行政院國家科學委員會專題研究計畫成果報告

在逐步設限隨機移除下柏拉圖分配之估計
Estimation in the Pareto Distribution Based on Progressive Censoring with Random Removals

計畫編號：NSC 90-2118-M-032-013
執行期限：90年8月1日至91年7月31日
主持人：吳碩傑　淡江大學統計學系
計畫參與人員：曾小喬 淡江大學統計學系碩士班

1 中文摘要

本研究將探討在逐步設限隨機移除下柏拉圖分配參數的估計問題。在此設限試驗中，我們假設在每一次故障發生時，隨機移除若干個未故障之測試產品，而每次移除的個數具有間斷型的均勻分配。我們將利用最大概似法來求出柏拉圖分配參數的估計式，並將推導出參數的信賴區間。此外，我們也將計算出在該實驗的期望執行時間，並舉例子來闡述我們所得到的結果。

關鍵詞：最大概似估計式；柏拉圖分配；逐步型 II 設限；隨機移 除；期望實驗時間。

Abstract This study considers the estimation problem for the Pareto distribution based on progressive Type II censoring with random removals, where the number of units removed at each failure time has a uniform distribution. We use the maximum likelihood method to obtain the estimators of parameters and the distributions of the estimators are derived. We also construct the confidence intervals for the parameters and percentile of the lifetime distribution. The expected time required to complete this censoring test is computed. Some numerical results of expected test times are carried out for this type of progressive censoring and other sampling schemes.

Keywords: maximum likelihood estimator; Pareto distribution; progressive type II censoring; random removals; expected test time.

2 Introduction

In Type II censoring, a total of n units is placed on test, but instead of continuing until all n units have failed, the test is terminated at the time of the m-th $(1 \leq m \leq n)$ unit failure. If an experimenter desires to remove live units at points other than the final termination point of the life test, the above described scheme will not be of use to the experimenter.

A generalization of Type II censoring is progressive Type II censoring. Under this scheme, n units are placed on test at time zero, and m failures are going to be observed.

When the first failure is observed, r_1 of the surviving units are randomly selected and removed. At the second observed failure, r_2 of the surviving units are randomly selected and removed. This experiment terminates at the time when the m-th failure is observed and the remaining $r_m = n - r_1 - r_2 - \cdots - r_{m-1} - m$ surviving units are all removed.

Note that, in this scheme, $r_1, r_2, \cdots, r_m$ are all pre-fixed. However, in some practical situations, these numbers may occur at random. Tse and Yuen (1998) indicated that, for example, the number of patients drop out from a clinical test at each stage is random and cannot be pre-determined. In some reliability experiments, an experimenter may decide that it is inappropriate or too dangerous to carry on the testing on some of the tested units even though these units have not failed. In these cases, the pattern of removal at each failure is random.

In this study, we assume that the lifetimes have a two-parameter Pareto distribution and discuss some results of progressively Type II censored order statistics arising from it. As mentioned by Balakrishnan and Aggarwala (2000), the Pareto distribution is not one of the distributions that is frequently used in life test studies, but it is an interesting distribution to discuss, particularly from a mathematical point of view. There are some results documented in the literature on the inference of the Pareto distribution. Some related works can be found, for example, in Vännman (1976), Arnold and Press (1989), and Ouyang and Wu (1994).

The maximum likelihood estimators for the parameters in an explicit form are derived in Section 3. We obtain the sampling distributions of the maximum likelihood estimators and derive the confidence intervals for the parameters in Section 4. The estimator and confidence interval for the percentile of lifetime distribution are also discussed in Section 4. In Section 5, we compare the expected test times for progressive Type II censoring, Type II censoring, and complete sampling schemes.

