On the reliability of Type II censored reliability analyses.

Chua, See Ju

On the Reliability of Type II Censored Reliability Analyses

by

See Ju Chua, BSc (Hons), University of Wales Swansea

Thesis submitted to the Swansea University

in candidature for the degree of

PHILOSOPHIÆ DOCTOR

School of Business and Economics
Swansea University
Singleton Park, Swansea SA2 8PP
United Kingdom

May 2009

Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

SEE JU CHUA
1 May 2009

Statement

This thesis is the result of my own investigation, except where acknowledgement of other sources is given.

SEE JU CHUA
1 May 2009

Statement

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan (subject to the law of copyright), and for the title and summary to be made available to outside organisations.

SEE JU CHUA
1 May 2009

Acknowledgements

I am deeply indebted to my supervisor Dr. Alan Watkins, without whose constant support, patience and guidance, this thesis would never have been possible. Thank you so much for everything you have done over the seven years, especially for the encouragement and inspiration to commence this thesis.

I am also grateful for the financial support from the ORS, and the School of Business and Economics. Moreover, I would like to thank my colleagues and friends, Jen Ning Tan and Hannah Finselbach, for all their help and stimulating suggestions in my research work. I have furthermore to thank everyone at the School of Business and Economics, for providing an environment that has been a pleasure to work in.

I owe my deepest gratitude to my mum and dad; I am deeply sorry for the time we spent apart. Finally, I would like to give my special thanks to my family, relatives and god-mother, who have always supported, encouraged and believed in me, in all my endeavours, throughout my life. Thank you each and every one.

SEE JU CHUA

Swansea University
May 2009

This thesis was typeset with LaTeX 2_ε by the author. LaTeX 2_ε is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The macros used in formatting this thesis were written by Anton Merlushkin of the European Business Management School, Swansea University.

Summary

This thesis considers the analysis of reliability data subject to censoring, and, in particular, the extent to which an interim analysis - here, using information based on Type II censoring - provides a guide to the final analysis. Under a Type II censored sampling, a random sample of n units is put on test simultaneously, and the test is terminated as soon as r ($1 \leq r \leq n$, although we are usually interested in $r < n$) failures are observed. In the case where all test units were observed to fail ($r = n$), the sample is complete. From a statistical perspective, the analysis of the complete sample is to be preferred, but, in practice, censoring is often necessary; such sampling plan can save money and time, since it could take a very long time for all units to fail in some instances. From a practical perspective, an experimenter may be interested to know the smallest number of failures at which the experiment can be reasonably or safely terminated with the interim analysis still providing a close and reliable guide to the analysis of the final, complete data. In this thesis, we aim to gain more insight into the roles of censoring number r and sample size n under this sampling plan.

Our approach requires a method to measure the precision of a Type II censored estimate, calculated at censoring level r, in estimating the complete estimate, and hence the study of the relationship between interim and final estimates. For simplicity, we assume that the lifetimes follow the exponential distribution, and then adopt the methods to the two-parameter Weibull and Burr Type XII distributions, both are widely used in reliability modelling. We start by presenting some mathematical and computational methodology for estimating model parameters and percentile functions, by the method of maximum likelihood. Expressions for the asymptotic variances and covariances of the estimators are given. In practice, some indication of the likely accuracy of these estimates is often desired; the theory of asymptotic Normality of maximum likelihood estimator is convenient, however, we consider the use of relative likelihood contour plots to obtain approximate confidence regions of parameters in relatively small samples.

Finally, we provide formulae of the correlations between the interim and final maximum likelihood estimators of model parameters and a particular percentile function, and discuss some practical implications of our work, based on the results obtained from published data and simulation experiments.

To my grandparents

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The term reliability usually refers to the probability that a piece of equipment, or a component of a larger system, will operate satisfactorily either at any particular instant at which it is required or for a certain length of time. Like survival analysis in medical studies, or duration analysis in economics, the quantity of interest in reliability analysis is the lifetime (also called the survival time, failure time, or time to failure) of a specimen; for instance, the lifetime of an electrical component (Epstein, 1960; Wingo, 1993), or the time to failure of a deep-groove ball bearing (Lieblein & Zelen, 1956), or the relief time of an arthritic patient after a fixed dosage of medication (Wingo, 1983). The methods for statistical analysis of data on reliability are widely discussed, with many textbooks covering solely this area; see, for instance, Mann *et al.* (1974), Lawless (1982), Nelson (1982), Bain & Engelhardt (1991) and Crowder *et al.* (1991).

The only way to measure reliability is to test specimens, under conditions that simulate real life, until failure occurs. Extensive testing, however, often results in undesirable expenditures of time and money. An important concept that arises naturally in this area is that of censored sampling plan; for example, not all electrical components may have failed at the close of a life test, and some arthritic patients may have left the clinical trial for unrelated reasons before completion. Such incomplete observation of the lifetime of a specimen is called censoring. Furthermore, in reality, it is not always feasible to examine all system requirements in reliability testing; some systems are prohibitively expensive to test, some failure modes may take years to observe, and some experiments may be hazardous to run over prolonged periods. In such cases, perhaps the most commonly used technique is to terminate the test after a certain number of failures, r, has been observed out of a sample of n test units; this gives rise to Type II censoring. The data observed thus consists of order statistics, and, formally, is said to be right censored. Other censoring regimes are possible (see Section 1.4) - for instance, Type I, left censoring, progressive censoring - but, for convenience only right censoring is discussed in any detail here. A comprehensive text on the subject is Cohen (1991).

Other life test plans pose different problems. Sequential plans are "accept-reject" tests under a given null hypothesis, H_o, versus an alternative hypothesis, H_1. The life test is continuously monitored and a decision made as soon as there is sufficient supporting evidence for one of the two hypotheses. These tests take less time than non-sequential plans but estimation is complicated and not very robust. The history and statistical theory of sequential test plans are illustrated well in Gosh & Sen (1991). Accelerated life testing concerns the collection of lifetime data more quickly than would be the case in the normal use of components. Often, in order to induce failure in a short time, it may be necessary to increase the severity of a condition such as temperature, load or vibration. The results of any of these tests have to be extrapolated back to the conditions of normal use, and care is needed in choosing the model on which to base this. The execution and analysis of accelerated life tests is in general a complex area. A comprehensive text on the subject is Nelson (1990), while Nelson (2005a,b) publishes an extensive bibliography of statistical plans on accelerated testing and test plans. In reality, various factors influence the choice of test plans, usually in relation to resources. These may be physical, time-related of financial.

This thesis considers some particular aspects of Type II censored reliability analysis. Suppose it is possible to conduct one or more interim analyses ($r < n$) in addition to a final analysis ($r = n$). For example, it may be possible to make inferences on model parameters at each of a sequence $r = r_1, r_2, \ldots < n$ of failures, until all items have failed and the data set is complete. In real life scenarios, we may also draw inferences about the percentile of a lifetime distribution, as a practitioner will typically wish to know the time at which a specified percentage of test units fails, either for monitoring purposes or to implement changes to the test at that time. In this case, we may consider the extent to which the final estimates of parameters are consistent with earlier estimates, or the rate at which interim estimates converge on their final values; more generally, we can consider the precision with which we can make statements on final estimates, based on interim estimates, as represented by the confidence limits for the final estimates given the interim estimates. This approach, of course, requires an evaluation of the relationship between final and interim results, and hence the extent to which an interim analysis - here, using information based on Type II censoring - provides a guide to the final analysis; this is the scenario outlined in Chua & Watkins (2007) and Chua & Watkins (2008a,b), and explains the title of our thesis. Some discussion on the corresponding analysis of reliability data under Type I censoring is given by Finselbach & Watkins (2006), Peng & MacKenzie (2007) and Finselbach (2007).

Our approach implicitly introduces the following question: as r is to be specified before testing commences, what is the smallest number of failures at which the experiment can be reasonably or safely terminated with the interim estimates still yielding close and reliable guides to the final estimates? This information is important for an experimenter, as he or she can then choose an acceptable censoring number and sample size, with the (expected) time required to complete a test generally directly linked to its cost. If the initial cost of test units is cheap compared to experiment time, he or she can increase the initial sample size to obtain results economically.

To address this question, we proceed on the basis of a parametric modelling of data, and assume that we have identified a distribution for the data, so that it remains to estimate the parameters and related quantities of that distribution. Statistical inference has been widely discussed from the classical, or frequentist, point of view. That is, estimators and test statistics are assessed by criteria relating to their performance in repeated sampling; see, for instance, Lawless (1982), Bain & Engelhardt (1991) and Cohen (1991), based on both complete and censored samples. On the other hand, in the Bayesian approach, direct probability statements are made about unknown quantities, conditional on the observed data. This necessitates the introduction of prior beliefs into the inference process. At the present time there is lively debate over the place of Bayesian statistics in reliability theory. Whether Bayesian statistics will eventually supplant classical statistics, as its more vigorous proponents have been proclaiming for the past fifty years, is something still to be seen, but reliability engineers certainly should have an awareness of the Bayesian approach. Dey & Rao (2005) provides a general overview of the area of Bayesian Thinking and describes what the current state is in the context of Bayesian theory, methodology, modelling, computation and applications.

In this thesis the classical approach to the statistical inference of reliability data is considered. Although there is much literature on the method of maximum likelihood estimation, authors like Nelson (1982) and Wingo (1993) have mentioned that exact mathematical expressions for the asymptotic variances and covariances of the maximum likelihood estimates are difficult to obtain. This may be regarded as a convenient starting point for our study, in which we attempt to derive analytical formulae for these variances and covariances. In addition, perhaps due to convenience, asymptotic theory of maximum likelihood is widely used to obtain approximate confidence limits for the maximum likelihood estimates. Such limits are essentially asymptotic ones, while real-life samples are, because of time or budget constraints, often of small to moderate sizes. Thus, we need to be able to establish confidence regions of estimates in relatively small samples subject to Type II censoring, using a more suitable and reliable statistical method; some work on this topic is presented in Chua *et al.* (2007).

In the framework outlined above, we obtain two sets of estimates, interim and final, of model parameters and a particular percentile, but we are also interested at the distributions of these quantities. We focus on conditional distributions of final estimators given interim counterparts; if these are Normal - as is the case asymptotically - then, in turn, we require the covariance between final and interim quantities. The classical asymptotic approach uses the relationship between the maximum likelihood estimators, the expected Fisher information matrix and the score vector.

For simplicity, we start by considering singly censored samples under Type II censoring. We start from the assumption that the lifetimes follow the exponential distribution, chiefly to exploit the familiar and extremely powerful lack-of-memory property of this distribution.

1.2	2.2	4.9	5.0	6.8	7.0	12.1	13.7	15.1	15.2
23.9	24.3	25.1	35.8	38.9	47.9	48.4	49.3	53.2	55.6
62.7	72.4	73.6	76.8	83.8	95.1	97.9	99.6	102.8	108.5
128.7	133.6	144.1	147.6	150.6	151.6	152.6	164.2	166.8	178.6
185.2	187.1	203.0	204.3	229.5	253.1	304.1	341.7	354.4	

Table 1.1: Failure times for 49 items placed on a life test; from Epstein (1960).

A natural extension of the exponential distribution is the Weibull distribution; the latter can model lifetimes with increasing, constant, or decreasing failure rate. The more flexible Burr Type XII distribution has in recent years assumed a position of some importance in the field of reliability and life testing. Unfortunately, the estimation of its parameters is not always straightforward.

Throughout, we illustrate the results using published data sets, and also validate asymptotic results for various combinations of sample size, censoring number and values of model parameters using extensive simulation experiments. We first outline some examples of lifetime data, then give some mathematical background, and finally introduce some key definitions occurring in the analysis of reliability data.

1.1 Some Examples of Reliability Data

We introduce some published data to illustrate various typical ways in which lifetime data arise, and also use these as the basis of worked examples to illustrate ideas and concepts.

1.1.1 Epstein's Failure Times Data

Manufactured items such as mechanical or electrical components are often placed on life tests in order to obtain information on their endurance. Table 1.1 presents failure times data on $n = 49$ items put on a life test, run until all items failed. This data set may be modelled, as in Epstein (1960), by the exponential distribution. Figure 1.1 shows the exponential P-P plot for these data, where a sample from an exponential distribution should form approximately a straight line. Departures from this straight line indicate departures from the exponential distribution. Hence, the linear plot in Figure 1.1 suggests that it is appropriate to model the Epstein's failure times data with the exponential distribution.

1.1.2 Ball Bearings Data

A second example is data arising in tests on the endurance for deep-groove ball bearings, given in Table 1.2. They were originally discussed by Lieblein & Zelen (1956), who assumed that the data follows a Weibull distribution. As shown in Figure 1.2, we see the Weibull P-P plot for these data deviates from the straight line in the middle but fits the line well at both ends. We also note that a review on this data set by Caroni (2002) points out

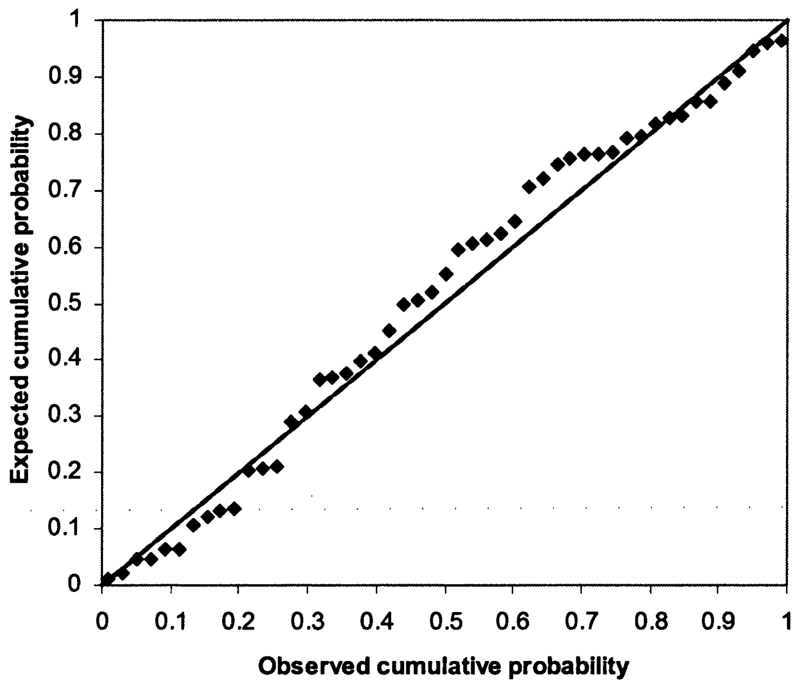

Figure 1.1: P-P plot for Epstein's failure times data based on exponential with $\widehat{\theta} = 104.8898$.

17.88	28.92	33.00	41.52	42.12	45.60	48.48	51.84	51.96	54.12
55.56	67.80	68.64	68.64	68.88	84.12	93.12	98.64	105.12	105.84
127.92	128.04	173.40							

Table 1.2: Lifetimes (in millions of revolutions) for 23 deep-groove ball bearings; based on Lieblein & Zelen (1956).

that they have been quoted incorrectly from Lieblein & Zelen (1956), firstly by Thoman *et al.* (1969), and subsequently by numerous authors such as Kalbfleisch (1979) and Lawless (1982). Nonetheless, like much of the later literature, we regard the version in Table 1.2 as a set of uncensored failure times, and assume that it can be modelled by the Weibull distribution.

1.1.3 Arthritic Patients Data

The ball bearings example is not the only well known lifetime data set in which the original data values have been changed. Table 1.3 shows data resulting from a clinical trial which was undertaken to test the efficacy of an analgesic, taken from Wingo (1983). This data represents relief times (in hours) of $n = 50$ arthritic patients receiving a fixed dosage of this medication, and, as indicated by the linear pattern in the P-P plot at Figure 1.3, is assumed to follow a Burr Type XII distribution. Watkins (1996) remarks that the last four

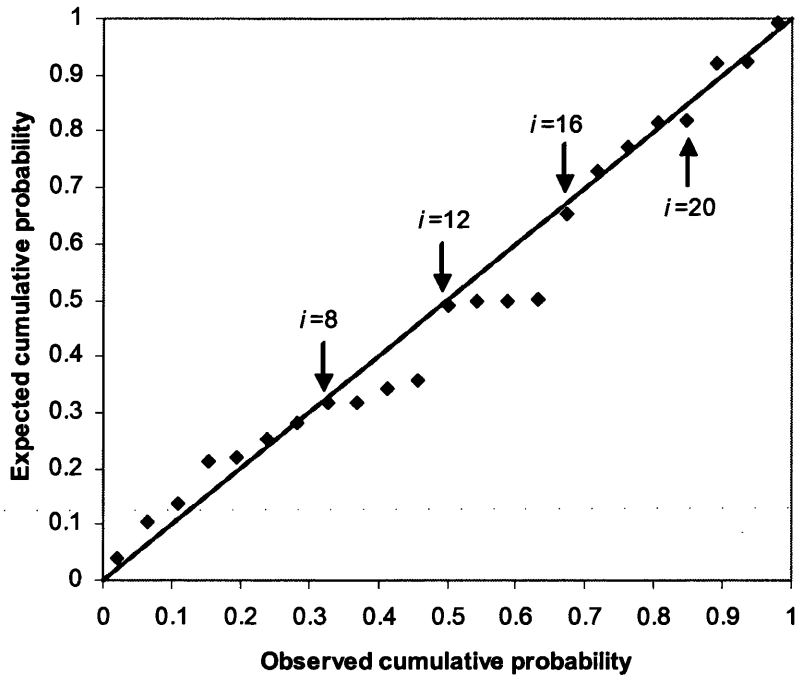

Figure 1.2: P-P plot for ball bearings data based on Weibull with $\widehat{\theta} = 81.8783, \widehat{\beta} = 2.1021$.

0.29	0.29	0.34	0.35	0.36	0.36	0.44	0.46	0.49	0.49
0.50	0.50	0.52	0.53	0.54	0.55	0.55	0.55	0.56	0.57
0.58	0.58	0.59	0.59	0.60	0.60	0.61	0.61	0.62	0.64
0.68	0.70	0.70	0.70	0.71	0.71	0.71	0.72	0.72	0.73
0.75	0.75	0.80	0.80	0.81	0.82	0.84	0.84	0.85	0.87

Table 1.3: Relief times (in hours) for 50 arthritic patients; from Wingo (1983).

places of the fourth column in Wingo (1983), namely $0.72, 0.53, 0.70, 0.58$, have been given as $0.36, 0.46, 0.34, 0.44$ in Wang *et al.* (1996).

1.1.4 Electronic Components Data

Wingo (1993) reports on a life test experiment conducted to assess the reliability of a certain electrical component, where $n = 30$ components were involved. However, for reasons of cost, the trial was terminated after the $r = 20^{th}$ component failed. Table 1.4 gives the failure times for these 20 components and 10 censored values (hereafter denoted by †), which, again, may be modelled by the Burr Type XII distribution.

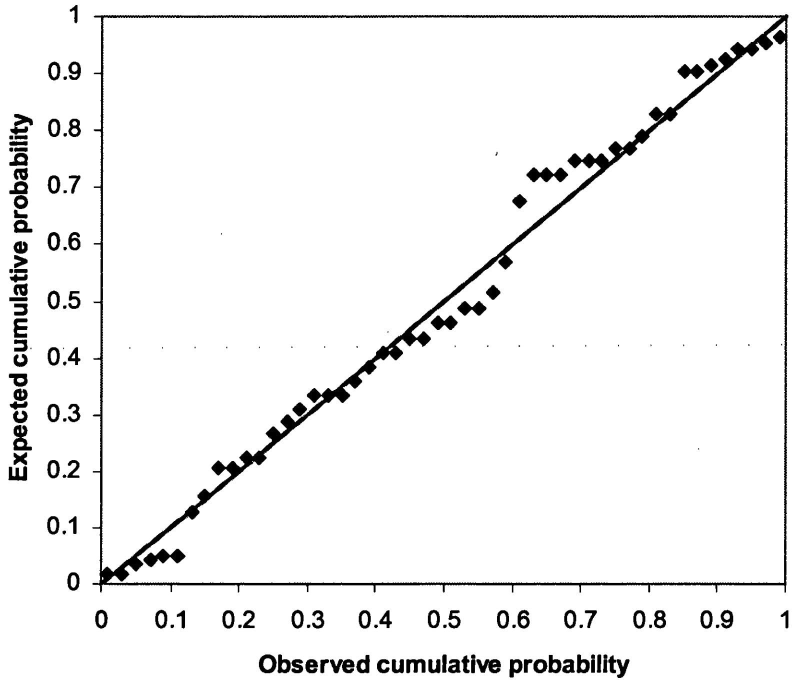

Figure 1.3: P-P plot for arthritic patients data based on Burr Type XII with $\widehat{\alpha} = 8.2681, \widehat{\tau} = 5.0006$.

0.1	0.1	0.2	0.3	0.4	0.5	0.5	0.6	0.7	0.8
0.9	0.9	1.2	1.6	1.8	2.3	2.5	2.6	2.9	3.1
$3.1^{\dagger}$	$3.1^{\dagger}$	$3.1^{\dagger}$	$3.1^{\dagger}$	$3.1^{\dagger}$	$3.1^{\dagger}$	$3.1^{\dagger}$	$3.1^{\dagger}$	$3.1^{\dagger}$	$3.1^{\dagger}$

Table 1.4: Failure times (in months) for 30 electronic components; from Wingo (1993); censored values are denoted by $\dagger$.

1.2 Mathematical Functions

1.2.1 Glossary of Functions and Notations

Table 1.5 summarises some standard mathematical functions and notations required throughout this thesis; we have adhered to the notation in Abramowitz & Stegun (1972). We list conventions like $\ln(x)$ here, where $\ln(x) \equiv \log_e(x)$ denotes the natural logarithm of the positive quantity x. Some specific notation is considered in more detail in Appendix A.

1.2.2 Useful Mathematical Properties

For later convenience, we denote the partial derivatives of an arbitrary function g with respect to (from now on, abbreviated to wrt) a as

$$g_a^k = \frac{\partial^k}{\partial a^k} g,$$

and, if g is univariate, the above reduces to

$$g^k = \frac{d^k}{da^k} g,$$

for $k = 1, 2, 3 \cdots$.

Gamma and Related Functions

The gamma function $\Gamma(a)$ satisfies the recurrence relation

$$\Gamma(1+a) = a\Gamma(a) = a! \tag{1.1}$$

for integer a. Its first and second derivatives with respect to wrt a are given by

$$\Gamma'(a) = \Gamma(a)\,\psi(a) \tag{1.2}$$

and

$$\Gamma''(a) = \Gamma(a)\left\{[\psi(a)]^2 + \psi'(a)\right\} \tag{1.3}$$

respectively. The psi or digamma function $\psi(a)$ satisfies the recursive relation given by

$$\psi(a+1) = \psi(a) + \frac{1}{a} \tag{1.4}$$

and has special values of

$$\begin{aligned} \psi(1) &= -\gamma, \\ \psi(a) &= -\gamma + \sum_{m=1}^{a-1} m^{-1}, \end{aligned} \tag{1.5}$$

$(a)_m$	Pochhammer's symbol	$\frac{\Gamma(a+m)}{\Gamma(a)}$
arg	argument	
$B(z,w)$	beta function	$\int_0^1 t^{z-1}(1-t)^{w-1}\,dt$ $= \int_0^\infty t^{z-1}(1+t)^{-(z+w)}\,dt$
$B_x(z,w)$	incomplete beta function	$\int_0^x t^{z-1}(1-t)^{w-1}\,dt$
cos, sin	cosine, sine function	
$e_n(z)$	truncated exponential	$\sum_{m=0}^{n} \frac{z^m}{m!}$
$\exp(z) = e^z$	exponential function	
$E_1(z)$	exponential integral	$\int_z^\infty \frac{e^{-t}}{t}dt$
$\mathrm{Ei}(z)$	exponential integral	$-\int_{-z}^\infty \frac{e^{-t}}{t}dt$
$F_{2,1}(a,b;c;z)$	hypergeometric function	$\sum_{m=0}^{\infty} \frac{(a)_m(b)_m}{(c)_m}\frac{z^m}{m!}$
$F_{3,2}(a,b,c;e,f;z)$	hypergeometric function	$\sum_{m=0}^{\infty} \frac{(a)_m(b)_m(c)_m}{(e)_m(f)_m}\frac{z^m}{m!}$
$F_{A,B}(a_\alpha;b_\beta;z)$	generalised hypergeometric function	$\sum_{m=0}^{\infty} \frac{(a_\alpha)_m}{(b_\beta)_m}\frac{z^m}{m!}$, where $a_\alpha = a_1, a_2, \ldots, a_A$ $b_\beta = b_1, b_2, \ldots, b_B$
$Li_p(z)$	polylogarithm function	$\sum_{m=1}^{\infty} \frac{z^m}{m^p}$
lim	limit	
ln	natural logarithm	$\log_e$
max	maximum	
min	minimum	
Pr	probability	
$\mathbb{R}$	real part	
γ	Euler's constant	$+0.5772156649\cdots$
$\gamma(a,x)$	normalised incomplete gamma function	$\int_0^x e^{-t}t^{a-1}dt$
$\Gamma(a)$	gamma function	$\int_0^\infty e^{-t}t^{a-1}dt$
$\Gamma(a,x)$	incomplete gamma function	$\int_x^\infty e^{-t}t^{a-1}dt$
$\zeta(p)$	Riemann zeta function	$\sum_{m=0}^{\infty} \frac{1}{(m+1)^p}$
$\Phi(z,p,q)$	Lerch transcendent	$\sum_{m=0}^{\infty} \frac{z^m}{(m+q)^p}$
$\psi(a)$	psi (digamma) function	$\frac{d}{da}\ln\Gamma(a) = \frac{\Gamma'(a)}{\Gamma(a)}$
$\psi^k(a)$	polygamma function	$\frac{d^k}{da^k}\psi(a) = \frac{d^{k+1}}{da^{k+1}}\ln\Gamma(a)$
$\binom{m}{n}$	binomial coefficient	$\frac{m!}{n!(m-n)!} = \frac{\Gamma(m+1)}{\Gamma(n+1)\Gamma(m-n+1)}$
$\lvert z\rvert$	absolute value or modulus of z	

Table 1.5: Notation and definitions of standard functions.

a	$\Gamma(a)$	$\psi(a)$	$\psi'(a)$
1	1	$-\gamma$	$\frac{\pi^2}{6}$
2	1	$1-\gamma$	$\frac{\pi^2}{6}-1$
3	2	$\frac{3}{2}-\gamma$	$\frac{\pi^2}{6}-\frac{5}{4}$
4	6	$\frac{11}{6}-\gamma$	$\frac{\pi^2}{6}-\frac{49}{36}$

Table 1.6: Some special values for the gamma and related functions.

for integer $a \geq 2$, where $\gamma = +0.5772\cdots$ is Euler's constant. For the polygamma functions $\psi^k(a)$, the following recursive formulae hold; with $k = 1$

$$\psi'(a+1) = \psi'(a) - \frac{1}{a^2}, \tag{1.6}$$

and, more generally, for $k = 1, 2, 3, \cdots$,

$$\psi^k(a+1) = \psi^k(a) + (-1)^k k! a^{-k-1}.$$

We also have

$$\psi^k(1) = (-1)^{k+1} k! \zeta(k+1),$$

so that

$$\psi'(1) = \zeta(2) = \frac{\pi^2}{6}. \tag{1.7}$$

Using (1.5) and (1.7), Table 1.6 gives the values of gamma and related functions evaluated at some integers, found from (1.1), (1.4) and (1.6).

The normalised incomplete gamma function is linked to a series expansion via

$$\gamma(a,x) = x^a \sum_{m=0}^{\infty} \frac{(-x)^m}{m!(a+m)} \tag{1.8}$$

and to the incomplete gamma function $\Gamma(a,x)$ via

$$\gamma(a,x) = \Gamma(a) - \Gamma(a,x). \tag{1.9}$$

Beta and Incomplete Beta Functions

With z, w positive and real, we can write the complete beta function in terms of the gamma functions as

$$B(z,w) = B(w,z) = \frac{\Gamma(z)\Gamma(w)}{\Gamma(z+w)}. \tag{1.10}$$

The incomplete beta function $B_x(z,w)$ is related to hypergeometric function (see below) via the following (see (6.6.8) in Abramowitz & Stegun, 1972):

$$B_x(z,w) = z^{-1} x^z F_{2,1}(z, 1-w; z+1; x). \tag{1.11}$$

Hypergeometric Functions

The generalised hypergeometric function is defined as (a, b and z may be real or complex)

$$F_{p,q}(a_1, a_2, \ldots, a_A; b_1, b_2, \ldots, b_B; z) = \sum_{m=0}^{\infty} \frac{(a_1)_m (a_2)_m \cdots (a_A)_m}{(b_1)_m (b_2)_m \cdots (b_B)_m} \frac{z^m}{m!}$$

where $(x)_m$ is Pochhammer's symbol. Two specific cases frequently used in this thesis have $p = 2, q = 1$ and $p = 3, q = 2$, which then give, respectively,

$$\begin{aligned} F_{2,1}(a, b; c; z) &= F_{2,1}(b, a; c; z) \\ &= \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{m=0}^{\infty} \frac{\Gamma(a+m)\Gamma(b+m)}{\Gamma(c+m)} \frac{z^m}{m!} \end{aligned} \tag{1.12}$$

and

$$F_{3,2}(a, b, c; e, f; z) = \frac{\Gamma(e)\Gamma(f)}{\Gamma(a)\Gamma(b)\Gamma(c)} \sum_{m=0}^{\infty} \frac{\Gamma(a+m)\Gamma(b+m)\Gamma(c+m)}{\Gamma(e+m)\Gamma(f+m)} \frac{z^m}{m!},$$

where we now write terms in the summation explicitly in terms of the gamma functions. Note that, for convenience, we sometimes write $F_{2,1}(a, b; c; z) \equiv F_{2,1}(z)$ and $F_{3,2}(a, b, c; e, f; z) \equiv F_{3,2}(z)$. Abramowitz & Stegun (1972) provide numerous linear transformation formulae for the $F_{2,1}(a, b; c; z)$ function; two relevant ones are

$$F_{2,1}(a, b; c; z) = (1-z)^{-a} F_{2,1}\left(a, c-b; c; \frac{z}{z-1}\right)$$

and

$$F_{2,1}(a, b; c; z) = (1-z)^{-b} F_{2,1}\left(b, c-a; c; \frac{z}{z-1}\right), \tag{1.13}$$

given, respectively, by (15.3.4) and (15.3.5) therein.

In Sections 4.2.2 and 4.4.2, it will be necessary to check that the hypergeometric series is convergent for a given set of arguments and variables. Slater (1966) considers various convergence tests on $F_{2,1}(a, b; c; z)$. Briefly, a series (1.12) is convergent for all values of z, real or complex, such that

$$|z| < 1. \tag{1.14}$$

When $|z| = 1$ and $z = 1$, the series is convergent if $\mathbb{R}(c-a-b) > 0$, and divergent if $\mathbb{R}(c-a-b) \leq 0$. When $|z| = 1$ but $z \neq 1$, the series is absolutely convergent if $\mathbb{R}(c-a-b) > 0$, convergent but not absolutely so if $-1 < \mathbb{R}(c-a-b) \leq 0$, and divergent if $\mathbb{R}(c-a-b) < -1$. However, when $\mathbb{R}(c-a-b) = -1$ the series is convergent if $\mathbb{R}(a+b) > \mathbb{R}(ab)$, and divergent if $\mathbb{R}(a+b) \leq \mathbb{R}(ab)$. While for any $F_{3,2}(a, b, c; e, f; z)$, the function is convergent if $|z| < 1$, or, if $z = 1$ then

$$\mathbb{R}(e+f-a-b-c) > 0, \tag{1.15}$$

or, if $z = -1$ then

$$\mathbb{R}(e + f - a - b - c) > -1.$$

In particular, when $z = 1$, we may also employ the generalised Dixon's theorem (found at (2.3.3.7) in Slater, 1966) to scale the arguments; this theorem states

$$F_{3,2}(a, b, c; e, f; 1) = \frac{\Gamma(e)\Gamma(f)\Gamma(s)}{\Gamma(a)\Gamma(s+b)\Gamma(s+c)} F_{3,2}(e-a, f-a, s; s+b, s+c; 1), \qquad (1.16)$$

where $s = e + f - a - b - c$, where $\mathbb{R}(s) > 0$ and $\mathbb{R}(a) > 0$ ensure convergency in both series.

Exponential Integral and Properties of Related Integrals

The exponential integrals have series representations given by

$$E_1(z) = -\gamma - \ln z - \sum_{m=1}^{\infty} \frac{(-1)^m z^m}{m \times m!}$$

for $|\arg(z)| < \pi$, and

$$\mathrm{Ei}(z) = \gamma + \ln z + \sum_{m=1}^{\infty} \frac{z^m}{m \times m!}$$

for $z > 0$. It is important to observe here that

$$\mathrm{Ei}(-z) = -E_1(z),$$

and

$$E_1(z) = \int_z^{\infty} e^{-t} t^{-1} dt = \Gamma(0, z). \qquad (1.17)$$

Geller & Ng (1969) provide an useful list of integrals of the exponential integral, frequently used in Chapter 4, including (the parameters a, b and c are real and positive)

$$\int_0^{\infty} x e^{-ax} E_1(bx) dx = \frac{1}{a^2}\left\{\ln\left(1 + \frac{a}{b}\right) - \frac{a}{a+b}\right\}, \qquad (1.18)$$

$$\begin{aligned} \int_c^{\infty} e^{-ax} E_1(bx) \frac{dx}{x} &= [\gamma + \ln ac + E_1(ac)]\, E_1(bc) + \frac{1}{2}\left[\zeta(2) + (\gamma + \ln bc)^2\right] \\ &\quad + e^{-bc} \sum_{m=0}^{\infty} \frac{e_m(bc)}{(m+1)^2}\left(-\frac{a}{b}\right)^{m+1} + \sum_{m=1}^{\infty} \frac{(-bc)^m}{m! m^2}, \end{aligned} \qquad (1.19)$$

$$\int_0^{\infty} e^{-ax} (\ln x) E_1(bx) dx = -\frac{1}{a}\left\{\ln\left(1 + \frac{a}{b}\right)[\gamma + \ln(a+b)] + \frac{a}{a+b}\Phi\left(\frac{a}{a+b}, 2, 1\right)\right\}, \qquad (1.20)$$

and

$$\int_0^\infty xe^{-ax}(\ln x)E_1(bx)dx = -\frac{1}{a^2}\left\{ \begin{array}{c} \left[\ln\left(1+\frac{a}{b}\right) - \frac{a}{a+b}\right][\gamma + \ln(a+b) - 1] \\ +\left(\frac{a}{a+b}\right)^2 \Phi\left(\frac{a}{a+b}, 2, 2\right) \end{array} \right\}, \quad (1.21)$$

defined, respectively, at (4.2.11), (4.2.29), (4.5.2) and (4.5.4) therein. Then, from Guillera & Sondow (2005), the above two Lerch transcendent functions, $\Phi(z,p,1)$ and $\Phi(z,p,2)$, are linked to the polylogarithm $Li_p(z)$ as follows:

$$Li_p(z) = z\Phi(z,p,1), \quad (1.22)$$

and

$$Li_p(z) - z = z^2\Phi(z,p,2). \quad (1.23)$$

1.3 Basic Concepts and Reliability Models

1.3.1 Basic Concepts

Suppose X is a nonnegative random variable representing the lifetime of an individual from a homogeneous population. The probability distribution of X can be specified in many ways, but the probability density function (pdf) and the cumulative distribution function (cdf) are particularly useful in reliability analysis. The pdf of X involving a vector of unknown model parameters $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)'$ defines the probability of a failure in a very small interval; it is given by

$$f(x;\boldsymbol{\pi}) = \lim_{\Delta x \to 0^+} \frac{\Pr(x \le X < x + \Delta x)}{\Delta x} = \frac{dF(x)}{dx},$$

at which $f(x;\boldsymbol{\pi}) \ge 0$ and $\int_0^\infty f(x)dx = 1$, so that, conversely, the cdf of X is defined as

$$F(x;\boldsymbol{\pi}) = \Pr(X \le x) = \int_0^x f(t)dt.$$

The hazard function specifies the instantaneous rate of failure or death at time $X = x$ (conditional upon survival to time x) and can be defined as

$$haz(x;\boldsymbol{\pi}) = \lim_{\Delta x \to 0} \frac{\Pr(x \le X < x + \Delta x | X \ge x)}{\Delta x} = \frac{f(x;\boldsymbol{\pi})}{S(x;\boldsymbol{\pi})},$$

where the survivor function $S(x;\boldsymbol{\pi}) = \int_x^\infty f(t;\boldsymbol{\pi})dt = 1 - F(x;\boldsymbol{\pi})$ is the probability of surviving until time x. These functions, unless stated otherwise, are defined over the interval $[0,\infty)$.

Other aspects of lifetime distribution are useful in certain circumstances; for instance, the $100q^{th}$ $(0 < q < 1)$ percentile function, also named as the q^{th} quantile function, written as B_q $(0 < q < 1)$, is the value x_q such that

$$\Pr(X \leq x_q) = q.$$

Thus, we may write B_q as

$$B_q = F^{-1}(q) = Q(q), \tag{1.24}$$

where the quantile function $Q(q)$ is the inverse of the cdf. The percentiles of a lifetime distribution specify times at which specified proportions of items fail; for example, in reliability analysis $B_{0.1}$ is commonly employed to determine a warranty period for the items under consideration, while $B_{0.9}$ is of particular relevance in survival analysis, especially in deciding the term of a life insurance contract.

The p^{th} moment about the origin, μ_p, of a pdf $f(x)$ is merely the expected value of X^p: that is,

$$\mu_p = E[X^p],$$

for $p = 1, 2, 3, \cdots$, while the p^{th} moment about the mean of X (or the p^{th} central moment) is defined as

$$\mu_p^* = E[(X - \mu)^p],$$

where $\mu \equiv \mu_1 = E[X]$ is the mean; these moments can be used to find some characteristics of the distribution of X. For example, the skewness and kurtosis of X are, respectively,

$$\gamma_1 = \frac{\mu_3^*}{\sigma^3} \tag{1.25}$$

and

$$\gamma_2 = \frac{\mu_4^*}{\sigma^4} \tag{1.26}$$

where $\sigma^2 \equiv \mu_2^* = Var(X)$ is the variance.

1.3.2 Lifetime Distributions

We have already mentioned three particular distributions, although other parametric models have been used throughout the literature on lifetime data. For a survey of the properties and theoretical bases of these distributions, see, for instance, Lawless (1982), Nelson (1982), and Bain & Engelhardt (1991). Throughout this thesis, we will consider the following particular distributions.

The Exponential Distribution

The exponential distribution, sometimes referred to as the negative exponential distribution, was widely used in early work on the reliability of, for example, electronic components and, to a more limited extent, in clinical studies. With only one parameter, it is rather sensitive even when modelling data with modest departures to this distribution, especially when such

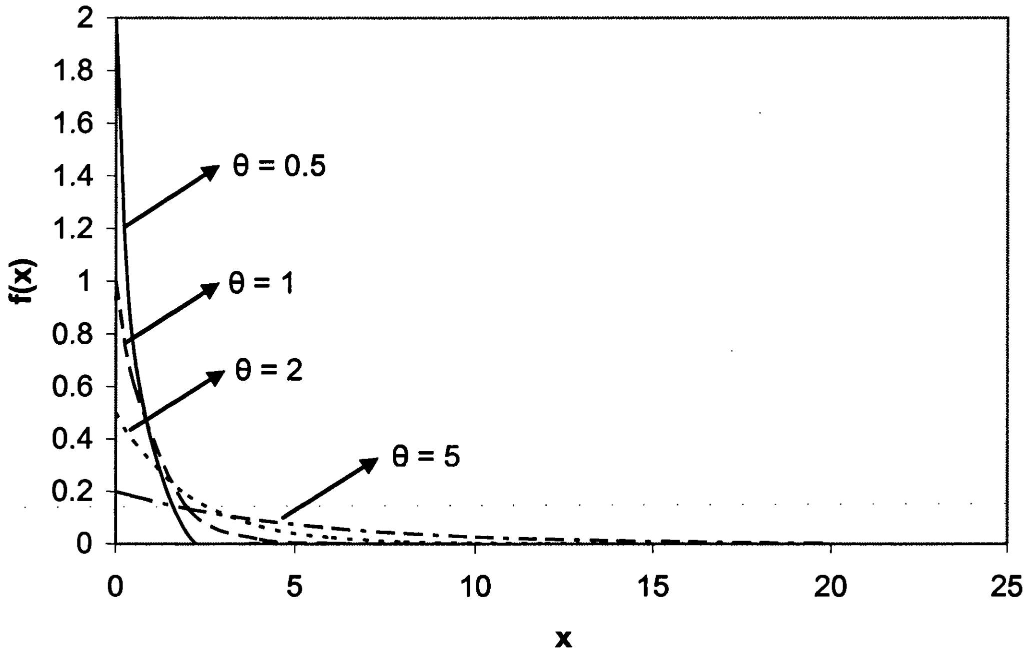

Figure 1.4: Pdf of the exponential distribution for varying θ.

departures occur in the tail. The exponential distribution has a one-parameter pdf given by

$$f(x;\theta) = \theta^{-1} \exp\left\{-\frac{x}{\theta}\right\}, \tag{1.27}$$

where $\theta > 0$ is the scale parameter, so that the cdf is

$$F(x;\theta) = 1 - \exp\left\{-\frac{x}{\theta}\right\}. \tag{1.28}$$

Figure 1.4 shows the exponential pdf for several values of θ. The special case of $\theta = 1$ is called the standard exponential distribution.

From (1.27) and (1.28), we see that the hazard function is

$$haz(x;\theta) = \theta^{-1}.$$

Hence, this distribution is characterised by a constant hazard function over the range of X, which implies that the instantaneous rate of failure or death is independent of x, so that the conditional chance of failure in a time interval of specified length is the same regardless of how long the individual has been on trial; this is referred to as the lack-of-memory (or memoryless) property. Moreover, the $100q^{th}$ percentile function can be expressed as

$$B_q = \theta\{-\ln(1-q)\}. \tag{1.29}$$

We also have

$$\mu_p = \theta^p \Gamma(1+p),$$

so that the mean, variance, skewness, and kurtosis are, respectively, $\mu = \theta, \sigma^2 = \theta^2, \gamma_1 = 2$, and $\gamma_2 = 9$.

The Weibull Distribution

The Weibull distribution, introduced by Weibull (1939, 1951), is the most widely used in reliability modelling, due to its flexibility in fitting failure time data in many applications, particularly when related to extreme-value characteristics. This distribution has pdf (with origin at zero) defined as

$$f(x;\theta,\beta) = \beta\theta^{-\beta}x^{\beta-1}\exp\left\{-\left(\frac{x}{\theta}\right)^{\beta}\right\}, \tag{1.30}$$

and cdf given by

$$F(x;\theta,\beta) = 1-\exp\left\{-\left(\frac{x}{\theta}\right)^{\beta}\right\}, \tag{1.31}$$

where $\theta > 0$ and $\beta > 0$ are the scale and shape parameters respectively. Hence, the elegance and utility of this model are further enhanced by having a closed form cdf. The hazard function is

$$haz(x;\theta,\beta) = \beta\theta^{-\beta}x^{\beta-1},$$

and the $100q^{th}$ percentile function is readily found to be

$$B_q = \theta\left\{-\ln(1-q)\right\}^{\frac{1}{\beta}}. \tag{1.32}$$

Figure 1.5 illustrates how varying β affects the shape of the Weibull pdf for $\theta = 10$. When $\beta > 1$, this distribution is bell-shaped, indicating increasing hazard over time. However, for $\beta \le 1$, it is reverse J-shaped, an indication of decreasing hazard over time. In particular, the Weibull distribution reduces to negative exponential when $\beta = 1$; it is known as the Rayleigh when $\beta = 2$, and for $\beta = 3.6023\cdots$, it is approximately Normal with $\gamma_1 = 0$ and $\gamma_2 = 2.72$ (as compared to 3 for the Normal distribution); see Cohen (1991).

In addition, the Weibull distribution has p^{th} moment about zero given by

$$\mu_p = \theta^p \Gamma\left(1+\frac{p}{\beta}\right);$$

then its mean and variance are

$$\mu = \theta\Gamma\left(1+\frac{1}{\beta}\right),$$

and

$$\sigma^2 = \theta^2\left\{\Gamma\left(1+\frac{2}{\beta}\right)-\Gamma^2\left(1+\frac{1}{\beta}\right)\right\}.$$

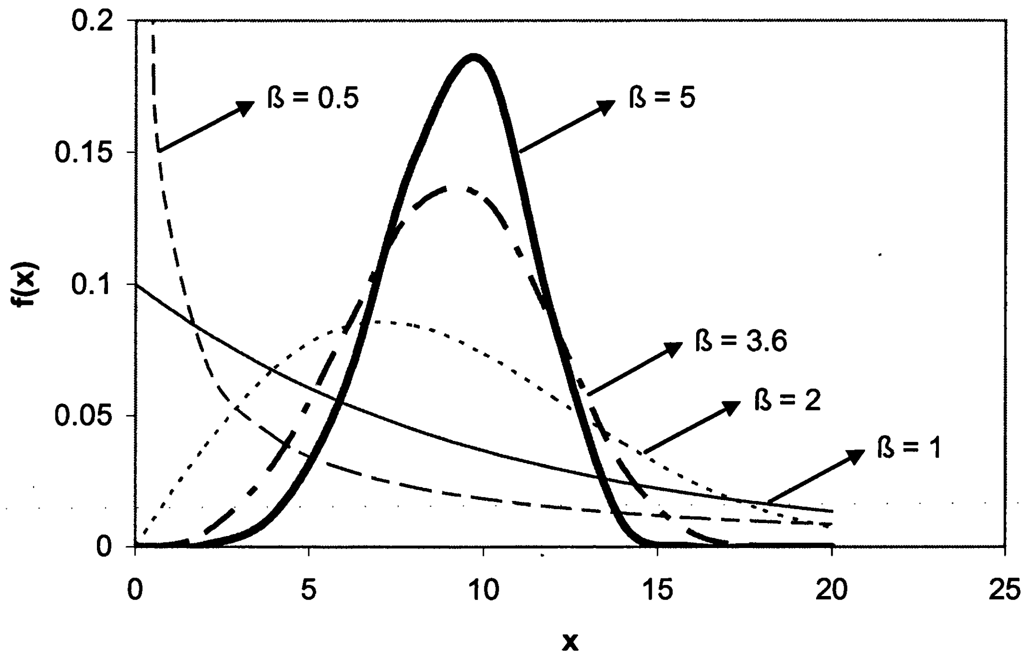

Figure 1.5: Pdf of the Weibull distribution for $\theta = 10$ and varying β.

The Burr Type XII Distribution

The Burr Type XII distribution (hereafter referred to simply as the Burr distribution), introduced by Burr (1942), has been shown empirically to provide a good fit to data in many different types of characteristics and applications. Examples include fitting a uranium survey data set (Cook and Johnson, 1986), data arising in actuarial science (Klugman, 1986), analysis of business failure data (Lomax, 1954), modelling the size distribution of incomes (Sinha and Moddola, 1976), the efficacy of analgesics in clinical trials (Wingo, 1983), and the time to failure of electronic components (Wingo, 1993; Wang *et al.*, 1996). In their discussion on the statistical and probabilistic properties of the Burr distribution, Zimmer *et al.* (1998) emphasise the advantages of this distribution in modelling failures over the other commonly used models, such as the log-normal and the log-logistic distributions. These advantages include the fact that the Burr XII covers the curve shape characteristics for the Normal, logistic and exponential (Pearson Type X) distributions, as well as a significant portion of the curve shape characteristics for the Pearson Type I (beta), II, III (gamma), V, VII, IX and XII; for instance, see Burr & Cislak (1968), Rodriguez (1977), and Tadikamalla (1980).

For simplicity, we initially focus on the basic two-parameter Burr distribution. This has positive shape parameters α and τ, with pdf

$$f\left(x;\alpha,\tau\right)=\alpha\tau x^{\tau-1}\left(1+x^{\tau}\right)^{-(\alpha+1)}, \tag{1.33}$$

a closed form cdf

$$F(x;\alpha,\tau) = 1 - (1 + x^{\tau})^{-\alpha}, \tag{1.34}$$

hazard function

$$haz(x;\alpha,\tau) = \alpha\tau x^{\tau-1}(1 + x^{\tau})^{-1},$$

and $100q^{th}$ percentile function given by

$$B_q = \left\{(1-q)^{-\frac{1}{\alpha}} - 1\right\}^{\frac{1}{\tau}}. \tag{1.35}$$

Figure 1.6 shows the effect of changing α when $\tau = 1$; larger values of α correspond to steeper density functions that tend to 1 more rapidly. In contrast, Figure 1.7 presents a similar comparison for varying τ, with $\alpha = 1$; we see that increasing τ produces a steeper density function that tends to 1 extremely quickly. The moment μ_p for this distribution exists provided that $\alpha\tau > p$; we have

$$\mu_p = \alpha B\left(\frac{p}{\tau} + 1, \alpha - \frac{p}{\tau}\right), \tag{1.36}$$

so, with $\alpha\tau > 2$, the mean and variance are

$$\mu = \alpha B\left(\frac{1}{\tau} + 1, \alpha - \frac{1}{\tau}\right),$$

and

$$\sigma^2 = \alpha B\left(\frac{2}{\tau} + 1, \alpha - \frac{2}{\tau}\right) - \alpha^2 B^2\left(\frac{1}{\tau} + 1, \alpha - \frac{1}{\tau}\right).$$

It is appropriate to mention here the connection between the Weibull and Burr distributions, by which the lower bound for the Burr region forms part of the Weibull curve in the (γ_1, γ_2) plane, the limiting distribution of (1.33) as $\alpha \to \infty$ is the Weibull distribution (1.30). This is shown by Rodriguez (1977) as follows:

$$\begin{aligned}
\Pr\left(x < \left(\frac{1}{\alpha}\right)^{\frac{1}{\tau}} y\right) &= 1 - \left(1 + \frac{y^{\tau}}{\alpha}\right)^{-\alpha} \quad \text{from (1.34)} \\
&= 1 - \exp\left\{-\alpha \log\left(1 + \frac{y^{\tau}}{\alpha}\right)\right\} \\
&= 1 - \exp\left\{-\alpha\left[\frac{y^{\tau}}{\alpha} - \frac{1}{2}\left(\frac{y^{\tau}}{\alpha}\right)^2 + \cdots\right]\right\} \\
&= 1 - \exp\{-y^{\tau}\} \quad \text{as } \alpha \to \infty.
\end{aligned}$$

Thus, in this limit, the Burr shape parameter τ corresponds to the Weibull shape parameter β.

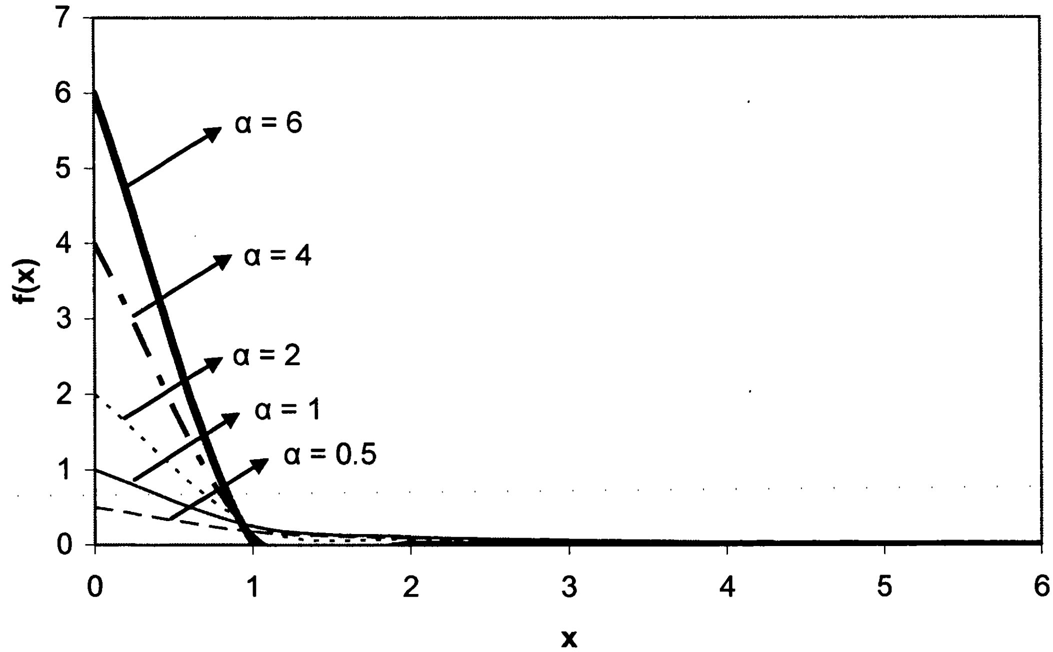

Figure 1.6: Pdf of the Burr distribution for $\tau = 1$ and varying α.

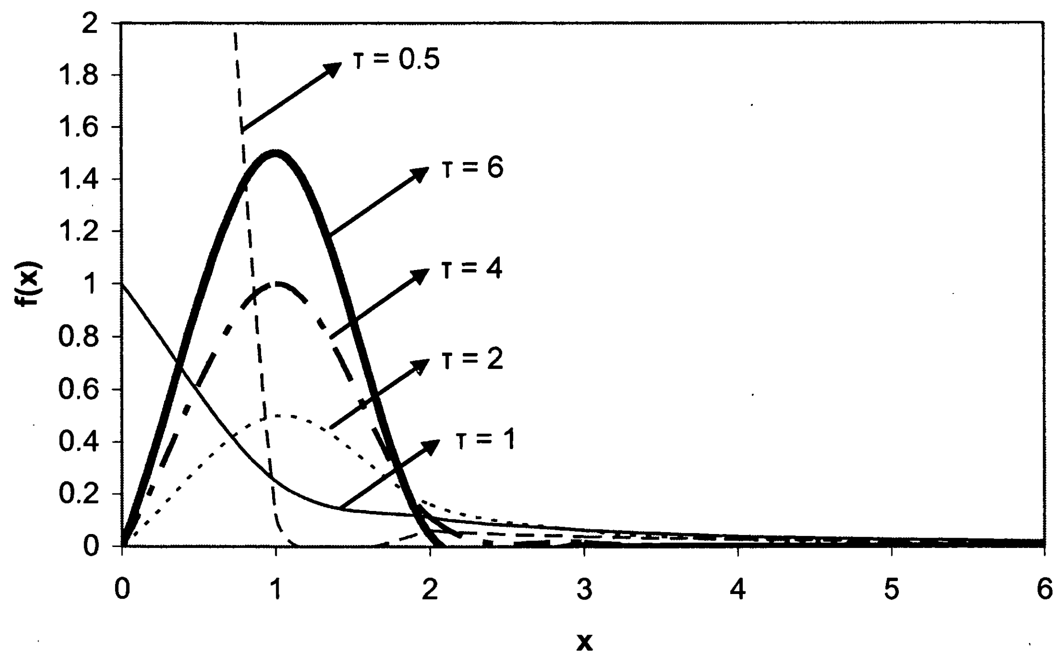

Figure 1.7: Pdf of the Burr distribution for $\alpha = 1$ and varying τ.

The Pareto Distribution

We briefly mention the Pareto distribution, noting its link to the negative exponential distribution. The two-parameter Pareto pdf is given by

$$f(x;\alpha,k) = \alpha k^{\alpha} x^{-(\alpha+1)}, \tag{1.37}$$

with corresponding cdf

$$F(x;\alpha,k) = 1 - \left(\frac{k}{x}\right)^{\alpha}, \tag{1.38}$$

for $x \geq k$, where $\alpha > 0, k > 0$ are, respectively, shape and location parameters. It is straightforward to show that $\ln(X/k)$ has an exponential distribution with mean α^{-1}. We shall exploit this important relationship later in the derivation of the expected Fisher information matrix for the Burr distribution.

1.4 Censoring Regimes

We have already mentioned that, at the close of a life-testing experiment in reliability, not all specimens may have failed. For example, suppose n light bulbs are selected at random and placed on test. Many, perhaps nearly all, may fail in the first year, but a few bulbs may last for several years. Similarly, some patients will survive to the end of a clinical trial. An individual who is observed to be failure-free for 30 days and then withdrawn from the study has a failure time which must exceed 30 days. Such incomplete observation of the failure time is called censoring.

Censored sampling is a key feature of failure time data (indeed reliability and survival analysis have been broadly defined as the analysis of censored data), and the mechanisms which give rise to censoring play a crucial part in statistical inference. Some of the commonest assumptions are right censoring, left censoring, Type I censoring and Type II censoring; these are not all mutually exclusive. For later, and practical utility, we only consider Type II singly censored sampling on the right, though many of the ideas transfer in an obvious way to the case of Type I and/or left censoring.

1.4.1 Right and Left Censoring

Formally, data are right censored if the censoring regime cuts short observations in progress. An example is the ending of an investigation at a fixed time. In contrast, data are left censored if the censoring mechanism prevents us from knowing when entry into the state which we wish to observe took place. Both forms of censoring can occur in practice. For instance, in medical studies in which patients are subject to regular examinations, discovery of a condition indicates only that the onset fell in the period since the previous examination; the time elapsed since onset is thus left censored. Right censoring is very common in life-testing of electromechanical items, but left censoring is fairly rare. We ignore left censoring here, so that the term "censoring" in the remainder of this thesis will generally mean "right censoring".

1.4.2 Type I Censoring

Suppose that n items are independently tested and entered into a trial at the same time. If the experiment is terminated after a pre-specified time t, this is referred to as Type I censored sampling (on the right). As a result, the number of observed failures m $(0 \leq m \leq n)$ is a random variable, and the remaining $n - m$ items are censored at the stopping time t. We use the ball bearings data from Table 1.2 to illustrate this experimental set-up. Suppose the trial is terminated at time $t = 60$, instead of allowing all of the items to fail, then the Type I censored sample would be as follows

17.88	28.92	33.00	41.52	42.12	45.60	48.48	51.84	51.96	54.12
55.56	$60^{\dagger}$	$60^{\dagger}$	$60^{\dagger}$	$60^{\dagger}$	$60^{\dagger}$	$60^{\dagger}$	$60^{\dagger}$	$60^{\dagger}$	$60^{\dagger}$
$60^{\dagger}$	$60^{\dagger}$	$60^{\dagger}$							

This form of censoring has the practical advantages of known experimental duration, but the statistical disadvantage of prior uncertainty over the exact number of failure times available for analysis.

1.4.3 Type II Censoring

In contrast, Type II censored sampling (on the right) occurs when the experiment is discontinued after the first r $(r \leq n)$ failure times are observed. The number of failures r is fixed in advance, and the remaining $n - r$ items will have a censored failure time equal to the time of failure of the r^{th} item. Using the ball bearings data again, suppose that the life test is stopped after $r = 12$ failures are obtained. Thus, lifetimes after the 12^{th} item are censored at the value of 67.80, and we would obtain the following Type II censored sample

17.88	28.92	33.00	41.52	42.12	45.60	48.48	51.84	51.96	54.12
55.56	67.80	$67.80^{\dagger}$	$67.80^{\dagger}$	$67.80^{\dagger}$	$67.80^{\dagger}$	$67.80^{\dagger}$	$67.80^{\dagger}$	$67.80^{\dagger}$	$67.80^{\dagger}$
$67.80^{\dagger}$	$67.80^{\dagger}$	$67.80^{\dagger}$							

Type II censoring has the significant advantage that an experimenter knows in advance how many failure times the experiment will yield, which helps enormously when planning tests with an adequate level of statistical precision. However, the experimental duration is not known precisely in advance, and it is possible for an experiment to continue for long periods until r failures are observed.

We aim to gain more insight into the roles of censoring number r and sample size n in a Type II censoring setting. In this example, we may wish to assess the difference between censoring at $r = 8$ and $r = 16$. For $r = 8$, testing stops after 51.84 million revolutions, while, with $r = 16$, we would need to wait roughly 30 million revolutions longer. We can also assess the changes due to waiting for the final few failures, by taking $r = 20$, when we intuitively expect estimates to be more consistent with final values than with $r = 8$ or 16. More generally, we can consider the precision in using a Type II censored estimate as an estimate to its complete counterpart. This approach will require an assessment of the relationship between interim and final estimates.

In our simulation experiments, we censor the data at various proportions of the sample size, typically taking $r = 0.2n, 0.4n, 0.6n, 0.8n, 1.0n$, so that the last case corresponds to a complete sample. In practice, however, factors such as the cost of testing units, the precision required and the value of saving time would be important in deciding the best choice for the sampling plan; this will be explored in further detail elsewhere.

The above outline indicates that, in Type II censored sampling, the data arrives already in a naturally ordered way due to the method of experimentation. Hence, it is now appropriate to review briefly some properties of order statistics.

1.5 Properties of Order Statistics

The theory of order statistics is well-established, but known to be analytically complicated, chiefly because the probability density functions of order statistics contain both the probability density and powers of cumulative distribution functions for the underlying population. Thus, relatively basic theoretical properties of order statistics, such as their expectations and joint expectations, can involve integrals of considerable complexity, even for well-known and widely-used lifetime models such as the Weibull and Burr distributions. David & Nagaraja (2003) is a standard reference for the theory of order statistics; we also note the two volumes by Balakrishnan & Rao (1998a,b), the first of which focuses on theory and methods, while the second one deals primarily with applications.

1.5.1 Notation and Basic Properties

Let X be a continuous random variable with probability distribution F and probability density function f. Suppose that a random sample $X_1, X_2, \ldots, X_n$ from this distribution is put in ascending order, and the re-ordered sample denoted (in a standard way) as

$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}.$$

That is, $X_{1:n}$ is the smallest sample value, $X_{2:n}$ is the next smallest, and so on. The set of these ordered quantities is referred to as the order statistics of this sample with size n. For a single order statistic $X_{i:n}$ $(1 \leq i \leq n)$, its cdf is given by

$$F_{(i)}(x) = \sum_{j=i}^{n} \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j}, \tag{1.39}$$

where the range of $X_{i:n}$ is that of X. Then, differentiating (1.39) wrt x yields the pdf of $X_{i:n}$ as

$$f_{(i)}(x) = c_{i:n} f(x) \left[F(x)\right]^{i-1} \left[1 - F(x)\right]^{n-i}, \tag{1.40}$$

where we write

$$c_{i:n} = \frac{n!}{(n-i)!(i-1)!} = n\binom{n-1}{i-1} = i\binom{n}{i}.$$

From (1.40), we see that the probability of event $x < X_{i:n} \leq x + \Delta x$ can also be found from the probability that, of the n values $X_1, X_2, \ldots, X_n$, $(i-1)$ of the X_i are less than x, one X_i is in $(x, x+\Delta x)$ and $(n-i)$ of the X_i are greater than $x + \Delta x$.

The formulae are greatly simplified when we consider the cdf and pdf of $X_{1:n}$; here, we obtain, respectively,

$$F_{(1)}(x) = 1 - \left[1 - F(x)\right]^n, \tag{1.41}$$

and

$$f_{(1)}(x) = n \left[1 - F(x)\right]^{n-1} f(x). \tag{1.42}$$

It is also known that, if $X_1, X_2, \ldots, X_n$ be independent and identically distributed continuous random variables from any member of the exponential family, then $X_{1:n}$ will follow the distribution at which X_i are taken. This is because both F and f have the exponential function, and hence the algebra simplifies; see Patel *et al.* (1976). For instance, when X follows the exponential, Weibull, Burr and Pareto distributions, (1.41) becomes

$$\text{Exponential} : 1 - \exp\left\{-\frac{nx}{\theta}\right\} = 1 - \exp\left\{-\frac{x}{\theta^*}\right\}, \tag{1.43a}$$

$$\text{Weibull} : 1 - \exp\left\{-n\left(\frac{x}{\theta}\right)^\beta\right\} = 1 - \exp\left\{-\left(\frac{x}{\theta^*}\right)^\beta\right\}, \tag{1.43b}$$

$$\text{Burr} : 1 - (1 + x^\tau)^{-\alpha n} = 1 - (1 + x^\tau)^{-\alpha^*}, \tag{1.43c}$$

$$\text{Pareto} : 1 - \left(\frac{k}{x}\right)^{\alpha n} = 1 - \left(\frac{k}{x}\right)^{\alpha^*}, \tag{1.43d}$$

in turn; that is, the same distribution, but with at least one different parameter, as follows:

	X	$X_{1:n}$
Exponential	θ	$\theta^* = \theta/n$
Weibull	θ, β	$\theta^* = \theta/n^{\frac{1}{\beta}}, \beta$
Burr	α, τ	$\alpha^* = \alpha n, \tau$
Pareto	α, k	$\alpha^* = \alpha n, k$

The joint distributions of order statistics can be similarly derived, although naturally more complicated. For $x < y$, the joint cdf of $X_{i:n}$ and $X_{j:n}$ $(1 \leq i < j \leq n)$ is

$$F_{(i,j)}(x, y) = \sum_{s=j}^{n} \sum_{r=i}^{s} \frac{n!}{r!(s-r)!(n-s)!} \left[F(x)\right]^r \left[F(y) - F(x)\right]^{s-r} \left[1 - F(y)\right]^{n-s}. \tag{1.44}$$

Also, for $x \geq y$, the inequality $X_{j:n} \leq y$ implies $X_{i:n} \leq x$ so that

$$F_{(i,j)}(x,y) = F_{(j)}(y).$$

If we extend the definition of $c_{i:n}$ to

$$c_{i,j:n} = \frac{n!}{(i-1)!(j-i-1)!(n-j)!},$$

then the joint pdf of $X_{i:n}$ and $X_{j:n}$ may be written as

$$f_{(i,j)}(x,y) = c_{i,j:n}\left[F(x)\right]^{i-1}\left[F(y)-F(x)\right]^{j-i-1}\left[1-F(y)\right]^{n-j} f(x)f(y) \tag{1.45}$$

for $x < y$, as obtained from (1.44) by differentiation.

1.5.2 Moments and Product Moments

In the general continuous case, the single moment of $X_{i:n}$ $(1 \leq i \leq n)$ is

$$E\left[X_{i:n}^p\right] = \int_x x^p f_{(i)}(x)dx = c_{i:n}\int_x x^p f(x)\left[F(x)\right]^{i-1}\left[1-F(x)\right]^{n-i} dx, \tag{1.46}$$

while the product moment of $X_{i:n}$ and $X_{j:n}$ $(1 \leq i < j \leq n)$, $E\left[X_{i:n}^p X_{j:n}^q\right]$, is defined as

$$\begin{aligned} &\int_y \int_{x<y} x^p y^q f_{(i,j)}(x,y)dxdy \\ = \quad & c_{i,j:n}\int_y \int_{x<y} x^p y^q \left[F(x)\right]^{i-1}\left[F(y)-F(x)\right]^{j-i-1}\left[1-F(y)\right]^{n-j} f(x)f(y)dxdy \end{aligned} \tag{1.47}$$

As with the distribution of X, we can use moments and product moments to compute summaries of the distribution and joint distribution of order statistics, as required. For instance, the covariance of $X_{i:n}$ and $X_{j:n}$ is simply

$$Cov(X_{i:n}, X_{j:n}) = E\left[X_{i:n}X_{j:n}\right] - E\left[X_{i:n}\right]E\left[X_{j:n}\right].$$

1.5.3 Recurrence Relations for Moments and Product Moments

Expectations and joint expectations of order statistics can be derived explicitly in some distributions such as exponential and Pareto, but need to be computed by numerical methods in most other models. Otherwise, one may use the recurrence relations between the moments of order statistics, chiefly to cut down the number of independent calculations required when evaluating an expectation. David & Nagaraja (2003) and Balakrishnan & Rao (1998a) provide several recursive relations and identities satisfied by the moments of order statistics from some specific continuous distributions, wherein the interrelationships between many of these results are presented. In general, for an arbitrary function g of a single order statistic, we have

$$(n-i)E\left[g\left(X_{i:n}\right)\right]+iE\left[g\left(X_{i+1:n}\right)\right]=nE\left[g\left(X_{i:n-1}\right)\right], \tag{1.48}$$

for $1 \leq i \leq n-1$, linking the expectations of order statistics from neighbouring sample sizes. As we have seen in (1.41) and (1.42), computation can be greatly simplified if we could express the moments of $X_{i:n}$ in terms of the simpler moments of the smallest in samples of $1, 2, \ldots, i$ for which the properties and results of $X_{1:n}$ are a lot more straightforward than the other order statistics. By repeated use of (1.48), Watkins & John (2006) obtained an expression of the expectation of $g\left(X_{i:n}\right)$ in terms of the first order statistic in various sample sizes; we have

$$E\left[g\left(X_{i:n}\right)\right]=\sum_{j=1}^{i}(-1)^{i-j}\binom{n}{j-1}\binom{n-j}{i-j}E\left[g\left(X_{1:n+1-j}\right)\right]. \tag{1.49}$$

As a result, we can exploit the connection between the distribution of $X_{1:n}$ and the underlying distribution, as illustrated in (1.43).

Similarly, for joint order statistics we have

$$\begin{aligned}(i-1)E[g\left(X_{i:n}\right)h(X_{j:n})] &= nE[g\left(X_{i-1:n-1}\right)h(X_{j-1:n-1})]-(j-i)E[g\left(X_{i-1:n}\right)h(X_{j:n})]\\ &\quad -(n-j+1)E[g\left(X_{i-1:n}\right)h(X_{j-1:n})] \end{aligned} \tag{1.50}$$

where $2 \leq i < j \leq n$ and g, h are arbitrary functions. Then, using this, we may state the joint expectation of $g\left(X_{i:n}\right)$ and $h(X_{j:n})$ in terms of the first order statistic with the j^{th} in different sample sizes (see John, 2003)

$$E[g\left(X_{i:n}\right)h(X_{j:n})]=\frac{n!}{(j-i-1)!}\sum_{s=1}^{i}\sum_{t=0}^{i-s}\left[\begin{array}{c}\frac{(-1)^{s+t-1}(n+t-j)!(s+j-i-2)!}{t!(n-j)!(i-t-s)!(s-1)!(n+t+s-i)!}\times\\ E[g\left(X_{1:n-i+s+t}\right)h(X_{j-i+s:n-i+s+t})]\end{array}\right]. \tag{1.51}$$

We note that this important result is independent of the underlying distribution.

It should be noted that results for higher order moments are possible; see, for example, Chapter 2 in Balakrishnan & Rao (1998b), for recurrence relations satisfied by the triple and quadruple moments of order statistics from the standard exponential distribution.

1.6 Numerical Considerations

In order to validate theoretical results developed, this thesis will rely heavily on computing software. One particular instance is to obtain, by running simulations, a sampling distribution of maximum likelihood estimator to check that asymptotic Normality holds for large sample sizes, but also to assess the extent to which asymptotic results apply in relatively small samples. Therefore, it will be of particular interest to consider various computational strategies for evaluating these results for specific values of sample size (choosing a range of sample sizes likely to be encountered in practice, but also assessing agreement with asymptotic formulae), and of a wide range of censoring levels and representative distribution parameter values.

Throughout this thesis, we will use Mathematica (Wolfram, 1999) for theoretical evaluations, and the standard statistical package SAS (SAS, 2004) for simulated counterparts. We also use Microsoft Excel and SPSS for simpler calculations and graphs.

1.6.1 Data Simulation

We often require to simulate data in order to validate the theoretical expressions. In general, if u represents an observation from a uniform distribution in $(0, 1)$, then a simulated observation x from a distribution with cdf F is given by

$$x = F^{-1}(u);$$

this is known as the inverse transformation method. We remark that x is effectively the u^{th} quantile function, and that the inverse transform method works best if the distribution has a closed form cdf. Therefore, to simulate a set of data from an exponential distribution with specified parameter θ, we use (1.28) and calculate

$$x = -\theta \ln(1-u),$$

while, for the Weibull distribution, we employ (1.31) and compute

$$x = \theta \{-\ln(1-u)\}^{\frac{1}{\beta}},$$

and, for the Burr distribution, we have, from (1.34),

$$x = \left\{(1-u)^{-\frac{1}{\alpha}} - 1\right\}^{\frac{1}{\tau}}.$$

In SAS, we generate independent and identically distributed uniform $(0, 1)$ random variates using the function **ranuni** and then find the corresponding x values from the above formulae, though one may also employ **ranexp** function to generate an exponential random value. We also generally take the number of replications, N, to be 10^4, so that inferences on the tails of a particular distribution are based on an acceptable number (≥ 100) of replicated values. This value of N generally provides a reliable representation of the distribution under consideration, and, perhaps importantly, is also feasible in term of computational times in SAS.

1.6.2 Computer Generation of Order Statistics

When considering censored sampling, the ordered observations are needed. If $U_1, U_2, \ldots, U_n$ denote a random sample from the uniform $(0, 1)$ distribution and $U_{1:n} \leq U_{2:n} \leq \cdots \leq U_{n:n}$

are the matching order statistics, then, using the inverse transformation method, we obtain $(i = 1, 2, \ldots, n)$

$$X_{i:n} = F^{-1}(U_{i:n}),$$

represent the order statistics from the distribution with cdf F. There is a direct correspondence between the order statistics of $X_1, X_2, \ldots, X_n$ and the order statistics of the associated uniform sample $U_1, U_2, \ldots, U_n$. Again, in SAS, we use the sort procedure (`proc sort`), or simply `sort` if within the IML procedure (`proc IML`) to obtain the desired ordered sample.

1.6.3 Numerical Iterative Methods for Solving Equations

When the maximum likelihood method is employed to estimate parameters, we will need to find the roots of $\frac{dl_r^*}{d\pi}$, where l_r^* is the Type II censored profile log-likelihood function; in most cases, only limited analytical progress is possible, so that a numerical procedure must be employed. We generally locate the roots of $\frac{dl_r^*}{d\pi}$ using the Newton-Raphson computational procedure. This method is well-known for its quick convergence, and, again importantly, is eminently suitable for implementation in SAS; see Nelson (1982) for more details. Given an initial value $\pi^{[0]}$, a sequence of (generally) better approximations can be obtained by the iterative process

$$\tilde{\pi}^{[j+1]} = \tilde{\pi}^{[j]} - \frac{\frac{dl_r^*}{d\pi}\Big|_{\text{at } \pi=\tilde{\pi}^{[j]}}}{\frac{d^2 l_r^*}{d\pi^2}\Big|_{\text{at } \pi=\tilde{\pi}^{[j]}}}.$$

We generally stop the iterative process when

$$\left| \frac{\frac{dl_r^*}{d\pi}\Big|_{\text{at } \pi=\tilde{\pi}^{[j]}}}{\sqrt{-\frac{d^2 l_r^*}{d\pi^2}\Big|_{\text{at } \pi=\tilde{\pi}^{[j]}}}} \right| < 10^{-9};$$

this criterion is deemed equivalent to regarding the maximum likelihood iterations as converging.

1.7 Outline of Future Chapters

In this chapter, we have defined all relevant mathematical functions required in reliability analysis, and presented fundamental results for specific reliability distributions that will be the focus of our work, namely, the exponential, Weibull and Burr distributions. We have also discussed some practical considerations of and various forms of censoring regimes used to overcome difficulties in industrial life-testing. We then summarised the theory of order statistics, and concluded by outlining some numerical considerations.

In particular, we have used the ball bearings data to distinguish between Type II and Type I censoring, but also to illustrate various practically-based problems, which form the motivation behind our work. As noted in Section 1.4.3, we are interested at the link between a Type II censored estimate, obtained at the r^{th} failure, and the corresponding complete estimate, obtained when all items have failed. We proceed on the basis of a parametric modelling of data, and assume that we have identified a distribution for the data, so that we estimate the parameters and related quantities of that distribution. For example, using the ball bearings data again, we obtain the following maximum likelihood estimates of θ, β and $B_{0.1}$ (see Section 2.3 for further details on maximum likelihood estimation with Type II censored Weibull data) under Weibull analysis, when the data is subject to Type II censoring at the r^{th} failure.

r	8	12	16	20	23
$\widehat{\theta}_r$	67.6415	75.2168	76.6960	78.9674	81.8783
$\widehat{\beta}_r$	3.2280	2.6241	2.4695	2.3539	2.1021
$\widehat{B}_{0.1,r}$	33.6860	31.9063	30.8329	30.3563	28.0694

In this example, we may consider the extent to which the final estimates (of either parameters or percentile $B_{0.1}$) are consistent with earlier estimates, or the rate at which interim estimates converge on their final values; more generally, we can determine the precision with which we can make statements on final estimates, based on interim estimates. This approach requires an assessment of the extent to which $\widehat{\theta}_r, \widehat{\beta}_r$ and $\widehat{B}_{0.1,r}$ can, respectively, be regarded as a reliable guide to $\widehat{\theta}_n, \widehat{\beta}_n$ and $\widehat{B}_{0.1,n}$, and hence we study the relationship between final and interim estimates.

As already noted, Chapter 2 considers in further details the method of maximum likelihood to obtain estimates of the model parameters and the percentile function for the aforementioned lifetime distributions under Type II censoring, and derives the expected Fisher information matrix analytically. This, in turn, yields asymptotically valid variances and covariances of the maximum likelihood estimators, and their large-sample properties.

In this thesis, we will consider three distinct problems regarding the maximum likelihood estimation under a Type II censoring regime:

- Asymptotic Normality of maximum likelihood estimators is well known, for example, see Cox & Hinkley (1974) and Bain & Engelhardt (1991). This large sample result is often used in making inferences from small to moderate samples, despite the drawback that it is not always accurate with such sample sizes. Chapter 3 assesses, by means of a detailed simulation study, the extent to which the assumption of Normality for parameters and $B_{0.1}$ holds for finite Type II censored samples, and the role of censoring in the convergence towards Normality.

- Although the large-sample result is, perhaps surprisingly, rather robust in some senses - for instance, the distribution of the maximum likelihood estimator of $B_{0.1}$ converges to Normality more rapidly than those of the model parameters - it is also the case that the large-sample result can be shown to be unrealistic in samples of small to moderate size, such as in the ball bearings data with $n = 23$ failure times. Hence, we also discuss in Chapter 3 the use of relative likelihood function and related contour plots as an alternative for assessing the precision in estimates of parameters in relatively small or highly censored samples.

- We then move on to establish a method to measure the precision in using a Type II censored analysis as a guide to the final analysis. Since the analysis of reliability of Type II censored data typically requires single and joint expectations of order statistics, Chapter 4 computes all necessary moments and product moments for various functions of order statistics; this involves a considerable amount of algebra. Chapter 5 then considers the correlations between final and interim estimates of model parameters and $B_{0.1}$; for large samples, this is then transformed into a study of the correlations of score functions. These results, in turn, give asymptotic 95% confidence limits for the final estimate given the interim estimate, which we will regard as a measure of precision. We illustrate these results using published data sets and simulation experiments, from which some practical implications are drawn.

Lastly, Chapter 6 presents summaries and conclusions, together with a brief outline of some possible future research.

Chapter 2

Maximum Likelihood Estimation Based on Type II Censored Samples

2.1 Introduction

As outlined above, we suppose we have identified a distribution for the lifetimes, so that it remains to estimate the parameters of that distribution. The method of maximum likelihood has theoretical support (see Crowder *et al.*, 1991, for instance); moreover, computer programmes for the appropriate calculations are widely available, for example, for implementing a numerical search for the root of an equation, estimating the model parameters by the method of maximum likelihood is also to be recommended on practical grounds. Maximum likelihood estimation (hereafter abbreviated as ML estimation) for lifetime models considered in Chapter 1 is widely discussed throughout the reliability literature; for instance, see Lawless (1982), Bain & Engelhardt (1991), and Cohen (1991), though discussion on the Burr distribution is relatively limited. However, these references focus on the theoretical maximum likelihood equations, with few details on computation or further interpretation. We provide formulae for the elements of the expected Fisher information (hereafter abbreviated to EFI) matrix; in particular, analytical expressions for the elements of this matrix for the Burr distribution with Type II censored data is obtained. This allows us to write down the asymptotic covariance matrix of the maximum likelihood estimators (from now on, abbreviated to MLEs), and hence, the confidence intervals for the MLEs based on their asymptotic Normality. In addition to estimating the model parameters, it is particularly relevant in practical applications to make inferences on either the running time for the experiment or some percentile of lifetimes based on Type II censored samples; for example, estimating the 10^{th} percentile of failure times. Some discussion on percentile estimation is given in Meeker & Nelson (1974, 1977), where the emphasis concentrated on singly censored Weibull data. As in Chua & Watkins (2008a,b), we extend some recent work (Chua & Watkins, 2007) on the Weibull case to the Burr distribution, and consider the asymptotic distribution of the estimator of $B_{0.1}$. Results under complete sampling are also presented as a special case of Type II censoring.

We begin with a brief discussion on likelihood, and state the asymptotic properties of both MLE and score function. In Section 2.2, we consider the exponential model, for which most results can be expressed explicitly. Then, we extend the discussion to the Weibull (Section 2.3) and Burr (Section 2.4) distributions, where the extra parameter makes inference more involved. While there are no analytical expressions for MLEs of parameters, we obtain profile likelihood functions and maximise these instead. We illustrate Type II censoring using published data sets, and also present results from simulation experiments to assess the extent to which asymptotic results apply in samples of finite size.

2.1.1 Statistical Background

We now provide further details of the reliability setting: when n (> 0) independent items are put on a life test at the same time, and the experiment is terminated after some (pre-specified) number r $(1 \leq r \leq n$, although we are usually interested in $r < n)$ of failures, the data available for analysis is said to be Type II censored, and comprises the r order statistics $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{r:n}$, and $n - r$ lifetimes censored at $X_{r:n}$. The distinction between Type II censoring and complete sampling decreases as $r \to n$, and vanishes when $r = n$. Ignoring the ordering constant, the likelihood of a Type II singly right censored sample is

$$L_r = \left\{ \prod_{i=1}^{r} f(X_{i:n}; \boldsymbol{\pi}) \right\} \left\{ \prod_{i=r+1}^{n} [1 - F(X_{i:n}; \boldsymbol{\pi})] \right\}, \tag{2.1}$$

and the corresponding log-likelihood is

$$l_r = \sum_{i=1}^{r} \ln f(X_{i:n}; \boldsymbol{\pi}) + (n-r) \ln [1 - F(X_{r:n}; \boldsymbol{\pi})] \,. \tag{2.2}$$

The principle of ML estimation, as suggested by its name, is to select as an estimate of $\boldsymbol{\pi}$ the value for which the observed sample would have been most likely to occur. Assuming that the partial derivatives of l_r exist, then the maximising value is the solution of the simultaneous equations $(i = 1, \ldots, k)$

$$U_{r,i} = \frac{\partial l_r}{\partial \pi_i} = 0,$$

where $\mathbf{U}_r = (U_{r,1}, \ldots, U_{r,k})'$ is the score function. We denote the MLE by $\widehat{\boldsymbol{\pi}}_r$, in which r represents the censoring number.

The asymptotic theory of maximum likelihood (see, for example, Cox & Hinkley, 1974) implies that, in general, $\widehat{\boldsymbol{\pi}}_r$ is asymptotically Normally distributed with mean vector $\boldsymbol{\pi}$ and covariance matrix equal to the inverse of the EFI matrix $\mathbf{A}_r$, which is symmetric, with

$(i,j)^{th}$ entry

$$E\left[-\frac{\partial^2 l_r}{\partial \pi_i \partial \pi_j}\right],$$

for $i,j = 1,\ldots,k$, so that we need only give the lower triangle of elements. This, in turn, yields the approximate confidence limits for the parameter $\boldsymbol{\pi}$; for instance, the $100\,(1-\lambda)\,\%$ confidence intervals for π_i is

$$\widehat{\pi}_i \pm Z_{\lambda/2}\sqrt{Var\,(\pi_i)},$$

where $Z_{\lambda/2}$ is the upper $100\left(1-\frac{\lambda}{2}\right)$ percentage point of the standard Normal distribution. Where the true parameters are unknown (as in practice, although not in simulation experiments), we evaluate these limits by replacing $\boldsymbol{\pi}$ by $\widehat{\boldsymbol{\pi}}_r$. We have also implicitly introduced the observed Fisher information matrix, $\mathbf{J}_r$, given by

$$-\frac{\partial^2 l_r}{\partial \pi_i \partial \pi_j}.$$

In addition, the EFI matrix also appears in the asymptotic distribution of the score function; since l_r involves $\sum_{i=1}^{r} \ln f(X_{i:n};\boldsymbol{\pi})$, $\mathbf{U}_r$ is a sum of independent and identically distributed random variables, and, under mild conditions (for example, see Section 9.2 in Cox & Hinkley, 1974 and Bain & Engelhardt, 1991), is asymptotically Normally distributed with mean $\mathbf{0}$ and covariance matrix $\mathbf{A}_r$.

For percentile estimation, we will consider $q = 0.1$ throughout; the details and principles for other values of q are similar. In general, $B_{0.1}$ is a non-linear function of $\boldsymbol{\pi}$. Consequently, we consider the Taylor series of $B_{0.1}$ about the true parameter $\boldsymbol{\pi}$ up to its first-order derivative to estimate $B_{0.1}$; this can be written as

$$\widehat{B}_{0.1,r} \simeq B_{0.1} + \mathbf{b}'_{\boldsymbol{\pi}}\left(\widehat{\boldsymbol{\pi}}_r - \boldsymbol{\pi}\right), \tag{2.3}$$

with which $(i = 1,\ldots,k)$

$$\mathbf{b}_{\boldsymbol{\pi}} = \frac{\partial B_{0.1}}{\partial \pi_i}.$$

We see that $\widehat{B}_{0.1,r}$ is now a linear combination of $(\widehat{\boldsymbol{\pi}}_r - \boldsymbol{\pi})$, and hence is asymptotically Normal with mean

$$E\left[\widehat{B}_{0.1,r}\right] \simeq B_{0.1}$$

as, for large samples, $E\left[\widehat{\boldsymbol{\pi}}_r - \boldsymbol{\pi}\right] = 0$, and variance

$$Var\left(\widehat{B}_{0.1,r}\right) \simeq \mathbf{b}'_{\boldsymbol{\pi}} Var\left(\widehat{\boldsymbol{\pi}}_r - \boldsymbol{\pi}\right)\mathbf{b}_{\boldsymbol{\pi}} = \mathbf{b}'_{\boldsymbol{\pi}}\mathbf{A}_r^{-1}\mathbf{b}_{\boldsymbol{\pi}}. \tag{2.4}$$

Approximate $100\,(1-\lambda)\,\%$ confidence intervals for $B_{0.1}$ then follow immediately.

For complete samples, we can drop the subscript n; for instance, we write $L \equiv L_n$.

2.2 ML Estimation in the Exponential Distribution

From (2.1), the likelihood function for Type II censored data drawn from the exponential distribution is given by

$$L_r = \left[\prod_{i=1}^{r} \theta^{-1} \exp\left\{-\frac{X_{i:n}}{\theta}\right\}\right] \left[\exp\left\{-\frac{X_{r:n}}{\theta}\right\}\right]^{n-r} = \theta^{-r} \exp\left\{-\theta^{-1} S_r\right\},$$

in which

$$S_r = \sum_{i=1}^{r} X_{i:n} + (n-r)X_{r:n}, \tag{2.5}$$

so that the log-likelihood function can be expressed as

$$l_r = -r \ln \theta - \theta^{-1} S_r, \tag{2.6}$$

with derivative

$$\frac{dl_r}{d\theta} = -r\theta^{-1} + \theta^{-2} S_r. \tag{2.7}$$

S_r is sometimes referred to as "the total sample time on test". Hence, on equating (2.6) to zero, the MLE of θ is

$$\widehat{\theta}_r = \frac{S_r}{r}. \tag{2.8}$$

2.2.1 Regularity and EFI

Following Bain & Engelhardt (1991), we can write $S_r = \sum_{i=1}^{r} W_i$, where

$$W_1 = nX_{1:n}, \text{ and } W_i = (n-i+1)(X_{i:n} - X_{i-1:n}), \tag{2.9}$$

for $i = 2, \ldots, r$. The lack-of-memory property, previously mentioned in Section 1.3.2.1, indicates that the W_i $(i \geq 1)$ are independent variables following (1.27); we then have $E\left[S_r\right] = r\theta$ and $Var\left(S_r\right) = r\theta^2$, and hence

$$\begin{aligned} E\left[\widehat{\theta}_r\right] &= \frac{E\left[S_r\right]}{r} = \theta, \\ Var\left(\widehat{\theta}_r\right) &= \frac{Var\left(S_r\right)}{r^2} = \frac{\theta^2}{r}. \end{aligned} \tag{2.10}$$

It follows that (2.8) is an unbiased estimator of θ. Moreover, we see that

$$E\left[\frac{dl_r}{d\theta}\right] = -r\theta^{-1} + \theta^{-2} E\left[S_r\right] = 0,$$

as expected from the regularity consideration. The second derivative of (2.6) is

$$\frac{d^2 l_r}{d\theta^2} = r\theta^{-2} - 2\theta^{-3} S_r,$$

so that the EFI is given by

$$E\left[-\frac{d^2 l_r}{d\theta^2}\right] = -r\theta^{-2} + 2\theta^{-3}E\left[S_r\right] = r\theta^{-2}.$$

In particular, $\widehat{\theta}_r = S_r/r$ is the minimum variance unbiased estimator of θ, since

$$Var\left(\widehat{\theta}_r\right) = E\left[-\frac{d^2 l_r}{d\theta^2}\right]^{-1}.$$

As previously noted, we are interested at the estimation of the 10^{th} percentile function with Type II censored data; since (1.29) indicates that $B_{0.1}$ is linearly related θ, we obtain its MLE as

$$\widehat{B}_{0.1,r} = \widehat{\theta}_r\left(-\ln 0.9\right), \tag{2.11}$$

with mean

$$E\left[\widehat{B}_{0.1,r}\right] = \theta\left(-\ln 0.9\right),$$

and variance equals to

$$Var\left(\widehat{B}_{0.1,r}\right) = \left(-\ln 0.9\right)^2 Var\left(\widehat{\theta}_r\right) = \frac{\left(-\ln 0.9\right)^2 \theta^2}{r}.$$

2.2.2 Asymptotic Properties of the MLEs

The asymptotic Normality of MLEs implies that $\widehat{\theta}_r$ and $\widehat{B}_{0.1,r}$ can, for large sample sizes, be regarded as Normally distributed. We therefore have

$$\widehat{\theta}_r \sim N\left(\theta, \frac{\theta^2}{r}\right),$$

from which the 95% confidence intervals for θ is

$$\widehat{\theta}_r \pm 1.96\theta r^{-1/2}. \tag{2.12}$$

Similarly, we see that

$$\widehat{B}_{0.1,r} \sim N\left(B_{0.1}, \frac{\theta^2\left(-\ln 0.9\right)^2}{r}\right),$$

which, in turn, gives the 95% limits of $B_{0.1}$ as

$$\widehat{B}_{0.1,r} \pm 1.96\theta\left(-\ln 0.9\right) r^{-1/2}. \tag{2.13}$$

2.2.3 Complete Sample

For later convenience, we briefly present some results under complete sampling, obtained simply by setting $r = n$. The likelihood here is

$$L = \theta^{-n} \exp\left\{-\theta^{-1} S\right\},$$

from which the log-likelihood is

$$l = -n \ln \theta - \theta^{-1} S, \tag{2.14}$$

with derivative

$$\frac{dl}{d\theta} = -n\theta^{-1} + \theta^{-2} S, \tag{2.15}$$

so that the complete MLE of θ is

$$\widehat{\theta} = \frac{S}{n}.$$

Since $E[S] = n\theta$, the second derivative of (2.14) yields the EFI as $n\theta^{-2}$. From (2.11), the complete MLE of $B_{0.1}$ is

$$\widehat{B}_{0.1} = \widehat{\theta} \left(-\ln 0.9\right),$$

with the following characteristics:

$$E\left[\widehat{B}_{0.1}\right] = \theta \left(-\ln 0.9\right)$$

and

$$Var\left(\widehat{B}_{0.1}\right) = \frac{\left(-\ln 0.9\right)^2 \theta^2}{n}.$$

We note that Type II censored results are very similar to their complete counterparts; if n items are placed on test and first r failures are observed, it is clear that the statistical procedures based on this data are equivalent to those gained by placing n items on test and obtaining all n failures.

2.2.4 Numerical Examples

Epstein's Failure Times Data

We use the failure times data from Table 1.1, modelled, as in Epstein (1960) and as re-enforced in Figure 1.1, by the exponential distribution, to illustrate this experimental set-up. If we had stopped the experiment at $r = 40$, then failure times after the 40^{th} item are

r	10	20	30	40	49
$X_{r:49}$	15.2	55.6	108.5	178.6	354.4
$\widehat{\theta}_r$	67.6000	104.9000	114.0100	112.1150	104.8898
$\widehat{sd}(\widehat{\theta}_r)$	21.3770	23.4564	20.8153	17.7269	14.9843
95% CIs	25.701,109.499	58.926,150.874	73.212,154.808	77.370,146.860	75.521,134.259
$\widehat{B}_{0.1,r}$	7.1224	11.0523	12.0122	11.8125	11.0512
$\widehat{sd}(\widehat{B}_{0.1,r})$	2.2523	2.4714	2.1931	1.8677	1.5787
95% CIs	2.708,11.537	6.208,15.896	7.714,16.311	8.152,15.473	7.957,14.146

Table 2.1: Summaries of the exponential MLEs calculated at various r for Epstein's failure times data.

censored at the value of $X_{40:49} = 178.6$, and we would obtain the following data set

1.2	2.2	4.9	5.0	6.8	7.0	12.1	13.7	15.1	15.2
23.9	24.3	25.1	35.8	38.9	47.9	48.4	49.3	53.2	55.6
62.7	72.4	73.6	76.8	83.8	95.1	97.9	99.6	102.8	108.5
128.7	133.6	144.1	147.6	150.6	151.6	152.6	164.2	166.8	178.6
$178.6^{\dagger}$	$178.6^{\dagger}$	$178.6^{\dagger}$	$178.6^{\dagger}$	$178.6^{\dagger}$	$178.6^{\dagger}$	$178.6^{\dagger}$	$178.6^{\dagger}$	$178.6^{\dagger}$	

with $\widehat{\theta}_{40}$ found to be 112.1150 and $\widehat{B}_{0.1,40} = 11.8125$, and via (2.12) and (2.13), we obtain the approximate 95% confidence intervals for θ and $B_{0.1}$ to be

$$112.1150 \pm 1.96 \times 112.1150 \times 40^{-1/2} = (77.370, 146.860)$$

and

$$11.8125 \pm 1.96 \times 112.1150 \times (-\ln 0.9) \times 40^{-1/2} = (8.152, 15.473)\,.$$

More generally, Table 2.1 presents summaries of the ML estimates of θ and $B_{0.1}$ when the data is subject to Type II censoring at the r^{th} failure; we see that the interim estimates $\widehat{\theta}_r$ and $\widehat{B}_{0.1,r}$ increase sharply when r doubles from 10 to 20, and then gradually converge to their complete counterparts. There are also some increases in estimated standard deviations at $r = 20$ (from $r = 10$), reflecting the consequence of swapping θ by a large estimate $\widehat{\theta}_{20}$ in (2.10). Otherwise, the standard deviation is generally decreasing with r, as expected. It should be noted that $\widehat{\theta}_{20}$ is the closest to $\widehat{\theta}$, but also has the largest standard deviation. One obvious point to consider is can we safely regard $\widehat{\theta}_{20}$ as a reliable guide to $\widehat{\theta}$? If so, this indicates that the experiment time would be cut from 354.4 to 55.6, an approximate 84% reduction in time. This gives us some motivation to investigate the extent to which $\widehat{\theta}_r$ provides a guide to $\widehat{\theta}$. We have used asymptotic Normality of MLE to compute the approximate 95% confidence intervals for θ, but we will also need to consider if such calculations are appropriate in a sample as small as $n = 49$; this will be further considered in Chapter 3.

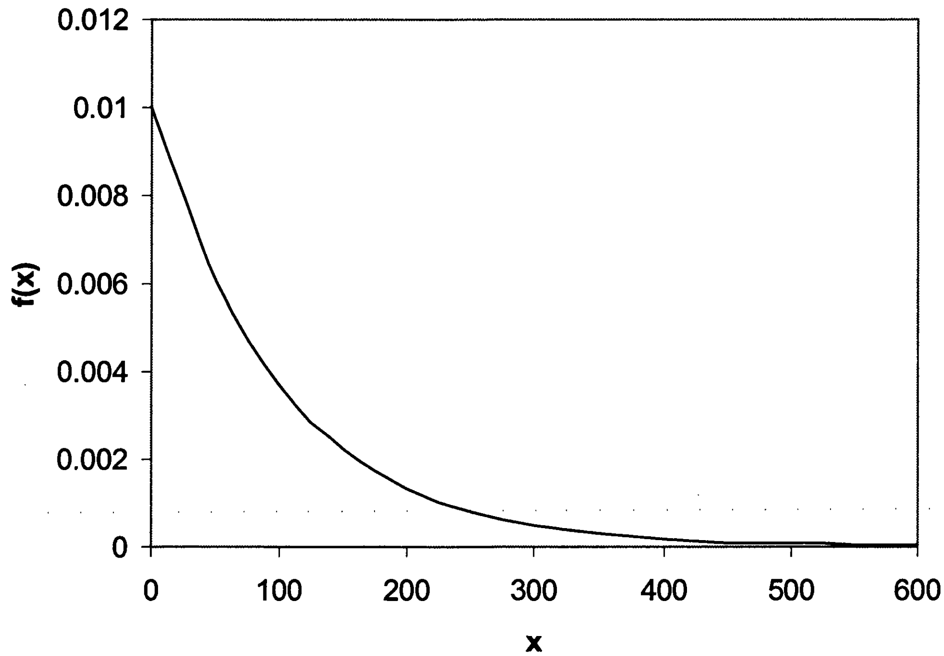

Figure 2.1: Pdf of the exponential distribution for $\theta = 100$.

Simulations

We now illustrate some results obtained from simulation experiments; this involves specifying the parameter value, sample size and censoring level, and then calculating the MLE for each sample. Here, we assume $\theta = 100$, and, for each combination of r and n, replicate 10^4 sets of data; this yields 10^4 estimates from the sampling distribution of $\widehat{\theta}_r$. Figure 2.1 shows the exponential pdf for such simulation with $\theta = 100$, while Table 2.2 summarises the observed means for $\widehat{\theta}_r$, where we see good agreement between $\widehat{\theta}_r$ and its true value, even for small n and r. In Table 2.3, we also noted good agreement between theoretical and observed standard deviations, with decreasing values when r and n increase. This is due to the fact high censoring levels imply relatively more complete failure times being observed, which provide more information about the lifetime distribution and hence a more precise estimation of θ. Moreover, it is of interest to look at the scatter plots of final estimates against interim estimates. Figure 2.2 (when $n = 50$) has wider scales than Figure 2.3 (when $n = 1000$), and both seem to suggest a link between $\widehat{\theta}$ and $\widehat{\theta}_r$. The evidence becomes clearer as r tends to n, and we will quantify the correlation between $\widehat{\theta}$ and $\widehat{\theta}_r$, and hence determine the extent to which $\widehat{\theta}_r$ can be regarded as a reliable guide to $\widehat{\theta}$.

Since $B_{0.1}$ is a linear function of θ in the exponential distribution, the study of the

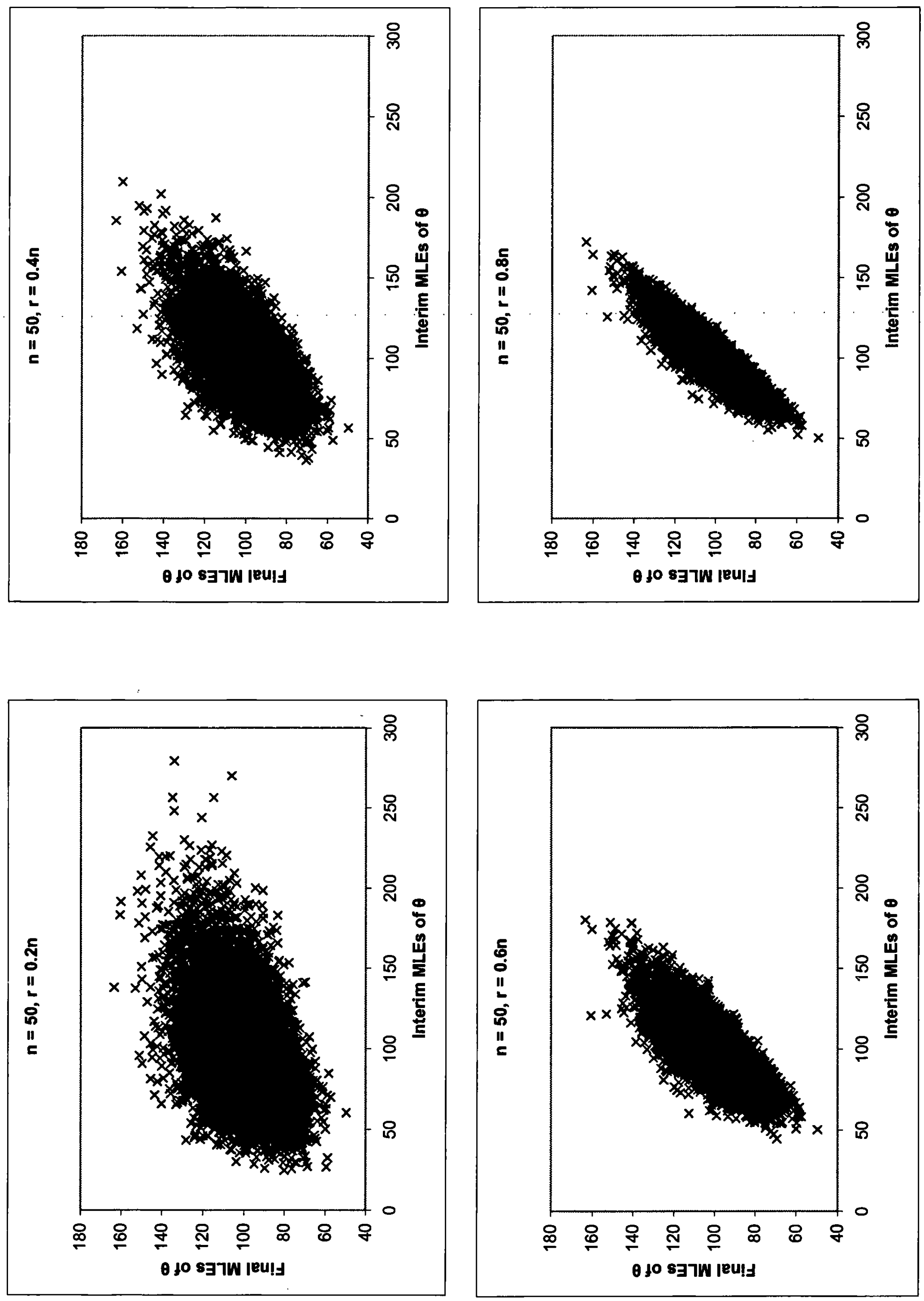

Figure 2.2: Scatter plots of $\hat{\theta}$ versus $\hat{\theta}_r$ for $n = 50$ and various r, for exponential data generated with $\theta = 100$.

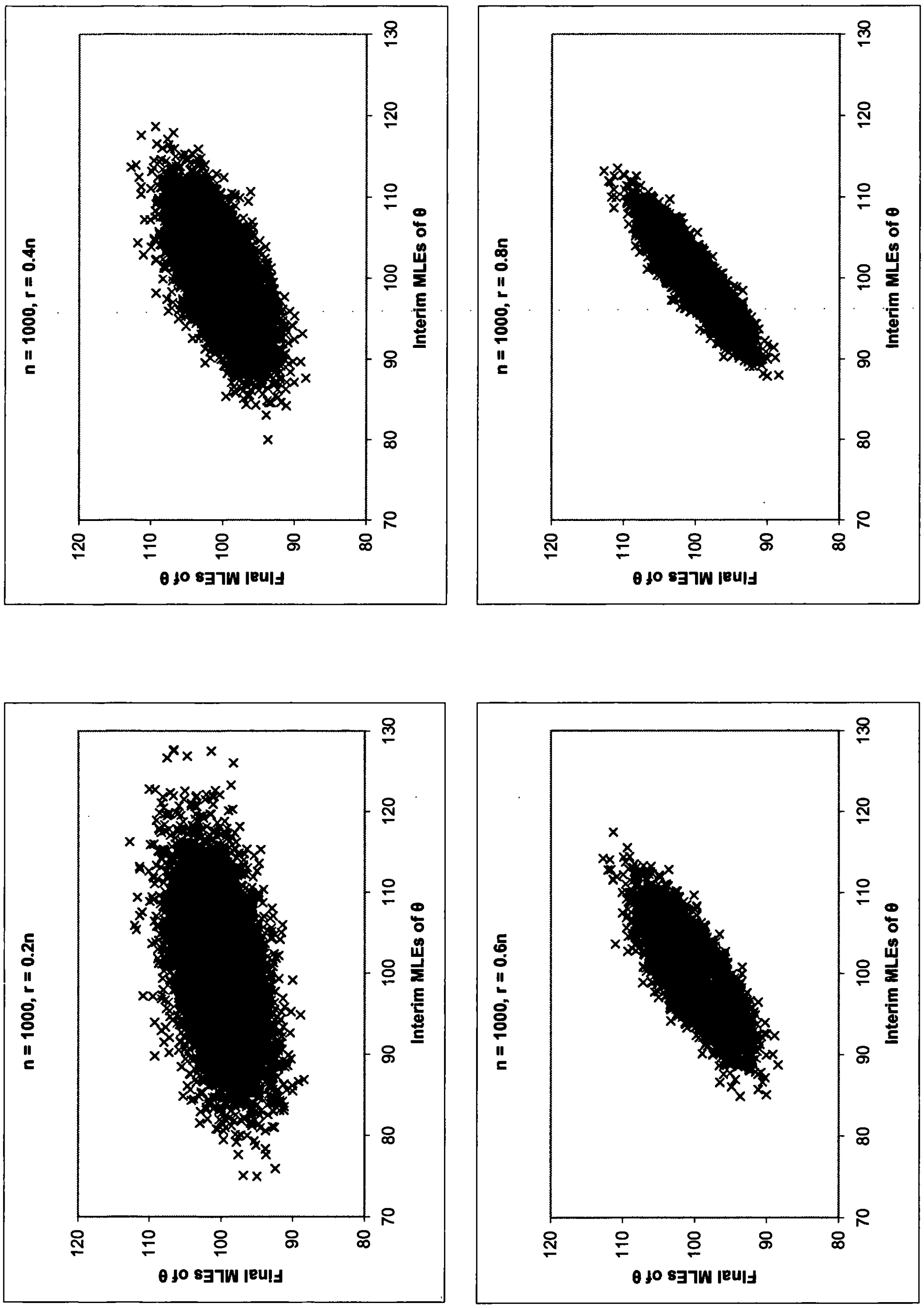

Figure 2.3: Scatter plots of $\hat{\theta}$ versus $\hat{\theta}_r$ for $n = 1000$ and various r, for exponential data generated with $\theta = 100$.

