4 AN UNBIASED ESTIMATOR FOR THE PARAMETER OF A HOMOGRAPHIC DISTRIBUTION USED IN ECONOMY

Poliana ŞTEFĂNESCU[*]
Ştefan ŞTEFĂNESCU[**]

Abstract

In a previous study, (Ştefănescu, P., Ştefănescu, Şt., 2006) we suggested two possible estimators for the unknown value of the parameter θ which characterize a homographic type $HG(\theta)$ distribution.

In the current paper we shall prove that the proposed estimators of θ based on the median of the r.v. X, $X \sim HG(\theta)$, always over- evaluate the real value of θ. For this reason we determined an adjusted multiplicative factor such that the median type estimators to become unbiased.

The theoretical results were confirmed experimentally by using a Monte Carlo stochastic simulation technique.

Key words: homographic distribution, unbiased estimator, Monte Carlo technique, generating random variables.

JEL Classification: C13, C15

1. Introduction

Other authors (Isaic-Maniu, Al., Vodă, V. Gh., 2005) have investigated the properties of a homographic type random variable (r.v.) X having $F(x;\theta)$ as cumulative distribution function (c.d.f.), where

$$F(x;\theta) = \frac{x}{x+\theta}, \quad x \geq 0 \quad , \quad \theta > 0 \qquad (1)$$

In the following we shall denote by $X \sim HG(\theta)$ if the r.v. X has the c.d.f. $F(x;\theta)$.

The distribution $HG(\theta)$ is well used for modeling some economic aspects, especially to describe the failure of different financial markets[1].

[*] Associate Professor, Ph.D., Faculty of Sociology and Social Work, University of Bucharest, e-mail : poliana@sas.unibuc.ro.

[**] Associate Professor, Ph.D., Faculty of Mathematics and Informatics, University of Bucharest, e-mail : stefanst@fmi.unibuc.ro.

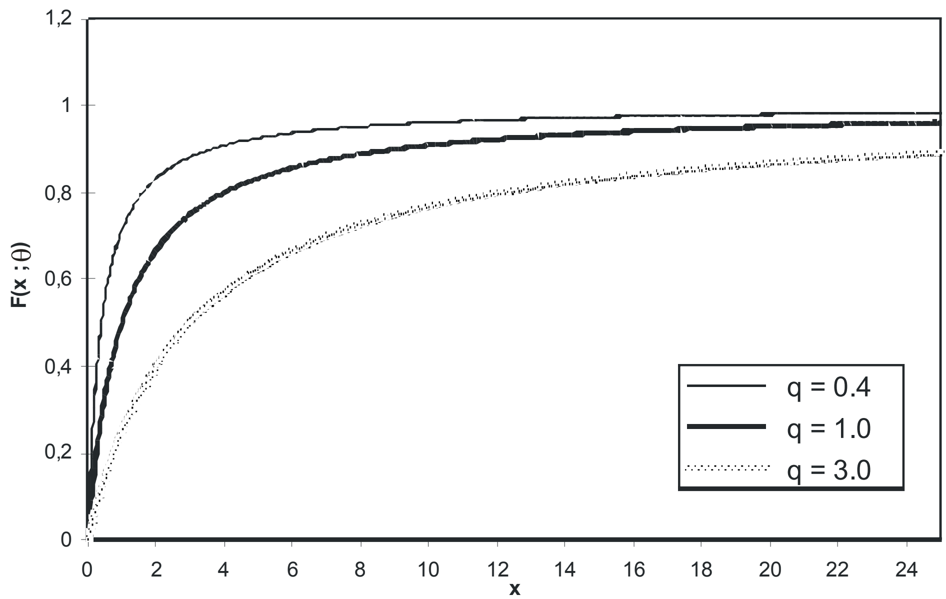
The probability density function (p.d.f.) $f(x;\theta)$, x ≥ 0, of the r.v. X, X ~ HG(θ), has the form

$$f(x;\theta) = \frac{\theta}{(x+\theta)^2}, \ x \geq 0, \ \theta > 0 \quad (2)$$

Depending on the values of the parameter θ, the c.d.f.-s $F(x;\theta)$ are enough different (see Figure 1), suitable to study peculiar situations.

