A Note on the Posterior Mean of a Population Mean

By W. A. Ericson

University of Michigan

[Received January 1968]

SUMMARY

It is a well-known result, see for example Lindley (1965) and Raiffa and Schlaifer (1961), that if $\bar{x}$ is the mean of a sample of independent observations distributed $N(\mu, \sigma^2)$ where σ^2 is known, and if μ has been assigned a normal prior distribution, $N(m, v)$, then the posterior expectation of μ, given the sufficient statistic $\bar{x}$, has the form $\{\bar{x}(n/\sigma^2) + m/v\}/\{(n/\sigma^2) + 1/v\}$, that is, has the intuitively appealing form of a weighted average of the prior mean and sample mean with weights inversely proportional to the prior variance and the conditional sampling variance of $\bar{X}$ respectively. In this note we present an extremely simple generalization of this result which seems not to have been noted explicitly before and which holds for a variety of distributions other than the normal.

1. INTRODUCTION

VERY often inference concerns a population mean or the common conditional expectation of the observations given some parameter. In this note we show that if a simple and easily checked condition holds then the form of the posterior mean of a population mean is a weighted average of the observed sample mean, $\bar{x}$, and the prior expectation of the population mean, m say, with weights respectively proportional to the inverses of the *prior expectation of the conditional sampling variance* of $\bar{X}$ and the prior variance of the population mean.

This basic result is given in Section 2. In Section 3 we give a simple example and also indicate that the condition holds for a wide variety of distributions other than normal with known variance. The well-known normal result turns out to be the simplest special case of an equally appealing, but more general, form. Some generalizations are also indicated very briefly in that section.

2. BASIC RESULT

Suppose observations $\mathbf{X} = (X_1, X_2, ..., X_n)$ have a joint distribution conditional on some unknown parameter $\boldsymbol{\theta} = (\theta_1, ..., \theta_s)$. The X_i's are not necessarily independent given $\boldsymbol{\theta}$ but do share common means denoted by

$$E(X_i | \boldsymbol{\theta}) = \mu(\boldsymbol{\theta}), \quad i = 1, 2, ..., n. \tag{1}$$

A joint prior distribution is assigned $\boldsymbol{\theta}$ such that the variance of the population mean, $\mu(\boldsymbol{\theta})$, exists. This variance is denoted by $V\{\mu(\boldsymbol{\theta})\}$. The prior expectation of $\mu(\boldsymbol{\theta})$ is denoted by

$$E_{\boldsymbol{\theta}}\{\mu(\boldsymbol{\theta})\} \equiv m. \tag{2}$$

Finally, we let $\bar{X}$ denote the mean of a sample of X_i's ($\bar{X} = n^{-1}\sum X_i$) and assume that the conditional (on $\boldsymbol{\theta}$) variance of $\bar{X}$, $V(\bar{X}|\boldsymbol{\theta})$, exists. Under these assumptions and

definitions a sufficient condition that the posterior expectation (given $\mathbf{X} = \mathbf{x}$) can be expressed in the weighted average form indicated above is that $E\{\mu(\boldsymbol{\theta})|\mathbf{x}\}$ be a linear function of $\bar{x}$, the observed sample mean. This condition is usually very simply checked whereas to show directly that the weighted average form holds often involves tedious manipulation. The basic result may be stated more formally as follows:

Theorem. If $\mathbf{X} = (X_1, ..., X_n)$ are any jointly distributed random variables conditional on an unknown parameter $\boldsymbol{\theta}$, which have common means, $\mu(\boldsymbol{\theta})$, and finite variances, $V(X_i|\boldsymbol{\theta}) < \infty$, and if a joint prior distribution is assigned to $\boldsymbol{\theta}$ such that $0 < V\{\mu(\boldsymbol{\theta})\} < \infty$ and if the posterior expectation of $\mu(\boldsymbol{\theta})$ given $\mathbf{X} = \mathbf{x}$ is given by

$$E\{\mu(\boldsymbol{\theta})|\mathbf{x}\} = \alpha\bar{x} + \beta, \qquad (3)$$

where α and β do not depend on $\mathbf{x}$, then

$$E\{\mu(\boldsymbol{\theta})|\mathbf{x}\} = \frac{\bar{x} V\{\mu(\boldsymbol{\theta})\} + m E_{\boldsymbol{\theta}} V(\bar{x}|\boldsymbol{\theta})}{V\{\mu(\boldsymbol{\theta})\} + E_{\boldsymbol{\theta}} V(\bar{x}|\boldsymbol{\theta})}. \qquad (4)$$

