INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

LIFE TESTING PROBLEMS WITH GAMMA TYPE INPUTS

THEODORO KOULIS

Department of Mathematics and Statistics, McGill University,

805 rue Sherbrooke Ouest, Montréal, Québec, Canada H3A 2K6

July 2000

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Master of Science.

© Theodoro Koulis 2000

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-70725-3

Canada

In loving dedication to my mother and father.

T. K.

Abstract

We examine the problem of finding the exact distributions of linear functions of k independent generalized gamma variables, $X_1, X_2, \ldots, X_k$. Special cases of generalized gamma distributions include the exponential, gamma and Weibull distributions. A linear function of such variables is often a quantity of interest in the analysis of survival data, reliability of certain systems and stochastic processes and hence we present this problem in the context of life testing. The exact distributions of these linear functions are needed to compute survival functions, hazard functions and other important functions in practical problems. Stacy (1962) obtained some exact results involving generalized gamma variables and Huzurbazar and Huzurbazar (1999) used saddlepoint approximations where the input variables are gamma or Weibull. We examine this problem where the k independent real scalar random variables, $X_1, X_2, \ldots, X_k$, are of gamma type with general parameters. For this case, various exact distributions are obtained and it is shown that most of these representations are easily computable. These exact results are compared with the usual saddlepoint approximations. We also examine numerically inverting the Laplace transform in this context, showing that it is one of the most efficient and accurate ways of estimating the exact distribution for certain cases. Results of this thesis are being published and presented in co-authorship with A.M. Mathai in Koulis and Mathai (2000).

Résumé

Nous étudions le problème de determination exacte de la loi de distribution d'une fonction linéaire de k variables aléatoires indépendantes, $X_1, X_2, \ldots, X_k$, de type gamma généralisé. Des cas particuliers de distributions de type gamma généralisé sont la distribution exponentielle, gamma et Weibull. Une fonction linéaire de ces variables est souvent une quantité importante dans l'analyse de processus stochastiques, la fiabilité d'un système ou dans l'analyse de la survie et donc nous présentons ce problème dans le contexte de l'analyse de données de faillite. La distribution exacte de ces fonctions linéaires est nécessaire pour calculer la fonction de survie, la fonction de risque et d'autres fonctions importantes. Stacy (1962) a obtenu quelques résultats éxacts et Huzurbazar and Huzurbazar (1999) ont utilisés des approximations de type "point de selle" (saddlepoint) pour des variables gamma ou Weibull. Nous étudions ce problème où les k variables aléatoires, $X_1, X_2, \ldots, X_k$, sont de type gamma avec des paramètres généraux. Dans ce cas, quelques distributions exactes sont obtenues et c'est montré que ces représentations sont facilement calculées. Les resultats exacts sont comparés avec l'approximation de type "point de selle". Nous étudions aussi l'inversion numerique de la transforme de Laplace, montrant que cette méthode est une façon efficace d'estimer la distribution exacte pour certains cas. Quelques résultats de cette thèse seront publiés et présentés en collaboration avec A. M. Mathai dans Koulis and Mathai (2000).

Acknowledgements

Many thanks go to Professor A. M. Mathai who taught me much of what I know in mathematical statistics and whose guidance, encouragement and inspiration provided the fundamental groundwork for my graduate studies. I owe thanks to Professor D. Wolfson for all his help and suggestions in survival analysis. I am indebted to Marcus Hum for his assistance in finding combinatorial algorithms necessary for computations found in this thesis, Alex Ghitza for his suggestions and Shabnam Beheshti whose love and inspiration made this thesis possible. I am also grateful to everyone in the Department of Mathematics and Statistics for providing a stimulating work environment.

Contents

Chapter 1

Preliminaries

1.1 Introduction

Often in industry, a manufacturer wishes to have some information about the items it produces. One of the prevalent themes in the computer industry is determining whether or not a producer's computer chip is reliable. In situations where such delicate equipment fails to work once in a while, a measure of frequency of failure is often needed. Another situation may involve the study of an assembly line with many stages where one may wish to know how long it takes to produce an item. In medicine, researchers wish to answer questions relating to the survival of patients afflicted with potentially deadly diseases. One way to obtain this information is by constructing life testing experiments designed to measure the 'lifetime' of a product or patient subjected to tests.

The analysis of life testing has its origins in the study of population life tables. Recently, methods have been developed to study diseases such as AIDS and cancer and to examine various problems in engineering and industry. In general, one wishes to study subjects or

groups of subjects each of whom have a predefined 'failure' time. For example, light bulbs may be examined, each of which have a failure time defined to be the waiting time until burn out or the subjects can be patients in a clinical trial where the failure time is defined as the time in months from the beginning of the trial until death due to heart failure.

The difficulty in studying this kind of data arises when considering censored data. Often in life testing experiments, subjects are still 'alive' at the time the experiment terminates. Because of the context, this censored data is often called survival data, failure time data or lifetime data.

Two fields concerned with the analysis of life testing are survival analysis and reliability. Survival analysis is mainly concerned with the probability of failure-free operation of an item in an interval $[t_0, t_0 + t)$ given that the item is still 'alive' at time t_0. Reliability is concerned with a quantitative measurement of the reliability of an item, usually in terms of average lifetime. Survival analysis is often used in medicine and biology and reliability usually deals with problems in engineering and industry. Even though the two fields deal with different real world problems, they can both be studied through life testing.

1.2 Failure Time

Before we begin discussing the analysis of failure time data, it is essential that we define precisely what is meant by failure time of a subject. In order to do so, we need to define a clear time of origin, a scale for measuring the passage of time and what is meant by a failure event. The waiting time between the time of origin and the failure event is often referred to as the failure time or survival time.

1.2.1 Time of Origin

The time of origin is defined according to the life testing problem at hand. If one is studying the engine failure of automobiles, then a natural measure of reliability or failure is mileage of the vehicle. The time of origin here is at 0 km. If we consider prostate cancer patients in a clinical trial then the time of origin can be chosen to be at the time of diagnosis of the disease. In this case, it would be appropriate to take the exact time of onset of the disease as the time of origin, but this is not always possible. Note that the time of origin may not be necessarily at the same calendar date for each subject. Two patients may be diagnosed at different times in a year, however their time of origin is defined in the same way.

When considering the time of origin, it is essential to assign a measuring scale for the passage of time. For the survival of patients in a clinical trial, time is measured in months or years. In industrial problems, other types of information can be used to indicate the passage of time, such as mileage for engines, amount of load for springs or distance traveled for particles.

1.2.2 Failure Event

Measurement of data for a subject in life testing ends with a failure event. In medical fields, a failure event is often defined as death or as death due to a specific cause. A failure event may or may not be arbitrary, depending on the problem. For instance, when testing the reliability of a washing machine, a choice for the failure event may be the instant at which the performance of the machine falls bellow a predefined level, whereas when testing the reliability of a lightbulb the failure event is defined as the instant at which the light bulb

burns out.

1.3 Life Testing and Related Functions

In life testing, the failure time of a subject can be represented by a non-negative random variable, X say, with a probability density function (pdf) $f(x)$, a cumulative probability density function (cdf) $F(x)$ and a moment generating function (mgf) $M_X(t)$, if it exists. The distribution of X is referred to as the life testing model and will be the main concern in our discussion. The probability that a subject survives beyond a time x is called the survival function in survival analysis and the reliability at time x in reliability. It is denoted by

$$S(x) = \mathrm{P}(X \geq x) = 1 - F(x) \,. \tag{1.1.1}$$

Throughout the text, we will refer to (1.1.1) as the survival function. Note that in life testing, it is usually the convention to define the survival function as in (1.1.1) rather than $S(x) = \mathrm{P}(X > x)$. For continuous random variables either definition will work, but for the discrete case, results vary slightly.

The probability that a subject is failure-free in $[x, x+h)$ given that it has survived until x is given by

$$\frac{\mathrm{P}(x \leq X < x+h)}{\mathrm{P}(X \geq x)} = \frac{F(x+h) - F(x)}{1 - F(x)} \,. \tag{1.1.2}$$

Assume that X is a continuous random variable. The instantaneous failure rate at time x conditional upon survival to time x is called the hazard function and is given by

$$\begin{aligned} h(x) &= \lim_{h \to 0} \frac{F(x+h) - F(x)}{h\,[1 - F(x)]} \\ &= \frac{f(x)}{1 - F(x)} \\ &= \frac{f(x)}{S(x)} \,. \end{aligned} \tag{1.1.3}$$

The hazard function $h(x)$ is often called the force of mortality in actuarial science. In most areas of application it is called the failure rate. The hazard function is said to be increasing if for any $x_1 < x_2$ we have $h(x_1) \leq h(x_2)$. Analogously, the hazard function is said to be decreasing if $h(x_1) \geq h(x_2)$. For continuous distributions, we have that

$$F(x) = \int_0^x f(t)\mathrm{d}t \ , \quad f(x) = \frac{\mathrm{d}}{\mathrm{d}x}F(x) \quad \text{and} \quad f(x) = -\frac{\mathrm{d}}{\mathrm{d}x}S(x)$$

and so (1.1.3) can be rewritten as

$$h(x) \ = \ \frac{-\frac{\mathrm{d}}{\mathrm{d}x}S(x)}{S(x)} \ = \ -\frac{\mathrm{d}}{\mathrm{d}x}\log\left[S(x)\right] \ . \tag{1.1.4}$$

Since $S(0) = 1$, integration of the above equality yields,

$$S(x) = \ \exp\left\{-\int_0^x h(t)\mathrm{d}t + \log\left[S(0)\right]\right\} = \ \exp\left\{-\int_0^x h(t)\mathrm{d}t\right\} \ . \tag{1.1.5}$$

The integral found in (1.1.5)

$$H(x) = \int_0^x h(t)\mathrm{d}t \tag{1.1.6}$$

is called the integrated hazard. Upon differentiation of $S(x)$ in (1.1.5), we have

$$f(x) = \ h(x)\exp\left\{-H(x)\right\} \ . \tag{1.1.7}$$

The mean residual life time of a subject with failure time X is defined as

$$\begin{aligned} m(t) \ &= \ E\left(X - t | X \geq t\right) \\ &= \ \int_t^\infty (x-t)\frac{f(x)}{S(t)}\mathrm{d}x \ . \end{aligned} \tag{1.1.8}$$

Given that $m(0) = E\left(X\right)$, the mean residual life time can be looked at as a generalization of the mean failure time. Note that, (1.1.8) can be rewritten as

$$m(t) = \ E\left(X | X \geq t\right) - t \ , \tag{1.1.9}$$

so that a possible interpretation of $m(t)$ is that it is the average amount of unused "life time" of a subject at time t. Using integration by parts with $u = x - t$ and $\mathrm{d}v = f(x)\mathrm{d}x$, we obtain

$$m(t) = \int_t^\infty \frac{S(x)}{S(t)}\mathrm{d}x \ . \tag{1.1.10}$$

The form of the mean residual life time in (1.1.10) gives another representation for the mean failure time by substituting $t = 0$. The representation is

$$E(X) = m(0) = \int_0^\infty S(x)\mathrm{d}x \ . \tag{1.1.11}$$

If X is discrete, then some modifications of the above definitions are necessary. If X takes on values $x_1 < x_2 < x_3 < \cdots$ then the pdf of X is defined as

$$f(x_i) = \mathrm{P}\,(X = x_i) \quad \text{for } i = 1, 2, 3, \cdots$$

The cdf and survival function of X are

$$F(x) = \sum_{j:x_j<x} f(x_j) \quad \text{and} \quad S(x) = 1 - F(x) = \sum_{j:x_j\geq x} f(x_j) \ ,$$

respectively. For discrete distributions, the hazard at x_i is defined to be the conditional probability of failure at x_i and we write

$$h(x_i) = \frac{f(x_i)}{1 - F(x_i)} \ . \tag{1.1.12}$$

Note that

$$\begin{aligned} S(x_i) &= 1 - F(x_i) = f(x_i) + f(x_{i+1}) + \cdots \qquad \text{and} \\ 1 - h(x_i) &= 1 - \frac{f(x_i)}{f(x_i) + f(x_{i+1}) + \cdots} = \frac{S(x_{i+1})}{S(x_i)}. \end{aligned}$$

Given that $S(x_1) = 1$, we have by recursion

$$S(x) = \prod_{x_i<x} (1 - h(x_i)) \quad \text{and} \tag{1.1.13}$$

$$f(x_i) = h(x_i) \prod_{j=1}^{i-1} (1 - h(x_j)) \ . \tag{1.1.14}$$

To be consistent with (1.1.5), the convention is to define the integrated hazard as

$$H(x) = \sum_{j:x_j<x} \log\left(1 - h(x_j)\right) \tag{1.1.15}$$

so that

$$S(x) = \exp\left[-H(x)\right] . \tag{1.1.16}$$

Example 1. Assume that a system has the following reliability (survival) function:

$$S(x) = \begin{cases} \exp\left(-2x - 5x^2\right) & \text{for } x \geq 0 \\ 1 & \text{elsewhere.} \end{cases} \tag{1.1.17}$$

Differentiation gives the pdf as

$$f(x) = \begin{cases} (2 + 10x)\exp\left(-2x - 5x^2\right) & \text{for } x \geq 0 \\ 0 & \text{elsewhere} \end{cases} \tag{1.1.18}$$

and the hazard function is

$$h(x) = \begin{cases} (2 + 10x) & \text{for } x \geq 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{1.1.19}$$

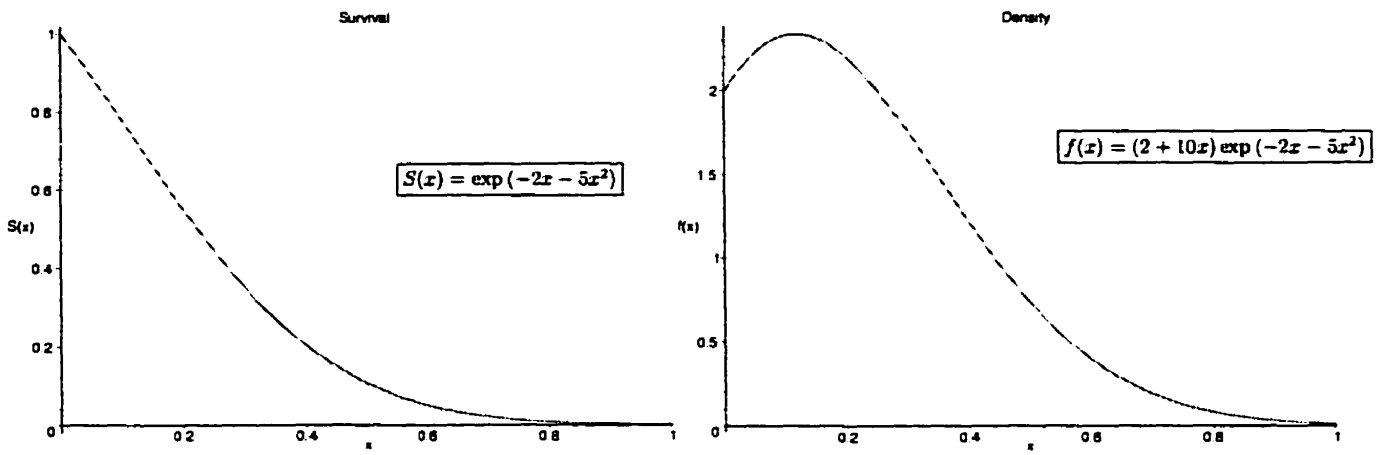

Figure 1.1: Survival and Density Functions: Example 1

1.4 Censoring

In a life testing experiment, it may not always be possible to obtain a complete sample of data, since this may take a long time. In industrial reliability, an experiment may be terminated after a certain time to minimize costs that increase with time. Also, subjects may fail due to reasons other than the failure event. For example, a patient in a clinical trial may die in a car accident. This means that some subjects may have incomplete information and such data are called time-censored. Generally speaking, censoring occurs when exact lifetimes are known for only a portion of the subjects in a life testing experiment. Because censoring gives partial information about the life of the subject in an experiment, it must be recorded and included in any analysis. Censoring, however, represents a problem in the analysis of life time data.

A subject is right censored at a time c if it is known only that its failure time, X, is greater or equal to c. Similarly, a subject is left censored at time c if its failure time, X, is less than or equal to c. Both right and left censoring are special cases of interval censoring where the failure time is observed to lie in an interval. Left and interval censoring are not as common as right censoring, and so we will restrict our discussion of life testing to right censoring.

The censoring time of a subject in a life testing experiment is defined to be the waiting time from the time of origin until the subject is censored. Like failure time, censoring time can be represented by a positive random variable, C, say. Even if a subject fails before the experiment is terminated, it has an inherent unobserved censoring time. In a similar fashion, if a subject is censored, then it is assumed that it has an unobserved failure time, represented

by a positive random variable X.

If a life testing experiment contains n subjects, then the data observed consist of waiting times, $y_1, y_2, y_3, \cdots, y_n$, where $y_i = \min(x_i, c_i)$, $i = 1, \cdots, n$, along with an indicator function, δ_i, where

$$\delta_i = \begin{cases} 1 & \text{if } x_i \leq c_i \\ 0, & \text{if } x_i > c_i \end{cases}$$

and x_i and c_i are the actual failure and censoring times of subject i. If for some i, $\delta_i = 1$ then the failure time of subject i was observed and $y_i = x_i$. If $\delta_i = 0$ then the subject i was censored and $y_i = c_i$.

Before we include censoring in our discussion, it is important to describe some specific censoring mechanisms.

1.4.1 Type I Censoring

A type I censored sample is one in which the censoring time for all subjects is predetermined and identical. In other words, we have $c_i = c \; \forall \; i$, $i = 1, 2, \cdots, n$. In real life, this type of censoring may be considered to save time and money since a full sample may take too long to obtain. Data obtained from such a life testing experiment is often called time-censored data. Figure 1.2 shows an example of type I censoring with $c = 6$.

If n subjects start operation at different time points in an interval $[0, c]$, and the life testing experiment is terminated at time c, then data obtained is called multiply censored data. An example of multiple censoring is displayed in figure 1.3 with $c = 6$.

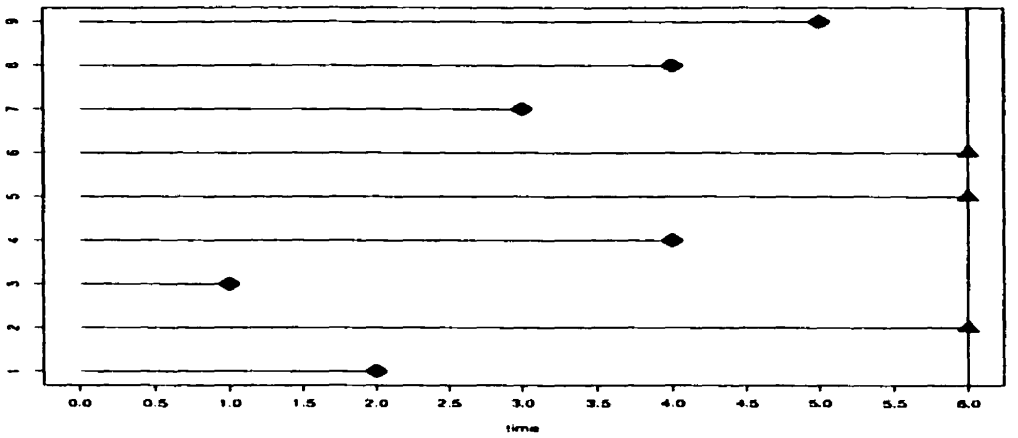

Figure 1.2: Type I Censoring : $c = 6$; Triangle: Censor, Diamond: Failure

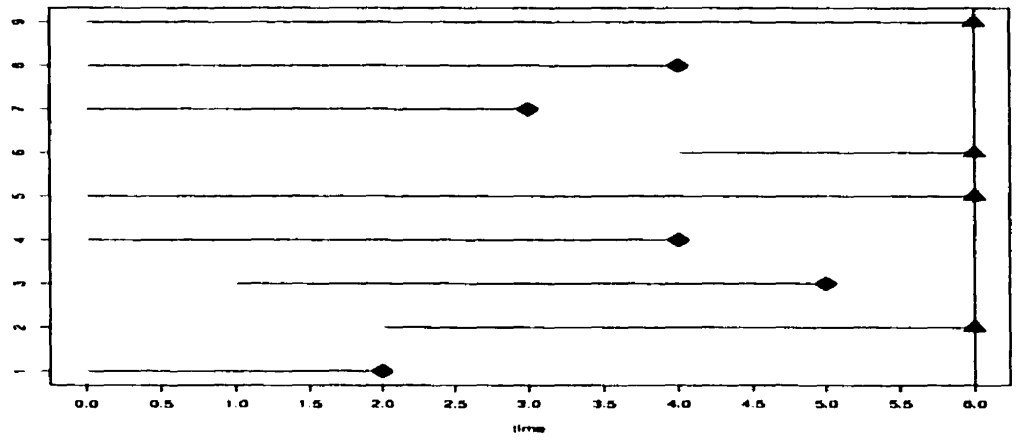

Figure 1.3: Multiple Censoring : $c = 6$; Triangle: Censor, Diamond: Failure

1.4.2 Type II Censoring

A type II censored sample is one in which only the r smallest failure times are observed in a life testing experiment of n subjects, with $1 \leq r < n$. This means that the life testing experiment terminates at the r^{th} failure. the data consists of the ordered sample $y_{(1)} < y_{(2)} < \cdots < y_{(r)}$ and $y_{(r)} = y_{(r+1)} = \cdots = y_{(n)}$. Data obtained from type II censoring is also referred to as failure-censored data. This type of censoring is often considered when testing high cost items such as microchips. Figure 1.4 shows an example of type II censoring with $r = 6$.

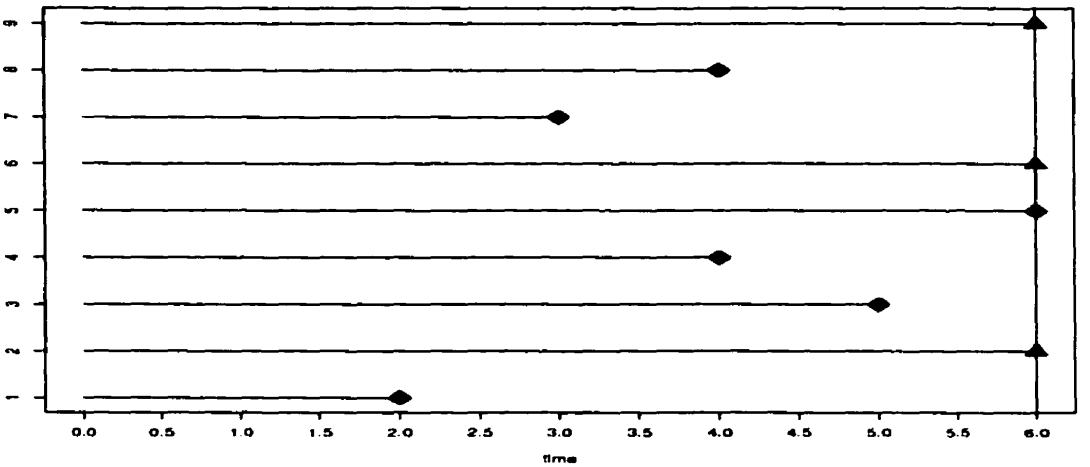

Figure 1.4: Type II Censoring : $r = 6$; Triangle: Censor, Diamond: Failure

Lawless (1982) also considers a more general type II censoring mechanism called progressive type II censoring. This involves the removal of n_1 subjects of a sample of size n, $n_1 \leq n$, after r_1 of them have failed, so that $n - n_1 - r_1$ items are now remaining. The process is then repeated by removing n_2 items after r_2 of the remaining $n - n_1 - r_1$ have failed, and so on.

Most of the analysis concerning type II censoring can be done using order statistics as will be seen in chapter 2.

1.4.3 Random and Independent Censoring

In real life situations, censoring times are often random variables. A random censoring process is one in which subjects have a failure time X, and a censoring time C, with X and C independent variables with survival functions $S(x)$ and $G(x)$ respectively. Type I censoring is a special case of random censoring. There is a weaker form of independence between the survival times and the censoring times, which, if satisfied by a life testing model, is sufficient for the establishment of most of the asymptotic results discussed in chapter 2. This weaker

form is called independent censoring (not to be confused with the independence of X and C in random censoring). Let $\mathbb{H}(x)$ be the history of a subject up to time x. The history may contain failure times and censoring times for all the subjects in a life testing model and even covariate information in the case of a regression model. We say we have independent censoring if

$$P(x \leq X \leq x + \mathrm{d}x | \mathbb{H}(x),\ X \geq x) = P(x \leq X \leq x + \mathrm{d}x | X \geq x) \ .$$

This means that independent censoring allows for censoring to depend on mechanisms external to the failure process. Note that random censoring is a special case of independent censoring. Note also, that in type II censoring, we do not have random censoring but we do have independent censoring.

1.5 Bibliographic Notes

There are numerous books available presenting life testing through various levels of mathematical and statistical sophistication. Books close to the spirit of our subject include Cox and Oakes (1984), Crowder, Kimber, Smith, and Sweeting (1991), Miller (1981), Sinha (1986) and Zacks (1992). Kalbfleisch and Prentice (1980) and Fleming and Harrington (1991) are more advanced texts, but provide more insight into the field. Also, there are many journals such as IEEE Transactions on Reliability which deals with issues in reliability in engineering. Useful sources for searching for articles in journals are the on-line databases MathSciNet (http://www.ams.org/mathscinet/) and Zentralblatt MATH Database (http://www.springer-ny.com/ZMATH/).