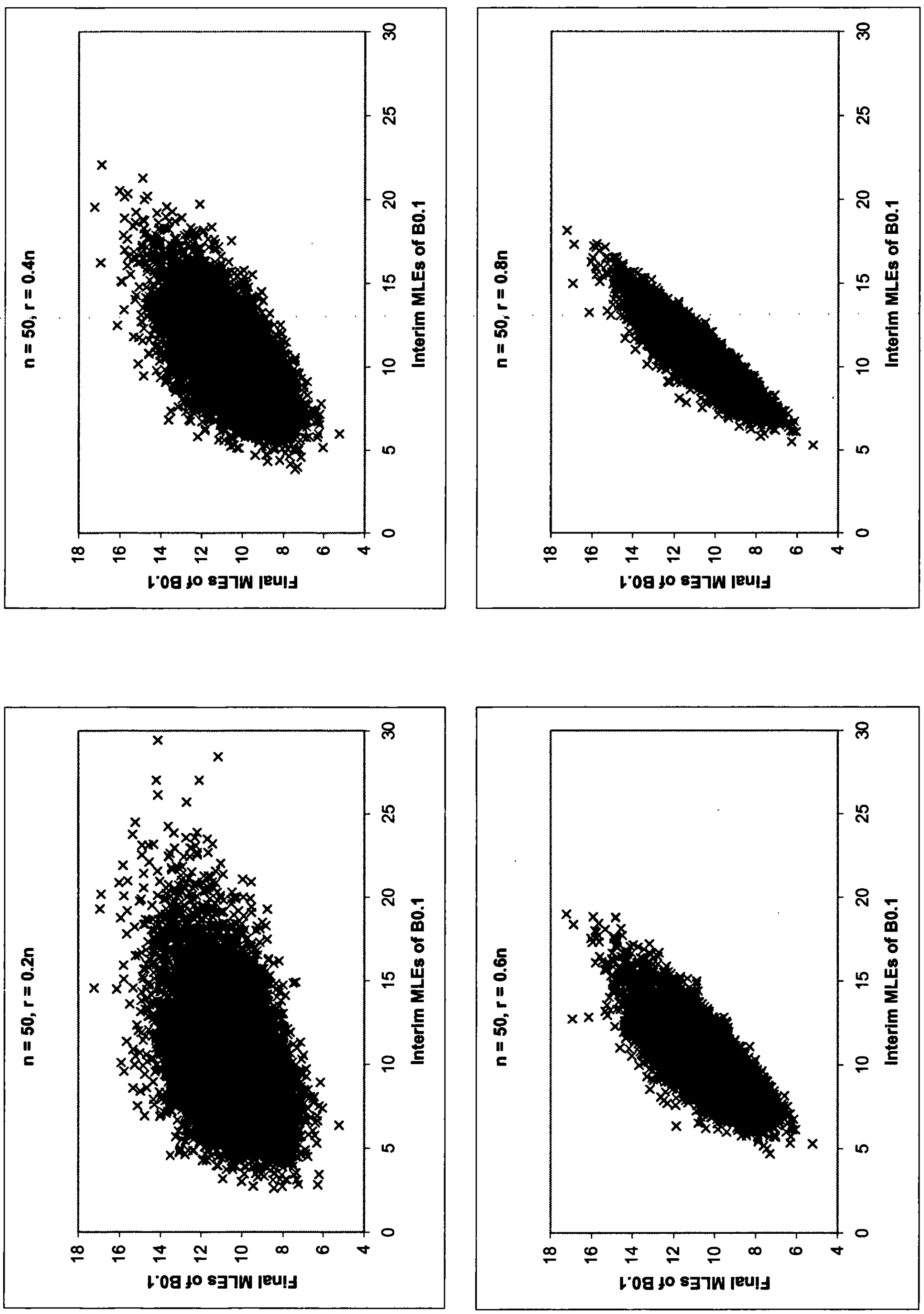

Figure 2.4: Scatter plots of $\hat{B}_{0.1}$ versus $\hat{B}_{0.1,r}$ for $n = 50$ and various r, for exponential data generated with $\theta = 100$.

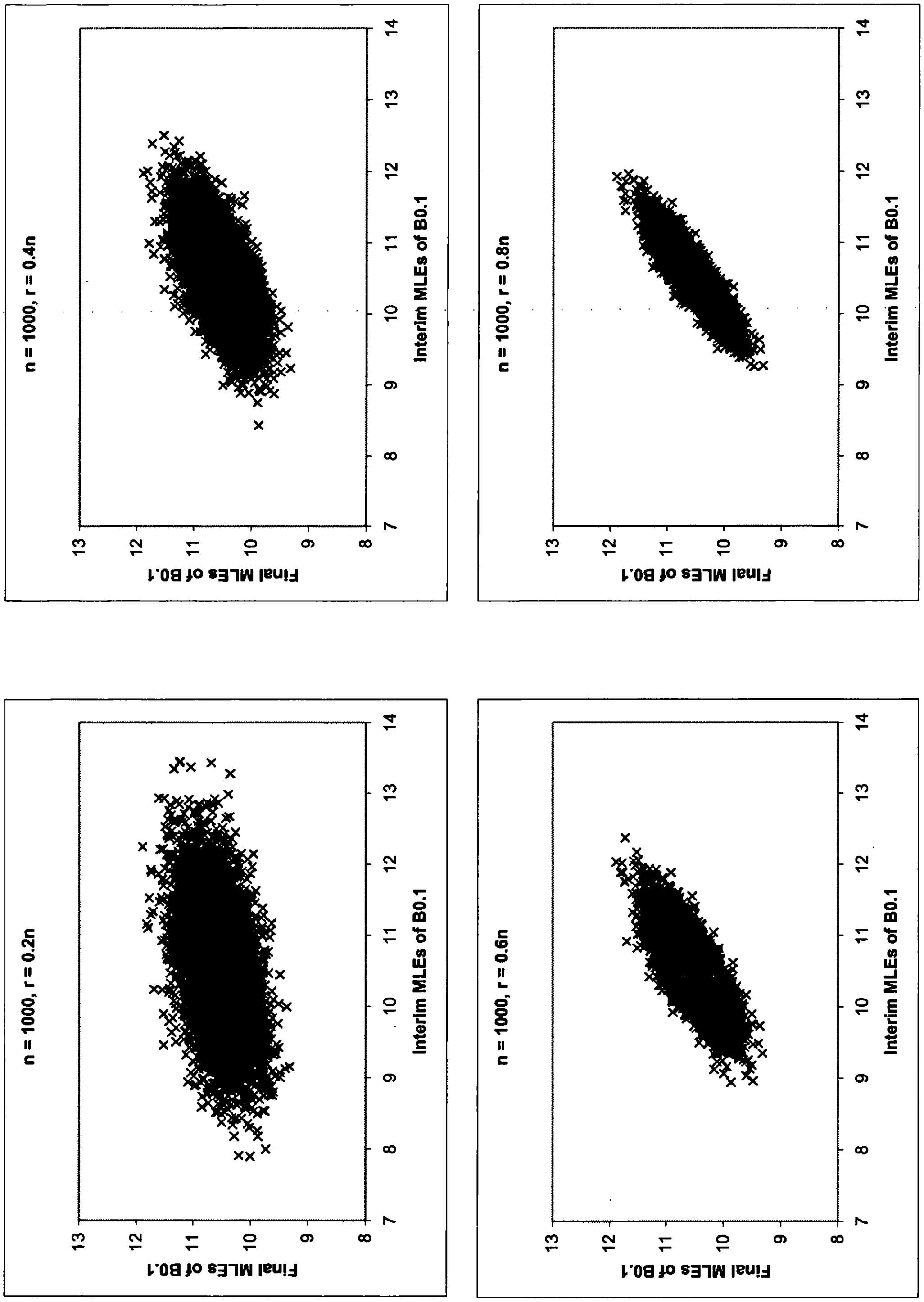

Figure 2.5: Scatter plots of $\hat{B}_{0.1}$ versus $\hat{B}_{0.1,r}$ for $n = 1000$ and various r, for exponential data generated with $\theta = 100$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	100.0000	100.1681	99.6900	100.0394	99.9731	99.9901
$0.4n$	99.7516	100.0787	99.9979	100.0798	99.9657	99.9561
$0.6n$	99.8007	99.9753	100.1715	100.0674	99.9816	99.9742
$0.8n$	99.8502	100.0242	100.2244	100.0558	99.9769	99.9704
$1.0n$	99.9283	100.1135	100.2409	100.0557	99.9924	99.9737

Table 2.2: Simulated means of $\hat{\theta}_r$ for various r, n, for exponential data generated with $\theta = 100$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	44.7214	31.6228	22.3607	7.0711	4.4721	3.1623
	45.5428	31.7004	22.2567	7.0799	4.4733	3.2014
$0.4n$	31.6228	22.3607	15.8114	5.0000	3.1623	2.2361
	32.0108	22.4806	15.8449	5.0272	3.1772	2.2415
$0.6n$	25.8199	18.2574	12.9099	4.0825	2.5820	1.8257
	25.9997	18.4661	12.9659	4.1167	2.5642	1.8266
$0.8n$	22.3607	15.8114	11.1803	3.5355	2.2361	1.5811
	22.3636	16.1072	11.2422	3.5559	2.2304	1.5835
$1.0n$	20.0000	14.1421	10.0000	3.1623	2.0000	1.4142
	20.0695	14.4386	10.0839	3.1786	1.9880	1.4244

Table 2.3: Theoretical (upper) and simulated (lower) standard deviations of $\hat{\theta}_r$ for various r, n, for exponential data generated with $\theta = 100$.

properties of $\widehat{B}_{0.1,r}$, whose true value is given by

$$100\,(-\ln 0.9) = 10.5361,$$

is essentially covered by the above study on θ. We display equivalent statistics for $\widehat{B}_{0.1,r}$ in Tables 2.4 and 2.5, together with scatter plots given in Figures 2.4 ($n = 50$) and 2.5 ($n = 1000$).

r	n					
	25	50	100	1000	2500	5000
$0.2n$	10.5361	10.5538	10.5034	10.5402	10.5332	10.5350
$0.4n$	10.5099	10.5443	10.5358	10.5445	10.5324	10.5314
$0.6n$	10.5150	10.5335	10.5541	10.5432	10.5341	10.5333
$0.8n$	10.5203	10.5386	10.5597	10.5419	10.5336	10.5329
$1.0n$	10.5285	10.5480	10.5614	10.5419	10.5352	10.5333

Table 2.4: Simulated means of $\hat{B}_{0.1,r}$ for various r, n, for exponential data generated with $\theta = 100$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	4.7119	3.3318	2.3559	0.7450	0.4712	0.3332
	4.7984	3.3400	2.3450	0.7459	0.4713	0.3373
$0.4n$	3.3318	2.3559	1.6659	0.5268	0.3332	0.2356
	3.3727	2.3686	1.6694	0.5297	0.3347	0.2362
$0.6n$	2.7204	1.9236	1.3602	0.4301	0.2720	0.1924
	2.7393	1.9456	1.3661	0.4337	0.2702	0.1925
$0.8n$	2.3559	1.6659	1.1780	0.3725	0.2356	0.1666
	2.3562	1.6971	1.1845	0.3746	0.2350	0.1668
$1.0n$	2.1072	1.4900	1.0536	0.3332	0.2107	0.1490
	2.1145	1.5213	1.0624	0.3349	0.2095	0.1501

Table 2.5: Theoretical (upper) and simulated (lower) standard deviations of $\hat{B}_{0.1,r}$ for various r, n, for exponential data generated with $\theta = 100$.

2.3 ML Estimation in the Weibull Distribution

From a computational point of view, the Weibull distribution is particularly appealing, since its cdf can be expressed explicitly as a simple function of the random variable. For accounts on the ML estimation for the Weibull parameters, see, for instance, Lawless (1982) and Cohen (1991), for both complete and censored samples. Using (2.1), the likelihood function for data drawn from a Weibull distribution is

$$L_r = \left[\prod_{i=1}^{r} \beta\theta^{-\beta} X_{i:n}^{\beta-1} \exp\left\{-\left(\frac{X_{i:n}}{\theta}\right)^{\beta}\right\}\right] \left[\exp\left\{-\left(\frac{X_{r:n}}{\theta}\right)^{\beta}\right\}\right]^{n-r}.$$

If we use subscripts f and c to indicate failed and censored items, and let, respectively,

$$\begin{aligned} S_{f,j}(k) &= \sum_{i=1}^{r} X_{i:n}^{k} \left(\ln X_{i:n}\right)^{j}, \\ S_{c,j}(k) &= (n-r) X_{r:n}^{k} \left(\ln X_{r:n}\right)^{j}, \end{aligned}$$

for real $k > 0$ and integer $j \geq 0$, then we have

$$\frac{\partial S_{*,j}(k)}{\partial k} = S_{*,j+1}(k),$$

for $* = f$ or c, and the log-likelihood function may be expressed as

$$l_r = r\ln\beta - r\beta\ln\theta + (\beta - 1) S_{f,1}(0) - \theta^{-\beta}\left\{S_{f,0}(\beta) + S_{c,0}(\beta)\right\}. \tag{2.16}$$

The MLEs can be obtained by maximising l_r, or equivalently, by finding the roots of the score functions, based on the two partial derivatives given by

$$\frac{\partial l_r}{\partial \theta} = -r\beta\theta^{-1} + \beta\theta^{-\beta-1}\left\{S_{f,0}(\beta) + S_{c,0}(\beta)\right\}, \tag{2.17}$$

and

$$\frac{\partial l_r}{\partial \beta} = r\beta^{-1} - r\ln\theta + S_{f,1}(0) - \theta^{-\beta}\left\{S_{f,1}(\beta) + S_{c,1}(\beta) - (\ln\theta)\left[S_{f,0}(\beta) + S_{c,0}(\beta)\right]\right\}. \quad (2.18)$$

Unlike exponential model, there are no analytical expressions for these roots. However, we note that if we equate (2.17) to zero, then θ_r can be expressed in terms of the data and the shape parameter β; we have

$$\theta_r = \left[\frac{S_{f,0}(\beta) + S_{c,0}(\beta)}{r}\right]^{\frac{1}{\beta}}. \quad (2.19)$$

Inserting this into (2.16) yields a profile log-likelihood given by

$$l_r^* = r\ln\beta + (\beta - 1)S_{f,1}(0) - r\ln\left[S_{f,0}(\beta) + S_{c,0}(\beta)\right] + r(\ln r - 1), \quad (2.20)$$

and a profile score function

$$\frac{dl_r^*}{d\beta} = r\beta^{-1} + S_{f,1}(0) - r\left\{\frac{S_{f,1}(\beta) + S_{c,1}(\beta)}{S_{f,0}(\beta) + S_{c,0}(\beta)}\right\}. \quad (2.21)$$

Since no closed form expression for $\widehat{\beta}_r$ exists, a numerical procedure must be used to locate the root of (2.21). As noted in Section 1.6.3, we use the Newton-Raphson approach, which requires the second-order derivative

$$\frac{d^2 l_r^*}{d\beta^2} = -r\beta^{-2} - r\left\{\frac{S_{f,2}(\beta) + S_{c,2}(\beta)}{S_{f,0}(\beta) + S_{c,0}(\beta)} - \left(\frac{S_{f,1}(\beta) + S_{c,1}(\beta)}{S_{f,0}(\beta) + S_{c,0}(\beta)}\right)^2\right\},$$

and an initial value. This starting value should be close to $\widehat{\beta}_r$, otherwise the Newton-Raphson process may fail to converge. Farnum & Booth (1997) suggest

$$\beta^{[0]} = \left[\left(1 - \frac{r}{2n}\right)V\right]^{-1}$$

as a quick initial approximation to β, where

$$V = \ln X_{r:n} - r^{-1}S_{f,1}(0)$$

may be interpreted as a measure of variation in data. With $\widehat{\beta}_r$ thus determined, θ_r is estimated from (2.19) with $\beta = \widehat{\beta}_r$. We will also require the EFI matrix of the Weibull

MLEs, which is based on second-order partial derivatives of (2.16), as listed below:

$$\frac{\partial^2 l_r}{\partial \theta^2} = r\beta\theta^{-2} - \beta(\beta+1)\theta^{-\beta-2}\{S_{f,0}(\beta) + S_{c,0}(\beta)\}, \tag{2.22}$$

$$\frac{\partial^2 l_r}{\partial \theta \partial \beta} = \frac{\partial^2 l_r}{\partial \beta \partial \theta} = -r\theta^{-1} + \theta^{-\beta-1}\left\{ \begin{array}{c} (1-\beta\ln\theta)[S_{f,0}(\beta) + S_{c,0}(\beta)] \\ +\beta[S_{f,1}(\beta) + S_{c,1}(\beta)] \end{array} \right\}, \tag{2.23}$$

$$\frac{\partial^2 l_r}{\partial \beta^2} = -r\beta^{-2} - \theta^{-\beta}\left\{ \begin{array}{c} (\ln\theta)^2[S_{f,0}(\beta) + S_{c,0}(\beta)] \\ -2(\ln\theta)[S_{f,1}(\beta) + S_{c,1}(\beta)] + S_{f,2}(\beta) + S_{c,2}(\beta) \end{array} \right\}. \tag{2.24}$$

2.3.1 Regularity and EFI Matrix

To consider the regularity of the log-likelihood function, we take expectations of the first- and second-order partial derivatives of (2.16). The form of these derivatives implies that we will need results on the expectation of various functions $g(X_{i:n})$, on the sum of these expectations, and, in particular, on expectations of the following expression:

$$\left\{\sum_{i=1}^{r} g(X_{i:n})\right\} + (n-r)g(X_{r:n}), \tag{2.25}$$

where $g(X_{i:n})$ can be any of

$$X_{i:n}, \ \ln X_{i:n}, \ X_{i:n}\ln X_{i:n}, \text{ and } X_{i:n}(\ln X_{i:n})^2.$$

Watkins & John (2006) outline a framework for deriving these expected values; the transformation

$$Z = \left(\frac{X}{\theta}\right)^{\beta} \tag{2.26}$$

links the Weibull pdf (1.30) to the standard negative exponential pdf, given by setting $\theta = 1$ in (1.27). Then, using the fact that $nZ_{1:n}$ follows the standard negative exponential distribution, Watkins & John (2006) obtain, based on Watkins (1998), the following results:

$$\sum_{i=1}^{r} E[Z_{i:n}] + (n-r)E[Z_{r:n}] = r,$$

$$\sum_{i=1}^{r} E[\ln Z_{i:n}] = -r(\gamma + \phi_1),$$

$$\sum_{i=1}^{r} E[Z_{i:n}\ln Z_{i:n}] + (n-r)E[Z_{r:n}\ln Z_{r:n}] = r(1-\gamma-\phi_1),$$

$$\sum_{i=1}^{r} E\left[Z_{i:n}(\ln Z_{i:n})^2\right] + (n-r)E\left[Z_{r:n}(\ln Z_{r:n})^2\right] = r\left\{\frac{\pi^2}{6} + \gamma^2 - 2\gamma - 2(1-\gamma)\phi_1 + \phi_2\right\},$$

where

$$\phi_k = r^{-1}\sum_{i=1}^{r}(-1)^{r-i}\binom{n}{i-1}\binom{n-i-1}{r-i}[\ln(n+1-i)]^k,$$

for $k = 1, 2$, with the convention $0^0 = 1$; see also Watkins (1998). It is then straightforward to see that

$$E[S_{f,1}(0)] = \sum_{i=1}^{r} \{\ln\theta + \beta^{-1}E[\ln Z_{i:n}]\} = r\ln\theta - r\beta^{-1}(\gamma + \phi_1),$$
$$E[S_{f,0}(\beta) + S_{c,0}(\beta)] = \theta^\beta \left\{\sum_{i=1}^{r} E[Z_{i:n}] + (n-r)E[Z_{r:n}]\right\} = r\theta^\beta,$$

and

$$\begin{aligned} E[S_{f,1}(\beta) + S_{c,1}(\beta)] &= \theta^\beta \ln\theta \left\{\sum_{i=1}^{r} E[Z_{i:n}] + (n-r)E[Z_{r:n}]\right\} \\ &\quad + \beta^{-1}\theta^\beta \left\{\sum_{i=1}^{r} E[Z_{i:n}\ln Z_{i:n}] + (n-r)E[Z_{r:n}\ln Z_{r:n}]\right\} \\ &= \theta^\beta \left[r\ln\theta + r\beta^{-1}(1-\gamma-\phi_1)\right], \end{aligned}$$

so that

$$E\left[\frac{\partial l_r}{\partial\theta}\right] = -r\beta\theta^{-1} + r\beta\theta^{-1} = 0,$$

and

$$\begin{aligned} E\left[\frac{\partial l_r}{\partial\beta}\right] &= r\beta^{-1} - r\ln\theta + r\ln\theta - r\beta^{-1}(\gamma+\phi_1) \\ &\quad -\theta^{-\beta}\left\{\theta^\beta\left[r\ln\theta + r\beta^{-1}(1-\gamma-\phi_1)\right] - r\theta^\beta\ln\theta\right\} \end{aligned}$$

also simplifies to 0, which confirms the known regularity of (2.16). For the expectations of the second-order partial derivatives in (2.22) to (2.24), Watkins & John (2006) obtain

$$\begin{aligned} E\left[\frac{\partial^2 l_r}{\partial\theta^2}\right] &= -r\beta^2\theta^{-2}, \\ E\left[\frac{\partial^2 l_r}{\partial\theta\partial\beta}\right] &= r\theta^{-1}\{1-\gamma-\phi_1\}, \\ E\left[\frac{\partial^2 l_r}{\partial\beta^2}\right] &= -r\beta^{-2}\left\{\frac{\pi^2}{6} + (1-\gamma)^2 - 2(1-\gamma)\phi_1 + \phi_2\right\}; \end{aligned}$$

these results yield the Type II censored EFI matrix as

$$\begin{aligned} \mathbf{A}_r &= \begin{pmatrix} A_{r,\theta\theta} & A_{r,\theta\beta} \\ A_{r,\theta\beta} & A_{r,\beta\beta} \end{pmatrix} \\ &= \begin{pmatrix} r\beta^2\theta^{-2} & \\ -r\theta^{-1}\{1-\gamma-\phi_1\} & r\beta^{-2}\left\{\frac{\pi^2}{6} + (1-\gamma)^2 - 2(1-\gamma)\phi_1 + \phi_2\right\} \end{pmatrix}, \end{aligned} \tag{2.27}$$

so that inverting this gives the asymptotic covariance matrix of $\left(\widehat{\theta}_r, \widehat{\beta}_r\right)$ as

$$\begin{aligned}
\mathbf{A}_r^{-1} &= \begin{pmatrix} A_r^{\theta\theta} & A_r^{\theta\beta} \\ A_r^{\theta\beta} & A_r^{\beta\beta} \end{pmatrix} \\
&= \frac{6}{r\left(\pi^2 - 6\phi_1^2 + 6\phi_2\right)} \begin{pmatrix} \beta^{-2}\theta^2\left\{\frac{\pi^2}{6} + (1-\gamma)^2 - 2(1-\gamma)\phi_1 + \phi_2\right\} & \\ \theta\left\{1-\gamma-\phi_1\right\} & \beta^2 \end{pmatrix}
\end{aligned} \tag{2.28}$$

We can now compute the asymptotic properties of the 10^{th} percentile function, defined at (1.32) as

$$B_{0.1} = \theta\left(-\ln 0.9\right)^{\frac{1}{\beta}}.$$

In contrast to the exponential distribution, consideration of $B_{0.1}$ and its estimator here are more complicated, as we need to linearise the above expression. From (2.3), we can obtain the linear approximation

$$\widehat{B}_{0.1,r} \simeq B_{0.1} + \begin{pmatrix} b_\theta & b_\beta \end{pmatrix} \begin{pmatrix} \widehat{\theta}_r - \theta \\ \widehat{\beta}_r - \beta \end{pmatrix},$$

where

$$\begin{pmatrix} b_\theta \\ b_\beta \end{pmatrix} = \begin{pmatrix} \frac{\partial B_{0.1}}{\partial\theta} \\ \frac{\partial B_{0.1}}{\partial\beta} \end{pmatrix} = \begin{pmatrix} \left(-\ln 0.9\right)^{\frac{1}{\beta}} \\ -\theta\beta^{-2}\left(-\ln 0.9\right)^{\frac{1}{\beta}}\ln\left(-\ln 0.9\right) \end{pmatrix}. \tag{2.29}$$

Thus, on taking expectation, we have, for large samples,

$$E\left[\widehat{B}_{0.1,r}\right] \simeq B_{0.1} + b_\theta E\left[\widehat{\theta}_r - \theta\right] + b_\beta E\left[\widehat{\beta}_r - \beta\right] = B_{0.1},$$

and variance, from (2.4), given by

$$Var\left(\widehat{B}_{0.1,r}\right) \simeq \begin{pmatrix} b_\theta & b_\beta \end{pmatrix} \mathbf{A}_r^{-1} \begin{pmatrix} b_\theta \\ b_\beta \end{pmatrix} = b_\theta^2 A_r^{\theta\theta} + 2b_\theta b_\beta A_r^{\theta\beta} + b_\beta^2 A_r^{\beta\beta}.$$

2.3.2 Asymptotic Properties of the MLEs

Here we are interested in the asymptotic properties of $\widehat{\theta}_r, \widehat{\beta}_r$ and $\widehat{B}_{0.1,r}$. From the asymptotic Normality of MLE, $\left(\widehat{\theta}_r, \widehat{\beta}_r\right)'$ is bivariate Normal with mean $(\theta, \beta)'$ and covariance matrix $\mathbf{A}_r^{-1}$ from (2.28). Consequently, individual approximate 95% confidence intervals for θ and β are, respectively,

$$\widehat{\theta}_r \pm 1.96\sqrt{A_r^{\theta\theta}},$$

and

$$\widehat{\beta}_r \pm 1.96\sqrt{A_r^{\beta\beta}}.$$

Since, asymptotically,

$$\begin{pmatrix} \theta - \widehat{\theta}_r \\ \beta - \widehat{\beta}_r \end{pmatrix}' \mathbf{A}_r \begin{pmatrix} \theta - \widehat{\theta}_r \\ \beta - \widehat{\beta}_r \end{pmatrix} \simeq \chi^2_2,$$

where the chi-square variate with 2 degrees of freedom, χ^2_2, is equivalent to an exponential variate with mean 2, Watkins (2004) illustrates that due to the convergence of observed and expected Fisher information matrices, an approximate $100\,(1-\lambda)\,\%$ confidence region for (θ, β) can be obtained by calculating the ellipse

$$\begin{pmatrix} \theta - \widehat{\theta}_r \\ \beta - \widehat{\beta}_r \end{pmatrix}' \widehat{\mathbf{J}}_r \begin{pmatrix} \theta - \widehat{\theta}_r \\ \beta - \widehat{\beta}_r \end{pmatrix} = -2\ln\lambda$$

where $\mathbf{J}_r$ is the observed Fisher information matrix. This result depends on unknown θ and β, hence in practice we use the estimates $\widehat{\theta}_r$ and $\widehat{\beta}_r$, and the notation $\widehat{\mathbf{J}}_r$.

2.3.3 Complete Sample

For later convenience, it is suitable to summarise here some results for the complete case; when all n items are observed to fail, we have $r = n$ so that all the terms associated with subscript c will disappear from the above consideration. The complete likelihood function is given by

$$L = \beta^n \theta^{-n\beta} \prod_{i=1}^{n} X_i^{\beta-1} \exp\left\{ -\sum_{i=1}^{n} \left(\frac{X_i}{\theta} \right)^{\beta} \right\}$$

from which its log-likelihood function is

$$l = n\ln\beta - n\beta\ln\theta + (\beta - 1)\,S_1\,(0) - \theta^{-\beta} S_0\,(\beta)\,, \tag{2.30}$$

where

$$S_j\,(k) = \sum_{i=1}^{n} X_i^k\,(\ln X_i)^j$$

and, similarly,

$$\frac{\partial S_j\,(k)}{\partial k} = S_{j+1}\,(k)\,.$$

As per previously, the relevant components of the score functions are

$$\frac{\partial l}{\partial \theta} = -n\beta\theta^{-1} + \beta\theta^{-\beta-1} S_0\,(\beta) \tag{2.31}$$

and

$$\frac{\partial l}{\partial \beta} = n\beta^{-1} - n\ln\theta + S_1\,(0) - \theta^{-\beta}\left\{ S_1\,(\beta) - (\ln\theta) S_0\,(\beta) \right\}. \tag{2.32}$$

Watkins (1998) computes expectations of second derivatives of (2.30) given by

$$\begin{aligned}
\frac{\partial^2 l}{\partial \theta^2} &= n\beta\theta^{-2} - \beta(\beta+1)S_0(\beta), \\
\frac{\partial^2 l}{\partial\theta\partial\beta} &= \frac{\partial^2 l}{\partial\beta\partial\theta} = -n\theta^{-1} + \theta^{-\beta-1}\{(1-\beta\ln\theta)S_0(\beta) + \beta S_1(\beta)\}, \\
\frac{\partial^2 l}{\partial\beta^2} &= -n\beta^{-2} - \theta^{-\beta}\left\{(\ln\theta)^2 S_0(\beta) - 2\ln\theta S_1(\beta) + S_2(\beta)\right\},
\end{aligned}$$

to give the complete EFI matrix as

$$\mathbf{A} = \begin{pmatrix} A_{\theta\theta} & A_{\theta\beta} \\ A_{\theta\beta} & A_{\beta\beta} \end{pmatrix} = \begin{pmatrix} n\beta^2\theta^{-2} & \\ -n\theta^{-1}\{1-\gamma\} & n\beta^{-2}\left\{\frac{\pi^2}{6} + (1-\gamma)^2\right\} \end{pmatrix}. \tag{2.33}$$

We invert $\mathbf{A}$ to obtain the complete covariance matrix:

$$\mathbf{A}^{-1} = \begin{pmatrix} A^{\theta\theta} & A^{\theta\beta} \\ A^{\theta\beta} & A^{\beta\beta} \end{pmatrix} = \frac{6}{n\pi^2}\begin{pmatrix} \beta^{-2}\theta^2\left\{\frac{\pi^2}{6} + (1-\gamma)^2\right\} & \\ \theta\{1-\gamma\} & \beta^2 \end{pmatrix}, \tag{2.34}$$

and use this result to compute the variance of $\widehat{B}_{0.1}$ from (2.4).

2.3.4 Numerical Examples

Ball Bearings Data

We can illustrate ML estimation of Weibull parameters using the classic ball bearings data from Table 1.2. Table 2.6 summarises the estimates calculated for various censoring numbers, and we note that $\widehat{\theta}_r, \widehat{\beta}_r$ and $\widehat{B}_{0.1,r} = \widehat{\theta}_r(-\ln 0.9)^{\frac{1}{\widehat{\beta}_r}}$ converge on their complete values in upwards, downwards, and downwards directions respectively, as r approaches $n = 23$. Note also that $\widehat{\beta}_r > 1$ for all r considered; this indicates that the failure rate is increasing over time. For the estimated standard deviations, $\widehat{sd}(\widehat{\theta}_r)$ decreases as r rises but shows a steep increase from $r = 20$ to 23, in part due to the increase in $\widehat{\theta}$ over $\widehat{\theta}_{20}$. $\widehat{sd}(\widehat{\beta}_r)$ decreases consistently. In contrast, $\widehat{sd}(\widehat{B}_{0.1,r})$ increases from $r = 8$ to 12 and then reduces slightly. Overall, in $\theta, \beta, B_{0.1}$, the percentages of change in the values of interim estimates for $r = 8$ to 12 seem to be more significant than any other jump in r, although the jump sizes are not the same throughout. We recall from Figure 1.2 that the P-P plot for the uncensored ball bearings data based on Weibull with $\widehat{\theta} = 81.8783$ and $\widehat{\beta} = 2.1021$ deviates from the straight line for data values around $X_{8:23}$ to $X_{16:23}$, but fits the line well at both ends. This might have some influence on the values of interim estimates we thus obtained, especially those calculated at $r = 12$ and 16. In addition, all interim 95% confidence limits appear to enclose their final estimates, showing consistency between interim and final results, but it remains to check if such calculations are appropriate for a sample of size $n = 23$; this will be discussed in more details in next chapter.

r	8	12	16	20	23
$X_{r:23}$	51.84	67.80	84.12	105.84	173.40
$\widehat{\theta}_r$	67.6415	75.2168	76.6960	78.9674	81.8783
$\widehat{sd}(\widehat{\theta}_r)$	9.6143	8.9694	7.8079	7.5906	8.5521
95% CIs	48.797,86.485	57.637,92.797	61.393,91.999	64.090,93.845	65.116,98.640
$\widehat{\beta}_r$	3.2280	2.6241	2.4695	2.3539	2.1021
$\widehat{sd}(\widehat{\beta}_r)$	1.0378	0.6797	0.5382	0.4381	0.3417
95% CIs	1.194,5.262	1.292,3.956	1.415,3.524	1.495,3.213	1.432,2.772
$\widehat{B}_{0.1,r}$	33.6860	31.9063	30.8329	30.3563	28.0694
$\widehat{sd}(\widehat{B}_{0.1,r})$	5.8179	6.6261	6.5860	6.5203	6.4367
95% CIs	22.283,45.089	18.919,44.893	17.924,43.741	17.576,43.136	15.454,40.685

Table 2.6: Summaries of the Weibull MLEs calculated at various r for the ball bearings data.

Simulations

The theoretical standard deviations obtained from the EFI matrix, see (2.28), need to checked against finite simulated samples to assure the suitability of the asymptotic approximations. These checks should also be extended to the asymptotic Normal distribution of the MLEs, which we will study in the next chapter.

We assume an increasing hazard function (so $\beta > 1$) since, as mostly encountered in practice, the electromechanical items are more likely to fail as time goes on. We generate Weibull data with $\theta = 100, \beta = 2$, and then compute the MLEs using the procedures described above. This is repeated 10^4 times to give 10^4 estimates from the sampling distribution of $\left(\widehat{\theta}_r, \widehat{\beta}_r\right)$. Figure 2.6 illustrates the Weibull pdf when $\theta = 100$ and $\beta = 2$; this distribution is bell-shaped, indicating increasing hazard over time. We note that other shape parameter values are possible; for example, as illustrated in Figure 1.5, increasing β gives a narrower pdf, implying that the items "wear out" sooner.

First, we assess the agreement between the simulated means of $\left(\widehat{\theta}_r, \widehat{\beta}_r\right)$ and their true values. As shown in Tables 2.7 and 2.8, we see some discrepancies between the true and observed values for small r and n, which improve as r and n increase. In addition, Tables 2.9 and 2.10 summarise the theoretical and simulated standard deviations for $\widehat{\theta}_r$ and $\widehat{\beta}_r$ respectively. As expected, the standard deviations reduce as r increases. We see, particularly for $\widehat{\beta}_r$, at early censoring levels there are large discrepancies between the theoretical and simulated values, but agreement improves as more items are left to fail. On the other hand, when we keep the censoring level fixed and vary the sample size, we see that the simulated standard deviations are closer to their theoretical counterparts as n increases. We also provide scatter plots of final estimates against interim estimates when $n = 50$ (Figures 2.7, 2.8, 2.9 and 2.10), where, in general, there are some connections among the four MLEs as r increases. The linear correlation is particularly evident between $\widehat{\theta}$ and $\widehat{\theta}_r$, and $\widehat{\beta}$ and $\widehat{\beta}_r$, but is less obvious between $\widehat{\theta}$ and $\widehat{\beta}_r$, and $\widehat{\beta}$ and $\widehat{\theta}_r$. Again, it would be useful to know, numerically, how much information about $\widehat{\theta}, \widehat{\beta}$ we could gain from the interim estimates

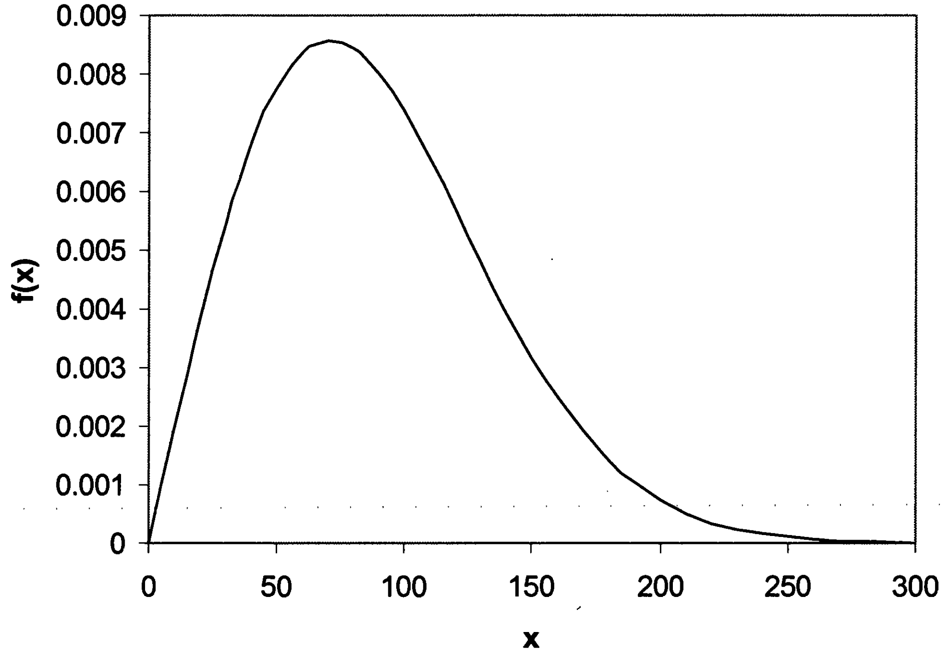

Figure 2.6: Pdf of the Weibull distribution for $\theta = 100$ and $\beta = 2$.

$\widehat{\theta}_r, \widehat{\beta}_r$.

As for the estimators $\widehat{\theta}_r$ and $\widehat{\beta}_r$, Tables 2.11 and 2.12 provide the corresponding statistics for $\widehat{B}_{0.1,r}$. Since we have used simulated data, we can compare these estimates with the true value given by

$$B_{0.1} = 100\,(-\ln 0.9)^{\frac{1}{2}} = 32.4593.$$

We see that, for all n and r, the similarity between the simulated sample mean of $\widehat{B}_{0.1,r}$ and its true value is generally good; the highest relative margin of error between theory and simulation is 7% (when $r = 10, n = 25$), but is generally less than 2% in most cases. We note that the standard deviations decrease as r increases. We also see excellent agreement between simulated and theoretical standard deviations of $\widehat{B}_{0.1,r}$, even with very early

r	n					
	25	50	100	1000	2500	5000
$0.2n$	88.1007	93.7863	97.2542	99.5719	99.9308	99.9617
$0.4n$	95.6636	97.5939	98.9468	99.8468	99.9850	99.9879
$0.6n$	98.0182	98.8529	99.5528	99.9658	99.9780	99.9831
$0.8n$	99.1820	99.4875	99.8805	99.9732	99.9844	99.9904
$1.0n$	99.8106	99.8111	100.0393	99.9995	99.9913	99.9934

Table 2.7: Simulated means of $\hat{\theta}_r$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 2$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	3.2060	2.4784	2.2102	2.0218	2.0060	2.0032
$0.4n$	2.4367	2.2038	2.0964	2.0100	2.0022	2.0012
$0.6n$	2.2505	2.1236	2.0611	2.0051	2.0017	2.0010
$0.8n$	2.1663	2.0824	2.0413	2.0042	2.0012	2.0007
$1.0n$	2.1137	2.0567	2.0296	2.0024	2.0006	2.0004

Table 2.8: Simulated means of $\hat{\beta}_r$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 2$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	37.1703	27.3761	19.8032	6.4021	4.0552	2.8690
	38.4271	27.9356	20.2001	6.4176	4.0930	2.9025
$0.4n$	19.2607	13.8102	9.8385	3.1332	1.9825	1.4021
	19.2340	13.7300	9.7953	3.1222	2.0001	1.4162
$0.6n$	13.3729	9.4857	6.7187	2.1280	1.3460	0.9518
	13.2336	9.4461	6.7349	2.1111	1.3430	0.9585
$0.8n$	11.1975	7.9158	5.5967	1.7696	1.1192	0.7914
	11.1167	7.9040	5.5842	1.7659	1.1289	0.7950
$1.0n$	10.5293	7.4454	5.2647	1.6648	1.0529	0.7445
	10.5112	7.4529	5.2606	1.6629	1.0653	0.7461

Table 2.9: Theoretical (upper) and simulated (lower) standard deviations of $\hat{\theta}_r$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 2$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	0.8079	0.5911	0.4262	0.1374	0.0870	0.0615
	2.0788	0.9105	0.5148	0.1399	0.0884	0.0625
$0.4n$	0.5756	0.4140	0.2955	0.0942	0.0596	0.0422
	0.8652	0.5044	0.3227	0.0941	0.0603	0.0426
$0.6n$	0.4590	0.3278	0.2331	0.0741	0.0469	0.0331
	0.5888	0.3727	0.2463	0.0737	0.0472	0.0336
$0.8n$	0.3807	0.2708	0.1921	0.0609	0.0385	0.0272
	0.4558	0.2947	0.1990	0.0613	0.0388	0.0273
$1.0n$	0.3119	0.2205	0.1559	0.0493	0.0312	0.0221
	0.3547	0.2334	0.1599	0.0497	0.0312	0.0221

Table 2.10: Theoretical (upper) and simulated (lower) standard deviations of $\hat{\beta}_r$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 2$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	33.9274	33.2509	32.9350	32.5044	32.4602	32.4616
$0.4n$	34.7536	33.6599	33.1188	32.5241	32.4672	32.4641
$0.6n$	34.6446	33.5890	33.0936	32.5117	32.4703	32.4666
$0.8n$	34.4116	33.4514	33.0200	32.5137	32.4714	32.4656
$1.0n$	34.1427	33.2987	32.9536	32.4975	32.4661	32.4634

Table 2.11: Simulated means of $\hat{B}_{0.1,r}$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 2$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	8.8795	6.2345	4.3937	1.3854	0.8760	0.6194
	8.9479	6.2477	4.3708	1.3824	0.8852	0.6243
$0.4n$	8.6325	6.1211	4.3355	1.3733	0.8687	0.6143
	8.8804	6.1997	4.3314	1.3732	0.8780	0.6196
$0.6n$	8.3738	5.9417	4.2094	1.3335	0.8435	0.5965
	8.6531	6.0280	4.2180	1.3380	0.8541	0.6021
$0.8n$	8.0205	5.6867	4.0269	1.2751	0.8065	0.5703
	8.3324	5.7571	4.0370	1.2852	0.8150	0.5733
$1.0n$	7.5037	5.3059	3.7519	1.1864	0.7504	0.5306
	7.7938	5.3526	3.7564	1.1983	0.7566	0.5337

Table 2.12: Theoretical (upper) and simulated (lower) standard deviations of $\hat{B}_{0.1,r}$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 2$.

censoring and regardless of the discrepancies we have observed in $\widehat{\beta}_r$. Figure 2.11 shows the scatter plots of $\widehat{B}_{0.1}$ against $\widehat{B}_{0.1,r}$ for various r when $n = 50$. Strikingly, the linear behaviour here is quite strong even for low censoring levels, and is also more evident than the plots for the MLEs of θ and β. We will investigate this behaviour in more details in Chapter 5.

2.4 ML Estimation in the Burr Distribution

The applicability of the Burr distribution in simulation modelling is enhanced by the fact that its cdf (and hence its quantile function) exists in simple closed form, from which random samples can be generated by the inverse transformation method. Some of the major contributors to the development of the theory underlying ML estimation for this distribution have been, for two-parameter case, Wingo (1983) and Watkins (1997) with complete data, Wingo (1993) and Wang *et al.* (1996) with Type II censored data; for three-parameter case, Watkins (1997) with complete data, Watkins (1999) with both complete data and censored data. Specifically, as far as our case of interest - two-parameter Burr subject to Type II censoring - is concerned, Wingo (1993) and Wang *et al.* (1996) have provided only the observed Fisher information; here, we will derive the EFI matrix explicitly. Without loss of generality, the likelihood function of a Type II censored sample drawn from the Burr distribution is

$$L_r = \left[\prod_{i=1}^{r} \alpha\tau X_{i:n}^{\tau-1}\left(1+X_{i:n}^{\tau}\right)^{-(\alpha+1)}\right]\left[1+X_{r:n}^{\tau}\right]^{-\alpha(n-r)},$$

and the log-likelihood is

$$l_r = r\ln\alpha + r\ln\tau + (\tau-1)\,S_{f,1}(0) - (\alpha+1)\,T_f - \alpha T_c, \tag{2.35}$$

where we now define

$$\begin{aligned} T_f &= \sum_{i=1}^{r}\ln(1+X_{i:n}^{\tau}), \\ T_c &= (n-r)\ln(1+X_{r:n}^{\tau}). \end{aligned}$$

The remaining notations are concerned with the derivatives of T_f and T_c:

$$\begin{aligned} T_{f,abc} &= \sum_{i=1}^{r}\frac{(X_{i:n}^{\tau})^a\,(\ln X_{i:n})^b}{(1+X_{i:n}^{\tau})^c}, \\ T_{c,abc} &= (n-r)\frac{(X_{r:n}^{\tau})^a\,(\ln X_{r:n})^b}{(1+X_{r:n}^{\tau})^c}, \end{aligned}$$

so that (for $* = f$ or c)

$$\frac{\partial^k T_*}{\partial\tau^k} = T_{*,11k}$$

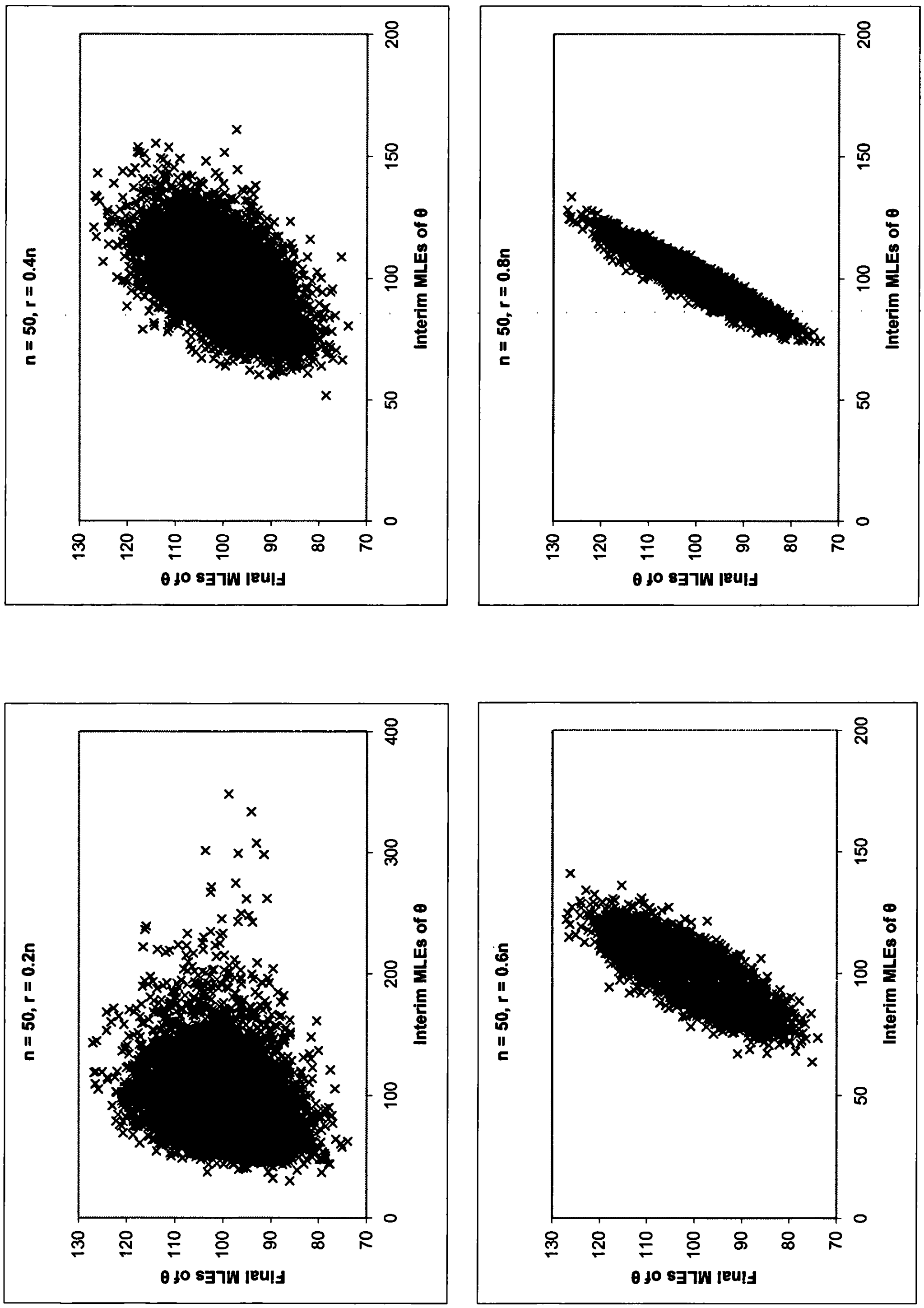

Figure 2.7: Scatter plots of $\hat{\theta}$ versus $\hat{\theta}_r$ for $n = 50$ and various r, for Weibull data generated with $\theta = 100, \beta = 2$.

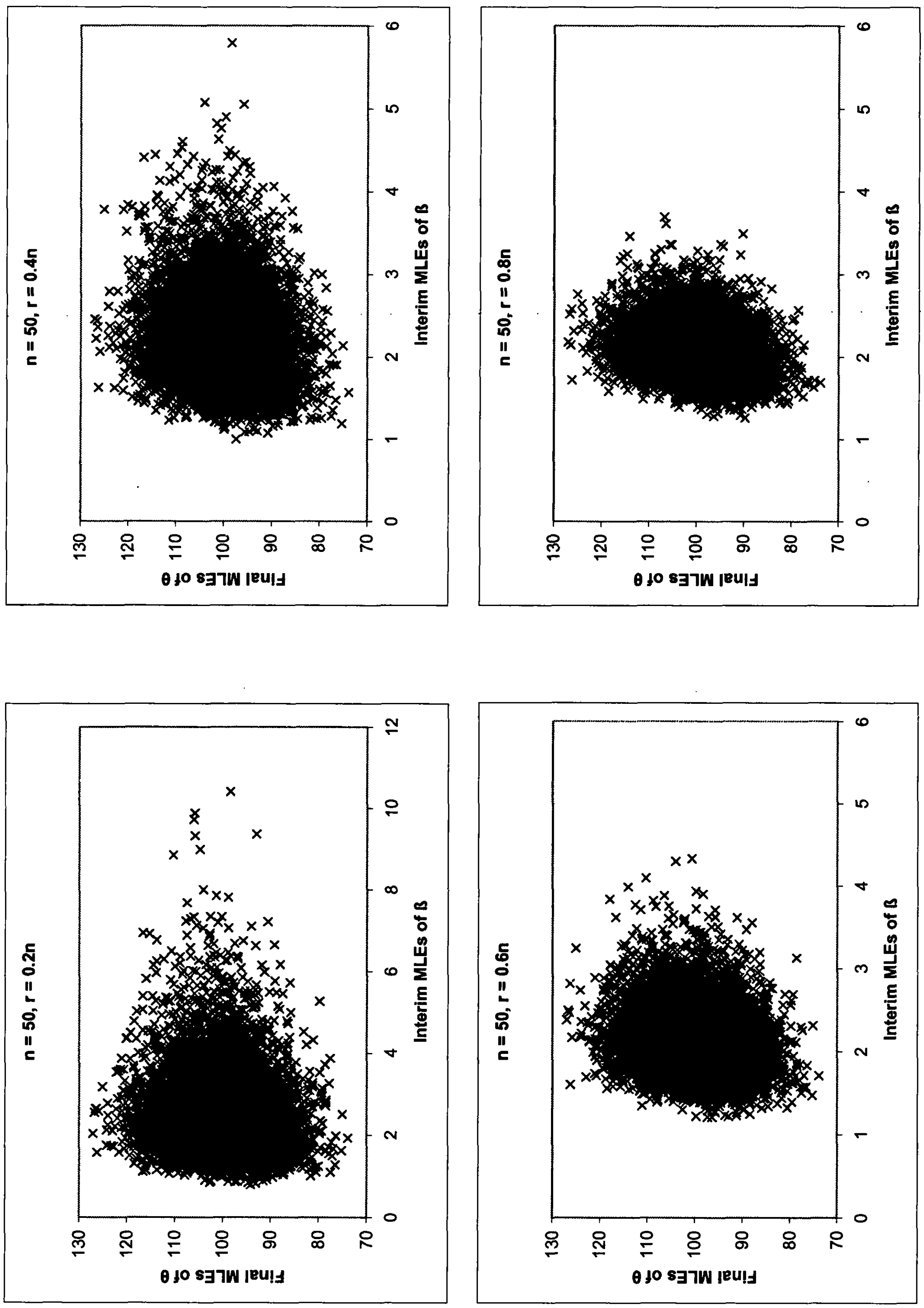

Figure 2.8: Scatter plots of $\hat{\theta}$ versus $\hat{\beta}_r$ for $n = 50$ and various r, for Weibull data generated with $\theta = 100, \beta = 2$.

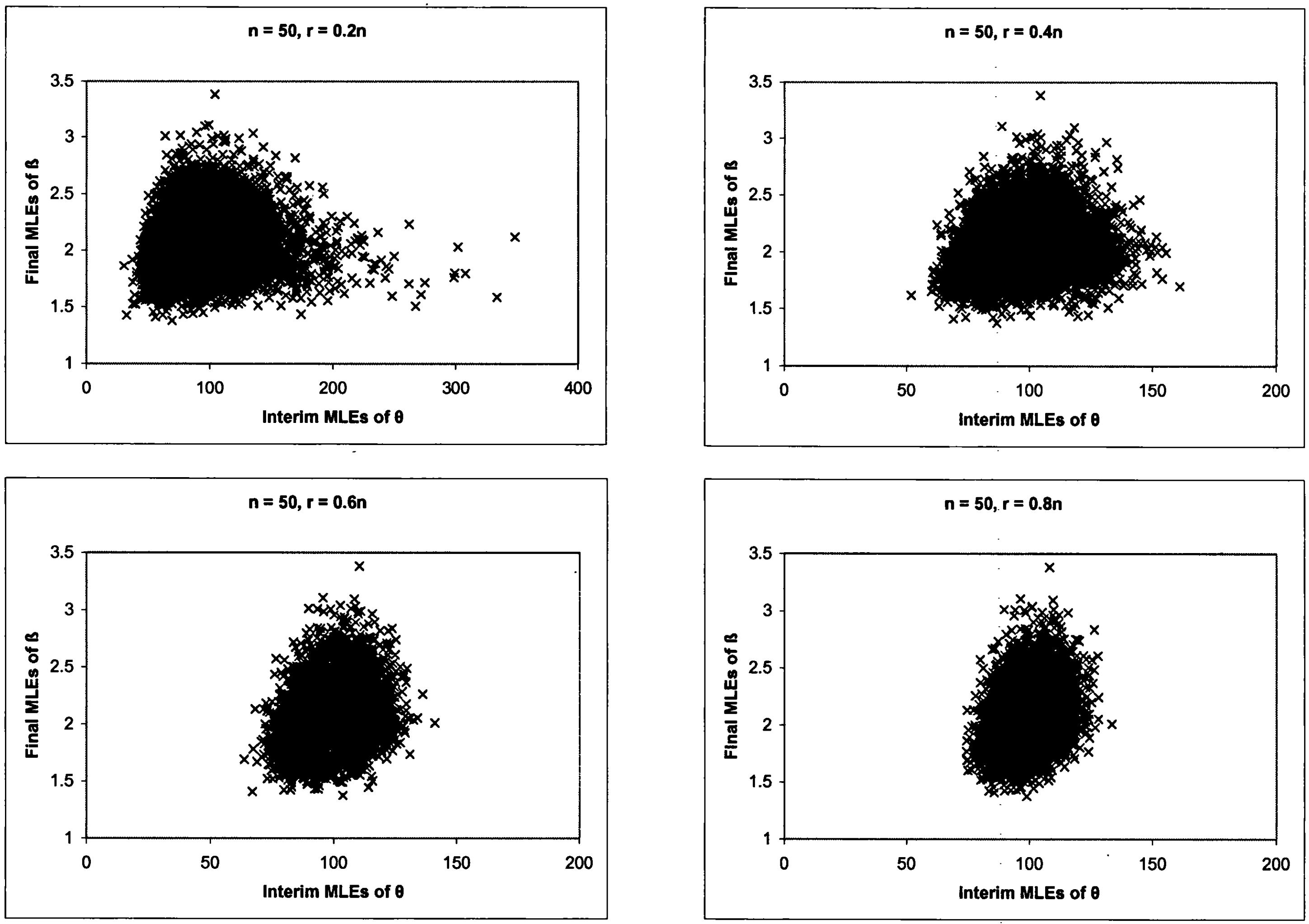

Figure 2.9: Scatter plots of $\hat{\beta}$ versus $\hat{\theta}_r$ for $n = 50$ and various r, for Weibull data generated with $\theta = 100, \beta = 2$.

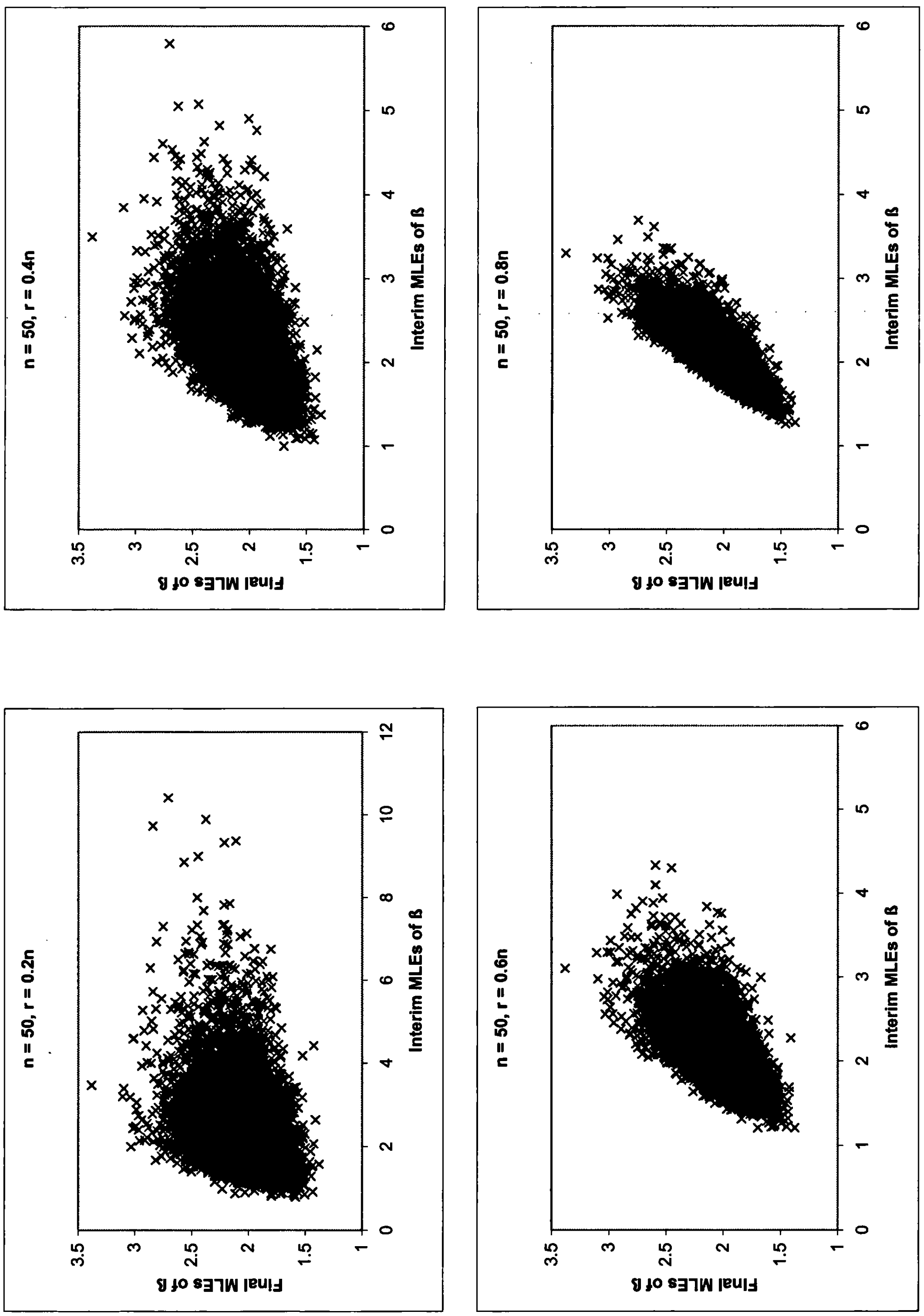

Figure 2.10: Scatter plots of $\hat{\beta}$ versus $\hat{\beta}_r$ for $n = 50$ and various r, for Weibull data generated with $\theta = 100, \beta = 2$.

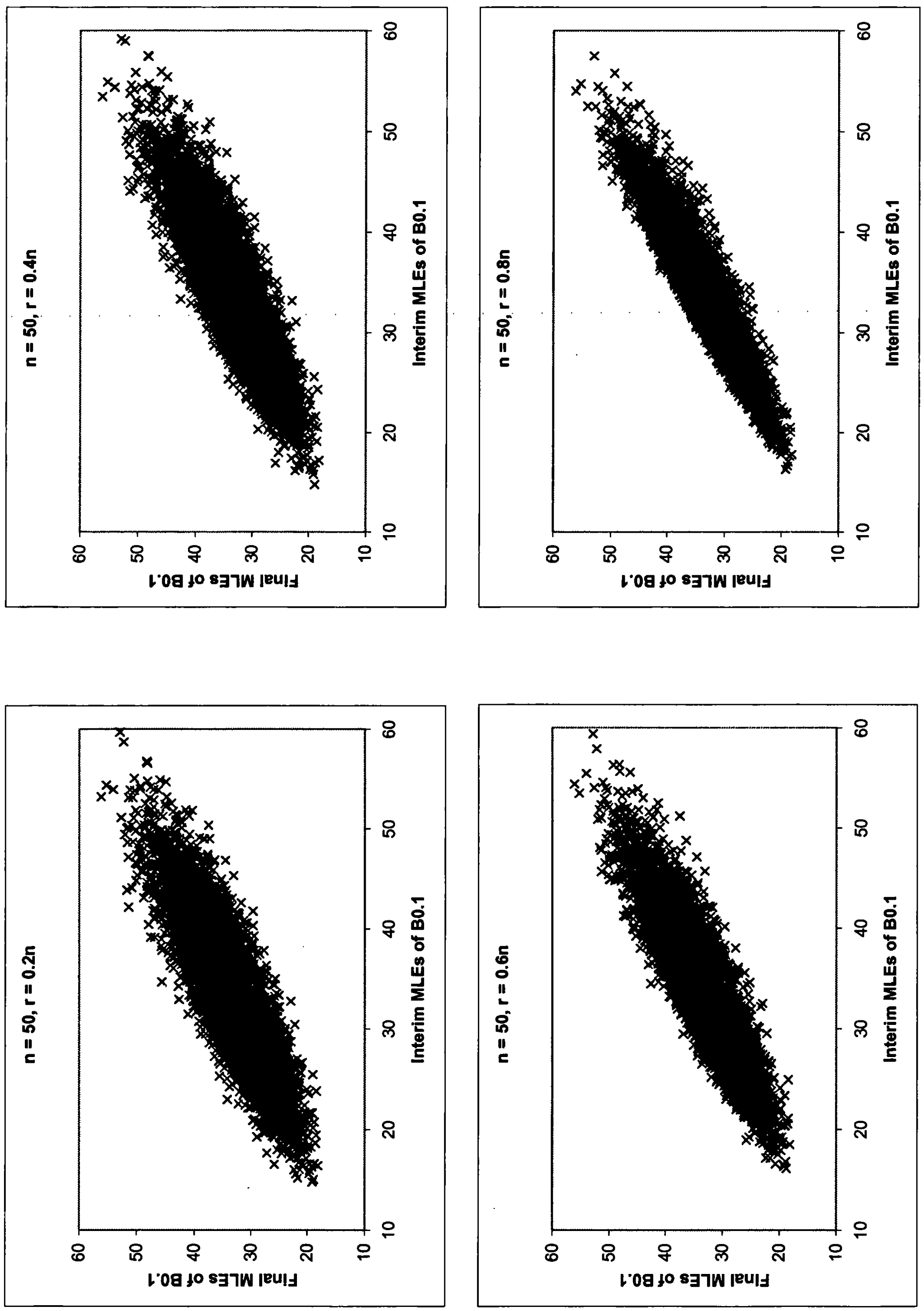

Figure 2.11: Scatter plots of $\hat{B}_{0.1}$ versus $\hat{B}_{0.1,r}$ for $n = 50$ and various r, for Weibull data generated with $\theta = 100, \beta = 2$.

when $k = 1, 2$. The conditions necessary for the existence of a stationary point of (2.35) require the score functions

$$\frac{\partial l_r}{\partial \alpha} = r\alpha^{-1} - T_f - T_c = 0, \tag{2.36}$$

$$\frac{\partial l_r}{\partial \tau} = r\tau^{-1} + S_{f,1}(0) - (\alpha+1)T_{f,111} - \alpha T_{c,111} = 0, \tag{2.37}$$

hold simultaneously. We see that the solution of (2.36) provides

$$\alpha_r = \frac{r}{T_f + T_c}, \tag{2.38}$$

so that inserting this into (2.35) yields the profile log-likelihood

$$l_r^* = r \ln \tau + (\tau - 1) S_{f,1}(0) - T_f - r \ln(T_f + T_c) + r \{\ln r - 1\}, \tag{2.39}$$

with first and second derivatives

$$\frac{dl_r^*}{d\tau} = r\tau^{-1} + S_{f,1}(0) - T_{f,111} - r \left\{ \frac{T_{f,111} + T_{c,111}}{T_f + T_c} \right\} \tag{2.40}$$

and

$$\frac{d^2 l_r^*}{d\tau^2} = -r\tau^{-2} - T_{f,122} - r \left\{ \frac{T_{f,122} + T_{c,122}}{T_f + T_c} - \left(\frac{T_{f,111} + T_{c,111}}{T_f + T_c} \right)^2 \right\}.$$

We can now find the roots of (2.40) using the Newton-Raphson approach. With $\widehat{\tau}_r$ thus calculated, MLE of α_r can be determined from (2.38) with $\tau = \widehat{\tau}_r$. As previously noted, Wingo (1993) computes the second-order partial derivatives of (2.35), which are given by

$$\frac{\partial^2 l_r}{\partial \alpha^2} = -r\alpha^{-2}, \tag{2.41}$$

$$\frac{\partial^2 l_r}{\partial \alpha \partial \tau} = \frac{\partial^2 l_r}{\partial \tau \partial \alpha} = -\left(T_{f,111} + T_{c,111}\right), \tag{2.42}$$

$$\frac{\partial^2 l_r}{\partial \tau^2} = -r\tau^{-2} - (\alpha+1)T_{f,122} - \alpha T_{c,122}, \tag{2.43}$$

and states that the exact mathematical expressions for the expected values of (2.42) and (2.43) are very difficult to obtain.

2.4.1 Regularity and EFI Matrix

We develop the discussion in Wingo (1993), adapting the work of Watkins (1997) and Watkins & John (2006). As with the Weibull distribution in Section 2.3.1 above, in order to write down the expected values of all derivatives we will require expectations of the form given at (2.25), where the $g(X_{i:n})$ now are

$$\ln X_{i:n},\ \ \ln\left(1+X_{i:n}^{\tau}\right),\ \ \frac{X_{i:n}^{\tau}\ln X_{i:n}}{1+X_{i:n}^{\tau}},\ \text{ and }\ \frac{X_{i:n}^{\tau}(\ln X_{i:n})^2}{(1+X_{i:n}^{\tau})^2}.$$

We can use (1.49) to state expectations of (2.25) in terms of $E\left[g\left(X_{1:j}\right)\right]$, for which they are usually the most direct to compute. Then, expectation of the first part in (2.25) may be written as

$$\sum_{i=1}^{r} E\left[g\left(X_{i:n}\right)\right] = \sum_{i=1}^{r}(-1)^{r-i}\binom{n}{i-1}\binom{n-i-1}{r-i}E\left[g\left(X_{1:n+1-i}\right)\right], \tag{2.44}$$

and expectation of the second part in (2.25) becomes

$$\begin{aligned}(n-r)E\left[g\left(X_{r:n}\right)\right] &= (n-r)\sum_{i=1}^{r}(-1)^{r-i}\binom{n}{i-1}\binom{n-i}{r-i}E\left[g\left(X_{1:n+1-i}\right)\right] \\ &= (n-i)\sum_{i=1}^{r}(-1)^{r-i}\binom{n}{i-1}\binom{n-i-1}{r-i}E\left[g\left(X_{1:n+1-i}\right)\right],\end{aligned}$$

so that combining these results gives expectation of (2.25) as

$$\begin{aligned}&(n-i+1)\sum_{i=1}^{r}(-1)^{r-i}\binom{n}{i-1}\binom{n-i-1}{r-i}E\left[g\left(X_{1:n+1-i}\right)\right] \\ =\ &n\sum_{i=1}^{r}(-1)^{r-i}\binom{n-1}{i-1}\binom{n-i-1}{r-i}E\left[g\left(X_{1:n+1-i}\right)\right];\end{aligned} \tag{2.45}$$

see Watkins & John (2006).

Expectations in Derivatives

We can now exploit the link between the distribution of the first order statistic and the underlying distribution to proceed. We first show that $Y_{1:n} = 1 + X_{1:n}^{\tau}$ follows a Pareto distribution with pdf $\alpha y^{-(\alpha+1)}, (y \geq 1)$: we have

$$\Pr\{Y_{1:n} \leq y\} = \Pr\{X_{1:n} \leq (y-1)^{\frac{1}{\tau}}\} = 1 - [1 + (y-1)]^{-\alpha n} = 1 - y^{-\alpha n}$$

from (1.43*c*), which gives (1.43*d*) on rearranging. Then, from Section 1.3.2.4, $\ln Y_{1:n}$ has an exponential distribution with mean $(\alpha n)^{-1}$. We thus obtain

$$E\left[\ln\left(1+X_{1:n}^{\tau}\right)\right] = \frac{1}{\alpha n}. \tag{2.46}$$

For $E[\ln X_{1:n}]$, we have, based on (1.36),

$$E\left[X_{1:n}^{p}\right] = \frac{\Gamma\left(\alpha n - \frac{p}{\tau}\right)\Gamma\left(1+\frac{p}{\tau}\right)}{\Gamma\left(\alpha n\right)}$$

so that, using (1.2), differentiating this wrt p yields

$$E\left[X_{1:n}^{p}\ln X_{1:n}\right]=\frac{\Gamma\left(\alpha n-\frac{p}{\tau}\right)\Gamma'\left(1+\frac{p}{\tau}\right)-\Gamma'\left(\alpha n-\frac{p}{\tau}\right)\Gamma\left(1+\frac{p}{\tau}\right)}{\tau\Gamma\left(\alpha n\right)}, \tag{2.47}$$

and evaluating this expectation at $p=0$ gives (from Table 1.6)

$$E[\ln X_{1:n}]=-\left\{\frac{\gamma+\psi(\alpha n)}{\tau}\right\}. \tag{2.48}$$

Next, as in Watkins (1997), we write expectations as $E_{\alpha n}$ to emphasise the role of αn in the pdf of the first Burr order statistic:

$$f_{(1)}(x;\alpha n,\tau)=\alpha n\tau x^{\tau-1}\left(1+x^{\tau}\right)^{-(\alpha n+1)}$$

so that

$$\begin{aligned}
E_{\alpha n}\left[\frac{X_{1:n}^{\tau}\ln X_{1:n}}{1+X_{1:n}^{\tau}}\right] &= \int_{0}^{\infty}\frac{x^{\tau}\ln x}{1+x^{\tau}}\alpha n\tau x^{\tau-1}\left(1+x^{\tau}\right)^{-(\alpha n+1)}dx \\
&= \frac{\alpha n}{\alpha n+1}\int_{0}^{\infty}x^{\tau}\left(\ln x\right)(\alpha n+1)\tau x^{\tau-1}\left(1+x^{\tau}\right)^{-(\alpha n+2)}dx \\
&= \frac{\alpha n}{\alpha n+1}\int_{0}^{\infty}x^{\tau}\left(\ln x\right)f_{(1)}(x;\alpha n+1,\tau)dx \\
&= \frac{\alpha n}{\alpha n+1}E_{\alpha n+1}\left[X_{1:n}^{\tau}\ln X_{1:n}\right].
\end{aligned}$$

Consequently, using (2.47) with $p=\tau$ and αn replaced by $\alpha n+1$, we arrive at

$$\begin{aligned}
E_{\alpha n+1}\left[X_{1:n}^{\tau}\ln X_{1:n}\right] &= \frac{\Gamma(\alpha n)\Gamma'(2)-\Gamma'(\alpha n)\Gamma(2)}{\tau\Gamma(\alpha n+1)} \\
&= \frac{\Gamma(\alpha n)\left(1-\gamma\right)-\psi\left(\alpha n\right)\Gamma(\alpha n)}{\tau\alpha n\Gamma(\alpha n)} \\
&= \frac{1-\gamma-\psi\left(\alpha n\right)}{\alpha n\tau},
\end{aligned}$$

and the expression for the third expectation is given by

$$E_{\alpha n}\left[\frac{X_{1:n}^{\tau}\ln X_{1:n}}{1+X_{1:n}^{\tau}}\right]=\frac{1-\gamma-\psi\left(\alpha n\right)}{\tau\left(\alpha n+1\right)}. \tag{2.49}$$

A similar approach is employed to obtain the final expectation; we have

$$E_{\alpha n}\left[\frac{X_{1:n}^{\tau}(\ln X_{1:n})^{2}}{(1+X_{1:n}^{\tau})^{2}}\right]$$

given by

$$\begin{aligned}
&\int_0^\infty \frac{x^\tau(\ln x)^2}{(1+x^\tau)^2}\alpha n\tau x^{\tau-1}\left(1+x^\tau\right)^{-(\alpha n+1)}dx \\
&= \frac{\alpha n}{\alpha n+2}\int_0^\infty x^\tau(\ln x)^2(\alpha n+2)\tau x^{\tau-1}\left(1+x^\tau\right)^{-(\alpha n+3)}dx \\
&= \frac{\alpha n}{\alpha n+2}\int_0^\infty x^\tau(\ln x)^2 f_{(1)}(x;\alpha n+2,\tau)dx \\
&= \frac{\alpha n}{\alpha n+2}E_{\alpha n+2}\left[X_{1:n}^\tau(\ln X_{1:n})^2\right] \\
&= \frac{\alpha n}{\tau^2(\alpha n+1)(\alpha n+2)}\left\{\begin{array}{c}\frac{\pi^2}{6}+\gamma^2-2\gamma-2(1-\gamma)\psi(\alpha n+1)\\ +\left[\psi(\alpha n+1)\right]^2+\psi'(\alpha n+1)\end{array}\right\}, \qquad (2.50)
\end{aligned}$$

at which, based on (1.3) and Table 1.6,

$$\begin{aligned}
E_{\alpha n+2}\left[X_{1:n}^\tau(\ln X_{1:n})^2\right] &= \frac{\Gamma(\alpha n+1)\Gamma''(2)-2\Gamma'(\alpha n+1)\Gamma'(2)+\Gamma''(\alpha n+1)\Gamma(2)}{\tau^2\Gamma(\alpha n+2)} \\
&= \frac{\Gamma''(2)-2\psi(\alpha n+1)\Gamma'(2)+\left[\psi(\alpha n+1)\right]^2+\psi'(\alpha n+1)}{\tau^2(\alpha n+1)} \\
&= \frac{\frac{\pi^2}{6}+\gamma^2-2\gamma-2(1-\gamma)\psi(\alpha n+1)+\left[\psi(\alpha n+1)\right]^2+\psi'(\alpha n+1)}{\tau^2(\alpha n+1)}
\end{aligned}$$

obtained upon replacing p by τ and αn by $\alpha n+2$ in (2.47).

Expectations of the Score

Having found the expressions for $E\left[g(X_{1:n})\right]$, we now check that the expectations of the score functions are in fact zero. For (2.36), we have

$$\begin{aligned}
E\left[T_f+T_c\right] &= \sum_{i=1}^r E\left[\ln(1+X_{i:n}^\tau)\right]+(n-r)E\left[\ln(1+X_{r:n}^\tau)\right] \\
&= n\sum_{i=1}^r(-1)^{r-i}\binom{n-1}{i-1}\binom{n-i-1}{r-i}E\left[\ln(1+X_{1:n+1-i}^\tau)\right] \quad \text{from (2.45)} \\
&= \alpha^{-1}\sum_{i=1}^r(-1)^{r-i}\binom{n-1}{i-1}\binom{n-i-1}{r-i}\frac{n}{(n+1-i)} \quad \text{from (2.46)} \\
&= r\alpha^{-1}
\end{aligned}$$

because the sum of indices is

$$\sum_{i=1}^r(-1)^{r-i}\binom{n-1}{i-1}\binom{n-i-1}{r-i}\frac{n}{(n+1-i)}=\sum_{i=1}^r(-1)^{r-i}\binom{n}{i-1}\binom{n-i-1}{r-i}=r; \qquad (2.51)$$

see Watkins & John (2006). Thus, taking expectation in (2.36), we obtain

$$E\left[\frac{\partial l_r}{\partial\alpha}\right]=r\alpha^{-1}-E\left[T_f+T_c\right]=r\alpha^{-1}-r\alpha^{-1}=0,$$

as required.

For (2.37), we first write

$$\frac{\partial l_r}{\partial \tau} = r\tau^{-1} + S_{f,1}(0) - \alpha\left(T_{f,111} + T_{c,111}\right) - T_{f,111},$$

and see that, from (2.44),

$$\begin{aligned} E\left[S_{f,1}(0)\right] &= \sum_{i=1}^{r}(-1)^{r-i}\binom{n}{i-1}\binom{n-i-1}{r-i}E\left[\ln X_{1:n+1-i}\right] \\ &= \sum_{i=1}^{r}(-1)^{r-i}\binom{n-1}{i-1}\binom{n-i-1}{r-i}\frac{n}{n+1-i}E\left[\ln X_{1:n+1-i}\right] \end{aligned}$$

and

$$\begin{aligned} E\left[T_{f,111}\right] &= \sum_{i=1}^{r}(-1)^{r-i}\binom{n}{i-1}\binom{n-i-1}{r-i}E\left[\frac{X^{\tau}_{1:n+1-i}\ln X_{1:n+1-i}}{1+X^{\tau}_{1:n+1-i}}\right] \\ &= \sum_{i=1}^{r}(-1)^{r-i}\binom{n-1}{i-1}\binom{n-i-1}{r-i}\frac{n}{n+1-i}E\left[\frac{X^{\tau}_{1:n+1-i}\ln X_{1:n+1-i}}{1+X^{\tau}_{1:n+1-i}}\right], \end{aligned}$$

while, from (2.45),

$$E\left[T_{f,111}+T_{c,111}\right] = n\sum_{i=1}^{r}(-1)^{r-i}\binom{n-1}{i-1}\binom{n-i-1}{r-i}E\left[\frac{X^{\tau}_{1:n+1-i}\ln X_{1:n+1-i}}{1+X^{\tau}_{1:n+1-i}}\right].$$

In particular, we express, using (2.48) and (2.49),

$$\begin{aligned} &\frac{n}{n+1-i}E\left[\ln X_{1:n+1-i}\right] - \alpha n E\left[\frac{X^{\tau}_{1:n+1-i}\ln X_{1:n+1-i}}{1+X^{\tau}_{1:n+1-i}}\right] - \frac{n}{n+1-i}E\left[\frac{X^{\tau}_{1:n+1-i}\ln X_{1:n+1-i}}{1+X^{\tau}_{1:n+1-i}}\right] \\ = & -\frac{n}{n+1-i}\times\frac{\gamma+\psi\left(\alpha\left(n+1-i\right)\right)}{\tau} - \frac{n}{n+1-i}\times\frac{1-\gamma-\psi\left(\alpha\left(n+1-i\right)\right)}{\tau} \\ = & -\frac{n}{\tau\left(n+1-i\right)}. \end{aligned}$$

Thus, taking expectation in (2.37) yields

$$\begin{aligned} E\left[\frac{\partial l_r}{\partial \tau}\right] &= r\tau^{-1} + E\left[S_{f,1}(0)\right] - \alpha E\left[T_{f,111}+T_{c,111}\right] - E\left[T_{f,111}\right] \\ &= r\tau^{-1} - \sum_{i=1}^{r}(-1)^{r-i}\binom{n-1}{i-1}\binom{n-i-1}{r-i}\frac{n}{\tau\left(n+1-i\right)} \\ &= r\tau^{-1} - r\tau^{-1} \end{aligned}$$

from (2.51), so that this expectation is 0, as expected from the regularity consideration.

Expectations of Second Derivatives

In addition to $E\left[T_{f,111}+T_{c,111}\right]$ in (2.42), we will need, for (2.43), $\alpha E\left[T_{f,122}+T_{c,122}\right] + E\left[T_{f,122}\right]$, which is given by

$$\begin{aligned}
&\alpha n \sum_{i=1}^{r}(-1)^{r-i}\binom{n-1}{i-1}\binom{n-i-1}{r-i} E\left[\frac{X_{1:n+1-i}^{\tau}(\ln X_{1:n+1-i})^2}{(1+X_{1:n+1-i}^{\tau})^2}\right] \\
&+\sum_{i=1}^{r}(-1)^{r-i}\binom{n-1}{i-1}\binom{n-i-1}{r-i}\frac{n}{n+1-i} E\left[\frac{X_{1:n+1-i}^{\tau}(\ln X_{1:n+1-i})^2}{(1+X_{1:n+1-i}^{\tau})^2}\right] \\
= \;& n\alpha\tau^{-2}\sum_{i=1}^{r}(-1)^{r-i}\binom{n-1}{i-1}\binom{n-i-1}{r-i}\frac{1}{\alpha(n+1-i)+2}\times \\
&\left\{\begin{array}{l} \frac{\pi^2}{6}+\gamma^2-2\gamma-2(1-\gamma)\psi\left(\alpha(n+1-i)+1\right) \\ +\left[\psi\left(\alpha(n+1-i)+1\right)\right]^2+\psi'\left(\alpha(n+1-i)+1\right)\end{array}\right\}.
\end{aligned}$$

Then, the expectations of (2.41) to (2.43) can now be expressed - using (2.49) and (2.50) - as

$$\begin{aligned}
E\left[\frac{\partial^2 l_r}{\partial\alpha^2}\right] &= -r\alpha^{-2}, \\
E\left[\frac{\partial^2 l_r}{\partial\alpha\partial\tau}\right] &= -n\tau^{-1}\left\{(1-\gamma)\rho_{0,0}-\rho_{0,1}\right\}, \\
E\left[\frac{\partial^2 l_r}{\partial\tau^2}\right] &= -r\tau^{-2}-n\alpha\tau^{-2}\left\{\left(\frac{\pi^2}{6}+\gamma^2-2\gamma\right)\rho_{1,0}-2(1-\gamma)\rho_{1,1}+\rho_{1,2}+\varphi_{1,1}\right\},
\end{aligned}$$

where we find it convenient to define

$$\begin{aligned}
\rho_{k,m} &= \sum_{i=1}^{r}(-1)^{r-i}\binom{n-1}{i-1}\binom{n-i-1}{r-i}\frac{\left[\psi\left(\alpha(n+1-i)+k\right)\right]^m}{\alpha(n+1-i)+k+1}, \\
\varphi_{k,m} &= \sum_{i=1}^{r}(-1)^{r-i}\binom{n-1}{i-1}\binom{n-i-1}{r-i}\frac{\left[\psi'\left(\alpha(n+1-i)+k\right)\right]^m}{\alpha(n+1-i)+k+1},
\end{aligned}$$

for $k=0,1$ and $m=0,1,2$. Furthermore, writing

$$\Omega_r = \left(\frac{\pi^2}{6}+\gamma^2-2\gamma\right)\rho_{1,0}-2(1-\gamma)\rho_{1,1}+\rho_{1,2}+\varphi_{1,1},$$

we immediately obtain the Type II censored EFI matrix as

$$\begin{aligned}
\mathbf{A}_r &= \begin{pmatrix} A_{r,\alpha\alpha} & A_{r,\alpha\tau} \\ A_{r,\alpha\tau} & A_{r,\tau\tau}\end{pmatrix} \\
&= \begin{pmatrix} r\alpha^{-2} & \\ n\tau^{-1}\left\{(1-\gamma)\rho_{0,0}-\rho_{0,1}\right\} & r\tau^{-2}+n\alpha\tau^{-2}\Omega_r\end{pmatrix},
\end{aligned} \tag{2.52}$$

and inverting the above yields the corresponding covariance matrix as

$$\begin{aligned}\mathbf{A}_r^{-1} &= \begin{pmatrix} A_r^{\alpha\alpha} & A_r^{\alpha\tau} \\ A_r^{\alpha\tau} & A_r^{\tau\tau} \end{pmatrix} \\ &= \frac{1}{r^2 + rn\alpha\Omega_r - n^2\alpha^2 \left\{(1-\gamma)\rho_{0,0} - \rho_{0,1}\right\}^2} \\ &\quad \begin{pmatrix} r\alpha^2 + n\alpha_r^3\Omega_r & \\ -n\tau\alpha^2\left\{(1-\gamma)\rho_{0,0} - \rho_{0,1}\right\} & r\tau^2 \end{pmatrix}.\end{aligned} \tag{2.53}$$

Using these results, we also can obtain the moments of the asymptotic distribution for the estimator of the 10^{th} percentile, based on a first order Taylor series expansion of (1.35); we have, from (2.3),

$$\widehat{B}_{0.1,r} \simeq \left(0.9^{-\frac{1}{\alpha}} - 1\right)^{\frac{1}{\tau}} + \begin{pmatrix} b_\alpha & b_\tau \end{pmatrix} \begin{pmatrix} \widehat{\alpha}_r - \alpha \\ \widehat{\tau}_r - \tau \end{pmatrix}$$

with which

$$\begin{pmatrix} b_\alpha \\ b_\tau \end{pmatrix} = \begin{pmatrix} \frac{\partial B_{0.1}}{\partial \alpha} \\ \frac{\partial B_{0.1}}{\partial \tau} \end{pmatrix} = \begin{pmatrix} \alpha^{-2}\tau^{-1}\left(0.9^{-\frac{1}{\alpha}} - 1\right)^{\frac{1}{\tau}-1}\left(0.9^{-\frac{1}{\alpha}}\ln 0.9\right) \\ -\tau^{-2}\left(0.9^{-\frac{1}{\alpha}} - 1\right)^{\frac{1}{\tau}}\ln\left(0.9^{-\frac{1}{\alpha}} - 1\right) \end{pmatrix}. \tag{2.54}$$

Therefore, on taking expected values, we have

$$E\left[\widehat{B}_{0.1,r}\right] \simeq B_{0.1} + b_\alpha E\left[\widehat{\alpha}_r - \alpha\right] + b_\tau E\left[\widehat{\tau}_r - \tau\right] = B_{0.1},$$

and variance given by

$$Var\left(\widehat{B}_{0.1,r}\right) \simeq \begin{pmatrix} b_\alpha & b_\tau \end{pmatrix} \mathbf{A}_r^{-1} \begin{pmatrix} b_\alpha \\ b_\tau \end{pmatrix} = b_\alpha^2 A_r^{\alpha\alpha} + 2b_\alpha b_\tau A_r^{\alpha\tau} + b_\tau^2 A_r^{\tau\tau},$$

obtained from appropriate application of (2.4).

2.4.2 Asymptotic Properties of the MLEs

We are now in the position to write down the asymptotic distribution of the Burr parameters; $(\widehat{\alpha}_r, \widehat{\tau}_r)'$ follows the bivariate Normal distribution with mean $(\alpha, \tau)'$ and covariance matrix given at (2.53). Thus, we may obtain an approximate 95% confidence region for (α, τ) by calculating the ellipse

$$\begin{pmatrix} \alpha - \widehat{\alpha}_r \\ \tau - \widehat{\tau}_r \end{pmatrix}' \widehat{\mathbf{J}}_r \begin{pmatrix} \alpha - \widehat{\alpha}_r \\ \tau - \widehat{\tau}_r \end{pmatrix} = -2\ln 0.05.$$

2.4.3 Complete Sample

For later convenience, we briefly present here some results for the complete sampling. Here, the likelihood is

$$L = \prod_{i=1}^{n} \alpha\tau X_i^{\tau-1} (1 + X_i^{\tau})^{-(\alpha+1)},$$

and the log-likelihood is

$$l = n \ln \alpha + n \ln \tau + (\tau - 1)\, S_1(0) - (\alpha + 1)\, T, \tag{2.55}$$

with two partial derivatives given by

$$\frac{\partial l}{\partial \alpha} = n\alpha^{-1} - T = 0, \tag{2.56}$$

$$\frac{\partial l}{\partial \tau} = n\tau^{-1} + S_1(0) - (\alpha + 1)\, T_{111} = 0, \tag{2.57}$$

where

$$\begin{aligned} T &= \sum_{i=1}^{n} \ln(1 + X_i^{\tau}), \\ T_{abc} &= \sum_{i=1}^{n} \frac{(X_i^{\tau})^a (\ln X_i)^b}{(1 + X_i^{\tau})^c}. \end{aligned}$$

We also list below the second-order partial derivatives of (2.55):

$$\begin{aligned} \frac{\partial^2 l}{\partial \alpha^2} &= -n\alpha^{-2}, \\ \frac{\partial^2 l}{\partial \alpha \partial \tau} &= \frac{\partial^2 l}{\partial \tau \partial \alpha} = -T_{111}, \\ \frac{\partial^2 l}{\partial \tau^2} &= -n\tau^{-2} - (\alpha + 1) T_{122}. \end{aligned}$$

Watkins (1997) computes the following results:

$$\begin{aligned} E[\ln X] &= -\left\{\frac{\gamma + \psi(\alpha)}{\tau}\right\}, \\ E[\ln(1 + X^{\tau})] &= \alpha^{-1}, \\ E\left[\frac{X^{\tau} \ln X}{1 + X^{\tau}}\right] &= \frac{1 - \gamma - \psi(\alpha)}{\tau(\alpha + 1)}, \\ E\left[\frac{X^{\tau} (\ln X)^2}{(1 + X^{\tau})^2}\right] &= \frac{\alpha}{\tau^2 (\alpha + 1)(\alpha + 2)} \left\{ \begin{array}{c} \frac{\pi^2}{6} + \gamma^2 - 2\gamma - 2(1 - \gamma)\psi(\alpha + 1) \\ + [\psi(\alpha + 1)]^2 + \psi'(\alpha + 1) \end{array} \right\}. \end{aligned}$$

Using these, we have, with

$$\Omega = \frac{\pi^2}{6} + \gamma^2 - 2\gamma - 2(1 - \gamma)\psi(\alpha + 1) + [\psi(\alpha + 1)]^2 + \psi'(\alpha + 1),$$

the uncensored EFI matrix given by

$$\mathbf{A}=\begin{pmatrix} A_{\alpha\alpha} & A_{\alpha\tau} \\ A_{\alpha\tau} & A_{\tau\tau} \end{pmatrix}=\begin{pmatrix} n\alpha^{-2} & \\ \frac{n\{1-\gamma-\psi(\alpha)\}}{\tau(\alpha+1)} & n\tau^{-2}\left\{1+\frac{\alpha}{\alpha+2}\Omega\right\} \end{pmatrix}, \tag{2.58}$$

and the associated covariance matrix given by

$$\begin{aligned}\mathbf{A}^{-1} &= \begin{pmatrix} A^{\alpha\alpha} & A^{\alpha\tau} \\ A^{\alpha\tau} & A^{\tau\tau} \end{pmatrix} \\ &= \frac{(\alpha+1)(\alpha+2)}{n\left\{(\alpha+1)^2(\alpha+2)+\alpha(\alpha+1)^2\Omega-\alpha^2(\alpha+2)\left[1-\gamma-\psi(\alpha)\right]^2\right\}} \\ &\quad\times\begin{pmatrix} \alpha^2(\alpha+1)\left\{1+\frac{\alpha}{\alpha+2}\Omega\right\} & \\ -\tau\alpha^2\{1-\gamma-\psi(\alpha)\} & \tau^2(\alpha+1) \end{pmatrix}. \end{aligned} \tag{2.59}$$

2.4.4 Numerical Examples

Arthritic Patients Data

We first use the arthritic patients data given in Table 1.3, and note that the Burr P-P plot for these data fits well to a straight line (see Figure 1.3); see Appendix B for details of the SAS IML algorithm used to locate the MLEs of the Burr parameters and $B_{0.1}$. Table 2.13 gives a summary of $\widehat{\alpha}_r, \widehat{\tau}_r$ and $\widehat{B}_{0.1,r}=\left(0.9^{-\frac{1}{\widehat{\alpha}_r}}-1\right)^{\frac{1}{\widehat{\tau}_r}}$ calculated at several censoring values for these $n=50$ relief times. It is observed that the interim estimates converge to their final values as r tends to 50, in which the convergence in α is the most volatile, followed by τ and then $B_{0.1}$. In fact, a plot of $\widehat{B}_{0.1,r}$ against r would be close to a flat line. The volatility is particularly high when r increases from 10 to 20, and then reduces gradually for each subsequent rise (of size 10) in r. Consequently, we see large $\widehat{sd}(\widehat{\alpha}_r)$ relative to $\widehat{\alpha}_r$ at the 10^{th} and 20^{th} failure, which, in turn, leads to a negative 95% confidence limit for α obtained by assuming that asymptotic Normality holds here. However, we will later investigate the suitability of Normality assumption for parameters and $B_{0.1}$ for sample sizes as small as the arthritic patients data. In contrast, we note steadily decreasing estimated standard deviations for $\widehat{\tau}_r$ and $\widehat{B}_{0.1,r}$.

Simulations

As previously mentioned at (1.36), the moment μ_p for the Burr distribution exists provided that $\alpha\tau > p$, and, since we are often interested at the first two moments we will require $\alpha\tau > 2$. Next, we take $\alpha = 4, \tau = 3$ and run some simulations to validate the theoretical expressions for means and standard deviations of the estimators $\widehat{\alpha}_r$ and $\widehat{\tau}_r$, based on 10^4 replications. Figure 2.12 shows the shape of the Burr pdf for such simulation. We note that other values of α and τ are possible; see, for instance, Figures 1.6 and 1.7 for the effect of varying α and τ on the shape of the Burr pdf. However, with $\alpha\tau = 12 > 2$, simulations

r	10	20	30	40	50
$X_{r:50}$	0.49	0.57	0.64	0.73	0.87
$\widehat{\alpha}_r$	4.5450	7.9878	8.9031	7.7911	8.2681
$\widehat{sd}(\widehat{\alpha}_r)$	4.0266	4.6839	3.6191	2.1342	1.6837
95% CIs	-3.347,12.437	-1.193,17.168	1.810,15.997	3.608,11.974	4.968,11.568
$\widehat{\tau}_r$	4.1860	4.8626	4.9997	4.8490	5.0006
$\widehat{sd}(\widehat{\tau}_r)$	1.1833	0.9587	0.7707	0.6053	0.5045
95% CIs	1.867,6.505	2.984,6.742	3.489,6.510	3.663,6.035	4.012,5.990
$\widehat{B}_{0.1,r}$	0.4080	0.4112	0.4112	0.4113	0.4185
$\widehat{sd}(\widehat{B}_{0.1,r})$	0.0380	0.0321	0.0303	0.0297	0.0272
95% CIs	0.333,0.483	0.348,0.474	0.353,0.472	0.354,0.471	0.365,0.472

Table 2.13: Summaries of the Burr MLEs calculated at various r for the arthritic patients data.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	134.7855	36.5905	8.6449	4.2097	4.0759	4.0405
$0.4n$	13.5455	5.9973	4.7773	4.0613	4.0227	4.0125
$0.6n$	6.1467	4.7217	4.3112	4.0291	4.0113	4.0064
$0.8n$	4.8061	4.3339	4.1539	4.0133	4.0069	4.0041
$1.0n$	4.3969	4.1808	4.0800	4.0082	4.0039	4.0025

Table 2.14: Simulated means of $\hat{\alpha}_r$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 3$.

are much more controlled than, say, with α and τ close to 0, and hence we stand at a good chance of getting asymptotically valid agreement between theory and simulation.

Results for simulated means are shown in Table 2.14 for $\widehat{\alpha}_r$ and Table 2.15 for $\widehat{\tau}_r$; and for theoretical and simulated standard deviations are shown in Table 2.16 for $\widehat{\alpha}_r$ and Table 2.17 for $\widehat{\tau}_r$. In general, for small samples with low censoring levels, $\widehat{\alpha}_r$ and $\widehat{\tau}_r$ do not agree with their true values very well at all, although the disagreement is less severe for $\widehat{\tau}_r$. In fact, for $\widehat{\alpha}_r$, it is only really for a sample size of 1000 that we begin to observe agreement between simulated and theoretical means, for any value of r we have considered. We also note in Figure 2.13 certain replications with large estimates of α; such values will clearly affect the sample mean and standard deviation. The effect is stronger when r is low, generally $\leq 0.4n$, because the lower the censoring level, the less information we have, thus increasing the chance of obtaining an unusually large $\widehat{\alpha}_r$. As shown in Table 2.16, there are quite large discrepancies between the simulated and theoretical standard deviation values at early censoring levels, but discrepancy reduces as r increases. Figures 2.13, 2.14, 2.15 and 2.16 show scatter plots for four combinations of final estimates against interim estimates when $n = 50$. In general, there is some link of varied strength between the two sets of estimates, although this is partly distorted by large standard deviations of $\widehat{\alpha}_r$ when $r \leq 0.4n$. As before, we wish to determine if we can make inferences on final estimates, $\widehat{\alpha}$ and $\widehat{\tau}$, given interim estimates, $\widehat{\alpha}_r$ and $\widehat{\tau}_r$.

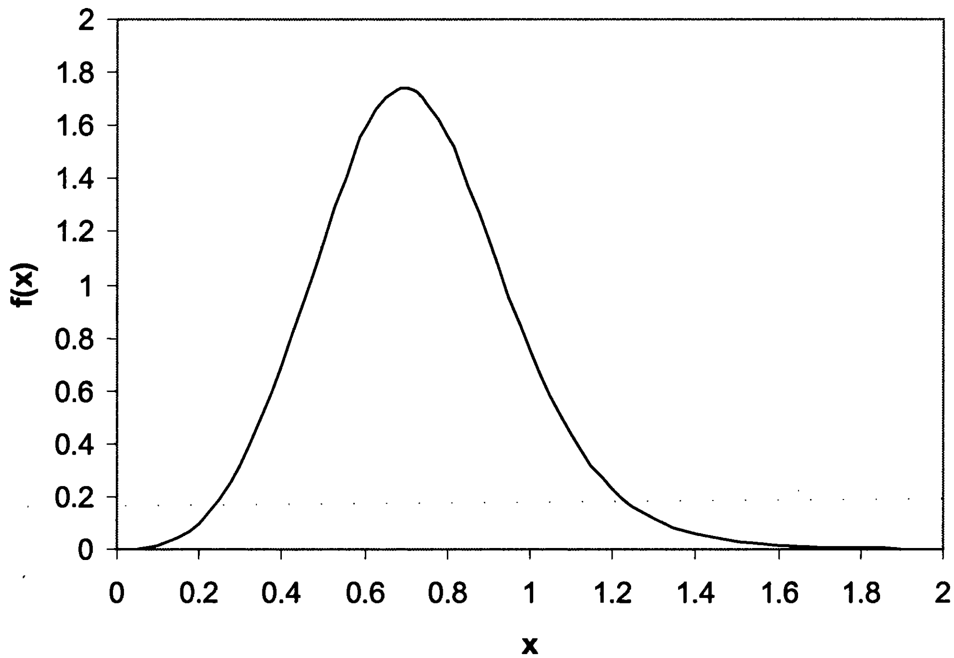

Figure 2.12: Pdf of the Burr distribution for $\alpha = 4$ and $\tau = 3$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	4.7524	3.6854	3.3058	3.0275	3.0103	3.0058
$0.4n$	3.6317	3.2695	3.1372	3.0121	3.0047	3.0027
$0.6n$	3.3519	3.1592	3.0806	3.0076	3.0032	3.0019
$0.8n$	3.2252	3.1004	3.0529	3.0042	3.0025	3.0015
$1.0n$	3.1473	3.0653	3.0352	3.0031	3.0016	3.0010

Table 2.15: Simulated means of $\hat{\tau}_r$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 3$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	4.5732	3.3881	2.4590	0.7976	0.5053	0.3575
	453.4663	206.5771	21.4551	0.9054	0.5242	0.3654
$0.4n$	2.4557	1.7754	1.2704	0.4062	0.2571	0.1819
	73.8326	5.9415	2.0405	0.4219	0.2586	0.1832
$0.6n$	1.5981	1.1426	0.8126	0.2584	0.1635	0.1156
	10.3940	1.9311	1.0123	0.2584	0.1640	0.1152
$0.8n$	1.1473	0.8148	0.5775	0.1830	0.1158	0.0819
	2.2064	1.0342	0.6598	0.1816	0.1158	0.0819
$1.0n$	0.8880	0.6279	0.4440	0.1404	0.0888	0.0628
	1.2002	0.7293	0.4744	0.1387	0.0888	0.0623

Table 2.16: Theoretical (upper) and simulated (lower) standard deviations of $\hat{\alpha}_r$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 3$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	1.1508	0.8448	0.6103	0.1971	0.1249	0.0883
	2.8447	1.3481	0.7538	0.2016	0.1252	0.0890
$0.4n$	0.8013	0.5775	0.4125	0.1317	0.0834	0.0590
	1.2326	0.6970	0.4543	0.1327	0.0830	0.0591
$0.6n$	0.6291	0.4496	0.3197	0.1016	0.0643	0.0455
	0.8130	0.5068	0.3376	0.1012	0.0643	0.0453
$0.8n$	0.5181	0.3683	0.2612	0.0828	0.0524	0.0370
	0.6160	0.3946	0.2714	0.0825	0.0527	0.0369
$1.0n$	0.4335	0.3065	0.2168	0.0685	0.0434	0.0307
	0.4914	0.3237	0.2214	0.0681	0.0435	0.0303

Table 2.17: Theoretical (upper) and simulated (lower) standard deviations of $\hat{\tau}_r$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 3$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	0.3064	0.3020	0.3011	0.2990	0.2989	0.2989
$0.4n$	0.3115	0.3046	0.3024	0.2991	0.2990	0.2989
$0.6n$	0.3108	0.3043	0.3021	0.2991	0.2990	0.2989
$0.8n$	0.3092	0.3034	0.3017	0.2990	0.2990	0.2989
$1.0n$	0.3071	0.3023	0.3011	0.2990	0.2989	0.2989

Table 2.18: Simulated means of $\hat{B}_{0.1,r}$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 3$.

When we examine results for $\widehat{B}_{0.1,r}$ in Tables 2.18 and 2.19, we observe simulated means converge to the true value of

$$B_{0.1} = \left(0.9^{-\frac{1}{4}} - 1\right)^{\frac{1}{3}} = 0.2988.$$

as n and r increase, together with decreasing standard deviations. It is somewhat surprising to note that large estimates of α do not seem to affect the estimates of $B_{0.1}$, and the largest relative margin of error between theoretical and simulated mean is just 4%, and 2.62% for standard deviation. As before, Figure 2.17 displays the relationship between $\widehat{B}_{0.1}$ and $\widehat{B}_{0.1,r}$ when n is 50, in which we see clear linear pattern even for low censoring levels.

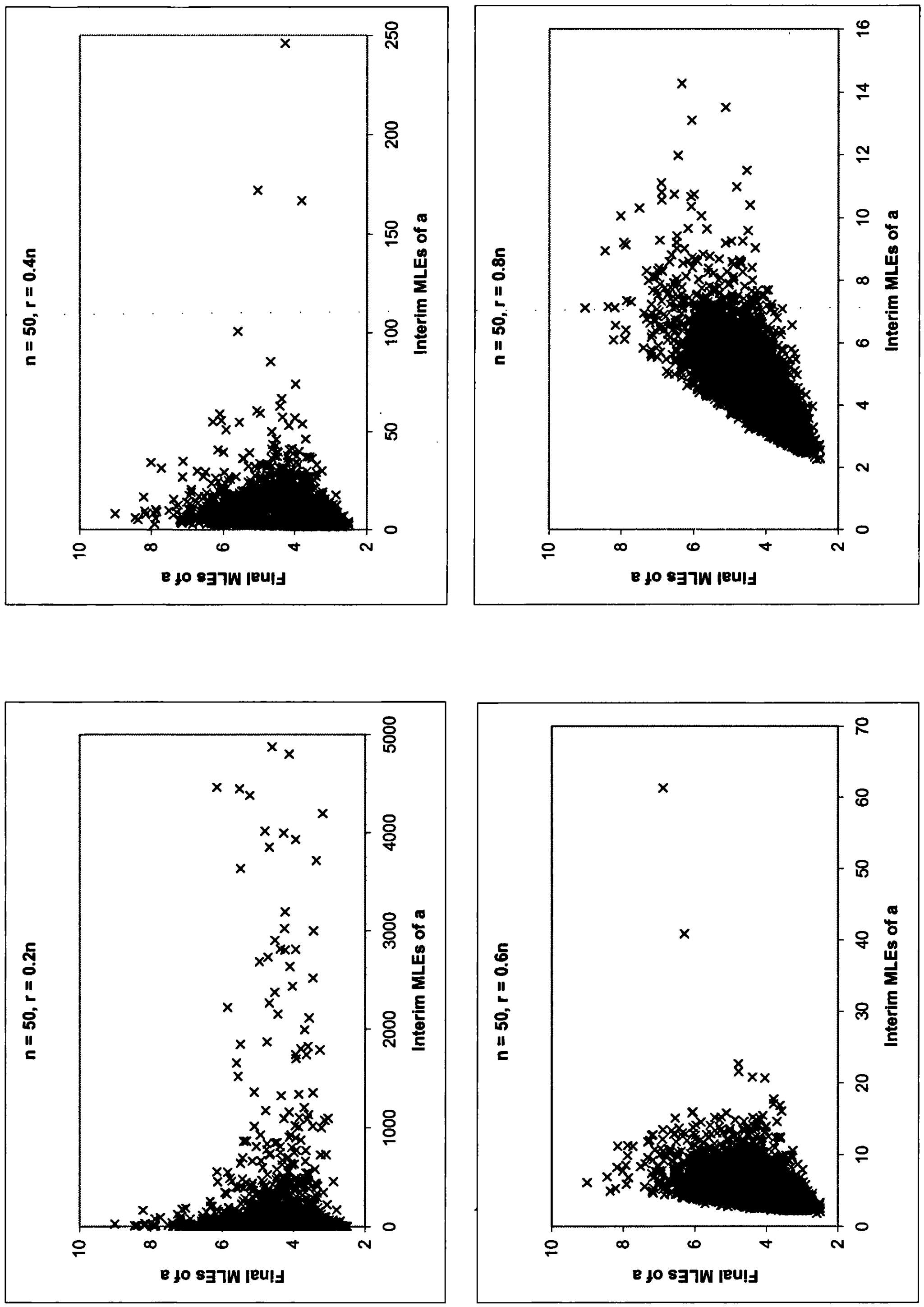

Figure 2.13: Scatter plots of $\hat{\alpha}$ versus $\hat{\alpha}_r$ for $n = 50$ and various r, for Burr data generated with $\alpha = 4, \tau = 3$.

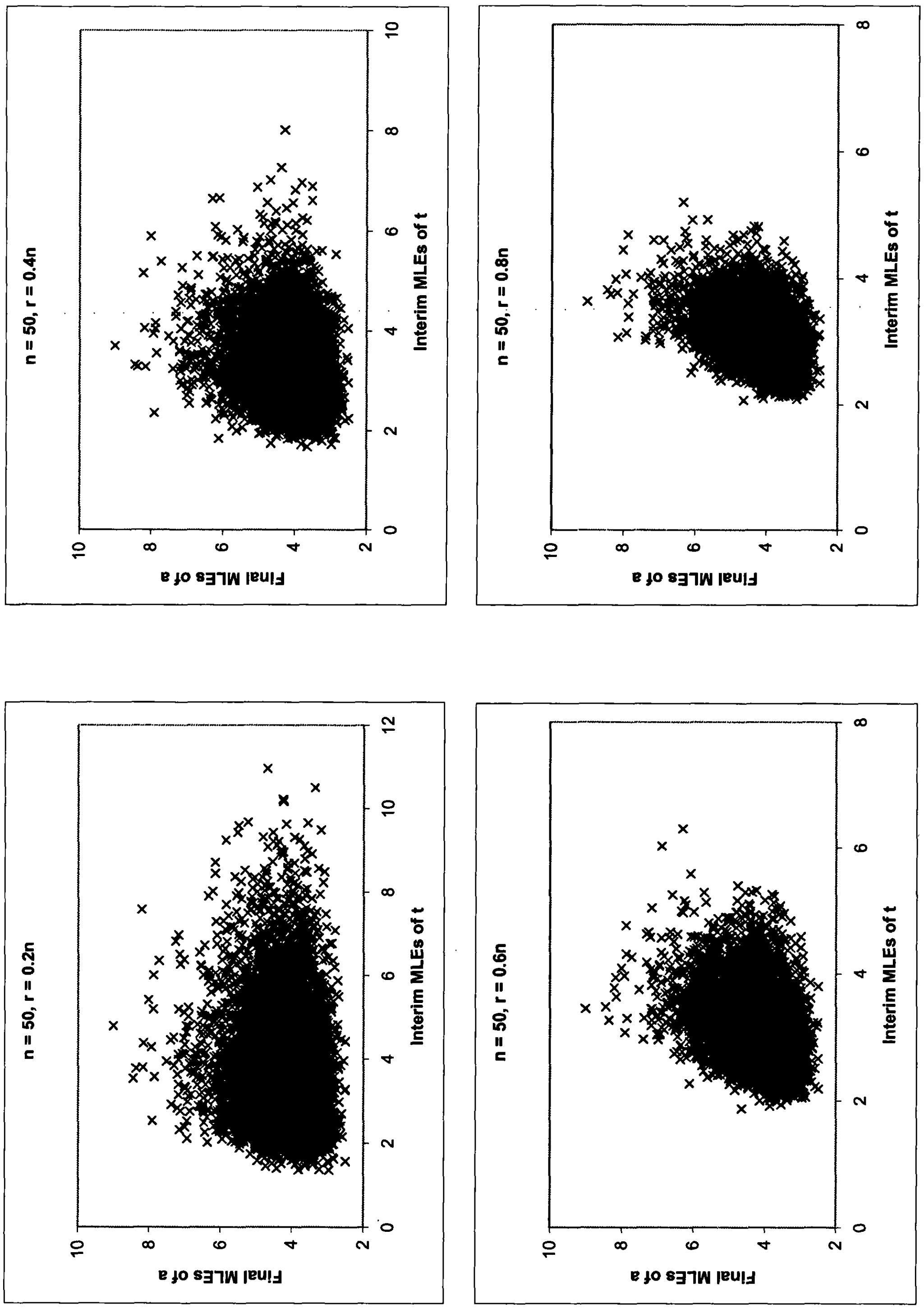

Figure 2.14: Scatter plots of $\hat{\alpha}$ versus $\hat{\tau}_r$ for $n = 50$ and various r, for Burr data generated with $\alpha = 4, \tau = 3$.

