

Reconstruction of Order Statistics in Exponential Distribution

M. Razmkhah, B. Khatib, Jafar Ahmadi

Department of Statistics, Ordered and Spatial Data Center of Excellence,
Ferdowsi University of Mashhad, Iran.

(razmkhah_m@um.ac.ir, khatib_b@yahoo.com, ahmadi-j@um.ac.ir.)

Abstract. In this article, a new censoring scheme is considered, namely, a middle part of a random sample is censored. A treatment for reconstructing the missing order statistics is investigated. The proposed procedure is studied in detail under exponential distribution which is widely used as a constant failure model in reliability. Different approaches are used to obtain point and interval reconstructors and then they are compared. A numerical example is presented for illustrating all the proposed inferential procedures. Eventually, we present some remarks including how the results of the paper can be used when the parameters of the exponential distribution are unknown.

Key words and phrases: Beta distribution, highest conditional density, Markov property, pivotal quantity.

1 Introduction

There are some situations in life testing and reliability experiments in which a middle part of the sample (or subjects) are lost or removed from the experiment. For example, in a life testing experiment, suppose n items are placed on test simultaneously. The first few observations may be observed at the beginning of the experiment, then some other data points may be censored due to negligence or problems, while the last few observations are recorded. In such situations, the experimenter may not obtain complete information on failure times for all experimental units. What motivated us to write the paper is that “How can one reconstruct the missing observations?”. This scheme is the complementary to the idea of double censoring in which the middle part of the sample is actually stored. The doubly censored data model has been studied by several authors, see for example, Fernández (2004) and Sun *et al.* (2008). Meanwhile, the proposed scheme can be considered as a special case of middle censoring model which was first introduced by Jammalamadaka and Mangalam (2003) for which all random intervals are the same. The later plan is studied by Jammalamadaka and Iyer (2004) and Mangalam *et al.* (2008) in nonparametric set up and Iyer *et al.* (2008) in the parametric set up.

Now, suppose n independent and identical units are placed on a life test with corresponding lifetimes X_1, \dots, X_n with probability density function (pdf) f and cumulative distribution function (cdf) F . Denote the i th order statistic of the sample X_1, \dots, X_n by Y_i . Assume that some order statistics are lost, that is we only observed the data set

$$\mathbf{Y} = \{Y_1, \dots, Y_r, Y_s, \dots, Y_n\},$$

where $0 \leq r < s \leq n + 1$, for convenience of notation, we let $Y_0 = 0$ and $Y_{n+1} = \infty$. If $r = 0$, we indeed observe $\{Y_s, \dots, Y_n\}$ which coincides with the left censoring model and in the case $s = n + 1$, the data set $\{Y_1, \dots, Y_r\}$ is observed where is the same as Type II censored data. In life testing and survival analysis, several researches have been done based on left and right censored data. The main goal of this paper is to reconstruct the value of Y_l for $r < l < s$ based on observed ordered data \mathbf{Y} while the underlying distribution is exponential.

A random variable X is said to have a two-parameter exponential distribution, which we shall write $X \sim \text{Exp}(\mu, \sigma)$, if its cdf is

$$F(x) = 1 - e^{-\frac{x-\mu}{\sigma}}, \quad x \geq \mu, \quad \sigma > 0, \quad (1)$$

where μ and σ are the location and scale parameters, respectively. It is well known that the exponential distribution is the simplest and most important distribution in reliability studies, and is applied in a wide variety of statistical procedures, especially in life testing problems. See Balakrishnan and Basu (1995) for some researchs based on this distribution.

The rest of this paper is as follows: In Section 2, some preliminaries are presented. Three point reconstructors for the censored data points of the exponential distribution are given in Section 3. Section 4 is focused on the interval reconstruction. In order to illustrate the proposed scheme, we present a numerical example in Section 5. At the end, we present some remarks including how the results of the paper can be used when the parameters of the exponential distribution are unknown.

2 Some Preliminaries

Here, we present some well known properties of order statistics which will be used to obtain the new results in the next sections. Let X_1, \dots, X_n be independent and identically distributed (iid) continuous random variables with cdf $F_\theta(x)$ and pdf $f_\theta(x)$. In order to reconstruct the l th ($r < l < s$) order statistic based on the data set \mathbf{Y} , we propose various methods and obtain some reconstructors of Y_l . First of all, notice that by Markove property of order statistics (see, David and Nagaraja, 2003), the conditional pdf of Y_l given \mathbf{Y} is equivalent to that of Y_l given Y_r and Y_s . Hence,

$$\begin{aligned} f_{Y_l|\mathbf{Y}}(y_l|\mathbf{y}) &= f_{Y_l|Y_r, Y_s}(y_l|y_r, y_s) \\ &= \frac{(F(y_l) - F(y_r))^{l-r-1} (F(y_s) - F(y_l))^{s-l-1}}{B(l-r, s-l) (F(y_s) - F(y_r))^{s-r-1}} f(y_l), \\ &\qquad\qquad\qquad y_r < y_l < y_s, \end{aligned} \tag{2}$$

where $\mathbf{y} = (y_1, \dots, y_r, y_s, \dots, y_n)$ is the observed value of \mathbf{Y} and $B(a, b)$ is the complete beta function.

From Eq. (2), it is deduced that the distribution of Y_l given \mathbf{Y} is just the distribution of the $(l-r)$ th order statistic in a sample of size $s-r-1$ drawn from a population with pdf $\frac{f(x)}{F(y_s)-F(y_r)}$, $y_r < x < y_s$ (i.e., from the parent distribution truncated in the tails at y_r and y_s);

Also for $0 \leq r < l < s \leq n + 1$, we have

$$\frac{F(Y_l) - F(Y_r)}{F(Y_s) - F(Y_r)} \mid \mathbf{Y} \sim \text{Beta}(l - r, s - l), \tag{3}$$

where $\text{Beta}(a, b)$ denotes a beta distribution with parameters $a > 0$ and $b > 0$. It is obvious that the conditional distribution of $\frac{F(Y_l) - F(Y_r)}{F(Y_s) - F(Y_r)}$ given \mathbf{Y} is identical to the unconditional distribution of the $(l - r)$ th order statistic in a sample of size $s - r - 1$ from the standard uniform distribution.

