TAMKANG JOURNAL OF MATHEMATICS
Volume 34, Number 2, Summer 2003

AN ASYMPTOTIC MEASURE OF ACCURACY EFFECT FROM CENSORSHIP IN PARAMETRIC ESTIMATION

CHUNG-SIUNG KAO

Abstract. An asymptotic measure is provided to evaluate the effect on loss of accuracy for censored data in parametric estimation of location and scale parameters. With this measure, it is shown that the amount of effect from censored data relative to noncensored data is invariant of the actual values of the location and scale parameters, but is only dependent on the form of underlying distributions which the data are originated. In addition, among the most well-known distributions, obtained results for the measure show that two censored data values together usually may possess more information than one noncensored data value in the parametric estimation for location and scale parameters.

1. Introduction

In the occurrences of strike by natural forces like earthquakes, etc., the after-math may only provide registered magnitude of the forces and the number of investigated subjects that are broken and not broken. Also, in qualification tests for quality control upon produced pipes or metal rods, etc., an economic consideration is to exert some preassigned magnitude of force upon said pieces of material to observe how many of them are broken and not broken instead of tediously obtaining the precise break point for each of the pieces. For either of the above frequently occurred examples, what are observed appear obviously to be censored data, which mean the registered or preassigned magnitudes of force here and whether the investigated subjects are broken or not.

Throughout the years a large quantity of published work covered estimation problems with censored data. Most of the one-sided censored data are right censored, especially those in survival analysis. Among many such examples, Englehardt and Bain (1974) dealt point estimation problem with censored samples. Lately, estimation of regression models appear to be the main subject for investigation with right-censored data where efficiency of the estimates is studied. For example, Gould and Lawiess (1988) and Young and Bakier (1993) investigated estimation efficiency for censored data. Recently attention has been directed to treatment method when the data are interval-censored. Among others, Rabinowitz, Tsiatis and Aragon (1995) and Leung and Elashoff (1996) touched on use of interval-censored data in regression.

Received May 13, 2002.
2000 *Mathematics Subject Classification.* 62F10.
Key words and phrases. Threshold value, distribution observation, qualification observation, maximum likelihood estimator, consistency, asymptotic normality, M-efficienhcy.

An important question regarding censored data should be how much information remain of such censored data as compared to that of the precise data which are the values of actual strength of the investigated subjects. It appears that this question has not be addressed and properly studied. In this work, we compare the maximum likelihood estimators (MLE) between using censored data and using actual data as basis for answering the aforesaid question, assuming the underlying distribution regarding investigated subjects is known. A ratio measure called in this work as M-efficiency is properly established to shed light upon how much information is lost when precise data are changed to censored data in the context of estimating location and scale parameters. To begin with, we obtain proofs to assert that the MLE's from using totally censored data are both consistent and asymptotically normal. Then it is shown that specifically for estimation of location and scale parameters, the M-efficiency is invariant to the true values of location and scale parameters. The M-efficiency is explicitly expressed in form of a formula in terms of the p.d.f. and c.d.f. of the underlying distribution. Finally, examples with the best known distributions including Normal, Exponential and Uniform are provided to show that despite censorship the censored data may still retain major portion of the information possessed by their corresponding precise data.

For uniformity in the presentation, we shall use X to denote a precise data value and Z to denote a 0 or 1 qualification value. The threshoulds shall be denoted by Q's. In addition, $F(\cdot;\underline{\theta})$ shall denote the cumulative distribution function (c.d.f.) of X, where $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_J)$ is the parameter vector of dimension J. The corresponding probability density function (p.d.f.) will be denoted by $f(\cdot;\underline{\theta})$. The inference problem considered here will be to estimate $\underline{\theta}$ given observations of X's and Z's and Q's using maximum likelihood method.

