嘉南藥理科技大學教師補助專題研究計畫 ■ 成 果 報 告
□期中進度報告

整合型計畫：推估設限資料的未來失敗時間與數目

子計畫（2）在 pareto 分配下估算設限資料的信賴信間

計畫類別：□ 個別型計畫　■ 整合型計畫
計畫編號：CNMI －93－01－子計畫(2)
執行期間：　93 年 01 月 01 日至　　93 年 12 月 31 日

計畫主持人：陸海林　教授
共同主持人：薛雅明　助理教授
計畫參與人員：

成果報告類型(依經費核定清單規定繳交)：■精簡報告　□完整報告

執行單位：嘉南藥理科技大學資管系

中　華　民　國　93　年　2　月　25　日

Prediction intervals of an ordered observation from Pareto Distribution based on Doubly Type II censored samples

ABSTRACT

Pareto Distribution is widely used in biostatistics and economics areas. In this paper, we provide some suitable pivotal quantities for estimating the prediction intervals of the *j*th future ordered observation in a sample of size *n* from the pareto distribution based on doubly type II censored samples. We also give simulation study to analyze its feasibility of pivotal quantities. Finally, two illustrative examples are also included.

***Keywords*:** pivotal quantity, doubly type II censored samples, pareto distribution, approximation maximum likelihood estimation.

1. INTRODUCTION

Pareto Distribution is widely used in many different areas, especially in biostatistics, insurance and management (see Johnson etc., 1995). The probability density function (pdf) of a pareto variable ***X*** is simply given by

$$f(x) = \frac{\beta}{\alpha}(1+\frac{y-\mu}{\sigma})^{-(\beta+1)} \qquad \begin{array}{c} \mu < y < \infty, \\ -\infty < \mu < \infty, \\ \sigma > 0, \beta > 0 \end{array}$$

and the corresponding cumulative distribution function (cdf) is

$$F(x) = 1-(1+\frac{y-\mu}{\sigma})^{-\beta} \qquad (1)$$

where μ and σ are location and scale parameters, respectively. The notation Y ~ P(μ, σ , β) will be used to indicate that a random variable ***X*** has c.d.f. (1).

In life testing experiments, the experiment may not be always in a position to observe the life times of all the items put on test. In this paper , we suppose the parameter β is known. This may be because of limitation and/or other restrictions (such as money, and material resources, etc.) on data collection. In this article, we consider doubly type II censored data that out of *n* items put on life test, the middle *k* items $X_{(r+1)} < X_{(r+2)} < \cdots < X_{(r+k)}$ have been observed only and the life times for the rest *r*+*s* components remain unobserved or missing (Upadhyay *et al*., 1996). Gupta *et al*. (1967) have worked out and established tables for the best linear unbiased estimation (BLUE) of μ and σ based on the doubly censored samples for selected sample size up to 25 and for some selected choices of censoring. This tables have been extended by Balakrishnan (1990) for sample sizes up to 40.

In this paper, we estimate the prediction intervals of unknown observations. We provide some suitable pivotal quantities and refer to AMLE estimators to give a pivotal quantity for estimating the prediction intervals. The fully structure of the paper is as follows. Section 2 lists our providing pivotal quantities. In section 3, we offer two examples to illustrate the methods in this paper. A brief discussion is included at last.

2. METHODS AND RESULTS

We provide some suitable pivotal quantities that are independent with μ and σ parameters to estimate the prediction intervals of unknown observations. Let

$$\hat{U}_t = \frac{X_{(j)} - X_{(n-s)}}{\hat{W}_t} \qquad n-s < j \le n \text{ , } 1 \le t \le 9 \qquad (2)$$

where

$$\hat{W}_1 = \sum_{i=2}^{k} (X_{(r+i)} - X_{(r+1)}) \Big/ (k-1) \qquad (3)$$

$$\hat{W}_2 = \left\{ \sum_{i=2}^{k-1} (X_{(r+i)} - X_{(r+1)}) + (s+1)(X_{(r+k)} - X_{(r+1)}) \right\} \Big/ (k+s-1) \qquad (4)$$

$$\hat{W}_3 = \left\{ (r+1)(X_{(r+2)} - X_{(r+1)}) + \sum_{i=2}^{k} (X_{(r+i)} - X_{(r+1)}) \right\} \Big/ (r+k) \qquad (5)$$

$$\hat{W}_4 = \left\{ (r+1)(X_{(r+2)} - X_{(r+1)}) + \sum_{i=2}^{k-1} (X_{(r+i)} - X_{(r+1)}) + (s+1)(X_{(r+k)} - X_{(r+1)}) \right\} \Big/ n \qquad (6)$$

