The Korean Communications in Statistics **Vol. 14** No. 2, 2007, pp. 309–316

Sequential Estimation in Exponential Distribution*

Sangun Park[1)]

Abstract

In this paper, we decompose the whole likelihood based on grouped data into conditional likelihoods and study the approximate contribution of additional inspection to the efficiency. We also combine the conditional maximum likelihood estimators to construct an approximate maximum likelihood estimator. For an exponential distribution, we see that a large inspection size does not increase the efficiency much if the failure rate is small, and the maximum likelihood estimator can be approximated with a linear function of inspection times.

Keywords: Conditional likelihood; Fisher information; life testing; maximum likelihood estimator; order statistics.

1. Introduction

Many lifetime experiments employ the intermittent inspection scheme rather than the continuous one for its convenience and saving costs. The data from this intermittent inspection contain only the numbers of failures in each inspection interval, and are called *grouped data* or *quantal response data.*

In this paper, we decompose the whole likelihood of grouped data into conditional likelihoods, which enables us to study the approximate contribution of additional inspection to the efficiency. Moreover, the decomposition of likelihoods also enables us to get an approximate maximum likelihood estimator since each conditional maximum likelihood estimator can be linearly combined.

For the mean of the exponential distribution, we see that the information obtained with additional inspection is close to the number of failures in additional

* This work was supported in part by the Yonsei University College of Business and Economics Research Fund.

1) Professor, Department of Applied Statistics, Yonsei University, Shinchon Dong 134, Seoul 120-749, Korea.
E-mail : sangun@yonsei.ac.kr

inspection interval and a large inspection size does not increase the efficiency much if the proportion of total failures is low. In estimating the mean of the exponential distribution based on these grouped data, the maximum likelihood estimator is the natural choice but does not have a closed form solution. Therefore, several approximate maximum likelihood estimators of closed form have been proposed and discussed by some authors including (Tallis, 1967; Kendell and Anderson, 1971; Seo and Yum, 1993). We provide another approximate maximum likelihood estimator, which is a linear function of inspection times.

2. Decomposition of Likelihood and Conditional Maximum Likelihood Estimator

Suppose that $t_1 < \cdots < t_k$ are the inspection times and x_i failures are observed in $(t_{i-1}, t_i]$. Then we have grouped data $(x_1, x_2, \ldots, x_{k+1})$ from k intermittent inspection. The likelihood based on these grouped data can be written as follows:

$$L(\theta; t_1, \ldots, t_k) \propto \prod_{i=1}^{k+1} (F(t_i;\theta) - F(t_{i-1};\theta))^{x_i}, \tag{2.1}$$

where $t_0 = 0$ and $t_{k+1} = \infty$.
We first note that (1) can be decomposed into conditional likelihoods as

$$L(\theta; t_1, \ldots, t_k) = L(\theta; t_1) \times L(\theta; t_2|t_1) \times \cdots \times L(\theta; t_k|t_{k-1}),$$

where

$$L(\theta; t_i|t_{i-1}) \propto \left(\frac{F(t_i;\theta) - F(t_{i-1};\theta)}{1 - F(t_{i-1};\theta)}\right)^{x_i} \left(\frac{1 - F(t_i;\theta)}{1 - F(t_{i-1};\theta)}\right)^{n_i - x_i}$$

and $n_i = \sum_{j=i}^{k+1} x_j$.
We denote $I_{t_i|t_{i-1}}(\theta)$ to be the Fisher information in the conditional likelihood, $L(\theta; t_i|t_{i-1})$, which can be written as

$$I_{t_i|t_{i-1}}(\theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{t_i} \left(\frac{\frac{\partial}{\partial\theta} F_{t_i|t_{i-1}} \cdot (x_i - n_i F_{t_i|t_{i-1}})}{F_{t_i|t_{i-1}}(1 - F_{t_i|t_{i-1}})}\right)^2 f_{t_{i-1}t_i} dt_{i-1} dt_i,$$

where $F_{t_i|t_{i-1}} = (F(t_i;\theta) - F(t_{i-1};\theta))/(1 - F(t_{i-1};\theta))$.
Thus we can study the contribution of additional inspection to the efficiency with $I_{t_i|t_{i-1}}(\theta)$.

