KYBERNETIKA — VOLUME 30 (1994), NUMBER 3, PAGES 271–278

GENERALIZED BAYESIAN-TYPE ESTIMATORS. ROBUST AND SENSITIVITY ANALYSIS

JAN HANOUSEK

Let $X_1, X_2, \dots, X_n$, be i.i.d. random variables with a density function $f(x,\theta)$ where $\theta \in \Theta \subset R^k$ is an unknown parameter that we are interested in estimating. Following up robustification procedure presented by Huber [10] we shall study one possible approach for using (non-sample) prior information for robust type estimators and prove some asymptotic properties of introduced estimators. We shall show that the Bayes-type estimators and maximum posterior probability estimators are asymptotically equivalent to the order $O_p(n^{-1})$ or $o_p(n^{-1})$, depending on some regularity conditions. Because of this asymptotic relation, one expects that with an appropriate choice of ρ (i. e such as we would use in generating an M-estimator) we can obtain a Bayesian type estimator with good robustness properties.

In addition, if $f(x,\theta) = \exp\{-\rho(X_i,\theta)\}$ then these results lead to relations of maximum likelihood and Bayes' estimators.

1. INTRODUCTION

Let $X_1, X_2, \dots, X_n$, be i.i.d. random variables with a density function $f(x,\theta)$ where $\theta \in \Theta \subset R^k$ is an unknown parameter that we are interested in estimating.

In this paper we shall deal with the problem of how we can use prior information about some unknown parameter in estimation procedure, if the data contain gross errors or are contaminated by a heavy-tailed distribution.

We shall concentrate on a possible modification of standard robust procedures that operates with a prior distribution and we shall show that the Bayes-type estimators and maximum posterior probability estimators are asymptotically equivalent to the orders $O_p(n^{-1})$ and $o_p(n^{-1})$ (Hanousek, Lachout [9]), depending on some regularity conditions.

Definition 1. The Bayes-type (or B-) estimator (see Hanousek [5], Hanousek, Jurečková, Saleh [6]) B_n^* is defined as

$$B_n^* = \frac{\int_\theta \theta \cdot \exp\{-\sum_{i=1}^n \rho(X_i,\theta)\} \cdot \pi(\theta) \mathrm{d}\theta}{\int_\theta \exp\{-\sum_{i=1}^n \rho(X_i,\theta)\} \cdot \pi(\theta) \mathrm{d}\theta}$$

if both integrals exist.

Definition 2. The maximum posterior likelihood-type estimator (MPL- or posterior M- estimators) M_n^* is defined as

$$M_n^* \in \operatorname*{argmax}_{\theta\in\Theta} \left(-\sum_{i=1}^{n} \rho(X_i,\theta) + \ln \pi(\theta) \right).$$

Remark. B-estimator is a generalization of Pitman type or P-estimator (see Johns [17]), Huber [12] and Hanousek [4]) P_n,

$$P_n = \frac{\int_\theta \theta \cdot \exp\{-\sum_{i=1}^n \rho(X_i,\theta)\}\,\mathrm{d}\theta}{\int_\theta \exp\{-\sum_{i=1}^n \rho(X_i,\theta)\}\,\mathrm{d}\theta}, \quad \text{if both integrals exist.}$$

Recalling a relation between Pitman and MLE estimators (see Janssen, Jurečková, Veraverbeke [16]) or Bayes' and MLE estimators (cf. Ibragimov, Khasminskii [13], [14]), we expect, because of similarity, corresponding relations between robustified versions of these estimates. Johns [17] has shown that there is a connection between P- and M-estimators; another proof could be found in Hanousek [4].

In Section 2, we shall show, under some regularity conditions, asymptotic equivalence of B- and MPL-estimators. Particularly, if

$$f(x,\theta) = \exp\{-\rho(X_i,\theta)\}$$

then these results lead to the well-known relations of maximum (posterior) likelihood and Bayes' estimators.

2. ASYMPTOTIC BEHAVIOUR OB BAYES-TYPE ESTIMATORS

We consider the following sets of regularity conditions (A), (B) and (C):

(A)

(A1) $\Theta \subset R'$ is an open set.