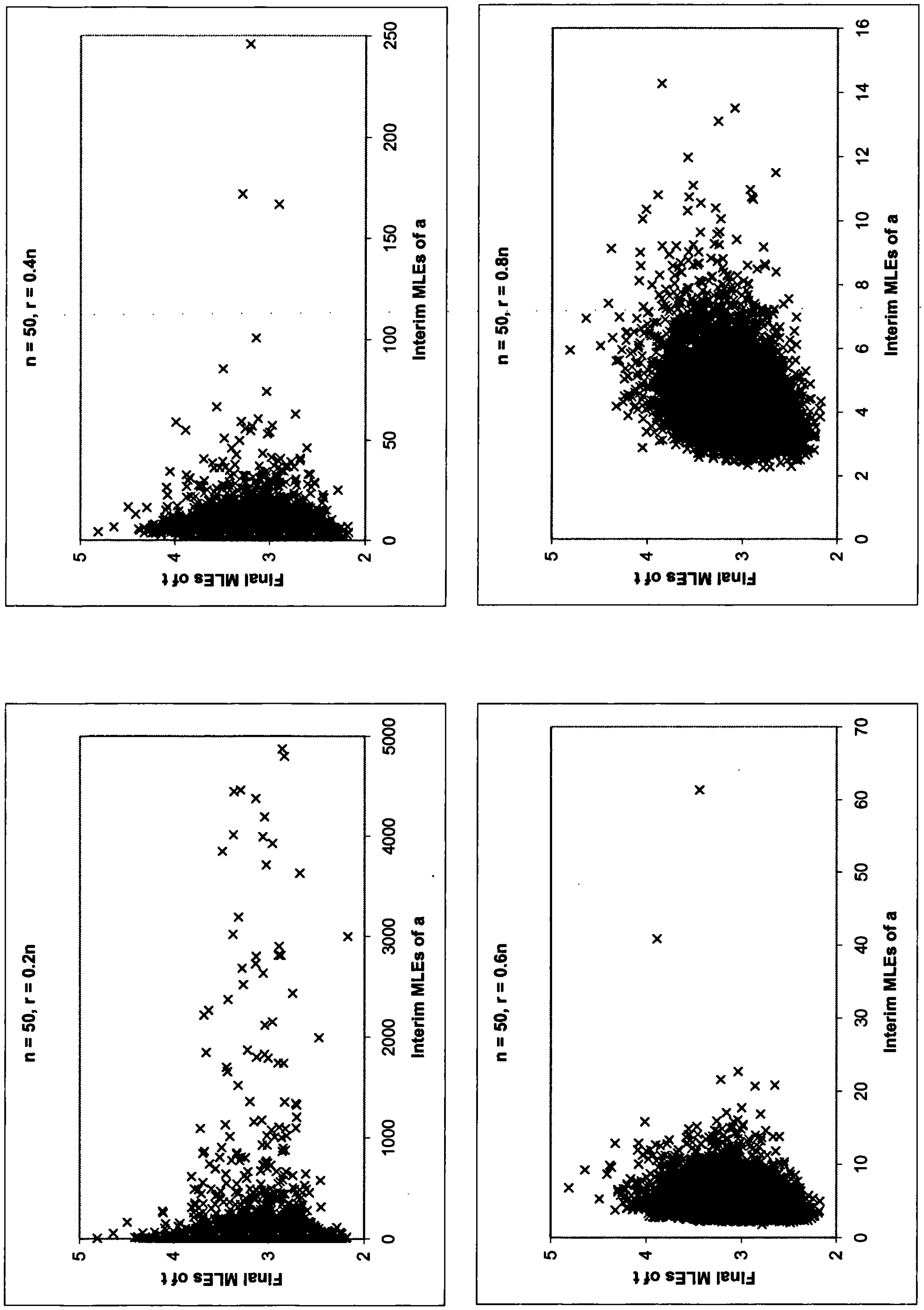

Figure 2.15: Scatter plots of $\hat{\tau}$ versus $\hat{\alpha}_r$ for $n = 50$ and various r, for Burr data generated with $\alpha = 4, \tau = 3$.

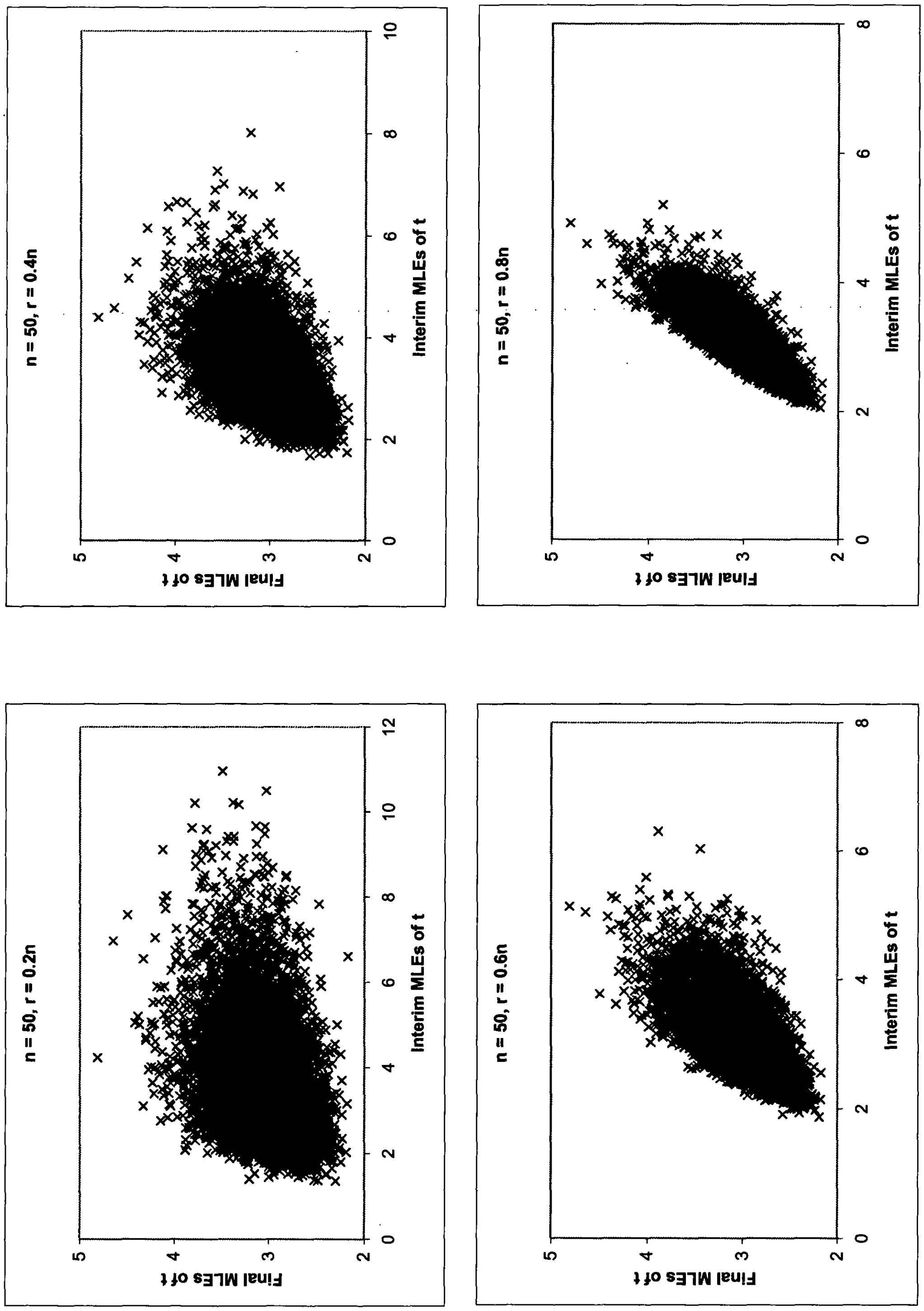

Figure 2.16: Scatter plots of $\hat{\tau}$ versus $\hat{\tau}_r$ for $n = 50$ and various r, for Burr data generated with $\alpha = 4, \tau = 3$.

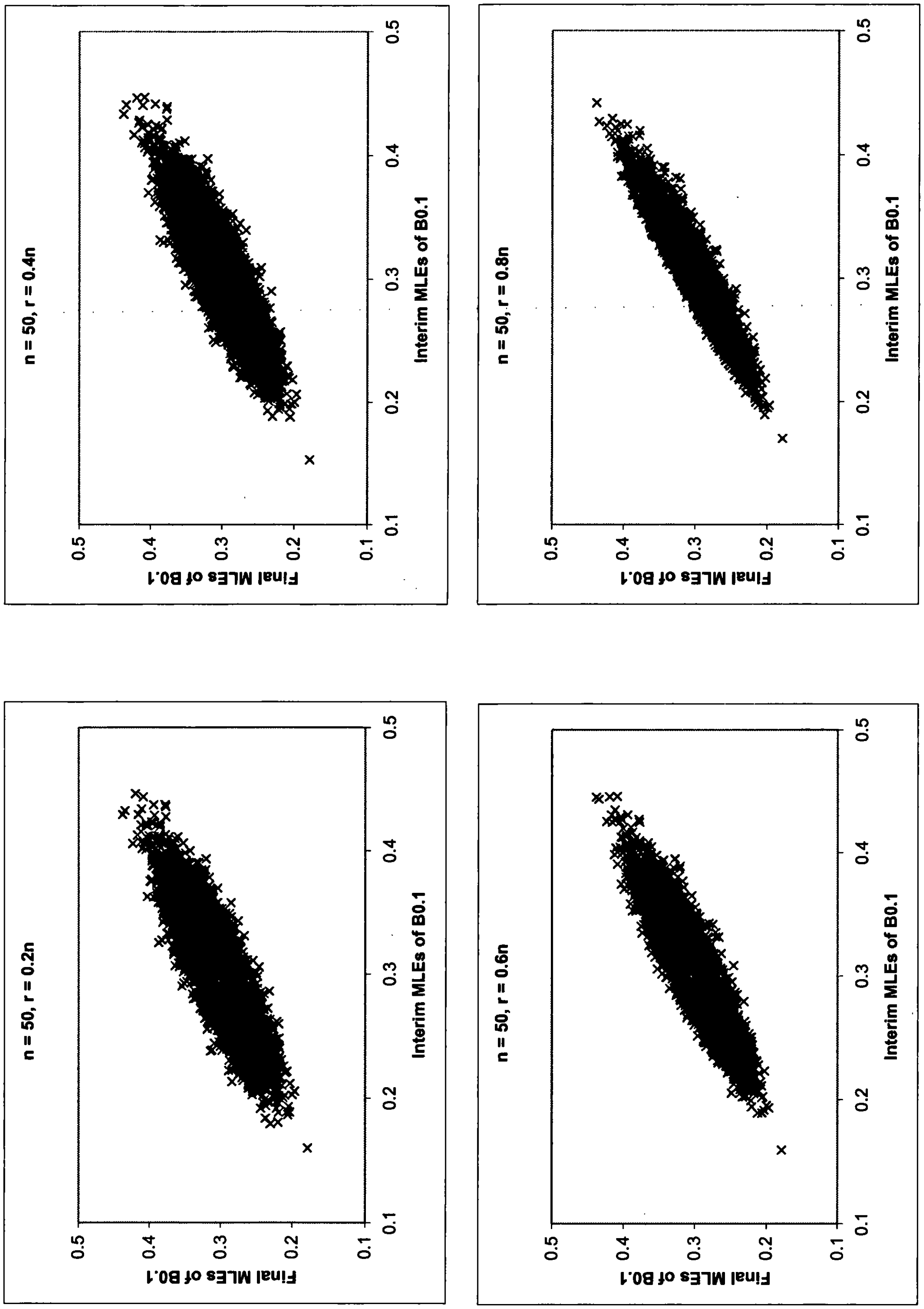

Figure 2.17: Scatter plots of $\hat{B}_{0.1}$ versus $\hat{B}_{0.1,r}$ for $n = 50$ and various r, for Burr data generated with $\alpha = 4, \tau = 3$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	0.0555	0.0390	0.0275	0.0087	0.0055	0.0039
	0.0557	0.0386	0.0274	0.0087	0.0055	0.0039
$0.4n$	0.0538	0.0382	0.0270	0.0086	0.0054	0.0038
	0.0548	0.0382	0.0271	0.0086	0.0054	0.0038
$0.6n$	0.0519	0.0369	0.0261	0.0083	0.0052	0.0037
	0.0531	0.0370	0.0262	0.0083	0.0052	0.0037
$0.8n$	0.0496	0.0352	0.0249	0.0079	0.0050	0.0035
	0.0509	0.0353	0.0251	0.0079	0.0050	0.0035
$1.0n$	0.0470	0.0332	0.0235	0.0074	0.0047	0.0033
	0.0481	0.0333	0.0237	0.0074	0.0047	0.0033

Table 2.19: Theoretical (upper) and simulated (lower) standard deviations of $\hat{B}_{0.1,r}$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 3$.

2.5 Chapter Summary and Conclusions

In this chapter, we outlined the theory necessary to fit exponential, Weibull and Burr distributions to Type II censored data using maximum likelihood techniques. In each case, we confirmed the regularity conditions and obtained suitable formulae for the elements of the EFI matrix analytically; the inverses of the matrices providing us with asymptotically valid variances and covariances of the MLEs of the model parameters, as well as variances of functions of the parameters, such as $B_{0.1}$. In particular, we have made progress from Wingo (1993) to derive analytical expressions for the elements of the EFI matrix for Type II censored Burr data. Naturally, it would be of interest to extend the two-parameter Burr to a three-parameter model by introducing a (natural) scale parameter ϕ in many different ways. In Tadikamalla (1980), one of these is to consider $Y = \phi X$ for $\phi > 0$, where X is a random variable following (1.34). Then, the random variable $Y \geq 0$ will have pdf and cdf defined, respectively, by

$$f(y; \alpha, \tau, \phi) = \frac{\alpha\tau}{\phi^{\tau}} y^{\tau-1} \left[1 + \left(\frac{y}{\phi}\right)^{\tau}\right]^{-(\alpha+1)} \tag{2.60}$$

and

$$F(y; \alpha, \tau, \phi) = 1 - \left[1 + \left(\frac{y}{\phi}\right)^{\tau}\right]^{-\alpha}. \tag{2.61}$$

Inevitably, here the statistical analysis, like the derivation of the Type II censored EFI matrix, will be much more involved, and hence is considered elsewhere.

We have constructed a set of simulations to check such approximations to the moments of the MLEs and $\widehat{B}_{0.1,r}$ for various sample sizes and censoring levels, and noted good agreement between the theoretical approximations and simulated values, which improves as n and r increase. We have also shown that, perhaps surprisingly, the agreement for the moments of $\widehat{B}_{0.1,r}$ is generally better than those of the MLEs.

We have computed the approximate 95% confidence intervals for parameters and $B_{0.1}$ for published data mentioned in Chapter 1, assuming that asymptotic Normality of MLEs holds for small samples. But are these large sample theory approximations suitable in the inference of small to moderate samples, such as the ball bearings data? In the following chapter, we consider the implications of asymptotic Normality, and more importantly, the extent to which this large sample result holds for samples of small size, subject to Type II censoring. Where the sample size is too small for Normality to be assumed, we also discuss the use of relative likelihood function as an alternative to measure the precision of the MLEs. As well as being asymptotically equivalent to the Normal confidence regions, studies by Watkins (2004) and Chua *et al.* (2007), for example, have shown that relative likelihood contours reflect more accurately the behaviour of the distributions of MLEs for relatively small samples.

Following the observation in Tables 2.1, 2.6 and 2.13, and, as discussed for all scatter plots of final estimates of parameters and $B_{0.1}$ against interim estimates, we wish to find the extent to which a censored estimate, obtained in an interim analysis, can be regarded as a reliable guide to complete estimate, obtained when the last item fails. In Chapter 5 we will be investigating the relationship between the two sets of estimates of parameters and $B_{0.1}$, using results on expectations of various functions of order statistics found from Chapter 4.

Chapter 3

Small Sample Properties of Maximum Likelihood Estimators for Type II Censored Data

3.1 Introduction

We have already mentioned asymptotic properties of the MLEs (for instance, see Cox & Hinkley, 1974), and like many other authors (see Meeker & Nelson, 1977 for example) we used these properties to obtain approximate confidence intervals for parameters and for $B_{0.1}$. In particular, this asymptotic theory implies symmetric confidence intervals for a single parameter or quantity, and elliptical confidence regions for two. Billmann *et al.* (1972) give confidence limits for the Weibull parameters from Type II censored samples, for sample sizes $n = 40, 60, 80, 100, 120$ with $r = 0.5n, 0.75n, 1.0n$, based on $N = 4000$ replications. They note that their sampling distributions of $\widehat{\theta}_r$ and $\widehat{\beta}_r$ are not close to Normal for small samples, say, where n is less than 100, but there is no mention on how large a sample needs to be for this large-sample approximation to hold. Hence, it is now appropriate to assess the progress of the MLEs of parameters to Normality. The relevance and importance of the percentile $B_{0.1}$ has been introduced in chapter one and two. Naturally, it is also of interest to extend the Normality checks to the sampling distribution of $\widehat{B}_{0.1,r}$.

Chua *et al.* (2007) consider two issues emerging from the above, with the first part focusing on the progress towards Normality (the problem), and the second part dealing primarily with the use of relative likelihood contour as an approximate confidence region (a possible solution). As outlined in Lawless (1982), likelihood function is usually used to examine the whole range of possible parameter values, and to investigate which values are plausible and which are implausible in the light of the data. In particular, relative likelihood function ranks possible parameter values according to their consistency with the observed data, and, as Kalbfleisch (1979) has discussed, contour plots of relative likelihood function may be used to obtain confidence regions for a sample, including the possibility of censoring.

In this chapter, we will not necessarily be looking to test Normality at any given sample size, but instead to show, by means of a detailed simulation study with $N = 10^4$, that at small sample sizes the MLEs of parameters and $B_{0.1}$ are non-Normal, and eventually when the sample size increases, the MLEs become Normally distributed. We also aim to illustrate the effects of varying r on the convergence to asymptotic Normality, and, on the shape and size of the relative likelihood contours. Section 3.2 investigates the extent to which univariate Normality of the MLE applies in finite samples based on Type II censored exponential, Weibull and Burr data. In Section 3.3, we extend the study to testing for bivariate Normality in Weibull and Burr MLEs. Then, in Section 3.4, we consider the use of relative likelihood function as a method for obtaining confidence regions of the sampling distribution of MLEs in small samples of varying sizes. For consistency, we use the six sample sizes and censoring proportions given in Chapter 2.

3.2 Tests of Univariate Normality

Numerous tests for assessing Normality, including both univariate and multivariate Normality, exist in the literature; each has its relative strengths and weaknesses. In summary, numerical analyses include moment-type tests, general goodness of fit tests (tests based on empirical distribution function, the Kolmogorov-Smirnov test, and so forth), and other tests specifically derived to detect outliers; see, for example, D'Agostino & Stephens (1986).

Recent reviews on testing for Normality (Thode, 2002 and Srivastava & Mudholkar, 2003, for example) tend to focus on procedures based on the sample moments, we will consider the skewness (γ_1) and kurtosis (γ_2) statistics of the distribution of MLE in this thesis. In particular, for a sample of N values $\widehat{\pi}_1, \ldots, \widehat{\pi}_N$ the sample estimates of skewness and kurtosis are, respectively, from (1.25) and (1.26),

$$g_1 = \frac{m_3^*}{S^3}$$

and

$$g_2 = \frac{m_4^*}{S^4},$$

where m_p^* is the p^{th} sample moment about the mean given by

$$m_p^* = \frac{\sum_{i=1}^{N} \left(\widehat{\pi}_i - \overline{\widehat{\pi}}\right)^p}{N}$$

so that

$$\overline{\widehat{\pi}} = \frac{\sum_{i=1}^{N} \widehat{\pi}_i}{N}$$

is the sample mean, and $S^2 \equiv m_2^*$ is the sample variance. Hence, values of g_1 and g_2 close to 0 and 3, respectively, are consistent with Normality. We further refer to D'Agostino & Pearson (1973) for the K^2 statistic, originally discussed in D'Agostino (1971), which combines g_1 and g_2 as an omnibus test for univariate Normality. By omnibus, we mean it is able to detect deviations from Normality due to either skewness or kurtosis; we have

$$K^2 = \{Z(g_1)\}^2 + \{Z(g_2)\}^2, \tag{3.1}$$

where $Z(g_1)$ and $Z(g_2)$ are suitably standardised and Normalised measures of skewness and kurtosis. Hence, under the hypothesis that the marginal distribution of a MLE is Normal, we have $K^2 \sim \chi^2_2$, so that we can assess the marginal Normality of $\widehat{\pi}_r$ via the critical value

$$-2\ln\lambda$$

for an upper tail probability of λ. Since $\chi^2_{2,0.95} = 5.9915$, $K^2 \leq 5.9915$ indicates the possibility of univariate Normality. This procedure has the computational advantages that skewness and kurtosis measures are readily supplied by many standard statistical packages (SAS and SPSS) as well as by Excel, and D'Agostino *et al.* (1990) also provide a simple SAS macro programme to implement the K^2 test.

It is always useful to include a graphical inspection of the data in conjunction with a formal test. Classic methods include probability plots, and regression and correlation tests. Since we are mainly concerned with progress towards Normality and symmetrical confidence limits in the context of single MLE, we use histogram overlaid with the best-fit Normal curve as a display of the distribution of the sample. This shows clearly the frequency of observations within bins, and also allows us to observe easily features like skewness, spread, outliers and multimodality in the sampling distribution. Thus, for each MLE, we will investigate the symmetry around the probability intervals, calculated from asymptotic Normality theory, with the focus on the effects of varying r on the rate at which the MLE approaches Normality. In our simulation experiments with $N = 10^4$, and suppose this large-sample result holds, we would expect to find $95\% \times 10^4$ of the estimates within the 95% limits, $2.5\% \times 10^4$ to lie below and $2.5\% \times 10^4$ to lie above the limits.

3.2.1 Simulation Study: the Exponential Distribution

As shown in Table 2.1 for the failure times data with $n = 49$, the approximate 95% confidence intervals for θ and $B_{0.1}$ are, respectively, $(77.370, 146.860)$ and $(8.152, 15.473)$ when $r = 40$, assuming that the asymptotic theory of MLE held for a sample of this size. However, the question is can we safely exploit Normality in inference of small to moderate samples, such as the failure times data.

For our investigation the parameter value was chosen to be $\theta = 100$, and we take $r = 0.8n$ so that the experiments were terminated after 80% of the items fails; we may again omit the analysis of $B_{0.1}$ since this percentile is linearly related to θ via (1.29). Figure

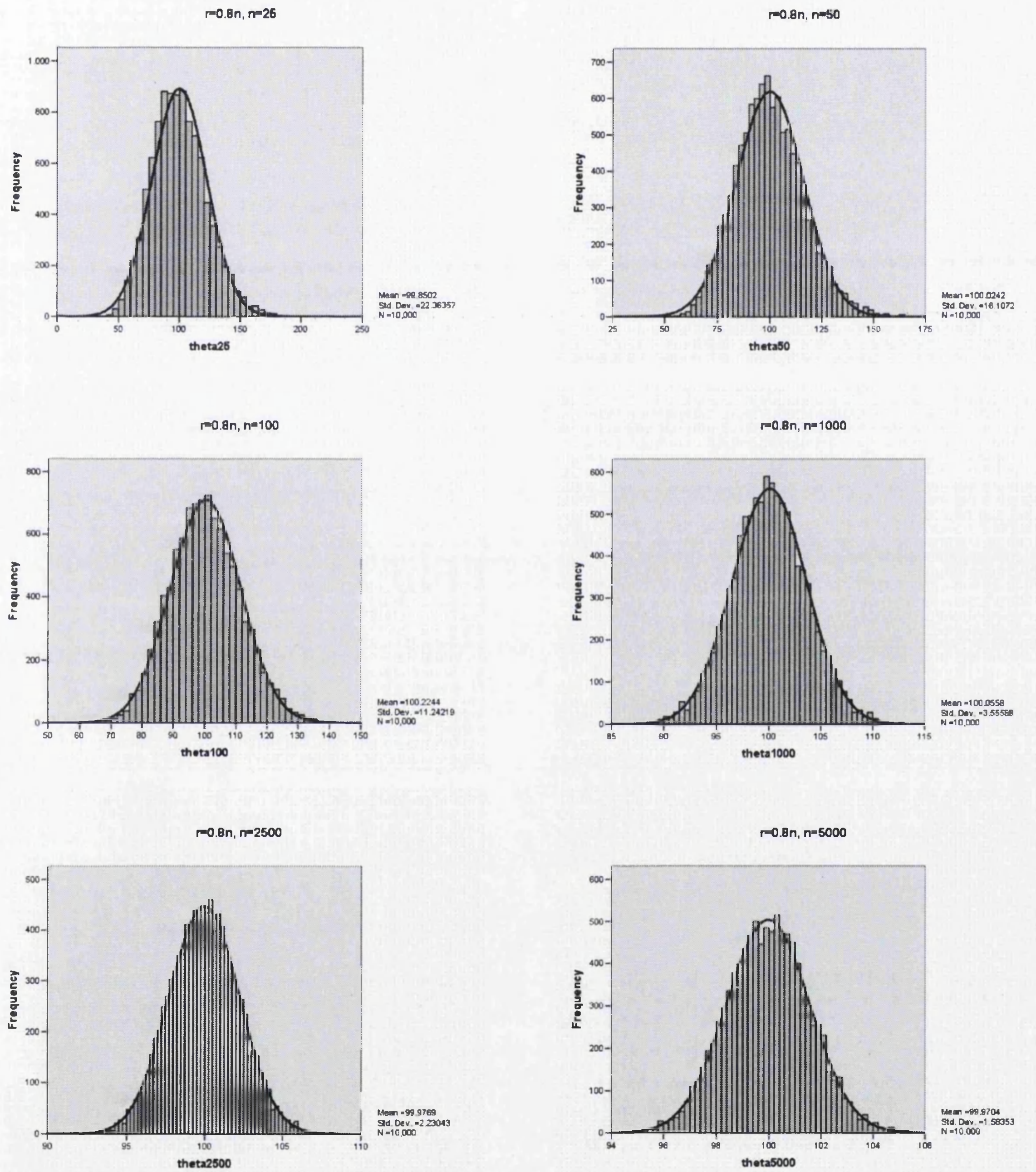

Figure 3.1: Histograms of $\hat{\theta}_{0.8n}$ for various n, for exponential data generated with $\theta = 100$.

n	g_1	$Z(g_1)$	g_2	$Z(g_2)$	K^2	95% prob. intervals		
						Below	Within	Above
25	0.4196	16.4672	3.2445	4.4995	291.4154	130	9518	352
50	0.3478	13.8130	3.1852	3.5058	203.0899	168	9474	358
100	0.2235	9.0210	3.0147	0.3343	81.4904	188	9474	338
1000	0.0711	2.8997	3.0740	1.4947	10.6421	240	9473	287
2500	0.0507	2.0709	3.0032	0.1022	4.2991	244	9505	251
5000	0.0472	1.9280	2.9677	-0.6340	4.1191	230	9524	246

Table 3.1: Summary statistics for $\hat{\theta}_{0.8n}$ for various n, for exponential data generated with $\theta = 100$.

n	$r = 0.2n$	$r = 0.4n$	$r = 0.6n$	$r = 0.8n$	$r = 1.0n$
25	1409.3486	788.5079	484.5088	291.4154	269.7868
50	807.6852	406.5420	276.7434	203.0899	143.3069
100	298.0283	160.4063	118.8980	81.4904	59.7064
1000	15.9070	6.3177	14.2387	10.6421	10.8260
2500	33.3457	8.5319	10.6261	**4.2991**	**4.3223**
5000	18.0613	13.9747	**3.8181**	**4.1191**	**4.2676**

Table 3.2: K^2 statistics for $\hat{\theta}_r$ for various r, n, for exponential data generated with $\theta = 100$.

3.1 together with summaries in Table 3.1 seem to suggest that as the sample size increases, the distributions of $\widehat{\theta}_{0.8n}$ become more and more Normal, as indicated by the K^2 values. At the same time, we see the distributions are less skewed and more centred around the expected value of 100. As a supplementary check of the distribution, we further examine how the MLEs spread about 100; this uses the 95% probability limits for θ given by

$$100 \pm 1.96 \times 100 \times (0.8n)^{-1/2},$$

obtained from (2.12), which generates symmetric confidence intervals for θ. We then plot the 10^4 simulated observations of $\widehat{\theta}_r$, and, if the large-sample result holds, we expect to find 9500 of $\widehat{\theta}_r$ within the corresponding limits, with 250 of $\widehat{\theta}_r$ below (above) the lower (upper) limit. Table 3.1 shows that, although approximately 95% of the $\widehat{\theta}_r$ are enclosed in the intervals, the remaining 5% are divided unequally between the two tails, with more in the upper tail. This implies right skewness in both sampling distributions of $\widehat{\theta}_{0.8n}$ and $\widehat{B}_{0.1,0.8n}$.

In addition, Table 3.2 tabulates the K^2 statistics for assessment of Normality in $\widehat{\theta}_r$ for varying r and n; we have highlighted any entry less than 5.9915. We see that we need large samples and almost complete data sets before we could formally accept the hypothesis of Normality. Therefore, with respect to the failure times data with $n = 50$, it is not really sensible to employ Normality in the calculation of confidence limits.

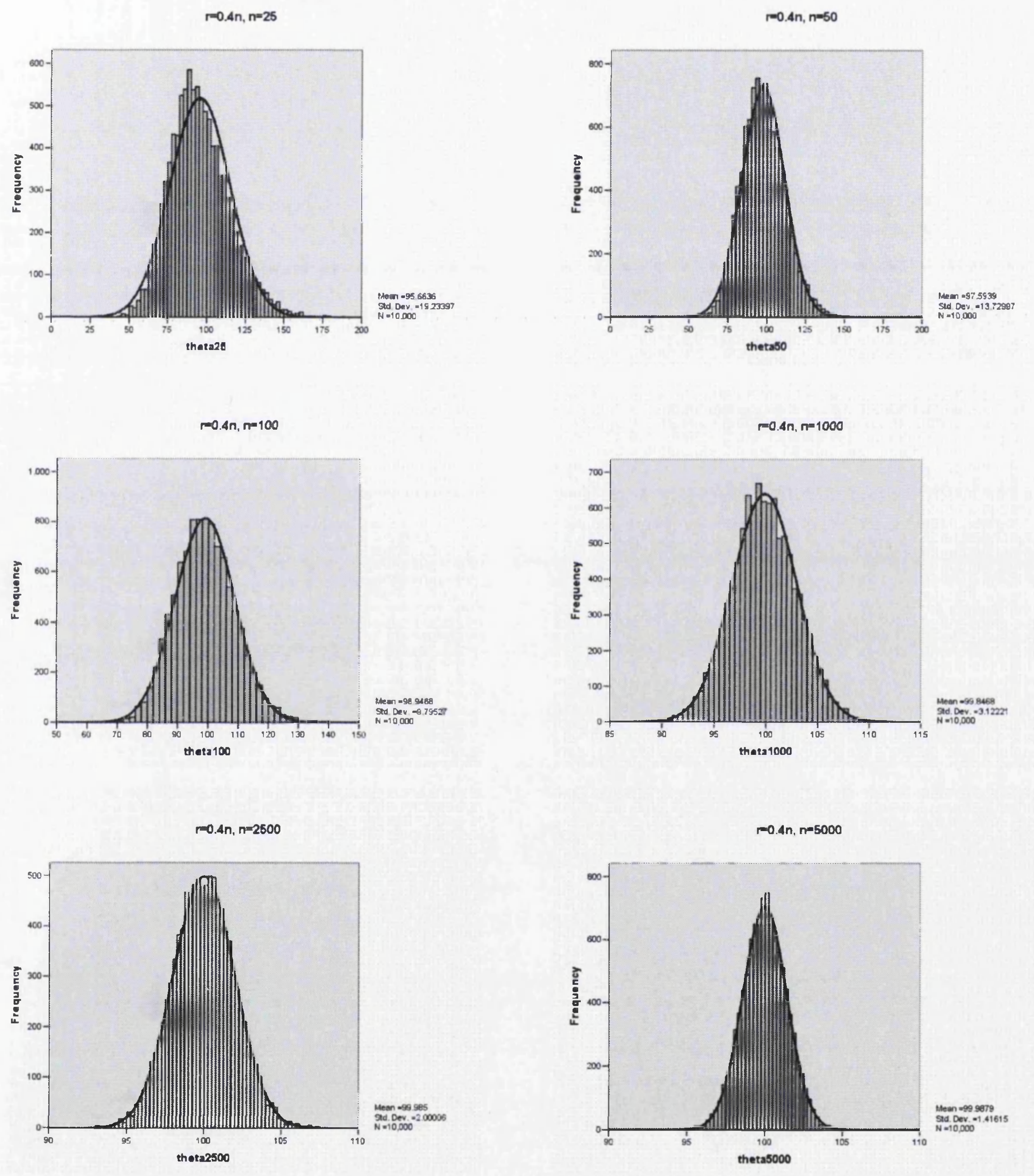

Figure 3.2: Histograms of $\hat{\theta}_{0.4n}$ for various n, for Weibull data generated with $\theta = 100, \beta = 2$.

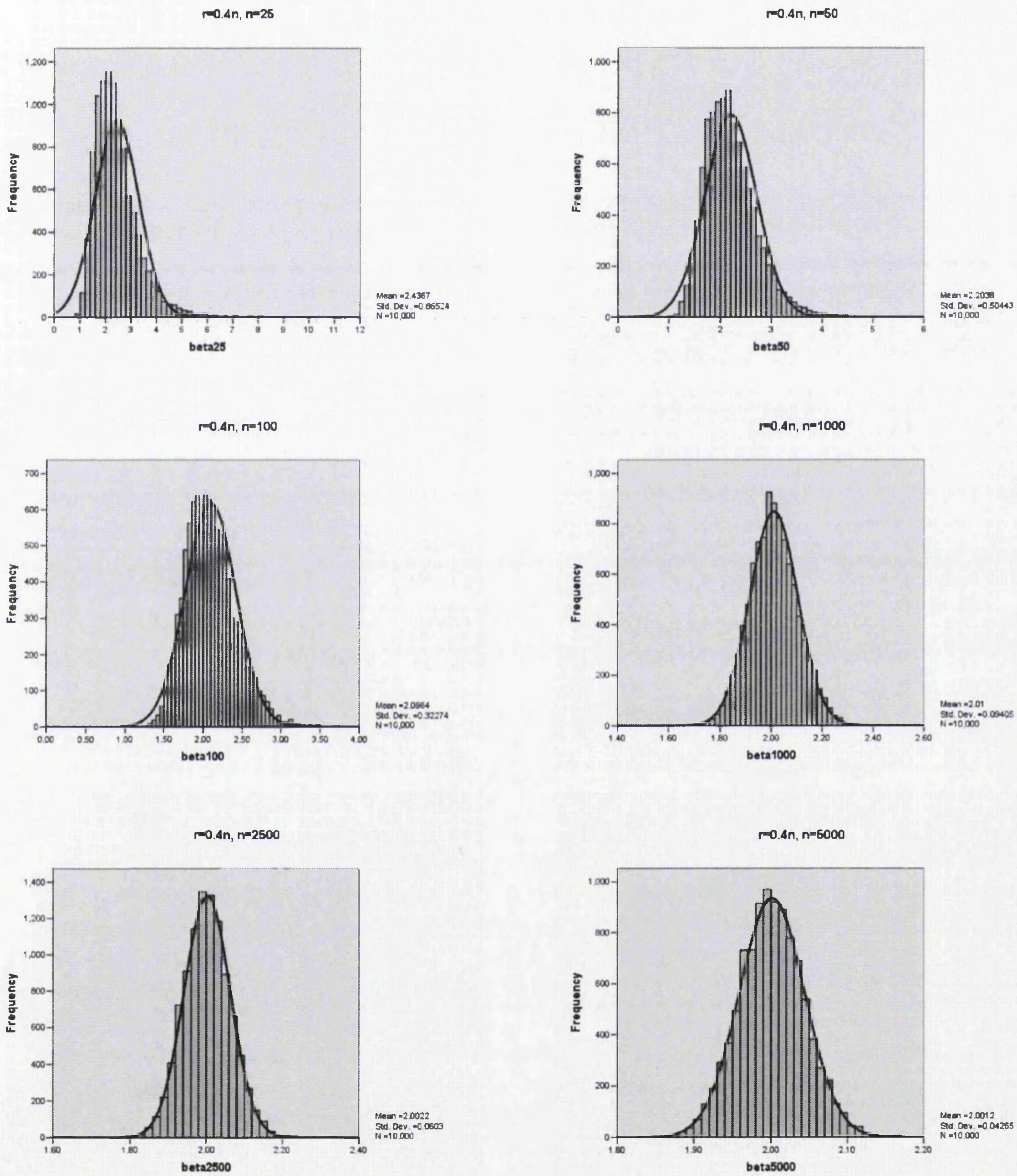

Figure 3.3: Histograms of $\hat{\beta}_{0.4n}$ for various n, for Weibull data generated with $\theta = 100, \beta = 2$.

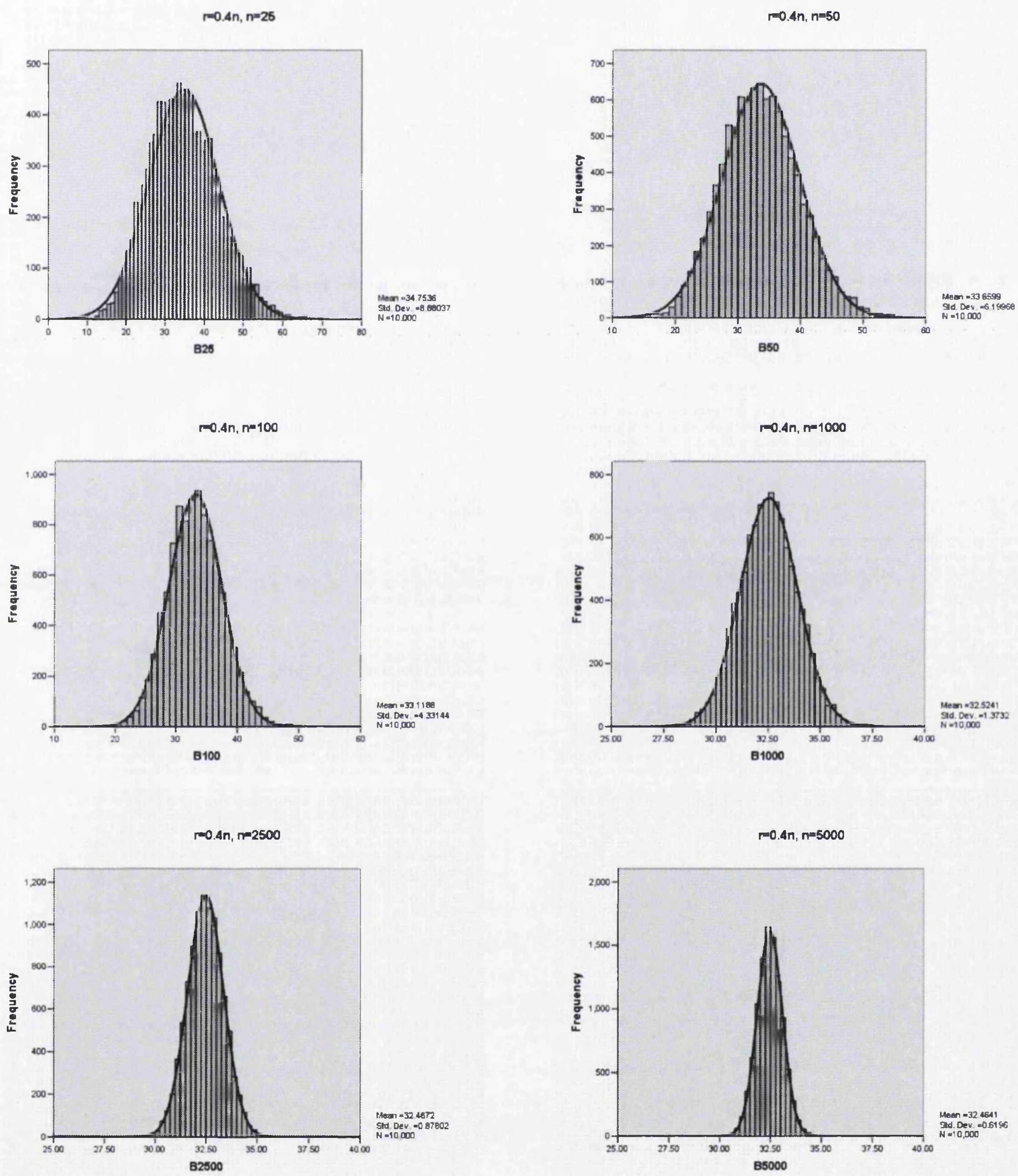

Figure 3.4: Histograms of $\hat{B}_{0.1,0.4n}$ for various n, for Weibull data generated with $\theta = 100, \beta = 2$.

n	$\widehat{\theta}_{0.4n}$					95% prob. intervals		
	g_1	$Z(g_1)$	g_2	$Z(g_2)$	K^2	Below	Within	Above
25	0.5541	21.1884	3.6228	9.8450	545.8730	225	9514	261
50	0.3911	15.4245	3.2575	4.7101	260.1012	231	9523	246
100	0.2495	10.0407	3.1888	3.5677	113.5438	239	9522	239
1000	0.0920	3.7497	3.0434	0.9050	14.8791	267	9495	238
2500	0.0632	2.5796	3.1016	2.0126	10.7047	256	9486	258
5000	0.0017	0.0714	3.0111	0.2618	0.0736	266	9480	254
n	$\widehat{\beta}_{0.4n}$					95% prob. intervals		
	g_1	$Z(g_1)$	g_2	$Z(g_2)$	K^2	Below	Within	Above
25	1.6682	48.8155	8.7945	35.0824	3613.7228	0	8291	1709
50	0.9562	33.2790	4.7187	19.7727	1498.4534	14	8854	1132
100	0.6578	24.5917	3.8467	12.4005	758.5249	43	9174	783
1000	0.1959	7.9283	3.0797	1.6039	65.4299	142	9476	382
2500	0.1565	6.3554	3.1164	2.2852	45.6132	214	9461	325
5000	0.1110	4.5192	3.1446	2.7947	28.2331	234	9458	308
n	$\widehat{B}_{0.1,0.4n}$					95% prob. intervals		
	g_1	$Z(g_1)$	g_2	$Z(g_2)$	K^2	Below	Within	Above
25	0.3535	14.0254	3.0843	1.6899	199.5689	60	9333	607
50	0.2481	9.9873	3.0312	0.6648	100.1875	92	9435	473
100	0.2183	8.8152	3.0285	0.6107	78.0799	119	9473	408
1000	0.0527	2.1529	3.0191	0.4233	4.8143	221	9511	268
2500	0.0265	1.0827	3.0116	0.2716	1.2460	250	9449	301
5000	0.0608	2.4829	3.0627	1.2785	7.7991	245	9471	284

Table 3.3: Summary statistics for $\hat{\theta}_{0.4n}, \hat{\beta}_{0.4n}$ and $\hat{B}_{0.1,0.4n}$ for various n, for Weibull data generated with $\theta = 100, \beta = 2$.

n	$\widehat{\theta}_r$				
	$r = 0.2n$	$r = 0.4n$	$r = 0.6n$	$r = 0.8n$	$r = 1.0n$
25	4910.1470	545.8730	64.2378	18.3207	19.6841
50	3199.5568	260.1012	28.9901	26.1206	30.5296
100	1460.4282	113.5438	25.4638	15.0040	19.6485
1000	94.0605	14.8791	15.4257	9.0157	9.8398
2500	48.9845	10.7047	**1.6297**	**0.0149**	**0.6302**
5000	34.8682	**0.0736**	**0.2704**	**1.3046**	**2.0840**
n	$\widehat{\beta}_r$				
	$r = 0.2n$	$r = 0.4n$	$r = 0.6n$	$r = 0.8n$	$r = 1.0n$
25	8509.0987	3613.7228	2655.9902	1696.8138	931.6946
50	3604.3806	1498.4534	975.7929	652.6939	470.9069
100	1263.8783	758.5249	345.4412	283.6994	301.5820
1000	137.5339	65.4299	26.9748	17.8407	6.4525
2500	40.0567	45.6132	10.7781	**5.5650**	**1.4498**
5000	21.1457	28.2331	13.6265	**4.6236**	**5.5368**
n	$\widehat{B}_{0.1,r}$				
	$r = 0.2n$	$r = 0.4n$	$r = 0.6n$	$r = 0.8n$	$r = 1.0n$
25	195.7168	199.5689	224.2453	253.0174	252.8119
50	92.8809	100.1875	113.7228	115.5903	128.4058
100	70.1326	78.0799	83.6202	103.4977	107.0821
1000	**5.0896**	**4.8143**	**5.0348**	8.6360	8.7274
2500	**1.8647**	**1.2460**	**0.1048**	**0.2265**	**0.1188**
5000	9.1868	7.7991	**5.7883**	**4.1265**	**4.0784**

Table 3.4: K^2 statistics for $\hat{\theta}_r, \hat{\beta}_r$ and $\hat{B}_{0.1,r}$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 2$.

3.2.2 Simulation Study: the Weibull Distribution

Table 2.6, based on the ball bearings data where $n = 23$, shows the approximate 95% confidence intervals for the Weibull parameters and $B_{0.1}$ assuming asymptotic theory for the MLEs. Now we consider the sampling distributions of $\widehat{\theta}_r, \widehat{\beta}_r$ and $\widehat{B}_{0.1,r}$, bearing in mind that the theory means symmetrical confidence intervals.

We begin with parameter values $\theta = 100, \beta = 2$ and set, say, $r = 0.4n$. The resultant summary statistics are given in Table 3.3, while the histograms are presented in Figures 3.2, 3.3 and 3.4 for θ, β and $B_{0.1}$ respectively. As in previous studies, we see the marginal distributions of MLEs become more Normal (but rarely to the extent that they would be regarded as acceptably so) as the sample size increases. The coverage of the probability intervals are good (close to 9500) for $\widehat{\theta}_{0.4n}$ and $\widehat{B}_{0.1,0.4n}$ for all n we have considered, but this is far from the case for $\widehat{\beta}_{0.4n}$, most clearly when $n \leq 100$. Moreover, there is a much larger number of the MLEs of β falling above the upper limit than below the lower limit, implying right skewness of the distributions of $\widehat{\beta}_r$. As a result, the distributions of $\widehat{B}_{0.1,0.4n}$ also seem to skew to the right, and we only approach symmetry when $n = 5000$. Besides these outputs, Table 3.4 furnishes the assessment of Normality when data is from a Weibull distribution. In general, and entirely as expected, we obtain smaller K^2 values with increasing r and n. More bold values are found in the distribution of $\widehat{B}_{0.1,r}$; however, these always correspond to large sample sizes.

Since β controls the shape of a Weibull distribution, it is often the quantity of interest in real-life situations, and we next consider the rate at which its distribution converges to Normality, for different shape parameter values and censoring levels.

Focus on $\widehat{\beta}_r$

We have already seen that, with $\beta = 2$, the non-Normality in the distribution of $\widehat{\beta}_r$ was partially attributable to the problem of right skewness, particularly in small samples ($n \leq 100$). But, because $\beta = 2 > 1$ implies an increasing failure rate, it might be the case that the line $\beta = 1$ acts as a lower limit to the simulated values of β, and consequently, we were more likely to observe large estimates of β, leading to a right skewed sample. This gives rise to the following question: would changing the nature of the data, as determined by the shape parameter β, reduce the (right) skewness of the distribution of $\widehat{\beta}_r$?

We now consider some alternative shape parameter values, keeping the scale parameter constant (at 100), to assess the extent to which these conclusions can be regarded as typical. We take $\beta = 0.5$ (negative aging/improvement over time) and 4 (positive aging/deterioration over time), and $r = 0.4n$ (as before); see Figure 1.5 for the effect of varying β on the shape of the Weibull pdf. The summary statistics for properties of $\widehat{\beta}_{0.4n}$ are listed in Tables 3.5 and 3.6 respectively. We observe values in striking resemblance between these tables and Table 3.3, and that the distributions remain right skewed; this is then confirmed by the associated histograms (see Figures 3.5 and 3.6). There is generally reasonable percentage of the 10^4

n	g_1	$Z(g_1)$	g_2	$Z(g_2)$	K^2	95% prob. intervals		
						Below	Within	Above
25	1.8038	51.1765	9.9896	37.4427	4020.9896	0	8257	1743
50	0.9506	33.1301	4.7138	19.7399	1487.2668	10	8908	1082
100	0.6346	23.8463	3.7010	10.7796	684.8439	45	9146	809
1000	0.1996	8.0768	3.0535	1.1009	66.4463	151	9487	362
2500	0.1005	4.0957	3.0038	0.1143	16.7879	202	9483	315
5000	0.0632	2.5795	2.9834	-0.3049	6.7467	205	9484	311

Table 3.5: Summary statistics for $\hat{\beta}_{0.4n}$ for various n, for Weibull data generated with $\theta = 100, \beta = 0.5$.

n	g_1	$Z(g_1)$	g_2	$Z(g_2)$	K^2	95% prob. intervals		
						Below	Within	Above
25	1.6642	48.7442	8.4814	34.3752	3557.6553	0	8315	1685
50	0.9134	32.1286	4.5802	18.8038	1385.8331	13	8832	1155
100	0.6012	22.7578	3.5640	9.1094	600.8985	46	9201	753
1000	0.1811	7.3386	3.0843	1.6897	56.7108	158	9485	357
2500	0.1091	4.4440	3.0943	1.8776	23.2744	214	9465	321
5000	0.0621	2.5352	3.0133	0.3062	6.5209	190	9511	299

Table 3.6: Summary statistics for $\hat{\beta}_{0.4n}$ for various n, for Weibull data generated with $\theta = 100, \beta = 4$.

replications of $\widehat{\beta}_{0.4n}$ within the 95% probability limits, derived from the known parameter values. However, for $n < 100$, nearly all of the excluded estimates are greater than the upper limit, revealing a severe non-symmetry in small samples. Indeed, when $n = 25$, we notice that for both β values considered, the remaining 5% all are above the upper limit. A reduction in skewness can be observed as n increases. Also tabulated in Tables 3.7 and 3.8 are the K^2 values for various r and n, gradually falling below 5.9915 as r and n approach infinity. We see that we might be prepared to accept that the sample distribution does, in fact, follow a Normal distribution at $n \geq 2500$ for $\beta > 1$, and at $n \geq 5000$ for $\beta < 1$.

Therefore, the sampling distributions of the MLEs become more Normally distributed when the shape parameter value increases. In general, these distributions seem to be (right) skewed, and the convergence to Normality is slow, and we can only sensibly assume Normality when $n \geq 2500$.

3.2.3 Simulation Study: the Burr Distribution

Table 2.13 shows the approximate 95% confidence intervals for α, τ and $B_{0.1}$ for the arthritic patients data where $n = 50$, assuming asymptotic theory for the MLEs. Nevertheless, are these asymptotic assumptions suitable in the inference of a sample as small as the arthritic patients data? To assess such assumptions when data is from a Burr distribution, we choose the values $\alpha = 4$ and $\tau = 3$. Table 3.9 presents, for $r = 0.6n$, the relevant summary statistics for $\widehat{\alpha}_{0.6n}, \widehat{\tau}_{0.6n}$ and $\widehat{B}_{0.1,0.6n}$ based on 10^4 replications. The coverage

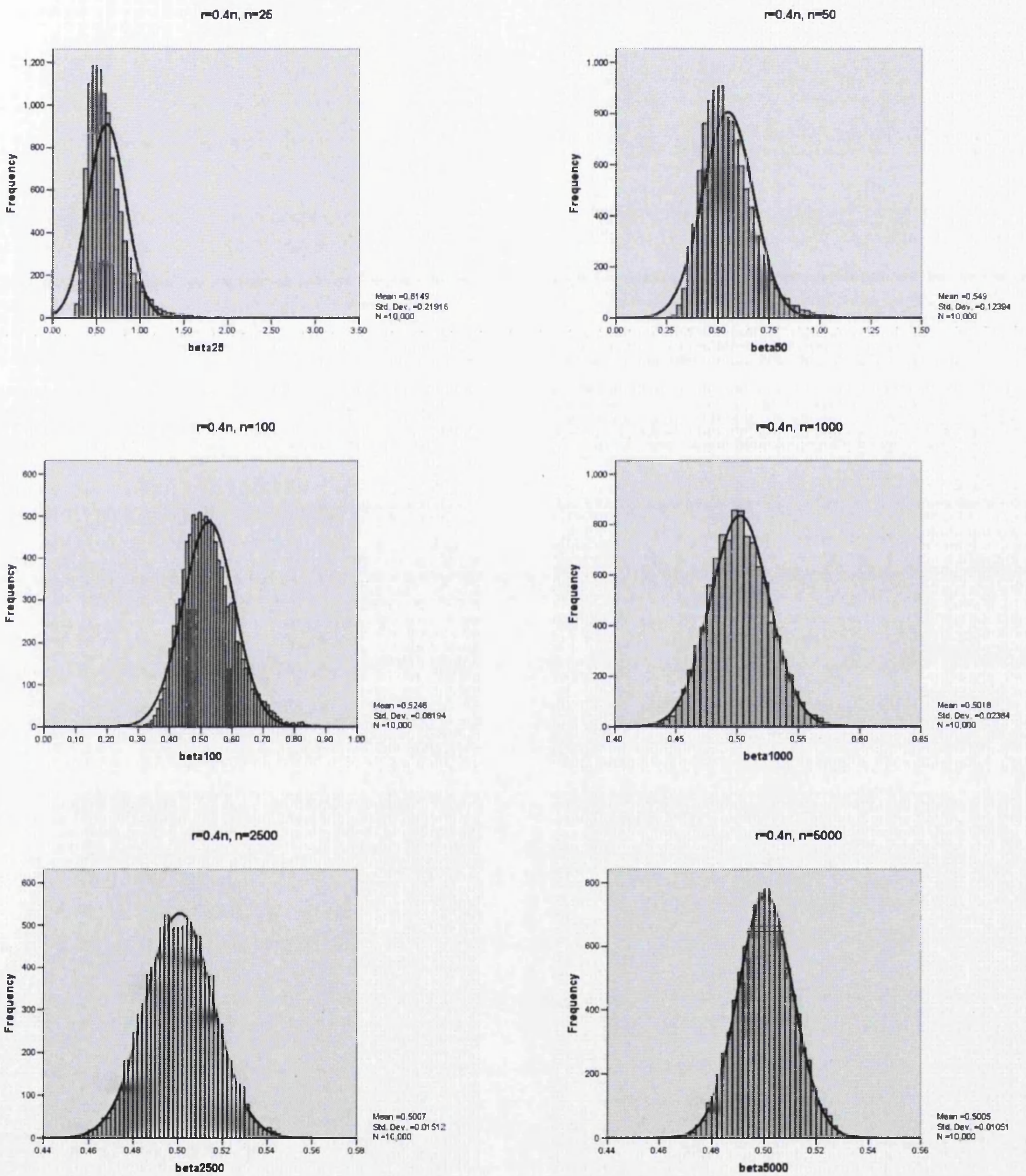

Figure 3.5: Histograms of $\hat{\beta}_{0.4n}$ for various n, for Weibull data generated with $\theta = 100, \beta = 0.5$.

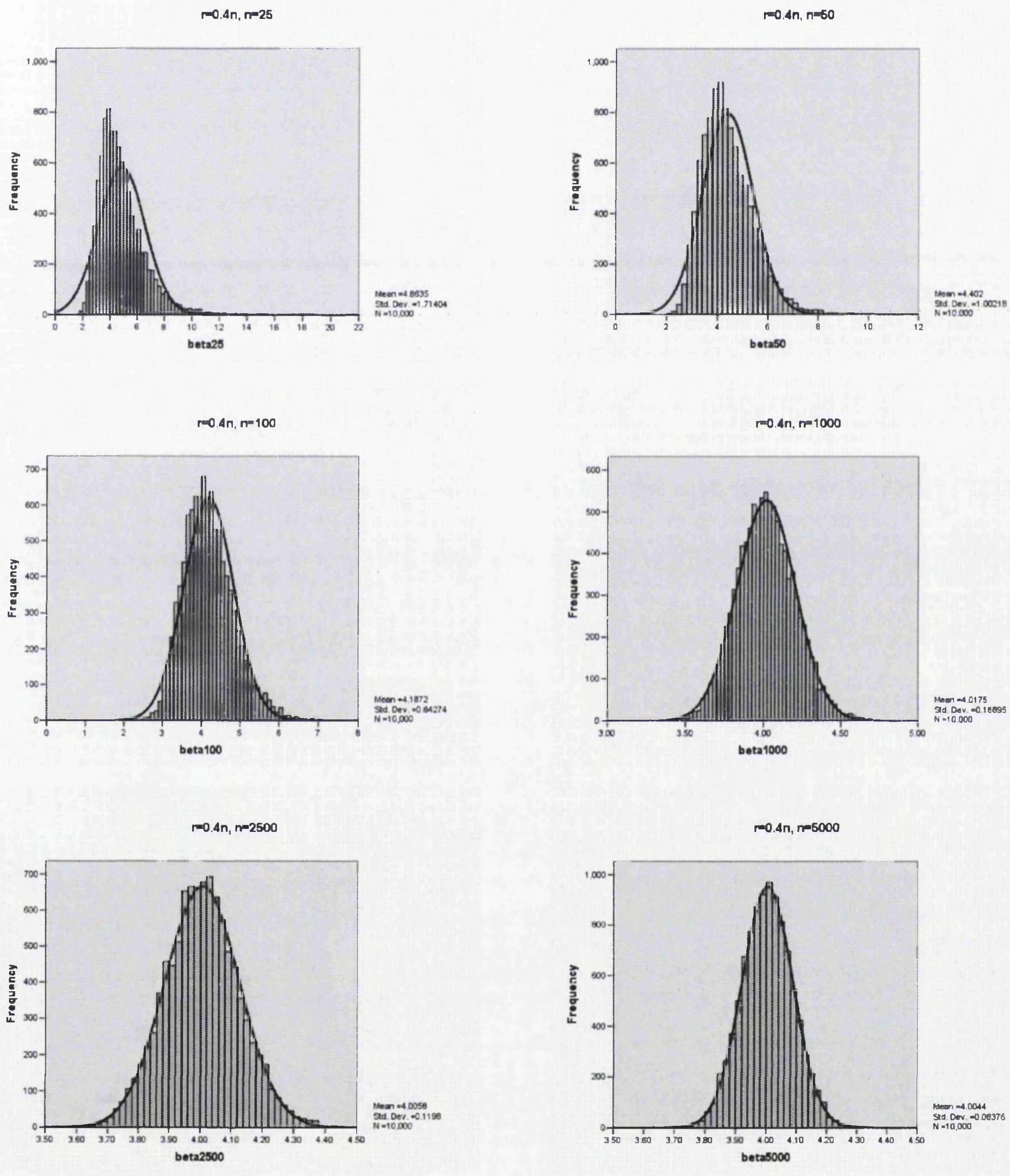

Figure 3.6: Histograms of $\hat{\beta}_{0.4n}$ for various n, for Weibull data generated with $\theta = 100, \beta = 4$.

n	$r=0.2n$	$r=0.4n$	$r=0.6n$	$r=0.8n$	$r=1.0n$
25	8450.8815	4020.9896	1998.0479	1466.5978	1140.0336
50	3234.4068	1487.2668	865.5300	667.0570	484.0441
100	1730.1939	684.8439	406.0410	271.7092	263.7248
1000	119.3940	66.4463	43.1624	26.8985	29.5334
2500	50.2340	16.7879	8.5018	8.5729	6.3537
5000	19.1902	6.7467	**4.4144**	**0.9676**	**2.1967**

Table 3.7: K^2 statistics for $\hat{\beta}_r$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 0.5$.

n	$r=0.2n$	$r=0.4n$	$r=0.6n$	$r=0.8n$	$r=1.0n$
25	6664.5081	3557.6553	1737.1188	1177.1974	1111.8832
50	3610.2131	1385.8331	947.6087	586.4931	571.7761
100	1436.0637	600.8985	433.5450	284.8207	234.3522
1000	157.7135	56.7108	29.1919	27.4829	25.4924
2500	61.2223	23.2744	9.1173	**3.4602**	**5.7275**
5000	20.5511	6.5209	7.7274	11.9036	7.3604

Table 3.8: K^2 statistics for $\hat{\beta}_r$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 4$.

n	$\widehat{\alpha}_{0.6n}$					95% prob. intervals		
	g_1	$Z(g_1)$	g_2	$Z(g_2)$	K^2	Below	Within	Above
25	39.3826	152.5583	2203.9356	75.0381	28904.7597	0	7916	2084
50	4.9860	83.7584	94.9171	62.0716	10868.3500	0	8558	1442
100	1.3586	42.7943	6.4711	28.4966	2643.4130	1	8944	1055
1000	0.3838	15.1543	3.3711	6.4622	271.4116	114	9489	397
2500	0.2498	10.0547	3.0741	1.4964	103.3366	143	9494	363
5000	0.1373	5.5848	2.9990	0.0162	31.1901	195	9485	320
n	$\widehat{\tau}_{0.6n}$					95% prob. intervals		
	g_1	$Z(g_1)$	g_2	$Z(g_2)$	K^2	Below	Within	Above
25	1.1069	37.1036	5.4731	24.1841	1961.5480	14	8680	1306
50	0.7171	26.4441	3.9511	13.4749	880.8608	35	9112	853
100	0.4502	17.5727	3.2013	3.7796	323.0863	59	9260	681
1000	0.1753	7.1087	3.1021	2.0228	54.6247	165	9512	323
2500	0.1165	4.7421	2.9595	-0.8088	23.1418	190	9486	324
5000	0.0414	1.6924	3.0010	0.0577	2.8676	207	9507	286
n	$\widehat{B}_{0.1,0.6n}$					95% prob. intervals		
	g_1	$Z(g_1)$	g_2	$Z(g_2)$	K^2	Below	Within	Above
25	0.1538	6.2490	3.0050	0.1391	39.0696	134	9375	491
50	0.1166	4.7476	2.9348	-1.3404	24.3365	142	9487	371
100	0.0668	2.7274	2.8715	-2.7720	15.1225	157	9480	363
1000	0.0751	3.0633	3.0256	0.5522	9.6889	203	9496	301
2500	0.0156	0.6370	2.9835	-0.3040	0.4982	242	9489	269
5000	-0.0182	-0.7434	2.9884	-0.2028	0.5937	227	9512	261

Table 3.9: Summary statistics for $\hat{\alpha}_{0.6n}, \hat{\tau}_{0.6n}$ and $\hat{B}_{0.1,0.6n}$ for various n, for Burr data generated with $\alpha = 4, \tau = 3$.

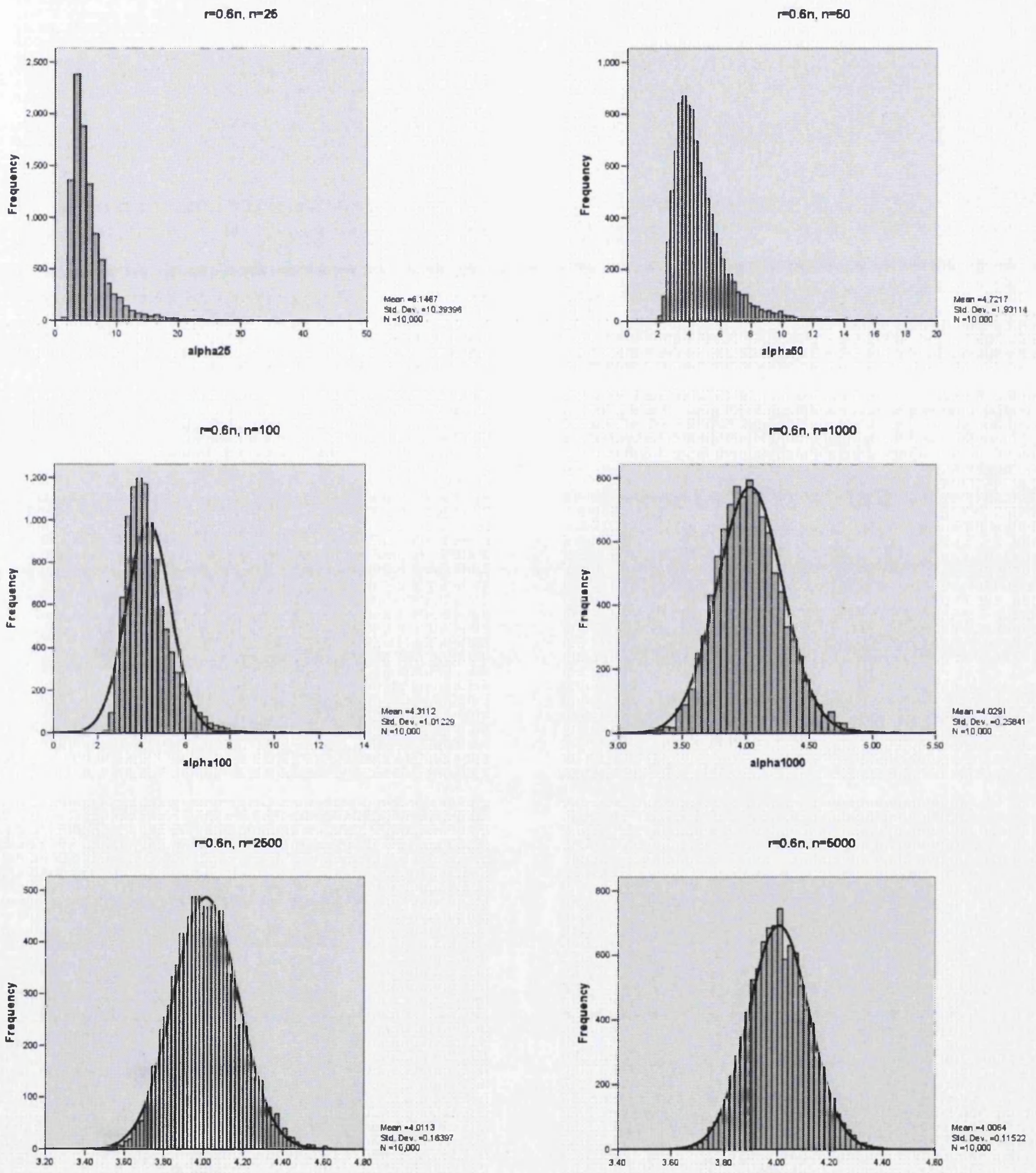

Figure 3.7: Histograms of $\hat{\alpha}_{0.6n}$ for various n, for Burr data generated with $\alpha = 4, \tau = 3$.

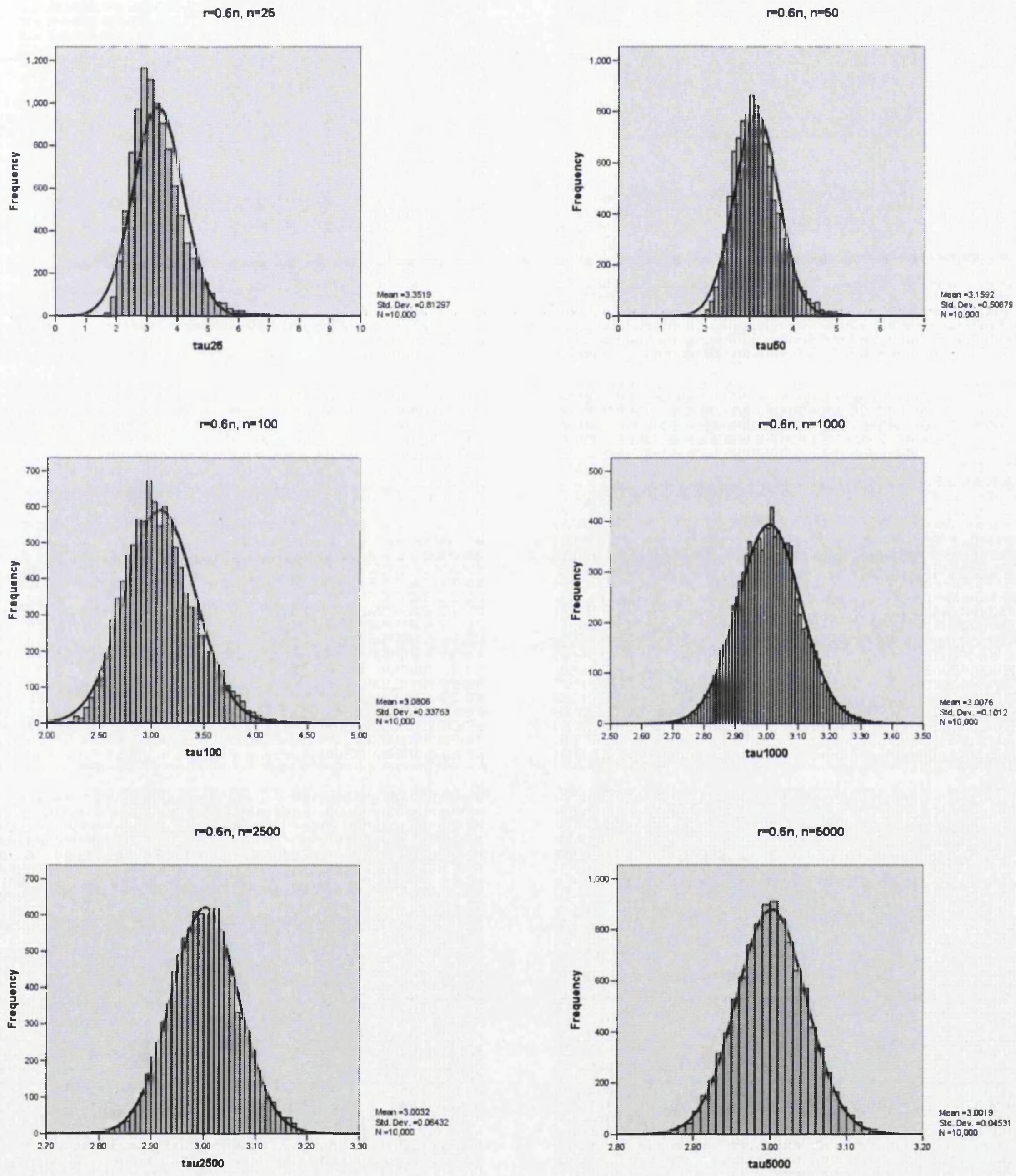

Figure 3.8: Histograms of $\hat{\tau}_{0.6n}$ for various n, for Burr data generated with $\alpha = 4, \tau = 3$.

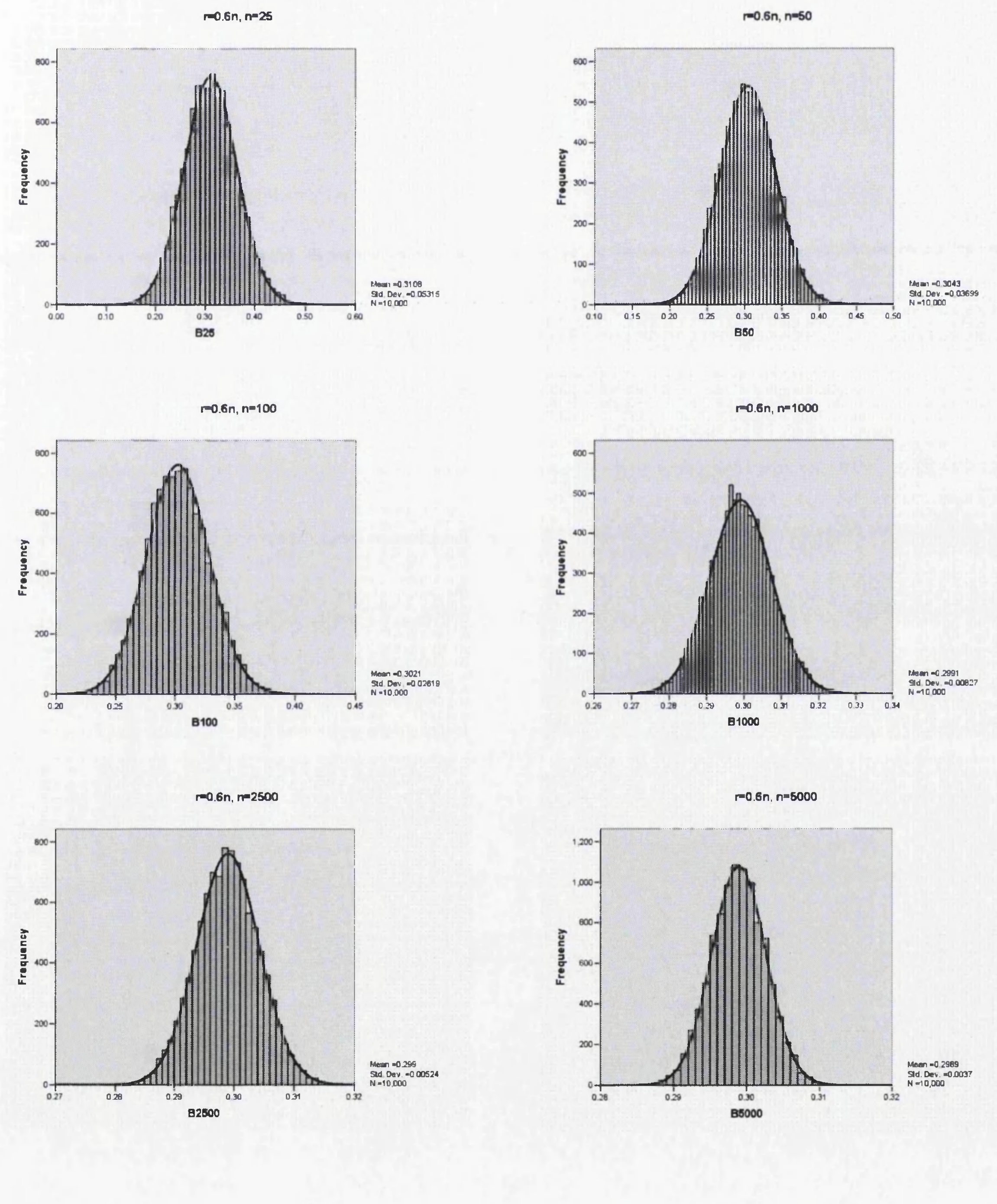

Figure 3.9: Histograms of $\hat{B}_{0.1,0.6n}$ for various n, for Burr data generated with $\alpha = 4, \tau = 3$.

of the probability intervals for $\widehat{\alpha}_{0.6n}$ and $\widehat{\tau}_{0.6n}$ improve as n increases, but more estimates outside the limits are above the upper limit, again suggesting right skewness of the MLEs. In contrast, the coverage and spread of the estimates of $B_{0.1}$ are much better than either parameters. Clearly, we are led to the same conclusions as for the previous two lifetime models, where the distributions of the MLEs becoming more Normal as the sample size increases. This is apparent from the histograms (Figures 3.7 to 3.9), and re-enforced by the K^2 statistics. Note also for graphical convenience, we truncate, in Figure 3.7, the α-axis at 50 when $n = 25$; this excludes the following 29 estimates

50.5658	53.2722	53.2722	55.1094	55.3440	57.3566	58.5072	59.8007
63.5590	64.4777	64.7645	64.7666	64.7967	64.7967	65.5133	66.8281
67.7875	71.6625	72.7258	73.9242	74.4186	77.0094	87.8581	91.6693
197.3220	209.7272	238.0291	478.6130	675.6811			

,

and at 20 when $n = 50$; excluding the following 6 estimates

20.6613	20.8016	21.5761	22.6671	40.8366	61.3055

.

As a result, the Normal curves are omitted in the first two plots, but would clearly not be a good fit in either case. Table 3.10 continues this study for various censoring levels using, as before, 10^4 estimates of α, τ and $B_{0.1}$. We note that for no sample size considered is the Normal distribution regarded as a suitable model for the distribution of $\widehat{\alpha}_r$, as indicated by K^2 statistics well above 5.9915 in Table 3.10, whereas the analysis of $\widehat{\tau}_r$ reports a few bold entries. More such entries are observed in the consideration of $\widehat{B}_{0.1,r}$, but all are associated with large r and n.

As with the Weibull distribution, it seems that right skewness is typical in the distributions of the MLEs for Burr shape parameters for small samples, typically, less than 1000. Tables 3.11 and 3.12 are based on $\alpha = 0.9, \tau = 3$, whilst Tables 3.13 and 3.14 are based on $\alpha = 4, \tau = 0.9$; we see that, regardless of the shape parameter values chosen, the progress towards Normality is quite slow, especially in the case of $\widehat{\alpha}_r$, as indicated by Figures 3.10 and 3.11. In fact, we should not formally accept the hypothesis of Normality even when $n = 5000$ for α, and when $n = 2500$ for τ.

n	$\widehat{\alpha}_r$				
	$r = 0.2n$	$r = 0.4n$	$r = 0.6n$	$r = 0.8n$	$r = 1.0n$
25	39737.9208	27080.4244	28904.7597	13652.1224	3849.8656
50	39780.0164	18911.5790	10868.3500	3347.9557	1686.0013
100	23125.6237	6267.3671	2643.4130	1703.6368	873.1987
1000	1734.0031	562.9673	271.4116	128.7738	53.2084
2500	599.1690	222.5042	103.3366	30.0695	22.1464
5000	268.9492	86.5960	31.1901	14.7940	13.4892
n	$\widehat{\tau}_r$				
	$r = 0.2n$	$r = 0.4n$	$r = 0.6n$	$r = 0.8n$	$r = 1.0n$
25	5593.8574	3582.4246	1961.5480	1284.3626	765.3068
50	6746.5910	1520.3691	880.8608	491.2848	340.5830
100	1785.8462	670.2142	323.0863	214.1388	134.5801
1000	158.5951	57.0847	54.6247	27.0103	15.1617
2500	47.0871	26.3294	23.1418	10.7947	**1.7895**
5000	20.1953	6.7533	**2.8676**	6.0837	**5.0929**
n	$\widehat{B}_{0.1,r}$				
	$r = 0.2n$	$r = 0.4n$	$r = 0.6n$	$r = 0.8n$	$r = 1.0n$
25	25.0947	26.3841	39.0696	56.7920	64.8937
50	15.8024	17.8928	24.3365	24.9160	27.2481
100	11.6343	13.7021	15.1225	8.1730	7.6625
1000	**4.1282**	**4.4379**	9.6889	6.2888	11.2315
2500	**0.6176**	**0.6931**	**0.4982**	**0.3401**	**1.2610**
5000	**2.1233**	**2.4685**	**0.5937**	**1.1200**	**1.1565**

Table 3.10: K^2 statistics for $\hat{\alpha}_r, \hat{\tau}_r$ and $\hat{B}_{0.1,r}$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 3$.

n	g_1	$Z(g_1)$	g_2	$Z(g_2)$	K^2	95% prob. intervals		
						Below	Within	Above
25	1.7930	50.9939	11.1209	39.2944	4144.4312	74	8938	988
50	0.9319	32.6276	5.2261	22.8785	1587.9825	98	9200	702
100	0.5151	19.8528	3.6995	10.7627	509.9690	122	9373	505
1000	0.1178	4.7960	3.0142	0.3247	23.1074	208	9433	359
2500	0.1050	4.2772	2.9894	-0.1812	18.3272	208	9520	272
5000	0.0862	3.5149	3.0606	1.2394	13.8904	213	9515	272

Table 3.11: Summary statistics for $\hat{\alpha}_{0.6n}$ for various n, for Burr data generated with $\alpha = 0.9, \tau = 3$.

n	$r = 0.2n$	$r = 0.4n$	$r = 0.6n$	$r = 0.8n$	$r = 1.0n$
25	39744.9162	15684.7942	4144.4312	1527.0629	612.6169
50	35684.9476	5450.1748	1587.9825	474.4686	303.1235
100	10009.8376	2561.5849	509.9690	167.7750	103.6142
1000	863.5744	138.5899	23.1074	18.3585	17.8973
2500	248.6987	79.2532	18.3272	14.6199	13.0677
5000	72.7707	24.0838	13.8904	7.9560	8.3689

Table 3.12: K^2 statistics for $\hat{\alpha}_r$ for various r, n, for Burr data generated with $\alpha = 0.9, \tau = 3$.

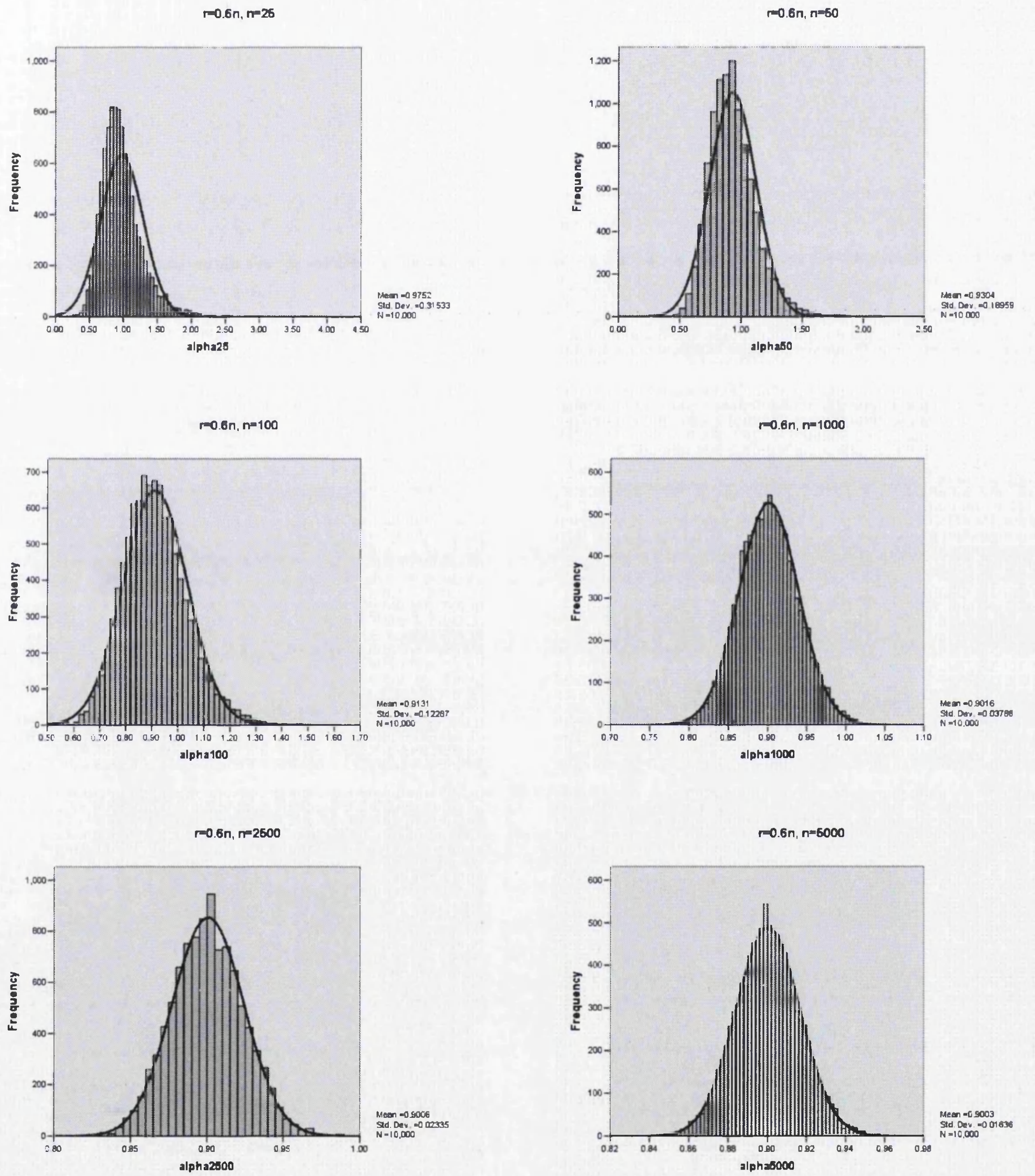

Figure 3.10: Histograms of $\hat{\alpha}_{0.6n}$ for various n, for Burr data generated with $\alpha = 0.9, \tau = 3$.

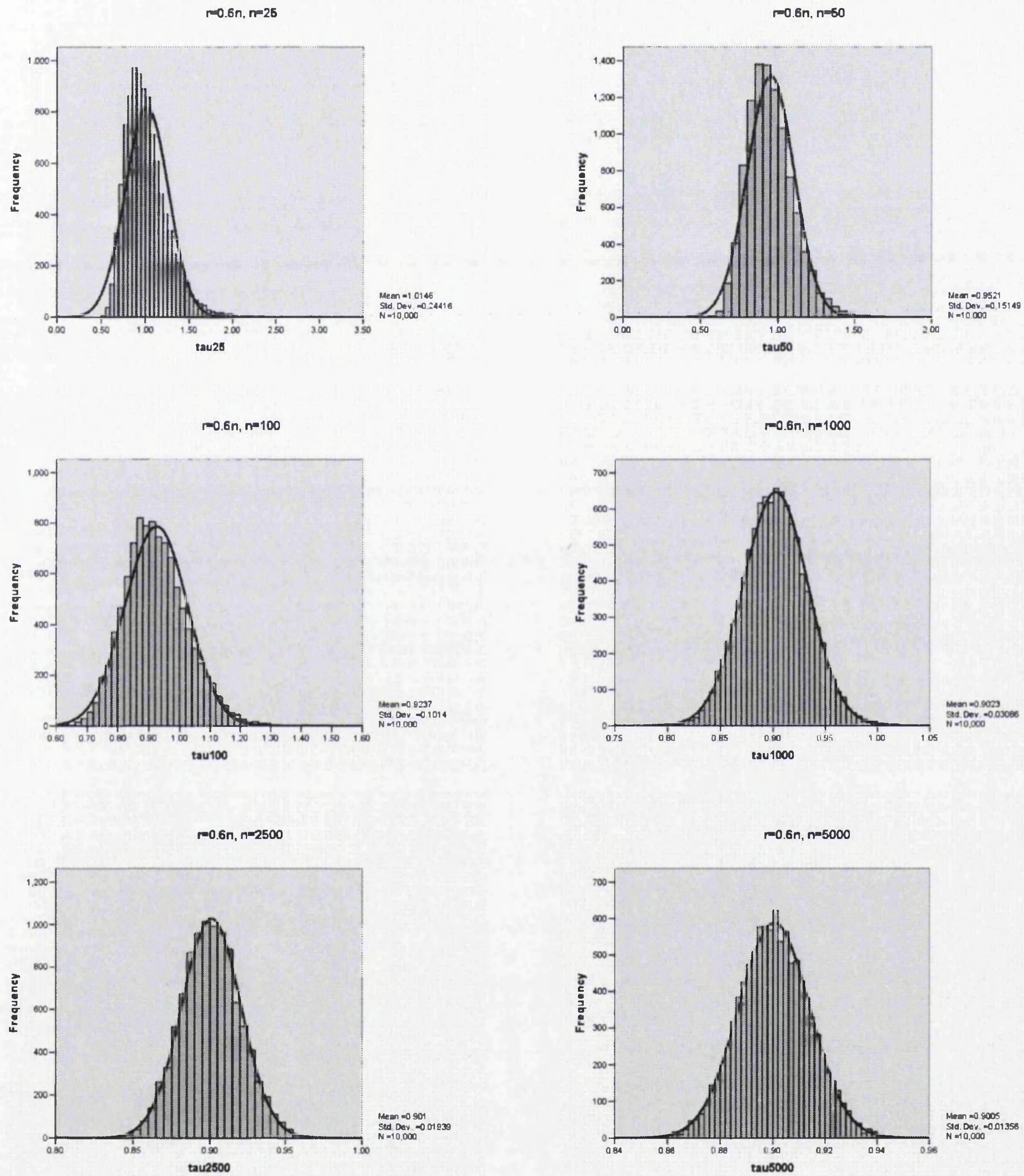

Figure 3.11: Histograms of $\hat{\tau}_{0.6n}$ for various n, for Burr data generated with $\alpha = 4, \tau = 0.9$.

n	g_1	$Z(g_1)$	g_2	$Z(g_2)$	K^2	95% prob. intervals		
						Below	Within	Above
25	1.3026	41.5977	6.7776	29.5870	2605.7590	5	8639	1356
50	0.7431	27.2386	4.0680	14.6034	955.1998	26	9098	876
100	0.4486	17.5128	3.3262	5.7894	340.2138	74	9274	652
1000	0.1663	6.7479	3.1102	2.1712	50.2485	178	9473	349
2500	0.1264	5.1427	2.9948	-0.0707	26.4528	191	9485	324
5000	0.0560	2.2851	3.0158	0.3567	5.3488	206	9512	282

Table 3.13: Summary statistics for $\hat{\tau}_{0.6n}$ for various n, for Burr data generated with $\alpha = 4, \tau = 0.9$.

n	$r = 0.2n$	$r = 0.4n$	$r = 0.6n$	$r = 0.8n$	$r = 1.0n$
25	5570.1977	3502.3529	2605.7590	1383.7825	895.7894
50	4281.4335	1350.6207	955.1998	579.8398	491.7778
100	1667.1451	703.7234	340.2138	206.5481	99.0376
1000	152.0832	74.7403	50.2485	14.5300	17.5078
2500	59.7999	43.9957	26.4528	7.0937	**3.2418**
5000	29.5796	13.3561	**5.3488**	**5.4923**	7.8803

Table 3.14: K^2 statistics for $\hat{\tau}_r$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 0.9$.

3.3 Tests of Bivariate Normality

A necessary, but not sufficient, condition for multivariate Normality is that each marginal distribution is univariate Normal. Hence, as we have proceeded here, it is usual to start with univariate tests for marginal Normality, at which detection of one non-Normal marginal implies that the joint distribution is non-Normal. A fair amount of work is available on tests of multivariate Normality, many of which are generalisation of univariate procedures. On the basis of power studies and the ease of implementation, perhaps the most widely referenced multivariate Normality test is due to Mardia; see Gnanadesikan (1977) and Thode (2002) for excellent summaries of the merits of this test. The sample estimates of multivariate skewness, denoted by $(g_{1,k})$ and kurtosis $(g_{2,k})$ (for k variates), were first presented by Mardia (1970), defined, respectively, at (2.23) and (3.12) therein; Mardia & Foster (1983) then proposed several omnibus tests based on these two measures, including the S_W^2 statistic, which, for $k = 2$, is

$$S_W^2 = \{W(g_{1,2})\}^2 + \{W(g_{2,2})\}^2,$$

in which $W(g_{1,2}), W(g_{2,2})$ are standardised bivariate measures of skewness and kurtosis using the Wilson-Hilferty approximation, as given, respectively, by (3.7) and (3.10) in Mardia & Foster (1983). The case $k = 2$ will apply to both the Weibull and Burr cases here; again, under the hypothesis that the joint distribution of the estimators is multivariate Normal, we have $S_W^2 \sim \chi_2^2$, with a corresponding assessment of joint Normality of $\left(\widehat{\theta}_r, \widehat{\beta}_r\right)$ in the case of Weibull, and, of $(\widehat{\alpha}_r, \widehat{\tau}_r)$ for Burr distribution, using the critical value $-2\ln\lambda$ for an upper tail probability of λ. Hence, $S_W^2 \leq 5.9915$ indicates that we can accept the hypothesis of

n	$g_{1,2}$	$W(g_{1,2})$	$g_{2,2}$	$W(g_{2,2})$	S_W^2	Within 95% prob. ellipse
25	1.7758	21.5755	12.2541	29.5184	1336.8399	8842
50	0.6014	13.8252	9.0716	11.0504	313.2491	9170
100	0.2410	9.1394	8.3841	4.4854	103.6476	9328
1000	0.0248	2.1522	8.1365	1.6955	7.5068	9495
2500	0.0060	-0.1593	7.9853	-0.1481	0.0473	9478
5000	0.0086	0.3270	8.0455	0.5996	0.4664	9464

Table 3.15: Summary statistics for $(\hat{\theta}_{0.6n}, \hat{\beta}_{0.6n})$ for various n, for Weibull data generated with $\theta = 100, \beta = 2$.

bivariate Normality at the 5% significance level. As before, by simply counting how many of the estimates of (θ, β) and (α, τ) within the corresponding probability ellipse derived from the true parameter values, we can judge the extent to which the Normal assumption is appropriate or justifiable.

3.3.1 Simulation Study: the Weibull Distribution

For our first example, we take parameter values to be $\theta = 100$ and $\beta = 2$. Knowing these true values, the scatter plots of $\widehat{\theta}_{0.6n}$ against $\widehat{\beta}_{0.6n}$, superimposed with the large-sample 95% probability ellipses, are presented in Figure 3.12 for varying n. These seem to suggest that as the sample size increases, the joint distributions of $\left(\widehat{\theta}_{0.6n}, \widehat{\beta}_{0.6n}\right)$ become more and more Normal - as indicated by the S_W^2 values in Table 3.15 - and are more uniformly spread around $(100, 2)$ - as shown by the number of replications enclosed in the probability regions. We also notice that, as a result of the right skewness in the distribution of $\widehat{\beta}_r$, as observed in the histograms, the sampling distribution of the Weibull MLEs is distinctly non-elliptical at $n = 25$ and 50. Moreover, from Table 3.16, which gives the S_W^2 values for various r and n, it can be deduced that the assumption of the χ_2^2 distribution as the null distributions of the S_W^2 statistics is inappropriate for small samples. In particular, the pattern observed here is entirely consistent with the findings in the corresponding univariate analyses, in which increasing censoring number and sample size leads to a lower S_W^2 value. In fact, we are in a position to accept formally the hypothesis of joint Normality only when $n \geq 1000$.

More numerical illustrations are shown in Table 3.17 (based on simulations with $\theta = 100, \beta = 0.5$) and Table 3.18 (based on simulations with $\theta = 100, \beta = 4$); a similar pattern is observed whether we have negative or positive aging over time, suggesting lack of Normality in small samples across the board. Furthermore, the rate at which the sampling distribution of $\left(\widehat{\theta}_r, \widehat{\beta}_r\right)$ approaches a Normal distribution increases when the shape parameter value increases. Specifically, we might be prepared to accept the hypothesis of joint Normality at $n \geq 1000$ for $\beta > 1$, but at $n \geq 2500$ for $\beta < 1$.

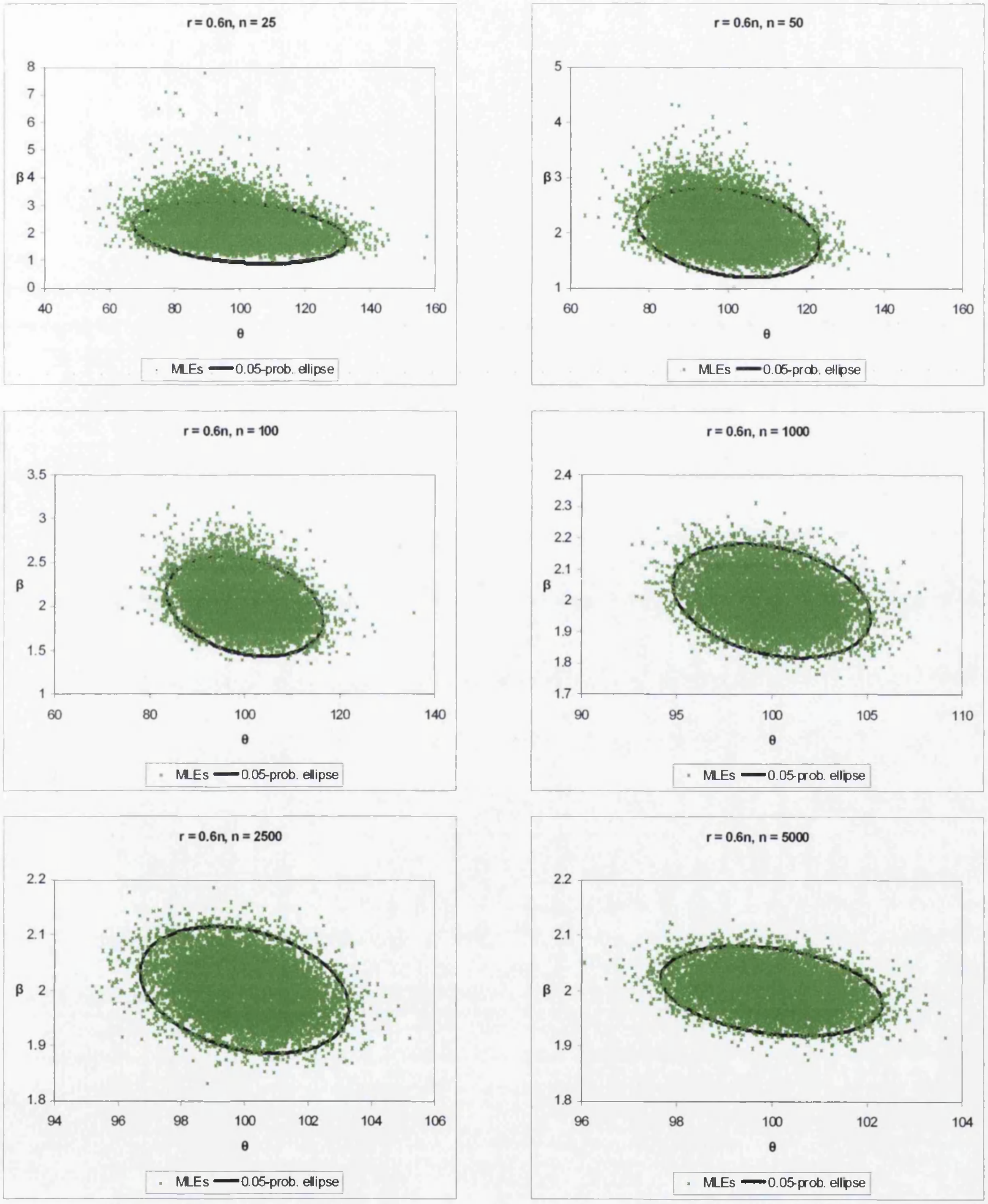

Figure 3.12: Scatter plots of $(\hat{\theta}_{0.6n}, \hat{\beta}_{0.6n})$, superimposed with asymptotic 0.05-probability ellipses, for various n, for Weibull data generated with $\theta = 100, \beta = 2$.

n	$r = 0.2n$	$r = 0.4n$	$r = 0.6n$	$r = 0.8n$	$r = 1.0n$
25	9509.1576	2551.6537	1336.8399	682.6428	343.1559
50	4897.6596	794.8631	313.2491	182.7758	148.8660
100	2030.4556	346.9915	103.6476	69.6449	101.4419
1000	140.3048	20.1564	7.5068	**5.0256**	**1.1946**
2500	61.0323	15.3622	**0.0473**	**0.6115**	**0.1085**
5000	32.8853	**4.8511**	**0.4664**	**0.6943**	**1.5280**

Table 3.16: S_W^2 statistics for the multivariate Normality of $(\hat{\theta}_r, \hat{\beta}_r)$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 2$.

n	$r = 0.2n$	$r = 0.4n$	$r = 0.6n$	$r = 0.8n$	$r = 1.0n$
25	18850.5749	5898.3124	1900.7956	1251.4989	1129.9458
50	12977.7339	2691.4239	774.7061	517.5551	505.2621
100	11739.0954	1554.1767	541.5757	259.7114	245.0951
1000	787.6916	81.0223	31.4821	20.0178	15.5689
2500	341.7878	28.1952	9.0086	**3.9388**	**4.0507**
5000	156.6740	11.0087	**0.8059**	**2.9406**	6.9522

Table 3.17: S_W^2 statistics for the multivariate Normality of $(\hat{\theta}_r, \hat{\beta}_r)$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 0.5$.

n	$r = 0.2n$	$r = 0.4n$	$r = 0.6n$	$r = 0.8n$	$r = 1.0n$
25	6466.0566	2230.1012	651.0079	361.8537	389.5068
50	4190.9002	693.4733	307.1663	171.9068	225.7919
100	1522.3346	220.8830	129.3909	75.8062	66.8740
1000	133.5082	14.1478	**5.7476**	**4.6976**	**4.0813**
2500	49.7458	7.9932	**0.2038**	**2.1638**	**1.0029**
5000	16.9615	**3.0223**	**0.3063**	**1.0016**	**0.2981**

Table 3.18: S_W^2 statistics for the multivariate Normality of $(\hat{\theta}_r, \hat{\beta}_r)$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 4$.

n	$g_{1,2}$	$W(g_{1,2})$	$g_{2,2}$	$W(g_{2,2})$	S_W^2	Within 95% prob. ellipse
25	90.8954	90.9781	297.3183	93.4697	17013.5927	8371
50	2.8237	25.8525	14.7995	37.9687	2109.9701	8937
100	1.1620	18.2027	10.3626	20.2367	740.8614	9137
1000	0.0902	5.4676	8.2545	3.0581	39.2468	9465
2500	0.0214	1.8572	7.9807	-0.2055	3.4913	9517
5000	0.0102	0.5813	8.0945	1.1952	1.7664	9493

Table 3.19: Summary statistics for $(\hat{\alpha}_{0.8n}, \hat{\tau}_{0.8n})$ for various n, for Burr data generated with $\alpha = 4, \tau = 3$.

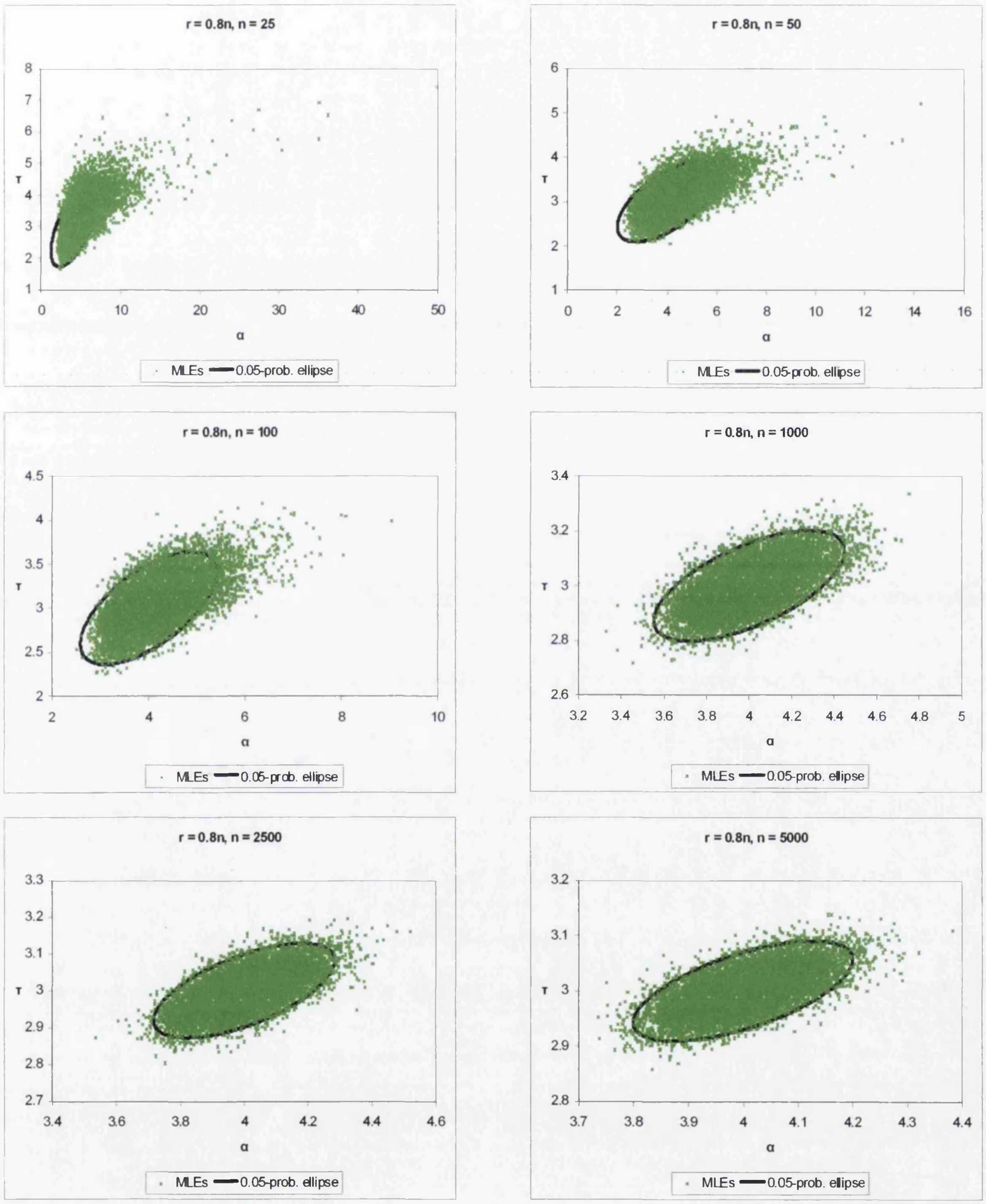

Figure 3.13: Scatter plots of $(\hat{\alpha}_{0.8n}, \hat{\tau}_{0.8n})$, superimposed with asymptotic 0.05-probability ellipses, for various n, for Burr data generated with $\alpha = 4, \tau = 3$.

n	$r = 0.2n$	$r = 0.4n$	$r = 0.6n$	$r = 0.8n$	$r = 1.0n$
25	215030.0009	59433.7760	85568.0385	17013.5927	2317.8424
50	215892.2746	36880.0460	15186.6148	2109.9701	761.1056
100	47954.9946	6781.8154	1548.0777	740.8614	312.7829
1000	1622.8513	268.1502	98.1737	39.2468	14.0501
2500	469.7332	100.6247	24.2774	**3.4913**	**1.1563**
5000	166.7399	26.5937	**3.4072**	**1.7664**	**2.0133**

Table 3.20: S_W^2 statistics for the multivariate Normality of $(\hat{\alpha}_r, \hat{\tau}_r)$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 3$.

n	$r = 0.2n$	$r = 0.4n$	$r = 0.6n$	$r = 0.8n$	$r = 1.0n$
25	215318.4273	17339.7582	3802.4482	2998.9998	2908.3681
50	151323.6236	3856.9657	1233.2114	601.9089	500.2051
100	10661.6125	1400.7537	375.8319	179.3075	143.6304
1000	357.7632	56.6306	15.7042	13.5700	10.0471
2500	84.0961	29.4680	7.9595	7.5043	9.6201
5000	21.5596	**5.5473**	**4.8758**	**1.3485**	**2.5061**

Table 3.21: S_W^2 statistics for the multivariate Normality of $(\hat{\alpha}_r, \hat{\tau}_r)$ for various r, n, for Burr data generated with $\alpha = 0.9, \tau = 3$.

