LINEAR AND QUASI-LINEAR BAYES ESTIMATORS

by

Stephen E. Fienberg

Technical Report #316

Department of Applied Statistics
School of Statistics
University of Minnesota
Saint Paul, MN 55108

April 23, 1978

Summary

Many statistical problems lead to posterior expectations that are linear combinations of the sample mean and the prior mean. Erickson [1969] proposed a characterization of such results for the univariate case. In this paper we discuss the natural multivariate generalization of his results, as well as extensions to quasi-linear posterior expectations resulting from two-stage prior structures. We also consider the applicability of these results for exponential family distributions in a form suggested by Jewell [1975].

1. Introduction

Following up on a series of results available in the statistical literature on posterior means which are linear combinations of a sample mean and a prior mean, Erickson [1969] proposed a simple characterization of such results in the univariate case. Suppose $\underset{\sim}{X}' = (X_1, X_2, \ldots, X_n)$ are jointly distributed random variables conditional on an unknown parameter vector $\underset{\sim}{\theta}$, with common mean, $m(\underset{\sim}{\theta})$, and finite variances $V(X_i|\underset{\sim}{\theta})$. Let have a prior distribution such that $E(m(\underset{\sim}{\theta})) = m$, $0 < V(m(\underset{\sim}{\theta})) < \infty$, and the posterior mean of $m(\underset{\sim}{\theta})$ given $\underset{\sim}{X} = \underset{\sim}{x}$ is of the form:

$$E(m(\underset{\sim}{\theta})|\underset{\sim}{x}) = \alpha\bar{x} + \beta \quad , \tag{1.1}$$

where α and β are independent of $\underset{\sim}{x}$. Then

$$E(m(\underset{\sim}{\theta})|\underset{\sim}{x}) = \frac{\bar{x}V(m(\underset{\sim}{\theta})) + mE_{\underset{\sim}{\theta}}V(\bar{x}|\underset{\sim}{\theta})}{V(m(\underset{\sim}{\theta}) + E_{\underset{\sim}{\theta}}V(\bar{x}|\underset{\sim}{\theta})} \quad . \tag{1.2}$$

Hartigan [1969] derived related results using a "linear" version of Bayes' theorem, where the structure in (1.1) is formally true.

Goldstein [1975], building on Erickson's result, noted that when (1.1) holds, the moments of the prior distribution are uniquely determined, and he gave a recursive expression for their derivation. Kagan, Linnik, and Rao [1973, Addendum B] also discuss linear characterizations in terms of moments of the prior distribution, but they use their characterizations primarily to develop results that hold off both the sampling and prior distributions are normal.

Parallel to the developments in the statistical literature, Jewell [1974a, 1974b, 1975] put foward a series of results on linear Bayes prediction in the context of the actuarial work on credibility theory.

These results provide multivariate generalizations of (1.1) and (1.2), and describe a class of distributions for which posterior means are linear. Diaconis and Ylvisaker [1977] give a rigorous treatment of several of Jewell's results.

In Section 2 we review the multivariate generalization of Erickson's result due to Jewell. Then we go on, in Section 3, to show how the linear structure of the Bayes predictors are modified when the parameters of the prior distribution are themselves viewed as random variables. In the univariate case, such structure leads to replacing α and β in (1.1) by quantities that are functions of the data, $\underset{\sim}{x}$. In Section 4, we consider the special case of exponential family distributions with conjugate priors. This is the most important class of distributions for which the results of Section 2 hold.

2. Multidimensional Linear Bayes Estimators

Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_n$ be independent identically distributed p-dimensional random variables such that

$$E_{\underset{\sim}{X}_i|\underset{\sim}{\theta},\underset{\sim}{\psi}}(\underset{\sim}{X}_i) = \underset{\sim}{m}(\underset{\sim}{\theta}) \quad , \tag{2.1}$$

where $\underset{\sim}{\theta}$ is of dimension s and $\underset{\sim}{\psi}$ is a vector of hyperparameters of dimension q. Moreover, let us denote the variance-covariance matrix of $\underset{\sim}{X}_i$ with respect to this conditional distribution by

$$\underset{\approx}{C}(\underset{\sim}{\theta},\underset{\sim}{\psi}) = \mathrm{Var}_{\underset{\sim}{X}_i|\underset{\sim}{\theta},\underset{\sim}{\psi}}(\underset{\sim}{X}_i) \quad . \tag{2.2}$$

We wish to find the posterior mean of $\underset{\sim}{m}(\underset{\sim}{\theta})$ (given the data and $\underset{\sim}{\psi}$), a parameter which is often of interest in its own right or because it is the mean of the predictive distribution of a new observation $\underset{\sim}{X}_{n+1}$ given the observed values for $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_n$.

