A Note on Bayesian Prediction Analysis for the Rayleigh Model in the presence of Outliers

Jeong Hwan Ko, Yeung-Hoon Kim[1)]

Abstract

This paper deals with the problem of predicting order statistics in samples from a Rayleigh population when an outlier is present. Bayesian predictive distribution and prediction bounds of the $p-th$ order statistics is obtained where an outlier of type $\theta\delta$ is present. In this connection, some identies are derived.

1. Introduction

The Rayleigh distribution with probability density function(pdf), denoted by $R(\sigma^2)$, is given by

$$f(x\ ;\ \sigma) = \frac{x}{\sigma^2}\exp(-\frac{x^2}{2\sigma^2}),\quad 0<x<\infty \tag{1.1}$$

and distribution function is

$$F(x:\sigma) = 1-\exp(-\frac{x^2}{2\sigma^2}),\quad \sigma>0,\ 0<x<\infty \tag{1.2}$$

The properties and application of Rayleigh distribution were discussed by Siddiqui(1962), Dyer and Whisenand (1973a,b), Sinha and Howlader(1983) obtained the Bayes estimator and credible intervals for the reliability function.

In this paper, few concepts like order statistics, prediction and outliers are dealt with. Kitagawa uses Bayes approach to analyse when outliers are present. Barnett and Lewis is a text devoted entirely to outliers. Regarding Bayesian prediction, Chhikara and Guttman(1982), Nigam and Hamdy(1987), Sinha(1989), Upadhyay and Pandey(1989) and Nigm and AL-Wahab(1996) suggested the Bayesian inference about prediction for inverse Gaussian, Pareto, lognormal, exponential and Burr distributions, respectively.

In this paper, a different situation is taken up. That is, in the samples now, outliers

1) 안동대학교 정보통계학과

are present. Predictive distribution of the $p-th$ order statistics a sample from a Rayleigh population is obtained when an outlier of type $\theta\delta$ is present. In this connection, some identies are derived.

2. Prediction of the Future Observations in the presence of Outlier for type $\theta\delta$

Suppose that $X_1, X_2, \cdots, X_m$ is a random sample of size m drawn from a population whose pdf is Rayleigh distribution and that $Y_1, Y_2, \cdots, Y_n$ is a second independent random sample of future observations from the same distribution.

The distribution of the $p-th$ order statistics in the a sample of size n when an outlier is present is given by $(y = y_{(p)})$

$$f(y) = a\,[(p-1)\{F(y)\}^{p-2}\{1-F(y)\}^{n-p}G(y)f(y) + \{F(y)\}^{p-1}\{1-F(y)\}^{n-p}g(y) + (n-p)\{F(y)\}^{p-1}\{1-F(y)\}^{n-p-1}\{1-G(y)\}f(y)] \tag{2.1}$$

where $a=\binom{n-1}{r-1}$ and $f(x)=f$ and $F(x)=F$ are the density and distribution function of all those x's which are not outliers while $g(x)$ and $G(x)$ are those of an outlier. If the outliers is of type $\theta\delta$, then from (1.1) and (1.2), we have

$$g(x:\sigma) = \frac{x}{(\sigma\delta)^2}\, e^{-\frac{x^2}{2(\sigma\delta)^2}} \tag{2.2}$$

and

$$G(x:\sigma) = 1 - e^{-\frac{x^2}{2(\sigma\delta)^2}} \tag{2.3}$$

and if the $p-th$ order statistics is $y = y_{(p)}$, then we have (2.1) as

$$f(y|\sigma) = a\Big[\frac{y}{\sigma^2}(p-1)\sum_{j=0}^{p-2}\binom{p-2}{j}(-1)^j\Big\{e^{-\frac{(n-p+j+1)y^2}{2\sigma^2}} - e^{-\frac{\{(n-p+j+1)\delta^2+1\}y^2}{2(\sigma\delta)^2}}\Big\} + \frac{y}{(\sigma\delta)^2}\sum_{j=0}^{p-1}\binom{p-1}{j}(-1)^j e^{\frac{-\{(n-p+j)\delta^2+1\}y^2}{2(\sigma\delta)^2}} + \frac{y}{\sigma^2}(n-p)\sum_{j=0}^{p-1}\binom{p-1}{j}(-1)^j e^{\frac{-\{(n-p+j-1)\delta^2+1\}y^2}{2(\sigma\delta)^2}}\Big] \tag{2.4}$$