Now, suppose that $Z_{r,s} \sim \text{Beta}(s - r, n - s + 1)$, then for $0 < m < 1$ we have

$$\begin{aligned} E[\log(1 - mZ_{r,s})] &= \frac{1}{B(s - r, n - s + 1)m^{n-r}} \\ &\times \sum_{i=0}^{s-r-1} \sum_{j=0}^{n-s} \binom{s-r-1}{i} \binom{n-s}{j} (-1)^i \\ &\times \frac{(m-1)^{n-s-j}}{(i+j+1)^2} \left\{ (1-m)^{i+j+1} \left(1 - (i+j+1) \log(1-m) \right) - 1 \right\} \\ &= \varphi_1(r, s, m), \text{ say.} \end{aligned} \tag{4}$$

Also

$$\begin{aligned} E[\log^2(1 - mZ_{r,s})] &= \frac{2}{B(s - r, n - s + 1)m^{n-r}} \\ &\times \sum_{i=0}^{s-r-1} \sum_{j=0}^{n-s} \binom{s-r-1}{i} \binom{n-s}{j} \frac{(-1)^i (m-1)^{n-s-j}}{(i+j+1)^3} \\ &\times \left\{ 1 - (1-m)^{i+j+1} \left(1 - (i+j+1) \log(1-m) \right) \right. \\ &\quad \left. + 0.5(i+j+1)^2 \log^2(1-m) \right\} \\ &= \varphi_2(r, s, m), \text{ say.} \end{aligned} \tag{5}$$

3 Point Reconstruction

To reconstruct the missing order statistics from a middle part of a random sample based on the data set \mathbf{Y} , in this section we propose three approaches and then compare them.

3.1 Convex Combination Reconstructor

If the data set $\mathbf{Y} = \{Y_1, \dots, Y_r, Y_s, \dots, Y_n\}$ is observed. Using the structure dependence properties of order statistics, it is logical to assume that a convex combination (CC) of Y_r and Y_s may be considered as a reconstructor for Y_l , denoted by $\hat{Y}_{l,CC}$, i.e.,

$$\hat{Y}_{l,CC} = wY_r + (1 - w)Y_s, \quad r < l < s, \quad 0 < w < 1.$$

The main question arises: “how can one choose w ”? It is reasonable to select the optimal value of w which can be determined by minimizing the mean squared error (MSE) of $\hat{Y}_{l,CC}$. Notice that

$$\begin{aligned} \text{MSE}(\hat{Y}_{l,CC}) &= \text{E}(\hat{Y}_{l,CC} - Y_l)^2 \\ &= \text{E}[(1 - w)W_{l,s} - wW_{r,l}]^2 \\ &= w^2\text{E}(W_{r,l}^2) + (1 - w)^2\text{E}(W_{l,s}^2) \\ &\quad - 2w(1 - w)\text{E}(W_{r,l})\text{E}(W_{l,s}), \end{aligned} \tag{6}$$

where $W_{r,l} = Y_l - Y_r$ and the last equality deduces from the independency of $W_{r,l}$ and $W_{l,s}$ in an exponential model (spacing of order statistics). We recall that if X_1, \dots, X_n are iid random variables from $\text{Exp}(\mu, \sigma)$, then $W_{r,l}$ can be expressed as

$$W_{r,l} = \sum_{i=r+1}^l \frac{Z_i}{n - i + 1}, \tag{7}$$

where Z_i 's are iid random variables from $\text{Exp}(0, \sigma)$. Using (7), we readily have

$$\sigma^{-1}\text{E}(W_{r,l}) = \sum_{i=r+1}^l \frac{1}{n - i + 1} = \varphi_3(r, l), \text{ say} \tag{8}$$

and

$$\sigma^{-2}\text{E}(W_{r,l}^2) = \sum_{i=r+1}^l \frac{1}{(n - i + 1)^2} + \varphi_3^2(r, l) = \varphi_4(r, l), \text{ say.} \tag{9}$$

Substituting (8) and (9) into (6), we have

$$\begin{aligned} \sigma^{-2}\text{MSE}(\hat{Y}_{l,CC}) &= w^2\varphi_4(r, l) + (1 - w)^2\varphi_4(l, s) \\ &\quad - 2w(1 - w)\varphi_3(r, l)\varphi_3(l, s). \end{aligned} \tag{10}$$

The optimal value of w may be obtained by minimizing (10) with respect to w which is given by

$$w_{opt} = \frac{\varphi_4(l, s) + \varphi_3(r, l)\varphi_3(l, s)}{\varphi_4(r, l) + \varphi_4(l, s) + 2\varphi_3(r, l)\varphi_3(l, s)}, \quad (11)$$

where $\varphi_3(r, l)$ and $\varphi_4(r, l)$ are defined in (8) and (9), respectively. For given n, r, s and l , one can easily find the values of w_{opt} from (11). We derive the values of $\sigma^{-2}\text{MSE}(\hat{Y}_{l,CC})$ for $n = 10$ and some selected values of r, s and l with corresponding w_{opt} , the results are presented in Table 1.

3.2 Conditional Median Reconstructor

In the context of prediction, the conditional median prediction approach has been studied by several authors, see for example Raqab and Nagaraja (1995), Raqab *et al.* (2007) and Ahmadi *et al.* (2009). Therefore, intuitively, the median of the conditional density of Y_l given \mathbf{Y} can be considered as a reconstructor of Y_l . So we say that $\hat{Y}_{l,CM}$ is a conditional median (CM) reconstructor of Y_l , if $P\{Y_l \leq \hat{Y}_{l,CM} | \mathbf{Y}\} = P\{Y_l \geq \hat{Y}_{l,CM} | \mathbf{Y}\}$. Using (3), we have

$$\hat{Y}_{l,CM} = F^{-1}\{F(Y_r) + \text{med}(V_{r,l,s})[F(Y_s) - F(Y_r)]\}, \quad (12)$$

where $V_{r,l,s} \sim \text{Beta}(l-r, s-l)$ and $\text{med}(X)$ stands for the median of X .

