ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XVI, 4 (1980)

J. KOGUT (Wrocław)

ON THE CONVERGENCE WITH PROBABILITY ONE FOR A SEQUENCE OF EMPIRICAL BAYES ESTIMATORS

0. Some sequences of empirical Bayes estimators for various problems of empirical Bayes estimation were proposed in papers [2] and [3]. In the construction of these sequences, Bayes estimators were used. Moreover, in [2] it was proved that the sequence of empirical Bayes estimators is asymptotically optimal, i.e., that the expected risks associated with this sequence of estimators are converging to a Bayes risk.

The aim of this paper is to obtain a sequence of empirical Bayes estimators uniformly converging with probability one to a Bayes estimator. In Section 1 we prove a theorem on convergence with probability one for certain sequences of estimators. In Section 2 we use this theorem in a problem of empirical Bayes estimation from [2].

1. Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with density function $f(x) > 0$. For every $j = 0, 1, \ldots$ let $f^{(j)}(x)$ denote the j-th derivative of $f(x)$ and let $f_n^{(j)}(x)$ be an estimator of $f^{(j)}(x)$ based on $X_1, X_2, \ldots, X_n$.

Let

$$d(x) = \frac{\sum_{j=0}^{m} w_j(x) f^{(j)}(x)}{f(x)}, \tag{1}$$

where $w_j(x)$ $(j = 0, 1, \ldots, m)$ are known real functions.

Consider the sequence $\{d_n(x)\}$ of estimators

$$d_n(x) = \frac{\sum_{j=0}^{m} w_j(x) f_n^{(j)}(x)}{f_n^*(x)}, \tag{2}$$

where

$$f_n^*(x) = \begin{cases} f_n(x) & \text{if } f_n(x) > \delta_n, \\ \delta_n & \text{if } f_n(x) \leqslant \delta_n, \end{cases} \tag{3}$$

with $\{\delta_n\}$ being a sequence of positive numbers such that

$$b_1 n^{-\delta} \leqslant \delta_n \leqslant b_2 n^{-\delta}, \quad 0 < b_1 \leqslant b_2 < \infty, \; \delta > 0. \tag{4}$$

Now, we prove that under some conditions the sequence $\{d_n(x)\}$ is uniformly convergent to $d(x)$ with probability one on a finite interval.

THEOREM. *Let the functions $w_j(x)$ and $f^{(j)}(x)$ $(j = 0, 1, \ldots, m)$ be bounded on a finite interval I and let*

$$\inf_{x\in I} f(x) > 0.$$

Let the sequences $\{f_n^{(j)}(x)\}$ of estimators of $f^{(j)}(x)$ satisfy the conditions

(5) $$\sum_{n=1}^{\infty} \mathrm{P}\{\sup_{x\in I} |f_n^{(j)}(x) - f^{(j)}(x)| > C\delta_n\} < \infty \qquad (j = 0, 1, \ldots, m)$$

for every positive constant C.

Then

(6) $$\mathrm{P}\{\lim_{n\to\infty} \sup_{x\in I} |d_n(x) - d(x)| = 0\} = 1.$$

Proof. It can be seen from equalities (1) and (2) that

(7)

$$d_n(x) - d(x) = \sum_{j=0}^{m} \frac{w_j(x)}{f_n^*(x)} \big(f_n^{(j)}(x) - f^{(j)}(x)\big) + \sum_{j=0}^{m} \frac{w_j(x) f^{(j)}(x)}{f_n^*(x) f(x)} \big(f(x) - f_n^*(x)\big).$$

From the assumptions of the Theorem we deduce that there exists a positive constant A such that for every $j = 0, 1, \ldots, m$

$$\sup_{x\in I} |w_j(x)| \leqslant A \quad \text{ and } \quad \sup_{x\in I} |w_j(x) f^{(j)}(x)/f(x)| \leqslant A.$$

For any arbitrary $\varepsilon > 0$ we put $\varepsilon_n = \varepsilon\delta_n/2(m+1)A$. Suppose that the inequalities

