Journal of the Korean Data & Information Science Society
2014, **25**(2), 431–438

http://dx.doi.org/10.7465/jkdi.2014.25.2.431
한국데이터정보과학회지

Estimation for the Rayleigh distribution based on Type I hybrid censored sample†

Byongwon Kwon[1] · Kyeongjun Lee[2] · Youngseuk Cho[3]

[1]Digital Marketing Team, Shinsegae Department Store Centumcity
[23]Department of Statistics, Pusan National University

Received 8 January 2014, revised 24 January 2014, accepted 3 February 2014

Abstract

Type I hybrid censoring scheme is the combination of the Type I and Type II censoring scheme introduced by Epstein (1954). Epstein considered a hybrid censoring sampling scheme in which the life testing experiment is terminated at a random time T^*, which is the time that happens first among the following two; time of the kth unit is observed or time of the experiment length set in advance. The likelihood function of this scheme from the Rayleigh distribution cannot be solved in a explicit solution and thus we approximate the function by the Taylor series expansion. In this process, we propose four different methods of expansion skill.

Keywords: Approximate maximum likelihood estimator, Rayleigh distribution, Taylor series expansion, Type I hybrid censoring scheme.

1. Introduction

Consider a life-testing experiment where n identical units are put. Assume that $X_1, X_2, \cdots,$ X_n denote the corresponding lifetimes from a distribution. The ordered lifetimes of these units are denoted by $X_{1:n}, X_{2:n}, \cdots, X_{n:n}$. The probability density function (p.d.f) $g(x)$ and cumulative density function (c.d.f) $G(x)$ of the random variable X having the Rayleigh distribution are given by

$$g(x) = \frac{x}{\sigma^2}\exp\left[-\frac{x^2}{2\sigma^2}\right],\ x > 0,\ \sigma > 0, \tag{1.1}$$

and

$$G(x) = 1 - \exp\left[-\frac{x^2}{2\sigma^2}\right],\ x > 0,\ \sigma > 0, \tag{1.2}$$

---

† This study was supported by the Research Fund Program of Research Institute for Basic Sciences, Pusan National University, Korea, 2011, Project No. RIBS-PNU-2013 302.

[1] Assistant manager, Digital Marketing Team, Shinsegae Department Store Centumcity, Busan 612-020, Korea.
[2] Instructor, Department of Statistics, Pusan National University, Busan 609-735, Korea.
[3] Corresponding author: Professor, Department of Statistics, Pusan National University, Busan 609-735, Korea. E-mail: choys@pusan.ac.kr

respectively, where σ is a scale parameter. The Rayleigh distribution is a suitable model for life testing studies. Polovko (1968), Dyer and Whisenand (1973) demonstrated the importance of this distribution in electro vacuum devices and communication engineering. One major application of this model is used in analyzing wind speed data. This statistical model was first introduced by Rayleigh (Rayleigh, 1880). Siddiqui (1962) discussed the origin and properties of the Rayleigh distribution. Several authors have contributed to this model. Balakrishnan studied approximate maximum likelihood estimator (MLE) of Rayleigh distribution from the Type II censored sample in 1989. Han and Kang (2006a) obtained a approximate MLE of known parameter Rayleigh distribution from the multiply Type II censored sample. Han and Kang (2006b) obtained a approximate MLE of two parameter Rayleigh distribution from the multiply Type II censored sample. Kang and Jung (2009) obtained a approximate MLE of double Rayleigh distribution from the progressive Type II censored sample. Kim and Han (2009) discussed estimation of the scale parameter of the Rayleigh distribution under general progressive censoring. Lee *et al*. (2011) obtained a Bayes estimator under the Rayleigh distribution with the progressive Type II right censored sample.

Type I hybrid censoring schemes are the combination of the Type I and Type II censoring scheme introduced by Epstein (1954). Epstein considered a hybrid censoring sampling scheme in which the life testing experiment is terminated at a random time T^*, which is the time that those happens first among the following two; time of the kth unit is observed or time of the experiment length set in advance. In other words, $T^* = \min\{x_{r+k:n}, T\}$, where $r \in \{0, 1, 2, 3, \cdots, n\}$, $T \in (0, \infty)$ is fixed in advance, $X_{r:n}$ denote the rth ordered failure time when the sample size is n, k is the number of failures wanted to be observed in advance.

