ON IMPLICATIONS OF CREDIBLE MEANS BEING EXACT BAYESIAN[1]

by

Prem K. Goel
Purdue University, W. Lafayette, Indiana

Mimeograph Series #81-4

Department of Statistics
Division of Mathematical Sciences

February 1981

[1] This research was supported in part by the National Science Foundation under Grant #MCS 80-02336.

ON IMPLICATIONS OF CREDIBLE MEANS BEING EXACT BAYESIAN

by

Prem K. Goel, Purdue University

ABSTRACT

In the literature on credibility theory, Jewell showed that the credible means are exact Bayesian for exponential families. In this paper we examine the implications of the statement that credible means are exact Bayesian for certain special forms of distributions. It is conjectured that this statement is valid only for exponential families.

1. INTRODUCTION

As discussed in Bühlman (1970) and Jewell (1974), the usual model of nonlife insurance assumes that each member of risk collective is characterized by an unknown parameter θ, called the risk parameter. For an individual with a given value of θ, the actual risk is a random variable X with probability distribution $p(x|\theta)$, and hence the fair premium of this individual is $m(\theta) = E\{X|\theta\}$. However, θ being unknown, the value of $m(\theta)$ is also unknown. The risk parameter θ is assumed to be a random variable with prior distribution $u(\theta)$, $\theta \in \Theta$. It is usually easy to obtain the information on the collective (marginal) distribution of X. In particular, the collective fair premium $(E(X) = E\{m(\theta)\})$ and collective variance $(\mathrm{Var}(X))$ are assumed to be known.

The actuarial problem is to estimate the fair premium of an individual risk, given the collective distribution, and n years individual experience $\underset{\sim}{X} = (X_1,\ldots,X_n)$, i.e. to estimate $E(X_{n+1}|\underset{\sim}{X})$. Given θ, X_i's are assumed iid $X_i \sim p(x|\theta)$. Since $E(X_{n+1}|\underset{\sim}{X}) = E[m(\theta)|\underset{\sim}{X}]$, the problem is to obtain the posterior expectation of $m(\theta)$ given $\underset{\sim}{X}$. The actuaries call the 'linear Bayes estimator' of $m(\theta)$ by 'credible mean'. Thus the problem of investigating the implications of the statement that credible means are exact Bayesian is exactly similar to that of the linearity of the posterior expectation of an unknown parameter in the observable random variable X.

There are many well known examples in which the posterior expectation of an unknown parameter in the probability distribution of the observations X_i, $i=1,\dots,n$ is a linear function of the observations, and thus for the squared error loss function, the Bayes estimator is linear. Jewell (1974a) showed that $E[m(\theta)|\underset{\sim}{X}]$ is a linear function of X_i's when $p(x|\theta)$ belongs to a regular exponential family and the prior distribution of θ belongs to a conjugate exponential family. Jewell (1974b) extends these results to the case of multidimensional risk vector $\underset{\sim}{X}$. Diaconis and Ylvisaker (1979) also provide a rigorous treatment of these results. Thus for the regular exponential family, credible means are known to be exact Bayesian.

We shall now discuss the implications of the statement

$$E\{m(\theta)|X_1,\dots,X_n\} = b \sum_1^n X_i + a \quad \text{a.e.} \tag{1}$$

for the prior distribution $u(\theta)$ and/or the distribution of the observable risk X given θ, i.e. $p(x|\theta)$.

There are a few papers which characterize the prior distributions $u(\theta)$ for which (1) holds in terms of the distribution $p(x|\theta)$. In the location parameter model, where $X = \theta + \varepsilon$ and θ and ε are independent, Goldstein (1975) proved that if the posterior expectation of θ, given X, is is linear in X then the moments of the $u(\theta)$ are uniquely determined by the moments of the random variable ε. However, in this situation, Rao

(1976) gives an explicit relationship between the characteristic function of ε and θ. In the context of accident proneness models, Johnson (1957) assumes X to be a Poisson random variable with mean θ and shows that (1) implies that $u(\theta)$ is a conjugate gamma distribution. For the Poisson variable X, Johnson (1967) also proves a one-to-one relationship between the prior distribution and $E(\theta|X)$. The most general result in this direction is by Diaconis and Ylvisaker (1979). They show that if the distribution of $\underset{\sim}{X}$ belongs to a regular exponential family, and a multidimensional analog of (1) holds then, under very mild conditions, $\underset{\sim}{\theta}$ has a conjugate prior distribution in the exponential family. They suggest that perhaps one should use linear posterior expectations, i.e. credible means being exact Bayesian, as the defining property for conjugate prior distributions.

