

ON THE ASYMPTOTIC BEHAVIOR OF GENERALIZED BAYES ESTIMATES

MATHEMATICS

1970

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-197001.59657>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

UDC 519.25

MATHEMATICS

I. A. IBRAGIMOV, R. Z. KHASHMINSKII

ON THE ASYMPTOTIC BEHAVIOR OF GENERALIZED BAYES ESTIMATES

(Presented by Academician A. N. Kolmogorov on 11 III 1970)

1. Let x_1, \dots, x_n be a sample with independent elements from a population with distribution P_θ , depending on an unknown parameter $\theta \in \Theta$. Suppose that P_θ , for all θ , are distributions in R_n , and that the parameter set $\Theta \subset R_m$. Suppose also that all measures P_θ are absolutely continuous with respect to some measure ν , and put $f(x, \theta) = dP_\theta/d\nu$.

Our aim will be to study the asymptotic behavior of the finite-dimensional distributions of the random field

$$p_n(\theta) = \prod_1^n f(x_i, \theta)\pi(\theta) / \int \prod_1^n f(x_i, \theta)\pi(\theta)\nu(d\theta) \quad (1)$$

and of estimates $\tilde{\theta}_n$ of the parameter θ , having the form

$$\tilde{\theta}_n = \int \theta p_n(\theta)\nu(d\theta) \quad (2)$$

for measurable functions $\pi(\theta)$ in R_m . If $\pi(\theta)$ is the density of a probability distribution in R_m with respect to the measure ν , then the field $p_n(\theta)$ and the estimate $\tilde{\theta}_n$ are naturally interpreted as the Bayesian posterior density and, respectively, the posterior mean of a random variable having the prior distribution with density $\pi(\theta)$. We shall retain the names posterior density and Bayesian estimate for the right-hand sides of (1) and (2), although in our considerations θ is not a random variable. We shall denote by θ_0 the true (unknown to the observer) value of the parameter θ .

We shall call a function $\varphi(n)$ a correct normalizing factor (c.n.f.) for (1), (2), if the finite-dimensional distributions of the random field $\eta_n(\theta) = \varphi^{-1}(n)p_n(\theta_0 + \theta/\varphi(n))$ and the distribution of the random variable $\zeta_n = \varphi(n)(\tilde{\theta}_n - \theta_0)$, as $n \rightarrow \infty$, converge to proper limiting distributions.

In the present note we are concerned with finding c.n.f.'s and the corresponding limiting distributions under various restrictions imposed on the family P_θ . To make the exposition less cumbersome, we shall restrict ourselves to the case where the random variables x_i and the parameter θ are one-dimensional, and $\nu(dy) = dy$, Lebesgue measure.*

2. Let $f(x, \theta)$ be a measurable function of (x, θ) , and let the "prior density" $\pi(\theta)$ be continuous at the point θ_0 , bounded on compact sets, and

$$\iint |\theta| f(x, \theta) f(x, \theta_0) |\pi(\theta)| dx d\theta < \infty.$$

With respect to the family P_θ , below one of the following conditions will be assumed to hold:

I. 1) $\lim_{|\theta| \rightarrow \infty} \int f(x, \theta) f(x, \theta_0) dx = 0$; 2) $\lim_{\varepsilon \rightarrow 0} \int f^{1-\varepsilon}(x, \theta_0) dx = 1$.

* We owe the formulation of some of the problems considered here to Yu. V. Linnik.

- II. Whatever the real numbers $\theta_1, \dots, \theta_s$, there exist intervals $[0, T_1], \dots, [0, T_s]$ such that, for all $t_j \in [0, T_j]$, as $\xi \downarrow 0$,

$$\int \prod_{j=1}^s \left(\frac{f(x, \theta_0 + \xi \theta_j)}{f(x, \theta_0)} \right)^{t_j} f(x, \theta_0) dx = 1 + \xi^\alpha a(t_1, \dots, t_s; \theta_1, \dots, \theta_s) + o(\xi^\alpha), \quad (3)$$

where the number α does not depend on $\theta_1, \dots, \theta_s, t_1, \dots, t_s$, and the function a is not identically zero.