Figure 1

The c.d.f. F(x ; θ) .



Therefore, it is very important to obtain good estimators of the parameter θ.

But, in our case, the classical moment method to estimate θ is in general unworkable since:

Proposition 1. For any a ≥ 1 and θ > 0, if X ~ HG(θ) then we have

$$Mean(X^a) = \infty \quad (3)$$

Proof: Indeed,

$$Mean(X^a) = \int_0^\infty x^a f(x;\theta)dx \geq \int_1^\infty x^a f(x;\theta)dx \geq \int_1^\infty x f(x;\theta)dx = \int_1^\infty \frac{x\theta}{(x+\theta)^2}dx =$$

[1] Isaic-Maniu, Alexandru, Vodă, Viorel Gh., “On a homographic distribution function”, *Economic Computation and Economic Cybernetics Studies and Research*, Vol. 39, No.1-4(2005), 11-18.

$$= \theta \left[-\frac{x}{x+\theta} + ln(x+\theta) \right] \Bigg|_{x=1}^{x=\infty} = \infty$$

Moreover, if $x_1, x_2, x_3, \dots, x_n$ are n independent observations from the r.v. X , X ~ HG(θ), then the maximum likelihood estimation[1] is reduced to find the real roots of a n-degree polynomial equation in θ[2]. For this reason, like the moment method also the standard maximum likelihood estimator procedure is not so easy to be applied[3].

But, for any X ~ HG(θ) , we get

$$Pr(X \le \theta) = F(\theta;\theta) = \frac{\theta}{\theta+\theta} = \frac{1}{2} \qquad (4)$$

Therefore, θ is just the median indicator of the r.v. X. This fact was used in a previous study[4] to estimate θ.

In the following, we shall try to establish some statistical properties of the estimators based on the median coefficient.

2. The Gross Median Type Estimator

For any sample $x_1, x_2, x_3, \dots, x_n$ obtained from the r.v. X, X ~ HG(θ) , the experimental median coefficient x^* has the form

$$x^* = \begin{cases} x_{(m+1)}; & if\ n = 2m+1 \\ (x_{(m)} + x_{(m+1)})/2; & if\ n = 2m \end{cases} \qquad (5)$$

where $x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n)}$ are just the values $x_1, x_2, x_3, \dots, x_n$ sorted in an increasing order, that is $x_{(1)} \le x_{(2)} \le x_{(3)} \le \dots \le x_{(n)}$.

By applying different statistical methods and using the computer we can generate independent random values $x_1, x_2, x_3, \dots, x_n$ of an arbitrary r.v. X[5]. More exactly:

Proposition 2. If U has an uniform distribution on (0, 1] interval, U ~ U((0 , 1]) , and

$$T = \frac{\theta(1-U)}{U} \qquad (6)$$

Then, T ~ HG(θ).

[1] *Papoulis, Athanasios,* Probability and statistics, *Prentice Hall, New Jersey, 1990.*

[2] See also Isaic-Maniu, Alexandru, Vodă, Viorel Gh., “On a homographic distribution function”, Economic Computation and Economic Cybernetics Studies and Research, Vol. 39, No.1-4(2005), 11-18.

[3] *More details in Ştefănescu, Poliana, Ştefănescu, Ştefan, “Estimating the parameter of a homographic distribution”,* Economic Computation and Economic Cybernetics Studies and Research*, Vol. 40, No. 1(2006), 10 pgs. (forthcoming).*

[4] *Ştefănescu, Poliana, Ştefănescu, Ştefan,* op cit.

[5] See for example Gentle, James E., *Random number generation and Monte Carlo methods,* Springer - Statistics and Computing, New York, 1998.

