Chapter 1

INTRODUCTION

1.1 Background

Survival data consist of a response variable that measures the duration of time until a specified event occurs, and a set of covariates that can be either discrete or continuous. Some observations are right-censored, that is, the time until the event occurs, is not observed due to withdrawal or termination of the study. Thus survival data consist of a response variable (duration of time), as well as an additional variable indicating which responses are **observed** survival times and which are **censored** times and some covariates.

In this thesis, a structured framework for the analysis of policy survival times has been developed. Such a framework must be embedded in the experimental design. The lifetime of a policy is measured from the inception date up to the lapsing date or a cut-off point. If the lapsing date is prior to a cut-off point of the study, determined in advance, then the lifetime is **observed** (an uncensored observation). If a policy is still in force (alive) when the cut-off point is reached, the lifetime of this policy is said to be **right-censored**. This thesis focuses on **grouped data** for lifetimes. This scenario has become extremely important, not only for application in the actuarial context, but also in other fields.

In this thesis, the analysis takes account of the actual lifetime (duration) of the policy rather than just recording the fact that the policy lapsed or was still alive after (say) twelve months. In other words, the response variable, lifetime, is a continuous one, and the whole distribution of lifetimes can be used. This is in contrast to a categorical response where a loglinear logit model can be used to express the relationship among the categorical response and some covariates at a fixed time point of twelve months, conditional on a restricted experimental design where all the policies must have an exposure of at least twelve months when investigating the lapses of policies in the first year. There is no such restrictions in the more general experimental design used in this thesis where all the policies can be used in the analysis, even those policies with inception dates very close to the cut-off point.

All the policies that are written during a certain month and are recorded at the end of that month as that month's business, can be considered as a group of policies with a common starting point or entry time. In this way policies written in different months or time-periods have different entry times. This is called **staggered entry** of policies.

The policy lifetimes (observed and censored lifetimes) are typically of a discrete nature and principles of grouped data have to be taken into account when fitting lifetime distributions to such survival data. Statistical analysing techniques for this kind of grouped data are currently insufficient.

1.2 The Research Problem

The aim of the research is the statistical modelling of parametric survival distributions of **grouped survival data** of long- and shortterm policies in the insurance industry, by means of a method of maximum likelihood estimation **subject to constraints**. The statistical methodology is developed from previous research on maximum likelihood estimation subject to constraints (refer to [11, 30]) and special attention is given to the **staggered entry** of policies and the fitting of **parametric** regression models.

Much literature exist on the fitting of survival distributions, mainly in connection with non-parametric analyses or the semi-parametric proportional hazards regression model of Cox [6, 23, 41]. Theoretical results for continuous data of lifetimes are well developed. Corresponding results for grouped data of lifetimes or the so-called interval-censored data are available only in special cases, and then again mainly in the non-parametric and semi-parametric case (refer to [33, 15, 14]). Only in recent years interval censoring in parametric model fitting were addressed by [27, 22, 14, 36]. This was made possible by modern computing power and sophisticated statistical programming languages.

The standard method of estimation to be used in the literature is to maximize the partial likelihood (at Cox's model) or the full likelihood (at parametric regression models). The estimates of the parameters are then found by maximizing the likelihood function numerically. The methodology of maximum likelihood estimation subject to constraints, used in this thesis, leads to **explicit expressions** for the estimates of the parameters, as well as for approximated variances and covariances of the estimates, which gives **exact** maximum likelihood estimates of the parameters. This makes direct extension to more complex designs feasible.

Once the parameters of the survival distributions have been estimated, estimated hazard and survivor functions, odds of a lapse, **odds ratios** and **hazard ratios** at time t can be directly calculated, as well as estimated percentiles for the fitted survival distributions. These estimates form the statistical foundation for scientific decisionmaking with respect to actuarial design, maintenance and marketing of insurance policies.

The intrinsic value of the methodology described in this thesis is of fundamental importance to the actuarial science. The statistical modelling offers parametric models for survival distributions, in contrast with non-parametric models that are used commonly in the actuarial profession. This leads to more accurate estimation procedures. When the parametric models provide a good fit to data, they tend to give more precise estimates of the quantities of interest such as odds ratios, hazard ratios or median lifetimes. Estimates of these quantities will tend to have smaller standard errors than they would in the absence of distributional assumptions.

Data from the insurance industry are extensive data sets with very large sample sizes. The focus in this thesis will be on the estimation of lifetime distributions, based on a large sample of lifetimes of policies that are grouped into intervals of lifetimes. The variance- covariance matrix can be estimated as part of this maximum likelihood estimation procedure subject to constraints, but no meaningful interpretation can be given to the exceedance probability values that are associated with the very small standard errors of the parameter estimates due to the large sample size. Therefore discrepancy values will be used to evaluate the fitting of the lifetime distributions.

1.3 Generalization of the Statistical Procedures

One important aspect of the research field of survival analysis is the wide range of types of application. Although the methodology in this thesis is developed specifically for the insurance industry, it may be applied in the normal context of research and scientific decisionmaking, that includes for example survival distributions for the medical, biological, engineering, econometric and sociological sciences.

1.4 Future Research

The potential for extending the methodology to other realistic practical application is unlimited. This can be to the advance of the insurance industry in general. The models should reflect an interactive adaptability for direct application in practice by salesforce (the marketing people on ground level), as well as for actuarial planning.

1.5 Outlay of chapters

In chapter two, basic survival functions are defined and a description of the survival distributions that are used in this thesis, together with their properties, are given. The construction of different likelihood functions is also discussed.

In chapter three, parametric models are fitted to a single sample of survival data. The standard method of maximum likelihood estimation is reviewed. This is followed by the methodology of maximum likelihood estimation subject to constraints, applied in two different scenarios: a fixed censoring time as well as staggered entry of policies. The chapter is concluded with simulation studies to compare the estimates of the parameters obtained by the standard and the proposed estimating procedures.

