Acta Universitatis Apulensis No. 22/2010
ISSN: 1582-5329 pp. 231-242

MONTE-CARLO DETECTION OF THE OUTLIERS DATA AND CRYPTOGRAPHIC APPLICATIONS

Mircea Andraşiu, Andrei Oprina,
Emil Simion and Gheorghe Simion

Abstract. The measure of the influence of an observation on a statistical estimator is the ϕ-divergence. We present a comparative computational study of the estimators of the ϕ-divergence, Monte-Carlo and bootstrap estimators, used for detection of the outliers data. We also give estimators for the prediction density of the Random Coefficient Autoregressive (RCA) models which have cryptographic implications.

2000 *Mathematics Subject Classification*: 62F12, 62F15.

1. Introduction

Practical problems which appeared in the last years in the mathematical statistics domain have a computational approach. We present new computational methods which solves the problem of outliers data detection for a large class of statistical models.

Outliers data detection has applications in regression models - time series forecast from the financial-economy domain which leads to better estimations and thus to smaller errors (see [4] and [5]), in symmetric cryptography - detection key exchange time for some cipher systems (see [3], [20] and [21]) or statistical test for hardware random generators (see [11]), in the theory of secure transmission of information over public communication channel (see [20]).

Section 2 presents the statistical model which is considered in this paper. This model includes the dynamic regression model and intervention models on the time series (see [15] for details). The intervention models are a combination of pulse interventions, gradual or abrupt permanent interventions. Section 3 presents some perturbed models which generalize the computational problem in the case of non perturbation of the a priori distribution. The perturbations are only in the observations that we make. The hardest point of this problem is to compute the a posteriori

probability. To avoid this inconvenient we can also use bootstrap techniques (more details can be found in [14], [17] and [18]). In conformity with this estimation technique the distribution function, which depends of one or more unknown parameters, is replaced by the empirical distribution function in which the parameter is replaced by with his estimation reported to the empirical distribution. Our problem is now transformed from a parametric problem into a non parametric problem. More exactly in *the real problem* we have the independent observations $x_1, ..., x_n$ with the distribution function $F(x)$. Parameter $\theta = g(F(.))$ is defined by distribution function and is estimated by $\hat{\theta} = \hat{\theta}(x_1, ..., x_n)$. This estimator can be obtained using several methods such as maximum likelihood method. *Bootstrap problem* is like the real problem with the difference that the distribution function F is replaced by the empirical distribution function F_n (or a sequence of estimators of the distribution F) and $\tilde{\theta} = g(F_n(.))$. After that we simulate a number m of random variables (called bootstrap samples) $X_1^*, ..., X_m^*$ with distribution F_n. Estimation of the parameter $\tilde{\theta}$ will be $\hat{\theta}^* = \hat{\theta}(X_1^*, ..., X_m^*)$. The properties of the bootstrap estimator allow us to derive properties of the estimator from the real problem. Therefore in section 4 we perform a comparative study of *Monte Carlo* and *bootstrap* methods (see [7]) in which the estimator of the a posteriori distribution is kernel estimator. Sections 5 is dedicated to estimators prediction density of RCA models in connection with cryptographic applications.

2. Statistical model

Let $\{f(\mathbf{y}|\theta, \mathbf{x}); \theta \in \Theta\}$ be a statistical model for the random variable $\mathbf{y} = (y_1, ..., y_n)$, θ a p-dimensional parameter, and $\mathbf{x}$ a covariance matrix. The considered model can be extended to the general dynamic regression model. In particular, in the regression model, the vector $\mathbf{x}$ can be seen like a vector of independent variables and $\mathbf{y}$ like a vector of dependent variables (see [1], [2] and [13] for details). We try to develop a method for the detection of *outliers data* which can influence the decision on parameter θ. A measure of the influence of one observation y_r on θ is the discrepancy $\pi(\theta|\mathbf{y})$ between $\pi(\theta|\mathbf{y}_{(r)})$ where $\mathbf{y}_{(r)} = (y_1, ...y_{r-1}, y_{r+1}, ..., y_n)$.

Following [12] and [16] we define the *discrepancy*, also known as ϕ*-divergence*, between two a posteriori densities $\pi(\theta|\mathbf{y})$ and $\pi_\delta(\theta|\mathbf{y})$ by the formula

$$D_\phi = D(\pi(\theta|\mathbf{y}), \pi_\delta(\theta|\mathbf{y})) = \int \phi(\frac{\pi_\delta(\theta|\mathbf{y})}{\pi(\theta|\mathbf{y})})\pi(\theta|\mathbf{y})d\theta,$$

where δ is a symbol for the perturbation of the a priori density and ϕ is a convex function with $\phi(1) = 0$.