Proof: Let G(t) be the c.d.f. of the r.v. V. Then, for any $t \geq 0$, we shall show that G(t) is just $F(t\,;\theta)$, that is $T \sim HG(\theta)$. Indeed,

$$G(t) = Pr(T \leq t) = Pr\left(\frac{\theta(1-U)}{U} \leq t\right) = Pr\left(U \geq \frac{\theta}{t+\theta}\right) =$$

$$= \int_{\theta/(t+\theta)}^{1} du = 1 - \frac{\theta}{t+\theta} = \frac{t}{t+\theta} = F(t;\theta)$$

Remark 1. Any programming language has implemented specialized procedures to generate U((0, 1]) random variables[1]. Specifically, in Microsoft Excel language the name of this procedure is *Rand*.

Taking into consideration all the previous aspects we suggest the algorithm ***A1*(n,θ)** to validate the quality of the estimates x* given by (5).

Algorithm *A1*(n,θ) (the median procedure).

Step 0. Inputs: n, θ ($n \geq 2$, $\theta > 0$).

m = *Int*(n/2) (*Int*(λ) is the integer part of the real number λ)

Step 1. Generate n independent U((0 , 1]) random values $u_1, u_2, u_3, \dots, u_n$.

Step 2. $x_i = \dfrac{\theta(1-u_i)}{u_i}$, $1 \leq i \leq n$.

Step 3. Sort the values $x_1, x_2, x_3, \dots, x_n$ in an ascending order, at last resulting the quantities

$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \dots \leq x_{(n-1)} \leq x_{(n)} .$$

Step 4. If $n = 2m$ then $x^* = \dfrac{x_{(m)} + x_{(m+1)}}{2}$

else $x^* = x_{(m+1)}$.

Step 5. Output: x* (x* estimates the unknown value of θ).

STOP.

Tables 1 and 2 show the estimates $x_1^*, x_2^*, x_3^*, \dots, x_p^*$, p = 20 resulted after running consecutively p = 20 times the algorithm ***A1*(n,θ)** for n = 9 , θ = 10 and n = 8 , θ = 11, respectively.

Table 1

The estimations x* of θ obtained by applying the algorithm *A1*(9,10)

5.09	36.33	21.59	6.00	28.04	5.27	7.74	10.88	42.19	15.77
28.60	2.45	6.59	3.34	7.82	18.35	6.79	15.13	10.80	9.26

[1] *Gentle, James E.,* op cit.

Table 2

The estimations x* of θ resulted by running the algorithm A1(8,11)

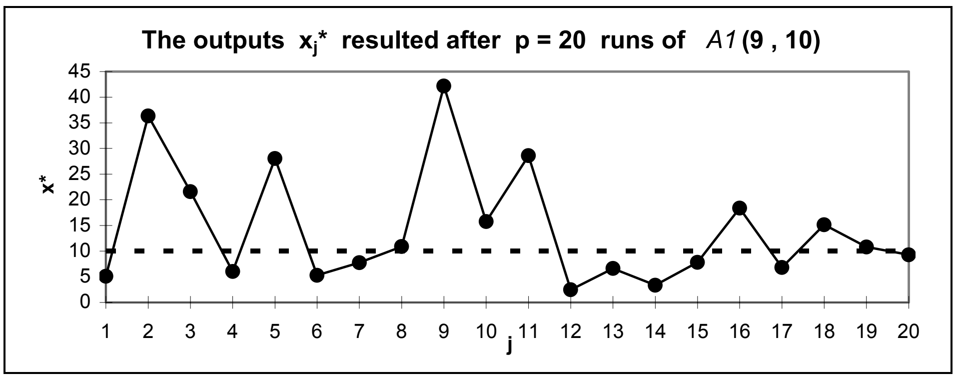
11.26	6.39	19.98	13.51	17.88	9.52	34.33	40.30	8.55	7.80
9.83	9.21	21.48	11.24	23.10	7.17	20.86	16.43	10.17	25.62

3. The Quality of the Estimates x*

Figures 2 and 3 present the variation of the x_j^* evaluation values for θ, $1 \le j \le p = 20$, which were taken from Tables 1 and 2.