By (12), the CM reconstructor of the l th order statistic in the exponential distribution (1) is given by

$$\hat{Y}_{l,CM} = Y_r - \sigma \log \left(1 - \text{med}(V_{r,l,s})(1 - e^{-W_{r,s}/\sigma}) \right).$$

It is obvious that the pdf of $W_{r,l}$ is

$$f_{W_{r,l}}(w) = \frac{e^{-(n-l+1)w/\sigma}(1 - e^{-w/\sigma})^{l-r-1}}{\sigma B(l-r, n-l+1)}, \quad w > 0. \quad (13)$$

Using (13), we obtain the MSE of $\hat{Y}_{l,CM}$ which is given by

$$\begin{aligned} \sigma^{-2}\text{MSE}(\hat{Y}_{l,CM}) &= \varphi_4(r, l) + \varphi_2(r, s, \text{med}(V_{r,l,s})) \\ &\quad + 2\varphi_5(r, l, s, \text{med}(V_{r,l,s})), \end{aligned}$$

where $\varphi_2(r, l)$ and $\varphi_4(r, s, \cdot)$ are defined in (5) and (9), respectively, and

$$\begin{aligned} &\varphi_5(r, l, s, m) \\ &= \frac{n!}{(r-1)!(l-r-1)!(s-l-1)!(n-s)!} \\ &\quad \int_0^1 \int_z^1 \int_y^1 \log\left(\frac{x}{y}\right) \log\left(1 - m\left(1 - \frac{z}{x}\right)\right) \\ &\quad (1-x)^{r-1} (x-y)^{l-r-1} (y-z)^{s-l-1} z^{n-s} dx dy dz \\ &= \frac{(n-r)!}{(l-r-1)!(s-l-1)!(n-s)!} \\ &\quad \sum_{i=0}^{l-r-1} \sum_{j=0}^{s-l-1} \binom{l-r-1}{i} \binom{s-l-1}{j} (-1)^{s-l-1+i-j} \\ &\quad \times \left\{ \frac{1}{(i+j+1)^2} \left(\varphi_6(r, l, s, m, -j-1) - \varphi_6(r, l, s, m, i) \right) \right. \\ &\quad \left. - \frac{1}{i+j+1} \left(\varphi_6(r, l, s, m, i) \log m + \varphi_7(r, l, s, m, i) \right) \right\}, \end{aligned} \tag{14}$$

where

$$\begin{aligned} \varphi_6(r, l, s, m, i) &= \frac{1}{m^{n-l+i+1}} \sum_{k=0}^{n-l+i} \binom{n-l+i}{k} \frac{(m-1)^{n-l+i-k}}{(k+1)^2} \\ &\quad \times \left\{ (1-m)^{k+1} [1 - (k+1) \log(1-m)] - 1 \right\} \end{aligned}$$

and

$$\begin{aligned} \varphi_7(r, l, s, m, i) &= \frac{1}{m^{n-l+i+1}} \sum_{k=0}^{n-l+i} \binom{n-l+i}{k} (m-1)^{n-l+i-k} \\ &\quad \times \int_0^{-\log(1-m)} x e^{-(k+1)x} \log(m-1+e^{-x}) dx. \end{aligned}$$

The numerical values of $\sigma^{-2} \text{MSE}(\hat{Y}_{l,CM})$ are presented in Table 1 for $n = 10$ and some selected values of r, s and l .

3.3 Unbiased Conditional Reconstructor

It is logical to consider $E[Y_l|\mathbf{Y}]$ as a reconstructor of Y_l having observed \mathbf{Y} . Whis this in mind, using (3), we consider

$$\hat{Y}_{l,UC} = F^{-1}\left(F(Y_r) + \frac{l-r}{s-r}[F(Y_s) - F(Y_r)]\right), \tag{15}$$

as an unbiased conditional (UC) reconstructor of Y_l . From (15), the UC reconstructor of the l th order statistic in $\text{Exp}(\mu, \sigma)$ is given by

$$\hat{Y}_{l,UC} = Y_r - \sigma \log\left(1 - \frac{l-r}{s-r}(1 - e^{-W_{r,s}/\sigma})\right).$$

Using (13), we obtain $\text{MSE}(\hat{Y}_{l,UC})$ which is stated as

$$\sigma^{-2}\text{MSE}(\hat{Y}_{l,UC}) = \varphi_4(r, l) + \varphi_2\left(r, s, \frac{l-r}{s-r}\right) + 2\varphi_5\left(r, l, s, \frac{l-r}{s-r}\right),$$

where $\varphi_2(r, s, \cdot)$, $\varphi_4(r, l)$ and $\varphi_5(r, l, s, \cdot)$ are defined in (5), (9) and (14), respectively.

We computed the numerical values of $\sigma^{-2}\text{MSE}(\hat{Y}_{l,UC})$ for $n = 10$ and some choices of r, s and l , which are calculated in 4 decimal places using the package R, the results are presented in Table 1.

Table 1. The numerical values of $\sigma^{-2}\text{MSE}(\hat{Y}_{l,CC})$, $\sigma^{-2}\text{MSE}(\hat{Y}_{l,CM})$ and $\sigma^{-2}\text{MSE}(\hat{Y}_{l,UC})$ for $n = 10$.

| s | $r = 3$ | | | | | $r = 4$ | | | |
|-----|---|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| | l | | | | | l | | | |
| | 4 | 5 | 6 | 7 | 8 | 5 | 6 | 7 | 8 |
| 5 | 0.0118 ^a 0.0097 ^b 0.0097 ^c | | | | | | | | |
| 6 | 0.0158 0.0165 0.0157 | 0.0221 0.0244 0.0234 | | | | 0.0165 0.0165 0.0165 | | | |
| 7 | 0.0180 0.0189 0.0176 | 0.0337 0.0256 0.0256 | 0.0375 0.0376 0.0375 | | | 0.0222 0.0230 0.0219 | 0.0331 0.0315 0.0325 | | |
| 8 | 0.0194 0.0203 0.0188 | 0.0414 0.0399 0.0392 | 0.0621 0.0488 0.0530 | 0.0670 0.0638 0.0634 | | 0.0253 0.0263 0.0245 | 0.0509 0.0488 0.0488 | 0.0622 0.0589 0.0609 | |
| 9 | 0.0207 0.0212 0.0195 | 0.0479 0.0447 0.0434 | 0.0823 0.0704 0.0704 | 0.1205 0.1008 0.1045 | 0.1409 0.1241 0.1301 | 0.0276 0.0282 0.0261 | 0.0637 0.0593 0.0582 | 0.1060 0.0942 0.0958 | 0.1335 0.1192 0.1254 |

| s | r = 5 | | |
|---|----------------------------|----------------------------|----------------------------|
| | l | | |
| | 6 | 7 | 8 |
| 5 | | | |
| 6 | | | |
| 7 | 0.0246 0.0246 0.0246 | | |
| 8 | 0.0332 0.0343 0.0325 | 0.0547 0.0528 0.0541 | |
| 9 | 0.0383 0.0391 0.0365 | 0.0859 0.0798 0.0798 | 0.1233 0.1121 0.1173 |

a , b and c indicate for $\sigma^{-2}\text{MSE}(\hat{Y}_{l,CC})$, $\sigma^{-2}\text{MSE}(\hat{Y}_{l,CM})$ and $\sigma^{-2}\text{MSE}(\hat{Y}_{l,UC})$, respectively.

