Libošvárová and P. Schreiber, Optimization of Technical System by Using FTA and Genetic Algo rithm, *Applied Mechanics and Materials*, vol. 693, pp. 135-140, 2014;

[7] F. De Carlo, M. Iacono and S. Marrone, *Combining Genetic Algorithm and Fault Tree Analysis in Reliability/Cost Optimization for Critical Complex Systems*, Proc. of 9th Modern Information Technology in the Innovation Processes of the Industrial Enterprises (MITIP), 2007;

[8] J.E. Núñez Mc Leod, S.S. Rivera and J.H. Barón, *Optimizing designs based on risk approach*, WCE '07, Proceedings of the World Congress on Engineering 2007 Vol II, July 2-4, London, U.K., pp. 1044-1049, 2007;

[9] *PRA Procedures Guide: A guide to the performance of probabilistic risk assessments for nuclear power plants review*, NUREG/CR 2300, Ed. NRC, Washington, 1983;

[10] J.D. Pintér, *Global Optimization: Software, Test Problems, and Applications*, Ch. 15 in Handbook of Global Optimization, Volume 2, P.M. Pardalos and H.E. Romeijn, Eds. Kluwer Academic Publishers, Dordrecht, pp. 515-569, 2002;

[11] K. Deb, *An efficient constraint handling method for genetic algorithms*, Computer Methods in Applied Mechanics and Engineering, vol. 186, no. 2-4, pp. 311-338, 2000;

[12] G. Anescu, *Gradual and Cumulative Improvements to the Classical Differential Evolution Scheme through Experiments*, Annals of West University of Timisoara, Mathematics and Computer Science, Timisoara, vol 54, no. 2, pp. 13-36, 2016;

[13] K. Price and R. Storn, *Differential Evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces*, Journal of Global Optimization, vol. 11, no. 4, pp. 341-359, 1997;

[14] G. Anescu and I. Prisecaru, *NSC-PSO, a novel PSO variant without speeds and coefficients*, Proceedings of 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2015, Timisoara, Romania, 21-24 September, pp. 460-467, 2015;

[15] E. Valian, S. Tavakoli, S. Mohanna and A. Haghi, *Improved cuckoo search for reliability optimization problems*, Computers and Industrial Engineering, vol. 64, no. 1, pp. 459-468, 2013;

[16] D. Zou, L. Gao, J. Wu, S. Li and Y. Li, *A novel global harmony search algorithm for reliability problems*, Computers and Industrial Engineering, vol 58, no. 2, pp. 307-316, 2010;

[17] P. Wu, L. Gao, D. Zou and S. Li, *An improved particle swarm optimization algorithm for reliability problems*, ISA Transactions, vol. 50, no. 1, pp. 71-81, 2010;

[18] W.C. Yeh and T.J. Hsieh, *Solving reliability redundancy allocation problems using an artificial bee colony algorithm*, Computers and Operations Research, vol. 38, no. 11, pp. 1465-1473, 2011;

[19] A.P. Ulmeanu, *Analytical method to determine uncertainty propagation in fault trees by means of binary decision diagrams*, IEEE Transactions on Reliability, vol 61, no. 1, pp. 84-94, 2012;

[20] J.E. Núñez Mc Leod and S.S. Rivera, *Human error management optimization in CAREM NPP*, WCE '09, Proceedings of the World Congress on Engineering 2009 Vol I, July 1-3, London, U.K., pp. 1044-1049, 2007.

DECOMPOSITION OF THE TIME SERIES AND OF SHOCKS USING THE SIMPLE FRACTIONS DECOMPOSITION AND APPLICATIONS

Daniel CIUIU

Technical University of Civil Engineering Bucharest (dciuiu@yahoo.com)

DOI: 10.19062/1842-9238.2017.15.2.4

Abstract: *In this paper we will use the decomposition of rational functions in simple fractions. The rational functions are build using the delay polynomials* $\varphi(L)$ *and* $\theta(L)$ *of an ARIMA time series.*

For decomposition of the time series X_t *we use the rational fraction* $\frac{\theta(L)}{\varphi(L)}$*, and for the decomposition of the white noise* a_t *we use the rational fraction* $\frac{\varphi(L)}{\theta(L)}$*.*

Finally, because for the decomposition of X_t *we do not take into account that the roots of* $\varphi(L)$ *are greater than one in absolute value, we eventually multiply in the first above case* $\varphi(L)$ *by* $(1-L)^d$ *for taking into account the possible trend and by* $\left(1-L^s\right)^{d_s}$ *for taking into account the possible seasonal components.*

Keywords: *ARMA and ARIMA time series, delay operator, delay polynomials.*

1. INTRODUCTION

The classical decomposition of time series is [3,4] in seasonal components, trend and a stationary component. The remaining stationary component, even we obtain it by removing seasonal components and trend, even we obtain it by seasonal and non-seasonal differentiation, is modeled as $AR(p)$, $MA(q)$ or $ARMA(p,q)$ time series.

For an $ARMA(p,q)$ we can write

$$\varphi(L)X_t=\theta(L)a_t, \text{ where} \tag{1}$$

$$\begin{cases}\varphi(L)=1-\sum_{i=1}^{p}\varphi_i L^i\\ \theta(L)=1-\sum_{i=1}^{q}\theta_i L^i\end{cases}. \tag{1'}$$

Using the above formula, we obtain [2,3,4]

$$\begin{cases}X_t=\frac{\theta(L)}{\varphi(L)}a_t\\ a_t=\frac{\varphi(L)}{\theta(L)}X_t\end{cases}. \tag{1''}$$

2. THE DECOMPOSITION OF TIME SERIES

For decomposition of X_t we use the decomposition in simple fractions of $\frac{\theta(L)}{\varphi(L)}$. Denote now the roots of $\varphi(L)$ x_1, ..., x_l with the multiplicities m_1, ..., m_l.

If X_t is $ARMA(p,q)$ with $p > q$, there exists the white noise a_t such that

$$X_t = \sum_{i=1}^{l}\sum_{j=1}^{m_i} \frac{A_{ij}}{(1 - x_i \cdot L)^j} a_t . \qquad (2)$$

If we have $p \le q$, the above formula becomes

$$X_t = \tilde{\theta}(L) a_t + \sum_{i=1}^{l}\sum_{j=1}^{m_i} \frac{A_{ij}}{(1 - x_i \cdot L)^j} a_t, \text{ where} \qquad (2')$$

$\tilde{\theta}(L)$ is the quote of $\frac{\theta(L)}{\varphi(L)}$.

In formulae (2) and (2') the roots of $\varphi(L)$ can be complex. In this case we can group the conjugate complex roots. We obtain at denominator $\left(1 - 2\,\mathrm{Re}(z_i) + |z|^2\right)^{m_i}$, and the numerator becomes a real polynomial of degree m_i. In the case $m_i = 1$ (simple complex roots), we obtain a linear numerator, and a second degree function at denominator. It results that if $p > q$, the $ARMA(p,q)$ is a sum of $AR(j)$ with $1 \le j \le m_i$ for real x_i, and a time series similar to $ARMA(2 \cdot m_i, m_i)$ for complex conjugate roots with the multiplicity m_i.All the above parts of X_t have the same white noise a_t, except multiplying by a constant. If $p = q$ we add to the above decomposition the term $\frac{\theta_p}{\varphi_p} a_t$, and if $p < q$ we add the term $\tilde{\theta}(L) a_t$, i.e. a polynomial of degree $q - p$ in lag L applied to the same white noise a_t.

If we consider the reverse in (1"), we decompose analogously a_t in terms of X_t. For forecasting we can forecast each term in the decomposition of X_t.In the above decomposition of X_t the fact that the roots of $\varphi(L)$ are in absolute value grater than one is used only for stationarity, not for decomposition. For instance, if the time series is $ARIMA(p,d,q)$ we use instead of $\varphi(L)$ $(1-L)^d \varphi(L)$. If we group the unit root and the roots of $\varphi(L)$, we obtain a decomposition in $ARIMA(0,j,0)$ with $j = \overline{1,d}$ and an $ARMA(p,q)$ time series. If we perform also the seasonal differentiation $\left(1 - L^s\right)^{d_s}$, we group also the complex roots of equation $L^s = 1$.

In the case of stationarizing using the removing trend by moving average method, we have to find the roots of $\sum_{j=0}^{2\cdot q} L^j - (2 \cdot q + 1) L^q = 0$, corresponding to the differences $\hat{m}_{t-q} - X_{t-q}$, i.e the reminding stationary time series after removing the moving average of order $2 \cdot q + 1$, with opposite sign. Using two times the scheme of Horner, we obtain $L = 1$ of multiplicity 2. The other roots are the roots of polynomial

$$\sum_{j=0}^{q-1}\frac{(j+1)(j+2)}{2}\left(L^j+L^{2\cdot q-j}\right)+\frac{q(q+1)}{2}L^q. \quad (3)$$

By multiplying $\sum_{j=0}^{2\cdot q}L^j-(2\cdot q+1)L^q=0$ by L-1, we can prove that the only multiple root is one (multiplicity is two), and we have no other root on the unit circle. Between the other $2\cdot q-2$ roots we can prove that we have at most two real roots. From the theory of symmetric polynomials, it results that mainly the other $2\cdot q-2$ roots are clustered in groups of four: L_j, $\overline{L_j}$, $\frac{1}{L_j}$ and $\frac{1}{\overline{L_j}}$. The two real roots appear if four does not divide $2\cdot q-2$, hence for even values of q. For odd values of q, these solutions are all simple and conjugated complex in the above groups of four. If we use a moving average with $q=1$, the roots are $L_1=L_2=1$. If $q=2$, the other two roots are the roots of second degree equation $L^2+3L+1=0$, having the roots $-\alpha_1^2$ and $-\alpha_2^2$, where $\alpha_j=\frac{1\pm\sqrt{5}}{2}$, from Fibonacci stream. The roots of polynomial involving moving average of order $2\cdot q+1$ with even and odd q are presented in Tables 6 and 7, Appendix A.

The following structure of solutions has not been proved, but it was checked for $q=4,6,\ldots,20$, $q=100$, $q=500$ and $q=1000$, and for $q=3,5,..,19$, $q=99$, $q=499$ and $q=999$. For even values of q the real negative roots make a circular crown with the radius the absolute values (the other roots have the absolute values between the two radius). The minimum absolute value (that of the real root ≥ -1) increases from 0.38197 for q=2 to 0.806351 for q=20, 0.94207 for q=100, 0.98493 for q=500 and 0.99174 for q=1000. For the minimum argument of complex roots expressed in degrees, the value $\frac{360}{\arg\min\cdot q}$ decreases from 1.08145 for q=4 to 1.0223052 for q=20, 1.0048508 for q=100, 1.0009924 for q=500, and 1.0004979 for q=1000. For odd values of q we have not real roots (all roots are complex in above groups of four). But the minimum absolute value is also increasing on q: from 0.47568 for q=3 to 0.79966 for q=19, 0.94161 for q=99, 0.9849 for q=499, and 0.99174 for q=999. The expression $\frac{180}{\arg\min\cdot q}$ decreases from 1.102567 for q=3 to 1.023379 for q=19, 1.0048986 for q=99, 1.0009944 for q=499, and 1.0004984 for q=999.

In the case of exponential smooth we multiply $\varphi(L)$ by $\frac{1-L}{1-\alpha L}$, where α is the ratio of decreasing the weights of exponential smooth. If α is the inverse of a root of $\varphi(L)$, we divide φ by $1-\alpha L$, otherwise we multiply θ by $1-\alpha L$. Of course, in both cases we multiply φ by $1-L$.

The effective decomposition of X_t is made starting from the moment t just before the first computed a_t. For instance, in an AR(p) model first t is p. We decompose this first X_t in $X_t^{(1)},\ldots,X_t^{(p)}$, and $b_t^{(j)}$ are the drifted white noises from the initially one multiplied by the constants from fractions decomposition. It results a linear regression with the coefficients $X_t^{(1)},\ldots,X_t^{(p)}$. The white noise starts in decomposition of X_t by multiplied by the constants, and b_t is decomposed in MA(1) like white noises ($X_t^{(j)}$ is revertible, but not necessary stationary).

3. APPLICATION

Consider the CPI (Consumer Prices Index) from Buletinul Institutului National de Statistică [6] expressed in percentage of current month related to previous, in the period January 1991 - February 2017.

We want to express the time series X_t as in ARIMA model, and next to decompose the time series X_t and the white noise a_t. First we notice that, using the Dickey - Fuller unit root test [2] that the time series is not stationary, but the difference $\Delta X_t = X_t - X_{t-1}$ is. In the case of AR (p) and MA(q)with $p,q = \overline{0,5}$, not both zero, the representations of X_t are presented in Table 1, that follows.

Table 1 – Representations of X_t for AR (p) and MA(q) time series

p|q	AR(p)	MA(q)
1	-0.38675 X_{t-1}+a_t	a_t-0.38675 a_{t-1}
2	-0.48401 X_{t-1} -0.25149 X_{t-2}+a_t	a_t-0.48401 a_{t-1}-0.0643 a_{t-2}
3	-0.52059 X_{t-1}-0.32189 X_{t-2}-0.14546 X_{t-3}+a_t	a_t-0.52059 a_{t-1}-0.06992 a_{t-2}+0.0125 a_{t-3}
4	-0.53818 X_{t-1} -0.36081 X_{t-2} -0.20841 X_{t-3} -0.12091 X_{t-4}+a_t	a_t-0.53818 a_{t-1}-0.08064 a_{t-2}+0.00386 a_{t-3}-0.02384 a_{t-4}
5	-0.55532 X_{t-1} -0.39035 X_{t-2} -0.25955 X_{t-3} -0.19719 X_{t-4} -0.14174 X_{t-5}+a_t	a_t-0.55532 a_{t-1}-0.09149 a_{t-2}-0.01155 a_{t-3}-0.04642 a_{t-4} -0.04043 a_{t-5}

In the AR(p) case we obtain the following results for $p = \overline{1,5}$.

Table 2 – Decomposition of X_t for ARIMA(p,1,0) time series

p	Simple fractions for AR(p)	Simple fractions for X_t
1		$\frac{0.72111}{1-L}+\frac{0.27819}{1+0.38675L}$
2	$\frac{0.5+0.27549i}{1+(0.24201-0.43923i)L}+\frac{0.5-0.27549i}{1+(0.24201+0.43923i)L}$	$\frac{0.5762}{1-L}$ $+\frac{0.21899-0.48207i}{1+(0.24201-0.43923i)L}+\frac{0.21899+0.48207i}{1+(0.24201+0.43923i)L}$
3	$\frac{0.44899}{1+0.48059L}$ $+\frac{0.2755+0.26718i}{1+(0.02-0.54979i)L}+\frac{0.2755-0.26718i}{1+(0.02+0.54979i)L}$	$\frac{0.5762}{1-L}$ $+\frac{0.21899-0.48207i}{1+(0.24201-0.43923i)L}+\frac{0.21899+0.48207i}{1+(0.24201+0.43923i)L}$
4	$\frac{0.15478+0.19809i}{1-(0.18476+0.57486i)L}+\frac{0.15478-0.19809i}{1-(0.18476-0.57486i)L}$ $+\frac{0.34525+0.07649i}{1+(0.45385-0.3544i)L}+\frac{0.34525-0.07649i}{1+(0.45385+0.3544i)L}$	$\frac{0.44877}{1-L}$ $+\frac{0.1424-0.0536i}{1-(0.18476+0.57486i)L}+\frac{0.1424+0.0536i}{1-(0.18476-0.57486i)L}$ $+\frac{0.13321-0.02782i}{1+(0.45385-0.3544i)L}+\frac{0.13321+0.02782i}{1+(0.45385+0.3544i)L}$
5	$\frac{0.28006}{1+0.64862L}$ $+\frac{0.10794+0.13626i}{1-(0.37203+0.58036i)L}+\frac{0.10794-0.13626i}{1-(0.37203-0.58036i)L}$ $+\frac{0.25203-0.11077i}{1+(0.32538+0.59495i)L}+\frac{0.25203+0.11077i}{1+(0.32538-0.59495i)L}$	$\frac{0.3943}{1-L}+\frac{0.10759}{1+0.64862L}$ $+\frac{0.12175-0.06692i}{1-(0.37203+0.58036i)L}+\frac{0.12175+0.06692i}{1-(0.37203-0.58036i)L}$ $+\frac{0.12731+0.02947i}{1+(0.32538+0.59495i)L}+\frac{0.12731-0.02947i}{1+(0.32538-0.59495i)L}$

In the above table, for instance in the AR(3) model X_t is decomposed in three AR(1) time series with the polynomial $\varphi_1(L)=1+0.48059L$, $\varphi_2(L)=1+(0.02-0.54979i)L$ and $\varphi_3(L)=1+(0.02+0.54979i)L$, and the white noises the white noise a_t of X_t multiplied by 0.44899, 0.2755+0.26178 i, respectively 0.44899, 0.2755-0.26178 i.

If we consider the non-zero expectation case, the above white noise a_t is substituted by the drifted noise $b_t = a_t + \varphi(1)\cdot m$, where $\varphi(L) = 1 + 0.52059L + 0.32189L^2 + 0.14546L^3$, according Table 1, hence φ(1)=1.98794. Because m=-0.04757 it results that the drift is -0.09457, hence we subtract from a_t the value 0.09457. Using this b_t we obtain the same three components for initial time series, but b_t is multiplied by other coefficients: 0.14574, 0.17561-0.0486 i, and 0.17561+0.0486 i. In addition, corresponding to the root L=1 in the ARIMA case, we have an ARIMA(0,1,0) component Y_t such that the difference is b_t multiplied by 0.50303. The decompositions of initial time series X_t and of the drifted noise b_t for ARIMA(0,1,q) are presented in the following table.

Table 3 – Decomposition of X_t a_t for ARIMA(0,1,q) time series

q	Simple fractions for X_t	Simple fractions for a_t
1	$0.38675 + \frac{0.61325}{1-L}$	$2.58565 - \frac{1.58565}{1-0.38675L}$
2	$0.54831 + 0.0643L + \frac{0.45169}{1-L}$	$-\frac{0.5812}{1-0.59253L} + \frac{1.5812}{1+0.10852L}$
3	$0.57801 + 0.05742L - 0.0125L^2 + \frac{0.42199}{1-L}$	$-\frac{0.6125}{1-0.60223L} + \frac{0.9556}{1+0.19056L} + \frac{0.657}{1-0.10892L}$
4	$0.6388 + 0.10062L + 0.01998L^2 + 0.02384L^3 + \frac{0.3612}{1-L}$	$-\frac{0.2872}{1-0.7105L} + \frac{0.5433}{1+0.33985L} + \frac{0.3719+0.2236i}{1-(0.08377+0.30284i)L} + \frac{0.3719-0.2236i}{1-(0.08377-0.30284i)L}$
5	$0.74521 + 0.18989L + 0.0984L^2 + 0.08685L^3 + 0.04043L^4 + \frac{0.25479}{1-L}$	$-\frac{0.1038}{1-0.84007L} + \frac{0.2381+0.196i}{1-(0.21154+0.47124i)L} + \frac{0.2381-0.196i}{1-(0.21154-0.47124i)L} + \frac{0.3138+0.1094i}{1+(0.35392-0.23478i)L} + \frac{0.3138-0.1094i}{1+(0.35392+0.23478i)L}$

For instance, the decomposition of ARIMA(0,1,3) is X_t=0.57801 b_t+0.05742 b_{t-1}+0.0125 b_{t-2}+Y_t, where Y_t is an ARIMA(0,1,0) time series with difference equal to 0.42199*b_t.

In the following we consider the model ARIMA(p,1,q), where the size of the ARMA model, the value of p+q, is constant. The values of $\varphi(L)$ for $p = 4,3,2,1$ are $1 - 0.05275L - 0.07812L^2 + 0.04559L^3 + 0.11309L^4$, $1 - 0.0486L - 0.48217L^2 - 0.16583L^3$, $1 + 0.55701L - 0.10944L^2$, respectively $1 + 0.92335L$. The corresponding values of $\theta(L)$ are $1 - 0.53855L$, $1 - 0.54042L - 0.41813L^2$, $1 + 0.07329L - 0.38463L^2 - 0.10279L^3$, and $1 + 0.61752L - 0.59303L^2 - 0.21334L^3 - 0.14483L^4$.

Consider now $p + q = 5$ with $1 \le p \le 4$. The results for decomposition of the ARMA(p,q) time series Y_t and of the initial ARIMA(p,1,q) time series X_t are presented in the following table.

Table 4 – Decomposition of X_t and Y_t for ARIMA(p,1,q) time series with p+q=5

p	Simple fractions for Y_t	Simple fractions for X_t
4	$\frac{0.38451-0.19825i}{1+(0.42136+0.35245i)L}+\frac{0.38451+0.19825i}{1+(0.42136-0.35245i)L}+$ $\frac{0.1155-0.15024i}{1-(0.44774-0.41748i)L}+\frac{0.1155+0.15024i}{1-(0.44774+0.41748i)L}$	$\frac{0.44896}{1-L}+$ $\frac{0.16224-0.00366i}{1+(0.42136+0.35245i)L}+\frac{0.16224+0.00366i}{1+(0.42136-0.35245i)L}+$ $\frac{0.11328+0.12348i}{1-(0.44774-0.41748i)L}+\frac{0.11328-0.12348i}{1-(0.44774+0.41748i)L}$
3	$\frac{0.54943+0.0823i}{1+(0.61666+0.29215i)L}+\frac{0.54943-0.0823i}{1+(0.61666-0.29215i)L}-$ $\frac{0.09886}{1-0.84791L}$	$\frac{0.13662}{1-L}+\frac{0.15613+0.0767i}{1+(0.61666+0.29215i)L}+$ $\frac{0.15613-0.0767i}{1+(0.61666-0.29215i)L}+\frac{0.55113}{1-0.84791L}$
2	$-\frac{7.64178}{1-0.15394L}+\frac{0.34688}{1+0.71095L}+8.2949+0.93924L$	$\frac{0.40473}{1-L}+\frac{1.39037}{1-0.15394L}+\frac{0.14414}{1+0.71095L}-0.93924$
1	$-\frac{0.2926}{1+0.92335L}+1.2926-0.576L-$ $0.06118L^2-0.15685L^3$	$\frac{0.34644}{1-L}-\frac{0.14047}{1+0.92335L}+0.79403+$ $0.21803L+0.15685L^2$

The corresponding decompositions of the white noise in the ARMA and ARIMA cases are presented in Table 6, that follows.

Table 5 – Decomposition of a_t for ARIMA(p,1,q) time series with p+q=5

p	Simple fractions for ARMA(p,q)	Simple fractions for X_t
4	$\frac{1.47457}{1-0.53855L}+0.42182+0.39645L+$ $0.25558L^2+0.11309L^3$	$-\frac{1.94412}{1-0.53855L}-1.58556-0.28694L-$ $0.14087L^2-0.14249L^3-0.11309L^4$
3	$\frac{0.18107}{1+0.43061L}+\frac{0.17836}{1-0.97103L}+$ $0.64057+0.3966L$	$\frac{0.60158}{1+0.43061L}-\frac{0.00532}{1-0.97103L}+0.40374-$ $0.24397L-0.3966L^2$
2	$\frac{0.117357+2.82097i}{1+(0.38772+0.02902i)L}+$ $\frac{0.117357-2.82097i}{1+(0.38772-0.02902i)L}+\frac{0.65287}{1-0.69415L}$	$-1.0647+\frac{1.17618+10.09684i}{1+(0.38772+0.02902i)L}+$ $\frac{1.17618-10.09684i}{1+(0.38772-0.02902i)L}-\frac{0.28767}{1-0.69415L}$
1	$\frac{0.20482+0.09338i}{1+(0.15477+0.38077i)L}+$ $\frac{0.20482-0.09338i}{1+(0.15477-0.38077i)L}+$ $\frac{0.6087}{1-0.78463L}-\frac{0.01835}{1+0.89667L}$	$\frac{0.60294-0.28271i}{1+(0.15477+0.38077i)L}+$ $\frac{0.60294+0.28271i}{1+(0.15477-0.38077i)L}-$ $\frac{0.16708}{1-0.78463L}-\frac{0.03881}{1+0.89667L}$

CONCLUSIONS

In [2,3,4] the decomposition of a time series in seasonal component, trend and stationary has been performed using for instance moving average. Analogously, if we use the differentiation and/ or seasonal differentiation we can group the root one and the complex unit root for seasonal differentiation. Other decompositions are performed due to economic reasons, as the decomposition of GDP in [1,5]. An open problem is if the economic decomposition can be naturally performed by grouping this paper decomposition of time series.

We have said "similar to ARMA(2*m,m)" instead of ARMA(2*m,m) in Section 2, because the roots of numerator are not necessary in absolute value greater than one. For instance, in the case of AR(5), if we add the corresponding AR(1) components $\frac{0.2755+0.26718i}{1+(0.02-0.54979i)L}$ and the conjugate, we obtain $\frac{0.21588-0.23847L}{(1+(0.02-0.54979i)L)(1+(0.02+0.54979i)L)}$, which has obviously roots for denominator greater than one in absolute value, but the numerator has the root L=0.90523! For MA(q) with $q=\overline{2,5}$ the quote of degree q-1 has in all four cases in Table 3 roots greater than one in absolute value. An open problem is if this is a rule, or it happens in our example and other ones.

REFERENCES

[1] L.-L. Albu, E. Pelinescu and C. Scutaru, *Modele si prognoze pe termen scurt. Aplicaţii pentru România*, Ed. Expert, Bucharest, 2003.

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

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

[4] Th. Popescu, *Serii de timp. Aplicatii în analiza sistemelor*, Ed. Tehnica, Bucharest, 2000.

[5] Steindel, Ch., "Chain weighting: the new approach to measuring GDP", *Current Issues in Economics and Finance*, 1(9), 1995, pp. 1-6.

[6] "Buletinul Institutului National de Statistică", www.insse.ro.

APPENDIX A. ROOTS FOR MOVING AVERAGE

Table 6 – Roots for even values of q

q	Real root ≥ -1	The other absolute values ≤ 1	The other angles in degrees
2	-0.38197		
4	-0.52031	0.55242	83.22129
6	-0.60296	0.65776; 0.61432	66.52096; 56.51432
8	-0.65882	0.72426; 0.68255; 0.66423	85.88092; 50,73123; 42.88701
10	-0.69947	0.76948; 0.73145; 0.71223; 0.70249	76.67192; 69.20316; 41.01439; 34.58262
12	-0.73058	0.8021; 0.76772; 0.74918; 0.73821; 0.73244	87.02103; 63.84469; 57.98187; 34.42691; 28.98308
14	-0.75264	0.82669; 0.79555; 0.77812; 0.76717; 0.76031; 0.75649	80.1066; 74.88039; 55.00913; 49.90537; 29.66504; 29.94881
16	-0.77539	0.84588; 0.8175; 0.80126; 0.79065; 0.78351; 0.77887; 0.77624	87.66116; 70.36931; 65.72646; 48.32439; 43.80988; 26.06156; 21.90285
18	-0.79215	0.86126; 0.83525; 0.82012; 0.81001; 0.80291; 0.79796; 0.79466; 0.79277	82.33238; 78.11198; 62.74497; 58.57341; 43.08966; 39.04454; 23.23933; 19.52106
20	-0.80635	0.87389; 0.84988; 0.83578; 0.82619; 0.81929; 0.81426; 0.81066; 0.80822; 0.80621	88.07264; 74.2845; 70.44565; 56.61257; 52.82802; 38.87899; 35.21603; 20.96897; 17.60727

Table 7 – Roots for odd values of q

q	The absolute values ≤ 1	The angles in degrees
3	0.47568	54.41846
5	0.61161; 0.57041	78.82098; 33.58896
7	0.69447; 0.65137; 0.63517	73.19168; 57.55775; 24.36856
9	0.70893; 0.68982; 0.68159; 0.74885	84.13503; 54.44701; 45.35669; 19.14011
11	0.78702; 0.75086; 0.73188; 0.72136; 0.71657	78.89304; 69.4227; 47.30958; 37.43245; 15.76617
13	0.8152; 0.7825; 0.7645; 0.7535; 0.74701; 0.74396	86.06313;67.06126; 59.0979; 40.22503; 31.86881; 13.40661
15	0.83685; 0.80714; 0.7903; 0.7795; 0.77244; 0.76812	81.68513; 74.92242; 58.32909; 51.45025; 34.99151; 27.74666; 11.66294
17	0.85397; 0.82683; 0.81115; 0.80078; 0.79364; 0.7888; 0.78577; 0.78431	87.0486; 72.27458; 66.33853; 51.6154; 45.55699; 30.96595; 24.56959; 10.32147
19	0.86786; 0.8439; 0.8283; 0.81844; 0.81442; 0.80643; 0.80294; 0.80073; 0.79966	83.3456; 78.10145; 64.81504; 59.52108; 46.29108; 40.87608; 27.7727; 22.04581; 9.25726

MATHEMATICAL ASPECTS REGARDING THE BASIS OF DECISION MAKING IN MILITARY ACTIONS

Traian ANASTASIEI[*], Gabriel RĂDUCANU[**], Radu-Mihai DINCĂ[*]

[*]"Henri Coandă" Air Force Academy, Braşov, Romania (a_traian03@yahoo.com, radu_mihai1989@yahoo.com)
[**]Romanian Air Force Staff, Bucharest, Romania (gabiraducanu@yahoo.ca)

DOI: 10.19062/1842-9238.2017.15.2.5

Abstract: *Modern military actions cannot be conceived in the absence of reasonable scientific approach. The selection of the best course of action (CoA) for the achievement of targeted objectives, while taking into account the available resources and the configuration of the internal / external background, is an extremely complex activity that is carried out under conditions of uncertainty and time constraints. Optimization of military actions as well as the implementation of efficient decision-making solutions is a matter of prime importance in the current battlefield configuration and it implies a rigorous mathematical apparatus. In this article, we analyze the importance of detecting/adjusting mathematical methods in accordance with the new technologies and armaments used in contemporary confrontations, using a particular case in the field of aviation.*

Keywords: *decision-making, aviation, military actions, mathematical methods*

1. INTRODUCTION

The military phenomenon is one of the social phenomena that best illustrates the idea of development - the forces and the whole complex of Tactics, Techniques, Procedures (TTPs) are subjected to dramatic changes in the context of the fundamental processes of social evolution. The causes of transformations are multiple and diverse, corresponding to the nature of society itself, and they can be found in any of its components.

Decryption of changes and trends occurring at geopolitical and geostrategic level and the awareness of possible developments in the field of military actions, especially those which are conducted in the airspace, constitute actions which are meant to provide inputs for developing a robust, resilient and versatile air force able to act within a fluid and unpredictable security environment. Nowadays military actions take place in an increasingly complex operation area which usually involves a lot of entities (armed forces, different factions, local population, NGOs, civilian agencies, economic agents etc.), many of them having divergent interests. In order to succeed, decision-makers need to adopt a comprehensive approach and they must act quickly and "surgically". There is a very short time for analysis, evaluation and decision-making and there is no place for error. [4, 5, 6]

Modern warfare, regardless of its form, type or magnitude, cannot be conceived without the active presence of the air force (aviation, radars, artillery / SAMs etc.) capable of executing rapid, surprising and destructive strikes, of generating strong and long-lasting effects on the enemy's ability to wage war. It can also disrupt enemy's actions by stealth, maneuverability and precise actions.

Another element which is strongly connected with the air force is represented by the advanced technologies. Air power can be considered one of the finest products of technological revolution initiated at the dawn of the 20th century. The intrinsic connection with technology allows air forces to exploit the benefits of outer space. The actions taken in a three-dimensional space provide air component with greater freedom of action than the other components (land and maritime).

Military actions are, by definition, planned, organized and conducted by military structures. Decision is the essence of military action management. In the complex and contradictory world in which we live, adopting and implementing high-performance decisions is increasingly difficult and at the same time necessary. [2] The greater our world's complexity and contradictions are, the more effective our decisions should be. Any mission aims to achieve a goal set by the higher echelon or resulting from the commander's plan.

2. ASPECTS REGARDING THE EFFECTIVENESS OF ACTIONS IN AVIATION

Generally speaking, the effectiveness of our actions obviously depends on the resources used and the quality of our decisions. In the broadest sense, this means that our ability to produce a certain effect on others' behavior (when we refer to human action, the effect is anticipated to become a *goal*) is based upon our ability to plan our future actions, to allot the needed resources and to act in accordance with our plan but also by taking into account the changes that occur in reality. We should take into account some restrictions or limitations: available resources, TTPs to use, time etc. Consequently, sometimes a normative and not a descriptive approach is needed, so formal methods and models are needed in order to set up your actions on scientific bases.

Most of the combat actions are designed to produce a desired level of attrition in the enemy's forces and logistics or to destroy them. Usually, many targets are located inside the enemy's territory (beyond the *Forward Edge of Battle Area - FEBA*) and this fact increases the uncertainty level regarding the final effects on target.

A course of action (CoA) which involves air strikes upon a major objective located in the depth of the enemy's territory implies the fact that the aircraft(s) will cross areas which pose different levels of hostility, as follows:

- the flight within territory controlled by own troops or allies ("the friendly area"), an area which has a very low probability of danger, basically 0 %;
- the flight within territory controlled by enemy forces ("the hostile area"), an area which comprises many Surface Based Air Defence (SBAD) threats, so the probability of having losses is getting higher;
- the flight within the district where the target is located ("the dangerous area"), an area which is strongly defended by SBAD systems, so this area has a high level of risk, and the probability of getting losses cannot be neglected (we should also notice that this phase includes preparatory maneuvers and the air strike itself).

If we analyze the losses suffered by our own forces in every phase of the mission we shall conclude that it is unprofitable, from the resource perspective, to get the aircraft to the target and not to use the load on board. [1] Friendly air forces may suffer certain losses caused by the enemy's reaction, which is carried out either with its own air forces or by GBAD means (SAMs or anti-aircraft artillery - AAA). Considering this, we can deduce a probability of reaching the target which is represented by the following relations:

$$Q = Q_{SAM} * Q_{AAA} * Q_{AF} \quad (1)$$

where:
Q - is the probability that the a/c will encounter the target and will launch the strike against it;
Q_{SAM} - is the probability that the a/c will get into the district where the target is located taking into account the SAMs reaction;
Q_{AAA} - is the probability that the a/c will get into the district where the target is located taking into account the antiaircraft artillery reaction;
Q_{AF} - is the probability that the a/c will get into the district where the target is located taking into account the enemy air forces reaction.

In order to establish which variables are relevant in the decision-making process, the air force staff must establish certain criteria for evaluating the strikes' efficiency. Based on these criteria the staff proceeds to a cost/benefit analysis or a similar method and then it will design the best CoAs. The commander will select the one which best fits his intention. Speaking of criteria, the most used criteria are:

- the amount of damage produced to the target (percentage and quantitative);
- the amount of resources (number of aircraft, manpower, fuel etc.) necessary to fulfill the mission.

