Robustification Process on Bayes Estimators

Anton Abdulbasah Kamil

School of Mathematical Sciences, Universiti Sains Malaysia
11800 USM, Minden, Penang, Malaysia
email: anton@cs.usm.my

Abstract. *The paper describes one possible robustification process on Bayes estimators and studies how a robust estimator can work with prior information. This robustification procedure, as one of possible sensitivity analysis, enables us to study the effect of the outlying observations together with sensitivity to a chosen prior distribution or to a chosen loss function. Consider i.i.d. d-dimensional random vectors $X_1,...,X_n$ with a distribution P_θ depending on an unknown parameter $\theta \in \Theta \subset R^l$. We deal with robust counterparts of maximum posterior likelihood estimators and Bayes estimators in the inference on θ. Asymptotic properties of these robust versions, including their asymptotic equivalence of order $o_p(n^{-1})$, are proven.*

Keywords: Bayes type estimators, robustification process, asymptotic theory.

1. Introduction

Let $X_1,...,X_n$ be i.d.d. d-dimensional random vectors with a common d.f. $F(x,\theta)$, where $\theta \in \Theta \subset R^l$ is an unknown parameter. Let $\pi(\theta)$ be a density of θ vanishing outside Θ. Denote by θ_0 the "true" value of θ; for the sake of simplicity we shall use the notation $\theta = \theta$.

In this paper we shall deal with Bayes and Bayes-type estimators. That is, we are interested in a suitable approximation of the integral ratio of the form

$$\frac{\int w(\theta).\exp\{L(\theta)\}d\theta}{\int v(\theta).\exp\{L(\theta)\}d\theta},\tag{1}$$

with appropriate w and v.

Specifically, if $w(\theta) = \theta.v(\theta)$ and $v(\theta)$ is a prior distribution for θ, then (1) becomes the posterior mean of θ. The question of the effect of distribution of $X_1,...,X_n$, or robustness with respect to data, can be discussed when we put $L(\theta) = -\sum_{i=1}^{n}\rho(X_i,\theta)$, where ρ varies over a class of appropriate functions (Hampel et al, 1986). This model also includes as a special case $\ln f(x,\theta) = -\rho(x,\theta)$. Before

starting to evaluate acceptable estimators for this theoretical setup, we note that this situation has important practical aspects.

Applied statisticians often wonder whether or not we can use prior information in the case when the data contain gross errors or when they are contaminated by a heavy tailed distribution. To solve this practical problem, we combine robust and Bayesian approaches. Namely, we want to find a class of robust estimators that take into account preliminary information. Analogously, from Bayesian point of view we look for Bayes-type estimators being sufficiently insensitive to deviation from a classical model. Accordingly, we need an estimator that puts Bayesian and robust viewpoints together.

Following these ideas, two main steps are to be taken:

1. Identification of the outliers and a subsequent application of the classical Bayes analysis to the remaining data set.
2. Modification of the robust methods, consisting either of plugging the prior information inside the estimators or of altering the Bayes method by using the robust approach.

One possible construction of a robust method is to start with the classical method and replace the density in the definition of the estimator with the function $c.\exp\{-\rho(X)\}$ that makes the resulting estimator more robust.

- The generalized maximum-likelihood type or M-estimator M_n is defined as

$$M_n \in \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} \rho(X_i, \theta) \tag{2}$$

- The Bayes type or B-estimator T_n is defined as

$$T_n = \frac{\int_{\Theta} \theta.\exp\{-\sum_{i=1}^{n} \rho(X_i, \theta)\}\pi(\theta)d\theta}{\int_{\Theta} \exp\{-\sum_{i=}^{n} \rho(X, \theta)\}\pi(\theta)d\theta} \tag{3}$$

if both the integrals exist (Akahira & Takeuchi, 1981; De Bruijn, 1961; Ibragimov & Khaminskii, 1981).

Analogously to (2), one can define a robust version of maximum posterior likelihood estimators $\hat{\theta}_n$ as

$$\hat{\theta}_n \in \arg\min_{\theta \in \Theta} \{\sum_{i=1}^{n} \rho(X_i, \theta) - \ln \pi(\theta)\} \tag{4}$$

Clearly, (4) is MLE with respect to posterior density (Berger, 1984; Tierney & Kadane, 1986).