3 Model

The exponential distribution is widely used as a model of lifetime data. This distribution is characterized by a constant failure rate, say λ. As noted by McNolty *et al.* (1980), in a population of units there could be a ubiquitous variation in λ-values because of small fluctuations in manufacturing tolerances so that a unit selected at random can be regarded as belonging to a random subpopulation. Let the lifetime of a particular unit have an exponential distribution with failure rate λ, and let λ follow a Gamma distribution with scale parameter θ and shape parameter ν. Then the failure time T of a unit selected at random from such a mixed population has a Pareto distribution with probability density function

$$f_T(t) = \frac{\nu}{\theta}\left(1 + \frac{t}{\theta}\right)^{-(\nu+1)}, \quad t > 0, \quad \theta, \nu > 0.$$

Set $X = T + \theta$. Then, the probability density function and cumulative distribution function are

$$f_X(x) = \frac{\nu}{\theta}\left(\frac{x}{\theta}\right)^{-(\nu+1)}, \quad x > \theta, \quad \theta, \nu > 0,$$

and

$$F_X(x) = 1 - \left(\frac{x}{\theta}\right)^{-\nu}, \quad x > \theta, \qquad (1)$$

respectively. The relation given by (1) is now more properly known as the Pareto distribution of the first kind.

Suppose n independent units are placed on a test with the corresponding life times being identically distributed with probability density function $f_X(x)$ and cumulative distribution function $F_X(x)$. For simplicity of notation, let $(X_1, X_2, \cdots, X_m)$ denote a progressively Type II censored sample. Then, the joint probability density function of all m progressively Type II censored order statistics is

$$f_{X_1,\cdots,X_m}(x_1,\cdots,x_m) = c\prod_{i=1}^{m} f_X(x_i)[1-F_X(x_i)]^{r_i}, \quad x_1 < \cdots < x_m, \tag{2}$$

where $c = n(n-r_1-1)\cdots(n-r_1-r_2-\cdots-r_{m-1}-m+1)$. Thus, for a progressive Type II censoring with pre-determined number of removals $\boldsymbol{R} = \boldsymbol{r}$, the conditional likelihood function can be written as

$$L(\nu,\theta|\boldsymbol{R}=\boldsymbol{r}) = c\nu^m\theta^{\nu n}\prod_{i=1}^{m} x_i^{-\nu(r_i+1)-1}, \tag{3}$$

for $\theta < x_1 < \cdots < x_m$, where c is defined in (2). Now suppose that the number of units removed at each failure time follows a discrete uniform distribution such that $P(R_1 = r_1) = \frac{1}{n-m+1}$, and $P(R_i = r_i|R_{i-1} = r_{i-1},\ldots,R_1 = r_1) = \frac{1}{n-m-(r_1+\cdots+r_{i-1})+1}$, where $0 \le r_1 \le n-m$ and $0 \le r_i \le n-m-(r_1+\cdots+r_{i-1})$, $i = 2,3,\ldots,m-1$. Suppose further that R_i is independent of X_i. Then the likelihood function can be expressed as $L(\nu) = L(\nu|\boldsymbol{R} = \boldsymbol{r})P(\boldsymbol{R}=\boldsymbol{r})$, where

$$P(\boldsymbol{R}=\boldsymbol{r}) = P(R_{m-1}=r_{m-1}|R_{m-2}=r_{m-2},\ldots,R_1=r_1)\cdots P(R_2=r_2|R_1=r_1)P(R_1=r_1). \tag{4}$$

It is clear that $P(\boldsymbol{R}=\boldsymbol{r})$ does not depend on the parameters ν and θ and, hence the maximum likelihood estimators (MLEs) of ν and θ can be obtained by maximizing (3) directly. Since this likelihood function is an increasing function of θ, and therefore, the MLE of θ is given by $\hat{\theta} = X_1$. We then find that (3) with $\theta = X_1$ is maximized for ν by $\hat{\nu} = \frac{m}{\sum_{i=1}^{m}(r_i+1)\log x_i - n\log\hat{\theta}}$.

