7th Seminar on Probability and

Stochastic Processes

Isfahan University of Technology

Aug. 13-14, 2009

A Note on Type II Censored Data

M. Razmkhah

Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran, 91775-1159.

email: razmkhah_m@um.ac.ir

B. Khatib Astaneh

Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran, 91775-1159.

Abstract

In the procedure of collecting type II censored data the test time is an open-ended random variable. Hence, in some situations, it may be very time consuming to perform an experiment sequentially to collect type II censored data. Therefore, we develop procedure based on k independent samples. In this paper, two mentioned sampling schemes are compared in the sense of Fisher information in some classes of distributions.

MSC 2000: Primary 62N01, Secondary 62B10.

keywords: Proportional (reversed) hazard rate, Location family, Scale family, Shape family.

1 Introduction and Preliminaries

Suppose that $X_1, \cdots, X_n$ be a random sample from a population with continuous cumulative distribution function (cdf) $F(x, \theta)$ and probability density function (pdf) $f(x, \theta)$. Let $f_{r:n}(x, \theta)$ be pdf of the rth order statistic, $X_{r:n}$. In Type II censored data, n units are on test and the experiment terminates when the rth outcome is observed, where $1 \leq r < n$ is fixed. See Lawless [1] for more methods of lifetime data. Park [2], showed that Fisher information contained in Type II censored data in a single sample of size n with r outcomes is as follows

$$I_{1,\ldots,r:n}(\theta) = nI_X(\theta) - (n-r)\int_{-\infty}^{\infty} g(w, \theta) f_{r:n}(w, \theta) dw, \tag{1}$$

where

$$g(w, \theta) = \int_{w}^{\infty} \left(\frac{\partial}{\partial \theta} \log \frac{f(x, \theta)}{1 - F(w, \theta)}\right)^2 \frac{f(x, \theta)}{1 - F(w, \theta)} dw. \tag{2}$$

Asymptotic Fisher information in Type II censored data has been investigated by Chernoff *et al.* [3] and Zheng and Gastwirth [4]. Takahashi and Sugiura [5] studied the rate of convergence of Fisher information for Type II censored data. The asymptotics of maximum likelihood and related estimators based on Type II censored data has been studied by Bhattacharyya [6].

In the procedure of collecting type II censored data the test time is an open-ended random variable. Hence, in some situations, it may be very time consuming to perform an experiment sequentially to collect type II censored data. Therefore, we develop procedure based on k independent samples. Consider the situation in which we are reliability testing n (non repairable) units taken randomly from a population at k periods of time, which we call it a multi sample scheme. In this paper, we compare Type II censored data in a single sample and multi sample scheme in the sense of Fisher information about the unknown parameter of population. In the single sample, let n units are put on test and experiment is terminated when the rth outcome is observed. We call this design as plan **A**, hence we set $I^A(\theta) = I_{1,\ldots,r:n}(\theta)$. In the multi sample scheme, suppose k independent samples of sizes n_i, such that $\sum_{i=1}^k n_i = n$, are considered and in the ith sample, r_i outcomes are observed, for which $\sum_{i=1}^k r_i = r$ $(1 \le i \le k)$. We call the later design as plan **B**. Razmkhah *et al.* [7] compared two sampling schemes A and B for extracting record data with regard to Fisher information. Also, Ahmadi and Razmkhah [8] showed that the nonparametric confidence intervals for quantiles and quantile intervals from inversely sampled record breaking data in plan B have more confidence coefficient than plan A.