For the first $r-th$ ordered failure times from the first sample of size m obtained from the Rayleigh population, the likelihood function is given, using (1.1) and (1.2), by

$$L(\sigma \mid x) \propto \frac{1}{\sigma^{2r}} \exp\left(-\frac{\sum_{i=1}^{r} x_i^2 + (m-r)x_{(r)}^2}{2\sigma^2}\right), \quad \sigma>0,\ 0<x_i<\infty,\ i=1,2,\cdots,m. \tag{2.5}$$

To obtain the posterior density of σ, we use a prior density which is given by

$$\Pi(\sigma) \propto \frac{1}{\sigma} \tag{2.6}$$

It follows, from (2.5) and (2.6), that the posterior density is given by

$$\Pi(\sigma \mid x) = \frac{z^{2r} e^{-\frac{z^2}{2\sigma^2}}}{\Gamma(r)\, 2^{r-1} \sigma^{2r+1}} \tag{2.7}$$

where $z^2 = \sum_1^r x_i^2 + (m-r)x_{(r)}^2$

Following Aitcheson and Dunsmore (1975), the Bayesian predictive density function of y given x, denoted by $f(y|x)$, is defined by

$$f(y \mid x) = \int_0^\infty f(y \mid \sigma)\, \Pi(\sigma \mid x)\, d\sigma$$

$$\begin{aligned}
&= 2az^{2r} ry \Big[(p-1) \sum_{j=0}^{p-2} \binom{p-2}{j} (-1)^j [\{(n-p+j+1)y^2 + z^2\}^{-(r+1)}] \\
&\quad - \delta^{2(r+1)} [\{(n-p+j+1)\delta^2 + 1\} y^2 + \delta^2 z^2]^{-(r+1)} \Big] \\
&\quad + \delta^{2r} \sum_{j=0}^{p-1} \binom{p-1}{j} (-1)^j [\{(n-p+j)\delta^2 + 1\} y^2 + \delta^2 z^2]^{-(r+1)} \\
&\quad + (n-p)\, \delta^{2(r+1)} \sum_{j=0}^{p-1} \binom{p-1}{j} (-1)^j [\{(n-p+j-1)\delta^2 + 1\} y^2 + \delta^2 z^2]^{-(r+1)}]
\end{aligned} \tag{2.8}$$

The prediction bounds for y are obtained by evaluating $P(Y \geq t \mid x)$ for some positive t. It follows from (2.8) that

$$\begin{aligned}
P(Y \geq t \mid x) &= az^{2r}(p-1) \sum_{j=0}^{p-2} \binom{p-2}{j} (-1)^j (n-p+j+1)^{-1} \{(n-p+j+1)t^2 + z^2\}^{-r} \\
&\quad - az^{2r}(p-1)\delta^{2(r+1)} \sum_{j=0}^{p-2} \binom{p-2}{j} (-1)^j \{(n-p+j+1)\delta^2 + 1\}^{-1} \\
&\qquad \times [\{(n-p+j+1)\delta^2 + 1\} t^2 + \delta^2 z^2]^{-r} \\
&\quad + az^{2r}\, r\, \delta^{2r} \sum_{j=0}^{p-1} \binom{p-1}{j} (-1)^j \{(n-p+j)\delta^2 + 1\}^{-1}
\end{aligned}$$

$$\times[\{(n-p+j)\delta^2+1\}t^2+\delta^2 z^2]^{-r}$$

$$+az^r(n-p)\delta^{2(r+1)}\sum_{j=0}^{p-1}\binom{p-1}{j}(-1)^j\{(n-p+j-1)\delta^2+1\}^{-1}$$