The rest of paper consists as follows; In section 2, we introduce Type I hybrid censoring scheme. In section 3, we derive approximate MLEs of the scale parameter σ for the Rayleigh distribution from the Type I hybrid censoring samples. The scale parameter is estimated with approximate MLE skill and four different way of Taylor series expansion is used. In Secton 4, the description of different estimators that are compared by performing the Monte Carlo simulation is presented.

2. Type I hybrid censoring scheme

Type I hybrid censoring scheme described as follows. Fix $1 < r < n$, and set terminating time T and number of units to be observed k in advance. If the $(r+k)$th unit occurs before time T, the experiment terminates at the $(r+k)$th point. This is the same as Type II censoring scheme. If the $(r+k)$th unit occurs after time T, the experiment terminates at the time T. In case of this, final observation is $X_{r+s:n}$. This is the same as Type I censoring scheme.

If the failure time of the units are following Rayleigh distribution, whose p.d.f is (1.1), the likelihood function based on the Type I hybrid censored data is given by

$$L = \frac{n!}{r!(n-r-D)!} \left[G(x_{r+1:n})\right]^r \left[1 - G(T^*)\right]^{n-r-D} \left[\prod_{i=r+1}^{r+D} g(x_{i:n})\right], \tag{2.1}$$

where $D = k$, if $x_{r+k:n} < T$ and $T = s$, otherwise.

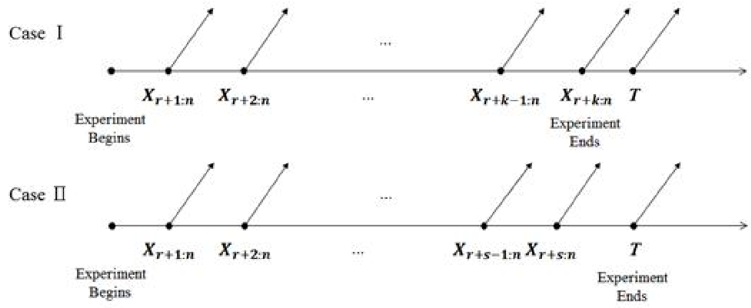

Figure 2.1 Scheme of Type I hybrid censoring data

Let $z_{i:n} = x_{i:n}/\sigma$ and $T^*_\sigma = T^*/\sigma$, where $\sigma > 0$. Then the p.d.f and c.d.f of Rayleigh function can be obtained as

$$f(z_{i:n}) = z_{i:n}\text{exp}\left[-\frac{z_{i:n}^2}{2}\right], \tag{2.2}$$

and

$$F(z_{i:n}) = 1 - \exp\left[-\frac{z_{i:n}^2}{2}\right], \tag{2.3}$$

respectively.

Also, the likelihood function (2.1) can be written as

$$L = \frac{n!}{r!(n-r-D)!}\sigma^{-D}\left[F(z_{r+1:n})\right]^r\left[1-F(T^*_\sigma)\right]^{n-r-D}\left[\prod_{i=r+1}^{r+D} f(z_{i:n})\right], \tag{2.4}$$

From (2.4), the log-likelihood function can be expressed as

$$\text{ln}L = C - D\text{ln}\sigma + r\text{ln}F(z_{r+1:n}) + (n-r-D)\text{ln}\left[1-F(T^*_\sigma)\right] + \sum_{i=r+1}^{r+D} f(z_{i:n}). \tag{2.5}$$

On differentiating the log-likelihood function (2.5) with respect to σ and equating to zero, we obtain the estimating equation as follows;

$$\begin{aligned}\frac{d\text{ln}L}{d\sigma} &= -\frac{1}{\sigma}\left[2D + rz_{r+1:n}\frac{f(z_{r+1:n})}{F(z_{r+1:n})} - (n-r-D)T^{*2}_\sigma - \sum_{i=r+1}^{r+D} z_{i:n}^2\right] \\ &= 0.\end{aligned} \tag{2.6}$$

However, the equation (2.6) cannot be solved in a explicit solution for σ unless $r = 0$. To obtain approximate, but explicit solution of the likelihood function (2.6), we may use the Taylor series expansion.

