Bayes prediction based on right censored data

Peer-reviewed author version

WANG, Lichun & VERAVERBEKE, Noel (2009) Bayes prediction based on right censored data. In: STATISTICAL PAPERS, 50(1). p. 137-149.

DOI: 10.1007/s00362-006-0044-1
Handle: http://hdl.handle.net/1942/9260

Bayes Prediction Under Random Censorship *

Lichun Wang † **Noel Veraverbeke**

Limburgs Universitair Centrum, Center for Statistics, B-3590 Diepenbeek, Belgium

Abstract: This paper is concerned with a Bayes prediction problem in the exponential distribution under random censorship. Using censored samples, we work out a prediction interval for a sum of interest which consists of some future samples. Differing from the general Bayes approach, we assume that the prior distribution of the parameter of the exponential distribution is of an unknown form, and only one moment condition of the prior distribution is given. Simulation studies are conducted to exhibit the coverage probabilities of the prediction interval.

Keywords: Bayes Prediction; Random Censorship; Prediction Interval; Exponential Distribution.

2000 Mathematics Subject Classification: 62C10, 62N01

1. Introduction

Prediction problems arise naturally in many important areas of statistical research and have been very useful in many fields of application such as quality control, life experiments and so on. The general prediction problem can be regarded as that of inferring the value of an unknown observable that belongs to a future sample from present available information. Enlightened by Robbins (1982, 1983), we assume that there are n workshops in a factory and that they manufacture the same type of electronic product. We now select one unit from each workshop and put them to use. After each of them is ineffective, we can obtain n survival data $X_1, X_2, \cdots, X_n$. If $X_i \leq a$, where a is a positive constant, then we again select one unit from the i-th workshop and denote

*Financial support from the IAP research network (#P5/24) of the Belgian Government (Belgian Science Policy) is gratefully acknowledged.

†Corresponding author: lichun.wang@luc.ac.be

its unknown life by Y_i. In practice, the electronic product life of the second round is an important and interesting problem. It can be used to evaluate the mechanical stability of a product, and it often appears in system failure analysis, product reliability testing and design. Our aim is to predict the sum of Y_i, i.e., $\sum_{i=1}^n I(X_i \leq a)Y_i$, where $I(A)$ is the indicator of the set A.

However, many statistical experiments result in incomplete samples, even under some well controlled conditions. In fact, right censored data often arise in the study of survival analysis, medical follow-up and reliability. In past decades, statistical inference with censorship attracted considerable attention and has been studied extensively. Let $V_1, V_2, \cdots, V_n$ be nonnegative independent and identically distributed (i.i.d.) random variables, which denote censoring times with a distribution function W. It is assumed that X_i's and V_i's are independent. In the random censorship model, the true survival times $X_1, X_2, \cdots, X_n$ are not always observable. Instead, we observe only $Z_i = \min\{X_i, V_i\}$ and $\delta_i = I(X_i \leq V_i)$.

Let us assume that X_i, given λ_i, has the following probability density function (pdf)

$$f(x_i|\lambda_i) = \frac{1}{\lambda_i}\exp(-\frac{x_i}{\lambda_i}), \quad x_i > 0, \tag{1}$$

where $1 \leq i \leq n$, and it is assumed that $\lambda_1, \lambda_2, \cdots, \lambda_n$ are i.i.d. and have a prior distribution function $G(\lambda)$ with support on $\Lambda = (0, \infty)$.

Define

$$S = \sum_{i=1}^{n} I(X_i \leq a)Y_i. \tag{2}$$

In the following, based on the censored observations (Z_i, δ_i) $(1 \leq i \leq n)$, we construct a prediction interval for S, firstly, under the condition that the censoring distribution W is known and secondly, that it is unknown.

As we know, Bayesian analysis is an important method of modern statistics(see Berger (1985) for more details). A main difference point with some classical statistical methods is that in Bayesian analysis, we use not only the sample information but also some information about the parameter. Usually, in the Bayesian framwork, given the states of a random variable, a conditional probability is attached to this variable and a prior density of the parameter is specified based on previous knowledge. But, in the present paper, it is unnecessary for us to specify the prior distribution of the parameter. We only make some moment assumption on the prior distribution $G(\lambda)$.