Chapter 2

Distributions and Techniques

2.1 Introduction

Now that we've discussed the nature of failure time data, we proceed to the discussion of its modeling and analysis. With any statistical inference based on the life testing experiment and the data acquired, one usually assumes that the data drawn is from a population with an underlying distribution function, $F(x)$, say. Before starting a life testing experiment, it is often necessary to make some assumptions about the underlying pdf, $f(x)$, or cdf, $F(x)$, of the population. These assumptions lead to different models for life data and such characterizations are discussed in detail in Kagan, Linnik, and Rao (1973), Galambos and Kotz (1978) and Mathai and Pederzoli (1977).

In this chapter we introduce the reader to elementary and complex statistical models and parameter estimation with and without censoring, as well as non-parametric results such as the product-limit estimator for the survival function. The likelihood function for different censoring mechanisms is also discussed.

2.2 Elementary Models

Given that failure time is defined to be a positive random variable, it is natural to begin with a certain class of distributions. It is evident that the most common distribution in applied statistics, the normal distribution, will not be particularly effective for explaining life data. From a stochastic point of view, it makes sense to describe life data with the exponential distribution in mind. Even though the exponential model is quite limited for life testing experiments, it leads to generalizations such as the gamma, Weibull and generalized gamma distributions which perform very well in applied settings.

In the next subsection, we will discuss briefly properties of these elementary models. We begin with the simplest life testing model.

2.2.1 Exponential Model

If we begin with the assumption that a population has a constant hazard rate

$$h(x) = \begin{cases} \frac{1}{\lambda} & \text{if } x \geq 0, \lambda > 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{2.2.1}$$

then from (1.1.5) we have that the survival function is

$$S(x) = \begin{cases} \exp\left(-\frac{1}{\lambda}\int_0^x dt\right) = \exp(-x/\lambda) & \text{for } x \geq 0 \\ 1 & \text{elsewhere.} \end{cases} \tag{2.2.2}$$

Application of (1.1.7) gives the density of X as

$$f(x) = \begin{cases} \frac{1}{\lambda}\exp(-x/\lambda) & \text{for } x \geq 0, \lambda > 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{2.2.3}$$

The pdf in (2.2.3) is the exponential density function with parameter λ. The moment generating function (mgf) of the exponential distribution is

$$M_X(t) = \frac{1}{1-\lambda t} \quad \text{for } t < 1/\lambda. \tag{2.2.4}$$

The exponential model is a natural distribution in life testing, and is a special case of the Weibull and gamma distributions as will be seen shortly. It is also the simplest model in life testing and, hence, one of the most exploited distributions. The constant hazard assumption implies that the exponential model is well suited for populations where there is no wearing or aging. This is referred to as the "no memory property" and can be expressed as

$$\mathrm{P}\,(X \geq x + t | X \geq x) = \mathrm{P}(X \geq t) = \exp\,(-t/\lambda)\ . \tag{2.2.5}$$

This means that under the exponential model, future survival of a subject does not depend on its past. Obviously, this model is poor for explaining life data, since in reality most objects age in some sense.

Example 2. Putting λ to be equal to 1, 2 and 0.3 gives the following survival functions:

$$S_1(x) = \exp\,(-x)\ , \quad S_2(x) = \exp\,(-x/2) \quad \text{and} \quad S_3(x) = \exp\,(-x/0.3)\ .$$

These are shown in figure 2.1. Their corresponding pdf and hazard functions are:

$$f_1(x) = \exp\,(-x)\ , \quad f_2(x) = \frac{1}{2}\exp\,(-x/2)\ , \quad f_3(x) = \frac{1}{0.3}\exp\,(-x/0.3)\ ,$$

$$h_1(x) = 1\ , \quad h_2(x) = \frac{1}{2} \quad \text{and} \quad h_3(x) = \frac{1}{0.3}\ .$$

Note that for the exponential model, it is easy to see the relationships between the hazard function, pdf and survival function (see chapter 1).

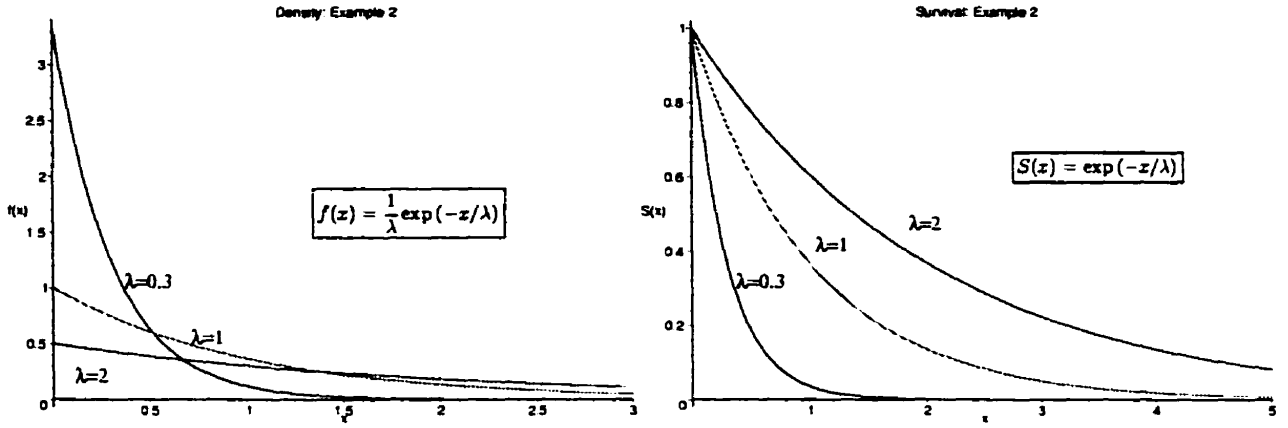

Figure 2.1: Exponential Density and Survival Functions

2.2.2 Gamma Model

A generalization of the exponential model discussed in section 2.2.1 is the gamma model with two parameters. The density of the gamma distribution is

$$f(x) = \begin{cases} \dfrac{1}{\lambda^{\alpha}\Gamma(\alpha)} x^{\alpha-1}\exp(x/\lambda) & \text{for } x > 0,\ \alpha > 0,\ \lambda > 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{2.2.6}$$

The parameters α and λ represent the shape parameter and scale parameter, respectively. An integral representation of the gamma function, $\Gamma(x)$, is given by

$$\Gamma(x) = \int_0^{\infty} t^{x-1}\exp(-t)\mathrm{d}t, \quad x > 0\ . \tag{2.2.7}$$

For example, $\Gamma(5) = 4!$ and $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. Properties of the gamma function are discussed in detail in Mathai (1993).

Putting $\alpha = 1$ in (2.2.6) shows that the density of the exponential model given in (2.2.3) is a special case of the gamma model. The cdf and survival function of X in this model do not have closed forms and are given in terms of incomplete gamma integrals. They are,

respectively

$$F(x) = \begin{cases} \gamma(x, \alpha, \lambda) & \text{for } x > 0,\ \lambda > 0,\ \alpha > 0 \\ 0 & \text{elsewhere,} \end{cases} \tag{2.2.8}$$

and

$$S(x) = \begin{cases} 1 - \gamma(x, \alpha, \lambda) & \text{for } x > 0,\ \lambda > 0,\ \alpha > 0 \\ 1 & \text{elsewhere.} \end{cases} \tag{2.2.9}$$

The incomplete gamma function $\gamma(x, \alpha, \lambda)$ is

$$\gamma(x, \alpha, \lambda) = \int_0^x \frac{1}{\lambda^\alpha \Gamma(\alpha)} t^{\alpha-1} \exp(-t/\lambda) dt, \quad \lambda > 0,\ \alpha > 0\ . \tag{2.2.10}$$

More on the incomplete gamma function can be found in Mathai (1993). The hazard function of X is given by

$$h(x) = \begin{cases} \dfrac{\frac{1}{\lambda^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp(-x/\lambda)}{1 - \gamma(x, \alpha, \lambda)} & \text{for } x > 0,\ \lambda > 0,\ \alpha > 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{2.2.11}$$

Unlike the cdf, survival function and hazard function for this model, the mgf of X has a useful closed form:

$$M_X(t) = \frac{1}{(1 - \lambda t)^\alpha} \quad \text{for } t < 1/\lambda. \tag{2.2.12}$$

The choice of parameters can greatly affect the shape of the hazard function. If one puts $\alpha = 1$ and $\lambda = 1$, then we are considering an exponential model and the hazard is constant. For $\alpha = 2.3$ and $\lambda = 0.3$ the hazard is monotone increasing and for $\alpha = 2$ and $\lambda = 1.5$, it is monotone decreasing. This can be seen in figure 2.2. These choices for the parameters produce the following density functions:

$$f_1(x) = \exp(-x)\ , \quad f_2(x) = \frac{1}{.07317159387} x^{1.3} \exp(-x/0.3)\ ,$$

$$f_3(x) = \frac{1}{2.25} x \exp(-x/1.5)\ ,$$

Example 3.

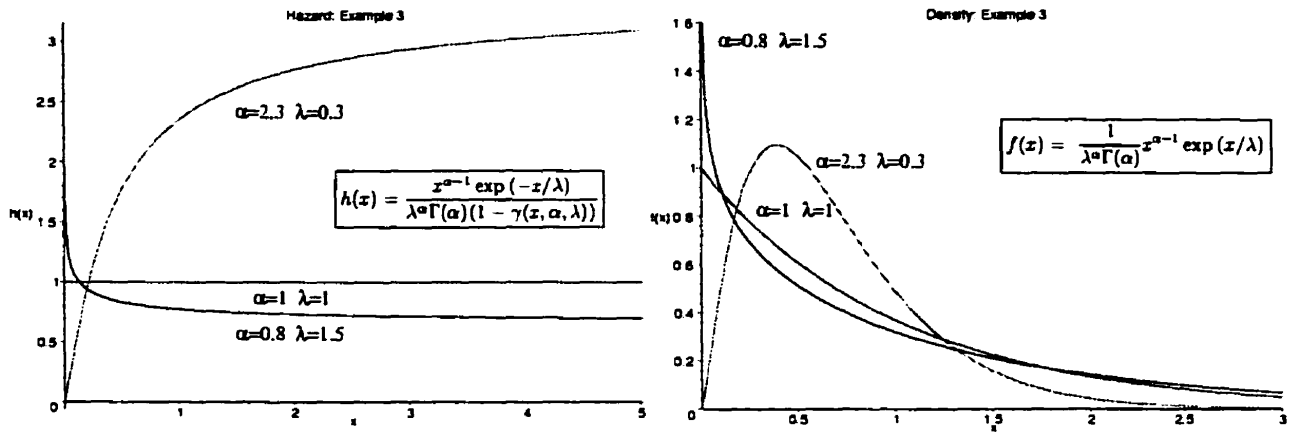

Figure 2.2: Gamma Hazard and Density Functions

2.2.3 Weibull Model

Another generalization of the exponential model is the Weibull model which, like the gamma model, has two parameters. The pdf of X in this model is

$$f(x) = \begin{cases} \frac{\beta}{\lambda} x^{\beta-1} \exp\left(-\frac{x^\beta}{\lambda}\right) & \text{for } x > 0, \beta > 0, \lambda > 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{2.2.13}$$

Note that by substituting $\beta = 1$ in (2.2.13), we obtain the exponential density in (2.2.3). The Weibull distribution is one of the most common and well suited distributions to model failure time data. Unlike the gamma distribution, the survival function, cdf and hazard function have closed forms in this case. They are

$$S(x) = \begin{cases} \exp\left(-\frac{x^\beta}{\lambda}\right) & \text{for } x > 0, \lambda > 0, \beta > 0 \\ 1 & \text{elsewhere,} \end{cases} \tag{2.2.14}$$

and

$$h(x) = \begin{cases} \frac{\beta}{\lambda} x^{\beta-1} & \text{for } x > 0, \lambda > 0, \beta > 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{2.2.15}$$

Unfortunately, the mgf of the Weibull distribution does not have a closed form. The moments of X are

$$\mu'_r = \lambda^{r/\beta}\Gamma\left(\frac{r}{\beta}+1\right) \quad \text{for } r = 1, 2, 3, \cdots \tag{2.2.16}$$

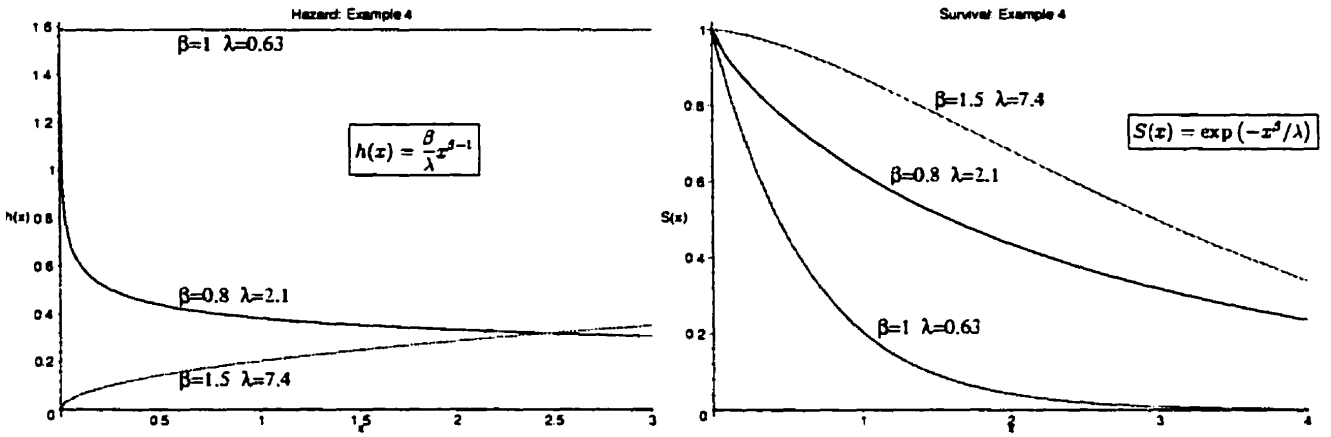

Figure 2.3: Weibull Hazard and Survival Functions

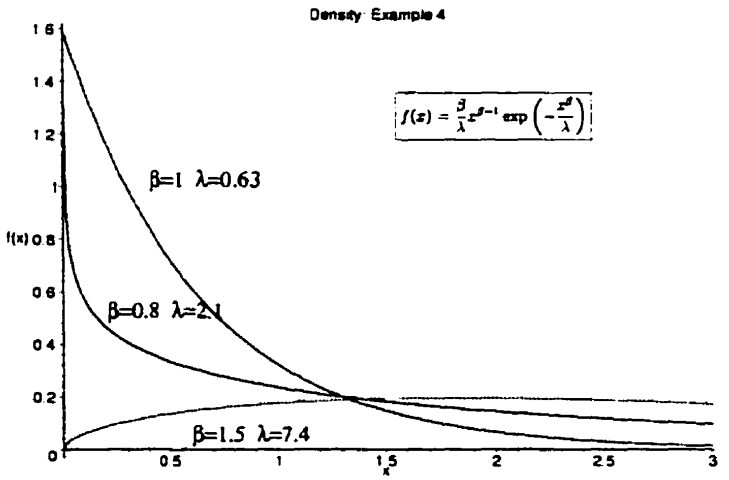

Figure 2.4: Weibull Density Functions

Example 4. Just as in the exponential model, it is easy to see the relationships between the pdf, hazard and survival functions in the Weibull model. However, like the gamma model, it depends on two parameters which determine the shape of the hazard function. The models

shown in figures 2.3 and 2.4 are those corresponding to the following pdfs:

$$f_1(x) = \frac{1.5}{7.4}x^{0.5}\exp\left(-\frac{x^{1.5}}{7.4}\right), \quad f_2(x) = \frac{1}{0.63}\exp\left(-\frac{x}{0.63}\right) \text{ and}$$
$$f_1(x) = \frac{0.8}{2.1}x^{-0.2}\exp\left(-\frac{x^{0.8}}{2.1}\right).$$

2.2.4 Generalized Gamma Model

We have seen that the exponential model is a special case for both the gamma and Weibull models. If we consider an even greater generalization of these models, then we access a wider, richer family with which to describe life testing data. This generalization is called the generalized gamma model and it has a density function defined as:

$$f(x) = \begin{cases} \dfrac{\beta\,\delta^{\alpha/\beta}}{\Gamma\left(\dfrac{\alpha}{\beta}\right)}x^{\alpha-1}\exp\left(-\delta x^{\beta}\right) & \text{for } x>0,\ \alpha>0, \beta>0,\ \delta>0 \\ 0 & \text{elsewhere.} \end{cases} \tag{2.2.17}$$

Putting $\beta = 1$ and $\alpha = 1$ gives the exponential distribution in (2.2.3) with parameter $\delta = \frac{1}{\lambda}$. With $\alpha = \beta$, we obtain the Weibull density in (2.2.13) with parameters $\delta = \frac{1}{\lambda}$ and α. For $\beta = 1$, we obtain the gamma density in (2.2.6) with parameters α and $\delta = \frac{1}{\lambda}$. In (2.2.17), the quantities α and β are referred to as the shape parameters and δ is called the scale parameter. Since the exponential, gamma and Weibull distributions are used often in life testing, it makes sense to analyze life data through generalized gamma distributions.

The survival function and hazard function for the generalized gamma model can be given in terms of incomplete gamma integrals. These are:

$$S(x) = \begin{cases} 1-\gamma(x^{\beta}, \alpha/\beta, 1/\delta) & \text{for } x>0,\ \delta>0,\ \alpha/\beta>0 \\ 1 & \text{elsewhere.} \end{cases} \tag{2.2.18}$$

and

$$h(x) = \begin{cases} \dfrac{\beta\, \delta^{\alpha/\beta} x^{\alpha-1} \exp\left(-\delta x^{\beta}\right)}{\Gamma\left(\dfrac{\alpha}{\beta}\right)\left(1 - \gamma(x^{\beta}, \alpha/\beta, 1/\delta)\right)} & \text{for } x > 0,\ \delta > 0,\ \alpha/\beta > 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{2.2.19}$$

The moment generating function of the generalized gamma, given in Stacy (1962), is equal to

$$M_X(t) = \left[\sum_{i=0}^{\infty} \frac{\left(\dfrac{t}{\delta^{1/\beta}}\right)^i}{i!} \frac{\Gamma\left(\dfrac{\alpha+i}{\beta}\right)}{\Gamma\left(\dfrac{\alpha}{\beta}\right)} \right] \tag{2.2.20}$$

defined for $t < c$, $0 \leq c$. If $\beta = 1$ then $t < \delta$. However, if $\beta < 1$, then $t \leq c = 0$ and if $\beta > 1$ then $c = \infty$. Thus, setting $\alpha = \beta$ and $\delta = \dfrac{1}{\lambda}$ produces the moment generating function for the Weibull distribution.

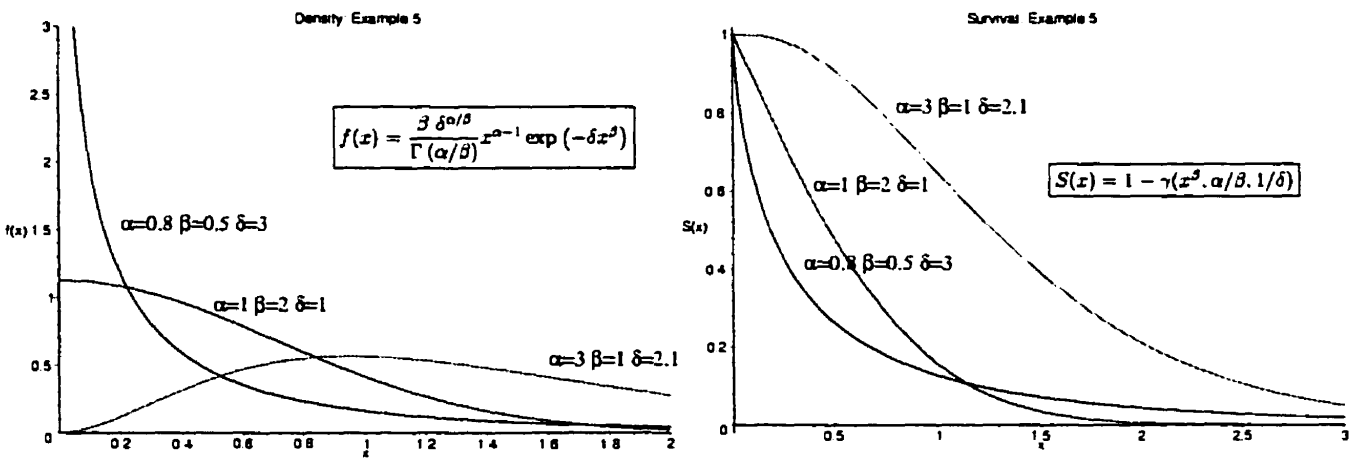

Figure 2.5: Generalized Gamma Density and Survival Functions

Example 5. If we put $\alpha = 3$, $\beta = 1$ and $\delta = 2.1$ the density is

$$f_1(x) = \frac{2.1^3}{2} x^2 \exp\left(-2.1x\right) . \tag{2.2.21}$$

For $\alpha = 1$, $\beta = 2$ and $\delta = 1$ we get

$$f_2(x) = \frac{2}{\sqrt{\pi}} \exp\left(-x^2\right) , \tag{2.2.22}$$

and for $\alpha = 0.8$, $\beta = 0.5$ and $\delta = 3$ the density is

$$f_3(x) = \frac{0.5\ 3^{1.6}}{\Gamma(1.6)} x^{-0.2} \exp\left(-3x^{0.5}\right) . \tag{2.2.23}$$

These examples are shown in figure 2.5 above.

2.2.5 Extreme Value Model

Another family of distributions worth mentioning is one related to the Weibull family. The model is called the extreme value distribution and is used to study the properties of the log-failure time, $Y = \log(X)$. The pdf of Y is

$$\widehat{f}(y) = \frac{1}{\sigma} \exp\left\{\frac{y-\mu}{\sigma} - \exp\left(\frac{y-\mu}{\sigma}\right)\right\} \tag{2.2.24}$$

for $-\infty < y < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$. The survival function of Y is

$$\widehat{S}(y) = \exp\left\{-\exp\left(\frac{y-\mu}{\sigma}\right)\right\} \tag{2.2.25}$$

for $-\infty < y < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$. Note that the extreme value distribution has support on $(-\infty, \infty)$. In life testing, the extreme value distribution often arises as the distribution of $Y = \log(X)$; this is equivalent to the assumption that X has a Weibull distribution. The relationships between parameters in (2.2.24) and (2.2.13) are

$$\mu = \frac{1}{\beta}\log(\lambda) \quad \text{and} \quad \sigma = \frac{1}{\beta} \quad . \tag{2.2.26}$$

Example 6. Figure 2.6 shows three examples of extreme value density functions. These are:

$$\begin{aligned}
\widehat{f}_1(y) &= \frac{1}{2}\exp\left\{\frac{y+1}{2} - \exp\left(\frac{y+1}{2}\right)\right\}, \\
\widehat{f}_2(y) &= \frac{1}{3.7}\exp\left\{\frac{y-2.6}{3.7} - \exp\left(\frac{y-2.6}{3.7}\right)\right\} \quad \text{and} \\
\widehat{f}_3(y) &= \exp\{y - \exp(y)\} .
\end{aligned}$$

Using (2.2.26) we find the density function of $X = \exp(Y)$. The Weibull densities corresponding to the pdfs above are

$$\begin{aligned} f_1(x) &= \frac{0.5}{e^{-1/2}} x^{-0.5} \exp\left(-\frac{x^{0.5}}{e^{-1/2}}\right), \\ f_2(x) &= \frac{0.27027}{2.0192} x^{-0.72973} \exp\left(-\frac{x^{0.27027}}{2.0192}\right) \quad \text{and} \\ f_3(x) &= x \exp(-x) . \end{aligned}$$

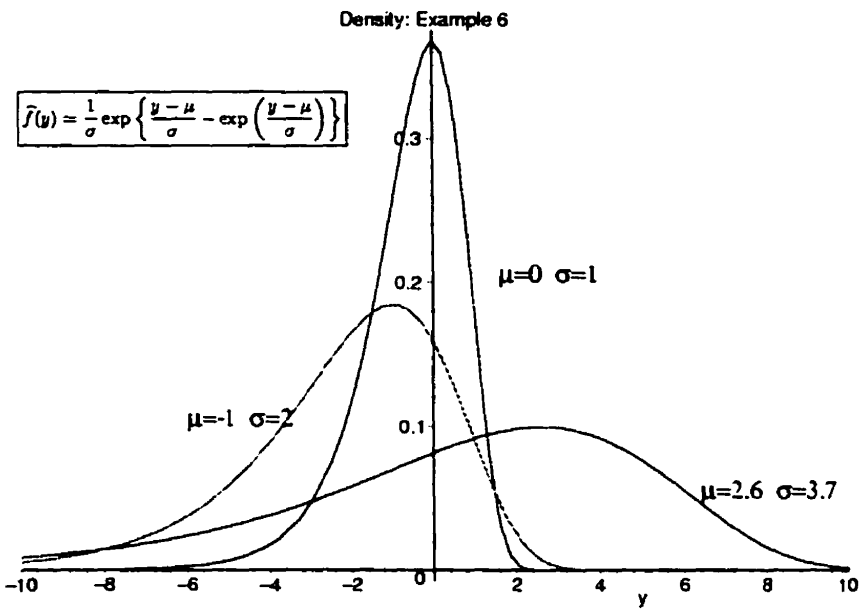

Figure 2.6: Extreme Value Density Functions

2.2.6 Normal and Log-Normal Models

Here we assume that $Y = \log(X)$, where Y is normally distributed. This is equivalent to assuming that X has a log-normal distribution. The pdf of the normal distribution is

$$\widehat{f}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \tag{2.2.27}$$

for $-\infty < y < \infty$, $-\infty < \mu < \infty$ and $\sigma^2 > 0$. The associated pdf for X is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2 x^2}} \exp\left(-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right) \tag{2.2.28}$$

where $x > 0$, $-\infty < \mu < \infty$ and $\sigma^2 > 0$. The survival and hazard functions for X involve the incomplete normal integral.