(8) $$\sup_{x\in I} |f_n^{(j)}(x) - f^{(j)}(x)| \leqslant \varepsilon_n \qquad (j = 0, 1, \ldots, m)$$

are satisfied. Since

$$\inf_{x\in I} f(x) > 0 \quad \text{ and } \quad \delta_n \to 0 \text{ as } n \to \infty,$$

it follows from (8) for $j = 0$ that for $x \in I$ and for n sufficiently large the relations $f_n(x) > \delta_n$ and $f_n^*(x) = f_n(x)$ hold. Hence, by (7) and (8) we have

$$\sup_{x\in I} |d_n(x) - d(x)| \leqslant \varepsilon$$

for n sufficiently large. To sum up, for n sufficiently large we obtain the following relation between the random events:

$$\{\sup_{x\in I} |f_n^{(j)}(x) - f^{(j)}(x)| \leqslant \varepsilon_n \text{ for } j = 0, 1, \ldots, m\} \subset \{\sup_{x\in I} |d_n(x) - d(x)| \leqslant \varepsilon\}.$$

Therefore, for n sufficiently large

$$P\{\sup_{x\in I}|d_n(x)-d(x)|>\varepsilon\}\leqslant\sum_{j=0}^{m}P\{\sup_{x\in I}|f_n^{(j)}(x)-f^{(j)}(x)|>\varepsilon_n\}.$$

Since $\varepsilon_n = C\delta_n$, using conditions (5), the definition of convergence with probability one, the first Borel-Cantelli lemma and the last inequality we obtain (6).

Now we give sequences $\{f_n^{(j)}(x)\}$ $(j = 0, 1, \ldots, m)$ of estimators of density functions and their derivatives satisfying conditions (5).

Let $f_n^{(j)}(x)$ be an estimator of $f^{(j)}(x)$ based on $X_1, X_2, \ldots, X_n$, as given in [4], i.e., let

$$f_n^{(j)}(x) = \frac{1}{na_n^{j+1}}\sum_{i=1}^{n}K^{(j)}\left(\frac{x-X_i}{a_n}\right), \tag{9}$$

where $\{a_n\}$ is a sequence of positive numbers converging to zero, and $K(u)$ is a probability density function such that $\int_{-\infty}^{\infty}|u|K(u)du$ is finite and $K^{(s)}(u)$ is a continuous function of bounded variation for $s = 0, 1, \ldots, j$. Schuster proved (see Lemma 2.4 in [4]) the following

LEMMA 1. *Let $f_n^{(j)}(x)$ be an estimator of $f^{(j)}(x)$ given by equality* (9). *Let $f(x)$ and its first $j+1$ derivatives be bounded and let $\{\varepsilon_n\}$ be a sequence of positive numbers such that $a_n = o(\varepsilon_n)$. Then there exist positive constants C_1 and C_2 such that*

$$P\{\sup_{-\infty<x<\infty}|f_n^{(j)}(x)-f^{(j)}(x)|>\varepsilon_n\}\leqslant C_1\exp[-C_2n\varepsilon_n^2a_n^{2j+2}]$$

for n sufficiently large.

Now, using Lemma 1 we obtain

COROLLARY 1. *For every $j = 0, 1, \ldots, m$ let $f_n^{(j)}(x)$ be an estimator of $f^{(j)}(x)$ given by* (9), *where*

$$d_1n^{-1/(2m+4)}\leqslant a_n\leqslant d_2n^{-1/(2m+4)}\qquad(0<d_1\leqslant d_2<\infty)$$

and

$$K(u) = 1/\sqrt{2\pi}\exp[-u^2/2].$$

Let the sequence $\{\delta_n\}$ satisfy (4) *with δ such that $0<\delta<1/(2m+4)$. If $f(x)$ and its first $m+1$ derivatives are bounded, then the sequences $\{f_n^{(j)}(x)\}$ $(j = 0, 1, \ldots, m)$ satisfy conditions* (5) *for any interval $I\subset(-\infty, \infty)$.*

Proof. We can easily verify that a_n and $K(u)$ in Corollary 1 satisfy the conditions concerning $f_n^{(j)}(x)$ given by (9) and $a_n = o(\delta_n)$. Substituting $\varepsilon_n = C\delta_n$, C being a positive constant, we infer from Lemma 1 that conditions (5) hold, since

$$\sum_{n=1}^{\infty} C_1 \exp(-C_3 n \delta_n^2 a_n^{2j+2})$$

is finite for all positive constants C_1, C_3 and $j = 0, 1, \ldots, m$.