The next step in obtaining characterization results in the Bayesian framework is to consider a particular type of model and show that the linear posterior expectation of the unknown parameter implies that $p(x|\theta)$ and $u(\theta)$ belong to some specified family. For example, in the location parameter model, where $X_i = \theta + \varepsilon_i$, $i=1,2,\ldots,n$ $(n \geq 2)$, ε_i's are independent random variables and θ is independent of ε_i's, along with the assumption that both ε_i's and θ have finite second moments, Kagan and Karpov show that [see Kagan, Linnik and Rao (1973) p.418, hereafter referred to as KLR], if θ has a linear posterior expectation given $X_1,\ldots,X_n$, then $p(x|\theta)$ and $u(\theta)$ are both normal distributions.

Rao (1976) and Goel and DeGroot (1980) consider the linear regression model $\underset{\sim}{X} = A\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$, where $\underset{\sim}{\beta}$ and $\underset{\sim}{\varepsilon}$ are independently distributed random variables. Under various conditions on the matrix A, they show that only normal distributions for $\underset{\sim}{\beta}$ and $\underset{\sim}{\varepsilon}$ have linear posterior expectation of $\underset{\sim}{\beta}$ given $\underset{\sim}{Y}$. Thus linear posterior expectations in linear regression provide a characterization of the normal distribution. In the actuarial terminology, these results imply that if one uses credible means in a location parameter model or in a linear regression credibility model, and believes that they are exact Bayesian, then he is implicitly assuming that the observations and the parameter are all normally distributed. Jewell (1974b) and Sundt (1979) mention the use of regression models in credibility theory and provide the best linear credibility estimators.

2. CREDIBLE MEANS FOR SCALE PARAMETER MODELS.

In this section, we consider the scale parameter model

$$X_i = \theta\varepsilon_i, \quad i=1,2,\dots,n \tag{2}$$

where, given θ, X_i's are independently distributed positive random variables and θ is a positive random variable, independent of the ε_i's. In the classical statistical terminology, θ is a scale parameter in the distribution of X_i's. It is well known [see DeGroot (1970), Jewell (1974a)] that if $p(x|\theta)$ is a gamma distribution and the prior distribution of $\beta = 1/\theta$ belongs to a conjugate gamma family, then the posterior expectation of θ given $X_1,\dots,X_n$ is of the form

$$E[\sigma|X_1,\dots,X_n] = b\sum_1^n X_i + a, \quad \text{a.e.}, \tag{3}$$

and hence (1) holds. In order to give a more general result, we shall not assume that given θ, X_i's are iid in the following theorem.

<u>Theorem 1.</u> Suppose that in the model (2), ε_i and σ are independent random variables with finite first moments and $\varepsilon_1,\dots,\varepsilon_n$ $(n \geq 2)$ are independent, positive non-degenerate random variables. If the support of the prior distribution of θ contains an open interval and

$$E[\sigma|X_1,\ldots,X_n] = \sum_1^n b_i X_i + a, \quad \text{a.e.} \tag{4}$$

holds then the distributions of ε_i's and the prior distribution of $\beta = 1/\theta$ must belong to the family of gamma distributions.

