IN THE NAME OF GOD

Shiraz University
Faculty of Science

Ph.D. Dissertation In Mathematical Statistics

INFERENCE BASED ON PROGRESSUVE CENSORING DATA

By

SAFAR PARSI

Supervised by

MOJTABA GANJALI
NAHID SANJARI FARSIPOUR

February 2009

دانشگاه شیراز

دانشکده علوم

پایان‌نامه دکتری در رشته آمار ریاضی

استنباط بر اساس داده های سانسور فزاینده

توسط

صفر پارسی

اساتید راهنما:

دکتر مجتبی گنجعلی

دکتر ناهید سنجری فارسی‌پور

بهمن ماه ۱۳۸۷

IN THE NAME OF GOD

INFERENCE BASED ON PROGRESSIVE CENSORING DATA

BY

SAFAR PARSI

THESIS

SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (Ph.D.)

IN

MATHEMATICAL STATISTICS
SHIRAZ UNIVERSITY
SHIRAZ
ISLAMIC REPUBLIC OF IRAN

EVALUATED AND APPROVED BY THE THESIS COMMITTEE AS: EXCELLENT

N. SANJARI FARSIPOUR, Ph.D., PROF. Of STATISTICS (CHAIRMAN)

M. GANJALI, Ph.D., ASSOC. PROF. Of STATISTICS (CHAIRMAN)

J. BEHBOODIAN, Ph.D., PROF. Of STATISTICS

M. ASADI, Ph.D., ASSOC. PROF. Of STATISTICS

A.R. BORHANI HAGHIGHI, Ph.D., ASSIST. PROF. Of STATISTICS

M. TOWHIDI, Ph.D., ASSIST. PROF. Of STATISTICS

Feb 2009

Dedicated to my family and friends

Acknowledgement

There are many people I want to acknowledge. First, I would like to give a special thank you to my co-advisers, Dr. M. Ganjali and Prof. Dr. N. Sanjari Farsipour for their guidance and encouragement throughout the preparation of this thesis. Without their guidance and encouragement, this work cannot be done. I owe a great debt to them.

My sincere thanks also go to Prof. Dr. J. Behboodian at the Department of Statistics in Shiraz university. His support and encouragement are greatly appreciated.

I would like to thank my other committee members: Dr. Majid Asadi, Dr. A.R. Borhani Haghighi and Dr. M. Towhidi, for helpful discussion, reviewing and critiquing my work.

My appreciation goes to all the Professors who taught me in the program, from the Departments of Statistics in Shiraz university. I also owe a debt of gratitude to the faculty, staff, and fellow students in the Department of Statistics during my time at Shiraz.

Next, I would like to thank Prof. Dr. I. Bayramoglu of Economics University of Izmir, for his support and his helpful comments about main part of this thesis during my visit in Turkey. Also, my thanks go to all friends in Economics University of Izmir.

I express my gratitude to the Mohaghegh Ardabili University and to the Ministry of Sciences, Researches and Technology of Islamic Republic of Iran for their financial support.

Finally, and foremost, I appreciate of all my friends for their support and friendship.

ABSTRACT

INFERENCE BASED ON PROGRESSUVE CENSORING DATA

BY

SAFAR PARSI

An increasing interest on inference based on progressively censored data have been recently observed in statistical literature. Numerous research, review papers and a monograph devoted to this subject following the early work of Cohen (1963), where the first description of the model was made, are appeared. The model of progressive Type-II right censoring is of importance in the field of reliability and life testing. In this thesis, we provide an overview of various developments that have been taken place in inferential procedures based on progressively censored samples and also discuss some potential problems relating to this case. We consider the progressive Type-II right censored sample from Pareto distribution and introduce a new approach for constructing the simultaneous confidence interval of the unknown parameters of this distribution under progressive Type-II censoring.

We also deal with a model where joint Type-II progressive censoring is implemented on two samples from different populations in a combined manner. In this study, we obtain the conditional maximum likelihood estimators of the two Weibull parameters under this scheme. Moreover, simultaneous approximate confidence region based on the asymptotic normality of the maximum likelihood estimators are also discussed and compared with two Bootstrap confidence regions.

The behavior of the probability of failure structure, with different schemes, is also studied.