3. Approximate maximum likelihood estimator

Balakrishnan (1989) suggested approximate MLEs which is the alternate of the MLE for the uncalculative likelihood function. See for example the work of Cho *et al*. (2013), Kang and Lee (2013), and Lee and Lee (2012).

We expand $f(z_{r+1:n})/F(z_{r+1:n})$ or $z_{r+1:n}f(z_{r+1:n})/F(z_{r+1:n})$ in (2.6) in a Taylor series expansion on the point of $\xi = F^{-1}(p) = \sqrt{-2\ln q}$, where $p = (r+1)/(n+1)$ and $q = 1-p$.

Since there cannot exist an explicit from of (2.6), Blakrishnan (1989) derived approximate MLE with the expansion of $f(z_{r+1:n})/F(z_{r+1:n})$. We use different ways of approxiamtion; approximate $F(z_{r+1:n}) \simeq p$ before the Taylor series expansion, and expand $z_{r+1:n}f(z_{r+1:n})/F(z_{r+1:n})$ at once. Four different expansion are as follows;

$$\begin{aligned}
\frac{f(z_{r+1:n})}{F(z_{r+1:n})} &\simeq \frac{f(\xi)}{F(\xi)} + \left[\frac{f(\xi)}{F(\xi)}\right]'(z_{r+1:n}-\xi) \\
&= \frac{\xi(1-p)}{p} + \left[\frac{f'(\xi)F(\xi) - f^2(xi)}{F^2(\xi)}\right](z_{r+1:n}-\xi) \\
&= \left[\frac{\xi q}{p} - \frac{q}{p}\left(1+\frac{2\ln q}{p}\right)\xi\right] + \left[\frac{q}{p}\left(1+\frac{2\ln q}{p}\right)\right] z_{r+1:n} \\
&= \alpha_1 + \beta_1 z_{r+1:n}, \qquad (3.1)
\end{aligned}$$

$$\begin{aligned}
f(z_{r+1:n}) &\simeq \frac{1}{p}[f(\xi) + f'(\xi)(z_{r+1:n}-\xi)] \\
&= \frac{\xi(1-p)}{p} + \left[\frac{f'(\xi)}{p}\right](z_{r+1:n}-\xi) \\
&= \left[\frac{q}{p}(-2\ln q)^{1.5}\right] + \left[\frac{q}{p}(1+2\ln q)\right] z_{r+1:n} \\
&= \alpha_2 + \beta_2 z_{r+1:n}, \qquad (3.2)
\end{aligned}$$

$$\begin{aligned}
z_{r+1:n}\frac{f(z_{r+1:n})}{F(z_{r+1:n})} &\simeq \frac{\xi f(\xi)}{F(\xi)} + \left[\frac{\xi f(\xi)}{F(\xi)}\right]'(z_{r+1:n}-\xi) \\
&= \frac{\xi^2 q}{p} + \frac{\xi q}{p}\left(2+\frac{2\ln q}{p}\right)(z_{r+1:n}-\xi) \\
&= \left[\frac{\xi^2 q}{p} - \frac{\xi^2 q}{p}\left(2+\frac{2\ln q}{p}\right)\right] + \left[\frac{\xi q}{p}\left(2+\frac{2\ln q}{p}\right)\right] z_{r+1:n} \\
&= \alpha_3 + \beta_3 z_{r+1:n}, \qquad (3.3)
\end{aligned}$$

$$\begin{aligned}
z_{r+1:n} f(z_{r+1:n}) &\simeq \frac{z_{r+1:n} f(z_{r+1:n})}{p} \\
&= \frac{1}{p}[\xi f(\xi) + [\xi f(\xi)]'(z_{r+1:n} - \xi)] \\
&= \frac{\xi^2 q}{p} + \frac{1}{p}\left[f(\xi) + \xi f'(\xi)\right](z_{r+1:n} - \xi) \\
&= \frac{\xi^2 q}{p} + \frac{\xi q}{p}(2 + 2\ln q)(z_{r+1:n} - \xi) \\
&= \left[\frac{\xi^2 q}{p} - \frac{\xi^2 q}{p}(2 + 2\ln q)\right] + \left[\frac{\xi q}{p}(2 + 2\ln q) z_{r+1:n}\right] \\
&= \alpha_4 + \beta_4 z_{r+1:n}, \qquad (3.4)
\end{aligned}$$