Papers that deal with this topic provide methods of calculation for some of these criteria. [3]

a) The average percentage of losses incurred by the enemy when acting against it with more than one aircraft.

a1) When the attack is carried out by aiming the constituent elements of the group of targets, the average percentage of enemy's losses, denoted by *W**, is calculated by the following equation:

$$W^* = 1 - e^{-\frac{N}{N_t} * W * N_{at} * Q} \quad (2)$$

where:
N - the number of aircraft performing the attack on the group of targets;
N_t - the number of targets (elementary objectives) within the group of targets;
W - the probability that an aircraft destroys, throughout a single attack, an element within the group of targets. This variable depends on the target's features (dimensions, degree of vulnerability) and the features of the ammunition used;
N_{at} - The number of attacks carried out by each aircraft;
Q - The probability that the aircraft will reach the target district in order to execute the attack. It depends on the enemy's retaliation means (air force, artillery/anti-aircraft missiles etc.). It is determined by the (1) equation.

a2) When the attack is conducted against a group of targets considered as a whole, so that the elements located in the vulnerable zone of the group of targets are considered to be destroyed, while the others are not, the average percentage of losses, denoted by M_N, is calculated by the relation:

$$M_N = 1 - (1 - M)^N \quad (3)$$

where:
N - The number of aircraft performing the attack on the group of targets;
M - The average percentage of damage when the group of targets is attacked by a single aircraft;

$$M = M_x * M_y \quad (4)$$

where:
M_x, M_y - mean values of the coverage percentage determined by area of destruction (first-order first moment), calculated as follows:

$$M_x = \frac{E_x}{T_x}\left[x_1 * \Phi(x_1) - x_2 * \Phi(x_2) + \frac{1}{\rho\sqrt{\pi}} * (e^{-\rho^2 * x_1^2} - e^{-\rho^2 * x_2^2})\right];$$

$$M_y = \frac{E_y}{T_y}\left[y_1 * \Phi(y_1) - x_2 * \Phi(y_2) + \frac{1}{\rho\sqrt{\pi}} * (e^{-\rho^2 * y_1^2} - e^{-\rho^2 * y_2^2})\right];$$

with:

$x_1 = \frac{T_x + L_x}{2 * E_x}$;

$x_2 = \frac{T_x - L_x}{2 * E_x}$;

$y_1 = \frac{T_y + L_y}{2 * E_y}$;

$y_2 = \frac{T_y + L_y}{2 * E_y}$;

where:

T_x, T_y - the dimensions of objective (target), approximated by a rectangle;

L_x, L_y - the dimensions of the area of destruction, approximated by a rectangle;

E_x, E_y - The probable deviations (probable error) of the center of the area of destruction relative to the center of the objective (target);

$\Phi(x)$ - the Laplace transformed function;

ρ - 0,476936.

b) The required number of aircraft in order to achieve an average predetermined percentage of losses for a group of targets, denoted by N, is calculated as follows:

b1) When the attack is carried out by aiming the constituent elements of the group of targets, we use the equation:

$$N = -\frac{N_t * \ln(1 - P_1)}{N_{at} * W * Q} \quad (5)$$

where:

P_1 - The predetermined average percentage of the damage caused to the objective (target);

N_t - The number of targets (elementary objectives) within the group of targets;

W - the probability that an aircraft destroys, on a single attack, an element within the group of targets;

N_{at} - The number of attacks carried out by each aircraft;

Q - As it was shown above.

b2) When the attack is conducted against a group of targets considered as a whole, we use:

$$N = \frac{\ln(1 - P_1)}{\ln(1 - M)} \quad (6)$$

where:

P_1 - The predetermined average percentage of the damage caused to the objective (target);

M - The average percentage of damages when the group of targets is attacked by a single aircraft.

3. AVIATION IN COMPLEX ACTIONS

Modern aircraft are characterized by advanced fighting capabilities, including fire power, stealth technologies, radar and electronic warfare (EW) equipment and above all, the technical and tactical performances which allows it to carry out complex actions in a complex environment. In order to determine some efficiency indicators, we will use a hypothetical situation:

3.1 Problem formulation

In a district (hereinafter referred to as **objective**) there are several constituent elements (hereinafter referred to as **targets**). The target is attacked by several aircraft, each of which is equipped with several types of weaponry (combat cargo).

For tactical (objective and target characteristics, attack procedure) and technical (launch / firing parameters) reasons, the battle load will not be launched all at once, the aircraft repeating the attack. It is required to determine **the main efficiency indicators** (average damage caused to the enemy, the number of aircraft required to carry out the mission).

3.2 Working hypotheses

In order to compute the calculations, we set out some simplifying hypotheses that will not influence the final result:

1. We consider that the targets are evenly distributed within the objective, which means that in a Z area (representing z % of S - the total area of the objective) there will be a number of T targets (representing z % of NT - total number of targets within the objective), so T is proportional to Z.

2. We consider that the attack will be carried out by means of 2 types of weapons / battle load:

- Type 1 produces effects on a large surface (all targets in the "vulnerable" area are destroyed);
- Type 2 produces point effects (only the directly hit target is destroyed).

From the probabilistic point of view we define:

- E_i - the event consisting of hitting and damaging the target, i.e. the destruction of the appropriate targets, using the type of weapon / battle load;
- $P(E_i)$ - the probability of producing E_i event.

3. We solve the problem using the above mentioned equations:

- for Type 1 weapons - equations (2) and (5);
- for Type 2 weapons - equations (3), (4) and (6).

We also add the following points:

- each aircraft performs successive attacks, each attack is carried out with a single type of weaponry;
- during the first attacks, type 1 weapons are used, but the following ones are carried out with type 2 weapons;
- events E_i are compatible, meaning that targets are destroyed whether they are hit by Type 1 weapons or by Type 2 weapons.

3.3 Solution

Whether or not the results of the attack are known immediately after the use of a specific type of weapons determines two situations that we will analyze in turn.

Case I - The results of the attack cannot be immediately known (the destroyed targets cannot be identified), so the following attacks can also be directed against targets that have already been hit.

In this case, events E_1 and E_2 are independent, meaning that even if a target was destroyed during the first attack, this does not mean that there will be no further attack on it. Thus,

$$P(E_1, E_2) = P(E_1) + P(E_2) - P(E_1) * P(E_2) \quad (7)$$

where:

P (E_1, E_2) - The probability of destroying the target using both types of weapons.

Considering the above and noting P (E_1, E_2) with P_p, it results:

$$P_p = M_N + W^* - M_N * W^* \quad (8)$$

By substituting and making calculations, based on the above equations, we obtain:

$$P_p = M_N + W^* - M_N * W^* =$$
$$= M_N + W^*(1 - M_N) =$$
$$= 1 - (1 - M)^N + W^*(1 - (1 - (1 - M)^N)) =$$
$$= 1 - (1 - M)^N + W^*(1 - M)^N =$$
$$= 1 - (1 - M)^N * (1 - W^*) =$$
$$= 1 - (1 - M)^N * (1 - (1 - e^{-\frac{N}{N_t} * W * N_{at} * Q})) =$$
$$= 1 - (1 - M)^N * e^{-\frac{N}{N_t} * W * N_{at} * Q} =$$
$$= 1 - ((1 - M) * e^{-\frac{W * N_{at} * Q}{N_t}})^N$$

We thus obtained a relation to determine the average percentage of enemy's losses when two types of armaments (Type 1 and Type 2) are used:

$$P_p = 1 - (1 - M)^N * e^{-\frac{N}{N_t} * W * N_{at} * Q} \quad (9)$$

The **average number of destroyed targets** is then determined by:

$$M_t = N_t * P_p \quad (10)$$

From equation (9) we obtain:

$1 - P_p = ((1 - M) * e^A)^N$,

where $A = -\frac{W * N_{at} * Q}{N_t}$

Then, by applying the logarithm,

$$\ln(1 - P_p) = \ln((1 - M) * e^A)^N =$$
$$= N * \ln((1 - M) * e^A) =$$
$$= N * (\ln(1 - M) + \ln e^A) =$$
$$= N * (\ln(1 - M) + A)$$

Thus, we obtain the N - the required number of aircraft in order to produce P_p average losses to the enemy.

$N = \frac{\ln(1 - p_p)}{\ln(1-M) + A}$, or

$$N = \frac{\ln(1 - p_p)}{\ln(1-M) - \frac{W^* \cdot N_{at} \cdot Q}{N_t}} \quad (11)$$

Case II - The conditions allow us to find out the outcome of the attack immediately, we can identify the destroyed targets, so the following attacks will be directed only on targets that have not yet been destroyed.

Identification of the main efficiency indicators will be done by means of the following algorithm:

- we calculate the average number of targets destroyed as a result of the attack carried out with Type 1 weapons (denoted by $M^{(1)}$). For the symmetry of the notations, we will note with $N^{(1)}$) - the initial number of targets (previously noted with N_t);

$$M_t^{(1)} = M_N * N_t^{(1)} \quad (12)$$

- we calculate the mean number of targets that have not been destroyed after the attack was carried out with Type 1 weapons (denoted by $N^{(2)}$). For these targets there will be used the Type 2 weapons.

$$N_t^{(2)} = N_t^{(1)} - M_t^{(1)} =$$

$$= N_t^{(1)} - M_N * N_t^{(1)} =$$

$$= N_t^{(1)} * (1 - M_N) =$$

$$= N_t^{(1)} * (1 - (1-(1 - M)^N)) =$$

$$= N_t^{(1)} * (1-M)^N$$

$$N_t^{(2)} = N_t^{(1)} * (1 - M)^N \quad (13)$$

- we calculate the average number of targets that were not destroyed after the attack carried out with Type 2 weapons (denoted by $M^{(2)}$);

$$M_t^{(2)} = N_t^{(2)} * W^* \quad (14)$$

- we calculate the average number of targets destroyed after the final attack (denoted M_t);

$$M_t = M_t^{(1)} + M_t^{(2)} \quad (15)$$

Using relations (12), (14) and (15), it results:

$$M_t = M_N * N_t^{(1)} + N_t^{(2)} * W^* =$$

$$= M_N * N_t^{(1)} + N_t^{(1)} * (1 - M)^N * W^* =$$

$$= N_t^{(1)} * (1 - (1 - M)^N + (1 - M)^N * W^*) =$$

$$= N_t^{(1)} * (1 - (1 - M)^N * (1 - W^*)) =$$

$$= N_t^{(1)} * (1 - (1 - M)^N * (1 - (1 - e^{-\frac{N}{N_t^{(2)}} * W * N_{at} * Q}))) =$$

$$= N_t^{(1)} * (1 - (1 - M)^N * e^{-\frac{N}{N_t^{(2)}} * W * N_{at} * Q})$$

$$M_t = N_t^{(1)} * (1 - (1 - M)^N * e^{-\frac{N}{N_t^{(1)} * (1-M)^N} * W * N_{at} * Q}) \quad (16)$$

We divide both members of this last relation by $N_t^{(1)}$). But M_t / $N_t^{(1)}$) = P_p. We obtain successively:

$$P_p = 1 - (1 - M)^N * e^{-\frac{N}{N_t^{(2)}} * W * N_{at} * Q}$$

$$1 - P_p = (1 - M)^N * e^{-\frac{N}{N_t^{(2)}} * W * N_{at} * Q}$$

$$\ln(1 - P_p) = \ln((1 - M)^N * e^{-\frac{N}{N_t^{(2)}} * W * N_{at} * Q})$$

$$\ln(1 - P_p) = \ln(1 - M)^N + \ln e^{-\frac{N}{N_t^{(2)}} * W * N_{at} * Q}$$

$$\ln(1 - P_p) = N * \ln(1 - M) - \frac{N}{N_t^{(2)}} * W * N_{at} * Q$$

Dividing by N and replacing $N_t^{(2)}$ with $N_t^{(1)} * (1 - M)^N$(equation (13), we obtain:

$$\frac{\ln(1 - P_p)}{N} - \ln(1 - M) = \frac{-W * N_{at} * Q}{N_t^{(1)} * (1-M)^N} \quad (17)$$

Since this equation is more difficult to solve, we will form 2 functions of variable N:

$$f_1(N) = \frac{\ln(1 - P_p)}{N} - \ln(1 - M) \quad (18)$$

$$f_2(N) = \frac{-W * N_{at} * Q}{N_t^{(1)} * (1 - M)^N} \quad (19)$$

Finding N can be done either by the graphical method, at the intersection of the graphs of the two functions *f1(N)* and *f2(N)*, or by numerical methods, using the computer. We used both methods and the results are highlighted in *Figure 1*. For the numerical determination of *N*, we considered 4) case from *Table 1* and we used the *Maple 13* software. This software also helped us to provide the graphs for the functions.

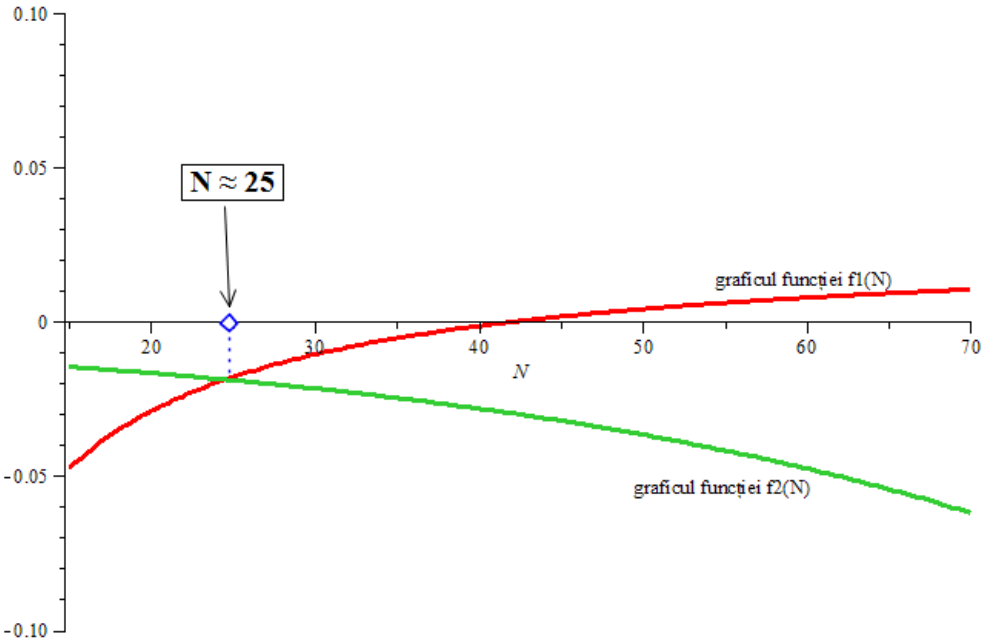


FIG. 1. Finding N by graphic method

Example (simplified, without affecting the final results)

There are 100 tanks in a district. We have to determine the required number of aircraft in order to destroy 2/3 (67) of them using 2 types of armament.

The results are shown in Table 1 (Pay attention to the order of magnitude!):

Table 1. The required number of aircrafts in certain conditions

The type of weapons used	**The required number of aircrafts**
1) Type I weapons (ammunition)	42
2) Type II weapons (ammunition)	112
3) Type I and II weapons (ammunition), without "fire transport"	31
4) Type I and II weapons (ammunition), with " fire transport "	25

CONCLUSIONS

The evolution of the air force is strongly related to the technological and scientific development as well as to the evolution of the regional and global security environment. Accordingly, the future of air force can only be designed in the context of the European and Euro-Atlantic security system. In such a system there are many interdependencies which are often difficult to quantify. It is the duty of military specialists to study, analyze and propose measures to decision-makers to effectively employ the air force in future operations.

Military actions are dynamic, and each mission is different and unrepeatable. In turn, advanced technologies radically change the means of military action. All of these require that decision makers use appropriate methods (simulation, modeling, conceptualization etc.) in order to base the decisions they make on a scientific foundation.

With regard to establishing the optimal need for aviation forces, we have found that the existing manuals / regulations contain methodologies that do not solve all the possible situations.

In this paper we have obtained a system of relations that allows us to find out the solution to the formulated problem, as follows:

- the percentage of average losses produced to the objective is given by relation (9);
- the average number of destroyed targets is given in **Case I** (without the immediate identification of damaged targets) by the relation (10), and in **Case II** (only the previously undamaged targets are attacked) by relation (16);
- the number of aircraft required to produce certain predetermined losses to the objective is presented in **Case I** by relation (11), and in **Case II** by relation (17).

REFERENCES

[1] Traian Anastasiei, *Aviaţia – alonja acţiunilor militare*, Editura Academiei Forţelor Aeriene „Henri Coandă", Braşov, 2002, pp. 170-173;
[2] Ovidiu Nicolaescu, *Sistemul decizional al organizaţiei*, Editura Economică, Bucureşti, 1998, p. 17;
[3] Valerian Nestorescu, *Calculul probabilităților în domeniul militar*, Editura Militară, București, 1963, pp. 140-214;
[4] ***Joint Air Power Competence Centre, *Present Paradox – Future Challenge*, March 2014, available at: https://www.japcc.org/portfolio/present-paradox-future-challenge/, accessed on 12 May 2017;
[5] ***Joint Air Power Competence Centre, *Air & Space Power in NATO – Future Vector Part I*, July 2014, available at: https://www.japcc.org/portfolio/air-space-power-in-nato-future-vector-part-i/, accessed on 12 May 2017;
[6] ***Joint Air Power Competence Centre, *Air & Space Power in NATO – Future Vector Part II*, October 2014, available at: https://www.japcc.org/portfolio/air-space-power-in-nato-future-vector-part-ii/, accessed on 12 May 2017.

SOLVING THE TRANSPORTATION PROBLEM WITH PIECEWISE-LINEAR CONCAVE COST FUNCTIONS

Tatiana PAȘA, Valeriu UNGUREANU

Universitatea de Stat, Chișinău, Moldova (pasa.tatiana@yahoo.com, v.a.ungureanu@gmail.com)

DOI: 10.19062/1842-9238.2017.15.2.6

Abstract: *In this work we expose the transportation problem on a network with piecewise-linear concave cost functions on edge flows. Some properties of the problem are highlighted. A series of results are presented, which concern the implementation of a method exposed in [1] in Wolfram Mathematica System. These results are compared with the results obtained by applying built-in Wolfram Language functions on a family of test problems. Tests are provided both on PC-s and a cluster.*

Keywords: *network transportation problem, optimal solution, flow in a network, concave function.*

1. INTRODUCTION

Transport have a very important role in the socio-economic development of a country because it facilitates the transportation of passengers and goods to the point of destination.

Air transportation is a civil or military aviation branch which deals with the transport of goods and passengers. This type of transport is preferred by those who have the primary purpose of moving as quickly as possible from one point to another, this being its essential feature, but also the precise schedule for any time of the year regardless of day or night. For this purpose both mixed-duty aircrafts for passengers and goods, as well as cargo aircrafts, are used. Airplanes can be used on a regular basis - regular traffic, or irregular traffic, which means contracts between airlines and different recipients who want to operate the rented airplanes for a specified period. The most important airports based on the number of passengers and quantity of goods are in London, Tokyo, New York, Los Angeles and Frankfurt.

If we are talking about the economic aspect of air transport, everything has a cost, therefore, for a quantity of cargo or for a certain number of passengers transported we have a price that varies if the quantity of the cargo or the number of passengers changes. This is why the price per passenger or per ton of product may vary. In this case, we can say that the cost of the transport can be described by a concave nonlinear function, i.e. the cost will depend nonlinearly on the number of passengers or on the weight of the cargo being transported.

The air transport network consists of paths and nodes where the connections between the nodes are made by means of air transport such as airplanes and helicopters.

The network that starts from a node (airport) and allows the transport of passengers and / or cargo to another node, but can also use other airports as transit nodes, may be formalized by graph theory means as a mathematical model of a standard single-source transport network shipping with one destination and intermediate points. A single-source transport network with several destinations describes the real situation when goods and passengers are transported to several nodes using nodes of transit.

2. PRELIMINARIES

First, let us expose notions that a necessary for studying the transportation problem with concave cost functions.

Transportation network: A transportation network [2] is an oriented graph $G = (V, E)$, without loops, which satisfies the following properties:

1. There is a vertex (source) $v_0 \in V$ which has only outgoing edges;

2. There is a vertex (destination, sink) $v_t \in V$ which has only incoming edges;

3. For each arc it is associated a value $c(e)$, for any $e \in E$, named capacity of the arc.

The cases with several sources and / or destinations can be easily modelled. To reduce such models to the standard transportation network we have to add a super-source connected to each of the original sources. Its capacity will be the total of the product amount from the original sources. We can also add a super-sink connected to the original destinations, its capacity will be equal to the total capacity of all destinations.

Generally, we can set for each edge a value that describes the maximal size or quantity of the goods which may be transported, distances between nodes, the time to cover the distance or the price to transport the cargo. In this paper, a piecewise function is assigned to each edge describing the transport costs and depending nonlinearly on the quantity of the goods being transported.

Flow: A flow in a network is a function $f: E \to \mathbb{R}$ which satisfies the following properties:

1. Capacity constraint: For all $e \in E$ the condition $f(e) \leq c(e)$ is satisfied;

2. Skew symmetry: For all $(v_1, v_2) \in V$ the condition $f(v_1, v_2) = -f(v_2, v_1)$ is satisfied;

3. Flow conservation: $\sum_{x \in E^+(v)} x(e) - \sum_{x \in E^-(v)} x(e) = 0$, where $E^+(t)$ is the set of edges that enter $t \in V$ and $E^-(s)$ - that exit $s \in V$.

Capacity constraints limit the quantity that can be transported along an arc. In the context of this paper, the capacities are equal to ∞, i.e. an unlimited quantity can be transported along an arc.

Skew symmetry assumes that for each node the quantity of product entering the node equals the quantity that exits it, i.e. their modulus (absolute values) are equal.

Flow conservation implies that the whole quantity of product that exits from the source reaches the destination, i.e. there are no losses along the arcs.

It is also impossible to have a surplus in the quantity when reaching the destination, because that would imply that the intermediate nodes also produce goods.

Concave function: A function f is concave on an interval, if for all x and y from this interval and for all $\alpha \in [0,1]$ the following is true:

$$f\big((1-\alpha)x + \alpha y\big) \geq (1-\alpha)f(x) + \alpha f(y).$$

In this paper, concave functions describe the cost of the product shipping along the arcs. These functions are non-increasing piecewise-linear functions defined on the interval [0; +∞] that initially describes a rise from 0 to some value which then become a constant value. This means that the transport cost raises with the increase of the product flow only up to a fixed value, then the expenses are the same for any product quantity being transported.

3. PROBLEM FORMULATION

Let us consider the network transportation problem described by the graph $G = (V, E)$, $|V| = n$, $|E| = m$ with the source $v_0 \in V$ and the sink $v_t \in V$. A real bounded function $q: V \to R$ is defined on the finite set of its vertices V and non–decreasing piecewise-linear concave cost functions $\varphi_e(x_e)$, defined for each arc $e \in E$, which depend on flow. We must find a flow x^* for which the function $F(x) = \sum_{e \in E} \varphi_e\big(x(e)\big)$ has the smallest value, i.e. $F(x^*) = \min_{x \in X} F(x)$, where X is the set of flows admissible in G, which means that it satisfies the system:

$$\begin{cases} \sum\limits_{x \in E^+(v)} x(e) - \sum\limits_{x \in E^-(v)} x(e) = q(v), & for\ all\ v \in V \\ x(e) \geq 0, & for\ all\ e \in E \end{cases}$$

where $E^+(v)$ is the set of edges that enter $v \in V$ and $E^-(v)$ - that exit $v \in V$.

1. In the formulated problem for every intermediary node the flow is conserved which means that:

$$q(v) = \begin{cases} -p, & v = v_0, \\ 0, & v \in V \backslash \{v_0, v_t\}, \\ p, & v = v_t, \end{cases}$$

where p is the net capacity.

Therefore, the transportation problem may be formulated as it follows:

$F(x^*) = \min\limits_{x \in X} F(x)$,

where X is defined by the function:

$$\begin{cases} \sum\limits_{x \in E^+(v)} x(e) - \sum\limits_{x \in E^-(v)} x(e) = \begin{cases} -p, & v = v_0 \\ 0, & v \in V \backslash \{v_0, v_t\}, \\ p, & v = v_t \end{cases} & for\ all\ v \in V, \\ x(e) \geq 0, & for\ all\ e \in E. \end{cases}$$

2. A particular case of this problem is the transportation network with a source and several destinations. The function will then be described by the system:

$$q(v) = \begin{cases} -\sum_{u \in V_t} p_u, & v = v_0 \\ 0, & v \in V/V_t \backslash \{v_0\} \\ p_v, & v \in V_t \end{cases}$$

where the set of vertices $V_t \subseteq V$ is the set of destinations. For every $v \in V_t$ it is known the necessity p_v of the point v.

The transportation problem may be formulated as it follows: $F(x^*) = \min_{x \in X} F(x)$, where X is defined by the system:

$$\begin{cases} \sum_{x \in E^+(v)} x(e) - \sum_{x \in E^-(v)} x(e) = q(v) = \begin{cases} -\sum_{u \in V_t} p_u, & v = v_0 \\ 0, & v \in V/V_t \backslash \{v_0\} \\ p_v, & v \in V_t \end{cases}, & for\ all\ v \in V \\ x(e) \geq 0, & for\ all\ e \in E \end{cases}$$

The formulated problem is NP-hard because the function is a sum of piecewise-linear concave functions, which means that there are no efficient (polynomial) algorithms to solve this problem. That's why our aim is to find an algorithm that will find a good solution as efficiently as possible.

4. WAYS OF SOLVING THE PROBLEM

4.1 Applying a known algorithm: In [1] an algorithm for solving the non-linear transport problem is described by reducing the solution to a problem of linear programming for which polynomial algorithms are known. Fig. 1 shows the block diagram describing the algorithm.

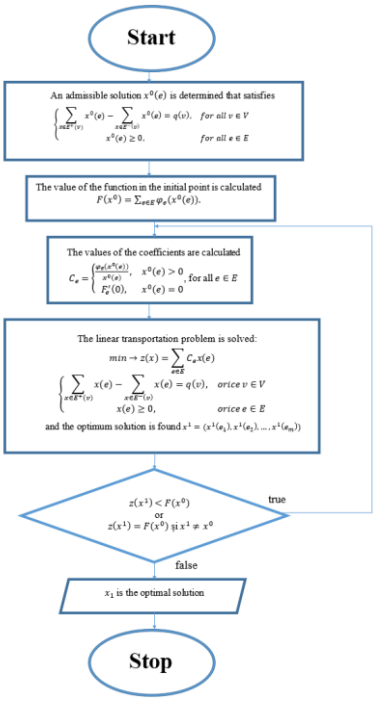


FIG.1. Block diagram for solving the TP

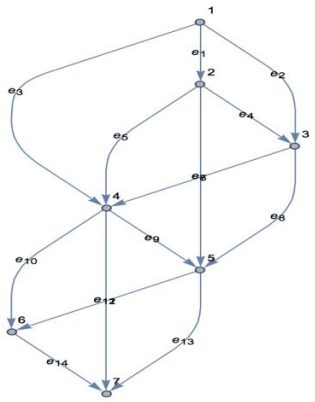
The Mathematica System and the Wolfram Language [3] make it much easier to implement the algorithm comparing with other languages and systems. The code is compact, easy-to-read, it is easy to define new variables and functions.

Next, we will explain how we applied the Wolfram Language to solve the problem based on the following example.

The description of the transport network is easily accomplished using the standard *Graph[]* function to which all pairs of connected vertices describing the direction of the edges are transmitted as parameters, see following Fig. 2 (a):

```
Graph[{Labeled[1 ↔ 2, "e1"], Labeled[1 ↔ 3, "e2"],
  Labeled[1 ↔ 4, "e3"],
  Labeled[2 ↔ 3, "e4"], Labeled[2 ↔ 4, "e5"],
  Labeled[2 ↔ 5, "e6"],
  Labeled[3 ↔ 4, "e7"], Labeled[3 ↔ 5, "e8"],
  Labeled[4 ↔ 5, "e9"], Labeled[4 ↔ 6, "e10"],
  Labeled[4 ↔ 7, "e11"],
  Labeled[5 ↔ 6, "e12"], Labeled[5 ↔ 7, "e13"],
  Labeled[6 ↔ 7, "e14"]}, VertexLabels → "Name"]
```

(a)



(b)

FIG. 2. Description of the transportation network

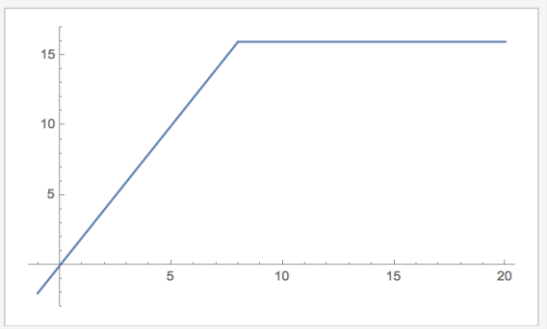
As a result of executing the *Graph[]* function we obtain the network represented in Fig. 2. (b), which is it graphical representation. Knowing the transportation network, the *IncidenceMatrix[]* function allows us to obtain the incidence matrix of the graph that describes the transportation network and permits to construct a system of constraints for the formulated problem.

As mentioned above, with each arc is associated a piecewise-linear concave function that is described generally as in Fig. 3 (a) and (b) using the Wolfram Language.

```
γ = {1, 1, 2, 2, 2, 2, 1, 2, 1, 2, 2};
δ = {1, 1, 2, 2, 2, 2, 1, 2, 1, 2, 2};

f[j_Integer, x_] := Piecewise[{{γ[[j]] x, x ≤ δ[[j]]},
                               {γ[[j]] δ[[j]], x > δ[[j]]}},
                              0, True]
```

(a)



(b)

FIG. 3. Description and graphic representation of the piecewise-linear function

In order to obtain an initial solution with which the program starts according to the block diagram of Fig. 1, we use the *FindInstace[]* standard function which solves the system of equations, provides the non-negative solutions, and as a result submits the set of possible solutions on the basis of which a linear objective function is obtained.

LinearPrograming[] is a standard function that solves the problem of linear programming that consists in minimizing the obtained linear objective function on the same system of constraints. In Fig. 4 we have the code of the implemented algorithm.

```
X1 = X0 = Values@FindInstance[{A.X == b[[All, 1]], X ≥ 0}, X][[1]];
F0 = Sum[f[j, X0[[j]]], {j, 1, m}];
Z1 = -∞;
i = 0;
While[Not[(Z1 > F0) || (Z1 == F0 && X0 == X1)],
 X0 = X1; F0 = Sum[f[j, X0[[j]]], {j, 1, m}];
 c = Table[fd[j, X0[[j]]], {j, 1, m}];
 X1 = LinearProgramming[c, A, b];
 Z1 = Sum[c[[j]] X1[[j]], {j, 1, m}];
 Print[++i];
 Print[Z1];
]
X1
Sum[f[j, X1[[j]]], {j, 1, m}]
```

FIG. 4. The algorithm code in the Wolfram Language

4.2 Using the standard Wolfram language functions: We can use the standard functions *Minimize[]* and *NMinimize[]*, that can find a global optimum, and *FindInstance[]*, for finding a local solution, in order to solve the formulated problem, which is in fact an optimization problem of nonlinear programming that consists in minimizing of an objective function that is described as a sum of concave piecewise-linear concave functions and satisfies the constraints described by a system of linear equations and constraints of non-negativity.

The prototype for the function *Minimize[{f, cons}, {x, y, ...}]* suggests that as input data we have the objective function and the system of linear constraints. It is used to obtain the exact global solution of the optimization problem that uses linear programming methods, the Lagrange multiplier method, integer programming methods and other symbolical and analytical methods, which involves obtaining exact solutions.

The prototype of the function *NMinimize [{f, cons}, {x, y, ...}]* suggests that, as input, we will have the objective function and the constraint system. It is used to obtain a global solution as a numerical value through the use of linear programming methods, Nelder-Mead, random search, i.e. numerical methods, which implies obtaining a result obtained by a series of approximations. We can use the attribute *Method* to select the method, e.g. *DifferentialEvolution*, *RandomSeach*, *SimulatedAnnealing*, *Neldermead*, by which the function *NMinimize[]* used to solve the problem.

The prototype of the function *FindMinimum{{f, cons}, {x, y, ...}]* suggests that as input it have the objective function and the system of constraints. It is used to obtain a local solution, that is a result that starts from a point in the region defined by the constraints. We can use the attribute *Method* to select the method, e.g. *PrincipalAxis*, *InteriorPoint*, *QuasiNewton*, *ConjugateGradient*, by which the function *FindMinimum[]* solves the problem.

Symbolic computation or, in other words, the use of computational algebra that is based on symbols that represent mathematical concepts, operates with polynomials, rational functions, trigonometric functions.

The results of symbolic algorithms are not affected by approximation errors that influence the result in numerical computation that operates with numbers.

Using the standard *Minimize[]* and *NMinimize[]* functions to solve the problem, the user does not know the method that was applied by the Wolfram Language. If we want to solve a series of problems and want to be sure which method was used then the *Method* option followed by the name of the requested method can be used. In Fig. 5 it is given the program that solves the problem using *Minimize[]* without specifying the method to be applied and as a result we have the solution and the value of the function at the point.

```
Minimize[
  {f1[x1] + f2[x2] + f3[x3] + f4[x4] + f5[x5] + f6[x6] +
    f7[x7] + f8[x8] + f9[x9] + f10[x10] + f11[x11] + f12[x12] +
    f13[x13] + f14[x14],
   -x1 - x2 - x3 == -10, (*1*)
   x1 - x4 - x5 - x6 == 0, (*2*)
   x2 + x4 - x7 - x8 == 0, (*3*)
   x3 + x5 + x7 - x9 - x10 - x11 == 0, (*4*)
   x6 + x8 + x9 - x12 - x13 == 0, (*5*)
   x10 + x12 - x14 == 0, (*6*)
   x11 + x13 + x14 == 10, (*7*)
   x1 ≥ 0, x2 ≥ 0, x3 ≥ 0, x4 ≥ 0, x5 ≥ 0, x6 ≥ 0, x7 ≥ 0,
   x8 ≥ 0, x9 ≥ 0, x10 ≥ 0, x11 ≥ 0, x12 ≥ 0, x13 ≥ 0, x14 ≥ 0},
  {x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x11, x12, x13, x14}] //
 AbsoluteTiming

{108.452,
 {4, {x1 → 0, x2 → 10, x3 → 0, x4 → 0, x5 → 0, x6 → 0, x7 → 10,
   x8 → 0, x9 → 10, x10 → 0, x11 → 0, x12 → 0, x13 → 10, x14 → 0}}}
```

FIG. 5. Application of the function *Minimize[]* to solve the TP in the Wolfram Language

In order to be able to control and compare the execution time of the algorithm, the *AbsoluteTiming[]* standard function, that returns the runtime of the program in seconds, is used.

We present a list of networks of different sizes and the computation time for each algorithm in Tab. 1.

Table 1. Computation time of the method (seconds)

Nr. Ord.	Nr. Nodes	Nr. Edges	Algorithm	Minimize	NMinimize	FindMinim
1	4	6	0.000279	0.17	0.2	0.18
2	6	10	0.000360	3.02	0.38	0.12
3	7	14	0.000492	108.45	0.61	0.19
4	8	16	0.000468	345.61	2.09	0.19
5	8	20	0.001135	10386.50	1.96	1.75

A substantial rise in the computation time can be observed when the function *Minimize[]* is used, though this method gives the best solution as can be seen in the diagram below.

FIG. 6. The number of optimal solutions for solving TP using different methods

CONCLUSIONS

We formulated a number of problems of different sizes, modelled on different networks with different number of edges and nodes. All were solved using the algorithm from the block scheme in Fig.1 and then using the standard functions *Minimize[]*, *NMinimize[]*, *FindMinimum[]*. The computation time and results of the programs has been compared for each of the used methods. With these data we can expose the following conclusions:

- The solutions obtained with the proposed algorithm is as good as *Minimize[]* or at least not worse than *NMinimize[]*;
- If we could choose another initial solution, the algorithm may give a better solution for the majority of problems;
- Half of the solutions obtained by applying the proposed algorithm are as good as the solutions from *Minimize[]*, while only 25% of the solutions obtained by *NMinimize[]* are as good as *Minimize[]*;
- The proposed algorithm's computation time is the fastest. It is smaller than that of *Minimize[]*, which needs more time for every new edge. E.g., for a network with 8 nodes and 20 edges it needs 3 hours, while our algorithm only about some milliseconds;
- *FindMinimum[]* returns a solution very fast, but it is rarely the optimal solution;
- If the problem is to get a good solution in a short time, our algorithm meets these conditions;
- This is an heuristic algorithm, it doesn't have a rigorous proof and must be improved to be used on a larger array of problems.

