第31卷 第1期 电子科技大学学报 Vol.31 No.1
2002年2月 Journal of UEST of China Feb.2002

Confidence Interval for Parameter of Exponential Distribution from Incomplete Observational Data

Zhu Hong Xiao Bailong
(Institute of Mathematics, Sichuan University Chengdu 610064) (Hunan University Changsha 410082)
Xu Daoyi
(Institute of Mathematics, Sichuan University Chengdu 610064)

Abstract A question of formation confidence interval for parameter of exponential distribution from incomplete observational data is discussed. A new confidence interval method is given. The pivot variable for confidence interval estimation of the parameter is constructed by making use of any two order statistics. The distribution function and probability density function of this pivot variable is obtained. An approximate confidence interval from large sample is obtained also.

Key words exponential distribution; confidence interval; parameter estimation; distribution function; probability density; incomplete observational data

1 Introduction

In some life tests of product, there are usually incomplete observational data. For instance, in fixed number censored life tests, the data we obtain are observational values of the lower order statistics for a sample. And in some special tests, we only record reaction strength data over a certain limit, which are upper order statistics. Sometimes, some observational data are lost because of unexpected events, and so on. The estimation for parameter of population from incomplete observational data is an important focal issue. But the available methods are mainly point estimation of parameter[1~4]. The approximate confidence intervals for Weibull distribution, log-normal distribution and Normal distribution have been given[5~9].

In this paper, a new estimation method for parameter of population from incomplete observational data is proposed.

2 Pivot Variable and it's Distribution

Let $X_1, X_2, \cdots, X_n$ be an independent and identically distributed sample from exponential distribution population X with probability density function

$$f_X(x) = \begin{cases} l\exp(-lx) & x > 0 \\ 0 & x \leqslant 0 \end{cases}$$

where $l(>0)$ is population parameter. Suppose $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ are order statistics from sample $X_1, X_2, \cdots, X_n$. We have

Proposition 1 Define

Received on Sep.20 2001

$$Z_{k,l}=\frac{X_{(k)}}{lX_{(k+l)}^2}$$

where k and l are any integers with $1\leqslant k<n$ and $1\leqslant l\leqslant n-k$. Then the distribution function of $Z_{k,l}$ is foreign to the value of parameter l .

Proof It is not difficult to prove that $lX_1, lX_2, \cdots, lX_n$ can be regarded as a sample of the population X whose l =1, and as $l>0$ corresponding order statistics are $lX_{(1)}, lX_{(2)}, \cdots, lX_{(n)}$. Since

$$Z_{k,l}=\frac{X_{(k)}}{lX_{(k+l)}^2}=\frac{lX_{(k)}}{[lX_{(k+l)}]^2}$$

for any positive number l , $Z_{k,l}$ has the same distribution as l =1. Thus, distribution function of $Z_{k,l}$ is foreign to the value of parameter l .

Proposition 1 proves that $Z_{k,l}$ is pivot variable for confidence interval estimation of the parameter l .

Now we discuss the distribution function of the pivot variable $Z_{k,l}$.

Lemma 1[7] The joint density function of the order statistics $(X_{(k)}, X_{(k+l)})$ is given by

$$f_{k,k+l}(x_1,x_2)=0 \qquad x_1\geqslant x_2$$

$$f_{k,k+l}(x_1,x_2)=\frac{n!}{(k-1)!(l-1)!(n-k-l)!}[F_X(x_1)]^{k-1}[F_X(x_2)-F_X(x_1)]^{l-1}[1-F_X(x_2)]^{n-k-l}f_X(x_1)f_X(x_2) \qquad x_1<x_2$$

where $F_X(x)$ and $f_X(x)$ are distribution function and density function of the population respectively.

From Lemma 1, we get

Lemma 2 For the population X , if l =1, then the joint density function of $(X_{(k)}, X_{(k+l)})$ is given by

$$f_{k,k+l}(x_1,x_2)=\frac{n!}{(k-1)!(l-1)!(n-k-l)!}[1-\exp(-x_1)]^{k-1}[\exp(-x_1)-\exp(-x_2)]^{l-1}\exp[-(n-k-l+1)x_2-x_1]$$

if $0<x_1<x_2$; else $f_{k,k+l}(x_1,x_2)=0$.