Proof. We first note that $E(\bar{X})_i = E_{\boldsymbol{\theta}} E(\bar{X}|\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\{\mu(\boldsymbol{\theta})\} = m$ and hence

$$m = E\{\mu(\boldsymbol{\theta})\} = E_{\mathbf{X}} E\{\mu(\boldsymbol{\theta})|\mathbf{X}\} = \alpha E(\bar{X}) + \beta = \alpha m + \beta, \qquad (5)$$

from which it follows that

$$\beta = (1-\alpha)m. \qquad (6)$$

Then equating

$$E\{\bar{X}\mu(\boldsymbol{\theta})\} = E_{\boldsymbol{\theta}}\{\mu(\boldsymbol{\theta}) E(\bar{X}|\boldsymbol{\theta})\} = E_{\boldsymbol{\theta}}\{\mu(\boldsymbol{\theta})\}^2 = V\{\mu(\boldsymbol{\theta})\} + m^2$$

and

$$E\{\bar{X}\mu(\boldsymbol{\theta})\} = E_{\mathbf{X}}[\bar{X}E\{\mu(\boldsymbol{\theta})|\mathbf{X}\}] = E_{\mathbf{X}}\{\alpha \bar{X}^2 + (1-\alpha)m\bar{X}\} = \alpha V(\bar{X}) + \alpha m^2 + (1-\alpha)m^2,$$

we have

$$\alpha = V\{\mu(\boldsymbol{\theta})\}/V(\bar{X}). \qquad (7)$$

The result (5) and existence of α follows immediately by noting that

$$V(\bar{X}) = E_{\boldsymbol{\theta}} V(\bar{X}|\boldsymbol{\theta}) + V_{\boldsymbol{\theta}} E(\bar{X}|\boldsymbol{\theta}) = E_{\boldsymbol{\theta}} V(\bar{X}|\boldsymbol{\theta}) + V\{\mu(\boldsymbol{\theta})\} > 0,$$

and substituting (6) and (7) and the above in (3).

3. EXAMPLES AND DISCUSSION

To illustrate the theorem suppose $X_1, X_2, ..., X_n$ are independent, identically distributed Poisson random variables having unknown parameter θ. Then $E(x_i|\theta) = \theta$ and $V(\bar{X}|\theta) = \theta/n$. Furthermore if θ is assigned a gamma prior distribution with density proportional to $\theta^{x'-1} e^{-n'\theta}$ for some $n' > 0$, $x' > 0$ then it is well known that the posterior distribution on θ given $X_i = x_i$, $i = 1, ..., n$ is also gamma with density proportional to $\theta^{n\bar{x}+x'-1} e^{-(n+n')\theta}$ and $E(\theta|x_1 ... x_n) = (n\bar{x} + x')/(n+n')$. Thus the posterior mean is linear in $\bar{x}$ and this posterior mean must have the interpretation or form (5), which seems not to have been previously recognized. This form is of course directly verifiable by noting that $E(\theta) = x'/n'$, $E_{\theta} V(\bar{X}|\theta) = E_{\theta}(\theta/n) = x'/(nn')$ and $V(\theta) = x'/n'^2$.

It is then readily checked that the following simple cases are among the many in which the conditions of the theorem hold and hence the form (4) persists: take the X_i's

to be conditionally independent normal (unknown mean and variance) or binomial, negative binomial, gamma or exponential and assign their unknown parameters a natural conjugate prior (see Raiffa and Schlaifer, 1961).

While we have not determined a general class of sampling distributions and priors for which (3) holds, it is clear that natural conjugate priors are not required, for, if (3) holds using any natural conjugate prior it must hold for mixtures of such distributions which, in turn, are not conjugate priors.

Finally, it can be shown that under a Bayesian model for sampling finite populations and under similar conditions to those of Section 2 the posterior expectation of the finite population mean also has the form (4). This model is discussed in detail by the author (Ericson, 1969).

REFERENCES

ERICSON, W. A. (1969). Subjective Baysian models in sampling finite populations (with Discussion). *J. R. Statist. Soc.* B, **31**, 195–233.

LINDLEY, D. V. (1965). *Introduction to Probability and Statistics.* Part 2: *Inference.* Cambridge University Press.

RAIFFA, H. and SCHLAIFER, R. O. (1961). *Applied Statistical Decision Theory.* Graduate School of Business Administration, Harvard University.