Rev. Soc. Chil. Estad.
9(2): 79–95, 1991

# Reporting Monte Carlo experiences in statistics: suggestions and an example

Oscar H. Bustos & Alejandro C. Frery [2]

**Abstract**

The minimal content of every report about statistical results obtained by stochastic simulation are suggested. A manner to present these results is shown. This is applied to a Monte Carlo experience designed to compare estimation procedures of the parameter of a Raighley distribution, a problem often encountered in synthetic aperture radar image processing.

*Key Words:* estimation, Monte Carlo methods, robustness, SAR image processing, stochastic simulation.

# 1. Introduction

Monte Carlo experiences are a powerful statistical technique used to provide approximate answers for questions about complex problems that may include a stochastic component. These simulation techniques rest upon the theory of controlled statistical sampling, They are particularly important when analytic and numeric techniques fail to supply an complete and/or exact answer to the problem at hand. Their range of applications is very wide indeed, including statistical mechanics, biology, games, combinatorial optimization, engineering, to name but a few areas.

There are several ways to summarize the behaviour of a sample of uni- or multi-dimensional data, either real or simulated data. Each technique is well suited for the enhancement of certain structures in the data: means, variances, quantiles, etc. and for the testing of different hypothesis about the underlying population.

It is commonplace, when simulation is used, to have several samples to be analyzed. Every sample could have been obtained, for example, by similar simulations of the same system and

[2] Facultad de Matemática, Astronomía y Física. Universidad Nacional de Córdoba. Ing. Medina Allende esq. Haya de la Torre. Ciudad Universitaria. 5000 Córdoba, Argentina. mafcor!bustos@uunet.uu.net
&
Instituto Nacional de Pesquisas Espaciais. Divisão de Processamento de Imagens. Avenida dos Astronautas, 1758. 12227-010 São José dos Campos, SP, Brazil. frery@dpi.inpe.br

with different parameter values. What is to be known is, for example, which are the parameter values that render the most adequate model to represent a real system, or, maybe, how does the system behave under different parameter values.

A simulation study must be carefully planned, in order to obtain meaningful and useful results. It is to be always remembered that this kind of study (nothing but an experience with numbers) and experiences with animals, crops, etc. are alike. From this viewpoint, Monte Carlo experiences have the advantage of being wholly controlable, as is usually not the case in other laboratories of Applied Sciences.

Therefore, a Monte Carlo experience should be planned obeying the rules of Experiments Design. Many good ideas could be borrowed from this area of Statistics (see, for example, the book by Box et al. (1978) for a comprehensive introduction). We should identify the critical hypothesis in the considered model, and we should also obtain every needed output in order to isolate the effects of every factor, besides considering groups of relevant factors.

In principle, a simulation model has two components: one is given by the parameters and interaction structures among the random variables: the *input*; the other is the response or *output*. There are several elements to be considered when planning the experience, for example:

1. Which are the combinations of factors levels to be analyzed in each simulation run? Some techniques useful for the determination of these combinations could be seen in Mauro (1986).
2. How to organize the outputs?
3. How to construct the model under study in order to assess how the factors influence the output?
4. How to control the inputs to make the ouputs more accurate?

The first three questions are of general interest, and applicable to every simulation or data analysis problem. A detailed discussion about these issues, specifically applied to simulation, could be seen in the work by Kleijnen (1975); Gruber and Freimann (1986) also treat this problem, in the context of comparison of estimators.

The two most relevant factors in stochastic simulation (whenever used as a study methodology in Statistics) are: the sample size and the distributions from which the *samples* are taken. Most of the works that use stochastic simulation (in Statistics) aim at comparing different techniques under different data distributions. Among these works we consider those devoted to the determination (or approximation) of the exact distribution of certain statistics.

Some common situations, when the interest is in comparing performances of different statistics, are:

a) The *t* statistic against other options to obtain test procedures and confidence intervals, which are dependable even when the data are not normally distributed.

b) Estimating the center of symmetry of a symmetric distribution using several estimation procedures: sample mean, sample median, trimmed means, robust estimators, etc.

c) Robust estimation vs. least squares in regression.

d) Parametric vs. nonparametric estimation.

The following questions could be answered in order to assess performances:

q1) How does the bias of an estimator varies with respect to variations of the sample size?

q2) How does the variance of an estimator varies with respect to variations of the sample size?

q3) How are the asymptotic distributions (if any) of the estimators under comparison?

q4) Is it possible to know something about the exact distribution of estimators? For instance: estimate the mean, the variance, the quantiles; or, even better, are there analytical results about this?

q5) If the estimators are asymptotically normal, which is the rate of convergence to normality?

A quite general setup for a Monte Carlo experiment, suited for the answering of these questions, could be:

s1) Generate the model under study, say $n$ times.

s2) Calculate (and save) the values of the statistics under comparison for the sample of size $n$ generated in s1).

s3) Repeat (*replication*) s1) and s2) $M$ times.

s4) Study the empirical distributions of the statistics based upon the *samples* obtained in step s3).

In order to have the study complete, the setup above could be modified or repeated for:

m1) Several values of $n$. These additional runs could give insight about the behaviour of *small size samples*, and about the convergence rate towards the asymptotic results.

m2) Different distributions for the input data, including

- several families of distributions,
- several values of the parameters for every distribution,
- possible dependence upon observations.

A first step towards increasing the accuracy of a certain estimator, could be the searching of efficient techniques for the calculation of the used quantities; or increasing the number of replications ($M$) without overweighting the computational cost, using, for instance, faster generation techniques. For example, if the programs were developed in a high level programming language, such as FORTRAN, the experimenter should try to write them down *without too much effort* in a lower level language, such as C or even ASSEMBLER. This could yield faster generations and calculations... though there might appear formidable programming problems, increasing the risk of bugs, mistakes, etc. The only rule about it we know is: be sensible!