Theorem 1 Let

$$h_{ij}(z)=z_1^3\{\exp(z_1^2)\sqrt{p}(1+2z_1^2)[\varPhi(\sqrt{2}z_2)-\varPhi(\sqrt{2}z_1)]+(2z_1-z_2)\exp(z_1^2-z_2^2)-z_1\}$$

where $z_1=(n-k-j)/2\sqrt{(k-i+j)z}$, $z_2=z_1+\sqrt{(k-i+j)/z}$, $i=0,1,\cdots,k-1$, $j=0,1,\cdots,l-1$; $\varPhi(x)$ is the distribution function of the standard normal distribution. Then the density function of the pivot variable $Z_{k,l}$ is given by

$$f_{kl}(z)=0 \qquad z\leqslant 0$$

$$f_{kl}(z)=\frac{4n!}{(k-1)!(l-1)!(n-k-l)!}\sum_{i=0}^{k-1}\sum_{j=0}^{l-1}\frac{(-1)^{k+l-i-j}}{(n-k-j)^3}C_{k-1}^i C_{l-1}^j h_{ij}(z) \qquad z>0$$

Proof Let($Z_{k,l}$, T)=($\frac{X_{(k)}}{[X_{(k+l)}]^2}$, $X_{(k+l)}$) and $z=\frac{x_1}{{x_2}^2}$, $t=x_2$ .We note that

$$(x_1,\ x_2)=(zt^2,\ t)$$

Therefore the Jacobian

$$|J|=\begin{vmatrix} t^2 & 2tz \\ 0 & 1 \end{vmatrix}=t^2$$

Suppose the joint density function of ($Z_{k,l}$, T) is $g(z,\ t)$, we know that $g(z,\ t)$ is given by[7]

$$g(z,\ t)=t^2\ f_{k,k+l}(zt^2,t)$$

From Lemma 2, $g(z,\ t)\neq 0$ if $0<zt^2<t$, else $g(z,\ t)=0$. We have:

if $z \leqslant 0$, then $zt^2 \leqslant 0$, thus $g(z,\ t) = 0$;

if $z > 0$ and $z \geqslant \frac{1}{t} > 0$, then $zt^2 \geqslant t$, thus $g(z,\ t) = 0$;

if $0 < z < \frac{1}{t}$, we get, $g(z,\ t) = \frac{n!}{(k-1)!(l-1)!(n-k-l)!} t^2 [1-\exp(-zt^2)]^{k-1} [\exp(-zt^2) - \exp(-t)]^{l-1}$ $\exp[-(n-k-l+1)t - zt^2]$.

Since $f_{kl}(z) = \int_{-\infty}^{+\infty} g(z,\ t)\,\mathrm{d}t$, then have

$$f_{kl}(z) = 0 \qquad z \leqslant 0$$

$$f_{kl}(z) = \int_0^{\frac{1}{z}} \frac{n!}{(k-1)!(l-1)!(n-k-l)!} t^2 [1-\exp(-zt^2)]^{k-1} [\exp(-zt^2) - \exp(-t)]^{l-1} \exp[-(n-k-l+1)t - zt^2]\,\mathrm{d}t \qquad z > 0$$

That is

$$f_{kl}(z) = \frac{4n!}{(k-1)!(l-1)!(n-k-l)!} \sum_{i=0}^{k-1} \sum_{j=0}^{l-1} \frac{(-1)^{k+l-i-j}}{(n-k-j)^3} C_{k-1}^i C_{l-1}^j h_{ij}(z)$$

From Theorem 1, get

Theorem 2 The distribution function of the pivot variable $Z_{k,\ l}$ is given by

$$F_Z(x) = \frac{n!}{(k-1)!(l-1)!(n-k-l)!} \sum_{i=0}^{k-1} \sum_{j=0}^{l-1} (-1)^{k+l-i-j} C_{k-1}^i C_{l-1}^j H(x)$$

where

$$H(x) = \frac{1}{(k-i+j)(n-k-j)} - \frac{1}{(k-i+j)(n-i)} \exp\left(-\frac{n-i}{x}\right) - \frac{1}{(k-i+j)} \sqrt{\frac{\pi}{(k-i+j)x}} \exp\left[\frac{(n-k-j)^2}{4(k-i+j)x}\right] \left\{ \Phi\left(\frac{n+k+j-2i}{\sqrt{2(k-i+j)x}}\right) - \Phi\left(\frac{n-k-j}{\sqrt{2(k-i+j)x}}\right) \right\}$$

From Theorem 1 and Theorem 2, it only needs the values of any two order statistics of a sample that we can construct a confidence interval for parameter l. Of course, we need to use numerical computation method to calculate the quantiles of the pivot variable $Z_{k,\ l}$.