Example 1. For different forms of ϕ we obtain different divergences.

(*i*) $\phi(x) = -\log x$, Kullback-Leibler divergence.

(*ii*) $\phi(x) = \dfrac{(\sqrt{x}-1)^2}{2}$, Hellinger divergence.

(*iii*) $\phi(x) = (x-1)^2$, χ^2 divergence.

(*iv*) $\phi(x) = \dfrac{1}{2}|x-1|$, variational divergence L_1.

(*v*) $\phi(x) = \dfrac{x^\lambda - 1}{\lambda(\lambda+1)}$, $\lambda \neq 0, -1$, pondered power divergence.

3. Perturbed models

Perturbation factor δ can depend of parameter θ, response variable $\mathbf{y}$ and covariance matrix $\mathbf{x}$. The model take under study allows perturbation both a priori density and likelihood. As in [17] we define:

$$\delta(\theta, \mathbf{y}, \mathbf{x}) = \frac{f_\delta(\mathbf{y}|\theta, \mathbf{x})\pi_\delta(\theta)}{f(\mathbf{y}|\theta, \mathbf{x})\pi(\theta)},$$

where f_δ is the likelihood and π_δ is the a priori density under some perturbations. In [17] was considered the case in which we have perturbation only on the likelihood i.e. $\pi_\delta(\theta) = \pi(\theta)$. In this paper we consider the case in which $\pi_\delta(\theta) \neq \pi(\theta)$. There are several methods to choose the a priori distributions; see for example [18] or [8], [9]. Proceed like in classic cases we consider the perturbations on the response variable $\mathbf{y}$ and covariance $\mathbf{x}$. More exactly we have

$$\begin{aligned}\delta_1(\theta, \mathbf{y}, \mathbf{x}) &= \frac{f_\delta(\mathbf{y}_{(r)}|\theta, \mathbf{x})\pi_\delta(\theta)}{f(\mathbf{y}|\theta, \mathbf{x})\pi(\theta)} = \frac{f_\delta(\mathbf{y}_{(r)}|\theta, \mathbf{x})\pi_\delta(\theta)}{f(y_r|\theta, \mathbf{x}, \mathbf{y}_{(r)})f(\mathbf{y}_{(r)}|\theta, \mathbf{x})\pi(\theta)} \\ &= \frac{\pi_\delta(\theta)}{f(y_r|\theta, \mathbf{x}_r, \mathbf{y}_{(r)})\pi(\theta)},\end{aligned}$$

and

$$\begin{aligned}\delta_2(\theta, \mathbf{y}, \mathbf{x}) &= \frac{f_\delta(\mathbf{y}|\theta, \mathbf{x}_{[r(s)]})\pi_\delta(\theta)}{f(\mathbf{y}|\theta, \mathbf{x})\pi(\theta)} = \frac{f_\delta(\mathbf{y}_{(r)}|\theta, \mathbf{x}_{[r(s)]}, y_r)f(y_r|\theta, \mathbf{x}_{[r(s)]})\pi_\delta(\theta)}{f(y_{(r)}|\theta, \mathbf{x}, y_r)f(\mathbf{y}_r|\theta, \mathbf{x})\pi(\theta)} \\ &= \frac{f_\delta(\mathbf{y}_{(r)}|\theta, \mathbf{x}_{[r]}, y_r)f(y_r|\theta, \mathbf{x}_{r(s)})\pi_\delta(\theta)}{f(y_{(r)}|\theta, \mathbf{x}_{[r]}, y_r)f(\mathbf{y}_r|\theta, \mathbf{x}_r)\pi(\theta)} = \frac{f(y_r|\theta, \mathbf{x}_{r(s)})\pi_\delta(\theta)}{f(\mathbf{y}_r|\theta, \mathbf{x}_r)\pi(\theta)}.\end{aligned}$$

The first case called *deletion observation case* is a measure of the effect of the deletion of the r-th observation from the model. The second case studies the *effect of the covariances* where $\mathbf{x}_{[r(s)]}$ is obtained by deleting the s-th component of the r-th

covariate vector form $\mathbf{x}$, $\mathbf{x}_{r(s)}$ is the r-th covariate vector with the s-th component deleted and $\mathbf{x}_{[r]}$ is the covariance matrix with the r-th covariate vector deleted. The problem can be generalized if we consider the case of the influence of a set of observations. We can choose different models to get the perturbations for example the pondered case. In [6] is studied a class of observations for linear models and for generalized linear models. The study can be performed for image recognition using Bayesian adaptive method (see [10] for details) and can be used also in the case of truncated a priori distribution (see [22], [23] and [24] for details). Thus the method is computational and can be used for every class of perturbation models.