$$\times[\{(n-p+j-1)\delta^2+1\}t^2+\delta^2 z^2]^{-r} \quad (2.9)$$

A $100\tau\%$ Bayesian prediction interval for y is such that

$$P[L(x)<Y<U(x)] = \tau, \quad (2.10)$$

where $L(x)$ and $U(x)$ are the lower and upper limits satisfying

$$P[Y>L(x)|x] = (1+\tau)/2$$

and

$$P[Y>U(x)|x] = (1+\tau)/2.$$

Also from (2.8), one has

$$E(Y^s)=\frac{a\Gamma(r-\frac{s}{2})\Gamma(\frac{s}{2}+1)z^s}{\Gamma(r)}\Big[(p-1)\sum_{j=0}^{p-2}\binom{p-2}{j}(-1)^j$$

$$\times\Big\{(n-p+j+1)^{-\frac{s}{2}-1}-\{(n-p+j+1)\delta^2+1\}^{-\frac{s}{2}-1}\delta^{s+2}\Big\}$$

$$+\sum_{j=0}^{p-1}\binom{p-1}{j}(-1)^j\delta^s\ \{\{(n-p+j)\delta^2+1\}^{-\frac{s}{2}-1}$$

$$-(n-p)\{(n-p+j-1)\delta^2+1\}^{-\frac{s}{2}-1}\delta^2\}\Big] \quad (2.11)$$

As another prior distribution for σ, we consider an inverse gamma prior distribution with the probability density function

$$\Pi(\sigma)=\frac{e^{-\frac{1}{2\beta\sigma^2}}}{\Gamma(a)\,\beta^a\,\sigma^{2(a+1)}},\qquad a,\ \beta>0,\ \sigma>0. \quad (2.12)$$

Then one can obtain easily the posterior density of σ given $X=x$, which is given by

$$\Pi(\sigma|x)=\frac{(\frac{\beta z^2+2}{\beta})^{r+a+\frac{1}{2}}}{\Gamma(r+a+\frac{1}{2})}\,\frac{e^{-\frac{\beta z^2+2}{2\beta\sigma^2}}}{2^{r+a-\frac{1}{2}}\,\sigma^{2(r+a+1)}} \quad (2.13)$$

Under the conditional predictive density and inverse gamma prior, the Bayesian prediction density function of y given x is

$$\Pi(y\,|\,x) = 2a(r+a+\frac{1}{2})\beta\,(\beta z^2+2)^{r+a+\frac{1}{2}}\Big[y(p-1)\sum_{j=0}^{p-2}\binom{p-2}{j}(-1)^j$$

$$\times\{(n-p+j+1)\beta y^2+(\beta z^2+2)\}^{-(r+a+\frac{3}{2})}-\delta^{2(r+a+\frac{3}{2})}[\{(n-p+j+1)\delta^2$$

$$+1\}\beta y^2+\delta^2(\beta z^2+2)]^{-(r+a+\frac{3}{2})}\Big\}+y\sum_{j=0}^{p-1}\binom{p-1}{j}(-1)^j\,\delta^{2(r+a+\frac{1}{2})}$$

$$\times[\{(n-p+j)\delta^2+1\}\beta y^2+\delta^2(\beta z^2+2)]^{-(r+a+\frac{3}{2})}$$

$$+(n-p)y\sum_{j=0}^{p-1}\binom{p-1}{j}(-1)^j\delta^{2(r+a+\frac{3}{2})}[\{(n-p+j-1)\delta^2+1\}\beta y^2$$

$$+\delta^2(\beta z^2+2)]^{-(r+a+\frac{3}{2})}\Big] \qquad (2.14)$$

Using (2.14), one can obtain the probability

$$P(Y\geq t\,|\,x) = a\,(\beta z^2+2)^{r+a+\frac{1}{2}}\Big[(p-1)\sum_{j=0}^{p-2}\binom{p-2}{j}(-1)^j$$

$$\times\{(n-p+j+1)^{-1}\{(n-p+j+1)\beta t^2+(\beta z^2+2)\}^{-(r+a+\frac{1}{2})}$$