2. Maximum Likelihood Estimator of $\underline{\theta}$

Assume that $X_1, X_2, \ldots, X_N$ are identically and independently distributed unknown random observations which fall upon K known censoring thresholds $Q_1, Q_2, \ldots, Q_K$. Without loss of generality, let there be sequentially n_k of X_i's falling on threshold Q_k, $1 \le k \le K$, and let $N_k = \sum_{j=1}^{k} n_j$ for $1 \le k \le K$ with $N_K = N$. Then for $N_{k-1} + 1 \le j \le N_k$ with $N_0 = 0$, define

$$\begin{aligned} Z_j &= 1 \quad \text{if } X_j > Q_k, \\ \text{and} \qquad &= 0 \quad \text{otherwise.} \end{aligned} \tag{2.1}$$

Although X_j's are not observable, we have the values of Z_j's which are Bernoulli observations, and the Q_k's. The common c.d.f. of X_i's is $F(\cdot;\underline{\theta})$, therefore for each Z_j, $N_{k-1} \le j \le N_k$, the p.d.f. of Z_j is $p_k^{Z_j}(1-p_k)^{1-Z_j}$ for $Z_j = 1$ or 0, where $p_k = p_k(\underline{\theta}) = 1 - F(Q_k;\underline{\theta})$. The likelihood function based on the qualification observations is $L_N(\underline{\theta})$, defined by

$$L_N(\underline{\theta}) = \prod_{k=1}^{K} \left(\prod_{i=N_{k-1}+1}^{N_k} p_k^{Z_i}(1-p_k)^{1-Z_i} \right), \tag{2.2}$$

where $N = N_K$. Then the log likelihood function becomes

$$\begin{aligned}\log L_N(\underline{\theta}) &= \log L_N(\underline{Z};\underline{\theta})\\ &= \sum_{k=1}^{K} S_k \log[1 - F(Q_k;\underline{\theta})] + \sum_{k=1}^{K}(n_k - S_k)\log F(Q_k;\underline{\theta}),\end{aligned} \tag{2.3}$$

where $S_k = \sum_{i=N_{k-1}+1}^{N_k} Z_i$ for $1 \le k \le K$ and $\underline{Z} = (Z_1, Z_2, \dots, Z_N)$ with N denoting the total sample size ($N = N_K$). Let

$$U_k(z;\underline{\theta}) = \bigtriangledown_{\underline{\theta}}(\log g_k(z;\underline{\theta})), \quad z = 0, 1 \quad \text{for} \quad 1 \le k \le K \tag{2.4}$$

where $g_k(z;\underline{\theta}) = (1 - F(Q_k;\underline{\theta}))^z F(Q_k;\underline{\theta})^{1-z}$, and $\bigtriangledown_{\underline{\theta}} = (\frac{\partial}{\partial\theta_1}, \frac{\partial}{\partial\theta_2}, \dots, \frac{\partial}{\partial\theta_J})$ with J being dim $\underline{\theta}$. Since $X_1, X_2, \dots, X_N$ are i.i.d., we may define

$$A_k(\underline{\theta}_0, \underline{\theta}) = E_{\underline{\theta}_0}[U_k(Z_i;\underline{\theta})] \quad \text{for} \quad N_{k-1} + 1 \le i \le N_k$$

and

$$\Psi_N(\underline{Z};\underline{\theta}) = \sum_{k=1}^{K} \sum_{N_{k-1}+1} \bigtriangledown'_{\underline{\theta}}(U_k(Z_i;\underline{\theta})) \quad \text{with} \quad N_0 \equiv 0, \tag{2.5}$$

where $\bigtriangledown'_{\underline{\theta}}$ is the transposed of vector operator $\bigtriangledown_{\underline{\theta}}$. Note that A_k and Ψ_N as defined above are an J-dimensional vector and a $J \times J$ matrix respectively. Since f and the g_k's are probability density function, it follows that

$$A_k(\underline{\theta}, \underline{\theta}) = \underline{0} \quad \text{for} \quad 0 \le k \le K, \tag{2.6}$$

where $\underline{0}$ denotes the zero vector. Let B_k be defined by

$$B_k(\underline{\theta}_0, \underline{\theta}) = E_{\underline{\theta}_0}[U'_k(Z_i;\underline{\theta})U_k(Z_i;\underline{\theta})] \quad \text{for} \quad N_{k-1} + 1 \le i \le N_k. \tag{2.7}$$