4 Some Further Results

4.1 Distributions of $\hat{\nu}$ and $\hat{\theta}$

Let $Y_i = \log\left(\frac{X_i}{\theta}\right)$, $i = 1,\ldots,m$. It is easy to show that $Y_1, Y_2, \ldots, Y_m$ is a progressively Type II censored sample from the exponential distribution with mean $\frac{1}{\nu}$. For a fixed set of $\boldsymbol{R} = (r_1, r_2, \ldots, r_m)$, it is easy to show that $2n\nu(\log\hat{\theta} - \log\theta)$ has a chi-square distribution with 2 degrees of freedom, and $\frac{2m\nu}{\hat{\nu}}$ has a chi-square distribution with $2(m-1)$ degrees of freedom. Because conditional distribution is independent of the values of $\boldsymbol{R}$, it must follow that the marginal distributions of $2n\nu(\log\hat{\theta} - \log\theta)$ and $\frac{2m\nu}{\hat{\nu}}$ are also chi-square distributions with 2 and $2(m-1)$ degrees of freedom, respectively. It is also easily seen that $\hat{\theta}$ and $\hat{\nu}$ are independent.

4.2 Confidence Intervals for ν and θ

Confidence intervals for ν and θ are easily to obtained from the above results. Confidence intervals for ν can be derived through the pivotal quantity $\frac{2m\nu}{\hat{\nu}} \sim \chi^2_{(2m-2)}$. Let $\chi^2_{p(2m-2)}$ denote the percentile of chi-square distribution with $2m-2$ degrees of freedom and right-tail probability p. Then

$$\left(\frac{\hat{\nu}}{2m}\chi^2_{1-\frac{\alpha}{2}(2m-2)}, \ \frac{\hat{\nu}}{2m}\chi^2_{\frac{\alpha}{2}(2m-2)}\right)$$

is an equitailed, two-sided $100(1-\alpha)\%$ confidence interval for ν.

To obtain confidence interval for θ we consider the pivotal quantity $\frac{n(m-1)\hat{\nu}(\log\hat{\theta}-\log\theta)}{m} \sim F_{(2,2m-2)}$. Thus, a $100(1-\alpha)\%$ confidence interval for θ is given by

$$\left(\hat{\theta}e^{-\frac{mF_{\alpha(2,2m-2)}}{n(m-1)\hat{\nu}}}\ ,\ \hat{\theta}\right), \tag{5}$$

where $F_{p(2,2m-2)}$ is the percentile of F distribution with 2 and $2m-2$ degrees of freedom and right-tail probability p. Since the exact percentiles of the $F_{(2,2m-2)}$ distribution can be written in closed form, that is, $F_{p(2,2m-2)} = (m-1)(p^{-\frac{1}{m-1}}-1)$, equation (5) can be simplified.

4.3 Confidence Interval for x_p

In reliability analysis, we are not only interested in deriving inference about parameter but also interested in making inference about percentiles of the lifetime distribution. Let x_p be the $100p$-th percentile of the lifetime distribution. One can obtain x_p by solving $P(X \le x_p) = F(x_p) = p$, where $F(\cdot)$ is as given in (1). Then, it is easy to see that $x_p = \theta(1-p)^{-\frac{1}{\nu}}$. Consequently, the MLE of x_p is given by $\hat{x}_p = \hat{\theta}(1-p)^{-\frac{1}{\hat{\nu}}}$. A confidence interval for x_p can be obtained using the pivotal quantity $U_p = \hat{\nu}(\log\hat{\theta} - \log x_p)$. If $u_{p,1-\alpha}$ is the $(1-\alpha)$-th percentile of U_p, then

$$\begin{aligned} 1-\alpha &= P(U_p \le u_{p,1-\alpha}) \\ &= P(x_p \ge \hat{\theta}e^{-\frac{u_{p,1-\alpha}}{\hat{\nu}}}), \end{aligned} \tag{6}$$

and hence, $\hat{\theta}e^{-\frac{u_{p,1-\alpha}}{\hat{\nu}}}$ is a lower confidence limit for x_p with confidence level $1-\alpha$. For any observed set of data, one must find the value of $u_{p,1-\alpha}$ to get the confidence limit for x_p. This has been discussed by Engelhardt and Bain (1978).