4. Monte Carlo estimation and bootstrap estimation

Using Bayes theorem (see for example [18])

$$\delta(\theta,\mathbf{y},\mathbf{x}) = \frac{f_\delta(\mathbf{y}|\theta,\mathbf{x})\pi_\delta(\theta)}{f(\mathbf{y}|\theta,\mathbf{x})\pi(\theta)} = \frac{f_\delta(\mathbf{y}|\mathbf{x})\pi_\delta(\theta|\mathbf{y})}{f(\mathbf{y}|\mathbf{x})\pi(\theta|\mathbf{y})},$$

we obtain

$$\frac{\pi_\delta(\theta|\mathbf{y})}{\pi(\theta|\mathbf{y})} = \delta(\theta,\mathbf{y},\mathbf{x})\frac{f(\mathbf{y}|\mathbf{x})}{f_\delta(\mathbf{y}|\mathbf{x})}$$

and

$$D_\phi = \int \phi(\delta(\theta,\mathbf{y},\mathbf{x})\frac{f(\mathbf{y}|\mathbf{x})}{f_\delta(\mathbf{y}|\mathbf{x})})\pi(\theta|\mathbf{y})d\theta.$$

Marginal distributions are easy to compute (see [8], [9] and [19]) only in particular cases. Let us consider:

$$\begin{aligned}\frac{f_\delta(\mathbf{y}|\mathbf{x})}{f(\mathbf{y}|\mathbf{x})} &= \frac{\int f_\delta(\mathbf{y}|\theta,\mathbf{x})\pi(\theta)d\theta}{f(\mathbf{y}|\mathbf{x})} = \frac{\int \frac{f_\delta(\mathbf{y}|\theta,\mathbf{x})}{f(\mathbf{y}|\theta,\mathbf{x})} f(\mathbf{y}|\theta,\mathbf{x})\pi(\theta)d\theta}{f(\mathbf{y}|\mathbf{x})} \\ &= \int \frac{f_\delta(\mathbf{y}|\theta,\mathbf{x})}{f(\mathbf{y}|\theta,\mathbf{x})}\frac{f(\mathbf{y}|\theta,\mathbf{x})\pi(\theta)}{f(\mathbf{y}|\mathbf{x})}d\theta = \int \frac{f_\delta(\mathbf{y}|\theta,\mathbf{x})}{f(\mathbf{y}|\theta,\mathbf{x})}\pi(\theta|\mathbf{y})d\theta \\ &= \int \frac{f_\delta(\mathbf{y}|\theta,\mathbf{x})\pi_\delta(\theta)}{f(\mathbf{y}|\theta,\mathbf{x})\pi(\theta)}\frac{\pi(\theta)}{\pi_\delta(\theta)}\pi(\theta|\mathbf{y})d\theta = \int \delta(\theta,\mathbf{y},\mathbf{x})\frac{\pi(\theta)}{\pi_\delta(\theta)}\pi(\theta|\mathbf{y})d\theta.\end{aligned}$$

We get

$$D_\phi = \int \phi(\frac{\delta(\theta,\mathbf{y},\mathbf{x})}{\int \delta(\theta,\mathbf{y},\mathbf{x})\frac{\pi(\theta)}{\pi_\delta(\theta)}\pi(\theta|\mathbf{y})d\theta})\pi(\theta|\mathbf{y})d\theta.$$

Remark 1. If $\pi(\theta) = \pi_\delta(\theta)$ we obtain the same result as in [17].

Remark 2. If $\pi(\theta) = 1$ (uniform distribution) then D_ϕ can be used like a measure of the deviation of the distribution $\pi_\delta(\theta)$ from the uniformity. This has applications in testing hardware key generators (see [11] for details).