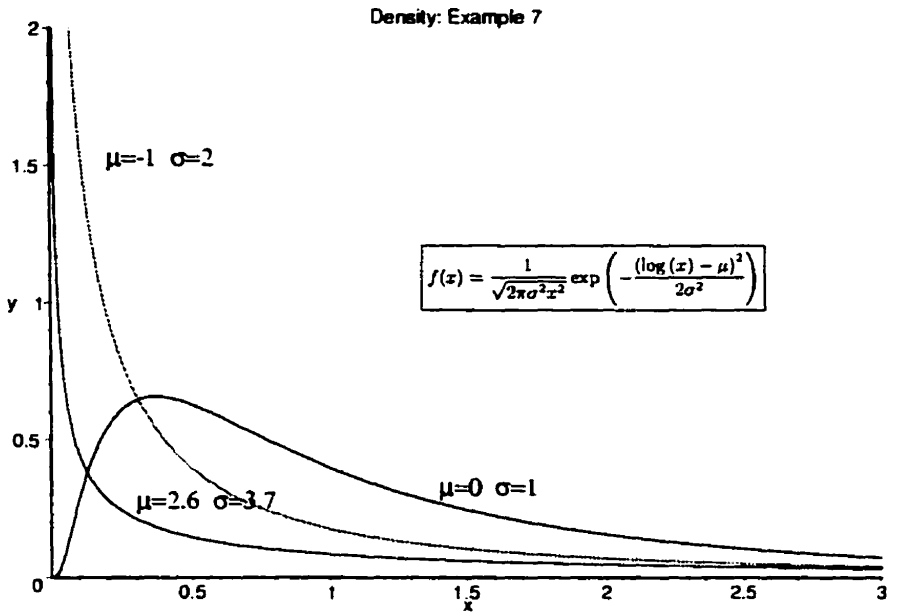

Figure 2.7: Log-Normal Density Functions

Example 7. In figure 2.7, we show three log-normal density functions. These are:

$$f_1(x) = \frac{1}{2x\sqrt{2\pi}} \exp\left(-\frac{(\log(x)+1)^2}{8}\right),$$

$$f_2(x) = \frac{1}{3.7x\sqrt{2\pi}} \exp\left(-\frac{(\log(x)-2.6)^2}{2(3.7)^2}\right) \quad \text{and}$$

$$f_3(x) = \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{(\log(x))^2}{2}\right).$$

2.3 Complex Models

In this section, we examine more complex models such as mixture, series, parallel and replacement models. Here the complexity of the models is increased to better explain life data and real life situations. Each type of model involves k failure times $X_1, X_2, \cdots, X_k$. For

each failure time X_i, $i = 1, 2, \cdots, k$, there is an associated pdf, cdf, survival function and hazard function, denoted by $f_i(x)$, $F_i(x)$, $S_i(x)$ and $h_i(x)$, respectively.

2.3.1 Mixture Models

Thus far, we have considered cases where subjects in a life testing experiment come from the same population. Assume now that we have a mixed population, with k different sub-populations. Each sub-population, i, has an inherent failure time X_i. Let m_i be the proportion of the i^{th} sub-population with respect to the whole population, so that $\sum_{i=1}^{k} m_i = 1$. The survival function for the whole population is

$$\begin{aligned} S(x) &= \mathrm{P}(X \geq x) = \sum_{i=1}^{k} \{m_i \mathrm{P}(X_i \geq x)\} \\ &= \sum_{i=1}^{k} \{m_i S_i(x)\} \ . \end{aligned} \tag{2.2.29}$$

Thus, the pdf of X is given by

$$f(x) = \sum_{i=1}^{k} \{m_i f_i(x)\} \ , \tag{2.2.30}$$

so that the hazard function is

$$h(x) = \frac{\sum_{i=1}^{k} \{m_i f_i(x)\}}{\sum_{i=1}^{k} \{m_i S_i(x)\}} = \frac{\sum_{i=1}^{k} \{m_i h_i(x) S_i(x)\}}{\sum_{i=1}^{k} \{m_i S_i(x)\}} \ . \tag{2.2.31}$$

The above equality indicates that the failure rate of the whole population is a weighted average of the failure rates of the sub-populations.

Example 8. Let $k = 4$ with $m_1 = 0.05$, $m_2 = 0.55$, $m_3 = 0.1$ and $m_4 = 0.3$ so that $\sum_{i=1}^{4} m_i = 1$. Let the failure times for the 4 populations be generalized gamma variables with parameters $\alpha_1 = 1$, $\beta_1 = 1$, $\delta_1 = 1$, $\alpha_2 = 2$, $\beta_2 = 1$, $\delta_2 = 0.5$, $\alpha_3 = 3$, $\beta_3 = 1$, $\delta_3 = 2$, and

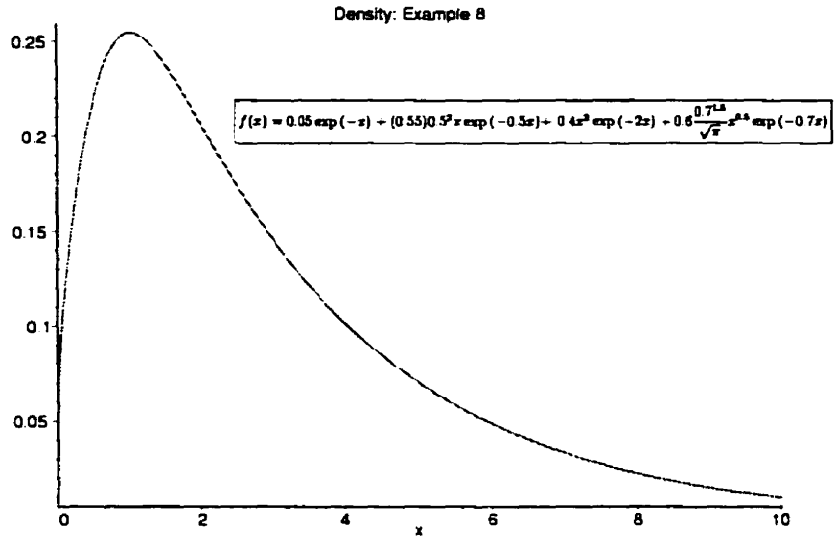

Figure 2.8: Mixtures Density Function

$\alpha_4 = 1.5$, $\beta_4 = 1$, $\delta_4 = 0.7$, respectively. The resulting density

$$\begin{aligned} f(x) &= 0.05\exp(-x) + (0.55)0.5^2 x\exp(-0.5x) + \\ &\quad 0.4x^2\exp(-2x) + 0.6\frac{0.7^{1.5}}{\sqrt{\pi}}x^{0.5}\exp(-0.7x) \end{aligned}$$

is plotted in figure 2.8.

2.3.2 Series Models

Consider a system of k transistors arranged in a series. The failure of any one of these transistors causes the entire system to fail. This model is known as a series model with k modes, where a mode represents a transistor, say. Let X_i be the failure time of the i^{th} mode. Assuming that the failure times are mutually independent, there are two scenarios to consider. If the mode of failure is known, we are in a competing risks model. This is discussed in more detail in Cox and Oakes (1984)) and Lawless (1982). If the mode of failure is not known and the failure time of the whole system is $X = \min(X_1, X_2, \cdots, X_k)$, then only X is observable. In this situation, the survival function of X can be derived through

the use of order statistics.

$$\begin{aligned} S(x) &= \mathrm{P}(X \geq x) = \mathrm{P}(\text{all } X_i \geq x;\ i = 1, 2, \cdots, k) \\ &= \prod_{i=1}^{k} \mathrm{P}(X_i \geq x) \\ &= \prod_{i=1}^{k} S_i(x) \ . \end{aligned} \tag{2.2.32}$$

Also, from (1.1.5), we have the identity $S_i(x) = \exp\left\{-\int_0^x h_i(t)\mathrm{d}t\right\}$. Substituting this in (2.2.32), we obtain

$$\begin{aligned} S(x) &= \exp\left\{-\sum_{i=1}^{k}\int_0^x h_i(t)\mathrm{d}t\right\} \\ &= \exp\left\{-\int_0^x \sum_{i=1}^{k} h_i(t)\mathrm{d}t\right\} \ . \end{aligned} \tag{2.2.33}$$

The representation of $S(x)$ in (2.2.33) gives the hazard function of X by comparison with (1.1.5):

$$h(x) = \sum_{i=1}^{k} h_i(x) \ . \tag{2.2.34}$$

Thus the hazard function in a series model is simply the sum of the hazard functions for each mode. This indicates that the survival or reliability of the system is less than the reliability of the individual components that make up the series.

Example 9. If we have a series with 3 modes where the failure time for each mode is a Weibull random variable, X_i for $i = 1, 2, 3$, with parameters $\beta_1 = 1$, $\lambda_1 = 2.3$, $\beta_2 = 1.7$, $\lambda_2 = 0.5$ and $\beta_3 = 2$, $\lambda_3 = 1.1$, respectively, then the survival function of $X = \min(X_1, X_2, X_3)$ is

$$S(x) = \exp\left(-\frac{x}{2.3} + \frac{x^{1.7}}{0.5} + \frac{x^2}{1.1}\right) \tag{2.2.35}$$

and its hazard function is

$$h(x) = \frac{1}{2.3} + 3.4x^{0.7} + \frac{2}{1.1}x \ . \tag{2.2.36}$$

These functions are plotted in figure 2.9.

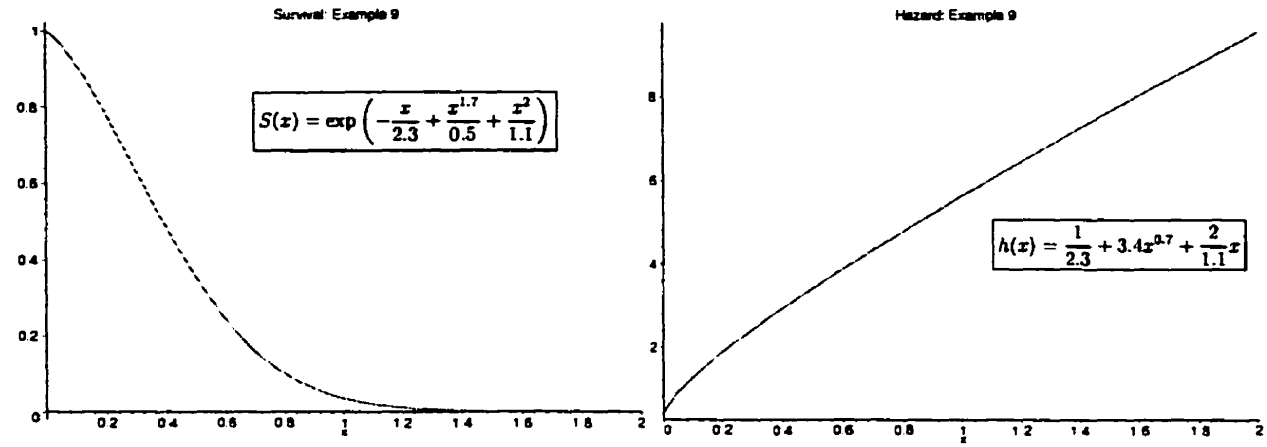

Figure 2.9: Series Survival and Hazard Functions

2.3.3 Parallel Models

Consider now a system where k components are connected in parallel. Each component i has a survival function denoted by $S_i(x)$. For each survival function $S_i(x)$ there is an associated cdf and pdf denoted by $F_i(x)$ and $f_i(x)$, respectively. The system fails when all of the components fail. This system is called a parallel model and the failure time of the system is $X = \max(X_1, X_2, \cdots, X_k)$. Let's assume that only X is observable and that the failure times X_i are mutually independent. The survival function of X is given in terms of its cdf.

$$\begin{aligned} S(x) &= \mathrm{P}(X \geq x) = 1 - \mathrm{P}(X < x) \\ &= 1 - \mathrm{P}(\text{all } X_i < x) \\ &= 1 - \prod_{i=1}^{k} \mathrm{P}(X_i < x) \\ &= 1 - \prod_{i=1}^{k} F_i(x) \ . \end{aligned} \tag{2.2.37}$$

Differentiation of this equality gives the pdf, $f(x)$, of X

$$f(x) = \sum_{j=1}^{k} \left[f_j(x) \prod_{\substack{i=1 \\ i \neq j}}^{k} \{F_i(x)\} \right] \ . \tag{2.2.38}$$

Hence, the hazard is simply the ratio of the pdf and the survival function as defined in (1.1.3):

$$h(x) = \frac{\sum_{j=1}^{k}\left[f_j(x)\prod_{\substack{i=1\\ i\neq j}}^{k}\{F_i(x)\}\right]}{1-\prod_{i=1}^{k}F_i(x)}. \qquad (2.2.39)$$

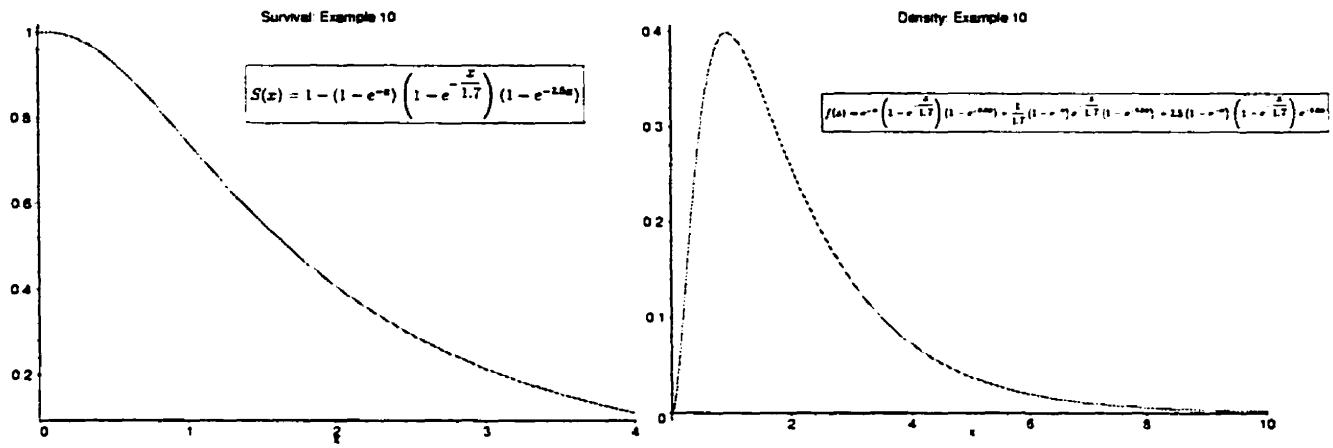

Figure 2.10: Parallel Survival and Density Functions

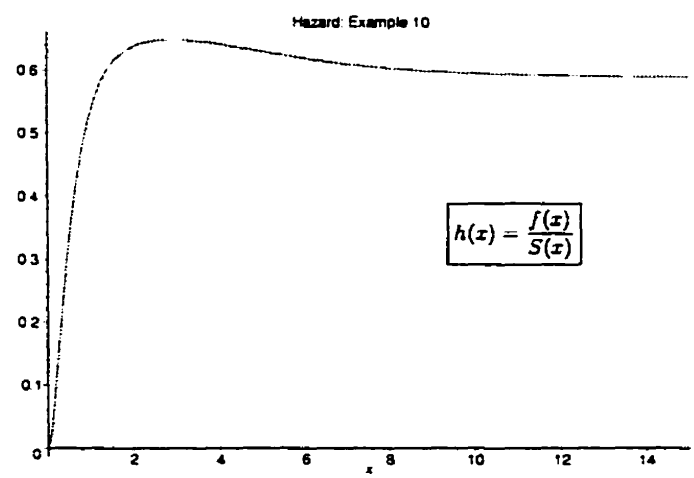

Figure 2.11: Parallel Hazard Function

Example 10. Examine a parallel system with three components all having exponential failure times with parameters $\lambda_1 = 1$, $\lambda_2 = 1.7$ and $\lambda_3 = 0.4$, respectively. Then the survival

function for the failure time of the system is

$$S(x) = 1 - (1 - e^{-x}) \left(1 - e^{-\frac{x}{1.7}}\right) (1 - e^{-2.5x}) \ .$$

and the pdf of the failure time is

$$\begin{aligned} f(x) \ = \ & e^{-x} \left(1 - e^{-\frac{x}{1.7}}\right) (1 - e^{-2.5x}) + \\ & \frac{1}{1.7} (1 - e^{-x}) e^{-\frac{x}{1.7}} (1 - e^{-2.5x}) + \\ & 2.5 (1 - e^{-x}) \left(1 - e^{-\frac{x}{1.7}}\right) e^{-2.5x} \ . \end{aligned}$$

The survival function and pdf are plotted in figure 2.10.

2.3.4 Replacement Models

In real world applications, it is often necessary to have spare parts available for a crucial component in a system. One might be interested in evaluating the reliability of a component, given that there are only k spare parts for it. Alternatively, one may wish to know what the number k should be to attain a certain level of reliability.

Here, we denote the failure time of the first component as X_1, and the failure time of replacement part i as X_i, $i = 2, 3, \cdots, k+1$. Also, we assume that the failure times X_i, $i = 1, 2, \cdots, k+1$, are mutually independent. The failure time for the whole process of failure and replacement is the sum of all the failure times, $X = X_1 + X_2 + \cdots + X_{k+1}$.

Depending on the type of model for each X_i, the survival function of X may not be so easy to calculate. If the failure times are all from the exponential model with the same parameter λ, then the survival function of X is easily computed from the moment generating

function in (2.2.4);

$$\begin{aligned} M_X(t) &= M_{\sum_{i=1}^{k+1} X_i}(t) \\ &= \prod_{i=1}^{k+1} M_{X_i}(t) \\ &= \left(\frac{1}{1-\lambda t}\right)^{k+1} \quad \text{for } t < 1/\lambda. \end{aligned} \tag{2.2.40}$$

$M_X(t)$ is the moment generating function of a gamma model (2.2.12) with parameters λ and $\alpha = k+1$. If we generalize further and let the failure times $X_1, X_2, \cdots, X_{k+1}$ be all from the gamma model, then the survival function is more difficult to compute. Results developed and discussed in chapter 3 will give computable forms for the survival function of a replacement model with gamma type failure times.

2.3.5 Linear Risk Models

In the exponential model, we begin with the assumption that X has a constant hazard rate. Assume now that X has a linear hazard rate

$$h(x) = \begin{cases} \lambda_1 + \lambda_2 x & \text{for } x > 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{2.2.41}$$

The integrated hazard function is obtained by integrating (2.2.41) with respect to x and it is

$$H(x) = \begin{cases} \lambda_1 x + \frac{\lambda_2}{2} x^2 & \text{for } x > 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{2.2.42}$$

Using results in chapter 1, one has

$$S(x) = \begin{cases} \exp\left(-\lambda_1 x - \frac{\lambda_2}{2} x^2\right) & \text{for } x > 0 \\ 1 & \text{elsewhere} \end{cases} \tag{2.2.43}$$

and

$$f(x) = \begin{cases} (\lambda_1 + \lambda_2 x)\exp\left(-\lambda_1 x - \frac{\lambda_2}{2}x^2\right) & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases} \tag{2.2.44}$$

for the survival and pdf of X, respectively. Clearly this model can be extended to the case where the hazard is a polynomial, $h(x) = \lambda_1 + \lambda_2 x + \cdots + \lambda_n x^n$.

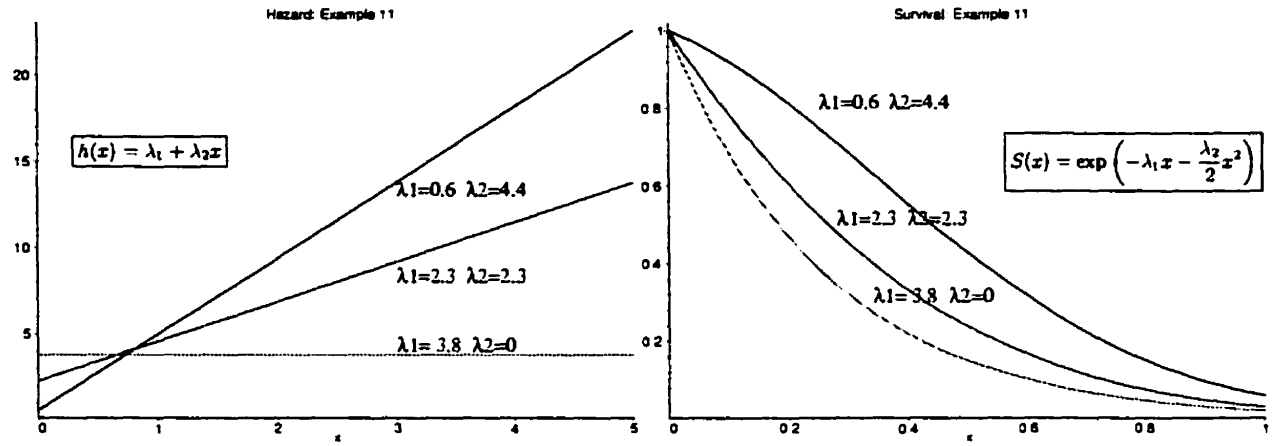

Figure 2.12: Linear Risk Hazard and Survival Functions

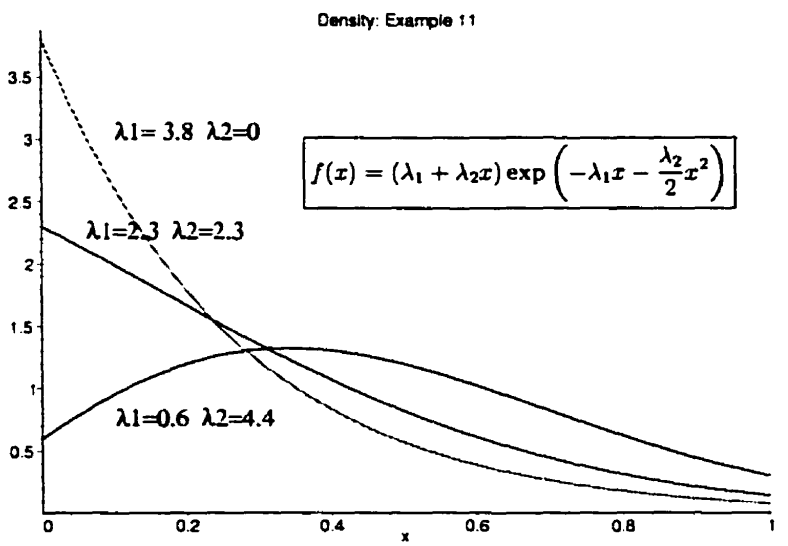

Figure 2.13: Linear Risk Density Function

Example 11. Figures 2.12 and 2.13 give three examples of linear risk models. The functions

plotted correspond to models with the following hazard functions:

$$h_1(x) = 3.8\,, \tag{2.2.45}$$

$$h_2(x) = 2.3 + 2.3x \quad \text{and} \tag{2.2.46}$$

$$h_3(x) = 0.6 + 4.4x\,. \tag{2.2.47}$$

2.3.6 Regression Models

The inclusion of regressor variables for the failure time X in a life testing experiment is another way to account for homogeneity in a population. These regressors are typically extra variables related to the failure. For example in a lung cancer study, regressors may be such variables as age and sex of a patient and the size and type of tumor. In a reliability study for an electrical component, one may have the amount of stress or voltage administered and the model of the component as regressors.

Regression models allow for the inclusion of regressor variables in life testing experiments. The survival function $S(x)$ no longer depends alone on the failure time X. If $Y_1, Y_2, \cdots, Y_l$ are the regressors of a model then $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_l)$ is the regressor vector. The survival function is now a function of X and the regressors $\mathbf{Y}$, and thus is denoted by $S(x|\mathbf{Y})$. Regression models are discussed in detail by Kalbfleisch and Prentice (1980) and Lawless (1982).

2.3.7 Remarks

There are many reasons to select a particular life testing model. For instance, past experience and knowledge of the underlying failure process may suggest the validity of a chosen model.

Often, one model is abandoned for another because of ease of use in computations. This is often the case with the exponential or Weibull model where the survival function and hazard function have closed forms. The gamma model however does not have closed forms for $S(x)$ and $h(x)$ but it has a closed form for its mgf $M_X(t)$ which makes it particularly useful for cases with replacement models. In chapter 3, the mgf of the gamma model will be used to develop techniques for gamma type models.

2.4 Estimation and Approximations

Suppose that the survival function and related life testing functions need to be estimated based on failure time data $y_1, y_2, \cdots, y_n$ from n subjects. There are two cases to consider. One is that the life testing experiment lasts as long as all the subjects survive. In this case, our random sample is complete in the sense that we have failure times for all the subjects. Here one uses standard elementary statistical methods to analyze failure time data and to check for the validity of a particular model. The other case to consider is that of a life testing experiment which yields only partial information on the failure times of the subjects. The analysis of a failure process when censoring is present varies depending on the form of the censoring mechanism. Note that, as outlined in chapter 1, determination of either the hazard function, pdf or survival function determines a given life testing model uniquely.

For the next portion of our discussion, we confine our results to the case where the failure times of subjects are continuous. Furthermore, we only examine type I, type II and independent censoring. For more general censoring mechanisms such as independent censoring, results are similar to those of random censoring and we refer the reader to Kalbfleisch and

Prentice (1980), Lawless (1982) and Fleming and Harrington (1991).

