8 HOW MUCH THE ROUNDING ERRORS COULD AFFECT THE COMPUTER RESULTS

Ştefan ŞTEFĂNESCU*

Abstract

The paper emphasizes the specific manner to perform the arithmetic operations in the computer as always it is possible to have "rounding errors".

Neglecting the influence of the computer rounding errors could affect a right conclusion, for example, in the case of an account balance verification.

Key words : rounding errors, Monte Carlo technique, specific distributions.

JEL Classification System : C87, C15, C16.

1. The Problem Formulation

The numerical values are approximated by the computer by retaining only a finite number q of significant digits. For this reason the results of the computer arithmetic operations are rounded to the first q important figures. For example, the *Excel* spreadsheet operates in the arithmetic calculus with $q = 15$ significant figures.

We can obtain wrong results if we don't take into consideration the possibility to have errors in the rounding process of computer outcome. We intend to give a quantitative image regarding the influence of the computer rounding errors.

For the simplicity of the exposure we'll treat the ordinary case of solving a linear system which has only two equations, the variables x, y being unknown, that is

$$\begin{cases} a_1 x + a_2 y = a_3 \\ a_4 x + a_5 y = a_6 \end{cases} \quad (1)$$

The model (1) is characterized by the input parameters a_i , $1 \le i \le 6$.

Obviously, the solution (x_0, y_0) of the system (1) has the form:

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

$$x_0 = \frac{a_5 a_3 - a_2 a_6}{a_1 a_5 - a_2 a_4} \qquad y_0 = \frac{a_1 a_6 - a_4 a_3}{a_1 a_5 - a_2 a_4} \tag{2}$$

The presence of computer rounding errors could induce an unexpected effect, that is the "solution" (x_0, y_0) obtained with formulas (2) don't necessary verify both equations (1).

We'll define the indicator W to determine the frequency of this "surprising" non concordance aspect . More precisely, the variable W takes the values 0 , 1 or 2 when, respectively, the quantities x_0, y_0 satisfy the both equations (1), verify only one of the relations (1), or none of the equalities (1) are fulfilled.

In the following we'll estimate the probability that the situations $W = k$, $k \in (0, 1, 2\}$ appear.

2. The Distribution of Input Data

We intend to determine the frequencies of the events $W = k$, $k \in \{0, 1, 2\}$, when n systems of the type (1) were taken into consideration, the values of the input parameters a_i , $1 \le i \le 6$, being chosen randomly.

So, we'll suppose that the quantities a_i are observations of the random variables (r.v.) A_i , $1 \le i \le 6$, the random vector $(A_1, A_2, A_3, A_4, A_5, A_6)$ having a known cumulative distribution function (c.d.f.) $F_A(a_1, a_2, a_3, a_4, a_5, a_6)$.

Below the random variables A_i will have a power type distribution Pow(λ) or also a standard normal distribution Nor$(0, 1)$.

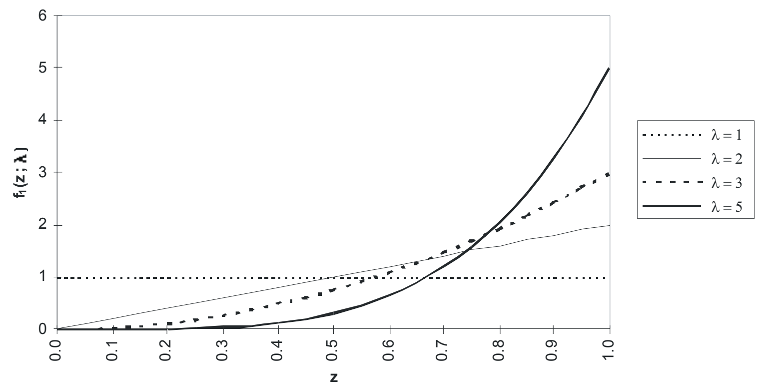
Definition 1. The r.v. Z has a power distribution Pow(λ) , Z ~ Pow(λ) , $\lambda > 0$, if its probability density function (p.d.f.) $f_1(z; \lambda)$ is given by the expression:

$$f_1(z; \lambda) = \lambda z^{\lambda - 1}, \quad 0 \le z \le 1 \tag{3}$$

The graphic **G1** presents the variation in the p.d.f. $f_1(z; \lambda)$, $0 \le z \le 1$, for different values of the shape power distribution parameter $\lambda \ge 1$ ($\lambda \ge 1$).