In practical applications, an experimenter may need to know the expected values of the number of failures for each population. This information is important for an experimenter when choosing an appropriate sampling plan, because, in statistical inference for parameters, the number of failures for a population is directly related to the efficiency of estimators. In this study, the formula allowing numerical computation of the expected value of the number of failures for two populations is given. Also, a detailed numerical study of this expected value is carried out for different parametric families of distributions.

Table of Contents

Content Page

List of Tables

List of Figures

Chapter 1

Progressive censoring methodology

In this chapter we will focus on the progressive Type-II right censoring situation and will present an overview of various developments relating to this case. For the sake of completeness, we will also present a brief discussion on the Type-I, Type-II and Type I progressive interval censoring schemes and some related works. In Monte Carlo studies relating to many statistical problems, one may wish to generate Type-II progressive censoring data from some continuous population with cumulative distribution function $F(x)$. For this, the algorithm described in Balakrishnan and Sandhu (1995), to generate progressively Type-II censored samples for different sizes and schemes, will be also explained.

1.1 Introduction

The importance of product reliability is greater than ever at the present time. As more and more products are introduced to the market, consumers now have the luxury of demanding high quality and long life in the products they purchase. In such a highly demanding and competitive market, one way by which manufactures (of computers, automobiles, and electronic items, for examples) attract consumers to their products is by providing warranties on product life-times. In order to design a cost-effective warranty, a manufacture must have sound knowledge about product failure-time distributions. To gain this knowledge, life-testing and reliability experiments are carried out before (and while) products are put on the market. Of course, the information gained through life-testing experiments is also used for other purposes in addition to determining effective warranties; for example, in pharmaceutical applications, the lifetimes of drugs may be studied in order to determine appropriate dosage administration and expiry dates. Furthermore, continuous improvement of products become essential and even critical in a competitive market. Life-testing experimentation is one way by which product improvement and product quality can be gauged.

There are several types of life testing experiments. In this chapter, we provide an overview of various sampling mechanisms in reliability experiments. In Sections 1.2, 1.3 and 1.4, we discuss briefly Type-I, Type-II and Type I progressive interval censoring, respectively. A versatile censoring method known as Progressive censoring, which may be employed in life-testing and reliability experimentation, is discussed in Section 1.5. In Section 1.6, the algorithm described in Balakrishnan and Sandhu (1995), to generate progressively Type-II censored samples for different sizes and schemes is given.

1.2 Type I censoring

Suppose that lifetimes for individuals in some population follow a distribution with probability density function (p.d.f.) $f(x)$ and cumulative distribution function (c.d.f.) $F(x)$. A Type-I censoring mechanism is said to apply when each individual has a fixed potential censoring time $C_i > 0$ such that X_i is observed if $X_i < C_i$; otherwise, we know only that $X_i > C_i$. Type-I censoring often arises when a study is conducted over a specified time period.

In our general notation, for Type I censoring, we have

$$x_i = min(X_i, C_i), \quad \delta_i = I(X_i \leq C_i). \tag{1.2.1}$$

The likelihood function for a Type-I censored sample is based on the probability distribution of $(x_i, \delta_i), i = 1, ..., n$. Both x_i and δ_i are random variables in (1.2.1), and their joint p.d.f., is

$$f(x_i)^{\delta_i} Pr(X_i > C_i)^{1-\delta_i}. \tag{1.2.2}$$

To see this, note that the C_i are fixed constants and that x_i can take on the values $\leq C_i$, with

$$Pr(x_i = C_i, \delta_i = 0) = Pr(X_i > C_i)$$

$$Pr(x_i, \delta_i = 1) = f(x_i) \qquad x_i \leq C_i$$

where Pr in the second expression denotes either a p.d.f. or a probability mass function according to whether the distribution of X_i is continuous or discrete at x_i. Assuming that the lifetimes $X_1, \ldots, X_n$ are statistically independent, we obtain the likelihood function from (1.2.1) as,

$$L = \prod_{i=1}^{n} f(x_i)^{\delta_i} \left[1 - F(x_i^+)\right]^{1-\delta_i}. \quad (1.2.3)$$

The term $1 - F(x_i^+)$ appears in (1.2.3) since it equals to $Pr(X_i > x_i)$ in general; if $F(x)$ is continuous at x_i, then $F(x_i^+) = F(x_i)$.