3.3.2 Simulation Study: the Burr Distribution

We now perform a similar series of investigations with data generated from the Burr distribution. Figure 3.13 together with summaries in Table 3.19 show the simulation results based on 10^4 repetitions assuming $\alpha = 4$ and $\tau = 3$; this is consistent with conclusions drawn in the univariate tests, at which we require a very large sample size in order for the distribution of $(\widehat{\alpha}_{0.8n}, \widehat{\tau}_{0.8n})$ to be Normal. Even when 80% of the failures are observed, the scatter plots are not entirely consistent with elliptical probability regions for small samples; rather, they extend across to large values of α in a systematic fashion. In Table 3.20, we see smaller S_W^2 values with increasing r and n, but only a few are lower than 5.9915. Furthermore, Table 3.21 tabulates S_W^2 statistics for data generated with $\alpha = 0.9, \tau = 3$, while Table 3.22 is for data generated with $\alpha = 4, \tau = 0.9$; these results, once more, confirm the lack of Normality in small data sets, seemingly independent of the choice of parameter values.

n	$r = 0.2n$	$r = 0.4n$	$r = 0.6n$	$r = 0.8n$	$r = 1.0n$
25	164369.6145	170832.8569	147506.2618	5589.2740	4630.4679
50	195577.2053	24340.0167	8657.3802	3087.4125	949.4046
100	62413.5502	10765.1581	1599.4672	641.2627	230.0604
1000	1978.5703	224.9025	77.3924	19.3295	6.4604
2500	386.3123	93.0322	23.7892	**5.0736**	**0.5012**
5000	203.7194	36.3863	6.3676	**1.4882**	**2.5661**

Table 3.22: S_W^2 statistics for the multivariate Normality of $(\hat{\alpha}_r, \hat{\tau}_r)$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 0.9$.

3.4 Relative Likelihood Contour Plots

We have observed that Normality was not reached until samples were very large; this covers both the univariate distributions of MLEs of parameters and functions of these MLEs, such as $\widehat{B}_{0.1,r}$, as well as the joint distributions of MLEs of parameters. Moreover, the Normal approximation is suitable only if the number of failures is large. Consequently, it does not seem appropriate to use the usual Normal critical values to establish confidence intervals from MLE in samples of small to moderate size.

As previously indicated at Section 3.1, we will use the relative likelihood function of $\boldsymbol{\pi}$, defined as

$$R(\boldsymbol{\pi}) = \frac{L(\boldsymbol{\pi})}{L(\widehat{\boldsymbol{\pi}})}, \tag{3.2}$$

so that $0 < R(\boldsymbol{\pi}) \leqslant 1$ for all $\boldsymbol{\pi}$, as an alternative for assessing the precision of MLEs in relatively small samples. Therefore, if $R(\boldsymbol{\pi}) \geq \lambda$, then the vector valued π is said to have at least $100(1-\lambda)\,\%$ of the maximum consistency possible under the model. For now, in a two-parameter case, a contour map of $R(\pi_1, \pi_2)$ portrays this consistency over the parameter space; for instance, points inside the 0.5-contour constitute fairly plausible parameter pairs, whereas values outside the 0.01-contour are very implausible. Kalbfleisch (1979) discusses the use of contour plots of $R(\boldsymbol{\pi})$ to obtain confidence limits for a single set of data. We adapt this approach to provide confidence regions for the sampling distribution of $(\widehat{\boldsymbol{\pi}}_1, \widehat{\boldsymbol{\pi}}_2)$; this involves specifying - for any parameter values, sample size and censoring regime - an *idealised* sample, and then calculating and plotting the contours for that idealised sample. One intuitive instance of an ideal sample can be obtained by taking the corresponding expected order statistics as data values, but we note that other methods may be possible; thus the ML estimates found from this sample will, naturally, possess maximum plausibility, and hence can be employed to produce the idealised or expected relative likelihood contour, as a counterpart for the large-sample probability ellipse. The contour plots are then validated for various r and n using simulation experiments.

3.4.1 Relative Likelihood Contour Plots in the Weibull Distribution

Suppose $\lambda_1, \lambda_2, \ldots,$ with $0 < \lambda_1 < \lambda_2 < \cdots < 1$, is a set of values for which contours on the relative likelihood surface are to be plotted. Watkins & Leech (1989) outline an algorithm for drawing relative likelihood contours for data from the Weibull distribution; we summarise the main stages as follows:

Stage 1 Location of the MLEs, in which we find $\left(\widehat{\theta}_r, \widehat{\beta}_r\right)$, the centre of all contours.

Stage 2 Defining the drawing area — by evaluating the relative likelihood at a series of fractions and multiples of $\widehat{\theta}_r$ and $\widehat{\beta}_r$, we search the $\theta - \beta$ plane for a rectangular within which the largest contour corresponding to λ_1 will lie. In practice, the transformations $\theta = a\widehat{\theta}_r$ and $\beta = b\widehat{\beta}_r$ introduce some numerical stability and flexibility over the range of possible parameter values.

Stage 3 Drawing contours — first find, and then join together a large number of points found from numerically solving the equation $l(\theta,\beta) - l\left(\widehat{\theta}_r, \widehat{\beta}_r\right) = \ln \lambda_1$ and its partial derivatives wrt a and b. We again benefit from the numerical stability introduced by these transformations. This process is repeated for each contour.

Illustration: Ball bearings data

We show here contour maps for the ball bearings data with censoring as in Table 2.6, with $\lambda = 0.01, 0.05, 0.1$ and 0.5. Thus, the first case yields approximate 99% confidence regions for (θ, β). Figure 3.14 shows the effect of r on the contours. In general, for given λ, we see that the contours get smaller as r increase; this is obvious because more failures would provide more information in estimating the parameters. We also note that contours extend over larger values in the β-axis, but over smaller values in the θ-axis. It is also clear that, as λ increases, contour areas drop dramatically and the contour shapes become more elliptical. Further, the shift in location, in line with the values in Table 2.6, is now more apparent.

It is interesting to compare and contrast the relative likelihood regions with the confidence regions based on asymptotic Normality. Watkins (2004) considers this issue for a sample of size 100 subject to Type I censoring. Figure 3.15 is an example using the ball bearings data at the 12^{th} failure. With r fixed, the two regions seem largely to coincide, and approach to complete overlap as λ increases. Moreover, the relative likelihood contours consistently appear tangential to the ellipses close to their minor axes. Intuitively, this may provide an alternative approach to locate the initial point in drawing a contour, by locating the first point on the minor axis of the ellipse, and then calculating the relative likelihood there. If this relative likelihood value is close to λ, then it could serve, possibly with further searching, as an initial point on the contour. The effectiveness of this procedure, in comparison with performing a numerical search as in Watkins & Leech (1989), will be explored elsewhere.

Expected Relative Likelihood Contours

Although the above discussion is based on a single set of data, we may adapt the approach to provide probability regions for the sampling distribution of $\left(\widehat{\theta}_r, \widehat{\beta}_r\right)$; as previously noted, this requires the expectations of order statistics for lifetimes drawn from the Weibull distribution, given by

$$E\left[X_{i:n}\right] = c_{i:n}\theta \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} \frac{\Gamma\left(\frac{1}{\beta}+1\right)}{(n-k)^{\frac{1}{\beta}+1}}. \tag{3.3}$$

Example: $r = 15, n = 25$ We assume $n = 25$ and $\theta = 100, \beta = 2$ to illustrate this experimental set-up; the corresponding (uncensored) idealised sample, calculated from (3.3),

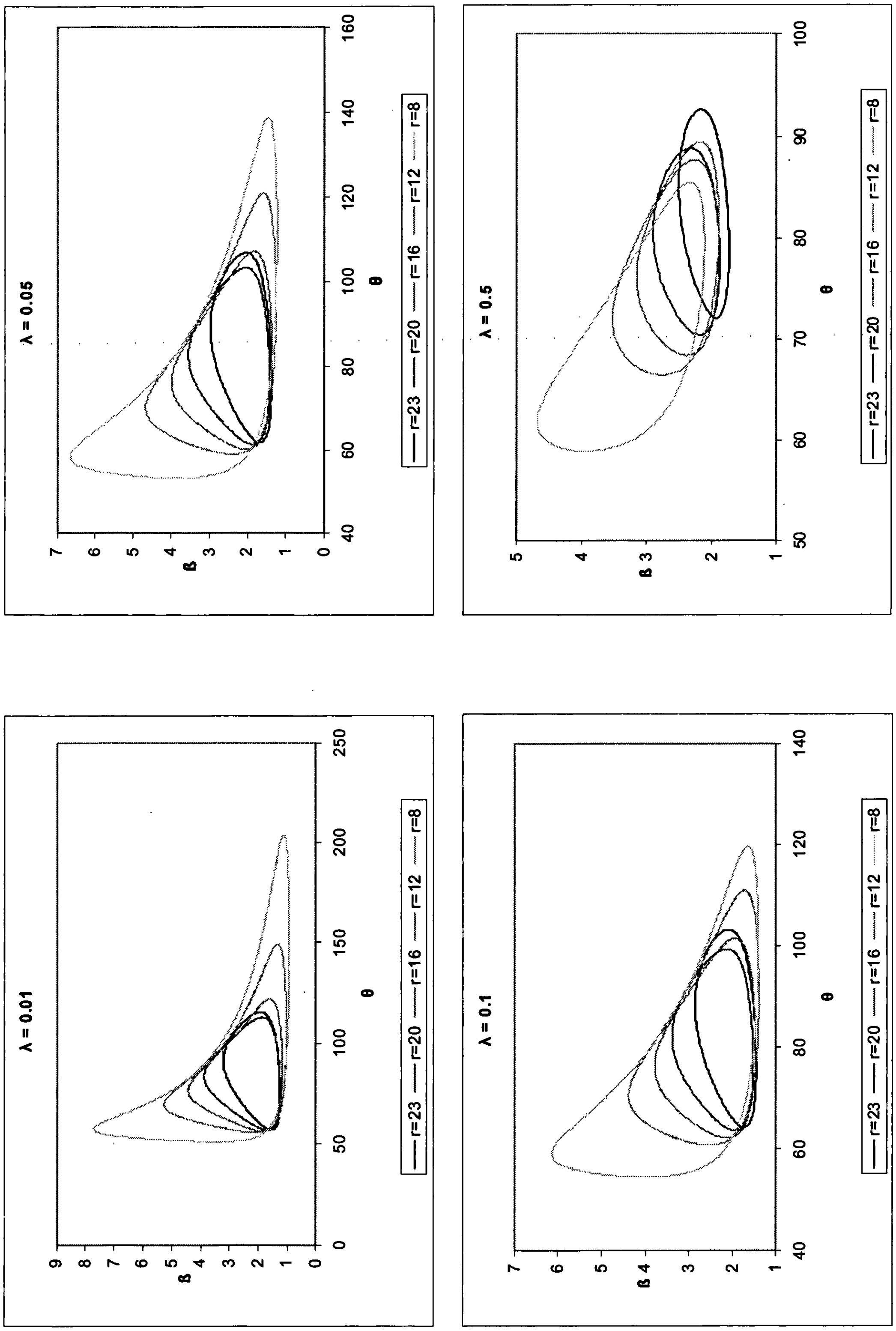

Figure 3.14: Four sets of relative likelihood contour plots using the ball bearings data.

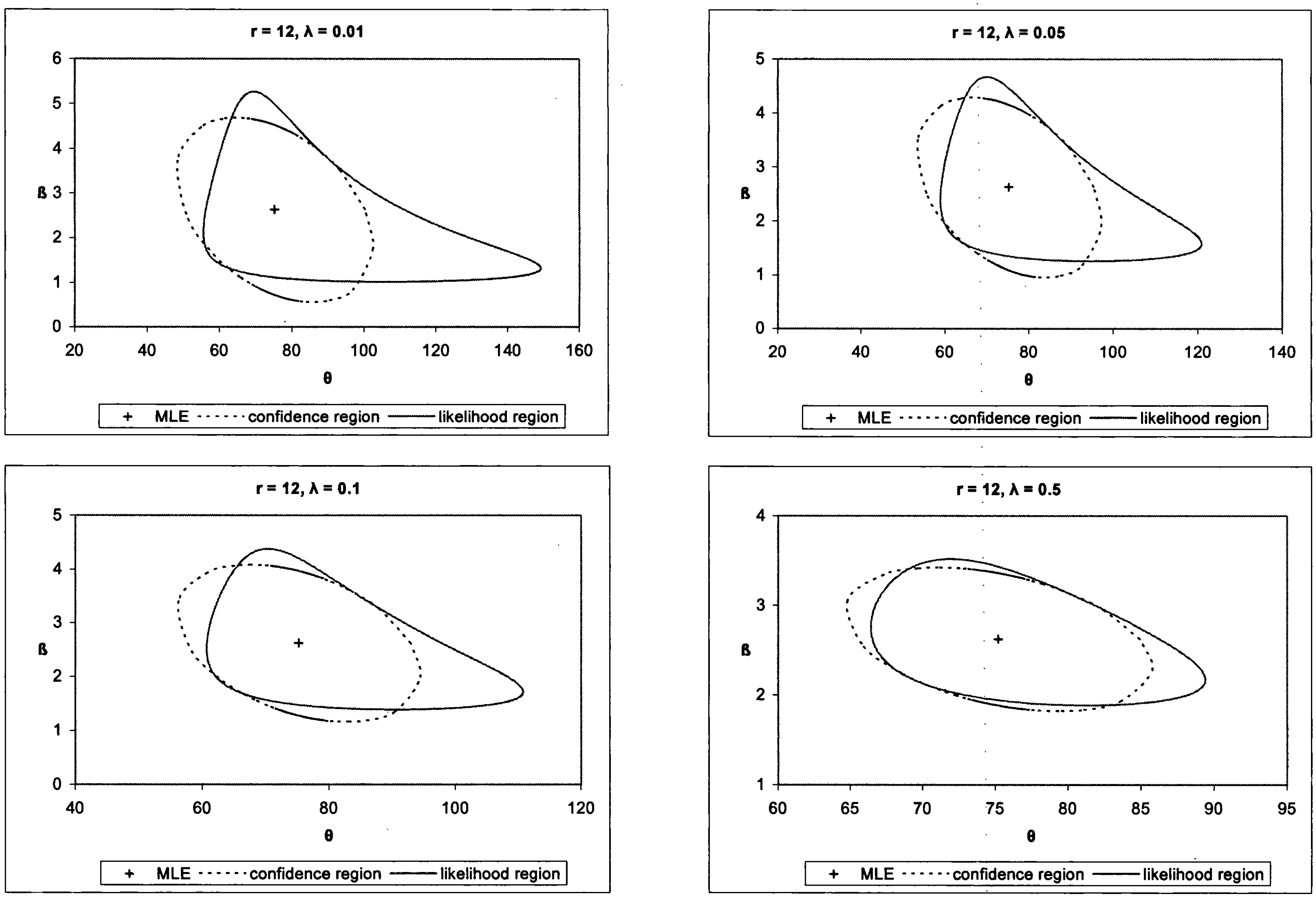

Figure 3.15: Four sets of relative likelihood regions versus the asymptotic confidence regions for $r = 12$ using the ball bearings data.

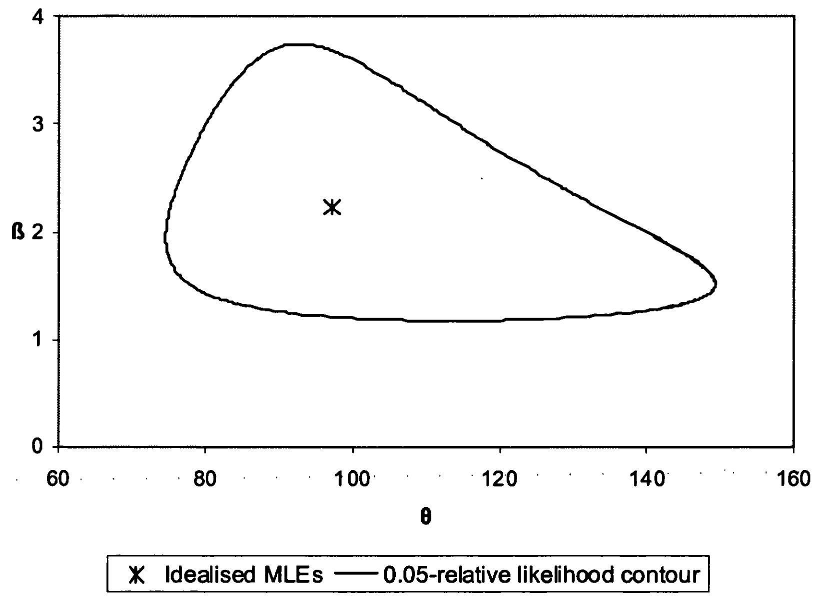

Figure 3.16: 0.05-relative likelihood contour plot for $r = 15, n = 25$, for ideal Weibull data generated with $\theta = 100, \beta = 2$.

is

17.7245	26.8619	33.9372	40.0338	45.5588
50.7180	55.6346	60.3916	65.0497	69.6576
74.2572	78.8869	83.5849	88.3912	93.3502
98.5144	103.9482	109.7355	115.9900	122.8762
130.6478	139.7332	150.9559	166.2711	192.8568

so that censoring at, say, $r = 15$ gives the following failure times

17.7245	26.8619	33.9372	40.0338	45.5588
50.7180	55.6346	60.3916	65.0497	69.6576
74.2572	78.8869	83.5849	88.3912	93.3502
93.3502†	93.3502†	93.3502†	93.3502†	93.3502†
93.3502†	93.3502†	93.3502†	93.3502†	93.3502†

at which $\widehat{\theta}^*_{15} = 97.2027$ and $\widehat{\beta}^*_{15} = 2.2306$ (we use $*$ to indicate ML estimates obtained from the idealised sample). Hence, the point $(97.2027, 2.2306)$ will act as the centre of the relative likelihood contour, as the most likely point to occur due to the method of experimentation. Figure 3.16 illustrates the 0.05-expected relative likelihood contour around $\left(\widehat{\theta}^*_{15}, \widehat{\beta}^*_{15}\right)$, which clearly is not elliptical.

r	n			
	25	50	100	1000
$0.2n : \widehat{\theta}^*_r$	78.3288	87.3439	92.8523	99.0496
$: \widehat{\beta}^*_r$	2.7710	2.3700	2.1892	2.0230
$0.4n : \widehat{\theta}^*_r$	92.6713	95.9604	97.8055	99.7268
$: \widehat{\beta}^*_r$	2.3566	2.1824	2.0958	2.0118
$0.6n : \widehat{\theta}^*_r$	97.2027	98.5117	99.2135	99.9081
$: \widehat{\beta}^*_r$	2.2306	2.1200	2.0634	2.0078
$0.8n : \widehat{\theta}^*_r$	99.2023	99.6069	99.8067	99.9820
$: \widehat{\beta}^*_r$	2.1685	2.0883	2.0466	2.0056
$1.0n : \widehat{\theta}^*_r$	100.2647	100.1811	100.1147	100.0196
$: \widehat{\beta}^*_r$	2.1377	2.0741	2.0400	2.0051

Table 3.23: Idealised MLEs $(\hat{\theta}^*_r, \hat{\beta}^*_r)$ for various r, n, for ideal Weibull data generated with $\theta = 100, \beta = 2$.

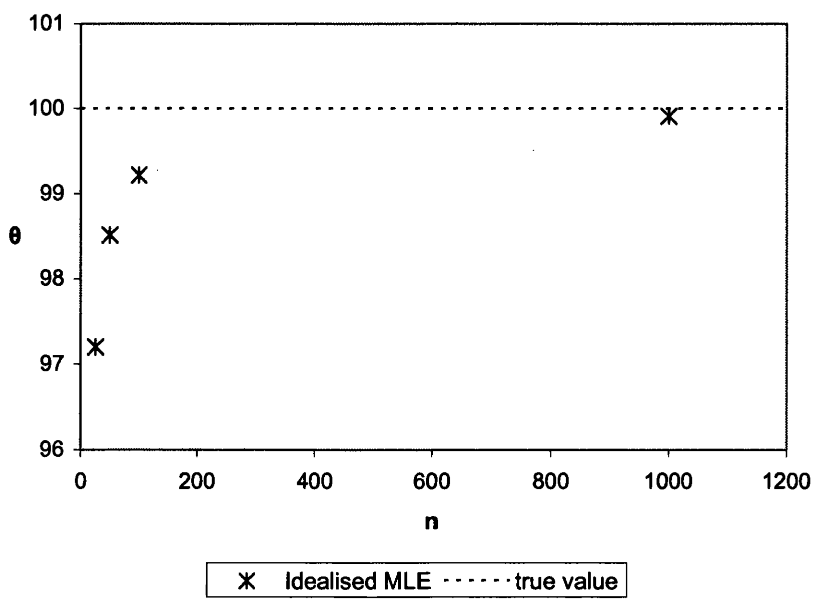

Figure 3.17: Plot of $\widehat{\theta}^*_{0.6n}$ versus n, for ideal Weibull data generated with $\theta = 100, \beta = 2$.

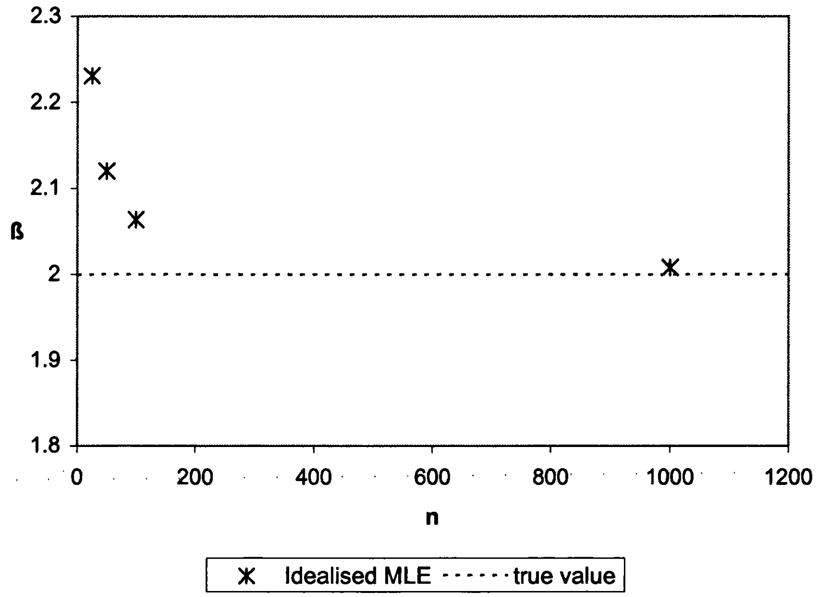

Figure 3.18: Plot of $\widehat{\beta}^*_{0.6n}$ versus n, for ideal Weibull data generated with $\theta = 100, \beta = 2$.

General: varying r and n Further illustrations are given in Table 3.23, which shows the idealised estimates of (θ, β) for various n when the ideal data, calculated with $\theta = 100, \beta = 2$, are subject to Type II censoring at the r^{th} failure. Note that these values can be compared with their average counterparts in Tables 2.7 and 2.8, where we notice a generally good match in the results. The agreement is better shown in Figures 3.17 and 3.18 for $r = 0.6n$, where we see both $\widehat{\theta}^*_r$ and $\widehat{\beta}^*_r$ gradually converge to their true values as n increases.

We assume $\lambda = 0.05$, and show in Figure 3.19, the contour maps for some ideal samples for various r and n; this yields the approximate 95% confidence regions for (θ, β), with centres $\left(\widehat{\theta}^*_r, \widehat{\beta}^*_r\right)$ as given in Table 3.23. When comparison is made with Figure 3.12 where $r = 0.6n$, the relative likelihood contours seem to move towards the probability ellipses in terms of both size and shape as n increases; we can expect similar convergence, perhaps at different rates, for other values of r.

A more detailed illustration is given assuming, say, $r = 5$ failures observed in a sample with $n = 25$; Figure 3.20 shows the joint distribution of $\left(\widehat{\theta}_5, \widehat{\beta}_5\right)$ is certainly not elliptical - stretching rightwards in the θ-direction and upwards in the β-direction - and hence it is non-Normal. On the other hand, the relative likelihood contour plot appears to capture more accurately than asymptotic probability ellipse the behaviour of the sampling distribution of $\left(\widehat{\theta}_5, \widehat{\beta}_5\right)$, especially the right skewed pattern observed in the distributions of $\widehat{\theta}_5$ and $\widehat{\beta}_5$.

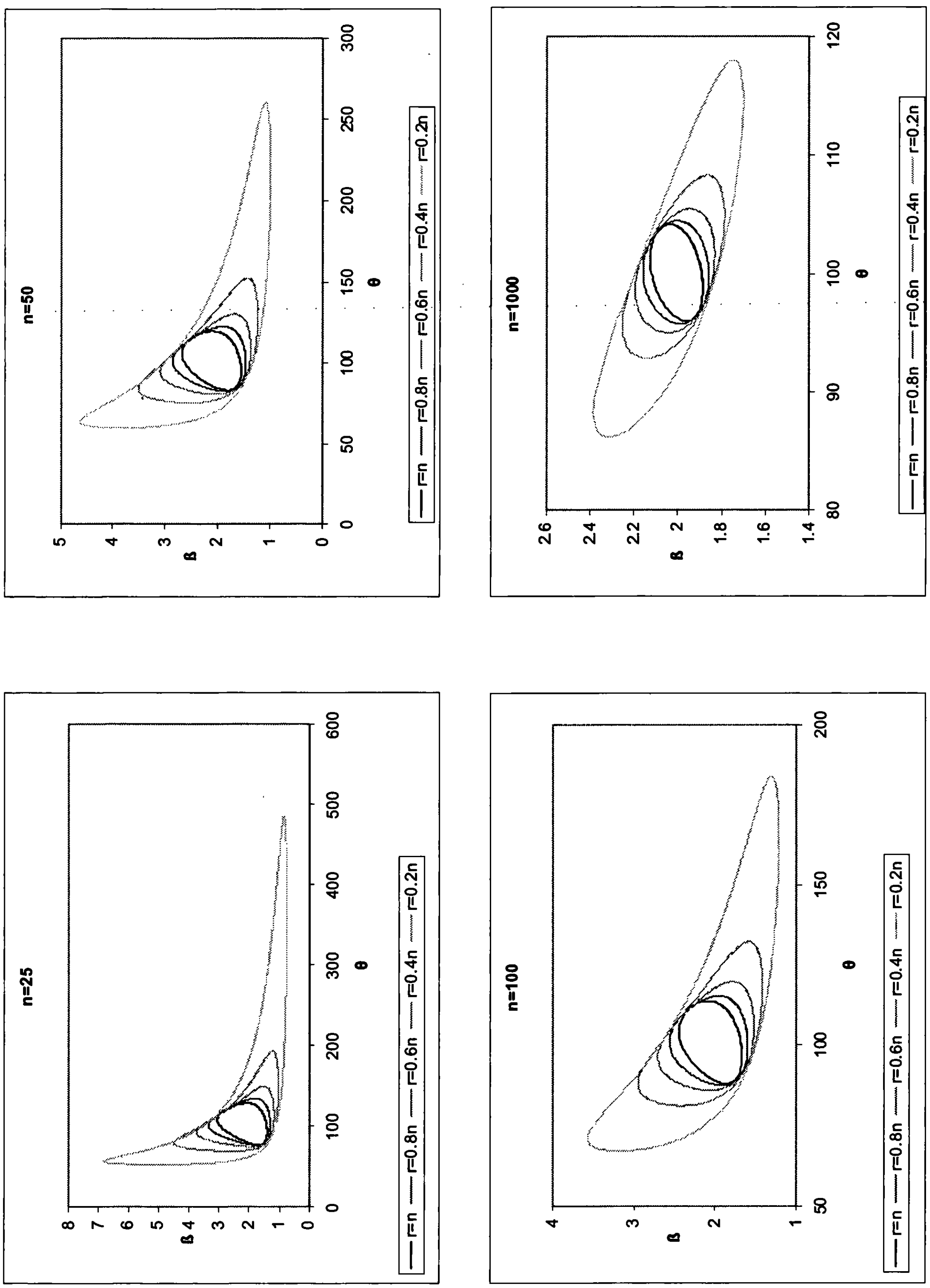

Figure 3.19: Four sets of 0.05-relative likelihood contour plots for ideal Weibull data generated with $\theta = 100, \beta = 2$.

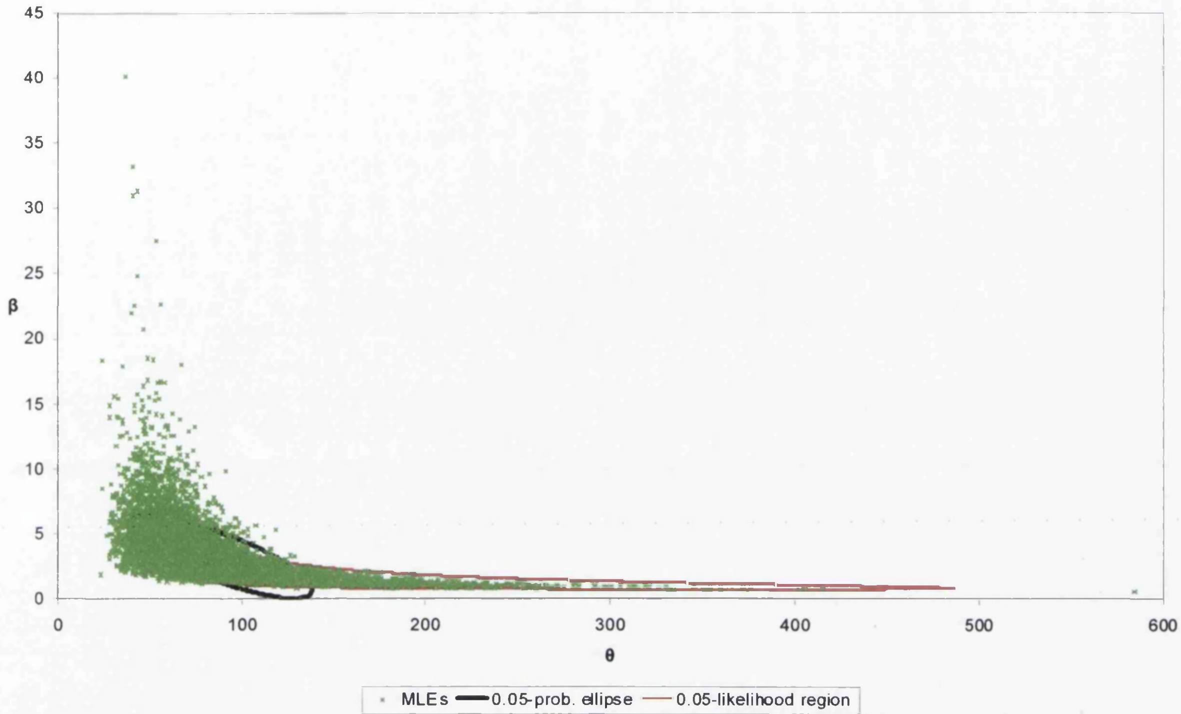

Figure 3.20: The MLEs (×) together with 0.05-relative likelihood contour and asymptotic 0.05-probability ellipse for $(\hat{\theta}_5, \hat{\beta}_5)$, for $n = 25$, for Weibull data generated with $\theta = 100, \beta = 2$.

Relative Likelihood Contour Validation and Comparison with Normal Theory Probability Region

This illustration suggests a method to validate the use of relative likelihood contours to obtain confidence regions of the sampling distribution of MLEs in small samples; this involves computing the relative likelihood for each simulated observations of $\left(\widehat{\theta}_r, \widehat{\beta}_r\right)$, and then counting the number of replications whose relative likelihood is ≥ 0.05. Note that, in conjunction with idealised samples, (3.2) is now defined as a ratio of the likelihood calculated using the ML estimates to the likelihood calculated using $\left(\widehat{\theta}^*_r, \widehat{\beta}^*_r\right)$. It is straightforward to show that an observed point $\left(\widehat{\theta}_r, \widehat{\beta}_r\right)$ is enclosed by the 0.05-relative likelihood contour if

$$l_r\left(\widehat{\theta}_r, \widehat{\beta}_r\right) - l_r\left(\widehat{\theta}^*_r, \widehat{\beta}^*_r\right) \geq \ln 0.05,$$

where $l_r\left(\widehat{\theta}_r, \widehat{\beta}_r\right)$ and $l_r\left(\widehat{\theta}^*_r, \widehat{\beta}^*_r\right)$ are obtained from (2.16) upon appropriate substitutions. Accordingly, we expect to find $95\% \times 10^4$ of $\left(\widehat{\theta}_r, \widehat{\beta}_r\right)$ within the 0.05-expected contour area. As discussed in Section 3.3, this procedure can be repeated for the large-sample probability ellipse derived from Normal theory, to find the number of replications of $\left(\widehat{\theta}_r, \widehat{\beta}_r\right)$ enclosed in the probability region.

r	n			
	25	50	100	1000
$0.2n$	7949	8826	9168	9450
	6834	7921	8600	9371
$0.4n$	8798	9138	9344	9466
	8403	8919	9197	9443
$0.6n$	9063	9267	9411	9492
	8842	9170	9328	9495
$0.8n$	9193	9352	9444	9491
	9022	9280	9369	9484
$1.0n$	9257	9415	9434	9508
	9091	9295	9392	9517

Table 3.24: Number of replications of $(\hat{\theta}_r, \hat{\beta}_r)$ within the 0.05-relative likelihood contour (upper) and the asymptotic 0.05-probability ellipse (lower) for Weibull data generated with $\theta = 100, \beta = 2$.

r	n			
	25	50	100	1000
$0.2n$	7709	8581	9004	9406
	5948	7036	7811	9185
$0.4n$	8666	9069	9259	9470
	7958	8679	9003	9434
$0.6n$	8958	9185	9282	9464
	8632	9069	9265	9458
$0.8n$	8992	9221	9325	9495
	8864	9199	9312	9502
$1.0n$	9047	9249	9336	9516
	8870	9208	9332	9514

Table 3.25: Number of replications of $(\hat{\theta}_r, \hat{\beta}_r)$ within the 0.05-relative likelihood contour (upper) and the asymptotic 0.05-probability ellipse (lower) for Weibull data generated with $\theta = 100, \beta = 0.5$.

Results for each combination of r, n and θ, β replicated are shown in Tables 3.24, 3.25 and 3.26. We see at early censoring levels there are quite large disagreements between observed and expected values, but agreement improves as more items are allowed to fail, to increase the precision of the estimates yielded. We also see that the results approach 9500 as n increase, and are reasonably consistent across the various values of the shape parameter considered here. Most importantly, for $n \leq 100$, the expected relative likelihood contours (upper entries) consistently contain more replications of $\left(\widehat{\theta}_r, \widehat{\beta}_r\right)$ than the elliptical probability regions (lower entries), indicating that the non-elliptical nature of the relative likelihood contours reflects more accurately the sampling distribution of $\left(\widehat{\theta}_r, \widehat{\beta}_r\right)$ for samples of small size.

r	n			
	25	50	100	1000
$0.2n$	7979	8797	9139	9447
	6947	8037	8706	9393
$0.4n$	8812	9199	9346	9493
	8397	8938	9219	9469
$0.6n$	9078	9280	9369	9524
	8839	9132	9299	9507
$0.8n$	9238	9337	9425	9516
	9069	9242	9356	9504
$1.0n$	9309	9413	9453	9510
	9108	9280	9404	9503

Table 3.26: Number of replications of $(\hat{\theta}_r, \hat{\beta}_r)$ within the 0.05-relative likelihood contour (upper) and the asymptotic 0.05-probability ellipse (lower) for Weibull data generated with $\theta = 100, \beta = 4$.

3.4.2 Relative Likelihood Contour Plots in the Burr Distribution

Here, we will adapt the contour drawing procedure proposed by Watkins & Leech (1989), giving the necessary formulae at each stage.

Stage 1

Location of the Burr MLEs has been covered in Section 2.4, where the classical Newton-Raphson iterative method is used to maximise the profile log-likelihood function at (2.39) wrt τ. With $\widehat{\tau}_r$ thus found, $\widehat{\alpha}_r$ is computed at (2.38).

Stage 2

We note that the transformations for Weibull parameters accommodate differing scales in θ and β. The corresponding transformations are thus less essential in the Burr case (since both shape parameters are, generally, of the same order), but are nevertheless retained here. We thus define two working variables a and b such that

$$\alpha = a\widehat{\alpha}_r \tag{3.4}$$

and

$$\tau = b\widehat{\tau}_r. \tag{3.5}$$

Thus, $a = b = 1$ implies $\alpha = \widehat{\alpha}_r$ and $\tau = \widehat{\tau}_r$. For τ-direction, we start with $b = 1$ so that the value of α which maximises the relative likelihood is just $r/(T_f + T_c)$. In order to locate $\tau_{\min}$, we take a series of values for b decreasing from 1 in steps of 0.1. Hence, we move down from $\widehat{\tau}_r$ until the relative likelihood is less than λ_1. We then take a series of values for b increasing from 1 in steps of 0.1 when searching for $\tau_{\max}$. In this case, we again stop when

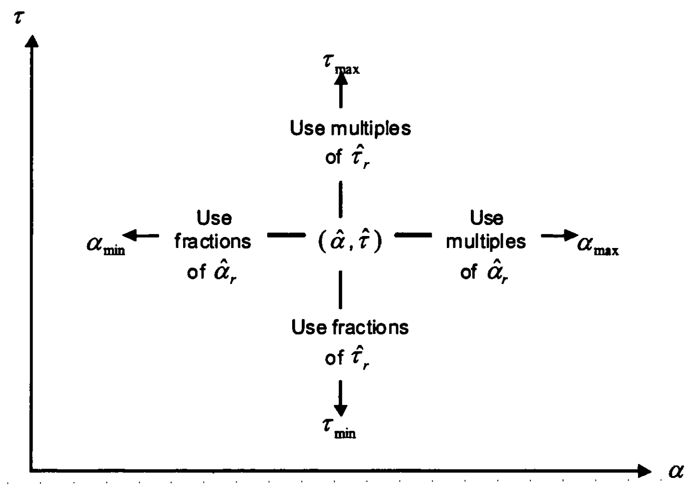

Figure 3.21: Defining the drawing area in the $\alpha - \tau$ plane about $(\hat{\alpha}_r, \hat{\tau}_r)$.

the relative likelihood is less than λ_1. In Figure 3.21, this is illustrated by the two vertical arrows moving away from $(\widehat{\alpha}_r, \widehat{\tau}_r)$.

Likewise, for α-direction, we consider a series of fractions and multiples of $\widehat{\alpha}_r$, and, for each value of α, find the τ which maximises the relative likelihood for that value of α. The Burr log-likelihood function is given by (2.35), which is to be maximised wrt τ, assuming α as fixed; the first- and second-order derivatives of (2.35) wrt τ are

$$r\tau^{-1} + S_{f,1}(0) - (\alpha + 1)T_{f,111} - \alpha T_{c,111} \tag{3.6}$$

and

$$-r\tau^{-2} - (\alpha + 1)T_{f,122} - \alpha T_{c,122}$$

respectively. Again, the Newton-Raphson method is used, with which the initial estimate of the root of (3.6) is $\widehat{\tau}_r$. This maximum value of the relative likelihood is computed, and hence one can search for the minimum and maximum values of α that need to be considered. This is illustrated by the two horizontal arrows moving away from $(\widehat{\alpha}_r, \widehat{\tau}_r)$ in Figure 3.21.

Stage 3

We use (3.2) to write

$$\lambda_1 = \frac{L_r(\alpha, \tau)}{L_r(\widehat{\alpha}_r, \widehat{\tau}_r)},$$

or equivalently,

$$l_r(\alpha, \tau) - l_r(\widehat{\alpha}_r, \widehat{\tau}_r) = \ln \lambda_1.$$

Thus, if we define

$$f(a,b) = l_r(a\widehat{\alpha}_r, b\widehat{\tau}_r) - l_r(\widehat{\alpha}_r, \widehat{\tau}_r) \tag{3.7}$$

then the first point on this contour can be obtained by solving the equation

$$f(a,b) - \ln(\lambda_1) = 0 \tag{3.8}$$

for b with $a = 1$. While there is no analytical expression for the root, we may employ the Newton-Raphson approach which we require the partial derivative of $f(a,b)$ wrt b:

$$f'_b = rb^{-1} + \widehat{\tau}_r S_{f,1}(0) - (a\widehat{\alpha}_r + 1)\widehat{\tau}_r T_{f,111}(b\widehat{\tau}_r) - a\widehat{\alpha}_r\widehat{\tau}_r T_{c,111}(b\widehat{\tau}_r), \tag{3.9}$$

together with $\tau_{\max}/\widehat{\tau}_r$ as the initial estimate.

To move around the contour, we compute the gradient of the tangent to the contour at this initial point, once more, with $a = 1$. This gradient is given by

$$-\frac{f'_a}{f'_b}, \tag{3.10}$$

where

$$f'_a = ra^{-1} - \widehat{\alpha}_r T_f(b\widehat{\tau}_r) - \widehat{\alpha}_r T_c(b\widehat{\tau}_r). \tag{3.11}$$

The second point on this contour is therefore achieved by moving a distance δ in the $a - b$ plane along this tangent

$$a \to a_{new} = a + \frac{\delta f'_b}{\sqrt{f'^2_a + f'^2_b}} \tag{3.12}$$

and

$$b \to b_{new} = b - \frac{\delta f'_a}{\sqrt{f'^2_a + f'^2_b}}. \tag{3.13}$$

When finding the subsequent contour points with these a_{new} and b_{new}, a is fixed at the value of previous a_{new} and an attempt is made to get a corresponding value of b solving (3.8). This is as discussed above, except that now b_{new} is taken to be the initial estimate of the solution of (3.8), and, more importantly, $a = 1$ in (3.8) and (3.9) no longer apply. Equation (3.11) can now be recomputed and a and b updated again.

It should also be noted, however, that when the values of f'_b and $f'_b/\sqrt{f'^2_a + f'^2_b}$ are near to zero, the iterating process (for estimating b) may produce a value of a for which no corresponding value of b solving (3.8) can be found; this indicates the contour is close to its extreme left or right edge. In such cases, we fix the value of b instead and try to find a corresponding value of a which solves (3.8). As before, all searches must be numerical; a_{new} is treated as the initial estimate of the solution in (3.8) and the derivative f'_a is used to improve this estimate. In other words, we are performing a change of direction (or a change in slope) to the ellipse. By repeating this process it should be possible to complete the λ_1-contour, after joining together numerous pairs of (α, τ) obtained.

We remark that the choice of δ used in (3.12) and (3.13) will reflect the smoothness and accuracy of a contour plot, and is directly linked to computational cost incurred; a small δ value will produce a large number of points used to draw a contour, but will, on the other hand, increase the computing time. Throughout this thesis, we have used $\delta = 0.01$; this seems to be sufficiently small in relation to the drawing area to allow us to regard the resulting contour plot as a smooth and accurate one.

Illustration: Arthritic patients data

The above stages are illustrated in a simple numerical example using the arthritic patients data. Here, we suppose that $r = n$ and $\lambda = 0.05$; this yields the approximate 95% confidence regions for (α, τ) under complete sampling. Appendix C presents the corresponding SAS code in more details.

Stage 1 From Table 2.13, the parameter estimates are

$$\widehat{\alpha} = \widehat{\alpha}_{50} = 8.2681 \text{ and } \widehat{\tau} = \widehat{\tau}_{50} = 5.0006.$$

Stage 2 Starting with $b = 1 \Leftrightarrow \tau = \widehat{\tau}$, rescaling is repeated until the relative likelihood is less than 0.05 at each end in the vertical direction to give

$$\tau_{\min} = 3.5004 = 0.7\widehat{\tau} \text{ and } \tau_{\max} = 7.0009 = 1.4\widehat{\tau},$$

while for the horizontal direction, we have

$$\alpha_{\min} = 4.1340 = 0.5\widehat{\alpha} \text{ and } \alpha_{\max} = 15.7094 = 1.9\widehat{\alpha},$$

found iteratively from an initial value of $\alpha = \widehat{\alpha}$ $(a = 1)$.

Stage 3 As outlined in Appendix C, the drawing stage consists of six separate iterative processes:

Process 1 Starting with $a = 1$ in (3.8), the Newton-Raphson approach yields $b = 1.1842$ on using $\tau_{\max}/\widehat{\tau}$ as an initial estimate. Hence, the first point on this contour is $(8.2681, 5.9219)$ with the corresponding updated a, b values of 0.9906, 1.1807 respectively (see below).

a	b	α	τ	a_{new}	b_{new}
1	1.1842	8.2681	5.9219	0.9906	1.1807

Process 2 The drawing is continued leftward and downward using the same algorithm. However, the initial value of b here for the second point is 1.1807, that is, the final value of b obtained in the previous process. This process terminates when for $a =$

0.5960, there is no corresponding value of b solving (3.8) can be found; thus, as indicated below, the contour has reached its extreme left at (4.9343, 4.1973) after 54 iterations. For convenience, we have omitted the details for intermediate iterations.

i	a	b	α	τ	a_{new}	b_{new}
1	0.9906	1.1806	8.1907	5.9040	0.9813	1.1770
2	0.9813	1.1770	8.1136	5.8857	0.9720	1.1733
⋮						
53	0.5988	0.8530	4.9508	4.2658	0.5968	0.8432
54	0.5968	0.8393	4.9343	4.1973	0.5960	0.8294

Process 3 We now find a for b fixed at 0.8294, using the last value of a computed in Process 2 - 0.5960 - as initial estimate. Only 12 iterations were observed below, implying that the contour has a sharp left edge.

i	a	b	α	τ	a_{new}	b_{new}
1	0.5965	0.8294	4.9316	4.1474	0.5967	0.8194
2	0.5973	0.8194	4.9383	4.0975	0.5987	0.8095
⋮						
11	0.6713	0.7627	5.5502	3.8141	0.6812	0.7617
12	0.6889	0.7617	5.6958	3.8090	0.6989	0.7615

Process 4 The procedure is continued in the rightward and upward direction, by solving (3.8) for b using a fixed at 0.6989. Note that here the starting value of b is 0.7615; we obtained 108 iterations as the follows.

i	a	b	α	τ	a_{new}	b_{new}
1	0.6989	0.7617	5.7785	3.8091	0.7089	0.7619
2	0.7089	0.7621	5.8611	3.8109	0.7189	0.7626
⋮						
107	1.6583	1.1854	13.7107	5.9276	1.6629	1.1943
108	1.6629	1.1965	13.7487	5.9834	1.6657	1.2061

Process 5 At the extreme right edge we search for a instead with b fixed at 1.2061, and 1.6657 acts as the initial value of a. There were 20 iterations observed at this stage:

i	a	b	α	τ	a_{new}	b_{new}
1	1.6647	1.2061	13.7640	6.0313	1.6655	1.2161
2	1.6642	1.2161	13.7597	6.0811	1.6623	1.2259
⋮						
19	1.4957	1.2747	12.3668	6.3741	1.4857	1.2749
20	1.4801	1.2749	12.2379	6.3755	1.4701	1.2750

λ	Process 1	Process 2	Process 3	Process 4	Process 5	Process 6	Total number of iterations
0.99	1	3	1	6	1	6	18
0.95	1	7	2	13	2	14	39
0.90	1	11	1	20	2	21	56
0.75	1	18	3	32	4	32	90
0.50	1	27	6	51	7	48	140
0.25	1	38	8	72	12	69	200
0.10	1	48	10	95	16	90	260
0.05	1	54	12	108	20	103	298
0.01	1	65	15	137	27	129	374

Table 3.27: Number of iterations required to complete the λ-relative likelihood contour for various λ, for arthritic patients data when $r = n$.

Process 6 To accomplish the 0.05-contour, we once more solve (3.8) for b using fixed a, and take the initial guess of b to be 1.2750; the 103 iterations are summarised in the following table.

i	a	b	α	τ	a_{new}	b_{new}
1	1.4701	1.2750	12.1552	6.3756	1.4601	1.2749
2	1.4601	1.2749	12.0726	6.3751	1.4501	1.2747
$\vdots$						
102	0.5990	0.8540	4.9525	4.2706	0.5969	0.8442
103	0.5969	0.8408	4.9354	4.2043	0.5960	0.8308

Based on $\delta = 0.01$, 298 iterations were needed to draw the 0.05-relative likelihood contour for the arthritic patients data under complete censoring. This is displayed in Figure 3.22, where different symbols have been used to differentiate each drawing process. Slight overlapping is observed at the beginning and the ending. Furthermore, with censored data, we can expect the number of iterations to increase in line with the severity of censoring, as summarised below

r	10	20	30	40	50
Number of iterations	3616	1188	646	398	298

In addition, Table 3.27 summarises the total number of iterations required to complete the λ-relative likelihood contour for various levels of λ for the arthritic patients data when $r = n$; we notice that the number of points increases as λ decreases, as we expected.

Figure 3.23 illustrates the effect of the amount of censoring on the contours using the arthritic patients data for censoring as in Table 2.13, and again, for $\lambda = 0.01, 0.05, 0.1$ and 0.5. In general, and entirely as expected, for given λ, we observe smaller surfaces with increasing r. We also note that contours stretch over larger values in the α-direction, but over smaller values in the τ-direction. It is also clear that, as λ increases, contour areas

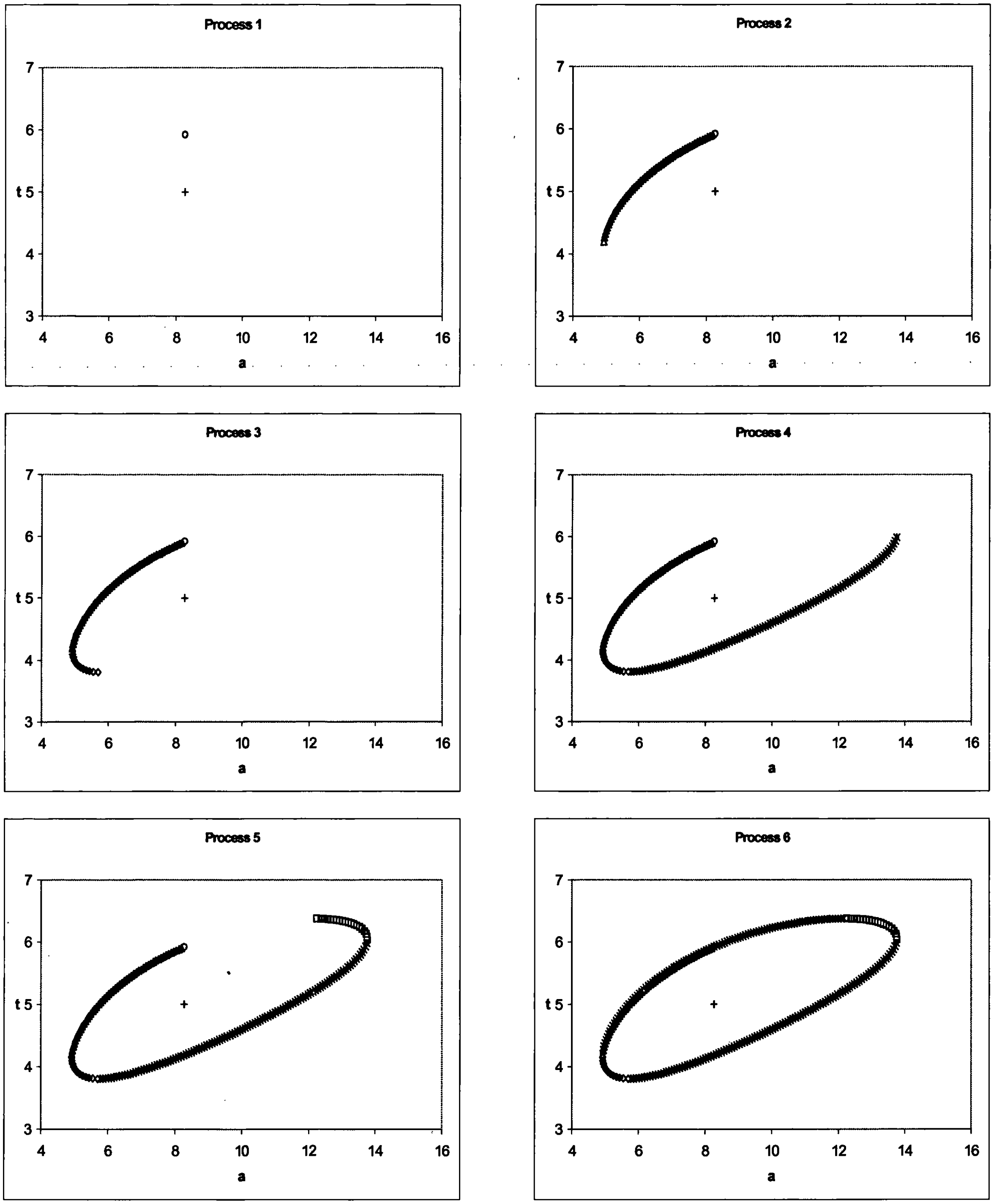

Figure 3.22: The six processes involved in constructing the 0.05-relative likelihood contour plot for arthritic patients data when $r = n$.

drop dramatically and the contour shapes become more elliptical.

Next, we take $r = 30$ and superimpose the relative likelihood regions with the confidence regions based on asymptotic theory of maximum likelihood. In Figure 3.24, the likelihood region shrinks, especially at the right edge, towards the ellipse as λ increases; the overlapping is almost perfect about the minor axis of the ellipse so that (as in the Weibull case) it may be possible to find the initial contour point by solving the two points on the minor axis; however, the non-symmetry in likelihood regions remains striking.

Expected Relative Likelihood Contours

To obtain confidence regions for the sampling distribution of $(\widehat{\alpha}_r, \widehat{\tau}_r)$ based on (3.2), we will require an idealised sample, again using the expected order statistics as data values. This is given by

$$E[X_{i:n}] = c_{i:n}\alpha \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} \frac{\Gamma\left[\frac{1}{\tau}+1\right]\Gamma\left[\alpha(n-k)-\frac{1}{\tau}\right]}{\Gamma\left[\alpha(n-k)+1\right]} \tag{3.14}$$

for lifetimes drawn from the Burr distribution.

Example: $r = 20, n = 25$ For example, the complete ideal sample, when $n = 25$ and $\alpha = 4, \tau = 3$, comprises

0.3195	0.4323	0.5122	0.5784	0.6374
0.6921	0.7444	0.7955	0.8464	0.8976
0.9501	1.0044	1.0614	1.1220	1.1873
1.2588	1.3384	1.4289	1.5343	1.6613
1.8209	2.0346	2.3513	2.9202	4.6881

so that, if the data is subject to Type II censoring at the $r = 20^{th}$ failure, then the resultant censored ideal sample is given by

0.3195	0.4323	0.5122	0.5784	0.6374
0.6921	0.7444	0.7955	0.8464	0.8976
0.9501	1.0044	1.0614	1.1220	1.1873
1.2588	1.3384	1.4289	1.5343	1.6613
$1.6613^{\dagger}$	$1.6613^{\dagger}$	$1.6613^{\dagger}$	$1.6613^{\dagger}$	$1.6613^{\dagger}$

from which $\widehat{\alpha}^*_{20} = 4.3696$ and $\widehat{\tau}^*_{20} = 3.1889$, and the corresponding 0.05-expected relative likelihood contour is shown in Figure 3.25.

General: varying r and n In addition, Table 3.28 presents the idealised ML estimates $(\widehat{\alpha}^*_r, \widehat{\tau}^*_r)$ computed from ideal samples of small to moderate size at a range of r, generated with $\alpha = 4, \tau = 3$; these values lie at the middle of all contours, and, as shown in Figures

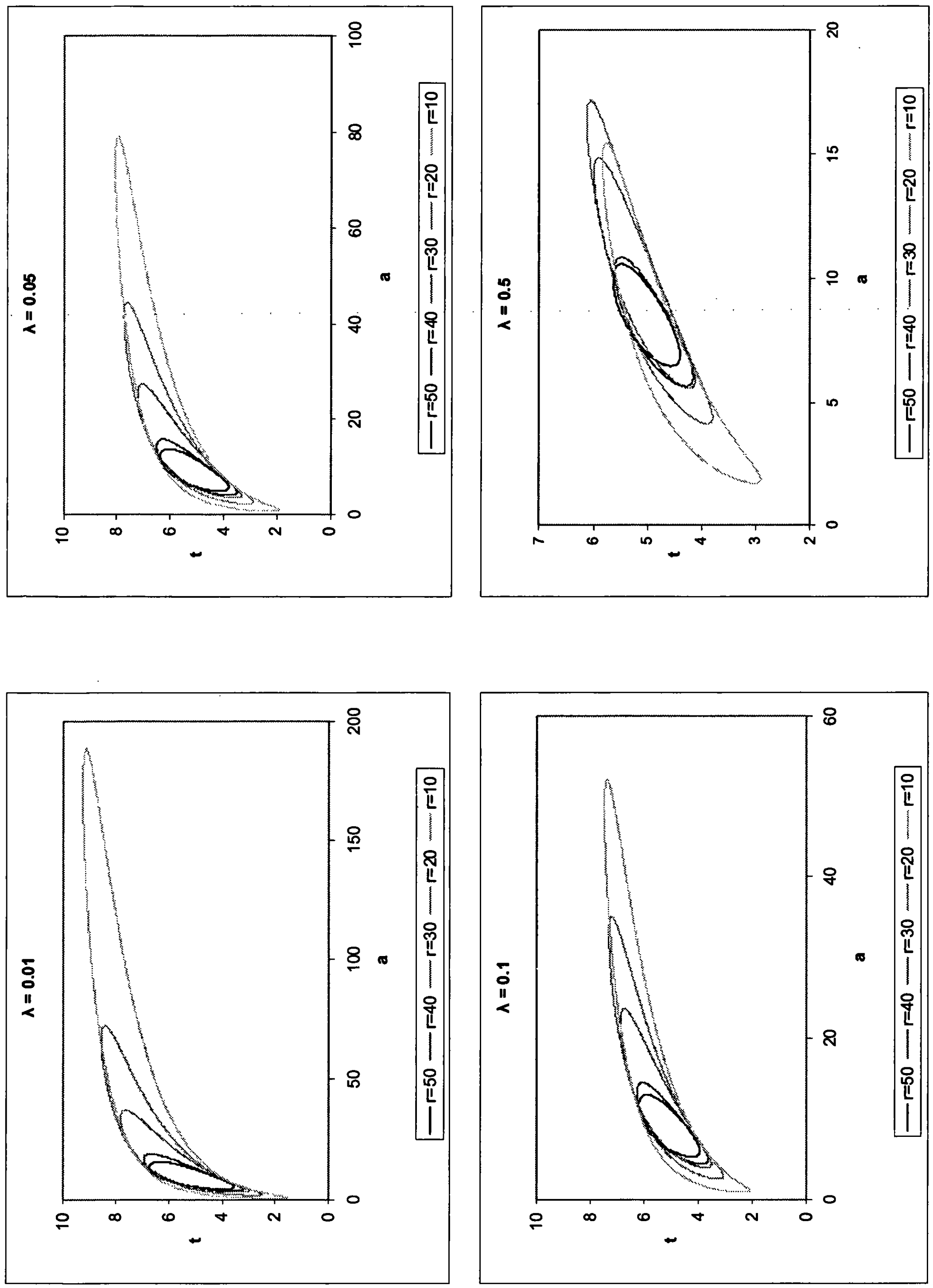

Figure 3.23: Four sets of relative likelihood contour plots using the arthritic patients data.

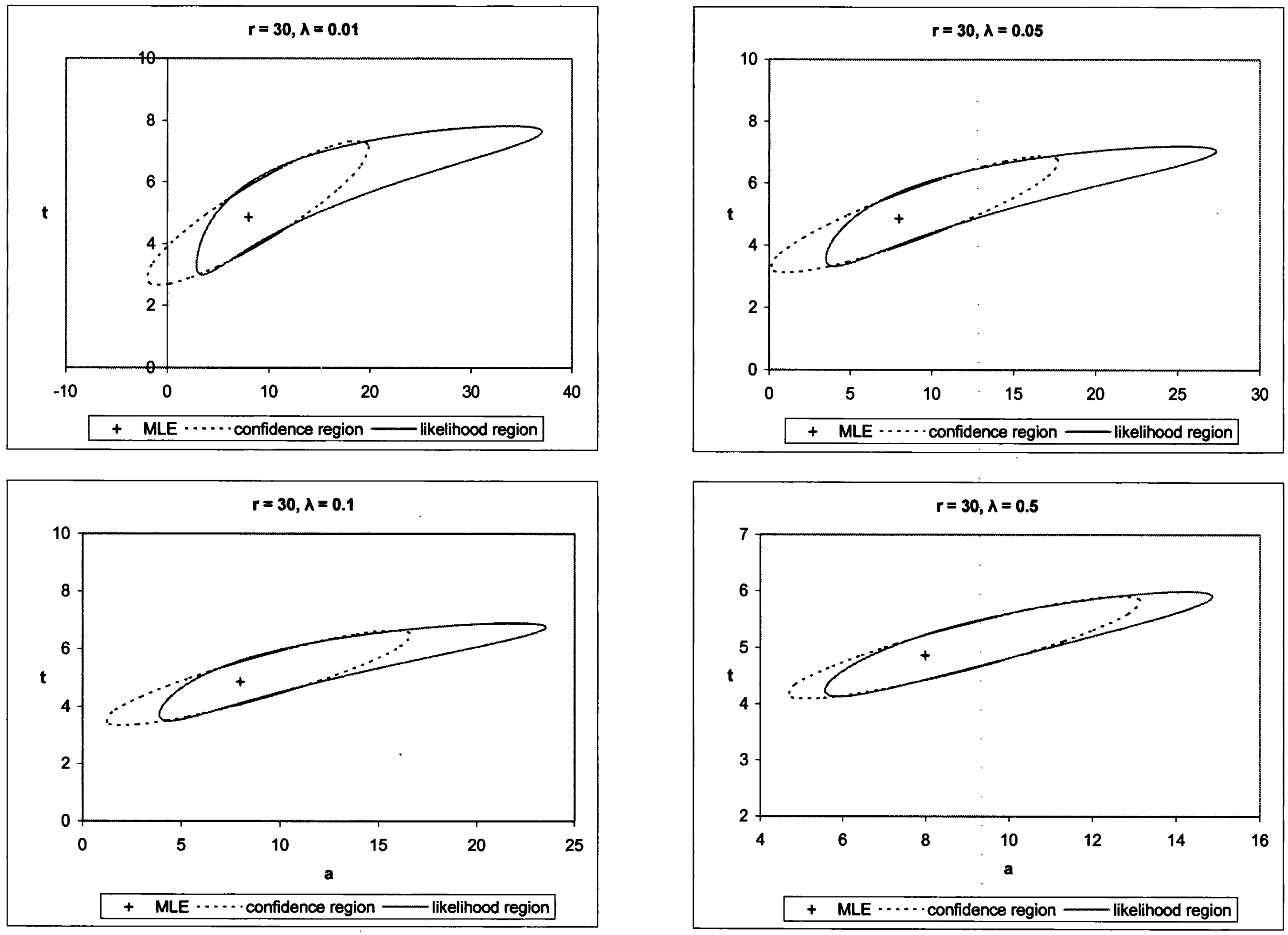

Figure 3.24: Four sets of relative likelihood regions versus the asymptotic confidence regions for $r = 30$ using the arthritic patients data.

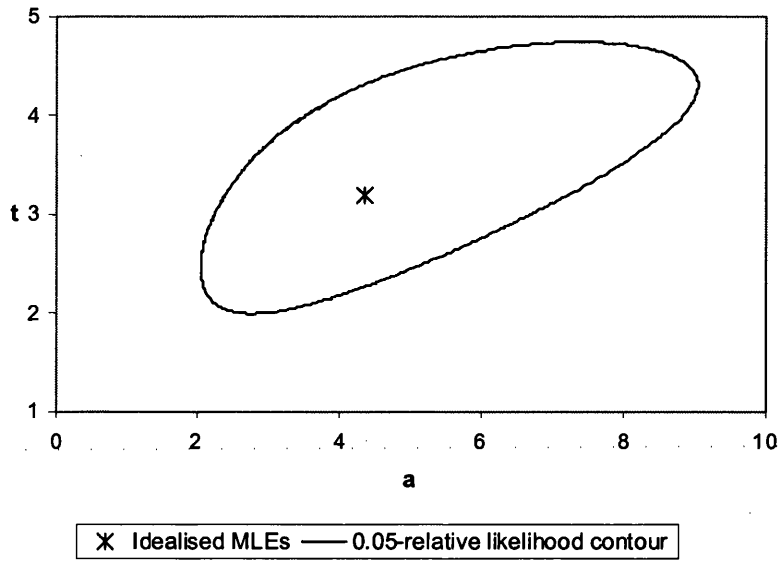

Figure 3.25: 0.05-relative likelihood contour plot for $r = 15, n = 25$, for ideal Burr data generated with $\alpha = 4, \tau = 3$.

3.26 and 3.27, as n increases, they generally converge to their respective true values quicker than the means of sampling distributions of $\widehat{\alpha}_r$ and $\widehat{\tau}_r$, which are given in Tables 2.14 and 2.15.

We continue to use the values $\alpha = 4, \tau = 3$ with $\lambda = 0.05$ in the following examples based on simulated data. Figure 3.28 shows the contour maps for some ideal samples for various r and n. As found for a single set of data, for given n, we see that the contours get smaller as r increase. It is also clear that, as the sample size increases, the contour shapes become more elliptical. It is useful to combine Figures 3.28 and 3.13; the resultant plots are displayed in Figure 3.29. For $r = 0.8n$, we see the relative likelihood contours tend to appear to the right of the large-sample probability ellipses, and capture the behaviour of the Type II censored MLEs more accurately. We will next compare the two confidence regions of (α, τ) for various censoring levels.

Relative Likelihood Contour Validation and Comparison with Normal Theory Probability Region

The method used to validate these contours is as before; we plot the 10^4 simulated observations of $(\widehat{\alpha}_r, \widehat{\tau}_r)$, and expect to find $95\% \times 10^4$ of $(\widehat{\alpha}_r, \widehat{\tau}_r)$ for which the following criterion holds:

$$l_r(\widehat{\alpha}_r, \widehat{\tau}_r) - l_r(\widehat{\alpha}_r^*, \widehat{\tau}_r^*) \geq \ln 0.05$$

r	n			
	25	50	100	1000
$0.2n : \widehat{\alpha}_r^*$	11.2485	6.4607	5.0755	4.1092
$: \widehat{\tau}_r^*$	3.9829	3.4611	3.2309	3.0266
$0.4n : \widehat{\alpha}_r^*$	5.5851	4.7184	4.3517	4.0387
$: \widehat{\tau}_r^*$	3.4305	3.2156	3.1110	3.0129
$0.6n : \widehat{\alpha}_r^*$	4.6986	4.3378	4.1700	4.0189
$: \widehat{\tau}_r^*$	3.2676	3.1366	3.0708	3.0082
$0.8n : \widehat{\alpha}_r^*$	4.3696	4.1833	4.0931	4.0103
$: \widehat{\tau}_r^*$	3.1889	3.0972	3.0506	3.0059
$1.0n : \widehat{\alpha}_r^*$	4.1969	4.0978	4.0490	4.0050
$: \widehat{\tau}_r^*$	3.1387	3.0715	3.0370	3.0041

Table 3.28: Idealised MLEs $(\hat{\alpha}_r^*, \hat{\tau}_r^*)$ for various r, n, for ideal Burr data generated with $\alpha = 4, \tau = 3$.

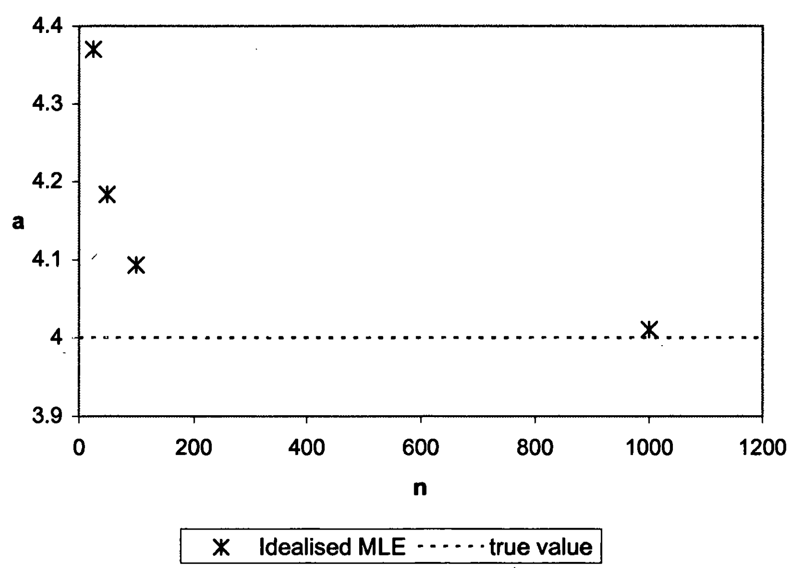

Figure 3.26: Plot of $\widehat{\alpha}_{0.8n}^*$ versus n, for ideal Burr data generated with $\alpha = 4, \tau = 3$.

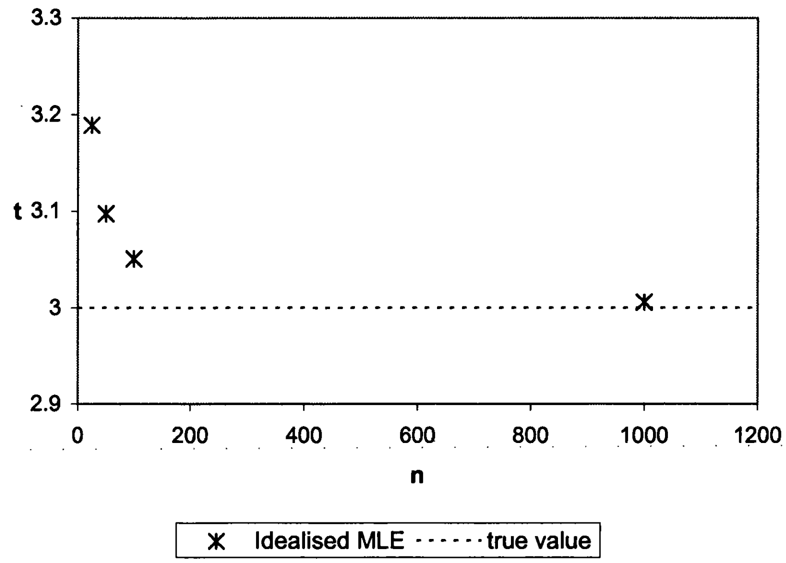

Figure 3.27: Plot of $\widehat{\tau}^*_{0.8n}$ versus n, for ideal Burr data generated with $\alpha = 4, \tau = 3$.

where $l_r(\widehat{\alpha}_r, \widehat{\tau}_r)$ and $l_r(\widehat{\alpha}^*_r, \widehat{\tau}^*_r)$ can be obtained from (2.35). We know that relative likelihood confidence regions are asymptotically equivalent to the Normal confidence regions (see, Cox & Hinkley, 1974, for example), but would like to investigate the extent to which relative likelihood approach outperforms the asymptotic Normality approach as a method to obtain the approximate 95% confidence regions for relatively small or highly censored samples.

Tables 3.29, 3.30 and 3.31, each assumes $(\alpha, \tau) = (4, 3), (0.9, 3)$ and $(4, 0.9)$ in the simulations, show some discrepancies between expected and observed values for small n and r, in part due to lack of information for estimating α and τ when the censoring level is low. We see the agreement improves, approaching 9500 as n and r increase, and is reasonably consistent across the various values of the parameters considered here. In particular, we also note the expected relative likelihood regions (upper entries) consistently record more replications of $(\widehat{\alpha}_r, \widehat{\tau}_r)$ than the elliptical probability regions (lower entries), and the upper entries converge to 9500 quicker than their lower counterparts, even at early censoring. Therefore, it transpires that relative likelihood approach provides a better measurement of precision in MLEs compared to probability regions obtained from asymptotic Normality.

We remark that it is possible to repeat the process of finding and validating expected Burr relative likelihood contours for various λ discussed earlier, such as 90%, or 99% confidence regions. As shown in Figure 3.29, the MLEs lying outside the expected relative likelihood contour are fairly informally spread around the contour. Hence, there is also scope to investigate the spread of the remaining λ% of simulated observations of $(\widehat{\alpha}_r, \widehat{\tau}_r)$ around the contour.

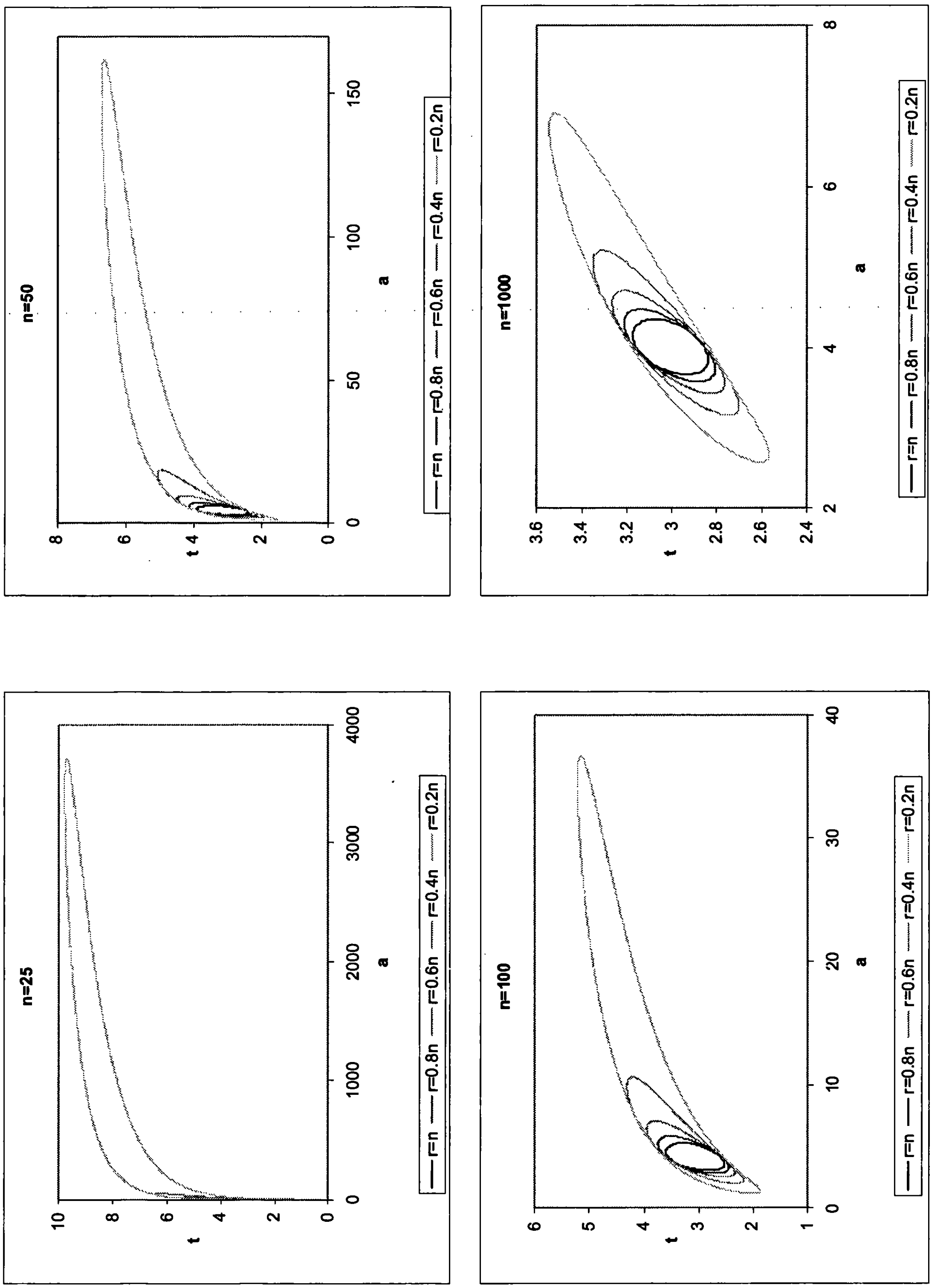

Figure 3.28: Four sets of 0.05-relative likelihood contour plots for ideal Burr data generated with $\alpha = 4, \tau = 3$.

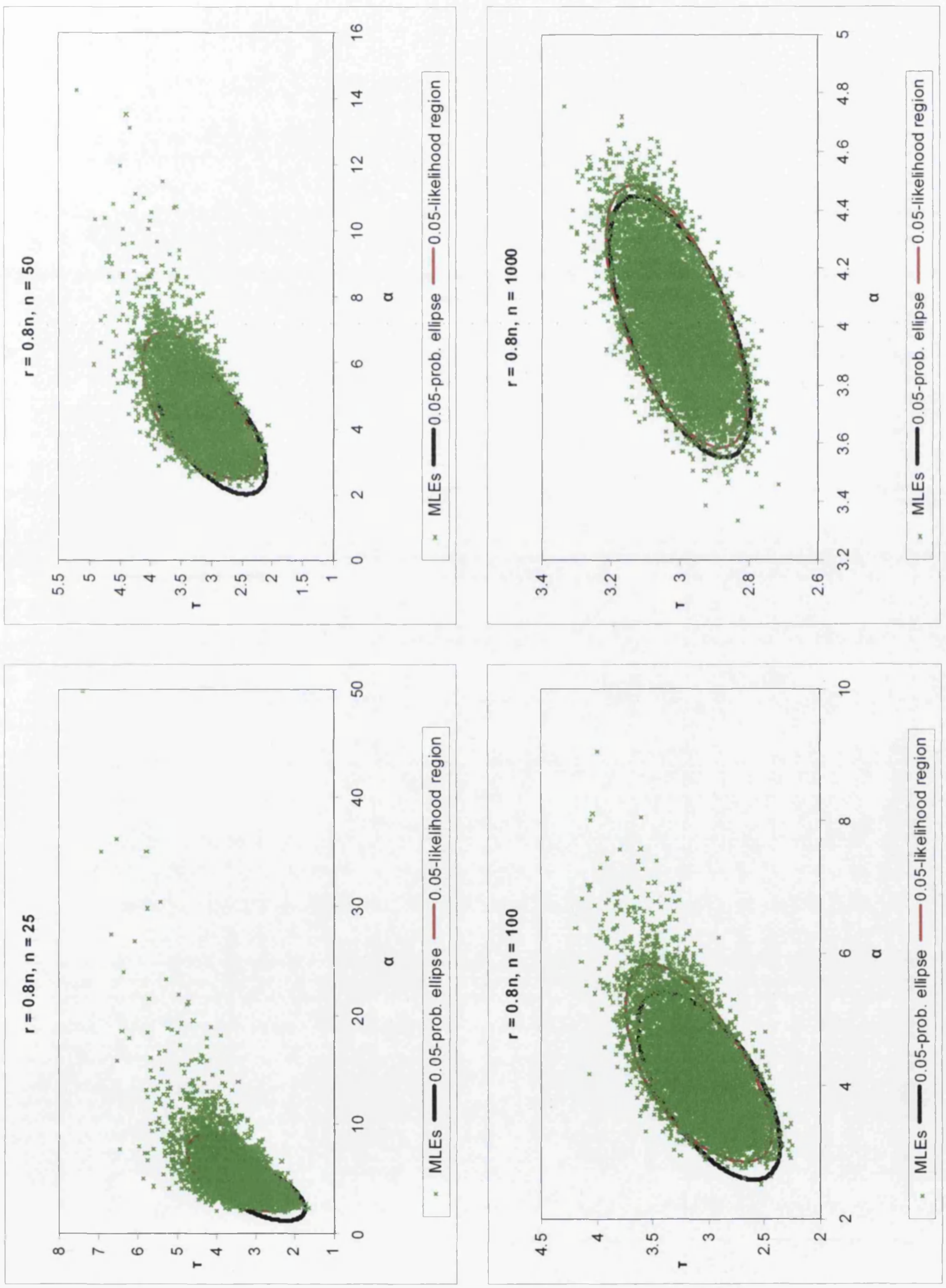

Figure 3.29: Four sets of MLEs (×) together with 0.05-relative likelihood contour and asymptotic 0.05-probability ellipse for $(\hat{\alpha}_{0.8n}, \hat{\tau}_{0.8n})$, for various n, for Burr data generated with $\alpha = 4, \tau = 3$.

r	n			
	25	50	100	1000
$0.2n$	7956	8722	9116	9497
	6199	6338	7312	9065
$0.4n$	8733	9183	9333	9488
	7018	7850	8480	9376
$0.6n$	9025	9257	9409	9517
	7910	8497	8877	9443
$0.8n$	9149	9378	9389	9499
	8371	8937	9137	9465
$1.0n$	9217	9376	9427	9518
	8736	9095	9288	9502

Table 3.29: Number of replications of $(\hat{\alpha}_r, \hat{\tau}_r)$ within the 0.05-relative likelihood contour (upper) and the asymptotic 0.05-probability ellipse (lower) for Burr data generated with $\alpha = 4, \tau = 3$.

r	n			
	25	50	100	1000
$0.2n$	7957	8685	9108	9464
	5820	7130	8114	9305
$0.4n$	8700	9125	9293	9462
	7673	8488	8926	9404
$0.6n$	8991	9239	9391	9445
	8384	8925	9230	9432
$0.8n$	9179	9334	9425	9457
	8736	9101	9298	9437
$1.0n$	9244	9355	9422	9468
	8904	9183	9349	9468

Table 3.30: Number of replications of $(\hat{\alpha}_r, \hat{\tau}_r)$ within the 0.05-relative likelihood contour (upper) and the asymptotic 0.05-probability ellipse (lower) for Burr data generated with $\alpha = 0.9, \tau = 3$.

r	n			
	25	50	100	1000
$0.2n$	7882	8816	9090	9461
	5995	6301	7245	9031
$0.4n$	8728	9139	9350	9459
	6772	7820	8561	9321
$0.6n$	8975	9237	9376	9490
	7700	8490	8959	9440
$0.8n$	9079	9303	9377	9472
	8300	8839	9155	9455
$1.0n$	9099	9291	9382	9476
	8669	9050	9296	9470

Table 3.31: Number of replications of $(\hat{\alpha}_r, \hat{\tau}_r)$ within the 0.05-relative likelihood contour (upper) and the asymptotic 0.05-probability ellipse (lower) for Burr data generated with $\alpha = 4, \tau = 0.9$.

3.5 Chapter Summary and Conclusions

Asymptotic Normality of MLE leads to symmetric confidence intervals for a single parameter, and elliptical confidence regions for two. This large sample result is often used in inference from small to moderate samples, despite the drawback that it is not always accurate with such sample sizes. Moreover, there appears to be no referenced information to which how large a sample should be before this large-sample assumption may hold.

The work reported in this chapter shows that, unless sample size is very large (generally larger than $n = 1000$), the hypothesis that the marginal distribution of a MLE is Normal should be regarded as implausible; clearly, we then reach the same conclusion about the hypothesis that the joint distribution of the MLEs is multivariate Normal. In general, the progress towards Normality is slow, but increasing the censoring number will expedite this progress. For a small sample size, typically less than $n = 100$, the distributions of MLEs tend to skew to the right, leading to a non-elliptical joint distribution; Billmann *et al.* (1972) argue that the slow convergence to Normality is due to the lack of symmetry when the samples are censored on one side (from the right). Then, it would be of interest to investigate whether the distribution of MLE would be left skewed if the data are left censored; this, however, is not our main focus, but is noted as a topic for further research. We have also shown that the sample distributions of $\widehat{B}_{0.1,r}$ approach symmetry and converge to Normality at earlier censoring and smaller sample size than the MLEs of parameters.

Despite these poor approximations to the Normal distribution, the corresponding probability intervals and ellipses still provided good coverage of the MLEs, but the shape of the distribution was not so well represented. We then considered, via an intuitive interpretation of (3.2), relative likelihood as an alternative method to asymptotic Normal theory to measure the precision of the MLEs, for Type II censored samples of small to moderate size. By extending the work by Watkins & Leech (1989) and Watkins (2004) for the Weibull case, we have derived an algorithm for drawing relative likelihood contours for Type II censored Burr data, and illustrated this procedure in detail for the arthritic patients data. We have also shown that the non-elliptical nature of the (expected) relative likelihood contours reflects more accurately than the large-sample probability ellipses the behaviour of the sampling distributions of the Type II censored MLEs for relatively small and/or highly censored samples. There is obvious scope to investigate the relative size of the two confidence regions, as well as the extent of the overlap in general; this will, nonetheless, be studied elsewhere.

We can now use these asymptotic theoretical results and move on to consider the link between the interim and final MLEs of parameters and $B_{0.1}$; this will involve taking joint expectations on the components of the Type II censored and complete score functions, which in turn requires various forms of moments and product moments of order statistics. Therefore, Chapter 4 aims to solve these moments, introducing the derivatives method, before moving on to Chapter 5 to look at the correlation between the two sets of MLEs.

Chapter 4

Moments and Product Moments of Order Statistics

4.1 Introduction

We have already seen that order statistics arise naturally in the analysis of reliability data subject to Type II censoring due to the method of experimentation. In considering the extent to which an interim analysis - here, using information based on Type II censored samples - provides a guide to the final analysis, we will require the study of the correlations between the complete and the Type II censored MLEs; for large samples, it can be shown that this is equivalent to a study of the correlations of score functions, which thus involves various forms of expectations and joint expectations of order statistics. We now outline here some useful preliminary work.

The moments of order statistics have generated considerable interest in statistical inference and, in fact, have been studied, and, where appropriate, tabulated quite extensively for many distributions; for instance, Joshi (1978, 1982) in the exponential distribution; Lieblein (1955) in the Weibull distribution; Malik (1966) in the Pareto distribution; Khan & Khan (1987) and Pawles & Szynal (2001) in the Burr distribution.

From (2.2), taking expected values on the products of the complete and Type II censored score functions means that we will require expectations of the form

$$E[g\,(X_{i:n})], \tag{4.1}$$

and joint expectations of the form

$$E[g\,(X_{i:n})\,h(X_{j:n})], \tag{4.2}$$

in which the arbitrary functions g and h usually involve logarithms and/or powers of $X_{i:n}$.

For instance, as seen at Section 2.3.1 for the Weibull distribution, (4.1) is typically of

$$E\left[X_{i:n}^{p}\left\{\ln X_{i:n}\right\}^{a}\right]$$

for some positive integers a and p; more specifically, these are

$$X_{i:n},\quad \ln X_{i:n},\quad X_{i:n}\ln X_{i:n},\quad \text{and } X_{i:n}(\ln X_{i:n})^{2}. \tag{4.3}$$

We consider two methods to solve (4.1) and (4.2). The first employs the conventional definition of expectation; as an illustration,

$$E\left[X_{i:n}^{p}\left\{\ln X_{i:n}\right\}^{a}\right]=\int_{0}^{\infty}x^{p}\left(\ln x\right)^{a}f_{(i)}(x)dx,$$

which we will refer to as the direct method. Therefore, depending on the form of $g\left(X_{i:n}\right)$ and $f_{(i)}(x)$, this approach may involve integrations of some complex functions. The second introduces an alternative based on repetitive partial differentiations of the moments and product moments of order statistics, as discussed below.

4.1.1 The Derivatives Method

In the exponential, Weibull and Burr distributions, μ_p is well defined so that expressions for $E[X_{i:n}^{p}]$ can be written down without too much difficulty. In particular, we have seen $E[X_{i:n}]$ at (3.3) and (3.14) for the Weibull and Burr distributions. Hence, differentiating $E[X_{i:n}^{p}]$ wrt p a times will yield $E[X_{i:n}^{p}\left\{\ln X_{i:n}\right\}^{a}]$; this effectively introduces the term $\ln X_{i:n}$ whose power depends on the order of differentiation, in addition to keeping the term $X_{i:n}^{p}$ in place. As a result, we can easily obtain expressions for the functions at (4.3) by replacing a and p by 0 and 1, 1 and 0, 1 and 1, 2 and 1, in turn. Similarly, for the joint expectations of $X_{i:n}^{p}$ and $X_{j:n}^{q}$, we can obtain a general expression for $E\left[X_{i:n}^{p}\left\{\ln X_{i:n}\right\}^{a}X_{j:n}^{q}\left\{\ln X_{j:n}\right\}^{b}\right]$ by applying the operator

$$\frac{\partial^{a+b}}{\partial p^{a}\partial q^{b}}$$

to $E[X_{i:n}^{p}X_{j:n}^{q}]$, and the formulae for specific expectations can be obtained on appropriate substitutions of a, b, p and q by positive integers. We note that such technique has been employed by Watkins (1997) and Watkins & Johnson (2002) to obtain results for the expectations of the first and second derivatives of the log-likelihood function for the Burr distribution under complete and Type I censoring regime respectively.

In Section 4.2 we begin with the assumption that the lifetimes follow the Weibull distribution, and, as mentioned in Section 2.3.1, we exploit the link between Weibull and standard exponential distributions to reduce the expectations to the standard exponential case. We will solve for (4.1) and (4.2) using the direct and derivatives methods. Results obtained from both methods are then validated for various combinations of i and j using simulation experiments. Then, in Section 4.3, we repeat the analysis for the Burr distribution, in which we benefit from the recurrence relationship given at (1.51).

4.2 Weibull and Std Exponential Order Statistics

Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the order statistics obtained from a random sample of size n drawn from the Weibull distribution. When considering the correlation of the complete and the Type II censored Weibull score functions, the form of the partial derivatives at (2.17), (2.18), (2.31) and (2.32) suggests that (4.1) will generally be of the form

$$E\left[\left(\frac{X_{i:n}^{\beta}}{\theta^{\beta}}\right)^{p}\left\{\ln\left(\frac{X_{i:n}^{\beta}}{\theta^{\beta}}\right)\right\}^{a}\right] \tag{4.4}$$

for $a = 0, 1, 2$ and $p = 0, 1$, while (4.2) will generally be of the form

$$E\left[\left(\frac{X_{i:n}^{\beta}}{\theta^{\beta}}\right)^{p}\left\{\ln\left(\frac{X_{i:n}^{\beta}}{\theta^{\beta}}\right)\right\}^{a}\left(\frac{X_{j:n}^{\beta}}{\theta^{\beta}}\right)^{q}\left\{\ln\left(\frac{X_{j:n}^{\beta}}{\theta^{\beta}}\right)\right\}^{b}\right] \tag{4.5}$$

for $a, b, p, q = 0, 1$. In some, but not all, cases, $a = b$ and $p = q$. We can, of course, obtain these expectations from Weibull pdf and cdf, but we can also exploit the connection between Weibull and standard exponential distributions.

4.2.1 Link between the Weibull and Standard Exponential Distributions

We have previously noted that a natural extension of the exponential distribution is the Weibull distribution; hence, it is often convenient to derive results for one case and then transfer to the other. In fact, we have already employed in Section 2.3.1 the transformation of Weibull random variable X into standard exponential random variable Z, given at (2.26), to obtain the elements of the Weibull EFI matrix. Therefore, using (2.26), we see (4.4) and (4.5) reduce, respectively, to

$$E\left[Z_{i:n}^{p}(\ln Z_{i:n})^{a}\right] \tag{4.6}$$

for $a = 0, 1, 2$ and $p = 0, 1$, and

$$E\left[Z_{i:n}^{p}(\ln Z_{i:n})^{a} Z_{j:n}^{q}(\ln Z_{j:n})^{b}\right] \tag{4.7}$$

for $a, b, p, q = 0, 1$. As before, in some, but not all, instances, $a = b$ and $p = q$. Next, we briefly present some results for the standard exponential order statistics.

4.2.2 Standard Exponential Order Statistics

For later convenience, it is suitable to summarise here some basic results on the moments of standard exponential order statistics. Suppose $Z_{1:n} \leq Z_{2:n} \leq \cdots \leq Z_{n:n}$ denote the order statistics in a random sample of size n from a standard (so $\theta = 1$ in (1.27) and (1.28)) exponential population, with pdf

$$f(z) = e^{-z}, \tag{4.8}$$

and cdf

$$F(z) = 1 - e^{-z}, \tag{4.9}$$

for $z \geq 0$. By writing

$$Z_{i:n} = (Z_{i:n} - Z_{i-1:n}) + (Z_{i-1:n} - Z_{i-2:n}) + \cdots + (Z_{2:n} - Z_{1:n}) + Z_{1:n} = \sum_{k=1}^{i} Z_{k:n} - Z_{k-1:n}$$

with the convention $Z_{0:n} = 0$, we see that we can exploit the lack-of-memory property to obtain

$$Z_{i:n} = \sum_{k=1}^{i} \frac{W_k}{n-k+1}$$

so that W_k, defined above at (2.9), are now independent and identically distributed variables with pdf (4.8). Using this result, it is straightforward to write down the mean and variance of $Z_{i:n}$; we have

$$E\left[Z_{i:n}\right] = \sum_{k=1}^{i} \frac{1}{n-k+1} \tag{4.10}$$

and

$$Var(Z_{i:n}) = \sum_{k=1}^{i} \frac{1}{(n-k+1)^2}.$$

Moreover, writing $Z_{j:n} = (Z_{j:n} - Z_{i:n}) + Z_{i:n}$, the covariance of $Z_{i:n}$ and $Z_{j:n}$ $(1 \leq i < j \leq n)$ is

$$Cov(Z_{i:n}, Z_{j:n}) = Cov(Z_{i:n}, Z_{j:n} - Z_{i:n}) + Cov(Z_{i:n}, Z_{i:n}) = Var(Z_{i:n})$$

since $Z_{i:n}$ and the increment $Z_{j:n} - Z_{i:n}$ are independent due to the lack-of-memory property. We further obtain the joint expectation of $Z_{i:n}$ and $Z_{j:n}$ $(1 \leq i < j \leq n)$ as

$$\begin{aligned} E\left[Z_{i:n}Z_{j:n}\right] &= Var(Z_{i:n}) + E\left[Z_{i:n}\right]E\left[Z_{j:n}\right] \\ &= \sum_{k=1}^{i} \frac{1}{(n-k+1)^2} + \left(\sum_{k=1}^{i} \frac{1}{n-k+1}\right) \times \left(\sum_{k=1}^{j} \frac{1}{n-k+1}\right) \end{aligned} \tag{4.11}$$

We can now move on to derive formulae for special cases of (4.6) and (4.7) using, first, the direct method, followed by the derivatives method.

4.2.3 Expectations of $g(Z_{i:n})$

It will be shown in next chapter (see Section 5.3.1.1) that we will require the following special cases of (4.6):

$$E[\ln Z_{i:n}], \quad (4.12a)$$
$$E[Z_{i:n} \ln Z_{i:n}], \quad (4.12b)$$
$$E\left[(Z_{i:n})^2 \ln Z_{i:n}\right], \quad (4.12c)$$
$$E\left[(\ln Z_{i:n})^2\right], \quad (4.12d)$$
$$E\left[Z_{i:n}(\ln Z_{i:n})^2\right], \quad (4.12e)$$
$$E\left[(Z_{i:n} \ln Z_{i:n})^2\right]. \quad (4.12f)$$

Direct Method

From (1.40), the marginal pdf of $Z_{i:n}$ is

$$f_{(i)}(z) = c_{i:n} e^{-(n-i+1)z} [1 - e^{-z}]^{i-1}$$

and we can use the Binomial Theorem to expand the square bracket as

$$\left[1 - e^{-z}\right]^{i-1} = \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} e^{-(i-1-k)z}$$

to give

$$f_{(i)}(z) = c_{i:n} \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} e^{-(n-k)z}.$$

Then, using (1.46), we have

$$\begin{aligned} E\left[Z_{i:n}^p (\ln Z_{i:n})^a\right] &= \int_0^\infty z^p (\ln z)^a f_{(i)}(z) dz \\ &= c_{i:n} \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} \int_0^\infty z^p (\ln z)^a e^{-(n-k)z} dz \\ &= c_{i:n} \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} A_{n-k}^{pa}, \end{aligned} \quad (4.13)$$

where

$$A_s^{pa} = \int_0^\infty z^p (\ln z)^a e^{-sz} dz$$

is related to gamma and polygamma functions defined in Section 1.2.2.1. As a result, for $E[\ln Z_{i:n}]$, we require

$$A_s^{01} = \int_0^\infty (\ln z) e^{-sz} dz = -\frac{\gamma + \ln s}{s},$$

so that, from (4.13), we obtain (4.12a) as

$$\begin{aligned} E\left[\ln Z_{i:n}\right] &= c_{i:n} \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} A_{n-k}^{01} \\ &= c_{i:n} \sum_{k=0}^{i-1} \frac{(-1)^{i-1-k} \binom{i-1}{k}}{(n-k)} \left\{-\gamma - \ln(n-k)\right\}. \end{aligned} \tag{4.14}$$

Likewise, in $E\left[Z_{i:n} \ln Z_{i:n}\right]$ and $E\left[(Z_{i:n})^2 \ln Z_{i:n}\right]$, the relevant integrals are, respectively,

$$A_s^{11} = \int_0^\infty z\,(\ln z)\, e^{-sz} dz = \frac{1-\gamma-\ln s}{s^2},$$

and

$$A_s^{21} = \int_0^\infty z^2\,(\ln z)\, e^{-sz} dz = \frac{3-2\gamma-2\ln s}{s^3}.$$

Next, $E\left[(\ln Z_{i:n})^2\right]$ needs

$$A_s^{02} = \int_0^\infty (\ln z)^2\, e^{-sz} dz = \frac{1}{s}\left\{\frac{\pi^2}{6} + (-\gamma - \ln s)^2\right\},$$

and for $E\left[Z_{i:n}(\ln Z_{i:n})^2\right]$, we want

$$A_s^{12} = \int_0^\infty z\,(\ln z)^2\, e^{-sz} dz = \frac{1}{s^2}\left\{\frac{\pi^2}{6} - 1 + (1-\gamma-\ln s)^2\right\}.$$

The final expectation is $E\left[(Z_{i:n} \ln Z_{i:n})^2\right]$, for which we require

$$A_s^{22} = \int_0^\infty z^2\,(\ln z)^2\, e^{-sz} dz = \frac{2}{s^3}\left\{\frac{\pi^2}{6} - \frac{5}{4} + \left(\frac{3}{2} - \gamma - \ln s\right)^2\right\}.$$

Derivatives Method

Basic expectation It will prove useful to begin with the following preliminary:

$$\begin{aligned} E[Z_{i:n}^p] &= c_{i:n} \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} \int_0^\infty z^p e^{-(n-k)z} dz \\ &= c_{i:n} \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} \frac{\Gamma(p+1)}{(n-k)^{p+1}}, \end{aligned} \tag{4.15}$$

which reduces to (4.10) when $p = 1$, as required. In particular, its first and second partial derivatives wrt p will introduce the term $\ln Z_{i:n}$, and, in turn, yield the expectations in (4.12), in terms of the digamma and polygamma functions.

Expectations in (4.12) The first partial derivative of (4.15) wrt p gives

$$E[Z_{i:n}^p \ln Z_{i:n}] = c_{i:n} \sum_{k=0}^{i-1} \frac{(-1)^{i-1-k} \binom{i-1}{k} \Gamma(p+1)}{(n-k)^{p+1}} \left\{\psi(p+1) - \ln(n-k)\right\}. \tag{4.16}$$

Hence, using the gamma and digamma values in Table 1.6 and setting $p = 0, 1, 2$ in (4.16) will yield, respectively, the following expectations:

$$E\left[\ln Z_{i:n}\right] = c_{i:n} \sum_{k=0}^{i-1} \frac{(-1)^{i-1-k} \binom{i-1}{k}}{(n-k)} \left\{-\gamma - \ln(n-k)\right\}, \tag{4.17}$$

$$E\left[Z_{i:n} \ln Z_{i:n}\right] = c_{i:n} \sum_{k=0}^{i-1} \frac{(-1)^{i-1-k} \binom{i-1}{k}}{(n-k)^2} \left\{1 - \gamma - \ln(n-k)\right\}, \tag{4.18}$$

$$E\left[(Z_{i:n})^2 \ln Z_{i:n}\right] = c_{i:n} \sum_{k=0}^{i-1} \frac{(-1)^{i-1-k} \binom{i-1}{k}}{(n-k)^3} \left\{3 - 2\gamma - 2\ln(n-k)\right\}.$$

Furthermore, the second partial derivative of (4.15) wrt p gives

$$E[Z_{i:n}^p (\ln Z_{i:n})^2] = c_{i:n} \sum_{k=0}^{i-1} \frac{(-1)^{i-1-k} \binom{i-1}{k} \Gamma(p+1)}{(n-k)^{p+1}} \left\{\psi'(p+1) + [\psi(p+1) - \ln(n-k)]^2\right\}$$

so that, likewise, setting $p = 0, 1, 2$ in this result will yield, in turn,

$$E\left[(\ln Z_{i:n})^2\right] = c_{i:n} \sum_{k=0}^{i-1} \frac{(-1)^{i-1-k} \binom{i-1}{k}}{(n-k)} \left\{\frac{\pi^2}{6} + [-\gamma - \ln(n-k)]^2\right\},$$

$$E\left[Z_{i:n} (\ln Z_{i:n})^2\right] = c_{i:n} \sum_{k=0}^{i-1} \frac{(-1)^{i-1-k} \binom{i-1}{k}}{(n-k)^2} \left\{\frac{\pi^2}{6} - 1 + [1 - \gamma - \ln(n-k)]^2\right\},$$

$$E\left[(Z_{i:n} \ln Z_{i:n})^2\right] = 2c_{i:n} \sum_{k=0}^{i-1} \frac{(-1)^{i-1-k} \binom{i-1}{k}}{(n-k)^3} \left\{\frac{\pi^2}{6} - \frac{5}{4} + \left[\frac{3}{2} - \gamma - \ln(n-k)\right]^2\right\}.$$

We note that these results are identical to those computed from direct integration; see, for instance, (4.14) and (4.17), and also can be compared with their counterparts for $E\left[Z^p (\ln Z)^a\right]$ given at Table 1 in Watkins (1998).

Some Numerical Details and Discussion

Being new results, it is important to check these expressions against simulation experiments. For illustration, we plot each expectation in (4.12) as a function of i for $n = 1000$, in Figures 4.1 to 4.6; for graphical convenience, simulated values are shown in steps of 50. These show that there is very little difference between theoretical (calculated from the direct method) and simulated results, based on 10^4 replications. Take, for instance, $E\left[Z_{i:n} \ln Z_{i:n}\right]$, Table 4.1 summarises the agreement between theoretical values, obtained from both direct (upmost entries) and derivatives (middle entries) methods, and their simulated counterparts (lowest entries) for varying i and n; this confirms that the theoretical and simulated data are indeed consistently the same up to 2 decimal places, and investigations for other expectations in (4.12) provide similar observations. We also note, from Figures 4.2, 4.3, 4.5 and 4.6, that $E\left[Z_{i:n} \ln Z_{i:n}\right]$, $E\left[(Z_{i:n})^2 \ln Z_{i:n}\right]$, $E\left[Z_{i:n} (\ln Z_{i:n})^2\right]$ and $E\left[(Z_{i:n} \ln Z_{i:n})^2\right]$ are relatively constant for $i \leq 800$.

i	n					
	25	50	100	1000	2500	5000
$0.2n$: direct	-0.3110	-0.3226	-0.3286	-0.3341	-0.3345	-0.3346
: deriv.	-0.3110	-0.3226	-0.3286	-0.3341	-0.3345	-0.3346
: simul.	-0.3109	-0.3224	-0.3287	-0.3341	-0.3345	-0.3346
$0.4n$: direct	-0.3223	-0.3325	-0.3378	-0.3426	-0.3429	-0.3430
: deriv.	-0.3223	-0.3325	-0.3378	-0.3426	-0.3429	-0.3430
: simul.	-0.3219	-0.3320	-0.3381	-0.3425	-0.3429	-0.3431
$0.6n$: direct	-0.0753	-0.0776	-0.0788	-0.0800	-0.0801	-0.0801
: deriv.	-0.0753	-0.0776	-0.0788	-0.0800	-0.0801	-0.0801
: simul.	-0.0728	-0.0789	-0.0784	-0.0798	-0.0801	-0.0802
$0.8n$: direct	0.7000	0.7323	0.7490	0.7642	0.7652	0.7656
: deriv.	0.7000	0.7323	0.7490	0.7642	0.7652	0.7656
: simul.	0.7030	0.7315	0.7492	0.7646	0.7646	0.7655
$1.0n$: direct	5.3073	6.9363	8.6885	15.1729	17.9755	20.1646
: deriv.	5.3073	6.9363	8.6885	15.1729	17.9755	20.1646
: simul.	5.2958	6.9279	8.7466	15.1135	17.9826	20.1639

Table 4.1: Numerical comparison of $E[Z_{i:n} \ln Z_{i:n}]$ for various i and n.

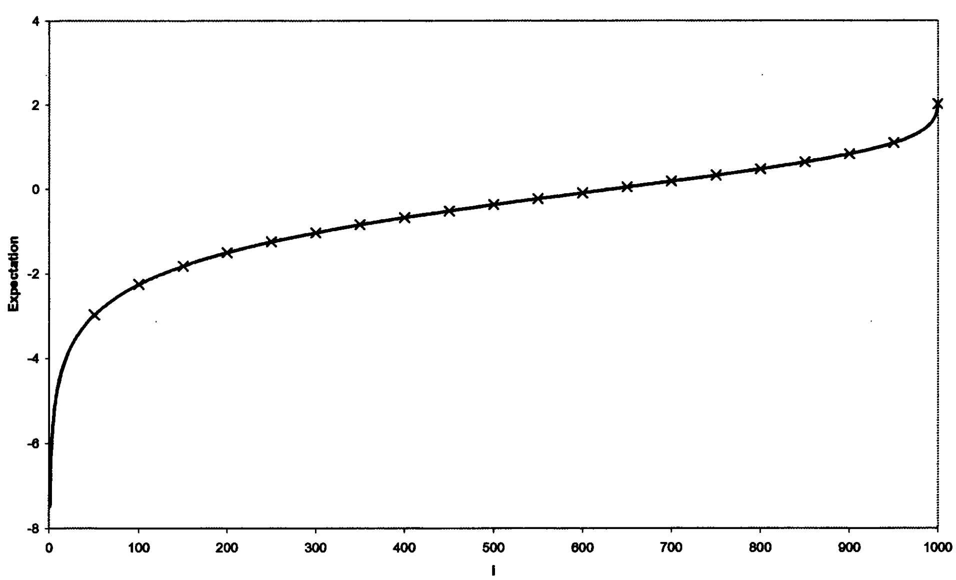

Figure 4.1: Theoretical (–) and simulated (×) values of $E\left[\ln Z_{i:n}\right]$ versus i, for $n = 1000$.

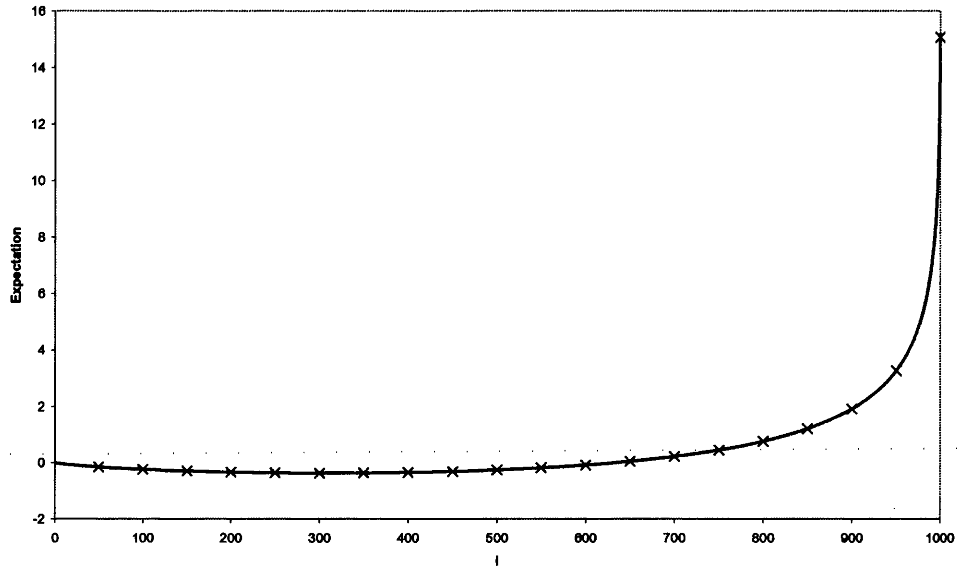

Figure 4.2: Theoretical (−) and simulated (×) values of $E\left[Z_{i:n} \ln Z_{i:n}\right]$ versus i, for $n = 1000$.