Remark 3. If $\pi(\theta) = 1$ (uniform distribution) we can take

$$\pi_\delta(\theta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad \delta = (\alpha, \beta).$$

For $\alpha \to 1$ and $\beta \to 1$ we have $\pi_\delta(\theta) \to \pi(\theta)$.

4.1. Calibration of the ϕ-divergence

From the above results we see that for every divergence we have no perturbation if and only if divergence is zero. We must indicate a cutoff point to decide if the perturbation is or not significant. Next we present a calibration method of the ϕ-divergence. For example let us consider the case of random binary variable with $Pr(X = 0) = p$ and $Pr(X = 1) = 1 - p$. Divergence between this biased variable and a binary unbiased random variable ($p = 0.5$) is:

$$D(f_0, f_1) = \int_{\mathcal{X}} \phi(\frac{f_0(x)}{f_1(x)}) f_1(x) dx,$$

where $\mathcal{X} = \{0, 1\}$ and $f_0(x) = p^x(1-p)^{1-x}$ and $f_1(x) = 0.5$. Observe that for $D(f_0, f_1) = d$ then d is a solution of the equation

$$d = \frac{\phi(2p) - \phi(2(1-p))}{2}.$$

For example for L_1 norm we get $d = \dfrac{|1-2p|}{2}$. It is obvious that for $p = 0$ or 1 we have $d = \frac{1}{2}$ and for $p = 0.5$ we have $d = 0$. Function d is symmetric around $p = 0.5$ and attains its minimum in the point for which $f_0 = f_1$. Let us consider the case in which $p > 0.5$ such that $d \in [0, 0.5]$. For $p \leq 0.5$ d will have the same property due to symmetry. Thus $p = 0.6$ can indicate an influent observation, which corresponds to $d = 0.1$. Similarly for $p = 0.75$ is corresponding d is 0.25. We can design a scale in which $d > 0.25$ indicates a influent observation and $d \in [0.1; 0.25]$ a medium influence. In general L_1 norm is between 0 and 1. For different divergences ϕ and different f_0 and f_1 we can decide in a similar way the cutoff point for detection of the outliers data.

4.2. Monte Carlo estimator

The evaluation of D_ϕ is hard even for more simple expressions of $\delta(\theta, \mathbf{y}, \mathbf{x})$. One method is based on the estimation of the integral using *Monte Carlo method* (see

[7]). Thus *Monte Carlo estimator* will be:

$$\hat{D}_{\phi}^{MC} = \frac{1}{N}\sum_{s=1}^{N}\phi\left(\frac{\delta(\theta^{(s)})}{\frac{1}{N}\sum\limits_{s=1}^{N}\delta(\theta^{(s)})\frac{\pi(\theta^s)}{\pi_\delta(\theta^s)}}\right) = \frac{1}{N}\sum_{s=1}^{N}\phi\left(\frac{\delta(\theta^{(s)})}{\frac{1}{N}\sum\limits_{s=1}^{N}\frac{f_\delta(\mathbf{y}|\theta^{(s)},\mathbf{x})}{f(\mathbf{y}|\theta^{(s)},\mathbf{x})}}\right),$$

where $\{\theta^s\}_{s=1}^N$ is a selection of size N from the a posteriori distribution $\pi(\theta|\mathbf{y})$.

For the case in which ϕ is *Kullback divergence* we have

$$\begin{aligned}\hat{D}_{KL,\delta}^{MC} &= -\frac{1}{N}\sum_{s=1}^{N}\log\left(\frac{\delta(\theta^{(s)})}{\frac{1}{N}\sum\limits_{s=1}^{N}\delta(\theta^{(s)})\frac{\pi(\theta^s)}{\pi_\delta(\theta^s)}}\right) = -\frac{1}{N}\log\frac{\prod\limits_{s=1}^{N}\delta(\theta^{(s)})}{\frac{1}{N}\sum\limits_{s=1}^{N}\delta(\theta^{(s)})\frac{\pi(\theta^s)}{\pi_\delta(\theta^s)}} \\ &= -\log\frac{(\prod\limits_{s=1}^{N}\frac{1}{\delta(\theta^{(s)})})^{\frac{1}{N}}}{\frac{1}{N}\sum\limits_{s=1}^{N}\delta(\theta^{(s)})\frac{\pi(\theta^s)}{\pi_\delta(\theta^s)}} = \log\frac{PAM(\delta(\theta),\frac{\pi(\theta)}{\pi_\delta(\theta)})}{GM(\delta(\theta))},\end{aligned}$$

where $GM(\delta(\theta))$ is the geometric mean of the values $\{\delta(\theta^s)\}_{s=1}^N$, and $PAM(\delta(\theta),\frac{\pi(\theta)}{\pi_\delta(\theta)})$ is the pondered arithmetic mean of the values $\{\delta(\theta^s)\}_{s=1}^N$ with ponders $\{\frac{\pi(\theta^s)}{\pi_\delta(\theta^s)}\}_{s=1}^N$.

