Journal of the Research Institute of Industrial Technology, Vol. 18, 2001

Bayesian Estimation of the Variance of a Normal Distribution

C. I. Park[1)], H. J. Kim[2)]

Abstract

Some new general forms of estimators of the variance of a normal distribution are derived using Bayesian methods, and the conditions under which they lead to previously proposed estimators are discussed.

1. Introduction

Let $x_1, x_2, \ldots, x_n$ be a random sample of n observations from a normal distribution whose mean μ and variance ω are both unknown. Several criteria of estimation have led to estimators of ω of the form $S/d(n)$, where $S=\sum(x_i-\bar{x})^2$, the most frequently occurring ones being the maximum-likelihood estimator with $d(n)=n$ and the unbiased estimator with $d(n)=(n-1)$. Other values of $d(n)$ have been derived by Goodman (1960) by determining estimators which, in the sense defined by Lehmann(1959), are unbiased with respect to certain loss functions. If $x_1, x_2, \ldots, x_n$ is a random sample from a distribution $p(x \mid \theta)$ depending on an unknown parameter θ, then an estimator $f(x_1, \ldots, x_n)$ of θ is defined to be unbiased with respect to the loss function $L\{f(x_1, \ldots, x_n), \theta\}$ if, for each θ,

$$E[L\{f(x_1, \ldots, x_n), \theta'\} \mid \theta] = \int L\{f(x_1, \ldots, x_n), \theta'\} \Pi p(x_i \mid \theta)\, dx_i$$

is a minimum when $\theta' = \theta$.

The purpose of this paper is to show how Bayesian methods lead to a class of estimators of ω, which includes the aforementioned ones as special cases.

2. Bayesian methods

Let x be a random variable whose distribution depends on k parameters $\theta_1, \theta_2, \ldots, \theta_k$ and let Ω denote the parameter space of possible values of θ, the k-dimensional vector $(\theta_1, \ldots, \theta_k)$. We now consider the general problem of estimating some specified real-valued function, $\gamma(\theta)$, of the unknown parameters θ, from the results of a random sample of n observations ; we shall assume that $\gamma(\theta)$ is defined for all θ in Ω.

Denoting the sample results $x_1, \ldots, x_n$ by x, let $f(\mathrm{x})$ be an estimate of $\gamma(\theta)$ and let $L\{f(\mathrm{x}), \gamma\}$ be the loss incurred by taking the value of γ to be $f(\mathrm{x})$. It should be noted that we

1) Department of Applied Mathematics, Korea Maritime University
2) Division of International Trade & Economics, Korea Maritime University

are restricting consideration here to loss functions which depend on θ through $\gamma(\theta)$ only.

If $\pi(\theta)$ is the prior density of θ, then by Bayes's theorem the posterior density of θ is $p(x \mid \theta)\pi(\theta)/p(x)$, where $p(x \mid \theta)$ is the likelihood of the sample results given θ, and

$$p(x) = \int_{\Omega} p(x \mid \theta)\, \pi(\theta)\, d\theta.$$

It follows that, for a given x, the expected loss of the estimator f is

$$\int_{\Omega} L\{f(x), \gamma(\theta)\} \frac{p(x \mid \theta)\pi(\theta)}{p(x)}\, d\theta. \qquad (1)$$

Following Lindley(1960), we define an *optimum estimator* of γ to be a function f for which (1) exists and is minimal. This definition has the advantage of enabling the optimum estimator to be determined for each set of observations x independently of any other set which may have been observed.

Assuming the existence of (1) and that sufficient regularity conditions prevail to permit differentiation under the integral sign, then the optimum estimator f or γ will be a solution of the equation

$$\int_{\Omega} \frac{\partial L}{\partial f} p(x \mid \theta)\, \pi(\theta)\, d\theta = 0. \qquad (2)$$

The validity of (2) and the desirability that it should lead to a unique solution necessarily impose restrictions on one's choice of loss function and prior density of θ.

It is evident form (2) that, for the squared-error loss function $L(f, \gamma) = c(f-\gamma)^2$, the optimum estimator is simply the mean of the posterior distribution of $\gamma(\theta)$.