From Table 1, it is observed that

1. For $l = r + 1 < s - 1$, $\text{MSE}(\hat{Y}_{l,CC}) < \text{MSE}(\hat{Y}_{l,CM})$, but for $r + 1 < l < s$ it is usually vice versa.
2. It is usually observed that $\text{MSE}(\hat{Y}_{l,UC}) < \text{MSE}(\hat{Y}_{l,CC})$.
3. For fixed r and s , the MSEs of $\hat{Y}_{l,CC}$, $\hat{Y}_{l,CM}$ and $\hat{Y}_{l,UC}$ increase as l increases.
4. For fixed r and l , the MSEs of $\hat{Y}_{l,CC}$, $\hat{Y}_{l,CM}$ and $\hat{Y}_{l,UC}$ increase as s increases.
5. For fixed s and l , the MSEs of $\hat{Y}_{l,CC}$, $\hat{Y}_{l,CM}$ and $\hat{Y}_{l,UC}$ decrease as r increases.
6. For fixed r and s , the MSE of $\hat{Y}_{l,UC}$ is usually less than, equal to and greater than that of $\hat{Y}_{l,CM}$ when l is less than, equal to and greater than $\frac{r+s}{2}$, respectively. Specially, when $s - r$ is even, $\text{MSE}(\hat{Y}_{\frac{r+s}{2},CM}) = \text{MSE}(\hat{Y}_{\frac{r+s}{2},UC})$.

4 Reconstruction Interval

In this section, we are interested in finding the reconstruction intervals for the l th ($r < l < s$) order statistic in terms of observed data set \mathbf{Y} . Two methods are proposed and then compared in view of shortest width criterion.

4.1 Conditional Reconstruction Interval

We say that the interval $[L, U]$ is the exact $100(1 - \alpha_1 - \alpha_2)\%$ conditional reconstruction interval (CRI) for Y_l given \mathbf{Y} if $P(Y_l \geq L|\mathbf{Y}) = 1 - \alpha_1$ and $P(Y_l \geq U|\mathbf{Y}) = \alpha_2$. Using (3), it can be shown that

$$L_{CRI} = F^{-1}\{F(Y_r) + B(l - r, s - l, \alpha_1)[F(Y_s) - F(Y_r)]\}, \quad (16)$$

and

$$U_{CRI} = F^{-1}\{F(Y_r) + B(l - r, s - l, 1 - \alpha_2)[F(Y_s) - F(Y_r)]\}, \quad (17)$$

where $B(l - r, s - l, \gamma)$ is the $100\gamma\%$ lower percentile of the Beta($l - r, s - l$) distribution, i.e., $P[V_{r,l,s} \leq B(l - r, s - l, \gamma)] = \gamma$.

For two-parameter exponential distribution, using the Eqs. (16) and (17), we obtain the lower and upper bounds of the exact $100(1 - \alpha_1 - \alpha_2)\%$ CRI, respectively, as

$$L_{CRI} = Y_r - \sigma \log \left(1 - B(l - r, s - l, \alpha_1)(1 - e^{-W_{r,s}/\sigma}) \right), \quad (18)$$

and

$$U_{CRI} = Y_r - \sigma \log \left(1 - B(l - r, s - l, 1 - \alpha_2)(1 - e^{-W_{r,s}/\sigma}) \right). \quad (19)$$

Hence, the expected width of the CRI, $E(W_{CRI}) = E(U_{CRI} - L_{CRI})$, is

$$\begin{aligned} E(W_{CRI}) &= \sigma E \left\{ \log \left(\frac{1 - B(l - r, s - l, \alpha_1)(1 - e^{-W_{r,s}/\sigma})}{1 - B(l - r, s - l, 1 - \alpha_2)(1 - e^{-W_{r,s}/\sigma})} \right) \right\} \\ &= \sigma \{ \varphi_1(r, s, B(l - r, s - l, \alpha_1)) \\ &\quad - \varphi_1(r, s, B(l - r, s - l, 1 - \alpha_2)) \}, \end{aligned} \quad (20)$$

where $\varphi_1(r, s, \cdot)$ is defined in (4). Similarly,

$$\begin{aligned} E[W_{CRI}^2] &= \sigma^2 \{ \varphi_2(r, s, B(l - r, s - l, \alpha_1)) \\ &\quad - \varphi_2(r, s, B(l - r, s - l, 1 - \alpha_2)) \\ &\quad - 2\varphi_1(r, s, B(l - r, s - l, \alpha_1)) \\ &\quad \varphi_1(r, s, B(l - r, s - l, 1 - \alpha_2)) \}, \end{aligned} \quad (21)$$

where $\varphi_2(r, s, \cdot)$ is defined in (5). Using (20) and (21), we can obtain variance of the width of the CRI.

Table 2 shows the numerical values of $\sigma^{-1}E(W_{CRI})$ and $\sigma^{-2}\text{Var}(W_{CRI})$ for $\alpha_1 = \alpha_2 = 0.1$, $n = 10$ and some choices of r , s and l .