2.4.1 Parametric Estimation

We now give some details on maximum likelihood for the purpose of estimating the parameters $\underline{\theta} = (\theta_1, \theta_2, \cdots, \theta_k,)^t$ of a certain life testing model. The method of maximum likelihood is used to fit a parametric model to an observed set of life testing data. Let X_i, $i = 1, \cdots, n$, be the failure times of n subjects with censoring times C_i, $i = 1, \cdots, n$. The observed life data is best described using the pair of random variables (Y_i, δ_i) where

$$Y_i = \min(X_i, C_i) \quad i = 1, 2, \cdots, n \tag{2.2.48}$$

and

$$\delta_i = \begin{cases} 1 & \text{if } x_i \leq c_i \\ 0 & \text{if } x_i > c_i \ . \end{cases} \tag{2.2.49}$$

The observed data is thus the pairs $(y_1, \delta_1), \cdots, (y_n, \delta_n)$ where $y_i = \min(x_i, c_i)$ for $i = 1, 2, \cdots, n$. Furthermore, assume that the failure times of the n subjects are independent and identically distributed with pdf $f(x)$ and survival function $S(x)$. We use this notation throughout the section.

Complete Sample

We begin with the assumption that the failure times have been observed for the n subjects in a life testing experiment. Hence, no censoring times have been observed. The likelihood of the n observations, $x_1, x_2, \cdots, x_n$ is simply

$$L(\underline{\theta}) = \prod_{i=1}^{n} f(x_i) \ . \tag{2.2.50}$$

The likelihood is a function of the unknown parameters $\underline{\theta} = (\theta_1, \theta_2, \cdots, \theta_k)^t$. In most cases, it is more convenient to work with the logarithm of the likelihood function, $\log(L(\underline{\theta}))$. The values of $\underline{\theta}$ which maximize $L(\underline{\theta})$ are the same values which maximize $\log(L(\underline{\theta}))$. Maximum likelihood estimators and confidence intervals for complete samples for many life models are discussed in detail in Zacks (1992).

Type II Censoring

In type II censoring, only the r smallest failure times are observed, $1 \leq r < n$. Consider the r failure times $y_1, y_2, \cdots, y_r$ and order them to find the ordered sample $x_{(1)} < x_{(2)} < \cdots < x_{(r)}$. From results in order statistics (see Balakrishnan and Cohen (1991)) we have that the joint distribution of the ordered failure time sample is

$$L = \frac{n!}{(n-r)!} f(x_{(1)}) f(x_{(2)}) \cdots f(x_{(r)}) \left[S(x_{(r)})\right]^{n-r} \tag{2.2.51}$$

The equation in (2.2.51) gives the likelihood function for type II censoring. Note that $y_{(i)} = x_{(r)}$ for $i = r+1, r+2, \cdots, n$. As a result, type II is not a case of random censoring.

Type I and Random Censoring

In this case, we assume random censoring which includes type I censoring ($C_i = C \ \forall \ i = 1, 2, \cdots, n$, C is a constant). Let the random censoring times denoted by $C_1, C_2, C_3, \cdots, C_n$ be independent and identically distributed with pdf $g(x)$ and survival function $G(x)$. Then the likelihood of the pair (y_i, δ_i) is easily obtained by

$$\begin{aligned} \mathrm{P}(Y_i = t, \delta_i = 0) &= \mathrm{P}(C_i = t, X_i > C_i) \\ &= g(t)S(t) \qquad \text{and} \end{aligned} \tag{2.2.52}$$

$$\begin{aligned} \mathrm{P}\left(Y_i = t, \delta_i = 1\right) &= \mathrm{P}\left(X_i = t, X_i \leq C_i\right) \\ &= f(t)G(t) \end{aligned} \tag{2.2.53}$$

so that

$$\mathrm{P}(Y_i = t, \delta_i) = \left[f(t)G(t)\right]^{\delta_i}\left[g(t)S(t)\right]^{1-\delta_i} . \tag{2.2.54}$$

Hence, the likelihood function for the pairs $(y_1, \delta_1), \cdots, (y_n, \delta_n)$ is given by

$$L = L\left(\underline{\theta} | (y_1, \delta_1), \cdots, (y_n, \delta_n)\right) = \prod_{i=1}^{n}\left[f(y_i)G(y_i)\right]^{\delta_i}\left[g(y_i)S(y_i)\right]^{1-\delta_i} . \tag{2.2.55}$$

Under the assumption that the censoring times have no connection to the failure times of the subjects, the products involving $g(y_i)$ and $G(y_i)$ do not involve parameters of the survival function of the failure times. This assumption is known as non-informative censoring. The censoring functions in this case do not need to be considered when maximizing the likelihood function (2.2.55). Note that for non-informative censoring, the likelihood function for Type II censoring is the same as that for Type I and random censoring. It also turns out that the likelihood above under non-informative censoring

$$L(\underline{\theta}) = \prod_{i=1}^{n}\left[f(y_i)\right]^{\delta_i}\left[S(y_i)\right]^{1-\delta_i} \tag{2.2.56}$$

is also the likelihood for independent censoring under non-informative censoring (see appendix A).

Numerical Solutions

As described in Miller (1981), the method of scoring and the Newton-Raphson method are useful in finding numerical solutions to $\max_{\underline{\theta}} L(\underline{\theta})$. This is equivalent to finding the solution $\widehat{\underline{\theta}}$ to the system of equations

$$\underline{0} = \frac{\partial}{\partial \underline{\theta}} \log\left(L(\underline{\theta})\right), \tag{2.2.57}$$

where $\frac{\partial}{\partial \underline{\theta}} \log(L(\underline{\theta})) = \left(\frac{\partial}{\partial \theta_1} \log(L(\underline{\theta})), \cdots, \frac{\partial}{\partial \theta_k} \log(L(\underline{\theta}))\right)^t$.

The Newton-Raphson method is the recursive solution of (2.2.57) given by

$$\hat{\underline{\theta}}^{l+1} = \hat{\underline{\theta}}^{l} + \left[-\left(\frac{\partial}{\partial \underline{\theta}}\right)\left(\frac{\partial}{\partial \underline{\theta}}\right)^t \log(L(\underline{\theta}))\Big|_{\underline{\theta}=\hat{\underline{\theta}}^l}\right]^{-1} \frac{\partial}{\partial \underline{\theta}} \log(L(\underline{\theta}))\Big|_{\underline{\theta}=\hat{\underline{\theta}}^l} . \tag{2.2.58}$$

The vector

$$\frac{\partial}{\partial \underline{\theta}} \log(L(\underline{\theta}))\Big|_{\underline{\theta}=\hat{\underline{\theta}}^l} \tag{2.2.59}$$

is called the score vector at $\hat{\underline{\theta}}^l$ and the matrix

$$\mathbf{i}\left(\hat{\underline{\theta}}^{l}\right) = -\left(\frac{\partial}{\partial \underline{\theta}}\right)\left(\frac{\partial}{\partial \underline{\theta}}\right)^t \log(L(\underline{\theta}))\Big|_{\underline{\theta}=\hat{\underline{\theta}}^l} \tag{2.2.60}$$

$$= \begin{pmatrix} -\frac{\partial^2}{\partial \theta_1^2} \log(L(\underline{\theta}))\Big|_{\underline{\theta}=\hat{\underline{\theta}}^l} & \cdots & -\frac{\partial^2}{\partial \theta_1 \partial \theta_k} \log(L(\underline{\theta}))\Big|_{\underline{\theta}=\hat{\underline{\theta}}^l} \\ \vdots & \ddots & \vdots \\ -\frac{\partial^2}{\partial \theta_k \partial \theta_1} \log(L(\underline{\theta}))\Big|_{\underline{\theta}=\hat{\underline{\theta}}^l} & \cdots & -\frac{\partial^2}{\partial \theta_k^2} \log(L(\underline{\theta}))\Big|_{\underline{\theta}=\hat{\underline{\theta}}^l} \end{pmatrix} \tag{2.2.61}$$

is called the sample information matrix at $\hat{\underline{\theta}}^l$.

If the operations of integration with respect to y_i and differentiation with respect to $\underline{\theta}$ can be interchanged then one can show that the score vector $\frac{\partial}{\partial \underline{\theta}} \log(L(\underline{\theta}))$ has expectation 0 and the covariance matrix

$$\mathrm{E}(\mathbf{i}(\underline{\theta})) = I(\underline{\theta}), \tag{2.2.62}$$

where $I(\underline{\theta})$ is the Fisher information matrix:

$$I(\underline{\theta}) = \mathrm{E}\left(\left[\frac{\partial}{\partial \underline{\theta}} \log(L(\underline{\theta}))\right]\left[\frac{\partial}{\partial \underline{\theta}} \log(L(\underline{\theta}))\right]^t\right) \tag{2.2.63}$$

If one uses $I(\theta)$ in (2.2.58), the recursive solution to (2.2.57) is

$$\hat{\underline{\theta}}^{l+1} = \hat{\underline{\theta}}^{l} + \left[I(\hat{\underline{\theta}}^{l})\right]^{-1} \frac{\partial}{\partial \underline{\theta}} \log(L(\underline{\theta}))\Big|_{\underline{\theta}=\hat{\underline{\theta}}^l} . \tag{2.2.64}$$

This is known as the method of scoring. The derivation for the Newton-Raphson method is given in appendix B where other optimization techniques are also discussed.

Confidence Intervals Under General Censoring

Miller (1981) gives three methods for confidence intervals under general independent censoring. He states that under smoothness conditions, the maximum likelihood estimator $\underline{\widehat{\theta}}$ is asymptotically distributed as a normal random variable with mean value $\underline{\theta}$ and covariance matrix $[I(\underline{\theta})]^{-1}$. There are three methods for testing $H_0 : \underline{\theta} = \underline{\theta^0}$ versus $H_1 : \underline{\theta} \neq \underline{\theta^0}$. These are based on:

- **Wald's Statistic**

$$\left(\underline{\widehat{\theta}} - \underline{\theta^0}\right)^t I(\underline{\theta^0}) \left(\underline{\widehat{\theta}} - \underline{\theta^0}\right)$$

- **Score Statistic (Rao's Method)**

$$\left[\frac{\partial}{\partial \underline{\theta}} \log\left(L(\underline{\theta})\right)\Big|_{\underline{\theta}=\underline{\theta^0}}\right]^t [I(\underline{\theta^0})]^{-1} \left[\frac{\partial}{\partial \underline{\theta}} \log\left(L(\underline{\theta})\right)\Big|_{\underline{\theta}=\underline{\theta^0}}\right]$$

- **Likelihood Ratio Statistic (Neyman-Pearson/Wilks)**

$$-2 \log\left\{\frac{L(\underline{\theta^0})}{L(\underline{\widehat{\theta}})}\right\} .$$

These three statistics are asymptotically distributed as a chisquare with k degrees of freedom under H_0. In most situations, particularly when censoring is present, it is difficult to calculate Fisher's information matrix $I(\underline{\theta})$. Instead we use the approximation $\mathbf{i}(\underline{\theta})$ in the statistics given above.

Let $U(\underline{\theta^0})$ denote one of the statistics above. To find a confidence interval for $U(\underline{\theta^0})$, one finds a constant $\chi^2_k(\alpha)$ such that

$$\mathrm{P}\left(U(\underline{\theta^0}) \leq \chi^2_k(\alpha)\right) \approx \mathrm{P}\left(Z \leq \chi^2_k(\alpha)\right) = 1 - \alpha \, , \tag{2.2.65}$$

and Z is distributed as a chisquare with k degrees of freedom. Thus one rejects the null hypothesis, H_0, when $U(\underline{\theta^0}) > \chi^2_k(\alpha)$. Cox and Oakes (1984) state that these three methods

usually yield the same results, but that the likelihood ratio statistic is strongly recommended. For more on properties of the maximum likelihood estimator, see Barndorff-Nielsen (1983), Barndorff-Nielsen and Cox (1984), Bartholomew (1963), Efron and Hinkley (1978) and Rao (1973).

Example 12 (Exponential Model: No Censoring). Assume that we have a complete sample. Let $(x_1, x_2, x_3, \cdots, x_n)$ be random failure times and assume that the failure times are independent and identically distributed under the exponential model. Given the observed failure times, we wish to estimate the mean life λ of the exponential model. The log-likelihood function for this model given the data is

$$l(\lambda) = -n \log(\lambda) - \sum_{i=1}^{n} \frac{x_i}{n} . \tag{2.2.66}$$

The maximum likelihood estimator of λ is the solution to

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} l(\lambda) = \frac{\sum_{i=1}^{n} \frac{x_i}{n}}{\lambda^2} - \frac{n}{\lambda} = 0 \tag{2.2.67}$$

which yields $\widehat{\lambda} = \sum_{i}^{n} \frac{x_i}{n}$. Note that $E(\widehat{\lambda}) = \lambda$ and $Var(\widehat{\lambda}) = \frac{\lambda^2}{n}$. The properties of $\widehat{\lambda}$ for the exponential model are known. Sinha (1986) shows that $\widehat{\lambda}$ is the uniformly minimum variance unbiased estimator (UMVUE) of λ.

Example 13 (Exponential Model: With Censoring). In this example, assume that we have right independent censoring. The failure time data consists of $(y_1, y_2, \cdots, y_n)$ along with the indicators $(\delta_1, \delta_2, \cdots, \delta_n)$. Further, let the failure times be independent and identically distributed under the exponential model. The log-likelihood function is given by

$$l(\lambda) = -\sum_{i=1}^{n} [\delta_i \log(\lambda)] - \frac{1}{\lambda} \sum_{i=1}^{n} y_i . \tag{2.2.68}$$

The maximum likelihood estimator under these assumptions is given by solving

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} l(\lambda) = -\sum_{i=1}^{n} \frac{\delta_i}{\lambda} + \sum_{i=1}^{n} \frac{y_i}{\lambda^2} = 0 \,. \tag{2.2.69}$$

The solution to (2.2.69) is $\widehat{\lambda} = \left(\sum_i^n y_i\right) \Big/ \left(\sum_i^n \delta_i\right)$. Note that this solution makes sense if there is at least one failure in our sample; that is there is at least one i such that $\delta_i = 1$. Now an estimate for the variance of $\widehat{\lambda}$ is given by looking at the second derivative of the log-likelihood function (2.2.68) with respect to λ.

$$-\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} l \Big|_{\widehat{\lambda}} = \frac{\left(\sum_{i=1}^{n} \delta_i\right)}{\widehat{\lambda}^2} \tag{2.2.70}$$

so that $s.e.(\widehat{\lambda}) \approx \dfrac{\widehat{\lambda}}{\left(\sqrt{\sum \delta_i}\right)}$, where $s.e.$ is the standard error.

Functions of the MLE

Let $\underline{\widehat{\theta}}$ be the calculated maximum likelihood estimator for $\underline{\theta}$ in a given model. A quantity of interest may be a function of $\underline{\theta}$. For instance, we may have

$$\phi = \mathbf{m}(\underline{\theta}) \,, \tag{2.2.71}$$

where $\mathbf{m}$ is a one-to-one function. From the theory of maximum likelihood estimators, the maximum likelihood estimator of ϕ is simply

$$\widehat{\phi} = \mathbf{m}(\underline{\widehat{\theta}}) \,. \tag{2.2.72}$$

For example, researchers in reliability and survival analysis are often interested in estimating a quantile, $q(p)$ say, of a particular life testing model. Here, $q(p)$ is the quantity such that

$$\mathrm{P}\left(X \geq q(p)\right) = S\left(q(p) \middle| \underline{\theta}\right) = p \,, \tag{2.2.73}$$

where $0 < p < 1$ is specified before hand. The maximum likelihood estimator of the pth quantile is given by solving $\widehat{q}(p)$ in

$$S\left(\widehat{q}(p)\middle|\widehat{\underline{\theta}}\right) = p\ . \tag{2.2.74}$$

The sample information matrix $\mathbf{i}\left(\widehat{\underline{\theta}}\right)$ is the estimated covariance matrix of the maximum likelihood estimator $\widehat{\underline{\theta}}$. Hence, if

$$\mathbf{i}\left(\widehat{\underline{\theta}}\right) = \{\widehat{\sigma}_{ij}\}\ , \tag{2.2.75}$$

where $i = 1, \cdots, k$ and $j = 1, \cdots, k$, then $\sqrt{(\widehat{\sigma}_{ii})}$ is an estimate of the standard error of the parameter $\widehat{\theta}_i$ such that $\widehat{\underline{\theta}} = (\widehat{\theta}_1, \cdots, \widehat{\theta}_k)$.

The standard error of the maximum likelihood estimator $\widehat{\phi}$ is needed to make any inference about $\phi = \mathbf{m}(\underline{\theta})$. Crowder, Kimber, Smith, and Sweeting (1991) give a standard procedure for estimating the variance of ϕ. It is called the delta method and it gives the standard error as

$$s.e.\left(\widehat{\phi}\right) \approx \left\{\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\frac{\partial}{\partial\theta_i}\mathbf{m}(\underline{\theta})\frac{\partial}{\partial\theta_j}\mathbf{m}(\underline{\theta})\right]\widehat{\sigma}_{ij}\right\}^{1/2}\ , \tag{2.2.76}$$

where the partial derivatives are evaluated at $\widehat{\underline{\theta}}$. If we have $k = 1$ then $\underline{\theta}$ is a scalar, θ say, then $s.e.(\widehat{\phi}) \approx \left|\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbf{m}\right|\Big|_{\theta=\widehat{\theta}}\sqrt{\widehat{\sigma}_{11}}$. A derivation of the delta method is given in appendix C.

Example 14 (Weibull Model: With Censoring). Assume that a life testing sample $(y_1, y_2, \cdots, y_n)$ along with indicators $(\delta_1, \delta_2, \cdots, \delta_n)$ arises from a Weibull model. The log-likelihood is

$$l(\lambda, \beta) = \sum_{i=1}^{n}\delta_i \log(\beta) - \sum_{i=1}^{n}\delta_i \log(\lambda) + (\beta - 1)\sum_{i=1}^{n}\delta_i log(y_i) - \frac{1}{\lambda}\sum_{i=1}^{n}y_i^{\beta}\ . \tag{2.2.77}$$

With the transformation $z_i = \log(y_i)$, it is equivalent to consider the log-likelihood from an extreme value model (see section (2.2.5)). The log-likelihood of the extreme value model is

$$l(\mu, \sigma) = -\sum_{i=1}^{n}\delta_i \log(\sigma) - \sum_{i=1}^{n}\delta_i\frac{\mu}{\sigma} + \frac{(1-\sigma)}{\sigma}\sum_{i=1}^{n}\delta_i z_i - \sum_{i=1}^{n}\exp\left(\frac{z_i - \mu}{\sigma}\right)\ . \tag{2.2.78}$$

The partial derivatives of $l(\mu,\sigma)$ are

$$\frac{\partial}{\partial\mu}l(\mu,\sigma) = -\frac{\sum_{i=1}^{n}\delta_i}{\sigma} + \frac{\sum_{i=1}^{n}\exp\left(\frac{z_i-\mu}{\sigma}\right)}{\sigma} \tag{2.2.79}$$

$$\frac{\partial}{\partial\sigma}l(\mu,\sigma) = -\frac{\sum_{i=1}^{n}\delta_i}{\sigma} - \frac{\sum_{i=1}^{n}\delta_i z_i}{\sigma^2} + \frac{\mu\sum_{i=1}^{n}\delta_i}{\sigma^2} + \frac{\sum_{i=1}^{n}(z_i-\mu)\exp\left(\frac{z_i-\mu}{\sigma}\right)}{\sigma^2} \tag{2.2.80}$$

$$\frac{\partial^2}{\partial\mu\partial\sigma}l(\mu,\sigma) = \frac{\sum_{i=1}^{n}\delta_i}{\sigma^2} - \sum_{i=1}^{n}\left[\frac{\exp\left(\frac{z_i-\mu}{\sigma}\right)}{\sigma^2} + \frac{(z_i-\mu)\exp\left(\frac{z_i-\mu}{\sigma}\right)}{\sigma^3}\right] \tag{2.2.81}$$

$$\frac{\partial^2}{\partial\mu^2}l(\mu,\sigma) = -\frac{\sum_{i=1}^{n}\exp\left(\frac{z_i-\mu}{\sigma}\right)}{\sigma^2} \tag{2.2.82}$$

$$\frac{\partial^2}{\partial\sigma^2}l(\mu,\sigma) = \frac{\sum_{i=1}^{n}\delta_i}{\sigma^2} + 2\frac{\sum_{i=1}^{n}\delta_i z_i}{\sigma^3} - 2\frac{\mu\sum_{i=1}^{n}\delta_i}{\sigma^3} - \sum_{i=1}^{n}\left[\frac{2(z_i-\mu)\exp\left(\frac{z_i-\mu}{\sigma}\right)}{\sigma^3} + \frac{(z_i-\mu)^2\exp\left(\frac{z_i-\mu}{\sigma}\right)}{\sigma^4}\right]. \tag{2.2.83}$$

The maximum likelihood estimators of (μ,σ) are given by solving

$$\widehat{\mu} = \widehat{\sigma}\log\left(\frac{\sum_{i=1}^{n}\exp\left(\frac{z_i}{\widehat{\sigma}}\right)}{\left(\sum_{i=1}^{n}\delta_i\right)}\right) \quad \text{and} \tag{2.2.84}$$

$$0 = \frac{\sum_{i=1}^{n}\delta_i z_i}{\sum_{i=1}^{n}\delta_i} + \widehat{\sigma} - \frac{\sum_{i=1}^{n} z_i\exp\left(\frac{z_i}{\widehat{\sigma}}\right)}{\sum_{i=1}^{n}\exp\left(\frac{z_i}{\widehat{\sigma}}\right)}. \tag{2.2.85}$$

Hence, solving for $(\widehat{\mu},\widehat{\sigma})$ simultaneously reduces to solving for $\widehat{\sigma}$ in (2.2.85) by numerical methods and then substituting back into (2.2.84) to find $\widehat{\mu}$.

To obtain the sample information matrix at $(\widehat{\mu},\widehat{\sigma})$, one needs only to evaluate the second order partial derivatives (2.2.81), (2.2.82) and (2.2.83) at $(\widehat{\mu},\widehat{\sigma})$. This simplifies the equations

somewhat:

$$\left.\frac{\partial^2}{\partial\mu^2}l(\mu,\sigma)\right|_{(\widehat{\mu},\widehat{\sigma})} = -\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\sigma}^2} \tag{2.2.86}$$

$$\begin{aligned}\left.\frac{\partial^2}{\partial\sigma^2}l(\mu,\sigma)\right|_{(\widehat{\mu},\widehat{\sigma})} &= \frac{\sum_{i=1}^{n}\delta_i}{\widehat{\sigma}^2} + 2\frac{\sum_{i=1}^{n}\delta_i z_i}{\widehat{\sigma}^3} \\ &\quad - \sum_{i=1}^{n}\left[\frac{2z_i\exp\left(\frac{z_i-\widehat{\mu}}{\widehat{\sigma}}\right)}{\widehat{\sigma}^3} + \frac{(z_i-\widehat{\mu})^2\exp\left(\frac{z_i-\widehat{\mu}}{\widehat{\sigma}}\right)}{\widehat{\sigma}^4}\right]\end{aligned} \tag{2.2.87}$$

$$\left.\frac{\partial^2}{\partial\mu\partial\sigma}l(\mu,\sigma)\right|_{(\widehat{\mu},\widehat{\sigma})} = \sum_{i=1}^{n}\left[\frac{(z_i-\widehat{\mu})\exp\left(\frac{z_i-\widehat{\mu}}{\widehat{\sigma}}\right)}{\widehat{\sigma}^3}\right] . \tag{2.2.88}$$

These equalities are used in constructing the sample information matrix $\mathbf{i}(\widehat{\mu},\widehat{\sigma})$ given by (2.2.61). Given that the parameters of the extreme value and Weibull models are related by $\mu = 1/\beta\log(\lambda)$ and $\sigma = 1/\beta$, one can use the delta method to find estimates for the standard errors of $\widehat{\lambda}$ and $\widehat{\beta}$, the maximum likelihood estimators of λ and β. The method gives

$$Var(\widehat{\beta}) \approx \frac{1}{\widehat{\sigma}^4}Var(\widehat{\sigma}) \quad \text{and} \tag{2.2.89}$$

$$\begin{aligned}Var(\widehat{\lambda}) &\approx \frac{\exp\left(\frac{2\widehat{\mu}}{\widehat{\sigma}}\right)}{\widehat{\sigma}^2}Var(\widehat{\mu}) + \frac{\widehat{\mu}^2\exp\left(\frac{2\widehat{\mu}}{\widehat{\sigma}}\right)}{\widehat{\sigma}^4}Var(\widehat{\sigma}) \\ &\quad - \frac{2\widehat{\mu}\exp\left(\frac{2\widehat{\mu}}{\widehat{\sigma}}\right)}{\widehat{\sigma}^3}Cov(\widehat{\mu},\widehat{\sigma}) .\end{aligned} \tag{2.2.90}$$

An estimate for the covariance matrix of $(\widehat{\mu},\widehat{\sigma})$ is obtained by examining the inverse of the sample information matrix $\mathbf{i}(\widehat{\mu},\widehat{\sigma})$.

Example 15 (Gamma Model). For the complete sample case, results for point estimation for the gamma and generalized gamma models have been discussed thoroughly in Lawless (1982) and Hager and Bain (1970). As in the example above for the Weibull model, we give

the general form for the log-likelihood under type I, type II and random censoring for the gamma model.