Let $X_{i,j}$, $(1 \le i \le k)$, $1 \le j \le n_i$, be k iid random samples with continuous cdf F. Also, Denote the first r_i order statistics obtained from the ith sample by $\mathbf{X}_i = (X_{1:n_i}, \cdots, X_{r_i:n_i})$, $(1 \le i \le k)$ such that $\sum_{i=1}^k n_i = n$ and $\sum_{i=1}^k r_i = r$. Note that the joint pdf of these variables is as follows

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_k; \theta) = \prod_{i=1}^k f(\mathbf{x}_i; \theta), \tag{3}$$

where $\mathbf{x}_i = (x_{1:n_i}, \ldots, x_{r_i:n_i})$ is the value of $\mathbf{X}_i$, $(1 \le i \le k)$. Therefore, amount of Fisher information contained in plan **B** is

$$\begin{aligned} I^B(\theta) &= \sum_{i=1}^k I_{\mathbf{X}_i}(\theta) \\ &= nI_X(\theta) - \sum_{i=1}^k (n_i - r_i) \int_{-\infty}^{\infty} g(w,\theta) f_{r_i:n_i}(w,\theta) dw, \end{aligned} \tag{4}$$

where $g(w,\theta)$ is defined in (2). According to the Eqs. (1) and (4), the following theorem deduces.

Theorem 1. *If $g(w,\theta)$, where is defined in (2), be a constant function in w, then*

$$I^B(\theta) = I^A(\theta).$$

The following lemma help us to simplify the results in the next sections.

Lemma 1. *Under regularity conditions, Eq. (2) can be rewrite as follows*

$$g(w,\theta) = -\int_w^{\infty} \left(\frac{\partial^2}{\partial\theta^2} \log \frac{f(x,\theta)}{1 - F(w,\theta)} \right) \frac{f(x,\theta)}{1 - F(w,\theta)} dx. \tag{5}$$

In this paper, we consider some classes of distributions and compare two sampling schemes A and B in collecting type II censored data with regard to Fisher information. In section 2, we study proportional hazard rate model. Proportional reversed hazard rate model is considered in Section 3. Location, scale and shape family of distributions are also investigated in Sections 4, 5 and 6, respectively.

2 Proportional hazard rate model

Let $X_{i,j}$'s, $(1 \le i \le k$ and $1 \le j \le n_i)$ be iid random variables with cdf

$$F(x;\theta) = 1 - [1 - G(x)]^{\alpha(\theta)}, \tag{6}$$

where $G(x)$ is an absolutely continuous distribution function. In the literature, the above model is well-known as proportional hazard rate model (PHRM), see for example [1]. Some lifetime distributions obey this model such as exponential, Weibull and Burr XII distributions.

Theorem 2. *In an PHRM, where is introduced by (6), both sampling schemes A and B provide same Fisher information about the unknown parameter θ.*

Proof. It is easy to investigate that for an PHRM, we have

$$g(w,\theta) = \left[\frac{\alpha'(\theta)}{\alpha(\theta)}\right]^2,$$

where $\alpha'(\theta)$ is the first derivative of $\alpha(\theta)$. Hence by Theorem 1, the result follows. □

3 Proportional reversed hazard rate model

Let $X_{i,j}$'s, $(1 \le i \le k$ and $1 \le j \le n_i)$ be iid random variables with cdf

$$F(x;\theta) = [G(x)]^{\beta(\theta)}, \tag{7}$$

where $G(x)$ is an absolutely continuous distribution function. In the literature, the above model is well-known as proportional reversed hazard rate model (PRHRM). A lot of lifetime distributions obey this model such as power distribution, exponentiated Weibull, exponentiated exponential, exponentiated Pareto and exponentiated Gamma family of distributions.

Hereafter, for more simplicity, we consider a sample of size nk in plan **A** with rk outcomes. Moreover, we assume that in plan **B**, $n_i = n$ and $r_i = r$, $(1 \le i \le k)$. Therefore, we have

$$I^B(\theta) - I^A(\theta) = k(n-r)\int_{-\infty}^{\infty} g(w,\theta)\big\{f_{rk:nk}(w,\theta) - f_{r:n}(w,\theta)\big\}\,dw. \tag{8}$$