$$-\{(n-p+j+1)\delta^2+1\}^{-1}\delta^{2(r+a+\frac{3}{2})}[\{(n-p+j+1)\delta^2+1\}\beta t^2$$

$$+\delta^2(\beta z^2+2)]^{-(r+a+\frac{1}{2})}\Big\}+\delta^{2(r+a+\frac{1}{2})}\sum_{j=0}^{p-1}\binom{p-1}{j}(-1)^j$$

$$\times\{(n-p+j)\delta^2+1\}^{-1}[\{(n-p+j)\delta^2+1\}\beta t^2+\delta^2(\beta z^2+2)]^{-(r+a+\frac{1}{2})}$$

$$+\delta^{2(r+a+\frac{3}{2})}(n-p)\sum_{j=0}^{p-1}\binom{p-1}{j}(-1)^j\{(n-p+j-1)\delta^2+1\}^{-1}$$

$$\times[\{(n-p+j-1)\delta^2+1\}\beta t^2+\delta^2(\beta z^2+2)]^{-(r+a+\frac{1}{2})} \qquad (2.15)$$

Also from (2.14), one can

$$E(Y^2)=\frac{a\Gamma(r+a-\frac{s}{2}+\frac{1}{2})\Gamma(\frac{s}{2}+1)}{\Gamma(r+a+\frac{1}{2})\beta^{\frac{s}{2}}}(\beta z^2+2)^{\frac{s}{2}}\Big[(p-1)\sum_{j=0}^{p-2}\binom{p-2}{j}(-1)^j$$

$$\times\Big\{(n-p+j+1)^{-\frac{s}{2}-1}-\delta^{s+2}\{(n-p+j+1)\delta^2+1\}^{-\frac{s}{2}-1}\Big\}$$

$$+\sum_{j=0}^{p-1}\binom{p-1}{j}(-1)^j\delta^s\{\{(n-p+j)\delta^2+1\}^{-\frac{s}{2}-1}+(n-p)\delta^s$$

$$\times\{(n-p+j-1)\delta^2+1\}^{-\frac{s}{2}-1}\}\Big] \qquad (2.16)$$

references

Aitchison, J. and Dunsmore, I. R. (1975), *Statistical Prediction Analysis*, Cambridge University Press, London.

Barnett V. and Lewis T. (1984), *Outliers in Statistical Data*, Edition II, John Wiley & Sons, New York.

Chhikara, R. S. and Guttman, I. (1982), Prediction Limits for the Inverse Gaussian distribution, *Technometrics*, 24, 319-324.

Dyer, D. D. and Whisenand, C. W. (1973a), Best Linear Unbiased Estimator of the Parameter of the Rayleigh Distribution - Part I : Snall Sample Theory for Censored Order Statistics, *IEEE Transactions on Reliability*, R-22, 27-34.

Dyer, D. D. and Whisenand, C. W. (1973b), Best Linear Unbiased Estimator of the Parameter of the Rayleigh Distribution - Part II : Optimum Theory for Selected Oder Statistics, *IEEE Transactions on Reliability*, R-22, 229-231.

Kitagawa, G. (1984), *Communications in Statistics Part B, 13, 107-126.*

Nigm, A. M. and Hamdy, H. I. (1987), Bayesian Prediction Bound the Pareto LIfetime Model, *Communications in Statistics Part A - Theory and Methods, 16, 1761-1772.*

Nigm, A. M. and AL-Wahab, N. Y. ABD (1996), Bayesian Prediction With Sample Size for the Burr Lifetime Distribution. *Communications in Statistics Part A - Theory and Methods, 25, 1289-1303.*

Sinha, S. K. (1989), Bayesian Inference about the Prediction Credible Intervals and Reliability Function for Lognormal Distribution, *Journal of the Indian Statistical Association*, 27, 73-78.

Upadhyay, S. K. and Pandey, M. (1989), Prediction Limits for an Exponential Distribution: A Bayes Predictive Distribution Approach, *IEEE Transactions on Reliability,* R-38, 599-602.