Graphic G1

The p.d.f. F$_1$(z; λ)



Remark 1. Interpreting the graphic **G1** we deduce that the r. v. Z ~ Pow(λ) with $\lambda > 1$ generates more frequently relative higher z random values.

In general we'll consider that the random variables A_i , $1 \le i \le 6$, are identically distributed and independent too. But we'll also treat the situation when two arbitrary input random variables, A_i and A_j, are correlated, that is $Cor(A_i, A_j) = \rho$ with $-1 \le \rho \le 1$.

Definition 2. We designate by (Z,V) ~ Nor2(ρ) if the random vector (Z,V) has a bivariate normal distribution with Z ~ Nor$(0,1)$, V ~ Nor$(0,1)$ and $Cor(Z,V) = \rho$, $-1 \le \rho \le 1$.

In the subsequent we'll estimate the probabilities $Pr(W = k)$, $k \in \{0,1,2\}$, considering the following variants :

Example 1. The input random variables A_i , $1 \le i \le 6$, of model (1) are independent and identically distributed with A_i ~ Pow(λ) , $\lambda > 0$.

Example 2. The random vectors (A_1, A_4), (A_2, A_5), (A_3, A_6) are independent and in addition (A_1, A_4) ~ Nor2(ρ), (A_2, A_5) ~ Nor2(ρ), $-1 \le \rho \le 1$, (A_3, A_6) ~ Nor2(0).

If $\rho = 0$ than all the random variables A_i, $1 \le i \le 6$, are independent and have a standard normal distribution Nor$(0,1)$.

3. A Monte Carlo Simulation

The difficulty of the theoretical evaluation of the probabilities $Pr(W = k)$, $k \in \{0,1,2\}$, arises from the relative complexity of the formulas (2) and also from the sophisticated form of the multivariate distribution $F_A(a_1, a_2, ..., a_6)$ which characterizes the random vector $(A_1, A_2, ..., A_6)$. Moreover, we'll impose a fixed number m of decimal figures for the input parameters a_i of the model (1).

For this reason we'll avoid a theoretical approach, preferring the computer Monte Carlo simulations. The stochastic simulation algorithm ***ASim*** determines the frequencies S_k to create the situations $W = k$, $k \in \{0, 1, 2\}$, when n linear systems of type (1) are solved.

Algorithm ASim (computes the frequencies S_k for appearing of the events $W = k$).

Step 0. Inputs : The number n of the type (1) systems used in the evaluation process.

C.d.f. $F_A(a_1, a_2, ..., a_6)$ for the random vector $(A_1, A_2, ..., A_6)$.

The number m of the decimal places imposed for the parameters a_i , $1 \le i \le 6$:

Step 1. $t = 0$ (t counts the number of systems (1) which are solved)

$$S_0 = 0 \qquad S_1 = 0 \qquad S_2 = 0$$

Step 2. $t = t + 1$

If $t > n$ then Print S_0, S_1, S_2. STOP

Step 3. Generate a random vector $(a_1, a_2, ..., a_6)$ having the c.d.f. $F_A(a_1, a_2, ..., a_6)$

Step 4. Determine the approximations a_i^* (with m decimal places) for a_i values , $1 \le i \le 6$

Step 5. $x_0 = \dfrac{a_5^* a_3^* - a_2^* a_6^*}{a_1^* a_5^* - a_2^* a_4^*}$ $\qquad$ $y_0 = \dfrac{a_1^* a_6^* - a_4^* a_3^*}{a_1^* a_5^* - a_2^* a_4^*}$