3. Application to a normal distribution

Consider the case of estimating the variance ω of a normal distribution of unknown mean μ. Here, $\theta = (\mu, \omega)$, Ω is the half-plane: $-\infty < \mu < \infty$, $0 < \omega < \infty$, and

$$\begin{aligned} p(x \mid \theta) &= (2\pi\omega)^{-n/2}\exp\{-\textstyle\sum(x_i-\mu)^2/2\omega\} \\ &= (2\pi\omega)^{-n/2}\exp[-\{S+n(\mu-\bar{x})^2\}/2\omega], \end{aligned} \qquad (3)$$

where $S = \sum(x_i - \bar{x})^2$. A mathematically convenient and widely applicable joint prior density for the problem under consideration is the class of natural conjugates

$$\pi(\mu, \omega) \propto \omega^{-(1+\nu/2)} \exp[-\{\eta + \zeta(\mu-\xi)^2\}/2\omega], \qquad (4)$$

where $\nu, \eta, \zeta \geq 0$ and $-\infty < \xi < \infty$. This is obtained by generalizing the likelihood (3) regarded as a function of the unknown parameters, and here is seen to be equivalent to assuming that the prior marginal density of ω is such that η/ω is distributed as χ^2 with $(\nu-1)$ degrees of freedom and that the prior conditional density of μ given ω is normal with mean μ and variance ω/ζ. The advantage of taking the prior distribution to be the natural conjugate lies in the fact that the likelihood $p(x \mid \mu, \omega)$, the prior density $\pi(\mu, \omega)$ and the posterior density $\pi(\mu, \omega \mid x)$ are all of the same functional form, thus ensuring mathematical tractability. Raiffa and Schlaifer (1961, Chapter 3 and 11) explained comprehensive account of natural conjugate prior distribution.

For the limiting case, when $\eta = \zeta = 0$ we have the subclass of prior densities given by

$$\pi(\mu, \omega) \propto \omega^{-(1+\nu/2)}, \tag{5}$$

which is equivalent to assuming the prior distributions of μ and ω to be independent, that of μ being uniform and that of ω being proportional to $\omega^{-(1+\nu/2)}$. The particular form of (5) corresponding to $\nu = 0$ is precisely the prior distribution advocated by Jeffreys (1961, see Section 3.1) when one is completely ignorant of the values of μ and ω apart, of course, from their admissible ranges.

Substituting from (3) and (4) in (2), the optimum estimator f of ω is a solution of

$$\int\int \frac{\partial L(f,\omega)}{\partial f}\,\omega^{-(n+\nu+2)/2}\exp[-\{S+n(\mu-\bar{x})^2+\zeta(\mu-\xi)^2+\eta\}/2\omega]\,d\mu\,d\omega=0,$$

the integration being over Ω : $-\infty < \mu < \infty$, $0 < \omega < \infty$. On noting that $L(f,\omega)$ is independent of μ and that

$$\int_{-\infty}^{\infty}\exp[-\{n(\mu-\bar{x})^2+\zeta(\mu-\xi)^2\}/2\omega]d\mu=\{2\pi\omega/(n+\zeta)^{\frac{1}{2}}\}\exp\{-n\zeta(\xi-\bar{x})^2/2(n+\zeta)\omega\},$$

we find that f is a solution of

$$\int_0^{\infty}\frac{\partial L(f,\omega)}{\partial f}\,\omega^{-(n+\nu+1)/2}\exp(-K/2\omega)d\omega=0, \tag{6}$$

where $K=S+\eta+\zeta(\xi-\bar{x})^2/(n+\zeta)$.

Of the six loss functions considered by Goodman (1960), four are particular cases of the form

$$L(f,\omega)=c\omega^a(f^b-\omega^b)^2, \tag{7}$$

where c is a positive constant. For the loss function given by (7), it follows from (6) that the optimum estimator f is given by

$$f^b=\frac{\int_0^{\infty}\omega^{a+b-(n+\nu+1)/2}\exp(-K/2\omega)d\omega}{\int_0^{\infty}\omega^{a-(n+\nu+1)/2}\exp(-K/2\omega)d\omega}$$
$$=(K/2)^b\,\Gamma\{\tfrac{1}{2}(n+\nu+1)-a-b-1\}/\Gamma\{\tfrac{1}{2}(n+\nu+1)-a-b-1\}.$$

Hence,

$$f=\frac{S+\eta+\zeta(\xi-\bar{x})^2/(n+\zeta)}{2\,[\Gamma\{\frac{1}{2}(n+\nu+1)-a-1\}/\Gamma\{\frac{1}{2}(n+\nu+1)-a-b-1\}]^{1/b}} \tag{8}$$

We note that this is a more general form than the proposed estimators $S/d(n)$; there appears to be no record in the literature of any estimator of ω having a correction term added to the usual numerator S in order to allow for prior knowledge of the mean μ. We note further, that this correction term vanishes if and only if is the prior distribution of μ is assumed to be uniform independently of ω.