$$\hat{W}_5 = \left[\prod_{i=2}^{k} (X_{(r+i)} - X_{(r+1)}) \right]^{1/(k-1)} \qquad (7)$$

$$\hat{W}_6 = \left\{ \left[\prod_{i=2}^{k-1} (X_{(r+i)} - X_{(r+1)}) \right] (X_{(r+k)} - X_{(r+1)})^{s+1} \right\}^{1/(s+k-1)} \qquad (8)$$

$$\hat{W}_7 = \left\{ (X_{(r+2)} - X_{(r+1)})^{r+1} \prod_{i=2}^{k} (X_{(r+i)} - X_{(r+1)}) \right\}^{1/(r+k)} \qquad (9)$$

$$\hat{W}_8 = \left\{ (X_{(r+2)} - X_{(r+1)})^{r+1} \left[\prod_{i=2}^{k-1} (X_{(r+i)} - X_{(r+1)}) \right] (X_{(r+k)} - X_{(r+1)})^{s+1} \right\}^{1/n} \qquad (10)$$

where

$$p_i = \frac{i}{n+1} \text{ , } q_i = 1 - p_i \text{ , } \gamma_i = p_i - p_i q_i \ln(\frac{p_i}{q_i}) \text{ , } \delta_i = p_i q_i \text{ , }$$

$$A = n - r - s \text{ , }$$

$$m = r\delta_{r+1} + s\delta_{n-s} + 2\sum_{i=r+1}^{n-s} \delta_i \text{ , }$$

From (3) ~ (11), we know the pivotal quantities $\hat{W}_1$ ~ $\hat{W}_4$ are arithmetic mean, $\hat{W}_5$ ~ $\hat{W}_8$ are geometric mean. Because $\frac{X_{(1)}-\mu}{\sigma}$, $\frac{X_{(2)}-\mu}{\sigma}$, … , $\frac{X_{(n)}-\mu}{\sigma}$ are derived from *n* order statistics of P(0,1), we can easily prove that $\hat{W}_1$ ~ $\hat{W}_9$ pivotal quantities do not depend on μ and σ. Hence, the distribution of $\hat{U}_t$ only depend on *n*, *j*, *r*, *s*, but not on μ and σ. Thus

$$1-\alpha = P\left\{0 < \frac{X_{(j)} - X_{(n-s)}}{\hat{W}_t} < \hat{u}_t(1-\alpha; n, j, r, s)\right\}$$

$$= P\left\{X_{(n-s)} < X_{(j)} < X_{(n-s)} + \hat{W}_t \hat{u}_t (1-\alpha; n, j, r, s)\right\} \quad t = 1,...,9 .$$

Therefore, $(X_{(n-s)}, X_{(n-s)} + \hat{W}_t \hat{u}_t (1-\alpha; n, j, r, s))$ are one-sided 100(1-α)% prediction intervals of $X_{(j)}$ based on $n-s-r$ observations.

3. ILLUSTRATED EXAMPLES

There are two illustrative examples as follows.

Example 1. We consider the following 20 observations data (Davis, 1952). Suppose the last 50% data were censored . The first 10 observations are given below:

785, 855, 905, 918, 919, 920, 929, 936, 948, 950

In this case, we have n=20, r=0, and s=10, and apply our methods to estimate the one-sided 95% prediction intervals of $X_{(11)}$ and $X_{(12)}$, respectively. The estimation results are presented in Table 1.

Table 1. **the one-sided prediction intervals and the percentiles of $\hat{U}_t$ (t=1,…,8)**

	$\hat{u}_t(0.95;20,11,0,10)$	$\hat{u}_t(0.95;20,12,0,10)$	**One-sided 95% prediction intervals of $X_{(11)}$**	**One-sided 95% prediction intervals of $X_{(12)}$**
$\hat{U}_1$	0.3310	0.5599	(950, 994)	(950, 1035)
$\hat{U}_2$	0.2543	0.4257	(950, 988)	(950, 1020)
$\hat{U}_3$	0.3572	0.6137	(950, 995)	(950, 1037)
$\hat{U}_4$	0.2643	0.4518	(950, 988)	(950, 1024)
$\hat{U}_5$	0.4041	0.6977	(950, 1003)	(950, 1041)
$\hat{U}_6$	0.2845	0.4793	(950, 992)	(950, 1030)
$\hat{U}_7$	0.4815	0.8423	(950, 1009)	(950, 1056)
$\hat{U}_8$	0.3130	0.5284	(950, 994)	(950, 1035)

From Table 1, we can find that using $\hat{U}_2$, and $\hat{U}_4$ could get the shorest prediction intervals in all of the other pivotal quantities in Example 1.