An improvement of this algorithm to give the optimum solution in all the cases depends on the initial solution generation, from which the algorithm starts, and may be realised efficiently through parallel computation. Applying an array of different initial solutions and taking the best one as the initial one, even if the computation time of the algorithm will increase it will increase considerably slower than using the function *Minimize[]*.

REFERENCES

[1] T. Pasha, D. Lozovanu. *An algorithm for solving the transport problem on network with concave cost functions on flow of edges.* Computer Science Journal of Moldova, vol. 10, no 3, Kishinev (2002), pp. 341-347.

[2] T. H. Cormen, C. E. Leiserson, R. L. Rivest. *Introduction to Algorithms*. MIT Press, Cambridge, Massachusetts London, England, 2-nd edition, 2002.

[3] S. Wolfram, *An elementary introduction to the Wolfram Language.* Champaign, IL: Wolfram Media, Inc., 2016.

STOCHASTIC GAMES WITH REAL TRANSITION COSTS AND FINAL SEQUENCE OF STATES

Alexandru LAZARI

Moldova State University, Republic of Moldova (lazarialexandru@mail.md)

DOI: 10.19062/1842-9238.2017.15.2.7

Abstract: *The stochastic games with unit transition time and final sequence of states were previously investigated and an efficient algorithm for determining the game duration was developed. The purpose of this paper is to generalize this problem for the case when the transition costs are real and to elaborate a method for determining the expectation of the game outcome. The elaborated approach is based on extended network method, that allows us to reduce the real transition costs case to unit transition time case.*

Keywords: *stochastic game, final sequence of states, game outcome, extended network.*

1. PROBLEM FORMULATION

Let L_{P} be a stochastic system with finite set of states V_{P}, $|V_{\mathrm{P}}| = \omega_{\mathrm{P}}$. The system L_{P} starts its evolution from the state $v \in V_{\mathrm{P}}$ with probability $p_{\mathrm{P}}^{*}(v)$, $\forall v \in V_{\mathrm{P}}$. Also, at every step t, the system L_{P} passes from the state $u \in V_{\mathrm{P}}$ to the state $v \in V_{\mathrm{P}}$ with probability $p_{\mathrm{P}}(u,v)$ and real transition cost $c_{\mathrm{P}}(u,v)$, that do not depend on t.

We assume that a sequence of states $X_{\mathrm{P}} = (x_1, x_2, \ldots, x_m)$ is given and the system L_{P} stops transitions as soon as the states $x_1, x_2, \ldots, x_m$ are reached consecutively in given order. We denote by C_{P} the evolution outcome of the system L_{P}, i.e. $C_{\mathrm{P}} = \sum_{(u,v)\in E_{\mathrm{P}}} c_{\mathrm{P}}(u,v)$, where E_{P} is the set of transitions that compose the evolution of the given system L_{P}.

The system L_{P} represents a Markov process with set of states V_{P}, initial distribution of the states $p_{\mathrm{P}}^{*} = (p_{\mathrm{P}}^{*}(v))_{v\in V_{\mathrm{P}}}$, probability transition matrix $p_{\mathrm{P}} = (p_{\mathrm{P}}(u,v))_{u,v\in V_{\mathrm{P}}}$ and real transition cost matrix $c_{\mathrm{P}} = (c_{\mathrm{P}}(u,v))_{u,v\in V_{\mathrm{P}}}$. One more thing that is specific to the system L_{P} is the property to have a stopping rule, represented by final sequence of states X_{P}.

Several interpretations of these extended Markov processes were analyzed in [1], [6] and [7]. Also, polynomial algorithms for determining the main probabilistic characteristics of the evolution outcome of the given stochastic system L_{P} were proposed.

Next, the following game Γ_{P}, that represents a generalization of the games defined in [2] and [3], is considered. Initially, each player Π_l, $l = \overline{0, r-1}$, defines his stationary strategy, represented by one transition matrix $p_{\mathrm{P}}^{(l)} = (p_{\mathrm{P}}^{(l)}(u,v))_{u,v\in V_{\mathrm{P}}}$, $l = \overline{0, r-1}$. The initial distribution of the states is $p_{\mathrm{P}}^{*} = (p_{\mathrm{P}}^{*}(v))_{v\in V_{\mathrm{P}}}$. The game is started by first player Π_0.

At every moment of time, the system L_{P} passes consecutively to the next state according to the strategy of the current player. After the last player Π_{r-1}, the first player Π_0 acts on the system evolution and the game continues in this way until the given final sequence of states X_{P} is achieved. The winner is the player who acts the last on the system evolution.

Our goal is to study the game outcome C_{P}, knowing the initial distribution of the states $p_{\mathrm{P}}^{*} = (p_{\mathrm{P}}^{*}(v))_{v \in V_{\mathrm{P}}}$, the real cost transition matrix $c_{\mathrm{P}} = (c_{\mathrm{P}}(u,v))_{u,v \in V_{\mathrm{P}}}$, the stationary strategy $p_{\mathrm{P}}^{(l)} = (p_{\mathrm{P}}^{(l)}(u,v))_{u,v \in V_{\mathrm{P}}}$ of each player Π_l, $l = \overline{0, r-1}$ and the final sequence of states X_{P} of the system L_{P}. We will show how this problem can be reduced to the unit transition costs case using extended network method and how to find the win probabilities of the players.

2. TRANSITION COSTS DISCRETIZATION

2.1 Reduction to the positive real transition costs case. First of all, our goal is to reduce the game network to the positive real transition costs case, for avoiding the problems that can be encountered when we deals with negative and zero numbers.

Let us consider a new game Γ_{P^+}, defined on stochastic system L_{P^+} with the same final sequence of states $X_{\mathrm{P}^+} = X_{\mathrm{P}}$, finite set of states $V_{\mathrm{P}^+} = V_{\mathrm{P}}$, initial distribution of the states $p_{\mathrm{P}^+}^{*} = p_{\mathrm{P}}^{*}$ and stationary strategy $p_{\mathrm{P}^+}^{(l)} = p_{\mathrm{P}}^{(l)}$ of each player Π_l, $l = \overline{0, r-1}$. The difference between the games Γ_{P^+} and Γ_{P} consists in the way how the transition costs are defined.

Let $c_{\mathrm{P}}^{\min} = \min\limits_{u,v \in V_{\mathrm{P}}} c_{\mathrm{P}}(u,v)$ be the minimal transition cost in the game Γ_{P}. We define the transition costs in the new game Γ_{P^+} in the following way:

$$c_{\mathrm{P}^+}(u,v) = c_{\mathrm{P}}(u,v) - c_{\mathrm{P}}^{\min} + 1 \geq 1, \ \forall u,v \in V_{\mathrm{P}^+}. \tag{1}$$

Next, we consider a new game Γ_1, defined on stochastic system L_1 with the same final sequence of states $X_1 = X_{\mathrm{P}}$, finite set of states $V_1 = V_{\mathrm{P}}$, initial distribution of the states $p_1^{*} = p_{\mathrm{P}}^{*}$ and stationary strategy $p_1^{(l)} = p_{\mathrm{P}}^{(l)}$ of each player Π_l, $l = \overline{0, r-1}$. The transition costs in the new game Γ_1 are all equal to 1.

It is easy to observe that the following relation holds:

$$C_{\mathrm{P}} = C_{\mathrm{P}^+} + (c_{\mathrm{P}}^{\min} - 1) \cdot C_1, \tag{2}$$

where C_1 is the outcome of the game Γ_1, C_{P^+} is the outcome of the game Γ_{P^+} and C_{P} is the outcome of the game Γ_{P}. Indeed, we have

$$\begin{aligned} C_{\mathrm{P}^+} &= \sum_{(u,v) \in E_{\mathrm{P}^+}} c_{\mathrm{P}^+}(u,v) = \sum_{(u,v) \in E_{\mathrm{P}}} \left(c_{\mathrm{P}}(u,v) - c_{\mathrm{P}}^{\min} + 1 \right) = \sum_{(u,v) \in E_{\mathrm{P}}} c_{\mathrm{P}}(u,v) - \sum_{(u,v) \in E_{\mathrm{P}}} \left(c_{\mathrm{P}}^{\min} - 1 \right) = \\ &= \sum_{(u,v) \in E_{\mathrm{P}}} c_{\mathrm{P}}(u,v) - (c_{\mathrm{P}}^{\min} - 1) \sum_{(u,v) \in E_{\mathrm{P}}} 1 = C_{\mathrm{P}} - (c_{\mathrm{P}}^{\min} - 1) \cdot C_1. \end{aligned}$$

2.2 Reduction to the positive rational transition costs case. At this step, our goal is to reduce the game network to the case when the transition costs of related stochastic system are positive rational values.

Similar to the approach elaborated in [4], we will approximate the real cost of each transition with a rational number, such that the accuracy for the expectation of the game outcome to be less than a given value $\varepsilon > 0$.

We consider the game Γ_{Θ^+}, defined on stochastic system L_{Θ^+} with the same final sequence of states $X_{\Theta^+} = X_{\mathrm{P}^+}$, finite set of states $V_{\Theta^+} = V_{\mathrm{P}^+}$, initial distribution of the states $p^*_{\Theta^+} = p^*_{\mathrm{P}^+}$ and stationary strategy $p^{(l)}_{\Theta^+} = p^{(l)}_{\mathrm{P}^+}$ of each player Π_l, $l = \overline{0, r-1}$. We define the transition costs in the new game Γ_{Θ^+} in the following way:

$$c_{\Theta^+}(u,v) = c_{\mathrm{P}^+}(u,v) + \varepsilon(u,v),\ \forall u,v \in V_{\Theta^+}, \tag{3}$$

choosing the nonnegative value $\varepsilon(u,v)$ such that $c_{\Theta^+}(u,v) \in \Theta^+$ for each $u,v \in V_{\Theta^+}$ and

$$\left| \mathrm{E}\left(C_{\Theta^+}\right) - \mathrm{E}\left(C_{\mathrm{P}^+}\right) \right| \le \varepsilon, \tag{4}$$

where the accuracy $\varepsilon > 0$ is given.

Consider that

$$0 \le \varepsilon(u,v) \le \varepsilon^* < 1,\ \forall u,v \in V_{\Theta^+}. \tag{5}$$

We obtain

$$\begin{aligned} C_{\Theta^+} &= \sum_{(u,v)\in E_{\Theta^+}} c_{\Theta^+}(u,v) = \sum_{(u,v)\in E_{\mathrm{P}^+}} \left(c_{\mathrm{P}^+}(u,v) + \varepsilon(u,v)\right) = \sum_{(u,v)\in E_{\mathrm{P}^+}} c_{\mathrm{P}^+}(u,v) + \sum_{(u,v)\in E_{\mathrm{P}^+}} \varepsilon(u,v) = \\ &= \sum_{(u,v)\in E_{\mathrm{P}^+}} c_{\mathrm{P}^+}(u,v) + \varepsilon(u,v) \cdot \sum_{(u,v)\in E_{\mathrm{P}^+}} 1 = C_{\mathrm{P}^+} + \varepsilon(u,v) \cdot C_1, \end{aligned}$$

that implies

$$\left| \mathrm{E}\left(C_{\Theta^+}\right) - \mathrm{E}\left(C_{\mathrm{P}^+}\right) \right| = \left| \mathrm{E}\left(C_{\Theta^+} - C_{\mathrm{P}^+}\right) \right| = \mathrm{E}(\varepsilon(u,v) \cdot C_1) = \varepsilon(u,v) \cdot \mathrm{E}(C_1) \le \varepsilon^* \mathrm{E}(C_1). \tag{6}$$

If we have $\varepsilon^* E(C_1) \le \varepsilon$, then also the inequality $\left| \mathrm{E}\left(C_{\Theta^+}\right) - \mathrm{E}\left(C_{\mathrm{P}^+}\right) \right| \le \varepsilon$ holds.

Consider an arbitrary value $\varepsilon^* \in \left(0, \min\left\{1, \varepsilon / \mathrm{E}(C_1)\right\}\right]$. If the natural number s satisfies the condition $\varepsilon^* > 10^{-s}$, then we can chose

$$\varepsilon(u,v) = 10^{-s}\left(1 - \{10^s c_{\mathrm{P}^+}(u,v)\}\right),\ \forall u,v \in V_{\Theta^+} \tag{7}$$

and these values verify the relations (3), (4) and (5), where by $\{a\}$ is denoted the fractional part of the arbitrary real number a. From relations (3) and (7) we obtain

$$c_{\Theta^+}(u,v) = 10^{-s}\left(1 + \left\lfloor 10^s c_{\mathrm{P}^+}(u,v) \right\rfloor\right),\ \forall u,v \in V_{\Theta^+}, \tag{8}$$

where by $\lfloor a \rfloor$ is denoted the integer part of the arbitrary real number a. So, in these conditions, we can consider $E\left(C_{\mathrm{P}^+}\right) \approx E\left(C_{\Theta^+}\right)$ with error of approximation ε.

2.3 Reduction to the positive integer transition costs case. At this step, our goal is to reduce the game network to the case when the transition costs of related system are positive integer values. We will do this transformation by changing the measurement unit, i.e. we will multiply each positive rational transition cost $c_{\Theta^+}(u,v)$, $\forall u,v \in V_{\Theta^+}$, by μ, where μ is the least common multiple of the denominators of these rational transition costs.

Let us consider a new game Γ_{N^*}, defined on stochastic system L_{N^*} with the same final sequence of states $X_{N^*} = X_{\Theta^+}$, finite set of states $V_{N^*} = V_{\Theta^+}$, initial distribution of the states $p^*_{N^*} = p^*_{\Theta^+}$ and stationary strategy $p^{(l)}_{N^*} = p^{(l)}_{\Theta^+}$ of each player Π_l, $l = \overline{0, r-1}$. We define the transition costs in the new game Γ_{N^*} in the following way:

$$c_{N^*}(u,v) = \mu \cdot c_{\Theta^+}(u,v),\ \forall u,v \in V_{N^*}. \tag{9}$$

In this way, we will have $c_{N^*}(u,v) \in N^*$, $\forall u,v \in V_{N^*}$. We obtain

$$C_{N^*} = \sum_{(u,v)\in E_{N^*}} c_{N^*}(u,v) = \sum_{(u,v)\in E_{N^*}} \left(\mu \cdot c_{\Theta^+}(u,v)\right) = \mu \cdot \sum_{(u,v)\in E_{\Theta^+}} c_{\Theta^+}(u,v) = \mu \cdot C_{\Theta^+}, \tag{10}$$

that implies $C_{\Theta^+} = \frac{1}{\mu} \cdot C_{N^*}$.

2.4 Players' states synchronization. Next step prepares the transition costs for extended network method application. It is necessary that, the application of extended network method do not change the player who acts in the state v at each transition (u, v). From this reason, we will ensure that, after this transformation, each positive integer transition cost is congruent to 1 modulo r, where r is the number of the players.

Let us consider a new game Γ_f, defined on stochastic system L_f with the same final sequence of states $X_f = X_{N^*}$, finite set of states $V_f = V_{N^*}$, initial distribution of the states $p^*_f = p^*_{N^*}$ and stationary strategy $p^{(l)}_f = p^{(l)}_{N^*}$ of each player Π_l, $l = \overline{0, r-1}$. We define the transition costs in the new game Γ_f in the following way:

$$c_f(u,v) = r \cdot c_{N^*}(u,v) + 1,\ \forall u,v \in V_f, \tag{11}$$

i.e. we apply the function $f : N^* \to N^*$, $f(t) = rt + 1$, to each positive integer transition cost.

In this way, we will have $c_f(u,v) \in N^*$ and $c_f(u,v) \equiv 1 (\mathrm{mod}\ r)$, $\forall u,v \in V_f$. Also,

$$C_f = \sum_{(u,v)\in E_f} c_f(u,v) = \sum_{(u,v)\in E_f} \left(r \cdot c_{N^*}(u,v) + 1\right) = r \cdot \sum_{(u,v)\in E_{N^*}} c_{N^*}(u,v) + C_1 = r \cdot C_{N^*} + C_1, \tag{12}$$

that implies $C_{N^*} = \frac{1}{r} \cdot \left(C_f - C_1\right)$.

3. EXTENDED NETWORK METHOD

3.1 Reduction to the unit transition costs case. At this step, our goal is to reduce the game network to the case when the transition costs of related stochastic system are equal to 1. Similar with approach presented in [5], we will do this transformation by applying extended network method, i.e. each edge (u,v) with positive integer transition cost $c_f(u,v)$ will be replaced with $c_f(u,v)$ edges with unit transition cost, $\forall u,v \in V_f$. In this way, the total transition cost from arbitrary state $u \in V_f$ to another state $v \in V_f$ will not be changed, it is necessary only to ensure that also the transition probability from the state u to the state v will remain unchanged after extended network method application.

Let us consider a new game Γ_{ext}, defined on stochastic system L_{ext} in the following way. For each state $v \in V_f$ we choose the state $u \in V_f$ such that $c_f(u,v) = \max\limits_{z \in V_f^-(v)} c_f(z,v)$, where $V_f^-(v) = \left\{ z \in V_f \;\middle|\; \sum\limits_{l=0}^{r-1} p_f^{(l)}(z,v) > 0 \right\}$. We split the transition (u,v) in $c_f(u,v)$ unit transitions using the new states $z_0(u,v) = u;\ z_1(u,v);\ z_2(u,v);\ \ldots;\ z_{c_f(u,v)-1}(u,v);\ z_{c_f(u,v)}(u,v) = v$. If $c_f(u,v) > 1$, then the stationary strategies $p_f^{(l)}$, $l = \overline{0, r-1}$, are updated in the following way:

1. new rows and columns, filled with zeros, that correspond to intermediate states $z_j(u,v)$, $j = \overline{1, c_f(u,v)-1}$, are added;
2. the element $p_f^{(l)}(z_j(u,v), z_{j+1}(u,v))$ is set to 1, $j = \overline{1, c_f(u,v)-1}$;
3. the element $p_f^{(l)}(z, z_{c_f(u,v)-c_f(z,v)+1}(u,v))$ is set to $p_f^{(l)}(z,v)$ and, after that, the element $p_f^{(l)}(z,v)$ is set to 0, for each $z \in V_f^-(v)$ that satisfies the inequality $c_f(z,v) > 1$.

We consider $p_f^*(z_j(u,v)) = 0$, $j = \overline{1, c_f(u,v)-1}$, because the game cannot be started from one intermediate state. The final sequence of states becomes

$$\begin{aligned} X_{ext} = (&x_1,\ z_1(x_1,x_2),\ z_2(x_1,x_2),\ldots, z_{c_f(x_1,x_2)-1}(x_1,x_2), x_2,\\ &z_1(x_2,x_3), z_2(x_2,x_3),\ldots, z_{c_f(x_2,x_3)-1}(x_2,x_3), x_3,\\ &\ldots,\\ &z_1(x_{m-1},x_m), z_2(x_{m-1},x_m),\ldots, z_{c_f(x_{m-1},x_m)-1}(x_{m-1},x_m), x_m). \end{aligned}$$

So, if we put all these things together, we obtain that the new game Γ_{ext} is defined on stochastic system L_{ext}, with a finite set of states V_{ext}, final sequence of states X_{ext}, an initial distribution of the states p_{ext}^* and stationary strategy $p_{ext}^{(l)}$ of each player Π_l, $l = \overline{0, r-1}$, defined in the way described above. Also, each transition cost in this stochastic system is equal to 1. This implies that the outcome C_{ext} of the game Γ_{ext} is

$$C_{ext} = \sum_{(u,v) \in E_{ext}} c_{ext}(u,v) = \sum_{(u,v) \in E_{ext}} 1 = |E_{ext}| = \sum_{(u,v) \in E_f} c_f(u,v) = C_f, \tag{13}$$

that implies $C_f = C_{ext}$.

3.2 Expectation of game outcome. Next, we will summarize the results obtained above. We have the following formula for calculating the outcome C_{P} of the game Γ_{P}:

$$\begin{aligned} C_{\mathrm{P}} &= C_{\mathrm{P}^+} + (c_{\mathrm{P}}^{\min} - 1)\cdot C_1 \approx C_{\Theta^+} + (c_{\mathrm{P}}^{\min} - 1)\cdot C_1 = \frac{1}{\mu}\cdot C_{\mathrm{N}^*} + (c_{\mathrm{P}}^{\min} - 1)\cdot C_1 = \\ &= \frac{C_f - C_1}{r\mu} + (c_{\mathrm{P}}^{\min} - 1)\cdot C_1 = \frac{C_{ext} - C_1}{r\mu} + (c_{\mathrm{P}}^{\min} - 1)\cdot C_1 = \frac{1}{r\mu}\cdot C_{ext} + \left(c_{\mathrm{P}}^{\min} - 1 - \frac{1}{r\mu} \right)\cdot C_1. \end{aligned}$$

So, the expectation of the outcome C_{P} of the game Γ_{P} is

$$\mathrm{E}(C_{\mathrm{P}}) \approx \frac{1}{r\mu} \cdot \mathrm{E}(C_{ext}) + \left(c_{\mathrm{P}}^{\min} - 1 - \frac{1}{r\mu} \right) \cdot \mathrm{E}(C_1). \tag{14}$$

The values $\mathrm{E}(C_{ext})$ and $\mathrm{E}(C_1)$ can be determined using the algorithm developed in [2], because the both games Γ_{ext} and Γ_1 are defined on stochastic systems with unit transition costs (which, in particular case, can be considered as unit transition time).

3.3 Win probabilities of the players. From the definition of the win probabilities of the players, it is easy to observe that they depend only on the network structure and the transition probabilities of the stochastic system on which the game is defined. Indeed, if we consider that each transition time in the game Γ_{P} is equal with 1, and T_{P} is the duration of the game Γ_{P}, then the win probability $w_{\mathrm{P}}^{(l)}$ of the player Π_l, $l = \overline{0, r-1}$, is given by the following relation:

$$w_{\mathrm{P}}^{(l)} = P(T_{\mathrm{P}} \equiv l \pmod r) = \sum_{k=0}^{\infty} P(T_{\mathrm{P}} = k + l),\ l = \overline{0, r-1}. \tag{15}$$

So, the win probabilities of the players do not depend on the transition costs. The win probabilities of the players in the game Γ_{P} are the same as the win probabilities of the players in the game Γ_1, that can be found using the algorithm from [2].

CONCLUSIONS

The purpose of this paper was to generalize the problem of determining the expectation of the game outcome for the case when the transition costs are real. The elaborated approach is based on extended network method, that allows us to reduce the real transition costs case to unit transition time case.

This reduced case was investigated in [2] and algorithms for determining the expectation of the game duration (outcome) and win probabilities of the players were developed. So, solving the problem for the unit transition time case and performing the transformations presented in this paper, we obtain the solution of the problem in the general case, when the transition costs are real.

REFERENCES

[1] A. Lazari, D. Lozovanu and M. Capcelea, *Dynamical deterministic and stochastic systems: evolution, optimization and discrete optimal control (in romanian)*, Chisinau, CEP USM, 2015;

[2] A. Lazari, Stochastic games on Markov processes with final sequence of states, *Bul. Acad. Ştiinţe Repub. Mold. Mat.*, no. 1 (83), pp. 77-94, 2017;

[3] A. Lazari, Determining the distribution of the duration of stationary games for zero-order Markov processes with final sequence of states, *Bul. Acad. Ştiinţe Repub. Mold. Mat.*, no. 3 (79), pp. 72-78, 2015;

[4] A. Lazari, Rationalization of discrete Markov processes with final sequence of states and real transition costs, *Proceedings of the international conference "Mathematical modelling, optimization and informational technologies"*, Chisinau, Evrica, pp. 120-126, 2014;

[5] A. Lazari, Extended network method for determining the main probabilistic characteristics of the evolution for stochastic discrete systems with positive rational transitions, *Scientific abstracts of International Conference of Young Researchers, VIII edition*, Chisinau, pp. 93, 2010;

[6] G. Zbaganu, Waiting for an ape to type a poem, *Bul. Acad. Ştiinţe Repub. Mold. Mat.*, no. 2 (8), pp. 66-74, 1992;

[7] L. Guibas and A. Odlyzko, Periods in strings, *J. Combin. Theory Ser. A*, no. 30, pp. 19-42, 1981.

QUALITATIVE ASPECTS OF THE MIN PARETO BINOMIAL DISTRIBUTION

Bogdan Gheorghe MUNTEANU

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

DOI: 10.19062/1842-9238.2017.15.2.8

Abstract: *Centered upon statistical models relating to qualitative aspects, the following paper sets out to demonstrate that by means of the Akaike information criterion (AIC) a statistical selection of the MinParB distribution for different parameter values can be obtained based on the statistical simulation algorithm of the power series distribution, called the Min Pareto Binomial [4], and the EM algorithm for the statistical estimation of the parameters of the MinParB distribution. The determining of the MinParB distribution from a unitary perspective regarding the class of the power series distributions [2] has also been taken into consideration.*

Keywords: *power series distributions, Pareto distribution, Binomial distribution, distribution of the minimum, EM algorithm, information criterion.*

1. INTRODUCTION

According to the Pareto principle, also known as the "80/20 rule", in the case of events, about 80% of the effects is generated by 20% of the causes. Management consultant Joseph M. Juran was the first to suggest this principle, which he named after the Italian economist Vilfredo Pareto, who identified the well-known 80/20 ratio. Basically, Pareto demonstrated that about 80% of the land in Italy was owned by 20% of the population. In business, the same basic rule applies (for example, 80% of sales come from 20% of clients) [1]. Similarly, for a given set of parameters, in the case of natural phenomena, the existence of an empirically obtained Pareto distribution has been observed [2].

The Pareto distribution is particularly used in situations in which there is a high probability of paying large sums in compensation, namely liability insurance.

Let $(X_i)_{i\geq 1}$ be a sequence of a number of independent and identically distributed random variables, $X_i \sim Par(\mu,\alpha),\ \mu,\alpha>0$ with the cumulative distribution function $F_{X_i}(x)=F_{Par}(x)=1-\left(\frac{\mu}{x}\right)^{\alpha},\ x\geq\mu$ and the probability density function $f_{X_i}(x)=f_{Par}(x)=\frac{\alpha\mu^{\alpha}}{x^{\alpha+1}},\ x\geq\mu$.

Also, we denote by $V_{Par}=\min\{X_1,X_2,...,X_Z\}$, where random variable $Z\sim Binom(n,p),\ n\in\{1,2,...\},\ p\in(0,1)$, and $(X_i)_{i\geq 1}$ are independent and Pareto distributed random variables. The cumulative distribution function, the probability density function and some reliability characteristics of the random variable V_{Par} are given in the paper [4].

The random variable V_{Par} generates two events:

- The event to minimize the amounts claimed $(X_i)_{i=\overline{1,Z}}$ regarding the civil liability insurance;
- The number of the claimed amounts Z represents the number of successes out of the n independent events with the probability of success p.

Therefore, we discuss the distribution $MinParB(\mu, \alpha, n, p)$, $\mu, \alpha > 0$, $p \in (0,1)$, $n \in \{1,2,...\}$. The numerical characteristics of this distribution are presented in the paper [4].

2. INFORMATION CRITERION

The common approach to model selection involves choosing a model that minimizes one or several information criteria applied to a set of statistical models [1],[5].

The commonly used information criteria are: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Hannan-Quinn Information Criterion (HQIC) and consistent Akaike information criteria (CAIC).

Each criterion is a sum of two terms: the first term characterizes the entropy rate or model prediction error, whereas the second one describes the number of the free parameters estimated based on the model [2].

2.1. Akaike Information Criterion. The Akaike Information Criterion (AIC) is a criterion for selecting from nested (overlapping) economic models. Basically, AIC is a measure estimating the quality of each studied economic model, since they relate to one another for a given set of data. Therefore, AIC is an ideal method for selecting the model.

Discovered and put forward by Professor Hirotugu Akaike in 1971, respectively in 1974, the AIC was defined as a measure of matching the statistical model.

AIC is an associated number for each separate model, as follows:

$$AIC = -2L(x, \hat{\Theta}) + 2q, \tag{1}$$

where $L(x, \hat{\Theta})$ represents the maximum likelihood function, $\hat{\Theta} = (\hat{\alpha}, \hat{p})$ the parameter vector estimated by applying the EM algorithm [4], and q represents the number of parameters of the statistical model. In our case, $q = 2$.

Therefore, for the set of AIC values corresponding to each particular economic model, the preferred one in terms of relative quality is the model with the minimum value. ($AIC_{\min}$).

The loss of information when the statistical model that has been studied and analysed in relation to the best estimated model is given by:

$$\Delta_i = AIC_{\min} - AIC_i, \tag{2}$$

where i is the number of statistical models to which AIC has been applied, and $AIC_{\min}$ stands for the minimal value AIC out of the values' vector.

2.2. Bayesian Information Criterion. Bayesian information criterion (BIC) is also a mathematical tool applied to statistical models from the economic field. It is a criterion similar to AIC. The BIC or the Schwartz criterion (1978) is a number characterized by the relation:

$$BIC = -2L(x,\hat{\Theta}) + q\ln(m), \tag{3}$$

where $L(x,\Theta)$ represents the maximum likelihood function, $\hat{\Theta} = (\hat{\alpha}, \hat{p})$ is the parameter vector estimated as a result of applying the EM algorithm [4], q represents the number of parameters of the statistical model ($q = 2$), and m characterizes the volume of the statistical data.

2.3. Hannan-Quinn Information Criterion. The Hannan-Quinn Information Criterion (HQIC) is an information criterion that is used alternatively with AIC and BIC. The criterion is represented by the number:

$$HQIC = -2L(x,\hat{\Theta}) + 2q\ln(\ln(m)), \tag{4}$$

where $L(x,\hat{\Theta})$, q and m have the same interpretations as the AIC and BIC criteria.

2.4. Consistent Akaike Information Criteria. Consistent Akaike Information Criteria (CAIC) is, essentially, a correction to the Akaike Information Criterion (AIC), this being characterized by the relation:

$$CAIC = AIC + \frac{2q(q-1)}{m-q-1} \tag{5}$$

or

$$CAIC = -2L(x,\hat{\Theta}) + \frac{2qm}{m-q-1}, \tag{6}$$

where $L(x,\hat{\Theta})$, q and m have the same interpretations as the AIC and BIC criteria.

3. APPLICATIONS

According to the studies in paper [4], the logarithm maximum likelihood function is defined as follows:

$$\ln L(x,\hat{\Theta}) = m\left(\ln n + \ln \hat{p} + \ln \hat{\alpha} + \hat{\alpha}\ln \mu\right) - m\ln\left[1-(1-\hat{p})^n\right] + \sum_{j=1}^{m}\left\{(n-1)\ln\left[1-\hat{p}+\hat{p}\left(\frac{\mu}{x_j}\right)^{\hat{\alpha}}\right] - (\hat{\alpha}+1)\ln x_j\right\}, \tag{4}$$

where $\hat{\Theta} = (\hat{\alpha}, \hat{p})$ the parameter vector estimated as a result of applying the EM algorithm.

The step-by-step description of the algorithm is included in paper [4] and implemented in the GUI Octave 1.5.4 programming environment.

The values of the estimated parameters, as well as the AIC values are shown in Tables 1 and 2 for sample values $m = 100$, the parameters of the Pareto distribution $\mu = 1$ and $\alpha \in \{0,5;1;3;10\}$, and for Binomial distribution parameters $n \in \{4;40\}$ and $p \in \{0,2;0,5;0,9\}$.

Also, in Tables 1 and 2 the values of the AIC, BIC, HQIC, CAIC are expressed. These values have been obtained by means of the EXCEL computing environment.

Based on the numerical values in Tables 1 and 2, the following situations are represented:

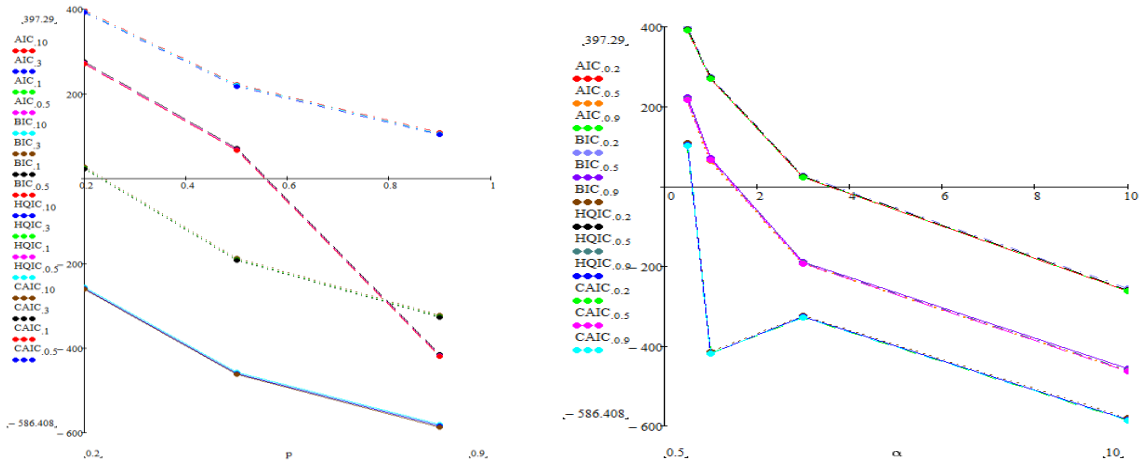
- The values of the information criteria (AIC, BIC, HQIC, CAIC) according to the estimated parameters $\hat{\Theta} = (\hat{\alpha}, \hat{p})$, (Fig. 1);
- The comparative graphical analysis (based on the value categories of α) of the information criteria values in relation to the values of the parameter p (Fig. 2).