Using the standard asymptotic expansion of the gamma function, it follows that for large N,

$$\frac{\Gamma(N+m)}{\Gamma(N)}\sim N^m\left(1+\frac{m^2-m}{2N}\right),$$

from which we deduce that, for large n, the denominator of (8) is approximately $n+\nu-2a-b-2$. Thus,

$$f \sim \frac{S+\eta+\zeta(\xi-\bar{x})^2/(n+\zeta)}{n+\nu-2a-b-2}.$$

The remaining two loss functions considered by Goodman(1960) were

$$L_5 = c(\log f - \log \omega)^2,$$

and

$$L_6 = \begin{cases} 0 & \text{if } |f-\omega| \le c\omega, \\ 1 & \text{if } |f-\omega| > c\omega. \end{cases}$$

Substituting L_5 in (6), the optimum estimator of ω in this case is

$$f = \frac{S+\eta+\zeta(\xi-\bar{x})^2/(n+\zeta)}{2\exp[\Psi\{(n+\nu-1)/2\}]}$$

where $\Psi(m) = d\log\Gamma(m)/dm$. The particular form of $d(n)$ obtained here was first proposed by Lendley(1953), his numerator being the customary S. Since, for large N,

$$\Psi(N) \sim \log N - \frac{1}{2N}$$

we have for large n,

$$f \sim \frac{S+\eta+\zeta(\xi-\bar{x})^2/(n+\zeta)}{n+\nu-2}.$$

Corresponding to L_6, the optimum estimator of ω is a solution of

$$\frac{\partial}{\partial f}\left\{\int_0^\infty L_6(f,\omega)\omega^{-(n+\nu+1)/2}\exp(-K/2\omega)d\omega\right\} = 0,$$

i. e. of

$$\frac{\partial}{\partial f}\left\{\int_{-f/(1+c)}^{f/(1+c)} \omega^{-(n+\nu+1)/2}\exp(-K/2\omega)\,d\omega\right\} = 0 \;;$$

it readily follows from this equation that

Loss function	$d(n)$	
	Here	Goodman
$L_1 = c(f-\omega)^2$	$n+\nu-3$	$n-1$
$L_2 = c(f-\omega)^2/\omega^2$	$n+\nu+1$	$n+1$
$L_3 = c(\sqrt{f}-\sqrt{\omega})^2$	$2\Gamma^2\{\frac{1}{2}(n+\nu-1)\}/\Gamma^2\{\frac{1}{2}(n+\nu-2)\}$ $\sim n+\nu-\frac{5}{2}$	$2\Gamma^2(\frac{1}{2}n)/\Gamma^2\{\frac{1}{2}(n-1)\}$ $\sim n-\frac{3}{2}$
$L_4 = c(\sqrt{f}-\sqrt{\omega})^2/\omega$	$2\Gamma^2\{\frac{1}{2}(n+\nu+1)\}/\Gamma^2\{\frac{1}{2}(n+\nu)\}$ $\sim n+\nu-\frac{1}{2}$	$2\Gamma^2\{\frac{1}{2}(n+1)\}/\Gamma^2(\frac{1}{2}n)$ $\sim n-\frac{1}{2}$
$L_5 = c(\log f-\log\omega)^2$	$2\exp[\Psi\{\frac{1}{2}(n+\nu-1)\}]$ $\sim n+\nu-2$	$2\exp[\Psi\{\frac{1}{2}(n-1)\}]$ $\sim n-2$
$L_6 = \begin{cases} 0 & \text{if } \|f-\omega\| \le c\omega \\ 1 & \text{if } \|f-\omega\| > c\omega \end{cases}$	$\frac{n+\nu-1}{2c}\log\left(\frac{1+c}{1-c}\right)$	$\frac{n-1}{2c}\log\left(\frac{1+c}{1-c}\right)$

$$f = \frac{S+\eta+\zeta(\xi-\bar{x})^2/(n+\zeta)}{\{(n+\nu-1)/2c\}\log\{(1+c)/(1-c)\}}$$

TABLE 1 *Comparison of the values of d(n) obtained here and by Goodman (1960)*

4. Conclusion

It is evident from the above table that, in every case considered, Goodman's "unbiased" estimator corresponds to assigning particular values to the parameters in $\pi(\mu,\omega)$: for identical numerators it is necessary to take $\eta=\zeta=0$. There is no one value of ν which may be taken for the denominators to be the same. We note, however, that for the loss functions L_2, L_4, L_5, L_6, Goodman's estimators are those derived here with $\eta=\zeta=\nu=0$, which are precisely the values to take, according to Jeffreys, when one has no prior knowledge of μ and ω apart from their independence.