Step 6. $W = 0$

If $a_1^* x + a_2^* y = a_3^*$ then $W = W + 1$

If $a_4^* x + a_5^* y = a_6^*$ then $W = W + 1$

Step 7. If $W = 0$ then $S_0 = S_0 + 1$

If $W = 1$ then $S_1 = S_1 + 1$

If $W = 2$ then $S_2 = S_2 + 1$

Go to *Step 2.*

Remark 2. The simulation algorithm ***ASim*** was implemented in *Excel* spreadsheet. To obtain observations from a uniform $[0, 1]$ distributed r.v. U , U ~ Uni$([0, 1])$, we run the *Excel* procedure *Rand*. The generation of the random values Z, Z ~ Uni$([0, 1])$, is based on the inverse method (Gentle [3], p. 42-43).

Applying the following proposition we reduce the generation of the random values Z, Z ~ Pow(λ) , $\lambda > 0$, by producing observations from U ~ Uni($[0,1]$) .

Proposition 1. If the r.v. U has a uniform $[0,1]$ distribution and $Z = U^{1/\lambda}$, $\lambda > 0$, then Z ~ Pow(λ) .

Proof. We'll compute the c.d.f. $F_1(z;\lambda)$ of the r.v. $Z = U^{1/\lambda}$ for an arbitrary $0 \le z \le 1$. So,

$$F_1(z) = Pr(Z \le z) = Pr(U^{1/\lambda} \le z) = Pr(U \le z^{\lambda}) = z^{\lambda}$$

Therefore the p.d.f. $f_1(z;\lambda)$ of the r.v. Z is given by the expression

$$f_1(z;\lambda) = \frac{\partial F_1(z;\lambda)}{\partial z} = \lambda z^{\lambda - 1}$$

that is Z ~ Pow(λ) .

The generation of the random vector (Z,V), (Z,V) ~ Nor2(ρ), $-1 \le \rho \le 1$, is based on the following result :

Proposition 2. Let $F_0(z)$, $x \in R$, be the c.d.f. of the r.v. Z, Z ~ Nor$(0,1)$ (Laplace function). If $-1 \le \rho \le 1$, U_1, U_2 are independent random variables having a uniform $[0,1]$ distribution and

$$Z = F_0^{-1}(U_1)$$

$$V = \rho Z + \sqrt{1-\rho^2}\, F_0^{-1}(U_2) \qquad (4)$$

then (Z,V) ~ Nor2(ρ) .

Proof. Obviously, the r.v.-s $Z = F_0^{-1}(U_1)$ and $F_0^{-1}(U_2)$ are independent and identically Nor$(0,1)$ distributed (Gentle [3], p.42-43, the inverse method). In this conditions the r.v. V defined by formula (4) is Nor$(0,1)$ distributed (Papoulis [6]) and, in addition, the random vector (Z,V) has a bivariate normal distribution (Papoulis [6], p.162-164 ; Gentle [3], p.105-109). More,

$$Cor(Z,V) = \frac{Cov(Z,V)}{\sqrt{Var(Z)}\sqrt{Var(V)}} = \frac{Cov(Z, \rho Z + \sqrt{1-\rho^2}\, F_0^{-1}(U_2))}{\sqrt{\rho^2 Var(Z) + (1-\rho^2) Var(F_0^{-1}(U_2))}} =$$

$$= \frac{\rho Cov(Z,Z) + \sqrt{1-\rho^2}\; Cov(Z, F_0^{-1}(U_2))}{\sqrt{\rho^2 + (1-\rho^2)}} = \rho Var(Z) = \rho$$

which prove that (Z,V) ~ Nor2(ρ).

To increase the accuracy of Monte Carlo estimations, the algorithm ***ASim*** will be run successively many times.