2. Prediction interval for S

Since X_i and Y_i (if necessary) come from the same workshop, (X_i, Y_i) $(1 \le i \le n)$ are i.i.d. with common marginal pdf

$$f(x, y) = \int_{\Lambda} f(x|\lambda) f(y|\lambda) dG(\lambda). \tag{3}$$

By Fubini's theorem, we have

$$\begin{aligned} E\left[I(X_i \le a)Y_i\right] &= \int_0^\infty \int_0^\infty I(x \le a) y f(x, y) dx dy \\ &= \int_{\Lambda} \left[\int_0^\infty I(x \le a) f(x|\lambda) dx \int_0^\infty y f(y|\lambda) dy\right] dG(\lambda) \\ &= \int_{\Lambda} \lambda[1 - \exp(-a/\lambda)] dG(\lambda). \end{aligned} \tag{4}$$

Note that

$$\begin{aligned} E[(X - a)I(X \le a)] &= \int_0^\infty \int_0^\infty (x - a) I(x \le a) f(x, y) dx dy \\ &= \int_{\Lambda} \lambda[1 - \exp(-a/\lambda)] dG(\lambda) - a, \end{aligned} \tag{5}$$

hence, we get

$$ES = E\left\{\sum_{i=1}^{n} [(X_i - a)I(X_i \le a) + a]\right\}. \tag{6}$$

Furthermore, we have

$$E[(X_i - a)I(X_i \le a)] = E\{E[(X_i - a)I(X_i \le a)|\lambda]\}, \tag{7}$$

where

$$\begin{aligned}
&E[(X_i - a)I(X_i \le a)|\lambda] \\
= &\int_0^\infty \frac{I(x \le a)(x-a)}{1-W(x)}[1-W(x)]f(x|\lambda)dx \\
= &\int_0^\infty \frac{I(x \le a)(x-a)}{1-W(x)} \int_x^\infty dW(v)f(x|\lambda)dx \\
= &\int\int_{x \le v} \frac{I(x \le a)(x-a)}{1-W(x)} dW(v)f(x|\lambda)dx \\
= &E\left[\frac{I(\min\{X_i, V_i\} \le a)(\min\{X_i, V_i\} - a)}{1-W(\min\{X_i, V_i\})} I(X_i \le V_i)|\lambda\right] \\
= &E\left[\frac{I(Z_i \le a)(Z_i - a)\delta_i}{1-W(Z_i)}|\lambda\right]. \qquad (8)
\end{aligned}$$

Put

$$T = \sum_{i=1}^n \left[\frac{I(Z_i \le a)(Z_i - a)\delta_i}{1-W(Z_i)} + a\right], \qquad (9)$$

then, by (6)-(9), it is easy to see that S has the same expectation as T. Thus,

$$S - T = \sum_{i=1}^n \left[I(X_i \le a)Y_i - \frac{I(Z_i \le a)(Z_i - a)\delta_i}{1-W(Z_i)} - a\right]$$

is the sum of n i.i.d. random variables with mean zero, and

$$\frac{S-T}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} N(0,1), \quad \text{as } n \to \infty, \qquad (10)$$

where $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution, and

$$\begin{aligned}
\sigma^2 &= E\left[I(X \le a)Y - \frac{I(Z \le a)(Z-a)\delta}{1-W(Z)}\right]^2 - a^2 \\
&= E[Y^2 I(X \le a)] + E\left[\frac{I(Z \le a)(Z-a)^2\delta}{(1-W(Z))^2}\right] \\
&\quad -2E\left[I(X \le a)Y\frac{I(Z \le a)(Z-a)\delta}{1-W(Z)}\right] - a^2 \\
&\hat{=} \sum_{i=1}^3 I_i - a^2. \qquad (11)
\end{aligned}$$

First, we easily get

$$I_1 = E[Y^2 I(X \le a)] = 2\int_\Lambda \lambda^2[1-\exp(-a/\lambda)]dG(\lambda). \tag{12}$$

Second, since

$$I_3 = -2E\left\{\left[I(X \le a)Y\frac{I(Z \le a)(Z-a)\delta}{1-W(Z)}|\lambda\right]\right\}, \tag{13}$$

where, using the independence of Y and X and V, and a similar discussion as in (8),

$$\begin{aligned}
& E\left[I(X \le a)Y\frac{I(Z \le a)(Z-a)\delta}{1-W(Z)}|\lambda\right] \\
= \;& \lambda\int\int_{x\le v}\frac{I(x\le a)(x-a)}{1-W(x)}dW(v)f(x|\lambda)dx \\
= \;& \lambda\int_0^\infty I(x\le a)(x-a)f(x|\lambda)dx \\
= \;& \lambda[\lambda-\lambda\exp(-a/\lambda)-a], \qquad (14)
\end{aligned}$$

we obtain that

$$I_3 = -2\int_\Lambda \lambda^2[1-\exp(-a/\lambda)]dG(\lambda) + 2a\int_\Lambda \lambda dG(\lambda). \tag{15}$$