Then assuming the true value of $\underline{\theta}$ to be $\underline{\theta}_0$, the variance-covariance matrix of $U_k(Z_i;\underline{\theta})$ for any i such that $N_{k-1} + 1 \le i \le N_k$ is equal to $B_k(\underline{\theta}_0, \underline{\theta}) - A_k(\underline{\theta}_0, \underline{\theta})A_k(\underline{\theta}_0, \underline{\theta})$. Due to the mutual independence of the Z_i's it follows that the variance-covariance matrix of $\log L_N(\underline{Z};\underline{\theta})$ given the true value of $\underline{\theta}$ being $\underline{\theta}_0$ is equal to $\sum_{k=0}^{K} n_k[B_k(\underline{\theta}_0, \underline{\theta}) - A'_k(\underline{\theta}_0, \underline{\theta})A_k(\underline{\theta}_0, \underline{\theta})]$. Let this sum of matrices be denoted by $C_N(\underline{\theta}_0, \underline{\theta})$. It is easily seen that $C_N(\underline{\theta}_0, \underline{\theta}) = -E_{\underline{\theta}_0}\Psi_N(\underline{Z};\underline{\theta})$ due to the fact that, for any random variable W,

$$E_{\underline{\theta}}\left[\frac{\partial}{\partial\theta_i}\log h(W;\underline{\theta})\right] = -E\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log h(W;\theta)\right],$$

where $h(\cdot;\theta)$ is the probability density function of W.

In what follows we will denote the maximum likelihood estimate (MLE) of $\underline{\theta}$ by $\hat{\underline{\theta}}_N$ based on likelihood function under the given Z_i's (depend on Q_k's). When the underlying cumulative distribution function $F(x;\theta)$ for X_i's are regular with respect to θ, it follows from the commonly known asymptotic properties of maximum likelihood estimators (see Serfling (1980)) that $\hat{\theta}$ is strongly consistent and asymptotically normal

$N(\underline{\theta}_0, C_N^{-1}(\underline{\theta}_0, \underline{\theta}_0))$ for large N, where $C_N(\underline{\theta}_0, \underline{\theta}) = -E_{\underline{\theta}_0}\Psi_N(\underline{Z};\theta)$ with $\Psi_N(\underline{Z};\theta)$ defined in above at (2.5).

3. Efficiency of Censored Observations

As it was mentioned earlier, the amount of information from censored observations is almost compatible to that from distribution observations provided that the censoring thresholds are properly chosen. It is important to note that the order of magnitude for estimation accuracy is maintained at $\frac{1}{\sqrt{n}}$ when the censored observations are used. This gives the essential reason for promoting the use of censored data. In this section, the loss of information is to be measured by comparing the widths of asymptotic confidence intervals obtained by using maximum likelihood method upon censored data and distribution data separately. For purpose of simplicity we shall restrict the comparison to only one-dimensional case for $\underline{\theta}(=\theta)$.

Following the notations in the preceding sections, let there be n random distribution observations $\{X_i, 1 \le i \le n\}$. Then for the distribution observation X_i, let the threshold be Q_i and the corresponding censored observation be Z_i. Accordingly,

$$\begin{aligned} Z_i &= 1 \quad \text{if } X_i > Q_i, \\ \text{and} \qquad\qquad &= 0 \quad \text{otherwise.} \end{aligned}$$

Again let the common cumulative distribution function of the X_i's be $F(\cdot;\theta)$, and the probability density function be $f(\cdot;\theta)$.

Now denote the maximum likelihood estimate (MLE) of θ using the X's by $\hat{\theta}_X$, and denote the MLE of θ using the Z's by $\hat{\theta}_z$. Then for large n let the asymptotic standard deviation derived from the asymptotic distribution of $\hat{\theta}_X$ be denoted by $V_{X,n}(\theta_0)$ and the one corresponding to $\hat{\theta}_Z$, be $V_{Z,n}(\underline{Q}, \theta_0)$, where $\underline{Q} = (Q_1, Q_2, \ldots, Q_n)$. Note that the components of $\underline{Q}$ have K distinct values for all $n \ge K$, and K is preassigned.