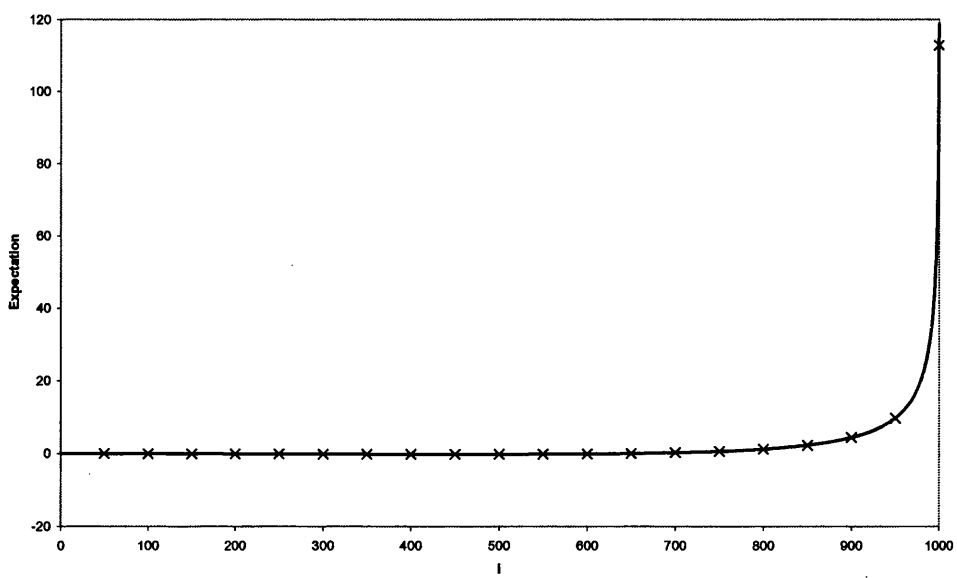

Figure 4.3: Theoretical (−) and simulated (×) values of $E\left[(Z_{i:n})^2 \ln Z_{i:n}\right]$ versus i, for $n = 1000$.

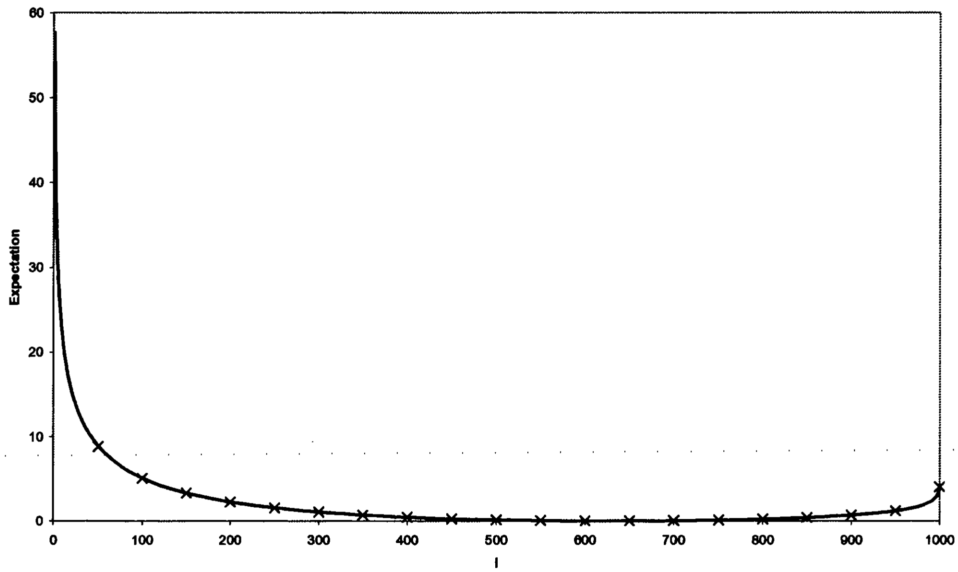

Figure 4.4: Theoretical (–) and simulated (×) values of $E\left[(\ln Z_{i:n})^2\right]$ versus i, for $n = 1000$.

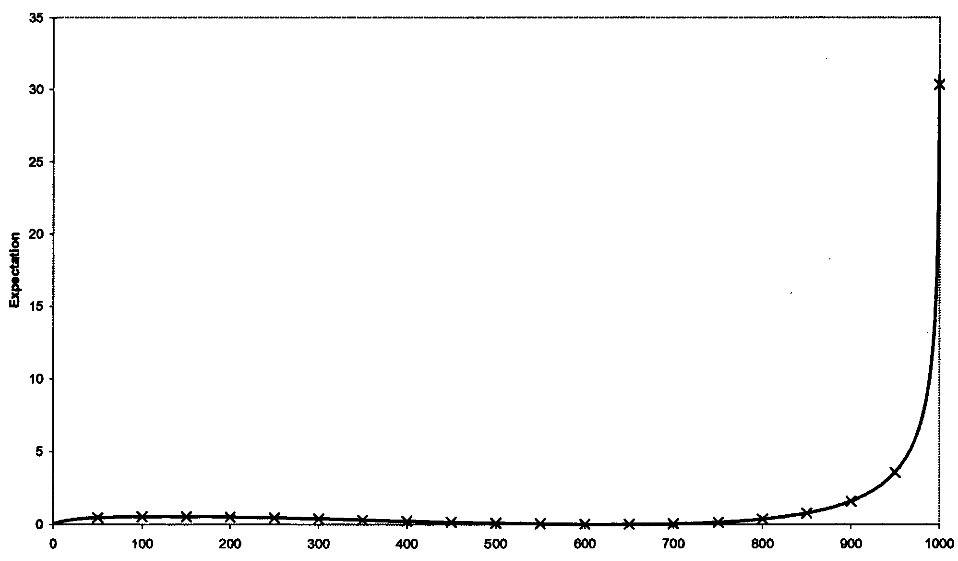

Figure 4.5: Theoretical (–) and simulated (×) values of $E\left[Z_{i:n}(\ln Z_{i:n})^2\right]$ versus i, for $n = 1000$.

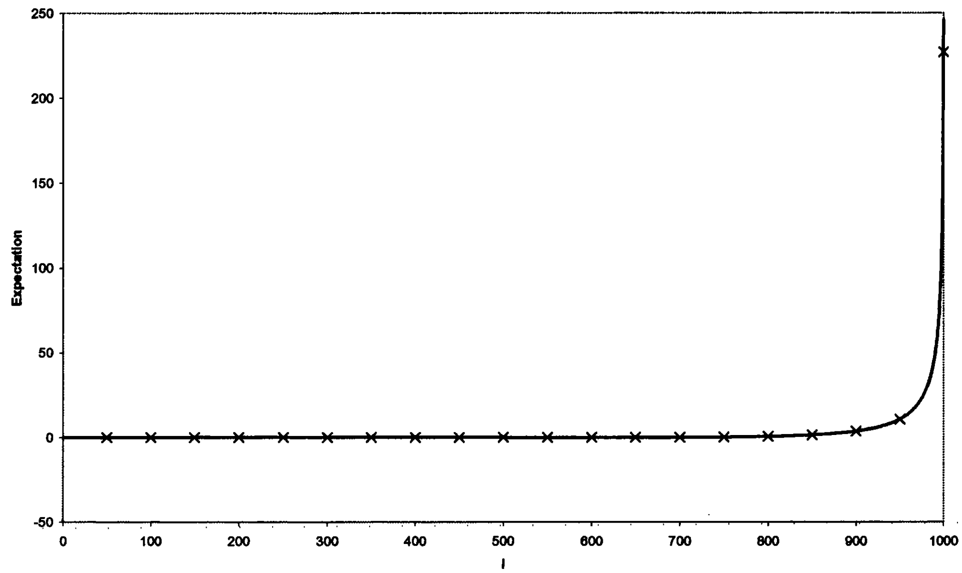

Figure 4.6: Theoretical (−) and simulated (×) values of $E\left[(Z_{i:n}\ln Z_{i:n})^2\right]$ versus i, for $n = 1000$.

4.2.4 Joint Expectations of $g(Z_{i:n})$ and $h(Z_{j:n})$

Similarly, we will consider the following special cases of (4.7):

$$E\left[Z_{i:n}\ln Z_{j:n}\right], \tag{4.19a}$$
$$E\left[(\ln Z_{i:n})Z_{j:n}\right], \tag{4.19b}$$
$$E\left[\ln Z_{i:n}\ln Z_{j:n}\right], \tag{4.19c}$$
$$E\left[Z_{i:n}Z_{j:n}\ln Z_{j:n}\right], \tag{4.19d}$$
$$E\left[Z_{i:n}(\ln Z_{i:n})Z_{j:n}\right], \tag{4.19e}$$
$$E\left[(\ln Z_{i:n})Z_{j:n}\ln Z_{j:n}\right], \tag{4.19f}$$
$$E\left[Z_{i:n}\ln Z_{i:n}\ln Z_{j:n}\right], \tag{4.19g}$$
$$E\left[Z_{i:n}(\ln Z_{i:n})Z_{j:n}(\ln Z_{j:n})\right]. \tag{4.19h}$$

As with single expectations, we consider two ways - direct and derivatives - to compute these expectations.

Direct Method

From (1.45), the joint pdf of $Z_{i:n}$ and $Z_{j:n}$ $(1 \le i < j \le n)$ can be defined as

$$f_{(i,j)}(x,y) = c_{i,j:n}\left[1-e^{-x}\right]^{i-1}\left[e^{-x}-e^{-y}\right]^{j-i-1}e^{-x}e^{-(n-j+1)y}$$

for $0 \le x < y < \infty$. We now expand both square brackets inside the integrals, writing

$$\left[1 - e^{-x}\right]^{i-1} = \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} e^{-(i-1-k)x}$$

and

$$\left[e^{-x} - e^{-y}\right]^{j-i-1} = \sum_{l=0}^{j-i-1} (-1)^{j-i-1-l} \binom{j-i-1}{l} e^{-lx} e^{-(j-i-1-l)y},$$

so that

$$f_{(i,j)}(x,y) = c_{i,j:n} \sum_{k=0}^{i-1} \sum_{l=0}^{j-i-1} (-1)^{j-k-l} \binom{i-1}{k} \binom{j-i-1}{l} e^{-(i+l-k)x} e^{-(n-i-l)y}.$$

Using (1.47), the joint expectation, $E\left[Z_{i:n}^p (\ln Z_{i:n})^a Z_{j:n}^q (\ln Z_{j:n})^b\right]$, is given by

$$\begin{aligned}
& c_{i,j:n} \int_{y=0}^{\infty} \int_{x=0}^{y} x^p (\ln x)^a y^q (\ln y)^b f_{(i,j)}(x,y) dx dy \\
= \; & c_{i,j:n} \sum_{k=0}^{i-1} \sum_{l=0}^{j-i-1} (-1)^{j-k-l} \binom{i-1}{k} \binom{j-i-1}{l} \times \\
& \int_{y=0}^{\infty} \int_{x=0}^{y} x^p (\ln x)^a y^q (\ln y)^b e^{-(i+l-k)x} e^{-(n-i-l)y} dx dy \\
= \; & c_{i,j:n} \sum_{k=0}^{i-1} \sum_{l=0}^{j-i-1} (-1)^{j-k-l} \binom{i-1}{k} \binom{j-i-1}{l} A_{i+l-k,n-i-l}^{pa,qb} \qquad (4.20)
\end{aligned}$$

at which we define

$$A_{s,t}^{pa,qb} = \int_{y=0}^{\infty} \int_{x=0}^{y} x^p (\ln x)^a e^{-sx} y^q (\ln y)^b e^{-ty} dx dy$$

where the parameters a, b, p, q, s and t are real and positive. Hence, we can anticipate some lengthy algebra here, involving functions like gamma and polygamma, exponential integrals and Lerch transcendent function, as well as the connections between these functions, as presented in Chapter 1. Next, we derive the different combinations of $A_{s,t}^{pa,qb}$ required when evaluating each function in (4.19) in turn:

$$A_{s,t}^{10,01}, A_{s,t}^{01,10}, A_{s,t}^{01,01}, A_{s,t}^{10,11}, A_{s,t}^{11,10}, A_{s,t}^{01,11}, A_{s,t}^{11,01}, A_{s,t}^{11,11}.$$

1. $E\left[Z_{i:n}\ln Z_{j:n}\right]$ We need

$$\begin{aligned}
A^{10,01}_{s,t} &= \int_{y=0}^{\infty}\int_{x=0}^{y} xe^{-sx}(\ln y)\,e^{-ty}dxdy \\
&= \int_{y=0}^{\infty}(\ln y)e^{-ty}\left[\int_{x=0}^{y} xe^{-sx}dx\right]dy \\
&= \int_{y=0}^{\infty}(\ln y)e^{-ty}\left[\frac{1}{s^2}\left(1-e^{-sy}-sye^{-sy}\right)\right]dy \\
&= \frac{1}{s^2}\left\{-\frac{\gamma+\ln t}{t}+\frac{\gamma+\ln(s+t)}{s+t}-\frac{s\left[1-\gamma-\ln(s+t)\right]}{(s+t)^2}\right\} \\
&= \frac{1}{s^2t(s+t)^2}\left\{-s\left(\gamma s+t\right)-(s+t)^2\ln t+t(2s+t)\ln(s+t)\right\}
\end{aligned}$$

for which the individual integrations wrt y are

$$\int_{y=0}^{\infty}(\ln y)e^{-ty}dy = -\frac{\gamma+\ln t}{t}, \tag{4.21}$$

$$\int_{y=0}^{\infty}(\ln y)e^{-(s+t)y}dy = -\frac{\gamma+\ln(s+t)}{s+t}, \tag{4.22}$$

$$\int_{y=0}^{\infty}y(\ln y)e^{-(s+t)y}dy = \frac{1-\gamma-\ln(s+t)}{(s+t)^2}. \tag{4.23}$$

2. $E\left[(\ln Z_{i:n})Z_{j:n}\right]$ The relevant integral is

$$\begin{aligned}
A^{01,10}_{s,t} &= \int_{y=0}^{\infty}\int_{x=0}^{y}(\ln x)e^{-sx}ye^{-ty}dxdy \\
&= \int_{y=0}^{\infty}ye^{-ty}\left[\int_{x=0}^{y}(\ln x)e^{-sx}dx\right]dy \\
&= \int_{y=0}^{\infty}ye^{-ty}\left[-\frac{1}{s}\left(\gamma+\ln s+E_1(sy)+e^{-sy}\ln y\right)\right]dy \\
&= -\frac{1}{s}\left\{\frac{\gamma+\ln s}{t^2}+\frac{1}{t^2}\left[\ln(1+\frac{t}{s})-\frac{t}{s+t}\right]+\frac{1-\gamma-\ln(s+t)}{(s+t)^2}\right\} \\
&= -\frac{1}{t^2(s+t)^2}\left\{(s+2t)\left[\gamma+\ln(s+t)\right]-t\right\},
\end{aligned}$$

obtained via

$$\int_{y=0}^{\infty}ye^{-ty}dy = \frac{1}{t^2}, \tag{4.24}$$

$$\int_{y=0}^{\infty}ye^{-ty}E_1(sy)dy = \frac{1}{t^2}\left[\ln(1+\frac{t}{s})-\frac{t}{s+t}\right] \text{ from (1.18)}, \tag{4.25}$$

and (4.23).

3. $E\left[\ln Z_{i:n}\ln Z_{j:n}\right]$ The internal integration of $A^{01,01}_{s,t}$ is the same as that in $A^{01,10}_{s,t}$; then, on dropping the constants, outer integration wrt y yields

$$\begin{aligned}
&\int_{y=0}^{\infty}(\ln y)e^{-ty}E_1(sy)dy \\
&= -\frac{1}{t}\left[\ln(1+\frac{t}{s})(\gamma+\ln(s+t))+\frac{t}{s+t}\Phi(\frac{t}{s+t},2,1)\right] \text{ from (1.20)} \\
&= -\frac{1}{t}\left[\ln(\frac{s+t}{s})(\gamma+\ln(s+t))+Li_2(\frac{t}{s+t})\right] \text{ from (1.22),} \qquad (4.26)
\end{aligned}$$

$$\int_{y=0}^{\infty}(\ln y)^2e^{-(s+t)y}dy = \frac{1}{s+t}\left[\frac{\pi^2}{6}+(\gamma+\ln(s+t))^2\right], \qquad (4.27)$$

and (4.21), for which

$$\begin{aligned}
A_{s,t}^{01,01} &= -\frac{1}{s}\left\{\begin{array}{c} -\frac{(\gamma+\ln s)(\gamma+\ln t)}{t} - \frac{1}{t}\left[\ln(\frac{s+t}{s})(\gamma+\ln(s+t))+Li_2(\frac{t}{s+t})\right] \\ +\frac{1}{s+t}\left[\frac{\pi^2}{6}+(\gamma+\ln(s+t))^2\right]\end{array}\right\} \\
&= -\frac{1}{st(s+t)}\left\{\begin{array}{c} -(s+t)\left[\gamma\ln t+\ln s\ln t-\ln s\ln(s+t)+Li_2(\frac{t}{s+t})\right] \\ -s\left[\gamma^2+(\ln(s+t))^2\right]+(-s+t)\gamma\ln(s+t)+\frac{t\pi^2}{6}\end{array}\right\}.
\end{aligned}$$

4. $E\left[Z_{i:n}Z_{j:n}\ln Z_{j:n}\right]$ This expectation is associated with $A_{s,t}^{10,11}$ whose inner integral is identical to that in $A_{s,t}^{10,01}$; hence the outer integrals (omitting the constants wrt y) are

$$\begin{aligned}
\int_{y=0}^{\infty}y(\ln y)e^{-ty}dy &= \frac{1-\gamma-\ln t}{t^2}, \qquad (4.28) \\
\int_{y=0}^{\infty}y^2(\ln y)e^{-(s+t)y}dy &= \frac{3-2\gamma-2\ln(s+t)}{(s+t)^3}, \qquad (4.29)
\end{aligned}$$

and (4.23), which lead to

$$\begin{aligned}
A_{s,t}^{10,11} &= \frac{1}{s^2}\left\{\frac{1-\gamma-\ln t}{t^2}-\frac{1-\gamma-\ln(s+t)}{(s+t)^2}-\frac{s\left[3-2\gamma-2\ln(s+t)\right]}{(s+t)^3}\right\} \\
&= \frac{1}{s^2t^2(s+t)^3}\left\{\begin{array}{c} -s\left[s(s+3t)(\gamma-1)+t^2\right] \\ -(s+t)^3\ln t+t^2(3s+t)\ln(s+t)\end{array}\right\}.
\end{aligned}$$

5. $E\left[Z_{i:n}(\ln Z_{i:n})Z_{j:n}\right]$ The interior integral in $A_{s,t}^{11,10}$ is

$$\begin{aligned}
&\int_{x=0}^{y}x(\ln x)e^{-sx}dx \\
&= -\frac{1}{s^2}\left\{-1+\gamma+e^{-sy}-\ln y+\ln sy+\Gamma(0,sy)+\Gamma(2,sy)\ln y\right\} \\
&= -\frac{1}{s^2}\left\{-1+\gamma+\ln s+e^{-sy}+E_1(sy)+e^{-sy}\ln y+sye^{-sy}\ln y\right\}
\end{aligned}$$

since, from (1.17),

$$\Gamma(0,sy)=E_1(sy)$$

and

$$\Gamma(2, sy) = \int_{t=sy}^{\infty} e^{-t} t dt = e^{-sy}(1+sy).$$

Then, integrating wrt y requires (neglect the constants)

$$\int_{y=0}^{\infty} y e^{-(s+t)y} dy = \frac{1}{(s+t)^2},$$

(4.23), (4.24), (4.25) and (4.29) so that

$$\begin{aligned} A_{s,t}^{11,10} &= -\frac{1}{s^2}\left\{ \begin{array}{c} \frac{-1+\gamma+\ln s}{t^2} + \frac{1}{(s+t)^2} + \frac{1}{t^2}\left[\ln(1+\frac{t}{s}) - \frac{t}{s+t}\right] \\ +\frac{1-\gamma-\ln(s+t)}{(s+t)^2} + \frac{s[3-2\gamma-2\ln(s+t)]}{(s+t)^3} \end{array} \right\} \\ &= \frac{1}{s^2t^2(s+t)^3}\left\{ s^2\left[s(1-\gamma)+t(4-3\gamma)\right] + \ln(s+t)\left[t^2(3s+t)-(s+t)^3\right] \right\}. \end{aligned}$$

6. $E\left[(\ln Z_{i:n})Z_{j:n}\ln Z_{j:n}\right]$ The appropriate integral is $A_{s,t}^{01,11}$ which has same inner part to that in $A_{s,t}^{01,10}$; we thus have

$$\begin{aligned} A_{s,t}^{01,11} &= -\frac{1}{s}\left\{ \begin{array}{c} \frac{(1-\gamma-\ln t)(\gamma+\ln s)}{t^2} \\ -\frac{1}{t^2}\left[\left(\ln(1+\frac{t}{s}) - \frac{t}{s+t}\right)(\gamma+\ln(s+t)-1) + Li_2(\frac{t}{s+t}) - \frac{t}{s+t}\right] \\ +\frac{6\gamma(-2+\gamma)+\pi^2+6\ln(s+t)(-2+2\gamma+\ln(s+t))}{6(s+t)^2} \end{array} \right\} \\ &= -\frac{1}{st^2(s+t)^2}\left\{ \begin{array}{c} -(s+t)^2\left[(\ln t)(\gamma+\ln s) + Li_2(\frac{t}{s+t})\right] \\ +\gamma s(s+3t) - s(s+2t)\left[\gamma^2 + (\ln(s+t))^2\right] \\ +\ln(s+t)\left[(s+t)^2(1-\gamma+\frac{t}{s+t}+\ln s) + 2t^2(\gamma-1)\right] + \frac{t^2\pi^2}{6} \end{array} \right\} \end{aligned}$$

by using the following results (ignoring the constants wrt y):

$$\begin{aligned} &\int_{y=0}^{\infty} y(\ln y)e^{-ty}E_1(sy)dy \\ &= -\frac{1}{t^2}\left[\left(\ln(1+\frac{t}{s}) - \frac{t}{s+t}\right)(\gamma+\ln(s+t)-1) + \left(\frac{t}{s+t}\right)^2 \Phi(\frac{t}{s+t},2,2)\right] \text{ from (1.21)} \\ &= -\frac{1}{t^2}\left[\begin{array}{c} \left(\ln(1+\frac{t}{s}) - \frac{t}{s+t}\right)(\gamma+\ln(s+t)-1) \\ +Li_2(\frac{t}{s+t}) - \frac{t}{s+t} \end{array} \right] \text{ from (1.23)}, \end{aligned} \tag{4.30}$$

$$\int_{y=0}^{\infty} y(\ln y)^2 e^{-(s+t)y}dy = \frac{6\gamma(-2+\gamma)+\pi^2+6\ln(s+t)\left(-2+2\gamma+\ln(s+t)\right)}{6(s+t)^2}, \tag{4.31}$$

and (4.28).

7. $E\left[Z_{i:n}\ln Z_{i:n}\ln Z_{j:n}\right]$ The internal integral of $A_{s,t}^{11,01}$ is studied before in $A_{s,t}^{11,10}$; by dropping the constants, the elements of external integration have already been given in

(4.21), (4.22), (4.26), (4.27) and (4.31) respectively. On rearranging, $A_{s,t}^{11,01}$ is

$$-\frac{1}{s^2}\left\{\begin{array}{c} -\frac{(\gamma+\ln t)(-1+\gamma+\ln s)}{t}-\frac{\gamma+\ln(s+t)}{s+t} \\ -\frac{1}{t}\left[\ln(\frac{s+t}{s})\left(\gamma+\ln(s+t)\right)+Li_2(\frac{t}{s+t})\right]+\frac{1}{s+t}\left[\frac{\pi^2}{6}+\left(\gamma+\ln(s+t)\right)^2\right] \\ +\frac{s\left[6\gamma(-2+\gamma)+\pi^2+6\ln(s+t)(-2+2\gamma+\ln(s+t))\right]}{6(s+t)^2} \end{array}\right\}$$

$$= -\frac{1}{s^2 t(s+t)^2}\left\{\begin{array}{c} -(s+t)^2\left[(\ln t)(\gamma-1+\ln s)+Li_2(\frac{t}{s+t})\right] \\ -\gamma s\left[-s+t\right]-s^2\left[\gamma^2+(\ln(s+t))^2\right] \\ +\ln(s+t)\left[(s+t)^2\ln s-3st-t^2\right] \\ +\gamma\ln(s+t)(-s^2+2st+t^2)+\frac{\pi^2}{6}t\,(2s+t) \end{array}\right\}.$$

8. $E\left[Z_{i:n}(\ln Z_{i:n})Z_{j:n}(\ln Z_{j:n})\right]$ $A_{s,t}^{11,11}$ has an identical interior part as $A_{s,t}^{11,10}$ but exterior part similar to $A_{s,t}^{10,11}$ and $A_{s,t}^{01,11}$. Accordingly, integration and simplification give $A_{s,t}^{11,11}$ as

$$-\frac{1}{s^2}\left\{\begin{array}{c} \frac{(1-\gamma-\ln t)(-1+\gamma+\ln s)}{t^2}+\frac{1-\gamma-\ln(s+t)}{(s+t)^2} \\ -\frac{1}{t^2}\left[\left(\ln(1+\frac{t}{s})-\frac{t}{s+t}\right)\left(\gamma+\ln(s+t)-1\right)+Li_2(\frac{t}{s+t})-\frac{t}{s+t}\right] \\ +\frac{6\gamma(-2+\gamma)+\pi^2+6\ln(s+t)(-2+2\gamma+\ln(s+t))}{6(s+t)^2} \\ +\frac{s\left[6+6\gamma(-3+\gamma)+\pi^2+6\ln(s+t)(-3+2\gamma+\ln(s+t))\right]}{3(s+t)^3} \end{array}\right\}$$

$$= -\frac{1}{s^2t^2(s+t)^3}\left\{\begin{array}{c} -(s+t)^3\left[(\ln t)(\gamma-1+\ln s)+Li_2(\frac{t}{s+t})\right] \\ -\gamma s\left(-2s^2-7st+t^2\right)-s^2(s+3t)\left[1+\gamma^2+(\ln(s+t))^2\right] \\ +\ln(s+t)\left[-6st^2-3t^2(s+t)+(s+t)^3+\left(\frac{t}{s+t}+\ln s\right)(s+t)^3\right] \\ +\gamma\ln(s+t)\left[-(s-t)(s^2+4st+t^2)\right]+\frac{\pi^2}{6}t^2\,(3s+t) \end{array}\right\}$$

based on (ignore the constants wrt y)

$$\int_{y=0}^{\infty} y^2(\ln y)^2 e^{-(s+t)y}dy=\frac{6+6\gamma(-3+\gamma)+\pi^2+6\ln(s+t)\left(-3+2\gamma+\ln(s+t)\right)}{3(s+t)^3},$$

and results in (4.23), (4.28), (4.30) and (4.31).

Therefore, we can now use the above results to write down the expectations in (4.19); for instance, using (4.20), (4.19a) is given by

$$\begin{aligned} E\left[Z_{i:n}\ln Z_{j:n}\right] &= c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}A_{i+l-k,n-i-l}^{10,01} \\ &= c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(i+l-k)^2(n-i-l)(n-k)^2}\times \\ &\quad \left\{\begin{array}{c} -(i+l-k)\left[\gamma(i+l-k)+n-i-l\right] \\ -(n-k)^2\ln(n-i-l) \\ +(n-i-l)(n+i-2k+l)\ln(n-k) \end{array}\right\}. \end{aligned} \tag{4.32}$$

Expressions for other expectations can similarly be written down from (4.20); we have included these expressions in Appendix D for ease of reading.

Basic Expectation

Before we move on to consider the derivatives method, it will prove useful to here consider in some detail $E\left[Z_{i:n}^{p}Z_{j:n}^{q}\right]$, which may be regarded as the preliminary result to achieve later expectations of the form $E\left[Z_{i:n}^{p}(\ln Z_{i:n})^{a}Z_{j:n}^{q}(\ln Z_{j:n})^{b}\right]$, using repetitive differentiations on $E\left[Z_{i:n}^{p}Z_{j:n}^{q}\right]$.

An approach due to John & Watkins (2006) Similarly, $E\left[Z_{i:n}^{p}Z_{j:n}^{q}\right]$ is given by

$$c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}A_{i+l-k,n-i-l}^{p0,q0}.$$

We refer to John & Watkins (2006) to proceed; in general, the inner integration in $A_{s,t}^{p0,q0}$ wrt x is given by

$$\int_{x=0}^{y}x^{p}e^{-sx}dx=\frac{1}{s^{p+1}}\int_{u=0}^{sy}u^{p}e^{-u}du=\frac{1}{s^{p+1}}\gamma(p+1,sy),$$

obtained by letting $u=sx$, so $0\leq x\leq y\Leftrightarrow 0\leq u\leq sy$, and $x=u/s$ so $dx=du/s$. Using (1.8), the normalised incomplete gamma function may be expressed as

$$\gamma(p+1,sy)=(sy)^{p+1}\sum_{m=0}^{\infty}\frac{(-sy)^{m}}{m!(p+1+m)}$$

so that

$$\begin{aligned}A_{s,t}^{p0,q0} &= \int_{y=0}^{\infty}y^{q}e^{-ty}y^{p+1}\sum_{m=0}^{\infty}\frac{(-sy)^{m}}{m!(p+1+m)}dy\\ &= \int_{y=0}^{\infty}\left[\sum_{m=0}^{\infty}\frac{(-s)^{m}y^{p+q+1+m}e^{-ty}}{m!(p+1+m)}\right]dy\\ &= \sum_{m=0}^{\infty}\left[\int_{y=0}^{\infty}\frac{(-s)^{m}y^{p+q+1+m}e^{-ty}}{m!(p+1+m)}dy\right]\end{aligned}$$

on reversing the order of integration and summation. We thus have

$$\begin{aligned}A_{s,t}^{p0,q0} &= \sum_{m=0}^{\infty}\left[\frac{(-s)^{m}}{m!(p+1+m)}\int_{y=0}^{\infty}y^{p+q+1+m}e^{-ty}dy\right]\\ &= \sum_{m=0}^{\infty}\left[\frac{(-s)^{m}}{m!(p+1+m)}\times\frac{\Gamma(p+q+2+m)}{t^{p+q+2+m}}\right]\end{aligned}$$

from the definition of gamma function. Hence,

$$A_{s,t}^{p0,q0} = \frac{1}{t^{p+q+2}} \sum_{m=0}^{\infty} \frac{(-\frac{s}{t})^m \Gamma(p+q+2+m)}{m!(p+1+m)}.$$

We also introduce a hypergeometric function, writing the summation as

$$\begin{aligned}
&\sum_{m=0}^{\infty} \left[\frac{\Gamma(p+1+m)\Gamma(p+q+2+m)}{\Gamma(p+2+m)} \times \frac{(-\frac{s}{t})^m}{m!} \right] \\
= \; & \frac{\Gamma(p+1)\Gamma(p+q+2)}{\Gamma(p+2)} F_{2,1}\left(p+1, p+q+2; p+2; -\frac{s}{t}\right) \\
= \; & \frac{\Gamma(p+q+2)}{p+1} F_{2,1}\left(p+1, p+q+2; p+2; -\frac{s}{t}\right).
\end{aligned}$$

As a result, we have

$$A_{s,t}^{p0,q0} = \frac{\Gamma(p+q+2)}{t^{p+q+2}(p+1)} F_{2,1}\left(p+1, p+q+2; p+2; -\frac{s}{t}\right).$$

Finally, we obtain an expression for $E\left[Z_{i:n}^p Z_{j:n}^q\right]$ in terms of hypergeometric functions as follows:

$$\frac{c_{i,j:n}\Gamma(p+q+2)}{(p+1)} \sum_{k=0}^{i-1} \sum_{l=0}^{j-i-1} \left[\begin{array}{c} \binom{i-1}{k}\binom{j-i-1}{l} \frac{(-1)^{j-k-l}}{(n-i-l)^{p+q+2}} \\ \times F_{2,1}\left(p+1, p+q+2; p+2; -\frac{i+l-k}{n-i-l}\right) \end{array} \right]; \tag{4.33}$$

for instance, when $p = q = 1$, this equation reduces to (4.11) as shown in John & Watkins (2006).

Convergence considerations Here, we consider the conditions under which the functions $F_{2,1}\left(p+1, p+q+2; p+2; -\frac{i+l-k}{n-i-l}\right)$ in (4.33) are convergent for $1 \le i < j \le n$. To illustrate this, we take $n = 6$, $i = 2$, $j = 4$, where (4.33) contains

$$\begin{aligned}
&F_{2,1}\left(p+1, p+q+2; p+2; -\frac{1}{2}\right), \quad F_{2,1}\left(p+1, p+q+2; p+2; -1\right), \\
&F_{2,1}\left(p+1, p+q+2; p+2; -\frac{1}{4}\right), \quad F_{2,1}\left(p+1, p+q+2; p+2; -\frac{2}{3}\right).
\end{aligned}$$

From (1.14), the condition for convergence requires

$$\left| -\frac{i+l-k}{n-i-l} \right| < 1,$$

meaning the second $F_{2,1}$ function in the above example is divergent. In general case, we thus need $i + l - k < n - i - l \Rightarrow 2i + 2l - k < n$. In particular, the term $2i + 2l - k$ is at its maximum when

$$2i + 2\max\{l\} - \min\{k\} = 2i + 2(j-i-1) - 0 = 2j-2$$

so the condition reduces to

$$\begin{aligned} 2j-2 &< n \\ j &< \frac{n}{2}+1. \end{aligned}$$

Therefore, $F_{2,1}\left(p+1, p+q+2; p+2; -\frac{i+l-k}{n-i-l}\right)$ is only convergent when $1 \le i < j < \frac{n}{2}+1$; more precisely, this is when

$$1 \le i < j \le \left\lfloor \frac{n}{2}+1 \right\rfloor \text{ if } n \text{ is odd,}$$

and

$$1 \le i < j \le \frac{n}{2} \text{ if } n \text{ is even.}$$

An alternative form To overcome the problem, we exploit (1.13) to express

$$F_{2,1}\left(p+1, p+q+2; p+2; -\frac{i+l-k}{n-i-l}\right)$$

as

$$\left(\frac{n-k}{n-i-l}\right)^{-(p+q+2)} F_{2,1}\left(p+q+2, 1; p+2; \frac{i+l-k}{n-k}\right).$$

It is easy to see that $\left|\frac{i+l-k}{n-k}\right|$ is strictly less than 1: we write

$$-n+k < i+l-k \Rightarrow 2k-l-i < n$$

where $\max\{2k-l-i\} = i$ is indeed $< n$, and,

$$i+l-k < n-k \Rightarrow i+l < n$$

where $\max\{i+l\} = j-1$ is also $< n$. Hence, the $F_{2,1}\left(p+q+2, 1; p+2; \frac{i+l-k}{n-k}\right)$ series is now absolutely convergent for the whole range of $1 \le i < j \le n$, and we have successfully rewritten $E\left[Z_{i:n}^p Z_{j:n}^q\right]$ in terms of the simpler hypergeometric functions;

$$\frac{c_{i,j:n}\Gamma(p+q+2)}{(p+1)} \sum_{k=0}^{i-1} \sum_{l=0}^{j-i-1} \left[\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(n-k)^{p+q+2}} F_{2,1}\left(p+q+2, 1; p+2; \frac{i+l-k}{n-k}\right)\right]. \tag{4.34}$$

Moreover, we will see that, in the derivatives method, the partial derivatives of $F_{2,1}\left(\frac{i+l-k}{n-k}\right)$ can be greatly simplified when p, q take values of 0 and 1, to give the specific expectations in (4.19).

Specific expectation	Partial derivative needed	p	q
$E[Z_{i:n} \ln Z_{j:n}]$	$E'_q[Z^p_{i:n} Z^q_{j:n}]$	1	0
$E[(\ln Z_{i:n}) Z_{j:n}]$	$E'_p[Z^p_{i:n} Z^q_{j:n}]$	0	1
$E[\ln Z_{i:n} \ln Z_{j:n}]$	$E''_{qp}[Z^p_{i:n} Z^q_{j:n}]$	0	0
$E[Z_{i:n} Z_{j:n} \ln Z_{j:n}]$	$E'_q[Z^p_{i:n} Z^q_{j:n}]$	1	1
$E[Z_{i:n} (\ln Z_{i:n}) Z_{j:n}]$	$E'_p[Z^p_{i:n} Z^q_{j:n}]$	1	1
$E[(\ln Z_{i:n}) Z_{j:n} \ln Z_{j:n}]$	$E''_{qp}[Z^p_{i:n} Z^q_{j:n}]$	0	1
$E[Z_{i:n} \ln Z_{i:n} \ln Z_{j:n}]$	$E''_{qp}[Z^p_{i:n} Z^q_{j:n}]$	1	0
$E[Z_{i:n} (\ln Z_{i:n}) Z_{j:n} (\ln Z_{j:n})]$	$E''_{qp}[Z^p_{i:n} Z^q_{j:n}]$	1	1

Table 4.2: Derivatives method: expectations in (4.18) and the partial derivatives needed.

Derivatives Method

As we have previously noted at Section 4.2.4.2, we are now in the position to differentiate the basic expectation, $E[Z^p_{i:n} Z^q_{j:n}]$ given at (4.34), partially wrt p and/or q, and then suitably replace p, q for $0, 1$ to obtain expectations of the form $E\left[Z^p_{i:n}(\ln Z_{i:n})^a Z^q_{j:n}(\ln Z_{j:n})^b\right]$; Table 4.2 lists the partial derivatives needed for each of the expectations in (4.19). In summary, we require $E'_p[Z^p_{i:n} Z^q_{j:n}]$, $E'_q[Z^p_{i:n} Z^q_{j:n}]$ and $E''_{qp}[Z^p_{i:n} Z^q_{j:n}]$, which contain the partial derivatives of $\Gamma(p+q+2)$ and $F_{2,1}\left(p+q+2, 1; p+2; \frac{i+l-k}{n-k}\right)$.

The partial derivatives of $\Gamma(p+q+2)$ are straightforward to obtain; we have, from (1.2),

$$\Gamma'_p(p+q+2) = \Gamma'_q(p+q+2) = \Gamma(p+q+2)\,\psi(p+q+2),$$

and, from (1.3),

$$\Gamma''_{qp}(p+q+2) = \Gamma(p+q+2)\left[\psi'(p+q+2) + \{\psi(p+q+2)\}^2\right].$$

Furthermore, when p and q are replaced by 0 or 1, these derivatives simplify to the values in Table 1.6. As a result, it is sensible to express $F_{2,1}\left(p+q+2, 1; p+2; \frac{i+l-k}{n-k}\right)$ in terms of gamma functions for which partial derivatives are easy to obtain; we refer to (1.12) to write

$$\begin{aligned} F_{2,1}(p+q+2, 1; p+2; z) &= \sum_{m=0}^{\infty} \frac{(p+q+2)_m (1)_m}{(p+2)_m} \times \frac{z^m}{m!} \\ &= \sum_{m=0}^{\infty} \frac{\Gamma(p+2)\,\Gamma(p+q+2+m)}{\Gamma(p+q+2)\,\Gamma(p+2+m)} \times z^m \end{aligned}$$

in which $z = \frac{i+l-k}{n-k}$, and it follows that (4.34) becomes

$$E\left[Z^p_{i:n} Z^q_{j:n}\right] = c_{i,j:n} \sum_{k=0}^{i-1} \sum_{l=0}^{j-i-1} \frac{(-1)^{j-k-l} \binom{i-1}{k} \binom{j-i-1}{l}}{(n-k)^{p+q+2}} \sum_{m=0}^{\infty} \frac{\Gamma(p+1)\,\Gamma(p+q+2+m)}{\Gamma(p+2+m)} \times z^m, \tag{4.35}$$

comprising of only gamma functions.

First partial derivatives of $E[Z^p_{i:n}Z^q_{j:n}]$ Consequently, the first derivative of (4.35) wrt p is

$$\begin{aligned}
E'_p\left[Z^p_{i:n}Z^q_{j:n}\right] &= E[Z^p_{i:n}(\ln Z_{i:n})Z^q_{j:n}] \\
&= c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(n-k)^{p+q+2}}\sum_{m=0}^{\infty}\frac{\Gamma(p+1)\Gamma(p+q+2+m)}{\Gamma(p+2+m)}\times z^m \\
&\quad\times\{-\ln(n-k)+\psi(p+1)+\psi(p+q+2+m)-\psi(p+2+m)\} \qquad (4.36)
\end{aligned}$$

and wrt q is

$$\begin{aligned}
E'_q\left[Z^p_{i:n}Z^q_{j:n}\right] &= E[Z^p_{i:n}Z^q_{j:n}\ln Z_{j:n}] \\
&= c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(n-k)^{p+q+2}}\sum_{m=0}^{\infty}\frac{\Gamma(p+1)\Gamma(p+q+2+m)}{\Gamma(p+2+m)}\times z^m \\
&\quad\times\{-\ln(n-k)+\psi(p+q+2+m)\}. \qquad (4.37)
\end{aligned}$$

Second partial derivatives of $E[Z^p_{i:n}Z^q_{j:n}]$ Then, second differentiation of (4.37) wrt p yields

$$E''_{qp}[Z^p_{i:n}Z^q_{j:n}] = E[Z^p_{i:n}(\ln Z_{i:n})Z^q_{j:n}(\ln Z_{j:n})]$$

as

$$\begin{aligned}
&c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(n-k)^{p+q+2}}\sum_{m=0}^{\infty}\frac{\Gamma(p+1)\Gamma(p+q+2+m)}{\Gamma(p+2+m)}\times z^m \\
&\times\begin{bmatrix}-\ln(n-k)\\ +\psi(p+q+2+m)\end{bmatrix}\begin{Bmatrix}-\ln(n-k)+\psi(p+1)+\psi(p+q+2+m)\\ +\frac{\psi'(p+q+2+m)}{-\ln(n-k)+\psi(p+q+2+m)}-\psi(p+2+m)\end{Bmatrix} \qquad (4.38)
\end{aligned}$$

Expectations in (4.19) Now we can obtain specific expressions for (4.19) by suitably replacing p and q, as summarised in Table 4.2. Firstly, setting $p=0, q=1$ and $p=q=1$ in (4.36) give

$$\begin{aligned}
E[(\ln Z_{i:n})Z_{j:n}] &= c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(n-k)^3} \\
&\quad\times\left\{\sum_{m=0}^{\infty}z^m(2+m)\left[-\ln(n-k)-\gamma+(2+m)^{-1}\right]\right\}
\end{aligned}$$

and

$$\begin{aligned}
E[Z_{i:n}(\ln Z_{i:n})Z_{j:n}] &= c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(n-k)^4} \\
&\quad\times\left\{\sum_{m=0}^{\infty}z^m(3+m)\left[-\ln(n-k)+1-\gamma+(3+m)^{-1}\right]\right\}
\end{aligned}$$

respectively. Similarly, setting $p = 1, q = 0$ and $p = q = 1$ in (4.37) give

$$E\left[Z_{i:n}\ln Z_{j:n}\right] = c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(n-k)^3}\left\{\sum_{m=0}^{\infty} z^m\left[-\ln(n-k)+\psi(3+m)\right]\right\} \tag{4.39}$$

and

$$\begin{aligned} E\left[Z_{i:n}Z_{j:n}\ln Z_{j:n}\right] &= c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(n-k)^4} \\ &\quad\times\left\{\sum_{m=0}^{\infty} z^m(3+m)\left[-\ln(n-k)+\psi(4+m)\right]\right\} \end{aligned}$$

respectively, while the remaining expectations can be obtained by setting $p = q = 0$, $p = 0, q = 1$, $p = 1, q = 0$, and $p = q = 1$ in (4.38) in each case, as given below:

$$\begin{aligned} E\left[\ln Z_{i:n}\ln Z_{j:n}\right] &= c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(n-k)^2}\sum_{m=0}^{\infty} z^m\left[-\ln(n-k)+\psi(2+m)\right] \\ &\quad\times\left\{-\ln(n-k)-\gamma+\frac{\psi'(2+m)}{-\ln(n-k)+\psi(2+m)}\right\}, \end{aligned}$$

$$\begin{aligned} E\left[(\ln Z_{i:n})Z_{j:n}\ln Z_{j:n}\right] &= c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(n-k)^3} \\ &\quad\times\sum_{m=0}^{\infty} z^m(2+m)\left[-\ln(n-k)+\psi(3+m)\right] \\ &\quad\times\left\{-\ln(n-k)-\gamma+(2+m)^{-1}+\frac{\psi'(3+m)}{-\ln(n-k)+\psi(3+m)}\right\}, \end{aligned}$$

$$\begin{aligned} E\left[Z_{i:n}\ln Z_{i:n}\ln Z_{j:n}\right] &= c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(n-k)^3}\sum_{m=0}^{\infty} z^m\left[-\ln(n-k)+\psi(3+m)\right] \\ &\quad\times\left\{-\ln(n-k)+1-\gamma+\frac{\psi'(3+m)}{-\ln(n-k)+\psi(3+m)}\right\}, \end{aligned}$$

and the final expectation $E\left[Z_{i:n}(\ln Z_{i:n})Z_{j:n}(\ln Z_{j:n})\right]$ is given by

$$\begin{aligned} &c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(n-k)^4}\sum_{m=0}^{\infty} z^m(3+m)\left[-\ln(n-k)+\psi(3+m)\right] \\ &\times\left\{-\ln(n-k)+1-\gamma+(3+m)^{-1}+\frac{\psi'(4+m)}{-\ln(n-k)+\psi(4+m)}\right\}. \end{aligned}$$

Unlike for single expectations, the expressions obtained here are not directly comparable to those found from the direct method, (see, for instance, (4.32) and (4.39)), so we will check this via numerical studies.

i,j	n			
	25	50	100	1000
$0.1n, 0.2n$: direct	0.0599	0.0741	0.0766	0.0791
: deriv.	0.0599	0.0741	0.0766	0.0791
: simul.	0.0598	0.0740	0.0769	0.0791
$0.3n, 0.4n$: direct	0.1106	0.1185	0.1222	0.1258
: deriv.	0.1106	0.1185	0.1222	0.1258
: simul.	0.1102	0.1183	0.1221	0.1258
$0.5n, 0.6n$: direct	0.0394	0.0307	0.0255	0.0209
: deriv.	0.0394	0.0307	0.0255	0.0209
: simul.	0.0401	0.0311	0.0254	0.0209
$0.7n, 0.8n$: direct	0.2266	0.2357	0.2042	0.1745
: deriv.	0.2266	0.2357	0.2042	0.1745
: simul.	0.2232	0.2348	0.2058	0.1742
$0.9n, 1.0n$: direct	8.5146	13.2134	16.5583	29.0923
: deriv.	8.5146	13.2134	16.5583	29.0923
: simul.	8.4176	13.2729	16.5480	29.0053

Table 4.3: Numerical comparison of $E[Z_{i:ni:n} Z_{j:n} \ln Z_{j:n}]$ for various i, j and n.

Some Numerical Details and Discussion

In this section, we validate the theoretical expressions using simulation experiments with 10^4 replications. We take $n = 10$, which yields $(10 + 11)/2 = 55$ distinct combinations of (i, j) with $1 \leq i < j \leq n$, and follow the graphical display at John & Watkins (2006); Figures 4.7 to 4.14 summarise the agreement between theoretical, obtained from both direct and derivatives methods, and simulated values for each expectation in (4.19) in turn, where we see excellent agreement between the three sets of values. Table 4.3 further presents such agreement for $E[Z_{i:n}(\ln Z_{i:n}) Z_{j:n}(\ln Z_{j:n})]$ for various i, j and n, in which the two theoretical evaluations (upmost and middle entries) are exactly equal and are often consistent with their simulated counterparts (lowest entries) to 3 decimal places. There is scope to check the theoretical results for larger sample sizes, in which case the computation time can increase considerably; for instance, Mathematica took over 7 days to evaluate $E[Z_{i:n}(\ln Z_{i:n}) Z_{j:n}(\ln Z_{j:n})]$ from the derivatives method when $i = 250, j = 500$ and $n = 2500$. It is obvious that the sample size reflects the number of calculations required, and hence the computational burden even for simple expectations is at best proportional to n. We have obtained accurate results for $n \leq 1000$, and noted that the direct method is more time-efficient than the derivatives approach, but we can expect the computation time to reduce given the advancement in computational capabilities available today.

4.3 Burr Order Statistics

In this section $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ represent the order statistics from a random sample of size n drawn from the Burr distribution. In order to determine the correlations

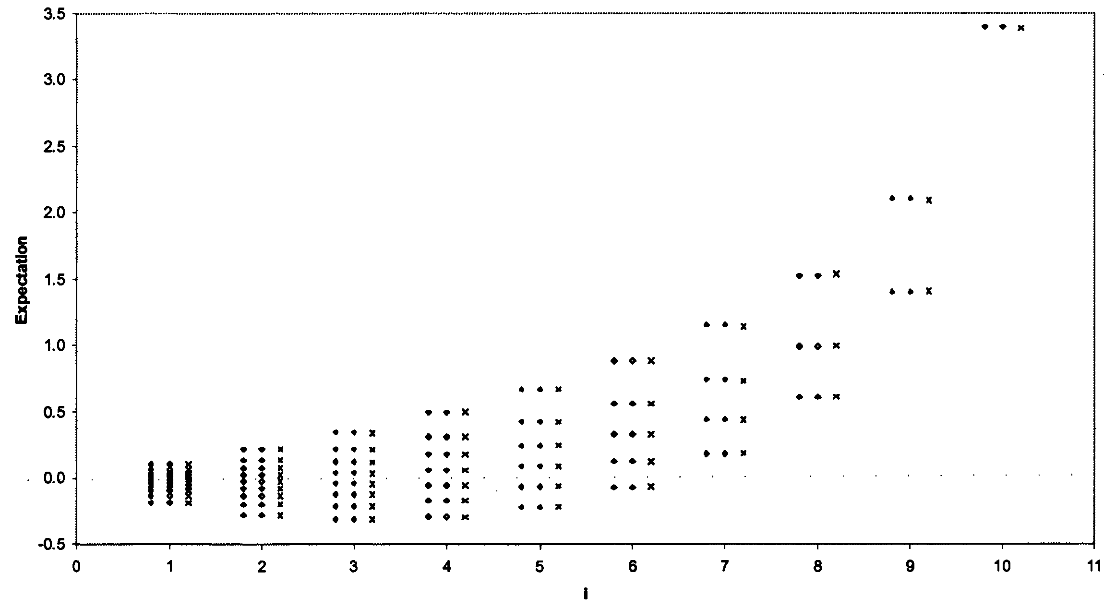

Figure 4.7: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E\left[Z_{i:n}\ln Z_{j:n}\right]$ for all $1 \leq i < j \leq n$, for $n = 10$.

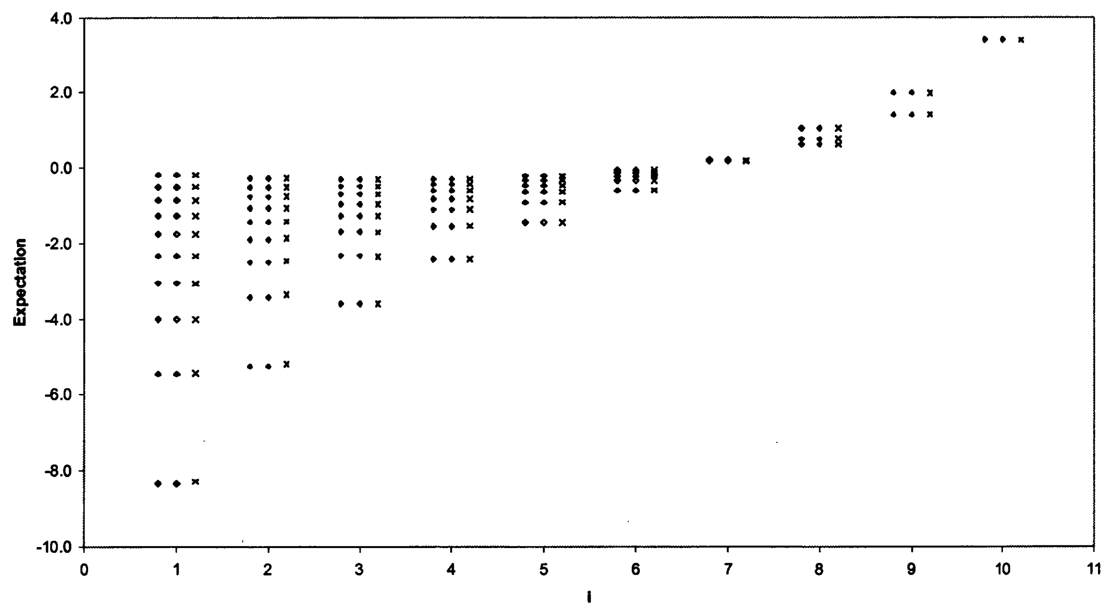

Figure 4.8: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E\left[(\ln Z_{i:n})Z_{j:n}\right]$ for all $1 \leq i < j \leq n$, for $n = 10$.

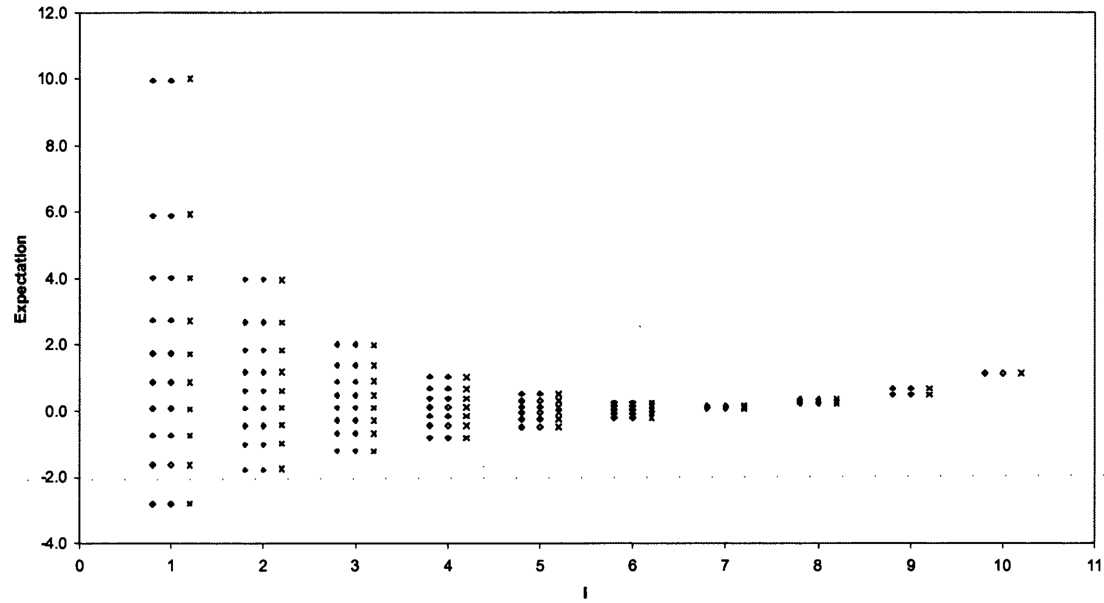

Figure 4.9: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E[\ln Z_{i:n} \ln Z_{j:n}]$ for all $1 \le i < j \le n$, for $n = 10$.

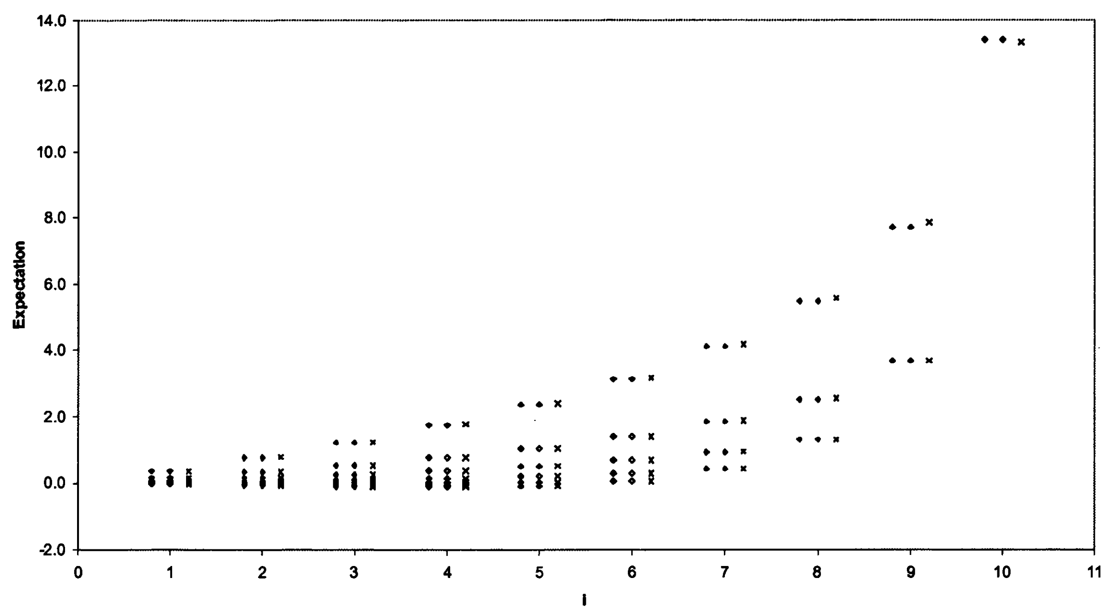

Figure 4.10: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E[Z_{i:n} Z_{j:n} \ln Z_{j:n}]$ for all $1 \le i < j \le n$, for $n = 10$.

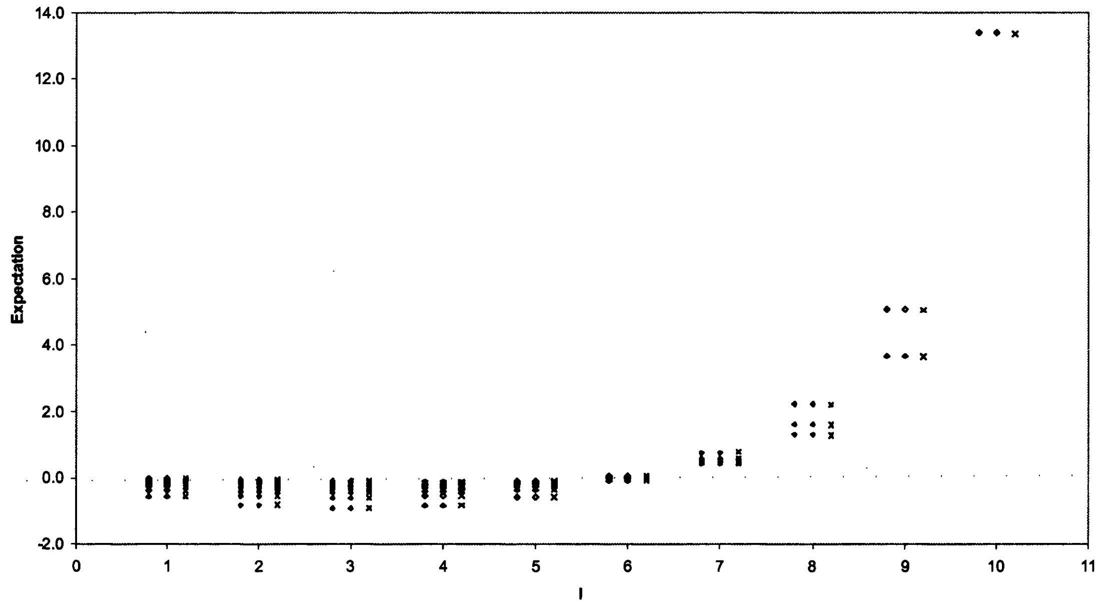

Figure 4.11: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E\left[Z_{i:n}(\ln Z_{i:n})Z_{j:n}\right]$ for all $1 \leq i < j \leq n$, for $n = 10$.

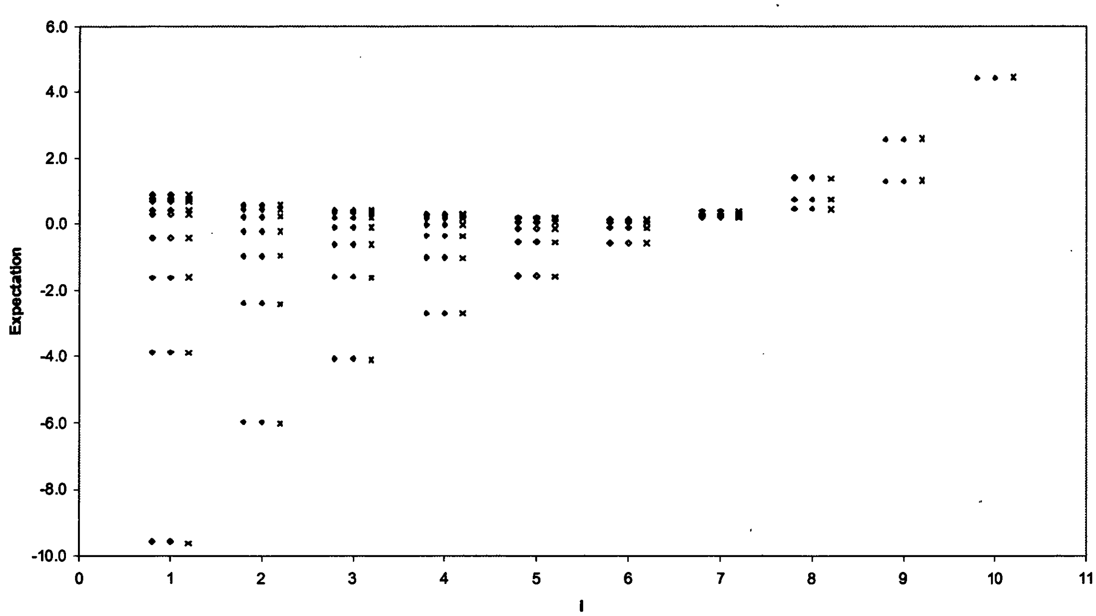

Figure 4.12: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E\left[(\ln Z_{i:n})Z_{j:n}\ln Z_{j:n}\right]$ for all $1 \leq i < j \leq n$, for $n = 10$.

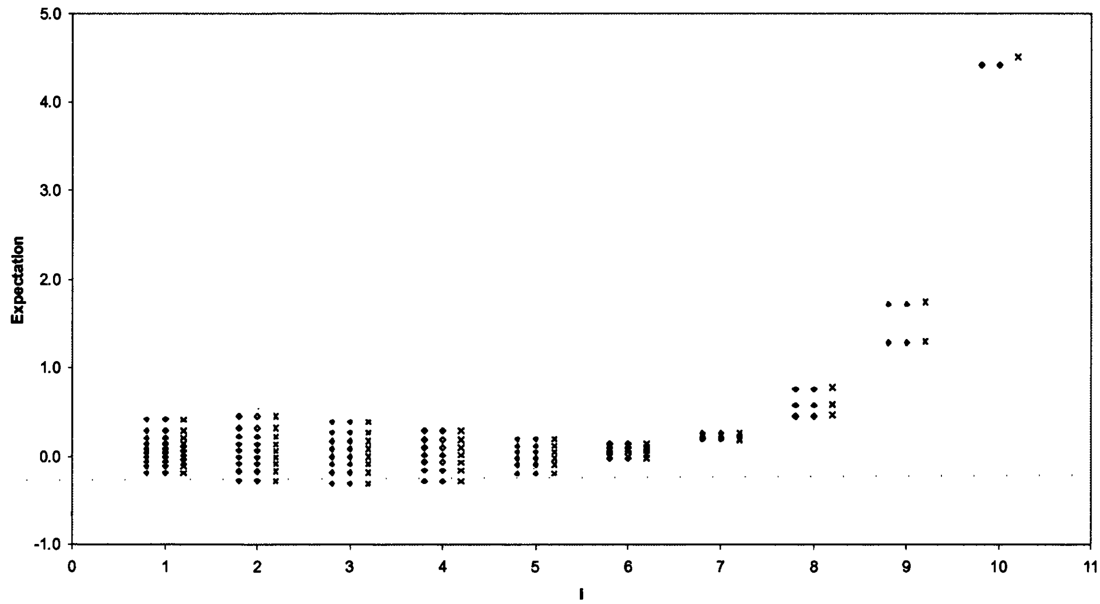

Figure 4.13: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E\left[Z_{i:n} \ln Z_{i:n} \ln Z_{j:n}\right]$ for all $1 \leq i < j \leq n$, for $n = 10$.

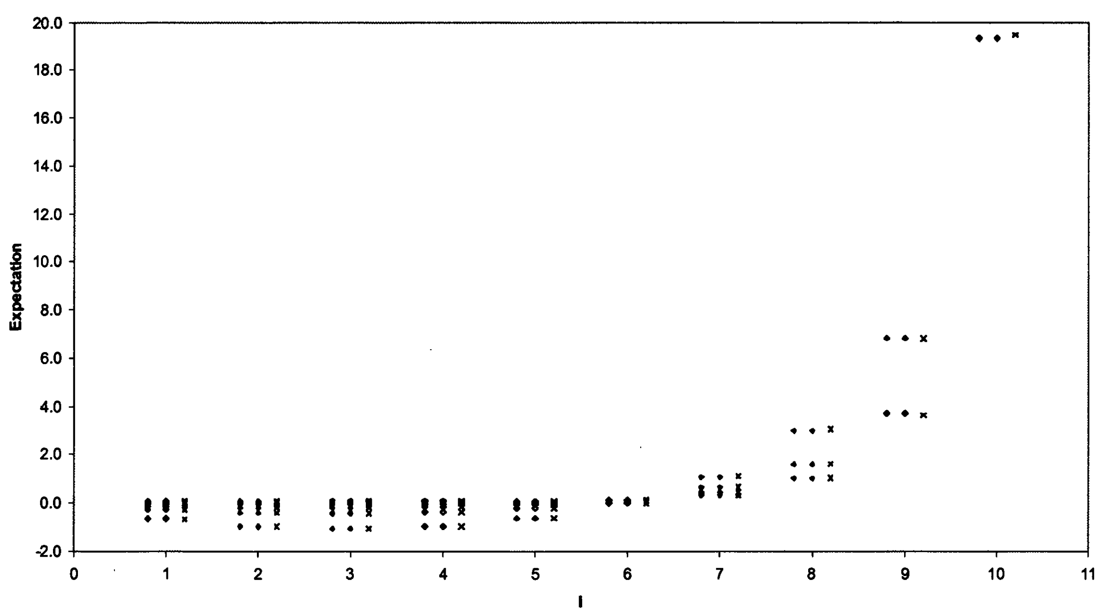

Figure 4.14: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E\left[Z_{i:n}(\ln Z_{i:n}) Z_{j:n}(\ln Z_{j:n})\right]$ for all $1 \leq i < j \leq n$, for $n = 10$.

between the final and interim Burr score functions, we will require, based on the form of (2.36), (2.37), (2.56) and (2.57), to take (4.1) as

$$E\left[\frac{X_{i:n}^{p}\left(\ln X_{i:n}\right)^{a}\left(\ln(1+X_{i:n}^{\tau})\right)^{b}}{\left(1+X_{i:n}^{\tau}\right)^{c}}\right], \tag{4.40}$$

and (4.2) as

$$E\left[\frac{X_{i:n}^{p}\left(\ln X_{i:n}\right)^{a}\left(\ln(1+X_{i:n}^{\tau})\right)^{b}}{\left(1+X_{i:n}^{\tau}\right)^{c}}\times\frac{X_{j:n}^{q}\left(\ln X_{j:n}\right)^{d}\left(\ln(1+X_{j:n}^{\tau})\right)^{e}}{\left(1+X_{j:n}^{\tau}\right)^{f}}\right], \tag{4.41}$$

for $a, b, c, d, e, f, p, q = 0, 1, 2$. Watkins (1997) considers some, but not all, expectations of these forms for the Burr distribution for the case of complete samples, while Watkins & Johnson (2002) give some corresponding discussion for samples obtained under a Type I censoring regime, and Pawles & Szynal (2001) provide recurrence relations for single and product moments of generalised order statistics from Pareto, generalised Pareto and Burr distributions. However, as for the Burr EFI matrix with Type II censored data, there appears to be no previous work about solving for expressions of the form given at (4.40) and (4.41) in terms of the Burr order statistics.

As with the Weibull case, we will look at direct and derivatives methods. We will also see that the latter is preferred to the former when deriving results for (4.41).

4.3.1 Expectations of $g\left(X_{i:n}\right)$

More specifically, we will need, for (4.40), the following expectations:

$$E\left[\left(\ln X_{i:n}\right)^{2}\right], \tag{4.42a}$$

$$E\left[\left(\ln\left(1+X_{i:n}^{\tau}\right)\right)^{2}\right], \tag{4.42b}$$

$$E\left[\ln X_{i:n}\ln\left(1+X_{i:n}^{\tau}\right)\right], \tag{4.42c}$$

$$E\left[\frac{X_{i:n}^{\tau}\ln X_{i:n}\ln(1+X_{i:n}^{\tau})}{1+X_{i:n}^{\tau}}\right], \tag{4.42d}$$

$$E\left[\frac{X_{i:n}^{\tau}\left(\ln X_{i:n}\right)^{2}}{1+X_{i:n}^{\tau}}\right], \tag{4.42e}$$

$$E\left[\left(\frac{X_{i:n}^{\tau}\ln X_{i:n}}{1+X_{i:n}^{\tau}}\right)^{2}\right]. \tag{4.42f}$$

In fact, these quantities are not entirely new to us; in Section 2.4.1 we have found explicitly expressions for

$$E[\ln X_{1:n}],\ E\left[\ln\left(1+X_{1:n}^{\tau}\right)\right],\ E\left[\frac{X_{1:n}^{\tau}\ln X_{1n}}{1+X_{1:n}^{\tau}}\right]\ \text{and}\ E\left[\frac{X_{1:n}^{\tau}(\ln X_{1:n})^{2}}{(1+X_{1:n}^{\tau})^{2}}\right],$$

benefiting from the fact that the properties and results of $X_{1:n}$ are a lot more straightforward than the other order statistics; then, expectations in terms of $X_{i:n}$ can be obtained from (1.49). However, in this section, we will be solving (4.40) in terms of $X_{i:n}$.

Direct Method

Using (1.40), the marginal pdf of $X_{i:n}$ is given by

$$\begin{aligned} f_{(i)}(x) &= c_{i:n}\alpha\tau x^{\tau-1}\left(1+x^{\tau}\right)^{-(\alpha+1)}\left[1-\left(1+x^{\tau}\right)^{-\alpha}\right]^{i-1}\left[\left(1+x^{\tau}\right)^{-\alpha}\right]^{n-i} \\ &= c_{i:n}\alpha\tau x^{\tau-1}\left(1+x^{\tau}\right)^{-\alpha(n-i+1)-1}\left[1-\left(1+x^{\tau}\right)^{-\alpha}\right]^{i-1}, \end{aligned}$$

and we can use the Binomial expansion to write

$$\left[1-\left(1+x^{\tau}\right)^{-\alpha}\right]^{i-1}=\sum_{k=0}^{i-1}(-1)^{i-1-k}\binom{i-1}{k}\left(1+x^{\tau}\right)^{-\alpha(i-1-k)}$$

so that

$$\left(1+x^{\tau}\right)^{-\alpha(n-i+1)-1}\left[1-\left(1+x^{\tau}\right)^{-\alpha}\right]^{i-1}=\sum_{k=0}^{i-1}(-1)^{i-1-k}\binom{i-1}{k}\left(1+x^{\tau}\right)^{-\alpha(n-k)-1},$$

which leads to

$$f_{(i)}(x)=c_{i:n}\alpha\tau\sum_{k=0}^{i-1}(-1)^{i-1-k}\binom{i-1}{k}x^{\tau-1}\left(1+x^{\tau}\right)^{-\alpha(n-k)-1}.$$

From this, we have

$$\begin{aligned} E\left[\frac{X_{i:n}^{p}\left(\ln X_{i:n}\right)^{a}\left(\ln(1+X_{i:n}^{\tau})\right)^{b}}{\left(1+X_{i:n}^{\tau}\right)^{c}}\right] &= \int_{0}^{\infty}\frac{x^{p}\left(\ln x\right)^{a}\left(\ln(1+x^{\tau})\right)^{b}}{\left(1+x^{\tau}\right)^{c}}f_{(i)}(x)dx \\ &= c_{i:n}\alpha\tau\sum_{k=0}^{i-1}(-1)^{i-1-k}\binom{i-1}{k}I_{c}^{pab} \qquad (4.43) \end{aligned}$$

at which we define

$$I_{c}^{pab}=\int_{0}^{\infty}x^{p+\tau-1}\left(\ln x\right)^{a}\left(\ln(1+x^{\tau})\right)^{b}\left(1+x^{\tau}\right)^{-\alpha(n-k)-c-1}dx.$$

Specific expectations	Partial derivatives needed	p	c
$E\left[(\ln X_{i:n})^2\right]$	$E''_{i,pp}$	0	0
$E\left[(\ln(1+X_{i:n}^{\tau}))^2\right]$	$E''_{i,cc}$	0	0
$E\left[\ln X_{i:n}\ln(1+X_{i:n}^{\tau})\right]$	$-E''_{i,pc}$	0	0
$E\left[\frac{X_{i:n}^{\tau}\ln X_{i:n}\ln(1+X_{i:n}^{\tau})}{1+X_{i:n}^{\tau}}\right]$	$-E''_{i,pc}$	τ	1
$E\left[\frac{X_{i:n}^{\tau}(\ln X_{i:n})^2}{1+X_{i:n}^{\tau}}\right]$	$E''_{i,pp}$	τ	1
$E\left[\left(\frac{X_{i:n}^{\tau}\ln X_{i:n}}{1+X_{i:n}^{\tau}}\right)^2\right]$	$E''_{i,pp}$	2τ	2

Table 4.4: Derivatives method: expectations in (4.41) and the partial derivatives needed.

Hence, here writing $s = \alpha(n-k)$, the expectations in (4.42) need, in turn, the following integrals:

$$\begin{aligned}
I_0^{020} &= \frac{1}{s\tau^3}\left\{\frac{\pi^2}{6} + [\gamma + \psi(s)]^2 + \psi'(s)\right\}, \\
I_0^{002} &= \frac{2}{s^3\tau}, \\
I_0^{011} &= \frac{1}{s^2\tau^2}\left\{-\gamma - \psi(s) + s\psi'(s)\right\}, \\
I_1^{\tau 11} &= \frac{1}{s^2(1+s)^2\tau^2}\left\{(1+2s)[1-\gamma-\psi(s)] + s(1+s)\psi'(s)\right\}, \\
I_1^{\tau 20} &= \frac{1}{s(1+s)\tau^3}\left\{\frac{\pi^2}{6} - 1 + [1-\gamma-\psi(s)]^2 + \psi'(s)\right\},
\end{aligned}$$

and

$$I_2^{(2\tau)20} = \frac{2}{s(1+s)(2+s)\tau^3}\left\{\frac{\pi^2}{6} - \frac{5}{4} + \left[\frac{3}{2} - \gamma - \psi(s)\right]^2 + \psi'(s)\right\}.$$

For example, using (4.43), we obtain (4.42a) as

$$\begin{aligned}
E\left[(\ln X_{i:n})^2\right] &= c_{i:n}\alpha\tau\sum_{k=0}^{i-1}(-1)^{i-1-k}\binom{i-1}{k}I_0^{020} \\
&= \frac{c_{i:n}\alpha}{\tau^2}\sum_{k=0}^{i-1}\frac{(-1)^{i-1-k}\binom{i-1}{k}}{s}\left\{\frac{\pi^2}{6} + [\gamma+\psi(s)]^2 + \psi'(s)\right\}, \qquad (4.44)
\end{aligned}$$

again, similarly for (4.42b) to (4.42f).

Derivatives Method

Basic expectation It is appropriate to here define, based on the form of (4.40), the basic expectation required in the derivatives method, given by

$$E_i = E\left[\frac{X_{i:n}^p}{(1+X_{i:n}^{\tau})^c}\right], \qquad (4.45)$$

for which the partial derivatives of (4.45) wrt p and c will respectively yield the terms $\ln X_{i:n}$ and $\ln(1+X_{i:n}^{\tau})$, leading to the expectations in (4.42); Table 4.4 summarises these partial derivatives in more details, in which we need $E''_{i,pp}$, $E''_{i,pc}$ and $E''_{i,cc}$.

As with (4.43), we can express E_i as

$$\begin{aligned} & c_{i:n}\alpha\tau \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} \int_0^{\infty} x^{p+\tau-1} (1+x^{\tau})^{-\alpha(n-k)-c-1}\, dx \\ &= c_{i:n}\alpha \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} B\left(\frac{p}{\tau}+1, \alpha(n-k)+c-\frac{p}{\tau}\right), \end{aligned}$$

obtained from the definition of beta function given in Table 1.5. Using (1.10) we can also express E_i in terms of the gamma functions; we have (as before $s=\alpha(n-k)$)

$$E_i = c_{i:n}\alpha \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} \frac{\Gamma\left(\frac{p}{\tau}+1\right)\Gamma\left(s+c-\frac{p}{\tau}\right)}{\Gamma(s+c+1)} \tag{4.46}$$

so that this introduces various digamma and polygamma functions.

First partial derivatives of E_i The first partial derivative of (4.46) wrt p gives

$$\begin{aligned} E'_{i,p} &= E\left[\frac{X_{i:n}^{p} \ln X_{i:n}}{(1+X_{i:n}^{\tau})^{c}}\right] \\ &= \frac{c_{i:n}\alpha}{\tau} \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} \frac{\Gamma\left(\frac{p}{\tau}+1\right)\Gamma\left(s+c-\frac{p}{\tau}\right)}{\Gamma(s+c+1)} \\ &\quad \times \left\{\psi\left(\frac{p}{\tau}+1\right) - \psi\left(s+c-\frac{p}{\tau}\right)\right\} \end{aligned} \tag{4.47}$$

and wrt c gives

$$\begin{aligned} E'_{i,c} &= E\left[-\frac{X_{i:n}^{p} \ln(1+X_{i:n}^{\tau})}{(1+X_{i:n}^{\tau})^{c}}\right] \\ &= c_{i:n}\alpha \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} \frac{\Gamma\left(\frac{p}{\tau}+1\right)\Gamma\left(s+c-\frac{p}{\tau}\right)}{\Gamma(s+c+1)} \\ &\quad \times \left\{\psi\left(s+c-\frac{p}{\tau}\right) - \psi(s+c+1)\right\}. \end{aligned} \tag{4.48}$$

Second partial derivatives of E_i It follows that the second differentiation of (4.47) wrt p yields

$$E''_{i,pp} = E\left[\frac{X_{i:n}^{p} (\ln X_{i:n})^2}{(1+X_{i:n}^{\tau})^{c}}\right]$$

as

$$\begin{aligned} & \frac{c_{i:n}\alpha}{\tau^2} \sum_{k=0}^{i-1} (-1)^{i-1-k} \binom{i-1}{k} \frac{\Gamma\left(\frac{p}{\tau}+1\right)\Gamma\left(s+c-\frac{p}{\tau}\right)}{\Gamma(s+c+1)} \\ & \times \left\{\left[\psi\left(\frac{p}{\tau}+1\right) - \psi\left(s+c-\frac{p}{\tau}\right)\right]^2 + \psi'\left(\frac{p}{\tau}+1\right) + \psi'\left(s+c-\frac{p}{\tau}\right)\right\} \end{aligned} \tag{4.49}$$

while wrt c yields

$$E''_{i,pc} = E\left[-\frac{X_{i:n}^{p}\ln X_{i:n}\ln\left(1+X_{i:n}^{\tau}\right)}{\left(1+X_{i:n}^{\tau}\right)^{c}}\right]$$

as

$$\frac{c_{i:n}\alpha}{\tau}\sum_{k=0}^{i-1}(-1)^{i-1-k}\binom{i-1}{k}\frac{\Gamma\left(\frac{p}{\tau}+1\right)\Gamma\left(s+c-\frac{p}{\tau}\right)\left[\psi\left(\frac{p}{\tau}+1\right)-\psi\left(s+c-\frac{p}{\tau}\right)\right]}{\Gamma\left(s+c+1\right)}$$
$$\times\left\{\psi\left(s+c-\frac{p}{\tau}\right)-\frac{\psi'\left(s+c-\frac{p}{\tau}\right)}{\psi\left(\frac{p}{\tau}+1\right)-\psi\left(s+c-\frac{p}{\tau}\right)}-\psi\left(s+c+1\right)\right\}. \tag{4.50}$$

Similarly, the second differentiation of (4.48) wrt c gives

$$E''_{i,cc} = E\left[\frac{X_{i:n}^{p}\left(\ln\left(1+X_{i:n}^{\tau}\right)\right)^{2}}{\left(1+X_{i:n}^{\tau}\right)^{c}}\right]$$

as

$$c_{i:n}\alpha\sum_{k=0}^{i-1}(-1)^{i-1-k}\binom{i-1}{k}\frac{\Gamma\left(\frac{p}{\tau}+1\right)\Gamma\left(s+c-\frac{p}{\tau}\right)\left[\psi\left(s+c-\frac{p}{\tau}\right)-\psi\left(s+c+1\right)\right]}{\Gamma\left(s+c+1\right)}$$
$$\times\left\{\psi\left(s+c-\frac{p}{\tau}\right)+\frac{\psi'\left(s+c-\frac{p}{\tau}\right)-\psi'\left(s+c+1\right)}{\psi\left(s+c-\frac{p}{\tau}\right)-\psi\left(s+c+1\right)}-\psi\left(s+c+1\right)\right\}. \tag{4.51}$$

It should be noted that, despite of being lengthy, these results simplify greatly when p and c take the values of 0 and 1, as shown below.

Expectations in (4.42) We can now suitably replace p and c (see Table 4.4) in the second derivatives of E_i to obtain expressions for the expectations in (4.42). For example, letting $p=c=0$, $p=\tau, c=1$, and $p=2\tau, c=2$ in (4.49) yields, respectively,

$$E\left[\left(\ln X_{i:n}\right)^{2}\right] = \frac{c_{i:n}\alpha}{\tau^{2}}\sum_{k=0}^{i-1}\frac{(-1)^{i-1-k}\binom{i-1}{k}}{s}\left\{\frac{\pi^{2}}{6}+\left[\gamma+\psi\left(s\right)\right]^{2}+\psi'\left(s\right)\right\}, \tag{4.52}$$

$$E\left[\frac{X_{i:n}^{\tau}\left(\ln X_{i:n}\right)^{2}}{1+X_{i:n}^{\tau}}\right] = \frac{c_{i:n}\alpha}{\tau^{2}}\sum_{k=0}^{i-1}\frac{(-1)^{i-1-k}\binom{i-1}{k}}{s\left(1+s\right)}\left\{\frac{\pi^{2}}{6}-1+\left[1-\gamma-\psi\left(s\right)\right]^{2}+\psi'\left(s\right)\right\},$$

and

$$E\left[\left(\frac{X_{i:n}^{\tau}\ln X_{i:n}}{1+X_{i:n}^{\tau}}\right)^{2}\right] = \frac{2c_{i:n}\alpha}{\tau^{2}}\sum_{k=0}^{i-1}\frac{(-1)^{i-1-k}\binom{i-1}{k}}{s\left(1+s\right)\left(2+s\right)}\left\{\frac{\pi^{2}}{6}-\frac{5}{4}+\left[\frac{3}{2}-\gamma-\psi\left(s\right)\right]^{2}+\psi'\left(s\right)\right\}.$$

Likewise, setting $p=c=0$ and $p=\tau, c=1$ in (4.50) gives, respectively,

$$E\left[\ln X_{i:n}\ln\left(1+X_{i:n}^{\tau}\right)\right] = \frac{c_{i:n}\alpha}{\tau}\sum_{k=0}^{i-1}\frac{(-1)^{i-1-k}\binom{i-1}{k}}{s^{2}}\left\{-\gamma-\psi\left(s\right)+s\psi'\left(s\right)\right\}$$

and

$$E\left[\frac{X_{i:n}^{\tau}\ln X_{i:n}\ln(1+X_{i:n}^{\tau})}{1+X_{i:n}^{\tau}}\right]=\frac{c_{i:n}\alpha}{\tau}\sum_{k=0}^{i-1}\frac{(-1)^{i-1-k}\binom{i-1}{k}}{s^2\left(1+s\right)^2}\left\{(1+2s)\left[1-\gamma-\psi\left(s\right)\right]+s\left(1+s\right)\psi'\left(s\right)\right\}.$$

While setting $p=c=0$ in (4.51) gives

$$E\left[\left(\ln\left(1+X_{i:n}^{\tau}\right)\right)^2\right]=2\alpha c_{i:n}\sum_{k=0}^{i-1}\frac{(-1)^{i-1-k}\binom{i-1}{k}}{s^3}.$$

We see that all expectations derived here are identical to those found from direct integration approach, as we expected; see, for instance, (4.44) and (4.52).

Some Numerical Details and Discussion

It is now appropriate to check the theoretical results above with some simulations based on 10^4 replications. Figures 4.15 to 4.20 show, for each expectation in (4.42), the agreement between theory and simulation for $1\leq i\leq n=1000$ when $\alpha=4$ and $\tau=3$; in all cases, the simulated values (shown in steps of 50 for graphical convenience) are close to the theoretical values obtained from the direct method. Again, we report in further details results only for $E\left[\frac{X_{i:n}^{\tau}\ln X_{i:n}\ln(1+X_{i:n}^{\tau})}{1+X_{i:n}^{\tau}}\right]$, although results for the remaining expectations at (4.42) show similar observations. Table 4.5 shows that the direct method (upmost entries) gives identical results to the derivatives method (middle entries) across all i and n. These values may be compared with the lowest entries obtained from 10^4 replications of samples of size n; we see almost perfect agreement between theory and simulation. We note that, as anticipated, the computational burden increases considerably with sample size; for example, Mathematica took up to 4 hours to evaluate $E\left[\frac{X_{i:n}^{\tau}\ln X_{i:n}\ln(1+X_{i:n}^{\tau})}{1+X_{i:n}^{\tau}}\right]$ when $n=5000$. We now move on to solve the joint expectations of Burr order statistics, in which case the algebra becomes much more involved than that discussed in this section.

i	n					
	25	50	100	1000	2500	5000
$0.2n$: direct	-0.0030	-0.0029	-0.0029	-0.0029	-0.0029	-0.0029
: deriv.	-0.0030	-0.0029	-0.0029	-0.0029	-0.0029	-0.0029
: simul.	-0.0030	-0.0029	-0.0029	-0.0029	-0.0029	-0.0029
$0.4n$: direct	-0.0099	-0.0100	-0.0101	-0.0102	-0.0102	-0.0102
: deriv.	-0.0099	-0.0100	-0.0101	-0.0102	-0.0102	-0.0102
: simul.	-0.0099	-0.0100	-0.0101	-0.0102	-0.0102	-0.0102
$0.6n$: direct	-0.0199	-0.0206	-0.0209	-0.0212	-0.0212	-0.0212
: deriv.	-0.0199	-0.0206	-0.0209	-0.0212	-0.0212	-0.0212
: simul.	-0.0200	-0.0206	-0.0209	-0.0212	-0.0212	-0.0212
$0.8n$: direct	-0.0280	-0.0295	-0.0303	-0.0311	-0.0312	-0.0312
: deriv.	-0.0280	-0.0295	-0.0303	-0.0311	-0.0312	-0.0312
: simul.	-0.0281	-0.0295	-0.0303	-0.0311	-0.0312	-0.0312
$1.0n$: direct	0.1369	0.2307	0.3514	0.9393	1.2485	1.5091
: deriv.	0.1369	0.2307	0.3514	0.9393	1.2485	1.5091
: simul.	0.1407	0.2311	0.3543	0.9422	1.2517	1.5034

Table 4.5: Numerical comparison of $E\left[\frac{X_{i:n}^{\tau}\ln X_{i:n}\ln(1+X_{i:n}^{\tau})}{1+X_{i:n}^{\tau}}\right]$ for various i and n, for Burr data generated with $\alpha = 4, \tau = 3$.

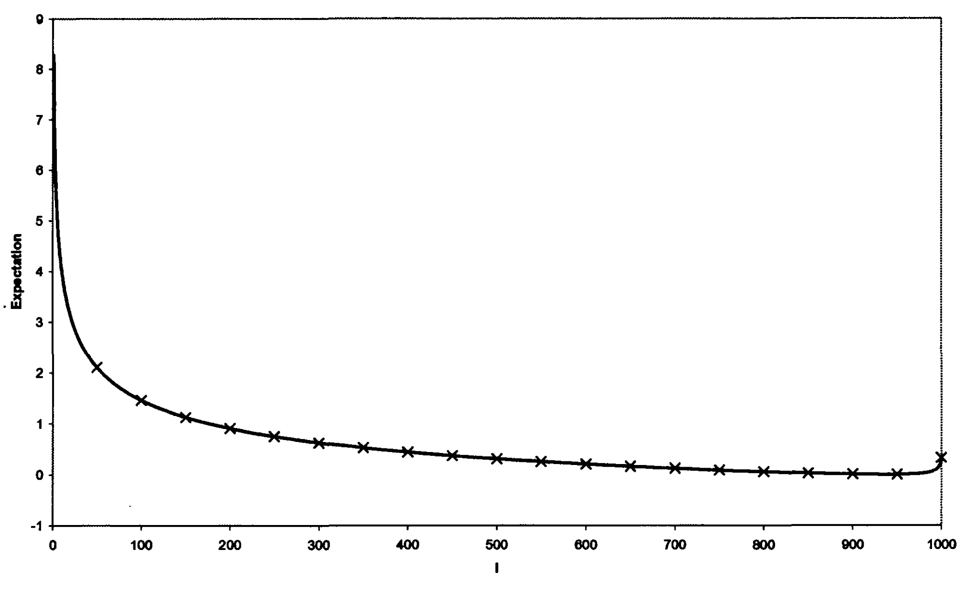

Figure 4.15: Theoretical (–) and simulated (×) values of $E\left[(\ln X_{i:n})^2\right]$ versus i, for $n = 1000, \alpha = 4, \tau = 3$.

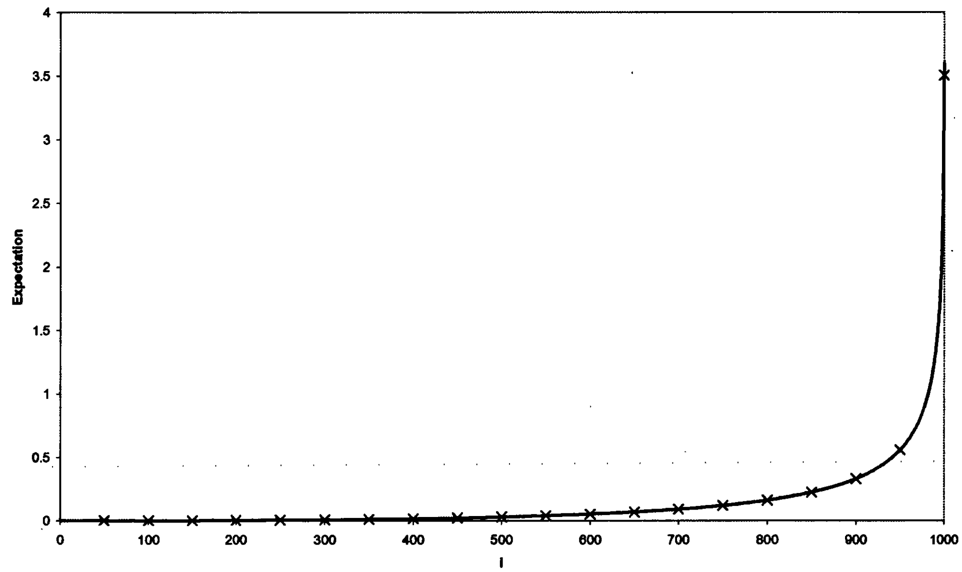

Figure 4.16: Theoretical (−) and simulated (×) values of $E\left[(\ln(1+X_{i:n}^{\tau}))^2\right]$ versus i, for $n = 1000, \alpha = 4, \tau = 3$.

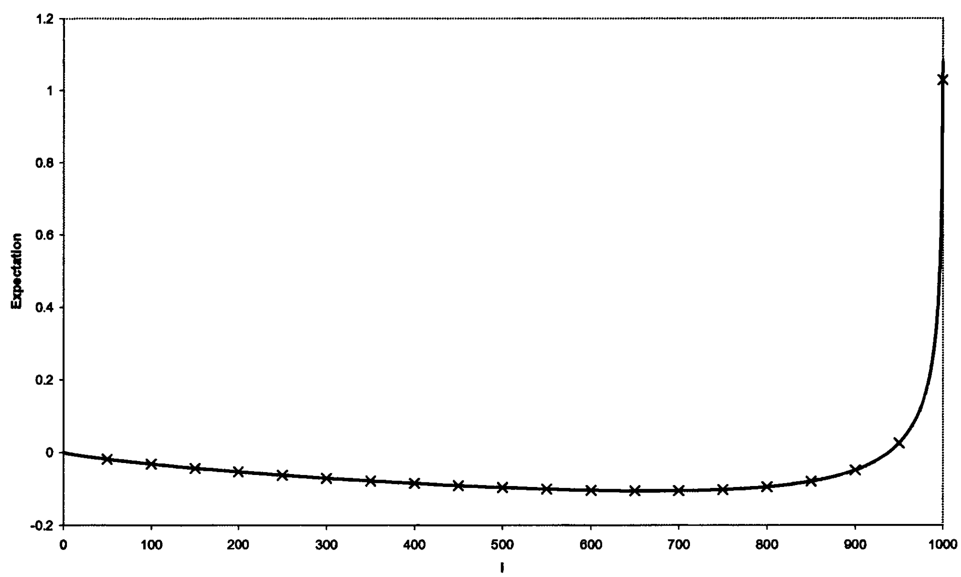

Figure 4.17: Theoretical (−) and simulated (×) values of $E\left[\ln X_{i:n} \ln(1+X_{i:n}^{\tau})\right]$ versus i, for $n = 1000, \alpha = 4, \tau = 3$.

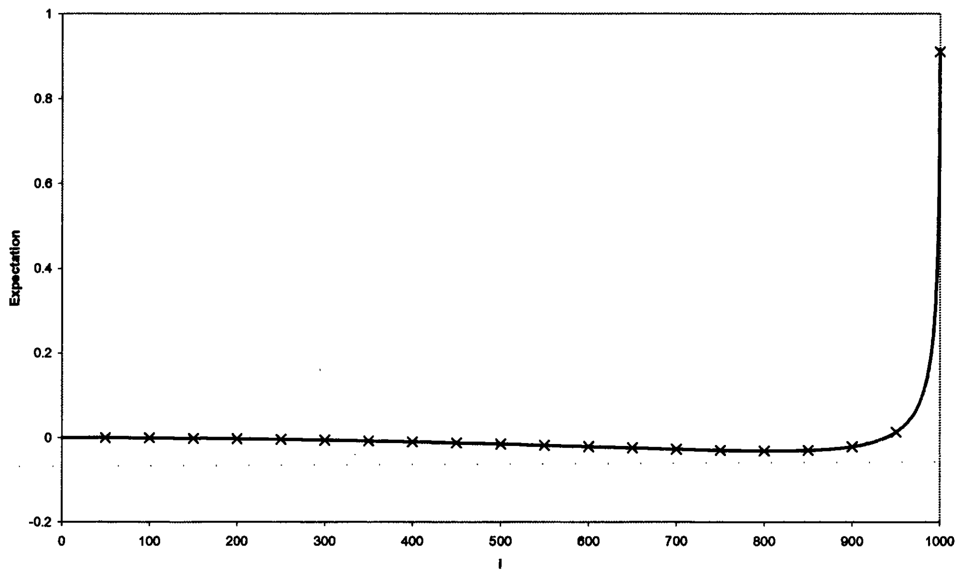

Figure 4.18: Theoretical (−) and simulated (×) values of $E\left[\frac{X_{i:n}^{\tau}\ln X_{i:n}\ln(1+X_{i:n}^{\tau})}{1+X_{i:n}^{\tau}}\right]$ versus i, for $n = 1000, \alpha = 4, \tau = 3$.

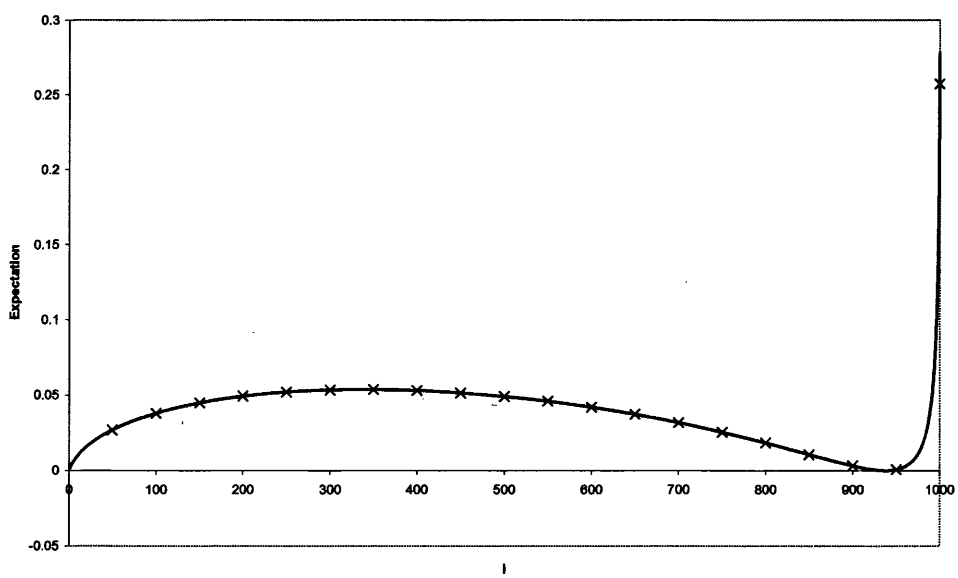

Figure 4.19: Theoretical (−) and simulated (×) values of $E\left[\frac{X_{i:n}^{\tau}(\ln X_{i:n})^2}{1+X_{i:n}^{\tau}}\right]$ versus i, for $n = 1000, \alpha = 4, \tau = 3$.

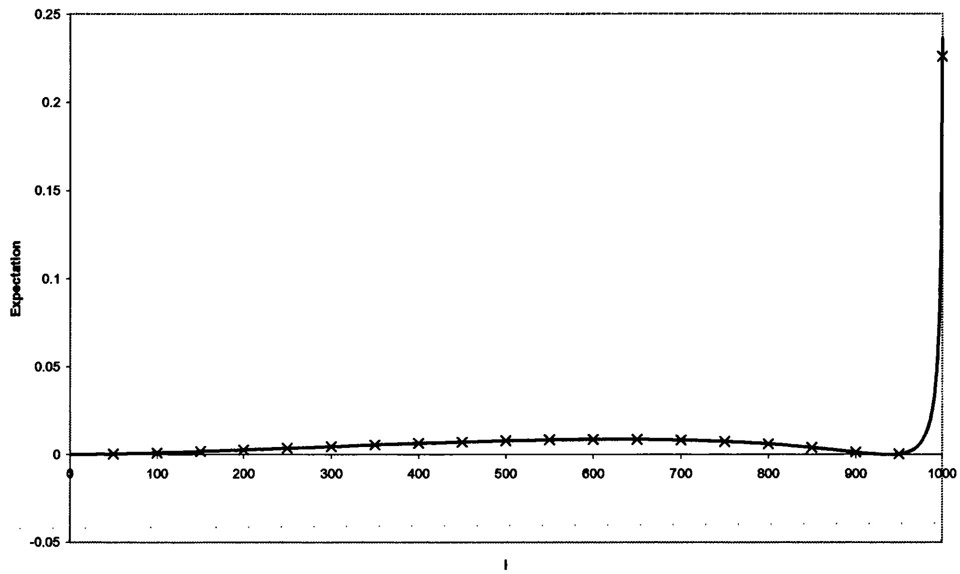

Figure 4.20: Theoretical (–) and simulated (×) values of $E\left[\left(\frac{X_{i:n}^{\tau}\ln X_{i:n}}{1+X_{i:n}^{\tau}}\right)^2\right]$ versus i, for $n = 1000, \alpha = 4, \tau = 3$.

4.3.2 Joint Expectations of $g(X_{i:n})$ and $h(X_{j:n})$

In particular, we will require (4.41) to be

$$E\left[\ln X_{i:n} \ln X_{j:n}\right], \tag{4.53a}$$

$$E\left[\ln(1+X_{i:n}^{\tau})\ln(1+X_{j:n}^{\tau})\right], \tag{4.53b}$$

$$E\left[\frac{X_{i:n}^{\tau}\ln X_{i:n}}{1+X_{i:n}^{\tau}}\frac{X_{j:n}^{\tau}\ln X_{j:n}}{1+X_{j:n}^{\tau}}\right], \tag{4.53c}$$

$$E\left[\ln X_{i:n}\ln(1+X_{j:n}^{\tau})\right], \tag{4.53d}$$

$$E\left[\ln(1+X_{i:n}^{\tau})\ln X_{j:n}\right], \tag{4.53e}$$

$$E\left[\ln X_{i:n}\frac{X_{j:n}^{\tau}\ln X_{j:n}}{1+X_{j:n}^{\tau}}\right], \tag{4.53f}$$

$$E\left[\frac{X_{i:n}^{\tau}\ln X_{i:n}}{1+X_{i:n}^{\tau}}\ln X_{j:n}\right], \tag{4.53g}$$

$$E\left[\ln(1+X_{i:n}^{\tau})\frac{X_{j:n}^{\tau}\ln X_{j:n}}{1+X_{j:n}^{\tau}}\right], \tag{4.53h}$$

$$E\left[\frac{X_{i:n}^{\tau}\ln X_{i:n}}{1+X_{i:n}^{\tau}}\ln(1+X_{j:n}^{\tau})\right], \tag{4.53i}$$

which can be deemed as some extended functions of (4.42).

Expectations in Terms of $X_{1:n}$ and $X_{j:n}$

For simplicity, we start from the expectations of $X_{1:n}$ and $X_{j:n}$ so that our problem reduces to a single summation, and then exploit the recurrence relationship given in (1.51) to give results in terms of $X_{i:n}$ and $X_{j:n}$. The joint pdf of $X_{1:n}$ and $X_{j:n}$ $(2 \le j \le n)$ is

$$f_{(1,j)}(x,y) = c_{1,j:n}(\alpha\tau)^2 (xy)^{\tau-1}\left(1+x^\tau\right)^{-\alpha-1}\left(1+y^\tau\right)^{-\alpha(n-j+1)-1}\left[\left(1+x^\tau\right)^{-\alpha}-\left(1+y^\tau\right)^{-\alpha}\right]^{j-2}$$

and from the Binomial expansion we see that

$$\left[\left(1+x^\tau\right)^{-\alpha}-\left(1+y^\tau\right)^{-\alpha}\right]^{j-2} = \sum_{k=0}^{j-2}(-1)^{j-k}\binom{j-2}{k}\left(1+x^\tau\right)^{-\alpha k}\left(1+y^\tau\right)^{-\alpha(j-2-k)}$$

so that $f_{(1,j)}(x,y)$ becomes

$$c_{1,j:n}(\alpha\tau)^2\sum_{k=0}^{j-2}(-1)^{j-k}\binom{j-2}{k}(xy)^{\tau-1}\left(1+x^\tau\right)^{-\alpha(1+k)-1}\left(1+y^\tau\right)^{-\alpha(n-k-1)-1}. \qquad (4.54)$$

Direct Method

In general, (4.41) may be stated, for $X_{1:n}$ and $X_{j:n}$, as

$$\begin{aligned}&\int_{y=0}^{\infty}\int_{x=0}^{y}\frac{x^p(\ln x)^a\left(\ln(1+x^\tau)\right)^b}{(1+x^\tau)^c}\frac{y^q(\ln y)^d\left(\ln(1+y^\tau)\right)^e}{(1+y^\tau)^f}f_{(1,j)}(x,y)dxdy\\ = \;& c_{1,j:n}(\alpha\tau)^2\sum_{k=0}^{j-2}(-1)^{j-k}\binom{j-2}{k}I_{c,f}^{pab,qde}\end{aligned}$$

where we have introduced the notation

$$I_{c,f}^{pab,qde} = \int_{y=0}^{\infty}\int_{x=0}^{y}\left\{\begin{array}{c} x^{p+\tau-1}(\ln x)^a\left(\ln(1+x^\tau)\right)^b\left(1+x^\tau\right)^{-\alpha(1+k)-c-1}\\ y^{q+\tau-1}(\ln y)^d\left(\ln(1+y^\tau)\right)^e\left(1+y^\tau\right)^{-\alpha(n-k-1)-f-1}\end{array}\right\}dxdy.$$

Then, the expectations in (4.53) require, in turn, integrals of the form

$$I_{0,0}^{010,010}, I_{0,0}^{001,001}, I_{1,1}^{\tau10,\tau10}, I_{0,0}^{010,001}, I_{0,0}^{001,010}, I_{0,1}^{010,\tau10}, I_{1,0}^{\tau10,010}, I_{0,1}^{001,\tau10}, I_{1,0}^{\tau10,001}.$$

Consequently, here the algebra becomes much more complicated than that discussed in previous sections, and, in some cases, involves integration of the $F_{3,2}$ series. We now consider each case in detail and, for convenience, let $\alpha(1+k)=s$ and $\alpha(n-k-1)=t$.

1. $E\left[\ln X_{1:n}\ln X_{j:n}\right]$ The relevant integral is

$$I_{0,0}^{010,010} = \int_{y=0}^{\infty}\int_{x=0}^{y} x^{\tau-1}\ln x\left(1+x^\tau\right)^{-s-1}y^{\tau-1}\ln y\left(1+y^\tau\right)^{-t-1}dxdy$$

at which solving the interior part gives rise to the $F_{3,2}(-y^\tau)$ series:

$$\begin{aligned}
&\int_{x=0}^{y} x^{\tau-1}\ln x\,(1+x^\tau)^{-s-1}\,dx \\
=\;& \frac{1}{\tau^2 s}\left\{\tau\ln y-\tau\ln y\,(1+y^\tau)^{-s}-sy^\tau F_{3,2}(1,1,1+s;2,2;-y^\tau)\right\}.
\end{aligned}$$

Hence, ignoring the constants wrt y, the exterior parts are

$$\begin{aligned}
\int_{y=0}^{\infty} y^{\tau-1}(\ln y)^2(1+y^\tau)^{-t-1}\,dy &= \frac{\frac{\pi^2}{6}+[\gamma+\psi(t)]^2+\psi'(t)}{\tau^3 t}, \\
\int_{y=0}^{\infty} y^{\tau-1}(\ln y)^2(1+y^\tau)^{-s-t-1}\,dy &= \frac{\frac{\pi^2}{6}+[\gamma+\psi(s+t)]^2+\psi'(s+t)}{\tau^3(s+t)},
\end{aligned}$$

but

$$\int_{y=0}^{\infty} y^{2\tau-1}\ln y\,(1+y^\tau)^{-t-1}F_{3,2}(1,1,1+s;2,2;-y^\tau)\,dy \tag{4.55}$$

is insolvable, primarily because there are too many functions of y (power, logarithm and algebraic) appearing simultaneously with the $F_{3,2}(-y^\tau)$ series of a power y argument. Alternatively, we write the hypergeometric function in terms of the gamma functions:

$$F_{3,2}(1,1,1+s;2,2;-y^\tau)=\sum_{m=0}^{\infty}\frac{\Gamma(1+s+m)}{\Gamma(s)(m+1)^2}\frac{(-y^\tau)^m}{m!},$$

so that the problem turns into

$$\begin{aligned}
&\sum_{m=0}^{\infty}\frac{\Gamma(1+s+m)(-1)^m}{\Gamma(s)(m+1)^2 m!}\int_{y=0}^{\infty} y^{\tau(2+m)-1}\ln y\,(1+y^\tau)^{-t-1}\,dy \\
=\;& \sum_{m=0}^{\infty}\frac{\Gamma(1+s+m)(-1)^m}{\Gamma(s)(m+1)^2 m!}\frac{\Gamma(2+m)\Gamma(t-1-m)[\psi(2+m)-\psi(t-1-m)]}{\tau^2\Gamma(t+1)}.
\end{aligned}$$

Nevertheless, since

$$\begin{aligned}
n-j+1 &\le n-k-1\le n-1 \quad (\text{as } 0\le k\le j-2) \\
&\Rightarrow 1\le n-k-1\le n-1 \quad (\text{as } 2\le j\le n) \\
&\Rightarrow \alpha\le\alpha(n-k-1)\le\alpha(n-1) \\
&\Rightarrow \alpha-1\le\alpha(n-k-1)-1\le\alpha(n-1)-1 \\
&\Rightarrow \alpha-1\le t-1\le\alpha(n-1)-1,
\end{aligned}$$

the functions $\Gamma(t-1-m)$ and $\psi(t-1-m)$ will soon become invalid (negative) in $\sum_{m=0}^{\infty}$, indicating that (4.55) remains insolvable.