The tools devoted to improve the accuracy of simulation results are generically known as *Variance Reduction Techniques*. This name could be justified by considering what is usually done in a simulation experiment: let $F\colon \mathbb{R} \longrightarrow [0,1]$ be a cummulative distribution function; let

$X$ be a random variable with distribution given by $F$ above, and let $g: \mathbb{R}^n \longrightarrow \mathbb{R}$ be a measurable function. Problem: estimate $\theta = \mathbb{E}_F(g(X_1, \ldots, X_n))$, where $X_1, \ldots, X_n$ are independent identically distributed random variables with common distribution $F$. The raw Monte Carlo estimator is:

m1) Choose a big $M$, the number of replications, in a more or less arbitrary fashion (say $M = 500$, 1000 or more).

m2) For every $m = 1, 2, \ldots, M$ generate a sample $x_{m,1}, \ldots, x_{m,n}$ from $X$ (with distribution $F$) and calculate

$$g_n(m) = g_n(x_{m,1}, \ldots, x_{m,n}).$$

m3) Define as estimator of $\theta$ the quantity

$$\overline{g}_n = \frac{1}{M} \sum_{1 \leq m \leq M} g_n(m).$$

m4) Define as estimator of the variance of $\overline{g}_n$ the quantity

$$S^2(\overline{g}_n) = \frac{1}{M(M-1)} \sum_{1 \leq m \leq M} \left(g_n(m) - \overline{g}_n\right)^2.$$

The variance reduction techniques consist of modifying this setup in order to obtain a variance reduction of the estimator of $\theta$. For instance, through modifying the way in which the random variables are generated, or by incorporating analytical knowledge about the distribution $F$.

In most problems, $\theta$ could be a vector, and $g$ and $F$ could have quite complicated forms; in such cases, only the use of some variance reduction technique would ensure dependable results.

# 2. Suggestions: How to write a report about a simulation experiment

Every research report has its own set of ideal formats, depending on the reported experience, the salient features and conclusions, etc. It would be interesting, though, to try to design our reports in a more or less standard and accepted form.

The main idea related to a *simulation report* is that the experience is nothing but a numerical experiment, in the same fashion as an experiment with animals, crops, etc. Some readers might be tempted to validate the results, or to repeat the experience, and the report must supply all the relevant information in order to help them. It is therefore a must to explicitely state the following items:

- Computer used (type, model, etc.)
- Operating system (UNIX, DOS, VM, etc. including version).

- Computational packages (SAS, S-PLUS, SOC, STATGRAPHICS, etc.)
- The programming language used in the development of the experimenter's programs (FORTRAN, C, PASCAL, etc.) and compiler (MICROSOFT, LAHEY, TURBO-PASCAL, etc.)
- Numerical subroutines involved in the calculations, including their origin or informing about the implemented algorithm(s). If their numerical precission and other details are known, state these facts.

Specifically related to the simulation, report the following:

- Pseudorandom uniform random number generator: type, tests results, either if it is a built-in subroutine of the programming language or not, seeds used, etc.
- Nonuniform pseudorandom number generators: the same as above; also state the algorithms or subroutines used.
- Results about the statistical quality of the pseudorandom sequences that will be involved in the study if any (there always should be).

In a forthcoming Section we recall a format for the report itself, suggested by Lewis and Orav (1989). This is neither the best nor the only possible form; again, every problem is a problem in its own right, but it might be useful to consider it as a reference.

**Notation:** For every $x \in \mathbb{R}^+$ we denote its integer part as $\lfloor x \rfloor$, i. e., $\lfloor x \rfloor = \max\{k \in \mathbb{N}: k \leq x\}$. Let $\mathbf{a} = (a_1, \ldots, a_N)$ be a real-valued vector. The evaluation of some of the following quantities will be called the *analysis of the empirical distribution* of $\mathbf{a}$.

- ⋆ Sample mean of $\mathbf{a}$ = Mean $(\mathbf{a}) = \bar{\mathbf{a}} = \frac{1}{N}\sum_{i=1}^{N} a_i$.
- ⋆ Sample variance of $\mathbf{a} = \widehat{\text{Var}}(\mathbf{a}) = S^2(\mathbf{a}) = \frac{1}{N-1}\sum_{i=1}^{N}(a_i - \bar{\mathbf{a}})^2$.
- ⋆ Standard deviation of $\mathbf{a} = S(\mathbf{a}) = \sqrt{S^2(\mathbf{a})}$.
- ⋆ Sample skewness of $\mathbf{a} = \widehat{\gamma_1}(\mathbf{a}) = \frac{\widehat{\mu_3}(\mathbf{a})}{S^3(\mathbf{a})}$, where

$$\widehat{\mu_3}(\mathbf{a}) = \frac{N}{(N-1)(N-2)} \sum_{1 \leq i \leq N} (a_i - \bar{\mathbf{a}})^3.$$

- ⋆ Sample kurtosis of $\mathbf{a} = \widehat{\gamma_2}(\mathbf{a}) = \frac{\widehat{\mu_4}(\mathbf{a})}{S^4(\mathbf{a})}$, where

$$\widehat{\mu_4}(\mathbf{a}) = \frac{N^2 - 2N + 3}{(N-1)(N-2)(N-3)} \sum_{1 \leq i \leq N} (a_i - \bar{\mathbf{a}})^4 - \frac{3(S^2(\mathbf{a}))^2(N-1)(2N-3)}{N(N-2)(N-3)}.$$

- ⋆ Sample mean variance of $\mathbf{a} = S^2(\bar{\mathbf{a}}) = \frac{1}{N} S^2(\mathbf{a})$.

Let now $\mathbf{a}_{(\bullet)}$ be the vector $\mathbf{a}$ sorted in ascending order. We write $\mathbf{a}_{(\bullet)} = (a_{N:1}, \ldots, a_{N:N})$ (i. e. $a_{N:1} \leq a_{N:2} \cdots \leq a_{N:N}$).

