The Annals of Statistics
1978, Vol. 6, No. 6, 1345–1351

UNIVERSAL BAYES ESTIMATORS[1]

BY A. L. RUKHIN
Purdue University

Let $x_1, \cdots, x_n$ be i.i.d. random variables with a distribution depending on the real parameter. Under what conditions is a generalized Bayes estimator independent of the choice of the even loss function? The known answer to this question is that this independence holds if the posterior density is symmetric and unimodal. The description of distributions and corresponding generalized prior densities on the real line, for which the posterior density is symmetric and unimodal, is presented. These families form an important subclass of all exponential laws with two-dimensional sufficient statistics.

1. Introduction. Let p_θ, $\theta \in R^1$ be a family of probability measures given on an abstract space X and such that each distribution P_θ is absolutely continuous with respect to some σ-finite measure μ on X. We assume throughout the paper that the coincidence of distributions P_{θ_1} and P_{θ_2} implies $\theta_1 = \theta_2$. Let λ be a generalized prior density on R^1. Suppose that the posterior density is defined almost everywhere $(\mu \times \cdots \times \mu)$ and is given by

$$\pi_x(\theta) = \left(\Pi_1^n p(x_j, \theta)\right)\lambda(\theta)\left[\int_{-\infty}^{\infty}\left(\Pi_1^n p(x_j, \theta)\right)\lambda(\theta)d\theta\right]^{-1},$$

where $p(u, \theta) = (dP_\theta/d\mu)(u)$, $u \in X$, $x = (x_1, \cdots, x_n) \in X^n$, $\theta \in R^1$. Also, let $W(\delta, \theta) = W(\delta - \theta)$ be the loss function depending only on the difference between the estimator δ and the true value of the parameter θ. This function is assumed to be nonincreasing on the negative half-line and nondecreasing on the positive half-line. The generalized Bayes estimator $\delta(x)$ of θ based on the random sample x satisfies the equation

$$\int_{-\infty}^{\infty} W(\delta(x) - \theta)\pi_x(\theta)d\theta = \inf_{t \in R^1}\int_{-\infty}^{\infty} W(t - \theta)\pi_x(\theta)d\theta. \tag{1}$$

In general this estimator depends on the choice of the loss function W. However, the exact form of this function is rarely known to the practical statistician. Therefore, it seems rather natural to investigate the situation where the estimator $\delta(x)$ is the same for every loss from a certain set of loss functions under consideration.

The above problem was solved by the author (Rukhin (1974)) for the case of a location parameter family and a constant density λ. A description was obtained of the distributions for which the best equivariant estimator of the location parameter is independent of the choice of the even loss function, for which integrals in (1) converge (cf. also Kagan, Linnik, Rao (1973) pages 255–258).

The following result is well known and often discussed in the literature (Britney, Winkler (1974), Sakrison (1970), Sherman (1958), Van Trees (1968) and Viterbi

Received June 1975; revised November 1977.
[1]This research was supported in part by National Science Foundation Grant MCS 77-19640.
AMS 1970 *subject classifications*. Primary 62F10; Secondary 62C10.
Key words and phrases. Generalized Bayes estimators, universal estimators, loss function, unimodal and symmetrical posterior density.

(1966)). If the continuous (in θ) posterior density $\pi_x(\theta)$ is unimodal and symmetrical with respect to the point $\delta(x)$ then $\delta(x)$ is optimal relative to every even loss function W, for which integrals in (1) converge. We call such estimators *universal.*

If a generalized Bayes estimator is the same for every even loss function from a sufficiently large set $\mathfrak{W}$ and, say, is the uniquely determined generalized maximum likelihood estimator, then the posterior density (assumed continuous) is unimodal and symmetrical. More precisely let $\mathfrak{W}$ be some set of even differentiable loss functions W for which the integrals in (1) are finite and the differentiation in t on the right side of (1) is allowed under the integral sign. Suppose also that the set $\mathfrak{W}$ possesses the following property: if for some continuous function g

$$\int_0^\infty W'(t)g(t)dt = 0$$

for all $W \in \mathfrak{W}$, then $g(t) = 0$.

Then (1) implies

$$\int_{-\infty}^\infty W'(t)\pi_x(\delta(x) - t)dt = 0$$

or

$$\int_0^\infty W'(t)[\pi_x(\delta(x) - t) - \pi_x(\delta(x) + t)]dt = 0.$$

Thus if the Bayes estimator $\delta(x)$ does not depend on W from $\mathfrak{W}$

$$\pi_x(\delta(x) - t) = \pi_x(\delta(x) + t)$$

i.e., the posterior density is symmetric. The unimodality follows from the uniqueness of the generalized maximum likelihood estimator. This property can be deduced also from the uniqueness of the Bayes estimator for some loss functions, for example, those corresponding to the problem of confidence estimation of θ.

