The Annals of Statistics
1975, Vol. 3, No. 3, 736–740

A NOTE ON SOME BAYESIAN NONPARAMETRIC ESTIMATES

By M. Goldstein

University of Oxford

With respect to a general quadratic loss function, the Bayes rule for the mean of a probability distribution of unknown form is obtained, in the class of linear functions of the sample. The associated Bayes risk is also obtained. A number of recent results in the literature are shown to be direct corollaries of this result, and applications are given for the empirical distribution function of the sample.

A sample, $\underline{x} = (x_1, \cdots, x_n)$, is drawn from $F(\cdot)$, a probability distribution on the real line. The form of $F(\cdot)$ is unknown, and it is required to estimate the mean of $F(\cdot)$, $\mu(F)$. The sample is not necessarily independent, but it is assumed that $E(x_i \mid F) = \mu(F)$ for all i, $E(x_i^2 \mid F)$ is independent of i, and $E(x_i x_j \mid F)$ is independent of i, j for $i \neq j$.

A prior probability measure, $P(\cdot)$, is assigned over $\mathscr{F}$, the space of all probability distributions on the real line, and the Bayes rule is sought with respect to the general quadratic loss function

$$L(F, d) = w(F)(\mu(F) - d)^2 , \tag{1}$$

where $w(\cdot)$ is a real nonnegative function on $\mathscr{F}$.

Direct, explicit evaluation of the Bayes rule for $\mu(F)$ will, typically, be very difficult. However, in Goldstein (1973), a general method is detailed for evaluating the Bayes rule for $\mu(F)$ in the class of linear combinations of a finite number of functions of $\underline{x}$, and also the associated Bayes risk (essentially, by solving the implied regression equations, using the general theory of least squares). This method is applied to derive the Bayes rule for $\mu(F)$, in the class of estimates of the form $a_1 x_1 + \cdots + a_n x_n + b$, and this rule will be shown to be equivalent to the Bayes rule for $\mu(F)$, in the class of estimates of the form $a\bar{x} + b$, where $\bar{x}$ is the sample mean (denote this estimate as the B.L.E. (Bayes linear estimate) in $\bar{x}$). A number of recent results in the literature are shown to be direct corollaries of this result and also some applications are given.

The loss function is normalized by assuming that

$$\int w(F)\, dP(F) = 1 .$$

Thus, the probability measure $P_w(\cdot)$ may be defined on $\mathscr{F}$, by the relation

$$dP_w(F) = w(F)\, dP(F) . \tag{2}$$

Assuming that, with respect to the prior measure $P_w(\cdot)$, the mean and variance of F exist a.e., the expected mean and the variance of the mean and

Received April 1974; revised May 1974.

the variance of $\bar{x}$ are defined, with respect to P_w, by

$$\begin{aligned} \bar{\mu}_w &= \int \mu(F)\, dP_w(F)\,, \\ V_w(\mu) &= \int (\mu(F) - \bar{\mu}_w)^2\, dP_w(F)\,, \\ V_w(\bar{x}) &= \int (\int (\bar{x} - \mu(F))^2\, d\bar{F}(\bar{x}))\, dP_w(F)\,, \end{aligned} \tag{3}$$

respectively, where $\bar{F}$ is the distribution of $\bar{x}$, when $\underline{x}$ is drawn from F.

THEOREM 1. (i) *Given observation* $\underline{x}$, *the Bayes estimate for* $\mu(F)$, *in the class of estimates of the form* $(a_1 x_1 + \cdots + a_n x_n + b)$, *with loss function* (1), *is given by*

$$(V_w(\bar{x})\bar{\mu}_w + V_w(\mu)\bar{x})/(V_w(\bar{x}) + V_w(\mu))\,. \tag{4}$$

(ii) *The Bayes risk of estimate* (4) *is*

$$((V_w(\bar{x}))^{-1} + (V_w(\mu))^{-1})^{-1}\,. \tag{5}$$

PROOF. By Theorem 2.1 (i) of Goldstein (1973), the Bayes rule for $\mu(F)$ of required form is

$$(1, x_1, \cdots, x_n)\underline{D}^{-1}\underline{b}\,, \tag{6}$$

where $\underline{D}$ is the $(n+1) \times (n+1)$ matrix whose (i, j)th entry, d_{ij}, is

$$d_{ij} = \int w(F)E(x_i x_j \mid F)\, dP(F)\,, \qquad i, j = 0, 1, 2, \cdots, n\,, \tag{7}$$

(where $x_0 = 1$), and $\underline{b}$ is the $(n+1)$ vector whose ith entry, b_i, is

$$b_i = \int w(F)\mu(F)E(x_i \mid F)\, dP(F)\,, \qquad i = 0, 1, 2, \cdots, n\,. \tag{8}$$