In chapter four parametric regression models are fitted and important indicators of the effect of the covariates are defined such as risk scores (hazard ratios) and indices (odds ratios).

The developed theoretical results are applied in chapter five to a typical data set from an insurance company.

Chapter six concludes with a summary of the most important results in this thesis and focuses on application of these results in practice, especially in the actuarial industry.

All computer programs that were written in the SAS/IML language to perform the new methodology are given in appendix A.

The abstract appears at the end of the thesis, just after the references.

Chapter 2

SURVIVAL FUNCTIONS AND DISTRIBUTIONS

2.1 Introduction

The aim of this chapter is to introduce notation and survival concepts and to summarize properties of the survival distributions to be used in this thesis.

2.2 Notation

2.2.1 Right-censored continuous survival data

Let X be a nonnegative continuous random variable denoting the lifetime (or survival time) of a policy. The lifetime of a policy is measured from inception date up to the lapsing date. If the lapsing date is prior to a fixed termination date of the study, determined in advance, then the lifetime is **observed** (an uncensored observation). If a policy is still in force (alive) when the termination point is reached, the lifetime of this policy is said to be **right-censored**.

A fixed censoring time

Consider the simplest scenario where all policies enter the study at the **same time**. Let C be the fixed termination date of the study, determined in advance. C is then the preassigned **fixed censoring time**. Instead of observing $X_1, X_2, ..., X_n$, only $T_1, T_2, ..., T_n$ are observed where $T_j = \begin{cases} X_j & \text{if} \quad X_j \leq C \\ C & \text{if} \quad X_j > C \end{cases}$

The survival data, based on a sample of size n, can then be represented by pairs of random variables (T_j, δ_j) where $T_1, T_2, ..., T_n$ are independent identically distributed random variables, each with distribution function F and density function f. δ_j is the survival status of the j^{th} policy and indicates whether the lifetime for the j^{th} policy corresponds to a lapse $(\delta_j = 1)$ or is censored $(\delta_j = 0)$. This type of censoring is known as **Type I right-censoring**.

Staggered entry of policies

Staggered entry of policies occur when policies enter the study at **different times**.

Define C_j as the potential censoring time for the j^{th} policy, associated with lifetime X_j. $C_1, C_2, ..., C_n$ are independent identically distributed random variables, each with distribution function G and density function g.

Only pairs $(T_1, \delta_1), (T_2, \delta_2), ..., (T_n, \delta_n)$ can be observed where

$$\begin{aligned} T_j &= min(X_j, C_j) \text{ for the } j^{th} \text{ policy} \\ \delta_j &= \begin{cases} 1 & \text{if} \quad X_j \leq C_j \quad \text{, that is, } X_j \text{ is not censored} \\ 0 & \text{if} \quad X_j > C_j \quad \text{, that is, } X_j \text{ is censored} \end{cases} \end{aligned}$$

$T_1, T_2, ..., T_n$ are independent identically distributed random variables with distribution function F if $T_j = X_j$ and distribution function G if $T_j = C_j$.

Random entries to the study are assumed. This type of censoring is known as **random right-censoring**. Type I right-censoring is a special case of the random censoring model by simply setting C_i=C. With random censoring the crucial **assumption** that X_i and C_i are independent, is made. This assumption means that censoring is not related to any factors associated with the actual survival time. This is called **non-informative censoring**. A graphical test to examine the assumption of non-informative censoring is given by [5, page

274].

2.2.2 Discrete data

Sometimes the lifetime T is a discrete random variable. Discrete lifetimes arise due to rounding off time measurements. Lifetimes of policies are usually measured on a discrete time-scale (to the nearest month).

If the lifetimes of the policies are distinct and $t_1 < t_2 < ... < t_n$ denotes the ordered lifetimes, then $R(t_i)$ is the risk set at time t_i, that is the set of policies that are still alive at a time **just prior** to t_i $\quad i = 1, 2, ..., n$.

To allow for possible ties in the data, suppose that the lapses occur at D distinct times $t_1 < t_2 < ... < t_D$. Define

$$\begin{aligned} d_i &= \text{number of lapses at time } t_i \\ Y_i &= \text{number of policies that are at risk at an instant just prior to time } t_i \\ &= \text{number of policies that are still alive at time } t_i \text{ or lapse at time } t_i \\ &= \text{number of policies in the risk set } R(t_i) \end{aligned}$$

2.2.3 Interval-censored data

Introduction

In the case where a large group of policies are followed from a common starting point (the inception date) over certain periods of time, the data consist of only the **number** of policies that lapse or are censored within various time-intervals. Interval censoring is used to describe this situation where a policy's lifetime is known only to lie between two values (interval boundaries). Interval-censored data can involve left and right censoring, as being outlayed in the scheme of [4, page 144-145].

Notation

Assume that the lifetime for the j^{th} policy is bounded between two known values, denoted $b_j \leq T < c_j$ and it is known whether the policy lapsed. The intervals do not necessarily coincide and may be overlapping. Left- and right-censoring are special cases of interval-censored data. Denote the left-censoring time by C_l and the right-censoring time by C_r. Observations that are left-censored have $b_j = 0$, $c_j = C_l$ and $\delta_j = 1$. Observations that are right-censored have $b_j = C_r$, $c_j = \infty$ and $\delta_j = 0$.

Grouped data

Grouped data arise due to the grouping of the continuous lifetimes of policies into k adjacent, non-overlapping fixed intervals

$$I_j = [a_{j-1}; a_j) \quad j = 1, 2, ..., k$$

with $a_0 = 0$ and $a_k = \infty$.

Thus grouped data consist of the **numbers** of observed and censored lifetimes falling into each of the k intervals. In the case of the last interval $I_k = [a_k; \infty)$ the policies will lapse some time in this open interval and all the lifetimes in this interval can be considered as observed.

