The Annals of Statistics
1988, Vol. 16, No. 3, 1335–1341

CONVERGENCE RATES FOR EMPIRICAL BAYES ESTIMATION IN THE UNIFORM $U(0,\theta)$ DISTRIBUTION

BY YOSHIKO NOGAMI

University of Tsukuba

Let $\{(X_i, \theta_i)\}$ be a sequence of independent random vectors where X_i has a uniform density $U(0, \theta_i)$ for $0 < \theta_i < m$ $(< \infty)$ and the unobservable θ_i are i.i.d. G in some class $\mathscr{G}$ of prior distributions. In the $(n + 1)$st problem we estimate θ_{n+1} by $t_n(X_1, \ldots, X_n, X_{n+1}) \doteq t_n(\mathbf{X})$, incurring the risk $R_n \doteq \mathbf{E}(t_n(\mathbf{X}) - \theta_{n+1})^2$, where $\mathbf{E}$ denotes expectation with respect to all random variables $\{(X_i, \theta_i)\}_{i=1}^{n+1}$. Let R be the infimum Bayes risk with respect to G.

In this paper the author exhibits empirical Bayes estimators with a convergence rate $O(n^{-1/2})$ of $R_n - R$ and shows that there is a sequence of empirical Bayes estimators for which $R_n - R$ has a lower bound of the same order $n^{-1/2}$.

1. Introduction. Since Robbins (1955, 1964), empirical Bayes (EB) problems have been developed in great detail in the literature; for examples, see Johns and Van Ryzin (1971, 1972), Van Ryzin and Susarla (1977) and a recent paper by Robbins (1983). Hannan and Macky (1971), Singh (1974, 1976) and many other authors have discussed (rates of) risk convergence under exponential families of distributions. For nonexponential families of distributions Susarla and O'Bryan (1979) have discussed EB interval estimation for the parameter θ of a uniform distribution $U(0, \theta)$ and Fox (1970, 1978) has considered EB squared-error loss estimation problems without rates.

In this paper the underlying distribution is $U(0, \theta)$ for $0 < \theta < m$ $(< \infty)$, where θ is distributed according to a prior G in some class $\mathscr{G}$ of distributions and the author exhibits EB estimators with a rate $O(n^{-1/2})$ of risk convergence and shows that there is a sequence of EB estimators for which a lower bound of the risk convergence has the same order $n^{-1/2}$. Independently of the author's work, Wei (1983) has established EB estimators for θ $[\in (0, \infty)]$ using kernel functions [see Parzen (1962)] with a rate near $O(n^{-1/2})$ under the assumption of infinite differentiability of the marginal pdf of X and with a rate near $O(n^{-1})$ under further strong assumptions on the marginal distribution of X. The rates in this paper are obtained without differentiability of the marginal pdf of X. In Section 5 one example of prior distributions is given which does not satisfy Wei's (1983, 1985) assumptions, but satisfies the assumptions in this paper.

2. The empirical Bayes problems. Let X be a random variable distributed according to cdf F_θ given θ. The θ_i are i.i.d. random variables distributed according to the unknown prior distribution G. Let $X_1, \ldots, X_n$ be n i.i.d. past observations with each X_i distributed according to the marginal cdf $K(x) =$

Received November 1984; revised December 1987.

AMS 1980 *subject classifications*. Primary 62C12; secondary 62F10, 62C25.

Key words and phrases. Empirical Bayes estimation, squared error loss, convergence rates, nonexponential family.