Remark 4. Let us observe that for $\pi(\theta)=\pi_\delta(\theta)$ we have $PAM(\delta(\theta),\frac{\pi(\theta)}{\pi_\delta(\theta)}) = AM(\delta(\theta))$ thus the arithmetic mean of the values $\{\delta(\theta^s)\}_{s=1}^N$. We obtain a generalization of the results from [17].

For the r-th deletion case we have:

$$\begin{aligned}\hat{D}_{KL,r,\delta}^{MC} &= \log\frac{PAM(\frac{\pi_\delta(\theta)}{f(y_r|\theta,\mathbf{x}_r,\mathbf{y}_{(r)})\pi(\theta)},\frac{\pi(\theta)}{\pi_\delta(\theta)})}{GM(\frac{\pi_\delta(\theta)}{f(y_r|\theta,\mathbf{x}_r,\mathbf{y}_{(r)})\pi(\theta)})} = \log\frac{AM(\frac{1}{f(y_r|\theta,\mathbf{x}_r,\mathbf{y}_{(r)})})}{GM(\frac{\pi_\delta(\theta)}{f(y_r|\theta,\mathbf{x}_r,\mathbf{y}_{(r)})\pi(\theta)})} \\ &= \log AM(\frac{1}{f(y_r|\theta,\mathbf{x}_r,\mathbf{y}_{(r)})}) - \log GM(\frac{\pi_\delta(\theta)}{f(y_r|\theta,\mathbf{x}_r,\mathbf{y}_{(r)})\pi(\theta)}) \\ &= \log(\prod_{s=1}^{N}f(y_r|\theta^s,\mathbf{x}_r,\mathbf{y}_{(r)})\frac{\pi(\theta^s)}{\pi_\delta(\theta^s)})^{\frac{1}{N}} - \log\frac{1}{\frac{1}{N}\sum\limits_{s=1}^{N}\frac{1}{f(y_r|\theta^s,\mathbf{x}_r,\mathbf{y}_{(r)})}}\end{aligned}$$

$$= \log(\prod_{s=1}^{N}f(y_r|\theta^s,\mathbf{x}_r,\mathbf{y}_{(r)}))^{\frac{1}{N}} - \log\frac{1}{\frac{1}{N}\sum\limits_{s=1}^{N}\frac{1}{f(y_r|\theta^s,\mathbf{x}_r,\mathbf{y}_{(r)})}} + \log(\prod_{s=1}^{N}\frac{\pi(\theta^s)}{\pi_\delta(\theta^s)})^{\frac{1}{N}}$$

$$= \hat{D}_{KL,r}^{MC} + \log(\prod_{s=1}^{N} \frac{\pi(\theta^s)}{\pi_\delta(\theta^s)})^{\frac{1}{N}}.$$

where $\hat{D}_{KL,r}^{MC}$ is the value of the discrepancy in case of no perturbation of the a priori distribution, the additional factor, which is an arithmetic mean, $\log(\prod_{s=1}^{N} \frac{\pi(\theta^s)}{\pi_\delta(\theta^s)})^{\frac{1}{N}}$ is due to perturbation of the a priori distribution. A big value of $\hat{D}_{KL,r,\delta}^{MC}$ suggests that the $r-$th observation is influent.

