数理解析研究所講究録
590巻 1986年 68-81

# Unbiased Estimate for $b$-value of Magnitude Frequency

*Yosihiko OGATA*

*Institute of Statistical Mathematics, Tokyo*

*and*

*Ken'ichiro Yamashina*

*Earthquake Research Institute, University of Tokyo*

## SUMMARY

If we assume that magnitudes of earthquakes are distributed identically and independently according to the negative-exponential, then the maximum likelihood estimate proposed by Utsu for the $b$-value is biased from the true value. We suggest an unbiased alternative estimate which is asymptotically equivalent to the maximum likelihood estimate. The relation between the unbiased estimate and the maximum likelihood estimate are presented from a Bayesian viewpoint. The two estimates are compared to show the superiority of the unbiased estimate to the maximum likelihood estimate, on a basis of the expected entropy maximization principle for the predictive distributions. From the same principle the posterior which derives the unbiased estimate is recommended for the inferential distribution of $b$-value rather than the standardized likelihood.

## 1. Bias of the maximum likelihood estimate

According to Gutenberg-Richter's law (Gutenberg and Richter, 1944), the number $N(M)$ of earthquakes having magnitude $M$ or larger can be expressed by the equation

$$\log_{10} N(M) = a - bM. \tag{1}$$

Utsu (1965) proposed the estimate for $b$ by

$$\hat{b} = \frac{N \log_{10} e}{\sum_{i=1}^{N} (M_i - M_0)} \tag{2}$$

where $N$ is the number of earthquakes and $M_0$ is the minimum magnitude in a given sample. Aki (1965) suggested that this is nothing but the maximum likelihood estimate which maximizes the likelihood function

$$L(b) = \prod_{i=1}^{N} f(M_i | \beta), \tag{3}$$

where

$$f(M|\beta) = \beta e^{-\beta(M-M_0)}, \ (M \geqslant M_0), \tag{4}$$

and $\beta = b/\log_{10} e$ . Using the large sample theory for the maximum likelihood Aki (1965) also provide the asymptotic error bands for large $N$ . It is further seen that the maximum likelihood estimate (2) tends to the true value $b_0$ and that $\hat{b}$ is

asymptotically unbiased estimate of $b_0$ ; i. e. $E(\hat{b})-b_0 \to 0$ as $N$ increases.

However for fixed $N$ , $\hat{b}$ is not unbiased. Indeed using the exact distribution density of $\hat{b}$ in (2)

$$f(b) = \frac{N^N}{\Gamma(N)} \frac{1}{b_0} \left(\frac{b_0}{b}\right)^{N+1} \exp\left(-\frac{Nb_0}{b}\right), \tag{5}$$

which is derived from (1.4) by Utsu (1966), we have

$$E(\hat{b}) = \int_0^\infty bf(b)db = \frac{Nb_0}{N-1} = b_0+\frac{b_0}{N-1}. \tag{6}$$

This suggests that the bias is not small when $N$ is small or $b_0$ is large.

A natural way for correcting the bias of $\hat{b}$ arises immediately based on (6); that is,

$$\tilde{b} = \frac{N-1}{N}\hat{b} = \frac{(N-1)\log_{10}e}{\sum_{i=1}^{N}(M_i-M_0)}, \tag{7}$$

is the unbiased estimate for $b_0$ . Although there are many unbiased estimators of $b_0$ besides $\tilde{b}$, we will show that the estimator (7) enjoyes certain optimality.

## 2. Relation between $\hat{b}$ and $\tilde{b}$ from a Bayesian viewpoint

Without loss of generality we consider $\hat{\beta}=N/\sum_{i=1}^{N}X_i$ and $\tilde{\beta}=(N-1)/\sum_{i=1}^{N}X_i$ instead of $\hat{b}$ and $\tilde{b}$, respectively, where $X_i=M_i-M_0$ is distributed according to $f(x|\beta)=\beta e^{-\beta x}$. Suppose a prior distribution $\pi(\beta)$ for the parameter $\beta$. Then by the Bayes theorem we have the posterior distribution

$$p(\beta) = \frac{\pi(\beta)L(\beta)}{\int_0^\infty \pi(\beta)L(\beta)d\beta}. \tag{8}$$