This exact Fisher information is difficult to obtain in most cases, and can be approximated with the asymptotic information in x_i/n_ith sample quantile (Zheng and Gastwirth, 2000),

$$\tilde{I}_{t_i|t_{i-1}}(\theta) = n_i \frac{(\frac{\partial}{\partial\theta} F_{t_i|t_{i-1}})^2}{x_i/n_i(1 - x_i/n_i)}, \tag{2.2}$$

where t_i can be approximated with $F^{-1}(1 - n_{i+1}/n_1; \theta)$.

If we denote $\hat{\theta}_{t_i|t_{i-1}}$ to be the maximum likelihood estimator of the conditional likelihood, $L(\theta; t_i|t_{i-1})$, we have k conditional maximum likelihood estimators. In view of Park (1996), we can combine $\hat{\theta}_{t_i|t_{i-1}}$'s, $i = 1, \ldots, k$, to produce an estimator as

$$\hat{\theta}_k = \sum_{i=1}^{k} \frac{I_{t_i|t_{i-1}}(\theta)}{I_{t_1 \cdots t_k}(\theta)} \hat{\theta}_{t_i|t_{i-1}}, \tag{2.3}$$

where $I_{t_i|t_{i-1}}(\theta)$ is the Fisher information about θ in $L(\theta; t_i|t_{i-1})$ and $I_{t_1 \cdots t_k}(\theta) = \sum_{i=1}^{k} I_{t_i|t_{i-1}}(\theta)$.

If the exact Fisher information is not available, we can use the asymptotic Fisher information.

3. Decomposition in Exponential Distribution

For an exponential distribution, $f(x; \theta) = \exp(-x/\theta)/\theta$, the conditional likelihood, $L(\theta; t_i|t_{i-1})$, can be written as

$$L(\theta; t_i|t_{i-1}) = \left(1 - \exp\left(-\frac{t_i - t_{i-1}}{\theta}\right)\right)^{x_i} \left(\exp\left(-\frac{t_i - t_{i-1}}{\theta}\right)\right)^{n_i - x_i},$$

and the asymptotic Fisher information in the conditional likelihood can be obtained in view of (2) as

$$\begin{aligned} \tilde{I}_{t_i|t_{i-1}}(\theta) &= E\left(\left(\frac{\theta}{\partial\theta} \log L(\theta; t_i|t_{i-1})\right)^2\right) \\ &= \frac{1}{\theta^2} n_i \left(\frac{n_i - x_i}{x_i}\right) \left(\log\left(\frac{n_i - x_i}{n_i}\right)\right)^2. \end{aligned}$$

If x_{k+1} (number of survivals) is relatively large so that $x_i/n_i, i = 1, \ldots, k$ is small, the above decomposition tells us two interesting facts: The first one is that $I_{t_i|t_{i-1}}(\theta) \approx x_i/\theta^2$. For example, $I_{t_i|t_{i-1}}(\theta) = 0.9609 x_i/\theta^2$ even for $x_i/n_i = 0.5$.

Thus we can see that additional information obtained by observing detailed failure times is not so much (see, Gersbakh, 1995; Park, 2003; Park and Kim, 2006). The second one is that $I_{t_1\cdots t_k}(\theta) \approx \sum_{i=1}^{k} x_i/\theta^2$, the number of total failures. Thus the total information does not increase much, though we increase the inspection size k with fixed t_k since fixed t_k gives the same number of total failures. We can also easily understand the fact in Shapiro and Gulati (1996) that an experiment even as few as three intervals (two inspections) does not result in a large loss of information.