Define

d_j = number of policies that lapsed in I_j

Y_j' = number of policies entering interval I_j that have not lapsed

W_j = number of policies with censored lifetimes in I_j
assuming that censored lifetimes are independent of the time those policies would have had they been observed until the lapse

$Y_j = Y_j' - \frac{W_j}{2}$
= number of policies at risk of lapsing in I_j (policies that are still alive at a_{j-1})
assuming that censored and observed lifetimes are uniformly distributed over the interva

2.3 Basic Survival Functions

2.3.1 Introduction

Four functions characterize the distribution of T, namely the **survivor function**, which is the probability of a policy surviving beyond time t, the **hazard rate function** which is the chance a policy, surviving up to time t, will lapse in the next instant, the **probability density (or mass) function** which is the unconditional probability of a lapse at time t, and the **mean residual life** at time t, which is the mean time to a lapse, given no lapse at t. If any one of these four functions is known, the other three then can be uniquely determined.

2.3.2 Probability distribution and distribution function

For a **continuous** random variable T the probability density function of T is

$$f(t) = \lim_{dt \to 0} \frac{P(t < T < t + dt)}{dt} = \lim_{dt \to 0} \frac{F(t + dt) - F(t)}{dt} \tag{2.3 .1}$$

with cumulative distribution function

$$F(t) = P(T \leq t) = \int_0^t f(x)\, dx \tag{2.3 .2}$$

$f(t)$ is the unconditional probability of a lapse at time t (probability per unit of time over a small interval of time) and is called the unconditional lapse rate.

For a **discrete** random variable T the probability mass function of T at time t_i is

$$p(t_i) = \begin{cases} P(T = t_i) & i = 1, 2, \ldots \\ 0 & \text{elsewhere} \end{cases} \tag{2.3 .3}$$

The distribution of T is characterized by either one of two equivalent functions, namely the survivor function and the hazard rate function.

2.3.3 Survivor function

The survivor function at time t is the probability of a policy surviving beyond time t, also called the survival rate at time t.

The survivor function for **continuous** T is defined as

$$S(t) = P(T > t) = \int_t^\infty f(x)\,dx = 1 - F(t) \quad \text{for} \quad t > 0 \tag{2.3 .4}$$

- The graph of $S(t)$ versus t is called the survival curve
- $S(t)$ is a continuous, strictly decreasing function with $S(0) = 1$ and $S(\infty) = 0$
- $f(t) = -\frac{d}{dt}S(t)$

The survivor function for **discrete** T is defined as

$$S(t) = P(T > t) = \sum_{t_j > t} p(t_j) \quad \text{for} \quad t > 0 \tag{2.3 .5}$$

- The graph of $S(t)$ versus t is for discrete T a nonincreasing step function with jumps downward at $t_1, t_2, ...$

2.3.4 Hazard rate function

The hazard rate function at time t is the **conditional** probability of a lapse in the next instant just beyond time t **given that the policy has survived up to time** t. This is also called the instantaneous lapse rate assuming that the policy was alive up to time t.

For a **continuous** random variable T the hazard rate function at time t is defined as

$$\begin{aligned}
h(t) &= \lim_{dt\to 0}\frac{P(t<T<t+dt \mid T>t)}{dt}\\
&= \lim_{dt\to 0}\frac{P[(t<T<t+dt)\cap(T>t)]}{P(T>t)dt}\\
&= \lim_{dt\to 0}\frac{P(t<T<t+dt)}{dt}\cdot\frac{1}{P(T>t)}\\
&= f(t)\cdot\frac{1}{S(t)}
\end{aligned}$$

$$\Rightarrow h(t)=\frac{f(t)}{S(t)}=\frac{-\frac{d}{dt}S(t)}{S(t)} \qquad (2.3\ .6)$$

For a **discrete** random variable T the hazard rate function at time t_j is defined as

$$\begin{aligned}
h(t_j) &= P(T=t_j \mid T\geq t_j) &&= \frac{P[(T=t_j)\cap(T\geq t_j)]}{P(T\geq t_j)}\\
&= \frac{P(T=t_j)}{P(T>t_{j-1})} &&= \frac{p(t_j)}{S(t_{j-1})}\\
&= \frac{S(t_{j-1})-S(t_j)}{S(t_{j-1})} &&= 1-\frac{S(t_j)}{S(t_{j-1})}
\end{aligned}$$

$$\Rightarrow \frac{S(t_j)}{S(t_{j-1})}=1-h(t_j)$$

Note that $S(t)=\frac{S(t_1)}{S(t_0)}\cdot\frac{S(t_2)}{S(t_1)}\cdot\ldots\cdot\frac{S(t)}{S(t_{j-1})}$

$$\Rightarrow S(t)=\prod_{t_j\leq t}\frac{S(t_j)}{S(t_{j-1})}=\prod_{t_j\leq t}[1-h(t_j)] \qquad (2.3\ .7)$$

The hazard rate is zero for a discrete random variable T except at points where a lapse could occur.