Example 1.2.1 Suppose that the lifetimes X_i for $i = 1, ..., n$ are independent and follow an exponential distribution with p.d.f. $f(x) = \lambda exp(-\lambda x)$ and survivor function $S(x) = exp(-\lambda x)$, then (1.2.3) gives

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{n} \left(\lambda e^{-\lambda x}\right)^{\delta_i} \left(e^{-\lambda x}\right)^{1-\delta_i} \\ &= \lambda^r exp\left(-\lambda \sum_{i=1}^{n} x_i\right) \quad (1.2.4) \end{aligned}$$

where $r = \sum \delta_i$ is the observed number of uncensored lifetimes, or failures. The log-likelihood function $\ell = \log L(\lambda)$ is

$$\ell = r \log \lambda - \lambda \sum_{i=1}^{n} x_i. \quad (1.2.5)$$

The maximum likelihood estimator is given by solving $d\ell/d\lambda = 0$, and is $\widehat{\lambda} = r/\sum_{i=1}^{n} x_i$. The exact distribution of $\widehat{\lambda}$ is rather intractable, as is the distribution of the minimal sufficient statistic $(r, \sum x_i)$.

1.3 Type-II censoring

The term Type-II censoring refers to the situation where only the r smallest lifetimes $x_{(1)} \leq \cdots \leq x_{(r)}$ in a random sample of size n are observed; here r is a specified integer between 1 and n. This censoring scheme arises when n individuals start on study at the same time, with the study terminating once r failures (or lifetimes) have been observed. Although some life tests are formulated with Type-II censoring, they have the practical disadvantage that the total time $x_{(r)}$ that the test will run is random and hence unknown at the start of the test. Type-I censoring is therefore much more common in planned experiments. The exact sampling properties of statistical procedures based on a Type-II censored sample are, however, tractable in many cases and this censoring scheme is often discussed in theoretical work.

With Type-II censoring the value of r is chosen before the data are collected, and the data consist of the r smallest lifetimes in a random sample $X_1 \dots X_n$. For continuous distributions we can ignore the possibility of ties and denote the r smallest lifetimes as $X_{(1)} < X_{(2)} < \dots < X_{(r)}$. If the X_i has p.d.f. $f(x)$ and distribution function $F(x)$, then from general results of order statistics the joint p.d.f. of $X_{(1)}, X_{(2)}, \dots, X_{(r)}$ is

$$\frac{n!}{(n-r)!}[1-F(x_{(r)})]^{n-r}\{\prod_{i=1}^{r} f(x_{(i)})\} \ , \qquad x_{(1)} < x_{(2)} < \dots < x_{(r)}. \quad (1.3.1)$$

The likelihood function is based on (1.3.1). By dropping the constant $n!/(n-r)!$ and noting that in terms of the (δ_i, x_i) notation we have $\delta_i = 0$ and $x_i = x_{(r)}$ for those individuals whose lifetimes are censored, we see that (1.3.1) gives a likelihood of the same form (1.2.3) as for Type-I censoring. The sampling properties are, however, different in finite samples.

Example 1.3.1 Consider the exponential distribution as in Example 1.2.1, but suppose lifetimes are Type-II censored. The log-likelihood is still of the form (1.2.5), but here it can be written as

$$rlog\lambda - \lambda \left\{ \sum_{i=1}^{r} x_{(i)} + (n-r)x_{(r)} \right\} \tag{1.3.2}$$

and the maximum likelihood estimate for λ can be written as $\widehat{\lambda} = r/W$, where

$$W = \sum_{i=1}^{r} x_{(i)} + (n-r)x_{(r)}.$$

Sincc r is fixed, the statistic W is sufficient for λ, and it is readily shown that with the data considered as random variables, $2\lambda W = 2r\lambda/\widehat{\lambda} \sim \chi^2_{2r}$, a chi-squared distribution with $2r$ degrees of freedom. This allows exact confidence intervals and tests for λ to be developed.

1.4 Type I Progressive Interval Censoring

Progressively Type I interval censoring is a union of Type I interval censoring and progressive censoring. A progressively Type I interval censored sample is collected as follows: n units are put on life test at time $T_0 = 0$. Units are observed at pre-set times $T_1, T_2, \ldots, T_m$ (m is also fixed). At these times, $r_1, r_2, \ldots, r_m$ live units are removed from experimentation, respectively. The values $r_1, r_2, \ldots, r_m$ may be prespecified as percentages of the remaining live units or, alternatively, r_i units available for removal. In this case, the number of live units removed at time T_i is $r_i = min(r_i$, number of units remaining$), i = 1, 2, \ldots, m-1$. Again r_m equals all remaining units at time T_m, when experimentation is scheduled to terminate. Suppose a progressively Type-I interval censored sample is collected as described above, beginning with a random sample of n units with a continuous lifetime distribution $F(x, \theta)$ and let $k_1, k_2, \ldots, k_m$ denote the number of units known to have failed in the intervals, $(0,T_1], (T_1,T_2], \ldots, (T_{m-1},T_m]$, respectively. Then, based on this observed data, the joint likelihood function will be