Table1. Estimated parameter values and the AIC, BIC, HQIC, CAIC values for $MinParB(1, \alpha, 4, p)$

$(\alpha; p)$	$\hat{\alpha}$	$\hat{p}$	$\hat{h}$	*AIC*	*BIC*	*HQIC*	*CAIC*
(10;0,2)	10,633	0,025	1159	-261,302	-256,091	-259,193	-261,178
(10;0,5)	10,625	0,446	712	-462,254	-457,044	-460,146	-462,131
(10;0,9)	11,033	0,724	457	-586,408	-581,197	-584,299	-586,284
(3;0,2)	3,188	0,025	855	22,376	27,586	24,485	22,500
(3;0,5)	3,187	0,446	588	-193,921	-188,711	-191,813	-193,798
(3;0,9)	3,310	0,724	392	-328,444	-323,233	-326,335	-328,320
(1;0,2)	1,062	0,027	646	269,888	275,098	271,996	270,011
(1;0,5)	1,062	0,446	474	65,457	70,668	67,566	65,581
(1;0,9)	1,103	0,724	333	-420,047	-414,837	-417,938	-419,923
(0,5;0,2)	0,531	0,027	646	392,080	397,290	394,189	392,204
(0,5;0,5)	0,530	0,447	426	217,308	222,518	219,417	217,432
(0,5;0,9)	0,551	0,724	302	102,860	108,070	104,968	102,983

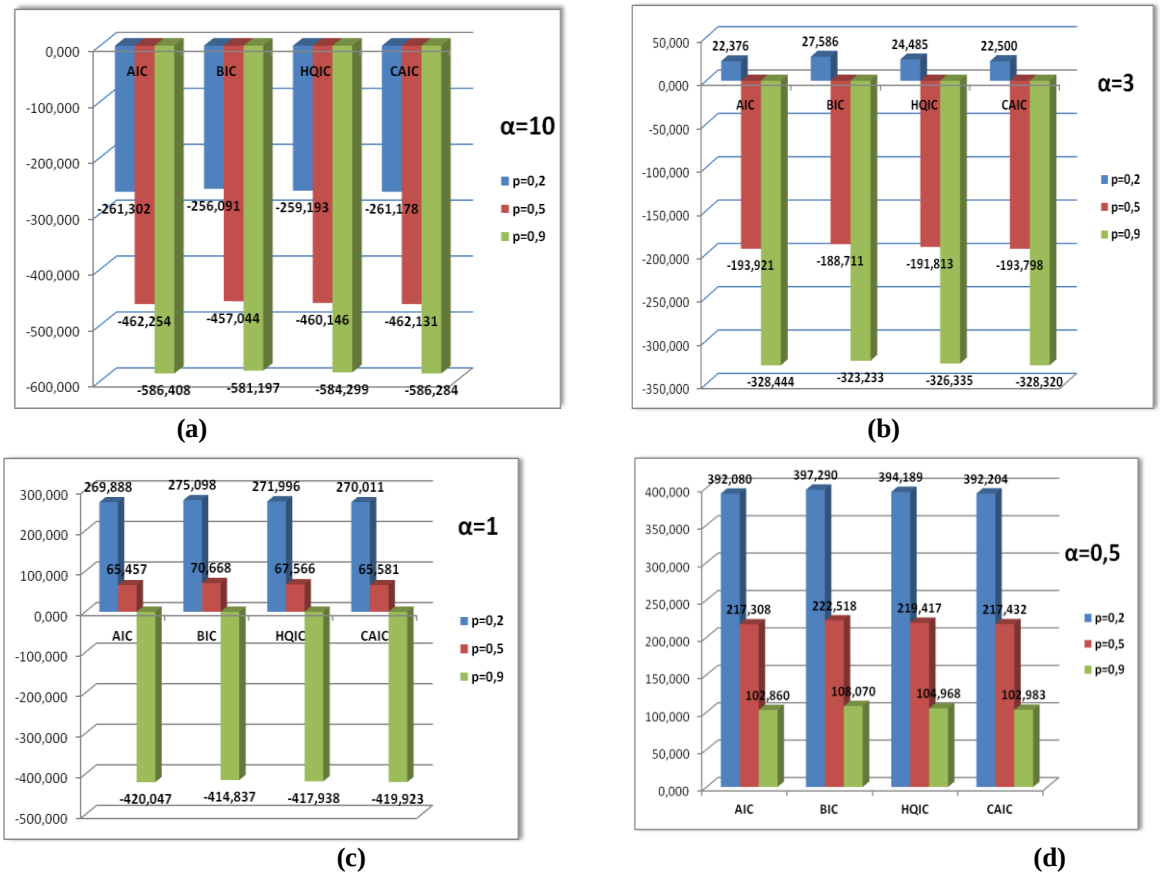
Table 2. Estimated parameter values and AIC, BIC, HQIC, CAIC values for $MinParB(1, \alpha, 40, p)$

$(\alpha; p)$	$\hat{\alpha}$	$\hat{p}$	$\hat{h}$	*AIC*	*BIC*	*HQIC*	*CAIC*
(10;0,2)	14,575	0,142	1339	-1,078E+71	-1,078E+71	-1,078E+71	-1,078E+71
(10;0,5)	8,831	0,585	5001	-7,134E+50	-7,134E+50	-7,134E+50	-7,134E+50
(10;0,9)	10,116	0,832	5001	-2,937E+53	-2,937E+53	-2,937E+53	-2,937E+53
(3;0,2)	4,371	0,142	1149	-584,960	-579,750	-582,852	-584,837
(3;0,5)	2,649	0,585	5001	-809,318	-804,108	-807,210	-809,195
(3;0,9)	3,035	0,832	5001	-910,978	-905,768	-908,869	-910,854
(1;0,2)	1,456	0,142	975	-345,206	-339,996	-343,097	-345,082
(1;0,5)	0,883	0,585	5001	-583,025	-577,815	-580,916	-582,901
(1;0,9)	1,101	0,832	5001	-703,705	-698,494	-701,596	-703,581
(0,5;0,2)	0,727	0,142	865	-184,485	-179,275	-182,376	-184,361
(0,5;0,5)	0,442	0,585	5001	-434,522	-429,312	-432,413	-423,398
(0,5;0,9)	0,506	0,832	5001	-576,236	-571,025	-574,127	-576,112



(a) **(b)**

FIG.1. (a) AIC depending on the probability p and $\alpha \in \{0,5;1;3;10\}$ established; (b) AIC depending on α and $p \in \{0,5;1;3;10\}$ for the distribution $MinParB(1,\alpha,4,p)$



(c) **(d)**

FIG.2. The values of the AIC, BIC, HQIC, CAIC depending on the probability p and $\alpha \in \{0,5;1;3;10\}$ established in situations (a), (b), (c), respectively (d) for the distribution $MinParB(1,\alpha,4,p)$ **.** Comparative graphic analysis

CONCLUSIONS

The main objective of this study is to perform a quantitative analysis of the statistical model as described in paper [4] in a unitary manner and from the perspective of the power series distribution class [3]. The values of the main information criteria (AIC, BIC HQIC, CAIC) as described in Section 2 have been determined. The values of the information criteria are closely related to the existence of the maximum likelihood function and the presence of the estimated parameters by means of the EM algorithm [4].

The findings of our analysis are, as follows: according to the representations in Fig. 1(a) a decrease in the parameter α determines an increase in the values of the information criteria AIC, BIC, HQIC, CAIC; this dependence decreases when the values of p increase. Also, it has been noted that compared to the threshold value $p = 0,5$ the values AIC, BIC HQIC, CAIC are equidistant, except when $\alpha = 1$. The following can be observed based on Fig. 1 (b): the higher the probability p, the smaller the values of AIC, BIC, HQIC, CAIC. Compared to the threshold value of the parameter $\alpha = 3$, the values of the information criteria are equidistant.

It can also be noted that the lowest values are characterized by the AIC information criterion in all the analyzed situations (Table 1, Table 2, Fig. 2). For example, based on Fig. 2(a), it can be concluded that the statistical model $MinParB(1,10,4,0.9)$ is selected as the best, providing us the best information, whereas Fig. 2(d) shows that the distribution $MinParB(1,0.5,4,0.9)$ is the model selected as being the best. From Fig. 2, for high probabilities (for example $p = 0,9$) we have low values for all the information criteria, and from Table 2, the values of the information criteria AIC, BIC HQIC, CAIC for the distribution $MinParB(1,10,4,p)$, $p \in \{0.2;0.5;0,9\}$ are very small, therefore a qualitative analysis of this distribution cannot be made in relation to the other distributions.

REFERENCES

[1] H. Akaike, A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19 (6), pp. 716–723, 1974.

[2] K. P. Burnham, D. R. Anderson, *Model Selection and Multimodel Inference*: *A Practical Information-Theoretic Approach* (2nd ed.), Springer-Verlag, 2002.

[3] 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;

[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] N. Enache-David, *On an application on entropy*, Bulletin of the "Transilvania" University of Brasov, Vol 6(55), No. 2 – 2013, Series III: Mathematics, Informatics, Physics, ISSN 2065-2151, pp. 87-94, 2013.

A NEW SUZUKI TYPE FIXED POINT THEOREM

Andreea FULGA

Transilvania University of Brasov, Romania (afulga@unitbv.ro)

DOI: 10.19062/1842-9238.2017.15.2.9

Abstract: *In this paper we prove a fixed point result for F-Suzuki contractions.*

Keywords: *fixed point, metric space, F-contraction.*

1. INTRODUCTION

Banach's contraction principle (BCP) [1] is one of the initial and also fundamental results in theory of fixed point. In the literature, there are plenty of extensions of this result.

Theorem 1.1.([1]). *Let* (X,d) *be a complete metric space and let* $T:X\to X$ *a contraction* $(d(Tx,Ty)\le c\cdot d(x,y),(\forall)x,y\in X,c\in[0,1))$. *Then T has a unique fixed point in X.*

Several authors have obtained many extensions and generalizations of the (BCP). So, in 1962, Edelstein [2] proved the next version of contraction principle.

Theorem 1.2.([2]). *Let* (X,d) *be a compact metric space and let* $T:X\to X$ *. Assume that* $d(Tx,Ty)<d(x,y)$ *for all* $x,y\in X$ *with* $x\neq y$. *Then T has a unique fixed point in X.*

In 2009, Suzuki [7] proved generalized versions of Edelstein's result in compact metric space as follows.

Theorem 1.3.([7]).*Let* (X,d) *be a compact metric space and let* $T:X\to X$ *. Assume that*

$\left[\frac{1}{2}d(x,Tx)<d(x,y)\Rightarrow d(Tx,Ty)<d(x,y)\right]$ *for all* $x,y\in X$ *with* $x\neq y$. *Then T has a unique fixed point in X.*

Later, in 2012,Wardowski [9] generalized the Banach contraction principle in a different manner, introducing a new type of contractions called *F*-contraction.

Definition 1.4. ([9]). *Let* (X,d) *be a metric space. An operator* $T:X\to X$ *is said to be an F-contraction if there exists* $\tau>0$ *such that*

$$d(Tx,Ty)>0\Rightarrow\tau+F(d(Tx,Ty))\le F(d(x,y)),(\forall)x,y\in X \tag{1}$$

where $F:(0,\infty)\to R$ *is a mapping satisfying the following conditions:*

(F1) F is strictly increasing, i.e. for all $\alpha,\beta\in(0,\infty)$, *such that* $\alpha<\beta,F(\alpha)<F(\beta)$;

(F2)For each sequence $\{\alpha_n\}_{n>0}$ *of positive numbers* $\lim_{n\to\infty}\alpha_n=0$ *if and only if* $\lim_{n\to\infty}F(\alpha_n)=-\infty$

(F3) There exists $k\in(0,1)$ *such that* $\lim_{\alpha\to0^+}\alpha^kF(\alpha)=0$.

Theorem 1.5.([9]). *Let* (X,d) *be a complete metric space and let* $T:X\to X$ *be an F-contraction. Then T has a unique fixed point* $x^*\in X$ *and for every* $x\in X$ *the sequence* $\{T^n x\}_{n\in}$ *converges to* x^*.

In 2014, Piri [5] proved the following result:

Theorem 1.6. ([5]). *Let* (X,d)*be a complete metric space and* $T:X\to X$ *be an F-Suzuki contraction. Then T has a unique fixed point* $x^*\in X$ *and for every* $x\in X$ *the sequence* $\{T^n x\}_{n\in}$ *converges to* x^*.

Definition 1.7. ([5]). *Let* (X,d)*be a metric space. A mapping* $T:X\to X$ *is said to be an F-Suzuki contraction if there exists* $\tau>0$ *such that for all* $x,y\in X$ *with* $Tx\neq Ty$

$$\frac{1}{2}d(x,Tx)<d(x,y)\Rightarrow \tau+F(d(Tx,Ty)\le F(d(x,y)), \tag{2}$$

where $F:\mathrm{R}_+\to\mathrm{R}$ *is a mapping satisfying the following conditions:*

(Fs1) F is strictly increasing, i.e. for all $\alpha,\beta\in(0,\infty)$, *such that* $\alpha<\beta, F(\alpha)<F(\beta)$;

(Fs2) inf $F=-\infty$;

(Fs3) F is continuous on $(0,\infty)$.

In this paper, using the idea from [4] we introduced a new type of *F*-contraction, and will prove a fixed point theorem which generalizes some known results.

2. MAIN RESULTS

First, let F denote the family of all functions $F:\mathrm{R}_+\to\mathrm{R}$ which satisfies the following conditions:

(F_E1) F is strictly increasing, that is, for all $x,y\in R_+$, if $x<y$ then $F(x)<F(y)$;

(F_E2) F is continuous on $(0,\infty)$.

Definition 2.1. *Let* (X,d)*be a complete metric space. A map* $T:X\to X$ *is said to be a* F_E ***-Suzuki contraction*** *on (X,d) if there exists* $F\in\mathfrak{F}$ *and* $\tau>0$ *such that for all* $x,y\in X$

$$\frac{1}{2}d(x,Tx)<d(x,y)\Rightarrow \tau+F(d(Tx,Ty))\le F(E(x,y)) \tag{3}$$

where

$$E(x,y)=d(x,y)+|d(x,Tx)-d(y,Ty)| \tag{4}$$

Theorem 2.2. *Let* (X,d)*be a complete metric space and* $T:X\to X$ *be an* F_E *-Suzuki contraction. Then T has a unique fixed point* $x^*\in X$ *and for every* $x_0\in X$ *the sequence* $\{T^n x_0\}_{n\in}$ *converges to* x^*.

Proof: Let $x_0\in X$ be arbitrary and fixed. We define a sequence $\{x_n\}_{n=1}^{\infty}$ by

$$x_1=Tx_0,\ x_2=Tx_1=T^2x_0,\ldots,x_{n+1}=Tx_n=T^{n+1}x_0,\quad \forall n>1 \tag{5}$$

Suppose that $x_{n_0}=x_{n_0+1}$ for some $n_0\in\mathrm{N}$. Then $Tx_{n_0}=x_{n_0}$. This proves that x_{n_0} is a fixed point of *T*.

From now, we assume that $x_n \neq x_{n+1},\ \forall n \in \mathrm{N}$. Then $0 < d(x_n, x_{n+1}) = d(x_n, Tx_n)$ and $\frac{1}{2} d(x_n, Tx_n) < d(x_n, Tx_n) = d(x_n, x_{n+1}), \forall n \in \mathrm{N}$. It follows from (3), that there exist $\tau > 0$ so that

$$\tau + F\left(d\left(Tx_n, T^2 x_n\right)\right) \leq F\left(E\left(x_n, Tx_n\right)\right) \Leftrightarrow \tau + F\left(d\left(x_{n+1}, x_{n+2}\right)\right) \leq F\left(E\left(x_n, x_{n+1}\right)\right) \tag{6}$$

where

$$\begin{aligned} E(x_n, x_{n+1}) &= d(x_n, x_{n+1}) + \left|d(x_n, Tx_n) - d(x_{n+1}, Tx_{n+1})\right| \\ &= d(x_n, x_{n+1}) + \left|d(x_n, x_{n+1}) - d(x_{n+1}, x_{n+2})\right| \end{aligned}$$

If we denote by $d_n = d(x_n, x_{n+1})$ we have $E(x_n, x_{n+1}) = d_n + |d_n - d_{n+1}|$ and (6) becomes

$$\tau + F(d_{n+1}) \leq F\left(d_n + |d_n - d_{n+1}|\right) \tag{7}$$

If there exists $n \in$ such that $d_{n+1} > d_n$, then $\tau + F(d_{n+1}) \leq F(d_{n+1}) \Rightarrow \tau \leq 0.$

This is a contradiction. Then, for $d_n < d_{n+1}$, because $\tau > 0$, we have

$$\tau + F(d_{n+1}) \leq F(2d_n - d_{n+1}) \Leftrightarrow F(d_{n+1}) \leq F(2d_n - d_{n+1}) - \tau < F(2d_n - d_{n+1}) \tag{8}$$

and using ($F_E 1$), $d_{n+1} < 2d_n - d_{n+1}$, so, the sequence $\{d_n\}$ is strictly increasing and bounded.

Now, let $d = \lim_{n\to\infty} d_n$ and we suppose that $d > 0$. Because $\{d_n\} \downarrow d$ it result that $(2d_n - d_{n+1}) \downarrow d$, and taking the limit as $n \to \infty$ in (8), we get

$\tau + F(d+0) \leq F(d+0) \Rightarrow \tau \leq 0.$

But, this is a contradiction. Therefore,

$$\lim_{n\to\infty} d_n = \lim_{n\to\infty} d(x_n, Tx_n) = 0. \tag{9}$$

In order to prove that $\{x_n\}_{n>0}$ is a Cauchy sequence in metric space (X, d), we suppose contrary, that is, there exist $\varepsilon > 0$ and the sequences $\{n(k)\}, \{m(k)\}$ of positiv integers with $n(k) > m(k) > k$ such that $d(x_{n(k)}, x_{m(k)}) \geq \varepsilon$ and $d(x_{n(k)-1}, x_{m(k)}) < \varepsilon$, $(\forall) k \in N$.

Then we have $\varepsilon \leq d(x_{n(k)}, x_{m(k)}) \leq d(x_{n(k)}, x_{n(k)-1}) + d(x_{n(k)-1}, x_{m(k)}) < d(x_{n(k)-1}, x_{n(k)}) + \varepsilon$.

Letting $k \to \infty$ and using (9) it follows that

$$\lim_{k\to\infty} d(x_{n(k)}, x_{m(k)}) = \varepsilon \tag{10}$$

From (9) and (10) it result there exist a natural number *N* such that

$$\frac{1}{2} d(x_{n(k)}, x_{n(k)+1}) = \frac{1}{2} d(x_{n(k)}, T(x_{n(k)})) < \frac{\varepsilon}{2} < d(x_{n(k)}, x_{m(k)}), \quad (\forall) k \geq N.$$

So, because the assumption of the theorem, we get

$$\frac{1}{2}d\left(x_{n(k)},T(x_{n(k)})\right)<d(x_{n(k)},x_{m(k)})\Rightarrow\tau+F\left[d\left(Tx_{n(k)},Tx_{m(k)}\right)\right]\leq F\left[E\left(x_{n(k)},x_{m(k)}\right)\right],\quad(\forall)k\geq N$$

$$\Leftrightarrow\tau+F\left[d\left(x_{n(k)+1},x_{m(k)+1}\right)\right]\leq F\left[d\left(x_{n(k)},x_{m(k)}\right)\right.$$

Taking the limit as $k\to\infty$ and using ($F_E 2$)

$$\tau+F(\varepsilon)\leq F(\varepsilon)\Rightarrow\tau\leq 0.$$

It is a contradiction. This shows that $\{x_n\}$ is a Cauchy sequences and by completeness of X there converges to some point $x^*\in X$. Therefore

$$\lim_{n\to\infty}d\left(Tx_n,x^*\right)=0. \tag{11}$$

Next, we show that x^* is a fixed point of T . For this, we claim that

$$\frac{1}{2}d\left(x_n,Tx_n\right)<d\left(x_n,x^*\right)or\ \frac{1}{2}d\left(Tx_n,T^2x_n\right)<d\left(Tx_n,x^*\right) \tag{12}$$

Assume that there exists $m\in\mathrm{N}$ such that

$$\frac{1}{2}d\left(x_m,Tx_m\right)\geq d\left(x_m,x^*\right)and\ \frac{1}{2}d\left(Tx_m,T^2x_m\right)\geq d\left(Tx_m,x^*\right) \tag{13}$$

Then,

$$d\left(x_m,x^*\right)\leq\frac{1}{2}d\left(x_m,Tx_m\right)\leq\frac{1}{2}\left[d\left(x_m,x^*\right)+d\left(x^*,Tx_m\right)\right]$$

which implies that

$$d\left(x_m,x^*\right)\leq d\left(x^*,Tx_m\right) \tag{14}$$

and from (13)

$$d\left(x_m,x^*\right)\leq d\left(x^*,Tx_m\right)\leq\frac{1}{2}d\left(Tx_m,T^2x_m\right) \tag{15}$$

Since $\frac{1}{2}d\left(x_m,Tx_m\right)<d\left(x_m,x_{m+1}\right)=d\left(x_m,Tx_m\right)$, by the assumption of theorem we get

$$F\left(d\left(Tx_m,T^2x_m\right)\right)\leq F\left[E\left(x_m,Tx_m\right)\right]-\tau\leq F\left[E\left(x_m,Tx_m\right)\right]$$

because $\tau>0$.

So, from ($F_E 1$) we get

$$d\left(Tx_m,T^2x_m\right)\leq E\left(x_m,Tx_m\right)=d\left(x_m,Tx_m\right)+\left|d\left(x_m,x_{m+1}\right)-d\left(Tx_m,T^2x_m\right)\right|$$

$$=2d\left(x_m,Tx_m\right)-d\left(Tx_m,T^2x_m\right)\Leftrightarrow d\left(Tx_m,T^2x_m\right)\leq d\left(x_m,Tx_m\right) \tag{16}$$

and from (13), (15), (16) it follows that

$$d\left(Tx_m, T^2x_m\right) < d\left(x_m, Tx_m\right) \le d\left(x_m, x^*\right) + d\left(x^*, Tx_m\right) \le d\left(Tx_m, T^2x_m\right)$$

This is a contradiction. Hence relations (12) holds.
We suppose now that $Tx^* \neq x^*$.
(1) If $\frac{1}{2}d\left(x_n, Tx_n\right) < d\left(x_n, x^*\right)$ from assumption of theorem,

$$\tau + F\left(d\left(Tx_n, Tx^*\right)\right) \le F\left(E\left(x_n, x^*\right)\right) \Leftrightarrow$$
$$\tau + F\left(d\left(x_{n+1}, Tx^*\right)\right) \le F\left(d\left(x_n, x^*\right) + \left|d(x_n, x_{n+1}) - d(x^*, Tx^*)\right|\right)$$

Taking the limit and using $(F_E 2)$ we have
$$\tau + F\left(d(x^*, Tx^*)\right) \le F\left(d(x^*, Tx^*)\right) \Rightarrow \tau \le 0$$

This is a contradiction.
(2) If $\frac{1}{2}d\left(Tx_n, T^2x_n\right) < d\left(Tx_n, x^*\right)$ then
$$\tau + F\left(d(T^2x_n, Tx^*)\right) \le F\left(E\left(Tx_n, x^*\right)\right) \Leftrightarrow$$
$$\tau + F\left(d(x_{n+2}, Tx^*\right) \le F\left(d(x_{n+1}, x^*) + \left|d(x_{n+1}, x_{n+2}) - d\left(x^*, Tx^*\right)\right|\right)$$

So, taking the limit when:
$$\tau + F\left(x^*, Tx^*\right) \le F\left(x^*, Tx^*\right) \Rightarrow \tau \le 0$$

Hence x^* is a fixed point of T.

Finally, we prove that the fixed point of T is unique. For this, let x^*, y^* be two fixed points of T and suppose that $Tx^* = x^* \neq y^* = Ty^*$, so $d(x^*, y^*) > 0$.

Because $E\left(x^*, y^*\right) = d(x^*, y^*) + \left|d(x^*, Tx^*) - d\left(y^*, Ty^*\right)\right| = d(x^*, y^*)$ it follows that

$$0 = \frac{1}{2}d(x^*, Tx^*) < d\left(x^*, y^*\right) \Rightarrow \tau + F\left(d\left(Tx^*, Ty^*\right)\right) \le F\left(E\left(x^*, y^*\right)\right) \Leftrightarrow$$
$$\Leftrightarrow \tau + F\left(d\left(x^*, y^*\right)\right) \le F\left(E\left(x^*, y^*\right)\right) \Rightarrow \tau \le 0.$$

It is a contradiction. Then, $d(x^*, y^*) = 0$, that is $x^* = y^*$. This proves that the fixed point of T is unique.

REFERENCES

[1] B.Banach, *Sur les operations dans les ensembles abstraits et leur applications aux equations integrals*, Fundam. Math.,3, 133-181, 1922;
[2] M.Edelstein, *On fixed and periodic points under contractive mappings*, J.London Math.Soc.37, 74-79, 1962;
[3] E.Karapinar, *Edelstein type fixed point theorems*, Fixed Point Theory, Appl.2012, Article ID 107, 2012;
[4] O.Popescu, *A new type of contractive mappings in complete metric spaces*, submitted;
[5] H.Piri and P.Kumam, *Some fixed point theorems concerning F-contraction in complete metric spaces*, Fixed Point theory and Applications, 2014:210, 2014;
[6] N.A.Secelean, *Iterated function systems consisting of F-contractions, Fixed Point Theory* Appl.,2013, Article ID 277, 2013;
[7] T.Suzuki, *A new type of fixed point theorem in metric spaces, Nonlinear Anal*. 71, 5313-5317, 2009;

[8] T.Suzuki, *A generalized Banach contraction principle that characterizes metric completeness*, Proc.Amer.Math.Soc., 136(2008), 1861-1869, 2008;
[9] D.Wardowski, *Fixed point theory of a new type of contractive mappings in complete metric spaces*, Fixed Point Theory Appl.2012, Aricle ID 94, 2012;
[10] D.Wardowski and N. Van Dung, *Fixed points of F-weak contractions on complete metric spaces*, Demonstratio Mathematica, vol.XLVII, no1, 145-155, 2014.

EXPONENTIATED POWER QUASI LINDLEY DISTRIBUTION. SUBMODELS AND SOME PROPERTIES

Manuela GHICA, Nicoleta DIMCEVICI POESINA, Irina PRASACU

Faculty of Pharmacy, University of Medicine and Pharmacy, Carol Davila, Bucharest, Romania (manuela.ghica@umfcd.ro)

DOI: 10.19062/1842-9238.2017.15.2.10

Abstract: *In this paper we introduce a new generalization of the Lindley distribution which generalizes the power Lindley distribution, proposed by Ghitany, and another form of generalized Lindley proposed by Nadarajah. This new kind of generalized Lindley distribution has four parameters and it allows more adaptability to analyze real lifetime data.*

Keywords: *Lindley distribution, order statistics*

1. INTRODUCTION

In this paper, we deal with a generalization of Lindley distribution because it forms a flexible family of distributions with an important selection of shape and hazard functions. The Lindley distribution was firstly proposed by Lindley (1958) in the context of Bayesian statistics, based on Bayes theorem [1], [2] as a counterexample of fiducial statistics. Mixing various distributions lead to the expansion of known families of distributions. In literature, there were introduced and studied some mixed data modeling distributions of life as Weibull Poisson, Weibull geometric, Exponential geometric.

Lindley distribution is a one-parameter distribution, given by its probability density function

$$f(x)=\frac{\theta^2}{\theta+1}(x+1)e^{-\theta x} \tag{1}$$

The cumulative distribution function corresponding to (1) is

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

The properties of the Lindley distribution were studied by M.E. Ghitany, B. Atieh, S. Nadarajah [4, 5, 6]. They discussed its applications to survival data and, also, showed in a numerical example that the Lindley distribution gives better modeling for waiting times and survival time data than the exponential distribution. Different forms of generalized Lindley distributions were been widely applied for reliability modeling and life testing data [10, 11]. There is a great development of another various quantitative techniques for solving optimization problems for biological and economical domains [8, 9].

Definition 1.1 Let X be a random variable and the parameters $\alpha, \theta > 0$. We say that X has a quasi Lindley distribution $X \sim QL(\alpha, \theta)$ if it has the probability density function

$$f(x) = \frac{\theta^2}{\alpha\theta + 1} e^{-\theta x} (\alpha + x)$$

and the cumulative distribution function

$$F(x) = 1 - \frac{\alpha\theta + 1 + \theta x}{\alpha\theta + 1} e^{-\theta x}, \quad x > 0, \alpha, \theta > 0$$

Because the Lindley distribution (having only one parameter) does not provide enough flexibility for analyzing different types of lifetime data, in statistic literature it were introduced some new compounding the Lindley distribution with Negative Binomial distribution [3], with Poisson distribution [4] or Exponential Poisson [7] offering some new distributions of lifetime case obtaining from Generalized Lindley distribution compounding with exponential and gamma distributions. The quasi Lindley distribution reduces of the one following known distribution:

1. For $\alpha = 1$, it becomes $Lindley(\theta)$
2. For $\alpha = 0$, it becomes $Gamma(2, \theta)$
3. For $\alpha \to \infty$, it becomes $Exp(\theta)$

The quasi Lindley distribution maybe can write as a two-component mixture of $Exp(\theta)$ and $Gamma(2, \theta)$:

$$f(x) = pf_1(x) + (1-p) f_2(x), \quad x > 0, \alpha, \theta > 0$$

where $p = \frac{\alpha\theta}{\alpha\theta + 1}$, $f_1(x) = \theta e^{-\theta x}$ and $f_2(x) = \theta^2 x e^{-\theta x}$.

Ghitany et al. proposed the power transformation, $Y = X^{1/\beta}, \beta > 0$, for Lindley distribution for generating a flexible family of probability distributions. The new parameter would offer more distributional flexibility with a form of the hazard rate what can be decreasing, unimodal and decreasing-increasing-decreasing for some particular cases of the parameters.

Definition 1.2 Let X be a random variable and the parameters $\alpha, \theta, \beta > 0$. We say that X has a quasi-power Lindley distribution $X \sim QPL(\alpha, \theta, \beta, b)$ if it has the probability density function

$$f(y) = \frac{\alpha\theta^2\beta}{\alpha\theta + 1} y^{\beta-1} e^{-\theta y^\beta} + \frac{\beta\theta^2}{\alpha\theta + 1} y^{2\beta-1} e^{-\theta y^\beta}$$

and the cumulative distribution function

$$F(y) = 1 - \left(1 + \frac{\theta y^\beta}{\alpha\theta + 1}\right) e^{-\theta y^\beta}, \quad y > 0, \alpha, \theta, \beta > 0$$

The quasi power Lindley distribution may be can write as a two-component mixture of $Weibull(\beta,\theta)$ and $Gamma(2,\beta,\theta)$:

$$f(y)=pf_1(y)+(1-p)f_2(y),\quad y>0,\alpha,\theta,\beta>0$$

where $p=\dfrac{\alpha\theta}{\alpha\theta+1}$, $f_1(y)=\theta\beta y^{\beta-1}e^{-\theta y^{\beta}}$ and $f_2(y)=\theta^2\beta y^{2\beta-1}e^{-\theta y^{\beta}}$.

We introduce a new four parameter distribution, denoted

$X\sim EQPL(\alpha,\theta,\beta,b),\alpha,\theta,\beta,b>0$

referred to as the exponentiated quasi power Lindley. This new distribution reduces to the quasi Lindley distribution, the exponential distribution and gamma distribution. On terms of reliability, the various shapes of the EQPL distribution give it an advantage, being more suitable to model many real systems which generally exhibit bath-tub shaped failure rate.

Definition 1.3 Let $X\sim EQPL(\alpha,\theta,\beta,b)$. The cumulative function of the $EQPL(\alpha,\theta,\beta,b)$, $\alpha,\theta,\beta,b>0$ is

$$F(x)=\left[1-\left(1+\frac{\theta x^{\beta}}{\alpha\theta+1}\right)e^{-\theta x^{\beta}}\right]^{b} \tag{3}$$

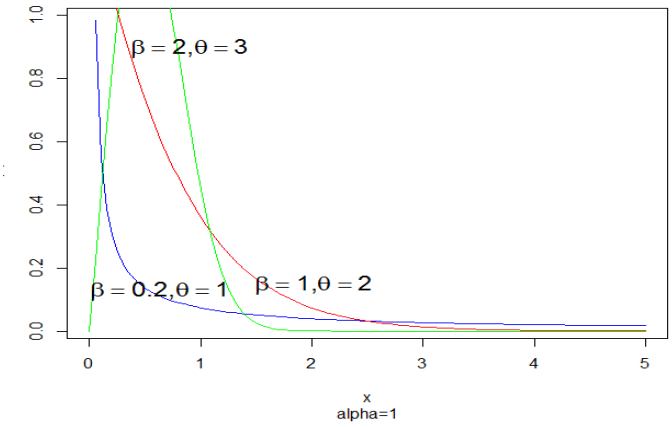
and the corresponding probability density is given by

$$f(x)=\frac{\beta\theta^2 b}{\alpha\theta+1}x^{\beta-1}\left(\alpha+x^{\beta}\right)e^{-\theta x^{\beta}}\left[1-\left(1+\frac{\theta x^{\beta}}{\alpha\theta+1}\right)e^{-\theta x^{\beta}}\right]^{b},\quad x>0 \tag{4}$$

The EQPL distribution reduces of the one following known distribution:

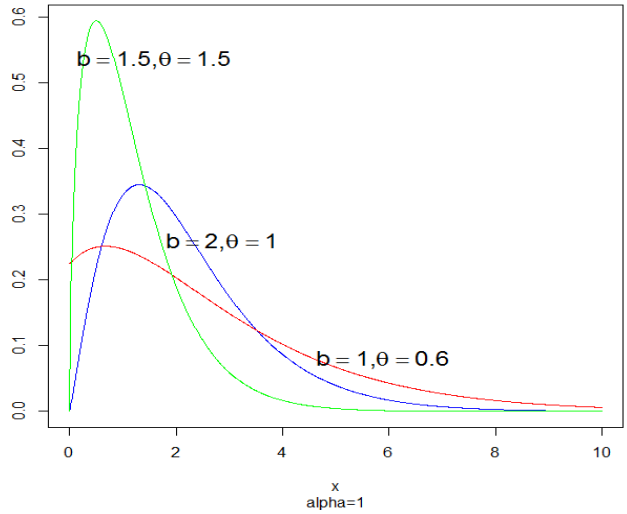
1. For b=1, it becomes $QPL(\alpha,\theta,\beta)$

$$f(x,\alpha,\theta,\beta)=\frac{\beta\theta^2 b}{\alpha\theta+1}x^{\beta-1}\left(\alpha+x^{\beta}\right)e^{-\theta x^{\beta}}$$



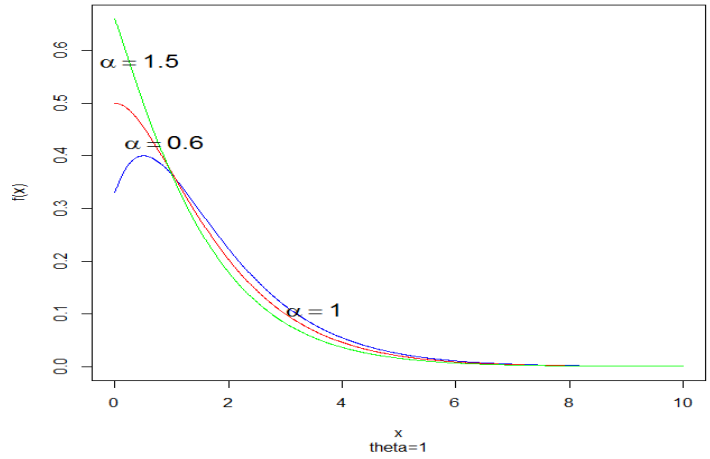
2. For $\beta=1$, it becomes PowerLindley (α,θ,b)

$$f(x,\alpha,\theta,b)=b\theta^2e^{-\theta x}\frac{\alpha+x}{\theta\alpha+1}\left(1-e^{-\theta x}\left(\frac{\theta x}{\theta\alpha+1}+1\right)\right)^{b-1}$$

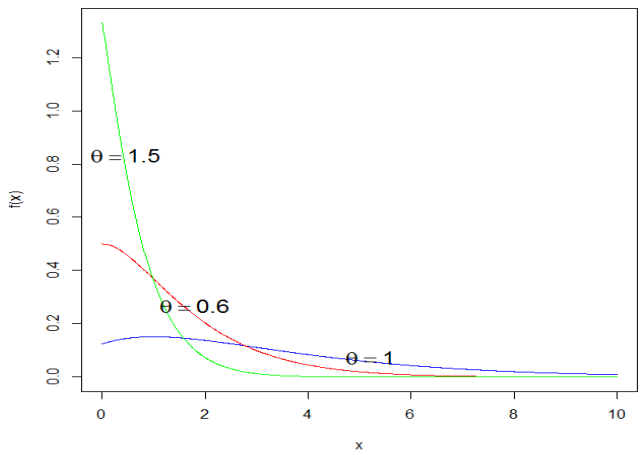


3. For $b=\beta=1$, it becomes $QL(\alpha,\theta)$

$$f(x,\alpha,\theta)=\theta^2e^{-\theta x}\frac{\alpha+x}{\theta\alpha+1}$$



4. For $b=\beta=\alpha=1$, it becomes $L(\theta)$



The cdf of X can also be represented in an extended form

$$F(x)=\sum_{i=1}^{\infty}\binom{b}{i}p^{b-i}(1-p)^{i}F_{W(\beta,\theta)}^{b-i}(x)F_{GGamma(2,\beta,\theta)}^{i}(x)$$

Definition 1.4 The corresponding hazard rate function is

$$h(x)=\frac{\beta\theta^{2}b}{\alpha\theta+1}x^{\beta-1}\left(\alpha+x^{\beta}\right)e^{-\theta x^{\beta}}\left[1-\left(1+\frac{\theta x^{\beta}}{\alpha\theta+1}\right)e^{-\theta x^{\beta}}\right]^{b}\left\{S(x)\right\}^{-1}$$

where

$$S(x)=S(x,\alpha,\theta,\beta,b)=1-\left[1-\left(1+\frac{\theta x^{\beta}}{\alpha\theta+1}\right)e^{-\theta x^{\beta}}\right]^{b}$$

2. STOCHASTIC ORDER

Let $X_1 \sim EQPL(\alpha_1,\theta_1,\beta_1,b_1)$ and $X_2 \sim EQPL(\alpha_2,\theta_2,\beta_2,b_2)$ be two exponentiated quasi power Lindley random variables with common shape β. Let F1 denote the cumulative distribution function of X_1 and F_2 the cumulative distribution function for X_2.