2.3.5 Cumulative hazard function

For **continuous** T the cumulative hazard function is defined as

$$\begin{aligned} H(t) &= \int_0^t h(x)\,dx \\ &= \int_0^t \frac{-\frac{d}{dx}S(x)}{S(x)}\,dx \\ &= -[\ln S(t) - \ln S(0)] \\ &= -\ln S(t) \end{aligned}$$

$$\Rightarrow h(t) = \frac{d}{dt}H(t) = -\frac{d}{dt}\ln S(t) \qquad (2.3\ .8)$$

For **discrete** T the cumulative hazard function is defined as

$$H(t) = \sum_{t_j \le t} h(t_j)$$

2.3.6 Median lifetime, mean lifetime and mean residual lifetime

The survivor function is used to determine the median lifetime (median time to a lapse) and other percentiles.

$$\begin{aligned} \text{median} \quad t_{50} &= \text{time point at which} \quad S(t_{50}) is equal to 0.5 & (2.3\ .9) \\ p^{th}\text{percentile} &= \quad t_p \quad \text{so that} \quad S(t_p) = \frac{100-p}{100} & (2.3\ .10) \end{aligned}$$

The mean lifetime is represented as the area under the survival curve.

$$\mu = E(T) = \int_0^\infty S(t)\,dt \qquad (2.3\ .11)$$

The variance of T is related to the survivor function by

$$var(T) = 2\int_0^\infty tS(t)\,dt - \left[\int_0^\infty S(t)\,dt\right]^2 \qquad (2.3\ .12)$$

The mean residual lifetime (mean amount of lifetime remaining after a particular time t) is defined as

$$mrl(t) = E(T - t \mid T > t) = \frac{\int_t^\infty (x-t)f(x)\,dx}{P(T > t)} = \frac{\int_t^\infty S(x)\,dx}{S(t)} \qquad (2.3\ .13)$$

2.4 Survival Models

2.4.1 Introduction

In this section analytical forms of survival distributions that are often used in fitting survival data are given. Properties of the Weibull survival model are discussed by [2, 25, 26], while [22] describes the log-logistic survival model and [34, 26] the lognormal survival model.

2.4.2 Weibull distribution

Consider $T \sim$ Weib(λ, α)

$$\begin{array}{lcll} \lambda & = & \text{scale parameter} & \lambda > 0 \\ \alpha & = & \text{shape parameter} & \alpha > 0 \end{array}$$

$$\begin{array}{rrcll} \text{density function} & f(t) & = & \lambda\alpha t^{\alpha-1}\,\exp\{-\lambda t^\alpha\} \quad t > 0 & (2.4\ .1) \\ \text{expected value} & E(T) & = & \Gamma(1+\frac{1}{\alpha}) \cdot \lambda^{\frac{-1}{\alpha}} & (2.4\ .2) \\ \text{variance} & var(T) & = & \left[\Gamma(1+\frac{2}{\alpha}) - \Gamma^2(1+\frac{1}{\alpha})\right] \cdot \lambda^{\frac{-2}{\alpha}} & (2.4\ .3) \\ \text{survivor function} & S(t) & = & \exp(-\lambda t^\alpha) & (2.4\ .4) \\ \text{hazard function} & h(t) & = & \lambda\alpha t^{\alpha-1} & (2.4\ .5) \\ p^{th}\ \text{percentile} & t_p & = & \left[\frac{1}{\lambda}\ln\left(\frac{100}{100-p}\right)\right]^{\frac{1}{\alpha}} \quad p = 1, 2, ..., 99 & \\ & & & & (2.4\ .6) \end{array}$$

$$\begin{array}{rrcll} \Rightarrow \quad \text{median lifetime} & t_{50} & = & \left[\frac{1}{\lambda}\ln\left(\frac{100}{100-50}\right)\right]^{\frac{1}{\alpha}} & \\ & & = & \left[\frac{1}{\lambda}\ln(2)\right]^{\frac{1}{\alpha}} & (2.4\ .7) \end{array}$$

Note that

$$-\ln S(t) \quad = \quad \lambda t^{\alpha}$$

$$\Rightarrow \ln(-\ln S(t)) = \ln\lambda + \alpha \ln t \qquad (2.4\ .8)$$

2.4.3 The log-logistic distribution

Consider $T \sim$ log-logistic(λ, α) $\quad \lambda \;=\;$ scale parameter $\;\lambda > 0$
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad \alpha \;=\;$ shape parameter $\;\alpha > 0$

density function $\quad f(t) = \dfrac{\lambda\alpha t^{\alpha-1}}{(1+\lambda t^{\alpha})^2}$ with $t > 0$ $\qquad$ (2.4 .9)

expected value $\quad E(T) = \dfrac{\pi csc\frac{\pi}{\alpha}}{\alpha\lambda^{\frac{1}{\alpha}}}$ if $\alpha > 1$ $\qquad$ (2.4 .10)

variance $\quad var(T) = \dfrac{2\pi csc\frac{2\pi}{\alpha}}{\alpha\lambda^{\frac{2}{\alpha}}} - E(T^2)$ if $\alpha > 2$(2.4 .11)

survivor function $\quad S(t) = (1+\lambda t^{\alpha})^{-1}$ $\qquad$ (2.4 .12)

hazard function $\quad h(t) = \dfrac{\lambda\alpha t^{\alpha-1}}{(1+\lambda t^{\alpha})}$ $\qquad$ (2.4 .13)

odds of a lapse at time t $\quad \dfrac{1-S(t)}{S(t)} = \lambda t^{\alpha}$ $\qquad$ (2.4 .14)

p^{th} percentile $\quad t_p = \left[\dfrac{1}{\lambda}\cdot\dfrac{p}{100-p}\right]^{\frac{1}{\alpha}}$ $\qquad$ (2.4 .15)

$\Rightarrow$ median lifetime $\quad t_{50} = \left(\dfrac{1}{\lambda}\right)^{\frac{1}{\alpha}}$ $\qquad$ (2.4 .16)

Note that

$$\ln\left(\frac{1-S(t)}{S(t)}\right) = \ln\lambda + \alpha\ln t \qquad (2.4\ .17)$$

2.4.4 The lognormal distribution

Consider $T \sim$ lognormal(μ, σ^2).

density function $\quad f(t) = \dfrac{\exp\left\{-\frac{1}{2}(\frac{ln(t)-\mu}{\sigma})^2\right\}}{t\sigma(2\pi)^{\frac{1}{2}}}$ with $t > 0$ $\qquad$ (2.4 .18) expected value $E(T) = \exp(\mu + 0.5\sigma^2)$ (2.4 .19)

variance $var(T) = \exp(2\mu + \sigma^2) \cdot [\exp(\sigma^2) - 1]$ (2.4 .20)

survivor function $S(t) = 1 - \Phi[\frac{ln(t) - \mu}{\sigma}]$ (2.4 .21)

with Φ the standard normal distribution function

hazard function $h(t) = f(t)/S(t)$

p^{th} percentile $t_p = \exp(\mu + \sigma z_p)$ (2.4 .22)

with z_p the p^{th} percentile of the n(0;1) distribution.