We begin with the life testing sample $(y_1, y_2, \cdots, y_n)$ along with the indicators $(\delta_1, \delta_2, \cdots, \delta_n)$. Define $\overline{y}$ and $\tilde{y}$ as

$$\overline{y} = \frac{\left(\sum_{i=1}^{n} \delta_i y_i\right)}{\left(\sum_{i=1}^{n} \delta_i\right)} \quad \text{and} \tag{2.2.91}$$

$$\tilde{y} = \left(\prod_{i=1}^{n} y_i^{\delta_i}\right)^{1/\left(\sum_{i=1}^{n} \delta_i\right)} . \tag{2.2.92}$$

These correspond to the analogous definitions of the arithmetic and geometric means in the complete sample case. We use $\overline{y}$ and $\tilde{y}$ to write the log-likelihood function as

$$\begin{aligned} l(\alpha, \lambda) &= -\sum_{i=1}^{n} \delta_i \alpha \log(\lambda) - \sum_{i=1}^{n} \delta_i \log(\Gamma(\alpha)) + (\alpha - 1) \sum_{i=1}^{n} \delta_i \log(\tilde{y}) \\ &- \sum_{i=1}^{n} \delta_i \frac{\overline{y}}{\lambda} + \sum_{i=1}^{n} (1 - \delta_i) \log(1 - \gamma(y_i, \alpha, \lambda)) . \end{aligned} \tag{2.2.93}$$

The incomplete gamma function used in (2.2.93) is defined in (2.2.10). The partial derivatives of (2.2.93) are

$$\begin{aligned} \frac{\partial}{\partial \alpha} l(\alpha, \lambda) &= -\sum_{i=1}^{n} \delta_i \log(\lambda) - \sum_{i=1}^{n} \delta_i \psi(\alpha) + \sum_{i=1}^{n} \delta_i \log(\tilde{y}) \\ &+ \sum_{i=1}^{n} \frac{(1 - \delta_i) \frac{\partial}{\partial \alpha} [1 - \gamma(y_i, \alpha, \lambda)]}{[1 - \gamma(y_i, \alpha, \lambda)]} \quad \text{and} \end{aligned} \tag{2.2.94}$$

$$\frac{\partial}{\partial \lambda} l(\alpha, \lambda) = -\sum_{i=1}^{n} \delta_i \frac{\alpha}{\lambda} + \sum_{i=1}^{n} \delta_i \frac{\overline{y}}{\lambda^2} + \sum_{i=1}^{n} \frac{(1 - \delta_i) \frac{\partial}{\partial \lambda} [1 - \gamma(y_i, \alpha, \lambda)]}{[1 - \gamma(y_i, \alpha, \lambda)]} \tag{2.2.95}$$

where $\psi(\alpha) = \dfrac{\dfrac{\mathrm{d}}{\mathrm{d}\alpha}\Gamma(\alpha)}{\Gamma(\alpha)}$ is the digamma function. Properties of the digamma function can be found in Mathai (1993)). Equating these partial derivatives to zero and solving for α and λ gives the maximum likelihood estimators for the gamma model. The equations (2.2.94) and

(2.2.95) involve terms which can be written in terms of incomplete gamma integrals. One can show that with a simple transformation of variables, the incomplete gamma function $\gamma(y_i, \alpha, \lambda)$ can be rewritten as

$$\gamma(y_i, \alpha, \lambda) = \int_0^{\frac{y_i}{\lambda}} \frac{u^{\alpha-1}e^{-u}}{\Gamma(\alpha)} du \ . \tag{2.2.96}$$

Thus

$$\begin{aligned} \frac{\partial}{\partial \alpha}[1 - \gamma(y_i, \alpha, \lambda)] &= -\frac{1}{\Gamma(\alpha)} \int_0^{\frac{y_i}{\lambda}} u^{\alpha-1} \log(u) \exp(-u) du \\ &+ (\gamma(y_i, \alpha, \lambda))\, \psi(\alpha) \qquad \text{and} \end{aligned} \tag{2.2.97}$$

$$\frac{\partial}{\partial \lambda}[1 - \gamma(y_i, \alpha, \lambda)] = \frac{1}{\Gamma(\alpha)\lambda} \left(\frac{y_i}{\lambda}\right)^{\alpha} \exp\left(-\frac{y_i}{\lambda}\right) . \tag{2.2.98}$$

Note that to find the maximum likelihood estimators for the parameters using the Newton-Raphson method, the second partial derivatives of the log-likelihood are required. This can be accomplished with the aid of advanced mathematical software. Lawless (1982) states that a good approach to solving the maximum likelihood estimators is to solve (2.2.95) for fixed α and then narrow down to the estimator for λ using graphical or numerical methods.

2.4.2 Non-Parametric Estimation

Methods of estimation which are said to be non-parametric or distribution-free do not require any assumption on the underlying failure time model. Non-parametric estimates of the survival and hazard functions are often useful to summarize and explain failure time data.

Full Sample

Let us assume that we have a complete sample of failure time data, $x_1, x_2, \cdots, x_n$. Here, the failure times for all subjects have been observed. In such a situation, one uses standard

elementary statistical methods to describe and analyze the data. The empirical survival function is as follows:

$$S_n(x) = \frac{(\text{Number of failures} \geq x)}{n} \; ; \quad x \geq 0 \, . \tag{2.2.99}$$

$S_n(x)$ is a step function and a non-parametric estimator of the survival function $S(x)$ in the sense that no assumption about the underlying model is necessary. When dealing with censored observations, some modification to (2.2.99) is necessary. A natural estimate of the integrated hazard when considering the form found in (1.1.5) is

$$H_n(x) = -\log(S_n(x)); \quad x \geq 0 \, . \tag{2.2.100}$$

Using sample moments, one can estimate the mean value and variance by the sample mean and sample variance. These are

$$\hat{\mu} = \sum_{i=1}^{n} \frac{x_i}{n} \quad \text{and} \tag{2.2.101}$$

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(x_i - \hat{\mu})^2}{n-1} \, . \tag{2.2.102}$$

Confidence intervals on $S_n(x)$ can be computed by using results from the binomial distribution. The standard error of $S_n(x)$ is

$$s.e.\,(S_n(x)) = \left(\frac{S_n(x)\,(1 - S_n(x))}{n} \right)^{\frac{1}{2}} \tag{2.2.103}$$

so that a $100(1-\alpha)\%$ confidence interval for $S_n(x)$ is

$$S_n(x) \pm z_{\alpha/2} s.e.\,(S_n(x)) \tag{2.2.104}$$

where $z_{\alpha/2}$ is the upper $100(1-\alpha/2)$ percentage point of a standard normal distribution.

Censoring

The non-parametric maximum likelihood estimator of the survival function $S(x)$ in the presence of censored data is called the product-limit estimator or the Kaplan-Meier estimator

(see Kaplan and Meier (1958)).

To begin, we examine the likelihood of the failure time data $((y_1, \delta_1), \cdots, (y_n, \delta_n))$. First, we assume that there are m failures at the times $a_1, a_2, \cdots, a_m$. By convention, we let $a_0 = 0$ and $a_{m+1} = \infty$. We also suppose that the failure time X is concentrated on these failure times and we allow for more than one individual to fail or be censored at a given time. Now, let d_j be the number of subjects who fail (but are not censored) at time a_j and r_j the number of subjects at risk just before a_j, which is the number of subjects still alive and uncensored just before a_j.

For the discrete case, the survival function given by (1.1.13) is

$$S(x) = \prod_{a_i < x} (1 - h(a_i)) \tag{2.2.105}$$

and so a non-parametric estimator of $S(x)$ is

$$\widehat{S}(x) = \prod_{a_i < x} \left(1 - \widehat{h}(a_i)\right) \tag{2.2.106}$$

where $\widehat{h}(a_i)$ is the maximum likelihood estimator of $h(a_i)$, for $i = 1, 2, \cdots, m$. If a subject fails at a_j then its contribution to the likelihood is $f(a_j)$; if a subject is censored at a time c then its contribution is

$$P(X > c) = S(c+) = \sum_{a_j > c} f(a_j) \ . \tag{2.2.107}$$

In terms of (1.1.13), we have

$$S(c+) = \prod_{a_j \leq c} (1 - h(a_j)) \quad \text{and} \tag{2.2.108}$$

$$f(a_j) = h(a_j) S(a_i) \ . \tag{2.2.109}$$

Cox and Oakes (1984) give the likelihood for the data as

$$L = \prod_{j=1}^{m} [h(a_j)]^{d_j} \, [(1 - h(a_j)]^{r_j - d_j} \ . \tag{2.2.110}$$

We now use the logarithm of the likelihood to find estimates for $h(a_j)$, $j = 1, 2, \cdots, m$.

$$l = \log(L) = \sum_{j=1}^{m} d_j \log(h(a_j)) + (r_j - d_j)\log(1 - h(a_j)) \tag{2.2.111}$$

The maximum likelihood estimator for $h(a_j)$ is then the solution to

$$\begin{aligned} 0 &= \frac{\partial}{\partial h(a_j)} \log(L) \\ &= \sum_{i=1}^{m} d_j \frac{\partial}{\partial h(a_j)} \log(h(a_i)) + (r_j - d_j)\frac{\partial}{\partial h(a_j)} \log(1 - h(a_i)) \\ &= \frac{d_j}{h(a_j)} - \frac{r_j - d_j}{1 - h(a_j)} . \end{aligned} \tag{2.2.112}$$

$$\text{Hence} \quad \widehat{h}(a_j) = \frac{d_j}{r_j} . \tag{2.2.113}$$

Substituting this result in (2.2.106), one obtains the product-limit estimator

$$\widehat{S}(x) = \prod_{a_j < x} \left(1 - \frac{d_j}{r_j}\right) . \tag{2.2.114}$$

If there is no censoring, the product-limit estimator is the same as the empirical survival function given by (2.2.99). If the failure times for all subjects are observed then $r_i = d_i + r_{i+1}$, $i = 1, 2, ..., m = n$. On expanding the product in (2.2.114), we see that

$$\begin{aligned} \widehat{S}(x) &= \prod_{a_j < x} \left(1 - \frac{d_j}{r_j}\right) \\ &= \frac{r_2}{r_1}\frac{r_3}{r_2} \cdots \frac{r_{s+1}}{r_s} \\ &= \frac{r_{s+1}}{r_1} \\ &= \frac{r_{s+1}}{n} \end{aligned} \tag{2.2.115}$$

where $a_s < x \leq a_{s+1}$. From the recursive relation $r_j = d_j + r_{j+1}$, we see that $r_1 = n = r_{s+1} + d_1 + d_2 + \cdots + d_s$. Hence, $n - (\#failures < x) = n - d_1 + d_2 + \cdots + d_s = r_{s+1}$, so that $\widehat{S}(x) = S_n(x)$.

Also, if the largest observation in the failure time data is y^* and y^* is censored then $\widehat{S}(x)$ is undefined for $x > y^*$. If y^* is an uncensored time, then $\widehat{S}(x) = 0$ for $x > y^*$.

Confidence Intervals for $\widehat{S}(x)$

It can be shown that under certain conditions, the product-limit estimator, $\widehat{S}(x)$, converges to the survival function, $S(x)$, as $n \to \infty$, where n is the number of observations. Also, we conclude that

$$\frac{\widehat{S}(x) - S(x)}{\sqrt{Var\,(S(x))}} \tag{2.2.116}$$

converges to a standard normal random variable in distribution. Using Greenwood's formula, the variance of $\widehat{S}(x)$ is estimated to be

$$\widehat{Var}\,(S(x)) = \left[\widehat{S}(x)\right]^2 \sum_{a_j < x} \frac{d_j}{r_j(r_j - d_j)} \,. \tag{2.2.117}$$

For a derivation of Greenwood's formula, refer to appendix D. This is particularly useful because it allows us to make approximate $100(1-\alpha)\%$ confidence intervals for $S(x)$ using

$$\widehat{S}(x) \pm z_{\alpha/2}\sqrt{\widehat{Var}\,(S(x))}\,, \tag{2.2.118}$$

where $z_{\alpha/2}$ is the upper $100(1-\alpha/2)$ percentage point of a standard normal distribution. The symmetric confidence interval given by (2.2.118) produce pointwise confidence intervals for each fixed failure time x.

Chapter 3

Additive Gamma Type Inputs

3.1 Introduction

In this chapter, we define inputs to be the survival times of the individual components that belong to a complex model. For example, if we have a mixture model (see section 2.3.1), then the lifetimes corresponding to each sub-population are considered to be inputs for the whole mixtures model. A gamma type input is one that follows the generalized gamma model discussed in section 2.2.4. An exponential input is one that comes from an exponential model and a gamma input is from a gamma model. These inputs are both of gamma type. As stated before, the generalized gamma model is well suited for explaining life data, however because of its numerical complexity, it is seldom used. These gamma type inputs offer wider variety and more flexibility to researchers and analysts compared to exponential type inputs, say. Thus, we now turn our discussion towards linear functions of gamma type inputs.

There exist a multitude of real-life situations in which one might consider the distributional properties of a linear function of gamma type inputs, $X_1, X_2, \ldots, X_k$. In survival

analysis, one is interested in the survival function or hazard function over the total waiting time of a process. In the case of a rapidly evolving disease such as AIDS or cancer, the individual inputs $X_1, X_2, \ldots, X_k$ may represent the waiting times of each progressive stage of the disease. In AIDS for instance, X_1 might be the length of time in the first stage where the test results show that a patient is HIV negative after receiving a contaminated blood transfusion; X_2 is the waiting time in the second stage where the patient is HIV positive; X_3 is when the patient shows pre-AIDS symptoms; X_4 is the duration of full-blown AIDS, and so forth.

On the other hand, our problem may involve k vehicles waiting to be serviced at a garage, where each automobile is serviced one at a time. The first vehicle may need an oil change, the second new tires, the third new brakes and so on. Thus the k individual inputs $X_1, X_2, \ldots, X_k$ are independently distributed waiting times and they need not be identically distributed.

Analogously, we can have a spare parts problem as discussed in section 2.3.4. An important component of the system may fail occasionally and thus it is replaced with an identical spare part; the spares are independent of one another, however, they are identically distributed.

Another popular scenario might be the study of a system failure in the case of a parallel port with k components. Here, X_i denotes the failure time for the i^{th} stage, $i = 1, \ldots, k$. In each of these examples, the quantity of interest is $X = X_1 + X_2 + \ldots + X_k$. For simplicity of calculations, these inputs are sometimes assumed to be exponentially distributed with the same or different means to obtain tractability. In reality however, it makes sense to consider these as being independent gamma type inputs with general parameters.

Researchers are now considering linear functions of gamma type inputs in their studies and are using traditional approximation techniques to obtain distributional properties of their functions. For example, Huzurbazar and Huzurbazar (1999) examine the survival function associated with AIDS with the help of data from the San Francisco Men's Health Study. A two stage situation is considered in their paper, $X = X_1 + X_2$, where X_1 is from an exponential model whereas X_2 is from a gamma model. The inputs X_1 and X_2 are assumed to be independent. Saddlepoint approximations are used to attempt the problem. In this fascinating paper, it is shown that the saddlepoint approximation is simple to apply and produces results close to the simulations arising from the exact distributions.

The saddlepoint approximation technique, widely used in physics, was introduced in statistics by Daniels (1954) for approximating a probability density function (pdf). Lugannani and Rice (1980) used the same technique to approximate cumulative probability density functions (cdf). Application of this technique to various problems may be seen, for example, from Barndorff-Nielsen and Cox (1979), Cox and Oakes (1984), Blæsild and Jensen (1985), Daniels (1987), Reid (1988), Butler, Huzurbazar, and Booth (1992a,b), Jensen (1995) and Butler and Huzurbazar (1997).

This chapter focuses mainly on $X = a_1X_1 + a_2X_2 + \ldots + a_kX_k$, where the inputs X_j are mutually independent and $a_j > 0$ for $j = 1, \ldots, k$. Since weighted gamma type inputs with positive weights are again gamma type inputs, then without loss of generality, one needs to consider only a sum of independent gamma type inputs, such as $X = X_1 + X_2 + \ldots + X_k$. We present several techniques for computing the exact pdf of X, $f(x)$, and exact cdf of X, $F(x)$, so that the exact forms for the survival function, $S(x)$, and hazard function, $h(x)$, can be computed. In order to discuss the properties of $S(x)$ and $h(x)$, it is necessary to study

exact forms of $f(x)$ and $F(x)$. We then compare the exact forms to the usual approximation results. Our aim is to compute $S(x)$ and $h(x)$ for general cases and then apply the results to specific problems. We begin with gamma inputs.

3.2 Gamma Inputs: Exact Forms

The exact density of X when the inputs are gamma can be written in terms of confluent hypergeometric functions. Before giving an explicit form for $f(x)$, we will consider a special case of the moment generating function of the gamma model. When the shape parameters of the model are integers, the pdf of X, $f(x)$, has a finite expression and can be handled easily.

The moment generating function of the gamma input X_j with parameters (α, λ) (see section 2.2.2) is denoted by

$$M_{X_j}(t) = (1 - \lambda_j t)^{-\alpha_j} \tag{3.3.1}$$

and since the components of $X = X_1 + X_2 + \ldots + X_k$ are assumed to be independently distributed, the moment generating function of X is simply given by

$$M_X(t) = \prod_{j=1}^{k} (1 - \lambda_j t)^{-\alpha_j} \ . \tag{3.3.2}$$

3.2.1 Integer Shape Parameters

Assume that the shape parameters of the gamma inputs are of the form $\alpha_j = m_j$, where $m_j = 1, 2, \ldots$ for $j = 1, 2, \ldots, k$. This means that all the α_j's are positive integers. In this case the moment generating function found in (3.3.2) can be written as an explicit finite sum

by using partial fraction techniques. That is,

$$\prod_{j=1}^{k}(1-\lambda_j t)^{-m_j} = \sum_{j=1}^{k}\sum_{i=1}^{m_j}\frac{a_{ji}}{(1-\lambda_j t)^i} \tag{3.3.3}$$

where the a_{ji}'s are free of t. Thus, inversion of the moment generating function given by (3.3.3) produces the pdf

$$f(x) = \sum_{j=1}^{k}\sum_{i=1}^{m_j} a_{ji}\left\{\frac{x^{i-1}e^{-x/\lambda_j}}{\lambda_j^i\,(i-1)!}\right\} \qquad \text{for } x>0 \tag{3.3.4}$$

and 0 elsewhere. The cdf $F(x)$ is thus available from (3.3.4) using term-by-term integration. The coefficients, a_{ji}, are given by the following:

$$\begin{aligned}
a_{ji} &= \left\{\prod_{n=1}^{k}(-\lambda_n)^{-m_n}\right\}(-\lambda_j)^i b_{ji}, \qquad (3.3.5)\\
b_{ji} &= \left\{\sum_{j_1=0}^{m_j-i-1}\binom{m_j-i-1}{j_1}A_j^{(m_j-i-1-j_1)}\sum_{j_2=0}^{j_1-1}\binom{j_1-1}{j_2}A_j^{(j_1-1-j_2)}\cdots\right\}\frac{\Delta_j}{(m_j-i)!},\\
A_j^{(s)} &= (-1)^{s+1}s!\sum_{\substack{r=1\\ r\neq j}}^{k} m_r\left(\frac{1}{\lambda_j}-\frac{1}{\lambda_r}\right)^{-(s+1)} \quad \text{and}\\
\Delta_j &= \prod_{\substack{r=1\\ r\neq j}}^{k}\left(\frac{1}{\lambda_j}-\frac{1}{\lambda_r}\right)^{-m_r} .
\end{aligned}$$

For a complete derivation of the coefficients a_{ji} refer to Mathai (1982).

Example 16. Consider a two stage process $X = X_1 + X_2$ where the inputs are assumed to be independent and the parameters are $(\alpha_1, \beta_1) = (2, 2)$ and $(\alpha_2, \beta_2) = (3, 2.5)$. The shape parameters are integers and we use the previous results to write the moment generating function of X as:

$$\begin{aligned}
M_X(t) &= (1-2t)^{-2}(1-2.5t)^{-3}\\
&= \frac{-960}{(1-2t)} + \frac{-64}{(1-2t)^2} + \frac{1200}{(1-2.5t)} + \frac{-200}{(1-2.5t)^2} + \frac{25}{(1-2.5t)^3},
\end{aligned}$$

Which gives the following pdf for X:

$$f(x) = -960\left(\frac{e^{-x/2}}{2}\right) - 64\left(\frac{xe^{-x/2}}{2^2}\right) + 1200\left(\frac{e^{-x/2.5}}{2.5}\right) \\ -200\left(\frac{xe^{-x/2.5}}{2.5^2}\right) + 25\left(\frac{x^2e^{-x/2.5}}{2.5^3\,2}\right) \quad .$$

The case considered in this section is not an isolated situation in statistics and other related fields. Problems of this type, usually arise in engineering and communication theory, where X is of the form $X = |X_1|^2 + |X_2|^2 + \ldots + |X_k|^2$, where $X_j = X_{j1} + iX_{j2}$, $i^2 = -1$. Here X_{j1} and X_{j2} are real with X_{j1}^2 and X_{j2}^2 independently distributed chisquares with the same degrees of freedom ν_j. So, if $|X_j|^2 \sim \chi^2_{2\nu_j}$ then $\alpha_j = 2\nu_j$, where $\nu_j = \frac{n_j}{2}$, $n_j = 1, 2, \ldots$ and thereby all α_j's are positive integers. Some problems of this type can be seen from Biyari and Lindsey (1993), and Divsalar, Simon, and Shahshahani (1990). Another example arises in testing of statistical hypotheses on the parameters of complex Gaussian distributions. The null distributions of the likelihood ratio test statistics, in many cases, can be written as that of a linear function of chisquares with even degrees of freedom, thereby making the α_j's in (3.3.2) positive integers.

3.2.2 General Shape Parameters

For general parameters, the density $f(x)$ arising from the moment generating function (3.3.2) can be written as a convex combination of gamma densities. We begin by stating the result. Without loss of generality, let $\lambda_1 < \lambda_2 < \ldots < \lambda_k$. The pdf of X thus looks like

$$f(x) = \begin{cases} C\sum_{r=0}^{\infty} \delta_r \dfrac{x^{\alpha+r-1}e^{-y/\lambda_1}}{\lambda_1^{\alpha+r}\Gamma(\alpha+r)} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases} \tag{3.3.6}$$

The normalizing constant C and the weights δ_r are given by

$$C = \prod_{j=1}^{k} \left(\frac{\lambda_1}{\lambda_j}\right)^{\alpha_j}$$

$$\delta_r = \sum_{r_2+\ldots+r_k=r} \frac{(\alpha_2)_{r_2}}{r_2!} \ldots \frac{(\alpha_k)_{r_k}}{r_k!} \left(1-\frac{\lambda_1}{\lambda_2}\right)^{r_2} \ldots \left(1-\frac{\lambda_1}{\lambda_k}\right)^{r_k},$$

where $\alpha = \alpha_1 + \alpha_2 + \ldots + \alpha_k$, $(r_2, \ldots, r_k)$ is a partitioning of r, $r = r_2 + \ldots + r_k$ and for example $(a)_m = a(a+1)\ldots(a+m-1)$, $(a)_0 = 1$, $a \neq 0$. Properties of $(a)_m$ are discussed in Mathai (1993). A recursive method of finding the weights δ_r is also discussed in the derivation below. The distribution function $F(x)$ is thus available from (3.3.6) using term-by-term integration. Justification for this can be found in Moschopoulos (1985). Each term of the integrated expression will involve the incomplete gamma function. Considering the power of computers at present time, these methods can easily be implemented using a statistical software package to obtain almost immediate results. We now discuss the derivation of this result.

Derivation 1

We follow the derivation given by Mathai (1982). Consider the moment generating function of X in (3.3.2). The j'th term of the moment generating function can be rewritten as

$$\begin{aligned}(1-\lambda_j t)^{-\alpha_j} &= \left(\frac{\lambda}{\lambda_j}\right)^{\alpha_j} (1-\lambda t)^{-\alpha_j} \left[1 - \frac{1-\lambda/\lambda_j}{1-\lambda t}\right]^{-\alpha_j} \\ &= \gamma_j^{\alpha_j} (1-\lambda t)^{-\alpha_j} \left[1 - \frac{1-\gamma_j}{1-\lambda t}\right]^{-\alpha_j} \qquad (3.3.7)\end{aligned}$$

where λ is arbitrary such that $|\lambda t| < 1$, $\gamma_j = \frac{\lambda}{\lambda_j}$. Application of (3.3.7) to the mgf of X in (3.3.2) gives us

$$\prod_{j=1}^{k} (1-\lambda_j t)^{-\alpha_j} = \left[\prod_{j=1}^{k} \gamma_j^{\alpha_j}\right] (1-\lambda t)^{-(\alpha_1+\alpha_2+\ldots+\alpha_k)} \left[\prod_{j=1}^{k} \left[1 - \frac{1-\gamma_j}{1-\lambda t}\right]^{-\alpha_j}\right] \qquad (3.3.8)$$

or

$$M_X(t) = C(1-\lambda t)^{-\alpha}\prod_{j=1}^{k}\left[1-\frac{1-\gamma_j}{1-\lambda t}\right]^{-\alpha_j} \tag{3.3.9}$$

where $C = \prod_{j=1}^{k}\gamma_j^{\alpha_j}$ and $\alpha = \alpha_1+\alpha_2+\ldots+\alpha_k$. From here, there are many ways to proceed to obtain the desired density $f(x)$.