In this paper we describe, under mild regularity restrictions, the densities $p(u, \theta)$ and generalized priors $\lambda(\theta)$ for which the posterior density is symmetrical and unimodal. Hence in these, and practically only in these, cases is the best estimator universal.

2. The main result.

We prove the following

THEOREM. *Let $\{p(u, \theta), \theta \in R^1, u \in X\}$ be a family of probability densities given on the pathwise connected topological space X. Assume that $p(u, \theta)$ is continuous in u for every fixed θ and continuous in θ for fixed u, the function $\lambda(\theta)$ is continuous, and both these functions are positive. Suppose further that the posterior density $\pi_x(\theta)$ is defined for all $x \in X^n$ for some $n \geqslant 3$, except for x from some nowhere dense set N, and that $\pi_x(\theta)$ is symmetrical with respect to the point $\delta(x)$, $x \in X^n - N$, and unimodal* (i.e. *increasing for $\theta < \delta(x)$*). *If $\delta(X^n - N) = R^1$ then either*

$$(1^0) \qquad \log p(u, \theta) + n^{-1}\log\lambda(\theta) = A_1(u)e^{\alpha\theta} + A_2(u)e^{-\alpha\theta} + A_3(u)$$

and

$$\delta(x_1, \cdots, x_n) = (2\alpha)^{-1}\log(\Sigma_1^n A_2(x_j)/\Sigma_1^n A_1(x_j))$$

or

$$(2^0) \qquad \log p(u, \theta) + n^{-1} \log \lambda(\theta) = B_1(u)\theta^2 - 2B_2(u)\theta + B_3(u)$$

and

$$\delta(x_1, \cdots, x_n) = \Sigma_1^n B_2(x_j)/\Sigma_1^n B_1(x_j).$$

PROOF. 1. The conditions of the theorem imply that for $x \in X^n - N$, $B \in R^1$ $\pi_x(\delta(x) - B) = \pi_x(\delta(x) + B)$. Moreover, if for some real B and $t \pi_x(t - B) = \pi_x(t + B)$, it follows from the unimodality and symmetry assumption that $t = \delta(x)$.

It is also clear that if the inequality $\pi_x(t - B) \leqslant \pi_x(t + B)$ is valid for some $B > 0$, then $\delta(x) \geqslant t$.

Let

$$\varphi(u, B) = \log p(u, B) + n^{-1} \log \lambda(B)$$

so that

$$\pi_x(t) = \exp(\Sigma_1^n \varphi(x_j, t))\left[\int_{-\infty}^{\infty} \exp(\Sigma_1^n \varphi(x_j, B))dB\right]^{-1}.$$

Define for a fixed t the function

$$R_t(u, B) = \varphi(u, t + B) - \varphi(u, t - B).$$

The relation $\delta(x) = t$ signifies that

$$(2) \qquad \Sigma_1^n R_t(x_j, B) = 0$$

for all B, $x \notin N$. The unimodality of $\pi_x(t)$ implies that if (2) holds for some $B \neq 0$ then this relation is valid for all B.

Note that if $x \notin N$ then the sum $\Sigma_1^n R_t(x_j, B)$ is positive for some $B > 0$ if and only if $\delta(x) > t$. This remark shows that since $\delta(X^n - N) = R^1$,

$$\alpha_t(B) = \inf_u R_t(u, B) < 0,$$

and that

$$\beta_t(B) = \sup_u R_t(u, B) > 0.$$

The connectedness of X and continuity in u of the function $R_t(u, B)$ for every fixed B show that the image $R_t(X, B)$ of X under the mapping $R_t(\cdot, B)$ forms an interval (closed or not) with boundary points $\alpha_t(B)$ and $\beta_t(B)$.

2. If the inequality

$$(3) \qquad (n-1)^{-1}\beta_t(B) \leqslant |\alpha_t(B)| \leqslant (n-1)\beta_t(B)$$

holds for some $B \neq 0$, then for every u there exist $x_2, \cdots, x_n$ with the property

$$R_t(u, B) + \Sigma_2^n R_t(x_j, B) = 0.$$

Thus if (3) is valid for some $B \neq 0$, the set

$$E_t = \{u : \exists\, x_2, \cdots, x_n, \delta(u, x_2, \cdots, x_n) = t\}$$

coincides with X and (3) holds for all $B \neq 0$.