Notice that T_n and $\hat{\theta}_n$ are obtained as generalizations of Bayes estimators and maximum posterior likelihood estimators, respectively. Indeed, putting,

$$\rho(X, \theta) = -\ln f(x, \theta) \tag{5}$$

The main technical tool is a Laplace type approximation and the fact that $\hat{\theta}_n$ is a saddle point of both integrals in (3). From this point of view, one can look at $\hat{\theta}_n$ as a computational version of T_n. Indeed, we can prove that asymptotically, under certain assumptions,

$$T_n = \hat{\theta}_n + o_p(n^{-1}) \tag{6}$$

(Strasser, 1977).

The choice $\rho = -\ln f$ again leads to a relation between Bayes posterior mean and maximum posterior likelihood estimator.

2. The Results

The asymptotic properties of the introduced estimators will be proven under the following sets of conditions:

(a)

(a1) $\Theta \subset R^l$ is an open connected set.

(a2) $\rho : R^d \times \Theta \to R_-$ is a continuous function.

Moreover, $\dfrac{\partial^2 \rho}{\partial \theta^2}$ exists and for every $\theta \in \Theta$ there exist $\delta > 0, \beta \geq 0$ such that for every $\xi, \eta \in \Theta$, $\|\xi - \theta\| < \delta, \|\eta - \theta\| < \delta$ and every $x \in R^d$ it holds,

$$\left\| \frac{\partial^2 \rho}{\partial \theta^2}(x,\xi) - \frac{\partial^2 \rho}{\partial \theta^2}(x,\eta) \right\| \leq \beta.\|\xi - \eta\|.$$

(a3) $\pi : \Theta \to R_+$ is bounded and $\ln \pi$ is well-defined with a continuous derivative $\dfrac{\partial \ln \pi}{\partial \theta}$.

(a4) The integral $\displaystyle\int_\Theta \|\theta\| \exp(-\rho(x,\theta))\pi(\theta)d\theta$ exists for every $x \in R^d$.

(b)

(b1) For every $n \in N$

$$\int M_n(x)dF(x,\theta_0) < +\infty, \quad \text{where } M_n(x) = \sup_{\substack{\|\theta\| \leq n \\ \theta \in \Theta}} \left\| \frac{\partial^2 \rho}{\partial \theta^2}(x,\theta) \right\|.$$

(b2) There exists a point $\theta^* \in R^l$ such that $\displaystyle\int \rho(x,\theta^*)dF(x,\theta_0)$ and

$$\int \frac{\partial \rho}{\partial \theta}\rho(x,\theta^*)dF(x,\theta_0) \text{ is finite.}$$

(c)

(c1) We assume that the function

$$h(\theta) = \int \rho(x,\theta)dF(x,\theta_0) \tag{7}$$

has a unique absolute minimum at $\theta = \theta_0$, i.e.

$$\theta_0 \in \arg\min_{\theta \in \Theta} h(\theta) \tag{8}$$

(c2) If $\sup\limits_{\theta \in \Theta}\|\theta\| = +\infty$ then

$$h(\theta_0) < \overline{\rho} \stackrel{df.}{=} \inf_{K>0} \lim_{\|\theta\|\to+\infty} \inf_{\|x\|\leq K} \rho(x,\theta) \tag{9}$$

(c3) $\dfrac{\partial^2 h}{\partial \theta^2}(\theta_0)$ is a positive definite matrix

(c4) $\int \dfrac{\partial \rho}{\partial \theta}(x,\theta_0) \left(\dfrac{\partial \rho}{\partial \theta}(x,\theta_0) \right)^T dF(x,\theta_0)$ is a real matrix.

The group (a) contains conditions on the loss function ρ and on prior information, which is represented by the function π and by the set Θ. Changing the order of derivation and integration is guaranteed by the conditions of the group (b). Finally, the group (c) requires θ_0 to be an optimal solution of (8) with nice properties. There is only the condition (a2) which needs a special discussion, the others are quite natural. The condition (a2) weakens an existence of continuous $\dfrac{\partial^3 \rho}{\partial \theta^3}$. But omitting (a2) we would lose the rate n^{-1} of $T_n - \hat{\theta}_n$.