For the exponential distribution, $\hat{\theta}_{t_i|t_{i-1}}$ can be obtained in a closed form as $(t_i - t_{i-1})/\log(n_i/(n_i - x_i))$. We note that there is no maximum likelihood solution if $x_i = 0$. Then we can combine the conditional maximum likelihood estimators to have an estimator in view of (3) as

$$\hat{\theta}_c = \frac{1}{\theta^2 I_{t_1\cdots t_k}(\theta)} \sum_{i=1}^{k} n_i \left(\frac{n_i - x_i}{x_i}\right) \log\left(\frac{n_i}{n_i - x_i}\right)(t_i - t_{i-1}). \tag{3.1}$$

We can approximate $n_i((n_i - x_i)/x_i)\log(n_i/(n_i - x_i))$ in (4) to be n_i for $x_i = 0$ since $((n_i - x_i)/x_i)\log(n_i/(n_i - x_i)) \to 1$ as $x_i \to 0$. In a similar way, we can approximate $n_i((n_i - x_i)/x_i)\log(n_i/(n_i - x_i))$ to be 0 for $n_i = x_i$.

We can also express $\hat{\theta}_k$ as a linear combination of t_i's as follows:

$$\hat{\theta}_k = \frac{1}{\theta^2 I_{t_1\cdots t_k}(\theta)} \sum_{i=1}^{k} w_i t_i,$$

where

$$w_i = n_i \left(\frac{n_i - x_i}{x_i}\right) \log\left(\frac{n_i}{n_i - x_i}\right) - n_{i+1}\left(\frac{n_{i+1} - x_{i+1}}{x_{i+1}}\right) \log\left(\frac{n_{i+1}}{n_{i+1} - x_{i+1}}\right).$$

Example 3.1 The data in table 1 comes from Nelson (1982):

$$\begin{aligned}(t_1, \ldots, t_8) &= (6.12, 19.92, 29.64, 35.40, 39.72, 45.24, 52.32, 63.48),\\ (x_1, \ldots, x_9) &= (5, 16, 12, 18, 18, 2, 6, 17, 73).\end{aligned}$$

The combined estimator can be obtained to be 82.70 which is close to the maximum likelihood estimate, 82.67. We can confirm from Table 3.1 that the conditional information obtained from i^{th} inspection is close to x_i since x_9 is relatively large. If we stop at the 7^{th} inspection time, the combined estimator will be obtained to be 89.18 and the total information to get is 76.80.

Table 3.1: Conditional maximum likelihood estimator, conditional information and combined estimator based on eight inspections

i	t_i	x_i	$\theta_{t_i\|t_{i-1}}$	$I_{t_i\|t_{i-1}}(\theta)(1/\theta^2)$	w_i	θ_i
1	6.12	5	201.33	5.00	10.75	201.33
2	19.92	16	32.71	15.99	13.89	149.06
3	29.64	12	113.33	11.99	15.26	136.06
4	35.40	18	39.93	17.97	18.07	102.16
5	39.72	18	25.62	17.96	9.50	82.21
6	45.24	2	267.71	2.00	4.06	87.44
7	52.32	6	109.70	6.00	12.03	89.18
8	63.48	17	53.31	16.94	80.91	82.70
9		73		$I_{t_1 \cdots t_8}(\theta) = 93.84$	$\sum_{i=1}^{8} w_i = 164.47$	

Table 3.2: Conditional maximum likelihood estimator, conditional information and combined estimator based on three inspections

i	t_i	x_i	$\theta_{t_i\|t_{i-1}}$	$I_{t_i\|t_{i-1}}(\theta)(1/\theta^2)$	w_i	θ_i
1	19.92	21	148.23	20.97	41.31	148.23
2	39.72	48	49.67	47.40	30.44	79.91
3	63.48	25	80.68	24.82	84.28	80.11
4		73		$I_{t_1 \cdots t_3}(\theta) = 93.11$	$\sum_{i=1}^{3} w_i = 156.03$	

Suppose now that the inspections are done only three times at $2^{nd}, 5^{th}$ and 8^{th} inspection times:

$$\begin{aligned}(t_1, t_2, t_3) &= (19.92, 39.72, 63.48),\\ (x_1, x_2, x_3, x_4) &= (21, 48, 25, 73).\end{aligned}$$

We can see that the total information in table 3.2 is still close to the number of failures, 94 and the loss of information is just 0.73 due to reducing the inspection size from 8 to 3 as can be expected.