- ⋆ Sample median of $\mathbf{a}$ = Median ($\mathbf{a}$) =

$$Q_2(\mathbf{a}) = \begin{cases} a_{N:\lfloor N/2\rfloor+1} & \text{if } N \text{ is odd} \\ \frac{1}{2}(a_{N:N/2} + a_{N:N/2+1}) & \text{if } N \text{ is even.} \end{cases}$$

- ⋆ Sample lower quartile of $\mathbf{a}$ =

$$Q_1(\mathbf{a}) = \begin{cases} a_{N:(\ell+1)/2} & \text{if } \ell \text{ is odd} \\ \frac{1}{2}(a_{N:\ell/2} + a_{N:\ell/2+1}) & \text{if } \ell \text{ is even.} \end{cases} \qquad (1)$$

- ⋆ Sample upper quartile of $\mathbf{a}$ =

$$Q_3(\mathbf{a}) = \begin{cases} a_{N:(N+1-(\ell+1)/2)} & \text{if } \ell \text{ is odd} \\ \frac{1}{2}(a_{N:N+1-\ell/2} + a_{N:N-\ell/2+1}) & \text{if } \ell \text{ is even.} \end{cases} \qquad (2)$$

In formulas (1) and (2) we wrote:

$$\ell = \begin{cases} \frac{N-1}{2} & \text{if } N \text{ is odd} \\ \frac{N}{2} & \text{if } N \text{ is even.} \end{cases}$$

Let $\alpha$ be a constant such that $0 < \alpha < 1$, we then define the

- ⋆ $\alpha$-sample quantile of $\mathbf{a} = \hat{\mathbf{a}}_\alpha$ =

$$\begin{cases} a_{N:\alpha N} & \text{if } \alpha N \text{ is integer} \\ a_{N:\lfloor \alpha N\rfloor+1} & \text{if } \alpha N \text{ is not.} \end{cases}$$

- ⋆ Sample coefficient of variation of $\mathbf{a} = \tilde{C}(\mathbf{a}) = \frac{s(\mathbf{a})}{\bar{\mathbf{a}}}$.

A fairly complete summary of the relevant properties of the aforementioned quantities, when $\mathbf{a}$ is a random vector such that $(a_1, \ldots, a_N)$ are the outcomes of independent identically distributed random variables, can be seen in Lewis and Orav (1989).

In the next Section we shall show how the previous suggestions could be applied to a problem of Image Processing: a Monte Carlo study that aims at comparing certain statistical procedures for estimating the parameter of a Raighley distribution. We will consider this distribution under a *pure* and a *contaminated* model. This kind of estimation problem appears in the statistical treatment of Synthetic Aperture Radar (SAR) images.

# 3. An example

SAR imagery is becoming more and more used. Among its advantages over visible range imagery, one could mention those related to the facts that there is no need to operate the satellite with daylight, that images could be obtained even in the presence of clouds, fog, etc., and that the returned signal carries information about the dielectrical properties of the soil. For a comprehensive summary of the statistical properties of SAR images, the reader is referred to Kelly et al. (1988), Derin et al. (1990), and the references therein.

A simple, though theoretically tractable and physically sensible, model for the marginal distribution of the observed values in every pixel, is the Raighley distribution. This assumption departs from the more classical hypothesis of normality, commonly used for observations in the visible range (e. g. SPOT and LANDSAT images).

Given a SAR image, it is often neccessary to estimate the parameters of the Raighley distributions whose outcomes we are seeing, for instance, to be used as input information for filters, segmentation algorithms, etc. (see Duda and Hart (1973) for details about these and other related procedures). These estimations are usually carried out by selecting a window (a subset of the image) of *homogeneous* observations, and then estimating the parameter using the data there contained. A common problem is that the chosen window might not be completely homogeneous, and some observations within it could have come from other population(s): we then say that there is a *contamination*. It is therefore interesting to evaluate the performance of several estimation procedures in order to compare them, either with and without the presence of contamination.

In Frery and Sant'Anna (1993) an implementation of the estimators that will illustrate this paper is shown. They are used as SAR image filters, and are applied in the reduction of the *speckle* noise present in a real image. The results presented there show that the overall noise is efectivelly reduced, and that details and sharp features are better preserved with the use of the robust estimators (filters) than with the use of the maximum likelihood and moments estimators (filters).

## 3.1. Description of the study

The indicator function of the set $A$ is defined as:

$$\mathbb{I}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{else.} \end{cases}$$

For $\xi$, a real positive number, we call the cummulative distribution of a Raighley random variable with parameter $\xi$ the function $F_\xi : \mathbb{R} \longrightarrow [0,1]$ defined by

$$F_\xi(y) = \left[1 - \exp\left\{-\frac{1}{2\xi^2}y^2\right\}\right] \mathbb{I}_{(0,+\infty)}(y).$$

It is easy to see that $F_\xi$ has a density given by

$$f_\xi(y) = \frac{y}{\xi^2}\exp\left\{-\frac{1}{2\xi^2}y^2\right\} \mathbb{I}_{(0,+\infty)}(y).$$

If $Y$ is a random variable with distribution $F_\xi$, then

$$\mathbb{E}(\mathrm{Y}) = \sqrt{\frac{\pi}{2}}\xi \quad \text{and} \quad \mathrm{Var}(\mathrm{Y}) = \left(2 - \frac{\pi}{2}\right)\xi^2.$$

Then

$$C(Y) = \text{coefficient of variation of } Y = \frac{\sigma(Y)}{\mathbb{E}(\mathrm{Y})} = 2\sqrt{\frac{1}{\pi} - \frac{1}{4}}.$$

Let $n$ be a positive integer and $\xi > 0$.

**Definition 3.1** *The random variables $Y_1, \ldots, Y_n$ satisfy a pure Raighley model with parameter $\xi$ if $Y_1, \ldots, Y_n$ are independent identically distributed random variables each with density $f_\xi$. We denote $\mathcal{R}^n(\xi)$ the cummulative distribution function of $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$.*

**Definition 3.2** *The random variables $Y_1, \ldots, Y_n$ satisfy a contaminated Raighley model with parameter $\xi$ and:*

- *lower contamination proportion $\alpha$ (with $\alpha > 0$),*
- *scale of lower contamination $\xi_0$ (with $0 < \xi_0 < \xi$);*
- *upper contamination proportion $\beta$ (with $\beta > 0$ and $\alpha + \beta < \frac{1}{2}$), and*
- *scale of upper contamination $\xi_2$ (with $\xi < \xi_2 < +\infty$)*

*if*

- *$Y_1, \ldots, Y_n$ are independent,*
- *$Y_1, \ldots, Y_{\lfloor \alpha n \rfloor}$ are identically distributed, each with density $f_{\xi_0}$,*
- *$Y_{\lfloor \alpha n \rfloor + 1}, \ldots, Y_{n - \lfloor \beta n \rfloor}$ are identically distributed, each with density $f_\xi$.*
- *$Y_{n - \lfloor \beta n \rfloor + 1}, \ldots, Y_n$ are identically distributed, each with density $f_{\xi_2}$.*

*We denote $\mathcal{RC}^n(\xi_0, \xi, \xi_2; \alpha, \beta)$ the cummulative distribution of $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ in this case.*

We could use stochastic simulation to study the behaviour of different estimation procedures (to be defined) of the parameter $\xi$, under several situations of the pure model (varying $\xi$ and $n$) and of the contaminated model (varying $\xi_0$, $\xi$, $\xi_2$, $\alpha$, $\beta$ and $n$). Thus, gaining some knowledge about the *robustness* of these procedures.