where

$$\begin{aligned}
\alpha_1 &= \frac{q}{p^2}(-2\ln q)^{3/2}, \quad \beta_1 = \frac{q}{p}\left(1 + \frac{2\ln q}{p}\right), \\
\alpha_2 &= \frac{q}{p}(-2\ln q)^{3/2}, \quad \beta_2 = \frac{q}{p}\left(1 + 2\ln q\right), \\
\alpha_3 &= 2\frac{q\ln q}{p}\left(1 + \frac{2\ln q}{p}\right), \quad \beta_3 = 2\frac{q}{p}\sqrt{-2\ln q}\left(1 + \frac{\ln q}{p}\right), \\
\alpha_4 &= \frac{q}{p}(2\ln q)(1 + 2\ln q), \quad \beta_4 = 2\frac{q}{p}\sqrt{-2\ln q}\left(1 + \ln q\right).
\end{aligned}$$

By substituting the equations (3.1) and (3.2) into (2.6), we may approximate the equation (2.6) by

$$\begin{aligned}
\frac{d\ln L}{d\sigma} &\simeq -\frac{1}{\sigma}\left[2D + r z_{r+1:n}\left(\alpha_j + \beta_j z_{r+1:n}\right) - (n - r - D)T_\sigma^{*2} - \sum_{i=r+1}^{r+D} z_{i:n}^2\right] \\
&= 0, \qquad (3.5)
\end{aligned}$$

where $j = 1, 2$.

By solving equation (3.5) for σ, we derive the approximate MLEs of σ as

$$\hat{\sigma}_j = \frac{-B_{1,j} + \sqrt{B_{1,j}^2 + 8DC_{1,j}}}{4D}, \qquad (3.6)$$

where $B_{1,j} = r\alpha_j x_{r+1:n}$, $C_{1,j} = -r\beta_j x_{r+1:n}^2 + (n - r - D)T^{*2} + \sum_{i=r+1}^{r+D} x_{i:n}^2$, $j = 1, 2$.

By substituting the equations (3.3) and (3.4) into (2.6), we may approximate the equation (2.6) by

$$\begin{aligned}
\frac{d\ln L}{d\sigma} &\simeq -\frac{1}{\sigma}\left[2D + r\left(\alpha_j + \beta_j z_{r+1:n}\right) - (n - r - D)T_\sigma^{*2} - \sum_{i=r+1}^{r+D} z_{i:n}^2\right] \\
&= 0, \qquad (3.7)
\end{aligned}$$

where $j = 3, 4$.

By solving equation (3.7) for σ, we derive the approximate MLEs of σ as

$$\hat{\sigma}_j = \frac{-B_{2,j} + \sqrt{B_{2,j}^2 + 4A_jC_{2,j}}}{2A_j}, \tag{3.8}$$

where $A_j = 2D + r\alpha_j$, $B_{2,j} = r\beta_j x_{r+1:n}$, $C_{2,j} = (n-r-D)T^{*2} + \sum_{i=r+1}^{r+D} x_{i:n}^2$, $j = 3, 4$.

4. Illustrative example and simulation results

4.1. Illustrative example

In this example, we analyze the ball bearing data, which was given by Caroni (2002) and represents the failure times of 25 ball bearings in the endurance test. The observed failure times are shown in Table 4.1. For this data set, Raqab and Wilson (2002) indicated that the one-parameter Rayleigh distribution provides a satisfactory fit.