Interpreting the Figures 2 and 3 we deduce that the estimations x* generally over evaluate the real value of θ. Thus, the quantities x_j^* are usually greater than θ = 10 in Figure 2 or they pass over the threshold θ = 11 in Figure 3.

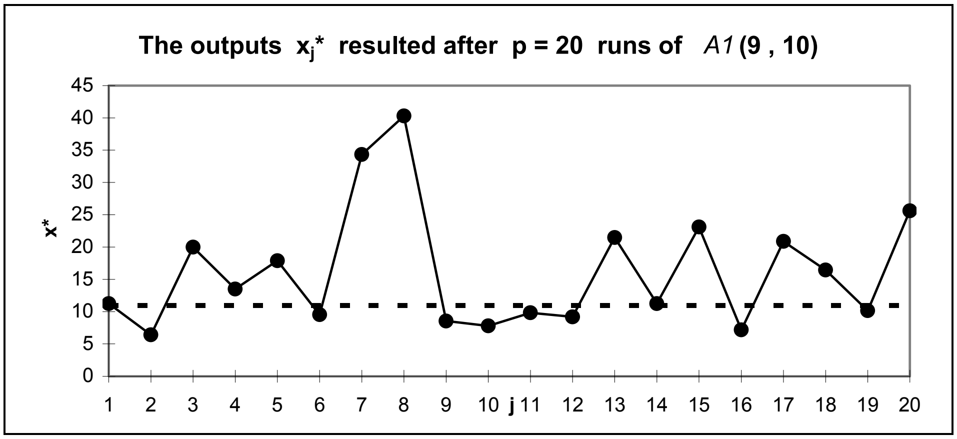
Figure 2



These aspects appear more clear and stable if instead of a single value x* we take into consideration the mean w of p consecutive estimations x_j^*, $1 \le j \le p$, resulted after running successively the algorithm *A1*(n,θ) , that is

$$w = \frac{x_1^* + x_2^* + x_3^* + \ldots + x_{p-1}^* + x_p^*}{p} \tag{7}$$

Figure 3



Reiterating q = 30 times the evaluation process of θ we got the quantities w_1, w_2, w_3, … , w_q presented in Tables 3 and 4.

Table 3

The means w_s, $1 \le s \le q$, given by (7) for n = 9, θ = 10, p = 20, q = 30

14.40	11.32	10.71	12.84	14.06	11.24	17.22	7.42	15.54	11.56
11.74	13.57	9.32	12.25	11.67	12.54	11.63	14.22	13.05	6.83
17.07	11.28	13.01	13.04	10.66	11.46	10.23	13.73	14.31	14.57

Table 4

The means w_s, $1 \le s \le q$, given by (7) for n = 8, θ = 11, p = 20, q = 30

16.23	15.15	22.10	11.98	11.65	10.16	14.15	13.28	15.36	15.00
13.57	12.10	18.31	15.80	14.08	16.31	11.82	10.28	17.01	15.79
13.52	13.90	15.35	13.65	14.76	13.49	16.62	17.22	13.64	14.32

We observe that almost always the quantities w_s, $1 \le s \le q = 30$, pass over the threshold θ (θ = 10 in Table 3 or θ = 11 in Table 4).
In conclusion, the brute estimates x* based on the median coefficient over evaluate the real value of the parameter θ.

4. Adjusting the Estimates x*

We saw that the x* estimations are biased. For this reason we shall try to determine theoretically the bias of the x_j^* quantities.

If X ~ HG(θ) , we denote by $X_{(k)}$ the k-order statistics considering samples of size n from X[1].