However, the total information is far from the number of failures if the failure rate is not low. We can show that the total information is 85.14 which is far from the total number of failures 94 if we assume x_9 in Table 3.1 to be 1 and the total information is 75.78 if we assume x_4 in Table 3.2 to be 1. Thus we can conclude that the intermittent inspection is effective if x_{k+1}/n_1 (survival rate) is large.

Remark 3.1 For the type II right censored case, where only $t_1 < \cdots < t_r$ are observed from a sample of size n, the conditional likelihood of t_i given t_{i-1}

can be written as

$$L(t_i|t_{i-1};\theta) \propto \frac{f(t_i;\theta)}{1-F(t_{i-1};\theta)} \left(\frac{1-F(t_i;\theta)}{1-F(t_{i-1};\theta)}\right)^{n-i}.$$

For an exponential distribution, the likelihood can be written as

$$L(\theta;t_i|t_{i-1}) \propto \frac{1}{\theta}\exp\left(-\frac{1}{\theta}(n-i+1)(t_i-t_{i-1})\right).$$

Then the conditional maximum likelihood estimator can be obtained as $\hat{\theta}_{t_i|t_{i-1}} = (n-i+1)(t_i-t_{i-1})$. We can here recall the well-known fact in Sukhatme (1937) that $(n-i+1)(t_i-t_{i-1})$'s, $i=1,\ldots,n$, are independently and exponentially distributed. Since the exact Fisher information in each conditional likelihood can be obtained to be 1, the combined estimator can be written as

$$\begin{aligned}\hat{\theta}_r &= \sum_{i=1}^{r}\frac{1}{r}(n-i+1)(t_i-t_{i-1}) \\ &= \frac{1}{r}\left(\sum_{i=1}^{r} t_i + (n-r)t_r\right),\end{aligned}$$

which is exactly equivalent to the maximum likelihood estimator.

4. Performance of the Combined Estimator

In this section, we study the performance of the combined estimator through Monte Carlo simulation for the sample size $n = 20, 30, 50, 100, 200$ and $k = 2, 3, 5, 7, 10$ cases. The scale parameter of an exponential distribution is assumed to be 1 without loss of generality and the inspection times are determined to be optimally spaced, which has been studied in (Kulldorff, 1961; Nelson, 1982). For comparison, we consider the mid-point estimator,

$$\hat{\theta}_0 = \frac{\left(\sum_{i=1}^{k} x_i \frac{t_i+t_{i-1}}{2} + t_k x_{k+1}\right)}{\sum_{i=1}^{k} x_i},$$

and the approximate maximum likelihood estimator in Seo and Yum (1993),

$$\hat{\theta}_a = \hat{\theta}_0 \left(1 - \frac{\sum_{i=1}^{k} x_i((t_i - t_{i-1})/\hat{\theta}_0)^2}{12 \sum_{i=1}^{k} x_i} \right).$$

The average mean squared errors based on 50,000 simulations for $\hat{\theta}_0$, $\hat{\theta}_a$ and the combined estimator, $\hat{\theta}_k$, have been calculated and are listed in Table 4.1. If $n_1 = x_1$, we have $\hat{\theta}_0 = t_1/2$ and $\hat{\theta}_a = t_1/3$ but the combined estimator and also the maximum likelihood estimator gives a little bit awkward estimate of 0. As a result, the combined estimator shows poor performance for small n and large k, but shows similar performance to $\hat{\theta}_a$ if n/k is large.