Averaging over the conditional distribution of $\underset{\sim}{\theta}$ given $\underset{\sim}{\psi}$ we get the quantities

$$\underset{\sim}{m}^*(\underset{\sim}{\psi}) = E_{\underset{\sim}{\theta}|\underset{\sim}{\psi}}(\underset{\sim}{m}(\underset{\sim}{\theta})) \quad , \tag{2.3}$$

$$\underset{\approx}{E}(\underset{\sim}{\psi}) = E_{\underset{\sim}{\theta}|\underset{\sim}{\psi}}(\underset{\approx}{C}(\underset{\sim}{\theta},\underset{\sim}{\psi})), \tag{2.4}$$

and

$$\underset{\approx}{D}(\underset{\sim}{\psi}) = \mathrm{Var}_{\underset{\sim}{\theta}|\underset{\sim}{\psi}}(\underset{\sim}{m}(\underset{\sim}{\theta})). \tag{2.5}$$

We assume that $\underset{\approx}{D}$ is positive definite so that $\underset{\approx}{D}^{-1}$ exists.

Suppose that the posterior expectation of $\underset{\sim}{m}(\underset{\sim}{\theta})$ given $\underset{\sim}{X}_1 = \underset{\sim}{x}_1$, $\underset{\sim}{X}_2 = \underset{\sim}{x}_2, \ldots, \underset{\sim}{X}_n = \underset{\sim}{x}_n$ has the linear form

$$E_{\theta|X_1,X_2,\ldots X_n,\psi}(m(\theta)) = A\bar{x} + B. \tag{2.6}$$

Then it follows (see Jewell [1974b]) that the posterior mean of θ given ψ and the sample $X_1 = x_1$, $X_2 = x_2$, ..., $X_n = x_n$ is

$$\begin{aligned} E_{\theta|X_1,X_2,\ldots X_n,\psi}(m(\theta)) &= E_{\theta|\bar{X},\psi}(m(\theta)) \\ &= Z\bar{x} + (I-Z)m^*(\psi) , \end{aligned} \tag{2.7}$$

where I is the p×p identity matrix, $\bar{x}$ is the vector of sample means, and Z is defined by

$$Z(E + nD) = nD . \tag{2.8}$$

Alternatively we can define

$$N = ED^{-1} , \tag{2.9}$$

and then Z is given by

$$Z = n(N + nI)^{-1} . \tag{2.10}$$

The linear property exhibited by (2.6) is referred to as <u>exact credibility</u> by those in the actuarial profession and Z is known as the credibility matrix.

Equations (2.6) and (2.7) provide the multivariate generalization of equations (1.1) and (1.2) when we assume that our observations are independent and identically distributed. Clearly they can be further generalized to the situation where X_1, X_2, ..., X_n have a common mean

$\underset{\sim}{m}(\underset{\sim}{\theta})$, but are not necessarily independent given $\underset{\sim}{\theta}$. The linear Bayes estimator is still of the form (2.7), but expression (2.8) for $\underset{\approx}{Z}$ is replaced by

$$\underset{\approx}{Z}(\underset{\approx}{E}^* + \underset{\approx}{D}) = \underset{\approx}{D} \ . \tag{2.11}$$

$$\underset{\approx}{E}^*(\underset{\sim}{\psi}) = E_{\underset{\sim}{\theta}|\underset{\sim}{\psi}}(\underset{\approx}{C}^*(\underset{\sim}{\theta},\psi)) \ , \tag{2.12}$$

and

$$\underset{\approx}{C}^*(\underset{\sim}{\theta},\underset{\sim}{\psi}) = \mathrm{Var}_{\underset{\sim}{X}_1,\underset{\sim}{X}_2,\ldots,\underset{\sim}{X}_n|\underset{\sim}{\theta}\ \underset{\sim}{\psi}}(\overline{\underset{\sim}{X}}) \ . \tag{2.13}$$

For the remainder of the paper, we work with the less general form of independent identically distributed random vectors.

3. Quasi-Linear Bayes Estimators

Expression (2.7) gives the posterior mean of the sampling distribution as a vector of linear combinations of the sample means $\bar{x}$ and the prior mean $m^*(\psi)$, given the hyperparameter ψ. The weights for the linear combination also depend on ψ through the matrices Z, E, and D.

