# 嘉南藥理科技大學專題研究計畫成果報告

## 推估設限資料的未來失敗時間與數目

計畫類別：□個別型計畫　　　　　■整合型計畫
計畫編號：CNMI93-01
執行期間：93 年 1 月 1 日至 93 年 12 月 31 日

整合型總主持: 陸海林

子計畫一：在 Weibull 分配下以 Bayesian 方法估算設限資料的未來失敗數目
計畫主持人：陸海林

子計畫二：在 Pareto 分配下估算社限資料的信賴信間
計畫主持人：薛雅明

執行單位：嘉南藥理科技大學

中華民國　94 年　2 月　24 日

# 推估設限資料的未來失敗時間與數目

## 在 **Weibull** 分配下以 **Bayesian** 方法估算設限資料的未來失敗數目


### 摘要

韋伯分配(Weibull distribution)被廣泛的應用在品質管制中可靠度的問題上，通常在設限資料下兩參數韋伯分配中參數的估計是較困難的，本文針對兩參數韋伯分配提供一個線性加權最小平方法估計參數。由此估計值來估算未來某特定時間內所發生之失敗數目。



**關鍵詞**： 韋伯分配、 加權最小平方法估計、次序統計量、蒙地卡羅模擬。


### 1. 緒論

通常韋伯分配廣泛的應用在可靠度與生命中存活壽命的研究領域中。如此之分配最先由 Weibull 於 1939[1]所提出，主要爲研究不同的損壞狀況可參考 Weibull 所作的論文[2]，其他也應用在鋼所產生的強度[3]，電子元件的損壞率－－－等等。機率密度函數與累積分配函數分別爲

$$f(x) = \beta\alpha^{-\beta}x^{\beta-1}\exp[-(\frac{x}{\alpha})^{\beta}]，$$

$$F(x) = 1-\exp[-(\frac{x}{\alpha})^{\beta}]，$$

其中 x>0 ， $\alpha$>0，$\beta$>0 此兩母數分別當作尺度(scale)和形狀(shape)母數。

若$\beta$>1 時韋伯分配的危險函數(hazard function)爲單調遞增(monotone increasing)，$\beta$<1時危險函數(hazard function)爲單調遞減(monotone decreasing)，$\beta$=1 時韋伯分配爲簡單的指數分配，近年很多學者欲估算未來某特定時間內所發生之失敗數目(參考 Nelson[4]，Nordman 和 Meeker[5]，後者採用以概似(Likelihood)爲基礎架構求出未來某特定時間內所發生之失敗數目的預估區間，前者採用以最大概似法(Maximum likelihood method)爲基礎架構先估算出$\alpha$，$\beta$在由母數值與樣本數配合機率估算未來的失敗數目。

近來估算 $\alpha$ ，$\beta$ 的文章包括的有 Bergman[6]；Faucher 和 Tyson[7]；Sinha 和 Guttman[8]；Chandhuri 和 Chandra[9]；Lockhart 和 Stephens[10]；Hossain 和 Howlader[11]；Drapella 和 Kosznik[12] 與 Hung[13]。這篇文章之動機爲來自 Johnson *et al.* [14]所提的用母數與次序統計量之線性關係的近似法，來估算 $\alpha$ ，$\beta$ 以及參考 Ananda *al.* [15]所提的適應性貝氏估計，其方式爲僅給事先分布一個參數，再利用使經驗分佈與累積分佈函數的距離最小來得到另一參數。以下精要說明兩種方法：

若令 T=－log(x)此 T 將具有極值分配(Extreme distribution)其 $\mu$=－log(1/ $\alpha$ )，b=1/ $\beta$ 分別爲位置與尺度母數

假設 *X* 爲兩參數極值分配機率密度函數與累積分配函數分別爲

$$f(x) = \frac{1}{\alpha} e^{\frac{x-\mu}{b}} e^{-e^{\frac{x-\varpi}{b}}}$$

$$F(x) = 1 - e^{-e^{\frac{x-\mu}{b}}} .$$

Then we have

$$\ln(-\ln(1-F(x))) = \frac{1}{b}x - \frac{\mu}{b}$$

$$\ln(Z_{(i)}) = \frac{1}{b}x - \frac{\mu}{b},$$

其中 $h(Z_{(i)}) = \ln(Z_{(i)})$ 。

再取 $w_i = \dfrac{1}{Z_{(i)}^2 \operatorname{var}(Z_{(i)})}$ 爲加權(weight)用最小平方法求估 $\mu$ ，b，再轉成 $\alpha$ ，$\beta$ 。