Tables 1a, 1b, 1c present the resulted frequencies S_k that the events $W = k$, $k \in \{0, 1, 2\}$ appear, where $S_0 + S_1 + S_2 = n = 1000$. The input parameters a_i of ***ASim*** simulation procedure respects one of the restrictions mentioned in *Examples 1* or *2*. Constantly, the simulation algorithm ***ASim*** was applied 10 times, the index s designating the current simulation execution, $1 \le s \le 10$.

The estimation of the probabilities $Pr(W = k)$, $k \in \{0, 1, 2\}$, taking into consideration the average results of all 10 simulations, are given in *Tables 2a, 2b, 2c.* In fact *Tables 2a, 2b, 2c* convert in percent the mean of the same frequency S_k for all simulation values listed in the similar *Tables 2a, 2b,* respectively *2c*.

Table 1a

The frequency of the event $W = k$, $k \in \{0, 1, 2\}$, at the simulation s

(Example 1, $n = 1000$, $m = 20$)

λ	W	s=1	s=2	s=3	s=4	s=5	s=6	s=7	s=8	s=9	s=10
1/5	0	576	563	563	591	589	598	629	549	549	582
	1	419	424	426	398	403	391	369	441	438	408
	2	5	13	11	11	8	11	2	10	13	10
1/4	0	619	588	622	647	610	614	613	640	647	628
	1	368	403	362	337	384	374	377	349	344	359
	2	13	9	16	16	6	12	10	11	9	13
1/3	0	678	692	651	652	672	689	669	679	682	677
	1	307	293	329	331	305	290	319	309	308	307
	2	15	15	20	17	23	21	12	12	10	16
1/2	0	761	750	720	759	791	758	760	776	734	743
	1	213	217	254	212	181	213	218	202	249	233
	2	26	33	26	29	28	29	22	22	17	24
1	0	863	838	821	848	843	853	840	856	826	843
	1	94	106	125	104	88	102	105	103	114	102
	2	43	56	54	48	69	45	55	41	60	55
2	0	806	821	812	811	814	826	826	810	809	810
	1	64	73	81	75	80	66	73	63	74	75
	2	130	106	107	114	106	108	101	127	117	115
3	0	746	735	742	751	741	761	742	739	753	764
	1	60	68	71	68	84	67	73	76	66	74
	2	194	197	187	181	175	172	185	185	181	162
4	0	690	677	698	671	665	676	711	696	653	699
	1	74	65	84	76	78	78	60	71	87	68

	2	236	258	218	253	257	246	229	233	260	233
5	0	654	615	611	653	642	630	636	625	647	644
	1	70	78	79	82	77	71	70	60	76	87
	2	276	307	310	265	281	299	294	315	277	269

Analyzing the values from the mentioned tables we conclude :

- The frequencies S_0, S_1, S_2 for the events $W=0, W=1, W=2$ are very distinct and depend on the specific form of the c.d.f. $F_A(a_1, a_2, ..., a_6)$ which characterizes the parameters a_i of the model (1) (*Tables 1a, 2a, 1b, 2b*).

Moreover, we also remark important variations of the frequency S_0 when we change the distribution of the random variables A_i , $1 \le i \le 6$, but preserving the independence of these variables (*Tables 1a, 1b* ; *Tables 2a, 2b* - case $\rho = 0$).

Table 1b

The estimation of $Pr(W=k)$, $k \in \{0, 1, 2\}$, from 10 simulations

(Example 1, $n = 1000$, $m = 20$)

W	λ=1/5	λ=1/4	λ=1/3	λ=1/2	λ=1	λ=2	λ=3	λ=4	λ=5
0	0.5789	0.6228	0.6741	0.7552	0.8431	0.8145	0.7474	0.6836	0.6357
1	0.4117	0.3657	0.3098	0.2192	0.1043	0.0724	0.0707	0.0741	0.0750
2	0.0094	0.0115	0.0161	0.0256	0.0526	0.1131	0.1819	0.2423	0.2893

Table 2a

The frequency of the event $W = k$, $k \in \{0, 1, 2\}$, at the simulation s

(Example 2, $n = 1000$, $m = 20$)