Definition 3.1. For given $\underline{Q}$, the M-efficiency $e(\underline{Q})$ of using censored observations versus using distribution observations is defined by

$$e(\underline{Q};\theta_0) = \lim_{n\to\infty} \frac{V_{X,n}(\theta_0)}{V_{Z,n}(\underline{Q}, \theta_0)}. \tag{3.1}$$

When $\underline{Q} = \underline{Q}_0$ maximizes $e(\underline{Q};\theta_0)$ for all Q, then we call $e^*(\theta_0) = e(\underline{Q}_0\theta_0)$ the optimal M-efficiency.

It is known that $\hat{\theta}_X$ is asymptotically normal with mean θ_0 and variance

$$n^{-1}E_{\theta_0}^{-1}\left[\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2_{\theta=\theta_0}\right].$$

According to known property of maximum likelihood estimator it is easily seen that $\hat{\theta}_Z$ is asymptotically normal with mean θ_0 and variance $v_n^2(\theta_0)$, where

$$v_n^2(\theta) = \left\{\sum_{i=1}^{n} \frac{[\frac{\partial}{\partial\theta}F(Q_i;\theta)]^2}{F(Q_i;\theta)[1-F(Q_i;\theta)]}\right\}^{-1}. \tag{3.2}$$

Therefore the optimal choice of Q_i's is to set $Q_i = q_0$ for all i such that $Q_i = q_0$ maximizes $[\frac{\partial}{\partial\theta}F(Q_i;\theta)]^2_{\theta=\theta_0}/[F(Q_i;\theta_0)(1-F(Q_i;\theta_0))]$. Then we have the optimal M-efficiency

$$e^*(\theta_0) = \left\{ \frac{[\frac{\partial}{\partial\theta}F(q_0;\theta)]^2_{\theta=\theta_0}}{E_{\theta_0}[(\frac{\partial}{\partial\theta}\log f(X;\theta))^2_{\theta=\theta_0})]F(q_0;\theta_0)[1-F(q_0;\theta_0)]} \right\}^{\frac{1}{2}}. \tag{3.3}$$

This is obvious, since it can be easily shown that

$$e(Q_i;\theta_0) = \left[\frac{\partial}{\partial\theta}F(Q_i;\theta)\right]^2_{\theta=\theta_0} \Big/ \left\{ E_{\theta_0}\left[\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2_{\theta=\theta_0}\right] F(Q_i;\theta_0)[1-F(Q_i;\theta_0)] \right\}$$

In case that θ is either a location parameter or a scale parameter, the optimal M-efficiency $e^*(\theta_0)$ is a constant e_0 which is invariant to value of θ_0. This is shown in the following:

When θ is a location parameter, we have $F(x;\theta) = F_0(x-\theta)$. Then

$$E_{\theta_0}\left[\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2_{\theta=\theta_0}\right] = \int_{-\infty}^{\infty} \frac{[f_0'(t)]^2}{f_0(t)}dt,$$

and

$$\max_Q \left\{ \frac{[\frac{\partial}{\partial\theta}F(Q;\theta)]^2_{\theta=\theta_0}}{F(Q;\theta_0)[1-F(Q;\theta_0)]} \right\} = \max_Q \left\{ \frac{f_0^2}{F_0(Q)[1-F_0(Q)]} \right\},$$

therefore $e^*(\theta_0)$ is actually a constant independent of θ_0. When θ is a scale parameter, we have $F(x,\theta) = F_0(z/\theta)$. Then

$$E_{\theta_0}\left[\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2_{\theta=\theta_0}\right] = \frac{1}{\theta_0^2}\int_{-\infty}^{\infty} \left[r\frac{f_0'(r)}{f_0(r)}+1\right]^2 f_0(r)dr,$$

and

$$\max_Q \left\{ \frac{[\frac{\partial}{\partial\theta}F(Q;\theta)]^2_{\theta=\theta_0}}{F(Q;\theta_0)[1-F(Q;\theta_0)]} \right\} = \frac{1}{\theta_0^2}\max_Q \left\{ \frac{Q^2 f_0^2}{F_0(Q)[1-F_0(Q)]} \right\},$$

therefore θ_0^2 is cancelled out in (3.3) for $e^*(\theta_0)$. Hence $e^*(\theta_0)$ is a constant also independent of θ_0. In such case, let $e^* = e^*(\theta_0)$ denote the constant. The above results prove for the following theorem.