3 Approximate Confidence Interval for Large Sample

Because of the complex form of the distribution function of the pivot variable $Z_{k,\ l}$, in this section we provide a way to construct approximate confidence interval for large sample.

Let $0 < a < b < 1$, $k = [na] + 1$, $k + l = [nb] + 1$, where $[na]$ and $[nb]$ stand for integer parts of na and nb respectively. Then $Z_{k,\ l} = X_{([na]+1)} / l X^2_{([nb]+1)}$, where $X_{([na]+1)}$ and $X_{([nb]+1)}$ are sample quantiles of the population. Suppose x_a and x_b are a and b-quantiles of the population respectively, then $X_{([na]+1)}$ and $X_{([nb]+1)}$ are asymptotically unbiased and consistent estimators of x_a and x_b respectively. We compute $x_a = -\ln(1-a)/l$, $x_b = -\ln(1-b)/l$.

Lemma 3[8] If $F(x)$ is absolutely continuous, x_p is relevant p-quantile. $f(x) = F'(x)$ is continuous in $x = x_p$ and nonzero. Then, as $n \to \infty$

$$\sqrt{n}\,[\,X_{([np]+1)} - x_p\,] \xrightarrow{L} N(0,\, p(1-p)/f^2(x_p))$$

From Lemma 3, we immediately get

Lemma 4 Under the conditions of Lemma 3, then

$$X_{([np]+1)} \xrightarrow{P} x_p$$

Lemma 5[8] Suppose both $\{\vec{w}_n\}$ and $\{\vec{e}_n\}$ are k-dimensional random vector series, $\vec{w}_n \xrightarrow{L} \vec{w}$,

$\vec{e}_n \xrightarrow{P} \vec{0}$, as $n \to \infty$; and $\{\vec{B}_n\}$ is $k \times k$-dimensional random matrix series, $\vec{B}_n \xrightarrow{P} \vec{B}$ as $n \to \infty$, where $\vec{w}$ is k-dimensional random vector, and $\vec{B}$ is $k \times k$-dimensional random matrix, $\vec{0}$ is k-dimensional zero vector, thus

$$\vec{B}_n \vec{w}_n + \vec{e}_n \xrightarrow{L} \vec{B}\vec{w} \quad \text{as} \quad n \to \infty$$

Theorem 3 Using the notations defined as above, we have

$$\frac{\sqrt{n}X_{([na]+1)}}{lX^2_{([nb]+1)}} - \frac{\sqrt{n}x_a}{lx_b^2} \xrightarrow{L} N(0, \frac{a(1-a)}{l^2 x_b^4 f^2(x_a)})$$

Proof By Lemma 4, $X_{([nb]+1)} \xrightarrow{P} x_b$, then $X^2_{([nb]+1)} \xrightarrow{P} x_b^2$, thus

$$\frac{\sqrt{n}x_a}{lX^2_{([nb]+1)}} - \frac{\sqrt{n}x_a}{lx_b^2} = \frac{\sqrt{n}x_a}{l}\frac{x_b^2 - X^2_{([nb]+1)}}{x_b^2 X^2_{([nb]+1)}} \xrightarrow{P} 0$$

Now noting that

$$\frac{\sqrt{n}X_{([na]+1)}}{lX^2_{([nb]+1)}} - \frac{\sqrt{n}x_a}{lx_b^2} = \frac{\sqrt{n}(X_{([na]+1)} - x_a)}{lX^2_{([nb]+1)}} + \left[\frac{\sqrt{n}x_a}{lX^2_{([nb]+1)}} - \frac{\sqrt{n}x_a}{lx_b^2}\right]$$

We obtain by Lemma 3

$$\sqrt{n}\,[\,X_{([na]+1)} - x_a\,] \xrightarrow{L} N(0, a(1-a)/f^2(x_a))$$

And by Lemma 5, we obtain the stated result.

