The Annals of Statistics
1977, Vol. 5, No. 6, 1235–1239

ON CONTRACTIONS OF BAYES ESTIMATORS FOR EXPONENTIAL FAMILY DISTRIBUTIONS

BY MICHAEL GOLDSTEIN

University of Hull

Sufficient conditions are provided to ensure that the Bayes estimator with squared error loss for the natural parameter of an exponential family distribution is a linear function of the sample value, given that various contractions of the Bayes estimator are also Bayes estimators, generalizing a result of Meeden (1976).

Let X be a normal random variable with mean θ and variance one. Meeden (1976) has shown that if a generalized Bayes estimator for θ, with squared error loss, has the property that any linear contraction of this estimator is also a Bayes estimator for θ, and the further property that there exists a constant M such that the estimator is bounded above by $(x + M)$ for positive x, and bounded below by $(x - M)$ for negative x, then the estimator must be a linear function of x.

We shall begin by providing a simple, elementary proof of Meeden's result in the general case for any exponential family distribution on the real line with a single natural parameter, requiring not that all linear contractions should be Bayes estimators but only that any subset tending to zero should have this property.

We then extend the result to the case of any contractions tending to zero smoothly, in a specified sense, motivating the result by considering a mixture of normal prior distributions. Finally, having noted a general result on the posterior mean of an exponential parameter, the linear bounds on the estimator are replaced by polynomials of any order.

THEOREM. *Let X be a random variable with* pdf $p(x, \theta)$ *of the form*

$$p(x, \theta) = \exp(\theta x + Q(\theta) + R(x)), \tag{1}$$

where $\theta, x \in (-\infty, \infty)$.

Let g be a generalized Bayes estimator for θ with squared error loss. Suppose that there exist values a, b such that

$$g(x) \leqq ax + b \quad \text{for} \quad x \geqq 0, \qquad \text{and} \tag{2}$$
$$g(x) \geqq ax - b \quad \text{for} \quad x < 0.$$

If there exists a sequence of positive values λ_i, converging to zero, for which $\lambda_i g$ is a Bayes estimator for each i, then $g(x) = cx + d$, for all x, for some constants c and d.

Received September 1976; revised March 1977.

AMS 1970 *subject classifications.* Primary 62F10; Secondary 62C10.

Key words and phrases. Exponential family, linear estimates, normal distribution, quadratic loss, Bayes estimation.

PROOF. For any prior distribution for θ, repeated differentiation of the expression for $E(\theta \mid x)$ gives the relation

$$m'''(x) + 4m(x)m''(x) + 6m'(x)m^2(x) + 3(m'(x))^2 + m^4(x) = E(\theta^4 \mid x), \tag{3}$$

where $m(x) = E(\theta \mid x)$.

In the particular case when $m(x) = \lambda_i g(x)$, substituting into (3) gives $g'''(x) \geqq 0$, for all x, as the values λ_i are converging to zero, and $E(\theta^4 \mid x)$ is always nonnegative.

If $g''(x)$ does not vanish identically, there must be a value x_0 for which $g''(x_0) = \varepsilon \neq 0$. Suppose $\varepsilon > 0$. From the above $g''(x) \geqq \varepsilon$ for all $x \geqq x_0$, and thus $g(x)$ increases as x^2 for large x, contradicting (2). A similar contradiction arises if the value ε defined above is negative, and thus $g''(x)$ vanishes identically, and the result follows.

The above theorem extends immediately to the multiparameter exponential family, with pdf $p(\mathbf{x}, \boldsymbol{\theta})$ of the form

$$p(\mathbf{x}, \boldsymbol{\theta}) = \exp(\boldsymbol{\theta}'\mathbf{x} + \sum (R_i(x_i) + Q_i(\theta_i)))\,.$$

If the prior distribution for $\boldsymbol{\theta}$ is such that the elements of $\boldsymbol{\theta}$ are a priori independent, then the posterior expectation for θ_i will be of the form $g_i(x_i)$. If, for each i, there exist values a_i, b_i such that $g_i(x_i)$ satisfies (2), and there also exists a sequence of positive vectors $\boldsymbol{\lambda}_i$ converging to zero, for which $\boldsymbol{\lambda}_i \mathbf{g}$ is a Bayes estimator for $\boldsymbol{\theta}$ for each j, then $\mathbf{g}$ is of the form $\mathbf{C}'\mathbf{x} + \mathbf{d}$, for some constant vectors $\mathbf{C}$ and $\mathbf{d}$.