Now let $\psi = (\psi_1, \psi_2)$ where ψ_1 has a density $\phi(\psi_1)$. Then

$$E_{\theta|\bar{X},\psi_2}(m(\theta)) = \int E_{\theta|\bar{X},\psi_1,\psi_2}(m(\theta)) \cdot \phi(\psi_1|X,\psi_2)d\psi$$

$$= E_{\psi_1|\bar{X},\psi_2}\left[Z\right] \cdot \bar{x} + E_{\psi_1|\bar{X},\psi_2}\left[I - Zm^*(\psi)\right] \qquad (3.1)$$

$$= W(x,\psi_2)\bar{x} + \left[I - W(x,\psi_2)\right] \lambda(x,\psi_2) \quad ,$$

where

$$W(\bar{x},\psi_2) = E_{\psi_1|\bar{X},\psi_2} Z(\psi) \quad , \qquad (3.2)$$

and

$$\lambda(\bar{x},\psi_2) = \left[E_{\psi_1|\bar{X},\psi_2}(I - Z(\psi))\right]^{-1} \cdot E_{\psi_1|\bar{X},\psi_2}\left[(I - Z(\psi))m^*(\psi)\right] . \qquad (3.3)$$

Thus we end up with a quasi-linear posterior mean of a form similar to (2.7), but where the new version of the credibility matrix Z is a matrix W which is the average of Z over the distribution of ψ_1 and now depends on the sample data through the mean $\bar{x}$.

If we actually did not know $\phi(\psi_1)$, we might choose to use the marginal likelihood $L(\bar{x}|\psi_1,\psi_2)$ to get a marginal maximum likelihood estimator of ψ_1 for fixed ψ_2, $\hat{\psi}_1 = \hat{\psi}_1(\bar{x},\psi_2)$. Then the compound Bayes estimator of $E_{\theta|\bar{X},\psi_2}(m(\theta))$ has the same form as the estimator in expression (3.1) with W and λ replaced by

$$W^*(\bar{x},\psi_2) = \left[Z(\psi)\right]_{\psi_1=\hat{\psi}_1} , \tag{3.4}$$

and

$$\lambda^*(\bar{x},\psi_2) = \left[(I-Z(\psi))m^*(\psi)\right]_{\psi_1=\psi_1} . \tag{3.5}$$

Examples of such approximations to quasi-linear Bayes estimators are given by Sutherland, Holland, and Fienberg [1975], and by Zellner and Vandaele [1975].

4. Multivariate Exponential Families

A general class of sampling distributions for which the results of Sections 2 and 3 can be shown to hold is the multivariate exponential family with natural parameter $\underset{\sim}{\theta}$. The results are exact in the case where $\underset{\sim}{\theta}$ has a "simple" conjugate prior distribution or an "enriched" conjugate prior distribution. The converse of the linear Bayes result, where we begin with the linear posterior mean (2.6) and ask if we can conclude that $\underset{\sim}{\theta}$ has a conjugate prior (or a mixture of such priors) has been explored by Diaconis and Ylvisaker [1977].

Suppose $\underset{\sim}{X}$ is a p-dimensional random variable from the multivariate exponential family with natural parameter $\underset{\sim}{\theta}$ and density

$$p(\underset{\sim}{x}|\underset{\sim}{\theta}) = \frac{a(\underset{\sim}{x})\ \exp(-\underset{\sim}{\theta}'\underset{\sim}{x})}{c(\underset{\sim}{\theta})} , \tag{4.1}$$

where the normalization constant $c(\underset{\sim}{\theta})$ is found by integrating the numerator of (4.1) with respect to $\underset{\sim}{x}$. We assume that the minimal representation (see Barnsdorff-Nielsen [1973]) for the density is as given in (4.1). Then $\underset{\sim}{x}$ is a minimal sufficient statistic for $\underset{\sim}{\theta}$,

$$\begin{aligned} E(\underset{\sim}{X}|\underset{\sim}{\theta}) &= \underset{\sim}{m}(\underset{\sim}{\theta}) \\ &= -\frac{\partial \log c(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} , \end{aligned} \tag{4.2}$$

and

$$Var(\underset{\sim}{X}|\underset{\sim}{\theta}) = \left\{\frac{\partial^2(-\log c(\underset{\sim}{\theta}))}{\partial\theta_i \partial\theta_j}\right\} = -\frac{\partial \underset{\sim}{m}(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} . \tag{4.3}$$