Consider the case where

$$\pi(\beta) = 1, \qquad 0 < \beta < \infty, \tag{9}$$

although this is not the probability density; this is called the *uniform improper prior*. Then (8) becomes the *standardized likelihood* which is conventionally used for the confidence probability distribution. For the present case this is given by

$$p_{SL}(\beta) = \frac{\{\Sigma X_i\}^{N+1}}{N!}\beta^N e^{-\beta\Sigma X_i} \tag{10}$$

where the sum $\Sigma$ is taken from $i=1$ up to $i=N$, and the mode is given by the maximum likelihood estimate $\hat{\beta}$. The choice of a prior (9) to characterize a situation where "nothing (or more realistically, little) is known a priori" does not seems to be

appropriate in general. Jeffereys (1961) suggested the rule that such *noninformative prior distribution* is given by the square root of Fisher's information measure. The rule is justified on the grounds of its invariance under parameter transformations; see also Box and Tiao (1973) and Akaike (1978) for example. For the present case the Fisher's information measure is

$$I(\beta) = E\left[-\frac{\partial^2}{\partial\beta^2}\log f(M|\beta)\right] = 1/\beta^2.$$

Thus the noninformative prior is improper and $\pi(\beta)=1/\beta$, $(0<\beta<\infty)$. Therefore the posterior probability is given by

$$p_{NI}(\beta) = \frac{(\Sigma X_i)^N}{(N-1)!}\beta^{N-1}e^{-\beta\Sigma X_i} \qquad (11)$$

where the sum $\Sigma$ is taken from $i=1$ up to $i=N$. It is easily seen that the mode of the posterior distribution (11) is $\tilde{\beta}$, while the posterior mean (*Bayes estimate*) is

$$\int_0^\infty \beta p_{NI}(\beta)d\beta = \frac{N}{\Sigma X_i} \qquad (12)$$

which is equal to the maximum likelihood estimate $\hat{\beta}$.

## 3. Entropy maximization principle and the performance of the estimates

Akaike (1977) introduced and formulated the *entropy maximization principle*, which may be specifically described for our purpose as follows. Denote by x the vector of observations $(x_1, x_2, \cdots, x_N)$. Assume that the true distribution of x is specified by an unknown density function $f(\mathbf{z})=f(z_1, z_2, \cdots, z_N)$ and consider a parameterized density $g(\mathbf{z}|\theta)$. Considering $\mathbf{y}=(y_1, y_2, \cdots, y_N)$ as the future observation with the same distribution as x, and supposing that we predict it based on the present data set x, we may call $g(\mathbf{y}|\tilde{\theta})$ *predictive distribution*, where $\tilde{\theta}=\tilde{\theta}(\mathbf{x})$ is an estimate of the true parameter $\theta_0$ satisfying $f(\mathbf{y})=g(\mathbf{y}|\theta_0)$.

As the measure of the goodness of fit of $g(\mathbf{y}|\tilde{\theta})$ to $f(\mathbf{y})$ we use the entropy of $f(\mathbf{y})$ with respect to $g(\mathbf{y}|\tilde{\theta})$ defined by

$$B(f;g) = -\int\left\{\frac{f(\mathbf{y})}{g(\mathbf{y}|\tilde{\theta})}\right\} \log \left\{\frac{f(\mathbf{y})}{g(\mathbf{y}|\tilde{\theta})}\right\} g(\mathbf{y}|\tilde{\theta})d\mathbf{y}. \tag{13}$$

Note that $B(f;g)$ is the function of x. Therefore we consider the *expected entropy*

$$\begin{aligned} J(\tilde{\theta}) &= \int B(f;g)f(\mathbf{x})d\mathbf{x} = E_X[B(f;g)] \\ &= -E_X E_Y[\log\{f(\mathbf{Y})/g(\mathbf{Y}|\tilde{\theta}(\mathbf{X}))\}], \end{aligned} \tag{14}$$

where $E_X$ and $E_Y$ are expectation with respect to the random vectors X and Y, respectively. Akaike (1977) justifies the use

of this measure based on the process of conceptual sampling experiments, and describes the natural relation to the log likelihood. Further the *AIC* was derived from this quantity (Akaike, 1973).