若 X1 表示到時間 t 時失敗的數目，X2 表示到時間 t 與 s 時的失敗的數目(t<s)， X3 表示到時間 s 時的未失敗的數目，則(X1，X2，X3)有一三度分配相關的機率爲(p、q、r)，其中 X1+X2+X3=n，p+q+r=1。

我們可能求的

$$p = 1 - \exp[-(\frac{t}{\alpha})^{\beta}]$$

$$q = \exp[-(\frac{t}{\alpha})^{\beta}] - \exp[-(\frac{s}{\alpha})^{\beta}]$$

$$r = \exp[-(\frac{s}{\alpha})^{\beta}]$$

其相關之預估值 $\hat{Y} = n\hat{q}$ ，其信賴區間可依 q 的區間估計而得。

另由 Bayesian 方式取 $\lambda = \dfrac{1}{\alpha^{\beta}}$ 得 $f(x) = \beta\lambda x^{\beta-1}\exp[-\lambda x^{\beta}]$ ，令 $\lambda$ 的事先分配(prior distribution) 取 gamma distribution $g(\lambda) \propto \lambda^{c-1}\exp(-\lambda a)$ ，0≦ $\lambda$ <∞，a，c>0 。再利用 Ananda *al.* [15]所提的適應性貝氏估計估算出參數的信賴區間，依前法可的 Y 的區間估計即爲未來的失敗數目的區間估計。

註：本文 WLSE 部份已投稿爲 Quality & Quantity (SCI)期刊接受並刊登在該期刊 2004，20；579-586。

## 參考文獻

[1] Wiebull W. A statistical theory of the strength of material. *Ingeniors Vetenskaps Akakemien Handlingar* 1939；**15**:293-297.

[2] Weibull W. A statistical distribution of wide applicability. Journal of Applied Mechanics 1951; **18**:293-297.

[3] Rao, S.S. *Reliability based design*, McGraw-Hill, Inc., U.S.A. (1992).

[4] Nellson W. Weibull prediction of a future number of failures., *Quality and Reliability Engineering International*, **16**, 23-26(2000)..

[5] Nordman D. J., and Meeker W. Q. Weibull prediction intervals for a future number of failures., *American Statistical Association and the American Society for Quality, Technometrics*, 2002, Vol 44, No. 1,15-23.

[6] Bergman B. “Estimation of Weibull parameters using a weight function”, *Journal of Materials Science Letters*, **5**, 611-614(1986).

[7] Faucher B., and Tyson W.R. "On the determination of Weibull parameters", *Journal of Materials Science Letters*, **7**, 1199-1203(1988).

[8] Sinha SK, Guttman I. Bayesian analysis of life-testing problems involving the Weibull distribution. Communications in Statistics-Theory and Methods 1988; 1792):343-356.

[9] Chaudhuri AR, Chandra NK. A test for Weibull population. Statistics and Probability Letters 1989;7(5):377-380.

[10] Lockhart RA., Stephens MA., Estimation and tests of fit for the three-parameter Weibull distribution. Journal of the Royal Statistical Society, Series B 1994; 56(3); 491-500.

[11] Hossain A, Howlader HA. Nweighted least-squares estimation of Weibull parameters. Jounal of Statistical computation and Simulations 1996; 54;265-271.

[12] Drapella A., and Kosznik S. "An alternative rule for placement of empirical points on Weibull probability paper", *Quality and Reliability Engineering International*, **15**, 57-59(1999).

[13] Hung W. L. “Weight least-squares estimation of the sharp parameter of the Weibull distribution”, *Quality and Reliability Engineering International*, **17**, 467-469(2001).

[14] Johnson NL., Kotz S, Balakrishnan N., Continuous Univeriate Distributions-I(2nd edn), Wiley: New York, 1994.

[15] Ananda M.M., Dalpatadu, R.J.D. and Singh A.K. Adaptive Bayes estimators for parameters of Gompertz survival model. Appl. Math & computation.,1996: 75,165-177.

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 $\mu$ and $\sigma$ are location and scale parameters, respectively. The notation Y ~ P( $\mu$ , $\sigma$ ， $\beta$ ) 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 $\beta$ 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 $\mu$ and $\sigma$ 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 $\mu$ and $\sigma$ parameters to estimate the prediction intervals of unknown observations. Let

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

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

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

$$A = n - r - s \ ,$$

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

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 $\mu$ and $\sigma$. Hence, the distribution of $\hat{U}_t$ only depend on *n*, *j*, *r*, *s*, but not on $\mu$ and $\sigma$. 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-$\alpha$)% 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 $\sigma$ 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。