2.4.5 Location-scale parameter survival models

Introduction

A univariate location-scale parameter distribution is described by [13] as a distribution with a probability density function of the form

$$f(y;\mu,\sigma) = \frac{1}{\sigma} \cdot g(\frac{y-\mu}{\sigma}) \quad -\infty < y < \infty \qquad (2.4\ .23)$$

where

μ = location parameter $-\infty < \mu < \infty$

σ = scale parameter $\sigma > 0$

g = a fully specified probability density function defined on $(-\infty, \infty)$

The survivor function corresponding to Equation 2.4 .23 is

$$G[\frac{y-\mu}{\sigma}] \quad \text{where} \quad G(x) = \int_x^{\infty} g(z)\, dz \qquad (2.4\ .24)$$

The extreme value, logistic and normal distributions are examples of location-scale parameter distributions that are used in fitting survival models to survival data.

The extreme value distribution

Consider $Y =\sim$ EV(σ, μ) μ = location parameter

σ = scale parameter $\sigma > 0$

$$\text{density function} \qquad f(y) = \frac{1}{\sigma} \cdot \exp\left\{ \left(\frac{y-\mu}{\sigma} \right) - \exp\left(\frac{y-\mu}{\sigma} \right) \right\} \qquad -\infty < y < \infty \tag{2.4 .25}$$

$$\text{expected value} \qquad E(Y) = \mu - 0.5772\,\sigma \qquad \text{Euler's constant} = 0.5772 \tag{2.4 .26}$$

$$\text{variance} \qquad var(Y) = \frac{\pi^2\sigma^2}{6} \tag{2.4 .27}$$

The extreme value distribution with $\mu = 0$ and $\sigma = 1$ is termed the standard extreme value distribution. A discussion of the exreme value distribution is given by [34].

The logistic distribution

Consider $Y \sim$ logistic(μ, σ) $\quad \mu =$ location parameter
$\sigma =$ scale parameter $\quad \sigma > 0$

$$\text{density function} \qquad f(y) = \frac{1}{\sigma} \cdot \frac{\exp\left(\frac{y-\mu}{\sigma} \right)}{\left\{ 1 + \exp\left(\frac{y-\mu}{\sigma} \right) \right\}^2} \qquad -\infty < y < \infty \tag{2.4 .28}$$

$$\text{expected value} \qquad E(Y) = \mu \tag{2.4 .29}$$

$$\text{variance} \qquad var(Y) = \frac{\pi^2\sigma^2}{3} \tag{2.4 .30}$$

The logistic distribution is a symmetrical distribution whose probability density function is very similar to that of the normal distribution.

The logistic distribution with $\mu = 0$ and $\sigma = 1$ is termed the standard logistic distribution.

The properties of the logistic distribution are discussed by [2].

The normal distribution

Consider $Y \sim$ normal(μ, σ) $\quad \mu =$ location parameter
$\sigma =$ scale parameter $\quad \sigma > 0$

$$\text{density function} \qquad f(y) = \frac{1}{\sigma(2\pi)^{\frac{1}{2}}} \cdot \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\} \qquad -\infty < y < \infty \tag{2.4 .31}$$

$$\text{expected value} \qquad E(Y) = \mu \tag{2.4 .32}$$

$$\text{variance} \qquad var(Y) = \sigma^2 \tag{2.4 .33}$$

The normal distribution with $\mu = 0$ and $\sigma = 1$ is termed the standard normal distribution.

2.5 Relationships between Survival Distributions

2.5.1 Introduction

In analyzing survival data it is often convenient to work with $Y = \ln(T)$, the logarithm of the lifetimes.

2.5.2 Relationship between the Weibull and extreme value distributions

The extreme value distribution arises when lifetimes are taken to be Weibull distributed. The following relationship between the Weibull and extreme value distributions exists:

$$T \sim \quad \text{Weib}(\lambda, \alpha) \qquad \Longleftrightarrow \qquad Y = \ln(T) \sim \quad \text{EV}\left(\frac{1}{\alpha}, \frac{-\ln\lambda}{\alpha}\right) \tag{2.5 .1}$$

This result can be proved in the following way. If $T \sim \text{Weib}(\lambda, \alpha)$ with density function

$$f_T(t) = \lambda\alpha t^{\alpha-1} \exp\{-\lambda t^{\alpha}\} \qquad t > 0$$

and survivor function

$$S_T(t) = \exp\{-\lambda t^{\alpha}\}$$

then, by using the transformation technique, $Y = ln(T)$ has density function

$$\begin{aligned} f_Y(y) &= f_T(t) \cdot \mid \frac{dt}{dy} \mid \\ &= \lambda\alpha(e^y)^{\alpha-1} \exp\{-\lambda(e^y)^{\alpha}\} \cdot \mid e^y \mid \end{aligned}$$

$$= \lambda\alpha e^{\alpha y}\exp\{-\lambda e^{\alpha y}\} \quad -\infty < y < \infty \qquad (2.5\ .2)$$

$$= \exp\{-\lambda e^{\alpha y}\}\cdot(-\lambda e^{\alpha y})\cdot\alpha\cdot(-1). \qquad (2.5\ .3)$$