Table 4.1 Failures of 25 ball bearing data for example

i	1	2	3	4	5	6	7	8	9	10
X_i	0.1788	0.2892	0.3300	0.4152	0.4212	0.4560	0.4848	0.5184	0.5196	0.5412
i	11	12	13	14	15	16	17	18	19	20
X_i	0.5556	0.6780	0.6780	0.6780	0.6864	0.6864	0.6888	0.8412	0.9312	0.9864
i	21	22	23	24	25					
X_i	1.0512	1.0584	1.2792	1.2804	1.7340					

We make the rule that whether number of units are 14 or experiment time is 1.7, the experiment is terminated. The observation of this endurance test started after 3 units are failed already (i.e., $r = 3$, $k = 14$, $T = 17$). From the Type I hybrid censoring scheme, we can obtain $\hat{\sigma}_1 = .5494$, $\hat{\sigma}_2 = .5201$, $\hat{\sigma}_3 = .5393$ and $\hat{\sigma}_4 = .5235$.

4.2. Simulation Results

To compare four different approximate MLEs of the scale parameter σ, we simulated the mean square error (MSE) and bias for four estimators by Monte Carlo simulations with 1,000 sets of data. Type I hybrid censored samples for sample size $n = 20$, 30 and 40 is used in this simulation.

k is the number of observed units set in advance, r is the number of units censored before the beginning of the observation and T is the length of time of experiment set in advance.

We compare several different conditions of k and T, and check the movement of MSEs and bias. k varies 50%, 60% and 70% of n, T varies 1.5, 1.7 and 1.9. We compare approximate MLEs in terms of MSE and bias. From the Table 4.1 as the number of full units are larger, the MSEs of approximate MLEs are getting smaller. As the number of observed units are smaller, the MSEs of approximate MLEs get larger. As the length of experiment time is longer, the MSEs of approximate MLEs get smaller.

From Table 4.1, the sizes of MSEs of four estimators are $\text{MSE}(\hat{\sigma_3}) < \text{MSE}(\hat{\sigma_1}) < \text{MSE}(\hat{\sigma_2}) < \text{MSE}(\hat{\sigma_4})$. Thus, σ_3 is the best approximate MLE for the Type I hybrid censoring scheme from the Rayleigh distribution.

Table 4.2 Simulation results of Type I hybrid censoring scheme

				MSE (bias)			
n	r	k	T	$\hat{\sigma_1}$	$\hat{\sigma_2}$	$\hat{\sigma_3}$	$\hat{\sigma_4}$
20	3	14	1.7	.0177 (.0118)	.0200 (.0078)	.0164 (-.084)	.0205 (.0120)
			1.9	.0161 (.0060)	.0178 (.0007)	.0153 (-.0127)	.0181 (.0045)
			2.1	.0143 (.0006)	.0157 (-.0058)	.0140 (-.0173)	.0158 (-.0023)
	4	12	1.7	.0182 (.0097)	.0215 (.0035)	.0168 (-.0188)	.0219 (.0074)
			1.9	.0164 (.0034)	.0189 (-.0049)	.0158 (-.0235)	.0191 (-.0014)
			2.1	.0148 (-.0006)	.0167 (-.0102)	.0147 (-.0269)	.0167 (-.0068)
	5	10	1.7	.0187 (.0055)	.0226 (-.0024)	.0177 (-.0309)	.0229 (.0005)
			1.9	.0168 (.0004)	.0196 (-.0098)	.0167 (-.0347)	.0197 (-.0071)
			2.1	.0156 (-.0020)	.0178 (-.0133)	.0159 (-.0367)	.0178 (-.0107)
30	4	22	1.7	.0112 (.0113)	.0125 (.0083)	.0106 (-.0042)	.0127 (.0117)
			1.9	.0103 (.0071)	.0114 (.0034)	.0099 (-.0072)	.0115 (.0064)
			2.1	.0095 (.0031)	.0103 (-.0012)	.0093 (-.0105)	.0103 (.0016)
	6	18	1.7	.0115 (.0100)	.0135 (.0062)	.0108 (-.0161)	.0137 (.0092)
			1.9	.0105 (.0051)	.0120 (-.0002)	.0102 (-.0195)	.0121 (.0025)
			2.1	.0098 (.0029)	.0110 (-.0031)	.0098 (-.0213)	.0110 (-.0005)
	7	16	1.7	.0119 (.0076)	.0139 (.0035)	.0113 (-.0239)	.0141 (.0060)
			1.9	.0107 (.0032)	.0121 (-.0028)	.0107 (-.0270)	.0122 (-.0005)
			2.1	.0103 (.0022)	.0115 (-.0042)	.0104 (-.0278)	.0116 (-.0019)
40	6	28	1.7	.0090 (.0059)	.0103 (.0046)	.0086 (-.0107)	.0104 (.0073)
			1.9	.0083 (.0025)	.0094 (.0003)	.0081 (-.0129)	.0094 (.0028)
			2.1	.0077 (-.0006)	.0086 (-.0035)	.0077 (-.0155)	.0086 (-.0012)
	8	24	1.7	.0092 (.0049)	.0108 (.0032)	.0088 (-.0197)	.0110 (.0056)
			1.9	.0084 (.0008)	.0097 (-.0024)	.0084 (-.0225)	.0097 (-.0002)
			2.1	.0080 (-.0009)	.0091 (-.0047)	.0082 (-.0239)	.0906 (-.0025)
	10	20	1.7	.0094 (.0029)	.0114 (.0000)	.0095 (-.0302)	.0115 (.0020)
			1.9	.0087 (-.0001)	.0101 (-.0045)	.0091 (-.0323)	.0102 (-.0026)
			2.1	.0085 (-.0007)	.0098 (-.0053)	.0090 (-.0328)	.0099 (-.0034)