Theorem 3. *Two sampling schemes A and B in collecting Type II censored data in an PRHRM can be compared in the sense of Fisher information about the unknown parameter θ as follows*

$$I^B(\theta) - I^A(\theta) = 2k(n-r)\left(\frac{\beta'(\theta)}{\beta(\theta)}\right)^2 \{\xi(n,r) - \xi(nk,rk)\}, \tag{9}$$

where

$$\xi(n,r) = r\binom{n}{r}\sum_{i=0}^{n-r-2}\binom{n-r-2}{i}\frac{(-1)^i}{(r+i+1)^3}. \tag{10}$$

Proof. By (5), in an PRHRM, we have

$$g(w,\theta) = -\frac{\partial^2}{\partial\theta^2}\log\beta(\theta) + \frac{\partial^2}{\partial\theta^2}\log\left(1-[G(w)]^{\beta(\theta)}\right) + \frac{\beta''(\theta)}{\beta(\theta)\left(1-[G(w)]^{\beta(\theta)}\right)}\left\{1+[G(w)]^{\beta(\theta)}\left(\log[G(w)]^{\beta(\theta)}-1\right)\right\}. \tag{11}$$

Therefore,

$$\int_{-\infty}^{\infty} g(w,\theta)f_{r:n}(w,\theta)\,dw = -\frac{\partial^2}{\partial\theta^2}\log\beta(\theta) + \frac{\beta''(\theta)}{\beta(\theta)}\psi(n,r) - 2\left(\frac{\beta'(\theta)}{\beta(\theta)}\right)^2\xi(n,r),$$

where $\xi(n,r)$ is defined in (10) and

$$\psi(n,r) = r\binom{n}{r}\sum_{i=0}^{n-r-1}\binom{n-r-1}{i}\frac{(-1)^i}{(r+i)(r+i+1)}.$$

Notice that for all n and r $(n > r)$, we have

$$\sum_{i=0}^{n-r-1}\binom{n-r-1}{i}\frac{(-1)^i}{r+i} = \int_0^1 u^{r-1}(1-u)^{n-r-1}\,du = \frac{(r-1)!(n-r-1)!}{(n-1)!}.$$

By some algebraic calculations, it can be deduced that, for all n and r $(n > r)$, $\psi(n,r) = 1$. Using (8), the result follows. □

To compare two sampling schemes A and B in the sense of Fisher information in an PRHRM, some numerical results of

$$\xi^*(n,r,k) = 2k(n-r)\{\xi(n,r) - \xi(nk,rk)\},$$

where $\xi(n,r)$ is defined in (10), are presented in Table 1 for some choices of n, r and k.

Table1: Values of $\xi^*(n,r,k)$ for some choices of n, r and k

n	r	k	$\xi^*(n,r,k)$
10	4	2	-0.227
5	2	4	-0.703
25	10	2	7587.106
10	4	5	7586.429
5	2	10	7585.238

It is observed from Table 1 that for different sample sizes and the number of outcomes, either sampling scheme A or B provide more Fisher information about the unknown parameter in an PRHRM.

4 Location family

In this section, a general formula is presented to compare two sampling schemes A and B in a location family.

Lemma 2. *Let $X_1,\cdots,X_n$ be a random sample from a population with cdf $F(x;\theta)=F_0(x-\theta)$, where F_0 dose not depend on θ. Then*

$$I^B(\theta)-I^A(\theta)=k(n-r)[\vartheta(nk,rk)-\vartheta(n,r)],$$

where

$$\begin{aligned}\vartheta(n,r) &= r\binom{n}{r}\left\{\int_{-\infty}^{\infty}\left(\frac{\partial^2}{\partial x^2}\log F_0(x)\right)[F_0(x)]^{r-1}f_0(x)[\bar{F}_0(x)]^{n-r}dx\right.\\ &\left.-\sum_{i=0}^{n-r-1}\binom{n-r-1}{i}\frac{(-1)^i}{r+i}\int_{-\infty}^{\infty}\left(\frac{\partial^2}{\partial x^2}\log f_0(x)\right)f_0(x)[F_0(x)]^{r+i}dx\right\}.\end{aligned}$$