ρ	W	s=1	s=2	s=3	s=4	s=5	s=6	s=7	s=8	s=9	s=10
0.0	0	801	822	804	810	817	821	814	814	836	808
	1	180	163	177	174	163	163	167	176	151	177
	2	19	15	19	16	20	16	19	10	13	15
0.4	0	790	809	785	787	810	820	814	813	817	788
	1	183	169	193	194	163	156	173	172	168	196
	2	27	22	22	19	27	24	13	15	15	16
0.8	0	769	737	750	731	768	780	765	758	748	776
	1	206	229	211	219	205	189	193	212	216	185
	2	25	34	39	50	27	31	42	30	36	39
0.9	0	718	695	738	703	737	730	721	713	713	705
	1	230	255	223	245	214	225	232	238	243	246
	2	52	50	39	52	49	45	47	49	44	49
0.95	0	696	681	689	684	687	684	702	670	635	651
	1	238	261	249	246	250	258	236	265	292	282

	2	66	58	62	70	63	58	62	65	73	67

- An imposed strong correlation coefficient ρ between the parameters a_i, $1 \le i \le 6$, affects essentially the frequencies S_0, S_1, S_2 of the events $W = 0, W = 1, W = 2$ (*Tables 2a* and *2b*).
- Contrary, the use of the random inputs a_i, $1 \le i \le 6$, which have an imposed number m of decimal places do not often affect much the frequencies S_k, $k \in \{0, 1, 2\}$ (*Tables 3a, 3b*). This aspect could be explained by the fact that the situations $W = k$, $k \in \{0, 1, 2\}$, depend basically by the complexity of the formulas (2), being secondary affected by the accuracy of the approximations a_i^* of the values a_i, $1 \le i \le 6$.

Table 2b

The estimation of $Pr(W = k)$, $k \in \{0, 1, 2\}$, from 10 simulations

(Example 2, $n = 1000$, $m = 20$)

	ρ = 0.0	ρ = 0.4	ρ = 0.8	ρ = 0.9	ρ = 0.95
W = 0	0.8147	0.8033	0.7582	0.7173	0.6779
W = 1	0.1691	0.1767	0.2065	0.2351	0.2577
W = 2	0.0162	0.0200	0.0353	0.0476	0.0644

Table 3a

The frequency of the event $W = k$, $k \in \{0, 1, 2\}$, at the simulation s

(Example 1, $n = 1000$, $\lambda = 1$)

m	W	s=1	s=2	s=3	s=4	s=5	s=6	s=7	s=8	s=9	s=10
12	0	838	812	824	819	858	822	810	834	834	833
	1	111	124	120	122	99	125	135	109	102	114
	2	51	64	56	59	43	53	55	57	64	53
13	0	819	825	830	830	837	814	825	824	835	823
	1	135	125	118	114	112	121	130	114	124	131
	2	46	50	52	56	51	65	45	62	41	46
14	0	828	831	829	844	825	832	839	834	836	834
	1	119	114	120	106	115	114	117	115	114	118
	2	53	55	51	50	60	54	44	51	50	48
15	0	835	821	843	820	821	841	845	821	822	838
	1	118	128	106	122	131	115	106	125	123	114
	2	47	51	51	58	48	44	49	54	55	48
20	0	843	863	831	864	845	853	860	844	839	839
	1	109	93	113	97	102	111	88	103	104	111