Here we use the expected entropy $J(\tilde{\theta})$ to measure the performance of the estimates; a larger value shows the better fit. Suppose $\{X_i\}$ and $\{Y_i\}$, $i=1, 2, \cdots, N$ are identically and independently distributed according to the density $g(x|\beta_0)=\beta_0 e^{-\beta_0 x}$. Then we have

$$J(\hat{\beta}) = N\left\{ \log N - \frac{1}{N-1} - \phi(N) \right\} \tag{15}$$

and

$$J(\tilde{\beta}) = N\{\log(N-1) - \phi(N)\}. \tag{16}$$

Here we have used the equalities

$$E[1/\sum_{i=1}^{N} X_i] = \beta_0/(N-1) \text{ and } E[\log \sum_{i=1}^{N} X_i] = \phi(N) - \log \beta_0,$$

and $\phi(N)$ is the digamma function such that $\phi(z)=\frac{d}{dz}\log \Gamma(z)$. Note that both $J(\hat{\beta})$ and $J(\tilde{\beta})$ are independent of $\beta_0$. Since

$$\{J(\tilde{\beta})-J(\hat{\beta})\}/N = \frac{N}{N-1} - 1 - \log \frac{N}{N-1} > 0,$$

we see the superiority of the estimates $\tilde{\beta}=(N-1)/\Sigma X_i$.

Furthermore, consider a family of estimators of the form $\xi/\Sigma X_i$ where $\xi$ is a positive constant, then we find the optimal $\xi$. Substitute this for $\tilde{\theta}(X)$ in (14), then we have

$$J = N\left[1 + \log \xi - \phi(N) - \frac{\xi}{N-1}\right].$$

From this we see that $\xi=N-1$ maximizes the expected entropy (14).

One may consider the expected square loss $E_X[(\theta(X)-\theta_0)^2]$ to measure the performance of the estimators. For the estimators of the form $\xi/\Sigma X_i$ we have

$$E[(\xi/\Sigma X_i - \beta_0)^2] = \frac{\beta_0^2}{(N-1)(N-2)}\{\xi^2 - 2(N-2)\xi + (N-1)(N-2)\}. \qquad (17)$$

The equality suggests that the unbiased estimator $\tilde{\beta} = (N-1)/\Sigma X_i$ is better than the maximum likelihood estimator $\hat{\beta} = N/\Sigma X_i$, although the minimum of (17) is attained by the estimator $(N-2)/\Sigma X_i$. Further interesting loss to see is absolute value of the difference, $|\theta(X) - \theta_0|$. By a Monte Carlo experiment the comparison has been performed for the estimator $N/\Sigma X_i$, $(N-1)/\Sigma X_i$ and $(N-2)/\Sigma X_i$. The trial was repeated 10000 times for $\beta_0$ such that $b_0 = \beta_0 \log_{10} e = 0.5, 1.0$ and $1.5$. The sample averages of the loss are shown in the Table 1, which shows that the unbiased estimator $\tilde{\beta}$ performs best.

## 4. Predictive distributions using posteriors

In the previous section we have only considered the case

where the predictive distribution is given in the form of $g(y|\tilde{\theta})$ by substituting a point estimate $\tilde{\theta}=\tilde{\theta}(x)$. But it is shown in Akaike (1978) that the predictive distribution of the form $\int g(y|\theta)p(\theta|x)d\theta$ provides a better fit on average than a particular $g(y|\tilde{\theta})$ with $\tilde{\theta}$ randomly chosen by the posterior distribution $p(\theta|x)$.

The implication of this is that the full use of posterior distribution performs better than the point estimates. For example the posteriors (10) and (11) for the estimate of $b$-value provides the so-called confidence limits or error bars of the inferential distributions. Incidentally the both distributions (10) and (11) tend to the normal distributions in a sense as the sample size increases, and the confidence limits provided in Aki (1965), for example, are obtained.