As we said before, a study like this has practical relevance since, as is usually the case in applications, we might not be sure about the pureness of the distribution of the observations $y_1, \ldots, y_n$ (in fact the contaminated model is the most frequent case). But the reader must keep in mind that this a mere example within this work, just aiming at showing how the suggestions presented in the previous Sections could be applied. Therefore, we shall restrict this study to a quite preliminar stage, ending this paper with some guidelines for its continuation.

The pure model will be studied for the following situations: $\xi = .5$ and $n(i) = 100i$ for $i = 1, 2, \ldots, 10$. The contaminated model will be studied for the following cases: $n(i) = 100i$ for $i = 1, 2, \ldots, 10$, $\xi_0 = .1$, $\xi = .5$, $\xi_2 = 1$ and $\alpha = \beta = .05$. Without loss of generality, we can suppose that every outcome $y_1, \ldots, y_n$ of the random vector $Y_1, \ldots, Y_n$, with $n$ any of the $n(i)$ above, will satisfy the following two conditions:

$$y_i > 0 \text{ for every } i = 1, \ldots, n \quad \text{and } y_i \neq y_j \text{ if } i \neq j.$$

We will consider the following estimation procedures for $\xi$ based in $y_1, \ldots, y_n$:

a. The ML estimator of $\xi$ defined by

$$\xi_{\mathrm{ML}}(y_1, \ldots, y_n) = \sqrt{\frac{1}{2n} \sum_{i=1}^{n} y_i^2}.$$

This is the maximum likelyhood estimator of $\xi$, under the pure model.

b. The MO estimator of $\xi$ defined by

$$\xi_{\mathrm{MO}}(y_1, \ldots, y_n) = \sqrt{\frac{2}{\pi}}\frac{1}{n}\sum_{i=1}^{n} y_i.$$

This is the estimator of $\xi$ based on the first sample moment, under the pure model.

c. The TML estimator of $\xi$ defined by

$$\xi_{\mathrm{TML}}(y_1, \ldots, y_n) = \sqrt{\frac{1}{2(n-2a)}\sum_{i=a+1}^{n-a} y_{n:i}^2}.$$

This is the *trimmed ML* estimator of $\xi$ with a proportion of deleted observations equal to $2\alpha_0$.

d. The T estimator of $\xi$ defined by

$$\xi_{\mathrm{T}}(y_1, \ldots, y_n) = \sqrt{\frac{2}{\pi}}\frac{1}{n-2a}\sum_{i=a+1}^{n-a} y_{n:i}.$$

This is the *trimmed mean* estimator of $\xi$ with a proportion of deleted observations equal to $2\alpha_0$. In both trimmed estimators we wrote $a = \lfloor n\alpha_0 \rfloor$ and $(y_{n:1}, \ldots, y_{n:n})$ denotes the vector $(y_1, \ldots, y_n)$ sorted in ascending order. The trimmed observations are the $\lfloor n\alpha_0 \rfloor$ smallest and the $\lfloor n\alpha_0 \rfloor$ biggest ones.

e. The MAD estimator of $\xi$ (Bustos, 1981) defined by

$$\xi_{\mathrm{MAD}}(y_1, \ldots, y_n) = \frac{1}{K}\mathrm{Med}\Big(|y_1 - \mathrm{Med}(y_1, \ldots, y_n)|, \ldots, |y_n - \mathrm{Med}(y_1, \ldots, y_n)|\Big),$$

where $\mathrm{Med}(y_1, \ldots, y_n) = \mathrm{Median}(y_1, \ldots, y_n)$, and $K = .4485$. This is the Median Absolute Deviation estimator of $\xi$, with the correction constant, $K$, determined using numerical tools.

## 3.2. Algorithm for the experience

Set the initial values $N = 30000$, $I = 10$. We make the conservative choice of $\alpha_0 = .05$. For every $i = 1, 2, \ldots, 10$ define $n(i) = 100i$ and $M(i) = \lfloor N/n(i) \rfloor$.

E1) Assign $d = 1$.

E2) Assign $i = 1$.

E3) Assign $m = 1$.

E4) Assign $a(m) = (m-1)n(i)$.

E5) Generate $z_{a(m)+1}, \ldots, z_{a(m)+n(i)}$, sampling from the random vector $\mathbf{Y} = (Y_1, \ldots, Y_{n(i)})^{\mathrm{T}}$ with cummulative distribution function $\mathcal{R}^{n(i)}(\xi)$ if $d = 1$, and with $\mathcal{RC}^{n(i)}(\xi_0, \xi, \xi_2; \alpha, \beta)$ if $d = 2$.

E6) Assign $\mathbf{z} = (z_{a(m)+1}, \ldots, z_{a(m)+n(i)})$ and calculate

- $\xi_{\mathrm{ML},n(i)}(m) := \xi_{\mathrm{ML}}(\mathbf{z})$,
- $\xi_{\mathrm{MO},n(i)}(m) := \xi_{\mathrm{MO}}(\mathbf{z})$,
- $\xi_{\mathrm{TML},n(i)}(m) := \xi_{\mathrm{TML}}(\mathbf{z})$,
- $\xi_{\mathrm{T},n(i)}(m) := \xi_{\mathrm{T}}(\mathbf{z})$ and
- $\xi_{\mathrm{MAD},n(i)}(m) := \xi_{\mathrm{MAD}}(\mathbf{z})$.

E7) If $m = M(i)$ continue in step E8. Else, assign $m = m + 1$ and return to step E4.

E8) If $i = I$ continue in step E9. Else, assign $i = i + 1$ and continue in step E3.

E9) If $d = 2$ continue in step E10. Else, assign $d = d + 1$ and continue in step E2.