Example 1. *Let $X_1,\cdots,X_n$ be a random sample from an extreme value distribution with cdf*

$$F(x;\theta)=e^{-e^{-(x-\theta)}},\quad -\infty<x<\infty. \tag{12}$$

Then

$$I^B(\theta)-I^A(\theta)=k(n-r)\{\vartheta_1(nk,rk)-\vartheta_1(n,r)\},$$

where

$$\vartheta_1(n,r)=1-2r\binom{n}{r}\sum_{i=0}^{n-r-2}\binom{n-r-2}{i}\frac{(-1)^i}{(r+i+1)^3}.$$

To compare two sampling schemes A and B in the sense of Fisher information in an extreme value distribution (12), some numerical results of

$$\vartheta^*(n,r,k)=k(n-r)\{\vartheta_1(nk,rk)-\vartheta_1(n,r)\}$$

are presented in Table 2.

5 Scale family

In this section, a scale family is considered and a general formula is presented in order to compare two sampling schemes A and B.

Lemma 3. *Let $X_1,\cdots,X_n$ be a random sample from a population with cdf $F(x;\theta)=F_0(\theta x)$, where F_0 dose not depend on θ. Then*

$$\theta^2[I^B(\theta)-I^A(\theta)]=k(n-r)\{\zeta(nk,rk)-\zeta(n,r)\},$$

where

$$\zeta(n,r) = 1 + r\binom{n}{r}\left\{\int_{-\infty}^{\infty} x^2\left(\frac{\partial^2}{\partial x^2}\log \bar{F}_0(x)\right)[F_0(x)]^{r-1} f_0(x)[\bar{F}_0(x)]^{n-r}dx - \sum_{i=0}^{n-r-1}\binom{n-r-1}{i}\frac{(-1)^i}{r+i}\int_{-\infty}^{\infty} x^2\left(\frac{\partial^2}{\partial x^2}\log f_0(x)\right) f_0(x)[F_0(x)]^{r+i}dx\right\}.$$

Example 2. *Let $X_1, \cdots, X_n$ be a random sample from an extreme value distribution with cdf*

$$F(x;\theta) = e^{-e^{-\theta x}}, \quad -\infty < x < \infty. \tag{13}$$

Then $F_0(x) = \exp(-e^{-x})$ and $f_0(x) = e^{-x}\exp(-e^{-x})$. Therefore, by Lemma 3, we have

$$\theta^2[I^B(\theta) - I^A(\theta)] = k(n-r)\{\zeta_1(nk, rk) - \zeta_1(n,r)\},$$

where

$$\zeta_1(n,r) = 1 + r\binom{n}{r}\left\{-\int_0^1 \log x[\log(-\log x)]^2(1 - x + \log x)x^r(1-x)^{n-r-2}dx - \sum_{i=0}^{n-r-1}\binom{n-r-1}{i}\frac{(-1)^i}{r+i}\int_0^1 \log x[\log(-\log x)]^2 x^{r+i}dx\right\}.$$

To compare two sampling schemes A and B in the sense of Fisher information in an extreme value distribution (13), some numerical results of

$$\zeta^*(n,r,k) = k(n-r)\{\zeta_1(nk, rk) - \zeta_1(n,r)\}$$

are presented in Table 2.

6 Shape family

In this section, two sampling schemes A and B are compared on the basis of a shape family with regard to Fisher information.