$$L(\theta;\mathrm{X}) = C\prod_{i=1}^{m}[F(T_i;\theta) - F(T_{i-1};\theta)]^{k_i}[1 - F(T_i;\theta)]^{r_i}$$

where C is constant.

1.5 Progressive Type-II right censoring

There are many scenarios in life-testing and reliability experiments in which units are lost or removed from experimentation before failure. The loss may occur unintentionally, or it may have been designed so in the study. Unintentional loss may occur, for example, in the case of accidental breakage of an experimental unit, or if an individual under study drops out, or if the experimentation itself must cease to some unforseen circumstances such as depletion of funds, unavailability of testing facilities, etc. More often, however, the removal of units from experimentation is pre-planned and intentional, and is done so in order to free up testing facilities for other experimentation, to save time and cost, or to exploit the straightforward analysis that often results. In some cases, when there are live units on test intentional removal of times or termination of the experiment may be due to ethical considerations. Specifically, consider the situation in which n identical units, with lifetime c.d.f. $F(x)$ and survivor function $S(x)$ and p.d.f. $f(x)$, are placed on a life-testing experiment. Then, immediately following the first failure, R_1 surviving units are removed from the test at random; next, immediately following the second observed failure, R_2 surviving units are removed from the test at random, and so on; finally, at the time of the mth observed failure, all the remaining $R_m = n - R_1 - \ldots - R_{m-1} - m$ surviving units are removed from the test. In this setup, the number of complete failures to be observed, m, and the progressive censoring scheme, $\mathbf{R} = (R_1, \ldots, R_m)$, to be carried out are all assumed to be pre-fixed. As the censoring times are all random here and the numbers of items to fail before each censoring time are all fixed, this scheme is said to be progressive Type-II right censoring, and the ordered values obtained as a consequence of this type of censoring are referred to as progressively Type-II right censored order statistics. These are denoted by $X^{\mathbf{R}}_{1:m:n} < X^{\mathbf{R}}_{2:m:n} < \ldots < X^{\mathbf{R}}_{m:m:n}$, and sometimes simply by $X_{1:m:n} < X_{2:m:n} < \ldots < X_{m:m:n}$ if there is no confusion in the context of the discussion. A book-length account of the developments on progressive Type-II censoring is due to Balakrishnan and Aggarwala (2000). This censoring scheme may be depicted pictorially as Figure 1.5.1.

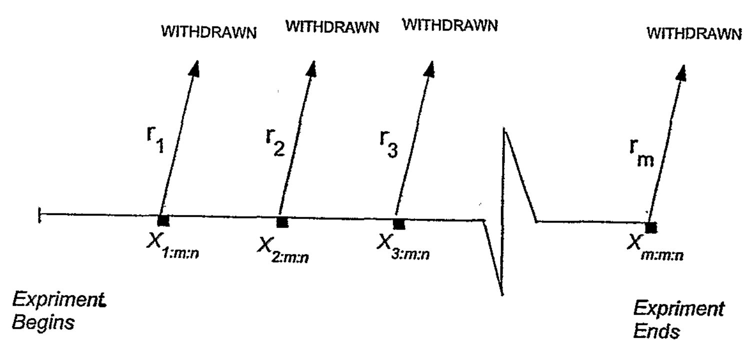

Figure 1.5.1: Progressive Type-II censoring

Then, the joint density of the progressively censored order statistics, $X^{\mathbf{R}}_{1:m:n}, X^{\mathbf{R}}_{2:m:n}, \ldots, X^{\mathbf{R}}_{m:m:n}$ is given by Balakrishnan and Aggarwalla (2000) as

$$f_{X^{\mathbf{R}}_{1:m:n} X^{\mathbf{R}}_{2:m:n} \ldots X^{\mathbf{R}}_{m:m:n}}(x_1, x_2, \ldots, x_m) = C \prod_{i=1}^{m} f(x_i)[1 - F(x_i)]^{R_i} \quad ,$$