E10) For every $d = 1, 2$ and every $i = 1, \ldots, I$ define the following vectors:

- $\xi_{\mathrm{ML},n(i)} = (\xi_{\mathrm{ML},n(i)}(1), \ldots, \xi_{\mathrm{ML},n(i)}(M(i)))^{\mathrm{T}}$,
- $\xi_{\mathrm{MO},n(i)} = (\xi_{\mathrm{MO},n(i)}(1), \ldots, \xi_{\mathrm{MO},n(i)}(M(i)))^{\mathrm{T}}$,
- $\xi_{\mathrm{TML},n(i)} = (\xi_{\mathrm{TML},n(i)}(1), \ldots, \xi_{\mathrm{TML},n(i)}(M(i)))^{\mathrm{T}}$,
- $\xi_{\mathrm{T},n(i)} = (\xi_{\mathrm{T},n(i)}(1), \ldots, \xi_{\mathrm{T},n(i)}(M(i)))^{\mathrm{T}}$ and
- $\xi_{\mathrm{MAD},n(i)} = (\xi_{\mathrm{MAD},n(i)}(1), \ldots, \xi_{\mathrm{MAD},n(i)}(M(i)))^{\mathrm{T}}$.

Analyze the empirical distributions of every vector defined above.

E11) Show the results in a table, for example following the presentation suggested in Lewis and Orav (1989), clearly stating:

- Distribution generated.
- Total size of the sample observed from the distribution, for every $i = 1, 2, \ldots, I$, say $N$.
- Sample size upon which the pocedures are based: $n_1, \ldots, n_I$.
- Number of observed procedures for every value of $i = 1, 2, \ldots, I$, say $M(1), \ldots, M(I)$. Note that

$$N \approx n(1)M(1) \approx n(2)M(2) \approx \ldots \approx n(I)M(I). \tag{3}$$

Formula (3) is required because under certain hypothesis (see, for example, Lehmann (1983)), the following asymptotic expansions hold:

$$\mathbb{E}(\Theta_n) = \theta + \frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_3}{n^3} + \cdots \tag{4}$$

$$\mathrm{Var}(\Theta_n) = \frac{b_1}{n} + \frac{b_2}{n^2} + \frac{b_3}{n^3} + \cdots. \tag{5}$$

If that is the case, one could carry on analyzing the regression model defined by:

$$Z_i = \theta + a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3} + \varepsilon_i, \qquad i = 1, \ldots, I, \tag{6}$$

where $Z_i = \overline{\Theta}_{n(i)}$ = estimator of $\mathbb{E}(\Theta)_{n(i)}$, "sample mean" over samples of $\Theta_{n(i)}$ of size $M(i)$, $x_{i1} = \frac{1}{n(i)}$, $x_{i2} = \frac{1}{n(i)^2}$, $x_3 = \frac{1}{n(i)^3}$, $i = 1, \ldots, I$. Here, $\Theta_{n(i)}$ represents any of the $\xi_{\mathrm{ML},n(i)}, \ldots, \xi_{\mathrm{MAD},n(i)}$ (100 in total).

Now, since $\mathrm{Var}(Z_i) = \frac{1}{M(i)}\mathrm{Var}(\Theta_{n(i)})$ we have from equation (5) that, for $n(i)$ large:

$$\mathrm{Var}(Z_i) \approx \frac{b_1}{n(i)M(i)} \approx \frac{b_1}{N} \text{ for every } i = 1, \ldots, I,$$

using equation (4). Thus, $\mathrm{Var}(Z_1) \approx \mathrm{Var}(Z_2) \approx \cdots \approx \mathrm{Var}(Z_I)$. Moreover, taking into account the simulation scheme, we might consider that $Z_1, \ldots, Z_I$ are independent (notice that, for each $i = 1, \ldots, I$ the generation proceeds, instead of going back to the beginning).

We can then apply a regression analysis to the model (6) considering the values $\overline{\theta}_{n(1)}, \ldots, \overline{\theta}_{n(I)}$ as the observed values of $Z_1, \ldots, Z_I$. Then, for example using least squares estimators, we could estimate the coefficients in (6) with

$$\widehat{\theta}_{\mathrm{reg}}, \quad \widehat{a}_{1,\mathrm{reg}}, \quad \widehat{a}_{2,\mathrm{reg}}, \quad \widehat{a}_{3,\mathrm{reg}}. \tag{7}$$

In order to assess the accuracy of the estimators in (7), the simulation could be carried on obtaining, say $R$, values like

$$\begin{array}{l} \widehat{\theta}_{\mathrm{reg}}(1), \ldots, \widehat{\theta}_{\mathrm{reg}}(R) \\ \widehat{a}_{1,\mathrm{reg}}(1), \ldots, \widehat{a}_{1,\mathrm{reg}}(R) \\ \vdots \end{array}$$

Then, one could analyze the empirical distributions of

$$\begin{array}{l} (\widehat{\theta}_{\mathrm{reg}}(1), \ldots, \widehat{\theta}_{\mathrm{reg}}(R)) \\ (\widehat{a}_{1,\mathrm{reg}}(1), \ldots, \widehat{a}_{1,\mathrm{reg}}(R)), \text{ etc.,} \end{array}$$

and, from this, we would obtain estimates of the accuracy of the least squares estimators in model (6). Remember that, in most of the situations, our main goal will be knowledge about $\theta$; but $a_1$ is also interesting since it says something about the asymptotic bias of the procedure.

Observe that, after performing those $R$ replications we will have at hand $RM(i)$ outcomes of the random variable $\Theta_{n(i)}$: $(\theta_{n(i)}(1), \ldots, \theta_{n(i)}(M(i)), \ldots, \theta_{n(i)}(RM(i)))$. This sample would allow us to study its empirical distribution. We omit this last part in this work.

## 3.3. The results

Tables 1, 2 and 3 summarize the results of the aforementioned experiencies. The values in Table 1 are rounded to the fourth decimal place, since they are used in Table 4; the values in Tables 2 and 3 are rounded to the third decimal place.

As previouslly stated in the definition of the algorithm, the values of $\xi_0$, $\xi$, $\xi_2$, $\alpha$ and $\beta$ are held constant during the whole experience; we can therefore write $\mathcal{R}^{n(i)}$ and $\mathcal{RC}^{n(i)}$ as short forms for $\mathcal{R}^{n(i)}(\xi)$ and $\mathcal{RC}^{n(i)}(\xi_0, \xi, \xi_2; \alpha, \beta)$, respectively.