From Equation 2.5 .3 and the relationship $f_Y(y) = -\frac{d}{dy}S_Y(y)$ follow that the survivor function of Y is

$$S_Y(y) = \exp\{-\lambda e^{\alpha y}\}. \qquad (2.5\ .4)$$

Equation 2.5 .2 implies that

$$\begin{aligned} f_Y(y) &= \lambda\alpha\exp\{\alpha y - \lambda e^{\alpha y}\} \\ &= \exp\{ln(\lambda)\}\cdot\alpha\cdot\exp\{\alpha y - \exp\{ln(\lambda)\}\exp\{\alpha y\}\} \\ &= \alpha\exp\{\alpha y + ln(\lambda) - \exp\{\alpha y + ln(\lambda)\}\} \\ &= \alpha\exp\left\{\alpha\left[y - \frac{-ln(\lambda)}{\alpha}\right] - \exp\left\{\alpha\left[y - \frac{-ln(\lambda)}{\alpha}\right]\right\}\right\}. \qquad (2.5\ .5)\end{aligned}$$

Let $\mu = \frac{-ln(\lambda)}{\alpha}$ and $\sigma = \frac{1}{\alpha}$ in Equation 2.5 .5.

$$\Rightarrow f_Y(y) = \frac{1}{\sigma}\exp\left\{\frac{y-\mu}{\sigma} - \exp\left\{\frac{y-\mu}{\sigma}\right\}\right\}. \qquad (2.5\ .6)$$

Equation 2.5 .6 is the density of the **extreme value distribution** with parameters σ and μ

$$\Rightarrow Y \sim \text{EV}(\sigma,\mu) \equiv \text{EV}\left(\frac{1}{\alpha}, \frac{-ln(\lambda)}{\alpha}\right). \qquad (2.5\ .7)$$

This leads to the relationship between the Weibull distribution and the extreme value distribution:

$$T \sim \text{Weib}(\lambda,\alpha) \Longleftrightarrow Y = ln(T) \sim \text{EV}\left(\frac{1}{\alpha}, \frac{-\ln\lambda}{\alpha}\right) \qquad (2.5\ .8)$$

2.5.3 Relationship between the log-logistic and logistic distributions

The log-logistic distribution is related to the logistic distribution by the relationship

$$T \sim \quad \text{log-logistic}(\lambda,\alpha) \quad \Longleftrightarrow \quad Y = \ln(T) \sim \text{logistic}(\ln\lambda, \frac{1}{\alpha}). \qquad (2.5\ .9)$$

2.5.4 Relationship between the lognormal and normal distributions

The lognormal distribution is related to the normal distribution by the relationship

$$T \sim \quad \text{lognormal}(\mu, \sigma^2) \qquad \Longleftrightarrow \qquad Y = \ln(T) \sim \text{normal}(\mu, \sigma^2). \tag{2.5 .10}$$

In view of the similarity of the normal and logistic distributions, the lognormal model will tend to be very similar to the loglogistic model.

2.5.5 A linear model representation in log-time

Consider the following log-linear model that describes the basic underlying distribution of lifetimes:

$$\boxed{Y = \ln T = \mu + \sigma W}$$

where W is the error distribution, μ is the location parameter and σ is the scale parameter. A variety of distributions for W can be assumed, for example the standard extreme value distribution, the standard logistic distribution or the standard normal distribution.

- If this linear model format $Y = \ln T = \mu + \sigma W$ is used where W has the standard extreme value distribution, that is $W \sim EV(1,0)$, with density

  $$f(w) = \exp\{w - \exp(w)\} \qquad -\infty < w < \infty$$

  then $Y = ln(T)$ has an extreme value EV $\left(\frac{1}{\alpha}, \frac{-\ln\lambda}{\alpha}\right)$ distribution and T has an underlying Weibull(λ, α) distribution with parameters

  $$\lambda = \exp\left\{\frac{-\mu}{\sigma}\right\} \quad \text{and} \quad \alpha = \frac{1}{\sigma}. \tag{2.5 .11}$$

- If this linear model format $Y = \ln T = \mu + \sigma W$ is used where W has the standard logistic distribution with density

  $$f(w) = \frac{\exp(w)}{(1+\exp(w))^2} \qquad -\infty < w < \infty$$

  then $Y = ln(T)$ has a logistic$\left(\frac{1}{\alpha}, \frac{-\ln\lambda}{\alpha}\right)$ distribution and T has an underlying log-logistic(λ, α) distribution also with parameters

  $$\lambda = \exp\left\{\frac{-\mu}{\sigma}\right\} \quad \text{and} \quad \alpha = \frac{1}{\sigma}. \tag{2.5 .12}$$

- If this linear model format $Y = \ln T = \mu + \sigma W$ is used where W has the standard normal distribution with density

$$f(w) = \frac{\exp(w^2)}{(2\pi)^{\frac{1}{2}}} \qquad -\infty < w < \infty$$

  then $Y = ln(T)$ has a normal(μ, σ^2) distribution and T has an underlying lognormal(μ, σ^2) distribution.

The first of the three results is now proven. The other two results follow in a similar way.

If the error distribution $W \sim$ EV$(1,0)$, with density

$$f(w) = \exp\{w - \exp(w)\} \qquad -\infty < w < \infty$$

then, by using the transformation technique and the fact that $W = \dfrac{ln(T) - \mu}{\sigma}$, the density function of T is

$$\begin{aligned}
f_T(t) &= f_W(w) \cdot \mid \frac{dw}{dt} \mid \\
&= \exp\left\{\frac{\ln(t) - \mu}{\sigma} - \exp\left\{\frac{\ln(t) - \mu}{\sigma}\right\}\right\} \cdot \mid \frac{1}{\sigma t} \mid \\
&= \exp\left\{\frac{1}{\sigma} ln(t) - \frac{\mu}{\sigma} - \exp\left\{\frac{ln(t)}{\sigma} - \frac{\mu}{\sigma}\right\}\right\} \cdot \frac{1}{\sigma t} \\
&= t^{\frac{1}{\sigma}} \cdot t^{-1} \cdot \frac{1}{\sigma} \cdot \exp\left(\frac{-\mu}{\sigma}\right) \cdot \exp\left(-t^{\frac{1}{\sigma}} \cdot \exp\left\{\frac{-\mu}{\sigma}\right\}\right) \\
&= \exp(\frac{-\mu}{\sigma}) \cdot \frac{1}{\sigma} \cdot t^{\frac{1}{\sigma} - 1} \cdot \exp\left(-\exp\left\{\frac{-\mu}{\sigma}\right\} \cdot t^{\frac{1}{\sigma}}\right) \qquad (2.5\ .13)
\end{aligned}$$