2. $E\left[\ln(1+X_{1:n}^{\tau})\ln(1+X_{j:n}^{\tau})\right]$ We need $I_{0,0}^{001,001}$ with an inner integral of

$$\begin{aligned} & \int_{x=0}^{y} x^{\tau-1}\ln(1+x^{\tau})(1+x^{\tau})^{-s-1}\,dx \\ = & \frac{1}{\tau s^2}\left\{1-(1+y^{\tau})^{-s}-(1+y^{\tau})^{-s}s\ln(1+y^{\tau})\right\} \end{aligned}$$

so that integration wrt y (dropping the constants) consists of

$$\int_{y=0}^{\infty} y^{\tau-1}\ln(1+y^{\tau})(1+y^{\tau})^{-t-1}\,dy = \frac{1}{\tau}\int_{u=1}^{\infty}(\ln u)\,u^{-t-1}du = \frac{1}{\tau t^2}, \tag{4.56}$$

$$\begin{aligned} \int_{y=0}^{\infty} y^{\tau-1}\ln(1+y^{\tau})(1+y^{\tau})^{-s-t-1}\,dy &= \frac{1}{\tau}\int_{u=1}^{\infty}(\ln u)\,u^{-s-t-1}du \\ &= \frac{1}{\tau(s+t)^2}, \end{aligned} \tag{4.57}$$

and

$$\int_{y=0}^{\infty} y^{\tau-1}(\ln(1+y^{\tau}))^2(1+y^{\tau})^{-s-t-1}\,dy = \frac{1}{\tau}\int_{u=1}^{\infty}(\ln u)^2 u^{-s-t-1}du = \frac{2}{\tau(s+t)^3},$$

obtained on letting $u = 1+y^{\tau}$ $(0 \le y \le \infty \Leftrightarrow 1 \le u \le \infty$ and $y = (u-1)^{\frac{1}{\tau}}$ so $dy = \frac{1}{\tau}(u-1)^{\frac{1}{\tau}-1}du)$. We thus have

$$I_{0,0}^{001,001} = \frac{1}{\tau s^2}\left\{\frac{1}{\tau t^2} - \frac{1}{\tau(s+t)^2} - s\frac{2}{\tau(s+t)^3}\right\}.$$

3. $E\left[\frac{X_{1:n}^{\tau}\ln X_{1:n}}{1+X_{1:n}^{\tau}}\frac{X_{j:n}^{\tau}\ln X_{j:n}}{1+X_{j:n}^{\tau}}\right]$ The inner integral in $I_{1,1}^{\tau 10,\tau 10}$ is

$$\begin{aligned} & \int_{x=0}^{y} x^{2\tau-1}\ln x\,(1+x^{\tau})^{-s-2}\,dx \\ = & \frac{1}{\tau^2 s(s+1)}\left\{\begin{array}{c} 1-(1+y^{\tau})^{-s}-sy^{\tau}F_{3,2}(1,1,1+s;2,2;-y^{\tau}) \\ +\tau s(s+1)\ln y B_{-y^{\tau}}(2,-1-s) \end{array}\right\} \end{aligned}$$

which solution involves the hypergeometric and incomplete beta functions. Then, neglecting the constants, integrate wrt y yields

$$\int_{y=0}^{\infty} y^{2\tau-1}\ln y\,(1+y^{\tau})^{-t-2}\,dy = \frac{1-\gamma-\psi(t)}{\tau^2 t(t+1)}, \tag{4.58}$$

$$\int_{y=0}^{\infty} y^{2\tau-1}\ln y\,(1+y^{\tau})^{-s-t-2}\,dy = \frac{1-\gamma-\psi(s+t)}{\tau^2(s+t)(s+t+1)}, \tag{4.59}$$

$$\int_{y=0}^{\infty} y^{3\tau-1}\ln y\,(1+y^{\tau})^{-t-2}F_{3,2}(1,1,1+s;2,2;-y^{\tau})\,dy \tag{4.60}$$

which, similar to (4.55), is insolvable, and

$$\begin{aligned}
&\int_{y=0}^{\infty} y^{2\tau-1} (\ln y)^2 (1+y^\tau)^{-t-2} B_{-y^\tau}(2,-1-s)\, dy \\
=\ & \int_{y=0}^{\infty} y^{2\tau-1} (\ln y)^2 (1+y^\tau)^{-t-2} \frac{y^{2\tau}}{2} F_{2,1}(2, 2+s; 3; -y^\tau)\, dy \quad \text{from (1.11)} \\
=\ & -\frac{1}{6\tau^3 (s+s^2)} \left\{ \begin{array}{c}
\frac{-12\gamma+6\gamma^2+\pi^2+6\psi^2(s+t+1)}{(s+t+1)(s+t+2)} + \frac{(s+1)\left(12-36\gamma+12\gamma^2+2\pi^2+12\psi^2(s+t)\right)}{(s+t)(s+t+1)(s+t+2)} \\
+\frac{6}{(s+t)^2(s+t+1)(s+t+2)} + \frac{12\gamma-6\gamma^2-\pi^2+6\psi^2(t+1)}{2+3t+t^2} \\
+\frac{12-12\gamma-12+36\gamma-12\gamma^2-2\pi^2+6\psi(t)(4\gamma+2t\gamma-6-2t)-12\psi^2(t)}{t(2+3t+t^2)} + \frac{6}{t^2(2+3t+t^2)} \\
+\frac{-12+12\gamma+12\psi(s+t)[-3-4s-t+\gamma(2+3s+t)]+6\psi'(s+t)(2+3s+t)}{(s+t)(2+3s+3t+2st+s^2+t^2)} - \frac{6\psi'(t)}{(t+t^2)}
\end{array} \right\}.
\end{aligned}$$

4. $E\left[\ln X_{1:n} \ln(1+X_{j:n}^\tau)\right]$ Solving the interior integration in $I_{0,0}^{010,001}$ gives rise to the $F_{3,2}(-y^\tau)$ series (see $I_{0,0}^{010,010}$) at which on omitting the constants, its exterior integrals become

$$\begin{aligned}
\int_{y=0}^{\infty} y^{\tau-1} \ln y \ln(1+y^\tau)(1+y^\tau)^{-t-1} dy &= \frac{-\gamma + t\psi'(t)}{\tau^2 t^2}, \\
\int_{y=0}^{\infty} y^{\tau-1} \ln y \ln(1+y^\tau)(1+y^\tau)^{-s-t-1} dy &= \frac{-\gamma + (s+t)\psi'(s+t)}{\tau^2 (s+t)^2}, \quad (4.61)
\end{aligned}$$

and

$$\begin{aligned}
&\int_{y=0}^{\infty} y^{2\tau-1} \ln(1+y^\tau)(1+y^\tau)^{-t-1} F_{3,2}(1,1,1+s;2,2;-y^\tau)\, dy \\
=\ & \frac{1}{\tau s t^2} \left\{ \begin{array}{c}
-\psi(1-t) + \psi(1-s-t) - t\psi'(1-t) + t\psi'(1-s-t) \\
+\pi \csc[\pi t] \csc[\pi(s+t)] \sin[\pi s] + \\
(\cot[\pi t] + \cot[\pi(s+t)])\, \pi^2 t \csc[\pi t] \csc[\pi(s+t)] \sin[\pi s]
\end{array} \right\}. \quad (4.62)
\end{aligned}$$

5. $E\left[\ln(1+X_{i:n}^\tau) \ln X_{j:n}\right]$ The internal integral of $I_{0,0}^{001,010}$ has been studied before in $I_{0,0}^{001,001}$; by dropping the constants, the external parts are

$$\begin{aligned}
\int_{y=0}^{\infty} y^{\tau-1} \ln y (1+y^\tau)^{-t-1} dy &= -\frac{\gamma+\psi(t)}{\tau^2 t}, \quad (4.63) \\
\int_{y=0}^{\infty} y^{\tau-1} \ln y (1+y^\tau)^{-s-t-1} dy &= -\frac{\gamma+\psi(s+t)}{\tau^2 (s+t)}, \quad (4.64)
\end{aligned}$$

and (4.61). We thus obtain

$$I_{0,0}^{001,010} = \frac{1}{\tau^3 s^2} \left\{ -\frac{\gamma+\psi(t)}{t} + \frac{\gamma+\psi(s+t)}{s+t} - s\frac{-\gamma+(s+t)\psi'(s+t)}{(s+t)^2} \right\}.$$

6. $E\left[\ln X_{i:n} \frac{X_{j:n}^\tau \ln X_{j:n}}{1+X_{j:n}^\tau}\right]$ $I_{0,1}^{010,\tau 10}$ has same inner part to that in $I_{0,0}^{010,010}$; we thus have outer

integrals (ignore the constants wrt y) as

$$\begin{aligned}
\int_{y=0}^{\infty} y^{2\tau-1} (\ln y)^2 (1+y^\tau)^{-t-2} dy &= \frac{\frac{\pi^2}{6} + [\gamma + \psi(t)]^2 - 2(\gamma + \psi(t)) + \psi'(t)}{\tau^3 t (1+t)}, \\
\int_{y=0}^{\infty} y^{2\tau-1} (\ln y)^2 (1+y^\tau)^{-s-t-2} dy &= \frac{\frac{\pi^2}{6} + [\gamma + \psi(s+t)]^2 - 2(\gamma + \psi(s+t)) + \psi'(s+t)}{\tau^3 (s+t)(s+t+1)},
\end{aligned}$$

and (4.60).

7. $E\left[\frac{X_{i:n}^\tau \ln X_{i:n}}{1+X_{i:n}^\tau} \ln X_{j:n}\right]$ We need $I_{1,0}^{\tau 10,010}$ whose inner integral is identical to that in $I_{1,1}^{\tau 10,\tau 10}$ whereas its outer integral is similar to $I_{0,0}^{001,010}$; we see that integration wrt y gives (neglect the constants)

$$\begin{aligned}
&\int_{y=0}^{\infty} y^{\tau-1} (\ln y)^2 (1+y^\tau)^{-t-1} B_{-y^\tau}(2, -1-s)\, dy \\
= {} & \int_{y=0}^{\infty} y^{\tau-1} (\ln y)^2 (1+y^\tau)^{-t-1} \frac{y^{2\tau}}{2} F_{2,1}(2, 2+s; 3; -y^\tau)\, dy \quad \text{from } (1.11) \\
= {} & -\frac{1}{6\tau^3 (s+s^2)} \left\{ \begin{array}{c} -\frac{6\psi'(t)}{t} + \frac{-6\gamma^2-\pi^2-6\psi^2(t+1)}{t+1} + \frac{-6\gamma^2-\pi^2-12\psi(t)(-1+\gamma+\gamma t)-6\psi^2(t)}{t(t+1)} + \\ \frac{6}{t^2(t+1)} + \frac{6\gamma^2+\pi^2+\psi^2(s+t+1)}{s+t+1} - \frac{6}{(s+t)^2(s+t+1)} + \\ \frac{(s+1)\left(6\gamma^2+\pi^2+6\psi^2(s+t)\right)-12\gamma s+12\psi(s+t)(-1-s+\gamma(1+2s+t))+6(1+2s+t)\psi'(s+t)}{(s+t)(s+t+1)} \end{array} \right\},
\end{aligned}$$

as well as (4.63), (4.64) and (4.55).

8. $E\left[\ln(1+X_{i:n}^\tau) \frac{X_{j:n}^\tau \ln X_{j:n}}{1+X_{j:n}^\tau}\right]$ $I_{0,1}^{001,\tau 10}$ has same inner part to that in $I_{0,0}^{001,001}$ but a similar outer part with $I_{1,1}^{\tau 10,\tau 10}$; its outer integrals (omit the constants wrt y) are

$$\begin{aligned}
&\int_{y=0}^{\infty} y^{2\tau-1} \ln y \ln(1+y^\tau)(1+y^\tau)^{-s-t-2} dy \\
= {} & \frac{(1+2s+2t)[1-\gamma-\psi(s+t)] + (s+t)(s+t+1)\psi'(s+t)}{\tau^2 (s+t)^2 (s+t+1)^2},
\end{aligned}$$

(4.58) and (4.59). Consequently, we have

$$I_{0,1}^{001,\tau 10} = \frac{1}{\tau^3 s^2} \left\{ \begin{array}{c} \frac{1-\gamma-\psi(t)}{t(t+1)} - \frac{1-\gamma-\psi(s+t)}{(s+t)(s+t+1)} \\ -s\frac{(1+2s+2t)[1-\gamma-\psi(s+t)]+(s+t)(s+t+1)\psi'(s+t)}{(s+t)^2(s+t+1)^2} \end{array} \right\}.$$

9. $E\left[\frac{X_{i:n}^\tau \ln X_{i:n}}{1+X_{i:n}^\tau} \ln(1+X_{j:n}^\tau)\right]$ $I_{1,0}^{\tau 10,001}$ has an identical interior part as $I_{1,1}^{\tau 10,\tau 10}$ but exterior part similar to $I_{0,0}^{001,001}$; we see that integration wrt y gives (ignore the constants)

(4.56), (4.57), (4.62) and

$$\begin{aligned}
&\int_{y=0}^{\infty} y^{\tau-1} \ln y \ln(1+y^{\tau}) (1+y^{\tau})^{-t-1} B_{-y^{\tau}} (2, -1-s)\, dy \\
=\ &\int_{y=0}^{\infty} y^{\tau-1} \ln y \ln(1+y^{\tau}) (1+y^{\tau})^{-t-1} \frac{y^{2\tau}}{2} F_{2,1} (2, 2+s; 3; -y^{\tau})\, dy \\
=\ &\frac{1}{2} \sum_{m=0}^{\infty} \frac{(2)_m (2+s)_m (-1)^m}{(3)_m\, m!} \int_{y=0}^{\infty} y^{\tau(3+m)-1} \ln y \ln(1+y^{\tau}) (1+y^{\tau})^{-t-1}\, dy \\
=\ &\frac{1}{2} \sum_{m=0}^{\infty} \frac{(2)_m (2+s)_m (-1)^m}{(3)_m\, m!} \times \frac{\pi \Gamma(-2-m+t)}{\tau^2 \Gamma(-2-m) \Gamma(1+t)} \times \\
&\left\{ \begin{array}{c}
-\pi^2 \cot[\pi(m-t)] \csc[\pi(m-t)] \csc[\pi t] + \pi \cot[mt] \csc[m\pi]\, \psi(-t) \\
+\psi(-2-m) \left(\pi \csc[\pi(m-t)] \csc[\pi t] + \csc[m\pi]\, \psi(3+m-t) - \csc[m\pi]\, \psi(-t)\right) \\
-\pi \csc[\pi(m-t)] \csc[\pi t]\, \psi(-2-m+t) + \csc[m\pi]\, \psi(-t)\, \psi(-2-m+t) \\
-\csc[m\pi]\, \psi(3+m-t) \left(\pi \cot[m\pi] + \psi(-2-m+t)\right) + \csc[m\pi]\, \psi'(3+m-t)
\end{array} \right\}.
\end{aligned}$$

However, owing to negative arguments in the gamma and polygamma functions the last integral has no solution.

In summary, using the direct integration approach, we are only able to find expressions for the second, fourth, fifth and eighth expectation in (4.53); others are insolvable here due to the inability to solve the integration of the following form:

$$\int_{y=0}^{\infty} y^{q-1} (\ln y)^d (1+y^{\tau})^{-f-1} F_{3,2} (1, 1, 1+s; 2, 2; -y^{\tau})\, dy.$$

We can, however, employ the derivatives method to obtain the expressions for (4.53) where we require a basic result under which differentiations could be applied to, as discussed in the following section.

Basic Expectation

As with (4.45) for single expectations, the preliminary for joint expectations here is

$$E_{1j} = E\left[\frac{X_{1:n}^p}{(1+X_{1:n}^{\tau})^c} \frac{X_{j:n}^q}{\left(1+X_{j:n}^{\tau}\right)^f} \right]$$

so that its partial derivatives wrt p, q, c, f will introduce $\ln X_{1:n}$, $\ln X_{j:n}$, $\ln(1+X_{1:n}^{\tau})$ and $\ln(1+X_{j:n}^{\tau})$ in turn, to give the functions in (4.53). Accordingly, E_{1j} may be expressed as

$$\begin{aligned}
&\int_{y=0}^{\infty} \int_{x=0}^{y} \frac{x^p}{(1+x^{\tau})^c} \times \frac{y^q}{(1+y^{\tau})^f} f_{(1,j)}(x,y)\, dx dy \\
=\ & c_{1,j:n} (\alpha\tau)^2 \sum_{k=0}^{j-2} (-1)^{j-k} \binom{j-2}{k} I_{c,f}^{p00,q00}.
\end{aligned}$$

It will prove more convenient to consider a reduced version of $I_{c,f}^{p00,q00}$, as follow, and then give result for $I_{c,f}^{p00,q00}$.

A result on integration Suppose here

$$S = -\alpha(1+k) - c$$

and

$$T = -\alpha(n-k-1) - f.$$

Due to the format of (4.54), the following integral

$$I_{S,T}^{p,q} = \int_{y=0}^{\infty}\int_{x=0}^{y} x^{p+\tau-1}(1+x^\tau)^{S-1} y^{q+\tau-1}(1+y^\tau)^{T-1}\,dxdy$$

often appears in the joint expectation of $X_{i:n}$ and $X_{j:n}$. Suppose further

$$u = x^\tau$$

then $0 \le x \le y \Leftrightarrow 0 \le u \le y^\tau$, and $x = u^{\frac{1}{\tau}}$ so $dx = \frac{1}{\tau}u^{\frac{1}{\tau}-1}du$. The inner integral wrt x takes the form

$$\int_{x=0}^{y} x^{p+\tau-1}(1+x^\tau)^{S-1}\,dx = \frac{1}{\tau}\int_{u=0}^{y^\tau} u^{\frac{p}{\tau}}(1+u)^{S-1}\,du,$$

but also can be written as

$$\frac{1}{\tau}\int_{v=0}^{\frac{y^\tau}{1+y^\tau}} v^{\frac{p}{\tau}}(1-v)^{-\frac{p}{\tau}-S-1}\,dv = \frac{1}{\tau}B_{\frac{y^\tau}{1+y^\tau}}\left(\frac{p}{\tau}+1, -\frac{p}{\tau}-S\right),$$

obtained by setting

$$v = \frac{u}{1+u}$$

where $0 \le u \le y^\tau \Leftrightarrow 0 \le v \le \frac{y^\tau}{1+y^\tau}$, and $u = \frac{v}{1-v}$ so $du = \frac{1}{(1-v)^2}dv$. Hence, $I_{S,T}^{p,q}$ is now

$$\frac{1}{\tau}\int_{y=0}^{\infty} y^{q+\tau-1}(1+y^\tau)^{T-1} B_{\frac{y^\tau}{1+y^\tau}}\left(\frac{p}{\tau}+1, -\frac{p}{\tau}-S\right)dy,$$

and if we let $w = y^\tau$, we have

$$\frac{1}{\tau^2}\int_{w=0}^{\infty} w^{\frac{q}{\tau}}(1+w)^{T-1} B_{\frac{w}{1+w}}\left(\frac{p}{\tau}+1, -\frac{p}{\tau}-S\right)dw$$

because $0 \le y \le \infty \Leftrightarrow 0 \le w \le \infty$, and $y = w^{\frac{1}{\tau}}$ so $dy = \frac{1}{\tau}w^{\frac{1}{\tau}-1}dw$.

We use two results from Chapter 1 to proceed; first, we use the relationship between the incomplete beta and hypergeometric functions, see (1.11), to write

$$B_{\frac{w}{1+w}}\left(\frac{p}{\tau}+1,-\frac{p}{\tau}-S\right)=\left(\frac{p}{\tau}+1\right)^{-1}\left(\frac{w}{1+w}\right)^{\frac{p}{\tau}+1}F_{2,1}\left(\frac{p}{\tau}+1,\frac{p}{\tau}+S+1;\frac{p}{\tau}+2;\frac{w}{1+w}\right)$$

and then use (1.12) to write the above as

$$\begin{aligned}
&= \left(\frac{p}{\tau}+1\right)^{-1}\left(\frac{w}{1+w}\right)^{\frac{p}{\tau}+1}\sum_{m=0}^{\infty}\frac{\Gamma\left(\frac{p}{\tau}+1+m\right)\Gamma\left(\frac{p}{\tau}+S+1+m\right)\Gamma\left(\frac{p}{\tau}+2\right)}{\Gamma\left(\frac{p}{\tau}+2+m\right)\Gamma\left(\frac{p}{\tau}+1\right)\Gamma\left(\frac{p}{\tau}+S+1\right)m!}\left(\frac{w}{1+w}\right)^{\frac{p}{\tau}+1+m}\\
&= \sum_{m=0}^{\infty}\frac{\Gamma\left(\frac{p}{\tau}+S+1+m\right)}{\left(\frac{p}{\tau}+1+m\right)\Gamma\left(\frac{p}{\tau}+S+1\right)m!}\left(\frac{w}{1+w}\right)^{\frac{p}{\tau}+1+m}.
\end{aligned}$$

It follows that

$$I_{S,T}^{p,q}=\frac{1}{\tau^2}\sum_{m=0}^{\infty}\frac{\Gamma\left(\frac{p}{\tau}+S+1+m\right)}{\left(\frac{p}{\tau}+1+m\right)\Gamma\left(\frac{p}{\tau}+S+1\right)m!}\int_{w=0}^{\infty}w^{\frac{p}{\tau}+\frac{q}{\tau}+1+m}(1+w)^{T-\frac{p}{\tau}-m-2}\,dw$$

in which

$$\begin{aligned}
\int_{w=0}^{\infty}w^{\frac{p}{\tau}+\frac{q}{\tau}+1+m}(1+w)^{T-\frac{p}{\tau}-m-2}\,dw &= B\left(\frac{p}{\tau}+\frac{q}{\tau}+2+m,-\frac{q}{\tau}-T\right)\\
&= \frac{\Gamma\left(\frac{p}{\tau}+\frac{q}{\tau}+2+m\right)\Gamma\left(-\frac{q}{\tau}-T\right)}{\Gamma\left(\frac{p}{\tau}-T+2+m\right)}
\end{aligned}$$

so that we obtain

$$I_{S,T}^{p,q}=\frac{1}{\tau^2}\times\frac{\Gamma\left(-\frac{q}{\tau}-T\right)}{\Gamma\left(\frac{p}{\tau}+S+1\right)}\sum_{m=0}^{\infty}\frac{\Gamma\left(\frac{p}{\tau}+S+1+m\right)\Gamma\left(\frac{p}{\tau}+\frac{q}{\tau}+2+m\right)}{\left(\frac{p}{\tau}+1+m\right)\Gamma\left(\frac{p}{\tau}-T+2+m\right)m!}. \tag{4.65}$$

Moreover, we can quote the above infinite summation in term of a hypergeometric function, as follows:

$$\begin{aligned}
&\sum_{m=0}^{\infty}\frac{\Gamma\left(\frac{p}{\tau}+S+1+m\right)\Gamma\left(\frac{p}{\tau}+\frac{q}{\tau}+2+m\right)}{\left(\frac{p}{\tau}+1+m\right)\Gamma\left(\frac{p}{\tau}-T+2+m\right)m!}\\
&= \sum_{m=0}^{\infty}\frac{\Gamma\left(\frac{p}{\tau}+1+m\right)\Gamma\left(\frac{p}{\tau}+S+1+m\right)\Gamma\left(\frac{p}{\tau}+\frac{q}{\tau}+2+m\right)}{\Gamma\left(\frac{p}{\tau}+2+m\right)\Gamma\left(\frac{p}{\tau}-T+2+m\right)}\times\frac{1^m}{m!}\\
&= \frac{\Gamma\left(\frac{p}{\tau}+S+1\right)\Gamma\left(\frac{p}{\tau}+\frac{q}{\tau}+2\right)}{\left(\frac{p}{\tau}+1\right)\Gamma\left(\frac{p}{\tau}-T+2\right)}\times\\
&\quad F_{3,2}\left(\frac{p}{\tau}+1,\frac{p}{\tau}+S+1,\frac{p}{\tau}+\frac{q}{\tau}+2;\frac{p}{\tau}+2,\frac{p}{\tau}-T+2;1\right).
\end{aligned}$$

Therefore, substituting this in (4.65) yields

$$\begin{aligned} I_{S,T}^{p,q} &= \frac{1}{\tau^2} \times \frac{\Gamma\left(-\frac{q}{\tau}-T\right)}{\Gamma\left(\frac{p}{\tau}+S+1\right)} \times \frac{\Gamma\left(\frac{p}{\tau}+S+1\right)\Gamma\left(\frac{p}{\tau}+\frac{q}{\tau}+2\right)}{\left(\frac{p}{\tau}+1\right)\Gamma\left(\frac{p}{\tau}-T+2\right)} \times \\ &\quad F_{3,2}\left(\frac{p}{\tau}+1, \frac{p}{\tau}+S+1, \frac{p}{\tau}+\frac{q}{\tau}+2; \frac{p}{\tau}+2, \frac{p}{\tau}-T+2; 1\right) \\ &= \frac{\Gamma\left(-\frac{q}{\tau}-T\right)\Gamma\left(\frac{p}{\tau}+\frac{q}{\tau}+2\right)}{\tau^2\left(\frac{p}{\tau}+1\right)\Gamma\left(\frac{p}{\tau}-T+2\right)} \times \\ &\quad F_{3,2}\left(\frac{p}{\tau}+1, \frac{p}{\tau}+S+1, \frac{p}{\tau}+\frac{q}{\tau}+2; \frac{p}{\tau}+2, \frac{p}{\tau}-T+2; 1\right). \end{aligned} \tag{4.66}$$

An expression for E_{1j} Using (4.66), we now obtain an expression for E_{1j} in term of hypergeometric function:

$$\begin{aligned} E_{1j} &= c_{1,j:n}\alpha^2 \sum_{k=0}^{j-2} (-1)^{j-k} \binom{j-2}{k} \frac{\Gamma\left(f+\alpha(n-k-1)-\frac{q}{\tau}\right)\Gamma\left(\frac{p}{\tau}+\frac{q}{\tau}+2\right)}{\left(\frac{p}{\tau}+1\right)\Gamma\left(\frac{p}{\tau}+f+\alpha(n-k-1)+2\right)} \times \\ &\quad F_{3,2}\left(\begin{array}{c} \frac{p}{\tau}+1, \frac{p}{\tau}-\alpha(1+k)-c+1, \frac{p}{\tau}+\frac{q}{\tau}+2; \\ \frac{p}{\tau}+2, \frac{p}{\tau}+\alpha(n-k-1)+f+2; 1 \end{array}\right). \end{aligned} \tag{4.67}$$

Nevertheless, a scrutiny on (4.67) unveils that the second argument in the $F_{3,2}(1)$ series therein would become negative under certain circumstances, leading to invalid gamma functions; take, for instance, $\alpha = 4, \tau = 3, c = 1, k = 0, p = 1$, we see

$$\frac{p}{\tau} - \alpha(1+k) - c + 1 = \frac{1}{3} - 4(1+0) - 1 + 1 = -\frac{11}{3}.$$

Hence, we must rescale the arguments in that $F_{3,2}(1)$ series. Using (1.16), we can obtain an alternative formula to (4.67), given by

$$\begin{aligned} & c_{1,j:n}\alpha^2 \sum_{k=0}^{j-2} (-1)^{j-k} \binom{j-2}{k} \frac{\Gamma\left(\alpha n+c+f-\frac{p}{\tau}-\frac{q}{\tau}\right)\Gamma\left(\frac{p}{\tau}+\frac{q}{\tau}+2\right)}{\left(\alpha(n-k-1)+f-\frac{q}{\tau}\right)\Gamma\left(\alpha n+c+f+2\right)} \times \\ & F_{3,2}\left(\begin{array}{c} 1, \alpha(n-k-1)+f+1, \alpha n+c+f-\frac{p}{\tau}-\frac{q}{\tau}; \\ \alpha(n-k-1)+f-\frac{q}{\tau}+1, \alpha n+c+f+2; 1 \end{array}\right). \end{aligned} \tag{4.68}$$

Convergence considerations In addition, as noted in Section 1.2.4.2, it is necessary to check for convergence in the $F_{3,2}(1)$ series at (4.68); from (1.15), this series is convergent provided that

$$1 + \frac{p}{\tau} > 0,$$

which, as far as our range of interest ($\tau > 0$ and $p = 0, 1$) is concerned, this is always the case. Next, we look at the derivatives method in detail.

Specific expectations	Partial derivatives needed	p	c	q	f
$E\left[\ln X_{1:n} \ln X_{j:n}\right]$	$E''_{1j,pq}$	0	0	0	0
$E\left[\ln(1+X^{\tau}_{1:n})\ln(1+X^{\tau}_{j:n})\right]$	$E''_{1j,cf}$	0	0	0	0
$E\left[\frac{X^{\tau}_{1:n}\ln X_{1:n}}{1+X^{\tau}_{1:n}}\frac{X^{\tau}_{j:n}\ln X_{j:n}}{1+X^{\tau}_{j:n}}\right]$	$E''_{1j,pq}$	τ	1	τ	1
$E\left[\ln X_{1:n}\ln(1+X^{\tau}_{j:n})\right]$	$-E''_{1j,pf}$	0	0	0	0
$E\left[\ln(1+X^{\tau}_{1:n})\ln X_{j:n}\right]$	$-E''_{1j,cq}$	0	0	0	0
$E\left[\ln X_{1:n}\frac{X^{\tau}_{j:n}\ln X_{j:n}}{1+X^{\tau}_{j:n}}\right]$	$E''_{1j,pq}$	0	0	τ	1
$E\left[\frac{X^{\tau}_{1:n}\ln X_{1:n}}{1+X^{\tau}_{1:n}}\ln X_{j:n}\right]$	$E''_{1j,pq}$	τ	1	0	0
$E\left[\ln(1+X^{\tau}_{1:n})\frac{X^{\tau}_{j:n}\ln X_{j:n}}{1+X^{\tau}_{j:n}}\right]$	$-E''_{1j,cq}$	0	0	τ	1
$E\left[\frac{X^{\tau}_{1:n}\ln X_{1:n}}{1+X^{\tau}_{1:n}}\ln(1+X^{\tau}_{j:n})\right]$	$-E''_{1j,pf}$	τ	1	0	0

Table 4.6: Derivatives method: expectations in (4.52) and the partial derivatives needed.

Derivatives Method

As previously mentioned, unlike for the Weibull case, the expectations in (4.53) cannot all be derived directly when the underlying distribution is Burr. Hence, we focus on the derivatives approach, where we have already obtained an expression for E_{1j}, given at (4.68), in terms of the gamma and hypergeometric functions. Table 4.6 summarises the corresponding partial derivatives of E_{1j} and the values of p, c, q, f needed; the relevant derivatives are $E''_{1j,pq}$, $E''_{1j,pf}$, $E''_{1j,cq}$ and $E''_{1j,cf}$. However, it will prove more appropriate to express E_{1j} in terms of just gamma functions so that we need only the digamma and polygamma functions, which have been shown to be more manageable than the derivatives of $F_{3,2}$; therefore, we write (4.68) as

$$E_{1j} = c_{1,j:n}\alpha^2 \sum_{k=0}^{j-2} (-1)^{j-k} \binom{j-2}{k} \frac{\Gamma(b_1)\Gamma(b_2)}{\Gamma(b_3)} \sum_{m=0}^{\infty} \frac{\Gamma(b_3+m)\Gamma(b_4+m)}{\Gamma(b_5+m)\Gamma(b_1+1+m)} \tag{4.69}$$

where we have introduced the following notations for convenience:

$$\begin{aligned} b_1 &= t+f-\frac{q}{\tau}, \\ b_2 &= \frac{p}{\tau}+\frac{q}{\tau}+2, \\ b_3 &= t+f+1, \\ b_4 &= \alpha n+c+f-\frac{p}{\tau}-\frac{q}{\tau}, \\ b_5 &= \alpha n+c+f+2, \end{aligned}$$

and, as before, $t = \alpha(n-k-1)$.

First partial derivatives of E_{1j} The two relevant first derivatives of (4.69) are

$$E'_{1j,p} = \frac{c_{1,j:n}\alpha^2}{\tau} \sum_{k=0}^{j-2} (-1)^{j-k} \binom{j-2}{k} \frac{\Gamma(b_1)\Gamma(b_2)}{\Gamma(b_3)} \times \left\{ \sum_{m=0}^{\infty} \frac{\Gamma(b_3+m)\Gamma(b_4+m)}{\Gamma(b_5+m)\Gamma(b_1+1+m)} \left[\psi(b_2) - \psi(b_4+m)\right] \right\} \tag{4.70}$$

and

$$E'_{1j,c} = c_{1,j:n}\alpha^2 \sum_{k=0}^{j-2} (-1)^{j-k} \binom{j-2}{k} \frac{\Gamma(b_1)\Gamma(b_2)}{\Gamma(b_3)} \times \left\{ \sum_{m=0}^{\infty} \frac{\Gamma(b_3+m)\Gamma(b_4+m)}{\Gamma(b_5+m)\Gamma(b_1+1+m)} \left[\psi(b_4+m) - \psi(b_5+m)\right] \right\}. \tag{4.71}$$

Second partial derivatives of E_{1j} It follows that the second differentiation of (4.70) wrt q is

$$E''_{1j,pq} = \frac{c_{1,j:n}\alpha^2}{\tau^2} \sum_{k=0}^{j-2} (-1)^{j-k} \binom{j-2}{k} \frac{\Gamma(b_1)\Gamma(b_2)}{\Gamma(b_3)} \times \sum_{m=0}^{\infty} \frac{\Gamma(b_3+m)\Gamma(b_4+m)\left[\psi(b_2) - \psi(b_4+m)\right]}{\Gamma(b_5+m)\Gamma(b_1+1+m)} \times \left\{ \begin{array}{c} -\psi(b_1) + \psi(b_2) - \psi(b_4+m) \\ + \frac{\psi'(b_2)+\psi'(b_4+m)}{\psi(b_2)-\psi(b_4+m)} + \psi(b_1+1+m) \end{array} \right\} \tag{4.72}$$

and wrt f is

$$E''_{1j,pf} = \frac{c_{1,j:n}\alpha^2}{\tau} \sum_{k=0}^{j-2} (-1)^{j-k} \binom{j-2}{k} \frac{\Gamma(b_1)\Gamma(b_2)}{\Gamma(b_3)} \times \sum_{m=0}^{\infty} \frac{\Gamma(b_3+m)\Gamma(b_4+m)\left[\psi(b_2) - \psi(b_4+m)\right]}{\Gamma(b_5+m)\Gamma(b_1+1+m)} \times \left\{ \begin{array}{c} \psi(b_1) - \psi(b_3) + \psi(b_3+m) + \psi(b_4+m) \\ - \frac{\psi'(b_4+m)}{\psi(b_2)-\psi(b_4+m)} - \psi(b_5+m) - \psi(b_1+1+m) \end{array} \right\}. \tag{4.73}$$

Whereas the second differentiation of (4.71) wrt q is

$$E''_{1j,cq} = \frac{c_{1,j:n}\alpha^2}{\tau} \sum_{k=0}^{j-2} (-1)^{j-k} \binom{j-2}{k} \frac{\Gamma(b_1)\Gamma(b_2)}{\Gamma(b_3)} \times \sum_{m=0}^{\infty} \frac{\Gamma(b_3+m)\Gamma(b_4+m)\left[\psi(b_4+m) - \psi(b_5+m)\right]}{\Gamma(b_5+m)\Gamma(b_1+1+m)} \times \left\{ \begin{array}{c} -\psi(b_1) + \psi(b_2) - \psi(b_4+m) \\ - \frac{\psi'(b_4+m)}{\psi(b_4+m)-\psi(b_5+m)} + \psi(b_1+1+m) \end{array} \right\} \tag{4.74}$$

and wrt f is

$$
\begin{aligned}
E''_{1j,cf} &= c_{1,j:n}\alpha^2 \sum_{k=0}^{j-2} (-1)^{j-k} \binom{j-2}{k} \frac{\Gamma(b_1)\Gamma(b_2)}{\Gamma(b_3)} \times \\
& \sum_{m=0}^{\infty} \frac{\Gamma(b_3+m)\Gamma(b_4+m)\left[\psi(b_4+m)-\psi(b_5+m)\right]}{\Gamma(b_5+m)\Gamma(b_1+1+m)} \times \\
& \left\{ \begin{array}{c} \psi(b_1)-\psi(b_3)+\psi(b_3+m)+\psi(b_4+m) \\ +\frac{\psi'(b_4+m)-\psi'(b_5+m)}{\psi(b_4+m)-\psi(b_5+m)} - \psi(b_5+m)-\psi(b_1+1+m) \end{array} \right\}. \qquad (4.75)
\end{aligned}
$$

Expectations in (4.53) We can now replace p, c, q and f according to Table 4.6, and will see the above derivatives simplify greatly to give us expressions for the functions in (4.53). Replacing $(p, c, q, f) = (0,0,0,0)$, $(\tau, 1, \tau, 1)$, $(0, 0, \tau, 1)$, and $(\tau, 1, 0, 0)$ in (4.72) gives, respectively,

$$
\begin{aligned}
E\left[\ln X_{1:n} \ln X_{j:n}\right] &= \frac{c_{1,j:n}\alpha^2}{\tau^2} \sum_{k=0}^{j-2} \frac{(-1)^{j-k}\binom{j-2}{k}}{t} \sum_{m=0}^{\infty} \frac{\left[\psi(2)-\psi(\alpha n+m)\right]}{(\alpha n+m)(\alpha n+1+m)} \times \\
& \left\{ \begin{array}{c} -\psi(t)+\psi(2)-\psi(\alpha n+m) \\ +\frac{\psi'(2)+\psi'(\alpha n+m)}{\psi(2)-\psi(\alpha n+m)} + \psi(t+1+m) \end{array} \right\}, \qquad (4.76)
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{X_{1:n}^{\tau} \ln X_{1:n}}{1+X_{1:n}^{\tau}} \frac{X_{j:n}^{\tau} \ln X_{j:n}}{1+X_{j:n}^{\tau}}\right] &= \frac{6c_{1,j:n}\alpha^2}{\tau^2} \sum_{k=0}^{j-2} \frac{(-1)^{j-k}\binom{j-2}{k}}{t(t+1)} \times \\
& \sum_{m=0}^{\infty} \frac{(t+1+m)\Gamma(\alpha n+m)\left[\psi(4)-\psi(\alpha n+m)\right]}{\Gamma(\alpha n+4+m)} \times \\
& \left\{ \begin{array}{c} -\psi(t)+\psi(4)-\psi(\alpha n+m) \\ +\frac{\psi'(4)+\psi'(\alpha n+m)}{\psi(4)-\psi(\alpha n+m)} + \psi(t+1+m) \end{array} \right\},
\end{aligned}
$$

$$
\begin{aligned}
E\left[\ln X_{1:n} \frac{X_{j:n}^{\tau} \ln X_{j:n}}{1+X_{j:n}^{\tau}}\right] &= \frac{2c_{1,j:n}\alpha^2}{\tau^2} \sum_{k=0}^{j-2} \frac{(-1)^{j-k}\binom{j-2}{k}}{t(t+1)} \times \\
& \sum_{m=0}^{\infty} \frac{(t+1+m)\Gamma(\alpha n+m)\left[\psi\left(\frac{p}{\tau}+\frac{q}{\tau}+2\right)-\psi(\alpha n+m)\right]}{\Gamma(\alpha n+3+m)} \times \\
& \left\{ \begin{array}{c} -\psi(t)+\psi(3)-\psi(\alpha n+m) \\ +\frac{\psi'(3)+\psi'(\alpha n+m)}{\psi(3)-\psi(\alpha n+m)} + \psi(t+1+m) \end{array} \right\},
\end{aligned}
$$

and

$$E\left[\frac{X_{1:n}^{\tau}\ln X_{1:n}}{1+X_{1:n}^{\tau}}\ln X_{j:n}\right] = \frac{2c_{1,j:n}\alpha^2}{\tau^2}\sum_{k=0}^{j-2}\frac{(-1)^{j-k}\binom{j-2}{k}}{t}\times \sum_{m=0}^{\infty}\frac{\Gamma(\alpha n+m)\left[\psi(3)-\psi(\alpha n+m)\right]}{\Gamma(\alpha n+3+m)}\times \left\{\begin{array}{c} -\psi(t)+\psi(3)-\psi(\alpha n+m) \\ +\frac{\psi'(3)+\psi'(\alpha n+m)}{\psi(3)-\psi(\alpha n+m)}+\psi(t+1+m)\end{array}\right\},$$

in which the exact values for $\psi(a)$ and $\psi'(a)$, for $a=2,3,4$, can be found at Table 1.6.

The expressions for other expectations can be similarly obtained. Setting $(p,c,q,f)=(0,0,0,0)$ and $(\tau,1,0,0)$ in (4.73) gives

$$E\left[\ln X_{1:n}\ln(1+X_{j:n}^{\tau})\right] = \frac{c_{1,j:n}\alpha^2}{\tau}\sum_{k=0}^{j-2}\frac{(-1)^{j-k}\binom{j-2}{k}}{t}\times \left\{\sum_{m=0}^{\infty}\frac{\psi(2)-\psi(\alpha n+m)}{(\alpha n+m)(\alpha n+1+m)}\left(\frac{1}{t}+\frac{\psi'(\alpha n+m)}{\psi(2)-\psi(\alpha n+m)}\right)\right\}$$

and

$$E\left[\frac{X_{1:n}^{\tau}\ln X_{1:n}}{1+X_{1:n}^{\tau}}\ln(1+X_{j:n}^{\tau})\right] = \frac{2c_{1,j:n}\alpha^2}{\tau}\sum_{k=0}^{j-2}\frac{(-1)^{j-k}\binom{j-2}{k}}{t}\times \sum_{m=0}^{\infty}\frac{\Gamma(\alpha n+m)\left[\psi(3)-\psi(\alpha n+m)\right]}{\Gamma(\alpha n+3+m)}\times \left\{\frac{1}{t}-\psi(\alpha n+m)+\frac{\psi'(\alpha n+m)}{\psi(3)-\psi(\alpha n+m)}+\psi(\alpha n+3+m)\right\}$$

respectively. While setting $(p,c,q,f)=(0,0,0,0)$ and $(0,0,\tau,1)$ in (4.74) gives, in turn,

$$E\left[\ln(1+X_{1:n}^{\tau})\ln X_{j:n}\right] = \frac{c_{1,j:n}\alpha^2}{\tau}\sum_{k=0}^{j-2}\frac{(-1)^{j-k}\binom{j-2}{k}}{t}\times \sum_{m=0}^{\infty}\frac{\left[\psi(\alpha n+m)-\psi(\alpha n+2+m)\right]}{(\alpha n+m)(\alpha n+1+m)}\times \left\{\begin{array}{c}\psi(t)-\psi(2)+\psi(\alpha n+m) \\ +\frac{\psi'(\alpha n+m)}{\psi(\alpha n+m)-\psi(\alpha n+2+m)}-\psi(t+1+m)\end{array}\right\}$$

and

$$E\left[\ln(1+X_{1:n}^{\tau})\frac{X_{j:n}^{\tau}\ln X_{j:n}}{1+X_{j:n}^{\tau}}\right] = \frac{2c_{1,j:n}\alpha^2}{\tau}\sum_{k=0}^{j-2}\frac{(-1)^{j-k}\binom{j-2}{k}}{t(t+1)}\times \sum_{m=0}^{\infty}\frac{(t+1+m)\Gamma(\alpha n+m)\left[\psi(\alpha n+m)-\psi(\alpha n+3+m)\right]}{\Gamma(\alpha n+3+m)}\times \left\{\begin{array}{c}\psi(t)-\psi(3)+\psi(\alpha n+m) \\ +\frac{\psi'(\alpha n+m)}{\psi(\alpha n+m)-\psi(\alpha n+3+m)}-\psi(t+1+m)\end{array}\right\}.$$

Lastly, setting $(p, c, q, f) = (0, 0, 0, 0)$ in (4.75) yields

$$E\left[\ln(1 + X_{1:n}^{\tau}) \ln(1 + X_{j:n}^{\tau})\right]$$

as

$$c_{1,j:n}\alpha^2 \sum_{k=0}^{j-2} \frac{(-1)^{j-k} \binom{j-2}{k}}{t} \sum_{m=0}^{\infty} \frac{\psi(\alpha n + m) - \psi(\alpha n + 2 + m)}{(\alpha n + m)(\alpha n + 1 + m)} \times$$
$$\left\{-\frac{1}{t} + \psi(\alpha n + m) + \frac{\psi'(\alpha n + m) - \psi'(\alpha n + 2 + m)}{\psi(\alpha n + m) - \psi(\alpha n + 2 + m)} - \psi(\alpha n + 2 + m)\right\}.$$

Joint Expectations of $g(X_{i:n})$ and $h(X_{j:n})$

As indicated at Section 4.3.2.1, having found the expressions for (4.53) in terms of $X_{1:n}$ and $X_{j:n}$ for various combinations of a, b, c, d, e, f, p, q from the derivatives method, we are now in the position to exploit the recurrence relationship for order statistics, given at (1.51), to obtain the corresponding expressions in terms of $X_{i:n}$ and $X_{j:n}$ $(2 \le i < j \le n)$. For instance,

$$E[\ln X_{i:n} \ln X_{j:n}] = \frac{n!}{(j-i-1)!} \sum_{s=1}^{i} \sum_{t=0}^{i-s} \left[\begin{array}{c} \frac{(-1)^{s+t-1}(n+t-j)!(s+j-i-2)!}{t!(n-j)!(i-t-s)!(s-1)!(n+t+s-i)!} \times \\ E[\ln X_{1:n-i+s+t} \ln X_{j-i+s:n-i+s+t}] \end{array} \right],$$

in which $E[\ln X_{1:n} \ln X_{j:n}]$ has been given at (4.76), and so forth.

Some Numerical Details and Discussion

The work above is from a theoretical viewpoint only, and we should seek some reassurance that this theory is in agreement with practice, as represented by simulations. We continue to use $\alpha = 4, \tau = 3$, and show in Figures 4.21 to 4.29 the theoretical, calculated from both direct and derivatives methods, and simulated values for each specific expectation in (4.53), for all combinations of i and j such that $1 \le i < j \le n = 10$. We see there is very little difference between the three sets of values, for all 55 pairs of (i, j) considered here. A more detailed comparison is given at Table 4.7 for $E\left[\ln(1 + X_{i:n}^{\tau}) \frac{X_{j:n}^{\tau} \ln X_{j:n}}{1+X_{j:n}^{\tau}}\right]$ evaluated at various i, j and n, which leads to the same conclusion. We remark that these checks should be extended to cover other values of α, τ and sample sizes, although the computation time will depend on the computer resources available.

4.4 Chapter Summary and Conclusions

In this chapter, we have obtained expressions for various expectations and joint expectations of order statistics, generally of the forms at (4.1) and (4.2). The derivation of these expectations is motivated by the study of the correlations between the complete and the Type II censored MLEs, which, as we will show in Chapter 5, involves the multiplication of

i, j	n			
	25	50	100	1000
$0.1n, 0.2n$: direct	-0.0011	-0.0014	-0.0014	-0.0014
: deriv.	-0.0011	-0.0014	-0.0014	-0.0014
: simul.	-0.0011	-0.0014	-0.0014	-0.0014
$0.3n, 0.4n$: direct	-0.0064	-0.0070	-0.0071	-0.0071
: deriv.	-0.0064	-0.0070	-0.0071	-0.0071
: simul.	-0.0063	-0.0070	-0.0071	-0.0071
$0.5n, 0.6n$: direct	-0.0143	-0.0156	-0.0158	-0.0163
: deriv.	-0.0143	-0.0156	-0.0158	-0.0163
: simul.	-0.0143	-0.0156	-0.0158	-0.0163
$0.7n, 0.8n$: direct	-0.0197	-0.0223	-0.0228	-0.0233
: deriv.	-0.0197	-0.0223	-0.0228	-0.0233
: simul.	-0.0203	-0.0223	-0.0228	-0.0233
$0.9n, 1.0n$: direct	0.0565	0.0972	0.1367	0.2758
: deriv.	0.0565	0.0972	0.1367	0.2758
: simul.	0.0573	0.0971	0.1388	0.2775

Table 4.7: Numerical comparison of $E\left[\ln(1+X_{i:n}^{\tau})\frac{X_{j:n}^{\tau}\ln X_{j:n}}{1+X_{j:n}^{\tau}}\right]$ for various i, j and n, for Burr data generated with $\alpha = 4, \tau = 3$.

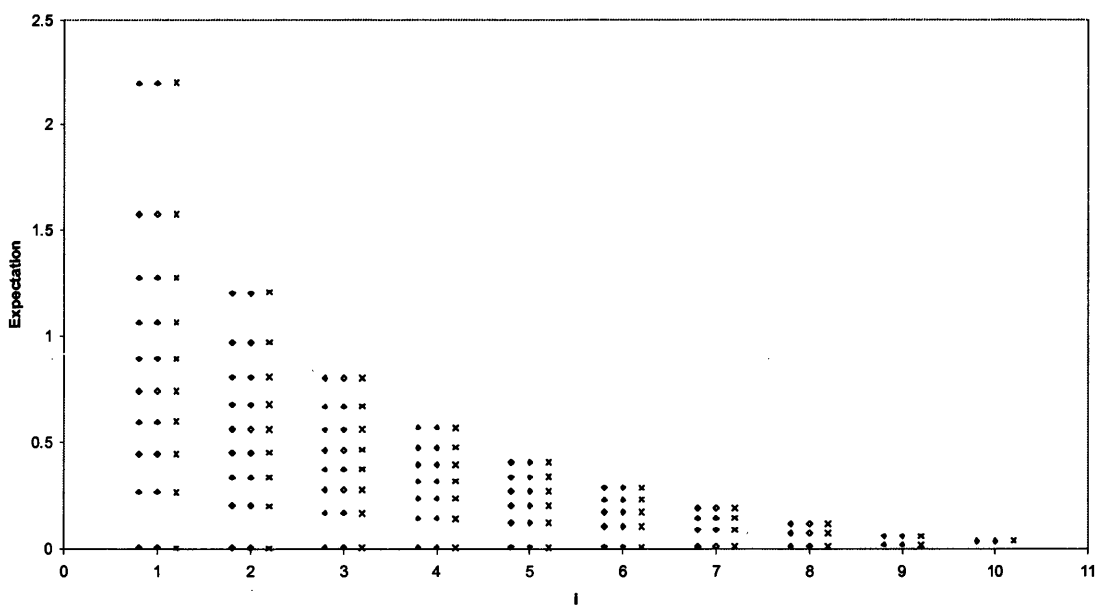

Figure 4.21: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E[\ln X_{i:n} \ln X_{j:n}]$ for all $1 \le i < j \le n$, for $n = 10, \alpha = 4, \tau = 3$.

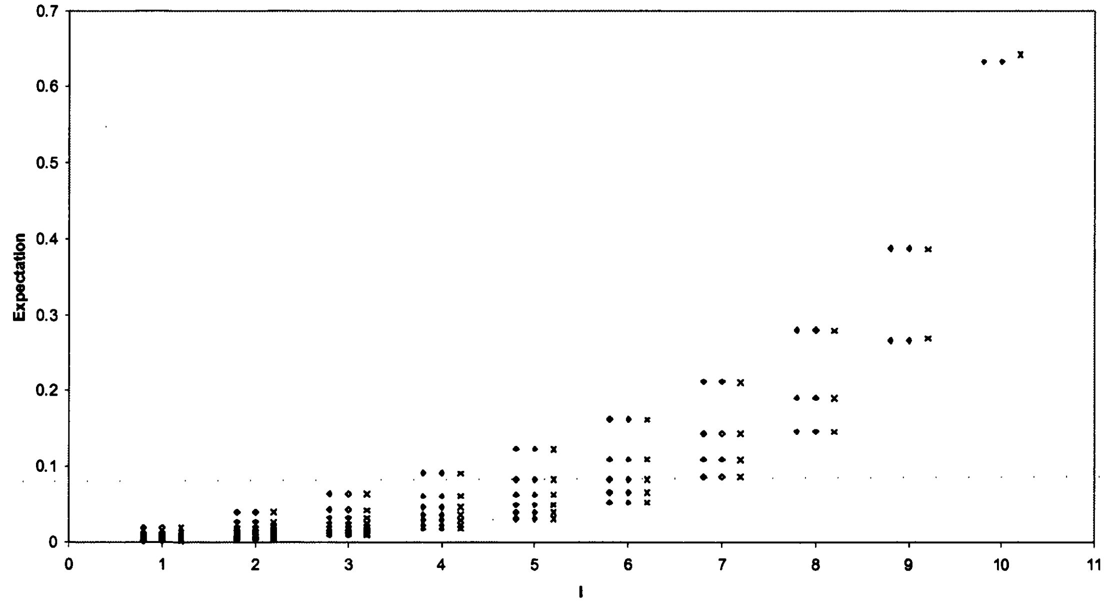

Figure 4.22: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E\left[\ln(1+X_{1:n}^{\tau})\ln(1+X_{j:n}^{\tau})\right]$ for all $1 \le i < j \le n$, for $n = 10, \alpha = 4, \tau = 3$.

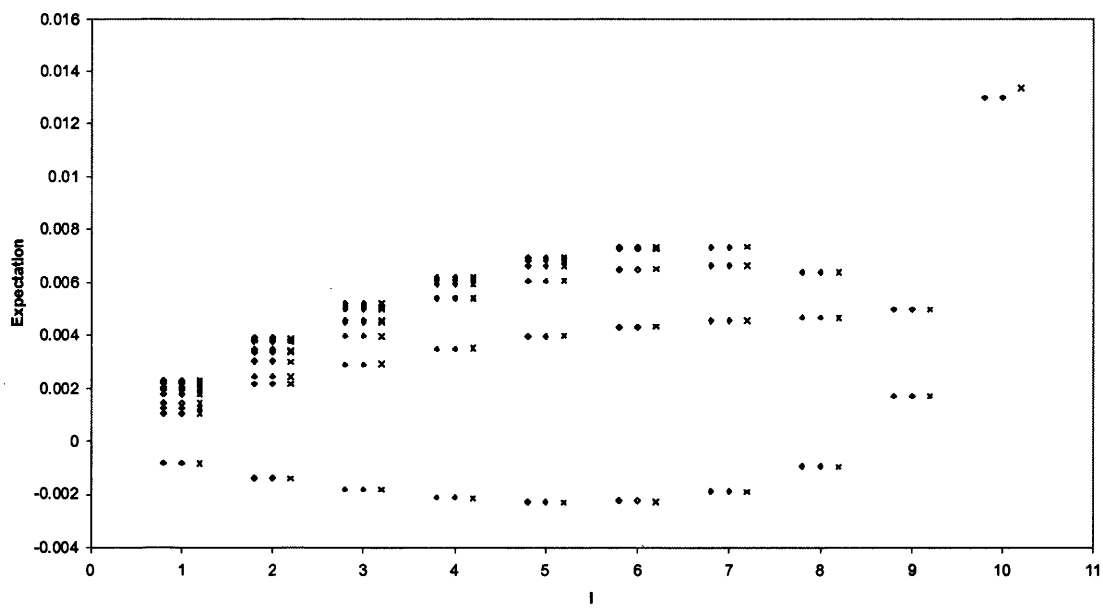

Figure 4.23: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E\left[\frac{X_{1:n}^{\tau}\ln X_{1:n}}{1+X_{1:n}^{\tau}}\frac{X_{j:n}^{\tau}\ln X_{j:n}}{1+X_{j:n}^{\tau}}\right]$ for all $1 \le i < j \le n$, for $n = 10, \alpha = 4, \tau = 3$.

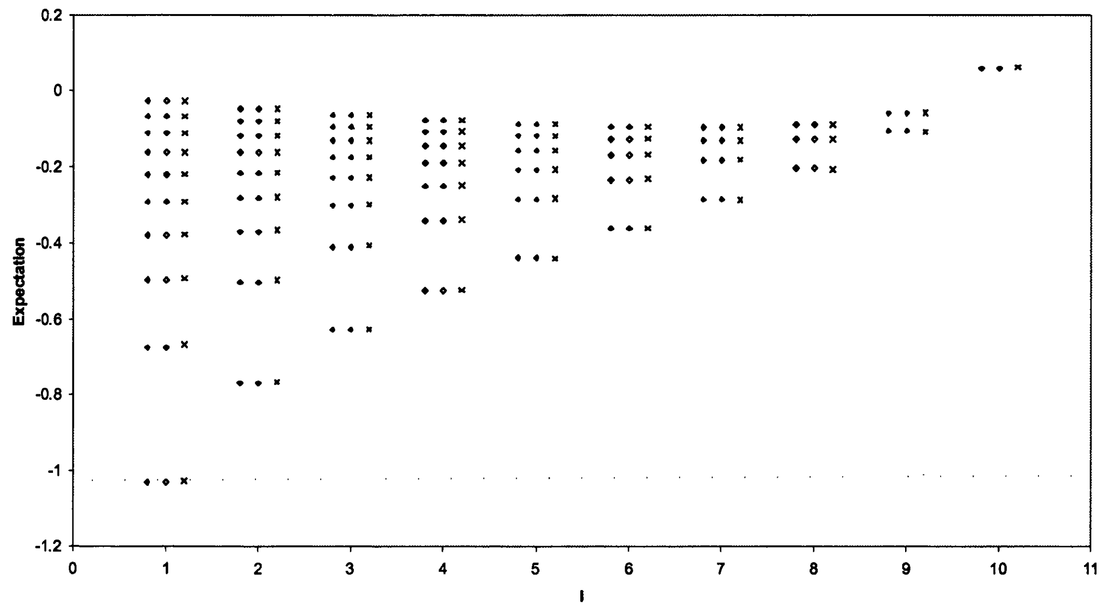

Figure 4.24: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E\left[\ln X_{1:n} \ln(1+X_{j:n}^{\tau})\right]$ for all $1 \le i < j \le n$, for $n = 10, \alpha = 4, \tau = 3$.

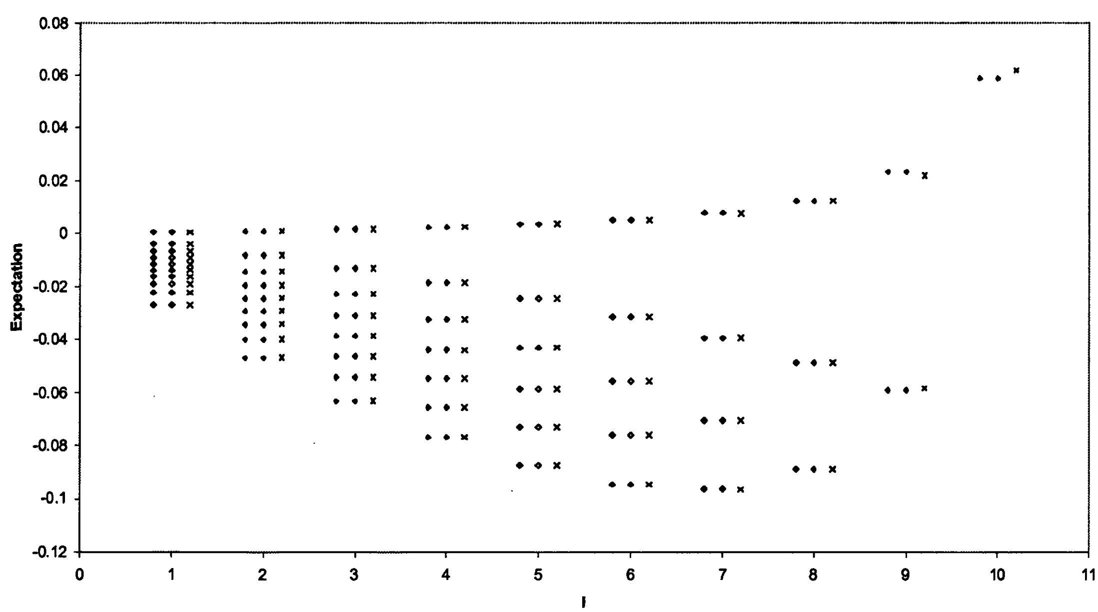

Figure 4.25: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E\left[\ln(1+X_{1:n}^{\tau}) \ln X_{j:n}\right]$ for all $1 \le i < j \le n$, for $n = 10, \alpha = 4, \tau = 3$.

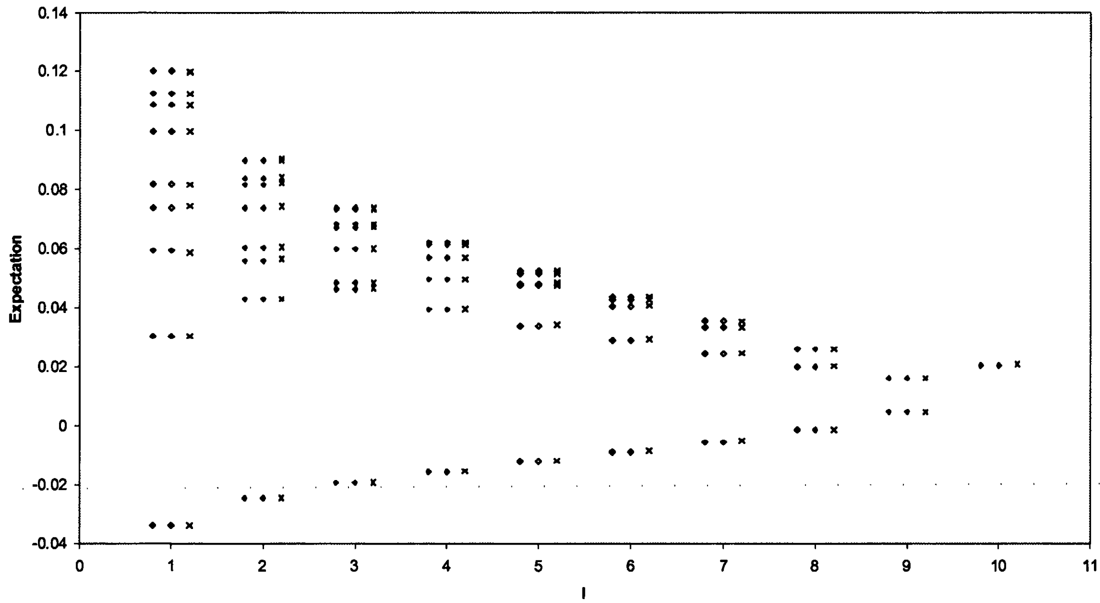

Figure 4.26: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E\left[\ln X_{1:n}\frac{X_{j:n}^{\tau}\ln X_{j:n}}{1+X_{j:n}^{\tau}}\right]$ for all $1 \le i < j \le n$, for $n = 10, \alpha = 4, \tau = 3$.

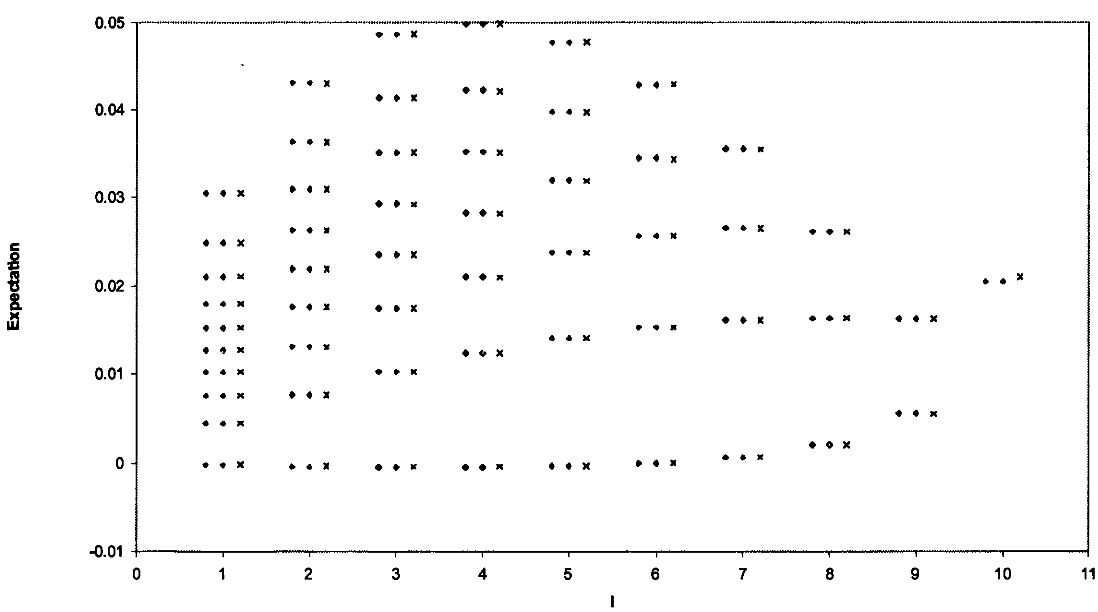

Figure 4.27: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E\left[\frac{X_{1:n}^{\tau}\ln X_{1:n}}{1+X_{1:n}^{\tau}}\ln X_{j:n}\right]$ for all $1 \le i < j \le n$, for $n = 10, \alpha = 4, \tau = 3$.

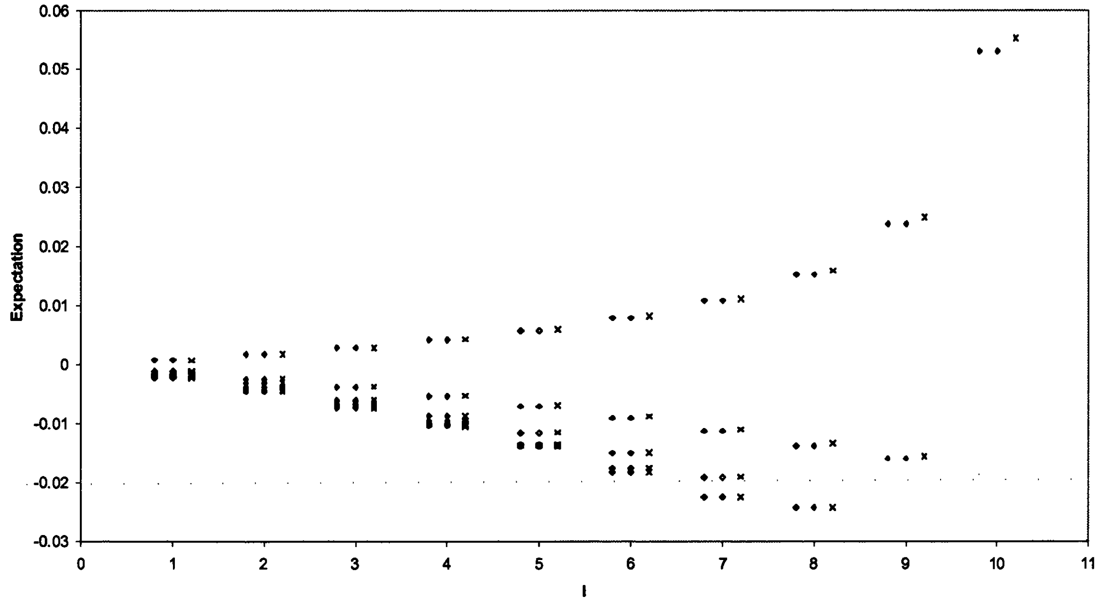

Figure 4.28: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E\left[\ln(1+X_{1:n}^{\tau})\frac{X_{j:n}^{\tau}\ln X_{j:n}}{1+X_{j:n}^{\tau}}\right]$ for all $1 \le i < j \le n$, for $n = 10, \alpha = 4, \tau = 3$.

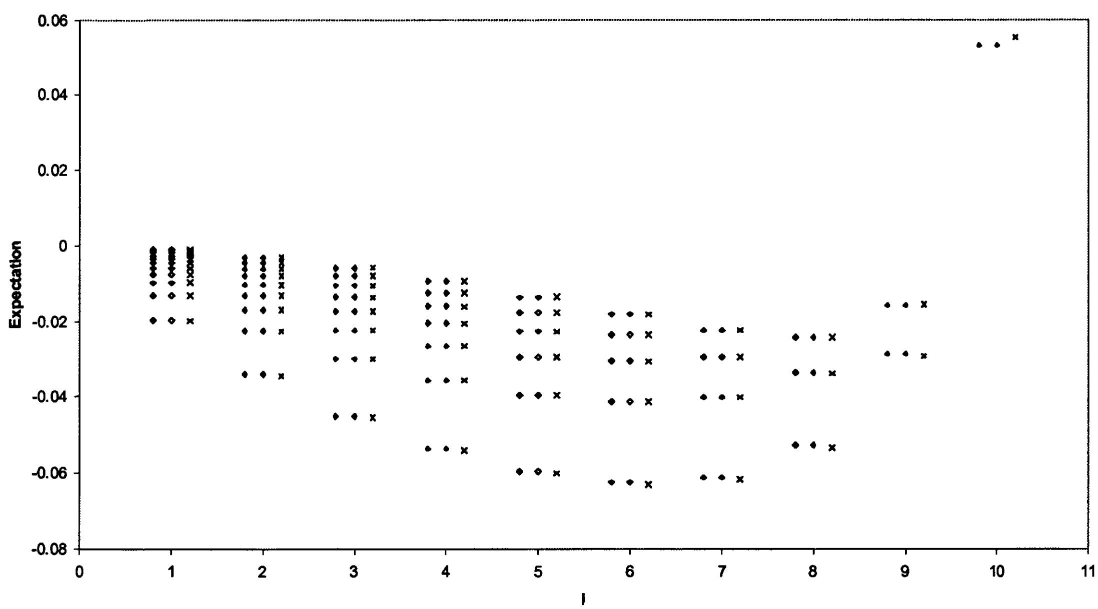

Figure 4.29: Theoretical (direct ♦, derivatives ◊) and simulated (×) values of $E\left[\frac{X_{1:n}^{\tau}\ln X_{1:n}}{1+X_{1:n}^{\tau}}\ln(1+X_{j:n}^{\tau})\right]$ for all $1 \le i < j \le n$, for $n = 10, \alpha = 4, \tau = 3$.

the two sets of score functions, and hence explaining the complexity in the formats of these expectations.

In summary, this chapter has involved a considerable amount of algebra. For each of the distributions considered, we first employed the direct method; this involved (double) integrations of some complex functions, including the exponential integrals and the hypergeometric series, and we noted that, however, certain integrals of the joint expectations from the Burr distribution could not be solved directly. We then considered the derivatives method; the technique is relatively straightforward, and has shown more generalisable in dealing with the logarithms and/or powers of order statistics; more importantly, it has provided expressions for all the joint expectations needed for the Burr distribution. Nevertheless, despite there being more complicated functions involved in the direct method, we note that this approach has generally consumed less computation time than the derivatives method when implemented in Mathematica, and hence is more useful in practical terms.

We have shown that the theoretical results agree with the behaviour observed in simulation experiments for various combinations of order statistics, parameter values and sample size; despite few computational problems for large sample sizes, we have covered most sample sizes and ranges of censoring encountered in practice, but we remark that results for larger sample sizes are possible with the computer resources available elsewhere. Most importantly, the agreement between theory and simulations indicates that our formulae can be employed as a building block in future evaluations. Therefore, we are now in the position to consider the link between the final and interim estimators.

Chapter 5

Correlations Between Final and Interim Estimates of Parameters and Percentiles

5.1 Introduction

As we have previously discussed, from a statistical perspective, the analysis of the complete data set is generally to be preferred (under complete sampling, the data available for analysis simply consisted of n independent failure times), but, in practice, some censoring - such as Type I or Type II - is often inevitable. The aim in this chapter is to establish a method to measure the precision of Type II censored MLEs, calculated at censoring level r, in estimating the complete MLEs.

Following the observations from Chapter 2, for example, as seen in Table 1.2 for the ball bearings data, we may wish to assess how useful are the MLEs calculated at $r = 8$

$$\widehat{\theta}_8 = 67.6415, \quad \widehat{\beta}_8 = 3.2280, \quad \widehat{B}_{0.1,8} = 33.6860$$

in predicting the complete estimates

$$\widehat{\theta} = 81.8783, \quad \widehat{\beta} = 2.1021, \quad \widehat{B}_{0.1} = 28.0694$$

which are obtained if all the $n = 23$ items were left to fail. We may also wish to quantify the increase in precision obtained on taking $r = 16$, where we have seen that the resultant estimates

$$\widehat{\theta}_{16} = 76.6960, \quad \widehat{\beta}_{16} = 2.4695, \quad \widehat{B}_{0.1,16} = 30.8329$$

are more consistent with final values than with $r = 8$. Because the percentiles feature times at which specified proportions of items fail, it is particularly relevant in practical applications to consider the agreement between $\widehat{B}_{0.1,r}$ and $\widehat{B}_{0.1}$, and the extent to which

$\widehat{B}_{0.1,r}$ can be regarded as a reliable guide to $\widehat{B}_{0.1}$.

Furthermore, the series of scatter plots of final estimates (of parameters or $B_{0.1}$) against interim estimates presented in Chapter 2 (see, for instance, Figures 2.7 to 2.11 under the Weibull analysis), seem to suggest a reasonably strong link between the two sets of estimates. Hence, it is of interest to consider the extent to which the final estimates are consistent with earlier estimates, and the rate at which interim estimates converge on their final values; more generally, we would like to determine the precision with which we can make statements on final estimates, based on interim estimates. We focus on the conditional distributions of final MLEs given interim counterparts; if these are Normal - as is the case asymptotically - then, in turn, we require the covariances of final and interim MLEs, equivalently, the study of the correlations between the two sets of MLEs. The classical asymptotic approach uses the relationship between the MLEs, the EFI matrix and the score vector. Chua & Watkins (2007) derive general expressions for correlations of exponential MLEs, and state (but do not prove) similar results for Weibull MLEs. Chua & Watkins (2008a,b) further extend this work to give a formula for correlation between interim and final estimates of Weibull percentiles. This chapter builds on these preliminaries, and presents the extension to the Burr distribution. Some discussion on the corresponding analysis of reliability data under Type I censoring are given by Finselbach & Watkins (2006) and Finselbach (2007).

We begin by showing that, for large samples, the study of the correlations between final and interim MLEs of parameters can be transformed into a study of the correlations between final and interim score functions, and can be further extended to the analysis of the precision in a sequence of Type II censored estimates of $B_{0.1}$, as an estimate to $\widehat{B}_{0.1}$. In Section 5.2, we assume that the lifetimes follow the negative exponential distribution, and present asymptotic 95% confidence limits for the final estimate given the interim estimate. This analysis uses results from the theory of order statistics, and hence exploits the familiar and extremely powerful lack-of-memory property of this distribution. These asymptotic results are then validated for various sample sizes and censoring fractions using simulation experiments. We then give details of how this analysis generalises to the Weibull (Section 5.3) and Burr (Section 5.4) distributions. In Section 5.5, we briefly consider some real life applications of this work.

5.1.1 Theoretical Considerations

The asymptotic theory of maximum likelihood, as outlined, for example, in Cox & Hinkley (1974) and Bain & Engelhardt (1991), implies that the asymptotic relationship between the MLE, the EFI matrix and the score vector is

$$(\widehat{\boldsymbol{\pi}}_r - \boldsymbol{\pi}) \simeq \mathbf{A}_r^{-1}\mathbf{U}_r, \tag{5.1}$$

which also covers the case $r = n$. Hence, the correlation between final $(r = n)$ and interim $(r < n)$ estimators of model parameters, written as

$$Corr\{\widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\pi}}_r\} = Corr\{\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}, \widehat{\boldsymbol{\pi}}_r - \boldsymbol{\pi}\},$$

can be approximated by

$$Corr\left\{\mathbf{A}^{-1}\mathbf{U}, \mathbf{A}_r^{-1}\mathbf{U}_r\right\}$$

(so that, where necessary, correlations can be found via the usual standardisations). With two or more parameters, it will prove more convenient to consider covariances; it follows that

$$Cov(\widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\pi}}_r) \simeq Cov\left(\mathbf{A}^{-1}\mathbf{U}, \mathbf{A}_r^{-1}\mathbf{U}_r\right) = \mathbf{A}^{-1}Cov(\mathbf{U}, \mathbf{U}_r)\,\mathbf{A}_r^{-1}. \tag{5.2}$$

We refer to Chapter 2 for the expressions of $\mathbf{A}^{-1}$ and $\mathbf{A}_r^{-1}$ for specific distributions, and due to the regularity condition, we will only require to take expectation on the product of $\mathbf{U}$ and $\mathbf{U}_r$; this, of course, then ties in with the various forms of moments and product moments of order statistics we have already derived in Chapter 4.

We are also interested in the agreement between $\widehat{B}_{0.1,r}$ and its counterpart for complete samples. Since we have a linear approximation to $B_{0.1}$ in terms of $(\widehat{\boldsymbol{\pi}}_r - \boldsymbol{\pi})$ in (2.3), the study of the correlation between the final and interim estimates of $B_{0.1}$ will also depend on the asymptotic relationship in (5.1); we have, from (2.3) with, first, $r\,(< n)$, and then with $r = n$,

$$\begin{aligned} Corr\left\{\widehat{B}_{0.1}, \widehat{B}_{0.1,r}\right\} &\simeq Corr\left\{B_{0.1} + \mathbf{b}'_{\boldsymbol{\pi}}(\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}), B_{0.1} + \mathbf{b}'_{\boldsymbol{\pi}}(\widehat{\boldsymbol{\pi}}_r - \boldsymbol{\pi})\right\} \\ &\simeq Corr\left\{\mathbf{b}'_{\boldsymbol{\pi}}\widehat{\boldsymbol{\pi}}, \mathbf{b}'_{\boldsymbol{\pi}}\widehat{\boldsymbol{\pi}}_r\right\}, \end{aligned}$$

so that, by first principles, the required correlation may be approximated by

$$\frac{\mathbf{b}'_{\boldsymbol{\pi}}Cov(\widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\pi}}_r)\,\mathbf{b}_{\boldsymbol{\pi}}}{\sqrt{\mathbf{b}'_{\boldsymbol{\pi}}Var(\widehat{\boldsymbol{\pi}})\,\mathbf{b}_{\boldsymbol{\pi}}} \times \sqrt{\mathbf{b}'_{\boldsymbol{\pi}}Var(\widehat{\boldsymbol{\pi}}_r)\,\mathbf{b}_{\boldsymbol{\pi}}}} = \frac{\mathbf{b}'_{\boldsymbol{\pi}}\mathbf{A}^{-1}Cov(\mathbf{U}, \mathbf{U}_r)\,\mathbf{A}_r^{-1}\mathbf{b}_{\boldsymbol{\pi}}}{\sqrt{\mathbf{b}'_{\boldsymbol{\pi}}\mathbf{A}^{-1}\mathbf{b}_{\boldsymbol{\pi}}} \times \sqrt{\mathbf{b}'_{\boldsymbol{\pi}}\mathbf{A}_r^{-1}\mathbf{b}_{\boldsymbol{\pi}}}}. \tag{5.3}$$

5.2 Correlation in the Exponential Distribution

5.2.1 Link Between $\widehat{\theta}$ and $\widehat{\theta}_r$

From our discussion of the exponential distribution in Section 2.2, $\widehat{\theta}_r$, given in (2.8), is the minimum variance unbiased estimator of θ so that factorisation of the score function gives

$$\sqrt{\frac{r}{\theta^2}}(\widehat{\theta}_r - \theta) = \sqrt{\frac{\theta^2}{r}}\frac{dl_r}{d\theta};$$

this is one version of the standard relationship (5.1) between the MLE, the EFI and the score function. We have

$$Corr(\widehat{\theta}, \widehat{\theta}_r) = Corr\left(\sqrt{\frac{n}{\theta^2}}(\widehat{\theta} - \theta), \sqrt{\frac{r}{\theta^2}}(\widehat{\theta}_r - \theta)\right)$$

which also equal to

$$Corr\left(\sqrt{\frac{\theta^2}{n}}\frac{dl}{d\theta}, \sqrt{\frac{\theta^2}{r}}\frac{dl_r}{d\theta}\right) = \frac{\theta^2}{\sqrt{nr}}Cov\left(\frac{dl}{d\theta}, \frac{dl_r}{d\theta}\right). \tag{5.4}$$

Then, via the usual regularity conditions, this is

$$\begin{aligned} \frac{\theta^2}{\sqrt{nr}}E\left[\frac{dl}{d\theta}\frac{dl_r}{d\theta}\right] &= \frac{\theta^2}{\sqrt{nr}}E\left[\left(-\frac{n}{\theta} + \frac{S}{\theta^2}\right)\left(-\frac{r}{\theta} + \frac{S_r}{\theta^2}\right)\right] \\ &= \frac{\theta^2}{\sqrt{nr}}\left\{\frac{nr}{\theta^2} - \frac{n}{\theta^3}E\left[S_r\right] - \frac{r}{\theta^3}E\left[S\right] + \frac{1}{\theta^4}E\left[SS_r\right]\right\} \end{aligned}$$

which involves the single and product moments of $X_{i:n}$. It is straightforward to obtain $E[S] = n\theta$ and $E[S_r] = r\theta$. In considering $E[SS_r]$, we may write

$$SS_r = (S - S_r + S_r)\, S_r = (S - S_r)\, S_r + S_r^2$$

in which it is convenient to express $S - S_r$ in terms of differences of order statistics:

$$\begin{aligned} S - S_r &= \sum_{i=1}^{n} X_{i:n} - \left\{\sum_{i=1}^{r} X_{i:n} + (n-r)X_{r:n}\right\} \\ &= \sum_{i=r+1}^{n} X_{i:n} - (n-r)X_{r:n} \\ &= \sum_{i=r+1}^{n} \left(X_{i:n} - X_{r:n}\right). \end{aligned}$$

Since S_r depends only on the first r order statistics while $S - S_r$ features differences from the r^{th} order statistic onwards, the lack-of-memory property implies that $S - S_r$ and S_r are independent. We thus obtain

$$E\left[SS_r\right] = E\left[S - S_r\right]E\left[S_r\right] + E\left[S_r^2\right] = r(n+1)\theta^2 \tag{5.5}$$

so that $Corr(\widehat{\theta}, \widehat{\theta}_r)$ becomes

$$\frac{\theta^2}{\sqrt{nr}}\left\{\frac{nr}{\theta^2} - \frac{n}{\theta^3} \times r\theta - \frac{r}{\theta^3} \times n\theta + \frac{1}{\theta^4} \times r(n+1)\theta^2\right\} = \sqrt{\frac{r}{n}}.$$

Therefore, the positive correlation between the complete and censored MLEs depends directly on the number of failures; increasing r will give a higher correlation value, and when $r = n$ we see that $Corr(\widehat{\theta}, \widehat{\theta}_r) = 1$, as expected.

A Remark

It is appropriate to here remark an useful observation from the above analysis, namely that $Cov\left(\frac{dl}{d\theta}, \frac{dl_r}{d\theta}\right)$ is, in fact, given by $Var\left(\frac{dl_r}{d\theta}\right)$; from regularity conditions,

$$\begin{aligned} Cov\left(\frac{dl}{d\theta}, \frac{dl_r}{d\theta}\right) &= E\left[\frac{dl}{d\theta}\frac{dl_r}{d\theta}\right] \\ &= E\left[\left(\frac{dl}{d\theta} - \frac{dl_r}{d\theta} + \frac{dl_r}{d\theta}\right)\frac{dl_r}{d\theta}\right] \\ &= E\left[\left(\frac{dl}{d\theta} - \frac{dl_r}{d\theta}\right)\frac{dl_r}{d\theta}\right] + E\left[\left(\frac{dl_r}{d\theta}\right)^2\right] \end{aligned}$$

and we see

$$\frac{dl}{d\theta} - \frac{dl_r}{d\theta} = -\frac{n-r}{\theta} + \frac{1}{\theta^2}(S - S_r)$$

is, again, via the lack-of-memory property, independent of $\frac{dl_r}{d\theta}$. We thus have

$$E\left[\left(\frac{dl}{d\theta} - \frac{dl_r}{d\theta}\right)\frac{dl_r}{d\theta}\right] = 0$$

so that

$$Cov\left(\frac{dl}{d\theta}, \frac{dl_r}{d\theta}\right) = Var\left(\frac{dl_r}{d\theta}\right) = r\theta^{-2}. \tag{5.6}$$

From this, we see that (5.4) reduces to

$$Corr(\widehat{\theta}, \widehat{\theta}_r) = \frac{\theta^2}{\sqrt{nr}} Var\left(\frac{dl_r}{d\theta}\right) = \frac{\theta^2}{\sqrt{nr}} r\theta^{-2} = \sqrt{\frac{r}{n}},$$

exactly as found from first principles.

A Possible Generalisation

In previous chapter, we have obtained the expressions required for the study of the correlations between complete and censored score functions for the Weibull and Burr distributions, and have seen a considerable amount of algebra being involved, even in the case of the transformed variable Z which has a standard exponential distribution. From the above remark, a possible generalisation to (5.6) is

$$Cov(\mathbf{U}, \mathbf{U}_r) = \mathbf{A}_r, \tag{5.7}$$

in which the covariance between final and interim scores is given by the censored EFI matrix. If this result holds, then the evaluation of the correlation between the two sets of MLEs would become much more straightforward; for now, the covariance at (5.2) would then simplify to

$$Cov(\widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\pi}}_r) \simeq \mathbf{A}^{-1}\mathbf{A}_r\mathbf{A}_r^{-1} = \mathbf{A}^{-1}, \tag{5.8}$$

suggesting that the correlation between complete and censored MLEs follows immediately from the complete and censored EFI matrices, rather than from the expectations of the forms at (4.1) and (4.2). In practical terms, this would also imply a substantial reduction in computational time for $Corr\{\widehat{\pi}, \widehat{\pi}_r\}$ in Mathematica.

Furthermore, another consequence of this result would be that we could write

$$\mathbf{b}'_{\pi}\mathbf{A}^{-1}Cov(\mathbf{U}, \mathbf{U}_r)\,\mathbf{A}_r^{-1}\mathbf{b}_{\pi} = \mathbf{b}'_{\pi}\mathbf{A}^{-1}\mathbf{b}_{\pi},$$

indicating

$$Cov\left(\widehat{B}_{0.1}, \widehat{B}_{0.1,r}\right) = Var\left(\widehat{B}_{0.1}\right) \tag{5.9}$$

so that we would then be able to write

$$Corr\left\{\widehat{B}_{0.1}, \widehat{B}_{0.1,r}\right\} \simeq \frac{Var\left(\widehat{B}_{0.1}\right)}{\sqrt{Var\left(\widehat{B}_{0.1}\right)} \times \sqrt{Var\left(\widehat{B}_{0.1,r}\right)}} = \sqrt{\frac{\mathbf{b}'_{\pi}\mathbf{A}^{-1}\mathbf{b}_{\pi}}{\mathbf{b}'_{\pi}\mathbf{A}_r^{-1}\mathbf{b}_{\pi}}}. \tag{5.10}$$

Hence, as with $Corr\{\widehat{\pi}, \widehat{\pi}_r\}$, correlation of the two sets of percentiles would follow immediately from the two sets of EFI matrices. We will later check (5.8) and (5.9) for the Weibull and Burr distributions, and show that, however, only limited analytical progress is possible.

5.2.2 Link between $\widehat{B}_{0.1}$ and $\widehat{B}_{0.1,r}$

From (5.3), $Corr\left(\widehat{B}_{0.1}, \widehat{B}_{0.1,r}\right)$ is exactly

$$\frac{(-\ln 0.9)^2\, Cov\left(\widehat{\theta}, \widehat{\theta}_r\right)}{\sqrt{(-\ln 0.9)^2\, Var\left(\widehat{\theta}\right)} \times \sqrt{(-\ln 0.9)^2\, Var\left(\widehat{\theta}_r\right)}} = Corr\left(\widehat{\theta}, \widehat{\theta}_r\right) = \sqrt{\frac{r}{n}},$$

as we would have expected from the linear relation between $B_{0.1}$ and θ in (1.29). Otherwise, we can use (5.10) to obtain this result; we first need to confirm that (5.9) holds for this distribution:

$$Cov\left(\widehat{B}_{0.1}, \widehat{B}_{0.1,r}\right) = (-\ln 0.9)^2\, Cov\left(\widehat{\theta}, \widehat{\theta}_r\right)$$

which equal to

$$(-\ln 0.9)^2\, Corr\left(\widehat{\theta}, \widehat{\theta}_r\right)\sqrt{Var\left(\widehat{\theta}\right) Var\left(\widehat{\theta}_r\right)} = (-\ln 0.9)^2 \sqrt{\frac{r}{n}}\sqrt{\frac{\theta^2}{r}\frac{\theta^2}{n}} = Var\left(\widehat{B}_{0.1}\right).$$

Consequently, from (5.10),

$$Corr\left(\widehat{B}_{0.1}, \widehat{B}_{0.1,r}\right) = \sqrt{\frac{\theta^2 (-\ln 0.9)^2 / n}{\theta^2 (-\ln 0.9)^2 / r}} = \sqrt{\frac{r}{n}},$$

r	Theory	n					
	$(=\sqrt{\frac{r}{n}})$	25	50	100	1000	2500	5000
$0.2n$	.4472	.4582	.4412	.4435	.4402	.4486	.4556
$0.4n$	.6325	.6414	.6401	.6256	.6291	.6336	.6352
$0.6n$	.7746	.7820	.7795	.7746	.7740	.7746	.7767
$0.8n$	.8944	.8959	.8985	.8973	.8945	.8934	.8950
$1.0n$	1	1	1	1	1	1	1

Table 5.1: Theoretical and simulated values of $Corr(\hat{\theta}, \hat{\theta}_r)$ for various r, n, for exponential data generated with $\theta = 100$.

as we required. Therefore, (and as previously), a numerical study on the link between $\widehat{\theta}$ and $\widehat{\theta}_r$ essentially covers all percentiles.

5.2.3 Numerical Results

We next validate these theoretical results with simulations. We revisit the simulated observations of $\widehat{\theta}_r$ obtained in Section 2.2.4, and summarise in Table 5.1 the theoretical and observed values of $Corr(\widehat{\theta}, \widehat{\theta}_r)$ for these 10^4 estimates. We see good agreement between theory and practice for varying censoring proportions and sample sizes. More specifically, the theoretical correlation coefficients found here for $n = 50$ and 1000 are consistent with the behaviour observed in the scatter plots at Figures 2.2 and 2.3.

5.2.4 Confidence Limits Considerations

We can now consider the precision with which we can make statements on final estimates, given the interim estimates. In particular, we can compute the 95% confidence limits for $\widehat{\theta}$ based on $\widehat{\theta}_r$. The asymptotic Normality of MLE implies that, for large samples, $\widehat{\theta} - \widehat{\theta}_r$ is also asymptotically Normal, with zero mean and variance, based on the above correlation, given by

$$\frac{\theta^2(n-r)}{nr}.$$

As a result, the 95% confidence intervals for $\widehat{\theta}$ given $\widehat{\theta}_r$ is

$$\widehat{\theta}_r \pm 1.96\theta\sqrt{\frac{n-r}{nr}}. \tag{5.11}$$

Now, when running simulations, we know the true parameter values and can also obtain a set of ML estimates for each of, say, 10^4 replications. However, in most practical circumstances, we will not know the true parameter values, but estimate them from the data; thus, we substitute $\widehat{\theta}_r$ for θ, giving

$$\widehat{\theta}_r \pm 1.96\widehat{\theta}_r\sqrt{\frac{n-r}{nr}}. \tag{5.12}$$

r	10	20	30	40	49
$\widehat{\theta}_r$	67.6000	104.9000	114.0100	112.1150	104.8898
$\widehat{sd}(\widehat{\theta}-\widehat{\theta}_r)$	19.0713	18.0452	12.9617	7.5973	0
$\widehat{B}_{0.1,r}$	7.1224	11.0523	12.0122	11.8125	11.0512
$\widehat{sd}(\widehat{B}_{0.1}-\widehat{B}_{0.1,r})$	2.0094	1.9013	1.3656	0.8005	0

Table 5.2: Standard deviations of $\hat{\theta}-\hat{\theta}_r$ and $\hat{B}_{0.1}-\hat{B}_{0.1,r}$ for the failure times data.

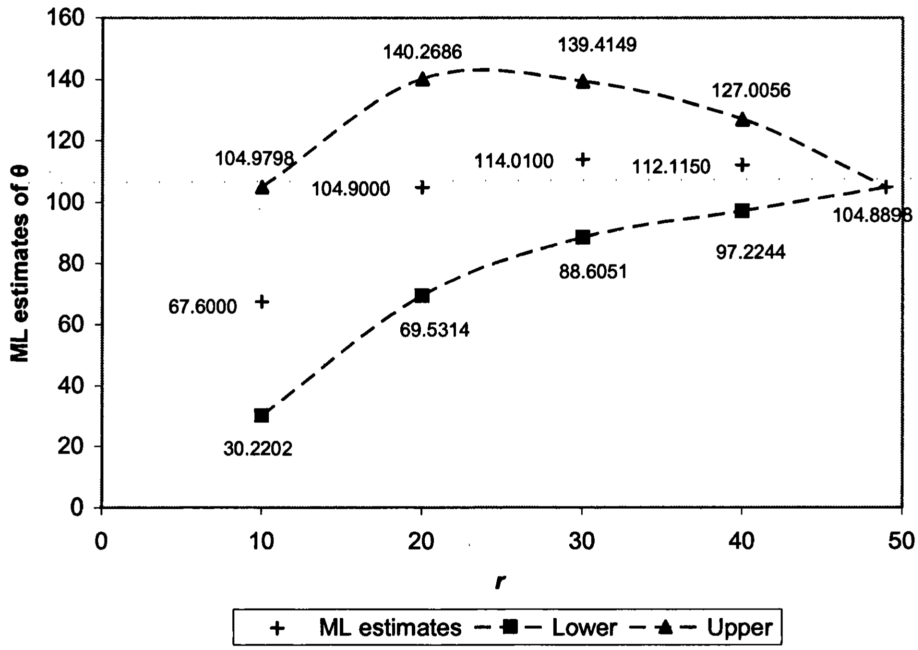

Figure 5.1: $\hat{\theta}_r$ and 95% confidence limits for $\hat{\theta}$ given $\hat{\theta}_r$ for the failure times data.

Similarly, $\widehat{B}_{0.1}-\widehat{B}_{0.1,r}$ has a Normal distribution with zero mean and variance

$$\frac{\theta^2\,(-\ln 0.9)^2\,(n-r)}{nr},$$

so that a 95% confidence interval for $\widehat{B}_{0.1}$ given $\widehat{B}_{0.1,r}$ is

$$\widehat{B}_{0.1,r}\pm 1.96\theta\,(-\ln 0.9)\sqrt{\frac{n-r}{nr}},$$

in which the usual substitution can be made in practice. Table 5.2, together with Figures 5.1 and 5.2 show these limits (based on (5.12)) for various r for the failure times data in Table 1.1. It is noted that the values of $\widehat{sd}(\widehat{\theta}-\widehat{\theta}_r)$ (and hence $\widehat{sd}(\widehat{B}_{0.1}-\widehat{B}_{0.1,r})$) are quite similar when $r=10$ and 20.

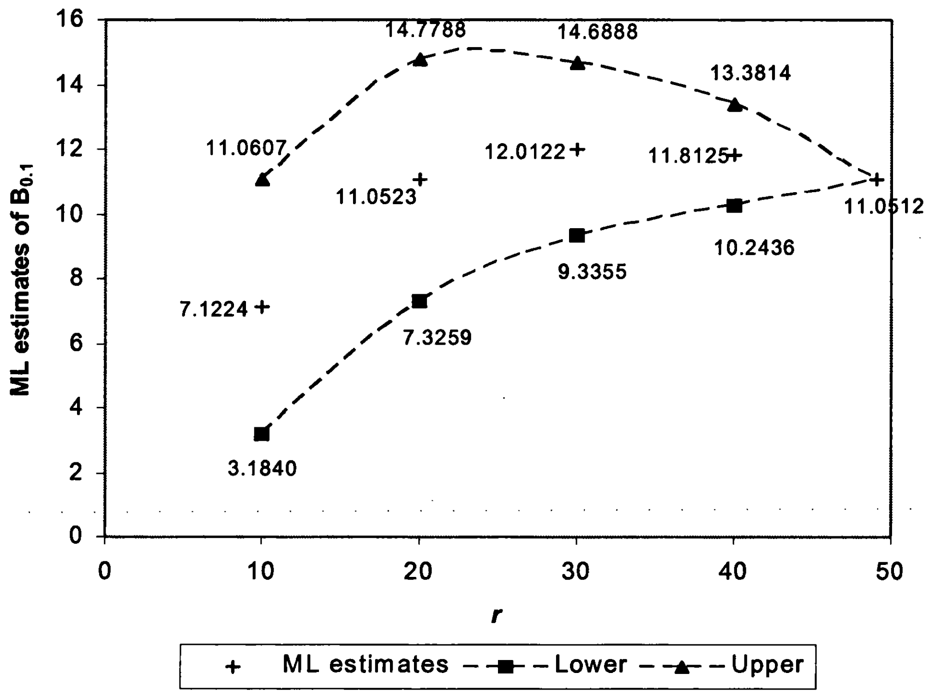

Figure 5.2: $\hat{B}_{0.1,r}$ and 95% confidence limits for $\hat{B}_{0.1}$ given $\hat{B}_{0.1,r}$ for the failure times data.

As above, we are also interested in the extent to which these limits apply in finite samples; we expect, in our simulations, to find 95% of the 10^4 simulated observations of $\widehat{\theta}$ within these limits based on $\widehat{\theta}_r$. Again, Table 5.3 shows generally good agreement between theory and practice, across all r and n considered. We also notice that the upper entries, obtained from (5.11), converge somewhat more quickly to 9500 than their lower counterparts, obtained from (5.12), reflecting the effect of replacing θ by $\widehat{\theta}_r$ in the confidence limits.

5.3 Correlation in the Weibull Distribution

5.3.1 Link Between Final and Interim MLEs

We now indicate the extent to which above results generalise to the Weibull distribution. We recall from (2.16) that the log-likelihood l_r is now a function of two parameters, and the main changes involve the introduction of matrix-vector versions of relationships, which are now approximate rather than exact; we have, from (5.1),

$$\begin{pmatrix} \widehat{\theta}_r - \theta \\ \widehat{\beta}_r - \beta \end{pmatrix} \simeq \mathbf{A}_r^{-1} \begin{pmatrix} \frac{\partial l_r}{\partial \theta} \\ \frac{\partial l_r}{\partial \beta} \end{pmatrix}.$$

r	n					
	25	50	100	1000	2500	5000
$0.2n$	9498	9507	9481	9500	9523	9470
	8738	9074	9263	9490	9508	9468
$0.4n$	9476	9498	9471	9471	9494	9478
	9095	9318	9383	9485	9503	9480
$0.6n$	9520	9469	9456	9485	9519	9506
	9314	9389	9413	9500	9528	9496
$0.8n$	9529	9513	9517	9532	9502	9485
	9374	9464	9518	9528	9495	9489

Table 5.3: Number of replications of $\hat{\theta}$ within the 95% confidence limits based on true θ (upper, based on (5.11)) and $\hat{\theta}_r$ (lower, based on (5.12)), for exponential data generated with $\theta = 100$.

We refer to (2.28) for the elements of $\mathbf{A}_r^{-1}$, but can express the above in general as

$$\begin{pmatrix} \widehat{\theta}_r - \theta \\ \widehat{\beta}_r - \beta \end{pmatrix} \simeq \begin{pmatrix} A_r^{\theta\theta} & A_r^{\theta\beta} \\ A_r^{\theta\beta} & A_r^{\beta\beta} \end{pmatrix} \begin{pmatrix} \frac{\partial l_r}{\partial \theta} \\ \frac{\partial l_r}{\partial \beta} \end{pmatrix} = \begin{pmatrix} A_r^{\theta\theta}\frac{\partial l_r}{\partial \theta} + A_r^{\theta\beta}\frac{\partial l_r}{\partial \beta} \\ A_r^{\theta\beta}\frac{\partial l_r}{\partial \theta} + A_r^{\beta\beta}\frac{\partial l_r}{\partial \beta} \end{pmatrix}.$$

We are again interested at the extent to which earlier estimates are consistent with final estimates; with two parameters, we have four combinations of correlation:

$$Corr\left(\widehat{\theta},\widehat{\theta}_r\right), Corr\left(\widehat{\theta},\widehat{\beta}_r\right), Corr\left(\widehat{\beta},\widehat{\theta}_r\right), Corr\left(\widehat{\beta},\widehat{\beta}_r\right).$$

Since (5.1) also covers the case $r = n$, we now have, for instance

$$Corr\left(\widehat{\theta},\widehat{\theta}_r\right) = Corr\left\{\widehat{\theta}-\theta,\widehat{\theta}_r-\theta\right\}$$

but, as mentioned in (5.2), it is more convenient to consider the corresponding covariance, in which case we require

$$\begin{aligned} Cov\left(\widehat{\theta},\widehat{\theta}_r\right) &\simeq Cov\left(\left\{A^{\theta\theta}\frac{\partial l}{\partial \theta} + A^{\theta\beta}\frac{\partial l}{\partial \beta}\right\}, \left\{A_r^{\theta\theta}\frac{\partial l_r}{\partial \theta} + A_r^{\theta\beta}\frac{\partial l_r}{\partial \beta}\right\}\right) \\ &\simeq A^{\theta\theta}A_r^{\theta\theta}Cov\left(\frac{\partial l}{\partial \theta},\frac{\partial l_r}{\partial \theta}\right) + A^{\theta\theta}A_r^{\theta\beta}Cov\left(\frac{\partial l}{\partial \theta},\frac{\partial l_r}{\partial \beta}\right) + \\ &\quad A^{\theta\beta}A_r^{\theta\theta}Cov\left(\frac{\partial l}{\partial \beta},\frac{\partial l_r}{\partial \theta}\right) + A^{\theta\beta}A_r^{\theta\beta}Cov\left(\frac{\partial l}{\partial \beta},\frac{\partial l_r}{\partial \beta}\right). \end{aligned} \quad (5.13)$$

Similarly, we obtain

$$\begin{aligned} Cov(\widehat{\theta},\widehat{\beta}_r) &\simeq A^{\theta\theta}A_r^{\theta\beta}Cov\left(\frac{\partial l}{\partial \theta},\frac{\partial l_r}{\partial \theta}\right) + A^{\theta\theta}A_r^{\beta\beta}Cov\left(\frac{\partial l}{\partial \theta},\frac{\partial l_r}{\partial \beta}\right) + \\ &\quad A^{\theta\beta}A_r^{\theta\beta}Cov\left(\frac{\partial l}{\partial \beta},\frac{\partial l_r}{\partial \theta}\right) + A^{\theta\beta}A_r^{\beta\beta}Cov\left(\frac{\partial l}{\partial \beta},\frac{\partial l_r}{\partial \beta}\right), \end{aligned} \quad (5.14)$$

$$\begin{aligned} Cov(\widehat{\beta},\widehat{\theta}_r) &\simeq A^{\theta\beta}A_r^{\theta\theta}Cov\left(\frac{\partial l}{\partial\theta},\frac{\partial l_r}{\partial\theta}\right)+A^{\theta\beta}A_r^{\theta\beta}Cov\left(\frac{\partial l}{\partial\theta},\frac{\partial l_r}{\partial\beta}\right)+ \\ & A^{\beta\beta}A_r^{\theta\theta}Cov\left(\frac{\partial l}{\partial\beta},\frac{\partial l_r}{\partial\theta}\right)+A^{\beta\beta}A_r^{\theta\beta}Cov\left(\frac{\partial l}{\partial\beta},\frac{\partial l_r}{\partial\beta}\right), \end{aligned} \tag{5.15}$$

$$\begin{aligned} Cov(\widehat{\beta},\widehat{\beta}_r) &\simeq A^{\theta\beta}A_r^{\theta\beta}Cov\left(\frac{\partial l}{\partial\theta},\frac{\partial l_r}{\partial\theta}\right)+A^{\theta\beta}A_r^{\beta\beta}Cov\left(\frac{\partial l}{\partial\theta},\frac{\partial l_r}{\partial\beta}\right)+ \\ & A^{\beta\beta}A_r^{\theta\beta}Cov\left(\frac{\partial l}{\partial\beta},\frac{\partial l_r}{\partial\theta}\right)+A^{\beta\beta}A_r^{\beta\beta}Cov\left(\frac{\partial l}{\partial\beta},\frac{\partial l_r}{\partial\beta}\right). \end{aligned} \tag{5.16}$$

We consider these covariances in two ways; first, we operate from basic principles, and we then consider the generalisation (5.8).