Also, note that

$$E\left[\frac{Z\delta}{1-W(Z)}\right] = \int_\Lambda \lambda dG(\lambda). \tag{16}$$

Therefore, combining (11)-(12) with (15)-(16) yields

$$\sigma^2 = E\left[\frac{I(Z\le a)(Z-a)^2\delta}{(1-W(Z))^2}\right] + 2aE\left[\frac{Z\delta}{1-W(Z)}\right] - a^2. \tag{17}$$

2.1. The distribution W is known

Define

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \frac{I(Z_i\le a)(Z_i-a)^2\delta_i}{[1-W(Z_i)]^2} + 2a\frac{1}{n}\sum_{i=1}^n\frac{Z_i\delta_i}{1-W(Z_i)} - a^2. \tag{18}$$

Then we have proved the following theorem.

Theorem 1. Let S and T be defined by (2) and (9), respectively, and $\hat{\sigma}^2$ given by (18). If $W(a) < 1$ and $E\lambda < \infty$, then for $n \to \infty$, we have

$$\frac{S-T}{\hat{\sigma}\sqrt{n}} \xrightarrow{\mathcal{L}} N(0,1).$$

Proof. Define

$$\phi(\lambda) = \int_0^\infty \frac{I(x \le a)(x-a)^2}{1-W(x)} f(x|\lambda)dx.$$

Note that

$$\begin{aligned} \phi(\lambda) &\le \frac{1}{\lambda[1-W(a)]}\int_0^a (x-a)^2 \exp(-x/\lambda)dx \\ &= \frac{a^2 + 2\lambda^2[1-\exp(-a/\lambda)] - 2a\lambda}{1-W(a)} \\ &\le \frac{a^2}{1-W(a)}, \end{aligned}$$

then, we have

$$E\left[\frac{I(Z \le a)(Z-a)^2\delta}{(1-W(Z))^2}\right] = E[\phi(\lambda)] \le \frac{a^2}{1-W(a)}.$$

Hence, under the conditions of Theorem 1, the conclusion is obvious.

Clearly, Theorem 1 can be used to construct prediction intervals for S. Denote

$$I_{n,\alpha} = \left\{S : |(S-T)/(\hat{\sigma}\sqrt{n})| \le N_\alpha\right\}, \tag{19}$$

where $P(|N(0,1)| \le N_\alpha) = 1-\alpha$. Then, $I_{n,\alpha}$ gives an approximate prediction interval for S with asymptotically correct coverage probability $1-\alpha$; that is

$$P(S \in I_{n,\alpha}) = 1-\alpha + o(1),$$

where $o(1)$ denotes terms converging to zero as $n \to \infty$.

Especially, when there is no censorship ($Z_i = X_i, \delta_i = 1$), (9) becomes

$$T_0 = \sum_{i=1}^n [(X_i - a)I(X_i \le a) + a], \tag{20}$$

and

$$S - T_0 = \sum_{i=1}^{n} [I(X_i \leq a)(Y_i - X_i + a) - a]$$

is the sum of n i.i.d. random variables with mean zero, and

$$\frac{S - T_0}{\sigma_0 \sqrt{n}} \xrightarrow{\mathcal{L}} N(0, 1), \quad \text{as} \quad n \to \infty, \tag{21}$$

where

$$\sigma_0^2 = E[I(X \leq a)(Y - X + a)]^2 - a^2. \tag{22}$$

we can define an estimator in (18) for σ_0^2 becomes

$$\hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \frac{1}{n} \sum_{i=1}^{n} I(X_i > a)(X_i - a)^2. \tag{23}$$

Hence, we have the following corollary.