Equation 2.5 .13 is the density of the **Weibull distribution** with parameters

$$\lambda = \exp\left\{\frac{-\mu}{\sigma}\right\} \quad \text{and} \quad \alpha = \frac{1}{\sigma}$$

$$\Rightarrow T \sim \text{Weibull}(\lambda, \alpha)$$

In a similar way, if $W \sim$ EV$(1,0)$, with density

$$f(w) = \exp\{w - \exp w\} \qquad -\infty < w < \infty$$

then, by using the transformation technique and the fact that $W = \dfrac{Y - \mu}{\sigma}$, the density function of Y is

$$\begin{aligned}
f_Y(y) &= f_W(w) \cdot \mid \frac{dw}{dy} \mid \\
&= \exp\left\{\frac{y - \mu}{\sigma} - \exp\left\{\frac{y - \mu}{\sigma}\right\}\right\} \cdot \mid \frac{1}{\sigma} \mid \\
&= \frac{1}{\sigma} \cdot \exp\left\{\frac{y - \mu}{\sigma} - \exp\left\{\frac{y - \mu}{\sigma}\right\}\right\} \\
& -\infty < y < \infty \qquad (2.5\ .14)
\end{aligned}$$

Equation 2.5 .14 is the density of the **extreme value distribution** with parameters σ and μ

$$\Rightarrow Y \sim \mathrm{EV}(\sigma, \mu) \equiv \mathrm{EV}\left(\frac{1}{\alpha}, \frac{-ln(\lambda)}{\alpha}\right)$$

To summarize, if the general linear model format $Y = \ln T = \mu + \sigma W$ is used, where the error distribution W is the standard extreme value $EV(1,0)$ distribution, then $Y = ln(T)$ has an extreme value EV $\left(\frac{1}{\alpha}, \frac{-\ln\lambda}{\alpha}\right)$ distribution and T has an underlying Weibull(λ, α) distribution with parameters

$$\lambda = \exp\left\{\frac{-\mu}{\sigma}\right\} \quad \text{and} \quad \alpha = \frac{1}{\sigma}.$$

2.6 Construction of Likelihood Functions

2.6.1 Introduction

The standard method of fitting survival models, specified in section 2.4, to survival data is the method of Maximum Likelihood Estimation (MLE). The first step is to create the specific likelihood function to be maximized. Likelihoods for different scenarios are now given.

2.6.2 Likelihood function for random right-censored continuous data

Assume the random right-censoring model where $T_1, T_2, ..., T_n$ are independent identically distributed random variables, each with distribution function F and density function f. δ_j is the survival status of the j^{th} policy and indicates whether the lifetime for the j^{th} policy corresponds to a lapse $(\delta_j = 1)$ or is censored $(\delta_j = 0)$.

Consider the pair (T_i, δ_i) for the i^{th} policy. The likelihood function is constructed by considering the contribution to the likelihood of the pairs $(t_i, \delta_i = 1)$ and $(t_i, \delta_i = 0)$ separately (refer to [25]).

- The contribution to the likelihood of the pair $(t_i, \delta_i = 1)$ is the probability that the i^{th} policy lapses at time $t_i = x_i$. This probability is

$$\begin{aligned} P(t_i \in (t_i, t_i + dt_i), \delta_i = 1) &= P(T_i \in (t_i, t_i + dt_i), C_i > t_i) \\ &= P(t_i < T_i < t_i + dt_i) \cdot P(C_i > t_i) \qquad (2.6\ .1) \end{aligned}$$

$$= \quad f(t_i)dt_i \cdot [1 - G(t_i)] \qquad (2.6\ .2)$$

Equation 2.6 .1 follows from the fact that the observed survival times are independent of the censoring times.

- The contribution to the likelihood of the pair $(t_i, \delta_i = 0)$ is the probability that the i^{th} policy survives at least time $t_i = C_i$. This probability is

$$\begin{aligned} P(C_i \in (t_i, t_i + dt_i), \delta_i = 0) &= P(C_i \in (t_i, t_i + dt_i), \boldsymbol{X_i} > t_i) \\ &= P(t_i < C_i < t_i + dt_i) \cdot P(X_i > t_i) \\ &= g(t_i)dt_i \cdot S(t_i) \qquad (2.6\ .3) \end{aligned}$$

The complete likelihood for the i^{th} policy under random censoring is

$$L(t_i, \delta_i) = \{f(t_i)dt_i[1 - G(t_i)]\}^{\delta_i} \cdot \{g(t_i)dt_i S(t_i)\}^{1-\delta_i} \qquad (2.6\ .4)$$

Under the assumption of n independent censored and observed survival times, the full likelihood function is obtained by multiplying the respective contributions of the n pairs $(t_i, \delta_i) \quad i = 1, 2, ..., n$ in the data set. This likelihood function of the full sample is

$$\begin{aligned} L(t_1, t_2, ..., t_n; \delta_1, \delta_2, ..., \delta_n) &= \prod_{i=1}^{n} L(t_i, \delta_i) \\ &= \prod_{i=1}^{n} [f(t_i)dt_i[1 - G(t_i)]]^{\delta_i} \cdot [g(t_i)dt_i S(t_i)]^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left\{[f(t_i)dt_i]^{\delta_i} [S(t_i)]^{1-\delta_i}\right\} \cdot \left\{[1 - G(t_i)]^{\delta_i} [g(t_i)dt_i]^{1-\delta_i}\right\} \\ &\qquad (2.6\ .5) \end{aligned}$$