It is not at all obvious why the optimum estimator as defined here should include Goodman's estimators as special cases. For, the optimum estimator is the one that minimizes

$$\underset{\mu,\omega}{E}[L\{f(\mathrm{x}),\omega\}]$$

with respect to the posterior distribution of μ and ω; on the other hand, the estimator f is unbiased with respect to the loss function $L\{f(\mathrm{x}),\omega\}$ if, for each ω,

$$\underset{\mathrm{x}}{E}[\{f(\mathrm{x}),\omega'\}\mid\omega]$$

is a minimum when $\omega'=\omega$. The former thus involves integration over the parameter space, whereas the latter involves integration over the sample space. We shall now show how these integrals are related when the prior density of μ and ω is taken to be the natural conjugate density. It follows directly from (6) that the optimum estimator is the one that minimizes

$$\underset{\omega}{E}\{L(f,\omega)\}$$

with respect to the marginal posterior distribution of ω, which, from the work leading to (6), is readily seen to be such that K/ω is distributed as χ^2 with $(n+\nu-1)$ degrees of freedom. Further, restricting consideration of "unbiased" estimators to those which depend on x through S only, $f(\mathrm{x})=\mathrm{g}(\mathrm{S})$ will be unbiased with respect to the loss function $L(f,\omega)$ if

$$\underset{\mathrm{S}}{E}[L\{g(S),\omega'\}\mid\omega]$$

is a minimum when $\omega'=\omega$, where S/ω is distributed as χ^2 with (n-1) degrees of freedom. These establish the similarity between the two integrals referred to above.

In what follows, we shall assume that $L(f,\omega)$ possesses first-order partial derivatives with respect to both f and ω, and that sufficient regularity conditions prevail to permit the operation of differentiation under the integral sign. Putting $\omega=K/2u$ in (6), the optimum estimator f is a solution of

$$\int_0^\infty \frac{\partial L(f,K/2u)}{\partial f}u^{(n+\nu-3)/2}e^{-u}du=0 \qquad (9)$$

Also, the estimator $f=S/d$ is unbiased with respect to $L(f,\omega)$ if

$$\int_0^{\infty}\left[\frac{\partial L(S/d,\omega')}{\partial\omega'}\right]_{\omega'=\omega} S^{(n-3)/2}e^{-S/2\omega}ds=0$$

which, on putting $u=S/2\omega$, becomes

$$\int_0^{\infty}\left[\frac{\partial L(2\omega u/d,\omega')}{d\omega'}\right]_{\omega'=\omega} u^{(n-3)/2}e^{-u}du=0 \qquad (10)$$

If d satisfies (10), then $f=K/d$ satisfies (9) if we can find a ν such that for all u,

$$\frac{\partial L(K/d,K/2u)}{\partial d}u^{\nu/2}\propto\left[\frac{\partial L(2\omega u/d,\omega')}{\partial\omega'}\right]_{\omega'=\omega} \qquad (11)$$

the only restriction on the proportionality constant being that it is independent of u. It is readily verified that ν can be found satisfying this condition for each of the loss functions L_1 to L_5 inclusive. Furthermore, (11) is satisfied by $\nu=0$ for any loss function of the form $F(f/\omega)$, which may well be appropriate in the problem under consideration since one is often interested in the proportional accuracy of the variance; L_2, L_4, L_5 and L_6 are of this form.

This analysis shows clearly that the generalization of Goodman's estimators obtained here is mainly coincidental, in that it depends to a large extent on the choice of loss function. It also emerges that if, having found a value of ν satisfying (11) for a given loss function, the constant of proportionality in (11) is also independent of K, then S/d is, in fact, an estimator which uniformly minimizes the risk

$$\underset{\mathrm{x}}{E}\{L[f(\mathrm{x}),\omega]\mid\omega\}$$

for all ω.

REFERENCES

[1] GOODMAN, L. A. (1960), "A note on the estimation of variance", *Sankhya*, 22, 221-228.
[2] JEFFREYS, H. (1961), Theory of Probability, 3rd ed. Oxford: Clarendon Press.
[3] LEHMANN, E. L. (1959), *Testing Statistical Hypotheses.* New York : Wiley.
[4] LENDLEY, D. V. (1953), "Statistical inference", J. R. *statist.* Soc. B, 15, 30-65.