(A2) $\rho : R^d \times \Theta \to R_+$ is a continuous function. Moreover, $\frac{\partial^2 \rho}{\partial\theta^2}$ exists and for every $\theta \in \Theta$ there exist $\rho > 0$, $\beta \geq 0$ such that for every

$$\xi, \eta \in \Theta : \|\xi - \theta\| < \delta, \; \|\eta - \theta\| < \delta,$$

and for every $x \in R^d$

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

(A3) $\pi : \Theta \to R_+$ is bounded and $\ln \pi$ is well-defined with a continuous derivative $\frac{\partial \ln \pi}{\partial\theta}$.

(A4) The integral $\int_\theta \|\theta\| \exp\{-\rho(x,\theta)\}\,\pi(\theta)\,\mathrm{d}\theta$ exists for every $x \in R^d$.

(B)

(B1) For every $n \in N$ $\int M_n(x)\,\mathrm{d}F(x,\theta_0) < +\infty$, where

$$M_n(x) = \sup_{\substack{\|\theta\| \le n \\ \theta \in \Theta}} \left\| \frac{\partial^2 \rho}{\partial \theta^2}(x,\theta) \right\|.$$

(B2) There exists a point $\theta^* \in R^\ell$ such that $\int \rho(x,\theta^*)\,\mathrm{d}F(x,\theta_0)$ and $\int \frac{\partial \rho}{\partial \theta}(x,\theta^*)\,\mathrm{d}F(x,\theta_0)$ are finite.

(C)

(C1) We assume that the function $h(\theta) = \int \rho(x,\theta)\,\mathrm{d}F(x,\theta_0)$ has a unique absolute minimum at $\theta = \theta_0$, i.e. $\theta_0 \in \operatorname*{argmin}_{\theta \in \Theta} h(\theta)$.

(C2) If $\sup_{\theta \in \Theta} \|\theta\| = +\infty$ then $h(\theta) < \overline{\rho} \overset{\text{df.}}{=} \inf_{K>0} \liminf_{\|\theta\| \to +\infty} \inf_{\|x\| \le K} \rho(x,\theta)$.

(C3) $\frac{\partial^2 h}{\partial \theta^2}(\theta_0)$ is a positive definite matrix.

(C4) $\int \frac{\partial \rho}{\partial \theta}(x,\theta_0) \cdot \left(\frac{\partial \rho}{\partial \theta}(x,\theta_0)\right)^{\mathrm{T}} \mathrm{d}F(x,\theta_0)$ is a real matrix.

Under these regularity conditions the following theorem can be proved:

Theorem 1. Suppose that conditions (A) – (C) are satisfied. Then, for $n \to +\infty$,

$$\sqrt{n}\|M_n^* - \theta_0\| = O_p(1),$$
$$\sqrt{n}\|B_n^* - \theta_0\| = O_p(1)$$

and

$$n\|B_n^* - M_n^*\| = O_p(1).$$

Moreover, if $\frac{\partial^3 \rho}{\partial \theta^3}$ exists and $\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^3 \rho}{\partial \theta^3}\left(X_i, \theta_0 + \frac{1}{\sqrt{n}}\theta\right) \to \frac{\partial^3 \rho}{\partial \theta^3}$, a.s. uniformly for $\|\theta\| \le \delta_0$, then

$$B_n^* = M_n^* + n^{-1} \cdot \frac{A}{B} + o_p(n^{-1}),$$

where

$$A = -\frac{1}{6} \int_{R^\ell} \theta \sum_{\ell_1,\ell_2,\ell_3=1}^{\ell} \theta_{\ell_1}\,\theta_{\ell_2}\,\theta_{\ell_3} \frac{\partial^3 h}{\partial \theta_{\ell_1} \partial \theta_{\ell_2} \partial \theta_{\ell_3}}(\theta_0) \exp\left(-\frac{1}{2}\theta^{\mathrm{T}} \frac{\partial^2 h}{\partial \theta^2}(\theta_0)\,\theta\right) \mathrm{d}\theta$$

and

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

P r o o f of Theorem 1 is rather tedious and technical. We give only main steps of the proofs (details can be found in Hanousek, Lachout [8]; technical reports are available upon request of the author). □