Corollary 1. Let S and T_0 be defined by (2) and (20), respectively. If $E\lambda < \infty$, then, for $n \to \infty$, we have

$$\frac{S - T_0}{\hat{\sigma}_0 \sqrt{n}} \xrightarrow{\mathcal{L}} N(0, 1).$$

2.2. The distribution W is unknown

Since $W(.)$ in the expression (17) is unknown, we can replace it by its Kaplan-Meier (1958) product limit estimator defined by

$$1 - \hat{W}_n(t) = \prod_{i=1}^{n} \left[\frac{n-i}{n-i+1} \right]^{I(Z_{(i)} \leq t, \delta_{(i)} = 0)}, \quad t < Z_{(n)}, \tag{24}$$

where $Z_{(1)} \leq Z_{(2)} \leq \cdots \leq Z_{(n)}$ are the order statistics of $(Z_1, Z_2, \cdots, Z_n)$ and $\delta_{(i)}$ is the concomitant of $Z_{(i)}$. Hence, we can present an estimator for σ^2 in this case as follows

$$\widetilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{I(Z_i \leq a)(Z_i - a)^2 \delta_i}{[1 - \hat{W}_n(Z_i)]^2} + 2a \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i \delta_i}{1 - \hat{W}_n(Z_i)} - a^2. \tag{25}$$

We now prove the following result.

Theorem 2. Let S and T be defined by (2) and (9), respectively, and $\widetilde{\sigma}^2$ defined in (25). If $W(a) < 1$ and $E\lambda < \infty$, then for $n \to \infty$, we have

$$\frac{S - T}{\widetilde{\sigma}\sqrt{n}} \xrightarrow{\mathcal{L}} N(0,1).$$

Proof. Under the conditions of Theorem 2, we know

$$\frac{1}{n}\sum_{i=1}^{n} \frac{I(Z_i \le a)(Z_i - a)^2 \delta_i}{[1 - W(Z_i)]^2} + 2a\frac{1}{n}\sum_{i=1}^{n} \frac{Z_i \delta_i}{1 - W(Z_i)} - a^2 \xrightarrow{p} \sigma^2, \tag{26}$$

where $\xrightarrow{p}$ denotes convergence in probability.

On the other hand, we have

$$\begin{aligned}
&\left| \frac{1}{n}\sum_{i=1}^{n} \frac{I(Z_i \le a)(Z_i - a)^2 \delta_i}{[1 - \hat{W}_n(Z_i)]^2} - \frac{1}{n}\sum_{i=1}^{n} \frac{I(Z_i \le a)(Z_i - a)^2 \delta_i}{[1 - W(Z_i)]^2} \right| \\
\le\ & \sup_{Z_i \le Z_{(n)}} \left| \frac{1}{[1 - \hat{W}_n(Z_i)]^2} - \frac{1}{[1 - W(Z_i)]^2} \right| \times \frac{1}{n}\sum_{i=1}^{n} I(Z_i \le a)(Z_i - a)^2 \delta_i \\
\le\ & C \sup_{Z_i \le Z_{(n)}} \left(\frac{|W(Z_i) - \hat{W}_n(Z_i)|}{[1 - \hat{W}_n(Z_i)]^2 [1 - W(Z_i)]^2} \right), \quad \text{in probability},
\end{aligned} \tag{27}$$

where C is a finite constant, not depending on n.

From (27) and the following result due to Zhou (1992)

$$\sup_{Z_i \le Z_{(n)}} |W(Z_i) - \hat{W}_n(Z_i)| \xrightarrow{p} 0, \tag{28}$$

we have

$$\frac{1}{n}\sum_{i=1}^{n} \frac{I(Z_i \le a)(Z_i - a)^2 \delta_i}{[1 - \hat{W}_n(Z_i)]^2} \xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n} \frac{I(Z_i \le a)(Z_i - a)^2 \delta_i}{[1 - W(Z_i)]^2}. \tag{29}$$

Similarly, we get

$$\frac{1}{n}\sum_{i=1}^{n} \frac{Z_i \delta_i}{1 - \hat{W}_n(Z_i)} \xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n} \frac{Z_i \delta_i}{1 - W(Z_i)}. \tag{30}$$

Hence, Theorem 2 is proved.