The conditions on $\underline{x}$ assumed above ensure that the coefficient of each x_i, $i = 1, 2, \cdots, n$, in the expansion of (6) above, is the same. Thus, the Bayes rule for $\mu(F)$ of required form is also the Bayes rule for $\mu(F)$ in the class of estimates of the form $(a\bar{x} + b)$. Applying Theorem 2.1 (i) again, the Bayes rule of required form is

$$(1, \bar{x})\begin{pmatrix} 1 & \bar{\mu}_w \\ \bar{\mu}_w & \bar{\mu}_{2(w)} \end{pmatrix}^{-1}\begin{pmatrix} \bar{\mu}_w \\ \bar{\mu}_w^{(2)} \end{pmatrix}, \tag{9}$$

where

$$\begin{aligned} \bar{\mu}_{2(w)} &= \int\int (\bar{x})^2\, d\bar{F}(\bar{x})\, dP_w(F)\,, \\ \bar{\mu}_w^{(2)} &= \int (\mu(F))^2\, dP_w(F)\,. \end{aligned} \tag{10}$$

Expanding (9) and noting that

$$\begin{aligned} \bar{\mu}_{2(w)} - \bar{\mu}_w^{(2)} &= V_w(\bar{x})\,, \\ \bar{\mu}_w^{(2)} - (\bar{\mu}_w)^2 &= V_w(\mu)\,, \end{aligned} \tag{11}$$

gives (4).

(ii) Similarly, applying Theorem 2.1 (ii), the Bayes risk of (4) is

$$\begin{vmatrix} 1 & \bar{\mu}_w & \bar{\mu}_w \\ \bar{\mu}_w & \bar{\mu}_{2(w)} & \bar{\mu}_w^{(2)} \\ \bar{\mu}_w & \bar{\mu}_w^{(2)} & \bar{\mu}_w^{(2)} \end{vmatrix} \Bigg/ \begin{vmatrix} 1 & \bar{\mu}_w \\ \bar{\mu}_w & \bar{\mu}_{2(w)} \end{vmatrix}\,. \tag{12}$$

Expanding (12) and applying (11) gives (5).

When the function $w(F)$ is constant (i.e., $L(F, d) = (\mu(F) - d)^2$), then the measure $P_w(\cdot)$, defined by (2) is the original measure $P(\cdot)$, and $\bar{\mu}_w$, $V_w(\mu)$, $V_w(\bar{x})$ are the prior expected mean, the prior variance of the mean and the prior variance of $\bar{x}$, respectively. Denote these quantities by $\bar{\mu}$, $V(\mu)$, $V(\bar{x})$, respectively.

The following corollary is immediate:

COROLLARY 1. (i) *If*

$$E(\mu(F) \mid \underline{x}) = a\bar{x} + b\,,$$

where a, b are constants, then $E(\mu(F) \mid \underline{x})$ is given by (4), *with $w(F) = 1$.*

(ii) *The posterior variance of the distribution of $\mu(F)$ (i.e., the posterior Bayes risk with quadratic loss) is not greater than* (5), *with $w(F) = 1$.*

Corollary 1 (i) is essentially the result given in Ericson (1969), and Corollary 1 (ii) is essentially the result given in Finucan (1971). The generalizations of Corollary 1 (i) and (ii), are immediate: that is, for a general set of functions in $\underline{x}$, $h_1(\cdot), \cdots, h_r(\cdot)$, if $E(\mu(F) \mid \underline{x}) = a_1 h_1(\underline{x}) + \cdots + a_r h_r(\underline{x})$, then we can apply Theorem 2.1 (i), Goldstein (1973), to identify the constants $a_1, \cdots, a_r$; similarly, by applying Theorem 2.1 (ii), to general sets of functions in $\underline{x}$, we may derive general upper bounds for the Bayes risk.

The nature of the B.L.E. in $\bar{x}$ is partly clarified in the following corollary.

COROLLARY 2. *When $\underline{x}$ is an independent sample, then for any prior distribution P on $\mathscr{F}$ with respect to which the mean and variance of F exist* a.e., *there exists another prior distribution P' on $\mathscr{F}$, such that the quantities $\bar{\mu}$, $V(\mu)$, $V(x)$ are the same with respect to P and P', and such that under quadratic loss, the* B.L.E. *in $\bar{x}$, for $\mu(F)$, with respect to P, is the unrestricted Bayes rule with respect to P', for every sample size.*

For example, with $\bar{\mu}$, $V(\mu)$, $V(x)$ defined for P ($V(x)$ is the value $V(\bar{x})$ for a sample size one), define a normal prior distribution for μ, with mean $\bar{\mu}$, variance $V(\mu)$, and define the conditional distribution for x given μ as normal, mean μ, variance $V(x)$. By (4), with $w(F) = 1$, this prior distribution satisfies Corollary 2. Thus, in this sense, the B.L.E. in $\bar{x}$ is the best normal approximation to the Bayes rule.