Lemma 4. *Let $X_1, \cdots, X_n$ be a random sample of a population with cdf $F(x;\theta) = F_0(x^\theta)$, where F_0 dose not depend on θ. Then*

$$\theta^2[I^B(\theta) - I^A(\theta)] = k(n-r)\{\delta(nk, rk) - \delta(n,r)\},$$

where

$$\delta(n,r) = 1 + r\binom{n}{r}\{\delta_1(n,r) - \delta_2(n,r)\}, \tag{14}$$

where

$$\delta_1(n,r) = \int_{-\infty}^{\infty} x\log^2 x\left(x\frac{\partial^2}{\partial x^2}\log \bar{F}_0(x) - \frac{f_0(x)}{\bar{F}_0(x)}\right)[F_0(x)]^{r-1} f_0(x)[\bar{F}_0(x)]^{n-r}dx \tag{15}$$

and

$$\delta_2(n,r) = \sum_{i=0}^{n-r-1}\frac{(-1)^i\binom{n-r-1}{i}}{r+i}\int_{-\infty}^{\infty} x\log^2 x\left(\frac{\partial}{\partial x}\log f_0(x) + x\frac{\partial^2}{\partial x^2}\log f_0(x)\right) f_0(x)[F_0(x)]^{r+i}dx \tag{16}$$

$$= \sum_{i=0}^{r-1}\frac{(-1)^i\binom{r-1}{i}}{n-r+i}\int_{-\infty}^{\infty} x\log^2 x\left(\frac{\partial}{\partial x}\log f_0(x) + x\frac{\partial^2}{\partial x^2}\log f_0(x)\right) f_0(x)\left(1 - [F_0(x)]^{n-r+i}\right)dx. \tag{17}$$

Example 3. *Let $X_1, \cdots, X_n$ be a random sample from a Weibull distribution with cdf*

$$F(x;\theta) = 1 - e^{-x^{\theta}}, \quad x > 0.$$

Then

$$\theta^2[I^B(\theta) - I^A(\theta)] = k(n-r)\{\delta_3(nk, rk) - \delta_3(n,r)\},$$

where

$$\begin{aligned}\delta_3(n,r) &= 1 + r\binom{n}{r}\Bigg\{(\gamma^2 - 2\gamma + \frac{\pi^2}{6})\sum_{i=0}^{r-1}\frac{(-1)^i\binom{r-1}{i}}{n-r+i+1} \\ &\quad - \sum_{i=0}^{r-1}\frac{(-1)^i\binom{r-1}{i}}{(n-r+i)(n-r+i+1)}\Big(\log^2(n-r+i+1) - 2(1-\gamma)\log(n-r+i+1)\Big)\Bigg\},\end{aligned}$$

and γ is Euler constant.

Proof. Note that $F_0(x) = 1 - e^{-x}$ and $f_0(x) = e^{-x}$. Therefore,

$$\frac{\partial^2}{\partial x^2}\log F_0(x) = \frac{\partial^2}{\partial x^2}\log f_0(x) = 0$$

and

$$\frac{\partial}{\partial x}\log f_0(x) = -1.$$

From Eq. (15), we have

$$\begin{aligned}\delta_1(n,r) &= -\int_0^{\infty} x(\log x)^2[1-e^{-x}]^{r-1}e^{-(n-r+1)x}dx \\ &= -\sum_{i=0}^{r-1}\binom{r-1}{i}(-1)^i\int_0^{\infty} x(\log x)^2 e^{-(n-r+i+1)x}dx. \qquad (18)\end{aligned}$$

Also, by Eq. (17),

$$\delta_2(n,r) = -\sum_{i=0}^{r-1}\binom{r-1}{i}\frac{(-1)^i}{n-r+i}\int_0^{\infty} x(\log x)^2 e^{-x}\Big(1 - e^{-(n-r+i)x}\Big)dx. \qquad (19)$$

Using the identities

$$\int_0^{\infty} e^{-x}\log x\, dx = -\gamma,$$

$$\int_0^{\infty} xe^{-x}\log x\, dx = 1 - \gamma,$$

$$\int_0^{\infty} xe^{-x}(\log x)^2\, dx = \gamma^2 - 2\gamma + \frac{\pi^2}{6}$$

and substituting (18) and (19) in (14), the result follows. □

In order to compare two sampling schemes A and B in the sense of Fisher information in a Weibull distribution, some numerical results of

$$\delta^*(n,r,k) = k(n-r)\{\delta_3(nk, rk) - \delta_3(n,r)\}$$

are presented in Table 2.