<u>Proof.</u> The model (2) can be expressed as

$$Y_i = \eta + e_i, \quad i=1,2,\ldots,n \ , \tag{5}$$

where $Y_i = \ell n\, X_i$, $\eta = \ell n\, \theta$ and $e_i = \ell n\, \varepsilon_i$. Now (4) is equivalent to

$$E[\exp(\eta)|X_1,\ldots,X_n] = \sum_1^n b_j X_j + a \ ,$$

or, equivalently

$$E[\exp(\eta) - \sum_1^n b_j X_j | Y_1,\ldots,Y_n] = a. \tag{6}$$

It follows from (6) and Lemma 1.1.1 in KLR, page 11, that

$$E\{[\exp(\eta) - \sum_1^n b_j X_j]\exp(i \sum_1^n t_j Y_j)\} = aE[\exp(i \sum_1^n t_j Y_j)] \ . \tag{7}$$

On substituting for X_j and Y_j in terms of η and e_j and using the mutual independence of $\eta, e_1,\ldots,e_n$, (7) can be written as

$$\varphi_0(\sum_1^n t_j - i)\prod_1^n \varphi_j(t_j) - \varphi_0(\sum_1^n t_j - i)[\sum_{j=1}^n b_j \varphi_j(t_j - i)\prod_{k\neq j} \varphi_k(t_k)] = \tag{8}$$

$$a\varphi_0(\sum_1^n t_j)\prod_1^n \varphi_j(t_j),$$

where $\varphi_0(t) = E[\exp(it\eta)]$ and $\varphi_j(t) = E[\exp(ite_j)]$, $j=1,\ldots,n$. Note that $|\varphi_j(t-i)| < \infty$, since θ and $\varepsilon_1,\ldots,\varepsilon_n$ have been assumed to have finite first moments.

For some $\delta > 0$, and all $|t_j| < \delta$, (8) implies that

$$\sum_{j=1}^{n} b_j\xi_j(t_j) + a\psi(\sum_1^n t_j) = 1 \ , \tag{9}$$

where $\psi(t) = \varphi_0(t)/\varphi_0(t-i)$ and $\xi_j(t) = \varphi_j(t-i)/\varphi_j(t)$, $j=1,\ldots,n$. Since ψ and ξ_j are continuous functions, it follows from the Corollary to Lemma 1.5.1 in KLR and (9) that

$$b_j\xi_j(t) = c_j + \gamma t \quad \text{for all } |t| < \delta, \quad j=1,\ldots,n \ . \tag{10}$$

If F_j is the distribution function of e_j, then (10) is equivalent to

$$\int \exp[e_j(1+it)]dF_j(e_j) = \frac{1}{b_j}(c_j+\gamma t)\int \exp(ite_j)dF_j(e_j), \tag{11}$$

for all $|t| < \delta$ and $j=1,\ldots,n$. It follows from Lemma 6.1.2 in KLR that (11) holds for all t. Now Corollary to Lemma 6.1.6 in KLR implies that the probability distribution of ε_j is a gamma distribution $G(c_j/\gamma,\gamma/b_j)$. Therefore, given θ, $X_1,\ldots,X_n$ are independent gamma random variables and the distribution of the sufficient statistic $\sum_1^n b_iX_i$ belongs to a gamma family with scale parameter σ. Hence $E[\sum_1^n b_iX_i|\sigma] = c\sigma$, where c is a constant. Since (4) implies that $E[\sigma|\sum_1^n b_iX_i] = \sum_1^n b_iX_i + a$, it follows from Theorem 3 in Diaconis and Ylvisaker (1979) that the prior distribution of $1/\theta$ must also be a gamma distribution.

Remark 1. If in (4) all the b_i's are equal, then each ε_i has a gamma distribution with the same scale, but they don't have to be identically distributed. However, if ε_i are identically distributed then all b_i's are equal.

Remark 2. Sections 7.12 and 7.13 in KLR give various theorems on the admissibility of $\sum_1^n b_iX_i$, for squared error loss, in the class of all unbiased estimators of θ implying that X_i, given θ are gamma variables. However they assume that $E(X_i^2) < \infty$.

Remark 3. Theorem 8.5.4 in KLR, page 285, shows that the sufficiency of $\bar{X} = \frac{1}{n}\sum_{1}^{n} X_i$ for the family of scale parameter models, under certain conditions, implies that X_i's have a gamma distribution. However, the random variable X, given θ, has been assumed to have all the moments (see Lemma 8.5.1, KLR p. 285). Furthermore, the sufficiency of $\bar{X}$ does not imply (3) and (3) does not imply the sufficiency of $\bar{X}$ in general.