Covariance from Basic Principles

The approach requires, for large samples, the following terms:

$$Cov\left(\frac{\partial l}{\partial\theta},\frac{\partial l_r}{\partial\theta}\right),Cov\left(\frac{\partial l}{\partial\theta},\frac{\partial l_r}{\partial\beta}\right),Cov\left(\frac{\partial l}{\partial\beta},\frac{\partial l_r}{\partial\theta}\right),Cov\left(\frac{\partial l}{\partial\beta},\frac{\partial l_r}{\partial\beta}\right),$$

and due to regularity conditions, these can be written in terms of joint expectations of complete and censored score functions. As discussed in Section 4.2.1, we proceed by writing the score functions in terms of the transformed variable Z, from (2.26), which follows the standard exponential distribution. It follows from (2.17), (2.18), (2.31) and (2.32) that

$$\frac{\partial l_r}{\partial\theta}=\beta\theta^{-1}\left\{\sum_{i=1}^{r}Z_{i:n}+(n-r)Z_{r:n}-r\right\}, \tag{5.17}$$

$$\frac{\partial l_r}{\partial\beta}=\beta^{-1}\left\{r+\sum_{i=1}^{r}\ln Z_{i:n}-\sum_{i=1}^{r}Z_{i:n}\ln Z_{i:n}-(n-r)Z_{r:n}\ln Z_{r:n}\right\}, \tag{5.18}$$

$$\frac{\partial l}{\partial\theta}=\beta\theta^{-1}\left\{\sum_{i=1}^{n}Z_i-n\right\}, \tag{5.19}$$

and

$$\frac{\partial l}{\partial\beta}=\beta^{-1}\left\{n+\sum_{i=1}^{n}\ln Z_i-\sum_{i=1}^{n}Z_i\ln Z_i\right\}. \tag{5.20}$$

Expectations involved It is clear from the above that, when taking joint expectations of the two sets of score functions, we can anticipate the need for the sum of the expectations given at (4.12) and (4.19), and, in particular, on expectations of the following expression:

$$\left\{\sum_{i=1}^{n}Z_i^p\left(\ln Z_i\right)^a\right\}\times\left\{\sum_{i=1}^{r}Z_{i:n}^p\left(\ln Z_{i:n}\right)^a+(n-r)\,Z_{r:n}^p\left(\ln Z_{r:n}\right)^a\right\}.$$

Hence, it is appropriate to here introduce the following expectations:

$$
\begin{aligned}
H_1 &= E\left[\sum_{i=1}^{n} Z_i\right] = nE[Z] = n, &(5.21)\\
H_2 &= E\left[\sum_{i=1}^{n} \ln Z_i\right] = nE\left[\ln Z\right] = -n\gamma, \\
H_3 &= E\left[\sum_{i=1}^{n} Z_i \ln Z_i\right] = nE\left[Z\ln Z\right] = n\left(1-\gamma\right), \\
H_4 &= E\left[\sum_{i=1}^{n} Z_{i:n} \sum_{i=1}^{r} \ln Z_{i:n}\right], \\
H_5 &= E\left[\sum_{i=1}^{n} \ln Z_i \sum_{i=1}^{r} \ln Z_{i:n}\right], \\
H_6 &= E\left[\sum_{i=1}^{n} Z_i \ln Z_i \sum_{i=1}^{r} \ln Z_{i:n}\right], \\
H_7 &= E\left[\sum_{i=1}^{n} Z_i \left(\sum_{i=1}^{r} Z_{i:n} + (n-r)Z_{r:n}\right)\right] = r\left(n+1\right) \quad \text{from (5.5)}, &(5.22)\\
H_8 &= E\left[\sum_{i=1}^{n} \ln Z_i \left(\sum_{i=1}^{r} Z_{i:n} + (n-r)Z_{r:n}\right)\right], \\
H_9 &= E\left[\sum_{i=1}^{n} Z_i \ln Z_i \left(\sum_{i=1}^{r} Z_{i:n} + (n-r)Z_{r:n}\right)\right], \\
H_{10} &= E\left[\sum_{i=1}^{n} Z_i \left(\sum_{i=1}^{r} Z_{i:n}\ln Z_{i:n} + (n-r)Z_{r:n}\ln Z_{r:n}\right)\right], \\
H_{11} &= E\left[\sum_{i=1}^{n} \ln Z_i \left(\sum_{i=1}^{r} Z_{i:n}\ln Z_{i:n} + (n-r)Z_{r:n}\ln Z_{r:n}\right)\right], \\
H_{12} &= E\left[\sum_{i=1}^{n} Z_i \ln Z_i \left(\sum_{i=1}^{r} Z_{i:n}\ln Z_{i:n} + (n-r)Z_{r:n}\ln Z_{r:n}\right)\right].
\end{aligned}
$$

We note that H_4 to H_{12} involve taking expected values on the products of summations with varied upper limits, and on expanding these products, the algebra can become considerably length; to illustrate this, we take, for example,

$$
\begin{aligned}
H_4 &= \sum_{i=1}^{r} E\left[Z_{i:n}\ln Z_{i:n}\right] + \sum_{i=1}^{r-1}\sum_{j=i+1}^{r} E\left[Z_{i:n}\ln Z_{j:n}\right] + \sum_{i=1}^{r-1}\sum_{j=i+1}^{r} E\left[(\ln Z_{i:n})Z_{j:n}\right] \\
&\quad + \sum_{i=1}^{r}\sum_{j=r+1}^{n} E\left[(\ln Z_{i:n})Z_{j:n}\right] \qquad (5.23)
\end{aligned}
$$

Expectation	Theoretical	Simulated
H_1	25.0000	25.0913
H_2	-14.4304	-14.3205
H_3	10.5696	10.7148
H_4	-473.9927	-473.8339
H_5	314.166249	311.0715
H_6	-194.3861	-196.4959
H_7	390.0000	392.1974
H_8	-195.7974	-194.2607
H_9	174.6429	177.0255
H_{10}	-103.5524	-103.2782
H_{11}	77.4108	76.8722
H_{12}	-35.7704	-36.2064

Table 5.4: Numerical checks of expectations H_1 to H_{12} calculated at $r = 15, n = 25$ using 10^4 replications.

while

$$\begin{aligned} H_{10} &= \sum_{i=1}^{r} E\left[Z_{i:n}^2 \ln Z_{i:n}\right] + \sum_{i=1}^{r-1}\sum_{j=i+1}^{r} E\left[Z_{i:n}Z_{j:n}\ln Z_{j:n}\right] \\ &\quad + \sum_{i=1}^{r-1}\sum_{j=i+1}^{r} E\left[Z_{i:n}(\ln Z_{i:n})Z_{j:n}\right] + \sum_{i=1}^{r}\sum_{j=r+1}^{n} E\left[Z_{i:n}(\ln Z_{i:n})Z_{j:n}\right] \\ &\quad + (n-r)\left\{ \begin{array}{c} \sum_{i=1}^{r-1} E\left[Z_{i:n}Z_{r:n}\ln Z_{r:n}\right] + E\left[Z_{r:n}^2 \ln Z_{r:n}\right] \\ + \sum_{j=r+1}^{n} E\left[Z_{r:n}(\ln Z_{r:n})Z_{j:n}\right] \end{array} \right\}, \end{aligned} \tag{5.24}$$

It follows that there are some structures embedded in these H equations; for instance, the similarity between H_4 and H_8, H_6 and H_{11}, H_9 and H_{10}; we will briefly discuss this later. We refer to Section 4.2.3 for expressions for the expectations of the form $E\left[Z_{i:n}^p(\ln Z_{i:n})^a\right]$ for $a = 0, 1, 2$ and $p = 0, 1$, and Section 4.2.4 for the expectations of the form $E\left[Z_{i:n}^p(\ln Z_{i:n})^a Z_{j:n}^q(\ln Z_{j:n})^b\right]$ for $a, b, p, q = 0, 1$. We can use Mathematica to compute H_1 to H_{12} (see Appendix E for more details on the Mathematica code), and compare these to their corresponding simulated counterparts obtained from 10^4 replications. Table 5.4 shows this comparison for $r = 15, n = 25$. We see good agreement between the theoretical and simulated values.

Covariances of the score functions Using the above expectations, and from (5.17) to (5.20), we can obtain the covariances of the score functions required in the covariances of the complete and censored MLEs, as shown below.

1. From (5.17) and (5.19),

$$\begin{aligned} Cov\left(\frac{\partial l}{\partial \theta}, \frac{\partial l_r}{\partial \theta}\right) &= E\left[\frac{\partial l}{\partial \theta} \times \frac{\partial l_r}{\partial \theta}\right] \\ &= E\left[\beta\theta^{-1}\left\{\sum_{i=1}^{n} Z_i - n\right\}\frac{\partial l_r}{\partial \theta}\right] \\ &= \beta\theta^{-1}E\left[\left\{\sum_{i=1}^{n} Z_i\right\}\frac{\partial l_r}{\partial \theta}\right] - n\beta\theta^{-1}E\left[\frac{\partial l_r}{\partial \theta}\right] \end{aligned}$$

and since $E\left[\frac{\partial l_r}{\partial \theta}\right] = 0$, we have

$$\begin{aligned} Cov\left(\frac{\partial l}{\partial \theta}, \frac{\partial l_r}{\partial \theta}\right) &= \beta\theta^{-1}E\left[\left\{\sum_{i=1}^{n} Z_i\right\}\left\{\beta\theta^{-1}\left(\sum_{i=1}^{r} Z_{i:n} + (n-r)Z_{r:n} - r\right)\right\}\right] \\ &= \beta^2\theta^{-2}\left\{H_7 - rH_1\right\}. \end{aligned} \tag{5.25}$$

2. Likewise, from (5.18) and (5.19), $Cov\left(\frac{\partial l}{\partial \theta}, \frac{\partial l_r}{\partial \beta}\right)$ is given by

$$\begin{aligned} & E\left[\frac{\partial l}{\partial \theta} \times \frac{\partial l_r}{\partial \beta}\right] \\ =\ & E\left[\beta\theta^{-1}\left\{\sum_{i=1}^{n} Z_i - n\right\}\frac{\partial l_r}{\partial \beta}\right] \\ =\ & \beta\theta^{-1}E\left[\left\{\sum_{i=1}^{n} Z_i\right\}\frac{\partial l_r}{\partial \beta}\right] - n\beta\theta^{-1}E\left[\frac{\partial l_r}{\partial \beta}\right] \\ =\ & \beta\theta^{-1}E\left[\left\{\sum_{i=1}^{n} Z_i\right\}\left\{\beta^{-1}\left(r + \sum_{i=1}^{r}\ln Z_{i:n} - \sum_{i=1}^{r} Z_{i:n}\ln Z_{i:n} - (n-r)Z_{r:n}\ln Z_{r:n}\right)\right\}\right] \\ =\ & \theta^{-1}\left\{rH_1 + H_4 - H_{10}\right\}. \end{aligned} \tag{5.26}$$

3. From (5.17) and (5.20), we have

$$\begin{aligned} Cov\left(\frac{\partial l}{\partial \beta}, \frac{\partial l_r}{\partial \theta}\right) &= E\left[\frac{\partial l}{\partial \beta} \times \frac{\partial l_r}{\partial \theta}\right] \\ &= E\left[\beta^{-1}\left\{n + \sum_{i=1}^{n}\ln Z_i - \sum_{i=1}^{n} Z_i\ln Z_i\right\}\frac{\partial l_r}{\partial \theta}\right] \\ &= n\beta^{-1}E\left[\frac{\partial l_r}{\partial \theta}\right] + \beta^{-1}E\left[\left\{\sum_{i=1}^{n}\ln Z_i\right\}\frac{\partial l_r}{\partial \theta}\right] - \beta^{-1}E\left[\left\{\sum_{i=1}^{n} Z_i\ln Z_i\right\}\frac{\partial l_r}{\partial \theta}\right] \\ &= \beta^{-1}E\left[\left\{\sum_{i=1}^{n}\ln Z_i\right\}\left\{\beta\theta^{-1}\left(\sum_{i=1}^{r} Z_{i:n} + (n-r)Z_{r:n} - r\right)\right\}\right] \\ &\quad -\beta^{-1}E\left[\left\{\sum_{i=1}^{n} Z_i\ln Z_i\right\}\left\{\beta\theta^{-1}\left(\sum_{i=1}^{r} Z_{i:n} + (n-r)Z_{r:n} - r\right)\right\}\right] \\ &= \theta^{-1}\left\{H_8 - rH_2 - H_9 + rH_3\right\}. \end{aligned} \tag{5.27}$$

4. From (5.18) and (5.20), $Cov\left(\frac{\partial l}{\partial\beta}, \frac{\partial l_r}{\partial\beta}\right)$ is

$$
\begin{aligned}
& E\left[\frac{\partial l}{\partial\beta} \times \frac{\partial l_r}{\partial\beta}\right] \\
= \; & E\left[\beta^{-1}\left\{n + \sum_{i=1}^{n} \ln Z_i - \sum_{i=1}^{n} Z_i \ln Z_i\right\} \frac{\partial l_r}{\partial\beta}\right] \\
= \; & n\beta^{-1}E\left[\frac{\partial l_r}{\partial\beta}\right] + \beta^{-1}E\left[\left\{\sum_{i=1}^{n} \ln Z_i\right\} \frac{\partial l_r}{\partial\beta}\right] - \beta^{-1}E\left[\left\{\sum_{i=1}^{n} Z_i \ln Z_i\right\} \frac{\partial l_r}{\partial\beta}\right] \\
= \; & \beta^{-1}E\left[\left\{\sum_{i=1}^{n} \ln Z_i\right\}\left\{\beta^{-1}\left(r + \sum_{i=1}^{r} \ln Z_{i:n} - \sum_{i=1}^{r} Z_{i:n} \ln Z_{i:n} - (n-r)Z_{r:n} \ln Z_{r:n}\right)\right\}\right] \\
& -\beta^{-1}E\left[\left\{\sum_{i=1}^{n} Z_i \ln Z_i\right\}\left\{\beta^{-1}\left(r + \sum_{i=1}^{r} \ln Z_{i:n} - \sum_{i=1}^{r} Z_{i:n} \ln Z_{i:n} - (n-r)Z_{r:n} \ln Z_{r:n}\right)\right\}\right] \\
= \; & \beta^{-2}\left\{rH_2 + H_5 - H_{11} - rH_3 - H_6 + H_{12}\right\}. \qquad (5.28)
\end{aligned}
$$

We can now use Mathematica (see Appendix E for further details) to calculate the covariances in (5.25) to (5.28) and set, as before, $\theta = 100, \beta = 2, r = 15, n = 25$; these are

$$
\begin{pmatrix} Cov(\frac{\partial l}{\partial\theta}, \frac{\partial l_r}{\partial\theta}) & Cov(\frac{\partial l}{\partial\theta}, \frac{\partial l_r}{\partial\beta}) \\ Cov(\frac{\partial l}{\partial\beta}, \frac{\partial l_r}{\partial\theta}) & Cov(\frac{\partial l}{\partial\beta}, \frac{\partial l_r}{\partial\beta}) \end{pmatrix} = \begin{pmatrix} 0.0060 & 0.0456 \\ 0.0456 & 5.0928 \end{pmatrix},
$$

and since

$$
\begin{pmatrix} A^{\theta\theta} & A^{\theta\beta} \\ A^{\theta\beta} & A^{\beta\beta} \end{pmatrix} = \begin{pmatrix} 110.8665 & 1.0281 \\ 1.0281 & 0.0973 \end{pmatrix}, \qquad (5.29)
$$

$$
\begin{pmatrix} A_{15}^{\theta\theta} & A_{15}^{\theta\beta} \\ A_{15}^{\theta\beta} & A_{15}^{\beta\beta} \end{pmatrix} = \begin{pmatrix} 178.8346 & -1.6012 \\ -1.6012 & 0.2107 \end{pmatrix},
$$

we obtain, from (5.13) to (5.16),

$$
\begin{aligned}
Cov\left(\widehat{\theta}, \widehat{\theta}_{15}\right) \; \simeq \; & 110.8665 \times 178.8346 \times 0.0060 - 110.8665 \times 1.6012 \times 0.0456 \\
& +1.0281 \times 178.8346 \times 0.0456 - 1.0281 \times 1.6012 \times 5.0928 \\
\simeq \; & 110.8665,
\end{aligned}
$$

$$
\begin{aligned}
Cov(\widehat{\theta}, \widehat{\beta}_{15}) \; \simeq \; & -110.8665 \times 1.6012 \times 0.0060 + 110.8665 \times 0.2107 \times 0.0456 \\
& -1.0281 \times 1.6012 \times 0.0456 + 1.0281 \times 0.2107 \times 5.0928 \\
\simeq \; & 1.0281,
\end{aligned}
$$

$$
\begin{aligned}
Cov(\widehat{\beta}, \widehat{\theta}_{15}) \; \simeq \; & 1.0281 \times 178.8346 \times 0.0060 - 1.0281 \times 1.6012 \times 0.0456 \\
& +0.0973 \times 178.8346 \times 0.0456 - 0.0973 \times 1.6012 \times 5.0928 \\
\simeq \; & 1.0281,
\end{aligned}
$$

$$\begin{aligned} Cov(\widehat{\beta}, \widehat{\beta}_{15}) &\simeq -1.0281 \times 1.6012 \times 0.0060 + 1.0281 \times 0.2107 \times 0.0456 \\ &\quad -0.0973 \times 1.6012 \times 0.0456 + 0.0973 \times 0.2107 \times 5.0928 \\ &\simeq 0.0973, \end{aligned}$$

which leads to

$$\begin{aligned} Corr\left(\widehat{\theta}, \widehat{\theta}_{15}\right) &\simeq \frac{110.8665}{\sqrt{110.8665} \times \sqrt{178.8346}} = 0.7874, \\ Corr\left(\widehat{\theta}, \widehat{\beta}_{15}\right) &\simeq \frac{1.0281}{\sqrt{110.8665} \times \sqrt{0.2107}} = 0.2127, \\ Corr\left(\widehat{\beta}, \widehat{\theta}_{15}\right) &\simeq \frac{1.0281}{\sqrt{0.0973} \times \sqrt{178.8346}} = 0.2465, \\ Corr\left(\widehat{\beta}, \widehat{\beta}_{15}\right) &\simeq \frac{0.0973}{\sqrt{0.0973} \times \sqrt{0.2107}} = 0.6795. \end{aligned}$$

Covariance from the Generalisation (5.8)

It is striking to note that the numerical values for the covariances of complete and censored MLEs are identical to those found at (5.29). This observation is consistent with the conjecture at (5.8), a result generalised from the exponential distribution, suggesting that it might be possible to extend (5.8) to the Weibull case, from which

$$\begin{pmatrix} Cov(\widehat{\theta}, \widehat{\theta}_r) & Cov(\widehat{\theta}, \widehat{\beta}_r) \\ Cov(\widehat{\beta}, \widehat{\theta}_r) & Cov(\widehat{\beta}, \widehat{\beta}_r) \end{pmatrix} = \begin{pmatrix} A^{\theta\theta} & A^{\theta\beta} \\ A^{\theta\beta} & A^{\beta\beta} \end{pmatrix} \tag{5.30}$$

or equivalently (from (5.7))

$$\begin{pmatrix} Cov(\frac{\partial l}{\partial \theta}, \frac{\partial l_r}{\partial \theta}) & Cov(\frac{\partial l}{\partial \theta}, \frac{\partial l_r}{\partial \beta}) \\ Cov(\frac{\partial l}{\partial \beta}, \frac{\partial l_r}{\partial \theta}) & Cov(\frac{\partial l}{\partial \beta}, \frac{\partial l_r}{\partial \beta}) \end{pmatrix} = \begin{pmatrix} A_{r,\theta\theta} & A_{r,\theta\beta} \\ A_{r,\theta\beta} & A_{r,\beta\beta} \end{pmatrix}, \tag{5.31}$$

written in terms of the score functions.

Simplifications of the covariances We would like to here show that (5.31) holds. Using (5.21) and (5.22), and from (5.25), $Cov\left(\frac{\partial l}{\partial \theta}, \frac{\partial l_r}{\partial \theta}\right)$ simplifies to

$$\beta^2\theta^{-2}\left\{r(n+1) - rn - nr - nr\right\} = r\beta^2\theta^{-2} = A_{r,\theta\theta}$$

as required by (5.31). This result is particularly relevant to (5.6); for $\beta = 1$ the Weibull distribution simplifies to an exponential model, and we see $Cov\left(\frac{\partial l}{\partial \theta}, \frac{\partial l_r}{\partial \theta}\right)$ reduces to $r\theta^{-2}$.

However, due to the forms of H_4 to H_{12} (but not H_7), the consideration of $Cov\left(\frac{\partial l}{\partial \theta}, \frac{\partial l_r}{\partial \beta}\right)$, $Cov\left(\frac{\partial l}{\partial \beta}, \frac{\partial l_r}{\partial \theta}\right)$ and $Cov\left(\frac{\partial l}{\partial \beta}, \frac{\partial l_r}{\partial \beta}\right)$ becomes much more involved than that of $Cov\left(\frac{\partial l}{\partial \theta}, \frac{\partial l_r}{\partial \theta}\right)$.

For instance, using (5.23) and (5.24), and from (5.26), $Cov\left(\frac{\partial l}{\partial \theta}, \frac{\partial l_r}{\partial \beta}\right)$ is given by

$$\theta^{-1}\left\{\begin{array}{c} rn + \sum_{i=1}^{r} E\left[Z_{i:n}\ln Z_{i:n}\right] + \sum_{i=1}^{r-1}\sum_{j=i+1}^{r} E\left[Z_{i:n}\ln Z_{j:n}\right] + \sum_{i=1}^{r-1}\sum_{j=i+1}^{r} E\left[(\ln Z_{i:n})Z_{j:n}\right] \\ +\sum_{i=1}^{r}\sum_{j=r+1}^{n} E\left[(\ln Z_{i:n})Z_{j:n}\right] - \sum_{i=1}^{r} E\left[Z_{i:n}^2\ln Z_{i:n}\right] - \sum_{i=1}^{r-1}\sum_{j=i+1}^{r} E\left[Z_{i:n}Z_{j:n}\ln Z_{j:n}\right] \\ -\sum_{i=1}^{r-1}\sum_{j=i+1}^{r} E\left[Z_{i:n}(\ln Z_{i:n})Z_{j:n}\right] - \sum_{i=1}^{r}\sum_{j=r+1}^{n} E\left[Z_{i:n}(\ln Z_{i:n})Z_{j:n}\right] \\ -(n-r)\left(\sum_{i=1}^{r-1} E\left[Z_{i:n}Z_{r:n}\ln Z_{r:n}\right] + E\left[Z_{r:n}^2\ln Z_{r:n}\right] + \sum_{j=r+1}^{n} E\left[Z_{r:n}(\ln Z_{r:n})Z_{j:n}\right]\right) \end{array}\right\} \quad (5.32)$$

which should be shown equal to

$$A_{r,\theta\beta} = -r\theta^{-1}\left\{1 - \gamma - r^{-1}\sum_{i=1}^{r}(-1)^{r-i}\binom{n}{i-1}\binom{n-i-1}{r-i}\ln(n+1-i)\right\}.$$

Moreover, since the majority of the terms in (5.32) involves at least one level of summations of varied lower and upper limits; see, for examples, using (4.18),

$$\sum_{i=1}^{r} E\left[Z_{i:n}\ln Z_{i:n}\right] = \sum_{i=1}^{r}\left\{c_{i:n}\sum_{k=0}^{i-1}\frac{(-1)^{i-1-k}\binom{i-1}{k}}{(n-k)^2}\left[1-\gamma-\ln(n-k)\right]\right\}$$

in which we require two levels of single summation, and, using (4.32),

$$\sum_{i=1}^{r-1}\sum_{j=i+1}^{r} E\left[Z_{i:n}\ln Z_{j:n}\right] = \sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\left\{\begin{array}{c} c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(i+l-k)^2(n-i-l)(n-k)^2}\times \\ \left[\begin{array}{c} -(i+l-k)\left[\gamma(i+l-k)+n-i-l\right] \\ -(n-k)^2\ln(n-i-l) \\ +(n-i-l)(n+i-2k+l)\ln(n-k) \end{array}\right] \end{array}\right\}$$

in which we require two levels of double summations, there is limited analytical progress we could make here, and hence a detailed proof for $Cov\left(\frac{\partial l}{\partial \theta}, \frac{\partial l_r}{\partial \beta}\right) = Cov\left(\frac{\partial l}{\partial \beta}, \frac{\partial l_r}{\partial \theta}\right) = A_{r,\theta\beta}$ and $Cov\left(\frac{\partial l}{\partial \beta}, \frac{\partial l_r}{\partial \beta}\right) = A_{r,\beta\beta}$ will be given elsewhere. Instead, we use a detailed simulation study to assess the extent to which (5.30) holds for the sampling distributions of $\left(\widehat{\theta}_r, \widehat{\beta}_r\right)$, for various combinations of n, r and parameter values.

Numerical check of (5.30) We use Mathematica to compute the elements of the complete covariance matrix, and compare these to simulated values of $Cov(\widehat{\theta}, \widehat{\theta}_r)$, $Cov(\widehat{\theta}, \widehat{\beta}_r)$, $Cov(\widehat{\beta}, \widehat{\theta}_r)$ and $Cov(\widehat{\beta}, \widehat{\beta}_r)$, which (throughout) are based on 10^4 replications. Tables 5.5 to 5.8 show this comparison for each covariance in turn with $\theta = 100, \beta = 2$, where we see generally good agreement between theory and practice for all r and n considered. We also observe similar agreement for different sets of integer values of θ, β, but there is further scope to check (5.30) for non-integer β values, where analytical progress may be even more

n	Theory	r				
	$(= A^{\theta\theta})$	$0.2n$	$0.4n$	$0.6n$	$0.8n$	$1.0n$
25	110.8665	103.0207	107.4784	108.5282	109.6782	110.4749
50	55.4332	49.1945	54.1417	55.0621	55.3464	55.5401
100	27.7166	27.7198	27.4541	27.6753	27.6516	27.6713
1000	2.7717	2.8197	2.7730	2.7391	2.7588	2.7648
2500	1.1087	1.1173	1.1297	1.1181	1.1323	1.1347
5000	0.5543	0.5540	0.5599	0.5577	0.5589	0.5566

Table 5.5: Theoretical and simulated values for $Cov(\hat{\theta}, \hat{\theta}_r)$ calculated at various r, n using Weibull data generated with $\theta = 100, \beta = 2$ and 10^4 replications.

n	Theory	r				
	$(= A^{\theta\beta})$	$0.2n$	$0.4n$	$0.6n$	$0.8n$	$1.0n$
25	1.0281	1.4808	1.2607	1.2009	1.1301	1.0734
50	0.5140	0.7391	0.5622	0.5239	0.5129	0.4991
100	0.2570	0.2515	0.2542	0.2509	0.2502	0.2491
1000	0.0257	0.0259	0.0272	0.0280	0.0273	0.0271
2500	0.0103	0.0113	0.0112	0.0114	0.0110	0.0108
5000	0.0051	0.0054	0.0053	0.0053	0.0053	0.0054

Table 5.6: Theoretical and simulated values for $Cov(\hat{\theta}, \hat{\beta}_r)$ calculated at various r, n using Weibull data generated with $\theta = 100, \beta = 2$ and 10^4 replications.

limited.

Implications of (5.30) on the H equations As we have previously mentioned, there are some structures embedded in the H equations, but we discuss this only briefly here. In particular, it follows from (5.31) that

$$\begin{aligned} rn + H_4 - H_{10} &= -r\{1 - \gamma - \phi_1\}, \\ rn + H_8 - H_9 &= -r\{1 - \gamma - \phi_1\}, \\ -rn + H_5 - H_6 - H_{11} + H_{12} &= r\left\{\frac{\pi^2}{6} + (1-\gamma)^2 - 2(1-\gamma)\phi_1 + \phi_2\right\}, \end{aligned}$$

n	Theory	r				
	$(= A^{\theta\beta})$	$0.2n$	$0.4n$	$0.6n$	$0.8n$	$1.0n$
25	1.0281	1.1986	1.1634	1.1674	1.0926	1.0734
50	0.5140	0.4412	0.5073	0.5075	0.5003	0.4991
100	0.2570	0.2195	0.2697	0.2510	0.2479	0.2491
1000	0.0257	0.0303	0.0335	0.0286	0.0274	0.0271
2500	0.0103	0.0099	0.0109	0.0106	0.0107	0.0108
5000	0.0051	0.0055	0.0054	0.0052	0.0053	0.0054

Table 5.7: Theoretical and simulated values for $Cov(\hat{\beta}, \hat{\theta}_r)$ calculated at various r, n using Weibull data generated with $\theta = 100, \beta = 2$ and 10^4 replications.

n	Theory	r				
	$(= A^{\beta\beta})$	$0.2n$	$0.4n$	$0.6n$	$0.8n$	$1.0n$
25	0.0973	0.1815	0.1415	0.1306	0.1278	0.1258
50	0.0486	0.0677	0.0580	0.0560	0.0550	0.0545
100	0.0243	0.0275	0.0258	0.0257	0.0257	0.0256
1000	0.0024	0.0023	0.0024	0.0024	0.0025	0.0025
2500	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
5000	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005

Table 5.8: Theoretical and simulated values for $Cov(\hat{\beta}, \hat{\beta}_r)$ calculated at various r, n using Weibull data generated with $\theta = 100, \beta = 2$ and 10^4 replications.

so that

$$H_4 - H_8 = H_{10} - H_9$$

and

$$H_5 - H_6 - H_{11} + H_{12} = r\left\{\frac{\pi^2}{6} + (1-\gamma)^2 - 2(1-\gamma)\phi_1 + \phi_2 + n\right\}.$$

Therefore, this serves as a convenient starting point to consider further the relationship between the H equations; these will, nonetheless, be considered elsewhere.

Implications of (5.30) on the correlations between final and interim MLEs If (5.30) holds, the correlations between final and interim MLEs would follow immediately from the complete and censored EFI matrices; we have

$$\begin{aligned} Corr\left(\widehat{\theta}, \widehat{\theta}_r\right) &\simeq \frac{Cov\left(\widehat{\theta}, \widehat{\theta}_r\right)}{\sqrt{Var\left(\widehat{\theta}\right)} \times \sqrt{Var\left(\widehat{\theta}_r\right)}} \\ &\simeq \frac{A^{\theta\theta}}{\sqrt{A^{\theta\theta}} \times \sqrt{A_r^{\theta\theta}}} \\ &\simeq \sqrt{\frac{A^{\theta\theta}}{A_r^{\theta\theta}}} \\ &\simeq \sqrt{\frac{r\left(\pi^2 - 6\phi_1^2 + 6\phi_2\right)\left[\frac{\pi^2}{6} + (1-\gamma)^2\right]}{n\pi^2\left[\frac{\pi^2}{6} + (1-\gamma)^2 - 2(1-\gamma)\phi_1 + \phi_2\right]}}, \end{aligned}$$

$$\begin{aligned} Corr\left(\widehat{\theta},\widehat{\beta}_r\right) &\simeq \frac{Cov\left(\widehat{\theta},\widehat{\beta}_r\right)}{\sqrt{Var\left(\widehat{\theta}\right)}\times\sqrt{Var\left(\widehat{\beta}_r\right)}} \\ &\simeq \frac{A^{\theta\beta}}{\sqrt{A^{\theta\theta}}\times\sqrt{A_r^{\beta\beta}}} \\ &\simeq (1-\gamma)\sqrt{\frac{r\left(\pi^2-6\phi_1^2+6\phi_2\right)}{n\pi^2\left[\frac{\pi^2}{6}+(1-\gamma)^2\right]}}, \end{aligned}$$

$$\begin{aligned} Corr\left(\widehat{\beta},\widehat{\theta}_r\right) &\simeq \frac{Cov\left(\widehat{\beta},\widehat{\theta}_r\right)}{\sqrt{Var\left(\widehat{\beta}\right)}\times\sqrt{Var\left(\widehat{\theta}_r\right)}} \\ &\simeq \frac{A^{\theta\beta}}{\sqrt{A^{\beta\beta}}\times\sqrt{A_r^{\theta\theta}}} \\ &\simeq (1-\gamma)\sqrt{\frac{r\left(\pi^2-6\phi_1^2+6\phi_2\right)}{n\pi^2\left[\frac{\pi^2}{6}+(1-\gamma)^2-2(1-\gamma)\phi_1+\phi_2\right]}}, \end{aligned}$$

$$\begin{aligned} Corr\left(\widehat{\beta},\widehat{\beta}_r\right) &\simeq \frac{Cov\left(\widehat{\beta},\widehat{\beta}_r\right)}{\sqrt{Var\left(\widehat{\beta}\right)}\times\sqrt{Var\left(\widehat{\beta}_r\right)}} \\ &\simeq \frac{A^{\beta\beta}}{\sqrt{A^{\beta\beta}}\times\sqrt{A_r^{\beta\beta}}} \\ &\simeq \sqrt{\frac{A^{\beta\beta}}{A_r^{\beta\beta}}} \\ &\simeq \sqrt{\frac{r\left(\pi^2-6\phi_1^2+6\phi_2\right)}{n\pi^2}}. \end{aligned}$$

5.3.2 Link between $\widehat{B}_{0.1}$ and $\widehat{B}_{0.1,r}$

We are again interested in the agreement between $\widehat{B}_{0.1,r}$ and $\widehat{B}_{0.1}$ for Weibull data. Correlation from basic principles is possible, in which we will also require

$$\begin{pmatrix} b_\theta \\ b_\beta \end{pmatrix}$$

n	Theory $\left(= Var\left(\widehat{B}_{0.1}\right)\right)$	r				
		$0.2n$	$0.4n$	$0.6n$	$0.8n$	$1.0n$
25	56.3056	60.2070	60.7714	60.6623	60.8367	60.7377
50	28.1528	28.7114	28.8534	28.7883	28.7353	28.6470
100	14.0764	14.0576	14.0841	14.1101	14.1354	14.1089
1000	1.4076	1.4209	1.4225	1.4280	1.4332	1.4359
2500	0.5631	0.5731	0.5738	0.5756	0.5740	0.5724
5000	0.2815	0.2851	0.2855	0.2859	0.2849	0.2848

Table 5.9: Theoretical and simulated values for $Cov(\hat{B}_{0.1}, \hat{B}_{0.1,r})$ calculated at various r, n using Weibull data generated with $\theta = 100, \beta = 2$ and 10^4 replications.

given at (2.29). For example, we take, as before, $\theta = 100, \beta = 2, r = 15$ and $n = 25$, and use Mathematica to compute

$$\begin{pmatrix} b_\theta \\ b_\beta \end{pmatrix} = \begin{pmatrix} 0.3246 \\ 18.2613 \end{pmatrix},$$

so that from (5.3)

$$Corr\left(\widehat{B}_{0.1}, \widehat{B}_{0.1,15}\right) \simeq 0.8961.$$

Alternatively, if the conjecture at (5.9) holds here, then we could use (5.10) to obtain, for samples of large size,

$$Corr\left(\widehat{B}_{0.1}, \widehat{B}_{0.1,r}\right) \simeq \sqrt{\frac{b_\theta^2 A^{\theta\theta} + 2b_\theta b_\beta A^{\theta\beta} + b_\beta^2 A^{\beta\beta}}{b_\theta^2 A_r^{\theta\theta} + 2b_\theta b_\beta A_r^{\theta\beta} + b_\beta^2 A_r^{\beta\beta}}}.$$

Table 5.9 provides some summaries of simulation experiments to check (5.9), based on 10^4 estimates of $\widehat{B}_{0.1,r}$; this shows generally good agreement between theory and simulation, and, as with (5.30), $Cov\left(\widehat{B}_{0.1}, \widehat{B}_{0.1,r}\right)$ is independent of r. Returning to the above example, the correlation is

$$\sqrt{\frac{0.3246^2 \times 110.8665 + 2 \times 0.3146 \times 18.2613 \times 1.0281 + 18.2613^2 \times 0.0973}{0.3246^2 \times 178.8346 - 2 \times 0.3146 \times 18.2613 \times 1.6012 + 18.2613^2 \times 0.2107}}$$

and hence also yields 0.8961, but with considerably lesser amount of computation as compared to using (5.3).

5.3.3 Numerical Results

We now consider these results for finite samples. We revisit the sampling distributions of $\widehat{\theta}_r, \widehat{\beta}_r$ and $\widehat{B}_{0.1,r}$ in Section 2.3.4 generated with $\theta = 100$ and $\beta = 2$. Tables 5.10 and 5.13 summarise, for $Corr\left(\widehat{\theta}, \widehat{\theta}_r\right)$ and $Corr\left(\widehat{\beta}, \widehat{\beta}_r\right)$ respectively, the theoretical results for these 10^4 estimates, together with an experimental counterpart shown underneath; we observe a

r	n					
	25	50	100	1000	2500	5000
$0.2n$	.2833	.2720	.2658	.2600	.2596	.2595
	.2551	.2363	.2609	.2643	.2563	.2558
$0.4n$	.5467	.5391	.5351	.5314	.5311	.5310
	.5317	.5292	.5328	.5342	.5303	.5300
$0.6n$	.7874	.7849	.7836	.7824	.7823	.7822
	.7803	.7822	.7812	.7804	.7815	.7799
$0.8n$	.9403	.9406	.9407	.9408	.9408	.9408
	.9387	.9396	.9414	.9396	.9417	.9422
$1.0n$	1	1	1	1	1	1
	1	1	1	1	1	1

Table 5.10: Theoretical (upper) and simulated (lower) values of $Corr(\hat{\theta}, \hat{\theta}_r)$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 2$.

good agreement between theory and practice, with values approaching to 1 as $r \to n$, as we expected. Moreover, equivalent statistics for $Corr\left(\widehat{\theta}, \widehat{\beta}_r\right)$ and $Corr\left(\widehat{\beta}, \widehat{\theta}_r\right)$ are tabulated in Tables 5.11 and 5.12; in complete samples (so $r = n$), we note that $Corr\left(\widehat{\theta}, \widehat{\beta}_n\right) = Corr\left(\widehat{\beta}, \widehat{\theta}_n\right)$ is given by

$$Corr\left(\widehat{\theta}, \widehat{\beta}\right) = \frac{1-\gamma}{\sqrt{\frac{\pi^2}{6} + (1-\gamma)^2}} = 0.3131.$$

This cross-parameter correlation thus has an upper limit of 0.3131, and it is independent of n and the model parameters. Finally, Table 5.14 presents some summaries of simulation experiments for $B_{0.1}$; we notice excellent agreement between theoretical and observed values of $Corr\left(\widehat{B}_{0.1}, \widehat{B}_{0.1,r}\right)$, for all r and n we have considered.

As further reassurance that this theory is in agreement with practice, we may superimpose the theoretical correlation values with the scatter plots of final estimates against interim estimates shown in Figures 2.7 to 2.11; it is clear that our formulae agree with the pattern observed in simulation experiments. We are now in the position to employ these formulae in future calculations like confidence limits.

5.3.4 Confidence Limits Considerations

We are interested at the precision with which we can make statements on final estimates, given earlier estimates of the parameters. However, evidence from the previous section suggests that the relationship between the MLEs of the shape and scale parameters is weak; as seen in Tables 5.11 and 5.12, the upper bound on the strength of correlation is 0.3131, so we will here only consider inference of $\widehat{\theta}$ based on $\widehat{\theta}_r$, and of $\widehat{\beta}$ based on $\widehat{\beta}_r$. Let

$$\Delta_\theta = \widehat{\theta} - \widehat{\theta}_r$$

r	n					
	25	50	100	1000	2500	5000
$0.2n$	.1209	.1168	.1146	.1124	.1122	.1122
	.0678	.1089	.0929	.1113	.1196	.1163
$0.4n$	.1696	.1668	.1652	.1638	.1637	.1637
	.1386	.1495	.1497	.1740	.1738	.1658
$0.6n$	.2127	.2106	.2095	.2085	.2084	.2084
	.1941	.1886	.1936	.2285	.2277	.2130
$0.8n$	.2565	.2550	.2542	.2535	.2534	.2534
	.2359	.2335	.2390	.2682	.2659	.2584
$1.0n$	.3131	.3131	.3131	.3131	.3131	.3131
	.2879	.2869	.2962	.3282	.3257	.3259

Table 5.11: Theoretical (upper) and simulated (lower) values of $Corr(\hat{\theta}, \hat{\beta}_r)$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 2$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	.0887	.0851	.0832	.0814	.0813	.0813
	.0879	.0677	.0680	.1049	.0776	.0857
$0.4n$	.1711	.1688	.1675	.1664	.1663	.1663
	.1705	.1583	.1722	.1951	.1738	.1716
$0.6n$	.2465	.2457	.2453	.2449	.2449	.2449
	.2487	.2302	.2322	.2721	.2531	.2465
$0.8n$	.2944	.2945	.2945	.2945	.2945	.2945
	.2771	.2712	.2777	.3119	.3032	.3035
$1.0n$	.3131	.3131	.3131	.3131	.3131	.3131
	.2879	.2869	.2962	.3282	.3257	.3259

Table 5.12: Theoretical (upper) and simulated (lower) values of $Corr(\hat{\beta}, \hat{\theta}_r)$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 2$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	.3860	.3731	.3659	.3590	.3585	.3583
	.2462	.3186	.3338	.3367	.3592	.3528
$0.4n$	.5418	.5326	.5278	.5233	.5229	.5228
	.4611	.4929	.4991	.5094	.5188	.5207
$0.6n$	.6795	.6727	.6691	.6658	.6656	.6655
	.6254	.6438	.6527	.6592	.6667	.6669
$0.8n$	.8193	.8145	.8119	.8096	.8094	.8093
	.7904	.8000	.8079	.8071	.8093	.8118
$1.0n$	1	1	1	1	1	1
	1	1	1	1	1	1

Table 5.13: Theoretical (upper) and simulated (lower) values of $Corr(\hat{\beta}, \hat{\beta}_r)$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 2$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	.8451	.8511	.8539	.8564	.8565	.8566
	.8634	.8587	.8563	.8578	.8557	.8558
$0.4n$	.8692	.8668	.8654	.8639	.8638	.8638
	.8781	.8696	.8657	.8645	.8638	.8634
$0.6n$	.8961	.8930	.8913	.8897	.8896	.8896
	.8996	.8923	.8906	.8907	.8908	.8899
$0.8n$	.9356	.9330	.9317	.9304	.9304	.9303
	.9369	.9326	.9322	.9307	.9309	.9314
$1.0n$	1	1	1	1	1	1
	1	1	1	1	1	1

Table 5.14: Theoretical (upper) and simulated (lower) values of $Corr(\hat{B}_{0.1}, \hat{B}_{0.1,r})$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 2$.

so Δ_θ is asymptotically Normally distributed with zero mean and variance

$$\begin{aligned} Var(\Delta_\theta) &= Var\left(\widehat{\theta} - \widehat{\theta}_r\right) \\ &= Var(\widehat{\theta}) + Var(\widehat{\theta}_r) - 2Cov\left(\widehat{\theta}, \widehat{\theta}_r\right) \end{aligned}$$

and if (5.30) holds, this becomes

$$\begin{aligned} Var(\Delta_\theta) &= Var(\widehat{\theta}) + Var(\widehat{\theta}_r) - 2Var(\widehat{\theta}) \\ &= Var(\widehat{\theta}_r) - Var(\widehat{\theta}) \\ &\simeq 6\beta^{-2}\theta^2 \left\{ \frac{\frac{\pi^2}{6} + (1-\gamma)^2 - 2(1-\gamma)\phi_1 + \phi_2}{r\left(\pi^2 - 6\phi_1^2 + 6\phi_2\right)} - \frac{\frac{\pi^2}{6} + (1-\gamma)^2}{n\pi^2} \right\}. \end{aligned}$$

This yields the 95% confidence limits for $\widehat{\theta}$ given $\widehat{\theta}_r$:

$$\widehat{\theta} = \widehat{\theta}_r \pm 1.96\sqrt{Var(\Delta_\theta)}. \tag{5.33}$$

Analogously, if

$$\Delta_\beta = \widehat{\beta} - \widehat{\beta}_r$$

then, for large samples,

$$\Delta_\beta \sim N\left\{0, Var(\Delta_\beta)\right\}$$

in which (assuming that (5.30) holds)

$$\begin{aligned} Var(\Delta_\beta) &= Var(\widehat{\beta}) + Var(\widehat{\beta}_r) - 2Var(\widehat{\beta}) \\ &= Var(\widehat{\beta}_r) - Var(\widehat{\beta}) \\ &\simeq 6\beta^2 \left\{ \frac{1}{r\left(\pi^2 - 6\phi_1^2 + 6\phi_2\right)} - \frac{1}{n\pi^2} \right\}, \end{aligned}$$

and the 95% confidence limits for $\widehat{\beta}$ given $\widehat{\beta}_r$ is

$$\widehat{\beta} = \widehat{\beta}_r \pm 1.96\sqrt{Var(\Delta_\beta)}. \tag{5.34}$$

In practice, we replace the unknown parameters θ and β by their respective MLEs.

Some indication of the precision with which we can make statements on $\widehat{B}_{0.1}$ given $\widehat{B}_{0.1,r}$ is also desired;

$$\Delta_{B_{0.1}} = \widehat{B}_{0.1} - \widehat{B}_{0.1,r}$$

has a Normal distribution with mean zero and variance

$$\begin{aligned} Var(\Delta_{B_{0.1}}) &= Var\left(\widehat{B}_{0.1} - \widehat{B}_{0.1,r}\right) \\ &= Var(\widehat{B}_{0.1}) + Var(\widehat{B}_{0.1,r}) - 2Cov\left(\widehat{B}_{0.1}, \widehat{B}_{0.1,r}\right) \end{aligned}$$

and provided that (5.9) holds, this could be approximated by

$$\begin{aligned} & Var(\widehat{B}_{0.1}) + Var(\widehat{B}_{0.1,r}) - 2Var(\widehat{B}_{0.1}) \\ = \; & Var(\widehat{B}_{0.1,r}) - Var(\widehat{B}_{0.1}) \\ = \; & b_\theta^2 Var(\Delta_\theta) + b_\beta^2 Var(\Delta_\beta) + 2b_\theta b_\beta Cov(\Delta_\theta, \Delta_\beta) \end{aligned}$$

where

$$\begin{aligned} Cov(\Delta_\theta, \Delta_\beta) &= Cov(\widehat{\theta} - \widehat{\theta}_r, \widehat{\beta} - \widehat{\beta}_r) \\ &\simeq 6\theta \left\{ \frac{1-\gamma-\phi_1}{r\left(\pi^2 - 6\phi_1^2 + 6\phi_2\right)} - \frac{1-\gamma}{n\pi^2} \right\}. \end{aligned}$$

We can now write down approximate 95% confidence intervals for $\widehat{B}_{0.1}$ given $\widehat{B}_{0.1,r}$.

We use the ball bearings data to illustrate these limits calculated for censoring as in Table 2.6. Table 5.15 shows that these limits converge to 0 as r increases to $n = 23$, but convergence for $\Delta_{B_{0.1}}$ is rather slow compared to that for Δ_θ and Δ_β. Figures 5.3 and 5.4 show that the upper (lower) 95% limit of $\widehat{\theta}$ given $\widehat{\theta}_r$ ($\widehat{\beta}$ given $\widehat{\beta}_r$) is rather flat, but its lower (upper) counterpart converges to 0 quite quickly. It is also clear that early censoring ($r \leq 12$) tend to give wider confidence limits, indicating a lower level of precision; this phenomenon is particularly apparent in the case of shape parameter. Furthermore, because $B_{0.1}$ is a function of θ and β, Figure 5.5 appears to combine the nature of Figures 5.3 and 5.4, resulting in slow converging upper and lower limits.

In addition to a single set of data, we are also interested in the extent to which these limits apply in finite samples; again, based on 10^4 replications, we expect to find 95% of the 10^4 final estimates within the corresponding confidence limits. Tables 5.16 to 5.18 provide, respectively, some summaries for Δ_θ, Δ_β and $\Delta_{B_{0.1}}$. The upper entries assume the true parameters are known, as in running simulation experiments, while the lower entries are

r	8	12	16	20	23
$\widehat{\theta}_r$	67.6415	75.2168	76.6960	78.9674	81.8783
$\widehat{sd}(\Delta_\theta)$	8.4421	6.3909	3.8013	1.8366	0
$\widehat{\beta}_r$	3.2280	2.6241	2.4695	2.3539	2.1021
$\widehat{sd}(\Delta_\beta)$	0.8953	0.5292	0.3585	0.2132	0
$\widehat{B}_{0.1,r}$	33.6860	31.9063	30.8329	30.3563	28.0694
$\widehat{sd}(\Delta_{B_{0.1}})$	2.9231	3.0910	2.6749	1.9673	0

Table 5.15: Standard deviations of $\Delta_\theta, \Delta_\beta$ and $\Delta_{B_{0.1}}$ for the ball bearings data.

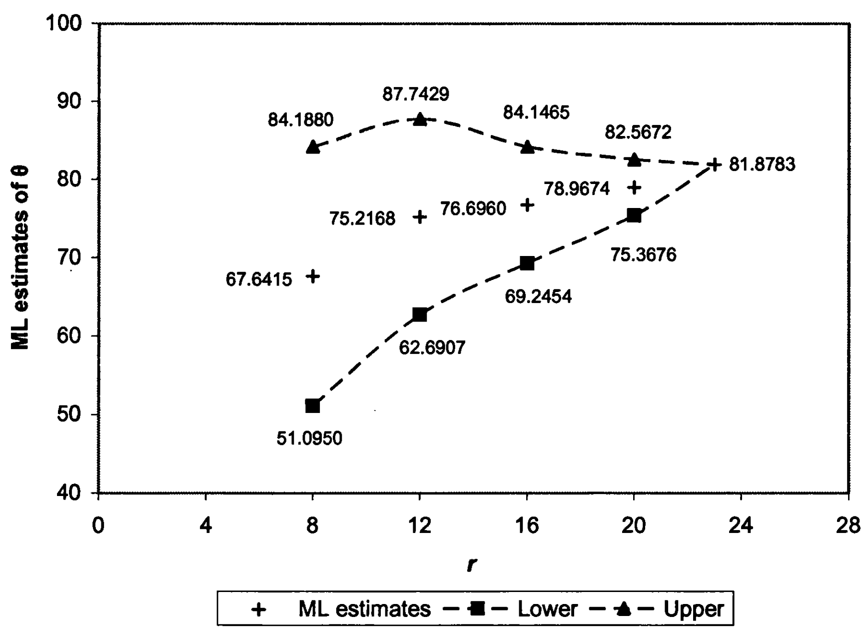

Figure 5.3: $\hat{\theta}_r$ and 95% confidence limits for $\hat{\theta}$ given $\hat{\theta}_r$ for the ball bearings data.

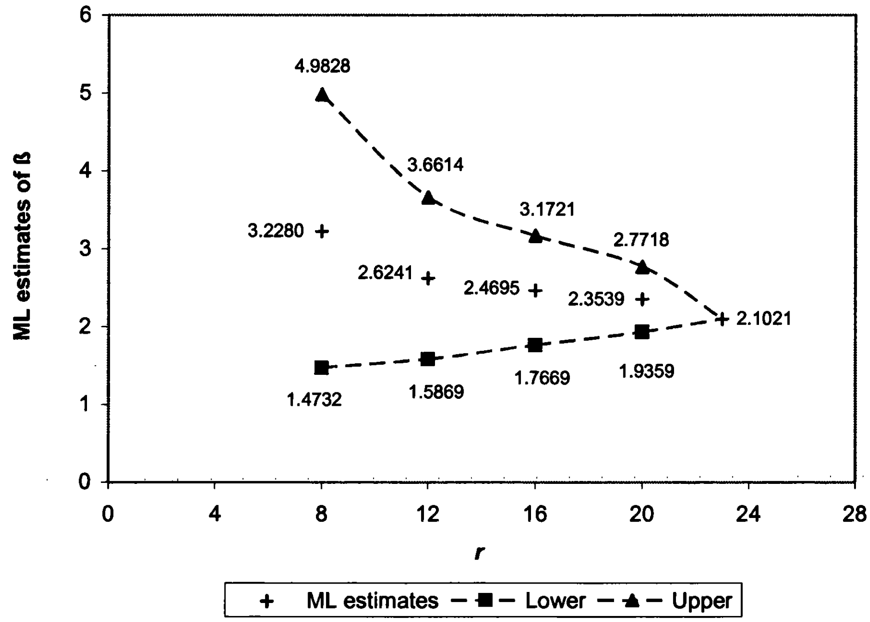

Figure 5.4: $\hat{\beta}_r$ and 95% confidence limits for $\hat{\beta}$ given $\hat{\beta}_r$ for the ball bearings data.

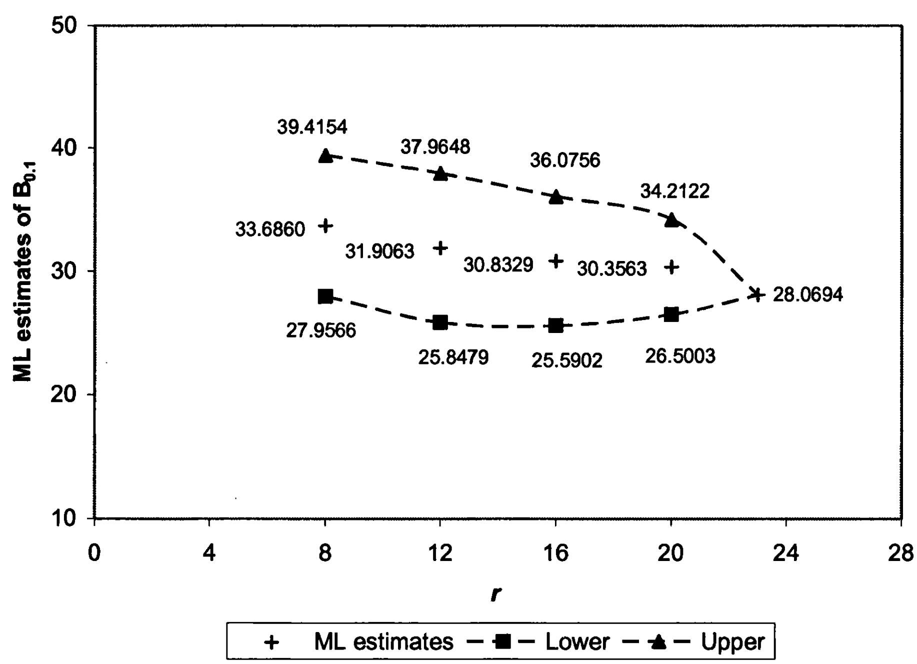

Figure 5.5: $\hat{B}_{0.1,r}$ and 95% confidence limits for $\hat{B}_{0.1}$ given $\hat{B}_{0.1,r}$ for the ball bearings data.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	9616	9610	9578	9465	9480	9491
	9644	9639	9613	9471	9500	9517
$0.4n$	9464	9510	9496	9499	9475	9447
	9461	9494	9489	9488	9495	9486
$0.6n$	9462	9478	9450	9502	9485	9471
	9395	9441	9475	9462	9426	9499
$0.8n$	9476	9466	9498	9478	9493	9528
	9350	9426	9468	9449	9457	9570

Table 5.16: Number of replications of $\hat{\theta}$ within the 95% confidence limits based on true θ, β (upper) and $\hat{\theta}_r, \hat{\beta}_r$ (lower), for Weibull data generated with $\theta = 100, \beta = 2$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	7192	8275	8847	9372	9458	9479
	9340	9258	9312	9408	9487	9515
$0.4n$	8352	8898	9218	9464	9448	9461
	9340	9374	9429	9454	9484	9487
$0.6n$	8782	9114	9282	9480	9474	9446
	9408	9451	9451	9448	9508	9479
$0.8n$	9047	9289	9402	9471	9468	9509
	9475	9511	9514	9478	9516	9520

Table 5.17: Number of replications of $\hat{\beta}$ within the 95% confidence limits based on true θ, β (upper) and $\hat{\theta}_r, \hat{\beta}_r$ (lower), for Weibull data generated with $\theta = 100, \beta = 2$.

based on practical consideration, where we use the MLEs of θ and β instead. All tables show a generally good agreement with expectation, and the difference between the two entries can be explained by the deviation between $\theta = 100, \beta = 2$ and their ML estimates. We recall from Table 2.10 that, in small samples, the estimates of standard deviation of $\widehat{\beta}_r$ are usually larger than their true values, leading to a large estimate of $Var(\Delta_\beta)$ and a wider confidence limits. In contrast, we notice only a slight difference between the two sets of values of $\Delta_{B_{0.1}}$ in Table 5.18, consistent to the observation in Table 2.12.

5.4 Correlation in the Burr Distribution

5.4.1 Link Between Final and Interim MLEs

Following the same approach as in the Weibull case in Section 5.3.1, we can now measure the effectiveness of $\widehat{\alpha}_r$ and $\widehat{\tau}_r$ as estimates of $\widehat{\alpha}$ and $\widehat{\tau}$. The general asymptotic relationship

r	n					
	25	50	100	1000	2500	5000
$0.2n$	9613	9543	9538	9498	9459	9479
	9440	9512	9479	9416	9492	9514
$0.4n$	9483	9476	9505	9475	9456	9463
	9478	9520	9494	9406	9508	9517
$0.6n$	9444	9463	9488	9512	9471	9457
	9471	9548	9496	9449	9538	9522
$0.8n$	9442	9510	9524	9487	9492	9488
	9501	9569	9538	9477	9535	9491

Table 5.18: Number of replications of $\hat{B}_{0.1}$ within the 95% confidence limits based on true θ, β (upper) and $\hat{\theta}_r, \hat{\beta}_r$ (lower), for Weibull data generated with $\theta = 100, \beta = 2$.

in (5.1) here is

$$\begin{pmatrix} \widehat{\alpha}_r - \alpha \\ \widehat{\tau}_r - \tau \end{pmatrix} \simeq \begin{pmatrix} A_r^{\alpha\alpha} & A_r^{\alpha\tau} \\ A_r^{\alpha\tau} & A_r^{\tau\tau} \end{pmatrix} \begin{pmatrix} \frac{\partial l_r}{\partial \alpha} \\ \frac{\partial l_r}{\partial \tau} \end{pmatrix} = \begin{pmatrix} A_r^{\alpha\alpha}\frac{\partial l_r}{\partial \alpha} + A_r^{\alpha\tau}\frac{\partial l_r}{\partial \tau} \\ A_r^{\alpha\tau}\frac{\partial l_r}{\partial \alpha} + A_r^{\tau\tau}\frac{\partial l_r}{\partial \tau} \end{pmatrix}.$$

From this, we can obtain results for

$$Corr\left(\widehat{\alpha}, \widehat{\alpha}_r\right), Corr\left(\widehat{\alpha}, \widehat{\tau}_r\right), Corr\left(\widehat{\tau}, \widehat{\alpha}_r\right), Corr\left(\widehat{\tau}, \widehat{\tau}_r\right)$$

from the corresponding covariances:

$$\begin{aligned} Cov\left(\widehat{\alpha}, \widehat{\alpha}_r\right) &\simeq Cov\left(\left\{A^{\alpha\alpha}\frac{\partial l}{\partial \alpha} + A^{\alpha\tau}\frac{\partial l}{\partial \tau}\right\}, \left\{A_r^{\alpha\alpha}\frac{\partial l_r}{\partial \alpha} + A_r^{\alpha\tau}\frac{\partial l_r}{\partial \tau}\right\}\right) \\ &\simeq A^{\alpha\alpha}A_r^{\alpha\alpha}Cov\left(\frac{\partial l}{\partial \alpha}, \frac{\partial l_r}{\partial \alpha}\right) + A^{\alpha\alpha}A_r^{\alpha\tau}Cov\left(\frac{\partial l}{\partial \alpha}, \frac{\partial l_r}{\partial \tau}\right) + \\ &\quad A^{\alpha\tau}A_r^{\alpha\alpha}Cov\left(\frac{\partial l}{\partial \tau}, \frac{\partial l_r}{\partial \alpha}\right) + A^{\alpha\tau}A_r^{\alpha\tau}Cov\left(\frac{\partial l}{\partial \tau}, \frac{\partial l_r}{\partial \tau}\right), \end{aligned} \tag{5.35}$$

$$\begin{aligned} Cov(\widehat{\alpha}, \widehat{\tau}_r) &\simeq Cov\left(\left\{A^{\alpha\alpha}\frac{\partial l}{\partial \alpha} + A^{\alpha\tau}\frac{\partial l}{\partial \tau}\right\}, \left\{A_r^{\alpha\tau}\frac{\partial l_r}{\partial \alpha} + A_r^{\tau\tau}\frac{\partial l_r}{\partial \tau}\right\}\right) \\ &\simeq A^{\alpha\alpha}A_r^{\alpha\tau}Cov\left(\frac{\partial l}{\partial \alpha}, \frac{\partial l_r}{\partial \alpha}\right) + A^{\alpha\alpha}A_r^{\tau\tau}Cov\left(\frac{\partial l}{\partial \alpha}, \frac{\partial l_r}{\partial \tau}\right) + \\ &\quad A^{\alpha\tau}A_r^{\alpha\tau}Cov\left(\frac{\partial l}{\partial \tau}, \frac{\partial l_r}{\partial \alpha}\right) + A^{\alpha\tau}A_r^{\tau\tau}Cov\left(\frac{\partial l}{\partial \tau}, \frac{\partial l_r}{\partial \tau}\right), \end{aligned} \tag{5.36}$$

$$\begin{aligned} Cov(\widehat{\tau},\widehat{\alpha}_r) &\simeq Cov\left(\left\{A^{\alpha\tau}\frac{\partial l}{\partial\alpha}+A^{\tau\tau}\frac{\partial l}{\partial\tau}\right\},\left\{A_r^{\alpha\alpha}\frac{\partial l_r}{\partial\alpha}+A_r^{\alpha\tau}\frac{\partial l_r}{\partial\tau}\right\}\right) \\ &\simeq A^{\alpha\tau}A_r^{\alpha\alpha}Cov\left(\frac{\partial l}{\partial\alpha},\frac{\partial l_r}{\partial\alpha}\right)+A^{\alpha\tau}A_r^{\alpha\tau}Cov\left(\frac{\partial l}{\partial\alpha},\frac{\partial l_r}{\partial\tau}\right)+ \\ &\quad A^{\tau\tau}A_r^{\alpha\alpha}Cov\left(\frac{\partial l}{\partial\tau},\frac{\partial l_r}{\partial\alpha}\right)+A^{\tau\tau}A_r^{\alpha\tau}Cov\left(\frac{\partial l}{\partial\tau},\frac{\partial l_r}{\partial\tau}\right), \end{aligned} \tag{5.37}$$

and

$$\begin{aligned} Cov(\widehat{\tau},\widehat{\tau}_r) &\simeq Cov\left(\left\{A^{\alpha\tau}\frac{\partial l}{\partial\alpha}+A^{\tau\tau}\frac{\partial l}{\partial\tau}\right\},\left\{A_r^{\alpha\tau}\frac{\partial l_r}{\partial\alpha}+A_r^{\tau\tau}\frac{\partial l_r}{\partial\tau}\right\}\right) \\ &\simeq A^{\alpha\tau}A_r^{\alpha\tau}Cov\left(\frac{\partial l}{\partial\alpha},\frac{\partial l_r}{\partial\alpha}\right)+A^{\alpha\tau}A_r^{\tau\tau}Cov\left(\frac{\partial l}{\partial\alpha},\frac{\partial l_r}{\partial\tau}\right)+ \\ &\quad A^{\tau\tau}A_r^{\alpha\tau}Cov\left(\frac{\partial l}{\partial\tau},\frac{\partial l_r}{\partial\alpha}\right)+A^{\tau\tau}A_r^{\tau\tau}Cov\left(\frac{\partial l}{\partial\tau},\frac{\partial l_r}{\partial\tau}\right). \end{aligned} \tag{5.38}$$

Therefore, the study of the covariances of the Type II censored and complete MLEs has now been transformed into a study of the correlations of score functions. We first use basic principles to compute these covariances, and then compare the results to those obtained from a generalised version of (5.6).

Covariance from Basic Principles

It follows from the above that we require

$$Cov\left(\frac{\partial l}{\partial\alpha},\frac{\partial l_r}{\partial\alpha}\right),Cov\left(\frac{\partial l}{\partial\alpha},\frac{\partial l_r}{\partial\tau}\right),Cov\left(\frac{\partial l}{\partial\tau},\frac{\partial l_r}{\partial\alpha}\right),Cov\left(\frac{\partial l}{\partial\tau},\frac{\partial l_r}{\partial\tau}\right);$$

we refer back to Section 2.4 for the expressions of these score functions, given at (2.36), (2.37), (2.56) and (2.57). We next consider the expectations emerging from the expansion of the these expressions.

Expectations involved Our calculations of the covariances of the complete and censored score functions in (5.35) to (5.38) will require the following expectations. By manipulating (2.46), (2.48) and (2.49), it is straightforward to compute

$$\begin{aligned} B_1 &= E\left[\sum_{i=1}^{n}\ln X_{i:n}\right]=nE\left[\ln X\right]=-n\left[\frac{\gamma+\psi(\alpha)}{\tau}\right], \\ B_2 &= E\left[\sum_{i=1}^{n}\ln(1+X_{i:n}^{\tau})\right]=nE\left[\ln(1+X^{\tau})\right]=\frac{n}{\alpha}, \\ B_3 &= E\left[\sum_{i=1}^{n}\frac{X_{i:n}^{\tau}\ln X_{i:n}}{1+X_{i:n}^{\tau}}\right]=nE\left[\frac{X^{\tau}\ln X}{1+X^{\tau}}\right]=n\left[\frac{1-\gamma-\psi(\alpha)}{\tau\,(\alpha+1)}\right]. \end{aligned} \tag{5.39}$$

In contrast, the remaining expectations, see below, involve products of summations of varying upper limits:

$$
\begin{aligned}
B_4 &= E\left[\sum_{i=1}^{n} \ln X_{i:n} \sum_{i=1}^{r} \ln X_{i:n}\right], \\
B_5 &= E\left[\sum_{i=1}^{n} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}} \sum_{i=1}^{r} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}}\right], \\
B_6 &= E\left[\sum_{i=1}^{n} \ln X_{i:n} \sum_{i=1}^{r} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}}\right], \\
B_7 &= E\left[\sum_{i=1}^{n} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}} \sum_{i=1}^{r} \ln X_{i:n}\right], \\
B_8 &= E\left[\sum_{i=1}^{n} \ln(1+X_{i:n}^{\tau}) \sum_{i=1}^{r} \ln X_{i:n}\right], \\
B_9 &= E\left[\sum_{i=1}^{n} \ln(1+X_{i:n}^{\tau}) \sum_{i=1}^{r} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}}\right], \\
B_{10} &= E\left[\sum_{i=1}^{n} \ln X_{i:n} \left\{\sum_{i=1}^{r} \ln(1+X_{i:n}^{\tau}) + (n-r)\ln(1+X_{r:n}^{\tau})\right\}\right], \\
B_{11} &= E\left[\sum_{i=1}^{n} \ln(1+X_{i:n}^{\tau}) \left\{\sum_{i=1}^{r} \ln(1+X_{i:n}^{\tau}) + (n-r)\ln(1+X_{r:n}^{\tau})\right\}\right], \\
B_{12} &= E\left[\sum_{i=1}^{n} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}} \left\{\sum_{i=1}^{r} \ln(1+X_{i:n}^{\tau}) + (n-r)\ln(1+X_{r:n}^{\tau})\right\}\right], \\
B_{13} &= E\left[\sum_{i=1}^{n} \ln X_{i:n} \left\{\sum_{i=1}^{r} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}} + (n-r)\frac{X_{r:n}^{\tau} \ln X_{r:n}}{1+X_{r:n}^{\tau}}\right\}\right], \\
B_{14} &= E\left[\sum_{i=1}^{n} \ln(1+X_{i:n}^{\tau}) \left\{\sum_{i=1}^{r} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}} + (n-r)\frac{X_{r:n}^{\tau} \ln X_{r:n}}{1+X_{r:n}^{\tau}}\right\}\right], \\
B_{15} &= E\left[\sum_{i=1}^{n} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}} \left\{\sum_{i=1}^{r} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}} + (n-r)\frac{X_{r:n}^{\tau} \ln X_{r:n}}{1+X_{r:n}^{\tau}}\right\}\right].
\end{aligned}
$$

We proceed to expand the terms in B_4 to B_{15}; take, for example,

$$
\begin{aligned}
B_{11} = & \sum_{i=1}^{r} E\left[(\ln(1+X_{i:n}^{\tau}))^2\right] + 2\sum_{i=1}^{r-1}\sum_{j=i+1}^{r} E\left[\ln(1+X_{i:n}^{\tau})\ln(1+X_{j:n}^{\tau})\right] \\
& + \sum_{i=1}^{r}\sum_{j=r+1}^{n} E\left[\ln(1+X_{i:n}^{\tau})\ln(1+X_{j:n}^{\tau})\right] \\
& + (n-r)\left\{\begin{array}{c} \sum_{i=1}^{r-1} E\left[\ln(1+X_{i:n}^{\tau})\ln(1+X_{r:n}^{\tau})\right] + E\left[(\ln(1+X_{r:n}^{\tau}))^2\right] \\ + \sum_{j=r+1}^{n} E\left[\ln(1+X_{r:n}^{\tau})\ln(1+X_{j:n}^{\tau})\right] \end{array}\right\}, \quad (5.40)
\end{aligned}
$$

in which we require $E\left[(\ln(1+X_{i:n}^{\tau}))^2\right]$ and $E\left[\ln(1+X_{i:n}^{\tau})\ln(1+X_{j:n}^{\tau})\right]$. More generally, we recall from Section 4.3 that the expectations required on expanding B_1 to B_{15} are given in (4.42) and (4.53). As illustrated in Appendix E for the H equations, we can use Mathematica to calculate B_4 to B_{15}, and compare these to their corresponding simulated counterparts

Expectation	Theoretical	Simulated
B_1	-15.2778	-15.2252
B_2	6.2500	6.2542
B_3	-1.3889	-1.3922
B_4	201.5476	203.9854
B_5	1.2870	1.2969
B_6	13.8388	13.8722
B_7	17.1134	17.8016
B_8	-80.8945	-80.9954
B_9	-5.8850	-5.9878
B_{10}	-55.3803	-55.5203
B_{11}	24.3747	24.5236
B_{12}	-5.1896	-5.2319
B_{13}	27.4351	27.3313
B_{14}	-11.3942	-11.4994
B_{15}	2.4440	2.5439

Table 5.19: Numerical checks of expectations B_1 to B_{15} calculated at $r = 15, n = 25$ using Burr data generated with $\alpha = 4, \tau = 3$ and 10^4 replications.

obtained from 10^4 replications. Table 5.19 shows this comparison for $\alpha = 4, \tau = 3, r = 15, n = 25$. We see good agreement between the theoretical and simulated values.

Covariances of the score functions Using the above expectations, and from (2.36), (2.37), (2.56) and (2.57), we can obtain the covariances of the score functions for all combinations of the final and interim MLEs, as shown below.

1. From (2.36) and (2.56), we have

$$\begin{aligned}
Cov\left(\frac{\partial l}{\partial \alpha}, \frac{\partial l_r}{\partial \alpha}\right) &= E\left[\frac{\partial l}{\partial \alpha} \times \frac{\partial l_r}{\partial \alpha}\right] \\
&= E\left[\left(n\alpha^{-1} - T\right) \times \frac{\partial l_r}{\partial \alpha}\right] \\
&= n\alpha^{-1} E\left[\frac{\partial l_r}{\partial \alpha}\right] - E\left[T \times \frac{\partial l_r}{\partial \alpha}\right] \\
&= -E\left[T \times \left(r\alpha^{-1} - T_f - T_c\right)\right] \\
&= -r\alpha^{-1} B_2 + B_{11}. \qquad (5.41)
\end{aligned}$$

2. Similarly, from (2.37) and (2.56), we have

$$
\begin{aligned}
Cov\left(\frac{\partial l}{\partial \alpha}, \frac{\partial l_r}{\partial \tau}\right) &= E\left[\frac{\partial l}{\partial \alpha} \times \frac{\partial l_r}{\partial \tau}\right] \\
&= E\left[\left(n\alpha^{-1} - T\right) \times \frac{\partial l_r}{\partial \tau}\right] \\
&= n\alpha^{-1} E\left[\frac{\partial l_r}{\partial \tau}\right] - E\left[T \times \frac{\partial l_r}{\partial \tau}\right] \\
&= -E\left[T \times \left\{r\tau^{-1} + S_{f,1}(0) - (\alpha+1)T_{f,111} - \alpha T_{c,111}\right\}\right] \\
&= -r\tau^{-1} B_2 - B_8 + B_9 + \alpha B_{14}.
\end{aligned}
$$

3. From (2.36) and (2.57) we may write

$$
\begin{aligned}
Cov\left(\frac{\partial l}{\partial \tau}, \frac{\partial l_r}{\partial \alpha}\right) &= E\left[\frac{\partial l}{\partial \tau} \times \frac{\partial l_r}{\partial \alpha}\right] \\
&= E\left[\left\{n\tau^{-1} + S_1(0) - (\alpha+1)\, T_{111}\right\} \times \frac{\partial l_r}{\partial \alpha}\right] \\
&= n\tau^{-1} E\left[\frac{\partial l_r}{\partial \alpha}\right] + E\left[\left\{S_1(0) - (\alpha+1)\, T_{111}\right\} \times \frac{\partial l_r}{\partial \alpha}\right] \\
&= E\left[\left\{S_1(0) - (\alpha+1)\, T_{111}\right\} \times \left(r\alpha^{-1} - T_f - T_c\right)\right] \\
&= r\alpha^{-1}\left[B_1 - (\alpha+1)\, B_3\right] - B_{10} + (\alpha+1)\, B_{12}.
\end{aligned}
$$

4. From (2.37) and (2.57), we have

$$
\begin{aligned}
Cov\left(\frac{\partial l}{\partial \tau}, \frac{\partial l_r}{\partial \tau}\right) &= E\left[\frac{\partial l}{\partial \tau} \times \frac{\partial l_r}{\partial \tau}\right] \\
&= E\left[\left\{n\tau^{-1} + S_1(0) - (\alpha+1)\, T_{111}\right\} \times \frac{\partial l_r}{\partial \tau}\right] \\
&= n\tau^{-1} E\left[\frac{\partial l_r}{\partial \tau}\right] + E\left[\left\{S_1(0) - (\alpha+1)\, T_{111}\right\} \times \frac{\partial l_r}{\partial \tau}\right] \\
&= E\left[\left\{S_1(0) - (\alpha+1)\, T_{111}\right\} \times \left\{r\tau^{-1} + S_{f,1}(0) - (\alpha+1)T_{f,111} - \alpha T_{c,111}\right\}\right] \\
&= r\tau^{-1}\left[B_1 - (\alpha+1)\, B_3\right] + B_4 - B_6 - \alpha B_{13} + (\alpha+1)\left[B_5 - B_7 + \alpha B_{15}\right].
\end{aligned}
$$

For illustration, we continue to use $\alpha = 4, \tau = 3, r = 15, n = 25$, and compute

$$
\begin{pmatrix} Cov(\frac{\partial l}{\partial \alpha}, \frac{\partial l_r}{\partial \alpha}) & Cov(\frac{\partial l}{\partial \alpha}, \frac{\partial l_r}{\partial \tau}) \\ Cov(\frac{\partial l}{\partial \tau}, \frac{\partial l_r}{\partial \alpha}) & Cov(\frac{\partial l}{\partial \tau}, \frac{\partial l_r}{\partial \tau}) \end{pmatrix} = \begin{pmatrix} 0.9375 & -1.8174 \\ -1.8174 & 6.0499 \end{pmatrix}
$$

so that using

$$\begin{pmatrix} A^{\alpha\alpha} & A^{\alpha\tau} \\ A^{\alpha\tau} & A^{\tau\tau} \end{pmatrix} = \begin{pmatrix} 0.7885 & 0.1671 \\ 0.1671 & 0.1879 \end{pmatrix},$$

$$\begin{pmatrix} A_{15}^{\alpha\alpha} & A_{15}^{\alpha\tau} \\ A_{15}^{\alpha\tau} & A_{15}^{\tau\tau} \end{pmatrix} = \begin{pmatrix} 2.5538 & 0.7672 \\ 0.7672 & 0.3957 \end{pmatrix},$$

and from (5.35) to (5.38) we see

$$\begin{pmatrix} Cov\,(\widehat{\alpha}, \widehat{\alpha}_{15}) & Cov\,(\widehat{\alpha}, \widehat{\tau}_{15}) \\ Cov\,(\widehat{\tau}, \widehat{\alpha}_{15}) & Cov\,(\widehat{\tau}, \widehat{\tau}_{15}) \end{pmatrix} = \begin{pmatrix} 0.7885 & 0.1671 \\ 0.1671 & 0.1879 \end{pmatrix}$$

which, in turn, gives the correlations for all combinations of the final and interim MLEs as

$$\begin{aligned} Corr\,(\widehat{\alpha}, \widehat{\alpha}_{15}) &\simeq \frac{0.7885}{\sqrt{0.7885} \times \sqrt{2.5538}} = 0.5557, \\ Corr\,(\widehat{\alpha}, \widehat{\tau}_{15}) &\simeq \frac{0.1671}{\sqrt{0.7885} \times \sqrt{0.3957}} = 0.2991, \\ Corr\,(\widehat{\tau}, \widehat{\alpha}_{15}) &\simeq \frac{0.1671}{\sqrt{0.1879} \times \sqrt{2.5538}} = 0.2411, \\ Corr\,(\widehat{\tau}, \widehat{\tau}_{15}) &\simeq \frac{0.1879}{\sqrt{0.1879} \times \sqrt{0.3957}} = 0.6891. \end{aligned}$$

As seen in the analysis for Weibull MLEs, we see here numerical values of the covariances of final and interim MLEs are identical to those found for the complete covariance matrix. Thus, it is suitable to next consider the extent to which the conjecture at (5.8) holds for the Burr MLEs.

Covariance from the Generalisation (5.8)

When extended to the Burr distribution, (5.8) would become

$$\begin{pmatrix} Cov\,(\widehat{\alpha}, \widehat{\alpha}_r) & Cov\,(\widehat{\alpha}, \widehat{\tau}_r) \\ Cov\,(\widehat{\tau}, \widehat{\alpha}_r) & Cov\,(\widehat{\tau}, \widehat{\tau}_r) \end{pmatrix} = \begin{pmatrix} A^{\alpha\alpha} & A^{\alpha\tau} \\ A^{\alpha\tau} & A^{\tau\tau} \end{pmatrix}. \tag{5.42}$$

Simplifications of the covariances Alternatively, we can check that (from (5.7))

$$\begin{pmatrix} Cov(\frac{\partial l}{\partial \alpha}, \frac{\partial l_r}{\partial \alpha}) & Cov(\frac{\partial l}{\partial \alpha}, \frac{\partial l_r}{\partial \tau}) \\ Cov(\frac{\partial l}{\partial \tau}, \frac{\partial l_r}{\partial \alpha}) & Cov(\frac{\partial l}{\partial \tau}, \frac{\partial l_r}{\partial \tau}) \end{pmatrix} = \begin{pmatrix} A_{r,\alpha\alpha} & A_{r,\alpha\tau} \\ A_{r,\alpha\tau} & A_{r,\tau\tau} \end{pmatrix} \tag{5.43}$$

n	Theory	r				
	$(= A^{\alpha\alpha})$	$0.2n$	$0.4n$	$0.6n$	$0.8n$	$1.0n$
25	0.7885	38.2848	5.5227	2.5260	1.5883	1.4211
50	0.3942	6.2904	0.8645	0.6358	0.5479	0.5306
100	0.1971	0.6786	0.2940	0.2490	0.2375	0.2250
1000	0.0197	0.0188	0.0184	0.0185	0.0189	0.0192
2500	0.0079	0.0071	0.0080	0.0080	0.0079	0.0079
5000	0.0039	0.0038	0.0038	0.0038	0.0039	0.0039

Table 5.20: Theoretical and simulated values for $Cov(\hat{\alpha}, \hat{\alpha}_r)$ calculated at various r, n using Burr data generated with $\alpha = 4, \tau = 3$ and 10^4 replications.

holds for the Burr distribution. From (5.41), and using (5.39) and (5.40), we may write the first covariance as

$$
\begin{aligned}
Cov\left(\frac{\partial l}{\partial \alpha}, \frac{\partial l_r}{\partial \alpha}\right) &= -\frac{rn}{\alpha^2} + \sum_{i=1}^{r} E\left[(\ln(1+X_{i:n}^{\tau}))^2\right] + 2\sum_{i=1}^{r-1}\sum_{j=i+1}^{r} E\left[\ln(1+X_{i:n}^{\tau})\ln(1+X_{j:n}^{\tau})\right] \\
&\quad + \sum_{i=1}^{r}\sum_{j=r+1}^{n} E\left[\ln(1+X_{i:n}^{\tau})\ln(1+X_{j:n}^{\tau})\right] \\
&\quad +(n-r)\left\{\begin{array}{c} \sum_{i=1}^{r-1} E\left[\ln(1+X_{i:n}^{\tau})\ln(1+X_{r:n}^{\tau})\right] + E\left[(\ln(1+X_{r:n}^{\tau}))^2\right] \\ + \sum_{j=r+1}^{n} E\left[\ln(1+X_{r:n}^{\tau})\ln(1+X_{j:n}^{\tau})\right] \end{array}\right\},
\end{aligned}
$$

which we want to show to equal

$$A_{r,\alpha\alpha} = r\alpha^{-2}.$$

However, due to the various levels of single $\left(\sum_{i=1}^{r}, \sum_{i=1}^{r-1}, \sum_{j=r+1}^{n}\right)$ and double $\left(\sum_{i=1}^{r-1}\sum_{j=i+1}^{r} \text{ and } \sum_{i=1}^{r}\sum_{j=r+1}^{n}\right)$ summations of expectations being involved, it is clear from the above that simplification of $Cov\left(\frac{\partial l}{\partial \alpha}, \frac{\partial l_r}{\partial \alpha}\right)$ to $r\alpha^{-2}$ is very tedious to obtain, and hence is considered elsewhere. Obviously, we then reach the same conclusion about the consideration of $Cov\left(\frac{\partial l}{\partial \alpha}, \frac{\partial l_r}{\partial \tau}\right)$, $Cov\left(\frac{\partial l}{\partial \tau}, \frac{\partial l_r}{\partial \alpha}\right)$ and $Cov\left(\frac{\partial l}{\partial \tau}, \frac{\partial l_r}{\partial \tau}\right)$. Instead, we check for (5.42) via extensive simulation experiments.

Numerical check of (5.42) Here, we assume $\alpha = 4, \tau = 3$ and find the simulated values of $Cov(\widehat{\alpha}, \widehat{\alpha}_r)$, $Cov(\widehat{\alpha}, \widehat{\tau}_r)$, $Cov(\widehat{\tau}, \widehat{\alpha}_r)$ and $Cov(\widehat{\tau}, \widehat{\tau}_r)$ based on 10^4 estimates of $(\widehat{\alpha}_r, \widehat{\tau}_r)$. Tables 5.20 to 5.23 compare these values to their theoretical counterparts, obtained from the complete covariance matrix given at (2.59). We observe generally good agreement between theory and practice across all combinations of r and n considered. This agreement improves as r and n increase. We remark that other values of α and τ were also considered; the results were not reported here because those cases exhibited similar conclusions.

n	Theory	r				
	$(= A^{\alpha\tau})$	$0.2n$	$0.4n$	$0.6n$	$0.8n$	$1.0n$
25	0.1671	0.3268	0.3131	0.3042	0.2946	0.2982
50	0.0835	0.1287	0.1133	0.1127	0.1099	0.1129
100	0.0418	0.0538	0.0496	0.0479	0.0480	0.0459
1000	0.0042	0.0035	0.0036	0.0037	0.0038	0.0040
2500	0.0017	0.0015	0.0017	0.0017	0.0017	0.0017
5000	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008

Table 5.21: Theoretical and simulated values for $Cov(\hat{\alpha}, \hat{\tau}_r)$ calculated at various r, n using Burr data generated with $\alpha = 4, \tau = 3$ and 10^4 replications.

n	Theory	r				
	$(= A^{\alpha\tau})$	$0.2n$	$0.4n$	$0.6n$	$0.8n$	$1.0n$
25	0.1671	6.3057	1.2919	0.4785	0.3260	0.2938
50	0.0835	1.3047	0.1816	0.1318	0.1143	0.1129
100	0.0418	0.1206	0.0596	0.0484	0.0480	0.0459
1000	0.0042	0.0036	0.0038	0.0038	0.0039	0.0040
2500	0.0017	0.0016	0.0017	0.0017	0.0017	0.0017
5000	0.0008	0.0009	0.0008	0.0008	0.0008	0.0008

Table 5.22: Theoretical and simulated values for $Cov(\hat{\tau}, \hat{\alpha}_r)$ calculated at various r, n using Burr data generated with $\alpha = 4, \tau = 3$ and 10^4 replications.

n	Theory	r				
	$(= A^{\tau\tau})$	$0.2n$	$0.4n$	$0.6n$	$0.8n$	$1.0n$
25	0.1879	0.3056	0.2657	0.2503	0.2424	0.2396
50	0.0940	0.1249	0.1097	0.1064	0.1044	0.1047
100	0.0470	0.0533	0.0505	0.0494	0.0496	0.0490
1000	0.0047	0.0045	0.0046	0.0046	0.0046	0.0046
2500	0.0019	0.0019	0.0019	0.0019	0.0019	0.0019
5000	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009

Table 5.23: Theoretical and simulated values for $Cov(\hat{\tau}, \hat{\tau}_r)$ calculated at various r, n using Burr data generated with $\alpha = 4, \tau = 3$ and 10^4 replications.

Implications of (5.42) on the B equations We discuss this only in passing here; the consequences of (5.43) would be that

$$\begin{aligned}-\frac{rn}{\alpha^2}+B_{11}&=r\alpha^{-2},\\ -\frac{rn}{\alpha\tau}-B_8+B_9+\alpha B_{14}&=n\tau^{-1}\left\{(1-\gamma)\,\rho_{0,0}-\rho_{0,1}\right\},\\ -\frac{rn}{\alpha\tau}-B_{10}+(\alpha+1)\,B_{12}&=n\tau^{-1}\left\{(1-\gamma)\,\rho_{0,0}-\rho_{0,1}\right\},\\ -\frac{rn}{\tau^2}+B_4-B_6-\alpha B_{13}+(\alpha+1)\,(B_5-B_7+\alpha B_{15})&=r\tau^{-2}+n\alpha\tau^{-2}\Omega_r,\end{aligned}$$

which lead to

$$\begin{aligned}B_{11}&=r\alpha^{-2}\,(n+1)\,,\\ -B_8+B_9+\alpha B_{14}=-B_{10}+(\alpha+1)\,B_{12}&=n\tau^{-1}\left\{r\alpha^{-1}+(1-\gamma)\,\rho_{0,0}-\rho_{0,1}\right\};\\ B_4-B_6-\alpha B_{13}+(\alpha+1)\,(B_5-B_7+\alpha B_{15})&=\tau^{-2}\,(r+rn+n\alpha\Omega_r)\,.\end{aligned}$$

As previously mentioned at Section 5.3.1 for the H equations, a detailed proof to these results will be given elsewhere.

Implications of (5.42) on the correlations between final and interim MLEs If (5.42) holds, then we have, for large samples,

$$Corr\,(\widehat{\alpha},\widehat{\alpha}_r)\simeq\frac{A^{\alpha\alpha}}{\sqrt{A^{\alpha\alpha}}\times\sqrt{A_r^{\alpha\alpha}}}=\sqrt{\frac{A^{\alpha\alpha}}{A_r^{\alpha\alpha}}}$$

is given by

$$\sqrt{\frac{\alpha^2\,(\alpha+1)^2\,(\alpha+2)\left(1+\frac{\alpha}{\alpha+2}\Omega\right)\left\{r^2+rn\alpha\Omega_r-n^2\alpha^2[(1-\gamma)\,\rho_{0,0}-\rho_{0,1}]^2\right\}}{n\,(r\alpha^2+n\alpha_r^3\Omega_r)\left\{(\alpha+1)^2\,(\alpha+2)+\alpha\,(\alpha+1)^2\,\Omega-\alpha^2\,(\alpha+2)\,[1-\gamma-\psi(\alpha)]^2\right\}}},$$

$$Corr\,(\widehat{\alpha},\widehat{\tau}_r)\simeq\frac{A^{\alpha\tau}}{\sqrt{A^{\alpha\alpha}}\times\sqrt{A_r^{\tau\tau}}}$$

is given by

$$-\left\{1-\gamma-\psi(\alpha)\right\}\times\sqrt{\frac{\alpha^2\,(\alpha+2)\left\{r^2+rn\alpha\Omega_r-n^2\alpha^2[(1-\gamma)\,\rho_{0,0}-\rho_{0,1}]^2\right\}}{rn\left(1+\frac{\alpha}{\alpha+2}\Omega\right)\left\{(\alpha+1)^2\,(\alpha+2)+\alpha\,(\alpha+1)^2\,\Omega-\alpha^2\,(\alpha+2)\,[1-\gamma-\psi(\alpha)]^2\right\}}},$$

$$Corr\,(\widehat{\tau},\widehat{\alpha}_r)\simeq\frac{A^{\alpha\tau}}{\sqrt{A^{\tau\tau}}\times\sqrt{A_r^{\alpha\alpha}}}$$

is given by

$$-\{1-\gamma-\psi(\alpha)\}\times\sqrt{\frac{\alpha^4(\alpha+2)\left\{r^2+rn\alpha\Omega_r-n^2\alpha^2[(1-\gamma)\rho_{0,0}-\rho_{0,1}]^2\right\}}{n\left(r\alpha^2+n\alpha_r^3\Omega_r\right)\left\{(\alpha+1)^2(\alpha+2)+\alpha(\alpha+1)^2\Omega-\alpha^2(\alpha+2)[1-\gamma-\psi(\alpha)]^2\right\}}},$$

and

$$\begin{aligned}Corr\left(\widehat{\tau},\widehat{\tau}_r\right) &\simeq \frac{A^{\tau\tau}}{\sqrt{A^{\tau\tau}}\times\sqrt{A_r^{\tau\tau}}}\\ &\simeq \sqrt{\frac{A^{\tau\tau}}{A_r^{\tau\tau}}}\\ &\simeq \sqrt{\frac{(\alpha+1)^2(\alpha+2)\left\{r^2+rn\alpha\Omega_r-n^2\alpha^2[(1-\gamma)\rho_{0,0}-\rho_{0,1}]^2\right\}}{rn\left\{(\alpha+1)^2(\alpha+2)+\alpha(\alpha+1)^2\Omega-\alpha^2(\alpha+2)[1-\gamma-\psi(\alpha)]^2\right\}}},\end{aligned}$$

in which we refer to Section 2.4.1 for the expressions for $\rho_{k,m}$ and Ω_r, and Section 2.4.3 for Ω.

5.4.2 Link between $\widehat{B}_{0.1}$ and $\widehat{B}_{0.1,r}$

We move on to consider the extent to which $\widehat{B}_{0.1,r}$ can be regarded as a reliable guide to $\widehat{B}_{0.1}$. It will prove more illuminating to here start with a worked example, taking, as before, $\alpha=4,\tau=3,r=15,n=25$ so that, from (2.54),

$$\begin{pmatrix} b_\alpha \\ b_\tau \end{pmatrix}=\begin{pmatrix} -0.0252 \\ 0.1203 \end{pmatrix},$$

and using (5.3), we can approximate $Corr\left(\widehat{B}_{0.1},\widehat{B}_{0.1,15}\right)$ by

$$\frac{\begin{pmatrix} -.0252 \\ .1203 \end{pmatrix}'\begin{pmatrix} .7885 & .1671 \\ .1671 & .1879 \end{pmatrix}\begin{pmatrix} 0.9375 & -1.8174 \\ -1.8174 & 6.0499 \end{pmatrix}\begin{pmatrix} 2.5538 & .7672 \\ .7672 & .3957 \end{pmatrix}\begin{pmatrix} -.0252 \\ .1203 \end{pmatrix}}{\sqrt{\begin{pmatrix} -.0252 \\ .1203 \end{pmatrix}'\begin{pmatrix} .7885 & .1671 \\ .1671 & .1879 \end{pmatrix}\begin{pmatrix} -.0252 \\ .1203 \end{pmatrix}}\times\sqrt{\begin{pmatrix} -.0252 \\ .1203 \end{pmatrix}'\begin{pmatrix} 2.5538 & .7672 \\ .7672 & .3957 \end{pmatrix}\begin{pmatrix} -.0252 \\ .1203 \end{pmatrix}}}$$

which can be shown equal to 0.9049.

Otherwise, the agreement between the simulated values of $Cov\left(\widehat{B}_{0.1},\widehat{B}_{0.1,r}\right)$ with their theoretical counterparts $Var\left(\widehat{B}_{0.1}\right)$, as shown in Table 5.24 for various r and n with 10^4 replications, suggests that it might be possible to extend (5.9) to the Burr case, so that we

n	Theory $\left(=Var\left(\widehat{B}_{0.1}\right)\right)$	r				
		$0.2n$	$0.4n$	$0.6n$	$0.8n$	$1.0n$
25	.002208	.002318	.002320	.002314	.002315	.002318
50	.001104	.001106	.001109	.001109	.001108	.001111
100	.000552	.000558	.000557	.000557	.000559	.000560
1000	.000055	.000055	.000055	.000055	.000055	.000055
2500	.000022	.000022	.000022	.000022	.000022	.000022
5000	.000011	.000011	.000011	.000011	.000011	.000011

Table 5.24: Theoretical and simulated values for $Cov(\hat{B}_{0.1}, \hat{B}_{0.1,r})$ calculated at various r, n using Burr data generated with $\alpha = 4, \tau = 3$ and 10^4 replications.

could use (5.10) to obtain, for large n,

$$Corr\left(\widehat{B}_{0.1}, \widehat{B}_{0.1,r}\right) \simeq \sqrt{\frac{b_\alpha^2 A^{\alpha\alpha} + 2b_\alpha b_\tau A^{\alpha\tau} + b_\tau^2 A^{\tau\tau}}{b_\alpha^2 A_r^{\alpha\alpha} + 2b_\alpha b_\tau A_r^{\alpha\tau} + b_\tau^2 A_r^{\tau\tau}}};$$

in the above example, this correlation is

$$\sqrt{\frac{(-0.0252)^2 \times 0.7885 - 2 \times 0.0252 \times 0.1203 \times 0.1671 + 0.1203^2 \times 0.1879}{(-0.0252)^2 \times 2.5538 - 2 \times 0.0252 \times 0.1203 \times 0.7672 + 0.1203^2 \times 0.3957}} = 0.9049,$$

exactly as before.

5.4.3 Numerical Results

Next, we provide some validation for the above expressions through simulation experiments; we revisit Section 2.4.4 for the 10^4 replications of $\widehat{\alpha}_r, \widehat{\tau}_r$ and $\widehat{B}_{0.1,r}$, generated with $\alpha = 4$ and $\tau = 3$. Tables 5.25 to 5.29 summarise the theoretical (upper) and practical (lower) values for $Corr(\widehat{\alpha}, \widehat{\alpha}_r), Corr(\widehat{\alpha}, \widehat{\tau}_r), Corr(\widehat{\tau}, \widehat{\alpha}_r), Corr(\widehat{\tau}, \widehat{\tau}_r)$ and $Corr\left(\widehat{B}_{0.1}, \widehat{B}_{0.1,r}\right)$ respectively, and consistently show a good agreement between theory and simulation, for all r and n considered. We also noted that, when $r = n$, $Corr(\widehat{\alpha}, \widehat{\tau}_n) = Corr(\widehat{\tau}, \widehat{\alpha}_n)$ becomes

$$Corr(\widehat{\alpha}, \widehat{\tau}) = -\frac{\alpha\{1 - \gamma - \psi(\alpha)\}}{(\alpha+1)\sqrt{1 + \frac{\alpha}{\alpha+2}\Omega}} = 0.4340$$

for all sample sizes. Therefore, this value acts as an upper bound on the strength of cross-parameter correlation there; it depends on α, though is independent of n and τ. In addition, the theoretical correlation values obtained here confirm the pattern observed in Figures 2.13 to 2.17. In particular, agreement so far means that we can now employ these expressions to compute the confidence limits for final estimates, given earlier estimates.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	.1942	.1835	.1806	.1760	.1757	.1756
	.0708	.0418	.0667	.1501	.1518	.1687
$0.4n$	.3616	.3537	.3495	.3456	.3454	.3453
	.0627	.1998	.3038	.3152	.3493	.3334
$0.6n$	.5557	.5495	.5464	.5434	.5432	.5431
	.2065	.4532	.5186	.5154	.5516	.5361
$0.8n$	.7740	.7706	.7689	.7672	.7671	.7671
	.6112	.7281	.7590	.7506	.7708	.7596
$1.0n$	1	1	1	1	1	1
	1	1	1	1	1	1

Table 5.25: Theoretical (upper) and simulated (lower) values of $Corr(\hat{\alpha}, \hat{\alpha}_r)$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 3$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	.1635	.1575	.1541	.1509	.1507	.1506
	.1343	.1420	.1505	.1240	.1331	.1382
$0.4n$	.2348	.2304	.2280	.2258	.2257	.2256
	.2185	.2243	.2299	.1960	.2310	.2076
$0.6n$	.2991	.2959	.2942	.2927	.2926	.2925
	.3175	.3055	.2992	.2611	.3021	.2757
$0.8n$	.3631	.3612	.3602	.3592	.3592	.3591
	.3953	.3824	.3726	.3354	.3623	.3431
$1.0n$	.4340	.4340	.4340	.4340	.4340	.4340
	.5035	.4789	.4369	.4217	.4350	.4191

Table 5.26: Theoretical (upper) and simulated (lower) values of $Corr(\hat{\alpha}, \hat{\tau}_r)$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 3$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	.0843	.0804	.0784	.0764	.0763	.0762
	.0284	.0195	.0254	.0583	.0715	.0822
$0.4n$	.1569	.1535	.1517	.1500	.1499	.1498
	.0357	.0945	.1320	.1306	.1534	.1466
$0.6n$	.2411	.2385	.2371	.2358	.2357	.2357
	.0953	.2114	.2161	.2179	.2422	.2293
$0.8n$	.3359	.3344	.3337	.3330	.3329	.3329
	.3055	.3420	.3286	.3182	.3420	.3192
$1.0n$	.4340	.4340	.4340	.4340	.4340	.4340
	.5035	.4789	.4369	.4217	.4350	.4191

Table 5.27: Theoretical (upper) and simulated (lower) values of $Corr(\hat{\tau}, \hat{\alpha}_r)$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 3$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	.3767	.3628	.3551	.3477	.3472	.3470
	.3365	.3103	.3194	.3264	.3459	.3490
$0.4n$	.5410	.5309	.5254	.5204	.5200	.5199
	.4722	.4891	.5024	.5047	.5284	.5153
$0.6n$	.6891	.6818	.6780	.6744	.6742	.6741
	.6362	.6496	.6611	.6646	.6827	.6704
$0.8n$	.8368	.8323	.8300	.8278	.8276	.8276
	.8060	.8181	.8249	.8209	.8331	.8211
$1.0n$	1	1	1	1	1	1
	1	1	1	1	1	1

Table 5.28: Theoretical (upper) and simulated (lower) values of $Corr(\hat{\tau}, \hat{\tau}_r)$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 3$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	.8470	.8530	.8558	.8582	.8583	.8584
	.8647	.8604	.8612	.8572	.8635	.8561
$0.4n$	.8736	.8706	.8688	.8670	.8669	.8668
	.8791	.8721	.8700	.8648	.8716	.8657
$0.6n$	.9049	.9015	.8997	.8979	.8978	.8977
	.9043	.8993	.8995	.8970	.9009	.8971
$0.8n$	.9469	.9446	.9433	.9422	.9421	.9421
	.9446	.9433	.9438	.9410	.9438	.9404
$1.0n$	1	1	1	1	1	1
	1	1	1	1	1	1

Table 5.29: Theoretical (upper) and simulated (lower) values of $Corr(\hat{B}_{0.1}, \hat{B}_{0.1,r})$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 3$.

5.4.4 Confidence Limits Considerations

We also wish to obtain confidence intervals for $\widehat{\alpha}, \widehat{\tau}, \widehat{B}_{0.1}$, given that we know the values of $\widehat{\alpha}_r, \widehat{\tau}_r, \widehat{B}_{0.1,r}$. We denote

$$\begin{aligned} \Delta_\alpha &= \widehat{\alpha} - \widehat{\alpha}_r, \\ \Delta_\tau &= \widehat{\tau} - \widehat{\tau}_r, \\ \Delta_{B_{0.1}} &= \widehat{B}_{0.1} - \widehat{B}_{0.1,r}, \end{aligned}$$

and assume that each of these differences is Normally distributed with zero mean; provided that (5.42) holds, the variances for Δ_α and Δ_τ are, respectively,

$$Var(\Delta_\alpha) = Var(\widehat{\alpha}_r) - Var(\widehat{\alpha})$$

and

$$Var(\Delta_\tau) = Var(\widehat{\tau}_r) - Var(\widehat{\tau}),$$

and if (5.9) holds, we could approximate the variance of $\Delta_{B_{0.1}}$ by

$$Var(\Delta_{B_{0.1}}) = Var(\widehat{B}_{0.1,r}) - Var(\widehat{B}_{0.1})$$

which, in turn, depends on $Var(\Delta_\alpha), Var(\Delta_\tau)$ and $Cov(\Delta_\alpha, \Delta_\tau)$. These, in turn, yield the approximate 95% confidence intervals for final estimate, given interim estimate; we have

$$\widehat{\Lambda} = \widehat{\Lambda}_r \pm 1.96\sqrt{Var(\Delta_\Lambda)}$$

for $\Lambda = \alpha, \tau, B_{0.1}$. In practice, we would then estimate the true value of α, τ with the MLEs $\widehat{\alpha}_r, \widehat{\tau}_r$ calculated at r.

As our first example, Table 5.30 presents these limits for various r for the arthritic patients data in Table 1.3. As r approaches $n = 50$, we notice fluctuating $\widehat{sd}(\Delta_\alpha)$, as shown in Figure 5.6, but a smoothly decreasing $\widehat{sd}(\Delta_\tau)$, as displayed in Figure 5.7. For $\Delta_{B_{0.1}}$, the interval size reduces steadily as r increases, as shown in Figure 5.8.

We have made checks throughout the theory developed so far, and now we need to validate the resulting confidence intervals using our simulation experiment set up. Again, we plot the 10^4 simulated observations of $\widehat{\alpha}, \widehat{\tau}, \widehat{B}_{0.1}$, and, in each case, record the number of $\widehat{\alpha}, \widehat{\tau}, \widehat{B}_{0.1}$ within the 95% confidence limits evaluated, firstly, at true parameter values $\alpha = 4, \tau = 3$ (corresponding to upper entries), and secondly, at the MLEs $\widehat{\alpha}_r, \widehat{\tau}_r$ (corresponding to lower entries); these results are summarised in Tables 5.31 to 5.33. The difference between entries is due to the penalty on replacing the true values by their MLEs in the calculation, and in particular, for small samples with low to mild censoring, the results are largely distorted by some large values of $\widehat{\alpha}_r$ as shown in Table 2.16. In general, and entirely as expected, the results approach 9500 with increasing n and r.

r	10	20	30	40	50
$\widehat{\alpha}_r$	4.5450	7.9878	8.9031	7.7911	8.2681
$\widehat{sd}(\Delta_\alpha)$	3.9577	4.4005	3.1000	1.4676	0
$\widehat{\tau}_r$	4.1860	4.8626	4.9997	4.8490	5.0006
$\widehat{sd}(\Delta_\tau)$	1.1043	0.8237	0.5823	0.3568	0
$\widehat{B}_{0.1,r}$	0.4080	0.4112	0.4112	0.4113	0.4185
$\widehat{sd}(\Delta_{B_{0.1}})$	0.0201	0.0167	0.0141	0.0109	0

Table 5.30: Standard deviations of $\Delta_\alpha, \Delta_\tau$ and $\Delta_{B_{0.1}}$ for the arthritic patients data.

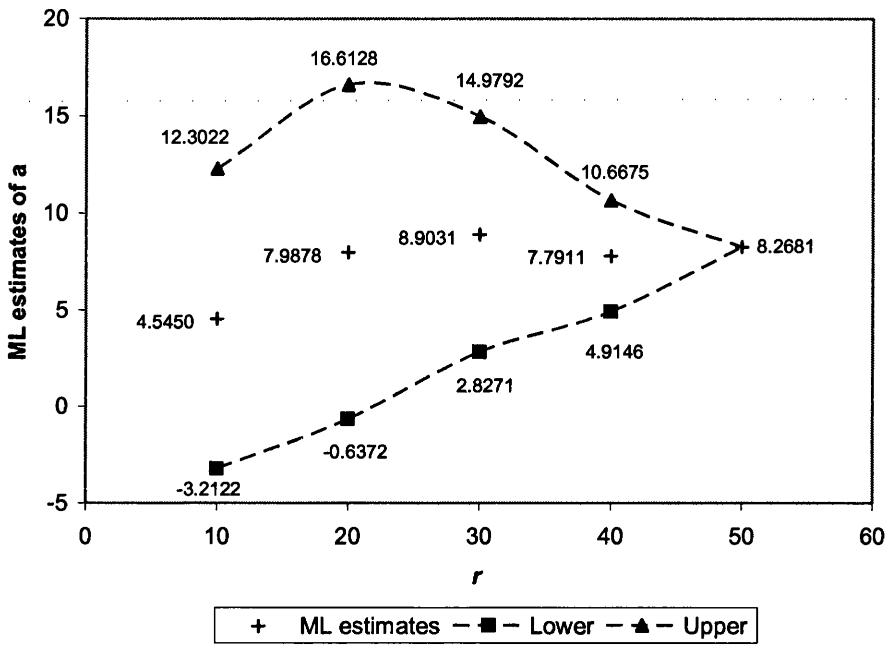

Figure 5.6: $\hat{\alpha}_r$ and 95% confidence limits for $\hat{\alpha}$ given $\hat{\alpha}_r$ for the arthritic patients data.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	6996	7532	8290	9216	9515	9443
	9682	9616	9543	9416	9482	9509
$0.4n$	7443	8094	8597	9384	9507	9439
	9662	9583	9415	9463	9494	9456
$0.6n$	7907	8563	8962	9459	9487	9487
	9636	9496	9426	9468	9520	9485
$0.8n$	8432	8943	9176	9442	9434	9461
	9596	9489	9548	9437	9517	9493

Table 5.31: Number of replications of $\hat{\alpha}$ within the 95% confidence limits based on true α, τ (upper) and $\hat{\alpha}_r, \hat{\tau}_r$ (lower), for Burr data generated with $\alpha = 4, \tau = 3$.

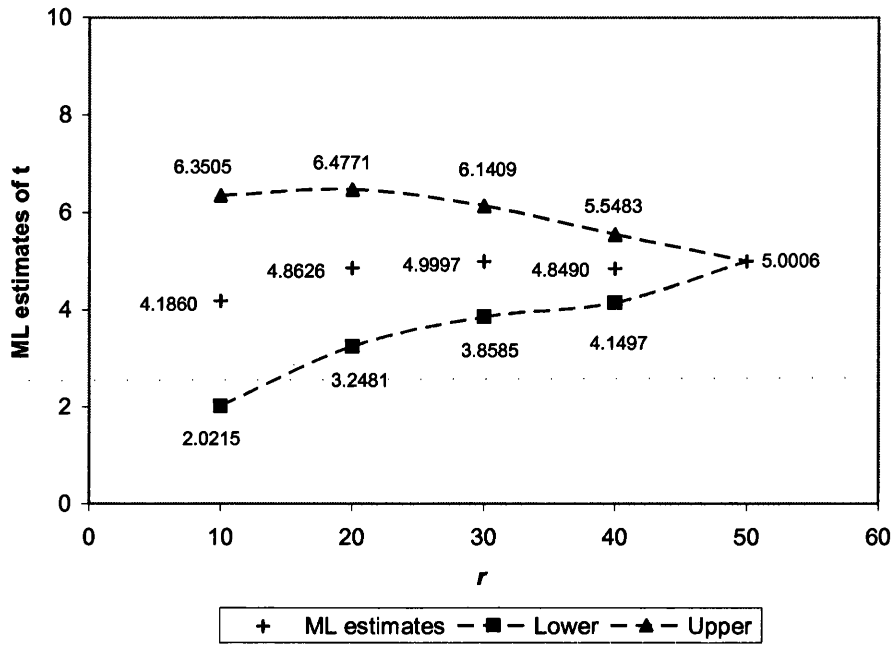

Figure 5.7: $\hat{\tau}_r$ and 95% confidence limits for $\hat{\tau}$ given $\hat{\tau}_r$ for the arthritic patients data.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	7796	8182	8849	9424	9491	9487
	9218	9268	9342	9479	9497	9513
$0.4n$	8296	8910	9155	9447	9523	9458
	9316	9365	9416	9748	9507	9468
$0.6n$	8724	9058	9287	9492	9519	9490
	9413	9406	9456	9480	9514	9495
$0.8n$	9006	9289	9376	9465	9527	9476
	9458	9479	9526	9467	9560	9503

Table 5.32: Number of replications of $\hat{\tau}$ within the 95% confidence limits based on true α, τ (upper) and $\hat{\alpha}_r, \hat{\tau}_r$ (lower), for Burr data generated with $\alpha = 4, \tau = 3$.

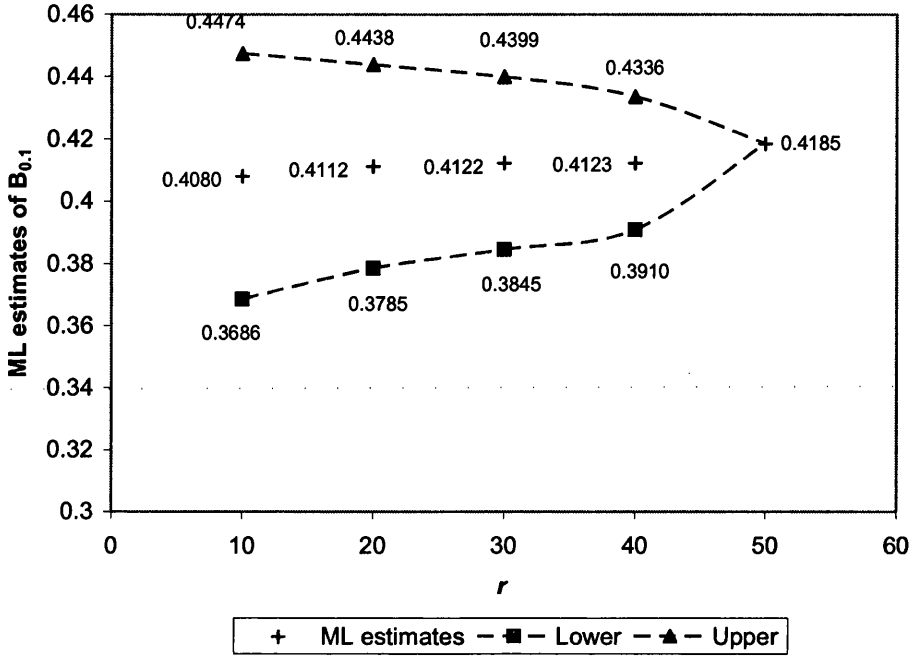

Figure 5.8: $\hat{B}_{0.1,r}$ and 95% confidence limits for $\hat{B}_{0.1}$ given $\hat{B}_{0.1,r}$ for the arthritic patients data.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	9604	9562	9555	9451	9569	9460
	9512	9468	9494	9466	9550	9497
$0.4n$	9485	9487	9512	9473	9535	9495
	9486	9458	9474	9487	9525	9543
$0.6n$	9429	9426	9497	9492	9543	9487
	9472	9440	9465	9499	9542	9522
$0.8n$	9426	9508	9497	9468	9531	9475
	9477	9499	9494	9499	9590	9505

Table 5.33: Number of replications of $\hat{B}_{0.1}$ within the 95% confidence limits based on true α, τ (upper) and $\hat{\alpha}_r, \hat{\tau}_r$ (lower), for Burr data generated with $\alpha = 4, \tau = 3$.