We next show that the set $T_u = \{t : u \in E_t\}$ contains an open interval except for u from a nowhere dense set. This remains to be proven only if (3) fails. Thus assume, for instance, that $(n-1)|\alpha_t(b)| < \beta_t(B)$ for some $B > 0$ and $\alpha_t(B) \in R_t(X, B)$. Then

$$E_f = \{u : R_t(u, B) \leqslant (n-1)|\alpha_t(B)|\}$$

and

$$T_u = \{t : R_t(u, B) \leqslant (n-1)|\alpha_t(B)|\}.$$

Suppose that there exists u, for which $(u, x_2, \cdots, x_n) \in N$ for some $x_2, \cdots, x_n$ and such that the set T_u does not contain a nonempty interval. Then because of continuity of $R_t(u, B)$ in t for fixed u one deduces that $F_u = \{t : R_t(u, B) = (n-1)|\alpha_t(B)|\}$, and for all t the inequality $R_t(u, B) \geqslant (n-1)|\alpha_t(B)|$ holds. This inequality implies that

$$R_t(u, B) + \Sigma_2^n R_t(x_j, B) \geqslant 0$$

for all t, i.e., $\delta(u, x_2, \cdots, x_n) \geqslant t$ for all t, which is impossible. However the set u for which $(u, x_2, \cdots, x_n) \notin N$ for some $x_2, \cdots, x_n$ is dense in X. Hence T_u contains an interval except for u from a nowhere dense set.

If $(n-1)|\alpha_t(B)| < \beta_t(B)$ for some $B > 0$ but $\alpha_t(B) \notin R_t(X, B)$, then $T_u = \{t : R_t(u, B) < (n-1)|\alpha_t(B)|\}$ and the nonempty set T_u is open and contains an interval. The cases $(n-1)|\alpha_t(B)| < \beta_t(B)$, $B < 0$ and $(n-1)\beta_t(B) < |\alpha_t(B)|$, $B \neq 0$ are treated analogously. (Note that $\beta_t(-B) = -\alpha_t(B)$).

Thus T_u always contains an interval except for u from a nowhere dense set.

3. For $u \in E_t$ and fixed t, $B_1 \neq 0$, the relation $R_t(u, B_1) = R_t(u_0, B_1)$ implies $R_t(u, B) = R_t(u_0, B)$ for all B. In fact, since for some $x_2, \cdots, x_n$

$$R_t(u, B_1) = -\Sigma_2^n R_t(x_j, B_1) = R_t(u_0, B_1),$$

one gets

$$R_t(u, B) = -\Sigma_2^n R_t(x_j, B) = R_t(u_0, B).$$

Thus for every B there exists a function g (dependent on t) defined on $R_t(X, B_1)$ and taking values in $R_t(X, B)$ such that

$$R_t(u, B) = g(R_t(u, B_1)), \qquad u \in E_t.$$

If $R_t(u_1, B_1) \neq R(u_2, B_1)$ for some u_1, u_2 and S in the path connecting u_1 and u_2, then the set $R_t(S, B_1)$ is a nondegenerate interval. The continuous function $R_t(\cdot, B)$ is bounded on the compact set S, and the function g is bounded on the interval $R_t(S, B_1)$.

The function g satisfies the following functional equation

$$g(z_1) + \cdots + g(z_n) = 0.$$

This equation is equivalent to the known Cauchy's functional equation, and since g is bounded on a interval there exists a real number m for which $g(z) = mz$. (cf. Aczel (1966) pages 34–35). Thus we have obtained

$$R_t(u, B) = m(B, t)R_t(u, B_1)$$

In other words, for all t, $B \in R^1$ and $u \in E_t$ the following equation holds

$$\varphi(u, t + B) - \varphi(u, t - B) = m(B, t)[\varphi(u, t + B_1) - \varphi(u, t - B_1)] = m(B, t)k(u, t).$$

Here the set $T_u = \{t : u \in E_t\}$ contains a nonempty interval except for u from a nowhere dense set, and the functions φ and k are continuous. It is clear that $\Sigma_1^n k(x_j, \delta(x)) = 0$ and that the relation $\Sigma_1^n k(x_j, t) = 0$ implies $t = \delta(x)$.

4. For any fixed t_0 there exists $u^* \in E_{t_0}$ such that the set $\{t : k(u^*, t) \neq 0\}$ is dense in T_{u^*}. (If this were not true, then for all $u \in E_{t_0}$ the set $\{t : k(u, t) = 0\}$ would contain a nonempty interval, contradicting the property of the function k established in 3.) Therefore, for all $u \in E_t$ one obtains with some u^*

$$\varphi(u, t + B) - \varphi(u, t - B) = l(u, t)[\varphi(u^*, t + B) - \varphi(u^*, t - B)], \tag{4}$$

where $l(u, t) = k(u, t)[k(u^*, t)]^{-1}$.