Lemma 1. Let conditions (A1), (A2), (B1) and (B2) be fulfilled. Then for every $\theta \in \Theta$ the integrals

$$h(\theta = \int \rho(x,\theta)\,\mathrm{d}F(x)$$
$$\frac{\partial h}{\partial \theta}(\theta) = \int \frac{\partial \rho}{\partial \theta}(x,\theta)\,\mathrm{d}F(x),$$
$$\frac{\partial^2 h}{\partial \theta^2}(\theta) = \int \frac{\partial^2 \rho}{\partial \theta^2}(x,\theta)\,\mathrm{d}F(x)$$

are finite.

Lemma 2. Let conditions (A1), (A2), (B) be fulfilled. Then, there exists a set $A \in \mathcal{A}$, $P(A) = 1$ such that for every $\omega \in A$ and $\theta \in \Theta$

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

and

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

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 $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0$.

Denote by α the smallest eigenvalue of $\frac{\partial^2 h}{\partial \theta^2}$ and define $Q = \sup_{\|\eta-\theta_0\|} \ln\left(\frac{\pi(\eta)}{\pi(\theta_0)}\right)$. Notice that $\alpha > 0$ by (C3) and π is continuous and positive by (A3).

By Lemmas 1–3 and Corollary 1 we get

$$\begin{aligned} &P\left(\sqrt{n}\|M_n^* - \theta_0\| > M\right) \\ &\leq P\left(\|M_n^* - \theta_0\| > \delta\right) \\ &+ P\left(\frac{\alpha}{4}n\|M_n^*-\theta_0\|^2-Q<-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sqrt{n}(M_n^*-\theta_0)\frac{\partial \rho}{\partial \theta}(X_i,\theta_0),\ \delta \geq \|M_n^*-\theta_0\| > \frac{1}{\sqrt{n}}H\right) \end{aligned}$$

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

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

$$\left\| \int_{\|\theta - \theta_0\| > \delta} \theta \exp\left\{ -\sum_{i=1}^n \rho(X_i, \theta) + \ln \pi(\theta) \right\} \mathrm{d}\theta \right\| \leq B_n(\Delta) \cdot O_p(1)$$
$$\left\| \int_{\|\theta - \theta_0\| > \delta} \exp\left\{ -\sum_{i=1}^n \rho(X_i, \theta) + \ln \pi(\theta) \right\} \mathrm{d}\theta \right\| \leq B_n(\Delta) \cdot O_p(1)$$
$$B_n(\Delta) = \exp(-n\Delta) \cdot \exp\left(-\sum_{i=1}^n \rho(X_i, M_n^*) + \ln \pi(M_n^*) \right).$$

Lemma 5. Let the groups of assumptions (A), (B), (C) be fulfilled. Then

$$\int \exp\left(-\sum_{i=1}^n \rho(X_i, \theta) + \ln \pi(\theta) \right) \mathrm{d}\theta =$$
$$= \left(\frac{1}{\sqrt{n}}\right)^{\ell} \exp\left(-\sum_{i=1}^n \rho(X_i, M_n^*) + \ln \pi(M_n^*) \right) \left(\int_{R^\ell} \exp\left(-\frac{1}{2} \theta^{\mathrm{T}} \frac{\partial^2 h}{\partial \theta^2}(\theta_0)\, \theta \right) \mathrm{d}\theta \right) + o_p(1).$$

Lemma 6. Let all the assumptions (A), (B), (C) hold. Then,

$$\int_\theta \theta \exp\left(-\sum_{i=1}^n \rho(X_i, \theta) + \ln \pi(\theta) \right) \mathrm{d}\theta =$$
$$= M_n^* \int_\theta \exp\left(-\sum_{i=1}^n \rho(X_i, \theta) + ln\pi(\theta) \right) \mathrm{d}\theta +$$
$$+ \frac{1}{n} \left(\frac{1}{\sqrt{n}}\right)^{\ell} \exp\left(-\sum_{i=1}^n \rho(X_i, M_n^*) + \ln \pi(M_n^*) \right) \cdot A_n,$$