One method involves expanding each factor in the product of (3.3.9) through binomial expansions. To do this, we require that $|\frac{1-\gamma_j}{1-\lambda t}| < 1$. Without loss of generality, let $\lambda_1 < \lambda_2 < \ldots < \lambda_k$; for convenience take $\lambda = \lambda_1$. As seen in (3.3.10), this choice for λ eliminates some of the complexity involved in the computable form of $f(x)$ by removing a summation term. The binomial expansions for each factor gives the following for the mgf of X.

$$\begin{aligned} M_X(t) &= C(1-\lambda_1 t)^{-\alpha}\sum_{r_2=0}^{\infty}\cdots\sum_{r_k=0}^{\infty}\frac{(\alpha_2)_{r_2}}{r_2!}\cdots\frac{(\alpha_k)_{r_k}}{r_k!}\left(\frac{1-\gamma_2}{1-\lambda_1 t}\right)^{r_2}\cdots\left(\frac{1-\gamma_k}{1-\lambda_1 t}\right)^{r_k} \\ &= C(1-\lambda_1 t)^{-\alpha}\sum_{r=0}^{\infty}\delta_r(1-\lambda_1 t)^{-r} \end{aligned} \tag{3.3.10}$$

where $(r_2,\ldots,r_k)$ is a partitioning of r, $r = r_2+\ldots+r_k$,

$$\delta_r = \sum_{r_2+\ldots+r_k=r}\frac{(\alpha_2)_{r_2}}{r_2!}\cdots\frac{(\alpha_k)_{r_k}}{r_k!}(1-\gamma_2)^{r_2}\cdots(1-\gamma_k)^{r_k} \tag{3.3.11}$$

and, as seen before, $(a)_m = a(a+1)\ldots(a+m-1)$, $(a)_0 = 1$, $a \neq 0$. The density of $X = X_1+X_2+\ldots+X_k$ is then given by inverting the mgf of X in (3.3.10). This gives

$$f(x) = \begin{cases} C\sum_{r=0}^{\infty}\delta_r\dfrac{x^{\alpha+r-1}e^{-x/\lambda_1}}{\lambda_1^{\alpha+r}\Gamma(\alpha+r)} & \text{for } x>0 \\ 0 & \text{elsewhere.} \end{cases} \tag{3.3.12}$$

It is easy to show that the weights δ_r are positive for all $r = 0,1,2,\cdots$ and from the pdf in (3.3.12) it can be observed that $C\sum_{r=0}^{\infty}\delta_r = 1$. Thus, the pdf $f(x)$ in (3.3.12) is a convex combination of gamma densities.

Derivation 2

Alternatively we can consider the recursive method found in Moschopoulos (1985). Taking logarithms in (3.3.9), one obtains

$$\log M_X(t) = \log C(1-\lambda_1 t)^{-\alpha} + \sum_{i=1}^{\infty} \sigma_i (1-\lambda_1 t)^{-i} \tag{3.3.13}$$

where $\sigma_i = \sum_{j=1}^{k} \alpha_j \frac{(1-\gamma_j)^i}{i}$. Note that the expansion of the logarithm in (3.3.13) is valid for $|\frac{1-\gamma_j}{1-\lambda_1 t}| < 1$. Hence, taking exponentials in (3.3.13) and expanding, one gets

$$M_X(t) = C(1-\lambda_1 t)^{-\alpha} \sum_{i=0}^{\infty} \rho_i (1-\lambda_1 t)^{-i} \quad \text{where} \tag{3.3.14}$$

$$\rho_{i+1} = \frac{1}{i+1} \sum_{n=1}^{i+1} n\sigma_n \rho_{i+1-n} \quad \text{for } i > 1 \quad , \rho_0 = 1 \ . \tag{3.3.15}$$

The density of $X = X_1 + X_2 + \ldots + X_k$ from (3.3.14) can be expressed as

$$f(x) = \begin{cases} C \sum_{s=0}^{\infty} \rho_s \frac{x^{\alpha+s-1} e^{-x/\lambda_1}}{\lambda_1^{\alpha+s} \Gamma(\alpha+s)} & \text{for } x > 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{3.3.16}$$

Note that the form of $f(x)$ in (3.3.16) is the same as the one found in (3.3.12) and so by comparison, $\delta_r = \rho_s$ for $s = r = 0, 1, 2, \cdots$. Hence (3.3.11) and (3.3.15) give us two ways to compute the weights used in calculating the pdf $f(x)$.

3.3 Gamma Inputs: Gamma Approximation

In the previous sub-section we saw that the density of X is given as a convex combination of gamma densities. We now use the expected value $E(X)$ and variance $Var(X)$ to approximate the density of X, $f(x)$, by a gamma density.

Since the inputs X_j are independent with parameters (α_j, β_j) for $j = 1, 2, \cdots, k$, we know

that

$$E(X) = \sum_{j=1}^{k} \alpha_j \beta_j \tag{3.3.17}$$

and

$$Var(X) = \sum_{j=1}^{k} \alpha_j \beta_j^2 \tag{3.3.18}$$

We approximate the pdf of X by using a gamma density with parameters $(\tilde{\alpha}, \tilde{\beta})$:

$$\tilde{f}(x) = \begin{cases} \frac{1}{\Gamma(\tilde{\alpha})} \tilde{\beta}^{-\tilde{\alpha}} x^{\tilde{\alpha}-1} \exp\left(-x/\tilde{\beta}\right) \\ 0 \qquad\qquad\qquad\qquad \text{elsewhere.} \end{cases} \tag{3.3.19}$$

The parameters $\tilde{\alpha}$ and $\tilde{\beta}$ are obtained by solving the following system of equations:

$$E(X) = \tilde{\alpha}\tilde{\beta} = \sum_{j=1}^{k} \alpha_j \beta_j \quad Var(X) = \tilde{\alpha}\tilde{\beta}^2 = \sum_{j=1}^{k} \alpha_j \beta_j^2$$

We obtain

$$\tilde{\alpha} = \frac{\left(\sum_{j=1}^{k} \alpha_j \beta_j\right)^2}{\left(\sum_{j=1}^{k} \alpha_j \beta_j{}^2\right)} \quad \text{and} \quad \tilde{\beta} = \frac{\left(\sum_{j=1}^{k} \alpha_j \beta_j^2\right)}{\left(\sum_{j=1}^{k} \alpha_j \beta_j\right)} \tag{3.3.20}$$

3.4 Gamma Type Inputs: Exact Forms

We now consider exact density of X when the inputs X_j are gamma type inputs. These include exponential and gamma inputs arising from the exponential and gamma models, respectively. They also include Weibull inputs.

We begin by considering two independent gamma type inputs X_1 and X_2 with parameters $(\alpha_1, \beta_1, \delta_1)$ and $(\alpha_2, \beta_2, \delta_2)$ (see section 2.2.4) to derive the density for $X = X_1 + X_2$. We then use the result to generalize the solution for $X = X_1 + X_2 + \cdots + X_k$. Let the pdfs of X_1 and X_2 be denoted by $C_1 f_1(x)$ and $C_2 f_2(x)$ respectively, where C_1 and C_2 are normalizing

constants. Using the assumption of independence between X_1 and X_2 and making a change of variables, the pdf of X is given by

$$\begin{aligned} f_X(x) &= C_1C_2\int_0^x t^{\alpha_1-1}(x-t)^{\alpha_2-1}\exp\left\{-\delta_1 t^{\beta_1}-\delta_2(x-t)^{\beta_2}\right\}\mathrm{d}t \\ &= C_1C_2x^{\alpha_1+\alpha_2-1}\times \\ &\quad \int_0^1 s^{\alpha_1-1}(1-s)^{\alpha_2-1}\exp\left\{-\delta_1 x^{\beta_1}s^{\beta_1}-\delta_2 x^{\beta_2}(1-s)^{\beta_2}\right\}\mathrm{d}s \end{aligned} \tag{3.3.21}$$

for $x > 0$ and zero elsewhere. Upon making a change of variables, expanding the exponentials and integrating, we obtain

$$\begin{aligned} f_X(x) &= Cx^{\alpha_1+\alpha_2-1}\times \\ &\quad \sum_{r_1=0}^{\infty}\sum_{r_2=0}^{\infty}\frac{(-\delta_1 x^{\beta_1})^{r_1}}{r_1!}\frac{(-\delta_2 x^{\beta_2})^{r_2}}{r_2!}\frac{\Gamma(\alpha_1+\beta_1 r_1)\,\Gamma(\alpha_2+\beta_2 r_2)}{\Gamma(\alpha_1+\alpha_2+\beta_1 r_1+\beta_2 r_2)} \end{aligned} \tag{3.3.22}$$

where

$$C=\frac{\beta_1\beta_2(\delta_1)^{\alpha_1/\beta_1}(\delta_2)^{\alpha_2/\beta_2}}{\Gamma\left(\frac{\alpha_1}{\beta_1}\right)\Gamma\left(\frac{\alpha_2}{\beta_2}\right)}\,.$$

We shall see in section 3.8 that the representation of $f_X(x)$ in (3.3.22) is not practical for large x. In such situations, we can still make use of numerical integration techniques to approximate the integral in (3.3.21) and easily obtain accurate values for $f_X(x)$.

Next, we generalize the result for $X = X_1 + X_2 + \cdots + X_k$, where each gamma type input has parameters $(\alpha_i, \beta_i, \delta_i)$, by following the same steps for (3.3.22). Since the inputs are assumed to be independent, we can form the joint density of $(X_1, X_2, \cdots, X_k)$. Upon making two changes of variables, we obtain the following for the pdf of X

$$\begin{aligned} f_X(x) &= Cx^{\alpha-1}\int_0^1\int_0^1\cdots\int_0^1\left\{z_{k-1}^{(\alpha-\alpha_k-1)}(1-z_{k-1})^{\alpha_k-1}\right. \\ &\quad z_{k-2}^{(\alpha-\alpha_k-\alpha_{k-1}-1)}(1-z_{k-2})^{\alpha_{k-1}-1}\cdots z_1^{\alpha_1-1}(1-z_1)^{\alpha_2-1} \\ &\quad \exp\Big(-\delta_1(z_1\cdots z_{k-1}x)^{\beta_1}-\delta_2(z_2\cdots z_{k-1}x)^{\beta_2}(1-z_1)^{\beta_2}-\cdots \\ &\quad \left.-\delta_k x^{\beta_k}(1-z_{k-1})^{\beta_k}\Big)\right\}\mathrm{d}z_{k-1}\cdots\mathrm{d}z_1 \quad \text{for} \quad x>0 \end{aligned} \tag{3.3.23}$$

and 0 elsewhere, where $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_k$ and C is the normalizing constant. Expanding the exponentials in (3.3.23) and integrating produces a representation for $f_X(x)$ in terms of infinite series:

$$f_X(x) = Cx^{\alpha-1} \sum_{r=0}^{\infty} (-1)^r \sum_{r_1+r_2+\cdots+r_k=r} M(x; r_1, r_2, \cdots, r_k) , \tag{3.3.24}$$

$$M(x; r_1, r_2, \cdots, r_k) = \frac{x^{\sum_{i=1}^k \beta_i r_i}}{\Gamma\left(\alpha + \sum_{i=1}^k \beta_i r_i\right)} \prod_{i=1}^k \frac{\Gamma(\alpha_i + \beta_i r_i)\delta_i^{r_i}}{r_i!} \quad \text{and} \tag{3.3.25}$$

$$C = \prod_{i=1}^k \frac{\beta_i \delta_i^{\alpha_i/\beta_i}}{\Gamma\left(\frac{\alpha_i}{\beta_i}\right)} , \tag{3.3.26}$$

where $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_k$, $(r_1, r_2, \cdots, r_k)$ is a partition of r such that $r_1 + r_2 + \cdots + r_k = r$. Term-by-term integration of (3.3.24) gives the cdf of X, which is

$$F_X(x) = Cx^{\alpha} \sum_{r=0}^{\infty} (-1)^r \sum_{r_1+r_2+\cdots+r_k=r} M'(x; r_1, r_2, \cdots, r_k) \tag{3.3.27}$$

where

$$M'(x; r_1, r_2, \cdots, r_k) = \frac{x^{\sum_{i=1}^k \beta_i r_i}}{\Gamma\left(\alpha + \sum_{i=1}^k \beta_i r_i + 1\right)} \prod_{i=1}^k \frac{\Gamma(\alpha_i + \beta_i r_i)\delta_i^{r_i}}{r_i!} . \tag{3.3.28}$$

These forms for the pdf and cdf of X have already been derived by Stacy (1962) where, by similar methods as those seen above, the moment generating function of each input X_j is also given. This is

$$M_{X_j}(t) = \sum_{i=0}^{\infty} \frac{\left(\frac{t}{\delta_j^{1/\beta_j}}\right)^i}{i!} \frac{\Gamma\left(\frac{\alpha_j + i}{\beta_j}\right)}{\Gamma\left(\frac{\alpha_j}{\beta_j}\right)} \tag{3.3.29}$$

for $j = 1, 2, \cdots, k$. The mgf in (3.3.29) is defined for $t < c$, $0 \leq c \leq \infty$. If $\beta_j = 1$ then $c = \delta_j$. However, if $\beta_j < 1$, then $t \leq c = 0$ and if $\beta_j > 1$ then $c = \infty$. Given that the inputs are independent, the mgf of X is

$$M_X(t) = \prod_{j=1}^k M_{X_j}(t) . \tag{3.3.30}$$

For certain values of β_j, (3.3.29) is numerically unstable. Direct numerical integration is required to evaluate the mgf of X in such cases. Similarly with the case where $X = X_1 + X_2$, the form of $f_X(x)$ in (3.3.24) is often impractical for large x. We will show in section 3.8 that using a numerical estimate for (3.3.23) gives more precise results for $f_X(x)$.

3.5 Saddlepoint Approximations

Originally used by physicists and introduced to statistics by Daniels (1954), this popular method can be utilized to solve a wide variety of problems. It is applied successfully by Huzurbazar (Aparna), her associates and research collaborators in a wide variety of problems in survival analysis. Daniels (1954, 1987) exploit the method to approximate densities, Lugannani and Rice (1980) used it to approximate distribution functions. Conditional distributions are examined by Reid (1988) and Jensen (1995). These approximations can be given in terms of a standard normal base or with other distributions such as gamma and inverse Gaussian as the base. Distributions which readily admit saddlepoint approximations include exponential, gamma, Raleigh, Weibull and inverse Gaussian, which are, as already discussed in chapter 2, commonly used to model survival data. In order to apply saddlepoint approximations, one needs to consider the moment generating function (mgf) of the random variable. If this random variable is denoted by X, its mgf by $M_X(t) = M(t)$ and its cumulant generating function (cgf) by $K_X(t) = K(t)$, then $K(t) = \ln M(t)$. Daniels (1954) explains that the largest interval (c_1, c_2) in which the mgf converges needs to be known. In other words, we need to find positive numbers c_1 and c_2 where $-c_1 < t < c_2$, $0 \leq c_1 \leq \infty$, $0 \leq c_2 \leq \infty$ but $c_1 + c_2 > 0$, so that c_1 or c_2 can be zero, but not both.

3.5.1 Gamma Inputs

Assume that we have $X = X_1 + \cdots + X_k$, where the inputs X_j, $j = 1, 2, \cdots, k$ are independent gamma inputs with parameters (α, λ). Then from (3.3.2), we have

$$M(t) = \prod_{j=1}^{k} (1 - \lambda_j t)^{-\alpha_j} \quad , 1 - \lambda_j t > 0, \; j = 1, \cdots, k \quad \text{and} \tag{3.3.31}$$

$$K(t) = \sum_{j=1}^{k} -\alpha_j \ln (1 - \lambda_j t) \quad . \tag{3.3.32}$$

The saddlepoint approximation also requires computation of the first three derivatives of $K(t)$ with respect to t and the solution of the equation

$$K'(t) = x \tag{3.3.33}$$

where $K'(t)$ denotes the first derivative of $K(t)$ with respect to t. For the cgf in (3.3.32), the above equation becomes

$$\frac{\alpha_1 \lambda_1}{1 - \lambda_1 t} + \frac{\alpha_2 \lambda_2}{1 - \lambda_2 t} + \cdots + \frac{\alpha_k \lambda_k}{1 - \lambda_k t} = x \quad . \tag{3.3.34}$$

It is evident that if the λ_j's are different then it is somewhat difficult to solve (3.3.34), even for moderately large values of k. For instance, when $k = 2$, $\lambda_1 \neq \lambda_2$, one must solve a quadratic equation. Once (3.3.34) is solved (usually by numerical methods), let the solution be denoted by $\hat{t}$. The saddlepoint approximations of the pdf $f(x)$ and cdf $F(x)$ of X, denoted by $\hat{f}(x)$ and $\hat{F}(x)$ respectively, are given by the following

$$\hat{f}(x) = C \frac{e^{K(\hat{t}) - x\hat{t}}}{[2\pi K''(\hat{t})]^{1/2}} \quad \text{and} \tag{3.3.35}$$

$$\hat{F}(x) = \Phi(r) + \phi(r) \left\{ \frac{1}{r} - \frac{1}{s} \right\} \tag{3.3.36}$$

where $s = \hat{t}[K''(\hat{t})]^{1/2}$, $r = sgn(s) \left\{ 2(\hat{t}\, x - K(\hat{t})) \right\}^{1/2}$ and C is the normalizing constant, determined numerically, $\phi(\cdot)$ is the standard normal density, and $\Phi(\cdot)$ is the standard normal

distribution function. Formula (3.3.36) is valid for all x except at the point $x = E(x)$, where $r = 0$ and $s = 0$. At $x = E(x)$ the formula in (3.3.36) is replaced by

$$\widehat{F}(x) = \frac{1}{2} + \frac{K'''(\hat{t})}{6(2\pi)^{1/2}[K''(\hat{t})]^{3/2}} \quad . \tag{3.3.37}$$

Another approximation for the distribution function is obtained by numerical integration of (3.3.35). In section 3.8, we demonstrate this procedure and compare results with those using the methods introduced in section 3.2. For a more thorough introduction to saddlepoint approximations, refer to Daniels (1954).

3.5.2 Gamma Type Inputs

Applying the saddlepoint approximation to the case when the inputs X_j, $j = 1, 2, \cdots, k$ are gamma type inputs is somewhat more difficult. To do this, we use the formulas for the saddlepoint approximation discussed in the previous section, but with the moment generating function of X replaced by (3.3.30). The cumulant generating function is now

$$K(t) = \sum_{j=1}^{k} K_{X_j}(t) \tag{3.3.38}$$

where $K_{X_j}(t) = \log\left(M_{X_j}(t)\right)$ and $M_{X_j}(t)$, the mgf of the j'th input, is given by (3.3.29). The derivatives of $K(t)$ are found by using the following technique. The first derivative of the cgf of X is

$$\frac{\mathrm{d}}{\mathrm{dt}}K(t) = \sum_{j=1}^{k} \frac{\mathrm{d}}{\mathrm{dt}}K_{X_j}(t) \tag{3.3.39}$$

where

$$\frac{\mathrm{d}}{\mathrm{dt}}K_{X_j}(t) = \frac{\frac{\mathrm{d}}{dt}M_{X_j}(t)}{M_{X_j}(t)} \, . \tag{3.3.40}$$

To find the derivative of the mgf of X_j, we use the fact that

$$M_{X_j}(t) = \frac{\beta_j \, \delta_j^{\alpha_j/\beta_j}}{\Gamma\left(\frac{\alpha_j}{\beta_j}\right)} \int_0^\infty x_j^{\alpha_j - 1} \exp\left(-\delta_j x_j^{\beta_j} + t x_j\right) \mathrm{d}x_j$$

$$= \sum_{i=0}^{\infty} \frac{\left(\frac{t}{\delta_j^{1/\beta_j}}\right)^i}{i!} \frac{\Gamma\left(\frac{\alpha_j + i}{\beta_j}\right)}{\Gamma\left(\frac{\alpha_j}{\beta_j}\right)} . \tag{3.3.41}$$

Hence,

$$\frac{\mathrm{d}}{\mathrm{dt}} M_{X_j}(t) = \frac{\beta_j \, \delta_j^{\alpha_j/\beta_j}}{\Gamma\left(\frac{\alpha_j}{\beta_j}\right)} \int_0^\infty x_j^{\alpha_j} \exp\left(-\delta_j x_j^{\beta_j} + t x_j\right) \mathrm{d}x_j$$

$$= \sum_{i=0}^{\infty} \frac{\left(\frac{t}{\delta_j^{1/\beta_j}}\right)^i}{i!} \frac{\Gamma\left(\frac{\alpha_j + 1 + i}{\beta_j}\right)}{\Gamma\left(\frac{\alpha_j + 1}{\beta_j}\right)} . \tag{3.3.42}$$

The second and third derivatives are calculated in similar ways. If for some β_j we have $0 < \beta_j < 1$, then the summation formulas for the mgf and cgf of X_j are numerically unstable. In this case it is better to evaluate $M_{X_j}(t)$ and $K_{X_j}(t)$ by numerical integration methods. It is apparent that computation of the saddlepoint approximation of X when the inputs are of gamma type can be quite intense and even impractical in certain cases.

3.6 Edgeworth Series

The Edgeworth series provides a fast way to approximate the density and distribution of a sum of independent and identically distributed input variables, $X = X_1 + X_2 + \cdots + X_k$ say. Barndorff-Nielsen and Cox (1989) give the Edgeworth expansion of $\overline{X} = \frac{X - \sum \mu_j}{\sum \sigma_j^2}$ when the inputs X_j are independent but not identically distributed. We use their results applied to our general problem of obtaining distributional properties of X. Let μ_j, σ_j^2 represent the

mean and variance of X_j, respectively. If we let the ith moment of X be denoted by m_i, then using independence we have:

$$E(X) = \sum_{j=1}^{k} E(X_j) = \mu \quad \text{and} \quad Var(X) = \sum_{j=1}^{k} Var(X_j) = \sigma^2.$$

where $\mu = m_1$ and $\sigma^2 = m_2 - m_1^2$. Putting $\rho_r = m_r/\sigma^r$ we have the following approximations from Barndorff-Nielsen and Cox (1989):

$$\widehat{f}(\overline{x}) = \phi(\overline{x})\left\{1 + \frac{\rho_3 H_3(\overline{x})}{6\sqrt{n}} + \frac{\rho_4 H_4(\overline{x})}{24n} + \frac{\rho_3^2 H_6(\overline{x})}{72n}\right\} \quad \text{and} \tag{3.3.43}$$

$$\widehat{F}(\overline{x}) = \Phi(\overline{x}) - \phi(\overline{x})\left\{\frac{\rho_3 H_2(\overline{x})}{6\sqrt{n}} + \frac{\rho_4 H_3(\overline{x})}{24n} + \frac{\rho_3^2 H_5(\overline{x})}{72n}\right\} \tag{3.3.44}$$

for the density and distribution of $\overline{X}$, where $H_r(x)$ is the Hermite polynomial of degree r (see Mathai (1993)), $\phi(\cdot)$ is the standard normal density function and $\Phi(\cdot)$ is the standard normal distribution function. The first few Hermite polynomials are

$$H_0(x) = 1,$$

$$H_1(x) = 2x, \qquad H_2(x) = 4x^2 - 2,$$

$$H_3(x) = 8x^3 - 12x, \qquad H_4(x) = 16x^4 - 48x^2 + 12,$$

$$H_5(x) = 32x^5 - 160x^3 + 120x, \quad \text{and} \quad H_6(x) = 64x^6 - 480x^4 + 720x^2 - 120\ .$$

If, for example, all the X_j's, $j = 1, 2, \cdots, k$ are gamma inputs with parameters (α_j, λ_j) then

$$m_1 = \sum_{j=1}^{k} \alpha_j \lambda_j = \mu \qquad m_2 = \sum_{j=1}^{k} (\alpha_j + 1)\alpha_j \lambda_j^2$$

$$m_3 = \sum_{j=1}^{k} (\alpha_j + 2)(\alpha_j + 1)\alpha_j \lambda_j^3 \qquad m_4 = \sum_{j=1}^{k} (\alpha_j + 3)(\alpha_j + 2)(\alpha_j + 1)\alpha_j \lambda_j^4\ .$$

If the inputs are of gamma type with parameters $(\alpha_j, \beta_j, \delta_j)$, then

$$m_n = \sum_{j=1}^{k} \left(\frac{\Gamma\left(\frac{\alpha_j + n}{\beta_j}\right)}{\Gamma\left(\frac{\alpha_j}{\beta_j}\right)} \delta_j^{\,n/\beta_j} \right) \quad \text{for } n = 1, 2, \cdots\ .$$

3.7 Numerical Inversion of the Laplace Transform

In this section we discuss methods to determine the pdf of X, $f(x)$, as accurately as possible given values of its Laplace transform $L_f(s)$, $s \geq 0$. The Laplace transform is related to the moment generating function of X by

$$L_f(s) = M_X(-s) \ . \tag{3.3.45}$$

If $X = X_1 + X_2 + \cdots + X_k$ where the inputs X_j, $j = 1, 2, \cdots, k$ are independent, then the moment generating function of X can be written as

$$M_X(t) = \prod_{i=1}^{k} M_{X_j}(t) \ , \tag{3.3.46}$$

where $M_{X_j}(t)$ is the moment generating function of X_j. In terms of the Laplace transform, we have

$$L_f(s) = \prod_{i=1}^{k} L_{f_j}(s) \ , \tag{3.3.47}$$

where f_j is the pdf of the jth input. If the inputs are of gamma type, then one can use numerical integration or equation (3.3.30) to calculate values for either (3.3.46) or (3.3.47).

As described in Bellman, Kalaba, and Lockett (1966), there is no specific method for inverting the Laplace transform and hence of the mgf which will work well in all cases. A reason for this is that arbitrarily small changes in the Laplace transform may produce large changes in the value of $f(x)$. This is often referred to in the literature as the ill-posedness of the inverse Laplace transform problem. We examine three methods of numerically inverting the Laplace transform to obtain approximations for $f(x)$.

