3. Least squares estimators

As a means of assessing the validity of the model or examining the effects of outlying observations, least squares estimates of k and m may be useful for comparison with the maximum likelihood estimates. Their simplicity also makes them attractive as seeds for the iterative solution of the maximum likelihood equation for $\hat{k}$.

From the properties of the model, we have

$$E(x_i|w_i) = mw_i,$$

$$E[\{x_i - E(x_i)\}^2|w_i] = mw_i(m+k)/k.$$

The expected variance for any element is thus directly proportional to its magnitude. Thence, using the reciprocals of the w_i as weighting factors, we obtain the weighted sum of squares from expectation,

$$S = \Sigma\,(x_i - mw_i)^2/w_i,$$

the summations in this and subsequent expressions extending over $i = 1, \ldots, n$.

Differentiating with respect to m and equating to zero, we have $n\hat{m} = \Sigma x_i$, as for maximum likelihood estimation.

Proceeding as for the case of the compound model (Bissell, 1972), we equate the sum of standardized squared deviations from expectation to its degrees of freedom, yielding

$$S' = \Sigma\,(x_i - \hat{m}w_i)^2/\{\hat{m}w_i(\hat{m}+\hat{k})/\hat{k}\} = n-1,$$

whence

$$\hat{k} = \frac{\hat{m}^2}{\left\{\dfrac{1}{n-1}\,\Sigma\,\dfrac{(x_i - \hat{m}w_i)^2}{w_i}\right\} - \hat{m}}.$$

My thanks are due to Dr C. Chatfield for suggesting this investigation of the generalized model and to the Directors of ICI for permission to publish.

References

Anscombe, F. J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika* **37**, 358–82.
Bissell, A. F. (1972). A negative binomial model with varying element sizes. *Biometrika* **59**, 435–41.
Luders, R. (1934). Die statistik der seltenen ereignisse. *Biometrika* **26**, 108–28.
Quenouille, M. H. (1949). A relation between the logarithmic, Poisson and negative binomial series. *Biometrics* **5**, 162–4.

(*Received January* 1972. *Revised June* 1972)

Shrinkage of the posterior mean in the normal case

By R. L. ANDREWS, J. C. ARNOLD and R. G. KRUTCHKOFF
Virginia Polytechnic Institute and State University

Summary

Bayes estimation with a normal conditional distribution and an unknown prior distribution is considered. It is shown that the posterior mean always shrinks the observation toward a local maximum of the marginal distribution of the observation. For any symmetric, unimodal prior distribution it is shown that the posterior mean lies between the observation and the prior mean.

Some key words: Bayes estimation for mean of normal distribution; Shrinkage of point estimates.

1. Introduction

Consider the Bayes estimation problem in which the observable random variable X comes from a normal distribution with known variance σ^2 and with unknown and unobservable mean θ, where θ has the unknown distribution $G(\theta)$. For a squared error loss function the Bayes estimator is known to be the posterior mean. It is also well known (see, for example, Rutherford & Krutchkoff, 1969) that the posterior mean is given by the equation

$$E(\theta|x) = x + \sigma^2 f'(x)/f(x), \tag{1}$$

where $f(x)$ is the marginal density of the observation and $f'(x)$ is its derivative. If the situation is such that the observable is a random sample, then the x used here actually represents a sufficient statistic with σ^2 being the variance of that statistic.

Clearly, if $f'(x) = 0$, then $E(\theta|x) = x$. If x is slightly to the left of one of the local maxima of $f'(x)$, then the term $\sigma^2 f'(x)/f(x)$ will be positive, and if x is slightly to the right the term will be negative. In each case the tendency will be to move $E(\theta|x)$ from x towards the local maximum. A question of interest is whether or not the move is a shrinkage type in which the estimator moves toward the local maximum but never past it.

2. Proof of shrinkage properties

Lemma 1. *For the normal conditional distribution and any prior distribution, $E(\theta|x)$ is a nondecreasing function of x.*

Proof. We have that

$$E(\theta|x) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma f(x)} \int \theta \exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\} dG(\theta). \tag{2}$$

From this it can be shown that

$$\partial/\partial x\, E(\theta|x) = \operatorname{var}(\theta|x)/\sigma^2 \geqslant 0 \tag{3}$$

and hence $E(\theta|x)$ is a nondecreasing function of x.

Lemma 2. *The value of $E(\theta|x)$ shrinks towards a local maximum but does not pass it.*

Proof. Assume that the observation x is such that $x_c < x$, where x_c is the local maximum value toward which $E(\theta|x)$ is moving. Since $x_c < x$, $E(\theta|x_c) \leqslant E(\theta|x)$. But for a local maximum value $E(\theta|x_c) = x_c$ and so

$$x_c \leqslant E(\theta|x); \tag{4}$$

this says that the estimator has only shrunk towards the local maximum value but did not pass it. Similar results hold when $x < x_c$.

Theorem 1. *If the observation is normal and the prior is symmetric about the mean, μ, then the prior mean is a critical point.*

Proof. Here

$$f(x) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \int \exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\} dG(\theta), \tag{5}$$

$$f'(x) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma^3} \int (\theta - x) \exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\} dG(\theta). \tag{6}$$

Since $dG(\theta)$ and the normal density are symmetrical about $\theta = \mu$, whereas $\theta - x$ is anti-symmetric about $\theta = \mu$, we obtain that $f'(\mu) = 0$, indicating that μ is a critical point.

Theorem 2. *If in addition to the requirements of Theorem 1 the prior distribution has a density function which is a strictly decreasing function of $|\theta - \mu|$, then the prior mean μ is the only critical value, and is a local maximum.*

Proof. As before,

$$\begin{aligned} f'(x) &= \frac{1}{(2\pi)^{\frac{1}{2}}\sigma^3} \int (\theta - x) \exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\} g(\theta)\, d\theta \\ &= \frac{1}{(2\pi)^{\frac{1}{2}}\sigma^3} \left\{\int_0^\infty z \exp\left(-\frac{z^2}{2\sigma^2}\right) g(x+z)\, dz - \int_0^\infty z \exp\left(-\frac{z^2}{2\sigma^2}\right) g(x-z)\, dz\right\}. \end{aligned} \tag{7}$$

Now, if $x = \mu - \epsilon$ $(\epsilon > 0)$, then

$$g(x+z) = g(\mu+\epsilon+z) = g(\mu-\epsilon-z), \quad g(\mu-\epsilon+z) = g(\mu+\epsilon-z) = g(x-z).$$

Thus the first integral of equation (7) is strictly larger than the second. Similar results hold when $x = \mu+\epsilon$ $(\epsilon > 0)$; then $f'(x) < 0$.

3. Conclusions

We have shown that the posterior mean always shrinks the observation toward a local maximum of the marginal distribution. For symmetric priors the prior mean is a critical value. For priors with density functions symmetric about the mean and decreasing from the mean, the prior mean is at the maximum of the marginal density; hence the posterior mean always lies between the observation and the prior mean. The condition of symmetric unimodality were shown to be necessary in the 1971 doctoral dissertation of R. L. Andrews at the Virginia Polytechnic Institute and State University.

Reference

Rutherford, J. R. & Krutchkoff, R. G. (1969). Some empirical Bayes techniques in point estimation. *Biometrika* **56**, 133–7.

[*Received January* 1972. *Revised March* 1972]