5 The Expected Test Time

In practical applications, it is often useful to have an idea of the duration of a lifetime test. Therefore it is important to compute the expected time required to complete a life test. In the case of progressively Type II censored sampling plan with uniform removals, one can compute the expectation of the m-th order statistic.

Conditionally on $\boldsymbol{R} = \boldsymbol{r}$, Balakrishnan and Aggarwala (2000) showed that the expected value of X_m is

$$E(X_m|\boldsymbol{R}=\boldsymbol{r}) = \theta\prod_{i=1}^{m}\frac{\beta_i}{\beta_i-1}, \tag{7}$$

for $\beta_i > 1$, where $\beta_1 = \nu n$ and $\beta_i = \nu(n - r_1 - r_2 - \cdots - r_{i-1} - i + 1)$, $i = 2, 3, \ldots, m$. Then, the expected test time under progressive Type II censoring with random removals can be computed by taking expectations on both sides of (7) with respect to the $\boldsymbol{R}$. That is,

$$\begin{aligned} E(X_m) &= E_{\boldsymbol{R}}[E(X_m|\boldsymbol{R})] \\ &= \sum_{r_1=0}^{g(r_1)}\sum_{r_2=0}^{g(r_2)}\cdots\sum_{r_{m-1}=0}^{g(r_{m-1})} E(X_m|\boldsymbol{R})P(\boldsymbol{R}=\boldsymbol{r}), \end{aligned}$$

where $g(r_1) = n-m$, $g(r_i) = n-m-r_1-\cdots-r_{i-1}$, $i = 2, 3, \ldots, m-1$, and $P(\boldsymbol{R}=\boldsymbol{r})$ is defined in (4).

Furthermore, it should be mentioned here that the expected test time of a Type II censoring can be obtained by setting $r_1 = r_2 = \cdots = r_{m-1} = 0$ in (7). Equivalently, the expected test time of a complete sampling plan can be found by substituting $m = n$ and setting $r_1 = r_2 = \cdots = r_{m-1} = 0$ in (7).

Up to this point, we have derived the expected test times of progressively Type II censored sample with uniform removals, Type II censored sample, and complete sample. It will be of interest to compare these three expected times in order to gain some idea about the roles of n and m on the duration of the lifetime test. Since the $E(X_m)/\theta$ is independent of θ, the scale parameter θ can be ignored for the purpose of comparisons. It is obvious that analytically comparing these three expected test times is very difficult. An alternative is to calculate them numerically for various n, m and ν. Table 1 gives the values of $E(X_m)/\theta$ under progressive Type II censoring for different values of n, m, and ν. The exact expected test time can be obtained by multiplying the entries in Table 1 by θ. Several combinations of n and m are considered. For $n = 8, 10, 12$, and 15, the corresponding choices of m are listed in this table. The cases for $m = n$ correspond to the complete sample. Various values of shape parameter ν are studied. Other values of ν were also considered and the results are not reported here since they have a similar pattern to the cases listed in Table 1.

For given values of n and m, the expected test time decreases significantly as the value of ν increases. For a fixed value of m, the expected test time decreases as the sample size n increases. In addition, the effect of m on expected test time is strongly influenced by the value of ν. As ν increases, this effect becomes smaller. Table 1 also shows the ratio of expected test times for progressive Type II censoring and complete sampling plans. Note that the ratio does not depend on the scale parameter θ. For a fixed value of m, the ratio decreases as n increases, but the size of decrement decreases as ν increases. Moreover, for any value of n, the change of ratios is affected by the value of ν. For larger value of ν, the ratio goes to 1 quickly as m increases.

References

Arnold, B. C. and Press, S. J. (1989) " Bayesian estimation and prediction for Pareto data". *Journal of the American Statistical Association*, **84**, 1079-1084.

Balakrishnan, N. and Aggarwala, R. (2000) "*Progressive Censoring – Theory, Methods, and Applications*". Birkhäuser, Boston.

Engelhardt, M. and Bain, L. J. (1978) "Tolerance limits and confidence limits on reliability for the two-parameter exponential distribution". *Technometrics*, **20**, 37-39.