3.7.1 Method Based on Gaussian Quadrature

In this section we follow the results for approximating the inverse Laplace transform found in Bellman, Kalaba, and Lockett (1966). The Laplace transform of $f(x)$, denoted by $L_f(s)$, is defined as

$$L_f(s) = \int_0^\infty \exp(-sx) f(x) \mathrm{d}x \ . \tag{3.3.48}$$

First, we make a change of variables by putting $y = \exp(-x)$. In doing this, we obtain a finite interval of integration for the integral in (3.3.48). This becomes

$$L_f(s) = \int_0^1 y^{s-1} g(y) \mathrm{d}y \ . \tag{3.3.49}$$

where $g(y) = f(-\log(y))$. From numerical analysis, one can obtain an approximation to (3.3.49) by using the Gaussian quadrature formula (see Abramowitz and Stegun (1992), Bellman, Kalaba, and Lockett (1966) and Whittaker and Robinson (1967)). With this method we can obtain an extremely accurate approximation to the Laplace transform in (3.3.49) given by

$$L_f(s) \approx \sum_{i=1}^{N} w_i y_i^{s-1} g(y_i) \ . \tag{3.3.50}$$

This formula consists of finding a good numerical estimate of an integral by picking optimal abscissas y_i in $[0,1]$ and weights w_i for $i = 1, 2, \cdots, N$. The N abscissas are the zeroes of the shifted Legendre polynomials of order N

$$P_N^*(y) = P_N(1-2y) \, , \tag{3.3.51}$$

where $P_N(X)$ is the Legendre polynomial of order N (see Mathai (1993)). The weights w_i are called the Christoffel weights for the interval $(-1,1)$. If $r_i = 2x_i - 1$ and $a_i = 2w_i$ for $i = 1, 2, \cdots, N$, then the r_i's are the N roots of the Legendre polynomial of degree N. The

new weights a_i are given by

$$a_i = \int_{-1}^{1} \frac{P_N(r)}{(r - r_i)P'_N(r_i)} \mathrm{d}r \quad i = 1, 2, \cdots, N \, . \tag{3.3.52}$$

A proof of this can be found in Bellman, Kalaba, and Lockett (1966).

The following is the method of of inverting the Laplace transform based on the Gaussian quadrature formula. If we evaluate $L_f(s)$ at N different values, we obtain a system of N equations in N unknowns. Hence by solving

$$L_f(j) = \sum_{i=1}^{N} w_i y_i^{j-1} g(y_i) \quad j = 1, \cdots, N \tag{3.3.53}$$

we can obtain approximations for $g(y_i)$ or equivalently for $f(x_i)$ by $x_i = -\log(y_i)$. The approximations are

$$f(x_i) \approx \sum_{j=1}^{N} a_{ij} L_f(j) \quad i = 1, 2, \cdots, N \tag{3.3.54}$$

where the coefficients are determined by solving the system in (3.3.53) . Given these solutions, we can then fit an interpolating polynomial for $f(x)$ through the points x_i, $i = 1, 2, \cdots, N$. For example the Lagrange interpolating polynomial is given by

$$\phi(x) = \sum_{i=1}^{N} \frac{\pi(x)}{(x - x_i)\pi'(x_i)} f(x_i) \, , \tag{3.3.55}$$

where $\pi(x)$ and $\pi'(x)$ are defined as

$$\pi(x) = \prod_{i=1}^{N} (x - x_i) \quad \text{and} \tag{3.3.56}$$

$$\pi'(x_i) = \frac{\mathrm{d}}{dx}\pi(x)\Big|_{x=x_i} = \prod_{\substack{j=1 \\ j \neq i}}^{N} (x_i - x_j) \quad . \tag{3.3.57}$$

One can show that the roots of $P_N^*(x)$ are uniformly distributed over $[0, 1]$ as N increases. The values of $x_i = -\log(y_i)$ however do not follow the same behavior. Because of this behavior, the method so far is not good for large values of x. To remedy this, we start by

noting that the Laplace transform of $f(ax)$ is $\frac{L_f(s/a)}{a}$. Hence, (3.3.53) becomes

$$\frac{L_f(j/a)}{a} = \sum_{i=1}^{N} w_i y_i^{j-1} f(-a \log(y_i)) \quad j = 1, \cdots, N \,. \tag{3.3.58}$$

This allows us to obtain approximations for $f(x)$ near the values $x_i = -a \log(y_i)$ from

$$f(x_i) \approx \sum_{j=1}^{N} a_{ij} \frac{L_f(j/a)}{a} \quad i = 1, 2, \cdots, N \tag{3.3.59}$$

where the a_{ij}, for $i, j = 1, \cdots, N$, are the same as before.

One must be aware that the solutions $f(x_i)$, $i = 1, \cdots, N$, are unstable functions of $L_f(j)$, $j = 1, \cdots, N$. The accuracy worsens as N increases. A way to see this is to start at $N = 3$ and then increase N until precision decreases.

3.7.2 Method of Papoulis

Papoulis (1957) gives a method of determining $f(x)$ in terms of an infinite sequence of equidistant points

$$s_i = (2i + 1)\rho \quad i = 1, 2, 3, \cdots \tag{3.3.60}$$

in the region of existence of $L_f(s)$. The parameter ρ is an arbitrary positive real number. The function of interest, $f(x)$, is given as a series of Legendre polynomials

$$f(x) = \sum_{i=1}^{\infty} C_i P_{2i}(\exp(-\rho x)) \,, \tag{3.3.61}$$

where the coefficients C_i, $i = 1, 2, \cdots, k, \cdots$, are determined by solving the following system:

$$\begin{aligned} \rho L_f(\rho) &= C_0 \\ \rho L_f(3\rho) &= \frac{C_0}{3} + \frac{2C_1}{15} \\ &\cdots \end{aligned}$$

$$\begin{aligned}\rho L_f\left((2k+1)\rho\right) &= \frac{C_0}{2k+1}+\frac{2kC_1}{(2k+1)(2k+3)}+\cdots \\ &\quad +\frac{2k(2k-2)\cdots(2)C_k}{(2k+1)(2k+3)\cdots(4k+1)} \qquad (3.3.62)\end{aligned}$$

$$\cdots$$

An approximation to $f(x)$ is thus given by

$$f(x) \approx \sum_{i=1}^{N} C_i P_{2i}\left(\exp\left(-\rho x\right)\right) \ . \qquad (3.3.63)$$

Papoulis (1957) explains that the parameter ρ is chosen depending on the interval in which $f(x)$ is to be described.

3.7.3 Method of Miller and Guy

Another method for approximating the pdf $f(x)$ by numerically inverting its Laplace transform is described in Miller and Guy (1966). Just as in the method of Papoulis, it determines functional values of $f(x)$ based on values of $L_f(s)$ at discrete points of s in the domain of existence. Evaluation of $L_f(s)$ at the points $s_i = (\beta+1+i)\sigma$, $i = 1, 2, \cdots$, where $\beta > -1$ and $\sigma > 0$ determine the coefficients in an infinite series expansion of $f(x)$ in terms of Jacobi polynomials, $P_n^{(\alpha,\beta)}(x)$ (see Mathai (1993)). The theory for this method requires that

$$\lim_{x\to 0} f(x) \quad \text{and} \quad \lim_{x\to\infty} f(x)$$

be finite, which is the case for our general problem. The series solution to $f(x)$ is

$$f(x) = \sum_{n=0}^{\infty} C_n P_n^{(0,\beta)}\left(2\exp\left(-\sigma x\right)-1\right), \qquad (3.3.64)$$

where the solution of

$$\sigma L_f\left((\beta+1)\sigma\right) = \frac{C_0}{(\beta+1)}$$

$$\begin{aligned}
\sigma L_f((\beta+2)\sigma) &= \frac{C_0}{\beta+2} + \frac{2C_1}{(\beta+2)(\beta+3)} \\
&\cdots \\
\sigma L_f((\beta+k+1)\sigma) &= \sum_{m=0}^{k} \frac{k(k-1)\cdots(k-(m-1))\, C_m}{(k+\beta+1)(k+\beta+2)\cdots(k+\beta+1+m)} \qquad (3.3.65) \\
&\cdots
\end{aligned}$$

gives the coefficients C_i for $i = 1, 2, 3, \cdots$. The pdf $f(x)$ can then be approximated using

$$f(x) \approx \sum_{n=0}^{N} C_n P_n^{(0,\beta)} \left(2\exp\left(-\sigma x\right) - 1\right) . \qquad (3.3.66)$$

The accuracy of the approximation (3.3.66) to $f(x)$ may be improved by picking appropriate parameters β and σ. The general guideline given by Miller and Guy (1966) is to select β and σ such that

$$-0.5 \leq \beta \leq 5.0 \quad \text{and} \quad 0.05 \leq \sigma \leq 2.0 . \qquad (3.3.67)$$

3.8 Examples and Comparisons

We now illustrate the methods developed in sections 3.2, 3.4, 3.5 and 3.7. All computations are made using Maple VI and the R (version 1.0.0) statistics package. Most calculations are straightforward and the algorithms used may be found in any standard numerical analysis book, such as Conte and de Boor (1980). To implement the task of summing over partitions of an integer as needed in sections 3.2 and 3.4, we refer the reader to Nijenhuis and Wilf (1978).

In the following sections, we refer to the method found in section 3.2.1 as the Integer Method and the one in section 3.2.2 is called the Exact Method. The method in section 3.3 is called the Gamma Method. Results obtained using equations (3.3.24) and (3.3.23) in

section 3.4 are called the Raw Method and Convolution Method, respectively. For sections 3.5 and 3.7.1, they are the Saddlepoint and Gaussian Quadrature Methods, respectively.

3.8.1 Gamma Inputs

Here we consider the case where the inputs X_i, $i = 1, \cdots, k$ are gamma inputs. We begin with the case where the shape parameters are integers, following with an example for general shape parameters.

Integer Shape Parameters

Consider a 6 stage process where $X = X_1 + X_2 + \cdots + X_6$, and the independent gamma inputs X_j have $(m_1, m_2, m_3, m_4, m_5, m_6) = (4, 3, 2, 1, 2, 5)$ and $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6) = (1, 1.2, 4.3, 2.5, 3.7, 4.9)$ as shape and scale parameters. The Integer Method gives the coefficients for equation (3.3.4) as:

$a_{11} = -8.494665 \quad a_{12} = -0.64293 \quad a_{13} = -0.03641015 \quad a_{14} = -0.001163430$

$a_{21} = 11.36933 \quad a_{22} = -0.8478953 \quad a_{23} = 0.1482047$

$a_{31} = -441483062 \quad a_{32} = -17846569$

$a_{41} = -563.4837$

$a_{51} = -11111622 \quad a_{52} = -373560.4$

$a_{61} = 517801841 \quad a_{62} = -51141923 \quad a_{63} = 4451650 \quad a_{64} = -309331.9 \quad a_{65} = 13140.54$.

The resulting density is

$$f(x) = \begin{cases} \sum_{j=1}^{6} \sum_{i=1}^{m_j} a_{ji} \left\{ \frac{x^{i-1} e^{-x/\lambda_j}}{\lambda_j^i (i-1)!} \right\} & \text{for } x > 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{3.3.68}$$

The survival function, $S(x)$ is obtained through direct numerical integration of $f(x)$. These two are plotted in figure 3.1. Note that $\sum_{j=1}^{6}\sum_{i=1}^{m_j} a_{ji} = 1$.

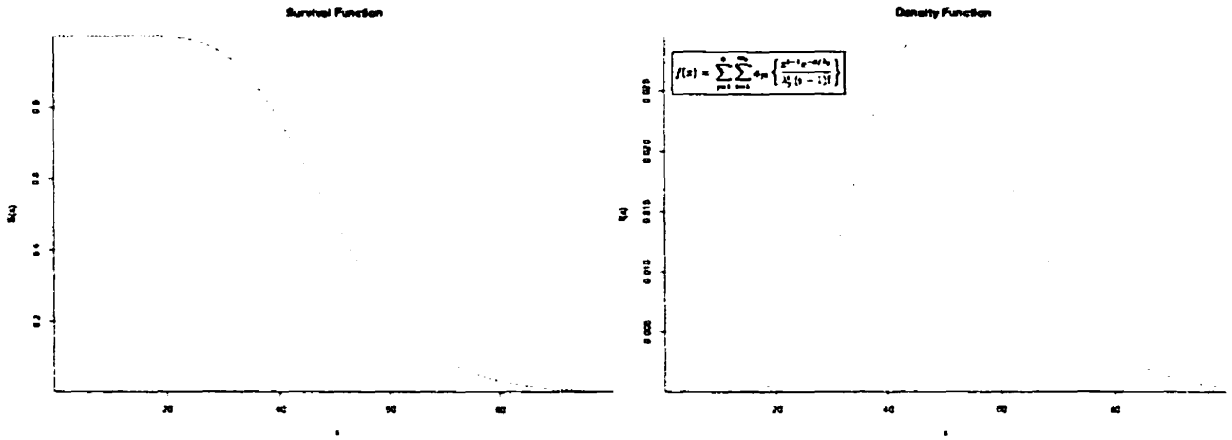

Figure 3.1: Survival and Density Functions: Integer Method

General Shape Parameters

We can apply the Exact Method to the previous example. Using the result in (3.3.68) obtained through the Integer Method, we compare it with results obtained through the Exact Method. The Exact Method gives the pdf in equation (3.3.6). We then approximate the density of X by summing up to a large enough integer, n say, so that

$$f(x) \approx C\sum_{r=0}^{n} \delta_r \frac{x^{\alpha+r-1}e^{-x/\lambda_1}}{\lambda_1^{\alpha+r}\Gamma(\alpha+r)} \quad \text{for } x > 0, \tag{3.3.69}$$

and 0 elsewhere. To obtain a "good" approximation, we find an integer n such that $C\sum_{r=0}^{n} \delta_r \approx 1$. For our example, the Exact Method approximated through the formula (3.3.69) for $n = 40$ yields $C\sum_{r=0}^{40} \delta_r = 0.7403136$ and for $n = 65$ we obtain $C\sum_{r=0}^{65} \delta_r = 0.9863663$. The weights δ_r are easily computed by using the recursion identity in (3.3.15). An approximation for the cdf of X can be obtained from direct integration of (3.3.69). The cdf allows us to obtain an approximation for the survival function $S(x) = 1 - F(x)$. For the

saddlepoint approximation of the distribution function of X, we solve the equation (3.3.34) to obtain $\hat{t}$. For our example, we use the more stable bisection method rather than the Newton-Raphson method to solve (3.3.34) numerically.

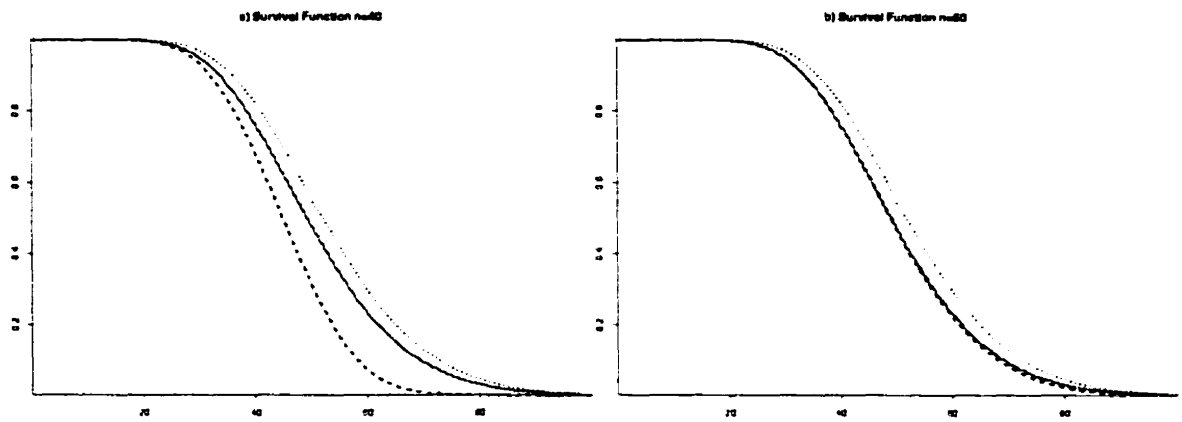

Figure 3.2: Survival Functions: Solid: Integer Method, dashed: Exact Method and dotted: Saddlepoint Method

Figure 3.2 shows the plots for the survival function of X obtained from the three methods. The solid curve shows the Integer Method, the dotted is for the Saddlepoint Method, and the dashed curve gives the Exact Method. The first plot (a) gives the Exact method for $n = 40$ and plot (b) is for $n = 65$. The Saddlepoint Method performs better than the Exact Method for $n = 40$. For $n = 65$ the Exact Method outperforms the Saddlepoint Method.

We can also use the Gaussian Quadrature Method to obtain an approximation for the density $f(x)$ for the failure time X. Putting $N = 11$ and $a = 40$ in equation (3.3.58) gives fairly good values for the density. For instance, at $x' = 40.28929229502032$ we get .026471344924404 for the approximation to $f(x')$. The real value for $f(x')$ is 0.02615254. Using spline interpolation we obtain an approximation for $f(x)$. The result is plotted in figure 3.3. The dotted curve gives the exact density of X obtained from the Integer Method and the solid curve gives the interpolated function for $f(x)$ at the estimated points found

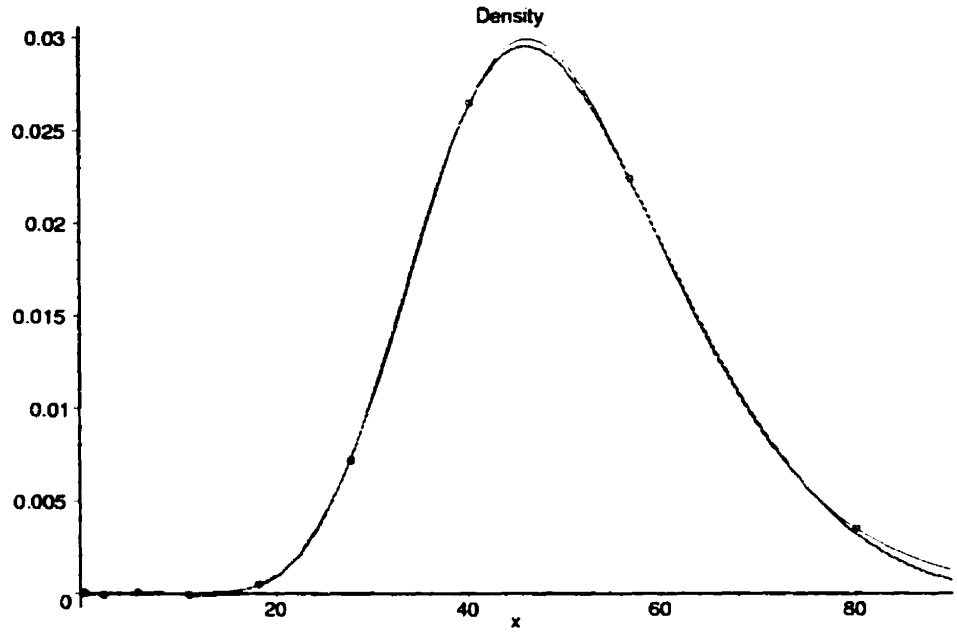

Figure 3.3: Density functions: dotted: Integer Method and solid: Gaussian Quadrature Method

using the Gaussian Quadrature Method. In general, the Gaussian Quadrature Method is extremely useful for making very fast approximations when $X_1, \cdots, X_k$ are all gamma inputs. When they are gamma type inputs, this method is not very stable.

Another method which yields fast results is the Gamma Method of section 3.3. It approximates the density of X by a gamma density with parameters $\tilde{\alpha} = 12.86742$ and $\tilde{\beta} = 3.932411$:

$$\tilde{f}(x) = \frac{1}{\Gamma(\tilde{\alpha})}\tilde{\beta}^{-\tilde{\alpha}}x^{\tilde{\alpha}-1}\exp\left(-x/\tilde{\beta}\right) \tag{3.3.70}$$

and 0 elsewhere. Figure 3.4 gives the results of the Integer and Gamma Methods. The solid curve represents the Integer Method and the dotted one is for the Gamma method.

3.8.2 Gamma Type Inputs

Most methods for calculating the pdf of $X = X_1 + X_2 + \cdots + X_k$ when the inputs are gamma type inputs are impractical. The Saddlepoint method requires the solution to

$$K'(t) = x \tag{3.3.71}$$

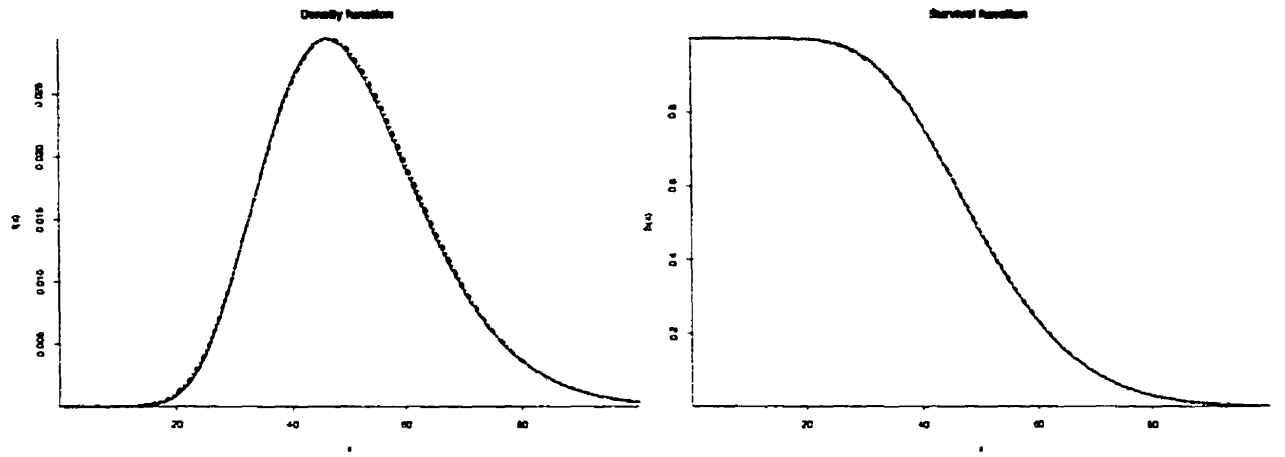

Figure 3.4: Density and Survival functions: Solid: Integer Method and dash-dotted: Gamma Method

where $K(t)$ is the cumulant generating function and the Raw Method requires the calculation of partitions of an integer, both of which can be quite long and exhaustive. The Raw Method utilizes the following approximation to (3.3.24)

$$f_X(x) \approx Cx^{\alpha-1}\sum_{r=0}^{N}(-1)^r \sum_{r_1+r_2+\cdots+r_k=r} M(x;r_1,r_2,\cdots,r_k) \qquad (3.3.72)$$

Of course, the larger N is the better the approximation. However, this means that one needs to calculate partitions for large integers, which can take some time. Also, for fixed N, the Raw Method yields poor results for the tail of the density of X. The best way to find fast and precise approximations for the pdf of X is to use the Convolution Method. We show its effectiveness by looking at two examples.

First, let $X = X_1 + X_2$ where the parameters of the inputs are $(\alpha_1, \beta_1, \delta_1) = (1, 1, 0.5)$ and $(\alpha_2, \beta_2, \delta_2) = (2, 2, 2)$. This means that X_1 is a gamma input with parameters $(1, 2)$ and X_2 is a Weibull input with parameters $(2, 1/2)$. Using the "integrate" package in R (version 1.0.0), we obtain values for the pdf of X. Figure 3.5 plots the results using both methods. The solid line represents the plot of the pdf using the Raw Method and the dashed line is for the Convolution Method. The abruptness in the solid curve is due to the numerical

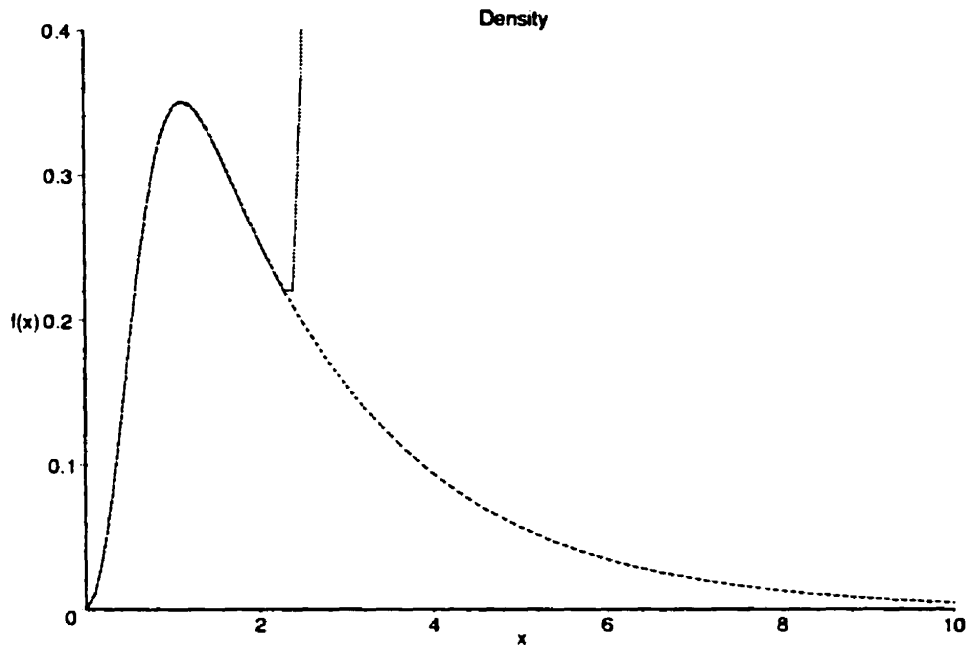

Figure 3.5: Density functions: Solid: Raw Method and dashed: Convolution Method

instability of the Raw Method.

Now let $X = X_1 + X_2 + X_3 + X_4$, where the parameters of the gamma type inputs are $(\alpha_1, \beta_1, \delta_1) = (1, 1, 1)$, $(\alpha_2, \beta_2, \delta_2) = (2, 1.2, 0.5)$, $(\alpha_3, \beta_3, \delta_3) = (1.6, 0.8, 2)$ and $(\alpha_4, \beta_4, \delta_4) = (3.2, 2, 0.7)$. The Convolution Method produces accurate plots for the density and survival function of X. These are shown in figure 3.6. The method depends on the numerical integration method chosen. For instance, R (version 1.0.0) uses the A. C. Genz's fortran ADAPT subroutine to do the calculations. For more on numerical evaluation of multiple integrals refer to Genz (1986) and Berntsen, Espelid, and Genz (1991).