Table 1: Sample Means, Mean Standard Deviations and Standard Deviations of estimators over $M(i) = \lfloor 30000/n(i) \rfloor$ vectors of size $n(i) = 100i$.

| — | $\xi_{\mathrm{ML}}$ | | $\xi_{\mathrm{MO}}$ | | $\xi_{\mathrm{TML}}$ | | $\xi_{\mathrm{T}}$ | | $\xi_{\mathrm{MAD}}$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ |
| 1 | .4995 | .5226 | .5021 | .5056 | .4742 | .4777 | .4921 | .4870 | .4983 | .5424 |
| | .0026 | .0026 | .0028 | .0026 | .0025 | .0024 | .0028 | .0026 | .0032 | .0037 |
| | .0454 | .0458 | .0482 | .0451 | .0436 | .0420 | .0480 | .0448 | .0553 | .0646 |
| 2 | .4990 | .5286 | .5021 | .5093 | .4732 | .4803 | .4921 | .4894 | .4941 | .5476 |
| | .0028 | .0029 | .0029 | .0028 | .0026 | .0031 | .0028 | .0025 | .0035 | .0033 |
| | .0337 | .0356 | .0354 | .0347 | .0318 | .0308 | .0349 | .0341 | .0428 | .0408 |
| 3 | .5000 | .5204 | .5005 | .5022 | .4723 | .4734 | .4895 | .4828 | .4981 | .5377 |
| | .0029 | .0031 | .0029 | .0029 | .0026 | .0027 | .0028 | .0029 | .0031 | .0034 |
| | .0290 | .0307 | .0292 | .0293 | .0262 | .0269 | .0282 | .0289 | .0314 | .0432 |
| 4 | .5005 | .5186 | .5018 | .5004 | .4724 | .4724 | .4906 | .4811 | .4993 | .5410 |
| | .0025 | .0030 | .0027 | .0029 | .0024 | .0027 | .0027 | .0028 | .0032 | .0038 |
| | .0218 | .0263 | .0238 | .0248 | .0211 | .0231 | .0237 | .0244 | .0273 | .0328 |
| 5 | .4998 | .5195 | .5000 | .5001 | .4711 | .4724 | .4886 | .4806 | .4980 | .5510 |
| | .0026 | .0029 | .0027 | .0030 | .0024 | .0027 | .0027 | .0030 | .0034 | .0034 |
| | .0205 | .0227 | .0213 | .0235 | .0185 | .0211 | .0206 | .0236 | .0266 | .0267 |
| 6 | .4974 | .5254 | .4979 | .5054 | .4690 | .4770 | .4867 | .4856 | .4944 | .5492 |
| | .0028 | .0025 | .0027 | .0024 | .0025 | .0022 | .0026 | .0027 | .0038 | .0041 |
| | .0199 | .0193 | .0194 | .0185 | .0173 | .0171 | .0186 | .0191 | .0265 | .0287 |
| 7 | .4994 | .5169 | .5007 | .4971 | .4716 | .4691 | .4899 | .4772 | .4958 | .5423 |
| | .0025 | .0032 | .0027 | .0029 | .0023 | .0026 | .0026 | .0029 | .0030 | .0035 |
| | .0165 | .0207 | .0172 | .0190 | .0152 | .0170 | .0167 | .0186 | .0197 | .0229 |
| 8 | .4999 | .5247 | .5005 | .5053 | .4712 | .4761 | .4890 | .4852 | .4979 | .5436 |
| | .0029 | .0035 | .0029 | .0033 | .0026 | .0029 | .0028 | .0031 | .0032 | .0040 |
| | .0174 | .0215 | .0175 | .0201 | .0155 | .0174 | .0168 | .0188 | .0192 | .0197 |
| 9 | .5037 | .5273 | .5035 | .5057 | .4746 | .4759 | .4920 | .4845 | .5039 | .5469 |
| | .0029 | .0025 | .0030 | .0024 | .0026 | .0025 | .0028 | .0025 | .0041 | .0047 |
| | .0168 | .0142 | .0171 | .0137 | .0149 | .0143 | .0163 | .0145 | .0233 | .0271 |
| 10 | .4989 | .5255 | .4992 | .5075 | .4705 | .4790 | .4882 | .4881 | .4981 | .5459 |
| | .0021 | .0027 | .0022 | .0026 | .0019 | .0023 | .0022 | .0026 | .0025 | .0031 |
| | .0113 | .0150 | .0121 | .0144 | .0104 | .0127 | .0119 | .0145 | .0136 | .0168 |

Table 2: Sample Minima, Medians and Maxima of estimators over $M(i) = \lfloor 30000/n(i) \rfloor$ vectors of size $n(i) = 100i$.