$\int F_\theta(x)\, dG(\theta)$. Hereafter we let X denote the $(n+1)$st observation X_{n+1} distributed according to $F_{\theta_{n+1}}$. Let $\doteq$ indicate a defining property. The EB estimation problem is to estimate $\theta \doteq \theta_{n+1}$ by using all $n+1$ observations $\mathbf{X} = (X_1, \dots, X_n, X)$. Let $\mathbf{E}$ denote expectation wrt the product measure on the space of $(X_1, \dots, X_n, (X, \theta))$, resulting from K^n and the joint distribution of (X, θ). With $X = x$, let $\phi_G(x)$ denote the Bayes estimator for the prior G given by

$$\phi_G(x) = \int \theta f_\theta(x)\, dG(\theta) \Big/ \int f_\theta(x)\, dG(\theta), \tag{2.1}$$

where $f_\theta(x)$ is the pdf of X, conditionally on θ. The risk of an EB estimator t_n for θ is

$$R_n \doteq R(t_n, G) \doteq \mathbf{E}\left((t_n(\mathbf{X}) - \theta)^2\right)$$

and the Bayes envelope is

$$R \doteq R(\phi_G, G) \doteq \inf_\phi R(\phi, G).$$

When R_n and R are both finite,

$$(0 \le)\ R_n - R = \mathbf{E}(\phi_G(X) - t_n(\mathbf{X}))^2. \tag{2.2}$$

We call a[illegible]B estimator asymptotically optimal (a.o.) when $R_n - R \to 0$ as $n \to \infty$. [illegible]ll find convergence rates for (2.2).

We use [illegible]e following notational conventions. $[A]$ denotes the indicator function of the event A. Let $\wedge$ and $\vee$ denote infimum and supremum. Let $\mathbf{E}_x$ and E denote expectations wrt the conditional product measure on the space of $(X_1, \dots, X_n, (\theta|x))$ given $X = x$ and the marginal probability measure of X, respectively. We denote $\mathbf{E}_x(Y - \mathbf{E}_x(Y))^2$ by $\mathrm{Var}_x Y$. Let $\to_d$ denote convergence in distribution and $\to_p$ denote convergence in probability.

3. An upper bound for $R_n - R$. Let m be a positive finite number and suppose that the support of G is included in the interval $(0, m)$. Let $f_\theta(x) = \theta^{-1}[0 < x < \theta]$ for $\theta \in (0, m)$. We shall exhibit a.o. estimators with an upper bound $O(n^{-1/2})$ for (2.2).

Let $k(x)$ be the marginal pdf of X, which is of the form, for $x > 0$,

$$k(x) = \int f_\theta(x)\, dG(\theta) = \int_x^m \theta^{-1}\, dG(\theta), \tag{3.1}$$

and assume $0 < k(0) < \infty$. Also let the prior distribution G of θ satisfy

$$E\left(\frac{K(X)(1 - K(X))}{k^2(X)}\right) = M\ (< \infty) \tag{3.2}$$

and define by $\mathscr{G}$ the class of priors satisfying the preceding assumptions.

Fox (1978) observed that $K(x) = xk(x) + G(x)$ because $F_\theta(x) = \theta^{-1}x[0 < x < \theta] + [x \ge \theta]$. Hence, from (2.1) we have the following Bayes estimator:

When $k(x) > 0$,

$$\phi_G(x) = (1 - G(x))/k(x) = x + \psi(x), \tag{3.3}$$

where

$$\psi(x) = (1 - K(x))/k(x).$$

Let h be a positive number depending on n such that $0 < h < 1$ and $h \to 0$ as $n \to \infty$. We also let $K_n(y) = n^{-1}\Sigma_{j=1}^{n}[X_j \le y]$ and

$$\begin{aligned} k_n(y) &= h^{-1}(K_n(y + h) - K_n(y)) \\ &= (nh)^{-1} \sum_{j=1}^{n} \left[y < X_j \le y + h\right]. \end{aligned}$$

If we define $G_n(x) \doteq K_n(x) - xk_n(x)$, then we get an estimate for $\phi_G(x)$:

$$\phi_n(x) = 0 \vee \{(1 - G_n(x))/k_n(x)\} \wedge m = 0 \vee (x + \psi_n(x)) \wedge m, \tag{3.4}$$

where

$$\psi_n(x) = (1 - K_n(x))/k_n(x).$$

Since we shall use Lemma A.2 of Singh (1974) to obtain an upper bound for $R_n - R$, we restate it here.