As a simple counterexample, for nonexponential families, consider the family of random variables, X, with probability distributions given by

$$P(X = 1) = \theta\,, \qquad P(X = -1) = k\theta\,, \qquad P(X = 0) = 1 - (k+1)\theta\,,$$

where k is some positive constant. Now $E(\theta \mid x)$ is not a linear function of x, for any nontrivial prior distribution for θ. But for any two prior distributions, denoting by $E_1(\theta \mid x)$, $E_2(\theta \mid x)$ the posterior mean for θ under the first and second prior distributions, then the requirement that $E_2(\theta \mid x) = \lambda E_1(\theta \mid x)$, for each value of x, where $0 < \lambda < 1$, is equivalent to the two requirements

$$E_2\theta = \frac{\lambda(1 - (k+1)(E_1\theta^2/E_1\theta))E_1\theta}{1 - \lambda(k+1)(E_1\theta^2/E_1\theta) - (1-\lambda)(k+1)E_1\theta}\,,$$

$$E_2\theta^2 = \lambda \frac{E_1\theta^2}{E_1\theta} E_2\theta\,.$$

(This follows by comparing the posterior mean under each prior distribution, for each value of x.)

But, for values $E_1\theta$, $E_1\theta^2$ derived from any prior distribution for θ, we may find, for any $\lambda \in (0, 1)$, a prior distribution for θ for which the values $E_2\theta$, $E_2\theta^2$ satisfy the above equations, as $(E_2\theta)^2 \leqq E_2\theta^2 \leqq (1/(k+1))E_2\theta$.

Thus, for any Bayes estimator g, with squared error loss, for θ, the rule λg is a Bayes rule for θ for each $\lambda \in (0, 1)$, although g is not a linear function of x for any nontrivial prior distribution for θ.

The proof of the above theorem actually implies a stronger conclusion. In fact the λ_i may be replaced, in the statement of the result, by a series of positive functions, $f_i(x)$, such that $f_i(x)g(x)$ is a Bayes estimator for θ, where the functions $f_i(x)$ have the properties

(i) for each x, $\lim_i f_i(x) = 0$;
(ii) for each x, $\lim_i (1/f_i(x)(d^r f_i(x)/dx^r) = 0$, $r = 1, 2, 3$.

Properties (i) and (ii) (which are trivially satisfied for the functions $f_i(x) = \lambda_i$ for all x, $\lambda_i \to 0$) simply express the property that $f_i(x)g(x)$ is a contraction of $g(x)$, and that for large i the contraction is smooth.

Consider, for example, the generalization from estimating the mean of a normal distribution when the prior distribution is normal, to that of estimating the normal mean when the prior distribution is a mixture of normals. For example, if X is a normal random variable, mean θ and variance 1, and the prior distribution for θ is a mixture of two normal distributions both with mean zero and one with precision r_i, one with precision kr_i, then the Bayes estimator for θ, with squared error loss, will be of the form $f_i(x)x$, and if $r_i \to \infty$, it may be checked that the functions $f_i(x)$ have properties (i) and (ii) above.

Thus we have shown that no nonlinear estimates for θ may possess this property and we note this as:

COROLLARY. *Suppose that all the conditions of the theorem above are satisfied, with the exception that the values λ_i are replaced by a sequence of positive functions $f_i(x)$, satisfying properties* (i) *and* (ii) *above, for which $f_i(x)g(x)$ is a Bayes estimator for each i. Then, as before, $g(x) = cx + d$, for all x, for some constants c and d.*

It is natural to enquire whether the very strong constraint (2) in the statement of the above theorem, namely that $g(x)$ should be bounded by a linear function of x, is necessary to derive the required conclusion. We will show that it is sufficient to suppose that $g(x)$ is bounded by polynomials of any order, and to do this we shall require a preliminary lemma, which is a consequence of a theorem of Sampson (1975).