- III. For all θ_1, θ_2 with $|\theta_1|, |\theta_2| \leq H < \infty$, as $\xi \rightarrow 0$,

$$\left| \int \ln \frac{f(x, \theta_0 + \xi \theta_1)}{f(x, \theta_0 + \xi \theta_2)} f(x, \theta_0 + \xi \theta_2) dx \right| \leq |\xi|^\alpha C_1(|\theta_2 - \theta_1|),$$

$$\left| \int \ln^2 \frac{f(x, \theta_0 + \xi \theta_1)}{f(x, \theta_0 + \xi \theta_2)} f(x, \theta_0 + \xi \theta_2) dx \right| \leq |\xi|^\alpha C_2(|\theta_2 - \theta_1|);$$

here the functions $C_i(h)$, $i = 1, 2$, (depending on H) tend to zero as $h \rightarrow 0$, and in both integrals the function $f(x, \theta_0 + \xi \theta_1)$ is to be replaced by one wherever it is equal to zero.

Denote by $Y(\theta)$ a (separable) random process whose finite-dimensional distributions are determined by the equalities

$$M \exp\{\Sigma t_j Y(\theta_j)\} = \exp\{a(t_1, \dots, t_s; \theta_1, \dots, \theta_s)\},$$

where here and below $M(\cdot) = M_{\theta_0}(\cdot)$, $P(\cdot) = P_{\theta_0}(\cdot)$.

Theorem 1. *Suppose that assumptions I-III are satisfied. Then, for n.h.m. $\varphi(n) = n^{1/\alpha}$, the finite-dimensional distributions of the process $\eta_n(\theta) = \varphi^{-1}(n)p_n(\theta_0 + \theta/\varphi(n))$, $\theta \in (-\infty, +\infty)$, converge to the corresponding distributions of the process*

$$\eta_0(\theta) = \left(\int e^{Y(\theta)} d\theta \right)^{-1} e^{Y(\theta)},$$

and the normalized differences $\xi_n = \varphi(n)(\tilde{\theta}_n - \theta_0)$ converge in distribution to the random variable

$$\int \theta \eta_0(\theta) d\theta.$$

3. If, for example, the left-hand side of (3) can be differentiated twice under the integral sign, we find that $\alpha = 2$,

$$\eta_\xi(\theta) = \frac{1}{\sqrt{2\pi I}} \exp \left\{ - \left(\theta - \frac{\xi}{\sqrt{I}} \right)^2 / 2I \right\}; \quad Y(\theta) = \sqrt{I} \theta \xi - \frac{\theta^2}{2} I, \quad (4)$$

where ξ is a standard normal random variable, and I is the Fisher information. Therefore the normalized posterior density is asymptotically Gaussian, and the difference $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ is also asymptotically normal with parameters $0, I^{-1/2}$ (see also item 5). Under similar assumptions one can also prove that $\tilde{\theta}_n$ is close to the maximum likelihood estimate $\hat{\theta}_n$ in the sense that $\sqrt{n}(\tilde{\theta}_n - \hat{\theta}_n) \rightarrow 0$ in probability, or with probability 1, depending on the restrictions on P_θ . We shall not dwell on these results in greater detail, because closely related theorems are available in the literature (see, for example, ⁽¹⁾, where one can also find the history of the question going back to S. N. Bernstein and R. Mises).

4. Let us turn to discontinuous densities $f(x, \theta)$. We shall assume that the function $f(x, \theta)$ is bounded, continuously differentiable with respect to θ on each of the intervals $(x_k(\theta), x_{k+1}(\theta))$, and has discontinuities of the first kind at the points $x_k(\theta)$, $k = 1, \dots, r$. Suppose that the functions $x_k(\theta)$ are continuously differentiable, and denote

$$p(k) = \lim_{x \downarrow x_k(\theta_0)} f(x, \theta_0), \quad q(k) = \lim_{x \uparrow x_k(\theta_0)} f(x, \theta_0).$$

Consider two independent Poisson random measures concentrated on the set $R = \{1, 2, \dots, r\}$. It will be convenient for us to denote these measures by the common symbol $\nu(\theta, A)$, $A \subseteq R$. This will not cause misunderstandings,

since for one of these measures the parameter $\theta \in (-\infty, 0]$, and for the other $\theta \in [0, \infty)$, $\nu(\theta, A) = 0$. The distributions of the measures $\nu(\theta, A)$ are determined by the equalities: $M\nu(\theta, k) = p(k)x'_k(\theta_0)\theta$, if $\theta x'_k(\theta_0) > 0$, and $M\nu(\theta, k) = -q(k)x'_k(\theta_0)\theta$, if $\theta x'_k(\theta_0) < 0$. Denote by R^* the subset of points k in R for which $p(k)q(k) = 0$, and put

$$\tau_+ = \inf_{\theta > 0} \{\theta : \nu(\theta, R^*) > 0\}, \quad \tau_- = \sup_{\theta < 0} \{\theta : \nu(\theta, R^*) > 0\}$$

(if R^* is empty, we shall take $\tau_+ = \infty$, $\tau_- = -\infty$).