To see the performance of the posteriors for fixed sample size we calculate the expected entropy. For the predictive distribution with the posterior (10), we have

$$J_{SL} = N + \log \frac{\Gamma(2N)}{\Gamma(N)} - (2N+1)\phi(2N) + (N+1)\phi(N) + \log 2, \qquad (18)$$

and for the predictive distribution with the posterior (11), we have

$$J_{NI} = N + \log \frac{\Gamma(2N)}{\Gamma(N)} - 2N\phi(2N) + N\phi(N). \qquad (19)$$

Note that these quantity also are independent of true value $\beta_0$. Therefore

$$J_{NI} - J_{SL} = \varphi(2N) - \varphi(N) - \log 2 = \sum_{k=N}^{2N-1} 1/k - \log 2 > 0, \qquad (20)$$

which suggests that the posterior distribution with the noninformative prior works better than one with the uniform prior from the predictive viewpoint.

To compare the values of (15), (16), (18) and (19), we carried out numerical calculation using the approximation of the digamma function for large $x$

$$\phi(x) \sim \log x - \frac{1}{2x} - \frac{1}{12x^2} + \frac{1}{120x^4} - \frac{1}{252x^6} + \frac{1}{240x^8} - \frac{1}{132x^{10}}$$

and also using the relation $\phi(x+1)=\phi(x)+1/x$ . Table 2 suggests that the predictive distribution by averaging the posterior is more effective than those based on the point estimates.

## 5. Concluding remarks

The entropy maximization principle suggests that the full use of posterior (11) is the most effective among the considered estimations for $b$-value. Thus the confidence limits, for example, should be made based on the posterior (11).

As far as the point estimation is concerned the unbiased estimate $\tilde{\beta}=(N-1)/\Sigma X_i$ is better than the maximum likelihood estimate, and further the best among the family of the type $\xi/\Sigma X_i$ .

## Acknowledgements

We are grateful to Sigeo Aki for the helpful information in numerical calculation of the digamma function.

## References

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, (B. N. Petrov and F. Csaki, Eds.) pp. 267-281. Akademiai Kiado, Budapest.

——— (1977) On entropy maximization principle. *Applications of Statistics*, Ed. P. R. Krishnaiah, pp. 27-41. North-Holland, Amsterdam.

——— (1978) A new look at the Bayes procedure. *Biometrika*, 65, pp. 53-59.

Aki, K. (1965) Maximum likelihood estimate of $b$ in the formula $\log N=a-bM$ and its confidence limits. *Bull. Earthq. Res. Inst.*, 43, pp. 237-239.

Box, G. E. P. and Tiao, G. C. (1973) *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, Massachusetts.

Gutenberg, B. and Richter, C. F. (1944) Frequency of earthquakes in California. *Bull. Seism. Soc. Am.*, 34, pp. 185-188.

Jeffereys, H. (1961) *Theory of Probability*, third edition,

Clarendon Press, Oxford.

Utsu, T. (1965) A method for determining the value of *b* in a formula log *n*=*a*-*bM* showing the magnitude-frequency relation for earthquakes, *Geophys. Bull. Hokkaido Univ.* 13, pp. 99-103, (in Japanese with English summary).

Utsu, T. (1966) A statistical significance test of the difference in *b*-value between two earthquake groups, *J. Phys. Earth.*, 14, pp. 37-40.