Keeping all the previous interpretations, the p.d.f. $f_k(x\ ;\ \theta)$ of an arbitrary k-order statistic $X_{(k)}$, $1 \le k \le n$, has the following expression[2]:

$$f_k(x;\theta) = \frac{n!}{(k-1)!(n-k)!}\left(F(x;\theta)\right)^{k-1}\left(1-F(x;\theta)\right)^{n-k} f(x;\theta) \tag{8}$$

When X ~ HG(θ) the formula (8) becomes

$$f_k(x;\theta) = \frac{n!}{(k-1)!(n-k)!}\frac{\theta^{n-k+1}x^{k-1}}{(x+\theta)^{n+1}}\ ,\ x \ge 0 \tag{9}$$

Remark 2. For any a, b ∈ **N**, the *B*(a,b) value for Euler's integral is given by the formula[3]:

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt = \frac{(a-1)!(b-1)!}{(a+b-1)!} \tag{10}$$

Proposition 3. If X ~ HG(θ) , then for any $1 \le k \le n$ we have

$$Mean(X_{(k)}) = \frac{k\theta}{n-k} \tag{11}$$

Proof: If $t = \dfrac{x}{x+\theta}$ then we deduce $x = \dfrac{t\,\theta}{1-t}$ and $x+\theta = \dfrac{\theta}{1-t}$.

Making the substitution $t = \dfrac{x}{x+\theta}$ and using Remark 2 we get successively

$$Mean(X_{(k)}) = \int_0^\infty x f_k(x;\theta)dx = \frac{n!\theta^{n-k+1}}{(k-1)!(n-k)!}\int_0^\infty \frac{x^k}{(x+\theta)^{n+1}}dx =$$

$$= \frac{n!\theta^{n-k+1}}{(k-1)!(n-k)!}\int_0^1 t^k(1-t)^{n-k-1}\theta^{k-n}dt = \frac{n!\theta}{(k-1)!(n-k)!}B(k+1,n-k) =$$

$$= \frac{n!\theta}{(k-1)!(n-k)!}\frac{k!(n-k-1)!}{n!} = \frac{k\theta}{n-k}$$

Proposition 4. If X ~ HG(θ) and the r.v. Y is given by

---

[1] *Mihoc, Gheorghe, Ciucu, George, Craiu, Virgil,* Teoria probabilităților şi statistică matematică, *Editura Didactică şi Pedagogică, Bucureşti, 1970.*

[2] *Mihoc, Gheorghe, Ciucu, George, Craiu, Virgil,* op cit, *p. 472.*

[3] *Fihtenholt, G.M.,* Calcul diferențial şi integral, *Editura Tehnică, Bucureşti, 1964 (translation from Russian), p. 690.*

$$Y=\frac{m}{m+1}X_{(m+1)} \tag{12}$$

with $n=2m+1$, then $Mean(Y)=\theta$.

Proof: Applying Proposition 3 for $n=2m+1$ we obtain

$$Mean(Y)=Mean\left(\frac{m}{m+1}X_{(m+1)}\right)=\frac{m}{m+1}Mean\left(X_{(m+1)}\right)=\frac{m}{m+1}\frac{(m+1)\theta}{n-(m+1)}=$$

$$=\frac{m}{m+1}\frac{(m+1)\theta}{(2m+1)-(m+1)}=\theta$$

Proposition 5. If X ~ HG(θ) and the r.v. Y has the form

$$Z=\frac{m-1}{m}\frac{X_{(m)}+X_{(m+1)}}{2} \tag{13}$$

with $n=2m$, then $Mean(Z)=\theta$.

Proof: Indeed, using Proposition 3 for $n=2m$ we deduce in order

$$Mean(Z)=Mean\left(\frac{m-1}{m}\frac{X_{(m)}+X_{(m+1)}}{2}\right)=\frac{m-1}{2m}Mean\left(X_{(m)}+X_{(m+1)}\right)=$$

$$=\frac{m-1}{2m}\left(Mean(X_{(m)})+Mean(X_{(m+1)})\right)=\frac{m-1}{2m}\left(\frac{m\theta}{n-m}+\frac{(m+1)\theta}{n-(m+1)}\right)=$$

$$=\frac{m-1}{2m}\left(\frac{m\theta}{2m-m}+\frac{(m+1)\theta}{2m-(m+1)}\right)=\frac{(m-1)\theta}{2m}\left(1+\frac{m+1}{m-1}\right)=\theta$$

Remark 3. Propositions 4 and 5 suggest that the r.v.-s Y and Z can be used as unbiased estimators for the unknown parameter θ when the size of the experimental sample is odd, respectively even.