where

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

If additionally, $\frac{\partial^3 \rho}{\partial \theta^3}$ is finite and

$$\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^3 \rho}{\partial \theta^3}(X_i,\theta_0)+\frac{1}{\sqrt{n}}\theta\right) \xrightarrow{\text{a.s.}} \frac{\partial^2 h}{\partial \theta^3}(\theta_0) \quad \text{uniformly for } \|\theta\| \le \delta_0$$

then $A_n = A + o_p(1)$, where

$$A = -\frac{1}{6}\int_{R^\ell} \theta \sum_{\ell_1,\ell_2,\ell_3=1}^{\ell} \theta_{\ell_1}\,\theta_{\ell_2}\,\theta_{\ell_3}\frac{\partial^3 h}{\partial\theta_{\ell_1}\partial\theta_{\ell_2}\partial\theta_{\ell_3}}(\theta_0)\cdot \exp\left(-\frac{1}{2}\theta^{\mathrm{T}}\frac{\partial^2 h}{\partial \theta^2}(\theta_0)\,\theta\right)\mathrm{d}\theta.$$

By Lemma 5 and Lemma 6, we obtain

$$\begin{aligned} B_n^* &= \frac{\int_\theta \theta \exp\left(-\sum_{i=1}^n \rho(X_i,\theta)\right)\pi(\theta)\,d\theta}{\int_\theta \exp\left(-\sum_{i=1}^n \rho(X_i,\theta)\right)\pi(\theta)\,d\theta} = \\ &= M_n^* + \frac{\frac{1}{n}\left(\frac{1}{\sqrt{n}}\right)^\ell \exp\left(-\sum_{i=1}^n \rho(X_i,M_n^*)+\ln\pi(M_n^*)\right)O_p(1)}{\left(\frac{1}{\sqrt{n}}\right)^\ell \exp\left(-\sum_{i=1}^n \rho(X_i,M_n^*)+\ln\pi(M_n^*)\right)(B+\sigma_p(1))} = \\ &= M_n^* + \frac{1}{n}\frac{O_p(1)}{B+\sigma_p(1)} = M_n^* + \frac{1}{n}O_p(1). \end{aligned}$$

Under the additional assumptions we have

$$B_n^* = M_n^* + \frac{1}{n}\cdot\frac{A}{B} + o_p\left(\frac{1}{n}\right).$$

□

Remarks.

1. Regularity conditions e. g. (A) – (C) are rather strong, they can be weakened, assuming Lipschitz conditions instead of existence of third derivative of ρ etc.

2. Let consider estimator of location, e. g. $\rho(X,\theta) = \eta(X-\theta)$, η is symmetric around zero and the true distribution is symmetric around θ_0, i. e. $F(X,\theta_0) = 1 - F(2\theta_0 - X,\,\theta_0)$, we get $\frac{\partial^3 h}{\partial \theta^3}(\theta_0) = 0$. Consequently, $A = 0$ and $B_n^* = M_n^* + o_p(n^{-1})$.

3. By linear substitution we get for B-estimator

$$B_n^* = \theta_0 + n^{-1/2}\cdot\frac{\int_{-\infty}^{\infty} t\cdot L_n(t)\,\mathrm{d}t}{\int_{-\infty}^{\infty} L_n(t)\,\mathrm{d}t}$$

where

$$L_n(t) = \exp\left(-\sum_{i=1}^{n}[\rho(X_i,\theta_0) - \rho(X_i,\theta_0 + n^{-1/2}\cdot t)]\right)\cdot \pi(\theta_0 + n^{-1/2}\cdot t).$$

We can study asymptotic behavior via Taylor expansion of exponential functions of the integrands; manipulation with Taylor expansion then brings desired equivalence. But the difficulty is, that the error of any expansion (plugged into integrals) has to be integrate out and this causes problems. A possible solution is to prove that the difference between the form with an expansion and that without has the proper order. It is easier to use if we apply the results of asymptotic behaviour of random processes (see Ibragimov, Khasminskii [13, 14], and Inagaki, Ogata [15]). We get

$$T_n^* = \theta_0 + n^{-1/2} \cdot \frac{\int_{-C}^{C} t \cdot L_n(t)\,\mathrm{d}t}{\int_{-C}^{C} L_n(t)\,\mathrm{d}t} \quad \text{where } C \text{ is independent on } n.$$