Let $W_\varepsilon(B)$ be an infinitely differentiable function such that $\int_\infty^\infty W_\varepsilon(B)dB = 1$ and $W_\varepsilon(B) = 0$ for $|B| > \varepsilon$, multiplying both sides of (4) by $W_\varepsilon(B - s)$ and integrating out B gives

$$\varphi_\varepsilon(u, t + s) - \varphi_\varepsilon(u, t - s) = l(u, t)[\varphi_\varepsilon(u^*, t + s) - \varphi_\varepsilon(u^*, t - s)] \tag{5}$$

with

$$\varphi_\varepsilon(u, s) = \int_\infty^\infty \varphi(u, B + s) W_\varepsilon(B)dB.$$

Note that $\varphi_\varepsilon(u, s)$ is an infinitely differentiable function of s. Therefore $l(u, t)$ is infinitely differentiable in t within the interval where $\varphi_\varepsilon(u^*, t + s) \neq \varphi_\varepsilon(u^*, t - s)$, ε arbitrary positive, i.e., is infinitely differentiable in t outside some nowhere dense set.

Differentiating twice by s on both sides of (5) we obtain

$$\varphi_\varepsilon''(u, t + s) - \varphi_\varepsilon''(u, t - s) = l(u, t)[\varphi_\varepsilon''(u^*, t + s) - \varphi_\varepsilon''(u^*, t - s)].$$

Differentiating (5) twice by t gives

$$\begin{aligned}\varphi_\varepsilon''(u, t + s) - \varphi_\varepsilon''(u, t - s) = {} & l''(u, t)[\varphi_\varepsilon''(u^*, t + s) - \varphi_\varepsilon(u^*, t - s)] \\ & + 2l'(u, t)[\varphi_\varepsilon'(u^*, t + s) - \varphi_\varepsilon'(u^*, t - s)] \\ & + l(u, t)[\varphi_\varepsilon(u^*, t + s) - \varphi_\varepsilon(u^*, t - s)].\end{aligned}$$

Here $\varphi_\varepsilon'(u, t) = (\partial/\partial t)\varphi_\varepsilon(u, t)$, $l'(u, t) = (\partial/\partial t)\, l(u, t)$ and the same notation holds for second derivatives.

The last two relations imply that

$$l''(u, t)[\varphi_\varepsilon(u^*, t + s) - \varphi_\varepsilon(u^*, t - s)] = -2l'(u, t)[\varphi_\varepsilon'(u^*, t + s) - \varphi_\varepsilon'(u^*, t - s)].$$

We can assume that $l'(u, t)$ does not vanish within some interval.

If it did, then $l(u, t) = l(u)$ for t taking values in some interval, and $u \in E_t$. It would follow that $k(u, t) = l(u)k(u^*, t)$ which contradicts the properties of the function k established in 3.

Thus the equation

$$-\frac{1}{2}\frac{l''(u,t)}{l'(u,t)} = \frac{\varphi_\varepsilon'(u^*, t+s) - \varphi_\varepsilon'(u^*, t-s)}{\varphi_\varepsilon(u^*, t+s) - \varphi_\varepsilon(u^*, t-s)}$$

is established. It implies, for some t_0 that

$$\begin{aligned} -\frac{1}{2}[\log l'(u,t) - \log l'(u,t_0)] \\ &= \log[\varphi_\varepsilon(u^*, t+s) - \varphi_\varepsilon(u^*, t-s)] \\ &\quad -\log[\varphi_\varepsilon(u^*, t_0+s) - \varphi_\varepsilon(u^*, t_0-s)]. \end{aligned}$$

Hence the following equation holds:

$$\varphi_\varepsilon(u^*, t+s) - \varphi_\varepsilon(u^*, t-s) = h(u,t)[\varphi_\varepsilon(u^*, t_0+s) - \varphi_\varepsilon(u^*, t_0-s)] \tag{6}$$

with $h(u, t) = [l'(u, t_0)/l(u, t)]^{\frac{1}{2}}$. From (6) it is clear that $h(u, t)$ does not depend on u.