2. Now we consider the problem of empirical Bayes estimation.

Assume that we observe a random variable X whose density function $f(x \mid \theta)$ depends on an unknown parameter $\theta \in \Omega$, with Ω being the parameter space. Let $\lambda(\theta)$ be a real function of θ and let $d(x)$ stand for a decision when $X = x$ is observed. We wish to estimate $\lambda(\theta)$ with respect to the loss function $L(d(x), \lambda(\theta))$.

In the Bayes framework it is assumed that θ has a distribution function $G(\theta)$ known a priori and we use the Bayes estimator $d_G(x)$ minimizing the expected risk

$$r(d, G) = \int_R \int_\Omega L(d(x), \lambda(\theta)) f(x \mid \theta) dG(\theta) dx.$$

In the empirical Bayes framework we suppose that the decision problem just described occurs repeatedly and independently with the same unknown $G(\theta)$. Thus we make the following assumptions: Let (X_1, Θ_1), $(X_2, \Theta_2), \ldots, (X_n, \Theta_n), \ldots$ be a sequence of independent random vectors, Θ_n having a common a priori distribution $G(\theta)$, and the conditional density function of X_n, given $\Theta_n = \theta_n$, being $f(x \mid \theta_n)$ which belongs to the family of density functions $\{f(x \mid \theta)\colon \theta \in \Omega\}$. Let the values of $\theta_1, \theta_2, \ldots$ and the distribution function $G(\theta)$ remain unknown. We know only the values $x_1, x_2, \ldots$ of the random variables $X_1, X_2, \ldots$, and the form of the family $\{f(x \mid \theta)\colon \theta \in \Omega\}$. On the base of known observations $x_1, x_2, \ldots, x_n$; $x_{n+1} = x$ we construct the empirical Bayes estimator $d_n(x) = d_n(x_1, x_2, \ldots, x_n; x)$ for the unknown value of the function $\lambda(\theta_{n+1})$ provided a loss function $L(d_n(x), \lambda(\theta_{n+1}))$ is given. Therefore, we have

Remark 1. The random variables $X_1, X_2, \ldots$ are independent and have a common marginal density function

$$f_G(x) = \int_\Omega f(x \mid \theta)\, dG(\theta).$$

Suppose that the Bayes estimator $d_G(x)$ of $\lambda(\theta)$ can be written in the form

$$d_G(x) = \frac{\sum_{j=0}^{m} w_j(x) f_G^{(j)}(x)}{f_G(x)},$$

where $w_j(x)$ are known real functions. Then, using Remark 1 and suitable estimators $f_n^{(j)}(x)$ $(j = 0, 1, \ldots, m)$ for $f_G^{(j)}(x)$ we estimate $d_G(x)$ by $d_n(x)$ from equality (2). If the assumptions of the Theorem from Section 1 are satisfied, then the sequence of empirical Bayes estimators $\{d_n(x)\}$ is uniformly convergent to $d_G(x)$ with probability one on a finite interval. This fact is proved in the sequel, where the problem (see [2]) of empirical Bayes estimation of $\lambda(\theta) = \theta$ with a squared loss function for the family of exponential densities is considered.

Let $\{f(x|\theta): \theta \in \Omega\}$ be a family of density functions given by

$$f(x|\theta) = \begin{cases} e^{-\theta x}\beta(\theta)h(x), & x > a, \\ 0, & x \leqslant a, \end{cases} \tag{10}$$

where a may be finite or $a = -\infty$, $h(x) > 0$ for $x > a$, $\theta \in \Omega$, with Ω being any interval of the real line.