Since constant C is independent on n, we can apply Taylor expansion. This technique was used by Hanousek [5], Hanousek, Jurečková and Saleh [7] to obtain asymptotic representation of B_n^* as

$$B_n^* = \theta_0 \cdot \gamma^{-1}(\theta_0) \cdot \sum_{i=1}^{n} \psi(X_i, \theta_0) + o_p(n^{-1/2})$$

where

$$\gamma(\theta_0) = E_{\theta_0}\left(\psi'(X_1, \theta_0)\right).$$

One can see that this asymptotic representation does not depend on a prior information. It leads to the problem to study a higher order asymptotics.

4. Some interesting applications of Theorem 1 could be:

   a) Use of the estimator which, in a particular case, is easier to compute (switching between multiple-integration and optimalization problems).

   b) Analogously as for MLE we can show that one and two-step estimators based on any $n^{-1/2}$-consistent estimator will be asymptotically equivalent to these estimators.

   c) Study attractive properties of k-step estimators (easy computation, robustness and bayesian properties).

(Received March 30, 1994.)

REFERENCES

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

[2] G. E. P. Box and C. Tiao: A further look at robustness via Bayes' theorem. Biometrika *49* (1962), 419–432.

[3] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel. Robust statistics: the approach based on influence functions. J. Wiley, New York 1986.

[4] J. Hanousek: Asymptotic relation of M- and P-estimators of location. Comput. Statist. Data Anal. *6* (1988), 277–284.

[5] J. Hanousek: Robust Bayesian type estimators and their asymptotic representation. Statist. Decisions *8* (1990), 61–69.
[6] J. Hanousek, J. Jurečková and A. K. Md. E. Saleh: Robustified Bayes estimation. Multiparameter case. Technical Report, Carleton University, Ottawa 1989.
[7] J. Hanousek, J. Jurečková and A. K. Md. E. Saleh: Finite sample asymptotic representation of the robustified Bayes estimators. Technical Report, Carleton University, Ottawa 1989.
[8] J. Hanousek and P. Lachout: Robust Bayes-type Version of Classical Estimators. Research Report No. 1745, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague 1992.
[9] J. Hanousek and P. Lachout: Robust Bayes-type Version of Classical Estimators. Working Paper No. 24, CERGE, Charles University, Prague 1993.
[10] P. J. Huber: Robust estimation of a location parameter. Ann. Math. Statist. *35*, (1964), 73–101.
[11] P. J. Huber: Robust Statistics. J. Wiley, New York 1981.
[12] P. J. Huber: Finite sample breakdown point of M- and P-estimators. Ann. Statist. *12* (1984), 1, 119–126.
[13] I. A. Ibragimov and R. Z. Khasminskii: Asymptotic behavior of statistical estimators in the smooth case I–II. Theory Probab. Appl. *17* (1972) 445–464; *18* (1973), 76–91.
[14] I. A. Ibragimov and R. Z. Khasminskii: Statistical Estimation. Asymptotic Theory. Springer–Verlag, Berlin 1981.
[15] N. Inagaki and Y. Ogata: The weak convergence of likelihood random fields and its applications. Ann. Inst. Statist. Math. *27* (1975), 391–419.
[16] P. Janssen, J. Jurečková and N. Veraverbeke: Rate of convergence of one- and two-step M-estimators with application to maximum likelihood and Pitman estimators. Ann. Statist. *13* (1985), 3, 1222–1229.
[17] M. O. Johns: Robust Pitman-like estimators. In: Robust Statistics (R. L. Launer and O. N. Wilkinson, eds.), Academic Press, New York 1979.
[18] A. H. Welsh: On M-processes and M-estimation. Ann. Statist. *17* (1989), 337–361.

Dr. Jan Hanousek, CERGE, Charles University, Politických vězňů 7, P. O. Box 882, 111 21 Praha 1. Czech Republic.