	2	48	44	56	39	53	36	52	53	57	50

Table 3b

The estimation of $Pr(W = k)$, $k \in \{0, 1, 2\}$, from 10 simulations

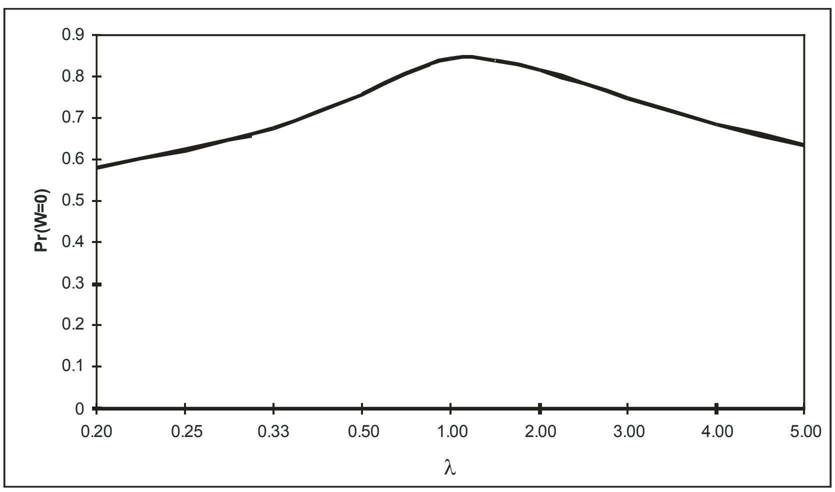
(Example 1, $n = 1000$, $\lambda = 1$)

W	m = 12	m = 13	m = 14	m = 15	m = 20
0	0.8284	0.8262	0.8332	0.8307	0.8481
1	0.1161	0.1224	0.1152	0.1188	0.1031
2	0.0555	0.0514	0.0516	0.0505	0.0488

- Graphic ***G2*** presents suggestively the variation in the probabilities $Pr(W = 0)$ depending on the shape parameter λ which defines the power distribution for the random values a_i , $1 \le i \le 6$ (*Table 1b*). Interpreting simultaneously the graphics ***G1*** and ***G2*** we remark a "monotony relation" between the probabilities $Pr(W = 0)$ and the λ values ($0 < \lambda \le 1$ or $\lambda \ge 1$; see also *Remark 1*).

Graphic G2

The estimation of Pr(W=0)



4. Concluding Remarks

The computer process which determines the solution (x_0, y_0) of the system (1) is affected by the rounding errors. The influence of the computer approximation errors is emphasized by the non-zero frequencies that the events $W = 1$ or $W = 2$ appear.

Surprising, but perfectly explained, is the fact that the probability $Pr(W = 0)$, to solve "correctly" with the computer the system (1) , is not closed to the maximal value 1. *Tables 1a, 1b, 1c* mention more situations where $Pr(W = 0) \approx 0.5$.

Obviously, the probability $Pr(W = 0)$ depends on the distribution $F_A(a_1, a_2, ..., a_6)$ of the parameters $a_1, a_2, ..., a_6$ which defines system (1). So, we studied the situation when the r.v. A_i were independent, being analyzed the correlation aspect too. The tables *1a-1c*, *2a-2c* and the graphs ***G1-G2*** give us a quantitative image regarding the influence of computer rounding errors.

The present paper warns about the possibility to produce errors even if we use a computer. Neglecting in practice the influence of the computer rounding errors could imply a lot of wrong decisions (see, for example, the computational errors in accounting, [8]).

References

Paul Bratley, Bennett L. Fox, Linus E. Schrage, *A guide to simulation.* Springer Verlag, New York, 1987 (second edition).

B.P. Demidovich, I.A. Maron, *Computational mathematics.* Mir Publishers, Moscow, 1981.

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

Donald E. Knuth, *The art of computer programming - Seminumerical algorithms* (volume 2)*.* Addison-Wesley Publishing Company, Reading, Massachusetts, 1981 (second edition).

V. Koroliouk, N. Portenko, A. Skorokhod, A. Tourbine, *Aide mémoire de théorie des probabilités et de statistique mathématique.* Editions Mir, Moscou, 1983.

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

Ron Person, *Special edition using Excel for Windows 95.* Teora Publishing House, Bucharest, 1998 (in Romanian).

Ştefan Ştefănescu, *Computational errors in accounting.* Studii şi Cercetări de Calcul Economic şi Cibernetică Economică, vol. XXXIII, nr. 2 (1999), 67-74 (in Romanian).