Table 1 Comparison of the average absolute loss

| N \ ξ | $b_0$=0.5 | | | $b_0$=1.0 | | | $b_0$=1.5 | | |
|---|---|---|---|---|---|---|---|---|---|
| | N | N-1 | N-2 | N | N-1 | N-2 | N | N-1 | N-2 |
| 2 | 0.656 | 0.374 | — | 1.191 | 0.694 | — | 1.673 | 1.020 | — |
| 3 | 0.391 | 0.275 | 0.294 | 0.794 | 0.573 | 0.618 | 1.154 | 0.829 | 0.902 |
| 4 | 0.277 | 0.224 | 0.241 | 0.571 | 0.456 | 0.492 | 0.872 | 0.682 | 0.714 |
| 5 | 0.235 | 0.197 | 0.207 | 0.500 | 0.416 | 0.428 | 0.658 | 0.552 | 0.595 |
| 6 | 0.220 | 0.186 | 0.187 | 0.410 | 0.351 | 0.364 | 0.593 | 0.507 | 0.524 |
| 7 | 0.181 | 0.157 | 0.161 | 0.371 | 0.320 | 0.324 | 0.562 | 0.488 | 0.493 |
| 8 | 0.154 | 0.139 | 0.146 | 0.313 | 0.278 | 0.289 | 0.468 | 0.431 | 0.459 |
| 9 | 0.154 | 0.139 | 0.142 | 0.303 | 0.274 | 0.279 | 0.453 | 0.416 | 0.431 |
| 10 | 0.144 | 0.131 | 0.133 | 0.284 | 0.259 | 0.263 | 0.405 | 0.374 | 0.387 |
| 11 | 0.137 | 0.124 | 0.125 | 0.274 | 0.249 | 0.249 | 0.420 | 0.387 | 0.386 |
| 12 | 0.131 | 0.119 | 0.117 | 0.259 | 0.239 | 0.239 | 0.368 | 0.347 | 0.354 |
| 13 | 0.125 | 0.115 | 0.115 | 0.258 | 0.241 | 0.242 | 0.370 | 0.336 | 0.334 |
| 14 | 0.113 | 0.107 | 0.109 | 0.243 | 0.228 | 0.228 | 0.337 | 0.318 | 0.325 |
| 15 | 0.112 | 0.104 | 0.104 | 0.216 | 0.207 | 0.213 | 0.327 | 0.310 | 0.313 |
| 16 | 0.106 | 0.101 | 0.103 | 0.214 | 0.207 | 0.212 | 0.326 | 0.311 | 0.313 |
| 17 | 0.102 | 0.096 | 0.096 | 0.200 | 0.188 | 0.190 | 0.322 | 0.302 | 0.299 |
| 18 | 0.101 | 0.096 | 0.097 | 0.197 | 0.186 | 0.188 | 0.306 | 0.288 | 0.287 |
| 19 | 0.094 | 0.090 | 0.092 | 0.202 | 0.191 | 0.189 | 0.281 | 0.268 | 0.273 |
| 20 | 0.091 | 0.088 | 0.090 | 0.198 | 0.186 | 0.185 | 0.278 | 0.267 | 0.269 |

Table 2 Comparison of the expected entropy

| $N$ * | $J(\hat{\beta})$ | $J(\tilde{\beta})$ | $J_{SL}$ | $J_{NI}$ |
|---|---|---|---|---|
| 2 | -1.45927 | -0.845569 | -0.527329 | -0.387143 |
| 3 | -0.972515 | -0.688911 | -0.464194 | -0.374007 |
| 4 | -0.812627 | -0.630022 | -0.433636 | -0.367259 |
| 5 | -0.733397 | -0.599116 | -0.415649 | -0.363163 |
| 6 | -0.686148 | -0.580077 | -0.403812 | -0.360415 |
| 7 | -0.654785 | -0.567174 | -0.395440 | -0.358451 |
| 8 | -0.632458 | -0.557851 | -0.389193 | -0.356970 |
| 9 | -0.615752 | -0.550799 | -0.384364 | -0.355816 |
| 10 | -0.602784 | -0.545278 | -0.380518 | -0.354893 |
| 11 | -0.592429 | -0.540844 | -0.377369 | -0.354132 |
| 12 | -0.583967 | -0.537191 | -0.374781 | -0.353516 |
| 13 | -0.576928 | -0.534152 | -0.372569 | -0.352963 |
| 14 | -0.570977 | -0.531564 | -0.370684 | -0.352512 |
| 15 | -0.565872 | -0.529339 | -0.369063 | -0.352119 |
| 16 | -0.561464 | -0.527409 | -0.367636 | -0.351769 |
| 17 | -0.557603 | -0.525722 | -0.366381 | -0.351463 |
| 18 | -0.554195 | -0.524224 | -0.365286 | -0.351204 |
| 19 | -0.551173 | -0.522897 | -0.364291 | -0.350956 |
| 20 | -0.548462 | -0.521698 | -0.363402 | -0.350758 |

*) $N$ is the sample size.