Example 4. *Let $X_1, \cdots, X_n$ be a random sample from a Burr XII distribution with cdf*

$$F(x;\theta) = 1 - (1 + x^{\theta})^{-1}.$$

Then

$$\theta^2[I^B(\theta) - I^A(\theta)] = k(n-r)\{\delta_4(nk, rk) - \delta_4(n,r)\},$$

where

$$\delta_4(n,r) = 1 + r\binom{n}{r}\Big\{ - \int_0^{\infty} \frac{x^r}{(1+x)^{n+3}} \log^2 x dx + 2\sum_{i=0}^{r-1}\binom{r-1}{i}\frac{(-1)^i}{n-r+i}\int_0^{\infty}\frac{x}{(1+x)^i}\log^2 x\Big(1 - \frac{1}{(1+x)^{n-r+i}}\Big)dx\Big\}.$$

To compare two sampling schemes A and B in the sense of Fisher information in a Burr XII distri tion, some numerical results of

$$\delta^b(n,r,k) = k(n-r)\{\delta_4(nk, rk) - \delta_4(n,r)\}$$

are presented in Table 2.

Table2: Values of $\vartheta^*(n,r,k)$, $\zeta^*(n,r,k)$, $\delta^*(n,r,k)$ and $\delta^b(n,r,k)$ for some choices of n, r and k.

n	r	k	$\vartheta^*(n,r,k)$	$\zeta^*(n,r,k)$	$\delta^*(n,r,k)$	$\delta^b(n,r,k)$
10	4	2	-0.227	0.253	0.186	0.443
5	2	4	-0.703	0.543	0.454	1.033
20	10	2	-0.128	0.242	9.112	-4697997
10	5	4	-0.402	0.773	8.361	-4697996
8	4	5	-0.547	1.003	9.452	-4697996
4	2	10	-1.318	1.822	10.852	-4697995
25	10	2	7587.106	-40490.72	-1191.372	-2170.455
10	4	5	7586.429	-40489.94	-1190.801	-2164.862
5	2	10	7585.238	-40489.21	-1190.130	-2163.388

Acknowledgement

The authors would like to thank the anonymous referee for his/her useful comments and careful readi the manuscript.

References

[1] LAWLESS, J. L., *Statistical methods and methods for lifetime data*, Wiley, New Yourk, 1982.

[2] PARK, S., *Fisher information in order statistics*, J. Amer. Statist. Assoc., 91 (1996) 385-390.

[3] CHERNOFF, H., GASTWIRTH, J. L., JOHNS JR., M. V., *Asymptotic distribution of linear comb nations of order statistics*, Ann. Math. Statist., 31 (1967) 52-72.

[4] ZHENG, G., GASTWIRTH, J. L., *Where is the Fisher information in an ordered sample?*, Statis Sinica, 10 (2000) 1267-1280.

[5] TAKAHASHI, H., SUGIURA, N., *The rate of convergence of Fisher information for Type II censor sample*, J. Japan Statist. Soc., 19 (1989) 139-144.

[6] BHATTACHARYYA, G. K., *The asymptotics of maximum likelihood and related estimators based o Type II censored data*, J. Amer. Statist. Assoc., 80 (1985) 398-404.

[7] RAZMKHAH, M., AHMADI, J., KHATIB, B., *Comparison of two sampling schemes for extracti record data with regard to Fisher information* (Farsi), J. Statist. Sci., 1 (2007) 19-44.

[8] AHMADI, J., RAZMKHAH, M., *Nonparametric confidence intervals for quantiles and quantile inte vals from inversely sampled record breaking data in a multi-sample plan*, J. Statist. Comput. Simu (2009), in press.