Theorem 2 can be used to present an approximate $1 - \alpha$ prediction interval for S as follows:

$$\widetilde{I}_{n,\alpha} = \left\{ S : |(S - \widetilde{T})/(\widetilde{\sigma}\sqrt{n})| \le N_\alpha \right\}, \tag{31}$$

where $P(|N(0,1)| \leq N_\alpha) = 1 - \alpha$ and

$$\widetilde{T} = \sum_{i=1}^{n} \left[\frac{I(Z_i \leq a)(Z_i - a)\delta_i}{1 - \hat{W}_n(Z_i)} + a \right].$$

3. Simulation

Let the prior density

$$g(\lambda) = \frac{dG(\lambda)}{d\lambda} = \frac{1}{\Gamma(3)}(\frac{1}{\lambda})^4 \exp(-\frac{1}{\lambda}), \tag{32}$$

and the censoring distribution be

$$W(v) = 1 - \exp(-cv), \quad c > 0, \quad v > 0, \tag{33}$$

where c can be used to describe the censoring proportion. Note that $EX = E\lambda = 1/2$ and take $a = 1/2$. Simulation studies are conducted to examine the coverage accuracies of the proposed prediction intervals $I_{n,\alpha}$ and $\widetilde{I}_{n,\alpha}$, respectively, for different sample sizes and the censoring distributions.

First, we generate n random values from the distribution (32) and denote them by $\lambda_1, \lambda_2, \cdots, \lambda_n$. Second, by (1) and (33), we can obtain $X_1, X_2, \cdots, X_n$ and $V_1, V_2, \cdots, V_n$, consequently, we get $Z_1, Z_2, \cdots, Z_n$. Third, we compute the prediction interval $I_{n,\alpha}$ and $\widetilde{I}_{n,\alpha}$, respectively, for c=2, c=1 and c=1/2. Repeating this process for 500 times, the results are presented in Table 1 and Table 2.

Table 1——coverage probability of $I_{n,\alpha}$

		nominal level is 0.90	nominal level is 0.95
c (censoring proportion)	n	coverage probability	coverage probability
2 (0.4453)	20	0.5480	0.6000
	50	0.6280	0.6900
	100	0.7400	0.7940
1 (0.2983)	20	0.7040	0.7560
	50	0.7900	0.8360
	100	0.8380	0.9040
1/2 (0.1828)	20	0.8240	0.8620
	50	0.8520	0.9060
	100	0.8800	0.9300

Table 2——coverage probability of $\widetilde{I}_{n,\alpha}$

		nominal level is 0.90	nominal level is 0.95
c (censoring proportion)	n	coverage probability	coverage probability
2 (0.4453)	20	0.4280	0.4840
	50	0.5980	0.6660
	100	0.7040	0.7620
1 (0.2983)	20	0.6700	0.7260
	50	0.7380	0.8140
	100	0.7900	0.8440
1/2 (0.1828)	20	0.8100	0.8720
	50	0.8160	0.8840
	100	0.8400	0.9020

First, from Table 1 and Table 2, for certain c (censoring proportion), we see the coverage accuracies under two nominal levels generally tend to increase as the sample size n gets larger. However, it is very difficult for both coverage probabilities to attain or exceed their respective nominal level even though the sample size is rather large. In fact, because we only select one sample from each workshop at the first round, it is not difficult to understand this result. Second, the performance of the coverage probability depends on c (censoring proportion), it generally decreases as c (censoring proportion) increases. Finally, we find the coverage probabilities of $I_{n,\alpha}$ are always better than those of $\widetilde{I}_{n,\alpha}$, as can be expected when the censoring distribution W is unknown. But, anyway, it gives us something like a lamppost before making a decision.

As a contrast, when there is no censorship , we report the prediction results for different sample sizes in Table 3.

Table 3

	nominal level is 0.90	nominal level is 0.95
n	coverage probability	coverage probability
20	0.9060	0.9320
50	0.9180	0.9460
100	0.8960	0.9480

Obviously, although it is still difficult to attain the nominal level, the coverage probabilities in Table 3 are better than those in Table 1, and it gets better as the sample size n becomes larger. Moreover, note that Theorem 1, Corollary 1 and Theorem 2 do not depend on any particular prior distribution, actually, we only make the assumption of a finite first moment. Hence, it is unnecessary to provide more simulation results for different prior distributions.

References

[1] Berger, J. O., 1985. Statistical Decision Theory and Bayesian Analysis, Springer, New York.

[2] Kaplan, E. T., Meier, P., 1958. Nonparametric estimation from incomplete observations, J. Amer. Statist. Assoc. 53, 457-481.

[3] Robbins, H., 1982. Estimating many variance, Statistical Decision Theory and Related Topics III, 2, 218-226. Academic Press, New York.

[4] Robbins, H., 1983. Some thoughts on empirical Bayes estimation, Ann. Statist. 11, 713-723.

[5] Zhou, M., 1992. Asymptotic normality of the synthetic data regression estimation for censored survival data, Ann. Statist. 20, 1002-1021.