Table 2. The numerical values of $\sigma^{-1}E(W_{CRI})$ and $\sigma^{-2}\text{Var}(W_{CRI})$ for the 80% CRI, when $n = 10$.

| s | $r = 3$ | | | | $r = 4$ | | | | |
|-----|--|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| | l | | | | l | | | | |
| | 4 | 5 | 6 | 7 | 8 | 5 | 6 | 7 | 8 |
| 5 | 0.2455 ^a 0.0348 ^b | | | | | | | | |
| 6 | 0.2885 0.0244 | 0.3443 0.0747 | | | | 0.2898 0.0474 | | | |
| 7 | 0.3024 0.0149 | 0.4375 0.0575 | 0.4461 0.1345 | | | 0.3397 0.0316 | 0.4200 0.1057 | | |
| 8 | 0.3037 0.0126 | 0.4775 0.0375 | 0.5889 0.1018 | 0.5891 0.2293 | | 0.3548 0.0185 | 0.5317 0.0759 | 0.5712 0.1996 | |
| 9 | 0.3133 0.0012 | 0.4988 0.0215 | 0.6570 0.0614 | 0.7920 0.1580 | 0.8314 0.3941 | 0.3612 0.0099 | 0.5782 0.0434 | 0.7488 0.1332 | 0.8159 0.3643 |

| s | $r = 5$ | | | |
|-----|---------|------------------|------------------|------------------|
| | l | | | |
| | 6 | 7 | 8 | |
| 5 | | | | |
| 6 | | | | |
| 7 | | 0.3534 0.0683 | | |
| 8 | | 0.4126 0.0418 | 0.5376 0.1606 | |
| 9 | | 0.4295 0.0215 | 0.6759 0.1005 | 0.7910 0.3246 |

^a and ^b indicate for the value of $\sigma^{-1}E(W_{CRI})$ and $\sigma^{-2}\text{Var}(W_{CRI})$, respectively.

From Table 2, it is observed that

1. For fixed r and s , $E(W_{CRI})$ and $\text{Var}(W_{CRI})$ increase as l increases.
2. For fixed r and l , $E(W_{CRI})$ increases and $\text{Var}(W_{CRI})$ decreases as s increases.
3. For fixed l and s , $E(W_{CRI})$ and $\text{Var}(W_{CRI})$ decrease as r increases.

4.2 Highest Conditional Density Reconstruction Interval

Similar to the idea of highest conditional density prediction interval, the highest conditional density reconstruction interval (HCDRI) may be considered such that the conditional pdf of Y_l given \mathbf{Y} for every point inside the interval is greater than that for every point outside of it. If the conditional pdf of Y_l given \mathbf{Y} is unimodal, it is sufficient to

derive an optimal $100(1 - \alpha)\%$ reconstruction interval $[L^*, U^*]$, such that

$$\begin{cases} P\{L^* < Y_l < U^* | \mathbf{Y} = \mathbf{y}\} = 1 - \alpha, \\ f_{Y_l | \mathbf{Y}}(L^* | \mathbf{y}) = f_{Y_l | \mathbf{Y}}(U^* | \mathbf{y}). \end{cases} \quad (22)$$

According to the Eqs. in (22) and using (3), the lower and upper bounds of the HCDRI can be determined such that the following equations are held

$$\begin{cases} B\left(l - r, s - l, \frac{F(U^*) - F(Y_r)}{F(Y_s) - F(Y_r)}\right) - B\left(l - r, s - l, \frac{F(L^*) - F(Y_r)}{F(Y_s) - F(Y_r)}\right) = 1 - \alpha, \\ \left(\frac{F(L^*) - F(Y_r)}{F(U^*) - F(Y_r)}\right)^{l-r-1} \left(\frac{F(Y_s) - F(L^*)}{F(Y_s) - F(U^*)}\right)^{s-l-1} = \frac{f(U^*)}{f(L^*)}. \end{cases}$$

Taking $v_1 = \frac{F(L^*) - F(Y_r)}{F(Y_s) - F(Y_r)}$ and $v_2 = \frac{F(U^*) - F(Y_r)}{F(Y_s) - F(Y_r)}$, the HCDRI with coefficient $1 - \alpha$ for Y_l given \mathbf{Y} is

$$\left(F^{-1}\left[F(Y_r) + v_1(F(Y_s) - F(Y_r))\right], F^{-1}\left[F(Y_r) + v_2(F(Y_s) - F(Y_r))\right] \right), \quad (23)$$

where v_1 and v_2 can be determined by solving the following two equations

$$\begin{cases} B(l - r, s - l, v_2) - B(l - r, s - l, v_1) = 1 - \alpha, \\ \left(\frac{v_1}{v_2}\right)^{l-r-1} \left(\frac{1-v_1}{1-v_2}\right)^{s-l-1} = \frac{f\left(F^{-1}\left[F(Y_r) + v_2(F(Y_s) - F(Y_r))\right]\right)}{f\left(F^{-1}\left[F(Y_r) + v_1(F(Y_s) - F(Y_r))\right]\right)}. \end{cases}$$

Now, let X_1, \dots, X_n be iid random variables with exponential distribution (1). Using (23), the HCDRI with coefficient $1 - \alpha$ for Y_l given \mathbf{Y} can be found as follows

$$\left[Y_r - \sigma \log\left(1 - v_1(1 - e^{-W_{r,s}/\sigma})\right), Y_r - \sigma \log\left(1 - v_2(1 - e^{-W_{r,s}/\sigma})\right) \right], \quad (24)$$

such that v_1 and v_2 are the solutions of the following two equations

$$\begin{cases} B(l - r, s - l, v_2) - B(l - r, s - l, v_1) = 1 - \alpha, \\ \left(\frac{v_1}{v_2}\right)^{l-r-1} \left(\frac{1-v_1}{1-v_2}\right)^{s-l-1} = \frac{1 - v_2(1 - e^{-W_{r,s}/\sigma})}{1 - v_1(1 - e^{-W_{r,s}/\sigma})}, \end{cases} \quad (25)$$

provided $1 > v_2 > v_1 > 0$, otherwise the HCDRI does not exist. Here we consider three special cases as follows.

Case 1. Suppose that only one order statistic in a sample of size n is missed and we are looking to reconstruct it. That is, we consider the reconstruction of Y_l on the basis of the data set \mathbf{Y} such that $r = l - 1$ and $s = l + 1$. In this case, the conditional pdf of Y_l given \mathbf{Y} is

$$f_{Y_l|\mathbf{Y}}(y_l|\mathbf{y}) = \frac{e^{-y_l/\sigma}}{\sigma} \{e^{-y_r/\sigma} - e^{-y_s/\sigma}\}^{-1}, \quad y_r < y_l < y_s,$$

which is a decreasing function of y_l . Hence, there is not any two-sided HCDRIs for Y_l given \mathbf{Y} . But, one can obtain a one-sided one with coefficient $1 - \alpha$ as follows

$$\left[Y_r, Y_r - \sigma \log \left(1 - (1 - \alpha)(1 - e^{-W_{r,s}/\sigma}) \right) \right]. \quad (26)$$

The expected width and variance of the one-sided HCDRI (26) are given by $-\sigma\varphi_1(r, s, 1 - \alpha)$ and $\sigma^2\{\varphi_2(r, s, 1 - \alpha) - \varphi_1(r, s, 1 - \alpha)^2\}$, respectively, where $\varphi_1(r, s, \cdot)$ and $\varphi_2(r, s, \cdot)$ are defined in (4) and (5), respectively.