LEMMA 3.1. *Let y, z and L be in $(-\infty, \infty)$ with $z \neq 0$ and $L > 0$. If Y and Z are two real random variables, then for every $\gamma > 0$,*

$$E\left(\left|\frac{y}{z} - \frac{Y}{Z}\right| \wedge L\right)^{\gamma}$$

$$\le 2^{\gamma + (\gamma - 1)^+} |z|^{-\gamma} \left\{E|y - Y|^{\gamma} + \left(\left|\frac{y}{z}\right|^{\gamma} + 2^{-(\gamma - 1)^+} L^{\gamma}\right) E|z - Z|^{\gamma}\right\},$$

where E here means the expectation wrt *the joint distribution of (Y, Z) and $a^+ = a$ if $a > 0$; $= 0$ if $a \le 0$.*

In (3.3) and (3.4), let $\phi_G(x) = v/w$ and $\phi_n(x) = 0 \vee (V/W) \wedge m$. In view of (2.2) with t_n replaced by ϕ_n and by applying Lemma 3.1, we find

$$\begin{aligned} (0 \le)\ R_n - R &\le E\{\mathbf{E}_X(|\psi(X) - \psi_n(X)| \wedge m)^2\} \\ &\le E\left[8k^{-2}(X)\{\mathbf{E}_X(K(X) - K_n(X))^2 \right. \\ &\qquad \left. + (3m^2/2)\mathbf{E}_X(k(X) - k_n(X))^2\}\right]. \end{aligned}$$

Since

$$\mathbf{E}_x(K(x) - K_n(x))^2 = n^{-1}K(x)(1 - K(x))$$

and

$$\mathbf{E}_x(k(x) - k_n(x))^2 = \mathrm{Var}_x(k_n(x)) + (\mathbf{E}_x(k_n(x)) - k(x))^2,$$

the preceding inequality reduces to

$$(0 \le)\ R_n - R \le 8n^{-1}E\left(\frac{K(X)(1-K(X))}{k^2(X)}\right) + 12m^2\Big\{E\big(\mathrm{Var}_X(k_n(X))/k^2(X)\big) + E\Big(\big(\mathbf{E}_X(k_n(X)) - k(X)\big)^2/k^2(X)\Big)\Big\}. \tag{3.5}$$

But, since k is a decreasing function, we have

$$\mathrm{Var}_x(k_n(x)) \le (nh)^{-1}k(x). \tag{3.6}$$

Also, since

$$(0 \le)\ (k(x) - \mathbf{E}_x(k_n(x)))/k(x) \le 1,$$

we can easily verify that

$$E\Big\{\big(k(X) - \mathbf{E}_X(k_n(X))\big)^2/k^2(X)\Big\} \le hk(0). \tag{3.7}$$

(3.5)–(3.7) and (3.2) give

THEOREM 1. *For any prior distribution $G \in \mathscr{G}$, we have*

$$(0 <)\ R_n - R \le 8Mn^{-1} + 12m^2\big\{m(nh)^{-1} + k(0)h\big\}.$$

From Theorem 1 we obtain that, with a choice of $h = n^{-1/2}$ and for some positive constant c_0,

$$\sup_{G\in\mathscr{G}} (R_n - R) \le c_0 n^{-1/2}. \tag{3.8}$$

4. A lower bound for $R_n - R$. Throughout this section, we assume that G is the degenerate distribution at $\theta \equiv 1$. Defining $0/0$ as 0 we have $\phi_G(x) = [0 < x < 1]$. For sufficiently large n, let δ be some positive number such that $1 > \delta > h$. Letting

$$B = [(1 - K_n(x))/k_n(x) \le 1]$$

and

$$\zeta_n(x) = 1 - x - \{(1 - K_n(x))/k_n(x)\}$$

we obtain from (2.2) that

$$R_n - R \ge E\big(\mathbf{E}_X(\zeta_n^2(x)B)[0 < x < 1 - \delta]\big). \tag{4.1}$$