In the derivation of (4), (5), the only property of $\bar{x}$ which was used was that it was an unbiased estimate for $\mu(F)$. Therefore these results can be applied to any unbiased estimator. As a specific example consider the following:

We wish to estimate the distribution $F(\cdot)$, over the real line, from an independent, identically distributed sample $\underline{x} = (x_1, \cdots, x_n)$.

Define the empirical distribution function of the sample by $F_n(\cdot)$, where

$$F_n(x_0) = r_0/n\,, \tag{13}$$

r_0 being the number of elements of the sample which are not greater than x_0.

For each x_0, define a real function $g_0(\cdot)$ on $\mathscr{F}$ by

$$g_0(F) = F(x_0)\,. \tag{14}$$

It is easily seen that $F_n(x_0)$ is unbiased for $g_0(F)$, and so Theorem 1 applies. Define

$$
\begin{aligned}
\bar{F}(x_0) &= \int F(x_0)\, dP(F) , \\
\bar{F}^{(2)}(x_0) &= \int F^2(x_0)\, dP(F) , \qquad (15)\\
d_0 &= (\bar{F}(x_0) - \bar{F}^{(2)}(x_0))/(\bar{F}^{(2)}(x_0) - (\bar{F}(x_0))^2) .
\end{aligned}
$$

Since r_0 can be thought of as the number of successes in n binomial trials with probability of success $F(x_0)$,

$$E(F_n^2(x_0)) = ((n-1)\bar{F}^{(2)}(x_0) + \bar{F}(x_0))/n . \qquad (16)$$

Theorem 1 then gives

COROLLARY 3. (i) *The Bayes rule for $F(x_0)$, in the class $(a\bar{F}(x_0) + bF_n(x_0))$, with respect to loss function*

$$L_0(F, F') = (F(x_0) - F'(x_0))^2 , \qquad (17)$$

is

$$(d_0\bar{F}(x_0) + nF_n(x_0))/(d_0 + n) . \qquad (18)$$

(ii) *The Bayes risk of* (18) *is*

$$(n(\bar{F}(x_0) - \bar{F}^{(2)}(x_0))^{-1} + (\bar{F}^{(2)}(x_0) - (\bar{F}(x_0))^2)^{-1})^{-1} . \qquad (19)$$

(As in Theorem 1, (18) is also the Bayes rule for $F(x_0)$ in the class of estimates of the form $a_1 I_1(x_0) + \cdots + a_n I_n(x_0) + b$, where I_j is the indicator of the set $[x_j, \infty)$.)

Note that (19) is less than $(\bar{F}(x_0) - \bar{F}^{(2)}(x_0))/n$, which is never greater than $(4n)^{-1}$.

When the prior distribution for F is Dirichlet, parameter α, (18) is the unrestricted Bayes rule for $F(x_0)$ and d_0 is $\alpha(\mathbb{R})$, (Ferguson (1973)). (The ratio of the Bayes risk before and after sampling is thus $(\alpha(\mathbb{R})/(\alpha(\mathbb{R}) + n))$.) The following extension of Corollary 2 is thus immediate.

COROLLARY 4. *For any given x_0, the Bayes rule for $F(x_0)$ under loss function* (17), *in the class of linear estimates of $F_n(x_0)$, against any prior distribution P on $\mathscr{F}$, is the unrestricted Bayes estimate against the Dirichlet prior distribution P' on $\mathscr{F}$, with parameter α defined by $\alpha(t) = \bar{F}(t)d_0$. There exists a single Dirichlet prior distribution such that the result holds for each x_0 in S, where S is a given subset of the real line, if and only if d_0 is constant on S.*

Acknowledgments. I would like to thank my supervisor, Dr. A. F. M. Smith, for many helpful discussions, and the Science Research Council for their financial support. A referee's comments were most useful.

REFERENCES

[1] ERICSON, W. A. (1969). A note on the posterior mean of a population mean. *J. Roy. Statist. Soc. Ser. B* **31** 332–334.

[2] FERGUSON, THOMAS S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

[3] FINUCAN, H. H. (1971). Posterior precision for non-normal distributions. *J. Roy. Statist. Soc. Ser. B* **33** 95–97.

[4] GOLDSTEIN, M. (1975). Approximate Bayes solutions to some nonparametric problems. *Ann. Statist.* **3** 512–517.

MATHEMATICAL INSTITUTE
UNIVERSITY OF OXFORD
24-29 ST. GILES
OXFORD OX1 3LB
ENGLAND