Case 2. Suppose that only two order statistics in a sample of size n are missed and we are looking to reconstruct the smallest one. That is, we consider the reconstruction of Y_l based on the data set \mathbf{Y} such that $r = l - 1$ and $s = l + 2$. Notice that in this case, the conditional pdf of Y_l given \mathbf{Y} is a decreasing function of y_l on the interval (y_r, y_s) . Similar to the Case 1, a one-sided HCDRI with coefficient $1 - \alpha$ can be found which is

$$\left[Y_r, Y_r - \sigma \log \left(1 - (1 - \sqrt{\alpha})(1 - e^{-W_{r,s}/\sigma}) \right) \right]. \quad (27)$$

The expected width and variance of the one-sided HCDRI (27) are $-\sigma\varphi_1(r, s, 1 - \sqrt{\alpha})$ and $\sigma^2\{\varphi_2(r, s, 1 - \sqrt{\alpha}) - [\varphi_1(r, s, 1 - \sqrt{\alpha})]^2\}$, respectively.

Case 3. Suppose that the assumption of Case 2 holds and we are attempting to reconstruct the largest one. That is, we consider the reconstruction of Y_l in terms of the data set \mathbf{Y} such that $r = l - 2$ and $s = l + 1$. In this case, the conditional pdf of Y_l given \mathbf{Y} is

$$f_{Y_l|\mathbf{Y}}(y_l|\mathbf{y}) = 2 \frac{e^{-y_r/\sigma} - e^{-y_l/\sigma}}{(e^{-y_r/\sigma} - e^{-y_s/\sigma})^2} \frac{e^{-y_l/\sigma}}{\sigma}, \quad y_r < y_l < y_s. \quad (28)$$

It can be shown that the conditional pdf in (28) is an increasing function of y_l on the interval (y_r, y_s) whenever $W_{r,s} = w_{r,s} < \sigma \log 2$, otherwise it is a unimodal pdf. Therefore, we consider the following two cases.

(i) $W_{r,s} = w_{r,s} < \sigma \log 2$. In this case, a one-sided HCDRI for Y_l given \mathbf{Y} with coefficient $1 - \alpha$ may be determined as follows

$$\left[Y_r - \sigma \log \left(1 - \sqrt{\alpha} (1 - e^{-W_{r,s}/\sigma}) \right), Y_s \right]. \tag{29}$$

The expected width of the one-sided HCDRI (29), denoted by $E(W_1)$, is

$$E(W_1) = \sigma \{ \varphi_1(r, s, \sqrt{\alpha}) + \varphi_3(r, s) \}, \tag{30}$$

where $\varphi_1(r, s, \cdot)$ and $\varphi_3(r, s)$ are defined in (4) and (8), respectively. Also,

$$E(W_1^2) = \sigma^2 \{ \varphi_4(r, s) + \varphi_2(r, s, \sqrt{\alpha}) + 2\varphi_8(r, s, \sqrt{\alpha}) \}, \tag{31}$$

where $\varphi_2(r, s, \cdot)$ and $\varphi_4(r, s)$ are defined in (5) and (9), respectively, and

$$\begin{aligned} \varphi_8(r, s, m) &= \frac{1}{B(s-r, n-s+1)m^{n-r}} \\ &\sum_{i=0}^{s-r-1} \sum_{j=0}^{n-s} \binom{s-r-1}{i} \binom{n-s}{j} (-1)^i \\ &\times (m-1)^{n-s-j} \left\{ \frac{\log m}{(i+j+1)^2} \right. \\ &\left[1 - (1-m)^{i+j+1} \left(1 - (i+j+1) \log(1-m) \right) \right] \\ &\left. - \int_0^{-\log(1-m)} x e^{-(i+j+1)x} \log(m-1+e^{-x}) dx \right\}. \tag{32} \end{aligned}$$

(ii) $W_{r,s} = w_{r,s} > \sigma \log 2$. In this case a two-sided HCDRI for Y_l given \mathbf{Y} can be found in the form of (24). By (25), we find v_1 and v_2 by solving the following two equations

$$\begin{cases} v_2^2 - v_1^2 = 1 - \alpha, \\ v_1 \left(1 - v_1 (1 - e^{-W_{r,s}/\sigma}) \right) = v_2 \left(1 - v_2 (1 - e^{-W_{r,s}/\sigma}) \right). \end{cases} \tag{33}$$

Then,

$$\begin{cases} v_1 = \frac{1-(1-\alpha)(1-e^{-W_{r,s}/\sigma})^2}{2(1-e^{-W_{r,s}/\sigma})}, \\ v_2 = \frac{1+(1-\alpha)(1-e^{-W_{r,s}/\sigma})^2}{2(1-e^{-W_{r,s}/\sigma})}. \end{cases} \quad (34)$$

It is obvious that $v_1 < v_2$. By investigating the conditions in (25), i.e., $0 < v_1, v_2 < 1$, we deduce that an HCDRI exists if and only if

$$1 - \alpha < (1 - 2e^{-W_{r,s}/\sigma})(1 - e^{-W_{r,s}/\sigma})^{-2},$$

provided $W_{r,s} = w_{r,s} > \sigma \log 2$. Then, by substituting (34) into (24), the HCDRI is given by

$$\left[Y_r - \sigma \log \left(\frac{1 + (1 - \alpha)(1 - e^{-W_{r,s}/\sigma})^2}{2} \right), Y_r - \sigma \log \left(\frac{1 - (1 - \alpha)(1 - e^{-W_{r,s}/\sigma})^2}{2} \right) \right]. \quad (35)$$