Let $u = \sum_{j=1}^n [X_j \le x]$ and $v = \sum_{j=1}^n [x < X_j \le x + h]$. Then, $\mathbf{E}_x u = nx$, $\mathrm{Var}_x(u) = nx(1-x)$, $\mathbf{E}_x v = nh$ and $\mathrm{Var}_x(v) = nh(1-h)$. Letting

$$Y = (u - nx)/\sqrt{nx(1-x)} \quad \text{and} \quad Z = (v - nh)/\sqrt{nh(1-h)},$$

we obtain

$$\sqrt{nh}\,\zeta_n(x) = \frac{(1-x)\sqrt{1-h}\,Z + h^{1/2}\sqrt{x(1-x)}\,Y}{(nh)^{-1/2}\sqrt{1-h}\,Z + 1}.$$

To get a lower bound for $R_n - R$ we shall use (4.1) and the fact that for fixed x, $\sqrt{nh}\,\zeta_n(x)B \to_d N(0,(1-x)^2)$ as $nh \to \infty$. Here, $N(c, d)$ denotes the normal distribution with mean c and variance d. We then apply a convergence theorem [cf. Loéve (1963), 11.4, A(i)]:

$$\text{(4.2)} \qquad \text{If } U_n \to_d U, \text{ then } \liminf EU_n^2 \ge EU^2.$$

Lemma 4.1 will allow us to prove the preceding convergence in distribution needed for the proof of Theorem 2. Let A^c denote the complement of a set A.

LEMMA 4.1. *If h is a function of n such that* $(1 > \delta >)\ h \to 0$ *and* $nh \to \infty$ *as* $n \to \infty$, *then we obtain that for* $0 < x < 1 - \delta$,

$$\mathbf{E}_x B^c \to 0 \quad \text{as } n \to \infty.$$

PROOF. Throughout the proof, assume that $0 < x < 1 - \delta$. Let

$$W_j = 1 - [X_j \le x] - h^{-1}[x < X_j \le x + h]$$

and

$$\overline{W} = n^{-1}\sum_{j=1}^{n} W_j.$$

Then, since $\mathbf{E}_x W_j = -x$ and $\operatorname{Var}_x(\overline{W}) = n^{-1}\operatorname{Var}_x(W_j) \le (nh)^{-1}$, Chebyshev's inequality leads to

$$\mathbf{E}_x B^c = \mathbf{E}_x[\overline{W} - \mathbf{E}_x\overline{W} > x] \le (x^2 nh)^{-1},$$

which tends to zero as $n \to \infty$. □

LEMMA 4.2. *When h is a function of n such that* $(1 > \delta >)\ h \to 0$ *and* $nh \to \infty$ *as* $n \to \infty$, *for* $0 < x < 1 - \delta$,

$$\text{(4.3)} \qquad \sqrt{nh}\,\zeta_n(x)B \to_d N\big(0,(1-x)^2\big).$$

PROOF. Since as $n \to \infty$, $h^{1/2}Y \to_p 0$, $(nh)^{-1/2}\sqrt{1-h}\,Z \to_p 0$ and $Z \to_d N(0,1)$ and since by Lemma 4.1, $\mathbf{E}_x B \to 1$ for $0 < x < 1 - \delta$ as $n \to \infty$, we obtain the asserted result by applying Slutsky's theorem [Serfling (1980), page 19]. □

THEOREM 2. *If G is the degenerate distribution at* $\theta \equiv 1$, *then for any* $\varepsilon > 0$ *and* $(1 > \delta >)\ h > 0$ *such that* $nh \to \infty$ *as* $n \to \infty$, *we have for sufficiently large* n,

$$\text{(4.4)} \qquad R_n - R \ge \{3^{-1}(1 - \delta^3) - \varepsilon\}(nh)^{-1}.$$