Theorem 1.
Suppose the groups of conditions (a), (b), and (c) are satisfied. Then, for $n \to +\infty$,
$$\sqrt{n}\|\hat{\theta}_n - \theta_0\| = o_p(1).$$

later this theorem as additional and supporting tool for Theorem 2.

Theorem 2.
Suppose the groups of conditions (a), (b), and (c) are satisfied. Then, for $n \to +\infty$,
$$\sqrt{n}\|T_n - \theta_0\| = o_p(1)$$
and
$$n\|T_n - \hat{\theta}_n\| = o_p(1).$$

Assuming, moreover, that $\dfrac{\partial^3 \rho}{\partial \theta^3}$ exists and that,

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^3 \rho}{\partial \theta^3}\left(X_i, \theta_0 + \frac{1}{\sqrt{n}}\theta \right) \xrightarrow{a.s.} \frac{\partial^3 h}{\partial \theta^3}(\theta_0)$$

uniformly for $\|\theta\| \leq \delta_0$, we have

$$T_n = \hat{\theta}_n + n^{-1} \cdot \frac{A}{B} + o_p(n^{-1}),$$

where

$$A = -\frac{1}{6}\int_{R^l}\theta \sum_{l_1,l_2,l_3=1}^{l} \theta_{l_1}\theta_{l_2}\theta_{l_3} \frac{\partial^3 h}{\partial\theta_{l_1}\partial\theta_{l_2}\partial\theta_{l_3}}(\theta_0).\exp\left(-\frac{1}{2}\theta^T \frac{\partial^2 h}{\partial\theta^2}(\theta_0)\theta\right)d\theta$$

and

$$B = \int_{R^l} \exp\left(-\frac{1}{2} \theta^T \frac{\partial^2 h}{\partial \theta^2}(\theta_0)\theta \right) d\theta .$$

Note that one may obtain $T_n = \hat{\theta}_n + o_p(n^{-1})$ in a number of cases. Consider an estimator of location, where $\rho(x,\theta) = \eta(x-\theta)$. Having η symmetric at zero and assuming that the true distribution is symmetric at θ_0, i.e. $F(x,\theta_0) = 1 - F(2\theta_0 - x, \theta_0)$, one deduces $\frac{\partial^3 h}{\partial \theta^3}(\theta_0) = 0$. Consequently, A = 0 and $T_n = \hat{\theta}_n + o_p(n^{-1})$.

3. Proof of Theorems

The steps of the proof of Theorems use the following auxiliary lemmas,

Lemma 1
Let conditions (a1) – (a2), (b1) – (b2) be fulfilled. Then for every $\theta \in \Theta$ the integrals

$$h(\theta) = \int \rho(x,\theta) dF(x) \tag{10}$$

$$\frac{\partial h}{\partial \theta}(\theta) = \int \frac{\partial \rho}{\partial \theta}(x,\theta) dF(x) \tag{11}$$

$$\frac{\partial^2 h}{\partial \theta^2}(\theta) = \int \frac{\partial^2 \rho}{\partial \theta^2}(x,\theta) dF(x) \tag{12}$$

are finite.

Lemma 2
Let conditions (a1), (a2), (b) be fulfilled. Then there exists a set $A \in \lambda$, $P(A) = 1$ such that for every $\omega \in A$ and $\theta \in \Theta$.

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \rho}{\partial \theta^2}(X_i(\omega),\theta) \rightarrow \frac{\partial^2 h}{\partial \theta^2}(\theta), \tag{13}$$

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \rho}{\partial \theta}(X_i(\omega),\theta) \rightarrow \frac{\partial h}{\partial \theta}(\theta), \tag{14}$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \rho(X_i(\omega),\theta) \rightarrow h(\theta) \tag{15}$$

hold.