Defnition 2.1 We say that X_1 is stochastically greater or equal than $X_2\left(X_1 \geq_{st} X_2\right)$ if $F_{X_1}(x) \leq F_{X_2}(x)$, for all x where $F_{X_1}(x)$ and $F_{X_2}(x)$ are the cdfs of X_1 and X_2, respectively.

Definition 2.2 We say that X_1 is stochastically greater than X_2 with respect to likelihood ratio $\left(X_1 \geq_{lr} X_2\right)$ if $\frac{f_{X_2}(x)}{f_{X_1}(x)}$ is an increasing function of x, where $F_{X_1}(x)$ and $F_{X_2}(x)$ are the cdfs of X_1 and X_2, respectively.

Definition 2.3 We say X_1 is stochastically greater than X_2 with respect to reverse hazard rate $\left(X_1 \geq_{hr} X_2\right)$ if $h_{X_1}(x) \leq h_{X_2}(x)$ and for all x.

For establishing stochastic order we have the following important results due to Shaked and Shantikumar

$$\left(X_1 \geq_{lr} X_2\right) \Rightarrow \left(X_1 \geq_{hr} X_2\right) \Rightarrow \left(X_1 \geq_{st} X_2\right)$$

The EQPL distribution is ordered with respect to the strongest one as shown in the following theorem.

Also, we have

$$\frac{f_{X_2}(x)}{f_{X_1}(x)}=\frac{\beta_2\theta_2^2 b_2 x^{\beta_2-1}\left(\alpha_2+x^{\beta_2}\right)}{\beta_1\theta_1^2 b_1 x^{\beta_1-1}\left(\alpha_1+x^{\beta_1}\right)}e^{-\theta_2 x^{\beta_2}+\theta_1 x^{\beta_1}}\times\left[1-\left(1+\frac{\theta_2 x^{\beta_2}}{\alpha_2\theta_2+1}\right)e^{-\theta_2 x^{\beta_2}}\right]^{b_2-1}\times$$

$$\times\quad\left[1-\left(1+\frac{\theta_1 x^{\beta_1}}{\alpha_1\theta_1+1}\right)e^{-\theta_1 x^{\beta_1}}\right]^{1-b_1} \tag{5}$$

Theorem 2.4 If we have $\alpha_1=\alpha_2,\theta_1=\theta_2$ and $\beta_1=\beta_2$ then X_2 is stochastically greater with respect to likelihood ratio than X_1 if and only if $b_1>b_2$.

Theorem 2.5 If we have $\alpha_1=\alpha_2, b_1=b_2=b\geq 1$ and $\beta_1=\beta_2$ then X_2 is stochastically greater with respect to likelihood ratio than X_1 if and only if $\theta_2\geq\theta_1$.

Theorem 2.6 If we have $\alpha_1=\alpha_2, b_1=b_2=b<1$ and $\beta_1=\beta_2$ then X_2 is stochastically greater with respect to likelihood ratio than X_1 if and only if $\theta_2\leq\theta_1$.

3. MOMENTS

We will obtain the moments of the EQPL distribution using the binomial series expansion.

Theorem 3.1 The rth moment of the exponentiated quasi power Lindley $E\left(X^r\right)$ is given by

$$E\left(X^r\right)=C_{i,k}\frac{\beta\theta^2}{\alpha\theta+1}\left[\alpha\frac{\Gamma\left(\frac{r+bk}{\beta}+1\right)}{\left[\theta(1+i)\right]^{\frac{r+bk}{\beta}+1}}+\alpha\frac{\Gamma\left(\frac{r+bk}{\beta}+2\right)}{\left[\theta(1+i)\right]^{\frac{r+bk}{\beta}+2}}\right]$$

Proof.

$$E\left(X^r\right)=\int_0^\infty x^r f(x)dx,$$

$$E\left(X^r\right)=\int_0^\infty\frac{\beta\theta^2}{\alpha\theta+1}x^{r+\beta-1}\left(\alpha+x^\beta\right)e^{-\theta x^\beta}\left[1-\left(1+\frac{\theta x^\beta}{\alpha\theta+1}\right)e^{-\theta x^\beta}\right]^{b-1}dx=$$

$$=\int_0^\infty\frac{\alpha\beta\theta^2 b}{\alpha\theta+1}x^{r+\beta-1}\left(\alpha+x^\beta\right)e^{-\theta x^\beta}\left[1-\left(1+\frac{\theta x^\beta}{\alpha\theta+1}\right)e^{-\theta x^\beta}\right]^{b-1}+$$

$$+\int_0^\infty\frac{\beta\theta^2 b}{\alpha\theta+1}x^{r+2\beta-1}e^{-\theta x^\beta}\left[1-\left(1+\frac{\theta x^\beta}{\alpha\theta+1}\right)e^{-\theta x^\beta}\right]^{b-1}dx.$$

Using the binomial series expansion of $\left[1-\left(1+\frac{\theta x^{\beta}}{\alpha\theta+1}\right)e^{-\theta x^{\beta}}\right]^{b-1}$ given by

$$\left[1-\left(1+\frac{\theta x^{\beta}}{\alpha\theta+1}\right)e^{-\theta x^{\beta}}\right]^{b-1}=\sum_{i=0}^{\infty}C_{b-1}^{i}(-1)^{i}\left(1+\frac{\theta x^{\beta}}{\alpha\theta+1}\right)^{i}e^{-i\theta x^{\beta}}$$

and using the following binomial series expansion of $\left(1+\frac{\theta x^{\beta}}{\alpha\theta+1}\right)^{i}$ given by

$$\left(1+\frac{\theta x^{\beta}}{\alpha\theta+1}\right)^{i}=\sum_{k=0}^{\infty}C_{i}^{k}\left(\frac{\theta}{\alpha\theta+1}\right)^{k}x^{k\beta}$$

We obtain the rth moment of X and we find

$$E\left(X^{r}\right)=C_{i,k}\int_{0}^{\infty}\frac{\alpha\beta\theta^{2}b}{\alpha\theta+1}x^{r+\beta-1+k\beta}e^{-\theta(1+i)x^{\beta}}dx+C_{i,k}\int_{0}^{\infty}\frac{\beta\theta^{2}b}{\alpha\theta+1}x^{r+2\beta-1+k\beta}e^{-\theta(1+i)x^{\beta}}dx.$$

Let $y=x^{\beta}$ and

$$C_{i,k}=\sum_{i=0}^{\infty}C_{b-1}^{i}(-1)^{i}\sum_{i=0}^{\infty}C_{i}^{k}\left(\frac{\theta}{\alpha\theta+1}\right)^{k}.$$

So, the rth moment can be rewritten

$$E\left(X^{r}\right)=C_{i,k}\frac{\beta\theta^{2}}{\alpha\theta+1}\left[\alpha\frac{\Gamma\left(\frac{r+bk}{\beta}+1\right)}{\left[\theta(1+i)\right]^{\frac{r+bk}{\beta}+1}}+\alpha\frac{\Gamma\left(\frac{r+bk}{\beta}+2\right)}{\left[\theta(1+i)\right]^{\frac{r+bk}{\beta}+2}}\right]$$

Theorem 3.2 The moment generating function of the exponentiated quasi power Lindley $M_{X}(t)$ is given by

$$M_{X}(t)=C_{i,k,j}\frac{\beta\theta^{2}}{\alpha\theta+1}\left[\alpha\frac{\Gamma\left(\frac{k\beta+j}{\beta}+1\right)}{\left[\theta(1+i)\right]^{\frac{k\beta+j}{\beta}+1}}+\alpha\frac{\Gamma\left(\frac{k\beta+j}{\beta}+2\right)}{\left[\theta(1+i)\right]^{\frac{k\beta+j}{\beta}+2}}\right]$$

Proof.

$$M_X(t)=\int_0^{\infty} e^{tx} f(x)dx$$

$$M_X(t)=\int_0^{\infty} e^{tx}\frac{\beta\theta^2}{\alpha\theta+1}x^{\beta-1}\left(\alpha+x^{\beta}\right)e^{-\theta x^{\beta}}\left[1-\left(1+\frac{\theta x^{\beta}}{\alpha\theta+1}\right)e^{-\theta x^{\beta}}\right]^{b-1}dx=$$

$$=\int_0^{\infty}\frac{\alpha\beta\theta^2 b}{\alpha\theta+1}x^{\beta-1}e^{-\theta x^{\beta}}e^{tx}\left[1-\left(1+\frac{\theta x^{\beta}}{\alpha\theta+1}\right)e^{-\theta x^{\beta}}\right]^{b-1}+$$

$$+\int_0^{\infty}\frac{\beta\theta^2 b}{\alpha\theta+1}x^{2\beta-1}e^{-\theta x^{\beta}}e^{tx}\left[1-\left(1+\frac{\theta x^{\beta}}{\alpha\theta+1}\right)e^{-\theta x^{\beta}}\right]^{b-1}dx.$$

Using the binomial series expansion like in the last theorem, we introduce the last two expansions in the moment generating function of X and we find

$$M_X(t)=C_{i,k}\int_0^{\infty}\frac{\alpha\beta\theta^2 b}{\alpha\theta+1}x^{\beta-1+k\beta}e^{-\theta(1+i)x^{\beta}}e^{tx}dx+C_{i,k}\int_0^{\infty}\frac{\beta\theta^2 b}{\alpha\theta+1}x^{2\beta-1+k\beta}e^{-\theta(1+i)x^{\beta}}e^{tx}dx.$$

Let consider

$$e^{tx}=\sum_{=0}^{\infty}\frac{t^j x^j}{j!}.$$

So, the moment generating function can be rewritten

$$M_X(t)=C_{i,k}\sum_{=0}^{\infty}\frac{t^j}{j!}\left(\int_0^{\infty}\frac{\alpha\beta\theta^2 b}{\alpha\theta+1}x^{\beta-1+k\beta}e^{-\theta(1+i)x^{\beta}}e^{tx}dx+C_{i,k}\int_0^{\infty}\frac{\beta\theta^2 b}{\alpha\theta+1}x^{2\beta-1+k\beta}e^{-\theta(1+i)x^{\beta}}e^{tx}\right)dx.$$

Let $y=x^{\beta}$

$$C_{i,k,j}=C_{i,k}\sum_{=0}^{\infty}\frac{t^j}{j!}.$$

The moment generating function has the last form

$$M_X(t)=C_{i,k,j}\frac{\beta\theta^2}{\alpha\theta+1}\left[\alpha\frac{\Gamma\left(\frac{k\beta+j}{\beta}+1\right)}{\left[\theta(1+i)\right]^{\frac{k\beta+j}{\beta}+1}}+\alpha\frac{\Gamma\left(\frac{k\beta+j}{\beta}+2\right)}{\left[\theta(1+i)\right]^{\frac{k\beta+j}{\beta}+2}}\right].$$

4. GENERATION ALGORITHMS

We consider simulating values of a random variable $X \sim EQPL(\alpha, \lambda, \beta, b)$

Algorithm 1

1. Generate $U_i \sim U(0,1), i = \overline{1,n}$

2. Set $X_i = \left\{ -\alpha - \frac{1}{\theta} - \frac{1}{\theta} W\left[-(\alpha\theta+1)\left(1-U_i^{1/b}\right)\exp(-\alpha\theta-1) \right] \right\}^{1/\beta}$, $i = \overline{1,n}.$

Algorithm 2

1. Generate $U_i \sim U(0,1), i = \overline{1,n}$
2. Generate $V_i \sim Exponential(0), i = \overline{1,n}$
3. Generate $G_i \sim \Gamma(2,0), i = \overline{1,n}$
4. If $U_i^{1/\beta} < p, p = \frac{\alpha\theta}{\alpha\theta+1}$, then set $X_i = V_i^{1/\beta}$ otherwise $X_i = G_i^{1/\beta}, i = \overline{1,n}$.

Simulation study

```
n=10
theta=seq(0,4,length=10)
beta=seq(0,4,length=10)
alpha=seq(0,4,length=10)
u=runif(n)
v=rexp(theta)
g=rgamma(2*beta,theta)
p=(alpha*theta)/(alpha*theta+1)
if (u^(1/b)<p)
{x=v^(1/b)}
else {x=g^(1/b)
x
[1] 0.86096886 0.02832007 0.89052860 0.16637737 0.27854706 0.78316390
[7] 0.30347644 0.42973307 0.21931648 0.76245404
```

REFERENCES

[1] Lindley D. V., *Fiducial distributions and Bayes Distributions*, Journal of the Royal Statistical Society, series B 20, 102-107, 1958;

[2] Lindley D. V., *Introduction to probability and statistics from a Bayesian viewpoint*, part II: Inference, Cambridge University Press, New York, 1965;

[3] Hassein Z., Noriszura I., *Negative Binomial Lindley Distribution and Its Application*, Journal of Mathematics and Statistics, 6 (1), 4-9, 2010;

[4] Ghitany M.E., Al-Mutairi D. K., Nadarajah S., *Zero-truncated Poisson-Lindley distribution and its application*, Math. Comput. Simul, 79, 279-287, 2008;

[5] Ghitany M.E., Al-Mutairi D. K., Balakrishnan N., Al-Nezi L.J., *Power Lind-ley distribution and associated inference*, Comput. Statist. Data Anal., 64,20-33, 2013;

[6] Ghitany ME.., Atieh B., Nadarajah S., *Lindley distribution and its application*, Mathematics and Computers in Simulation, Vol. 78, 4, 493-506, 2008;

[7] Nadarajah S., Bakouch SH, Tahmasbi R, *A generalized Lindley distribution*, 26, 145-149, 26, 2011;

[8] Danciulescu, D., Colhon, M., Splitting the structured paths in stratified graphs. Application in Natural Language Generation, An. Sti. U. Ovid. Co.-Mat.,22, 2, 59-69, 2014;

[9] Grigoras, G., Danciulescu, D., Sitnikov, C., Assessment Criteria of E-learning Environments Quality, Procedia Economics and Finance, 16, 40-46, Elsevier, 2014;
[10] Bancescu I., Dedu S., Stativa M., *Generalized Lindley populations under joint type II censored Scheme*, International Journal of Risk Theory, 6, 1, 1-13, 2016;
[11] Preda V., Băncescu I., *A new family of distributions with a general generic distribution for reliability studies. Properties and Application*, International Journal of Risk Theory, 6, 1, 13-39, 2016.

DATA ANALYSIS – BETWEEN THEORY AND PRACTICE

Nicoleta ENACHE-DAVID, Livia SANGEORZAN,
George-Alex STELEA

"Transilvania" University, Braşov, Romania (nicoleta.enache@unitbv.ro, sangeorzan@unitbv.ro, george@stelea.ro)

DOI: 10.19062/1842-9238.2017.15.2.11

Abstract: *In our paper we highlight some aspects of the text classification problem using the Naïve Bayes Multinomial classifier. We use Weka software for modeling this problem and we study the conditions that allow the classifier to obtain the highest prediction.*

Keywords: *classifier, prediction, text classification.*

1. INTRODUCTION

Document classification is a very actual issue and is a continuous challenge; it is based on different techniques of machine learning including Bayesian classification [6], SVM classifiers (Support Vector Machine) [9],[11] k-NN (k-Nearest-Neighbor) classifier [10], classification based on association rules [20], decision tree [16] etc.

Such machine learning techniques can be applied on complex information systems like in [1],[2],[3],[5], and for mathematical models like [6],[7],[8]. In the economic field, quantitative analysis of risk represents one of the phases that have to be followed in order to evaluate risks that an organization may face while developing its business [4], [12],[13]. This kind of analysis aims to numerical assessment of the probability and impact of each risk upon the organization's objectives. For this purpose there are used quantitative techniques such as the decision tree method [14],[15].

In literature there exist also statistical machine learning methods that can be applied to document clustering, document classification and predictive modeling. For testing the model inference, one can use the Monte Carlo method [18].

Our application for text classification takes into account the training set with synonyms and without synonyms. Synonyms are words having similar meaning. In our study we use the Naïve Bayes Multinomial classifier and we study the conditions that allow the classifier to obtain the highest prediction.

There are many studies on extracting synonyms automatically including the use of machine learning. Some studies analyze synonyms using similarity without machine learning. In [21] the authors present an automatic selection of synonyms using machine learning.

In our paper we are looking after synonyms for the four training sets for different categories (spam, sport, social-media, and travel). It is important to obtain high performance for automatic selection of synonyms using machine learning for the language of interest.

2. NAÏVE BAYES CLASSIFIER

2.1 Bayes' theorem

Let us consider an experiment denoted by E. We denote by S the sample space as the set of all possible outcomes of E. [17]

Definition 1

Let be A and B two events associated with an experiment E.

We denote by $P(B \mid A)$ the conditional probability of the event B, given that A has occurred.

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}, \text{ provided that } P(A) > 0$$

Definition 2

Events $B_1, B_2, ..., B_k$ represent a partition of the sample space S if

(a) $B_i \cap B_j = \varnothing$ for all $i \neq j$

(b) $\bigcup_{i=1}^{k} B_i = S$

(c) $P(B_i) > 0$ for all $i = \overline{1,k}$.

Definition 2 means that when the experiment E is performed one and only one of the events B_i occurs.

Let A be an arbitar event generated from S and let $B_1, B_2, ..., B_k$ the partition of S. Let $B_1 + B_2 + ... + B_k = S$, where $B_1, B_2, ..., B_k$ are mutually exclusive and exhaustive events. Each term $P(A \cap B_j)$ may be expressed as $P(A \mid B_j)P(B_j)$ and hence we obtain what is called the theorem on total probability:

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + ... + P(A \mid B_k)P(B_k) \quad (1)$$

Bayes' theorem

Let $B_1, B_2, ..., B_k$ be a partition of the sample space S and let A be an event associated with S. From the definition of conditional probability we obtain

$$P(B \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{k} P(A \mid B_j)P(B_j)} \quad (2)$$

The previous formula is called the formula for the probability of "causes". Bayes's theorem is the basis of statistical decision theory in some situations. [19]

Let us consider the event $B_i, i = \overline{1,k}$. The probability $P(B_i)$ is called prior probability, $P(B_i \mid A)$ is called posteriori probability and $P(A \mid B_i)$ is called the likelihood.

The Bayes' theorem provides a way of calculating the posterior probability, $P(B_i \mid A)$ if $P(B_i)$, $P(A)$ and $P(A \mid B_i)$ are known.

2.2 The classification problem

We consider a set of n classes $c_1, c_2, ..., c_n$. The problem is to determine which class(es) a given object belongs to. If this collection will increase, we must repeat the task but we want that repetitive task be automated. This process is called standing query, it is like any other query except that it is periodically executed on a collection to which new documents are incrementally added over time. [5]

If the standing query serves to divide the collection into the two classes, we refer to this as two-class classification.

Many systems support standing queries. When we use a classification with standing queries it is called routing or filtering.

2.3 Bayes classifiers

The Bayes classifiers are also called: Idiot Bayes, Naïve Bayes, Simple Bayes. The Bayes classifiers use Bayes' theorem.

The Naive Bayes classifier is a simple probabilistic classifier which is based on Bayes theorem with strong and naïve independence assumptions. It is one of the most basic text classification techniques with various applications.

Naive Bayes performs well in many complex real-world problems, even if it has a naïve design and oversimplified assumptions. This classifier is superior in terms of memory consumption and in several cases its performance is very close to more complicated and slower classification techniques. Overall, the Naive Bayes classifier is used as a baseline in many researches.

There are several types of Naïve Bayes classifiers: Multinomial Naive Bayes, Binarized Multinomial Naive Bayes and the Bernoulli Naive Bayes. Naïve Bayes and multinomial Naïve Bayes model are both supervised learning methods. They are also probabilistic learning methods.

Each type of Naïve Bayes classifiers can have as output different results since they use completely different models.

Multinomial Naive Bayes is used when the multiple occurrences of the words is very important in the classification problem. The Binarized Multinomial Naive Bayes is used when the frequencies of the words don't have a very important role in the classification. Bernoulli Naive Bayes can be used when is important the absence of a particular word matters. Bernoulli is usually used in spam detection and good results are obtained.

2.4 Text classification problem

Text classification is intended to assigning subjects to categories and can be used for spam detection, age or gender identification, language identification, etc.

In practice, it is possible to have more than two classes and the naïve Bayesian classifiers estimate the probability of class c_j generating instance d. Generally, the Naïve Bayes attributes have independent distributions. The assumption to have all attributes independent because of the meaning of the word naïve does not fit in real world situations. Though, the classifier works well in many practical situations.

A text classification definition can be: we have as input a document d, a fixed set of classes $C = \{c_1, c_2, ..., c_n\}$ and as output a predicted class $c \in C$ [16].

A method that we can use to predict the class c is using a Supervised Machine Learning Method. This means we have as input a document d, a fixed set of classes $C = \{c_1, c_2, ..., c_n\}$, a training set of m hand-labeled documents $(d_1, c_1), (d_2, c_2), ..., (d_m, c_m)$. The output will be a learned classifier $g : d \to c$.

We denote by X the document space. In text classification, we are given a description $d \in X$ of a document and a fixed set of classes $C = \{c_1, c_2, ..., c_n\}$. Classes are called categories or labels.

The Naive Bayes classifiers can be represented as this type of graph:

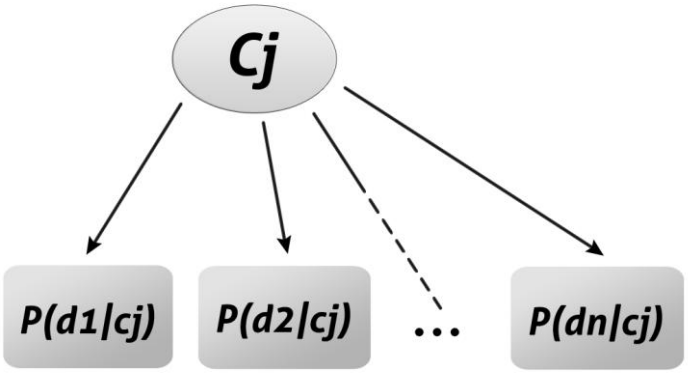


FIG. 1. Naïve Bayes classifier

The directions of the arrows indicate which state that each class causes certain features, with a certain probability.

In text classification, the goal is to find the best class for the document. The best class in Naïve Bayes classification is the most likely or MAximum a Posteriori (MAP) class noted with c_{MAP}. We refer at c_{MAP} as "MAximum a Posteriori"; the most likely class $c \in C$.

$$c_{MAP} = \arg\max P(c \mid d) = \frac{\arg\max(P(d \mid c)P(c))}{P(d)} = \frac{\arg\max (\mathrm{P}(\mathrm{t}_1, \mathrm{t}_2, \ldots, \mathrm{t}_n \mid \mathrm{c})\mathrm{P}(\mathrm{c}))}{P(d)} =$$

$$= \frac{\arg\max (\mathrm{P}(\mathrm{c})\mathrm{P}(\mathrm{t}_1 \mid \mathrm{c})\,\mathrm{P}(\mathrm{t}_2 \mid \mathrm{c}) \ldots \mathrm{P}(\mathrm{t}_n \mid \mathrm{c}))}{\mathrm{P}(\mathrm{d})} \tag{3}$$

There will be used only argmax(P(d|c)P(c) because it analyzes the same document d, which is test set. The document d is consisting of up of $t_k, k = \overline{1, n_d}$ terms, where n_d is the number of terms in document d. These t_k terms are tokens in document d [5].

The notation $P(t_n \mid c)$ represents the relative frequency of term t_n in document d belonging to class c. The situation when there is a term with zero frequencies is not possible and one use Laplace smoothing (add-1) for Naïve Bayes which adds one. This way one eliminates zeros.

We compare the calculated probabilities that the document belongs to a certain class and we choose the class with the higher probability.

3. APPLICATION - DOCUMENTS CLASSIFICATION WITH NAÏVE BAYES CLASSIFIER

In our application we use WEKA (Waikato Environment for Knowledge Analysis) that is a collection of machine learning algorithms for solving real-world data mining problems. Features of Weka are: machine learning, data mining, preprocessing, classification, regression, clustering, association rules, attribute selection, visualization [22].

WEKA software allows us to calculate the probability that a document belongs to a particular class, in which case we have the four categories of spam, sport, social media and travel but can not analyze the synonyms in a test set or training set for a foreign language.

The paper proposes to study the belonging of a document to a so-called class from two points of view, namely:

1. Pairs of words that are synonymous;
2. Words totally, without taking into account pairs of synonyms.

The scheme below suggests the proposed layout.

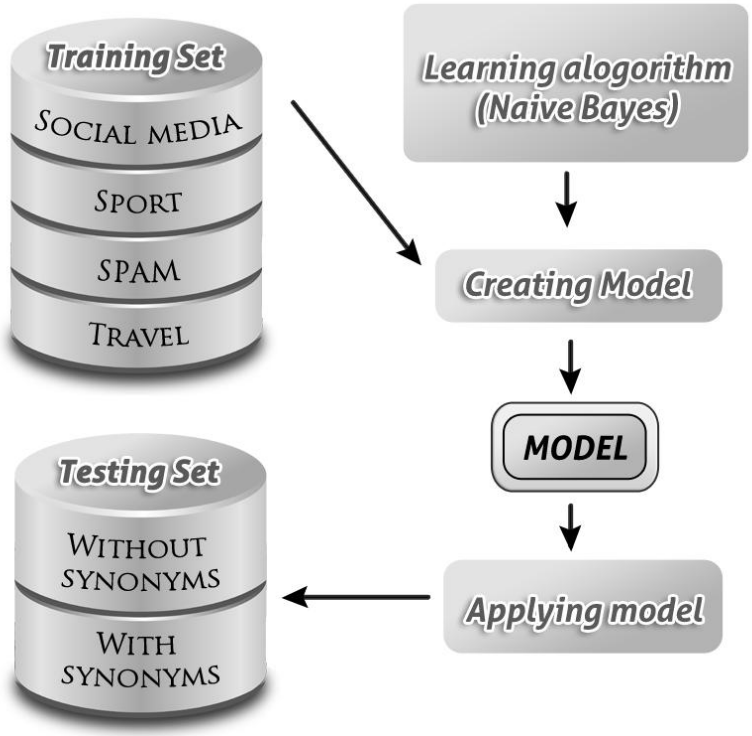


FIG. 2. Creating and testing the model

In our study we use the Naïve Bayes classifier from WEKA that has as output different files with results.

Our training dataset consists of 40 documents having 2131 words, respectively 10 documents for each category: sport, social media, spam and travel. We want to study which class the following document that includes synonyms (*holiday* and *vacation* belongs to: 'I want a great holiday and vacation'.

In WEKA we made the following settings: we have chosen Meta - Filtered Classifier and classifier: Naïve Bayes Multinomial, filter: StringToWordVector.

In *testHV* dataset *holiday* and *vacation* were considered distinct tokens and in *testH* and *testV* datasets they were considered as synonyms. The training set contains both tokens several times. We have obtained the results from the Table 1.

Table 1. Output for three datasets-Naïve Bayes

Test set	**Document content**	**Category**	**Prediction**	**Time (sec)**
testHV	I want a great *holiday* and *vacation*	Travel	0.429	0.02
testH	I want a great *holiday*	Travel	0.352	0.02
testV	I want a great *vacation*	Sport	0.515	0.01

In Table 2 we highlight the number of appearance of tokens holiday and vacation in training dataset.

Table 2. Number of appearance of words *holiday* and *vacation* in training dataset

Category	**Number of appearance in training dataset**	
	holiday	*vacation*
Sport	0	1
Social media	2	0
Spam	0	0
Travel	8	3

Weka output of probability of word given the class from the training dataset given by class is highlighted in Table 3. Table 3 contains only the number of occurrence of words in training set from the test datasets.

Table 3. Weka output: probability of word given by class

Word	**Sport**	**Social media**	**Spam**	**Travel**
I	0.00205	0.00071	0.00069	0.00067
want	0.00068	0.00213	0.00069	0.00135
a	0.00410	0.00498	0.00484	0.00610
great	0.00136	0.00071	0.00069	0.00067
holiday	0.0068	0.00142	0.00069	0.00203
and	0.00546	0.00570	0.00553	0.00542
vacation	0.00136	0.00071	0.00069	0.00013

In the *testH* document we have two synonyms *holiday* and *vacation*, and we have replaced all the occurrences of *vacation* with *holiday*. Applying the Naïve Bayes Multinomial classifier we have obtained that this document belongs to *Travel* category with prediction 0.352. This prediction is for *Travel* because the word *holiday* has the highest probability of word given the class in *Travel* (0.00203).

In the *testV* document we have have two synonyms *holiday* and *vacation*, and we have replaced all the occurrences of *holiday* with *vacation*. Following the application of Naïve Bayes Multinomial classifier we have obtained that this document belongs to the *Sport* category with prediction 0.515. This prediction is for *Sport* because the word *vacation* has the best probability of word given the class in *Sport* (0.00136).

The synonym selection in the dataset test is very important and should be made according to the significance of the information in the document, because in our training dataset we have both the words *holiday* and *vacation*. If we have replaced *vacation* with *holiday*, the *Travel* category has the highest probability of word given by the class, which has classified this document in *Travel*. Instead, when we replaced *holiday* with *vacation*, the probability of word given by class was for the *Sport* category.

Our proposal to do these two types of analysis is important if it is desired to classify documents in *Travel* or *Sport*, to take into account the meaning of synonyms for that language.

We want to emphasize that *holiday* and *vacation* in English language are synonymous, but have different meanings depending on the context: *vacation* means "planned time spent not working"; *holiday* means "celebration day or time off". In this analysis of the ambiguous intent document, the human factor must work to specify which synonyms will be used. Such an analysis is useful for books translations, newspaper articles and another related domains.

CONCLUSIONS

In this paper we highlighted some theoretical aspects regarding the text classification problem using the Naïve Bayes Multinomial classifier. In our application we uses WEKA software for modeling this problem.

We applied the Naïve Bayes Multinomial classifier on a training dataset containing a pair of words that are synonymous. We have studied the conditions for obtaining the highest prediction, taking into account the meaning of synonyms. From this point of view, our conclusion is that the human factor is decisive in choosing the proper synonyms.

REFERENCES

[1] C. Carstea, *Control and management in complex information systems*, Bulletin of the Transilvania University of Braşov, Series III Mathematics • Informatics • Physics, pp 73-87, 2013;
[2] C. Carstea, *Optimization Techniques in Project Controlling*, Ovidius University Annals Economic Sciences Series, 1: 428-432, 2013;
[3] C.G.Carstea, *Modeling System's Process for Control Of Complex Information Systems*, In Proceedings of The 25th International Business Information Management Association Conference (IBIMA), Soliman KS (ed). Amsterdam, Netherlands, May 2015, pp 566-574, 2015;
[4] C. Carstea, *Optimization Techniques in Project Controlling*, OVIDIUS UNIVERSITY ANNALS, ECONOMIC SCIENCES SERIES, Volume XIII Issue 1, pp. 428-432, Ovidius University Press, 2013;
[5] C.G. Carstea, *IT Project Management – Cost, Time and Quality*, ISSN 2067-5046, Economy Transdisciplinary Cognition International , Volume17, Issue 1/2014, pg.28-34, 2014;
[6] O. Florea, I. Rosca, *The Mechanical Behavior and the Mathematical Modeling of an Intervertebral Disc*, Acta Technica Napocensis Series-Applied Mathematics Mechanics and Engineering, 58: 213-218, 2015;
[7] O. Florea, I.C Rosca,, *Analytic study of a rolling sphere on a rough surface*, AIP ADVANCES, Volume: 6, Issue: 11, 2016;
[8] O. Florea, I.C. Rosca, *Stokes' Second Problem for a Micropolar Fluid with Slip*, PLOS ONE, Volume: 10, Issue: 7, Article Number: e0131860, 2015;
[9] A. Khashman, N.I. Nwulu, *Support vector machines versus back propagation algorithm for oil price prediction*, Advances in Neural Networks–ISNN 2011, pp 530-538, 2011;
[10] A. Khashman, *Blood Cell Identification using Emotional Neural Networks*, J. Inf. Sci. Eng. 25 (6), pp 1737-1751, 2009;
[11] A. Khashman, *IBCIS: Intelligent blood cell identification system*, Progress in Natural Science 18 (10), PP1309-1314, 2008;
[12] L. Mandru, *How to Control Risks? Towards A Structure of Enterprise Risk Management Process*, Journal of Public Administration, Finance and Law, pp 80-92, 2016;
[13] M. Popescu, L. Mandru**,** *Relationship between Quality Planning and Innovation*, Bulletin of the Transilvania University of Braşov, Series V, Economic Sciences, Vol. 9 (58) No. 2, pp.203-212, 2016;
[14] L. Mandru, D. Pauna, *Application of 'Small Steps Strategy' in the Management of European Companies*, Acta Universitatis Danubius Journal, Vol 8, no.3, October, 2012;
[15] L. Mandru, *Managementul integrat calitate-risc pentru societăţile comerciale cu profil industrial*, Editura Universităţii Transilvania, Brasov, 2011;

[16] C.D. Manning, P. Raghavan, H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008;
[17] P.L. Meyer, *Introductory probability and statistical applications. Second edition*, Addison Wesley Publishing Company, 1970;
[18] B.G. Munteanu, A. Leahu, S. Cataranciuc, *On the limit theorem for life time distribution connected with some reliability systems and their validation by means of the Monte Carlo method*, AIP Conference Proceedings 1557, pp. 582-588, 2013;
[19] H. Robbins, J. Ryzin, *Introduction to statistics*, Science research associates Inc., 1975;
[20] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Amsterdam, 1990;
[21] T. Wang, G. Hirst, *Exploring patterns in dictionary definitions for synonym extraction*, Natural Language Engineering, 18: 313–342, 2012;
[22] http:// http://www.cs.waikato.ac.nz/ml/weka/.