For the cases in which ϕ is χ^2 *divergence* and *variational divergence* we get

$$\begin{aligned}
\hat{D}_{\chi^2,\delta}^{MC} &= \frac{1}{N}\sum_{s=1}^{N}(\frac{\delta(\theta^{(s)})}{\frac{1}{N}\sum\limits_{s=1}^{N}\delta(\theta^{(s)})\frac{\pi(\theta^s)}{\pi_\delta(\theta^s)}} - 1)^2 \\
&= \frac{1}{N}\sum_{s=1}^{N}(\frac{\delta(\theta^{(s)})}{\frac{1}{N}\sum\limits_{s=1}^{N}\delta(\theta^{(s)})\frac{\pi(\theta^s)}{\pi_\delta(\theta^s)}})^2 - \frac{2}{N}\sum_{s=1}^{N}\frac{\delta(\theta^{(s)})}{\frac{1}{N}\sum\limits_{s=1}^{N}\delta(\theta^{(s)})\frac{\pi(\theta^s)}{\pi_\delta(\theta^s)}} + 1 \\
&= \frac{AM(\delta^2(\theta))}{PAM^2(\delta(\theta),\frac{\pi(\theta)}{\pi_\delta(\theta)})} - 2\frac{AM(\delta(\theta))}{PAM(\delta(\theta),\frac{\pi(\theta)}{\pi_\delta(\theta)})} + 1,
\end{aligned}$$

$$\hat{D}_{L_1,\delta}^{MC} = \frac{\sum\limits_{s=1}^{N}|\delta(\theta^{(s)}) - \frac{1}{2N}\sum\limits_{s=1}^{N}\delta(\theta^{(s)})\frac{\pi(\theta^s)}{\pi_\delta(\theta^s)}|}{\sum\limits_{s=1}^{N}\delta(\theta^{(s)})\frac{\pi(\theta^s)}{\pi_\delta(\theta^s)}}.$$

As an alternative, the Laplace method can be used to obtain the posterior divergence. In multidimensional problems and in more complicated models the implementation of this method is difficult while Monte Carlo method is straightforward.

4.3. Bootstrap estimator

Let us suppose that we have the selection $X_1, ..., X_n$ of random variables with unknown density f and we want to generate the variables $Y_1, ..., Y_m$ with the same density f. We construct the estimation $f_n(x) = f_n(x, X_1, ..., X_n)$ of the density $f(x)$ and we take a sample of size m from f_n. The new selection depends of the original selection $X_1, ..., X_n$. We want the new selection to be distributed like the original one. In this case we say about this two selections that are indistinguishable

Because $Y_1, ..., Y_m$ depends conditionally of $X_1, ..., X_n$ we define:

$$D_n = \sup_{A,B} |P(Y \in A, X \in B) - P(Y \in A)P(X \in B)|,$$

for every $A \in \mathbf{R}^d, B \in \mathbf{R}^{nd}$ borel sets where $Y = Y_1$ and $X = (X_1, ..., X_n)$. We say that the selections are *asymptotic independent* if $\lim_{n\to\infty} D_n = 0$.

In case in which $X_1, ..., X_n$ are used to *construct a system* and $Y_1, ..., Y_m$ are used for *testing* it, the dependence of the series produce optimist results. Without the asymptotic independence we cannot hope that increasing n this bias becomes smaller.

First we recall a result due to *Scheffe*.

Theorem 1. *If f and g are densities on $\mathbf{R}^d$ then we have*

$$\int |f - g| = 2 \sup_B |\int_B f - \int_B g|.$$

Theorem 2. *If f_n is an estimation of the density which is it self a density then $D_n \leq E(\int |f - f_n|)$.*

We say that the sequence of estimations f_n is *consistent* if and only if $\lim_{n\to\infty} E(\int |f - f_n|) = 0$ for all densities f. Observe that if the sequence is consistent then using the above theorem the samples are asymptotic independent. Consistency does not imply asymptotic independence.

As we see the Monte Carlo estimation is based on a selection from the a posteriori distribution (unknown) $\pi(\theta|\mathbf{y})$. It is known that computing a posteriori density is quite difficult even in particular cases (see [8], [9] and [18]) . From this reason if the parameter θ is 1 dimensional, we approximate the a posteriori density $\pi(\theta|\mathbf{y})$ with a *kernel distribution* estimation:

$$\pi_N(\theta|\mathbf{y}) = \frac{1}{Nh^d} \sum_{i=1}^{N} K(\frac{\theta - \theta^{(i)}}{h}),$$

where $\{\theta^{(i)}\}_{i=1}^N$ is a selection from the a posteriori density $\pi(\theta|\mathbf{y})$. The sequences of estimations $\pi_N(\theta|\mathbf{y})$ is consistent if and only if $\lim_{N\to\infty, h\to 0} Nh^d = \infty$ in probability. Consistence is the base requirement in theory of bootstrap estimations.