Lemma 3
Let assumptions (a), (b), (c) be fulfilled. Then for every $\delta > 0$ there exists $\Delta > 0$ such that,

$$\inf_{\|\theta-\theta_0\|\geq\delta} \left\{ \frac{1}{n} \sum_{i=1}^{n} \rho(X_i,\theta) \right\} \geq \frac{1}{n} \sum_{i=1}^{n} \rho(X_i,\theta_0) - \frac{1}{n} \ln \pi(\theta_0) + \Delta$$

holds for n sufficiently large with probability 1.

Corollary 1

Under assumptions (a), (b), (c) we have $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$.

Proof of Theorem 1: Let $A \in \lambda$, $P(A) = 1$ be such that the assertions of Lemma 2 and Lemma 3 hold. Denote by α the smallest eigenvalue of $\dfrac{\partial^2 h}{\partial \theta^2}(\theta^2)$; $\alpha > 0$ by assumption (c3). Θ being open, there is $\delta_1 > 0$ such that,

$$u_1 = \left\{ \theta \in R^l \mid \|\theta - \theta_0\| < \delta_1 \right\} \subset \Theta.$$

For every $\omega \in A$,

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \rho}{\partial \theta^2}(X_i(\omega),\theta) \to \frac{\partial^2 h}{\partial \theta^2}(\theta)$$

uniformly on u_1 because of (a2).

Therefore, by (a2) there is $0 < \delta < \delta_1$ such that for every $y \in R^k$ and a sufficiently large n,

$$\inf_{\|\theta-\theta_0\|<\delta} y^T \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \rho}{\partial \theta^2}(X_i(\omega),\theta) y > \frac{\alpha}{2}\|y\|^2 .$$

Thus,

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(X_i(\omega), \theta_0 + \frac{\theta}{\sqrt{n}} \right) + \frac{1}{n}\ln \pi\left(\theta_0 + \frac{\theta}{\sqrt{n}} \right) =$$

$$\frac{1}{n}\sum_{i=1}^{n} \rho(X_i(\omega),\theta_0) - \frac{1}{n}\ln \pi(\theta_0) + \frac{1}{n}\theta^T \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\partial \rho}{\partial \theta}(X_i(\omega),\theta_0)$$

$$+ \frac{1}{n}\int_0^1\!\!\int_0^s \theta^T\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \upsilon}{\partial \theta^2}\left(X_i(\omega), \theta_0 + \frac{\theta}{\sqrt{n}}t \right)\theta\, dt\, ds - \frac{1}{n}\left[\ln \pi\left(\theta_0 + \frac{\theta}{\sqrt{n}} \right) - \ln \pi(\theta_0) \right] \right)$$

$$\geq \frac{1}{n}\sum_{i=1}^{n} \rho(X_i(\omega),\theta_0) - \frac{1}{n}\sum_{i=1}^{n} \rho(X_i(\omega),\theta_0) - \frac{1}{n}\ln \pi(\theta_0)$$

$$+ \frac{1}{n}\left(\frac{\alpha}{4}\|\theta\|^2 + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \theta^T \frac{\partial \rho}{\partial \theta}(X_i(\omega),\theta_0) - \ln\left(\frac{\pi\left(\theta_0 + \dfrac{\theta}{\sqrt{n}}\right)}{\pi(\theta_0)} \right) \right)$$

$$\geq \frac{1}{n}\sum_{i=1}^{n} \rho(X_i(\omega),\theta_0) - \frac{1}{n}\ln \pi(\theta_0) + \frac{1}{n}\left(\frac{\alpha}{4}\|\theta\|^2 + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \theta^T \frac{\partial \rho}{\partial \theta}(X_i(\omega),\theta_0) - Q \right)$$

for every θ with the property $\|\theta\| \leq \sqrt{n}\delta$;

$$Q = \sup_{\|\eta-\theta_0\|\leq\delta} \ln\left(\frac{\pi(\eta)}{\pi(\theta_0)} \right)$$ is finite since π is continuous and positive by (a3).