| — | $\xi_{\mathrm{ML}}$ | | $\xi_{\mathrm{MO}}$ | | $\xi_{\mathrm{TML}}$ | | $\xi_{\mathrm{T}}$ | | $\xi_{\mathrm{MAD}}$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ |
| 1 | .392 | .389 | .383 | .355 | .373 | .331 | .373 | .332 | .375 | .397 |
| | .499 | .525 | .502 | .508 | .475 | .481 | .492 | .488 | .498 | .539 |
| | .636 | .651 | .647 | .622 | .607 | .595 | .642 | .593 | .687 | .784 |
| 2 | .431 | .445 | .435 | .430 | .413 | .409 | .427 | .411 | .408 | .455 |
| | .495 | .529 | .497 | .509 | .469 | .479 | .488 | .487 | .490 | .549 |
| | .640 | .629 | .654 | .607 | .608 | .569 | .639 | .590 | .690 | .637 |
| 3 | .454 | .453 | .455 | .439 | .429 | .420 | .443 | .421 | .433 | .447 |
| | .495 | .520 | .495 | .501 | .469 | .472 | .486 | .481 | .496 | .539 |
| | .588 | .633 | .595 | .607 | .561 | .575 | .585 | .585 | .583 | .660 |
| 4 | .453 | .461 | .452 | .454 | .432 | .429 | .446 | .436 | .442 | .472 |
| | .498 | .515 | .495 | .496 | .468 | .471 | .484 | .476 | .498 | .536 |
| | .549 | .608 | .552 | .585 | .516 | .547 | .541 | .563 | .576 | .639 |
| 5 | .464 | .473 | .462 | .453 | .441 | .424 | .451 | .430 | .450 | .486 |
| | .498 | .522 | .500 | .500 | .470 | .471 | .489 | .479 | .500 | .548 |
| | .567 | .584 | .565 | .574 | .528 | .545 | .548 | .558 | .587 | .628 |
| 6 | .460 | .487 | .461 | .464 | .437 | .443 | .452 | .445 | .441 | .499 |
| | .497 | .522 | .499 | .504 | .470 | .475 | .489 | .486 | .491 | .542 |
| | .544 | .567 | .538 | .555 | .510 | .528 | .524 | .540 | .559 | .601 |
| 7 | .465 | .482 | .466 | .465 | .441 | .442 | .457 | .445 | .456 | .495 |
| | .500 | .516 | .499 | .497 | .469 | .465 | .486 | .474 | .495 | .542 |
| | .541 | .570 | .539 | .550 | .511 | .518 | .528 | .533 | .565 | .594 |
| 8 | .466 | .488 | .463 | .468 | .440 | .444 | .452 | .450 | .447 | .497 |
| | .501 | .523 | .504 | .505 | .473 | .473 | .490 | .483 | .497 | .545 |
| | .536 | .569 | .533 | .538 | .501 | .512 | .518 | .517 | .545 | .589 |
| 9 | .473 | .504 | .470 | .486 | .448 | .455 | .461 | .466 | .478 | .489 |
| | .500 | .528 | .500 | .503 | .472 | .473 | .489 | .483 | .500 | .544 |
| | .537 | .552 | .542 | .530 | .506 | .499 | .530 | .511 | .567 | .592 |
| 10 | .478 | .504 | .479 | .479 | .453 | .454 | .470 | .456 | .472 | .518 |
| | .498 | .521 | .497 | .507 | .470 | .479 | .486 | .489 | .499 | .544 |
| | .528 | .561 | .537 | .535 | .501 | .505 | .526 | .520 | .542 | .591 |

Table 3: Sample Lower and Upper Quartiles of estimators over $M(i) = \lfloor 30000/n(i) \rfloor$ vectors of size $n(i) = 100i$.

| — | $\xi_{\text{ML}}$ | | $\xi_{\text{MO}}$ | | $\xi_{\text{TML}}$ | | $\xi_{\text{T}}$ | | $\xi_{\text{MAD}}$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ |
| 1 | .469 | .486 | .470 | .472 | .445 | .448 | .460 | .458 | .460 | .498 |
| | .527 | .553 | .531 | .536 | .499 | .505 | .521 | .519 | .532 | .579 |
| 2 | .477 | .505 | .479 | .489 | .452 | .461 | .470 | .468 | .464 | .521 |
| | .519 | .551 | .525 | .531 | .493 | .499 | .513 | .511 | .519 | .577 |
| 3 | .477 | .498 | .480 | .483 | .454 | .456 | .471 | .467 | .476 | .513 |
| | .519 | .539 | .519 | .521 | .490 | .490 | .511 | .500 | .517 | .560 |
| 4 | .483 | .502 | .484 | .486 | .456 | .459 | .474 | .466 | .482 | .521 |
| | .518 | .534 | .521 | .514 | .489 | .485 | .510 | .494 | .516 | .563 |
| 5 | .472 | .501 | .482 | .486 | .458 | .461 | .472 | .466 | .481 | .529 |
| | .501 | .535 | .514 | .515 | .482 | .484 | .501 | .494 | .513 | .561 |
| 6 | .483 | .510 | .484 | .492 | .455 | .465 | .470 | .473 | .478 | .528 |
| | .510 | .535 | .510 | .516 | .481 | .489 | .499 | .497 | .509 | .575 |
| 7 | .486 | .504 | .487 | .485 | .460 | .456 | .477 | .463 | .485 | .525 |
| | .511 | .527 | .513 | .506 | .482 | .478 | .500 | .486 | .506 | .558 |
| 8 | .486 | .508 | .487 | .490 | .460 | .465 | .476 | .471 | .485 | .532 |
| | .512 | .540 | .515 | .524 | .484 | .494 | .504 | .501 | .531 | .554 |
| 9 | .493 | .517 | .494 | .494 | .463 | .464 | .482 | .472 | .486 | .530 |
| | .520 | .540 | .514 | .521 | .488 | .489 | .502 | .500 | .516 | .568 |
| 10 | .492 | .515 | .492 | .499 | .462 | .471 | .481 | .477 | .490 | .536 |
| | .504 | .538 | .504 | .520 | .475 | .489 | .578 | .493 | .502 | .553 |

## 3.4. Conclusions of the simulation study

For the sake of comparison, we use the $\widehat{\mu} \pm 2\widehat{\sigma}$ confidence interval, using the values obtained by simulation and presented in Table 1. Table 4 summarizes that information in the following way: every entry has either a "Y" or a "N", indicating weather the corresponding confidence interval for the estimator/distribution situation *hits* the true value (.5) or not. Also, the number "$(j)$" indicates that the corresponding estimator, besides hitting the confidence interval around the true value, has the $j$ smallest confidence interval for that distribution and that sizes.

The quality of this Monte Carlo study, when measured by the Mean Standard Deviations (Table 1), is reasonably good. Tables 2 and 3 suggest that this study could be continued in order to determine the asymptotic behaviour of the estimators; this could be the goal of a forthcoming study.

Within the considered size ranges and situations (the experimenter does not know, a priori, if the data come from the pure or contaminated distribution, henceforth all the situations must be considereded globally), the *best* estimator is the TML. Also, a very poor performance of the MAD estimator has been detected suggesting, thus, not to use it in these situations.