The expected width of the two-sided HCDRI (35), denoted by $E(W_2)$, is

$$E(W_2) = \sigma \{ \varphi_9(r, s, \alpha - 1) - \varphi_9(r, s, 1 - \alpha) \}, \quad (36)$$

where

$$\begin{aligned} \varphi_9(r, s, m) &= \frac{1}{B(s - r, n - s + 1)} \\ &\sum_{i=0}^{n-s} \binom{n-s}{i} (-1)^i \int_0^1 \log(1 - mx^2) x^{s-r+i-1} dx. \end{aligned}$$

Also,

$$E(W_2^2) = \sigma^2 \{ \varphi_{10}(r, s, \alpha - 1) + \varphi_{10}(r, s, 1 - \alpha) - 2\varphi_{11}(r, s, 1 - \alpha) \}, \quad (37)$$

where

$$\begin{aligned} \varphi_{10}(r, s, m) &= \frac{1}{B(s - r, n - s + 1)} \\ &\sum_{i=0}^{n-s} \binom{n-s}{i} (-1)^i \int_0^1 x^{s-r+i-1} \log^2(1 - mx^2) dx \end{aligned}$$

and

$$\varphi_{11}(r, s, m) = \frac{1}{B(s-r, n-s+1)} \sum_{i=0}^{n-s} \binom{n-s}{i} (-1)^i \int_0^1 x^{s-r+i-1} \log(1-mx^2) \log(1+mx^2) dx.$$

Now, we can determine the expected value and the variance of the width of the HCDRI for the Case 3. Denote the width of the HCDRI in this case by W_{HCDRI} , then we have, for $k \geq 1$,

$$E(W_{HCDRI}^k) = p(n, r, s)E(W_1^k) + (1 - p(n, r, s))E(W_2^k), \quad (38)$$

where from the relation between beta and binomial distributions, we have

$$p(n, r, s) = P(W_{r,s} < \sigma \log 2) = 2^{r-n} \sum_{i=s}^n \binom{n-r}{i-r}.$$

Using (30), (31), (36), (37) and (38), the exact values of $\sigma^{-1}E(W_{HCDRI})$ and $\sigma^{-2}\text{Var}(W_{HCDRI})$ can be obtained. The results are presented in Table 3 for 80% HCDRI when $n = 10$.

Table 3. The numerical values of $\sigma^{-1}E(W_{HCDRI})$ and $\sigma^{-2}\text{Var}(W_{HCDRI})$ for 80% HCDRI when $n = 10$.

| s | r = 3 | | r = 4 | | r = 5 | |
|---|---------------------|--------|--------|--------|--------|--------|
| | l | | l | | l | |
| | 4 | 5 | 5 | 6 | 6 | 7 |
| 5 | 0.2347 ^a | | | | | |
| | 0.0245 ^b | | | | | |
| 6 | 0.2389 | 0.3109 | 0.2745 | | | |
| | 0.0135 | 0.4152 | 0.0328 | | | |
| 7 | | | 0.2787 | 0.3838 | 0.3315 | |
| | | | 0.0170 | 0.5168 | 0.0457 | |
| 8 | | | | | 0.3342 | 0.5029 |
| | | | | | 0.0217 | 0.6297 |

^a and ^b indicate for $\sigma^{-1}E(W_{HCDRI})$ and $\sigma^{-2}\text{Var}(W_{HCDRI})$, respectively.

Remark 4.1. Comparing Tables 2 and 3, it is observed that for fixed r, l and s , the expected width of the HCDRI is less than that of the corresponding CRI.

5 Numerical Example

To illustrate the performance of the proposed reconstructors in Sections 3 and 4, we use the data presented in Table 4 denoting the ordered observations of a random sample of size 10 generated from $\text{Exp}(2, 5)$.

Table 4. Ordered observations of a random sample of size 10 generated from $\text{Exp}(2, 5)$.

| | | | | | | | | | | |
|-------|--------|--------|--------|--------|--------|--------|---------|---------|---------|---------|
| i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Y_i | 2.2293 | 2.5244 | 3.0421 | 4.1165 | 5.9887 | 6.3241 | 10.7144 | 13.3795 | 14.7893 | 18.3202 |

Suppose that we only observe $\{Y_1, \dots, Y_4, Y_7, \dots, Y_{10}\}$. Then, for $l = 5, 6$, we reconstruct Y_l using different mentioned approaches in the previous sections. As shown in subsection 4.4, by (27) we find a one-sided HCDRI for Y_5 . Also, we can obtain a two-sided HCDRI for Y_6 with reconstruction coefficient at most $\frac{1-2e^{-w_{4,7}/\sigma}}{(1-e^{-w_{4,7}/\sigma})^2} = 0.867$ provided $W_{4,7} = w_{4,7} > 5 \log 2 = 3.466$, for which $\sigma = 5$ is known and $w_{4,7} = y_7 - y_4 = 6.5979$.

In the case that the scale parameter σ is unknown, one can plug in the common estimator of σ , see the next section, so using Eq. (43), we can obtain $\text{MLE}(\sigma) \simeq 5.5056$. In this case an HCDRI for Y_6 can be found with coefficient at most 0.8134.

Table 5 contains the values of different reconstructors of Y_l (The reconstruction coefficient $1 - \alpha = 0.80$ is considered).

Table 5. The values of different reconstructors of Y_l .

| l | Y_l (exact value) | $\hat{Y}_{l,CC}$ | $\hat{Y}_{l,CM}$ | $\hat{Y}_{l,UC}$ | $CRI(Y_l)$ | $HCDRI(Y_l)$ |
|-----|---------------------|------------------|---------------------|------------------|-------------------|-------------------|
| 5 | 5.9887 | 5.8037 | 5.3244 ^a | 5.5167 | (4.3081, 7.5926) | (4.1165, 6.7129)* |
| | | | 5.3763 ^b | 5.5753 | (4.3174, 7.6903) | (4.1165, 6.8021)* |
| 6 | 6.3241 | 7.9144 | 7.7669 | 7.4686 | (5.4345, 10.0562) | (5.7955, 10.3889) |
| | | | 7.8647 | 7.5658 | (5.4903, 10.0964) | (6.1192, 10.6552) |

^a, ^b and * indicate for σ is known, σ is unknown and one-sided HCDRI, respectively.