Notice that $\frac{\partial \rho}{\partial \theta}(X_1, \theta_0), \ldots, \frac{\partial \rho}{\partial \theta}(X_n, \theta_0)$ are i.i.d. random vectors with zero mean and a finite variance according to (c1), (c4). Therefore $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \rho}{\partial \theta}(X_i, \theta_0) \xrightarrow{d} Y$, ($d$: total differentiation), where Y is a Gaussian random vector with zero mean and $\mathrm{var}(Y) = \int \frac{\partial \rho}{\partial \theta}(x, \theta_0) \cdot \left(\frac{\partial \rho}{\partial \theta}(x, \theta_0) \right)^T dF(x, \theta_0)$. Finally, we have,

$$P\left(\sqrt{n} \left\| \hat{\theta}_n - \theta_0 \right\| > H \right) \le P\left(\left\| \hat{\theta}_n - \theta_0 \right\| > \delta \right) +$$

$$P\left(\frac{\alpha}{4} n \left\| \hat{\theta}_n - \theta_0 \right\|^2 - Q < -\frac{1}{n} \sum_{i=1}^{n} \sqrt{n}(\hat{\theta}_n - \theta_0) \frac{\partial \rho}{\partial \theta}(X_i, \theta_0), \delta \ge \left\| \hat{\theta}_n - \theta_0 \right\| > \frac{1}{\sqrt{n}} H \right)$$

$$\le P\left(\left\| \hat{\theta}_n - \theta_0 \right\| > \delta \right) + P\left(\frac{\alpha}{4} n \left\| \hat{\theta}_n - \theta_0 \right\|^2 - Q < \sqrt{n} \left\| \hat{\theta}_n - \theta_0 \right\| \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \rho}{\partial \theta}(X_i, \theta_0) \right\|, \left\| \hat{\theta}_n - \theta_0 \right\| > \frac{1}{\sqrt{n}} H \right)$$

$$\le P\left(\left\| \hat{\theta}_n - \theta_0 \right\| > \delta \right) + P\left(\frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{n} \frac{\partial \rho}{\partial \theta}(X_i, \theta_0) \right\| > \frac{\alpha}{4} H - \frac{Q}{H} \right)$$

therefore $\varlimsup P\left(\sqrt{n} \left\| \hat{\theta}_n - \theta_0 \right\| > H \right) \le P\left(\left\| Y \right\| > \frac{\alpha}{4} H - \frac{Q}{H} \right)$ because of Corollary 1, $\left\| Y \right\|$ is a random variable, and H can be chosen such that the right-hand side is arbitrarily small. Theorem 1 is thus completely proven.

Theorem 1 is a first step in the proof of Theorem 2. The remaining part of the proof is divided into the following auxiliary lemmas,

Lemma 4
Let the groups of assumptions (a), (b), (c) be fulfilled. Then for every $\delta > 0$ there exists $\Delta > 0$ such that

$$\left\| \int_{\substack{\|\theta - \theta_0\| > \delta \\ \theta \in \Theta}} \theta \cdot \exp\left\{ -\sum_{i=1}^{n} \rho(X_i, \theta) + \ln \pi(\theta) \right\} d\theta \right\| \le B_n(\Delta) . o_p(1)$$

and

$$\left\| \int_{\substack{\|\theta - \theta_0\| > \delta \\ \theta \in \Theta}} \exp\left\{ -\sum_{i=1}^{n} \rho(X_i, \theta) + \ln \pi(\theta) \right\} d\theta \right\| \le B_n(\Delta) . o_p(1)$$

where

$$B_n(\Delta) = \exp(-n\Delta) . \exp\left(-\sum_{i=1}^{n} \rho(X_i, \hat{\theta}_n) + \ln \pi(\hat{\theta}_n) \right)$$

Lemma 5
Under the assumptions of Theorem 2,

$$\int_{\Theta} \exp\left(-\sum_{i=1}^{n} \rho(X_i, \theta) + \ln \pi(\theta)\right) d\theta =$$

$$\left(\frac{1}{\sqrt{n}}\right)^l \exp\left(-\sum_{i=1}^{n} \rho(X_i, \hat{\theta}_n) + \ln \pi(\hat{\theta}_n)\right) \cdot \left(\int_{R^l} \exp\left(-\frac{1}{2}\theta^T \frac{\partial^2 h}{\partial \theta^2}(\theta_0)\theta\right) d\theta + o_p(1)\right)$$