Thus, the statistical quality of the initial estimations x* is clearly improved if the values of all these estimations are adjusted by a multiplicative coefficient γ, where

$$\gamma=\begin{cases} m/(m+1); if\ n=2m+1 \\ (m-1)/m; if\ n=2m \end{cases} \tag{14}$$

Practically, for estimating θ it is better to use the transformed values $\gamma x*$ instead of x*.

Tables 5 and 6 contains the values $v_s=\gamma w_s$ resulted after applying a correction operation to the initial means w_s of the median type estimations x*. The values v_s in Table 5 or Table 6 are more suitable to estimate θ = 10, respectively θ = 11 (for example, compare the v_s adjusted estimations with the corresponding value of θ; see also Figures 5-6 which present the variation of the v_s transformed quantities, 1 ≤ s ≤ 30).

Table 5

The values v_s obtained after the correction of the w_s quantities from Table 3

($v_s = \gamma\, w_s$, $1 \le s \le q = 30$, $\gamma = 0.8$, $n = 9$)

11.52	9.06	8.57	10.27	11.25	8.99	13.78	5.94	12.43	9.25
9.39	10.86	7.46	9.80	9.34	10.03	9.30	11.38	10.44	5.46
13.66	9.02	10.41	10.43	8.53	9.17	8.18	10.98	11.45	11.66

Table 6

The values vs obtained after the correction of the ws quantities from Table 4

($v_s = \gamma\, w_s$, $1 \le s \le q = 30$, $\gamma = 0.75$, $n = 8$)

12.17	11.36	16.58	8.99	8.74	7.62	10.61	9.96	11.52	11.25
10.18	9.08	13.73	11.85	10.56	12.23	8.87	7.71	12.76	11.84
10.14	10.43	11.51	10.24	11.07	10.12	12.47	12.92	10.23	10.74

Figure 4

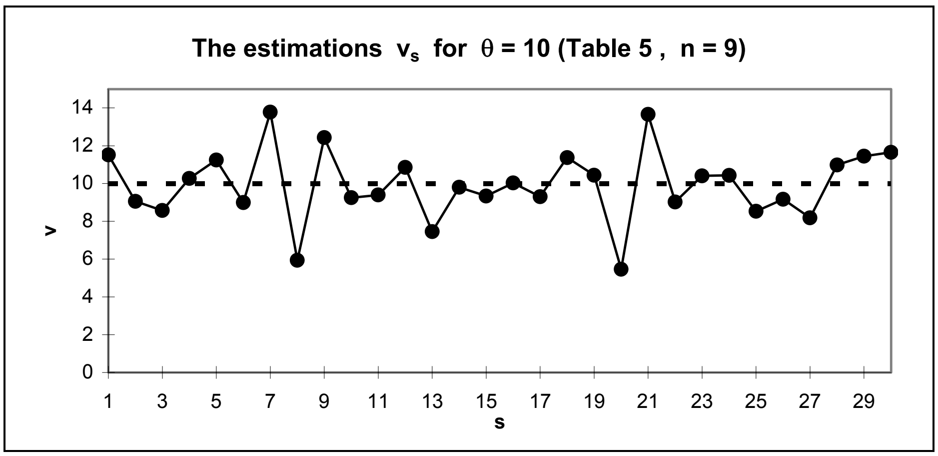
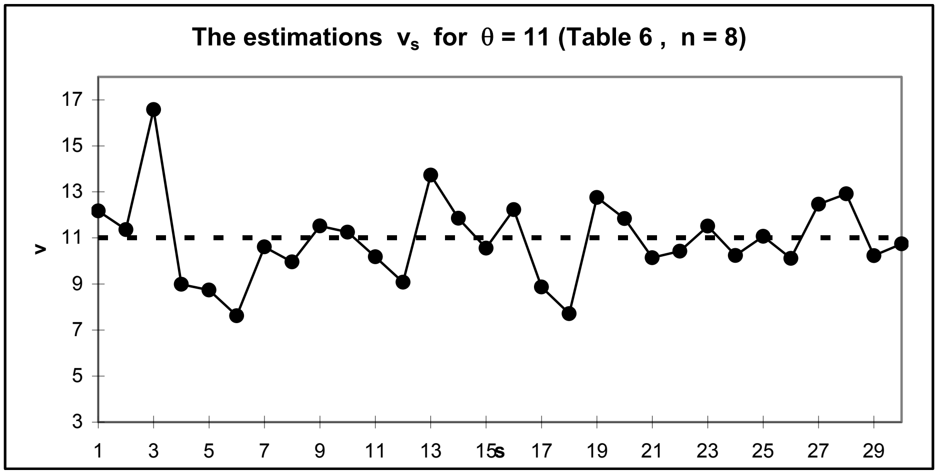