LEMMA. *Let X be a random variable with* pdf *$p(x, \theta)$ of form as given in* (1). *The Bayes estimator for θ given x, with squared error loss, cannot be a polynomial of degree greater than one, for any prior distribution for θ.*

PROOF. Sampson (1975) has shown that if a pdf $p(x, \theta)$, with respect to a general σ-finite measure μ, has form as given in (1), then the function $h(\theta) = E(X \mid \theta)$ cannot be a polynomial of degree greater than one. But, if the likelihood has form (1), then the posterior distribution for given x will be of the form

$$\exp(\theta x + Q(\theta) + R(x))p(\theta)/p(x) = \exp(\theta x + \hat{Q}(\theta) + \hat{R}(x)) ,$$

if the prior distribution is absolutely continuous, or, more generally, will be a pdf from the exponential family with respect to the prior measure for θ. Thus Sampson's result implies that $h(x) = E(\theta \mid x)$ cannot be a polynomial of degree greater than one.

This lemma has the consequence that any attempt to estimate the mean of the posterior distribution for θ by a polynomial in x (see, for example, Goldstein (1975)), will never yield an exact solution unless the mean is linear in x. However, if the mean is linear in x, it is an immediate consequence of equation (5) below that the kth moment of the posterior distribution for θ will be a polynomial of order k, so that in this case polynomial approximation for each moment will be exact.

The above lemma enables us to weaken the bounds imposed on $g(x)$ as follows:

THEOREM. *Suppose that all the conditions of the theorem above are satisfied, with the exception that there exists a positive integer n, and polynomials $A(x)$, $B(x)$ of order n, such that*

$$g(x) \leqq A(x) \quad \text{for} \quad x \geqq 0 \qquad \text{and} \qquad g(x) \geqq B(x) \quad \text{for} \quad x < 0 \,. \tag{4}$$

Then, as before, $g(x) = cx + d$, for all x, for some constants c and d.

PROOF. Differentiating the expression for $E(\theta^r \mid x)$ gives the following relation, for a general prior distribution for θ, namely

$$E(\theta^{r+1} \mid x) = \frac{d}{dx} E(\theta^r \mid x) + E(\theta \mid x) E(\theta^r \mid x) \,. \tag{5}$$

Thus, when the prior distribution for θ is such that $\lambda_i g$ is the Bayes estimator for θ, we have from (5), by induction, that for each integer r

$$E(\theta^{r+1} \mid x) = \lambda_i \frac{d^r g(x)}{dx^r} + \lambda_i^2 p_r(\lambda_i, x) \tag{6}$$

where $p_r(\lambda_i, x)$ is a polynomial in λ_i whose coefficients are derivatives of $g(x)$. As the values λ_i are converging to zero, we have from (6) that for all x and for all nonnegative integers r,

$$\frac{d^{2r+1} g(x)}{dx^{2r+1}} \geqq 0 \,. \tag{7}$$

Thus, for any even integer m, if $d^m g(x)/dx^m$ does not vanish identically, then, from (7), $g(x)$ behaves as x^m for large positive or negative x. Thus $d^{2n} g(x)/dx^{2n}$ must vanish identically, by (4), and so $g(x)$ must be a polynomial. But, by the above lemma, $E(\theta \mid x)$ cannot be a polynomial of degree greater than one, and the result follows.

The following corollary is proved as above.

COROLLARY. *Let X be a random variable with* pdf *of form* (1). *Let g be a Bayes estimator for θ with squared error loss and suppose that there exists a positive integer n, and polynomials $A(x)$ and $B(x)$ of order n, such that relation* (4) *is satisfied.*

Suppose also that there exists a sequence of positive functions $f_i(x)$, *satisfying the properties*

(i) *for each* x, $\lim_i f_i(x) = 0$,

(ii) *for each* x, $\lim_i (1/f_i(x))(d^r f_i(x)/dx^r) = 0$, $r = 1, 2, \cdots, n + 2$,

for which $f_i(x)g(x)$ *is a Bayes estimator for each* i. *Then* $g(x) = cx + d$, *for all* x, *for some constants* c *and* d.

As an illustration, in the example on estimating the mean of a normal distribution with the prior distribution for the mean being a mixture of two normal distributions, the functions $f_i(x)$ satisfy the conditions of the above corollary for any nonnegative integer n.

REFERENCES

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

[2] MEEDEN, G. (1976). A special property of linear estimates of the normal mean. *Ann. Statist.* **4** 649–650.

[3] SAMPSON, ALLAN R. (1975). Characterizing exponential family distributions by moment generating functions. *Ann. Statist.* **3** 747–753.

DEPARTMENT OF MATHEMATICAL STATISTICS
THE UNIVERSITY
HULL, ENGLAND