Theorem 1 shows that in the scale parameter credibility models, credible means are exact Bayesian implies that $p(x|\theta)$ is a gamma distribution and that the $u(\theta)$ is an inverted gamma distribution (i.e., $1/\theta$ has apriori a gamma distribution).

3. CONCLUDING REMARKS

Given θ, let $X_1,\ldots,X_n$ ($n \geq 2$) be iid random variables such that the cdf of X_i belongs to a family $\mathfrak{F} = \{F_\theta(\cdot), \theta \in \Theta\}$ where $\Theta \subset R$. This rules out the family of distributions defined by Dirichlet Process prior, since it cannot be indexed by a real valued parameter θ. Furthermore, let $E[X_i|\theta] = m(\theta)$ be finite. If (1) holds and $m(\theta)$ is a location or a scale parameter in this family, then it follows from the discussion in Sections 1 and 2 that $\bar{X}$ is the minimal sufficient statistic for the family F (since $p(x|\theta)$ is either normal or gamma). In general, if the posterior expectation of $m(\theta)$ depends only on $\bar{X}$, it is clear that $\bar{X}$ is a function of the minimal sufficient statistic for $\mathfrak{F}$. However, an interesting open problem is "Does (1) imply that $\bar{X}$ is a sufficient statistic itself for $\mathfrak{F}$?" If the answer is yes then it will also imply that $\mathfrak{F}$ is a regular exponential family. Thus, from Diaconis and Ylvisaker (1979) result, $u(\theta)$ is also in the conjugate exponential family. After searching the literature on Bayesian estimation extensively, we have not found any example of a non-exponential family distribution $p(x|\theta)$ in which $E[m(\theta)|X_1,\ldots,X_n]$ is a linear function of $\bar{X}$. So we believe that the following conjecture is valid.

CONJECTURE: If (1) holds, then the distribution of X_i's belongs to a regular exponential family. In the credibility language, credible means are exact Bayesian only for regular exponential families.

REFERENCES

Bühlman, H. (1970) Mathematical Methods in Risk Theory, Springer-Verlag, New York.

DeGroot, Morris H. (1970) Optimal Statistical Decisions. McGraw-Hill, New York.

Diaconis, P. and Ylvisaker, D. (1979) Conjugate priors for exponential families. Ann. Statist., 7, 269-281.

Ericson, W. A. (1969) A note on the posterior mean of a population mean. Jour. Roy. Statist. Soc., B, 31, 332-334.

Goel, Prem K. and DeGroot, Morris H. (1980) Only normal distributions have linear posterior expectation in linear regression. Jour. American Statist. Assoc., 75, 895-900.

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

Jewell, W. S. (1974a) Credible means are exact Bayesian for exponential families. ASTIN Bull. 8, 77-90.

Jewell, W. S. (1974b) Exact multidimensional credibility. Sonderabdruck aus den Mittelungen der Ver. Schw. Versicherungsmathematiker, #2, Druck Stämpflit + Cie AG, Bern.

Jewell, W. S. (1975) Bayesian Regression and Credibility Theory. RM-75-63. International Institute of Applied Systems Analysis, Laxenburg, Austria.

Johnson, N. L. (1957) Uniqueness of a result in the theory of accident proneness. Biometrika, 44, 530-531.

Johnson, N. L. (1967) Note on a uniqueness relation in certain accident proneness models. Jour. Amer. Stat. Assoc., 62, 288-289.

Kagan, A. M., Linnik, Yu. V. and Rao, C. R. (1973) Characterization Problems in Mathematical Statistics. John Wiley, New York.

Rao, C. R. (1976) Characterization of prior distributions and solution to a compound decision problem. Ann. Statist., 4, 823-835.

Sundt, B. (1979) A hierarchical credibility regression model. Scand. Actuarial J., 107-114.

Prem K. Goel
Department of Statistics
Purdue University
W. Lafayette, IN. 47907