AN APPLICATION OF THE HIPERGEOMETRIC SCHEME

Cristian-George CONSTANTINESCU[*], Ion CIOACĂ[**]

[*]"Henri Coandă" Air Force Academy, Braşov, Romania
(c.g.constantinescu@gmail.com)
[**]Secondary School, Buneşti, Romania (puiu_cioaca2007@yahoo.com)

DOI: 10.19062/1842-9238.2017.15.2.12

***Abstract**: The purpose of this scientific approach is to report an error. We probably are not the first to notice it, but it's very possible that we are the first to make it public. This paper presents the problem VIII.8, from the book "Problems of mathematics for engineers", by Rodica Trandafir, with the author's solution. It is demonstrated that this is wrong and the right solution is proposed using hypergeometric scheme.*

***Keywords:** hypergeometric scheme, probabilistic approach, distribution, Mathcad*

1. INTRODUCTION

1.1 The hypergeometric scheme (of the urn with the unreturned ball):

In a urn there are *A* white balls and *B* black balls, from which *n* balls are extracted, $n < N := A + B$, one by one, without returning the drawn ball into the urn (or equivalently, all the *n* balls are removed simultaneously). The probability that among the *n* balls we have *a* white balls and $n - a := b$ balls is [1]:

$$P(a,b) = \frac{C_A^a C_B^b}{C_{A+B}^{a+b}}, \text{ where } \max\{0, A+n-N\} \le a \le \min\{A,n\} \quad (1)$$

1.2 Generalization:

If there are balls colored in *m* colors in the urn: A_1 balls of c_1 color, A_2 balls of c_2 color, ..., A_m balls of c_m color and *n* balls are extracted one by one without returning the ball drawn into the urn (or, equivalently, all the *n* balls are removed simultaneously), then the probability that form the *n* balls we have a_1 balls of c_1 color, a_2 balls of c_2 color,... a_m balls of c_m color,...where $a_1 + a_2 + \ldots + a_m = n$ is:

$$P(a_1, a_2, \ldots, a_m) = \frac{C_{A_1}^{a_1} C_{A_2}^{a_2} \ldots C_{A_m}^{a_m}}{C_{A_1+A_2+\ldots+A_m}^{a_1+a_2+\ldots+a_m}} \quad (2)$$

2. THE WRONG SOLUTION

The problem in discussion is formulated as follows: *A urn contains 36 white balls and 12 black balls. One person pulls the balls one by one until 10 white balls are obtained. Calculate the mean value of the number of extracted black balls* [2].

The solution proposed by the author is detailed below.

We are in the case of the unbalanced ball. The probability of removing *10* white balls and *k* black balls is:

$$P = \frac{C_{36}^{10} \cdot C_{12}^{k}}{C_{48}^{10+k}} \tag{3}$$

The mean value of the number of extracted black balls is:

$$M = C_{36}^{10} \cdot \sum_{k=0}^{12} k \frac{C_{12}^{k}}{C_{48}^{10+k}} \tag{4}$$

It is true that we are in the case of the unreturned ball, but the probability computed above is wrong. It respects the condition of having 10 white balls, but does not respect what is meant by the context: the last ball drawn out is white. Probability P also includes the solution that the last ball is black, or when the 10th white ball is extracted, the experience ends. For example, we calculated the sum of the probabilities associated with the random variable Ω that counts the black balls drawn from the urn, according to the hypothesis of the problem:

$$\Omega : \begin{pmatrix} 0 & 1 & 2 & 3 & \dots & 12 \\ \frac{C_{36}^{10} \cdot C_{12}^{0}}{C_{48}^{10+0}} & \frac{C_{36}^{10} \cdot C_{12}^{1}}{C_{48}^{10+1}} & \frac{C_{36}^{10} \cdot C_{12}^{2}}{C_{48}^{10+2}} & \frac{C_{36}^{10} \cdot C_{12}^{3}}{C_{48}^{10+3}} & \dots & \frac{C_{36}^{10} \cdot C_{12}^{12}}{C_{48}^{10+12}} \end{pmatrix} \tag{5}$$

$$\sum_{k=0}^{12} p_k = C_{36}^{10} \cdot \sum_{k=0}^{12} \frac{C_{12}^{k}}{C_{48}^{10+k}} = \frac{36}{10 \cdot 26} \cdot \sum_{k=0}^{12} \frac{\frac{12}{k! \cdot (12-k)!}}{\left(\frac{48}{(10+k)! \cdot (38-k)!} \right)} = 1.324 \tag{6}$$

3. THE CORRECT SOLUTION

Below the correct solution is presented. It can be observed that we have modified the data of the problem, which obviously does not influence the correctness of the solution:

In a urn we have 20 white balls and 80 black balls. Balls are extracted from the urn (without turning back the extracted ball) until 10 white balls are obtained. Calculate the mean and variance of the number of extracted black balls.

The correct solution is detailed below.

It is considered a urn that contains N_A white balls and N_B black balls, from which we extract the balls until n_A white balls are obtained, where $0 \le n_A \le N_A$

Let Ω be the variable describing the number of extracted black balls. It can be seen that ***the last ball*** drawn in the experience ***can only be white***. It follows that event A: $\Omega = k$ (we extracted *k* black balls to get the n_A white balls, so a total amount of $k + n_A$ balls were extracted from the urn) is written as the intersection of the following two events, obviously independent ones:

A_1: in the first $k + n_A - 1$ extractions $n_A - 1$ white balls and *k* black balls were obtained (regardless of their order);

A_2 : in the last extraction a white ball was obtained.

$$\left.\begin{array}{r} A = A_1 \cap A_2 \\ A_1, A_2 \text{ independente} \end{array}\right\} \Rightarrow P(A) = P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) = P(\Omega = k) \tag{7}$$

Event A_1 is described by the hypergeometric scheme associated with the urn containing N_A white balls and N_B black balls, from which we extract $k + n_A - 1$ balls, such that $n_A - 1$ balls are white and k balls are black.

If $N = N_A + N_B$, it results:

$$P(A_1) = \frac{C_{N_A}^{n_A - 1} \cdot C_{N_B}^{k}}{C_N^{n_A + k - 1}} \tag{8}$$

Event A_2 is described by the hypergeometric scheme associated with the urn with $N_A - n_A + 1$ white ball and $N_B - k$ black balls, from which we draw a ball and it must be white.

$$P(A_2) = \frac{C_{N_A - n_A + 1}^{1} \cdot C_{N_B - k}^{0}}{C_{N_A - k - n_A + 1}^{1}} = \frac{N_A - n_A + 1}{N_A - k - n_A + 1} \tag{9}$$

$$P(\Omega = k) = \frac{C_{N_A}^{n_A - 1} C_{N_B}^{k}}{C_N^{n_A + k - 1}} \cdot \frac{N_A - n_A + 1}{N - k - n_A + 1}; k = \overline{0, N_B} \tag{10}$$

In the proposed problem we have the particular case: $N_A = 20$; $N_B = 80$; $n_A = 10$. The following random variable Ω distribution array results:

$$\Omega : \begin{pmatrix} 0 & 1 & 2 & \dots & k & \dots & 80 \\ \frac{C_{20}^{9} C_{80}^{0}}{C_{100}^{9}} \cdot \frac{11}{91} & \frac{C_{20}^{9} C_{80}^{1}}{C_{100}^{10}} \cdot \frac{11}{90} & \frac{C_{20}^{9} C_{80}^{2}}{C_{100}^{11}} \cdot \frac{11}{89} & \dots & \frac{C_{20}^{9} C_{80}^{k}}{C_{100}^{n_A + k - 1}} \cdot \frac{11}{91 - k} & \dots & \frac{C_{20}^{9} C_{80}^{80}}{C_{100}^{89}} \cdot \frac{11}{11} \end{pmatrix} \tag{11}$$

It may be observed that in this case the sum of probabilities is equal to 1:

$$\sum p_k = \sum_{k=0}^{Nb} \frac{\frac{Na!}{(na-1)! \cdot (Na - na + 1)!} \cdot \frac{Nb!}{k! \cdot (Nb - k)!}}{\left(\frac{(Na + Nb)!}{(na + k - 1)! \cdot (Na + Nb - na - k + 1)!}\right)} \cdot \frac{Na - na + 1}{Na + Nb - k - na + 1} = 1 \tag{12}$$

$$Na := 80 \qquad Nb := 20 \qquad na := 10$$

$$m := M(\Omega) = 11 \cdot C_{20}^{9} \cdot \sum_{k=0}^{80} k \frac{C_{80}^{k}}{C_{100}^{9+k}(91 - k)} \tag{13}$$

$$m = 11 \cdot \frac{20}{9! \cdot 11} \cdot \sum_{k=0}^{80} \frac{k \cdot \frac{80}{k! \cdot (80-k)!}}{(91-k) \cdot \frac{100}{(9+k)! \cdot (91-k)!}} = 38.095 \tag{14}$$

$$M\left(\Omega^2\right) = 11 \cdot C_{20}^{9} \cdot \sum_{k=0}^{80} k^2 \frac{C_{80}^{k}}{C_{100}^{9+k}(91-k)} \tag{15}$$

$$M\left(\Omega^2\right) = 11 \cdot \frac{20}{9! \cdot 11} \cdot \sum_{k=0}^{80} \frac{k^2 \cdot \frac{80}{k! \cdot (80-k)!}}{(91-k) \cdot \frac{100}{(9+k)! \cdot (91-k)!}} = 1.543\,10^{3} \tag{16}$$

$$\sigma^2 := D^2(\Omega) = M\left(\Omega^2\right) - m^2 \tag{17}$$

$$\sigma^2 = 11 \cdot \frac{20}{9! \cdot 11} \cdot \left[\sum_{k=0}^{80} \frac{k^2 \cdot \frac{80}{k! \cdot (80-k)!}}{(91-k) \cdot \frac{100}{(9+k)! \cdot (91-k)!}} \right] - 11 \cdot \frac{20}{9! \cdot 11} \cdot \left[\sum_{k=0}^{80} \frac{k \cdot \frac{80}{k! \cdot (80-k)!}}{(91-k) \cdot \frac{100}{(9+k)! \cdot (91-k)!}} \right]^2 = 91.61 \tag{18}$$

In formulas 12-18 the calculations were performed by using the Mathcad program.

CONCLUSIONS

Different extensions of this problem can be formulated and some of them can be reducible to the case of the two-color urn. One of them might be the following: *In a urn we have N_1 balls of c_1 color, N_2 balls of c_2 color, ... and N_m balls of c_m color. Balls from the urn (without returning the extracted ball) are extracted until n_p balls of c_p color are obtained. Calculate the average and dispersion of the number of other color balls extracted.* The problem is reduced to that presented at point 3, considering the urn with N_p balls of c_p color and $N - N_p$ balls of another color.

Another observation that can be made: the probabilistic approach is a method of calculating sums with combinations. For example, this problem has shown that:

$$\sum_{k=0}^{N_B} \frac{C_{N_A}^{n_A-1} C_{N_B}^{k}}{C_{N_A+N_B}^{n_A+k-1}} \cdot \frac{N_A - n_A + 1}{N - k - n_A + 1} = C_{N_A}^{n_A-1}(N_A - n_A + 1) \sum_{k=0}^{N_B} \frac{C_{N_B}^{k}}{C_{N_A+N_B}^{n_A+k-1}} \cdot \frac{1}{N - k - n_A + 1} = 1 \tag{19}$$

REFERENCES

[1] Gh. Mihoc, N. Micu, *Probability Theory and Mathematical Statistics*, Didactic and Pedagogical Publishing House, 1980.

[2] R. Trandafir, *Problems of Mathematics for Engineers*, Technical Publishing House, 1977.

A GENERALIZATION OF DVORETZKY'S STOCHASTIC APPROXIMATION THEOREM

Neculai CRÎŞMARU

"G. Bacovia" University, Bacau, Romania (crismaru_nicolae@yahoo.com)

DOI: 10.19062/1842-9238.2017.15.2.13

Abstract: *This presentation contains a generalization of one of Dvoretzk's well-known stochastic approximation theorem. This generalization is founded on the extension of one of Derman and Sacks's lemmas wich was the basis of a result given by the two authors in relation to Dvoretzky's stochastic approximation theorem.*

Keywords: *stochastic approximation theorems, Dvoretzky' theorem, Derman-Sacks lema.*

1. INTRODUCTION

For some time, the stochastic approximation has become a subject of interest for many researchers in different fundamental and applied fields. There are two famous algorithms wich are very often utilized in the research of stochastic approximations (among other algorithms): the Robbins-Monro algorithm ([5], [6]) and the Kiefer-Wolfowitz algorithm ([4]). A short time after these stochastic approximation algorithms appeared, A. Dvoretzky ([3]) discovered an approximation algorithm wich generalizesthe two previously mentioned algorithms. As a result of the particular interest generated by this new algorithm, many researchers have started studying it. Two of these researchers are worth mentioning: Derman and Sacks ([2]), who found a demonstration (based on a technical lemma) wich is shorter and easier than that Dvoretzky's initial theorem, wich has a rather lengthy and more difficult to apprehend demonstration. This presentation contains a generalization of Dvoretzky's initial theorem realized by means of a generalization (among many others) of Derman and Sacks's lemma. Firstly, we will provide Dvoretzky's initial theorem without its demonstration and Derman and Sacks's lemma for a better understanding of the starting point of this research.

Dvoretzky's Theorem ([3], pp. 39-55)

Consider a probability space (Ω, K, P).

Let be $(\alpha_n)_{n\in N^*}, (\beta_n)_{n\in N^*}, (\gamma_n)_{n\in N^*}$ such that $\alpha_n > 0$, $\beta_n \geq 0$, $\gamma_n \geq 0$,

$\forall n \in N^*$, $\alpha_n \xrightarrow[n\to+\infty]{} 0$, $\sum_{n=1}^{+\infty} \beta_n < +\infty$, $\sum_{n=1}^{+\infty} \gamma_n = +\infty$. Let be θ a real number and $(T_n)_{n\in N^*}$ a sequence of real and measurable functions wich satisfies:

$|T_n(r_1, r_2, \ldots, r_n) - \theta| \leq \max\{\alpha_n; (1+\beta_n)|r_n - \theta| - \gamma_n\}$, $\forall n \in (N^* - \{1\})$ for every real numbers $r_1, r_2, \ldots, r_n$.

Let be the sequences of random variables: $(X_n)_{n\in N^*}$, $(Y_n)_{n\in N^*}$, such that

$X_{n+1}(\omega) = T_n(X_1(\omega), X_2(\omega) \ldots, X_n(\omega)) + Y_n(\omega)$ (a.s.), $\forall \omega \in \Omega$,

$\forall n \in (N^* - \{1\})$.

Suppose that the following conditions hold:

$E[Y_1^2]<+\infty$, $\sum_{n=1}^{+\infty} E[Y_n^2]<+\infty$, and $E[Y_n \mid X_1, X_2, \ldots, X_n]=0$ (a.s.), $\forall n \in N^*$. In the conditons above there exist:

$E[(X_n - \theta)^2] \xrightarrow[n\to+\infty]{} 0$ with the probability one, and

$X_n \xrightarrow[n\to+\infty]{} \theta$ with the probabilty one.

Lema (Derman şi Sacks)([2])

Let be $(a_n)_{n\in N^*}, (b_n)_{n\in N^*}, (c_n)_{n\in N^*}, (\delta_n)_{n\in N^*}, (x_n)_{n\in N^*}$,sequences of real numbers wich satisfy the following conditions:

$(a_n)_{n\in N^*}, (b_n)_{n\in N^*}, (c_n)_{n\in N^*}, (x_n)_{n\in N^*} \subset [0,+\infty)$, $\lim_{n\to+\infty}(a_n)=0$; $\sum_{n=1}^{+\infty} b_n<+\infty$;

$\sum_{n=1}^{+\infty} c_n=+\infty$; $\sum_{n=1}^{+\infty} \delta_n<+\infty$ şi $\exists\, N_0 \in N^*$, such that for every $n> N_0$ we have:

$x_{n+1} \le \max\{a_n; (1+b_n)x_n + \delta_n - c_n\}$. Hence $\lim_{n\to+\infty}(x_n)=0$.

2. THE MAIN RESULT

Before we provide the generalization of Dvoretzky's theorem, we wil offer a series of helping lemmas, some of wich are generalization of Derman and Sack's lemma.

Lemma 1.

Let be the sequences of real number: $(a_n)_{n\in N^*}, (b_n)_{n\in N^*}, (c_n)_{n\in N^*}, (d_n)_{n\in N^*}$, $(e_n)_{n\in N^*}$ $(\delta_n)_{n\in N^*}, (x_n)_{n\in N^*}$, such that we have:

$0 \le x_{n+2} \le \max\{a_{n+1}; (1+b_{n+1})x_{n+1} + e_n x_n + d_n\}, \forall n \in N^*,$ (1)

$a_n, b_n, c_n, e_n, x_n \ge 0, \delta_n, a_n \le 1, \forall n \in N^*,$ (2)

$a_n \to 0$, $\sum_{n=1}^{+\infty} b_n<+\infty$, $\sum_{n=1}^{+\infty} e_n<+\infty$, $\sum_{n=1}^{+\infty} c_n=+\infty$, $\sum_{n=1}^{+\infty} \delta_n<+\infty$, and

$d_n = \delta_n - c_n, \forall n \in N^*, d_n \in R,$ (3)

$|\delta_n| \le c_n, \forall n \in N^*, x_1, x_2 \in R$, such that ($x_2 \ge 1$ or $x_1 \ge 1$), (4)

note that: $K=\max\{x_1, x_2\}$, (5)

suppose that the following conditions holds: $b_n + e_{n-1} \le \frac{-d_{n-1}}{K}$, (6)

$\forall n \in (N^*-\{1\})$.

Hence $(x_n)_{n\in N^*}$ isbounded and $0 \le x_n \le K$.

PROOF: From (3) and (4) it result that $d_n = \delta_n - c_n \le |\delta_n| - c_n \le 0$, so we have:

$d_n \le 0, \forall n \in N^*$ (7)

From (5) it result $K \ge 1$ (8)

For n=1 from (1) we obtain:

$0 \le x_3 \le \max\{a_2; (1+b_2)x_2 + e_1 x_1 + d_1\},$ (9)

We have the inequality :

$(1+b_2)x_2 + e_1 x_1 + d_1 \le (1+b_2)K + e_1 K + d_1 \le K(1+b_2+e_1+\frac{d_1}{K})$ (10)

For n=2 from (6) we obtain: $b_2+e_1+\frac{d_1}{K}\le 0$ and from (10) we have:

$$(1+b_2)x_2+e_1x_1+d_1\le K. \tag{11}$$

From (2), (8), (9) and (11) it result: $0\le x_3\le \max\{a_2; K\}=K$, so we have:

$$0\le x_3\le K \tag{12}$$

By induction we have:

$$(1+b_{n+1})x_{n+1}+e_nx_n+d_n\le K(1+b_{n+1}+e_n+\frac{d_n}{K}),\forall n\in N^*, \tag{13}$$

in (6) we replace n with (n+1) and we obtain:

$b_{n+1}+e_n+\frac{d_n}{K}\le 0$ and from (13) it result: $(1+b_{n+1})x_{n+1}+e_nx_n+d_n\le K$, and from (1) and (2) we obtain:

$0\le x_{n+2}\le \max\{a_{n+1}; K\}=K$, so $0\le x_n\le K$, $\forall n\in N^*$, so

$(x_n)_{n\in N^*}$ is bounded and $0\le x_n\le K$ (q.e.d.).

Lemma 2.

Let be the sequences $(a_n)_{n\in N^*}$, $(b_n)_{n\in N^*}$, $(c_n)_{n\in N^*}$, $(d_n)_{n\in N^*}$, $(e_n)_{n\in N^*}$, $(\delta_n)_{n\in N^*}$ and $(x_n)_{n\in N^*}$, such that we have:

$$0\le x_{n+2}\le \max\{a_{n+1};(1+b_{n+1})x_{n+1}+e_nx_n+d_n\},\forall n\in N^*, \tag{1}$$

$$a_n, b_n, c_n, e_n, x_n\ge 0,\ \delta_n, a_n\le 1,\ \forall n\in N^*, \tag{2}$$

$$a_n\to 0, \sum_{n=1}^{+\infty} b_n<+\infty, \sum_{n=1}^{+\infty} e_n<+\infty, \sum_{n=1}^{+\infty} c_n=+\infty, \sum_{n=1}^{+\infty} \delta_n<+\infty, \tag{3}$$

$$d_n=\delta_n-c_n, \forall n\in N^*, d_n\in R, \tag{4}$$

$$|\delta_n|\le c_n, \forall n\in N^*, \tag{5}$$

$x_1, x_2\in R$, such that ($x_2\ge 1$ or $x_1\ge 1$), and we note that:

$$K=\max\{x_1,x_2\}, \tag{6}$$

Suppose that the following holds: $b_n+e_{n-1}\le \frac{-d_{n-1}}{K}, \forall n\in(N^*-\{1\})$. (7)

Hence $x_n \xrightarrow[n\to+\infty]{} 0$.

PROOF: From the previous lemma, (lemma 1.) we have that the sequence $(x_n)_{n\in N^*}$ is bounded and

$0\le x_n\le K$ and from (1) we have:

$$0\le x_{n+2}\le \max\{a_{n+1}; (1+b_{n+1})x_{n+1}+e_nK+d_n\},\forall n\in N^*, \tag{8}$$

Note that:

$$e_nK+d_n=\tilde{d}_n \tag{9}$$

so we have $0\le x_{n+2}\le \max\{a_{n+1}; (1+b_{n+1})x_{n+1}+\tilde{d}_n\},\forall n\in N^*$, and from (4) we have:

$$\text{with } \tilde{d}_n=e_nK+\delta_n-c_n,\ \forall n\in N^*, \tag{10}$$

Facem acum altă notaţie:

$$e_nK+\delta_n=\tilde{\delta}_n \tag{11}$$

Using these notes (10) becomes:

$$0 \le x_{n+2} \le \max\{a_{n+1}; (1+b_{n+1})x_{n+1} + \tilde{\delta}_n - c_n\}, \forall n \in N^*, \text{ with } \tilde{d}_n = \tilde{\delta}_n - c_n, \quad (12)$$

$\forall n \in N^*$

and from (3) it result:

$$\sum_{n=1}^{+\infty} \tilde{\delta}_n = \sum_{n=1}^{+\infty} (e_n K + \delta_n) = K \sum_{n=1}^{+\infty} e_n - \sum_{n=1}^{+\infty} \delta_n < +\infty, \quad (13)$$

At this stage all the hypotheses of Derman and Sacks's lemma ([2], pp. 602) or of lemma 1. in this paper, so $x_n \to 0$ (q.e.d.).

Lemma 3.

Let be the real sequences $(a_n)_{n\in N^*}$, $(b_n)_{n\in N^*}$, $(c_n)_{n\in N^*}$, $(d_n)_{n\in N^*}$, $(e_n)_{n\in N^*}$, $(\delta_n)_{n\in N^*}$ $(x_n)_{n\in N^*}$, such that we have:

$$0 \le x_{n+2} \le \max\{a_{n+1}; (1+b_{n+1})x_{n+1} + e_n x_n + d_n\}, \forall n \in N^*, \quad (1)$$

$$a_n, b_n, c_n, e_n, x_n \ge 0, \delta_n, a_n \le 1, \forall n \in N^*, \quad (2)$$

$$a_n \to 0, \sum_{n=1}^{+\infty} b_n < +\infty, \sum_{n=1}^{+\infty} e_n < +\infty, \sum_{n=1}^{+\infty} c_n = +\infty, \sum_{n=1}^{+\infty} \delta_n < +\infty, \quad (3)$$

$$d_n = \delta_n - c_n, \forall n \in N^*, d_n \in R, \quad (4)$$

$$(x_n)_{n\in N^*} \text{ is bounded, } 0 \le x_n \le K, \forall n \in N^*. \quad (5)$$

Hence $x_n \xrightarrow[n\to+\infty]{} 0$.

PROOF: From (1) and (5) we have: $0 \le x_{n+2} \le \max\{a_{n+1}; (1+b_{n+1})x_{n+1} + e_n K + d_n\}$, $\forall n \in N^*$,

Note that: $e_n K + d_n = \tilde{d}_n$ (6)

so we have $0 \le x_{n+2} \le \max\{a_{n+1}; (1+b_{n+1})x_{n+1} + \tilde{d}_n\}, \forall n \in N^*$,

with $\tilde{d}_n = e_n K + \delta_n - c_n, \forall n \in N^*$, (7)

we add another note: $e_n K + \delta_n = \tilde{\delta}_n$ (8)

Using these notes (7) becomes:

$$0 \le x_{n+2} \le \max\{a_{n+1}; (1+b_{n+1})x_{n+1} + \tilde{\delta}_n - c_n\}, \forall n \in N^*, \quad (9)$$

with $\tilde{d}_n = \tilde{\delta}_n - c_n, \forall n \in N^*$, in wich, from (3) results that:

$$\sum_{n=1}^{+\infty} \tilde{\delta}_n = \sum_{n=1}^{+\infty} (e_n K + \delta_n) = K \sum_{n=1}^{+\infty} e_n - \sum_{n=1}^{+\infty} \delta_n < +\infty.$$

At this stage all hypotheses of Derman and Sack's lemma hold ([2], pag. 602), so $x_n \to 0$ (q.e.d.).

Lemma 4.

Let be the real sequences $(a_n)_{n\in N^*}$, $(b_n)_{n\in N^*}$, $(c_n)_{n\in N^*}$, $(d_n)_{n\in N^*}$, $(e_n)_{n\in N^*}$, $(\delta_n)_{n\in N^*}$ and $(x_n)_{n\in N^*}$, such that we have (for a chosen and fixed K):

1) $(x_n)_{n\in N^*}$ is bounded, $0 \le x_n \le K, \forall n \in N^*, K>0$,

2) $0 \le x_{n+2} \le \min\{\max\{a_{n+1}; (1+b_{n+1})x_{n+1} + e_n x_n + d_n\}; K\}, \forall n \in N^*$,

3) $a_n, b_n, c_n, e_n, x_n \ge 0, \delta_n, a_n \le 1, \forall n \in N^*$,

4) $a_n \to 0$, $\sum_{n=1}^{+\infty} b_n <+\infty$, $\sum_{n=1}^{+\infty} e_n <+\infty$, $\sum_{n=1}^{+\infty} c_n =+\infty$, $\sum_{n=1}^{+\infty} \delta_n <+\infty$,

5) $d_n = \delta_n - c_n$, $\forall n \in N^*$, $d_n \in R$.

Hence $x_n \xrightarrow[n\to+\infty]{} 0$.

PROOF: See lemma 3.

We now provide the main result of this presentation, result wich is a generalization of one of Dvoretzky's stochastic approximation theorem.

Theorem

Let be sequences $(X_n)_{n\in N^*}, (T_n)_{n\in N^*}, (Y_n)_{n\in N^*}$, such that the following conditions hold : X_n, Y_n are random variables on a probability space (Ω, K, P),

and X_i are random variables, $i \in \{1,2\}$, and $T_n : R^n \to R$, X_1, X_2 are measurables applications such that (X_1 or X_2) ≥ 1 (a.s.), and the following conditions hold:

$$X_{n+1} = T_n(X_1, X_2, \ldots, X_n) + Y_n \text{ (a.s.)} \quad (1)$$

$$E[Y_n \mid X_1, X_2, \ldots, X_n] = 0 \text{ (a.s.)}, \ \forall n \in N^*, \quad (2)$$

$$\sum_{n=1}^{+\infty} E[Y_n^2] <+\infty, \quad (3)$$

$$|T_n(x_1, x_2, \ldots, x_n)| \leq \max\{\alpha_n; (1+\beta_n)|x_n| + e_{n-1}|x_{n-1}| - \gamma_n\}, \quad (4)$$

$\forall n \in (N^* - \{1\})$.

Let be the sequences $(\alpha_n)_{n\in N^*}, (\beta_n)_{n\in N^*}, (\gamma_n)_{n\in N^*}, (e_n)_{n\in N^*}$, such that:

$\alpha_n \xrightarrow[n\to+\infty]{} 0$, $\sum_{n=1}^{+\infty} \beta_n <+\infty$, $\sum_{n=1}^{+\infty} e_n <+\infty$, $\sum_{n=1}^{+\infty} \gamma_n =+\infty$, and

$$\alpha_n > 0, \ \beta_n \geq 0, \ \gamma_n \geq 0, \ e_n \geq 0, \ \forall n \in N^*, \ \alpha_n \leq 1, \quad (5)$$

$$|Y_n| \leq \gamma_n \text{ (a.s.). Note that } K = \max\{X_1, X_2\}, \quad (6)$$

and suppose that:

$$\beta_n + e_n \leq \frac{1}{K}(|Y_{n-1}| - \gamma_n) \text{ (a.s.)}, \ \forall n \in (N^* - \{1\}). \quad (7)$$

So we have:

a) $(X_n)_{n\in N^*}$ is bounded (a.s.), $|X_n| \leq K$, (a.s.) $\forall n \in (N^* - \{1\}$ and

b) $X_n \xrightarrow[n\to+\infty]{} 0$ (a.s.).

PROOF:

As in the demonstration of theorem 1. from ([2], pp. 603), we suppose that:

$$\sum_{n=1}^{+\infty} \frac{1}{\alpha_n^2} E[Y_n^2] <+\infty \quad (8)$$

We define the random variables:

$Z_n = \text{sign}(T_n(X_1, X_2, \ldots, X_n))\, Y_n$

Hence for Z_n (2), (3) and (4) hold.

But from (2) and (3) we have (as in [2], pp. 603):

$$\sum_{n=1}^{+\infty} Z_n = \text{convergent (a.s.) and } |Z_n| = |Y_n| \quad (9)$$

From (8), the Cebisev inequality and the Borell-Cantelli lemma we have:

$$|Z_n| \le \alpha_n \quad \text{(a.s.)} \tag{10}$$

But from (1) and (10) we have:

$$|X_{n+1}| \le 2\alpha_n \text{, dacă } |T_n| \le \alpha_n \text{ and } |X_{n+1}| = |T_n| + Z_n \text{, if } |T_n| > \alpha_n \tag{11}$$

So we have (a.s.):

$$0 \le |X_{n+1}| \le \max\{2\alpha_n; |T_n| + Z_n\} \le$$
$$\le \max\{2\alpha_n; (1+\beta_n)|X_n| + e_{n-1}|X_{n-1}| + Z_n - \gamma_n\} \tag{12}$$

If we note:

$|X_n| = x_n$, $a_n = 2\alpha_n$, $b_n = \beta_n$, $c_n = \gamma_n$, $\delta_n = Z_n$, $d_n = \delta_n - c_n$,

then all the conditions from the hypoteses in lemma 1. hold, we have:

$(|X_n|)_{n \in N^*}$ is bounded and

$$|X_n| \le K, \ \forall n \in N^* \tag{13}$$

But from (12) and (13) we have:

$$0 \le |X_{n+1}| \le \max\{2\alpha_n; (1+\beta_n)|X_n| + e_{n-1}K + Z_n - \gamma_n\} = \tag{14}$$
$$= \max\{2\alpha_n; (1+\beta_n)|X_n| + (e_{n-1}K + Z_n) - \gamma_n\}$$

But $\sum_{n=2}^{+\infty} (e_{n-1}K + Z_n) = K\sum_{n=2}^{+\infty} e_{n-1} + \sum_{n=2}^{+\infty} Z_n < +\infty$ (see. (9)).

If we note:

$$a_n = 2\alpha_n, \ b_n = \beta_n, \ c_n = \gamma_n, \ \delta_n = Ke_n + Z_n, \ x_n = |X_n| \tag{15}$$

then, using the notes from (15) it results that all the conditions from the hypotheses in lemma 2. (or in lemma 1 of Derman and Sacks from [2], pp. 602), so we have:

$x_n \xrightarrow[n \to +\infty]{} 0$, that is $|X_n| \xrightarrow[n \to +\infty]{} 0$ (a.s.), so $X_n \xrightarrow[n \to +\infty]{} 0$ (a.s.) (q.e.d.).

REFERENCES

[1] Chen, Han-Fu. , *Stochastic Approximation and Its Applications,* Kluwer Academic Publishers, 2003, pp. 5;

[2] Derman, C. and Sacks, J. "*On Dvoretzky's stochastic approximation theorem"*. Anals of Math. Stat. (1959) 30 pp. 601-605;

[3] Dvoretzky, A. (1956). *On Stochastic Approximation*. Proc. Third Berkeley Symp. Math. Stat. Prob. 1 pp.39-55;

[4] Kiefer J. and Wolfowitz J. " *Stochastic estimation of the maximum of a regression function* ", Annals of Mth. Stat. , vol. 33(1952), pp. 462-466;

[5] Orman G. V. , *Handboock of Limit Theorems and Stochastic Approximation*, "Transilvania" University Press, Braşov, 2003;

[6] Robbins, H. , Monro, S. ,*A Stochastic Approximation Method* . Anals of Math. Stat. Vol. 22. (1951), pp. 400-407.

CUMULATIVE ENTROPIES: A SURVEY

Vasile PREDA[*], Costel BĂLCĂU[*], Doru CONSTANTIN[*], Ioana Ileana PANAIT[**]

[*]Department of Mathematics, University of Bucharest, Romania (preda@fmi.unibuc.ro, costel.balcau@upit.ro, doru.constantin@upit.ro)
[**]Doctoral School of Mathematics, University of Bucharest, Romania (yoana_panait@yahoo.com)

DOI: 10.19062/1842-9238.2017.15.2.14

Abstract: *We present a review of cumulative entropies from reliability theory.*
2010 Mathematics Subject Classification: *Primary 94A17.*

Keywords*: cumulative entropy, residual entropy, past entropy, interval entropy.*

1. INTRODUCTION

Consider X a non-negative absolutely continuous random variable with the probability density function (PDF) $f(x)$, the cumulative distribution function (CDF) $F(x) = P(X \le x)$, and the reliability function (RF) $\overline{F}(x) = P(X > x)$.

Let $t \in [0,\infty)$ and w a real nonnegative measurable function defined on $[0,\infty)$.

Let $D = \{(t_1,t_2) \mid 0 \le t_1 < t_2 \le \infty, F(t_1) < F(t_2)\}$.

For any given pair $(t_1,t_2) \in D$, the conditional random variable $(X \;\; t_1 \le X \le t_2)$ has the PDF $f(x;t_1,t_2)$, the CDF $F(x;t_1,t_2)$ and the RF $\overline{F}(x;t_1,t_2)$ given by

$$f(x;t_1,t_2) = \frac{f(x)}{F(t_2)-F(t_1)},\ F(x;t_1,t_2) = \frac{F(x)}{F(t_2)-F(t_1)},$$
$$\overline{F}(x;t_1,t_2) = \frac{\overline{F}(x)}{F(t_2)-F(t_1)} = \frac{\overline{F}(x)}{\overline{F}(t_1)-\overline{F}(t_2)},\ \text{for all}\ t_1 \le x \le t_2. \tag{1}$$

The **Shannon entropy** proposed by Shannon [45] (1948) is defined as

$$H(X) = -\int_0^\infty f(x)\ln f(x)dx. \tag{2}$$

Remark 1.1 *In all definition from this paper we assume that the integrals exist, and we use the notational convention* $0\ln 0 = 0$ *and* $0/0 = 0$.

The concept of entropy was proposed as a measure of the amount of information supplied by a random variable X or a probabilistic experiment. It has numerous extensions to other entropy-type measures, some of these will be presented below.

In the next sections we present the residual and past entropy-types measures based on CDF and RF, the cumulative entropies, the cumulative relative entropies and inaccuracy measures.