PROOF. Applying (4.2) to (4.3) gives that $\liminf \mathbf{E}_x((nh)\zeta_n^2(x)B) \ge (1-x)^2$ for $0 < x < 1 - \delta$. Thus, by Fatou's lemma

$$\begin{aligned}\liminf E\{\mathbf{E}_x\{((nh)\zeta_n^2(X)B)[0 < X < 1-\delta]\}\} &\ge E\{\liminf \mathbf{E}_X\{((nh)\zeta_n^2(X)B)[0 < X < 1-\delta]\}\} \\ &\ge \int_0^{1-\delta}(1-x)^2\,dx \\ &= 3^{-1}(1-\delta^3).\end{aligned}$$

Finally, (4.1) and the definition of lim inf give us (4.4) □

From Theorem 2 we can see that there exists a $G \in \mathscr{G}$ such that with a choice of $h = n^{-1/2}$ and for some positive constant c_1,

$$c_1 n^{-1/2} \le R_n - R,$$

which is the same order as the upper bound (3.8). Hence, the rate of risk convergence for the priors in $\mathscr{G}$ cannot be improved beyond $n^{-1/2}$.

5. Example.

Let $g(\theta)$ be the pdf of $G(\theta)$. For $0 < m < +\infty$, we define

$$g(\theta) = 4m^{-2}[0 < \theta < 2^{-1}m] + 4m^{-1}(1 - \theta m^{-1})[2^{-1}m \le \theta < m],$$

which is the triangular distribution on $(0, m)$.

Acknowledgments. The author wishes to thank the referees and Professor Ryoichi Shimizu at the Institute of Statistical Mathematics in Tokyo for valuable comments and discussion.

REFERENCES

FOX, R. (1970). Estimating the empiric distribution function of certain parameter sequences. *Ann. Math. Statist.* **41** 1845–1852.

FOX, R. (1978). Solutions to empirical Bayes squared error loss estimation problems. *Ann. Statist.* **6** 846–853.

HANNAN, J. and MACKY, D. W. (1971). Empirical Bayes squared-error loss estimation of unbounded functionals in exponential families. Report No. RM-290, Dept. Statistics and Probability, Michigan State Univ.

JOHNS, M. V., JR. and VAN RYZIN, J. R. (1971). Convergence rates for empirical Bayes two-action problems I. Discrete case. *Ann. Math. Statist.* **42** 1521–1539.

JOHNS, M. V., JR. and VAN RYZIN, J. R. (1972). Convergence rates for empirical Bayes two-action problems II. Continuous case. *Ann. Math. Statist.* **43** 934–947.

LOÉVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton, N.J.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

ROBBINS, H. (1955). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 157–163. Univ. California Press.

ROBBINS, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.

ROBBINS, H. (1983). Some thoughts on empirical Bayes estimation. *Ann. Statist.* **11** 713–723.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

SINGH, R. S. (1974). Estimation of derivatives of average of μ-densities and sequence-compound estimation in exponential families. Report No. RM-318, Dept. Statistics and Probability, Michigan State Univ.

SINGH, R. S. (1976). Empirical Bayes estimation with convergence rates in non-continuous Lebesgue-exponential families. *Ann. Statist.* **4** 431–439.

SUSARLA, V. and O'BRYAN, T. (1979). Empirical Bayes interval estimates involving uniform distributions. *Comm. Statist. A—Theory Methods* **8** 385–397.

VAN RYZIN, J. and SUSARLA, V. (1977). On the empirical Bayes approach to multiple decision problems. *Ann. Statist.* **5** 172–181.

WEI, L. (1983). Empirical Bayes estimation of parameter with convergence rates about uniform distribution families $U(0, \theta)$. *Acta Math. Appl. Sinica* **6** 485–493.

WEI, L. (1985). Empirical Bayes estimation of location parameter with convergence rates in one-sided truncation distribution families. *Chinese Ann. Math.* **6A** 193–202.

INSTITUTE FOR SOCIO-ECONOMIC PLANNING
UNIVERSITY OF TSUKUBA
TSUKUBA-SHI, IBARAKI 305
JAPAN