Theorem 4.1. *If θ is either a location parameter or a scale parameter, then the optimal M-efficiency is independent of the true value of the parameter θ.*

Example 1. Let the sequence $\{X_i, 1 \le i \le n\}$ be a simple random sample from $\mathcal{N}(\mu,\sigma^2)$, where σ^2 is known variance and, without loss of generality, σ^2 is assumed to be 1. Then we have the optimal threshold $q_0 = \mu$. Thus the optimal choice of the threshold is the unknown true mean. It follows that the optimal M-efficiency $e^*(\mu)$ is independent of μ. In fact, we have $e^* = \sqrt{\frac{2}{\pi}} = 0.798$. It is worth noting that if $\theta = \sigma$ is unknown and μ is known, then it may be shown that $q_0 = \mu + 1.58\sigma$. Furthermore, this gives $e^* = 0.554$. In fact, q_0 can also be $\mu - 1.58\sigma$, which yields the same value for e^*.

Example 2. Let the sequence $\{X_i, 1 \le i \le n\}$ be random observations from a negative exponential distribution, where

$$f(x;\lambda) = \frac{1}{\lambda} e^{-x/\lambda}, \quad x > 0 \ (\theta = \lambda > 0).$$

Then we have $F(x;\lambda) = 1 - e^{-x/\lambda}$. It is obtained that $q_0 = 1.59\lambda$, which yields $e^* = 0.805$.

Example 3. Let $\{X_i, 1 \le i \le n\}$ be random observations from a uniform distribution, where

$$f(x;\lambda) = \frac{1}{\lambda}, \quad 0 < x \le \lambda \ (\theta = \lambda > 0).$$

It follows that $F(x;\lambda) = \frac{1}{\lambda} I^+(\lambda - I^+(x))$, where I^+ is defined by

$$\begin{aligned} I^+(t) &= t \quad \text{if } t > 0 \\ \text{and} \qquad &= 0 \quad \text{otherwise.} \end{aligned}$$

Then we have $q_0 = \lambda$ and $e^* = 1$ according to the calculations. In fact, this is predictable and seems obvious, since for this example the optimal qualification observation and the distribution observation yield equivalent information about λ.

In the above examples it appears that $q_0 = g(\theta)$ for some function g, which normally has an obtainable analytic form. Thus in practice one may select the first threshold to be $g(\theta_p)$ where θ_p is an estimate of true θ based on prior information. Then the optimal threshold can be approached by successive approximations.

For example, when $e^* = 0.5$, it means that the information from four optimally censored observations is equivalent to information from one distribution observation. If $e^* = 0.8$, it means the information from two optimally censored observations is better than that from one distribution observation in the context of maximum likelihood estimation. The preceding examples strongly justify use of censored data in light of the e^* values.

References

[1] Engelhardt, M. and Bain, L. J., *Some results on point estimation for the two-parameter Weibull or extreme-value distribution*, Technometrics **16**(1974), 49-56.

[2] Gould, A. and Lawless, J. F., *Estimation efficiency in lifetime regression models when responses are censored or grouped*, Comm. Statist, B-Simulation Comput. **17**(1988), 689-712.

[3] Leung, K-M. and Elashoff, R. M., *A three-state disease model with interval-censored data: estimation and applications to AIDS and cannncer*, Lifetime Data Anal. **2**(1996), 175-194.

[4] Rabinowitz, D., Tsiatis, A. and Aragon, J., *Regression with interval-censored data*, Biometrika **82**(1995), 501-513.

[5] Serfling, R. J., Approximation Theorems of Mathematical Statistics. John Wiley and Sons, 1980.

[6] Young, D. H. and Bakir, S. T., *Estimation efficiency of grouped and censored observation schemes for a generalized regression model*, Comm. Statists. A-Theorey Methods **22**(1993), 403-423.

Department of Mathematics and Institute of Statistical Science, National Chung Cheng University, Minhsiung, Chiayi, Taiwan.

E-mail: cskao@math.ccu.edu.tw