2. WEIGHTED, RESIDUAL AND PAST ENTROPIES

The **weighted entropy**, referred by Di Crescenzo and Longobardi [9] (2006) in agreement with Belis and Guiasu [4] (1968), is defined as

$$H_w(X) = -\int_0^\infty w(x) f(x) \ln f(x) dx. \tag{3}$$

The **residual entropy (RE)** proposed by Ebrahimi and Pellerey [13] (1996) is defined as

$$RE(X;t) = -\int_t^\infty f(x;t,\infty) \ln f(x;t,\infty) dx = -\int_t^\infty \frac{f(x)}{\overline{F}(t)} \ln \frac{f(x)}{\overline{F}(t)} dx. \tag{4}$$

The **past entropy (PE)** proposed by Di Crescenzo and Longobardi [7] (2002) is defined as

$$PE(X;t) = -\int_0^t f(x;0,t) \ln f(x;0,t) dx = -\int_0^t \frac{f(x)}{F(t)} \ln \frac{f(x)}{F(t)} dx. \tag{5}$$

The **weighted residual entropy (WRE)** proposed by Di Crescenzo and Longobardi [9] (2006) is defined as

$$WRE(X;t) = -\int_t^\infty x\ f(x;t,\infty) \ln f(x;t,\infty) dx = -\int_t^\infty x \frac{f(x)}{\overline{F}(t)} \ln \frac{f(x)}{\overline{F}(t)} dx. \tag{6}$$

The **weighted past entropy (WPE)** proposed by Di Crescenzo and Longobardi [9] (2006) is defined as

$$WPE(X;t) = -\int_0^t x\ f(x;0,t) \ln f(x;0,t) dx = -\int_0^t x \frac{f(x)}{F(t)} \ln \frac{f(x)}{F(t)} dx. \tag{7}$$

These weighted entropies are suitable to describe dynamic information of random lifetimes, in analogy with the entropies of residual and past lifetimes introduced in [13] and [7], respectively.

3. CUMULATIVE ENTROPIES

The **cumulative residual entropy (CRE)** proposed by Rao, Chen, Vemuri, Wang [41] (2004) is defined as

$$CRE(X) = -\int_0^\infty \overline{F}(x) \ln \overline{F}(x) dx. \tag{8}$$

The CRE is an alternative measure of uncertainty in the random variable X that enjoys many of the properties of Shannon entropy and has some advantages such us is always non-negative, can be easily computed from sample data and these computations asymptotically converge to the true values.

Like as the Shannon entropy the CRE can be used to construct probability distributions by applying the Maximum Entropy Principle introduced in 1957 by Jaynes [18, 19]. For example, Rao [42] (2005) obtains a general result for characterization of MAX-CRE distributions and applies this result to construct the uniform distribution and the Weibull distribution.

Drissi, Chonavel and Boucher [12] (2008) generalize the definition of CRE to the case of random variables with supports that are not restricted to positive values.

The **dynamic cumulative residual entropy (DCRE)** proposed by Asadi and Zohrevand [2] (2007) is defined as

$$DCRE(X;t) = -\int_t^\infty \overline{F}(x;t,\infty)\ln \overline{F}(x;t,\infty)dx = -\int_t^\infty \frac{\overline{F}(x)}{\overline{F}(t)}\ln\frac{\overline{F}(x)}{\overline{F}(t)}dx. \tag{9}$$

The DCRE is a measure of the information in the residual life distribution. The authors show that the CRE and the DCRE is connected with some well-known reliability measures such as the mean residual lifetime and the hazard rate. Also, they prove that if the $DCRE(X;t)$ is an non-decreasing function on t then it characterizes the underlying distribution function uniquely.

The **cumulative past entropy (CPE)** proposed by Di Crescenzo and Longobardi [10] (2009) is defined as

$$CPE(X) = -\int_0^\infty F(x)\ln F(x)dx. \tag{10}$$

The CPE is also non-negative and it is useful to measure information on the inactivity time of a system, being appropriate for the systems whose uncertainty is related to the past.

The **dynamic cumulative past entropy (DCPE)** proposed by Di Crescenzo and Longobardi [10] (2009) and by Navarro, Del Aguila and Asadi [37] (2010) is defined as

$$DCPE(X;t) = -\int_0^t F(x;0,t)\ln F(x;0,t)dx = -\int_0^t \frac{F(x)}{F(t)}\ln\frac{F(x)}{F(t)}dx. \tag{11}$$

The **interval entropy (IH)** proposed by Sunoj, Sankaran and Maya [47] (2009) is defined as

$$IH(X;t_1,t_2) = -\int_{t_1}^{t_2} f(x;t_1,t_2)\ln f(x;t_1,t_2)dx = -\int_{t_1}^{t_2} \frac{f(x)}{F(t_2)-F(t_1)}\ln\frac{f(x)}{F(t_2)-F(t_1)}dx. \tag{12}$$

The **weighted cumulative residual entropies (WCRE)** proposed by Misagh, Panahi, Yari, Shahi [33] (2011) is defined as

$$WCRE(X) = -\int_0^\infty x\,\overline{F}(x)\ln \overline{F}(x)dx. \tag{13}$$

The **weighted cumulative past entropies (WCPE)** proposed by Misagh, Panahi, Yari, Shahi [33] (2011) is defined as

$$WCPE(X) = -\int_0^\infty x\,F(x)\ln F(x)dx. \tag{14}$$

The authors present various properties of this measure, including its connection with weighted residual and past entropies and obtain some upper and lower bounds.

The **interval cumulative residual entropies (ICRE)** proposed by Khorashadizadeh, Rezaei Roknabadi and Mohtashami Borzadaran [25] (2013) is defined as

$$ICRE(X;t_1,t_2) = -\int_{t_1}^{t_2} \overline{F}(x;t_1,t_2) \ln \overline{F}(x;t_1,t_2) dx = -\int_{t_1}^{t_2} \frac{\overline{F}(x)}{\overline{F}(t_1)-\overline{F}(t_2)} \ln \frac{\overline{F}(x)}{\overline{F}(t_1)-\overline{F}(t_2)} dx. \quad (15)$$

The **interval cumulative past entropies (ICPE)** proposed by Khorashadizadeh, Rezaei Roknabadi and Mohtashami Borzadaran [25] (2013) is defined as

$$ICPE(X;t_1,t_2) = -\int_{t_1}^{t_2} F(x;t_1,t_2) \ln F(x;t_1,t_2) dx = -\int_{t_1}^{t_2} \frac{F(x)}{F(t_2)-F(t_1)} \ln \frac{F(x)}{F(t_2)-F(t_1)} dx. \quad (16)$$

The authors present some properties and characterization of this measures, including its connections with doubly truncated Shannon entropy and mean residual life.

The **weighted cumulative residual entropies (WCRE)** proposed by Suhov and Yasaei Sekeh [46] (2015) is defined as

$$WCRE_w(X) = -\int_0^{\infty} w(x)\, \overline{F}(x) \ln \overline{F}(x) dx. \quad (17)$$

The **weighted cumulative past entropies (WCPE)** proposed by Suhov and Yasaei Sekeh [46] (2015) is defined as

$$WCPE_w(X) = -\int_0^{\infty} w(x)\, F(x) \ln F(x) dx. \quad (18)$$

The **interval weighted cumulative residual entropy (IWCRE)** of the random variable X at interval $[t_1,t_2]$ with the weight function w is defined by Sekeh, Borzadran and Roknabadi ([51]) (2015)

$$\begin{aligned} IWCRE_w(X;t_1,t_2) &= -\int_{t_1}^{t_2} w(x) \overline{F}(x;t_1,t_2) \ln \overline{F}(x;t_1,t_2) dx \\ &= -\int_{t_1}^{t_2} w(x) \frac{\overline{F}(x)}{\overline{F}(t_1)-\overline{F}(t_2)} \ln \frac{\overline{F}(x)}{\overline{F}(t_1)-\overline{F}(t_2)} dx. \end{aligned} \quad (19)$$

The **interval weighted cumulative (past) entropy (IWCE)** of the random variable X at interval $[t_1,t_2]$ with the weight function w is defined by Sekeh, Borzadran and Roknabadi ([51]) (2015)

$$\begin{aligned} IWCE_w(X;t_1,t_2) &= -\int_{t_1}^{t_2} w(x) F(x;t_1,t_2) \ln F(x;t_1,t_2) dx \\ &= -\int_{t_1}^{t_2} w(x) \frac{F(x)}{F(t_2)-F(t_1)} \ln \frac{F(x)}{F(t_2)-F(t_1)} dx. \end{aligned} \quad (20)$$

4. CUMULATIVE RELATIVE ENTROPIES AND INACCURACY MEASURES

Consider X and Y two non-negative absolutely continuous random variables with the probability density functions (PDFs) $f(x)$ and $g(y)$, the cumulative distribution functions (CDFs) $F(x) = P(X \le x)$ and $G(y) = P(Y \le y)$, and the reliability functions (RFs) $\overline{F}(x) = P(X > x)$ and $\overline{G}(y) = P(Y > y)$ respectively. Let $t \in [0,\infty)$ and $D = \{(t_1,t_2) \mid 0 \le t_1 < t_2 \le \infty, F(t_1) < F(t_2), G(t_1) < G(t_2)\}$.

For any given pair $(t_1,t_2) \in D$, consider the conditional random variable $(X \;\; t_1 \le X \le t_2)$ with the PDF $f(x;t_1,t_2)$, the CDF $F(x;t_1,t_2)$ and the RF $\overline{F}(x;t_1,t_2)$ defined in the first section and the conditional random variable $(Y \;\; t_1 \le Y \le t_2)$ has the PDF $g(y;t_1,t_2)$, the CDF $G(y;t_1,t_2)$ and the RF $\overline{G}(y;t_1,t_2)$ given by

$$g(y;t_1,t_2) = \frac{g(y)}{G(t_2)-G(t_1)}, G(y;t_1,t_2) = \frac{G(y)}{G(t_2)-G(t_1)},$$
$$\overline{G}(y;t_1,t_2) = \frac{\overline{G}(y)}{G(t_2)-G(t_1)} = \frac{\overline{G}(y)}{\overline{G}(t_1)-\overline{G}(t_2)}, \text{for all} t_1 \le y \le t_2. \tag{21}$$

The **relative entropy, Kullback-Leibler divergence, Kullback-Leibler discrimination information** proposed by Kullback and Leibler [27] (1951) is defined as

$$D(X,Y) = \int_0^\infty f(x) \ln \frac{f(x)}{g(x)} dx. \tag{22}$$

Developing the Shannon entropy, the authors have the idea to compare the entropy inside a family of probability measures, instead of considering the entropy corresponding to only one probability measure.

The **Kerridge measure of inaccuracy** proposed by Kerridge [24] (1961) is defined as

$$H(X;Y) = -\int_0^\infty f(x) \ln g(x) dx = D(X,Y) + H(X). \tag{23}$$

The **weighted inaccuracy measure** proposed by Taneja and Tuteja [48] (1986) is defined as

$$WH(X;Y) = -\int_0^\infty x f(x) \ln g(x) dx. \tag{24}$$

The **residual relative entropy** proposed by Ebrahimi and Kirmani, [15] (1996) is defined as

$$RD(X,Y;t) = \int_t^\infty f(x;t,\infty) \ln \frac{f(x;t,\infty)}{g(x;t,\infty)} dx = \int_t^\infty \frac{f(x)}{\overline{F}(t)} \ln \frac{f(x)/\overline{F}(t)}{g(x)/\overline{G}(t)} dx. \tag{25}$$

The **past relative entropy** proposed by Crescenzo and Longobardi, [8] (2004) is defined as

$$PD(X,Y;t) = \int_0^t f(x;0,t) \ln \frac{f(x;0,t)}{g(x;0,t)} dx = \int_0^t \frac{f(x)}{F(t)} \ln \frac{f(x)/F(t)}{g(x)/G(t)} dx. \tag{26}$$

The **dynamic measure of inaccuracy** proposed by Taneja, Kumar and Srivastava [49] (2009) is defined as

$$RI(X,Y;t) = -\int_t^\infty f(x;t,\infty)\ln g(x;t,\infty)dx = -\int_t^\infty \frac{f(x)}{\overline{F}(t)}\ln\frac{g(x)}{\overline{G}(t)}dx \tag{27}$$
$$= RD(X,Y;t) + RE(X;t).$$

The **weighted residual inaccuracy measure** proposed by Kumar and Taneja [48] (2012) is defined as

$$WRI(X,Y;t) = -\int_t^\infty xf(x;t,\infty)\ln g(x;t,\infty)dx = -\int_t^\infty x\frac{f(x)}{\overline{F}(t)}\ln\frac{g(x)}{\overline{G}(t)}dx \tag{28}$$
$$= WRD(X,Y;t) + WRE(X;t),$$

where

$$WRD(X,Y;t) = \int_t^\infty xf(x;t,\infty)\ln\frac{f(x;t,\infty)}{g(x;t,\infty)}dx = \int_t^\infty x\frac{f(x)}{\overline{F}(t)}\ln\frac{f(x)/\overline{F}(t)}{g(x)/\overline{G}(t)}dx. \tag{29}$$

represents the **weighted residual relative entropy**.

The **past inaccuracy measure (PI)** proposed by Kumar and Taneja [48] (2012) is defined as

$$PI(X;t) = -\int_0^t f(x;0,t)\ln g(x;0,t)dx = -\int_0^t \frac{f(x)}{F(t)}\ln\frac{g(x)}{G(t)}dx = PD(X,Y;t) + PE(X;t). \tag{30}$$

The **weighted past inaccuracy measure (WPI)** proposed by Kumar and Taneja [48] (2012) is defined as

$$WPI(X;t) = -\int_0^t xf(x;0,t)\ln g(x;0,t)dx = -\int_0^t x\frac{f(x)}{F(t)}\ln\frac{g(x)}{G(t)}dx \tag{31}$$
$$= WPD(X,Y;t) + WPE(X;t),$$

where

$$WPD(X,Y;t) = \int_0^t xf(x;0,t)\ln\frac{f(x;0,t)}{g(x;0,t)}dx = \int_0^t x\frac{f(x)}{F(t)}\ln\frac{f(x)/F(t)}{g(x)/G(t)}dx. \tag{32}$$

represents the **weighted past relative entropy**.

The **cumulative residual inaccuracy (CRI)** proposed by Taneja and Kumar [50] (2012) is defined as

$$CRI(X,Y) = -\int_0^\infty \overline{F}(x)\ln\overline{G}(x)dx = CRD(X,Y) + CRE(X), \tag{33}$$

where

$$CRD(X,Y) = \int_0^\infty \overline{F}(x)\ln\frac{\overline{F}(x)}{\overline{G}(x)}dx \tag{34}$$

represents the **cumulative residual relative entropy**.

The **dynamic cumulative residual inaccuracy (DCRI)** proposed by Taneja and Kumar [50] (2012) (a version was also introduced by Chamany and Baratpour (2014)) is defined as

$$DCRI(X,Y;t) = -\int_t^\infty \overline{F}(x;t,\infty)\ln \overline{G}(x;t,\infty)dx = -\int_t^\infty \frac{\overline{F}(x)}{\overline{F}(t)}\ln\frac{\overline{G}(x)}{\overline{G}(t)}dx \tag{35}$$

$$= DCRD(X,Y;t) + DCRE(X;t),$$

where

$$DCRD(X,Y;t) = \int_t^\infty \overline{F}(x;t,\infty)\ln\frac{\overline{F}(x;t,\infty)}{\overline{G}(x;t,\infty)}dx = \int_t^\infty \frac{\overline{F}(x)}{\overline{F}(t)}\ln\frac{\overline{F}(x)/\overline{F}(t)}{\overline{G}(x)/\overline{G}(t)}dx \tag{36}$$

represents the **dynamic cumulative residual relative entropy**.

The **interval relative entropy** proposed by Misagh and Yari, [35] (2012) is defined as

$$ID(X,Y;t_1,t_2) = \int_{t_1}^{t_2} f(x;t_1,t_2)\ln\frac{f(x;t_1,t_2)}{g(x;t_1,t_2)}dx$$

$$= \int_{t_1}^{t_2} \frac{f(x)}{F(t_2)-F(t_1)}\ln\frac{f(x)/[F(t_2)-F(t_1)]}{g(x)/[G(t_2)-G(t_1)]}dx. \tag{37}$$

Here was proposed a measure of discrepancy between two lifetime distributions at the interval of time in base of Kullback-Leibler discrimination information. They studied various properties of this measure, including its connection with residual and past measures of discrepancy and interval entropy, and they obtained its upper and lower bounds.

The **cumulative past inaccuracy (CPI)** proposed by Kumar and Taneja [29] (2015) is defined as

$$CPI(X,Y) = -\int_0^\infty F(x)\ln G(x)dx = CPD(X,Y) + CPE(X), \tag{38}$$

where

$$CPD(X,Y) = \int_0^\infty F(x)\ln\frac{F(x)}{G(x)}dx \tag{39}$$

represents the **cumulative past relative entropy**.

The **dynamic cumulative past inaccuracy (DCPI)** proposed by Kumar and Taneja [29] (2015) (received at 15 April 2014, accepted at 12 march 2015 and was published in december 2015 in J.T.S.A.) and Kundu, Di Crescenzo and Longobardi [31] (2016) (received at 28 March 2014, accepted at 2 August 2015 and was published in 2016 in Metrika) is defined as

$$DCPI(X,Y;t) = -\int_0^t F(x;0,t)\ln G(x;0,t)dx = -\int_0^t \frac{F(x)}{F(t)}\ln\frac{G(x)}{G(t)}dx \tag{40}$$

$$= DCPD(X,Y;t) + DCPE(X;t),$$

where

$$DCPD(X,Y;t)=\int_0^t F(x;0,t)\ln\frac{F(x;0,t)}{G(x;0,t)}dx=\int_0^t \frac{F(x)}{F(t)}\ln\frac{F(x)/F(t)}{G(x)/G(t)}dx \tag{41}$$

represents the **dynamic cumulative past relative entropy**.

The **interval inaccuracy measure (II)** proposed by Kundu [30] (2015) is defined as

$$II(X,Y;t_1,t_2)=-\int_{t_1}^{t_2} f(x;t_1,t_2)\ln g(x;t_1,t_2)dx=-\int_{t_1}^{t_2}\frac{f(x)}{F(t_2)-F(t_1)}\ln\frac{g(x)}{G(t_2)-G(t_1)}dx. \tag{42}$$

The **weighted interval inaccuracy measure (WII)** proposed by Kundu [30] (2015) is defined as

$$\begin{aligned} WII(X,Y;t_1,t_2)&=-\int_{t_1}^{t_2} xf(x;t_1,t_2)\ln g(x;t_1,t_2)dx \\ &=-\int_{t_1}^{t_2} x\frac{f(x)}{F(t_2)-F(t_1)}\ln\frac{g(x)}{G(t_2)-G(t_1)}dx. \end{aligned} \tag{43}$$

The **interval cumulative residual inaccuracy (ICRI)** proposed by Kundu, Di Crescenzo and Longobardi [30] (2016) is defined as

$$\begin{aligned} ICRI(X,Y;t_1,t_2)&=-\int_{t_1}^{t_2}\overline{F}(x;t_1,t_2)\ln\overline{G}(x;t_1,t_2)dx \\ &=-\int_{t_1}^{t_2}\frac{\overline{F}(x)}{\overline{F}(t_1)-\overline{F}(t_2)}\ln\frac{\overline{G}(x)}{\overline{G}(t_1)-\overline{G}(t_2)}dx. \end{aligned} \tag{44}$$

The **interval cumulative past inaccuracy (ICPI)** proposed by Kundu, Di Crescenzo and Longobardi [30] (2016) is defined as

$$\begin{aligned} ICPI(X,Y;t_1,t_2)&=-\int_{t_1}^{t_2}F(x;t_1,t_2)\ln G(x;t_1,t_2)dx \\ &=-\int_{t_1}^{t_2}\frac{F(x)}{F(t_2)-F(t_1)}\ln\frac{G(x)}{G(t_2)-G(t_1)}dx. \end{aligned} \tag{45}$$

CONCLUSIONS

In this paper we present a review of cumulative entropies from reliability theory. First we present the Shannon entropy concept proposed by Shannon [45] and then we present the residual and past entropy-types measures based on CDF and RF, the cumulative entropies, the cumulative relative entropies and inaccuracy measures.

REFERENCES

[1] J. Ahmadi, A. Di Crescenzo, M. Longobardi, On dynamic mutual information for bivariate lifetimes, Advances in Applied Probability 47 (2016), no. 4, 1157–1174.

[2] M. Asadi, Y. Zohrevand, On the dynamic cumulative residual entropy, Journal of Statistical Planning and Inference 137 (2007), no. 6, 1931–1941.

[3] S. Baratpour, Z. Bami, On the Discrete Cumulative Residual Entropy, JIRSS, 2 (2012), 203–215.

[4] M. Beliș S. Guiașu A Quantitative and qualitative measure of information in cybernetic systems, IEEE Trans. on Inf. Theory, 14 (1968), 593–594.

[5] N.D. Cahill, J.A. Schnabela, J.A. Noblea and D.J. Hawkesc, Overlap Invariance of Cumulative Residual Entropy Measures for Multimodal Image Alignment, Medical Imaging 2009: Image Processing, Proc. of SPIE (2002), vol. 7259.

[6] A. Chamany, S. Baratpour, A Dynamic Discrimination Information Based On Cumulative Residual Entropy And Its Properties, Commun Stat Theory Methods 43 (2014), no. 6, 1041–1049.

[7] A. Di Crescenzo, M. Longobardi, Entropy-based measure of uncertainty in past life time distributions, Journal of Applied Probability 39 (2002), no. 2, 434–440.

[8] A. Di Crescenzo, M.A. Longobardi, A measure of discrimination between past lifetime distributions, Stat. Prob. Lett. 67 (2004), no.2, 173–182.

[9] A. Di Crescenzo, M.A. Longobardi, On weighted residual and past entropies, Scientiae Math. Japon 64 (2006), 255–266.

[10] A. Di Crescenzo, M. Longobardi, *On cumulative entropies*, Journal of Statistical Planning and Inference 139 (2009), no. 12, 4072–4087.

[11] S. Das, *On Weighted Generalized Entropy*, Theoretical Statistics and Mathematics Unit, Indian Statistical Institute (2013).

[12] N. Drissi, T. Chonavel, J.M Boucher, *Generalized Cumulative Residual Entropy for Distributions with Unrestricted Supports*, Hindawi Publishing Corporation, Research Letters in Signal Processing (2008).

[13] N. Ebrahimi, F. Pellerey, *New partial ordering of survival functions based on notion of uncertainty*, Journal of Applied Probability 32 (1995) no. 1, 202–211.

[14] N. Ebrahimi, *How to measure uncertainty about residual lifetime*, Sankhya Ser. A 58 (1996), 48–57.

[15] N. Ebrahimi, S.N.U.A. Kirmani, *Some results on ordering of survival functions through uncertainty*, Statist. Prob. Lett. 29 (1996), no. 2, 167–176.

[16] S. Guiașu, *Weighted entropy*, Report on Math. Physics, 2 (1971), 165–179.

[17] S. Guiașu, M. Malița, *Coalition and Connection in Games*, Pergamon Press, Oxford, 1980.

[18] E.T. Jaynes, *Information theory and statistical mechanics*, Phys. Rev., 106 (1957), 620–630.

[19] E.T. Jaynes, *Information theory and statistical mechanics. II*, Phys. Rev., 108 (1957), 171–190.

[20] O. Kamari, *SOME PROPERTIES OF THE CUMULATIVE RESIDUAL ENTROPY*, International Journal of Multidisciplinary Academic Research 3 (2015), no. 1, 54–58.

[21] O. Kamari, *On Dynamic Cumulative Residual Entropy of Order Statistics*, Journal of Statistics Applications & Probability 5 (2016), no. 3, 515–519.

[22] D.T. Kang, *Further results on closure properties of LPQE order*, Statistical Methodology 25 (2015), 23–35.

[23] D.T. Kang, *Some Results on DDCRE Class of Life Distributions*, The Indian Journal of Statistics, Indian Statistical Institute Part 2 (2014), vol. 77-A, 351–363.

[24] D.F. Kerridge, *Inaccuracy and inference*, Journal of Royal Statistical Society, Series B, 23 (1961), 184–194.

[25] [25] M. Khorashadizadeh, A.H. Rezaei Roknabadi, G.R. Mohtashami Borzadaran, *Doubly truncated (interval) cumulative residual and past entropy*, Statistics and Probability Letters 83 (2013), no. 5, 1464–1471.

[26] I. Klein, B. Mangold, M. Doll, *Cumulative Paired ϕ-Entropy*, Entropy 18 (2016).

[27] S. Kullback, R.A. Leibler, *On information and sufficiency*, Ann. Math. Statist. 25 (1951), 745–751.

[28] V. Kumar, H.C. Taneja, *On length biased dynamic measure of past inaccuracy*, Metrika, 75 (2012), 73–78.

[29] V. Kumar, H.C. Taneja, *Dynamic Cumulative Residual and Past Inaccuracy Measures*, Atlantis Press, Journal of Statistical Theory and Applications 14 (2015), no. 4, 399–412.

[30] C. Kundu, *On Weighted Measure of Inaccuracy for Doubly Truncated Random Variables*, Communications in Statistics-Theory & Methods (2015).

[31] C. Kundu, A. Di Crescenzo, M.A. Longobardi, *On cumulative residual (past) inaccuracy for truncated random variables*, Springer-Verlag Berlin Heidelberg (2016).

[32] M. Mirali, S. Baratpourb, C.V. Fakoorc, *On weighted cumulative residual entropy*, Communications in Statistics - Theory and Methods (2016).
[33] F. Misagh, Y. Panahi, G.H. Yari, R. Shahi, *Weighted cumulative entropy and its estimation*, Proceedings of the IEEE International Conference on Quality and Reliability, 477–480, Bangkok, 2011.
[34] F. Misagh, G.H. Yari, *On weighted interval entropy*, Statistics and Probability Letters 81 (2011), no. 2, 188–194.
[35] F. Misagh, G.H. Yari, *Interval entropy and informative distance*, Entropy, 14, (2012), no. 2, 480–490.
[36] F. Misagh, *On Shift-Dependent Cumulative Entropy Measures*, Hindawi Publishing Corporation, International Journal of Mathematics and Mathematical Sciences (2016).
[37] J. Navarro, Y. del Aguila, M. Asadi, *Some new results on the cumulative residual entropy*, Journal of Statistical Planning and Inference 140 (2010), no. 1, 310–322.
[38] M. Nourbakhsh, G. Yari, Y. Mehrali, *Weighted Entropies and Their Estimations*, Communications in Statistics - Simulation and Computation (2016).
[39] V. Preda, C. Bălcău, *Entropy Optimization with Applications*, Ed. Academiei Române, Bucharest, 2010.
[40] G. Rajesh, E.I. Abdul-Sathar, K.V. Reshmi, K.R. Muraleedharan-Nair, *Bivariate Generalized Cumulative Residual Entropy*, The Indian Journal of Statistics, Indian Statistical Institute Part 1 (2014), vol. 76-A, 101–122.
[41] M. Rao, Y. Chen, B.C. Vemuri, F. Wang, *Cumulative residual entropy: a new measure of information*, IEEE Transactions Informations Theory 50 (2004), no. 6, 1220–1228.
[42] M. Rao, *More on a New Concept of Entropy and Information*, Journal of Theoretical Probability 18 (2005), no. 14, 967–981.
[43] S. Ravi, A. Saeb, *On information theory and its applications*, ProbStat Forum, (2014), no. 7, 45–54.
[44] M. Serrurier, H. Prade *Entropy evaluation based on confidence intervals of frequency estimates: Application to the learning of decision trees*, Proceedings of the 32nd International Conference on Machine Learning, Lille, 2015.
[45] C.E. Shannon, *A mathematical theory of communication*, Bell System Technical Journal 27 (1948), 379–423, 623–656.
[46] Y. Suhov, S. Yasaei Sekeh, *Weighted cumulative entropies: An extension of CRE and CE*, http://arxiv.org/abs/1507.07051v1, 2015.
[47] S.M. Sunoj, P.G. Sankaran, S.S. Maya, *Characterizations of life distributions using conditional expectations of doubly (interval) truncated random variables*, Communications in Statistics-Theory and Methods 38 (2009), no.9, 1441–1452.
[48] H.C. Taneja, R.K. Tuteja, *Characterization of a quantitative-qualitative measure of inaccuracy*, Kybernetika, 22 (1986), 393–402.
[49] H.C. Taneja, V. Kumar, R. Srivastava, *A dynamic measure of inaccuracy between two residual lifetime distributions*, International Mathematical Forum 4 (2009), no. 25–28, 1213–1220.
[50] H.C. Taneja, V. Kumar, *On Dynamic Cumulative Residual Inaccuracy Measure*, Proceedings of the World Congress on Engineering, London, 2012.
[51] S. Yasaei Sekeh, G.R. Mohtashami Borzadran, A.H. Rezaei Roknabadi, *A note on double truncated (interval) weighted cumulative entropies*, http://arxiv.org/abs/1508.00246v3, 2015.
[52] V. Zardasht, S. Parsi, M. Mousazadeh, *On empirical cumulative residual entropy and a goodness-of-fit test for exponentiality*, Springerlink.com, 2014.

ON HODRICK-PRESCOTT FILTER. A SHORT SURVEY AND APPLICATIONS

Bogdan Corneliu BIOLAN

University of Bucharest, Bucharest, Romania (bbiolanc@gmail.com)

DOI: 10.19062/1842-9238.2017.15.2.15

Abstract: *In this paper, the Hodrick-Prescott filter and some new related results are presented. These results are very useful in macroeconomics and Digital Signal Processing (DSP). Furthermore, an application in the DSP field is given.*

Keywords: *Hodrick-Prescott filter, DSP, macroeconomics.*

1. INTRODUCTION

The Hodrick-Prescott (HP) filter is a tool commonly used in macroeconomics, and it is very usefull to extract a trend component from a time series.In this survey we point out the main results from [1] obtained until now for this type of filter and draw some lines of future research in order to obtain new results. Sakarya and de Jong [1] obtained a new representation of the transformation of the data which is implied by the HP filter. This representation is useful to analyze the properties of the HP filter without relying on ARMA based approximation, which was used in the literature before. Also, we focus on the characterization of the large T behavior of the HP filter and some conditions under where it is asimptotically equivalent to a symmetric weighted average with weights independent of sample size.

Sakarya and de Jong [1] found that the cyclical component of the HP filter also has a weak dependence property when the HP filter is applied to a stationary mixing process, a linear deterministic trend process and/or a process with a unit root. The HP filter is a good tool to achieve weak dependence in time series. One of the techiques that we will pay an important role in our research is the large smoothing parameter approximation of the HP filter, which is derived in [1]. Using this approximation, authors found an alternative justification for the procedure given in 2002 in [2] (for more information, see [1]) for adjusting the smoothing parameter for the data frequency.

2. MAIN RESULTS REGARDING THE HODRICK-PRESCOTT FILTER

2.1 The Hodrick-Prescott filter. The Hodrick-Prescott filter represents a standard tool in macroeconomics, very useful for separating the long trend in data series from short run fluctuations.

The HP filter smoothed series $\hat{\tau}_T = (\hat{\tau}_{T1}, \hat{\tau}_{T2}, ..., \hat{\tau}_{TT})$ as defined and described in economics by Hodrick and Prescott [3,4] results from minimizing, over all $\tau \in \mathbf{R}^T$, the function

$$\sum_{t=1}^{T}(y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1}(\tau_{t+1} - 2\tau_t + \tau_{t-1})^2, \tag{1}$$

where T denotes the sample size, λ is the nonnegative smoothing parameter so that for quarterly data is often chosen to be equal to 1600, and $y = (y_1, y_2, ..., y_t)^t$ represents the data series to be smoothed.

In [5] a similar filtering technique has been introduced, according to [1]. Usually, $\hat{\tau}_{Tt}$ is referred to as the *trend component*, while $\hat{c}_{Tt} = y_t - \hat{\tau}_{Tt}$ is called the *cyclical component*. As stated in [1], there exists a unique minimizer to the minimization problem descried by equation (1), so that, for a known positive defined $(T \times T)$ matrix F_T, by letting I_T denote the $(T \times T)$ identity matrix, $y = (\lambda F_T + I_T)\hat{\tau}_{Tt}$ and

$$\hat{\tau}_{Tt} = (\lambda F_T + I_T)^{-1} y \tag{2}$$

Therefore, the trend component $\hat{\tau}_{Tt}$ and the cyclical component $\hat{c}_{Tt}$ are both weighted averages of y_t and according to [1] we have: $\hat{\tau}_{Tt} = \sum_{s=1}^{T} w_{Tts} y_s$.

For notational convenience, the dependence of w_{Tts} and $\hat{\tau}_{Tt}$ with respect to λ is suppressed. One of the purposes set in [1] is to find a new representation for w_{Tts} and provide immediate consequences of this representation. This approach eliminates the inability to derive a simple analytical formula for the elements of $(\lambda F_T + I_T)^{-1}$, fact that prevented other researchers to find a simple expression for the weights that are implicit for the HP filter (for more information, see [1]).

We shall notice first of all, that: $\hat{\tau}_{Tt}(y_1 + 1, y_2 + 1, ..., y_T + 1) = \hat{\tau}_{Tt}(y_1, y_2, ..., y_T) + 1$, so this means that $\sum_{s=1}^{T} w_{Tts} = 1$ for $t \in \{1,2,...,T\}$. Also, we have that: $\hat{\tau}_{Tt}(1,2,...,T) = t$ and therefore we have that: $\sum_{s=1}^{T} w_{Tts}\, s = t$, for $t \in \{1,2,...,T\}$.

Sakarya and de Jong [1] remarked that a quadratic trend is not absorbed in $\hat{\tau}_{Tt}$ in this way. They also noticed that previous literature regarding the HP filter is only based on the observation that the first order condition for $\hat{\tau}_{Tt}$, $t \in \{3,...,T-2\}$, is given by:

$$\begin{aligned} &-2(y_t - \hat{\tau}_{Tt}) - 4\lambda(\hat{\tau}_{T,t+1} - 2\hat{\tau}_{Tt} + \hat{\tau}_{T,t-1}) + 2\lambda(\hat{\tau}_{Tt} - 2\hat{\tau}_{T,t-1} + \hat{\tau}_{T,t-2}) \\ &+ 2\lambda(\hat{\tau}_{T,t+2} - 2\hat{\tau}_{T,t+1} + \hat{\tau}_{Tt}) = 0 \end{aligned} \tag{3}$$

Let $\overline{B}$ denote the forward operator and B the backward operator, then according to [1], this simplifies to the following relation:

$$y_t = (\lambda \overline{B}^2 - 4\lambda \overline{B} + (1 + 6\lambda) - 4\lambda B + \lambda B^2)\, \hat{\tau}_{Tt}, \tag{4}$$

which can also be re-written as:

$$y_t = (\lambda\,|1-B|^4 +1)\;\hat{\tau}_{Tt}. \tag{5}$$

Papers that analyze the HP filter based on the first order condition, according to [1], are for example [6-10].