5. Conclusions

In this paper, we use the decomposition of likelihoods of grouped data to see how much information we can get with additional inspection. For the exponential distribution, the additional information is close to the number of failures in the inspection interval if the failure rate is low. Thus we can conclude that if the failure rate is low, we do not need a large inspection size. We can also combine

Table 5.1: The average mean square errors based on 10,000 simulations

n/k		2	3	5	7	10
20	$\hat{\theta}_0$	0.0740	0.0620	0.0555	0.0531	0.0511
	$\hat{\theta}_k$	0.0696	0.0654	0.0622	0.0604	0.0589
	$\hat{\theta}_a$	0.0642	0.0583	0.0544	0.0526	0.0507
30	$\hat{\theta}_0$	0.0527	0.0428	0.0372	0.0355	0.0347
	$\hat{\theta}_k$	0.0419	0.0403	0.0389	0.0381	0.0378
	$\hat{\theta}_a$	0.0410	0.0381	0.0358	0.0348	0.0343
50	$\hat{\theta}_0$	0.0372	0.0281	0.0226	0.0214	0.0207
	$\hat{\theta}_k$	0.0247	0.0228	0.0220	0.0218	0.0216
	$\hat{\theta}_a$	0.0242	0.0227	0.0211	0.0207	0.0204
100	$\hat{\theta}_0$	0.0256	0.0170	0.0122	0.0110	0.0103
	$\hat{\theta}_k$	0.0123	0.0110	0.0106	0.0105	0.0103
	$\hat{\theta}_a$	0.0119	0.0110	0.0105	0.0103	0.0101
200	$\hat{\theta}_0$	0.0204	0.0118	0.0071	0.0059	0.0053
	$\hat{\theta}_k$	0.0062	0.0056	0.0053	0.0052	0.0051
	$\hat{\theta}_a$	0.0060	0.0055	0.0053	0.0051	0.0051

the conditional maximum likelihood estimators to construct a linear approximate maximum likelihood estimator.

Acknowledgements

The author is grateful to an anonymous referee for his careful comments.

References

Gersbakh, I. (1995). On the Fisher information in the Type-I censored and quantal response data. *Statistics and probability Letters,* **23**, 297–306.

Kendell, P. J. and Anderson, R. L. (1971). An estimation problem in life-testing. *Technometrics,* **13**, 289–301.

Kulldorff, G. (1961). *Estimation from Grouped and Partially Grouped Samples.* John Wiley & Sons, New York.

Meeker, W. Q. (1986). Planning life tests in which units are inspected for failure. *IEEE Transactions on Reliability,* **35**, 571–578.

Nelson, W. (1977). Optimum demonstration tests with grouped inspection data from exponential distribution. *IEEE Transactions on Reliability,* **36**, 226–231.

Nelson, W. (1982). *Applied Life Data Analysis.* John Wiley & Sons, New York.

Park, S. (1996). An asymptotic relation arising in the decomposition of the likelihood of order statistics. *Statistics and Probability Letters,* **29**, 101–106.

Park, S. (2003). On the asymptotic Fisher information in order statistics. *Metrika,* **57**, 71–80.

Park, S. and Kim, C. E. (2006). A note on the Fisher information in exponential distribution. *Communications in Statistics: Theory and Methods,* **35**, 13–19.

Seo, S. and Yum, B. (1993). Estimation methods for the mean of the exponential distribution based on grouped and censored data. *IEEE Transactions on Reliability,* **42**, 87–96.

Shapiro, S. S. and Gulati, S. (1996). Selecting failure monitoring times for an exponential life distribution. *Journal of Quality Technology,* **28**, 429–438.

Sukhatme, P. V. (1937). Tests of significance for samples of the χ^2 population with two degrees of freedom. *Annals of Eugenics,* **8**, 52–56.

Tallis, G. M. (1967). Approximate maximum likelihood estimates from grouped data. *Technometrics,* **9**, 599–606.

Zheng, G. and Gastwirth, J. L. (2000). Where is the Fisher information in an ordered sample?. *Statistica sinica,* **10**, 1267–1280.

[Received January 2007, Accepted April 2007]