3.9 Conclusion

We have seen various methods of computing the distribution functions of $X = X_1 + X_2 + \cdots + X_k$ where the inputs X_j are assumed to be independent gamma type inputs with parameters

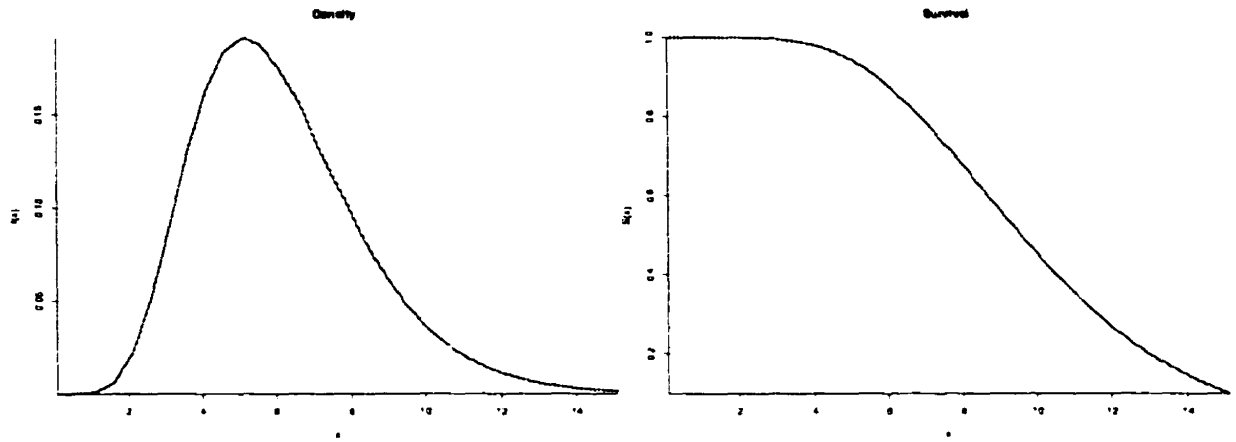

Figure 3.6: Density and Survival functions: Convolution Method

$(\alpha_j, \beta_j, \gamma_j)$. For the special case where $\beta_j = 1$ for all j, the inputs are gamma inputs. Here, we have shown that the Integer and Exact Methods yield better results than the Saddlepoint Method. Also for this special case, the Gamma and Gaussian Quadrature Methods can yield faster results and are quite useful for making preliminary approximations. For the general case, it seems that the only method worth considering is the Convolution Method. It is relatively fast and gives accurate results. For the examples given in the previous section, results were obtained in about 10 to 15 minutes on a Pentium 450MHz with 64 megabytes of ram memory. The Convolution Method involves no matrix inversion or the numerical solution to a function, as in the case with the Saddlepoint Method. In the end, the choice of method is highly variant on the problem being considered.

Appendix A

Independent Censoring

The independent censoring assumption has been discussed by many to encompass a wide variety of censoring mechanisms which include type I and type II censoring. Results can be found in Kalbfleisch and Prentice (1980), Fleming and Harrington (1991) and Lawless (1982). We shall follow the reasoning of Lawless (1982).

Suppose that we have data from n subjects and assume that we have independent censoring. Our life testing sample consists of the waiting times $(y_1, y_2, \cdots, y_n)$ along with indicators $(\delta_1, \delta_2, \cdots, \delta_n)$ that come from a model with survival function $S(x)$. The survival function has unknown parameters $\underline{\theta} = (\theta_1, \cdots, \theta_l)$. We will show that the likelihood for this data under non-informative censoring is the same as the one given in (2.2.56). First, let us assume that our data lies in the interval $[0, T]$. To obtain the result, we discretize the problem and then take limits. Let the failure time axis be partitioned into intervals $I_j = [a_{j-1}, a_j)$, $j = 1, 2, \cdots, k$ where $a_0 = 0$, $a_k = T$ and $a_{k+1} = \infty$. Let $R_j =$ the risk set for I_j, $D_j =$ the set of individuals who die in I_j and $C_j =$ the set of individuals who are censored in I_j with $n_j = |R_j|$, $d_j = |D_j|$ and $c_j = |C_j|$. Also, let $\Delta_j = a_j - a_{j-1}$ for $j = 1, 2, \cdots, k$. The

data now consists of the sets $D_1, D_2, \cdots, D_k$ and $C_1, C_2, \cdots, C_k$. The probability distribution of the D_i's and C_i's denoted by $\mathrm{P}(D_1, C_1, \cdots, D_k, C_k)$ can be written as a product of conditional probabilities

$$\begin{aligned}\mathrm{P}(D_1, C_1, \cdots, D_k, C_k) &= \mathrm{P}(D_1)\,\mathrm{P}(C_1|D_1) \times \\ &\quad \prod_{j=2}^{k} \Big[\mathrm{P}(D_j|D_1, D_2, \cdots, D_{j-1}, C_{j-1}) \times \\ &\quad\quad \mathrm{P}(C_j|D_1, D_2, \cdots, C_{j-1}, D_j) \Big] \\ &= \mathrm{P}(D_1)\, Q_1 \prod_{j=2}^{k} \mathrm{P}(D_j|D_1, D_2, \cdots, D_{j-1}, C_{j-1})\, Q_j \qquad \text{(A-1)}\end{aligned}$$

where $Q_1 = \mathrm{P}(C_1|D_1)$ and $Q_j = \mathrm{P}(C_j|D_1, D_2, \cdots, C_{j-1}, D_j)$ for $j = 2, 3, \cdots, k$. Let us examine the term $\mathrm{P}(D_j|D_1, D_2, \cdots, D_{j-1}, C_{j-1})$ in (A-1). The assumption of independent censoring can be sub-divided into smaller assumptions.

Assumption 1. The mechanisms of failure and censoring for different subjects act independently in I_j.

Assumption 2. For each subject in R_j, we have

$$\mathrm{P}(\text{Dying in } I_j|D_1, C_1, \cdots, D_{j-1}, C_{j-1}) = \mathrm{P}(\text{Dying in } I_j|\text{Survival beyond } I_{j-1}) \ . \qquad \text{(A-2)}$$

If we have covariate information for each subject, assumption 2 should change to the following assumption

Assumption 3. For each subject in R_j and conditional on the covariates, we have

$$\mathrm{P}(\text{Dying in } I_j|D_1, C_1, \cdots, D_{j-1}, C_{j-1}) = \mathrm{P}(\text{Dying in } I_j|\text{Survival beyond } I_{j-1}) \ . \qquad \text{(A-3)}$$

Given these assumptions, we have the following:

$$\mathrm{P}(D_j|D_1, D_2, \cdots, D_{j-1}, C_{j-1}) = \prod_{i \in D_j} \left(1 - \frac{S_i(a_j)}{S_i(a_{j-1})}\right) \prod_{t \in R_j \setminus D_j} \left(\frac{S_t(a_j)}{S_t(a_{j-1})}\right) . \qquad \text{(A-4)}$$

where $S_i(x)$ is the survival function of subject i. Given that the survival function is the same for all subjects, $S_i(x) = S(x)$ for $i = 1, 2, \cdots, n$, (A-4) becomes

$$\mathrm{P}\left(D_j | D_1, D_2, \cdots, D_{j-1}, C_{j-1}\right) = \left(1 - \frac{S(a_j)}{S(a_{j-1})}\right)^{d_j} \left(\frac{S(a_j)}{S(a_{j-1})}\right)^{n_j - d_j} . \qquad \text{(A-5)}$$

We can now write (A-1) in terms of (A-5):

$$\prod_{j=1}^{k} \left(1 - \frac{S(a_j)}{S(a_{j-1})}\right)^{d_j} \left(\frac{S(a_j)}{S(a_{j-1})}\right)^{n_j - d_j} Q \qquad \text{(A-6)}$$

where $Q = \prod_{j=1}^{k} Q_j$. Noting that $n_{j+1} = n_j - d_j - c_j$ and $S(a_0) = 1$, we can expand the product in (A-6) and rearrange terms to get

$$\prod_{j=1}^{k} \left(S(a_{j-1}) - S(a_j)\right)^{d_j} \left(S(a_j)\right)^{c_j} Q \, . \qquad \text{(A-7)}$$

Under the assumption of non-informative censoring, we can ignore the term Q in the product above. In chapter 1 we saw that $f(x) = -\frac{\mathrm{d}}{\mathrm{d}x} S(x)$, so by the Mean Value Theorem there exists $\eta_j \in I_j$ such that $S(a_{j-1}) - S(a_j) = f(\eta_j)\Delta_j$ for $j = 1, 2, \cdots, k$. We get the likelihood of the life testing sample by taking the limit of

$$\frac{\prod_{j=1}^{k} \left(f(\eta_j)\Delta_j\right)^{d_j} \left(S(a_j)\right)^{c_j}}{\prod_{j=1}^{k} \Delta_j} \qquad \text{(A-8)}$$

as $k \to \infty$ with $\max_j (\Delta_j) \to 0$ and $T \to \infty$. This gives

$$L(\underline{\theta}) = \prod_{i=1}^{n} \left[f(y_i)\right]^{\delta_i} \left[S(y_i)\right]^{1-\delta_i} . \qquad \text{(A-9)}$$

Appendix B

Newton-Raphson Method

There are many methods for optimizing a general system of n equations in n unknowns subject to constraints. One must be aware of the fact that there is no one particular method that works for all cases. However, the most widely used techniques are the Newton-Raphson method and method of scoring. These two rely on quadratic approximations to the function that is to be optimized. Often in statistics, we wish to optimize the likelihood or, equivalently, the log-likelihood function.

We begin by taking a first order Taylor expansion of the log-likelihood, $\log(L(\underline{\theta}))$, about a point $\widehat{\underline{\theta}}$:

$$\log(L(\underline{\theta})) \approx \log\left(L(\widehat{\underline{\theta}})\right) + \left(\frac{\partial}{\partial \underline{\theta}} \log(L(\underline{\theta}))\Big|_{\underline{\theta}=\widehat{\underline{\theta}}}\right)^t \left(\underline{\theta} - \widehat{\underline{\theta}}\right) . \tag{B-1}$$

The right-hand side of (B-1) is maximized by equating its gradient to zero:

$$\left(\frac{\partial}{\partial \underline{\theta}} \log(L(\underline{\theta}))\Big|_{\underline{\theta}=\widehat{\underline{\theta}}}\right) + \left[\left(\frac{\partial}{\partial \underline{\theta}}\right)\left(\frac{\partial}{\partial \underline{\theta}}\right)^t \log(L(\underline{\theta}))\Big|_{\underline{\theta}=\widehat{\underline{\theta}}}\right]\left(\underline{\theta} - \widehat{\underline{\theta}}\right) = 0 \tag{B-2}$$

Let the solution for $\underline{\theta}$ in (B-2) be denoted by $\widehat{\underline{\theta}}^1$. This is given by

$$\widehat{\underline{\theta}}^1 = \widehat{\underline{\theta}} + \left[-\left(\frac{\partial}{\partial \underline{\theta}}\right)\left(\frac{\partial}{\partial \underline{\theta}}\right)^t \log(L(\underline{\theta}))\Big|_{\underline{\theta}=\widehat{\underline{\theta}}}\right]^{-1} \frac{\partial}{\partial \underline{\theta}} \log(L(\underline{\theta}))\Big|_{\underline{\theta}=\widehat{\underline{\theta}}} . \tag{B-3}$$

This gives the following iteration scheme:

$$\underline{\widehat{\theta}}^{l+1} = \underline{\widehat{\theta}}^{l} + \left[-\left(\frac{\partial}{\partial \underline{\theta}}\right)\left(\frac{\partial}{\partial \underline{\theta}}\right)^{t} \log\left(L(\underline{\theta})\right)\Big|_{\underline{\theta}=\underline{\widehat{\theta}}^{l}} \right]^{-1} \frac{\partial}{\partial \underline{\theta}} \log\left(L(\underline{\theta})\right)\Big|_{\underline{\theta}=\underline{\widehat{\theta}}^{l}} . \qquad \text{(B-4)}$$

There are a few problems of which one must be aware when using the Newton-Raphson Method. Besides being computationally expensive, it is not an ascent method. In other words we may not have $L\left(\underline{\widehat{\theta}}^{l+1}\right) > L\left(\underline{\widehat{\theta}}^{l}\right)$. The iteration scheme can be made into an ascent method by replacing the sample information matrix in (B-4) by a positive definite matrix A^{l}. These methods are called quasi-Newton methods and further discussion may be found in Conte and de Boor (1980) and Lange (1999).

Appendix C

The Delta Method

Suppose that a random variable $\underline{Y} = (Y_1, Y_2, \cdots, Y_k)$ has mean $\underline{\mu} = (\mu_1, \mu_2, \cdots, \mu_k)$ and a non-singular covariance matrix $W = (\sigma_{ij})$, where $i = 1, 2, \cdots, k$ and $j = 1, 2, \cdots, k$. Let g be some function of $\underline{Y}$ that admits a Taylor's series expansion about $\underline{\mu}$

$$\begin{aligned} g(\underline{Y}) &= g(\underline{\mu}) + \left(\frac{\partial}{\partial \underline{Y}} g(\underline{Y})\Big|_{\underline{Y}=\underline{\mu}}\right)^t (\underline{Y} - \underline{\mu}) \\ &+ \frac{1}{2}(\underline{Y} - \underline{\mu})^t \left(\left(\frac{\partial}{\partial \underline{Y}}\right)\left(\frac{\partial}{\partial \underline{Y}}\right)^t g(\underline{Y})\Big|_{\underline{Y}=\underline{\mu}}\right)(\underline{Y} - \underline{\mu}) + \cdots \end{aligned} \tag{C-1}$$

Assuming that $\left(\frac{\partial}{\partial \underline{Y}} g(\underline{Y})\Big|_{\underline{Y}=\underline{\mu}}\right) \neq 0$ and ignoring higher terms, one gets estimates for the mean and variance of g in terms of $\underline{\mu}$ and W. These are

$$E(g(\underline{Y})) \approx g(\underline{\mu}) \quad \text{and} \tag{C-2}$$

$$Var(g(\underline{Y})) \approx W' = (\underline{Y} - \underline{\mu})^t W (\underline{Y} - \underline{\mu}) \ . \tag{C-3}$$

If furthermore $\underline{Y}$ is asymptotically normal with mean $\underline{\mu}$ and covariance matrix W, then $g(\underline{Y})$ is asymptotically normal with mean $g(\underline{\mu})$ and covariance matrix W' given in (C-3). If $\underline{Y} = Y$ a one dimensional random variable, then $\underline{\mu} = \mu$ and $W = \sigma^2$. Thus, $E(g(Y)) \approx g(\mu)$ and $Var(g(Y)) \approx \sigma^2 (g'(\mu))^2$.

Appendix D

Greenwood's Formula

We now give a derivation of Greenwood's method for estimating the variance of the product-limit estimator, also known as the Kaplan-Meier estimator. We follow the derivation found in Lawless (1982).

The product-limit estimator for the survival function $S(x)$ given the failure time data $((y_1, \delta_1), \cdots, (y_n, \delta_n))$ is

$$\widehat{S}(x) = \prod_{a_i < x} \left(1 - \frac{d_i}{r_i}\right) . \tag{D-1}$$

where $a_1, a_2, \cdots, a_m$ are the m uncensored failure times of the sample. Further, we define $a_0 = 0$ and $a_{m+1} = \infty$. d_i is the number of subjects who fail at time a_i and r_i is the number of subjects at risk just before a_i. If $a_k < x \leq a_{k+1}$ then the product-limit estimator can be rewritten as

$$\widehat{S}(x) = \prod_{i=1}^{k} \left(1 - \frac{d_i}{r_i}\right) \tag{D-2}$$

$$= \prod_{i=1}^{k} \widehat{p}_i \tag{D-3}$$

where $\widehat{p}_i = \dfrac{r_i - d_i}{r_i}$, $i = 1, 2, \cdots, m$. Now, $\widehat{p}_i$ is the maximum likelihood estimator of p_i, the

probability that a subject survives past the time a_i. If there are r_i subjects at risk before a_i then the number of subjects that survive past a_i is $r_i - d_i$. Thus $r_i - d_i$ is a binomial random variable with parameters (r_i, p_i). Thus the variance of $r_i - d_i$ is

$$Var(r_i - d_i) = r_i p_i (1 - p_i) \ . \tag{D-4}$$

The variance of $\widehat{p}_i = \dfrac{r_i - d_i}{r_i}$ is

$$\begin{aligned} Var(\widehat{p}_i) &= Var\left(\frac{r_i - d_i}{r_i}\right) \\ &= \frac{p_i(1-p_i)}{r_i} \ . \end{aligned} \tag{D-5}$$

A reasonable estimate for (D-5) is obtained by substituting $\widehat{p}_i$ for p_i. That is

$$Var(\widehat{p}_i) \approx \frac{\widehat{p}_i(1-\widehat{p}_i)}{r_i} \ . \tag{D-6}$$

Taking the logarithm of the product-limit estimator in (D-1), we have

$$\log\left(\widehat{S}(x)\right) = \sum_{i=1}^{k} \log\left(\widehat{p}_i\right) \quad \text{and} \tag{D-7}$$

$$Var\left(\log\left(\widehat{S}(x)\right)\right) = \sum_{i=1}^{k} Var\left(\log\left(\widehat{p}_i\right)\right) \ . \tag{D-8}$$

Using the delta method (see appendix C) we have an estimate for $Var\left(\log\left(\widehat{p}_i\right)\right)$

$$\begin{aligned} Var\left(\log\left(\widehat{p}_i\right)\right) &\approx \frac{Var(\widehat{p}_i)}{\widehat{p}_i^{\,2}} \\ &= \frac{1}{r_i}\frac{1-\widehat{p}_i}{\widehat{p}_i} \\ &= \frac{d_i}{r_i}\frac{1}{(r_i - d_i)} \ . \end{aligned} \tag{D-9}$$

Applying the delta method once more, we have

$$\begin{aligned} Var\left(\log\left(\widehat{S}(x)\right)\right) &\approx \frac{1}{\left(\widehat{S}(x)\right)^2} Var\left(\widehat{S}(x)\right) \quad \text{or} \\ Var\left(\widehat{S}(x)\right) &\approx \left(\widehat{S}(x)\right)^2 Var\left(\log\left(\widehat{S}(x)\right)\right) \\ &= \left(\widehat{S}(x)\right)^2 \sum_{i=1}^{k} \frac{d_i}{r_i}\frac{1}{(r_i - d_i)} \ . \end{aligned} \tag{D-10}$$

Bibliography

ABRAMOWITZ, MILTON and STEGUN, IRENE A., eds. (1992). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, New York, xiv+1046 pp.

BALAKRISHNAN, N. and COHEN, A. C. (1991). *Order Statistics and Inference*. Academic Press, Boston, MA, xx+377 pp.

BARNDORFF-NIELSEN, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343–365.

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1979). Edgeworth and saddle-point approximations with statistical applications. *Journal of the Royal Statistical Society. Series B. Methodological* **41**, 279–312.

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1984). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *Journal of the Royal Statistical Society. Series B. Methodological* **46**, 483–495.

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1989). *Asymptotic Techniques for Use in Statistics*. Chapman & Hall, London, x+252 pp.

BARTHOLOMEW, D. J. (1963). The sampling distribution of an estimate arising in life testing. *Technometrics* **5**, 361–374.

BELLMAN, RICHARD, KALABA, ROBERT E., and LOCKETT, JO ANN (1966). *Numerical Inversion of the Laplace Transform: Applications to Biology, Economics, Engineering and Physics*. American Elsevier Publishing, New York, viii+249 pp.

BERNTSEN, JARLE, ESPELID, TERJE O., and GENZ, ALAN (1991). An adaptive algorithm for the approximate calculation of multiple integrals. *Association for Computing Machinery. Transactions on Mathematical Software* **17**, 437–451.

BIYARI, KHALED H. and LINDSEY, WILLIAM C. (1993). Statistical distributions of Hermitian quadratic forms in complex Gaussian variables. *Institute of Electrical and Electronics Engineers. Transactions on Information Theory* **39**, 1076–1082.

BLÆSILD, PREBEN and JENSEN, JENS LEDET (1985). Saddlepoint formulas for reproductive exponential models. *Scandinavian Journal of Statistics. Theory and Applications* **12**, 193–202.

BUTLER, RONALD W. and HUZURBAZAR, APARNA V. (1997). Stochastic network models for survival analysis. *Journal of the American Statistical Association* **92**, 246–257.

BUTLER, RONALD W., HUZURBAZAR, S., and BOOTH, JAMES G. (1992a). Saddlepoint approximations for the Bartlett-Nanda-Pillai trace statistic in multivariate analysis. *Biometrika* **79**, 705–715.

BUTLER, RONALD W., HUZURBAZAR, S., and BOOTH, JAMES G. (1992b). Saddlepoint approximations for the generalized variance and Wilks' statistic. *Biometrika* **79**, 157–169.

CONTE, S. D. and DE BOOR, C. (1980). *Elementary Numerical Analysis: An Algorithmic Approach*. McGraw-Hill, New York, x+278 pp.

COX, D. R. and OAKES, D. (1984). *Analysis of Survival Data*. Chapman & Hall, London, ix+201 pp.

CROWDER, M. J., KIMBER, A. C., SMITH, R. L., and SWEETING, T. J. (1991). *Statistical Analysis of Reliability Data*. Chapman and Hall, London, xii+250 pp.

DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Annals of Mathematical Statistics* **25**, 631–650.

DANIELS, H. E. (1987). Tail probability approximations. *International Statistical Review* **55**, 37–48.

DIVSALAR, D., SIMON, M. K., and SHAHSHAHANI, M. (1990). The performance of trellis-coded mdpsk with multiple symbol detection. *Institute of Electrical and Electronics Engineers. Transactions on Communications* **38**, 1391–1403.

EFRON, BRADLEY and HINKLEY, DAVID V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika* **65**, 457–487.

FLEMING, THOMAS R. and HARRINGTON, DAVID P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York, xiv+429 pp.

GALAMBOS, JANOS and KOTZ, SAMUEL (1978). *Characterizations of Probability Distributions*. Springer, Berlin, viii+169 pp.

GENZ, ALAN (1986). Fully symmetric interpolatory rules for multiple integrals. *SIAM Journal on Numerical Analysis* **23**, 1273–1283.

HAGER, HAROLD W. and BAIN, LEE J. (1970). Inferential procedures for the generalized gamma distribution. *Journal of the American Statistical Association* 65, 1601–1609.

HUZURBAZAR, APARNA V. and HUZURBAZAR, S. (1999). Survival and hazard functions for progressive diseases using saddlepoint approximations. *Biometrics* **55**, 198–203.

JENSEN, JENS LEDET (1995). *Saddlepoint Approximations*. The Clarendon Press Oxford University Press, New York, xii+332 pp.

KAGAN, A. M., LINNIK, YU. V., and RAO, C. R (1973). *Characterization Problems in Mathematical Statistics*. Wiley, New York, xii+499 pp.

KALBFLEISCH, JOHN D. and PRENTICE, ROSS L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York, xi+321 pp.

KAPLAN, E. L. and MEIER, PAUL (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457–481.

KOULIS, T. and MATHAI, A.M. (2000). Survival analysis and reliability problems with gamma type inputs. To be presented at MMR'2000: Methodology, Practice, and Inference, Université Victor Segalen-Bordeaux 2.

LANGE, KENNETH (1999). *Numerical Analysis for Statisticians*. Springer, New York, xvi+356 pp.

LAWLESS, JERALD F. (1982). *Statistical Models and Methods for Lifetime Data*. Wiley, New York, xi+580 pp.

LUGANNANI, ROBERT and RICE, STEPHEN (1980). Saddle point approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability* **12**, 475–490.

MATHAI, A. M. (1982). Storage capacity of a dam with gamma type inputs. *Annals of The Institute of Statistical Mathematics* **34**, 591–597.

MATHAI, A. M. (1993). *A Handbook of Generalized Special Functions for Statistical and Physical Sciences*. The Clarendon Press, Oxford, xi+235 pp.

MATHAI, A. M. and PEDERZOLI, G. (1977). *Characterizations of the Normal Probability Law*. Wiley, New York, x+149 pp.

MATHAI, A. M. and PROVOST, SERGE B. (1992). *Quadratic Forms in Random Variables:Theory and Applications*. Marcel Dekker, New York, xxii+367 pp.

MILLER, MAX K. and GUY, JR., W. T. (1966). Numerical inversion of the Laplace transform by use of Jacobi polynomials. *SIAM Journal on Numerical Analysis* **3**, 624–635.

MILLER, JR., RUPERT G. (1981). *Survival Analysis*. Wiley, New York, xi+238 pp.

MOSCHOPOULOS, P. G. (1985). The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics* **37**, 541–544.

NIJENHUIS, ALBERT and WILF, HERBERT S. (1978). *Combinatorial Algorithms*. Academic Press, New York, second edn., xv+302 pp.

PAPOULIS, ATHANASIOS (1957). A new method of inversion of the Laplace transform. *Quarterly of Applied Mathematics* **14**, 405–414.

RAO, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley, New York, second edn., xx+625 pp.

REID, N. (1988). Saddlepoint methods and statistical inference. *Statistical Science* **3**, 213–238.

SINHA, S. K. (1986). *Reliability and Life Testing*. Wiley, New York, xii+252 pp.

STACY, E. W. (1962). A generalization of the gamma distribution. *Annals of Mathematical Statistics* **33**, 1187–1192.

WHITTAKER, EDMUND and ROBINSON, G. (1967). *The Calculus of Observations: An Introduction to Numerical Analysis*. Dover Publications, New York, fourth edn., xiv+397 pp.

ZACKS, SHELEMYAHU (1992). *Introduction to Reliability Analysis*. Springer-Verlag, New York, xiv+212 pp.