Example 1. Suppose $\underset{\sim}{X}$ is multivariate normal with mean $\underset{\sim}{\mu}$ and covariance $n^{-1}\underset{\approx}{I}$, i.e.

$$\underset{\sim}{X} \sim \mathcal{N}_p(\underset{\sim}{\mu}, n^{-1}\underset{\approx}{I}) .$$

Then $\underset{\sim}{X}$ is exponential family form with $\underset{\sim}{\theta} = N\underset{\sim}{\mu}$,

$$c(\underset{\sim}{\theta}) = (2\pi)^{p/2} n^{-\frac{1}{2}} \exp\{\tfrac{n}{2}\underset{\sim}{\theta}'\underset{\sim}{\theta}\} ,$$

and $\underset{\sim}{m}(\underset{\sim}{\theta}) = \underset{\sim}{\mu} = -\underset{\sim}{\theta}/n$.

Example 2. Suppose $\underset{\sim}{X}$ is multinomial with sample size n and cell probabilities $\underset{\sim}{\pi}' = (\pi_1, \ldots, \pi_p)$, i.e.

$$\underset{\sim}{X} \sim \mathcal{M}_p(n, \underset{\sim}{\pi}) .$$

Then, ignoring the redundant parameter (since $\sum_i \pi_i = 1$), we again have $\underset{\sim}{X}$ in exponential family form with $\underset{\sim}{\theta}$ defined by

$$\pi_i = \frac{e^{-\theta_i}}{\sum_{j=1}^{p} e^{-\theta_j}} ,$$

and

$$c(\underset{\sim}{\theta}) \propto \left[\sum_{j=1}^{p} e^{-\theta_j}\right]^n .$$

The mean is $\underset{\sim}{m}_1(\underset{\sim}{\theta}) = n\underset{\sim}{\pi}$.

Both examples have been chosen to allow for the interpretation of the components of $\underset{\sim}{X}$ as averages, or sums. Thus we are in effect dealing with a sample of independent identically distributed random variables from a multivariate exponential family.

We wish to characterize the class of prior distributions for the natural exponential family parameter $\underset{\sim}{\theta}$ which yield posterior means for $\underset{\sim}{m}(\underset{\sim}{\theta})$ that are linear in $\bar{\underset{\sim}{x}}$, the average for a sample of size n. A natural starting point for inquiry is the class of "simple" conjugate prior densities

$$\underset{\sim}{u}(\underset{\sim}{\theta}) \propto \left[c(\underset{\sim}{\theta})\right]^{-n_0} \exp(-\underset{\sim}{\theta}'\underset{\sim}{x}_0) , \tag{4.4}$$

with $n_0 > 0$. The posterior density of $\underset{\sim}{\theta}$ for a sample of size n is of the same form as (4.1) with n_0 and $\underset{\sim}{x}_0$ replace by n_0+n and $\underset{\sim}{x}_0 + \sum_{i=1}^{n} \underset{\sim}{x}_i$. If we assume suitable regularity conditions and the like for the density $\underset{\sim}{u}(\underset{\sim}{\theta})$ (see e.g. Jewell [1974b] or Diaconis and Ylvisaker, [1977]), it follows that

$$E_{\underset{\sim}{\theta}|\underset{\sim}{X}}(\underset{\sim}{m}(\underset{\sim}{\theta})) = \frac{n}{n_0+n}\,\bar{\underset{\sim}{x}} + \frac{n_0}{n_0+n}\,\underset{\sim}{m}^{*} , \tag{4.5}$$

where $\underset{\sim}{m}^{*}$ is the prior mean, i.e.

$$\underset{\sim}{m}^{*} = E_{\underset{\sim}{\theta}}\,(\underset{\sim}{m}(\underset{\sim}{\theta})) . \tag{4.6}$$

Thus each component of this posterior mean is the same linear combination of the sample mean and the prior mean. The specialization of this result to the two examples above is well known (e.g. see Sutherland, Holland, and Fienberg, [1975]).