McNolty, F., Doyle, J. and Hansen, E. (1980) "Properties of the mixed exponential failure process". *Technometrics*, **22**, 555-565.

Ouyang, L.-Y. and Wu, S.-J. (1994) "Prediction intervals for an ordered observation from a Pareto distribution". *IEEE Transactions on Reliability*, **43**, 264-269.

Tse S.-K. and Yuen H.-K. (1998) "Expected experiment times for the Weibull distribution under progressive censoring with random removals". *Journal of Applied Statistics*, **25**, 75-83.

Vännman, K. (1976) "Estimators based on order statistics from a Pareto distribution". *Journal of the American Statistical Association*, **71**, 704-708.

Table 1: $E(X_m)/\theta$ under progressive Type II censoring and the ratio of expected test times between progressive Type II censoring and complete sampling plans

		ν							
n	m	1.2		1.5		1.8		2.1	
		$E(X_m)/\theta$	ratio	$E(X_m)/\theta$	ratio	$E(X_m)/\theta$	ratio	$E(X_m)/\theta$	ratio
8	4	9.8631	0.3105	4.4177	0.4065	3.0898	0.4809	2.4979	0.5394
	5	15.5477	0.4894	6.2291	0.5732	4.0733	0.6339	3.1477	0.6797
	6	21.3527	0.6722	7.9607	0.7326	4.9752	0.7743	3.7268	0.8048
	7	26.8007	0.8437	9.5077	0.8750	5.7560	0.8958	4.2173	0.9107
	8	31.7661	1.0000	10.8666	1.0000	6.4256	1.0000	4.6309	1.0000
10	5	13.5700	0.3553	5.6074	0.4459	3.7388	0.5156	2.9281	0.5703
	6	19.0661	0.4992	7.2875	0.5796	4.6277	0.6382	3.5051	0.6827
	7	24.3994	0.6389	8.8350	0.7026	5.4196	0.7474	4.0073	0.7805
	8	29.3603	0.7688	10.2177	0.8126	6.1089	0.8425	4.4366	0.8641
	9	33.9350	0.8886	11.4534	0.9109	6.7126	0.9257	4.8072	0.9363
	10	38.1900	1.0000	12.5742	1.0000	7.2512	1.0000	5.1341	1.0000
12	6	17.4694	0.3934	6.8049	0.4801	4.3742	0.5463	3.3414	0.5979
	7	22.6989	0.5112	8.3471	0.5890	5.1717	0.6459	3.8509	0.6891
	8	27.6449	0.6226	9.7451	0.6876	5.8751	0.7337	4.2918	0.7680
	9	32.2426	0.7261	11.0019	0.7763	6.4938	0.8110	4.6737	0.8364
	10	36.5264	0.8226	12.1416	0.8567	7.0451	0.8798	5.0098	0.8965
	11	40.5601	0.9134	13.1913	0.9307	7.5455	0.9423	5.3117	0.9505
	12	44.4040	1.0000	14.1728	1.0000	8.0076	1.0000	5.5882	1.0000
15	7	20.8592	0.3905	7.8079	0.4756	4.8941	0.5410	3.6741	0.5924
	8	25.7743	0.4825	9.2200	0.5617	5.6121	0.6204	4.1274	0.6655
	9	30.3938	0.5690	10.5003	0.6397	6.2481	0.6907	4.5226	0.7292
	10	34.7135	0.6499	11.6633	0.7105	6.8150	0.7534	4.8701	0.7853
	11	38.7775	0.7260	12.7316	0.7756	7.3277	0.8101	5.1809	0.8354
	12	42.6392	0.7983	13.7265	0.8362	7.7988	0.8622	5.4638	0.8810
	13	46.3444	0.8676	14.6646	0.8933	8.2380	0.9107	5.7255	0.9232
	14	49.9283	0.9347	15.5580	0.9478	8.6521	0.9565	5.9704	0.9627
	15	53.4157	1.0000	16.4154	1.0000	9.0458	1.0000	6.2018	1.0000