From Table 5, it is observed that

1. Among the point reconstructors of Y_5 , $\hat{Y}_{5,CC}$ is the closest to Y_5 .
2. Among the point reconstructors of Y_6 , $\hat{Y}_{6,UC}$ is the closest to Y_6 .
3. The width of HCDRI for Y_5 and Y_6 are less than those of CRIs in the both cases that σ is known or unknown.

6 Concluding Remark

In this paper, we have tackled the problem of reconstruction of missing order data. Several methods proposed and we applied them to reconstruct the missing order statistics based on the data set $\mathbf{Y} = \{Y_1, \dots, Y_r, Y_s, \dots, Y_n\}$ from a two-parameter exponential distribution. Notice that by (1) and (2), we have

$$f_{Y_l|\mathbf{Y}}(y_l|\mathbf{y}) = \frac{(e^{-y_r/\sigma} - e^{-y_l/\sigma})^{l-r-1} (e^{-y_l/\sigma} - e^{-y_s/\sigma})^{s-l-1} e^{-y_l/\sigma}}{B(l-r, s-l)(e^{-y_r/\sigma} - e^{-y_s/\sigma})^{s-r-1} \sigma}. \quad (39)$$

The density function in (39) does not depend on the location parameter μ . The reconstructors were obtained in the case that σ is known. If the scale parameter is unknown, as mentioned in Balakrishnan *et al.* (2009), we can substitute a common estimator of σ , for example maximum likelihood estimator (MLE), in the corresponding formulas.

Let X_1, \dots, X_n be iid random variables with cdf $F_\theta(x)$ and pdf $f_\theta(x)$, where θ is an unknown parameter. Then, the likelihood function of θ on the basis of \mathbf{Y} is

$$L(\theta) = \frac{n!}{(s-r-1)!} \prod_{i \in \Delta_{r,s}} f_\theta(y_i) [F_\theta(y_s) - F_\theta(y_r)]^{s-r-1}, \quad (40)$$

where $\Delta_{r,s} = \{1, \dots, r, s, \dots, n\}$.

For the exponential distribution (1), we consider three cases as follows:

Case 1. $0 = r < s \leq n$ (left censored sample)

As mentioned in Section 1, this case coincides with the left censored sample. Then, $\text{MLE}(\mu) = Y_s$ and hence using (40),

$$\text{MLE}(\sigma) = \frac{1}{n-s+1} \sum_{i=s+1}^n W_{s,i}.$$

Case 2. $1 \leq r < s = n+1$ (right censored sample)

This special case coincides with the Type II censored sample. Then $\text{MLE}(\mu) = Y_1$ and hence using (40),

$$\text{MLE}(\sigma) = \frac{1}{r} \left\{ \sum_{i=1}^{r-1} W_{1,i} + (n-r+1)W_{1,r} \right\}.$$

Case 3. $1 \leq r < s \leq n$.

Using (1), we have $\text{MLE}(\mu) = Y_1$ and by (40) the MLE of σ is the solution of the following equation

$$(e^{W_{r,s}/\sigma} - 1)(A - (n - (s - r) + 1)\sigma) = (s - r - 1)W_{r,s}, \quad (41)$$

where

$$A = \sum_{i \in \Delta_{r,s}} W_{1,i} + (s - r - 1)W_{1,r}. \quad (42)$$

The solution Eq. (41) in terms of σ has no any closed form. Expanding the exponential function e^x in a Taylor series, we obtain an approximation for the MLE of σ . By using the first two terms of the Taylor series, we find

$$\text{MLE}(\sigma) = \frac{1}{n} \left\{ \sum_{i \in \Delta_{r,s}} W_{1,i} + (s - r - 1)W_{1,r} \right\}. \quad (43)$$

If more precision is needed in the approximation, then we use more terms of Taylor series expansion. Using the first j ($j > 2$) terms of this series, the MLE of σ can be approximately found by solving the following polynomial equation

$$\sum_{i=2}^{j-1} \frac{\sigma^{j-i}}{i!} (iAW_{r,s}^{i-2} - (n - (s - r) + 1)W_{r,s}^{i-1}) + \frac{AW_{r,s}^{j-2}}{(j-1)!} = n\sigma^{j-1},$$

where A is defined in (42).

Acknowledgments

The authors thank the referees and the associate editor for their useful comments and constructive criticisms on the original version of this manuscript which led to this considerably improved version. “Partial support from Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad” is acknowledged.

References

Ahmadi, J., Jafari Jozani, M., Marchand, E., and Parsian, A. (2009), Prediction of k-records from a general class of distributions under balanced type loss functions. *Metrika*, **70**, 19–33.

- Balakrishnan, N., Doostparast, M., and Ahmadi, J. (2009), Reconstruction of past records. *Metrika*, **70**, 89–109.
- Balakrishnan, N. and Basu, A. P. (1995), *Exponential Distribution: Theory, Methods and Applications*. New York: Taylor & Francis.
- David, H. A. and Nagaraja, H. N. (2003), *Order Statistics*. Third edition, New York: Wiley.
- Fernández, A. J. (2004), One and two sample prediction based on doubly censored exponential data and prior information. *Test*, **13**, 403–416.
- Iyer, S. K., Jammalamadaka, S. R., and Kundu, D. (2008), Analysis of middle-censored data with exponential lifetime distributions. *Journal of Statistical Planning and Inference*, **138**, 3550–3560.
- Jammalamadaka, S. R. and Iyer, S. K. (2004), Approximate self consistency for middle-censored data. *Journal of Statistical Planning and Inference*, **124**, 75–86.
- Jammalamadaka, S. R. and Mangalam, V. (2003), Nonparametric estimation for middle-censored data. *Journal of Nonparametric Statistics*, **15**, 253–265.
- Mangalam, V., Nair, G. M., and Zhao, Y. (2008), On computation of NPMLE for middle-censored data. *Statistics & Probability Letters*, **78**, 1452–1458.
- Raqab M. and Nagaraja H. N. (1995), On some predictors of future order statistics. *Metron*, **53**, 185–204.
- Raqab, M., Ahmadi, J., and Doostparast, M. (2007), Statistical inference based on record data from Pareto model. *Statistics*, **41**, 105–118.
- Sun, X., Zhou, X., and Wang, J. (2008), Confidence intervals for the scale parameter of exponential distribution based on type II doubly censored samples. *Journal of Statistical Planning and Inference*, **138**, 2045–2058.