To achieve the full generality of the structure of expression (2.7) we need to go beyond "simple" conjugate priors of the form (4.4) to ones which allow a full prior covariance structure $\underset{\approx}{N}_0$ where we had $n_0\underset{\approx}{I}$ in (4.4). Thus we now take the "enriched" conjugate prior

$$u(\theta) \propto \prod_{i=1}^{p} \left[d_i \left(\sum_h A_{ii}\theta_h \right) \right]^{-n_{i0}} e^{-\theta' x_0} , \tag{4.7}$$

where $n_{i0} > 0$ $(i=1,2,\ldots,p)$, and we assume the existence of an invertible $p\times p$ matrix, $\underset{\approx}{A}$, such that the linear transformation $\underset{\approx}{A}\,\underset{\sim}{X}$ yields a set of p independent random variables, each of which has a univariate exponential distribution, with natural parameters $\underset{\sim}{\phi} = \underset{\approx}{A}'\underset{\sim}{\theta}$ and normalization constants $\{d_i(\phi_i)\}$. This formulation includes as a special case the well-known example of a multivariate normal random vector with known covariance matrix $\underset{\approx}{\Sigma}$.

Jewell [1974b] proves that the posterior mean of $\underset{\sim}{m}(\underset{\sim}{\theta})$ when the prior is (4.7) is given by expression (2.7) where $\underset{\sim}{m}^x = \underset{\sim}{x}_0$,

$$\underset{\approx}{N} = \underset{\approx}{A}\underset{\approx}{N}_0\underset{\approx}{A}^{-1} \tag{4.8}$$

and $\underset{\approx}{N}_0 = \text{diag}(n_{10}, n_{20}, \ldots, n_{p0})$.

By extending the exponential family density in (4.1) to allow for functions of $\underset{\sim}{x}$ as the minimal sufficient statistics, Jewell also shows how we get linear Bayes predictors of the desired form, (2.7). These extensions include the well-known case of a Normal-Wishart prior for independent identically distributed normal random vectors with unknown mean and unknown covariance matrix.

Mixtures of the priors considered in this section also yield linear Bayes estimators.

5. Discussion

The exponential family distributions with suitable conjugate priors and their extensions do not determine a general class of sampling distributions and priors for which the linear structure of (2.7) holds. Diaconis and Ylvisaker [1977] explore the general location parameter problem, and provide an extension to the theorem of Goldstein [1975], which goes well beyond the exponential family class of distributions.

Because the structure of linear predictors is so intuitively appealing, a natural class of extensions worthy of further exploration involves the possible existence of linear Bayes structures similar to (2.7) for exponential response models. In these models the natural parameter vector $\underset{\sim}{\theta} = \underset{\sim}{\beta}\underset{\approx}{y}$, where $\underset{\approx}{y}$ can change from one observation to the next (see e.g. Dempster [1971] and Haberman [1977]). Some of the calculations in Hartigan [1969] seem to be relevant here, and it may be that the posterior linearity for the general exponential response model will be approximate at best.

References

Barndorff-Nielsen, O. [1973]. Exponential Families and Conditioning. Aarhus Universitet, Aarhus, Denmark.

Dempster, A.P. [1971]. An overview of multivariate analysis. J. Multivariate Analysis 1, 316-346.

Diaconis, P. and Ylvisaker, D. [1977]. Priors with linear posterior expectation. Technical Report No. 102, Stanford University.

Erickson, W.A. [1969]. A note on the posterior mean of a population mean. J. Roy. Statist. Soc. (B) 31, 332-334.

Goldstein, M. [1975]. Uniqueness relations for linear posterior expectations. J. Roy. Statist. Soc (B) 37, 402-405.

Haberman, S.J. [1977]. Maximum likelihood estimates in exponential family response models. Am. Statist. 5, 815-841.

Hartigan, J.A. [1969]. Linear Bayes methods. J. Roy. Statist. Soc. (B) 31, 446-454.

Jewell, W.S. [1974a]. Credible means are exact Bayesian for exponential families. Astin Bull. 8, 77-90.

Jewell, W.S. [1974b]. Exact multidimensional credibility. Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker 74 (No. 2), 193-214.

Jewell, W.S. [1975]. Regularity conditions for exact credibility. Astin Bull. 8, 336-341.

Kagan, A.M., Linnik, Y.V., and Rao, C.R. [1973]. Characterization Problems in Mathematical Statistics. Wiley, New York.

Sutherland, M., Holland, P.W., and Fienberg, S.E. [1975]. Combining Bayes and frequency approaches to estimate a multinomial parameter. In S.E. Fienberg and A. Zellner (eds): Studies in Bayesian Econometrics and Statistics. North Holland, Amsterdam, 275-307.

Zellner, A. and Vandaele, W. [1975]. Bayes-Stein estimators for k-means, regression, and simultaneous equation models. In S.E. Fienberg and A. Zellner (eds.): Studies in Bayesian Econometrics and Statistics. North Holland, Amsterdam.