Table 4: Does the estimator hit the true value?

| — | $\xi_{\mathrm{ML}}$ | | $\xi_{\mathrm{MO}}$ | | $\xi_{\mathrm{TML}}$ | | $\xi_{\mathrm{T}}$ | | $\xi_{\mathrm{MAD}}$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ |
| 1 | Y(2) | Y(4) | Y(4) | Y(3) | Y(1) | Y(1) | Y(3) | Y(2) | Y(5) | Y(5) |
| 2 | Y(2) | Y(4) | Y(4) | Y(3) | Y(1) | Y(1) | Y(3) | Y(2) | Y(5) | Y(5) |
| 3 | Y(3) | Y(4) | Y(4) | Y(3) | Y(1) | Y(1) | Y(3) | Y(2) | Y(5) | Y(5) |
| 4 | Y(2) | Y(4) | Y(4) | Y(3) | Y(1) | Y(1) | Y(3) | Y(2) | Y(5) | Y(5) |
| 5 | Y(2) | Y(2) | Y(4) | Y(3) | Y(1) | Y(1) | Y(3) | Y(4) | Y(5) | Y(5) |
| 6 | Y(4) | Y(4) | Y(3) | Y(2) | Y(1) | Y(1) | Y(2) | Y(3) | Y(5) | Y(5) |
| 7 | Y(2) | Y(4) | Y(4) | Y(3) | Y(1) | Y(1) | Y(3) | Y(2) | Y(5) | Y(5) |
| 8 | Y(3) | Y(4) | Y(4) | Y(3) | Y(1) | Y(1) | Y(2) | Y(2) | Y(5) | N |
| 9 | Y(3) | Y(2) | Y(4) | Y(1) | Y(1) | Y(3) | Y(2) | Y(4) | Y(5) | Y(5) |
| 10 | Y(1) | Y(4) | Y(3) | Y(2) | N | Y(1) | Y(2) | Y(3) | Y(4) | N |

Table 5 presents the regression coefficients for the models already introduced in equation (6). Notice that, in every regression, ten points are considered; this is not enough to make a detailed regression analysis, but we include these results just to illustrate this important part of every simulation study in Statistics. Forward Regression was used, with the $F$-to-include and $F$-to-remove values set to 4. All the regressions discarded the use of explanatory variables; i. e. all the $\widehat{a}_{\ell,\mathrm{reg}}$, for $1 \leq \ell \leq 3$ were set to zero and the corresponding explanatory variable $(n(i)^{-\ell})$ excluded from the model. A residual analysis was performed for every regression (10 models), showing no significant structures and, thus, validating the results.

Table 5: Regressions coefficients.

| — | $\xi_{\text{ML}}$ | | $\xi_{\text{MO}}$ | | $\xi_{\text{TML}}$ | | $\xi_{\text{T}}$ | | $\xi_{\text{MAD}}$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| — | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ | $\mathcal{R}^{n(i)}$ | $\mathcal{RC}^{n(i)}$ |
| $\widehat{\theta}_{\text{reg}}$ | .4998 | .5230 | .5008 | .5039 | .4720 | .4754 | .4899 | .4842 | .4978 | .5448 |

## 3.5. Computational Information

**Hardware-software:**

- Computer used: SUN SPARC station 2.
- Operating system/version/support: UNIX 4.1.1, OpenWindows Version 2.
- Computational package used: STATGRAPHICS Version 2.6.
- Programming language and compiler: C++, SUN C++ Version 2.
- Execution time (the whole experience with screen outputs): 64 seconds.
- Since all the data is "well behaved", there was no need to use specific or sofisticated numerical subroutines in this study.

Simulation details:

- Pseudorandom uniform random number generator: the generalized feedback shift register generator reported in Bustos (1990), with slight modifications in order to improve generation speed. Under test.
- Seeds used: 1234567890 1098765432 1029384756.
- Nonuniform pseudorandom number generators: the Raighley variables were generated using the inversion method (Bustos and Frery, 1992).

# References

[1] Box, G.E.; Hunter, W.G.; Hunter, J.S. *Statistics for experimenters.* Wiley, New York, 1978.

[2] Bustos, O.H. *Estimação robusta no modelo de posição.* Colóqio Brasileiro de Matemática, 13., Poços de Caldas, MG, Brazil, 1981.

[3] Bustos, O.H. A TURBO C implementation of a GFSR[0, 1] pseudorandom number algorithm. IMPA, Rio de Janeiro, Brazil. 1990. (Informes de Matemática B-62)

[4] Bustos, O.H.; Frery, A.C. *Simulação estocástica: teoria e algoritmos (versão completa).* IMPA, Rio de Janeiro, Brazil. (Monografias de Matemática, 49)

[5] Derin, H.; Kelly, P.; Vézina, G.; Labbit, S. Modelling and segmentation of speckled images using complex data. *IEEE Transactions on Geoscience and Remote Sensing*, 28:76–87, 1990.

[6] Duda, R.; Hart, E. *Pattern classification and scene analysis.* John Wiley & Sons, 1973.

[7] Frery, A.C.; Sant'Anna, S.J.S. Non-adaptive robust filters for speckle noise reduction. In: Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens, 6., Recife, PE, Brazil, 19–22 out. 1993. *Anais*. Recife, PE, Brazil, SBC/UFPe, 1993, p. 165–174.

[8] Gruber, J.; Freimann, K. Combined response surface regressions in Monte Carlo studies of small sample properties of estimators: theory and application. Discussion Paper # 108, Department of Economics, University of Hogen, West Germany, 1986.

[9] Kelly, P.; Derin, H.; Hartt, K. Adaptive segmentation of speckled images using a hierarchical random field model. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36:1628–1641, 1988.

[10] Kleijnen, J.P. *Statistical techniques in simulation: part II.* Dekker, New York, 1975.

[11] Lehmann, D.H. *Theory of point estimation.* Wiley, New York, 1983.

[12] Lewis, P.A.; Orav, E.J. *Simulation methodology for statisticians, operation analysts and engineers, vol. 1.* Cole Advanced Books & Software, Wadsworth & Books, Pacific Grove, California, 1989.

[13] Mauro, C.A. Efficient Identification of Important Factors in Large Scale Simulations. In: *Proc. 1986 Winter Simulation Conference*, p. 296–305. Wilson, J.; Hendriksen, J. and Roberts, S., eds. Institute of Electrical and Electronics Engineers. Piscataway, NJ, 1986.

## Resumen

Se sugieren contenidos mínimos que todo informe sobre los resultados estadísticos obtenidos por simulación estocástica debiera tener y se muestra una manera de presentar los resultados. Esto se aplica luego a una experiencia de Monte Carlo diseãda para comparar los procedimientos de estimación de los parámetros de la distribución de Raighley, un problema de común ocurrencia en el procesamiento de imágenes por radar.

*PalabrasClaves:* estimación, métodos de Monte Carlo, robustez, proceso de imágenes, simulación estocástica.