Figure 5



5. Conclusions

The homographic HG(θ) distribution has many applications, especially to simulate the failure of different financial markets[1].

In another study[2] we proposed two effective procedures to estimate the unknown value of the parameter θ which characterizes the homographic HG(θ) distribution. One of these estimation procedures was based on the median coefficient.

The present work shows that the median type estimators suggested in the above-mentioned study[3], as for example $X_{(m+1)}$ for $n = 2m + 1$ or $\frac{X_{(m)} + X_{(m+1)}}{2}$ when $n = 2m$, are always biased (see Tables 3-4 and Figures 2-3).

We also proved that the r.v.-s $Y = \frac{m}{m+1} X_{(m+1)}$, $Z = \frac{m-1}{m} \frac{X_{(m)} + X_{(m+1)}}{2}$ can be used as unbiased estimators for the parameter θ when $n = 2m + 1$, respectively $n = 2m$ (Propositions 4-5).

Therefore, adjusting the initial median type estimator proposed in our previous study[4] with a multiplicative factor γ given by (14) we obtain invariable an unbiased estimator for θ.

---

[1] Isaic-Maniu, Alexandru, Vodă, Viorel Gh., "On a homographic distribution function", *Economic Computation and Economic Cybernetics Studies and Research*, Vol. 39, No.1-4(2005), 11-18.

[2] *Ştefănescu, Poliana, Ştefănescu, Ştefan, "Estimating the parameter of a homographic distribution",* Economic Computation and Economic Cybernetics Studies and Research*, Vol. 40, No. 1(2006), 10 pgs. (forthcoming).*

[3] *Ştefănescu, Poliana, Ştefănescu, Ştefan,* op cit.

[4] Op cit.

The theoretical results were confirmed experimentally by applying a Monte Carlo stochastic simulation technique (compare Tables 3-4 with Tables 5-6; see also the Figures 5-6 with the graphic representation of v_s adjusted quantities, $1 \le s \le q = 30$).

References

Fihtenholt, G.M., *Calcul diferenţial şi integral,* Editura Tehnică, Bucureşti, 1964 (translation from Russian).

Gentle, James E., *Random number generation and Monte Carlo methods,* Springer - Statistics and Computing, New York, 1998.

Isaic-Maniu, Alexandru, Vodă, Viorel Gh., "On a homographic distribution function", *Economic Computation and Economic Cybernetics Studies and Research*, Vol. 39, No.1-4(2005), 11-18.

Mihoc, Gheorghe, Ciucu, George, Craiu, Virgil, *Teoria probabilitaţilor şi statistică matematică,* Editura Didactică şi Pedagogică, Bucureşti, 1970.

Papoulis, Athanasios, *Probability and statistic.* Prentice Hall, New Jersey, 1990.

Ştefănescu, Poliana, Ştefănescu, Ştefan, "Estimating the parameter of a homographic distribution", *Economic Computation and Economic Cybernetics Studies and Research*, Vol. 40, No. 1(2006), 10 pgs. (forthcoming).