Bootstrap estimation of D_ϕ will be:

$$\hat{D}_\phi^{bstrp} = \frac{1}{m} \sum_{s=1}^{m} \phi(\frac{\delta(\theta^{(s)})}{\frac{1}{m} \sum_{s=1}^{m} \delta(\theta^{(s)}) \frac{\pi(\theta^s)}{\pi_\delta(\theta^s)}}),$$

where $\{\theta^s\}_{s=1}^m$ is a selection of size m from the kernel estimation of the a posteriori distribution $\pi_N(\theta|\mathbf{y})$. This estimation of D_ϕ depends of the number of bootstrap

samples and the kernel estimator $\pi_N(\theta|\mathbf{y})$ of the a posteriori distribution $\pi(\theta|\mathbf{y})$. Proceed similarly as in Monte Carlo estimation we obtain simple formulas for different cases for ϕ.

Remark 5. We have the following inequality:

$$|\int \phi(\frac{\delta(\theta,\mathbf{y},\mathbf{x})}{\int \delta(\theta,\mathbf{y},\mathbf{x})\frac{\pi(\theta)}{\pi_\delta(\theta)}\pi(\theta|\mathbf{y})d\theta})\pi(\theta|\mathbf{y})d\theta - \int \phi(\frac{\delta(\theta,\mathbf{y},\mathbf{x})}{\int \delta(\theta,\mathbf{y},\mathbf{x})\frac{\pi(\theta)}{\pi_\delta(\theta)}\pi(\theta|\mathbf{y})d\theta})\pi_N(\theta|\mathbf{y})d\theta|$$

$$\leq 2\sup\{\phi(\frac{\delta(\theta,\mathbf{y},\mathbf{x})}{\int \delta(\theta,\mathbf{y},\mathbf{x})\frac{\pi(\theta)}{\pi_\delta(\theta)}\pi(\theta|\mathbf{y})d\theta})\}\int |\pi(\theta|\mathbf{y}) - \pi_N(\theta|\mathbf{y})|d\theta,$$

which is proved with Scheffe's result and Theorem 2.

Remark 6. Mean square error of the bootstrap sample is smaller then mean square error of the Monte Carlo estimator.

5. Applications to the RCA models

In [17] are presented some applications of the Monte Carlo estimations for the following models probit, logistic, overdisipated generalized linear and nonlinear regression. We present a new application more exactly autoregressive model of first order with random coefficients. The model **RCA**(1) has the form

$$Y_t = \theta_t Y_{t-1} + e_t,$$

$$\theta_t = \mu + \varepsilon_t,$$

where $e_t \sim N(0, \frac{1}{\tau})$ and $\varepsilon_t \sim N(0, \frac{1}{\tau\xi})$ are independent. Parameter μ is a real number. We have the following theorem.

Theorem 3. *For the above model consider the following a priori distributions on the parameter space: $\tau \sim Gamma(\alpha, \beta)$, $\mu \sim U(0,1)$, and for ξ the improper distribution Jeffreys ([18]). Then we find the following a posteriori distributions:*

$$\tau|\theta,\xi,\mu,Y \sim Gamma(n+\alpha, \frac{1}{2}(\sum_{t=1}^{n}(Y_t - \theta_t Y_{t-1})^2 + \xi\sum_{t=1}^{n}(\theta_t - \mu)^2 + 2\beta)),$$

$$\xi|\tau,\theta,\mu,Y \sim Gamma(\frac{n}{2}, \frac{1}{2}\tau(\sum_{t=1}^{n}(\theta_t - \mu)^2)),$$

$$\mu|\theta,\tau,\xi,Y \sim N(\frac{1}{n}\sum_{t=1}^{n}\theta_t, \frac{(\tau\xi)^{-1}}{n}),$$

$$\theta_t|\mu,\tau,\xi,Y \sim N(\frac{Y_tY_{t-1}+\xi\mu}{Y_{t-1}^2+\xi}, (\tau(Y_{t-1}^2+\xi))^{-1}),$$

and the prediction density:

$$Y_{n+1}|\theta,\theta_{n+1},\tau,\xi,\mu,Y \sim N(\theta_{n+1}Y_n,\tau^{-1}).$$

To construct a Monte Carlo estimator we must draw a sample from the prediction distribution which is made by *Gibbs sample*. Monte Carlo estimator results by applying the corresponding formulas.

A cryptographic application of RCA model is in the *cryptanalysis of a stream cipher system*: the plain text can be viewed like the noise parameters e_t which a priori distribution is known. Of course this distribution is not necessary a normal one. The computations are done in a similar way. In robust techniques the normality assumption does not influence the obtained results. The cipher algorithm is assumed to be a memory algorithm.