For the squared loss function the Bayes estimator of $\lambda(\theta) = \theta$ (see [2]) is given by

$$d_G(x) = \frac{\int_\Omega \theta f(x|\theta)\,dG(\theta)}{f_G(x)}. \tag{11}$$

LEMMA 2 (see Lemma 2 in [1]). *Let $f(x|\theta)$ be given by* (10) *and let*

$$f_G(x) = \int_\Omega f(x|\theta)\,dG(\theta),$$

where $G(\theta)$ is any distribution function. Then the existence and continuity of $h^{(j)}(x)$ for $x > a$ imply the existence and continuity of $f_G^{(j)}(x)$ for $x > a$. Moreover, for $x > a$

$$f_G^{(1)}(x) = \frac{h^{(1)}(x)}{h(x)} f_G(x) - \int_\Omega \theta f(x|\theta)\,dG(\theta). \tag{12}$$

Therefore, if $h^{(2)}(x)$ exists and is continuous for all $x > a$, then:
1° by equalities (11) and (12) for $x > a$ we have

$$d_G(x) = \frac{[h^{(1)}(x)/h(x)]f_G(x) - f_G^{(1)}(x)}{f_G(x)},$$

thus $d_G(x)$ is of the form (1), where $m = 1$, $w_0(x) = h^{(1)}(x)/h(x)$ and $w_1(x) = -1$;

2° the functions $w_0(x), w_1(x)$ (of the form given above) and $f_G(x)$, $f_G^{(1)}(x)$, $f_G^{(2)}(x)$ (from Lemma 2) are continuous for $x > a$, and so they are bounded for any finite interval $I_1 \subset (a, \infty)$.

To obtain

$$\inf_{x\in I_1} f_G(x) > 0$$

it suffices to suppose that

$$\inf_{x\in I_1}\inf_{\theta\in\Omega} f(x|\theta) > 0.$$

Let

$$d_n(x) = \frac{[h^{(1)}(x)/h(x)]f_n(x) - f_n^{(1)}(x)}{f_n^*(x)} \quad \text{for } x > a,$$

where the estimators $f_n(x)$, $f_n^{(1)}(x)$ and $f_n^*(x)$ are given by (9) and (3), respectively, with

$$K(u) = 1/\sqrt{2\pi}\exp[-u^2/2],$$

$$d_1 n^{-1/6} \leqslant a_n \leqslant d_2 n^{-1/6} \quad (0 < d_1 \leqslant d_2 < \infty), \quad 0 < \delta < 1/6.$$

By Corollary 1 the sequences $\{f_n^{(j)}(x)\}$ $(j = 0, 1)$ of these estimators satisfy conditions (5) for any finite interval $I_1 \subset (a, \infty)$.

Finally, since all the assumptions of the Theorem from Section 1 of this paper are valid, we obtain

$$\mathrm{P}\{\lim_{n\to\infty}\sup_{x\in I_1} |d_n(x) - d_G(x)| = 0\} = 1.$$

The same result we can show for Case II from paper [1] by a quite similar detailed analysis.

References

[1] M. V. Jr. Johns and J. Van Ryzin, *Convergence rates in empirical Bayes two-action problems, II. Continuous case*, Ann. Math. Statist. 43 (1972), p. 934-947.

[2] P. E. Lin, *Rates of convergence in empirical Bayes estimation problems: Continuous case*, Ann. Statist. 3 (1975), p. 155-164.

[3] J. R. Rutherford and R. G. Kruchkoff, *Some empirical Bayes techniques in point estimation*, Biometrika 56 (1969), p. 133-137.

[4] E. F. Schuster, *Estimation of a probability density function and its derivatives*, Ann. Math. Statist. 40 (1969), p. 1187-1196.

INSTITUTE OF MATHEMATICS
TECHNICAL UNIVERSITY
50-370 WROCŁAW

Received on 4. 3. 1977

J. KOGUT (Wrocław)

O ZBIEŻNOŚCI Z PRAWDOPODOBIEŃSTWEM JEDEN DLA CIĄGU EMPIRYCZNYCH ESTYMATORÓW BAYESOWSKICH

STRESZCZENIE

W pracy rozważono zagadnienie empirycznej bayesowskiej estymacji. Korzystając z odpowiednich estymatorów funkcji gęstości i jej pochodnych zaproponowano ciąg empirycznych bayesowskich estymatorów. Dowiedziono, że ciąg ten jest jednostajnie zbieżny z prawdopodobieństwem jeden do bayesowskiego estymatora. Podano również przykład takiego ciągu estymatorów.