The equation of type (6) is treated in Aczel (1966) page 176. Its solutions having probabilistic sense are of the form

$$\begin{aligned} \varphi_\varepsilon(u^*, s) &= D_\varepsilon \cosh(\alpha s + F_\varepsilon) + C_\varepsilon \\ &= D_\varepsilon s^2 + F_\varepsilon s + C_\varepsilon \end{aligned}$$

where α, as seen from (6), is independent of ε. From (5) it follows that

$$\begin{aligned} \varphi_\varepsilon(u, s) &= D_\varepsilon(u)\cosh(\alpha s + F_\varepsilon(u)) + C_\varepsilon(u) \\ &= D_\varepsilon(u)s^2 + F_\varepsilon(u)s + C_\varepsilon(u). \end{aligned}$$

Letting ε go to zero shows that

$$\begin{aligned} \varphi(u, s) &= A_1(u)e^{\alpha s} + A_2(u)e^{-\alpha s} + A_3(u) \\ &= B_1(u)s^2 - 2B_2(u)s + B_3(u). \end{aligned}$$

The form of the corresponding estimators follows easily and this completes the proof.

REMARK. The same method can be used in the case when the parameter space is an open interval.

3. Discussion. It is worthwhile noting that the function $\varphi(u, \theta)$ in 2^0 of the theorem can be considered as a limiting case of the solution 1^0 as α tends to zero. Therefore, we will speak here mostly about distributions corresponding to 1^0. The density in this case can be represented in the form

$$p(u, \theta) = C(\theta)B(u)\exp\{A_1(u)e^{\alpha\theta} + A_2(u)e^{-\alpha\theta}\}$$

where $C(\theta) = [\lambda(\theta)]^{1/n}$, $B(u) = \{\exp A_3(u)\}$. Distributions of this form are of exponential type and they have a two dimensional sufficient statistic for the parameter θ (namely $\Sigma_1^n A_1(x_j)$ and $\Sigma_1^n A_2(x_j)$).

With the prior density

$$\lambda(\theta) = [C(\theta)]^{-n} \exp\{a_1 e^{\alpha\theta} + a_2 e^{-\alpha\theta}\},$$

the Bayes estimator based on the sample of size n is a universal one. This density belongs to the class of conjugate prior distributions.

If θ is a location parameter, then the distributions in 1^0 have the form

$$p(u, \theta) = p(u - \theta) = C \exp\{-\beta \cosh \alpha(u - \theta)\}. \tag{7}$$

In this case the prior density

$$\lambda(\theta) = \exp\{-\gamma \cosh \alpha(\theta - \theta_0)\}$$

can be chosen in such a way that the universal estimator exists for all sample sizes. Statistical properties of the distribution in (7) were investigated by the author (Rukhin (1974)).

The author is not aware of nonlocation parameter distributions for which the universal estimator exists for all sample sizes. However, it seems that the estimators given in the theorem may be of use in other estimation problems. Because of the relatively simple form of these estimators their distribution can be calculated but resulting formulas can be very messy. The asymptotical distribution can be derived from Rao (1965), page 321.

Note that the equation (6) has a solution of the form $D_\varepsilon \cos(\alpha s + F_\varepsilon) + C_\varepsilon$. This solution corresponds to the case of estimating the parameter θ with values in the circle. The analogue of the distributions (7) is the known von Mises distribution with the density proportional to $\exp\{D \cos \alpha(u - \theta)\}$, $0 \leqslant u, \theta < 2\pi\alpha^{-1}$. With the conjugate prior density λ also corresponding to the von Mises distribution, the universal estimator exists for all sample sizes.

REFERENCES

[1] ACZEL, J. (1966). *Lectures on Functional Equations and Their Applications*. Academic Press, New York.

[2] BRITNEY, R. R. and WINKLER, R. L. (1974). Bayesian point estimation and prediction. *Ann. Inst. Statist. Math.* **26** 15–34.

[3] KAGAN, A. M., LINNIK, YU. V. and RAO, C. R. (1973). *Characterizations Problems in Mathematical Statistics*. Wiley, New York.

[4] RAO, C. R. (1965). *Linear Statistical Inference and its Applications*. Wiley, New York.

[5] RUKHIN, A. L. (1974). Strongly symmetrical families and statistical analysis of their parameters. *Zap. Naucvn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* **43** 59–87. *(English Translation (1978): J. Soviet Math.* **9** 886–910.)

[6] SAKRISON, D. J. (1970). *Notes on Analog Communication*. Van Nostrand Reinhold, New York.

[7] SHERMAN, S. (1958). Non mean-square error criteria. *I.R.E. Trans. Information Theory* **IT-4** 125–126.

[8] VAN TREES, H. L. (1968). *Detection, Estimation and Modulation Theory*. Wiley, New York.

[9] VITERBI, J. (1966). *Principles of Coherent Communication*. McGraw Hill, New York.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
MATHEMATICAL SCIENCES BLDG.
WEST LAFAYETTE, INDIANA 47907