2.2 On the weights of a Hodrick-Prescott filter. In this subsection we shall refer to section 2 from [1] in order to show how the exact weights w_{Tts} implied by the HP filter have been derived. The approach is based on minimizing the function provided in (6) over $(\theta_1,\theta_2,...,\theta_T)$, for a basis of functions $p_j:[0,1]\to\mathbf{R}$, for $t,j\in\mathbf{N}^+$, rather than minimizing the function from (1) over τ:

$$\begin{aligned}&\sum_{t=1}^{T}\left[y_t-\sum_{j=1}^{T}\theta_j p_j\left(\frac{t}{T}\right)\right]^2+\lambda\sum_{t=2}^{T-1}\left[\sum_{j=1}^{T}\theta_j p_j\left(\frac{t+1}{T}\right)-2\sum_{j=1}^{T}\theta_j p_j\left(\frac{t}{T}\right)+\sum_{j=1}^{T}\theta_j p_j\left(\frac{t-1}{T}\right)\right]^2\\&=\sum_{t=1}^{T}(y_t-\theta' p_{Tt})^2+\lambda\sum_{t=2}^{T-1}(\theta'\Delta^2 p_{T,t+1})^2,\end{aligned} \tag{6}$$

where $\theta'=(\theta_1,\theta_2,...,\theta_T)$ and $p_{Tt}^{'}=(p_1(\frac{t}{T}),p_2(\frac{t}{T}),...,p_T(\frac{t}{T}))$. By differentiating (6) with respect to θ, the minimizer $\hat{\theta}$ derived in [1] is given by:

$$0=-2\sum_{t=1}^{T}y_t p_{Tt}+2\sum_{t=1}^{T}p_{Tt}p_{Tt}^{'}\hat{\theta}+2\lambda\sum_{t=2}^{T-1}(\Delta^2 p_{T,t+1})(\Delta^2 p_{T,t+1})'\hat{\theta}, \tag{7}$$

If an inverse exists, the minimize $\hat{\theta}$ can be expressed as follows:

$$\hat{\theta}=\left(T^{-1}\sum_{t=1}^{T}p_{Tt}p_{Tt}^{'}+\lambda T^{-1}\sum_{t=2}^{T-1}(\Delta^2 p_{T,t+1})(\Delta^2 p_{T,t+1})'\right)^{-1}T^{-1}\sum_{t=1}^{T}y_t p_{Tt}. \tag{8}$$

In [1], authors have chosen $p_1\left(\frac{t}{T}\right)=1$ and $p_j\left(\frac{t}{T}\right)=\sqrt{2}\cos\left(\frac{\pi(j-1)(t-\frac{1}{2})}{T}\right)$, $j=2,...,T$.
In what follows, I_T denotes the identity matrix of dimension $(T\times T)$.

Lemma [1]. Let $p_{Tt}=(p_1(\frac{t}{T}),p_2(\frac{t}{T}),...,p_T(\frac{t}{T}))'$, where $p_1\left(\frac{t}{T}\right)=1$ and $p_j\left(\frac{t}{T}\right)=\sqrt{2}\cos\left(\frac{\pi(j-1)(t-\frac{1}{2})}{T}\right)$, $j=2,...,T$. Then, we have:

$$T^{-1}\sum_{t=1}^{T}p_{Tt}p_{Tt}^{'}=I_T \tag{9}$$

and

$$T^{-1}\sum_{t=2}^{T-1}(\Delta^2 p_{T,t+1})(\Delta^2 p_{T,t+1})' = D_T - 32T^{-1}q_{T1}q_{T1}' - 32T^{-1}q_{T2}q_{T2}', \tag{10}$$

where $D_T = \text{diag}\left(\left\{16\sin\left(\frac{\pi(j-1)}{2T}\right)^4,\ j=1,...,T\right\}\right)$ and

$q_{T1} = (q_{T11}, q_{T12}, ..., q_{T1T})'$, $q_{T2} = (q_{T21}, q_{T22}, ..., q_{T2T})'$, where, for $j = 1,2,...,T$,

$q_{T1j} = \sin\left(\frac{\pi(j-1)}{2T}\right)^2 \cos\left(\frac{\pi(j-1)}{2T}\right)$ and $q_{T2j} = \sin\left(\frac{\pi(j-1)}{2T}\right)^2 \cos\left(\frac{\pi(j-1)(T-\frac{1}{2})}{2T}\right)$.

The proof of this result can be found in the Mathematical Appendix of [1].

As authors in [1] point out, the importance of this result that the matrix to be inverted is now 'close' to an easily invertible diagonal matrix (in the way that two matrices of rank 1 have been added to the diagonal matrix).

Remark. The difference from the classical approaches which can be found in literature is the fact that they minimize equation (1) over τ, and no such structure occurs, as the one mentioned above.

It is well known from the literature that explicit formulas can be obtained for the inverse of the sum of a matrix plus another matrix of rank 1, and such results can be adapted to deal with the inverse of a matrix plus a matrix of rank 2 as well. The main purpose of [1] and also of [11] is to use such a result, to obtain a tractable expression for $\hat{\tau}_{\pi}$.

For $m \in \mathbf{Z}$, $\lambda \in [0,\infty)$ and $T \geq 1$ we define:

$$f_{T\lambda}(m) = \frac{1}{2T} + \frac{(-1)^m}{1+16\lambda}(2T)^{-1} + T^{-1}\sum_{j=2}^{T}\cos\left(\frac{\pi(j-1)m}{T}\right)\left(1+16\lambda\sin\left(\frac{\pi(j-1)}{2T}\right)^4\right)^{-1} \tag{11}$$

and for $m \geq 1$, $\lambda \in [0,\infty)$ and $T \geq 1$,

$$\begin{aligned} g_{T\lambda} &= T^{-1}p_{Tm}'(1+\lambda D_T)^{-1}q_{T1} \\ &= T^{-1}\sum_{j=1}^{T}\sqrt{2}\cos\left(\frac{\pi(j-1)(m-\frac{1}{2})}{T}\right)q_{T1j}\left(1+16\lambda\sin\left(\frac{\pi(j-1)}{2T}\right)^4\right)^{-1}. \end{aligned} \tag{12}$$

We also define, similarly to [1], the following sequences:

$$\delta_{T\lambda} = T^{-1}q_{T1}'(I_T + \lambda D_T)^{-1}q_{T1}, \tag{13}$$

$$\eta_{T\lambda} = T^{-1}q_{T1}'(I_T + \lambda D_T)^{-1}q_{T2}, \tag{14}$$

$$\begin{aligned} \xi_{T\lambda} &= 32\lambda(1-64\lambda\delta_{T\lambda})(1-64\lambda\delta_{T\lambda}+32^2\lambda^2(\delta_{T\lambda}^2-\eta_{T\lambda}^2))^{-1} \\ &+32^2\lambda^2(1-64\lambda\delta_{T\lambda}+32^2\lambda^2(\delta_{T\lambda}^2-\eta_{T\lambda}^2))^{-1}\delta_{T\lambda} \end{aligned} \tag{15}$$

and

$$\phi_{T\lambda} = 32^2\lambda^2(1-64\lambda\delta_{T\lambda}+32^2\lambda^2(\delta_{T\lambda}^2-\eta_{T\lambda}^2))^{-1}\eta_{T\lambda}. \tag{16}$$

It is clear that: $f_{T\lambda}(m) = f_{T\lambda}(-m), \forall m \in \mathbf{Z}$.
We state now the following result, from [1]:

Theorem [1]. For any $\lambda \in [0,\infty)$ we have $\hat{\tau}_{Tt} = \sum_{s=1}^{T} w_{Ts} y_s$, where:

$$\begin{aligned} w_{Ts} &= f_{T\lambda}(t-s) + f_{T\lambda}(T)I(t+s-1=T) + f_{T\lambda}(t+s-1)I(t+s-1<T) \\ &+ f_{T\lambda}(2T-t-s+1)I(t+s-1>T) + \xi_{T\lambda}g_{T\lambda}(t)g_{T\lambda}(s) + \phi_{T\lambda}g_{T\lambda}(T-t+1)g_{T\lambda}(s) \\ &+ \phi_{T\lambda}g_{T\lambda}(t)\xi_{T\lambda}g_{T\lambda}(T-s+1)g_{T\lambda}(s) + \xi_{T\lambda}g_{T\lambda}(T-t+1)g_{T\lambda}(T-s+1) \\ &= \sum_{j=1}^{T} w_{Ts}^{j}, \end{aligned} \tag{17}$$

with $|f_{T\lambda}(0)| \leq 1$ and for any $m \in \{1,2,...,T\}$, $|f_{T\lambda}(m)| \leq Cm^{-3}$, for some constant C not depending on T.

3. FUTURE WORK ON HP FILTER

Future work will be continued for the mathematical properties of the Hodrick-Prescott filter and find other original calculations of the explicit weights of this filter. We aim to obtain a weak law of large numbers result for functions of the cyclical component.Also, for future research we want to analyze what happens when the sample size and the smoothing parameter are large, independent of what the authors did in [1], and we want to obtain a new procedure for adjusting the smoothing parameter for the data frequency (a similar procedure was obtained also in [2]). We keep in mind that it is possible to utilize the HP filter in combination with other types of filters used in DSP, optimizing the whole filtering process. Also we will review the properties of the HP filter that will helps us in our future work:

- The cut-off region is not steep; this means that leakage from cycles just outside the target region can be significant.
- In engineering applications filter leakage represents a sign of a poor filter.
- Still, in business cycle analysis, researchers have arguments to support at least a small degree of desirable leakage.
- The frequency band of 1.5 to 8 years has been selected based on the expert decision.
- The boundaries 1.5 and 8 years should not be regarded as carved in stone.
- The filter leakage for example will allow strong 9 year cycles to appear in the filtered series.
- The HP filter is asymmetric. Except the central values the double HP filtered series are phase shifted compared to the underlying ideal cycle. Also the phase shifts fade out for a given data as newer data arrive.

CONCLUSIONS

This survey paper points out the new mathematically rigorous results and properties of the HP filter obtained by the authors in [1] and also establishes new lines of research in this field, having set for the future, clear objectives for what kind of results to be obtained. We also discuss about future possible applications of the Hodrick-Prescott filter in DSP, justifying this in section 3.

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.

ON THE SCALE PARAMETER OF EXPONENTIAL DISTRIBUTION

Anca Ileana LUPAŞ

Military Technical Academy, Bucharest, Romania (lupas_a@yahoo.com)

DOI: 10.19062/1842-9238.2017.15.2.16

Abstract: *Exponential distribution is one of the widely used continuous distributions in various fields for statistical applications. In this paper we study the exact and asymptotical distribution of the scale parameter for this distribution. We will also define the confidence intervals for the studied parameter as well as the fixed length confidence intervals.*

1. INTRODUCTION

Exponential distribution is used in various statistical applications. Therefore, we often encounter exponential distribution in applications such as: life tables, reliability studies, extreme values analysis and others.

In the following paper, we focus our attention on the exact and asymptotical repartition of the exponential distribution scale parameter estimator.

2. SCALE PARAMETER ESTIMATOR OF THE EXPONENTIAL DISTRIBUTION

We will consider the random variable X with the following cumulative distribution function:

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

where θ is an unknown scale parameter

Using the relationships between $M(X) = \theta$; $\sigma^2(X) = \theta^2$; $\sigma(X) = \theta$, we obtain a theoretical variation coefficient $\gamma = \frac{\sigma(X)}{M(X)} = 1$. This is a useful indicator, especially if you have observational data which seems to be exponential and with variation coefficient of the selection closed to 1.

If we consider $x_1, x_2, ...x_n$ as a part of a population that follows an exponential distribution, then by using the maximum likelihood estimation method we obtain the following estimate

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2}$$

Since $M\left(\hat{\theta}\right)=\theta$, it follows that $\hat{\theta}$ is an unbiased estimator for θ. Similarly, because $D^2(\hat{\theta})=\frac{\theta^2}{n}$ and $\lim_{n\to\infty}\frac{\theta^2}{n}=0$, we obtain that $\hat{\theta}$ is an absolutely correct estimator.

The efficiency of the estimator $\hat{\theta}$ can be calculated using the following formula:

$$e_n(\hat{\theta})=\frac{\dfrac{1}{n\int_0^{\infty}\left[\frac{\partial \ln f(x,\theta)}{\partial\theta}\right]^2 f(x,\theta)\,dx}}{D(\hat{\theta})} \tag{3}$$

where $f(x;\theta)=\frac{1}{\theta}e^{-\frac{x}{\theta}}$ $(x>0,\ \theta>0)$ is the probability density function of the exponential distribution. After the calculation we obtain $n\int_0^{\infty}\left[\frac{\partial \ln f(x;\theta)}{\partial\theta}\right]^2\cdot f(x;\theta)\,dx=\frac{n}{\theta^2}$ and thus $e_n(\hat{\theta})=1$ which implies that $\hat{\theta}$ is also an efficient estimator.

Taking into account the reproducibility property of the exponential distribution we can calculate the exact distribution of the random variable $\hat{\theta}$.

A sum $X_1+X_2+...+X_n$, consisting of n randomly selected variables, all exponentially distribuited, has a distribution density function equal to

$$f_n(x)=\frac{1}{\theta\,\Gamma(n)}\left(\frac{x}{\theta}\right)^{n-1}\cdot e^{-\frac{x}{\theta}}. \tag{4}$$

This leads us to the cumulative distribution function

$$H_n(x)=P(\hat{\theta}\le x)=P(X_1+...+X_n\le nx)=\frac{1}{\theta\,\Gamma(n)}\int_0^{nx}\left(\tfrac{u}{\theta}\right)^{n-1}e^{-\frac{u}{\theta}}\,du=F_n(nx) \tag{5}$$

from which we obtain that the distribution density function of the estimator $\hat{\theta}$ is equal to

$$h_n(x)=f_n(nx)\cdot n=\frac{n}{\theta\,\Gamma(n)}\left(\frac{nx}{\theta}\right)^{n-1}e^{-\frac{nx}{\theta}} \tag{6}$$

From the above we obtain that $M(\hat{\theta})=\int_0^{\infty}x\,h_n(x)\,dx=\theta$ and $D^2(\hat{\theta})=\int_0^{\infty}(x-\theta)^2\,h_n(x)=\frac{\theta^2}{n}$ which is consistent with previously known results.

By using the central limit theorem we can state that the random variable $\overline{X}$ follows a normal distribution with the parameters θ and $\frac{\theta^2}{n}$, which means that $\overline{X}\approx N\left(\theta;\frac{\theta^2}{n}\right)$.

Based on this we can calculate the asymptotical cumulative distribution function for the estimator $\hat{\theta}$,

$$G_n(x) = \lim_{n\to\infty} P(\hat{\theta} \le x) = \lim_{n\to\infty} P(\overline{X} \le x). \tag{7}$$

This shows that the estimator $\hat{\theta}$ has the same asymptotical distribution as $\overline{X}$ and a distribution density function equal to:

$$g_n(x) = \frac{\sqrt{n}}{\theta\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{(x-\theta)\sqrt{n}}{\theta}\right]^2} \tag{8}$$

Using the asymptotical distribution of the estimator $\hat{\theta}$ we can determine the confidence intervals for the scale parameter θ. To do this we need to consider the reduced normal random variable $Z = \frac{(\hat{\theta}-\theta)\sqrt{n}}{\theta}$ and the significance level α.

By definition we will get

$$P(|Z| \le Z_{1-\frac{\alpha}{2}}) = 1-\alpha \tag{9}$$

where $Z_{1-\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}$ quartile of the standard normal distribution. Equation (9) can be rewritten such as

$$P(-Z_{1-\frac{\alpha}{2}} \le \frac{(\hat{\theta}-\theta)\sqrt{n}}{\theta} \le Z_{1-\frac{\alpha}{2}}) = 1-\alpha. \tag{10}$$

We may assume that $1-\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} > 0$ because we deal with asymptotical distribution and thus *n* is sufficiently high. After carrying out the calculation we obtain for the parameter θ the following confidence interval which depends on the maximum likelihood estimation $\hat{\theta}$

$$\frac{\hat{\theta}}{1+\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}} \le \theta \le \frac{\hat{\theta}}{1-\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}}. \tag{11}$$

3. SET LENGTH CONFIDENCE INTERVALS FOR THE SCALE PARAMETER

Using a similar method to that which Stein proposed for the double selection method we find a set length confidence interval δ for the parameter θ.

Let *p* be a system of independent selections, each with a volume *n*, taken from a population which has the same exponential distribution as the random variable *X*

$$\begin{cases} x_{11},...,x_{1n} \\ ... \\ x_{p1},...,x_{pn} \end{cases} \tag{12}$$

These selections allow us to obtain independent and identical assigned estimators

$$\hat{\theta}_i \approx N\left(\theta;\frac{\theta^2}{n}\right) \quad i=1,...,p. \tag{13}$$

Let $\bar{\hat{\theta}} = \frac{1}{p}\sum_{i=1}^{p}\hat{\theta}_i$ and $s^2_{\hat{\theta}} = \frac{1}{p-1}\sum_{i=1}^{p}(\hat{\theta}_i - \bar{\hat{\theta}})^2$. Because $\bar{\hat{\theta}}$ is the selection mean for the random variables $\hat{\theta}_i$; i=1,...,p we know that $\bar{\hat{\theta}} \approx N(\theta;\frac{\theta^2}{np})$.

Next let us consider a second system with *m* independent selections, each of volume *n*, taken from a population which has the same exponential distribution as the random variable *X*

$$\begin{cases} x_{p+1;1},...,x_{p+1;n} \\ ... \\ x_{p+m;1},...,x_{p+m;n} \end{cases} \tag{14}$$

These new selections allow us to calculate *m* independent and identically assigned estimators

$$\hat{\theta}_{p+j} \approx N(\theta;\frac{\theta^2}{n}) \qquad j=1,...,m. \tag{15}$$

Let $\bar{\bar{\hat{\theta}}}$ be the calculated average of both selection systems:

$$\bar{\bar{\hat{\theta}}} = \frac{1}{p+m}\sum_{i=1}^{p+m}\hat{\theta}_i \tag{16}$$

It is know that $\bar{\bar{\hat{\theta}}} \approx N\left(\theta;\frac{\theta^2}{n(m+p)}\right)$. Note that for *m*=0 we obtain $\bar{\bar{\hat{\theta}}} = \bar{\hat{\theta}}$. Due to the fact that $\hat{\theta} \approx N(\theta;\frac{\theta^2}{n})$ and s^2_θ is the variance of random variable $\hat{\theta}$ considered in equation 2, we deduce the following equation

$$\frac{s^2_{\hat{\theta}}}{\frac{\theta^2}{n}} \approx \chi^2_{p-1}\Big/(p-1) \tag{17}$$

Also the random variable $\hat{\bar{\theta}}$ matches with the reduced random variable through the following equation:

$$\frac{\sqrt{n(m+p)}}{\theta}(\hat{\bar{\theta}}-\theta)\approx N(0,1) \tag{18}$$

The random variables from equations (17) and (18) are independent and thus we can consider the following ratio

$$\frac{\frac{\sqrt{n(m+p)}}{\theta}(\hat{\bar{\theta}}-\theta)}{\frac{s_{\hat{\theta}}\cdot\sqrt{n}}{\theta}}=\frac{\sqrt{(m+p)}}{s_{\hat{\theta}}}(\hat{\bar{\theta}}-\theta)\approx t_{p-1} \tag{19}$$

where t_{p-1} is a Student random variable with *p-1* degrees of freedom.

For the significance level α we have the following equation:

$$P\left(\left|\frac{\sqrt{m+p}}{s_{\hat{\theta}}}(\hat{\bar{\theta}}-\theta)\right|\le t_{\frac{\alpha}{2};p-1}\right)=1-\alpha \tag{20}$$

from which we deduce the following confidence interval for θ

$$\hat{\bar{\theta}}-\frac{s_{\hat{\theta}}t_{\frac{\alpha}{2};p-1}}{\sqrt{m+p}}\le\theta\le\hat{\bar{\theta}}+\frac{s_{\hat{\theta}}t_{\frac{\alpha}{2};p-1}}{\sqrt{m+p}} \tag{21}$$

The length of this interval can be easily calculated and is equal to

$$l=\frac{2\cdot s_{\hat{\theta}}\cdot t_{\frac{\alpha}{2};p-1}}{\sqrt{m+p}}. \tag{22}$$

The length of this interval must not be greater than the considered length δ.

We will start with m=0 and compare the length l with δ. If $l\le\delta$ then the interval $\left(\hat{\bar{\theta}}-\frac{1}{2}\delta;\hat{\bar{\theta}}+\frac{1}{2}\delta\right)$ is a confidence interval for θ of size δ, that has a confidence coefficient greater or at least equal to $1-\alpha$, build only with the help of the first selection system.

If $l>\delta$, we need to carry out the second selection system, where m is equal with the smallest integer for which $\frac{2\cdot s_{\hat{\theta}}\cdot t_{\frac{\alpha}{2};p-1}}{\sqrt{m+p}}\le\delta$. In this case $\left(\hat{\bar{\theta}}-\frac{1}{2}\delta;\hat{\bar{\theta}}+\frac{1}{2}\delta\right)$ will be the confidence interval for θ of size δ, build on both selection systems and having a confidence coefficient greater then 1-α.

REFERENCES

[1] Gheorghe, F.C. , Obreja , G. Teste statistice pentru selecţii de volum mic (n ≤ 50) *Ed. Univ. Piteşti 1999*.

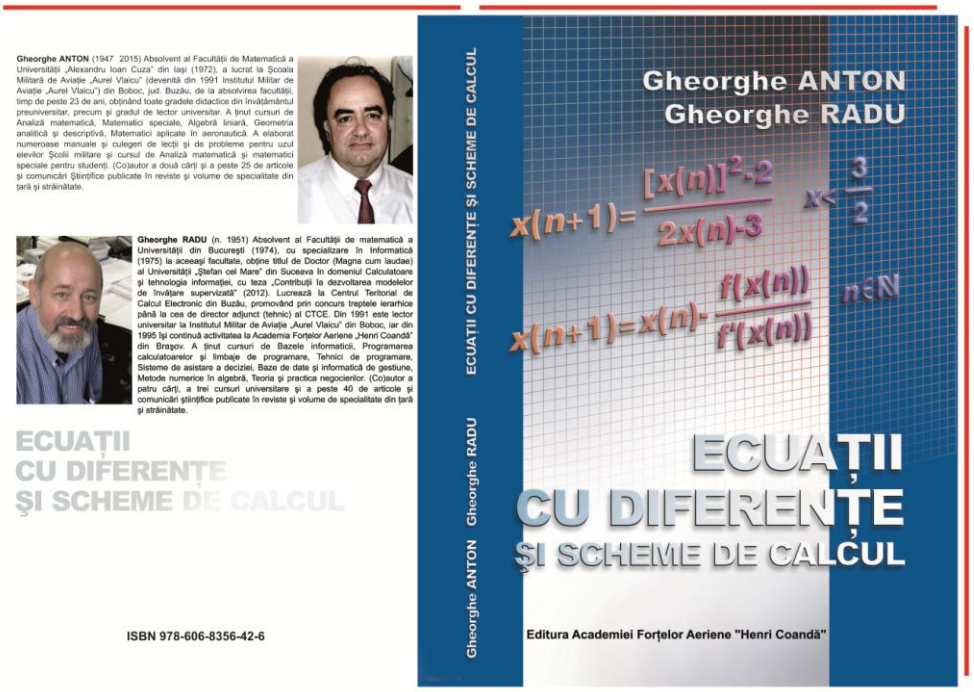
[2] Margolin,H.B., Maurer,W. Tests of the Kolmogorov Smirnov type for exponential data with unknow scale and related problems. *Biometrica (1976), 63,1 pg.149-160*

A USEFUL STUDY OF DIFFERENCE EQUATIONS – FROM THEORETICAL AND PRACTICAL POINTS OF VIEW

Review of the volume *Ecuații cu diferențe și scheme de calcul* by Gh. Anton and Gh. Radu, Editura Academiei Forțelor Aeriene "Henri Coandă", 2016

prof.univ.dr. **Bianca-Renata SATCO**

University "Ștefan cel Mare" Suceava, Romania



Due to the amazing rate of growing of computer science, the theory of difference equations, used to numerically approximate and to obtain properties of solutions for various kinds of differential equations, has received much attention in the last decades.

Thus, the idea of elaborating of a comprehensive material on this subject, containing the basic notions related to difference equations, is more than welcome.

Besides, for the beginners in the field and not only, putting, together with theoretical elements, of a section concerning a large number of applications to different branches of elementary and advanced mathematics, of computer science and many other areas makes the subject of this book even more interesting.

Combining their teaching and research experience, the authors provide a well organized, pedagogical approach, although many theoretical results are only stated without proof given the amplitude of the theory and the large number of notions taken into consideration.

The material is divided into two main sections: theory, respectively applications.

The first chapter, entitled "Ecuații cu diferențe", introduces the difference and antidifference operators, gives their basic properties and examples, then presents Cauchy problems and problems with conditions at the boundary associated to difference equations.

Afterwards, the authors focus on classical methods as well as on operational methods, based on MacLaurin or Z transforms, for solving linear (or reducible to linear) difference equations. In this part, the discussion is given in full detail, with definitions, proofs and tables of properties together with useful examples.

The next part of this chapter is concerned with the matter of stability for dynamical systems; namely, different types of stability are obtained for linear difference equations and also for nonlinear such equations.

A list of solved problems and a number of problems proposed to the reader ends this theoretical section.

Chapter 2 represents, in the opinion of the reviewer, the strong point of the book. It contains applications of the theory previously described to a large number of practical problems.

For example, elementary exercises containing recurrences for sequences of real numbers can be solved in a very elegant manner by regarding them as difference equations and applying specific methods (such as, operational methods).

Also, the problem of calculating powers of square matrices reduces to the solving of appropriate difference equations.

A whole subchapter is devoted to an interesting application to the novel theory of automatic voice recognition using Fourier transform. It can be seen from the study developed in this part that the calculus of Mel Frequency cepstral coefficients can be reduced to the solving of a difference equation. Moreover, these coefficients can be calculated more accurately with this method comparing to classical methods (such as, Linear Predictive Coding).

Of special practical importance, coming from the remark that ODEs are an essential feature in the mathematical modeling, is the discussion concerning the solving of ODEs via numerical schemes that lead to difference equations (see also [3]). Such numerical schemes are described and also their convergence, stability and the precision of the obtained approximation are rigorously studied.

Therefore, the case of ODEs with impulses at preassigned moments (which are, in fact, a mix between differential equations and difference equations) can also be studied through the theory of difference equations. Here the general theoretical results are only recalled from literature ([1]), emphasis being then put on the framework of linear differential equations with impulses where numerical schemes are detailed. It seems worthwhile to notice that impulsive problems occur in a huge number of fields (such as, biology, automatics, electronics and so on, see [2]).

The second chapter ends in the same (pedagogical) manner: with a list of solved or proposed to be solved exercises.

To conclude, the book under review has got all the premises to become a very useful tool for students in engineering, at master or PhD level, but also for researchers working in applied mathematics and related fields.

REFERENCES

[1] D. Bainov, P. Simionov, *Impulsive Differential Equations, Periodic Solutions and Applications*, Monographs Applied Mathematics, vol. 66, Langman Scientific Technical, John Wiley & Sons Inc., New-York, 1993.

[2] A.N. Sesekin, S.T. Zavalishchin, *Dynamic Impulse Systems*, Dordrecht, Kluwer Academic (1997).

[3] C.I. Gheorghiu, *Metode numerice pentru sisteme dinamice*, Casa cărții de știință, Cluj-Napoca, 2006.

NOTES FOR AUTHORS

Constantin POPESCU[*], Viorel VASILESCU[**]

[*]"Henri Coandă" Air Force Academy, Braşov, Romania (cpop@afahc.ro)
[**]Politehnica University, Bucharest, Romania (vvasil@edpub.ro)

Abstract: *Papers must be prefaced by an abstract up to 250 words. The text will be written in 11pt high, Italic, justified, left-right alignment. A number of maximum 8 keywords will be written 12pt below the abstract. The words will be 11pt high, Italic, left alignment, and separated by a comma.*

Keywords: *first keyword, second keyword, third keyword…*

1. INTRODUCTION

Page setup: Mirrored Margins: Top – 25.4 mm, Bottom – 25.4 mm, Inside – 31.75 mm, Outside – 25.4 mm, Header – 12.5 mm, Footer – 12.5 mm, Different odd and even pages, Different first page. Apply to whole document; Paper size: A4; Orientation: Portrait; Columns: One; Align to margin.

2. MAIN TITLE

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

3. AUTHORS

Names, style. Font: Times New Roman, 12 pt., bold, centered; first name − regular, family name − all caps; alignment: centered; spacing: after − 12 pt. A comma separates the names.

For a single affiliation, no superscript is necessary. In case of different authors, from different institutions, one marks every name by a superscript asterisks (like exponents). The legends of these superscript numbers (the affiliations and email address), will be indicated under authors lines. For affiliation use the style: Font: Times New Roman, 12 pt., regular, centered, after – two lines of 12 pt.

4. ABSTRACT AND KEYWORDS

Font: Times New Roman, 11 pt., italic. Paragraph: justify; line spacing: single.

The **Abstract** is followed by the **Keywords** and the spacing between abstract and the keywords is 11 pt.

5. CHAPTERS TITLES

The chapters are countered beginning with Arabic figures and printed in capitals (**1. XXXXXXXX)**, using the style: font: Times New Roman, 12 pt., bold, all caps; paragraph: alignment: centered; spacing: after − 12 pt.

6. PAPER TEXT

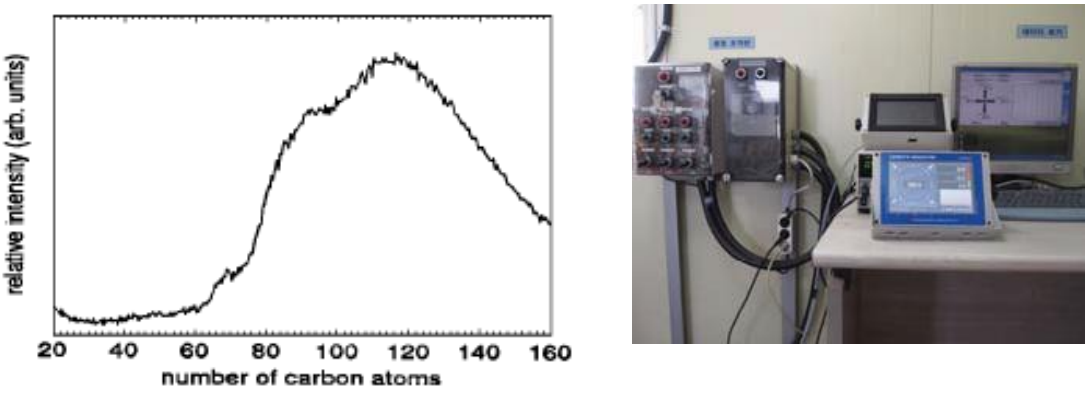
Font: Times New Roman, 12 pt. Paragraph: alignment: justified. Paragraphs will be 6 mm indented. Line spacing: single.

6.1 The main part of the text. Original and high-standard scientific papers shall be drawn up in a concise style, avoiding any oversized introduction.

7. FIGURES, TABLES

Figures and Tables shall be introduced at their appropriate place in the text and shall not be larger than a page width each. The legend of figures is included bellow the figure (centered) and for tables before (align text right), both with the style: Font: Times New Roman, 10 pt., regular; paragraph: spacing: before − 10 pt., after − 10 pt.

Landscape tables are not accepted. If you need to arrange a number of figures, a good tip is to place them in a table, which gives you additional control of the layout. Leave a line space (12 pt.) between your figure and any text above it, like this one:



(a) (b)

FIG. 1. The text "**FIG. 1.,**" which labels the caption, should be bold and in upper case. If figures have more than one part, each part should be labeled (a), (b), etc. Using a table, as in the above example, helps you control the layout

Cite all figures in the text consecutively. The word "Figure" should be spelled out if it is the first word of the sentence and abbreviated as "Fig." elsewhere in the text. Place the figures as close as possible to their first mention in the text at the top or bottom of the page with the figure caption positioned below, all centered. Figures must be inserted in the text and may not follow the Reference section.

Set figure captions in 10 point size, Times Roman font. Type the word "**FIG. 1**." in bold uppercase, followed by a period.

Authors are welcome to use color figures within their article.

Table 1. Example of table

	f_1	f_2	f_3	f_4
First set of values	0.8	0.6	0.4	0.2
Second set of values	1.1	1.0	0.9	0.8

8. EQUATIONS AND FORMULAS

It is strongly recommended to use a table with one row and two columns: in the first column, one writes the equation and in the second, the equation's number. Table: Insert table: number of columns: 2; number of rows: 1; alignment column 1: align text left, alignment column 2: align text right, format border: none, spacing after: 12 pt.,

$$f(x) = a_0 + \sum_{n=1}^{\infty}(a_n \cos\frac{n\pi x}{L} + b_n \sin\frac{n\pi x}{L}) \quad (1)$$

Table 2 – For formulas and equation, use only a Microsoft Equation Editor 3.0 with these settings:

Format + Spacing	Line spacing	150%
	Matrix row spacing	150%
	Matrix column spacing	100%
	Superscript height	45%
	Subscript height	25%
	Limit height	25%
	Limit depth	100%
	Spacing adjustment	100%
	Embellishment gap	1.5 pt.
Style + Define	Text	Times New Roman
	Function	Times New Roman
	Variable	Times New Roman italic
	LC. Greek	Symbol italic
	UC. Greek	Symbol
	Symbol	Symbol
	Matrix + Vector	Times New Roman bold
	Number	Times New Roman
Size + Define	Full	12 pt.
	Subscript/Superscript	7 pt.
	Sub-Subscript/Superscript	5 pt.
	Symbol	18 pt.
	Sub-Symbol	12 pt.

CONCLUSIONS

As the author and creator of your article (doc, docx), you have the most intimate knowledge of *exactly* what the file should display. We ask all authors to carefully check their article prior to submission. Perform visual inspections to detect subtle font errors and ensure that all fonts are embedded.

AKNOWLEDGMENT

Submission of an article implies that the work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder.

REFERENCES

The title, REFERENCES, will be printed as chapter title. For the references will be used the style: font: Times New Roman, regular, 10 pt.; paragraph alignment: justified, line spacing: single, hanging: 6mm, numbered.

The titles will be printed with *italic*, the volume number and the revue number will be printed with regular.

Inside the text, the books or the revues from REFERENCES are referred between square brackets: [1], [3, 4, 7].

The formats for references without authors and for referencing articles in journals, books, articles in books and proceedings are:

[1] *** European Aviation Safety Agency. *CS-25, Airworthiness codes for large aero-planes*, October 2003. Available at www.easa.eu.int, accessed on 10 Oct. 2015;
[2] C. Ohtar, A. Fujita, P. N. Nikiforov and M. K Santa, Active flutter suppression for two-dimensional airfoils, *Journal of Guidance, Control and Dynamics*, vol. 14, no. 2, pp. 283-293, 1991;
[3] T. R. Nail, *A disturbance-rejection problem for a 3-D airfoil exhibiting flutter*, Thesis, Virginia Tech., 2000;
[4] J. J. Black and H. Rilliatt, Active control of a structure, *AIAA Paper 99-0036;*
[5] R. Sind and M. Krenner, *Robust aeroservoelastic stability analysis*, Springer, London, 2000;
[6] P. Santorini, *Nonlinear Control Systems*, London, UK, Springer-Verlag, 3rd ed., 2005;
[7] R. Ceaikovski and P. Soznovski, *Neural approximation: a control perspective*, in Neural Network Engineering in Dynamic Control Systems, K. Runt, G. Irwin and F. Warwick (Eds.), Springer, 1995;
[8] F. Rudin, *The method of statistical linearization for nonlinear stochastic vibration*, in F. Ziegler and G. I. Schuller (Eds), *Nonlinear Stochastic Dynamic Engineering Systems*, pp. 45-56, IUTAM Symposium on Nonlinear stochastic dynamic engineering, Innsbruck, Austria, June 21-26, 1987, Springer Verlag, 1988.

Summary: Points to Consider when Preparing Your Paper

Well prepared papers enable rapid publication and reduce unnecessary work. The following points summarize the key issues you need to comply with when preparing your paper for Review of the Air Force Academy:

Do not add page numbers or footers. Our article templates **deliberately** do not include these, so please *do not add them*.

Do not alter the margins of our templates.

Prepare and format references with care. Please prepare and **format** your references in accordance with the examples supplied with the author templates and documentation.

Use clear, legible graphics and diagrams. Readers of your paper will be grateful. If they cannot read it, they are unlikely to cite it.

Do not use copyrighted material without permission. **Papers** using copyrighted material without appropriate permission and acknowledgment will be excluded.

Avoid large PDF files (10 MB maximum, ideally). For **the** benefit of your readers, we recommend keeping your article file below 10 MB. This is a recommendation, *not a requirement*.