Lemma 6

Let all the assumptions (a), (b), (c) hold. Then,

$$\int_{\Theta} \theta . \exp\left(-\sum_{i=1}^{n} \rho(X_i, \theta) + \ln \pi(\theta)\right) d\theta = \hat{\theta}_n \int_{\Theta} \exp\left(-\sum_{i=1}^{n} \rho(X_i, \theta) + \ln \pi(\theta)\right) d\theta +$$

$$\frac{1}{n} . \left(\frac{1}{\sqrt{n}}\right)^l \exp\left(-\sum_{i=1}^{n} \rho(X_i, \hat{\theta}_n) + \ln \pi(\hat{\theta}_n)\right) . A_n$$

where

$$A_n = o_p(1) \text{ as } n \to \infty .$$

Additionally if $\dfrac{\partial^3 \rho}{\partial \theta^3}$ is finite and $\dfrac{1}{n}\sum_{i=1}^{n}\left(\dfrac{\partial^3 \rho}{\partial \theta^3}(X_i, \theta_0) + \dfrac{1}{\sqrt{n}}\theta\right) \xrightarrow{a.s.} \dfrac{\partial^3 h}{\partial \theta^3}(\theta^0)$ uniformly

for $\|\theta\| \le \delta_0$ then $A_n = A + o_p(1)$, where

$$A = -\frac{1}{6}\int_{R^l} \theta \sum_{l_1, l_2, l_3 = 1}^{l} \theta_{l_1}\theta_{l_2}\theta_{l_3} \frac{\partial^3 h}{\partial \theta_{l_1} \partial \theta_{l_2} \partial \theta_{l_3}}(\theta_0) . \exp\left(-\frac{1}{2}\theta^T \frac{\partial^2 h}{\partial \theta^2}(\theta_0)\theta\right) d\theta$$

Proof of Theorem 2: Using Lemma 5 and Lemma 6 one can derive

$$T_n = \frac{\int_{\Theta} \theta . \exp\left(-\sum_{i=1}^{n} \rho(X_i, \theta)\right)\pi(\theta)d\theta}{\int_{\Theta} \exp\left(-\sum_{i=1}^{n} \rho(X_i, \theta)\right)\pi(\theta)d\theta} = \hat{\theta}_n + \frac{\frac{1}{n}\left(\frac{1}{\sqrt{n}}\right)^l \exp\left(-\sum_{i=1}^{n} \rho(X_i, \hat{\theta}_n) + \ln \pi(\hat{\theta}_n)\right)A_n}{\left(\frac{1}{\sqrt{n}}\right)^l \exp\left(-\sum_{i=1}^{n} \rho(X_i, \hat{\theta}_n) + \ln \pi(\hat{\theta}_n)\right)(B + o_p(1))}$$

$$= \hat{\theta}_n + \frac{1}{n}\frac{o_p(1)}{B + o_p(1)} = \hat{\theta}_n - \frac{1}{n}o_p(1).$$

Under the additional assumptions we have,

$$T_n = \hat{\theta}_n + \frac{1}{n} . \frac{A}{B} + o_p\left(\frac{1}{n}\right)$$

References

[1] Akahira, M & Takeuchi, K., *Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency*, Lecture Notes in Statistics, 7, Springer-Verlag, New York, 1981.

[2] Berger, J., *The robust Bayesian viewpoint*, In: Robustness in Bayesian Statistics (J.Kadane, ed.), North-Holland, Amsterdam, 1984.

[3] De Bruijn, N.G., *Asymptotic Methods in Analysis*, Noth-Holland, Amsterdam, 1961.

[4] Hampel, F. R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A., *Robust Statistics: The Approach Based on Influence Functions*, J. Wiley, New York, 1986.

[5] Ibragimov, I.A., Khaminskii, R.Z., *Statistical Estimation. Asymptotic Theory*, Springer-Verlag, Berlin, 1981.

[6] Strasser, H., *Asymptotic expansions for Bayes procedures*, In: Recent Development in Statistics (J.R. Barra et al.Ed.), Amsterdam: North-Holland, Amsterdam, (1977), 9-35.

[7] Tierney, L.& Kadane, J., *Accurate approximations for posterior moments and marginal desities*, JASA, 81(1986), 82-86.