Example 2. The following data (Sarhan and Greenberg, 1962) presented the measurements results of an experiment to measure the strontium-90 concentrations in samples of milk. Because the measurement error known was existed at the extremes, the two smallest and the three largest observations were censored. The remaining five observations are given below:

8.2, 8.4, 9.1, 9.8, 9.9

In this case, we have n=10, r=2, and s=3, and make use of our methods to estimate the one-sided 95% prediction intervals of $X_{(8)}$ and $X_{(9)}$, respectively. The estimation results is listed in Table 2.

Table 2. the one-sided prediction intervals and the percentiles of $\hat{U}_t$ (t=1,…,8)

	$\hat{u}_t(0.95;10,8,2,3)$	$\hat{u}_t(0.95;10,9,2,3)$	**One-sided 95% prediction intervals of $X_{(8)}$**	**One-sided 95% prediction intervals of $X_{(9)}$**
$\hat{U}_1$	1.9108	3.8040	(9.9, 12.01)	(9.9, 14.10)
$\hat{U}_2$	1.4787	2.9011	(9.9, 11.90)	(9.9, 13.86)
$\hat{U}_3$	2.7583	5.5957	(9.9, 11.87)	(9.9, 13.86)
$\hat{U}_4$	1.8874	3.7828	(9.9, 11.80)	(9.9, 13.68)
$\hat{U}_5$	2.7638	5.6030	(9.9, 12.21)	(9.9, 14.66)
$\hat{U}_6$	1.8359	3.6672	(9.9, 11.98)	(9.9, 14.05)
$\hat{U}_7$	5.7004	12.4846	(9.9, 12.48)	(9.9, 15.55)
$\hat{U}_8$	3.2393	6.8155	(9.9, 12.08)	(9.9, 14.48)

From Table 2, we can see that using $\hat{U}_2$, $\hat{U}_3$, and $\hat{U}_4$ could get the shortest prediction intervals in all of the other pivotal quantities in Example 2.

5. DISCUSSION

Since all providing $\hat{U}_1 \sim \hat{U}_8$ statistics are pivotal quantities and make use of simple algebraic expressions to represent ordered observations directly, the calculations are not difficult.

From the Table 1 and 2, the length of prediction intervals of unknown observations, we find the length of $\hat{U}_1 \sim \hat{U}_4$ are smaller than $\hat{U}_5 \sim \hat{U}_8$. The former use arithmetic mean expressions $(\hat{W}_1 \sim \hat{W}_4)$ as denominators, and the latter take geometric mean expressions $(\hat{W}_5 \sim \hat{W}_8)$ as denominators. As a whole, the smallest is $\hat{U}_9$ which use AMLE of σ pivotal quantity as denominator.

The performance of $\hat{U}_2$ is the best from the results of $\hat{U}_1 \sim \hat{U}_8$. From the simulation outputs presented, the error $\hat{U}_2$'s results is less than 1%. However, it is very easily for making tables of different doubly censored data because of the convenience and efficiency computers. Hence, it is very conveniently for estimating the prediction intervals of unknown ordered observations when we adopt $\hat{U}_2$.

REFERENCES

Balakrishnan, N. (1990) "Best linear unbiased estimates of the location and scale parameters of pareto distribution for complete and censored samples of size 2(1)25(5)40," submitted for publication.

Davis, D. J. (1952), "An analysis of some failure data," *Journal of the American Statistical Association*, 47, 113-150.

Fisk, P. R. (1961) "The graduation of income distribution," *Econometrica*, 29, 171-185.

Gupta, S. S., Qureishi, A. S., and Shah, B. K. (1968) "Best linear unbiased estimators of the parameters of the Pareto distribution using order statistics," *Technometrics*, 9, 43-56.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995) *Distribution in statistics: Continuous Univariate Distribution*, Vol. 2. John Wiley & Sons, New York.

Sarhan, A. E. and Greenberg, B. G. (1962) *Contributions to order statistics*, John Wiley & Sons, New York.

Upadhyay, S. K., Singh,. U. and Shastri, V., (1996) "Estimation of exponential parameters under multiply type II censoring," *Statistics Simulation*, 25(3), 801-815.

在雙邊型二設限資料的柏拉圖分配中
求未知觀測值的預測區間

摘　要

羅吉斯分配廣泛的應用在生物統計與經濟領域中，本文中我們對來自雙邊型二設限資料之柏拉圖分配中的未知觀測值，提出以一些適當的基準量，預測樣本大小爲 n 中第 j 個元件順序觀測值之信賴區間，同時對這些基準量以模擬方式評估其可行性。最後，並提供兩個實際例子來討論。

關鍵詞：基準量、雙邊型二設限資料、柏拉圖分配、近似最大概似估計。