Theorem 2. Suppose that, in addition to the assumptions of item 4, the assumptions of item 2 are satisfied and $|f'_\theta(x, \theta)| \leq q(x)$, where $\int q(x) dx < \infty$. Then the conclusion of Theorem 1 is valid with $r(n) = n$ and with limiting process

$$Y(\theta) = \theta \sum_{k=1}^r (p(k) - q(k))x'_k(\theta_0) + \int_R \text{sign}(x'_z(\theta_0)\theta) \ln \frac{q(z)}{p(z)} \nu(\theta, dz)$$

$$\text{for } \theta \in (\tau_-, \tau_+), \quad Y(\theta) = -\infty \quad \text{for } \theta \notin [\tau_-, \tau_+).$$

Moreover, the limiting distribution for $n(\bar{\theta} - \theta_0)$ can be written in the form

$$\left(\int_{\tau_-}^{\tau_+} \exp Y(\theta) d\theta \right)^{-1} \int_{\tau_-}^{\tau_+} \theta \exp Y(\theta) d\theta.$$

The conditions of this theorem are satisfied, for example, by the cases considered in (2), when: 1) $f(x, \theta) = \beta$ for $0 \leq x \leq \theta$, $f(x, \theta) = \gamma$ for $\theta < x \leq 1$, and $f(x, \theta) = 0$ for $x \notin [0, 1]$; 2) $f(x, \theta)$ has one point of discontinuity.

5. In this section we shall assume that $f(x, \theta) = f(x - \theta)$, so that the parameter to be estimated is a shift parameter. We also put $\pi(\theta) \equiv 1$. The estimator (2) is now the Pitman estimator for the shift parameter (see (3)); without loss of generality one may put $\theta_0 = 0$. Suppose, moreover, that

$$\int |x|f(x) dx < \infty. \tag{5}$$

Theorem 3. Suppose condition (5) is satisfied, the function $f(x)$ is absolutely continuous, and the Fisher information

$$I = \int \frac{|f'(x)|^2}{f(x)} dx$$

is finite. Then the finite-dimensional distributions of the process

$$\eta_n(\theta) = n^{-1/2} p_n(\theta_0 + \theta/\sqrt{n})$$

converge to the corresponding distributions of the process $\eta_0(\theta)$ defined by formula (4), and the difference $\sqrt{n}(\bar{\theta}_n - \theta_0)$ is asymptotically normal with parameters $0, I^{-1/2}$.

Theorem 4. Suppose the conditions of Theorem 3 are satisfied and $M|x_i|^p < \infty$ for some $p > 1$. Then

$$\lim_{n \rightarrow \infty} n^{p/2} M|\bar{\theta}_n - \theta_0|^p = I^{-p/2} \pi^{-1/2} 2^{(p+1)/2} \Gamma((p+1)/2). \quad (6)$$

By imposing further restrictions on the moments of the random variables $l^{(k)}(x_i)$, where $l(x) = \ln f(x)$, one can obtain further terms of the asymptotic expansion of the left-hand side of (6). For example, suppose the conditions of Theorem 3 are satisfied and, in addition, the moments are finite:

$$M|l^{(k)}(x_i)|^s, \quad k = 1, \dots, 5, \quad s \leq 20; \quad M \left(\max_{|\theta| \leq \varepsilon} |l^{(5)}(x_i - \theta)| \right)^s, \quad s \leq 20.$$

Then

$$M(\bar{\theta}_n - \theta_0)^2 = \frac{1}{In} + \frac{C_1}{n^2} + o(n^{-2}).$$

Similarly, if $\tilde{\theta}_n$ is the general estimate (2), but with $\pi(\theta) \equiv 1$, $\nu(dy) \equiv dy$, and the moments $M|l_\theta^{(k)}(x_i, \theta_0)|^s$, $k = 1, \dots, 5$; $s \leq 20$, are finite, as is

$$M \max_{|\theta| \leq \varepsilon} |l_\theta^{(5)}(x_0, \theta_0 - \theta)|^s, \quad s \leq 20,$$

where $l(x, \theta) = \ln f(x, \theta)$, and if $\max M\tilde{\theta}_n^2 < \infty$, then

$$M(\tilde{\theta}_n - \theta_0) = \frac{c(\theta_0)}{n} + o\left(\frac{1}{n}\right); \quad c(\theta_0) = \frac{1}{I^2} [Ml_\theta'''(x_i, \theta_0) + Ml'(x_i, \theta_0)l''(x_i, \theta_0)],$$

$$M(\tilde{\theta}_n - \theta_0)^2 = \frac{1}{In} + \frac{c_1(\theta_0)}{n^2} + o\left(\frac{1}{n^2}\right), \quad I = M|l'(x_i, \theta_0)|^2.$$