Another cryptographic application is the *reconstruction problem of the equivalent linear complexity sequence and the plain text* (here the cipher algorithm consists in bitwise of a pseudorandom sequence with the plain text). The parameters θ_t can be seen like algorithm settings.

References

[1] R. Azencott and D. Dacunha-Castelle, *Series of irregular Observations. Forecasting and Model Building*, Springer-Verlag, 1986.

[2] I.V.Basawa and B.L.S. Prakasa Rao *Statistical Inference for Stochastic Processes*, Academic Press, 1980.

[3] H. Becker and F. Piper, *Cipher Systems*, John Wiley & Sons, New York, 1982.

[4] E.P. Box and G.M. Jenkins, *Time series analysis, forecasting and control*, Holden-Day, 1970.

[5] P. Brockwell and R. Davis, *Time Series Theory and Methods*, Springer-Verlag, 1987.

[6] R. D. Cook, *Assessment of local influence (with discussion)*, J. Roy. Statist. Soc. Ser. B, 48, (1986), 133-169.

[7] L. Devroye, *Non uniform random generators*, Springer-Verlag, 1986.

[8] N. Johnson, S. Kotz, *Discrete Distributions,* John-Wiley &Sons, 1994.

[9] N. Johnson, S. Kotz, *Continuous Univariate Distributions*, John-Wiley &Sons, 1994.

[10] R. Klein, *Image reconstruction by adaptive Bayesian classification with a locally proportional prior.* Commun. Statist.- Theory and Methods, 22, 10, (1993).

[11] U. Maurer, *A Universal Test for Random Bit Generators*, J. of Cryptology, 5, 2, (1992).

[12] M. L. Menendez, *The minimum ϕ-divergence estimates of fuzzy observations: Statistical applications*, Fuzzy Sets and Systems, 96, (1998), 101-109.

[13] Gh. Mihoc and M. Craiu, *Inferenţă pentru variabile dependente*, Ed. Academiei, 1972.

[14] Gh. Mihoc and V. Craiu, *Tratat de Statistică Matematică*, Editura Academiei R.S.R., Bucuresti, 1985.

[15] A. Pankratz, *Forecasting with Dynamic Regression Models*, Wiley&Sons, 1991.

[16] J. A. Pardo and M. L. Menendez, *Some statistical applications of generalized Jensen difference measures for fuzzy information systems*, Fuzzy Sets and Systems, 52, (1992), 169-180.

[17] F. Peng, *Bayesian analysis of outlier problems using divergence measures*, The Canadian Journal of Statistics., 23, 2, (1995).

[18] V. Preda, *Teoria Deciziilor Statistice*, Ed. Academiei, 1991.

[19] J. Santer and E. Duffy, *The Statistical Analysis of Discrete Data*, Springer-Verlag, 1989.

[20] B. Schneier, *Applied Cryptography with Source Code in C++*, Addison-Wesley, 1996.

[21] B. Schneier, *Cryptanalitic Attacks on Pseudorandom Number Generators*, Proc. Fast Software Encryption, 1998.

[22] O.A. Shalaby, *Bayesian Inference in Truncated and in Censored Exponential Distribution and Reliability Estimation*, Commun. Statist.- Theory and Methods, 22, 1, 1993.

[23] E. Simion, *Remarks on the Bayesian truncated estimation of Shannon entropy using priori truncated Dirichlet distribution*, 51, 2, (1999).

[24] E. Simion, *Truncated Bayesian estimation of Renyi entropy and cryptographic applications*, Math. Reports, 52, 2, (2000).

Mircea Andraşiu
Independent cryptologist
email: *mircea_andrasiu@yahoo.com*

Andrei Oprina
"Simion Stoilow" Institute of Mathematics of the Romanian Academy
P.O. Box 1-764, 014700 Bucharest, Romania
email: *oandrei22@yahoo.com*

Emil Simion
"Simion Stoilow" Institute of Mathematics of the Romanian Academy (associate researcher)
P.O. Box 1-764, 014700 Bucharest, Romania
email: *esimion@fmi.unibuc.ro*

Gheorghe Simion
Faculty of Applied Sciences
University Politehnica of Bucharest
313 Splaiul Independenţei St., 060042 Bucharest, Romania
email: *gheorghesimion@hotmail.com*