From Theorem 3, as n is enough large, the pivot variable $Z_{k,l} = X_{([na]+1)} / lX^2_{([nb]+1)}$ is approximately normal distribution

$$N\left(\frac{a(1-a)}{nl^2 x_b^4 f^2(x_a)}, \frac{x_a}{lx_b^2}\right) \tag{1}$$

Also note that $x_a = -\ln(1-a)/l$, $x_b = -\ln(1-b)/l$, $f(x_a) = l(1-a)$, substituting these quantiles in (1), we obtain that the approximate distribution of $X_{([na]+1)} / lX^2_{([nb]+1)}$ is

$$N\left(\frac{-\ln(1-a)}{[\ln(1-b)]^2}, \frac{a}{n(1-a)[\ln(1-b)]^4}\right) \tag{2}$$

In large sample case, for simplicity, let $a = (k-1)/n$, $b = (k+l-1)/n$. If the values of k and $k+l$ is near to $n/2$, and $a/n(1-a)[\ln(1-b)]^4 \to 0$ as $n \to \infty$. That is, the accuracy of confidence interval is increasing in n.

By expression (2), we can construct approximate confidence intervals for l. For instance, an approximate intervals with confidence level $(1-a)$ for l is

$$\left(\frac{X_{([na]+1)}[\ln(1-b)]^2\sqrt{n(1-a)}}{X^2_{([nb]+1)}[\sqrt{a}u_{a/2} - \sqrt{n(1-a)}\ln(1-a)]}, \frac{X_{([na]+1)}[\ln(1-b)]^2\sqrt{n(1-a)}}{X^2_{([nb]+1)}[-\sqrt{a}u_{a/2} - \sqrt{n(1-a)}\ln(1-a)]}\right)$$

where $u_{a/2}$ is the level $a/2$ upper quantile of the standard normal distribution.

4 Conclusion

In this article, we suggest a new confidence interval method for parameter of exponential distribution from highly incomplete observational data. Our method is still suitable even if there are outliers in the sample observational data. In fact, outliers are usually the several largest or smallest data of observation. Obviously, if k and $k+l$ near to $n/2$ are chosen, we can make the result to avoid the disturbance of outliers. Furthermore, as we have pointed out, the accuracy of confidence interval for large sample is very high.

References

1 Cheng R C H, Iles T C. Corrected maximum likelihood in non-regular problems. Journal of Royal Statistical Society, 1987, B49(1): 95-101

2 Gourdin E, Hansen P, Jaumard B. Finding maximum likelihood estimators for the three-parameter Weibull distribution. Journal of Global Optimization, 1994, 5 : 373-397

3 Zanakis S H. A simulation study of some simple estimator of the three-parameter Weibull distribution. Journal of Statistical Computation and Simulation, 1979, 9(2): 101-116

4 Schmid U. Percentile estimators for the three-parameter Weibull distribution for use when all parameters are unknown. Communications in Statistics: Theory and Method, 1997, 26 : 765-785

5 Fei H L, Xu X L. Simultaneous confidence region of three parameter for the three-parameter Weibull distribution. Chinese Journal of Applied Probability and Statistics, 1992, 8(4): 398-402

6 Xu X L, Fei H L. The confidence bounds of location parameter for three-parameter log-normal distribution. Mathematical Statistics and Applied Probability (Chinese), 1991, 6(3): 382-392

7 Zhu Hong, Lü Shu. Confidence interval estimation for the location parameter of Normal distribution from incomplete data. Journal of applied sciences (Chinese), 1996, 14 (2):173-178

8 Zhu Hong, Lü Shu. A result of Normal distribution from order statistic. Journal of University Electronic Science and Technology of China, 1995, 24(3): 313-318

9 Luo Yan. Confidence interval estimation for parameter of exponential distribution from incomplete data. Journal of University Electronic Science and Technology of China, 1998, 27(4): 445-448

指数分布参数基于不完全样本的区间估计

朱　宏*　　　　肖百龙　　　　徐道义

(四川大学数学学院　成都　610064)　(湖南大学统计系　长沙　410082) (四川大学数学学院　成都　610064)

【摘要】对于不完全样本讨论指数分布总体参数的区间估计问题，利用完全样本中的任意两个顺序统计量构造出区间估计所需的枢轴变量，并讨论了相应的分布函数和密度函数。即使只知道样本观测值中任意的两个顺序统计量值，也可以计算出总体参数的置信区间。在大样本的情况下，给出了枢轴变量的近似分布，可以构造总体参数的大样本近似置信区间。

关　键　词　指数分布；置信区间；参数估计；分布函数；密度函数；不完全观测值

中图分类号　O212.1

2001年9月20日收稿

* 男　38岁　在职博士生　教授