6. In order to determine what happens when the regularity conditions are violated, consider the following class of densities $f(x, \theta) = f(x - \theta)$. The function $f(x)$ is twice continuously differentiable in each of the intervals (x_k, x_{k+1}) , $k = 0, \dots, r$, $x_0 = -\infty$, $x_{r+1} = \infty$, and, as $x \rightarrow x_k$, $k = 1, \dots, r$, its derivative $f'(x)$ satisfies the conditions:

$$f'(x) \sim a_k^- \alpha_k^- |x - x_k|^{\alpha_k^- - 1}, \quad x \rightarrow x_k - 0; \quad f'(x) \sim a_k^+ \alpha_k^+ |x - x_k|^{\alpha_k^+ - 1},$$

$$x \rightarrow x_k + 0,$$

where $-1 < \alpha_k^\pm < 1$, $\alpha_k^\pm \neq 0$. It will be convenient for us also to include here the case $\alpha_k^\pm = 0$, assuming that the equality $\alpha_k^+ = 0$ ($\alpha_k^- = 0$) means that $f(x_k + 0) = a_k^+$ ($f(x_k - 0) = a_k^-$).

Introduce the notation

$$\gamma = \min_k(\alpha_k^+, \alpha_k^-), \quad A^\pm = \{k : \alpha_k^\pm = \gamma\}, \quad C = \{k : f(x_k) = 0\},$$

$$\gamma_1 = \min\{2 \min_{k \in C} \alpha_k^\pm, \min_{k \in C} \alpha_k^\pm\}, \quad A_1^\pm = \{k : k \in C, 2\alpha_k^\pm = \gamma_1\},$$

$$A_2^\pm = \{k : k \in C, \alpha_k^\pm = \gamma_1\}.$$

The following assertions are valid:

- 1) If $\gamma \leq 0$, then the n.n.m. $\varphi(n) = n^{1/(1+\gamma)}$; in this case the only parameters determining the limiting distribution for $\eta_n(\theta)$ and ξ_n are the numbers γ and α_k^\pm for $k \in A^\pm$.
- 2) If $\gamma > 0$, then the n.n.m. $\varphi(n) = n^{1/(1+\gamma_1)}$, and the only parameters determining the limiting distributions for $\eta_n(\theta)$, ξ_n are the numbers γ_1 and α_k^\pm for $k \in A_1^\pm \cup A_2^\pm$.

These assertions make it possible to order the singularities of the function $f(x)$ according to their “informativeness,” if one restricts oneself to the class of singularities of power type. They can also be extended to the case of singularities of the form $|x - x_k|^\alpha l(x - x_k)$, where $l(x)$ is a slowly varying function in the sense of Karamata (already in (3) the factor ξ^α may be replaced by $\xi^{\alpha l(\xi)}$). It is interesting to note that slowly varying functions in the n.n.m. may also appear at the “extreme” points of power singularities of the function $f(x)$. For example, for a function $f(x)$ having only singularities of the form $a_k |x - x_k| + c_k |x - x_k|^{1/2}$, $c_k > 0$, the n.n.m. is $\varphi(n) = \sqrt{n \ln n}$.

We express our deep gratitude to A. N. Kolmogorov and Yu. V. Linnik for their attention to the work and for useful discussion.

Leningrad State University
named after A. A. Zhdanov

Institute for Problems of Information Transmission
Academy of Sciences of the USSR
Moscow

Received
1 III 1970

REFERENCES

1. L. Le Cam, Univ. Calif. Public. Stat., 1, 277 (1953); Collected Transl., Mathematics, 4, 2, 69 (1960).
2. H. Chernoff, H. Rubin, Proc. III Berkeley Symp., 1, Univ. Calif. Press, 1956, p. 19.
3. A. M. Kagan, Tr. Mat. Inst. im. V. A. Steklova AN SSSR, 104, 19 (1968).

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.