5.5 Practical Implications

Our work in the last two chapters so far has been primarily concerned with establishing and validating theoretical results. In this section, we briefly consider some practical implications of our work, based on published and simulated data. The relevance and importance of the percentile $B_{0.1}$ have been discussed in chapter one and two, and we have further shown in chapter three that the statistical properties of the sampling distributions of $\widehat{B}_{0.1,r}$ are more desirable (in the sense that asymptotic results apply in samples of smaller size) than those of the MLEs of parameters. Therefore, our discussion here focus mainly on the reliability of $\widehat{B}_{0.1,r}$, calculated at censoring level r, can be regarded as a reliable guide to $\widehat{B}_{0.1}$. We have already seen the agreement between $\widehat{B}_{0.1,r}$ and its counterparts for complete samples, but in practical terms, we would like to know the smallest r at which the experiment can be reasonably or safely terminated with the interim analysis still providing a close and reliable guide to the analysis of the final, complete data, as represented by the standard deviation of final estimate, given interim estimate.

5.5.1 Published Data

Epstein's Failure Times Data

We recall from Table 2.1 that $\widehat{B}_{0.1,20} = 11.0523$ is the closest to $\widehat{B}_{0.1} = 11.0512$, but also has the largest (estimated) standard deviation. In contrast, $\widehat{B}_{0.1,10} = 7.1224$ is the farthest from $\widehat{B}_{0.1}$, with $\widehat{sd}(\widehat{B}_{0.1,10}) = 2.2523$, only slightly less than $\widehat{sd}(\widehat{B}_{0.1,20}) = 2.4714$. Hence, intuitively, an experimenter may prefer $\widehat{B}_{0.1,20}$ to $\widehat{B}_{0.1,10}$ as a guide to $\widehat{B}_{0.1}$; in this case, the experiment time would be cut from $X_{49:49} = 354.4$ to $X_{20:49} = 55.6$, with, approximately, a 84% reduction in time. However, in the analysis of the reliability of a sequence of Type II censored estimates, we could also take into account the link between the interim and final estimates before a conclusion can be drawn. Table 5.2 shows that the variation between $\widehat{B}_{0.1,r}$ and $\widehat{B}_{0.1}$ gradually converges to 0 as r approaches $n = 49$. Strikingly, we see that the pattern on standard deviations when $r = 10$ and 20 has reversed; $\widehat{sd}(\widehat{B}_{0.1} - \widehat{B}_{0.1,10}) = 2.0094$ is now slightly larger than $\widehat{sd}(\widehat{B}_{0.1} - \widehat{B}_{0.1,20}) = 1.9013$, but the two values remain similar. Thus, $\widehat{B}_{0.1,10}$ and $\widehat{B}_{0.1,20}$ seem to provide similar amount of information concerning $\widehat{B}_{0.1}$. Statistically, this suggests that it may make no practical difference whether to terminate the experiment after $r = 10$ or $r = 20$, because the resultant censored estimates would be equally reliable in providing a guide to the final estimate. However, from a practical perspective, censoring at $r = 10$ is certainly not the same as $r = 20$, particularly in terms of the experiment time and costs; the former would give an extra saving in experiment time, which is cut from $X_{49:49} = 354.4$ to $X_{10:49} = 15.2$, an additional reduction of 40.4 units compared to $r = 20$. We remark that this information is obtained with hindsight, of course, but may be proven useful to an experimenter when planning a life test; if the precision level is set prior to an experiment, he or she could save the experiment time and costs by terminating the experiment at or around the smallest r for which the data is likely to yield the required level of precision.

Ball Bearings Data

Table 5.15 shows that the interim estimates slowly converge to $\widehat{B}_{0.1}$; the last three failures have a significant effect on the value of $\widehat{B}_{0.1}$, implying that the precise value relies heavily on the last few failures. More strikingly, we see that the precision levels associated with $\widehat{B}_{0.1,8}$ and $\widehat{B}_{0.1,16}$ are quite similar ($\widehat{sd}(\Delta_{B_{0.1}}) = 2.9231$ for $r = 8$ and $\widehat{sd}(\Delta_{B_{0.1}}) = 2.6749$ for $r = 12$); this provides partial answers to some questions posed in Section 5.1, again, we emphasise with the benefit of hindsight. In real life scenarios, censoring often leads to earlier termination of a life test; if the tolerance level is set prior to an experiment, an experimenter could terminate the test sooner than might have been thought. In this example, censoring at $r = 8$ would save the experiment time by roughly 70%, an extra saving of 19% compared to $r = 16$, but, notably, it would also yield interim estimates which are as consistent with the final values as those obtained by censoring at $r = 16$. Moreover, we could also plot the 95% confidence limits for $2 \leq r \leq n = 23$ for the ball bearings data. Figure 5.9 shows that the limits are generally quite flat for censoring values around $r = 5$ to 8, indicating that the precision obtained on censoring about this range of r would be approximately similar, as shown below:

r	2	3	4	5	6	7	8
$X_{r:23}$	28.92	33.00	41.52	42.12	45.60	48.48	51.84
$\widehat{sd}(\Delta_{B_{0.1}})$	3.6921	2.8068	3.7321	2.8481	2.8824	2.8619	2.9232
r	9	10	11	12	13	14	15
$X_{r:23}$	51.96	54.12	55.56	67.80	68.64	68.64	68.88
$\widehat{sd}(\Delta_{B_{0.1}})$	2.5951	2.5435	2.4218	3.0910	2.8689	2.6212	2.4016
r	16	17	18	19	20	21	22
$X_{r:23}$	84.12	93.12	98.64	105.12	105.84	127.92	128.04
$\widehat{sd}(\Delta_{B_{0.1}})$	2.6748	2.6174	2.4468	2.2492	1.9674	1.7195	1.2655

Arthritic Patients Data

Table 5.30 shows that $\widehat{B}_{0.1,r}$ converges to $\widehat{B}_{0.1}$ in an almost horizontal line, and the limits decrease with r at a steady rate up to $r = 40$, after which we see a sharp convergence to 0, corresponding to the case $r = n$. In this example, the case for early censoring is less obvious; as in previous example, the last 10 relief times contain important information regarding the precise value of $\widehat{B}_{0.1}$.

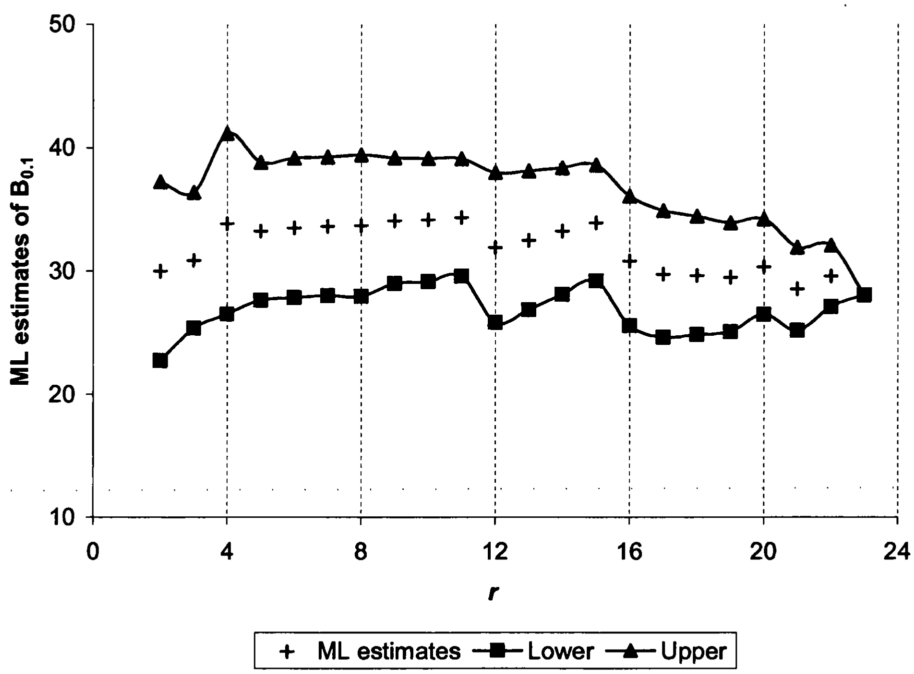

Figure 5.9: $\hat{B}_{0.1,r}$ and 95% confidence limits for $\hat{B}_{0.1}$ given $\hat{B}_{0.1,r}$, for $2 \leq r \leq n = 23$, for the ball bearings data.

5.5.2 Simulation Experiments

The results from the analysis when all the failure times are observed seem to suggest that, for a specified level of precision, it may be possible to design experiments in which early stopping is a viable option. In contrast, as r is to be specified before testing commences, this general conclusion may be less useful in practical terms. However, the practitioner could follow this method to establish a confidence interval for the final estimates, based on the censored estimates calculated at that r; if the precision level meets the required level set prior to running the test, then further tests can be terminated with even smaller r. Otherwise, one could increase the sample size or the censoring number to meet the tolerance level. This information is important for an experimenter, as he or she can then choose an acceptable censoring number and sample size, with the (expected) time required to complete a test generally directly linked to its cost. If the initial cost of test units is cheap compared to experiment time, he or she can increase the initial sample size to obtain results economically.

We use simulations to assess how the increase in censoring level increases the precision of $\widehat{B}_{0.1,r}$ as a guide to $\widehat{B}_{0.1}$, when, as we have seen above, the final few failure times may have a considerable effect on the precise values of the final estimates, and the expected time required to complete the test may also increase considerably. Tables 5.34 to 5.36 give a

r	n					
	25	50	100	1000	2500	5000
$0.2n$	4.2144	2.9800	2.1072	0.6664	0.4214	0.2980
	4.2660	2.9976	2.1019	0.6698	0.4212	0.3003
$0.4n$	2.5808	1.8249	1.2904	0.4081	0.2581	0.1825
	2.5881	1.8198	1.3025	0.4117	0.2590	0.1824
$0.6n$	1.7205	1.2166	0.8603	0.2720	0.1721	0.1217
	1.7075	1.2187	0.8641	0.2746	0.1708	0.1212
$0.8n$	1.0536	0.7450	0.5268	0.1666	0.1054	0.0745
	1.0468	0.7449	0.5230	0.1675	0.1056	0.0744

Table 5.34: Theoretical (upper) and simulated (lower) standard deviations of $\Delta_{B_{0.1}}$ for various r, n, for exponential data generated with $\theta = 100$.

summary of theoretical (upper) and simulated (lower) standard deviations of $\Delta_{B_{0.1}}$ from the exponential, Weibull and Burr distributions respectively, based on 10^4 replications. We see good agreement between theory and simulation. The conclusions reached for a single data set are confirmed here: the standard deviations decrease as r increases, meaning the experimenter would need to compromise between saving time (or cost) and additional information obtained from extra failure times. We also note that the ratio of change in standard deviations to change in censoring proportions decreases with r, suggesting that, if the censoring level has to be small relative to n, say $r \leq 0.4n$, then the experimenter may not need to consider too closely the exact value of r to use.

We can try to place this in a more practical context: in Table 5.35, based on the Weibull distribution with $\theta = 100, \beta = 2$, suppose there are $n = 100$ specimens put on a life test. If this experiment was to run to completion it would take, on average, $E[X_{100:100}] = 226$ units, (we may assume, somewhat arbitrary, that time units are hours - of course, they could be days or even months), obtained on setting $i = n = 100, \theta = 100, \beta = 2$ in (3.3). If, instead, we terminated the experiment after 40 failure times have been recorded, for which the expected experiment time is given by $E[X_{40:100}] = 71$ units, there would be a reduction in duration of 69%, together with a standard deviation $sd(\widehat{B}_{0.1} - \widehat{B}_{0.1,40})$ of 2.1726. Alternatively, we may consider to stop the experiment as soon as 20 failure times have been observed, for which $E[X_{20:100}] = 47$ units and $sd(\widehat{B}_{0.1} - \widehat{B}_{0.1,20}) = 2.2865$, to trade just 5% increase in the standard deviation value for a further 10% reduction in experiment time. In particular, in the case where the initial cost of test units is expensive, the penalty of replacing $\widehat{B}_{0.1,40}$ by $\widehat{B}_{0.1,20}$ as a guide to $\widehat{B}_{0.1}$ may be regarded as less important, compared to the value of test units and/or time saved.

In real life scenarios, censoring often leads to earlier termination of a life test; for a given tolerance level, an experimenter may repeat the analysis described in this chapter at each of a sequence $r = r_1, r_2, \ldots < n$, and examine the pattern of trade off between precision and censoring number, to give the smallest r needed to achieve that level of precision.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	4.7476	3.2735	2.2865	0.7153	0.4521	0.3196
	4.5148	3.2023	2.2575	0.7106	0.4580	0.3230
$0.4n$	4.2678	3.0520	2.1726	0.6917	0.4377	0.3095
	4.2485	3.0616	2.1681	0.6902	0.4424	0.3126
$0.6n$	3.2717	2.6742	1.9085	0.6088	0.3853	0.2725
	3.7794	2.7211	1.9180	0.6082	0.3881	0.2747
$0.8n$	2.8326	2.0459	1.4627	0.4673	0.2957	0.2092
	2.9132	2.0778	1.4608	0.4702	0.2976	0.2087

Table 5.35: Theoretical (upper) and simulated (lower) standard deviations of $\Delta_{B_{0.1}}$ for various r, n, for Weibull data generated with $\theta = 100, \beta = 2$.

r	n					
	25	50	100	1000	2500	5000
$0.2n$	0.0295	0.0203	0.0142	0.0044	0.0028	0.0020
	0.0280	0.0197	0.0139	0.0045	0.0028	0.0020
$0.4n$	0.0262	0.0188	0.0134	0.0043	0.0027	0.0019
	0.0261	0.0187	0.0134	0.0043	0.0027	0.0019
$0.6n$	0.0221	0.0160	0.0114	0.0036	0.0023	0.0016
	0.0227	0.0162	0.0114	0.0037	0.0023	0.0016
$0.8n$	0.0160	0.0116	0.0083	0.0026	0.0017	0.0012
	0.0167	0.0117	0.0083	0.0027	0.0017	0.0012

Table 5.36: Theoretical (upper) and simulated (lower) standard deviations of $\Delta_{B_{0.1}}$ for various r, n, for Burr data generated with $\alpha = 4, \tau = 3$.

5.6 Chapter Summary and Conclusions

We initially examined the relationships between the final and interim estimates of model parameters and a specific percentile in Chapter 2. We have now further studied these relationships, and, in particular, we are able to quantify the correlations between the two sets of ML estimates of model parameters and $B_{0.1}$. Our formulae are relatively straightforward and computationally tractable; provided that the conjectures, obtained from the exponential distribution, at (5.8) and (5.9) hold, we see correlations would follow immediately from the complete and censored EFI matrices. The extension of (5.8) and (5.9) to the Weibull case could be regarded as natural, but it turns out that similar extension might also hold in the Burr distribution. There is obvious scope to assess the extent to which (5.8) and (5.9) hold in other lifetime distributions and censoring regimes; for instance, Finselbach (2007) proves these results for Weibull data obtained from a Type I censoring. This, in turn, yields approximate 95% confidence limits for the final estimate given earlier estimate. We have also shown that these asymptotic results agree with the behaviour observed in simulation experiments for various combinations of censoring number r and (finite) sample size n, and validated the asymptotic 95% confidence limits, from which some issues on practical applications have been discussed.

The main focus of this chapter has been on the effect of censoring has on the precision of a Type II censored estimate as an estimate to its complete counterpart, and to establish some guidelines on an optimal censoring number to stop an experiment, which maximises the practical benefits while minimising the loss of statistical information. The decision to concentrate on $B_{0.1}$ was due to its widespread use in practice, and more favourable theoretical properties (in the sense that asymptotic results apply in samples of smaller size) against those of the MLEs of parameters. Based on published examples, and using hindsight, we see that if the tolerance level is set prior to an experiment, the practitioners could terminate the test sooner than might have been thought. While with simulated data, we noted the reduction in standard deviation of $\Delta_{B_{0.1}}$ changes with the reduction in censoring level at a varying rate, notably slow when $r \leq 0.4n$. Therefore, if r/n is small, perhaps due to high initial costs of running the experiment, it transpires that, as r decreases, there may be little loss in information extractable from the observed failure times, in comparison with the benefits gained from the reduction in the test duration and the costs of running the test.

Overall, the results are encouraging, suggesting that for a given set of interim estimates and the precision required, it may be possible to design experiments in which early stopping was a viable option. However, further work is required before a firm conclusion can be reached, especially with different combinations of parameter values to cover, as wide as possible, the whole range of real life scenarios.

Chapter 6

Summary and Conclusions

In this final chapter, we provide an overview of our work and present our conclusions. We begin by summarising our interests and aims, and discuss the extent to which each of these was achieved. We then present an overall conclusion, and finish by considering further areas of investigation.

6.1 Summary

Reliability Distributions

Our work has centred around three reliability models, namely, the exponential, Weibull and Burr distributions; some basic concepts for each of these models were given in Chapter 1, along with a list of relevant mathematical functions and properties of order statistics. We saw that, due to its lack-of-memory property, results for the exponential lifetime data are relatively straightforward to obtain. Then, by exploiting the relationship between Weibull and exponential random variables, these results transformed easily to the Weibull distribution. In contrast, despite of the limiting relationship between Burr and Weibull, it was not as simple to obtain results for the Burr distribution; throughout, we have seen that the analysis of Type II censored Burr data was considerably more complicated.

ML Estimation for Model Parameters and Percentiles

Although ML estimation for both complete and censored samples is widely discussed in the literature, discussion has focused in detail on both the theoretical and (to a somewhat lesser extent) the numerical aspects of this method. In Chapter 2, we considered the mathematical and computational methodology involved in ML estimation of parameters and percentile functions for some reliability models mentioned in Chapter 1, under a Type II censoring regime; some corresponding results under complete censoring were briefly presented, obtained simply by setting $r = n$. We have concentrated on $B_{0.1}$ throughout this thesis, but the details and principles can be easily adopted to other percentiles; depending on the form of the cdf, $B_{0.1}$ is often a non-linear function of model parameters, and we can linearise this relationship by considering a first order Taylor series expansion of $B_{0.1}$. The theoretical EFI (matrix) for exponential and Weibull distributions was first considered, before moving on to the Burr distribution. Here, we have established closed-form expressions for the elements of the Type II censored EFI matrix for the Burr distribution, previously unobtained by Wingo (1993). This, in turn, yielded asymptotically valid variances and covariances of the MLEs of parameters and percentile function.

Numerical examples were presented using published data to illustrate the relevant calculations involved. In addition, asymptotic results were validated through extensive simulations with 10^4 replications, for various combinations of sample of finite size, censoring number and parameter values. We found generally good agreement between theory and practice (simulations), and this improved as the n and r increased. We also observed smaller standard deviations by increasing n and r, as expected.

Asymptotic Normality of MLEs

In many situations, it is not enough to have merely an estimate of the parameter or $B_{0.1}$, but some indication of the likely accuracy of these estimates is also desirable. The asymptotic Normality of MLEs is widely known, and is often used in practice to obtain approximate confidence regions around parameters; this leads to symmetrical confidence intervals for a single-parameter case, and elliptical confidence regions for two. However, there appears to be no detailed information on how large a sample needs to be for this large-sample approximation to hold. We investigated, by means of extensive simulation studies, the distributions of the MLEs of parameters and $B_{0.1}$, with particular emphasis on the rate at which the MLEs approach Normality, and the effect of Type II censoring has on the progress towards Normality; this extended the work introduced by Chua *et al.* (2007), where the emphasis focused only on Type II censored Weibull data.

The formal tests in Chapter 3 revealed that, unless the sample size is very large, the hypothesis that the distribution of the MLE is Normal is unlikely to be formally accepted; this covered both univariate distributions of parameters and percentiles, as well as the joint distributions of parameters. In general, the progress towards Normality was slow, and censoring further impairs this progress. Furthermore, univariate tests showed that the non-Normality in the distributions of the MLEs was partially attributable to the problem of right skewness, and hence the scatter plots for joint distributions did not become elliptical until samples were very large. On the other hand, the distribution of $\widehat{B}_{0.1,r}$ exhibited milder right skewness, and converged to Normality at both earlier censoring (smaller r) and smaller sample size than in the case of model parameters. Despite these poor approximations to the Normal distribution, the corresponding probability regions obtained were shown to provide a good coverage of the ML estimates of parameters, but the non-elliptical shape of the distribution was not well represented. This resulted in the investigation of an alternative method to assess the precision in estimates of parameters in relatively small and/or highly censored samples.

An Alternative Measure of Precision - Relative Likelihood Contour Plots

Where the asymptotic Normality assumption is implausible in samples of small to moderate sizes, we have proposed an alternative measure of precision using the relative likelihood function and its contour plots; this is essentially the second part of Chapter 3, in which we considered the effects of varying r on the shape and the size of the relative likelihood contours. As an extension to Watkins & Leech (1989), in which the algorithm concentrated on Weibull data, we have outlined an automatic algorithm for drawing relative likelihood contours using the IML procedure in SAS for Burr data subject to Type II censoring, and we saw smaller and more elliptical contours as r increases. In addition to applying this method to single sets of data, we have adapted the method to provide approximate confidence regions for the MLEs of parameters, by introducing what we term idealised samples; we computed the expected order statistics in Mathematica, subjected these to censoring as appropriate, and then used the resulting values as data for plotting the expected relative likelihood contours.

When comparing with Normal theory probability ellipses, we note that the two curves overlap, and as n and r increase, the overlap increases, with the relative likelihood contour moving towards the asymptotic Normal ellipse. We have also shown that in small or highly censored samples, the non-elliptical nature of the relative likelihood contours captured more accurately the behaviour of the Type II censored MLEs, where the asymptotic Normality assumption seems to be invalid. Despite the more complicated computations involved in the relative likelihood approach as a measure of precision, with the use of the algorithm described and the computational advances today, we can reasonably recommend the use of relative likelihood contour as an alternative to quantify the precision in estimates of parameters in small to moderate samples, where large-sample Normal theory fails.

Moments and Product Moments of Order Statistics

We have seen much of the theoretical development in this thesis involved taking expectations and joint expectations of order statistics, such as the derivation of the elements of the EFI matrix. Chapter 4 then outlined some useful preliminary work for studying the correlations between interim and final estimates of parameters and $B_{0.1}$, equivalently, the correlations of the two sets of score functions. Unlike the Weibull distribution, which is linked to the standard exponential distribution (with a relatively simple pdf and cdf), corresponding analysis for the Burr distribution proved to be considerably more involved.

We considered two approaches, via direct integration (direct method) and repetitive differentiations (derivatives method), to obtain the moments and product moments of order statistics required in Chapter 5. Despite the more complicated functions (integrations of exponential integrals and the hypergeometric series) being involved in the direct method, this approach generally consumed less computation time than the derivatives method when implemented in Mathematica, and hence is more feasible in practice. On the other hand, due to its flexibility in dealing with the logarithms and/or powers of order statistics, the derivatives method was shown more useful in establishing the joint expectations of Burr order statistics. We sought to validate these new results by means of simulation experiments. We observed generally excellent agreement between theoretical and simulated results for various combinations of order statistics, sample size and parameter values; despite some computational problems for large sample sizes, we have covered most sample sizes and ranges of censoring likely to be encountered in practice.

The Reliability of Type II Censored Reliability Analyses

We have observed in Chapter 2 some linear relationship between the interim and final estimates of parameters and $B_{0.1}$, increasingly evident as r converged to n. Chapter 5 essentially carried on where Chapter 2 left off, by extending the work by Chua & Watkins (2007) and Chua & Watkins (2008a,b). We established a method to quantify the link between censored and complete estimates, and used this to measure the precision in using a Type II censored analysis as a guide to the final analysis.

We began with the exponential distribution, for which we benefited from its powerful lack-of-memory property, and employed the usual asymptotic relationship linking the MLE, the EFI and the score function to calculate the correlation between the final and interim estimates of parameter and $B_{0.1}$. We saw that our problem could be transformed into a study of the correlations between final and interim score functions, in which the covariance of the two sets of score functions was shown to simplify to the censored EFI, see (5.6). This correlation, in turn, provided the approximate 95% confidence limits for the final estimate given earlier estimate, as a measure of precision of the censored estimate in estimating the complete estimate.

We then followed the same approach as in the exponential distribution in the consideration of the Weibull and Burr distributions. With two parameters, the analysis proved to be more detailed, but the same concepts held, and the corresponding relationships between final and interim estimates were found. We first considered correlations from basic principles, using various expectations and joint expectations of order statistics outlined in Chapter 4. We then considered a possible generalisation to (5.6), in which we saw correlations between censored and complete MLEs might follow immediately from the two sets of EFI matrices, with the theory presented in Chapter 2. Nonetheless, further work is needed to establish analytically the simplification in the correlations.

We sought to validate these new theoretical expressions with simulation experiments. We established that these asymptotic results agreed with the behaviour observed in simulations for various combinations of censoring number and sample size, providing confirmation to our results, but the agreement was generally good even for early censoring and samples of small to moderate sizes. Moreover, the confidence intervals for the final estimate given interim estimate have shown to provide a reasonable coverage of the final estimates.

Practical Implications - Planning the Experiments

Confidence limits for the final estimate given earlier estimate were presented for publishes examples, and some practical issues on experimental design were identified and discussed, with particular stress on providing a guide to experimenters wishing to know the smallest number of failures at which a trial can be reasonably or safely terminated, but where the censored analysis still provides a reliable guide to the analysis of the final, complete data.

The relevance and importance of the percentile $B_{0.1}$ have been discussed in Chapters 1 and 2, and we have further shown that the statistical properties of $\widehat{B}_{0.1,r}$ were more desirable than those of the MLEs of parameters; for instance, as seen in Chapter 3, the sampling distribution of $\widehat{B}_{0.1,r}$ converged to Normality more rapidly, while in Chapter 5 we found the values of $Corr\left(\widehat{B}_{0.1}, \widehat{B}_{0.1,r}\right)$ were notably greater than $Corr\left(\widehat{\pi}, \widehat{\pi}_r\right)$, for each combination of r and n considered. Therefore, since we have established a link between censored and complete estimates of $B_{0.1}$, this provided a suitable ground to identifying an optimum number of failures to censor.

In published data, we spotted, more than once, situations where two distinct censoring numbers produced roughly equal values of $\widehat{sd}(\Delta_{B_{0.1}})$. Hence, if the tolerance level is set prior to an experiment, an experimenter could terminate the test sooner than might have been thought. This information may be viewed as a consequence of hindsight, but results obtained from simulation study were equally, if not more, encouraging. We saw a trade off between censoring level and $sd(\Delta_{B_{0.1}})$, where the extent of trade-off varied with the ratio of censoring number to sample size. It suggested that, if the censoring number was expected to be small in relative to the sample size, say $r/n \leq 0.4$, then it might be viable to forgo the precision obtained in using $\widehat{B}_{0.1,r}$ as a guide to $\widehat{B}_{0.1}$, for a reduction in the test duration and the cost of running the test. The reason has been the rate of decrease in the (expected) experiment time was larger than that of the $sd(\Delta_{B_{0.1}})$, when censoring number is relatively small. Therefore, our analysis indicated that the combination of r and n, specifically, the value of r/n, may have some role in the final decision-making process. Our results showed that, for low censoring values (relative to the sample size), the reduction in expected experiment time outweighs the loss in information. This transpires that it may be possible to design experiments in which early stopping is a viable option.

We remark that the scope of these practical investigations was rather narrow, but, real-life experiments are many and varied, and we can only partially cover the wide range of possible parameter combinations that are used in the real world. Nevertheless, our methods and results have provided practitioners with some insight into the roles of censoring number r and sample size n in a Type II censoring setting.

6.2 Conclusions

The principle aim of this thesis was to consider the relationship between final ($r = n$) and interim ($r < n$) results, and hence the extend to which an interim estimate - here, using information based on Type II censoring - can be regarded as a reliable guide to the final estimate. Our investigations, although based on limited parameter values, illustrated useful conclusions on the conduct of experiments under such censoring plan, and, consequently, are of potential value to a practitioner who, prior to carrying out an experiment, would like to know what combination of censoring level and sample size would return the most information about the final results.

Therefore, throughout this thesis, our primary focus has been on the development of theoretical results, and the validation of these through extensive simulation experiments. Having laid down the necessary groundwork, we first considered the computational and numerical aspects of maximum likelihood estimation, which often overlooked in published discussions. On a whole, Type II censoring, in turn, induced the study on expectations and joint expectations of order statistics, has not caused any special difficulty in the derivation of EFI matrix, owing to the connection between the distribution of first order statistic and the underlying distribution. Besides MLEs of parameters, we also obtained an estimate for the 10^{th} percentile of failure times, $B_{0.1}$, since practitioners would typically wish to make inferences on the running time of the experiment.

In order to assess our ability to make small sample theoretical inspections, we then proceeded to investigate, by means of a detailed simulation study, the extent to which asymptotic Normality of MLE applies in samples of finite size, subject to Type II censoring. We concluded that asymptotic Normality assumption is improbable in small samples, and recommended the use of relative likelihood contour plots to obtain approximate confidence regions of parameters in relatively small and/or highly censored samples (Chua *et al.*, 2007).

We have obtained general expressions for the correlations between the interim and final MLEs of model parameters, and a particular percentile. We noted that the evaluation of these expressions via basic principles involved some lengthy algebra, chiefly due to various moments and product moments of order statistics required. But the derivatives method provided an alternative, and has shown to be useful particularly for the Burr distribution. Furthermore, a possible generalisation from the exponential distribution suggests that correlations between the two sets of estimates might follow immediately from the EFI matrix. These, in turn, gave us approximate 95% confidence limits for the final estimate given interim estimate, as an indication of the precision with which we can make statements on final estimates, based on interim estimates.

Overall, our results are reasonably encouraging. The standard deviation of final estimate given interim estimate decreases with censoring number at varying rates, depending on the ratio of censoring number to sample size, and, in particular, the final few failure times carry important information regarding the precise values of the final estimates. The practical consequences of our work are as follows: for any r specified before testing commences, an experimenter is now able to gain, from the resultant interim estimates, some information concerning the final failure time. If the precision level is also set prior to the experiment, as always in practice, he or she could save the experiment time and costs by terminating the experiment at or around the smallest r for which the data is likely to yield the required level of precision. The trade-off between early censoring and precision can be formalised by considering the relationship between r and the standard deviation of final estimate given interim estimate, which can then be used as the criterion for assessing competing experiment designs. Practitioners can use the measure of precision discussed in this thesis to decide whether an interim experiment results are sufficient to make inferences from, or whether the experiment should continue to allow more items to fail.

Finally, we remark that with only one parameter, consideration of negative exponentially distributed lifetimes is clearly of limited practical value; it does, however, provide an useful insight in extending the analysis to other more widely used two-parameter lifetime models. In addition, since our discussion is completely general, the principles followed in this thesis have provided not only some guidance to practitioners wishing to conduct experiments subject to their own circumstances, but also as a basis for further investigation.

6.3 Areas for Future Research

Throughout our work, we have concentrated on the exponential, Weibull and Burr distributions, but other lifetime distributions could well prove to be even more fruitful in terms of quality of fit to data sets, robustness and practical applicability. Depending on the forms of their pdfs and cdfs, we may find the corresponding analysis to be more complex, since, as we have seen in Chapter 4, relatively basic theoretical properties of order statistics, such as their expectations and joint expectations, will involve both the pdf and powers of cdf for the underlying population.

Although attention is restricted primarily to models with two parameters, much of our discussion also applies when there are three or more. Lemon (1975) considers ML estimation for the three-parameter Weibull distribution based on censored samples. As mentioned in Chapter 2, one could extend the two-parameter Burr to a three-parameter model by including a scale parameter ϕ in many different ways. Naturally, its statistical analysis, like the derivation of the EFI matrix, will be more involved; we refer to Watkins (1999) for more details.

Analysis based on an accelerated framework could also be performed; see Nelson (1990) for details on accelerated life testing. Our interest would be extended to cover the effect of a certain combination of accelerated testing techniques, censoring number, sample size and parameter values on the final decision-making process, and, specifically, to determine an optimal censoring value, for a given set of values for the accelerated testing factors. This requires the accelerated version of EFI matrix for complete and censored data, and will, naturally, involve more formidable algebra. For example, Watkins & John (2008) consider constant stress accelerated life tests terminated by a Type II censoring regime at one of the stress levels for data assumed to follow the Weibull distribution.

The mechanism which gives rise to censoring also has much room for further investigation. So far we have considered singly censored samples (on the right) under Type II censoring setting. But, in real life, some test units may have to be removed at different stages in the study for various reasons. Consequently, it is of interest to look at doubly censored samples, and even samples subject to progressive (or multiple) censoring. Some of the recent contributors to the development of the theory underlying ML estimation for this censoring regime have been Tse *et al.* (2000) and Wu (2002) for data assumed to follow a Weibull distribution; Soliman (2005) and Wu *et al.* (2007) for the Burr distribution. Under this extension, we may wish to investigate the extent to which $\widehat{\pi}_i$, conditional on m_i (the number of observations censored at the i^{th} failure), can be regarded as a reliable guide to $\widehat{\pi}$. We can also assess the trade-off between shortening the test duration, by collecting more failure times in the early stage of the test, and the level of precision obtained.

Similarly, our approach can be carried on to the analysis of Type I singly and progressively censored samples, where the length of the experiment t, rather than the number of failure r, is fixed. In contrast to Type II censoring, Type I likelihood consists of independent components with identical or non-identical distributions, based on whether the censoring times are equal or not. Some discussion on the corresponding analysis of reliability data are given by Finselbach & Watkins (2006) and Finselbach (2007) for lifetimes drawn from a Weibull distribution.

In Chapter 3, the problem of right skewness in the distribution of MLE of parameter appears to be consistent with the believe by Billmann *et al.* (1972), that slow convergence to Normality was a consequence of lack of symmetry when the samples were censored on one side (from the right). It follows that it would be of interest to investigate whether the distribution of MLE would be left skewed when the data are censored from the left. We have observed the overlap between relative likelihood contour and the large-sample Normal theory probability ellipse, but the relative size of the two regions, as well as the extent of the overlap in general, could be examined in much further detail.

In Chapter 4 the analysis of reliability of exponential data showed that the covariance of interim and final score functions simplified to the censored EFI. Hence, there is scope to assess the extent to which this simplification holds in other lifetime distributions and censoring regimes.

From a practical perspective, we may wish to quantify the relationship between the value r/n and the standard deviation of final estimate given interim estimate. We also remark that our approach can be easily adopted in the analysis of claim time or survival time data. In this case, typically, life insurers would be interested at drawing inference of $B_{0.9}$, to determine the duration of, say, an endowment policy. Similar comments apply to duration analysis in economics.

Lastly, from a programming perspective, we could use alternative computing packages, such as Matlab, to compute the expectations and joint expectations of order statistics required, to see how its computation time compared to Mathematica. We could also use statistical softwares other than SAS and SPSS; for instance, the R programming language, widely used by statisticians and other practitioners requiring an environment for statistical computing and graphics.

Bibliography

Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. John Wiley and Sons.

Bain, L. J. & Engelhardt, M. (1991). *Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods*. Marcel Dekker, second edn.

Balakrishnan, N. & Rao, C. R. (1998a). *Order Statistics: Theory and Methods. Handbook of Statistics Volume 16*. Elsevier.

Balakrishnan, N. & Rao, C. R. (1998b). *Order Statistics: Applications. Handbook of Statistics Volume 17*. Elsevier.

Billmann, B. R., Antle, C. E. & Bain, L. J. (1972). Statistical inference from censored Weibull samples. *Technometrics* **14**, 831–840.

Burr, I. W. (1942). Cumulative frequency functions. *The Annals of Mathematical Statistics* **13**, 215–232.

Burr, I. W. & Cislak, P. J. (1968). On a general system of distributions: I. Its curve-shaped characteristics; II. The sample median. *Journal of the American Statistical Association* **63**, 627–635.

Caroni, C. (2002). The correct "ball bearings" data. *Lifetime Data Analysis* **8**, 395–399.

Chua, S. J., Finselbach, H. K. & Watkins, A. J. (2007). Small sample properties of maximum likelihood estimators for Type II censored data. *Proceedings of the 22nd International Workshop on Statistical Modelling, Barcelona, Spain* , 172–176.

Chua, S. J. & Watkins, A. J. (2007). The reliability of Type II censored reliability analyses. *Proceedings of the 5th International Mathematical Methods in Reliability Conference, Glasgow, Scotland* .

Chua, S. J. & Watkins, A. J. (2008a). The reliability of Type II censored reliability analyses for Weibull data. *Proceedings of the 23rd International Workshop on Statistical Modelling, Utrecht, Netherlands* , 173–177.

Chua, S. J. & Watkins, A. J. (2008b). The reliability of Type II censored reliability analyses for Weibull data. *Program of the 7th Bernoulli Society-Institute of Mathematical Statistics World Congress in Probability and Statistics, Singapore* , 85–86.

Cohen, A. C. (1991). *Truncated and Censored Samples: Theory and Applications*. Marcel Dekker.

Cook, R. D. & Johnson, M. E. (1986). Generalized Burr-Pareto-logistic distributions with applications to a uranium exploration data set. *Technometrics* **28**, 123–131.

Cox, D. R. & Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall.

Crowder, M. J., Kimber, A. C., Smith, R. L. & Sweeting, T. J. (1991). *Statistical Analysis of Reliability Data*. Chapman and Hall / CRC.

D'Agostino, R. B. (1971). An omnibus test of Normality for moderate and large size samples. *Biometrika* **58**, 341–348.

D'Agostino, R. B., Belanger, A. & D'Agostino Jr, R. B. (1990). A suggestion for using powerful and informative tests of Normality. *The American Statistician* **44**, 316–321.

D'Agostino, R. B. & Pearson, E. S. (1973). Testing for departures from Normality. I. Fuller empirical results for the distribution of b_2 and $\sqrt{b_1}$. *Biometrika* **60**, 613–622.

D'Agostino, R. B. & Stephens, M. A. (1986). *Goodness-of-Fit Techniques*. Marcel Dekker.

David, H. A. & Nagaraja, H. N. (2003). *Order Statistics*. John Wiley and Sons, third edn.

Dey, D. K. & Rao, C. R. (2005). *Bayesian Thinking: Modeling and Computation. Handbook of Statistics Volume 25*. Elsevier.

Epstein, B. (1960). Tests for the validity of the assumption that the underlying distribution of life is exponential: Part II. *Technometrics* **2**, 167–183.

Farnum, N. R. & Booth, P. (1997). Uniqueness of maximum likelihood estimators of the 2-parameter Weibull distribution. *IEEE Transactions on Reliability* **46**, 523–525.

Finselbach, H. (2007). *Some Properties of Maximum Likelihood Estimators of Weibull Parameters With Type I Censored Data*. Ph.D. thesis, University of Wales Swansea.

Finselbach, H. K. & Watkins, A. J. (2006). The reliability of censored reliability analyses. *Proceedings of the 21st International Workshop on Statistical Modelling, Galway, Ireland* , 182–189.

Geller, M. & Ng, E. W. (1969). A table of integrals of the exponential integral. *Journal of Research of the National Bureau of Standards Section B Mathematics and Mathematical Science* **73B**, 191–210.

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. John Wiley and Sons.

Gosh, B. K. & Sen, P. K. (1991). *Handbook of Sequential Analysis*. Marcel Dekker.

Guillera, J. & Sondow, J. (2005). Double integrals and infinite products for some classical constants via analytical continuations of Lerch's transcendent. *Available at http://arxiv.org/abs/math.NT/0506319* .

John, A. (2003). *Maximum Likelihood Estimation in Mis-Specified Reliability Distributions*. Ph.D. thesis, University of Wales Swansea.

John, A. M. & Watkins, A. J. (2006). On the product moments of order statistics from the Weibull distribution. *International Journal of Pure and Applied Mathematics* **30**, 119–131.

Joshi, P. C. (1978). Recurrence relations between moments of order statistics from exponential and truncated exponential distributions. *Sankhya: The Indian Journal of Statistics Series B* **39**, 362–371.

Joshi, P. C. (1982). A note on the mixed moments of order statistics from exponential and truncated exponential distributions. *Journal of Statistical Planning and Inference* **6**, 13–16.

Kalbfleisch, J. G. (1979). *Probability and Statistical Inference II*. Springer, New York.

Khan, A. H. & Khan, I. A. (1987). Moments of order statistics from Burr distribution and its characterizations. *Metron* **45**, 21–29.

Klugman, S. A. (1986). Loss distributions. *Proceedings of Symposia in Applied Mathematics: Actuarial Mathematics* **35**, 31–55.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons.

Lemon, G. H. (1975). Maximum likelihood estimation for the three-parameter Weibull distribution based on censored samples. *Technometrics* **17**, 247–254.

Lieblein, J. (1955). On moments of order statistics from the Weibull distribution. *The Annals of Mathematical Statistics* **26**, 330–333.

Lieblein, J. & Zelen, M. (1956). Statistical investigation of the fatigue life of deep-grove ball bearings. *Journal of Research of the National Bureau of Standards* **57**, 273–316.

Lomax, K. S. (1954). Business failure: Another example of the analysis of failure data. *Journal of American Statistical Association* **49**, 847–852.

Malik, H. J. (1966). Exact moments of order statistics from the Pareto distribution. *Skandinavisk Aktuarietidskift* **49**, 144–157.

Mann, N. R., Schafer, R. E. & Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data.* John Wiley and Sons.

Mardia, K. & Foster, K. (1983). Omnibus tests of multinormality based on skewness and kurtosis. *Communications in Statistics - Theory and Methods* **12**, 207–221.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519–530.

Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979). *Multivariate Analysis.* Academic Press.

Meeker, W. Q. & Nelson, W. (1974). Estimation of Weibull distribution percentiles from censored data. *IEEE Transactions on Reliability* **R-25**, 20–24.

Meeker, W. Q. & Nelson, W. (1977). Weibull variances and confidence limits by maximum likelihood for singly censored data. *Technometrics* **19**, 473–476.

Nelson, W. (1982). *Applied Life Data Analysis.* John Wiley and Sons.

Nelson, W. (1990). *Accelerated Testing - Statistical Models, Test Plans, and Data Analyses.* John Wiley and Sons.

Nelson, W. (2005a). A bibliography of accelerated test plans. *IEEE Transactions on Reliability* **54**, 194–197.

Nelson, W. (2005b). A bibliography of accelerated test plans part II - references. *IEEE Transactions on Reliability* **54**, 370–373.

Patel, J. K., Kapadia, C. H. & Owen, D. B. (1976). *Handbook of Statistical Distributions.* Marcel Dekker.

Pawles, P. & Szynal, D. (2001). Recurrence relations for single and product moments of generalized order statistics from Pareto, generalized Pareto, and Burr distribution. *Communications in Statistics: Theory and Methods* **30**, 739–746.

Peng, D. & MacKenzie, G. (2007). On the analysis of censored reliability data. *Proceedings of the 22nd International Workshop on Statistical Modelling, Barcelona, Spain* , 481–484.

Rodriguez, R. N. (1977). A guide to the Burr type XII distributions. *Biometrika* **64**, 129–134.

SAS (2004). *SAS/IML 9.1 User's Guide.* SAS Institute Inc.

Sinha, S. K. & Moddola, G. S. (1976). A function for size distribution of incomes. *Econometrica* **44**, 963–969.

Slater, L. J. (1966). *Generalized Hypergeometric Functions.* Cambridge University Press.

Soliman, A. A. (2005). Estimation of parameters of life from progressively censored data using Burr XII model. *IEEE Transactions on Reliability* **54**, 34–42.

Srivastava, D. K. & Mudholkar, G. S. (2003). Goodness-of-fit tests for univariate and multivariate Normal models. *Handbook of Statistics 22: Statistics in industry* **22**, 869–906.

Tadikamalla, P. R. (1980). A look at the Burr and related distributions. *International Statistical Review* **48**, 337–344.

Thode, H. C. (2002). *Testing for Normality.* Marcel Dekker.

Thoman, D. R., Bain, L. J. & Antle, C. E. (1969). Inferences on the parameters of the Weibull distribution. *Technometrics* **11**, 445–460.

Tse, S. K., Yang, C. & Yuen, H.-K. (2000). Statistical analysis of Weibull distributed lifetime data under Type II progressive censoring with binomial removals. *Journal of Applied Statistics* **27**, 1033–1043.

Wang, F. K., Keats, J. B. & Zimmer, W. J. (1996). Maximum likelihood estimation of the Burr XII parameters with censored and uncensored data. *Microelectronics Reliability* **36**, 359–362.

Watkins, A. J. (1996). Ramarks on maximum likelihood estimation for the Burr XII distribution. *Microelectronics Reliability* **36**, 1313–1314.

Watkins, A. J. (1997). A note on expected Fisher information for the Burr XII distribution. *Microelectronics Reliability* **37**, 1849–1852.

Watkins, A. J. (1998). On expectations associated with maximum likelihood estimation in the Weibull distribution. *Journal of the Italian Statistical Society* **7**, 15–26.

Watkins, A. J. (1999). An algorithm for maximum likelihood estimation in the three parameter Burr XII distribution. *Computational Statistics Data Analysis* **32**, 19–27.

Watkins, A. J. (2004). On precision in maximum likelihood estimates for the Weibull distribution. *International Journal of Pure and Applied Mathematics* **17**, 175–180.

Watkins, A. J. & John, A. M. (2006). On the expected Fisher information for the Weibull distribution with Type II censored data. *International Journal of Pure and Applied Mathematics* **26**, 93–106.

Watkins, A. J. & John, A. M. (2008). On constant stress accelerated life tests terminated by Type II censoring at one of the stress levels. *Journal of Statistical Planning and Inference* **138**, 768–786.

Watkins, A. J. & Johnson, R. (2002). Two results on hypergeometric functions with applications in reliability analysis. *International Journal of Pure and Applied Mathematics* **3**, 71–89.

Watkins, A. J. & Leech, D. J. (1989). Towards automatic assessment of reliability for data from a Weibull distribution. *Reliability Engineering and System Safety* **24**, 343–350.

Weibull, W. (1939). A statistical theory of the strength of materials. *Ingeniors Vetenskaps Akademien, Handlingar* **151-3**, 1–55.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics* **18**, 293–297.

Wingo, D. R. (1983). Maximum likelihood methods for fitting the Burr Type XII distribution to life test data. *Biometrical Journal* **25**, 77–84.

Wingo, D. R. (1993). Maximum likelihood estimation of Burr XII distribution parameters under Type II censoring. *Microelectronics Reliability* **33**, 1251–1257.

Wolfram, S. (1999). *The Mathematica Book.* Wolfram Media/Cambridge University Press, fourth edn.

Wu, S. J. (2002). Estimations of the parameters of the Weibull distribution with progressively censored data. *Journal of Japan Statistical Society* **32**, 155–163.

Wu, S. J., Chen, Y. J. & Chang, C. T. (2007). Statistical inference based on progressively censored samples with random removals from the Burr Type XII distribution. *Journal of Statistical Computation and Simulation* **77**, 19–27.

Zimmer, W. J., Keats, J. B. & Wang, F. K. (1998). The Burr XII distribution in reliability analysis. *Journal of Quality Technology* **30**, 386–394.

Appendix A : List of Specific Notations

As previously mentioned at Section 1.2.1, this appendix summarises some specific notations used throughout this thesis; for convenience of the readers, these are listed in the order of the chapters, and, where relevant, some remarks have been inserted.

Chapter one:

r	pre-specified number of failures in a Type II censoring regime	$1 \leq r \leq n$
n	sample size	> 0
†	censored value	
g	arbitrary function	
g_a^k	$\frac{\partial^k}{\partial a^k} g$	
g^k	$\frac{d^k}{da^k} g$ where g is univariate	
wrt	with respect to	
$F_{2,1}(z)$	$F_{2,1}(a, b; c; z)$	
$F_{3,2}(z)$	$F_{3,2}(a, b, c; e, f; z)$	
pdf	probability density function	
cdf	cumulative distribution function	
$\boldsymbol{\pi}$ (bold)	vector of π	
$\boldsymbol{\pi}'$	transpose of $\boldsymbol{\pi}$	
π	unknown model parameter	
f	probability density function	
F	cumulative distribution function	
haz	hazard function	
S	survivor function	
B_q	$100q^{th}$ percentile function	$0 < q < 1$
Q	quantile function	
μ_p	p^{th} moment about the origin	$p = 1, 2, 3, \cdots$
$E[g(X)]$	expected value operator for the function $g(x)$	

μ_p^*	p^{th} moment about the mean / p^{th} central moment	$p = 1, 2, 3, \cdots$
μ	mean	$\equiv \mu_1$
γ_1	skewness	
γ_2	kurtosis	
σ^2	variance	$\equiv \mu_2^*$
Var	variance	
θ	Exponential and Weibull scale parameter	> 0
β	Weibull shape parameter	> 0
Burr	Burr Type XII distribution	
α	Burr Type XII and Pareto shape parameter	> 0
τ	Burr Type XII shape parameter	> 0
k	Pareto location parameter	> 0
t	pre-specified stopping time in a Type I censoring regime	> 0
m	unknown number of failures obtained in a Type I censoring regime	$0 \leq m \leq n$
$X_{i:n}$	i^{th} order statistic of a random sample $X_1, X_2, \ldots, X_n$ of size n	$1 \leq i \leq n$
$F_{(i)}$	cumulative distribution function of $X_{i:n}$	$1 \leq i \leq n$
$f_{(i)}$	probability density function of $X_{i:n}$	$1 \leq i \leq n$
$c_{i:n}$	$\frac{n!}{(n-i)!(i-1)!}$	$1 \leq i \leq n$
$F_{(i,j)}$	joint cumulative distribution function of $X_{i:n}$ and $X_{j:n}$	$1 \leq i < j \leq n$
$c_{i,j:n}$	$\frac{n!}{(i-1)!(j-i-1)!(n-j)!}$	$1 \leq i < j \leq n$
$f_{(i,j)}$	joint probability density function of $X_{i:n}$ and $X_{j:n}$	$1 \leq i < j \leq n$
Cov	covariance	
h	arbitrary function	
N	number of replications	10^4
$\pi^{[0]}$	initial value used in the Newton-Raphson method	

Chapter two:

ML estimation	maximum likelihood estimation	
EFI	expected Fisher information matrix	
MLE	maximum likelihood estimator	
L_r ($L \equiv L_n$)	likelihood function	
l_r ($l \equiv l_n$)	log-likelihood function	
$\mathbf{U}_r$ ($\mathbf{U} \equiv \mathbf{U}_n$)	vector of score function	
$\widehat{\pi}_r$ ($\widehat{\pi} \equiv \widehat{\pi}_n$)	maximum likelihood estimator of π	
$\mathbf{A}_r$ ($\mathbf{A} \equiv \mathbf{A}_n$)	expected Fisher information matrix	
$Z_{\lambda/2}$	upper $100\left(1 - \frac{\lambda}{2}\right)$ percentage point of the standard Normal distribution	$0 < \lambda < 1$

$\mathbf{J}_r$ ($\mathbf{J} \equiv \mathbf{J}_n$)	observed Fisher information matrix	
$\mathbf{b}_\pi$	$\frac{\partial B_{0,1}}{\partial \pi}$	
S_r ($S \equiv S_n$)	$\sum_{i=1}^{r} X_{i:n} + (n-r)X_{r:n}$	
W_i	$(n-i+1)(X_{i:n} - X_{i-1:n})$	$i > 1$
sd	standard deviation	
$\widehat{sd}$	estimated standard deviation	
ML estimate	maximum likelihood estimate	
f	failed item	
c	censored item	
$S_{f,j}(k)$	$\sum_{i=1}^{r} X_{i:n}^{k} (\ln X_{i:n})^j$	
$S_{c,j}(k)$	$(n-r)X_{r:n}^{k} (\ln X_{r:n})^j$	
l_r^* ($l^* \equiv l_n^*$)	profile log-likelihood function	
Z	$\left(\frac{X}{\theta}\right)^\beta$ where X follows the Weibull distribution with parameters θ, β	
V	$\ln X_{r:n} - r^{-1} \sum_{i=1}^{r} \ln X_{i:n}$	
ϕ_k	$r^{-1} \sum_{i=1}^{r} (-1)^{r-i} \binom{n}{i-1} \binom{n-i-1}{r-i} [\ln(n+1-i)]^k$	
χ_k^2	chi-square variate with k degrees of freedom	
$S_j\,(k)$	$\sum_{i=1}^{n} X_i^k (\ln X_i)^j$	
T_f	$\sum_{i=1}^{r} \ln(1 + X_{i:n}^{\tau})$	
T_c	$(n-r)\ln(1 + X_{r:n}^{\tau})$	
$T_{f,abc}$	$\sum_{i=1}^{r} \frac{(X_{i:n}^{\tau})^a (\ln X_{i:n})^b}{(1+X_{i:n}^{\tau})^c}$	
$T_{c,abc}$	$(n-r)\frac{(X_{r:n}^{\tau})^a (\ln X_{r:n})^b}{(1+X_{r:n}^{\tau})^c}$	
$\rho_{k,m}$	$\sum_{i=1}^{r} (-1)^{r-i} \binom{n-1}{i-1} \binom{n-i-1}{r-i} \frac{[\psi((n+1-i)\alpha+k)]^m}{(n+1-i)\alpha+k+1}$	
$\varphi_{k,m}$	$\sum_{i=1}^{r} (-1)^{r-i} \binom{n-1}{i-1} \binom{n-i-1}{r-i} \frac{[\psi'((n+1-i)\alpha+k)]^m}{(n+1-i)\alpha+k+1}$	
Ω_r	$\left(\frac{\pi^2}{6} + \gamma^2 - 2\gamma\right) \rho_{1,0} - 2(1-\gamma)\rho_{1,1} + \rho_{1,2} + \varphi_{1,1}$	
T	$\sum_{i=1}^{n} \ln(1 + X_i^{\tau})$	
T_{abc}	$\sum_{i=1}^{n} \frac{(X_i^{\tau})^a (\ln X_i)^b}{(1+X_i^{\tau})^c}$	
Ω	$\frac{\pi^2}{6} + \gamma^2 - 2\gamma - 2(1-\gamma)\psi(\alpha+1) + [\psi(\alpha+1)]^2 + \psi'(\alpha+1)$	
ϕ	three-parameter Burr Type XII (natural) scale parameter	> 0

Chapter three:

g_1	sample estimate of skewness
g_2	sample estimate of kurtosis

m_p^*	p^{th} sample moment about the mean	$p = 1, 2, 3, \cdots$
$\overline{\overline{\pi}}$	sample mean	
S^2	sample variance	
K^2	test statistic for univariate Normality	
$Z(g_1)$	standardised and Normalised skewness	
$Z(g_2)$	standardised and Normalised kurtosis	
$\chi^2_{k,1-\lambda}$	$100(1-\lambda)$ percentage point of the chi-square distribution with k degrees of freedom	$0 < \lambda < 1$
$g_{1,k}$	sample estimates of multivariate skewness	
$g_{2,k}$	sample estimates of multivariate kurtosis	
S^2_W	test statistic for multivariate Normality	
$W(g_{1,k})$	standardised and Normalised multivariate skewness	
$W(g_{2,k})$	standardised and Normalised multivariate kurtosis	
R	relative likelihood function	
λ-contour	$100(1-\lambda)\,\%$ relative likelihood contour	$0 < \lambda < 1$
$\widehat{\pi}^*_r$	MLE of π obtained from idealised sample	
δ	step size used to draw $100(1-\lambda)\,\%$ relative likelihood contour	$= 0.01$
λ-probability ellipse	$100(1-\lambda)\,\%$ probability ellipse	$0 < \lambda < 1$

Chapter four:

A^{pa}_s	$\int_0^\infty x^p (\ln x)^a e^{-sx} dx$
$A^{pa,qb}_{s,t}$	$\int_{y=0}^\infty \int_{x=0}^y x^p (\ln x)^a e^{-sx} y^q (\ln y)^b e^{-ty} dxdy$
I^{pab}_c	$\int_0^\infty x^{p+\tau-1} (\ln x)^a (\ln(1+x^\tau))^b (1+x^\tau)^{-\alpha(n-k)-c-1} dx$
s	$\alpha(n-k)$
E_i	$E\left[\frac{X^p_{i:n}}{(1+X^\tau_{i:n})^c}\right]$
$I^{pab,qde}_{c,f}$	$\int_{y=0}^\infty \int_{x=0}^y \left\{ \begin{array}{c} x^{p+\tau-1} (\ln x)^a (\ln(1+x^\tau))^b (1+x^\tau)^{-\alpha(1+k)-c-1} \\ y^{q+\tau-1} (\ln y)^d (\ln(1+y^\tau))^e (1+y^\tau)^{-\alpha(n-k-1)-f-1} \end{array} \right\} dxdy$
t	$\alpha(n-k-1)$
E_{ij}	$E\left[\frac{X^p_{i:n}}{(1+X^\tau_{i:n})^c} \frac{X^q_{j:n}}{(1+X^\tau_{j:n})^f}\right]$
$I^{p,q}_{s,t}$	$\int_{y=0}^\infty \int_{x=0}^y x^{p+\tau-1} (1+x^\tau)^{s-1} y^{q+\tau-1} (1+y^\tau)^{t-1} dxdy$
b_1	$t + f - \frac{q}{\tau}$
b_2	$\frac{p}{\tau} + \frac{q}{\tau} + 2$
b_3	$t + f + 1$
b_4	$\alpha n + c + f - \frac{p}{\tau} - \frac{q}{\tau}$
b_5	$\alpha n + c + f + 2$

Chapter five:

$Corr$	correlation
H_1	$E\left[\sum_{i=1}^{n} Z_i\right]$
H_2	$E\left[\sum_{i=1}^{n} \ln Z_i\right]$
H_3	$E\left[\sum_{i=1}^{n} Z_i \ln Z_i\right]$
H_4	$E\left[\sum_{i=1}^{n} Z_{i:n} \sum_{i=1}^{r} \ln Z_{i:n}\right]$
H_5	$E\left[\sum_{i=1}^{n} \ln Z_i \sum_{i=1}^{r} \ln Z_{i:n}\right]$
H_6	$E\left[\sum_{i=1}^{n} Z_i \ln Z_i \sum_{i=1}^{r} \ln Z_{i:n}\right]$
H_7	$E\left[\sum_{i=1}^{n} Z_i \left(\sum_{i=1}^{r} Z_{i:n} + (n-r)Z_{r:n}\right)\right]$
H_8	$E\left[\sum_{i=1}^{n} \ln Z_i \left(\sum_{i=1}^{r} Z_{i:n} + (n-r)Z_{r:n}\right)\right]$
H_9	$E\left[\sum_{i=1}^{n} Z_i \ln Z_i \left(\sum_{i=1}^{r} Z_{i:n} + (n-r)Z_{r:n}\right)\right]$
H_{10}	$E\left[\sum_{i=1}^{n} Z_i \left(\sum_{i=1}^{r} Z_{i:n} \ln Z_{i:n} + (n-r)Z_{r:n} \ln Z_{r:n}\right)\right]$
H_{11}	$E\left[\sum_{i=1}^{n} \ln Z_i \left(\sum_{i=1}^{r} Z_{i:n} \ln Z_{i:n} + (n-r)Z_{r:n} \ln Z_{r:n}\right)\right]$
H_{12}	$E\left[\sum_{i=1}^{n} Z_i \ln Z_i \left(\sum_{i=1}^{r} Z_{i:n} \ln Z_{i:n} + (n-r)Z_{r:n} \ln Z_{r:n}\right)\right]$
Δ_π	$\widehat{\pi} - \widehat{\pi}_r$
B_1	$E\left[\sum_{i=1}^{n} \ln X_{i:n}\right]$
B_2	$E\left[\sum_{i=1}^{n} \ln(1 + X_{i:n}^{\tau})\right]$
B_3	$E\left[\sum_{i=1}^{n} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}}\right]$
B_4	$E\left[\sum_{i=1}^{n} \ln X_{i:n} \sum_{i=1}^{r} \ln X_{i:n}\right]$
B_5	$E\left[\sum_{i=1}^{n} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}} \sum_{i=1}^{r} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}}\right]$
B_6	$E\left[\sum_{i=1}^{n} \ln X_{i:n} \sum_{i=1}^{r} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}}\right]$
B_7	$E\left[\sum_{i=1}^{n} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}} \sum_{i=1}^{r} \ln X_{i:n}\right]$
B_8	$E\left[\sum_{i=1}^{n} \ln(1 + X_{i:n}^{\tau}) \sum_{i=1}^{r} \ln X_{i:n}\right]$
B_9	$E\left[\sum_{i=1}^{n} \ln(1 + X_{i:n}^{\tau}) \sum_{i=1}^{r} \frac{X_{i:n}^{\tau} \ln X_{i:n}}{1+X_{i:n}^{\tau}}\right]$
B_{10}	$E\left[\sum_{i=1}^{n} \ln X_{i:n} \left\{\sum_{i=1}^{r} \ln(1 + X_{i:n}^{\tau}) + (n-r)\ln(1 + X_{r:n}^{\tau})\right\}\right]$

B_{11} $E\left[\sum_{i=1}^{n}\ln(1+X_{i:n}^{\tau})\left\{\sum_{i=1}^{r}\ln(1+X_{i:n}^{\tau})+(n-r)\ln(1+X_{r:n}^{\tau})\right\}\right]$

B_{12} $E\left[\sum_{i=1}^{n}\frac{X_{i:n}^{\tau}\ln X_{i:n}}{1+X_{i:n}^{\tau}}\left\{\sum_{i=1}^{r}\ln(1+X_{i:n}^{\tau})+(n-r)\ln(1+X_{r:n}^{\tau})\right\}\right]$

B_{13} $E\left[\sum_{i=1}^{n}\ln X_{i:n}\left\{\sum_{i=1}^{r}\frac{X_{i:n}^{\tau}\ln X_{i:n}}{1+X_{i:n}^{\tau}}+(n-r)\frac{X_{r:n}^{\tau}\ln X_{r:n}}{1+X_{r:n}^{\tau}}\right\}\right]$

B_{14} $E\left[\sum_{i=1}^{n}\ln(1+X_{i:n}^{\tau})\left\{\sum_{i=1}^{r}\frac{X_{i:n}^{\tau}\ln X_{i:n}}{1+X_{i:n}^{\tau}}+(n-r)\frac{X_{r:n}^{\tau}\ln X_{r:n}}{1+X_{r:n}^{\tau}}\right\}\right]$

B_{15} $E\left[\sum_{i=1}^{n}\frac{X_{i:n}^{\tau}\ln X_{i:n}}{1+X_{i:n}^{\tau}}\left\{\sum_{i=1}^{r}\frac{X_{i:n}^{\tau}\ln X_{i:n}}{1+X_{i:n}^{\tau}}+(n-r)\frac{X_{r:n}^{\tau}\ln X_{r:n}}{1+X_{r:n}^{\tau}}\right\}\right]$

Chapter six:

m_i number of observations censored at the i^{th} failure in a progressive censoring regime

Appendix B : SAS Code: Fitting Burr MLEs to Arthritic Patients Data

In this appendix, we give details of the SAS IML algorithm used to locate the MLEs of the Burr parameters and the 10^{th} percentile for the arthritic patients data (see Table 1.3) where $n = 50$, when the data is subject to Type II censoring at the $r = 30^{th}$ failure. Throughout, comments will be inserted and italicised.

```
proc iml;

start burrmle;
n=nrow(bdata);
r=30;
c=n-r;
one=j(r,1,1);
zero=j(c,1,0);
ind=insert(one,zero,r+1);
cdata=(ind#bdata);
t=max(cdata);
lnt=log(t);
lntx2=log(t)*log(t);
lnx=log(bdata);
lncx=(ind#lnx);
lncx2=lncx#lncx;
se=sum(lncx);
```

$\tau^{[0]}$ *is set as* 1

```
tau=1;
```

Stopping criterion is set as $\left| \frac{\frac{dl_r^*}{d\pi}|_{\text{at } \pi=\hat{\pi}^{[j]}}}{\sqrt{-\frac{d^2 l_r^*}{d\pi^2}|_{\text{at } \pi=\hat{\pi}^{[j]}}}} \right| < 10^{-9}$

```
do iter=1 to 500 until (abs(plt/(-pltt)**0.5)<0.000000009);
    term=exp(tau*lncx);
    sfstar=sum(ind#log(term+1));
    sf111=sum(ind#term#lncx/(term+1));
    sf122=sum(ind#term#lncx2/((term+1)#(term+1)));
    termt=exp(tau*lnt);
    scstar=c*sum(log(termt+1));
    sc111=c*sum(termt#lnt/(termt+1));
    sc122=c*sum(termt#lntx2/((termt+1)#(termt+1)));
    sstar=sfstar+scstar;
    s111=sf111+sc111;
    s122=sf122+sc122;
    pl=r*log(tau)+(tau-1)*se-sfstar-r*log(sstar)+r*(log(r)-1);
    plt=r/tau+se-sf111-r*s111/sstar;
    pltt=-r/(tau**2)-sf122-r*(s122/sstar-(s111/sstar)**2);
    tau=tau-plt/pltt;
end;
```

We can now find $\widehat{\alpha}_r$*, the maximised log-likelihood,* $\widehat{B}_{0.1,r}$*, the score functions and the elements of the EFI matrix*

```
alpha=r/sstar;
loglike=r*log(tau*alpha)+(tau-1)*se-(alpha+1)*sfstar-alpha*scstar;
b10=(((0.9)**(-1/alpha))-1)**(1/tau);
da=r/alpha-sstar;
dt=r/tau+se-alpha*s111-sf111;
daa=-r/alpha**2;
dtt=-r/tau**2-alpha*s122-sf122;
dat=-s111;
print alpha tau loglike b10 da dt daa dtt dat;
finish burrmle;
```

Main programme

```
do;
    data= {0.29,0.29,0.34,0.35,0.36,0.36,0.44,0.46,0.49,0.49,
           0.50,0.50,0.52,0.53,0.54,0.55,0.55,0.55,0.56,0.57,
```

```
            0.58,0.58,0.59,0.59,0.60,0.60,0.61,0.61,0.62,0.64,
            0.68,0.70,0.70,0.70,0.71,0.71,0.71,0.72,0.72,0.73,
            0.75,0.75,0.80,0.80,0.81,0.82,0.84,0.84,0.85,0.87};
    bdata=data;
    run burrmle;
end;

quit;
```

Appendix C : SAS Code: Drawing Relative Likelihood Contours for Arthritic Patients Data

In this appendix, we give details of the SAS IML algorithms used to draw the relative likelihood contour for data drawn from the Burr distribution. We continue to use the arthritic patients data, and assume $r = n = 50$ and $\lambda = 0.05$; this yields the approximate 95% confidence regions for (α, τ) under complete sampling. Relative likelihood contours for other combinations of r, n and λ values can be similarly obtained.

Stage 1 The location of $(\widehat{\alpha}_r, \widehat{\tau}_r)$ has been given in Appendix B.

```
alpha=r/sstar;
loglike=r*log(tau*alpha)+(tau-1)*se-(alpha+1)*sfstar-alpha*scstar;
b10=(((0.9)**(-1/alpha))-1)**(1/tau);
```

Instead of the score functions and the EFI matrix, we require $L(\widehat{\alpha}_r, \widehat{\tau}_r)$

```
maxlike=exp(loglike);
print alpha tau loglike maxlike b10;
finish burrmle;
```

Stage 2 Defining the drawing area.

Set contour level as $\lambda = 0.05$

```
p=0.05;
```

Locate $\tau_{\min}$

```
do i=1 to 10;
    j=i*0.1;
    mint=tau*(1-j);
    termmint=exp(mint*lncx);
    sfstarmint=sum(ind#log(termmint+1));
    termtmint=exp(mint*lnt);
    scstarmint=c*log(termtmint+1);
    sstarmint=sfstarmint+scstarmint;
    mintalpha=r/sstarmint;
    loglikemint=r*log(mint*mintalpha)+(mint-1)*se-(mintalpha+1)*sfstarmint
                -mintalpha*scstarmint;
```

This yields $L(\alpha, \tau)$

```
    likemint=exp(loglikemint);
```

This defines $R(\alpha, \tau) = L(\alpha, \tau)/L(\widehat{\alpha}_r, \widehat{\tau}_r)$

```
    relmint=likemint/maxlike;
    print i mint likemint relmint;
    if relmint < p then stop;
end;
```

Locate $\tau_{\max}$

```
do i=1 to 100;
    k=i*0.1;
    maxt=tau*(1+k);
    termmaxt=exp(maxt*lncx);
    sfstarmaxt=sum(ind#log(termmaxt+1));
    termtmaxt=exp(maxt*lnt);
    scstarmaxt=c*log(termtmaxt+1);
    sstarmaxt=sfstarmaxt+scstarmaxt;
    maxtalpha=m/sstarmaxt;
    loglikemaxt=r*log(maxt*maxtalpha)+(maxt-1)*se-(maxtalpha+1)*sfstarmaxt
                -maxtalpha*scstarmaxt;
    likemaxt=exp(loglikemaxt);
    relmaxt=likemaxt/maxlike;
    print i maxt likemaxt relmaxt;
    if relmaxt < p then stop;
end;
```

Locate $\alpha_{\min}$

```
do i=1 to 10;
    j=i*0.1;
    mina=alpha*(1-j);
    minatau=tau;
    do iter=1 to 15;
        termmina=exp(minatau*lncx);
        sfstarmina=sum(ind#log(termmina+1));
        sf111mina=sum(ind#termmina#lncx/(termmina+1));
        sf122mina=sum(ind#termmina#lncx2/((termmina+1)#(termmina+1)));
        termtmina=exp(minatau*lnt);
        scstarmina=c*log(termtmina+1);
        sc111mina=c*(termtmina#lnt/(termtmina+1));
        sc122mina=c*(termtmina#lntx2/((termtmina+1)#(termtmina+1)));
        sstarmina=sfstarmina+scstarmina;
        s111mina=sf111mina+sc111mina;
        s122mina=sf122mina+sc122mina;
        ltmina=r/minatau+se-(mina+1)*sf111mina-mina*sc111mina;
        lttmina=-r/(minatau**2)-(mina+1)*sf122mina-mina*sc122mina;
        minatau=minatau-ltmina/lttmina;
    end;
    loglikemina=r*log(minatau*mina)+(minatau-1)*se-(mina+1)*sfstarmina
                -mina*scstarmina;
    likemina=exp(loglikemina);
    relmina=likemina/maxlike;
    print i mina likemina relmina;
    if relmina < p then stop;
end;
```

Locate $\alpha_{\max}$

```
do i=1 to 100;
    k=i*0.1;
    maxa=alpha*(1+k);
    maxatau=tau;
    do iter=1 to 15;
        termmaxa=exp(maxatau*lncx);
        sfstarmaxa=sum(ind#log(termmaxa+1));
        sf111maxa=sum(ind#termmaxa#lncx/(termmaxa+1));
        sf122maxa=sum(ind#termmaxa#lncx2/((termmaxa+1)#(termmaxa+1)));
```

```
        termtmaxa=exp(maxatau*lnt);
        scstarmaxa=c*log(termtmaxa+1);
        sc111maxa=c*(termtmaxa#lnt/(termtmaxa+1));
        sc122maxa=c*(termtmaxa#lntx2/((termtmaxa+1)#(termtmaxa+1)));
        sstarmaxa=sfstarmaxa+scstarmaxa;
        s111maxa=sf111maxa+sc111maxa;
        s122maxa=sf122maxa+sc122maxa;
        ltmaxa=r/maxatau+se-(maxa+1)*sf111maxa-maxa*sc111maxa;
        lttmaxa=-r/(maxatau**2)-(maxa+1)*sf122maxa-maxa*sc122maxa;
        maxatau=maxatau-ltmaxa/lttmaxa;
    end;
    loglikemaxa=r*log(maxatau*maxa)+(maxatau-1)*se-(maxa+1)*sfstarmaxa
            -maxa*scstarmaxa;
    likemaxa=exp(loglikemaxa);
    relmaxa=likemaxa/maxlike;
    print i maxa likemaxa relmaxa;
    if relmaxa < p then stop;
end;
```

Stage 3 Drawing the 0.05-relative likelihood contour.

Set δ *as* 0.01

```
delta=0.01;
```

Process 1: Find initial point on contour

```
a=1;
b=maxt/tau;
do until (abs(f/(-fb)**0.5)<0.000000009);
    term2=exp(b*tau*lncx);
    sfstar2=sum(ind#log(term2+1));
    sf1112=sum(ind#term2#lncx/(term2+1));
    termt2=exp(b*tau*lnt);
    scstar2=c*log(termt2+1);
    sc1112=c*(termt2#lnt/(termt2+1));
    f=r*log(b*tau*a*alpha)+(b*tau-1)*se-(a*alpha+1)*sfstar2-a*alpha*scstar2
      -r*log(tau*alpha)-(tau-1)*se+(alpha+1)*sfstar+alpha*scstar-log(p);
    fb=r/b+tau*se-(a*alpha+1)*tau*sf1112-a*alpha*tau*sc1112;
    b=b-f/fb;
end;
```

```
alphah=a*alpha;
tauh=b*tau;
fa=r/a-alpha*sfstar2-alpha*scstar2;
gradient=-fa/fb;
anew=a+delta*fb/SQRT(fa**2+fb**2);
bnew=b-delta*fa/SQRT(fa**2+fb**2);
a1=a1//a; b1=b1//b; anew1=anew1//anew; bnew1=bnew1//bnew;
alphah1=alphah1//alphah; tauh1=tauh1//tauh;
anew11=anew1[1:nrow(anew1)]; bnew11=bnew1[1:nrow(bnew1)];
alphah11=alphah1[1:nrow(alphah1)]; tauh11=tauh1[1:nrow(tauh1)];
alphah11=alphah1[1:nrow(alphah1)]; tauh11=tauh1[1:nrow(tauh1)];
matrix=a11||b11||anew11||bnew11||alphah11||tauh11;
varnames='a'//'b'//'anew'//'bnew'//'alphah'//'tauh';
create file1 from matrix[colname=varnames];
append from matrix;
close file1;
```

Process 2: Continue drawing leftward and downward

```
do i=1 to 10000;
    a=anew;
    b=bnew;
    do until (abs(f/(-fb)**0.5)<0.000000009);
        term2=exp(b*tau*lncx);
        sfstar2=sum(ind#log(term2+1));
        sf1112=sum(ind#term2#lncx/(term2+1));
        termt2=exp(b*tau*lnt);
        scstar2=c*log(termt2+1);
        sc1112=c*(termt2#lnt/(termt2+1));
        f=r*log(b*tau*a*alpha)+(b*tau-1)*se-(a*alpha+1)*sfstar2-a*alpha*scstar2
          -r*log(tau*alpha)-(tau-1)*se+(alpha+1)*sfstar+alpha*scstar-log(p);
        fb=r/b+tau*se-(a*alpha+1)*tau*sf1112-a*alpha*tau*sc1112;
        b=b-f/fb;
    end;
    alphah=a*alpha;
    tauh=b*tau;
    fa=r/a-alpha*sfstar2-alpha*scstar2;
    gradient=-fa/fb;
    anew=a+delta*fb/SQRT(fa**2+fb**2);
    bnew=b-delta*fa/SQRT(fa**2+fb**2);
```

```
    a1=a1//a; b1=b1//b; anew1=anew1//anew; bnew1=bnew1//bnew;
    alphah1=alphah1//alphah; tauh1=tauh1//tauh;
    a11=a1[1:nrow(a1)]; b11=b1[1:nrow(b1)];
    anew11=anew1[1:nrow(anew1)]; bnew11=bnew1[1:nrow(bnew1)];
    alphah11=alphah1[1:nrow(alphah1)]; tauh11=tauh1[1:nrow(tauh1)];
    matrix=a11||b11||anew11||bnew11||alphah11||tauh11;
    varnames='a'//'b'//'anew'//'bnew'//'alphah'//'tauh';
    create file2 from matrix[colname=varnames];
    append from matrix;
    close file2;
end;
```

Process 3: Contour is around its extreme left edge

```
do i=1 to 10000;
    b=bnew;
    a=anew;
    do until (abs(f/(-fa)**0.5)<0.000000009);
        term2=exp(b*tau*lncx);
        sfstar2=sum(ind#log(term2+1));
        sf1112=sum(ind#term2#lncx/(term2+1));
        termt2=exp(b*tau*lnt);
        scstar2=c*log(termt2+1);
        sc1112=c*(termt2#lnt/(termt2+1));
        f=r*log(b*tau*a*alpha)+(b*tau-1)*se-(a*alpha+1)*sfstar2-a*alpha*scstar2
          -r*log(tau*alpha)-(tau-1)*se+(alpha+1)*sfstar+alpha*scstar-log(p);
        fa=r/a-alpha*sfstar2-alpha*scstar2;
        a=a-f/fa;
    end;
    alphah=a*alpha;
    tauh=b*tau;
    fb=r/b+tau*se-(a*alpha+1)*tau*sf1112-a*alpha*tau*sc1112;
    gradient=-fa/fb;
    anew=a+delta*fb/SQRT(fa**2+fb**2);
    bnew=b-delta*fa/SQRT(fa**2+fb**2);
    a1=a1//a; b1=b1//b; anew1=anew1//anew; bnew1=bnew1//bnew;
    alphah1=alphah1//alphah; tauh1=tauh1//tauh;
    a11=a1[1:nrow(a1)]; b11=b1[1:nrow(b1)];
    anew11=anew1[1:nrow(anew1)]; bnew11=bnew1[1:nrow(bnew1)];
    alphah11=alphah1[1:nrow(alphah1)]; tauh11=tauh1[1:nrow(tauh1)];
```

```
    matrix=a11||b11||anew11||bnew11||alphah11||tauh11;
    varnames='a'//'b'//'anew'//'bnew'//'alphah'//'tauh';
    create file3 from matrix[colname=varnames];
    append from matrix;
    close file3;
end;
```

Process 4: Continue drawing rightward and upward

```
do i=1 to 10000;
    a=anew;
    b=bnew;
    do until (abs(f/(-fb)**0.5)<0.000000009);
        term2=exp(b*tau*lncx);
        sfstar2=sum(ind#log(term2+1));
        sf1112=sum(ind#term2#lncx/(term2+1));
        termt2=exp(b*tau*lnt);
        scstar2=c*log(termt2+1);
        sc1112=c*(termt2#lnt/(termt2+1));
        f=r*log(b*tau*a*alpha)+(b*tau-1)*se-(a*alpha+1)*sfstar2-a*alpha*scstar2
          -r*log(tau*alpha)-(tau-1)*se+(alpha+1)*sfstar+alpha*scstar-log(p);
        fb=r/b+tau*se-(a*alpha+1)*tau*sf1112-a*alpha*tau*sc1112;
        b=b-f/fb;
    end;
    alphah=a*alpha;
    tauh=b*tau;
    fa=r/a-alpha*sfstar2-alpha*scstar2;
    gradient=-fa/fb;
    anew=a+delta*fb/SQRT(fa**2+fb**2);
    bnew=b-delta*fa/SQRT(fa**2+fb**2);
    a1=a1//a; b1=b1//b; anew1=anew1//anew; bnew1=bnew1//bnew;
    alphah1=alphah1//alphah; tauh1=tauh1//tauh;
    a11=a1[1:nrow(a1)]; b11=b1[1:nrow(b1)];
    anew11=anew1[1:nrow(anew1)]; bnew11=bnew1[1:nrow(bnew1)];
    alphah11=alphah1[1:nrow(alphah1)]; tauh11=tauh1[1:nrow(tauh1)];
    matrix=a11||b11||anew11||bnew11||alphah11||tauh11;
    varnames='a'//'b'//'anew'//'bnew'//'alphah'//'tauh';
    create file4 from matrix[colname=varnames];
    append from matrix;
    close file4;
```

```
end;
```

Process 5: Contour is around its extreme right edge

```
do i=1 to 10000;
    b=bnew;
    a=anew;
    do until (abs(f/(-fa)**0.5)<0.000000009);
        term2=exp(b*tau*lncx);
        sfstar2=sum(ind#log(term2+1));
        sf1112=sum(ind#term2#lncx/(term2+1));
        termt2=exp(b*tau*lnt);
        scstar2=c*log(termt2+1);
        sc1112=c*(termt2#lnt/(termt2+1));
        f=r*log(b*tau*a*alpha)+(b*tau-1)*se-(a*alpha+1)*sfstar2-a*alpha*scstar2
          -r*log(tau*alpha)-(tau-1)*se+(alpha+1)*sfstar+alpha*scstar-log(p);
        fa=r/a-alpha*sfstar2-alpha*scstar2;
        a=a-f/fa;
    end;
    alphah=a*alpha;
    tauh=b*tau;
    fb=r/b+tau*se-(a*alpha+1)*tau*sf1112-a*alpha*tau*sc1112;
    gradient=-fa/fb;
    anew=a+delta*fb/SQRT(fa**2+fb**2);
    bnew=b-delta*fa/SQRT(fa**2+fb**2);
    a1=a1//a; b1=b1//b; anew1=anew1//anew; bnew1=bnew1//bnew;
    alphah1=alphah1//alphah; tauh1=tauh1//tauh;
    a11=a1[1:nrow(a1)]; b11=b1[1:nrow(b1)];
    anew11=anew1[1:nrow(anew1)]; bnew11=bnew1[1:nrow(bnew1)];
    alphah11=alphah1[1:nrow(alphah1)]; tauh11=tauh1[1:nrow(tauh1)];
    matrix=a11||b11||anew11||bnew11||alphah11||tauh11;
    varnames='a'//'b'//'anew'//'bnew'//'alphah'//'tauh';
    create file5 from matrix[colname=varnames];
    append from matrix;
    close file5;
end;
```

Process 6: Accomplish the contour

```
do i=1 to 10000;
```

```
    a=anew;
    b=bnew;
    do until (abs(f/(-fb)**0.5)<0.000000009);
        term2=exp(b*tau*lncx);
        sfstar2=sum(ind#log(term2+1));
        sf1112=sum(ind#term2#lncx/(term2+1));
        termt2=exp(b*tau*lnt);
        scstar2=c*log(termt2+1);
        sc1112=c*(termt2#lnt/(termt2+1));
        f=r*log(b*tau*a*alpha)+(b*tau-1)*se-(a*alpha+1)*sfstar2-a*alpha*scstar2
          -r*log(tau*alpha)-(tau-1)*se+(alpha+1)*sfstar+alpha*scstar-log(p);
        fb=rm/b+tau*se-(a*alpha+1)*tau*sf1112-a*alpha*tau*sc1112;
        b=b-f/fb;
    end;
    alphah=a*alpha;
    tauh=b*tau;
    fa=r/a-alpha*sfstar2-alpha*scstar2;
    gradient=-fa/fb;
    anew=a+delta*fb/SQRT(fa**2+fb**2);
    bnew=b-delta*fa/SQRT(fa**2+fb**2);
    a1=a1//a; b1=b1//b; anew1=anew1//anew; bnew1=bnew1//bnew;
    alphah1=alphah1//alphah; tauh1=tauh1//tauh;
    a11=a1[1:nrow(a1)]; b11=b1[1:nrow(b1)];
    anew11=anew1[1:nrow(anew1)]; bnew11=bnew1[1:nrow(bnew1)];
    alphah11=alphah1[1:nrow(alphah1)]; tauh11=tauh1[1:nrow(tauh1)];
    matrix=a11||b11||anew11||bnew11||alphah11||tauh11;
    varnames='a'//'b'//'anew'//'bnew'//'alphah'//'tauh';
    create file6 from matrix[colname=varnames];
    append from matrix;
    close file6;
end;
```

Appendix D : Expressions for Joint Expectations of Standard Exponential Order Statistics

This appendix gives expressions of the expectations at (4.19), obtained from (4.20).

1. $E\left[Z_{i:n}\ln Z_{j:n}\right]$ (4.19a) is given by

$$
\begin{aligned}
& c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}A^{10,01}_{i+l-k,n-i-l} \\
= \; & c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(i+l-k)^2(n-i-l)(n-k)^2}\times \\
& \left\{\begin{array}{c} -(i+l-k)\left[\gamma(i+l-k)+n-i-l\right] \\ -(n-k)^2\ln(n-i-l)+(n-i-l)(n+i-2k+l)\ln(n-k)\end{array}\right\}
\end{aligned}
$$

2. $E\left[(\ln Z_{i:n})Z_{j:n}\right]$ (4.19b) is given by

$$
\begin{aligned}
& c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}A^{01,10}_{i+l-k,n-i-l} \\
= \; & -c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(n-i-l)^2(n-k)^2}\times \\
& \left\{(2n-i-k-l)\left[\gamma+\ln(n-k)\right]-(n-i-l)\right\}
\end{aligned}
$$

3. $E\left[\ln Z_{i:n}\ln Z_{j:n}\right]$ (4.19c) is given by

$$
\begin{aligned}
& c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}A^{01,01}_{i+l-k,n-i-l} \\
= \; & -c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(i+l-k)(n-i-l)(n-k)}\times \\
& \left\{\begin{array}{c} -(n-k)\left[\begin{array}{c}\gamma\ln(n-i-l)+\ln(i+l-k)\ln(n-i-l) \\ -\ln(i+l-k)\ln(n-k)+Li_2\left(\frac{n-i-l}{n-k}\right)\end{array}\right] \\ -(i+l-k)\left[\gamma^2+\ln^2(n-k)\right]+\gamma(n-2i+k-2l)\ln(n-k)+\frac{\pi^2}{6}(n-i-l)\end{array}\right\}
\end{aligned}
$$

4. $E\left[Z_{i:n}Z_{j:n}\ln Z_{j:n}\right]$ (4.19d) is given by

$$
\begin{aligned}
& c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}A^{10,11}_{i+l-k,n-i-l} \\
= \; & c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(i+l-k)^2(n-i-l)^2(n-k)^3}\times \\
& \left\{\begin{array}{c} -(i+l-k)\left[(\gamma-1)(i+l-k)(3n-2i-k-2l)+(n-i-l)^2\right] \\ -(n-k)^3\ln(n-i-l)+(n-i-l)^2(n+2i-3k+2l)\ln(n-k) \end{array}\right\}
\end{aligned}
$$

5. $E\left[Z_{i:n}(\ln Z_{i:n})Z_{j:n}\right]$ (4.19e) is given by

$$
\begin{aligned}
& c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}A^{11,10}_{i+l-k,n-i-l} \\
= \; & c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(i+l-k)^2(n-i-l)^2(n-k)^3}\times \\
& \left\{\begin{array}{c} (i+l-k)^2\left[(i+l-k)(1-\gamma)+(n-i-l)(4-3\gamma)\right] \\ +\ln(n-k)\left[(n-i-l)^2(n+2i-3k+2l)-(n-k)^3\right] \end{array}\right\}
\end{aligned}
$$

6. $E\left[(\ln Z_{i:n})Z_{j:n}\ln Z_{j:n}\right]$ (4.19f) is given by

$$
\begin{aligned}
& c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}A^{01,11}_{i+l-k,n-i-l} \\
= \; & -c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(i+l-k)(n-i-l)^2(n-k)^2}\times \\
& \left\{\begin{array}{c} -(n-k)^2\left[\ln(n-i-l)\left(\gamma+\ln(i+l-k)\right)+Li_2\left(\frac{n-i-l}{n-k}\right)\right] \\ +\gamma(i+l-k)(3n-2i-k-2l)-(i+l-k)(2n-i-k-l)\left[\gamma^2+\ln^2(n-k)\right] \\ +\ln(n-k)\left[(n-k)^2(1-\gamma+\frac{n-i-l}{n-k}+\ln(i+l-k))+2(\gamma-1)(n-i-l)^2\right] \\ +\frac{\pi^2}{6}(n-i-l)^2 \end{array}\right\}
\end{aligned}
$$

7. $E\left[Z_{i:n}\ln Z_{i:n}\ln Z_{j:n}\right]$ (4.19g) is given by

$$
\begin{aligned}
& c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}A^{11,01}_{i+l-k,n-i-l} \\
= \; & -c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(i+l-k)^2(n-i-l)(n-k)^2}\times \\
& \left\{\begin{array}{c} -(n-k)^2\left[\ln(n-i-l)\left(\gamma-1+\ln(i+l-k)\right)+Li_2\left(\frac{n-i-l}{n-k}\right)\right] \\ -\gamma(i+l-k)(n-2i+k-2l)-(i+l-k)^2\left[\gamma^2+\ln^2(n-k)\right] \\ +\ln(n-k)\left[(n-k)^2\ln(i+l-k)-3(i+l-k)(n-i-l)-(n-i-l)^2\right] \\ +\gamma\ln(n-k)\left[-(i+l-k)^2+2(i+l-k)(n-i-l)+(n-i-l)^2\right] \\ +\frac{\pi^2}{6}(n-i-l)(n+i-2k+l) \end{array}\right\}
\end{aligned}
$$

8. $E\left[Z_{i:n}(\ln Z_{i:n})Z_{j:n}(\ln Z_{j:n})\right]$ (4.19h) is given by

$$
\begin{aligned}
& c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}A^{11,11}_{i+l-k,n-i-l} \\
= \;& -c_{i,j:n}\sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}\frac{(-1)^{j-k-l}\binom{i-1}{k}\binom{j-i-1}{l}}{(i+l-k)^2(n-i-l)^2(n-k)^3}\times \\
& \left\{
\begin{array}{c}
-(n-k)^3\left[\ln(n-i-l)(\gamma-1+\ln(i+l-k))+Li_2\left(\frac{n-i-l}{n-k}\right)\right] \\
-\gamma(i+l-k)\left[-2(i+l-k)^2-7(i+l-k)(n-i-l)+(n-i-l)^2\right] \\
-(i+l-k)^2(3n-2i-k-2l)\left[1+\gamma^2+\ln^2(n-k)\right] \\
+\ln(n-k)\left[\begin{array}{c}-6(i+l-k)(n-i-l)^2-3(n-k)(n-i-l)^2 \\ +(n-k)^3+(n-k)^3\left(\frac{n-i-l}{n-k}+\ln(i+l-k)\right)\end{array}\right] \\
+\gamma\ln(n-k)\left[(n-2i+k-2l)((i+l-k)^2+4(i+l-k)(n-i-l)+(n-i-l)^2)\right] \\
+\frac{\pi^2}{6}(n+2i-3k+2l)(n-i-l)^2
\end{array}
\right.
\end{aligned}
$$

Appendix E : Mathematica Code: Computing Covariances of Final and Interim Weibull Score Functions

This appendix gives details of the Mathematica code used to compute the expectations H_1 to H_{12} (defined at Section 5.3.1.1) required in the covariances of final and interim Weibull score functions, given at (5.25) to (5.28). This requires the single and joint expectations of the forms at (4.12) and (4.19); here, we calculate these expectations using the direct method.

We first define some useful notations:

γ = `EulerGamma;`

`ci[n_,i_] :=` $\frac{\texttt{n!}}{\texttt{(n-i)! (i-1)!}}$

`cij[n_,i_,j_] :=` $\frac{\texttt{n!}}{\texttt{(i-1)! (j-i-1)! (n-j)!}}$

We then define the single expectations at (4.12):

`Elnzi[n_,i_] := ci[n,i] *` $\sum_{k=0}^{i-1} \frac{(-1)^{i-1-k}\ \mathrm{Binomial}[n,i]}{(n-k)} (-\gamma - \mathrm{Log}[n-k])$

`Ezilnzi[n_,i_] := ci[n,i] *` $\sum_{k=0}^{i-1} \frac{(-1)^{i-1-k}\ \mathrm{Binomial}[n,i]}{(n-k)^2} (1-\gamma - \mathrm{Log}[n-k])$

`Ez2ilnzi[n_,i_] := ci[n,i] *` $\sum_{k=0}^{i-1} \frac{(-1)^{i-1-k}\ \mathrm{Binomial}[n,i]}{(n-k)^3} (3-2\gamma - 2\mathrm{Log}[n-k])$

`Eln2zi[n_,i_] := ci[n,i] *` $\sum_{k=0}^{i-1} \frac{(-1)^{i-1-k}\ \mathrm{Binomial}[n,i]}{(n-k)} \left(\frac{\pi^2}{6} + (-\gamma - \mathrm{Log}[n-k])^2\right)$

`Eziln2zi[n_,i_] := ci[n,i] *` $\sum_{k=0}^{i-1} \frac{(-1)^{i-1-k}\ \mathrm{Binomial}[n,i]}{(n-k)^2}$
$\left(\frac{\pi^2}{6} - 1 + (1-\gamma - \mathrm{Log}[n-k])^2\right)$

`Ez2iln2zi[n_,i_] := ci[n,i] *` $\sum_{k=0}^{i-1} \frac{(-1)^{i-1-k}\ \mathrm{Binomial}[n,i]}{(n-k)^3}$
$\left(\frac{\pi^2}{6} - \frac{5}{4} + (\frac{3}{2} - \gamma - \mathrm{Log}[n-k])^2\right)$

For the joint expectations at (4.19), it will prove more convenient to first define the relevant functions $A_{s,t}^{pa,qb}$:

$$\texttt{A1st[s\_,t\_]} := -\frac{\mathtt{s}(\gamma\mathtt{s}+\mathtt{t})+(\mathtt{s}+\mathtt{t})^2\ \mathrm{Log}[\mathtt{t}]-\mathtt{t}(2\mathtt{s}+\mathtt{t})\ \mathrm{Log}[\mathtt{s}+\mathtt{t}]}{\mathtt{s}^2\mathtt{t}(\mathtt{s}+\mathtt{t})^2}$$

$$\texttt{A2st[s\_,t\_]} := -\frac{(\mathtt{s}+2\mathtt{t})(\gamma+\mathrm{Log}[\mathtt{s}+\mathtt{t}])-\mathtt{t}}{\mathtt{t}^2(\mathtt{s}+\mathtt{t})^2}$$

$$\texttt{A3st[s\_,t\_]} := -\frac{1}{\mathtt{st}(\mathtt{s}+\mathtt{t})}\left\{\begin{array}{c} -(\mathtt{s}+\mathtt{t})\left(\begin{array}{c}\gamma\ \mathrm{Log}[\mathtt{t}]+\mathrm{Log}[\mathtt{s}]\ \mathrm{Log}[\mathtt{t}]\\ -\mathrm{Log}[\mathtt{s}]\ \mathrm{Log}[\mathtt{s}+\mathtt{t}]+\mathrm{PolyLog}[2,\frac{\mathtt{t}}{\mathtt{s}+\mathtt{t}}]\end{array}\right)\\ -\mathtt{s}(\gamma^2+\mathrm{Log}[\mathtt{s}+\mathtt{t}]^2)\\ +(\mathtt{t}-\mathtt{s})\gamma\ \mathrm{Log}[\mathtt{s}+\mathtt{t}]+\frac{\pi^2}{6}\mathtt{t}\end{array}\right\}$$

$$\texttt{A4st[s\_,t\_]} := \frac{1}{\mathtt{s}^2\mathtt{t}^2(\mathtt{s}+\mathtt{t})^3}\left\{\begin{array}{c} -\mathtt{s}((\mathtt{s}^2+3\mathtt{st})(\gamma-1)+\mathtt{t}^2)-(\mathtt{s}+\mathtt{t})^3\ \mathrm{Log}[\mathtt{t}]\\ +\mathtt{t}^2(3\mathtt{s}+\mathtt{t})\ \mathrm{Log}[\mathtt{s}+\mathtt{t}]\end{array}\right\}$$

$$\texttt{A5st[s\_,t\_]} := \frac{1}{\mathtt{s}^2\mathtt{t}^2(\mathtt{s}+\mathtt{t})^3}\left\{\mathtt{s}^2(\mathtt{s}-\gamma\mathtt{s}+4\mathtt{t}-3\gamma\mathtt{t})+\mathrm{Log}[\mathtt{s}+\mathtt{t}](\mathtt{t}^2(3\mathtt{s}+\mathtt{t})-(\mathtt{s}+\mathtt{t})^3)\right\}$$

$$\texttt{A6st[s\_,t\_]} := -\frac{1}{\mathtt{st}^2(\mathtt{s}+\mathtt{t})^2}\left\{\begin{array}{c} -(\mathtt{s}+\mathtt{t})^2(\mathrm{Log}[\mathtt{t}](\gamma+\mathrm{Log}[\mathtt{s}])+\mathrm{PolyLog}[2,\frac{\mathtt{t}}{\mathtt{s}+\mathtt{t}}])\\ +\gamma\mathtt{s}(\mathtt{s}+3\mathtt{t})-\mathtt{s}(\mathtt{s}+2\mathtt{t})(\gamma^2+\mathrm{Log}[\mathtt{s}+\mathtt{t}]^2)\\ +\mathrm{Log}[\mathtt{s}+\mathtt{t}]\ ((\mathtt{s}+\mathtt{t})^2(1-\gamma+\frac{\mathtt{t}}{\mathtt{s}+\mathtt{t}}+\mathrm{Log}[\mathtt{s}])\\ +2\mathtt{t}^2(\gamma-1))+\frac{\pi^2}{6}\mathtt{t}^2\end{array}\right\}$$

$$\texttt{A7st[s\_,t\_]} := -\frac{1}{\mathtt{s}^2\mathtt{t}(\mathtt{s}+\mathtt{t})^2}\left\{\begin{array}{c} -(\mathtt{s}+\mathtt{t})^2(\mathrm{Log}[\mathtt{t}](\gamma-1+\mathrm{Log}[\mathtt{s}])+\mathrm{PolyLog}[2,\frac{\mathtt{t}}{\mathtt{s}+\mathtt{t}}])\\ -\gamma\mathtt{s}(\mathtt{t}-\mathtt{s})-\mathtt{s}^2(\gamma^2+\mathrm{Log}[\mathtt{s}+\mathtt{t}]^2)\\ +\mathrm{Log}[\mathtt{s}+\mathtt{t}]\ (\mathrm{Log}[\mathtt{s}](\mathtt{s}+\mathtt{t})^2-3\mathtt{st}-\mathtt{t}^2)\\ +\gamma\ \mathrm{Log}[\mathtt{s}+\mathtt{t}]\ (-\mathtt{s}^2+2\mathtt{st}+\mathtt{t}^2)+\frac{\pi^2}{6}\mathtt{t}(2\mathtt{s}+\mathtt{t})\end{array}\right\}$$

$$\texttt{A8st[s\_,t\_]} := -\frac{1}{\mathtt{s}^2\mathtt{t}^2(\mathtt{s}+\mathtt{t})^3}\left\{\begin{array}{c} -(\mathtt{s}+\mathtt{t})^3(\mathrm{Log}[\mathtt{t}](\gamma-1+\mathrm{Log}[\mathtt{s}])+\mathrm{PolyLog}[2,\frac{\mathtt{t}}{\mathtt{s}+\mathtt{t}}])\\ -\gamma\mathtt{s}(-2\mathtt{s}^2-7\mathtt{st}+\mathtt{t}^2)-\mathtt{s}^2(\mathtt{s}+3\mathtt{t})(1+\gamma^2+\mathrm{Log}[\mathtt{s}+\mathtt{t}]^2)\\ +\mathrm{Log}[\mathtt{s}+\mathtt{t}]\ \left(\begin{array}{c}-6\mathtt{st}^2-3\mathtt{t}^2(\mathtt{s}+\mathtt{t})+(\mathtt{s}+\mathtt{t})^3\\ +(\mathtt{s}+\mathtt{t})^3(\frac{\mathtt{t}}{\mathtt{s}+\mathtt{t}}+\mathrm{Log}[\mathtt{s}])\end{array}\right)\\ +\gamma\ \mathrm{Log}[\mathtt{s}+\mathtt{t}]\ (-(\mathtt{s}-\mathtt{t})(\mathtt{s}^2+4\mathtt{st}+\mathtt{t}^2))+\frac{\pi^2}{6}\mathtt{t}^2(3\mathtt{s}+\mathtt{t})\end{array}\right\}$$

Hence, we can now define the joint expectations at (4.19):

$$\texttt{Ezilnzj[n\_,i\_,j\_]} := \texttt{cij[n,i,j]} * \sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}(-1)^{i-1-k}\ \texttt{Binomial[i-1,k]}\ \texttt{Binomial[j-i-1,l]} * \texttt{A1st[i+l-k,n-i-l]}$$

$$\texttt{Elnzizj[n\_,i\_,j\_]} := \texttt{cij[n,i,j]} * \sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}(-1)^{i-1-k}\ \texttt{Binomial[i-1,k]}\ \texttt{Binomial[j-i-1,l]} * \texttt{A2st[i+l-k,n-i-l]}$$

$$\texttt{Elnzilnzj[n\_,i\_,j\_]} := \texttt{cij[n,i,j]} * \sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}(-1)^{i-1-k}\ \texttt{Binomial[i-1,k]}\ \texttt{Binomial[j-i-1,l]} * \texttt{A3st[i+l-k,n-i-l]}$$

$$\texttt{Ezizjlnzj[n\_,i\_,j\_]} := \texttt{cij[n,i,j]} * \sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}(-1)^{i-1-k}\ \texttt{Binomial[i-1,k]}\ \texttt{Binomial[j-i-1,l]} * \texttt{A4st[i+l-k,n-i-l]}$$

$$\texttt{Ezilnzizj[n\_,i\_,j\_]} := \texttt{cij[n,i,j]} * \sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1}(-1)^{i-1-k}\ \texttt{Binomial[i-1,k]}\ \texttt{Binomial[j-i-1,l]} * \texttt{A5st[i+l-k,n-i-l]}$$

$$\texttt{Elnzizjlnzj[n\_,i\_,j\_]} := \texttt{cij[n,i,j]} * \sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1} (-1)^{i-1-k}\ \texttt{Binomial[i-1,k]}\ \texttt{Binomial[j-i-1,l]} * \texttt{A6st[i+l-k,n-i-l]}$$

$$\texttt{Ezilnzilnzj[n\_,i\_,j\_]} := \texttt{cij[n,i,j]} * \sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1} (-1)^{i-1-k}\ \texttt{Binomial[i-1,k]}\ \texttt{Binomial[j-i-1,l]} * \texttt{A7st[i+l-k,n-i-l]}$$

$$\texttt{Ezilnzizjlnzj[n\_,i\_,j\_]} := \texttt{cij[n,i,j]} * \sum_{k=0}^{i-1}\sum_{l=0}^{j-i-1} (-1)^{i-1-k}\ \texttt{Binomial[i-1,k]}\ \texttt{Binomial[j-i-1,l]} * \texttt{A8st[i+l-k,n-i-l]}$$

We then define the expectations H_1 to H_{12}:

$$\texttt{H1[n\_]} := \texttt{n}$$

$$\texttt{H2[n\_]} := -\texttt{n}\gamma$$

$$\texttt{H3[n\_]} := \texttt{n}(1-\gamma)$$

$$\texttt{H4[n\_,r\_]} := \sum_{i=1}^{r}\texttt{Ezilnzi[n\_,i\_]}+\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\texttt{Ezilnzj[n,i,j]}+\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\texttt{Elnzizj[n,i,j]} +\sum_{i=1}^{r}\sum_{j=r+1}^{n}\texttt{Elnzizj[n,i,j]}$$

$$\texttt{H5[n\_,r\_]} := \sum_{i=1}^{r}\texttt{Eln2zi[n\_,i\_]}+2\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\texttt{Elnzilnzj[n,i,j]} +\sum_{i=1}^{r}\sum_{j=r+1}^{n}\texttt{Elnzilnzj[n,i,j]}$$

$$\texttt{H6[n\_,r\_]} := \sum_{i=1}^{r}\texttt{Eziln2zi[n\_,i\_]}+\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\texttt{Elnzizjlnzj[n,i,j]} +\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\texttt{Ezilnzilnzj[n,i,j]}+\sum_{i=1}^{r}\sum_{j=r+1}^{n}\texttt{Elnzizjlnzj[n,i,j]}$$

$$\texttt{H7[n\_,r\_]} := \texttt{r(n+1)}$$

$$\texttt{H8[n\_,r\_]} := \sum_{i=1}^{r}\texttt{Ezilnzi[n\_,i\_]}+\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\texttt{Ezilnzj[n,i,j]} +\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\texttt{Elnzizj[n,i,j]}+\sum_{i=1}^{r}\sum_{j=r+1}^{n}\texttt{Ezilnzj[n,i,j]} +\texttt{(n-r)}\left(\sum_{i=1}^{r-1}\texttt{Elnzizj[n,i,r]}+\texttt{Ezilnzi[n,r]}+\sum_{j=r+1}^{n}\texttt{Ezilnzj[n,r,j]}\right)$$

$$\texttt{H9[n\_,r\_]} := \sum_{i=1}^{r}\texttt{Ez2ilnzi[n\_,i\_]}+\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\texttt{Ezizjlnzj[n,i,j]} +\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\texttt{Ezilnzizj[n,i,j]}+\sum_{i=1}^{r}\sum_{j=r+1}^{n}\texttt{Ezizjlnzj[n,i,j]} +\texttt{(n-r)}\left(\sum_{i=1}^{r-1}\texttt{Ezilnzizj[n,i,r]}+\texttt{Ez2ilnzi[n,r]}+\sum_{j=r+1}^{n}\texttt{Ezizjlnzj[n,r,j]}\right)$$

$$\texttt{H10[n\_,r\_]} := \sum_{i=1}^{r}\texttt{Ez2ilnzi[n\_,i\_]}+\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\texttt{Ezizjlnzj[n,i,j]} +\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\texttt{Ezilnzizj[n,i,j]}+\sum_{i=1}^{r}\sum_{j=r+1}^{n}\texttt{Ezilnzizj[n,i,j]}$$

$$+(\text{n-r})\left(\sum_{i=1}^{r-1}\text{Ezizjlnzj[n,i,r]+Ez2ilnzi[n,r]+}\sum_{j=r+1}^{n}\text{Ezilnzizj[n,r,j]}\right)$$

$$\text{H11[n\_,r\_]} := \sum_{i=1}^{r}\text{Eziln2zi[n\_,i\_]}+\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\text{Ezilnzilnzj[n,i,j]}$$
$$+\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\text{Elnzizjlnzj[n,i,j]}+\sum_{i=1}^{r}\sum_{j=r+1}^{n}\text{Ezilnzilnzj[n,i,j]}$$
$$+(\text{n-r})\left(\begin{array}{c}\sum_{i=1}^{r-1}\text{Elnzizjlnzj[n,i,r]+Eziln2zi[n,r]}\\ +\sum_{j=r+1}^{n}\text{Ezilnzilnzj[n,r,j]}\end{array}\right)$$

$$\text{H12[n\_,r\_]} := \sum_{i=1}^{r}\text{Ez2iln2zi[n\_,i\_]}+2\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\text{Ezilnzizjlnzj[n,i,j]}$$
$$+\sum_{i=1}^{r}\sum_{j=r+1}^{n}\text{Ezilnzizjlnzj[n,i,j]}$$
$$+(\text{n-r})\left(\begin{array}{c}\sum_{i=1}^{r-1}\text{Ezilnzizjlnzj[n,i,r]+Ez2iln2zi[n,r]}\\ +\sum_{j=r+1}^{n}\text{Ezilnzizjlnzj[n,r,j]}\end{array}\right)$$

Finally, we are in the position to compute the covariances in (5.25) to (5.28):

$$\text{covdtdt[n\_,r\_,}\theta\text{\_,}\beta\text{\_]} := \beta^2\,\theta^{-2}\text{(H7[n,r]-rH1[n])}$$
$$\text{covdtdb[n\_,r\_,}\theta\text{\_,}\beta\text{\_]} := \theta^{-1}\text{(rH1[n]+H4[n,r]-H10[n,r])}$$
$$\text{covdbdt[n\_,r\_,}\theta\text{\_,}\beta\text{\_]} := \theta^{-1}\text{(H8[n,r]-rH2[n]-H9[n,r]+rH3[n])}$$
$$\text{covdbdb[n\_,r\_,}\theta\text{\_,}\beta\text{\_]} := \beta^{-2}\left(\begin{array}{c}\text{rH2[n]} + \text{H5[n,r]} - \text{H11[n,r]}\\ -\text{rH3[n]} - \text{H6[n,r]} + \text{H12[n,r]}\end{array}\right)$$

For example, we set $\theta = 100, \beta = 2, r = 15, n = 25$; we have

```
In[1]:= N[covdtdt[25,15,100,2], 10]
Out[1]:= 0.006000000000
In[2]:= N[covdtdb[25,15,100,2], 10]
Out[2]:= 0.04559706435
In[3]:= N[covdbdt[25,15,100,2], 10]
Out[3]:= 0.04559706435
In[4]:= N[covdbdb[25,15,100,2], 10]
Out[4]:= 5.092796735
```