References

Balakrishnan, N. (1989). Approximate MLE of the scale parameter of the Rayleigh distribution with censoring. *IEEE Transaction on Reliability*, **38**, 355-357.

Caroni, C. (2002). The correct ball bearingdata. *Lifetime Data Analysis*, **8**, 395-399.

Cho, Y. S., Lee, C. S. and Shin, H. J. (2013). Estimation for the generalized exponential distribution under progressive type I interval censoring. *Journal of the Korean Data & Information Science Society*, **24**, 1309-1317.

Dyer, D. D. and Whisenand C. W. (1973). Best linear unbiased estimator of the parameter of the Rayleigh distribution. *IEEE Transaction on Reliability*, **22**, 27-34.

Epstein, B. (1954). Truncated life tests in the exponential case. *The Annals of Mathematical Statistics*, **25**, 555-564.

Han, J. T. and Kang, S. B. (2006a). Estimation for the Rayleigh distribution with known parameter under multiply Type II censoring. *Journal of the Korean Data & Information Science Society*, **17**, 933-943.

Han, J. T. and Kang, S. B. (2006b). Estimation for two-parameter Rayleigh distribution based on multiply Type II censored sample. *Journal of the Korean Data & Information Science Society*, **17**, 1319-1328.

Kang, J. H. and Lee, C. S. (2013). Estimations of the skew parameter in a skewed double power function distribution. *Journal of the Korean Data & Information Science Society*, **24**, 901-909.

Kang, S. B. and Jung, W. T. (2009). Estimation for the double Rayleigh distribution based on progressive Type II censored samples. *Journal of the Korean Data & Information Science Society*, **20**, 1199-1206.

Kim, C. and Han, K. (2009). Estimation of the scale parameter of the Rayleigh distribution under general progressive censoring. *Journal of the Korean Statistical Society*, **38**, 239-246.

Lee, J. C and Lee, C. S. (2012). An approximate maximum likelihood estimator in a weighted exponential distribution. *Journal of the Korean Data & Information Science Society*, **23**, 219-225.

Lee W. C., Wu J. W., Hong M. L., Lin L. S. and Chan R. L. (2011). Assessing the lifetime performance index of Rayleigh products based on the Bayesian estimation under progressive type II right censored samples. *Journal of Computational and Applied Mathematics*, **235**, 1676-1688.

Polovko A. M. (1968). *Fundamentals of reliability theory*, Academic Press, New York.

Raqab, M. Z. and Wilson, W. M. (2002). Bayesian prediction of the total time on test using doubly censored Rayleigh data. *Journal of Statistical Computation and Simulation*, **72**, 781-789.

Rayleigh, J. (1980). On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *Philosophical Magazine*, **10**, 73-78.

Siddiqui M. M. (1962). Some problems connected with Rayleigh distributions. *Journal of Research of the National Bureau of Standards*, **60D**, 167-174.