Let $\boldsymbol{\theta}$ be the vector of parameters of the survival model. The survival model specifies the distribution of T_i, independent from the distribution of C_i. Only the first term of the product in Equation 2.6 .5 involves the unknown lifetime parameters $\boldsymbol{\theta}$, so that the last term of the product, namely

$$\left\{[1 - G(t_i)]^{\delta_i} [g(t_i)dt_i]^{1-\delta_i}\right\}$$

can be treated like constants when maximizing $L(t_1, t_2, ..., t_n; \delta_1, \delta_2, ..., \delta_n)$. Thus the likelihood function, up to a multiple constant, is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} [f(t_i)]^{\delta_i} \cdot [S(t_i)]^{1-\delta_i} \qquad (2.6\ .6)$$

The log-likelihood function

$$\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \delta_i \cdot \ln[f(t_i)] + \sum_{i=1}^{n} (1 - \delta_i) \cdot \ln[S(t_i)] \qquad (2.6\ .7)$$

is maximized to obtain the maximum likelihood estimators of the unknown parameters $\boldsymbol{\theta}$. The procedure to obtain the values of the MLE involves taking derivatives of $lnL(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, setting these equations equal to zero, and solving for $\boldsymbol{\theta}$.

2.6.3 Likelihood function for interval-censored data

Consider the lifetime of the j^{th} policy that is bounded between two known values, that is $b_j \leq T_j < c_j$. It is also known whether this policy lapsed. Observations that are left-censored have $b_j = 0$, $c_j = C_l$ and $\delta_j = 1$, while observations that are right-censored have $b_j = C_r$, $c_j = \infty$ and $\delta_j = 0$, with C_l and C_r the left- and right-censoring times respectively.

In constructing a likelihood function for interval-censored data, the information each observation provides, needs to be considered as the contribution of that observation to the likelihood (refer to [4]).

- An observation corresponding to an **exact lifetime** t_i provides information on the probability that the lapse occurs at this time t_i, which is approximately equal to the density function of T at this time. This probability is $f(t_i)$.
- For a **right-censored observation**, it is known that the lifetime is larger than C_r with C_r the right-censoring time. Thus the information provided is the survival function evaluated at C_r, that is $S(C_r)$.
- For a **left-censored observation**, it is known that the lapse has already occurred, so that the contribution to the likelihood is the cumulative distribution function evaluated at C_l, that is $F(C_l) = 1 - S(C_l)$ with C_l the left-censoring time.
- For **interval-censored data**, it is known that the lapse occurred within the interval, so that the information is the probability that the lifetime is in this interval. This probability is $S(b_i) - S(c_i)$.

The likelihood function may be constructed by putting together the above-mentioned components.

$$L(\boldsymbol{\theta}) \quad \propto \quad \prod_{i \in D} f(t_i) \cdot \prod_{i \in RC} S(C_r) \cdot \prod_{i \in LC} [1 - S(C_l)] \cdot \prod_{i \in I} [S(b_i) - S(c_i)] \qquad (2.6\ .8)$$

where

D is the set of observed lifetimes (lapses)
RC is the set of right-censored observations
LC is the set of left-censored observations
I is the set of interval-censored observations

Consider the n pairs $(t_i, \delta_i) \quad i = 1, 2, ..., n$ in the data set. Some of the responses are observed, while other responses are left, right or interval-censored. By making use of the principles of construction of the likelihood for interval-censored data in Equation 2.6 .8 it follows that

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} [f(t_i)]^{\delta_i} \cdot [S(t_i)]^{1-\delta_i} \cdot [1 - S(t_i)]^{\delta_i} \cdot [S(b_i) - S(t_i)]^{1-\delta_i} \qquad (2.6\ .9)$$

with b_i the lower end of a censoring interval.

2.6.4 Likelihood function for right-censored grouped data

Consider the grouped data case as a special case of interval-censored data where the n lifetimes of policies are grouped into k adjacent, non-overlapping fixed intervals

$$I_j = [a_{j-1}; a_j) \quad j = 1, 2, ..., k$$

with $a_0 = 0$ and $a_k = \infty$.

The likelihood function for right-censored grouped data is stated by [24].

For **complete data** with no censored lifetimes, the n observed lifetimes are grouped into k intervals so that
$n = d_1 + d_2 + ... + d_k$ with d_j=number of lapses in I_j.

The unconditional probability of a lapse in I_j is

$$\pi_j = S(a_{j-1}) - S(a_j) \quad j = 1, 2, ..., k$$

. Then $(d_1, d_2, ..., d_k)$ has a multinomial probability function

$$\frac{n!}{d_1!d_2!...d_k!} {\pi_1}^{d_1} {\pi_2}^{d_2} ... {\pi_k}^{d_k}$$

. The likelihood function can thus be taken as

$$L(\boldsymbol{\theta}) = n! \prod_{j=1}^{k} \left\{ \frac{[S(a_{j-1}) - S(a_j]^{d_j}}{d_j!} \right\} \qquad (2.6\ .10)$$

For **incomplete data**, where the n censored and observed lifetimes are grouped into k intervals, it is **further assumed** that the W_j censored lifetimes in I_j occur at the midpoint of the interval $a_j^\star = a_{j-1} + \frac{1}{2}h_j$ with $h_j = a_j - a_{j-1}$ the length of interval I_j.

For interval $I_j = [a_{j-1}; a_j)$, **conditional** on surviving till a_{j-1},

- the probability of a lapse is
$$q_j = \frac{S(a_{j-1}) - S(a_j)}{S(a_{j-1})}$$
- the probability of surviving until $a_j^\star$ is
$$p_j^\star = \frac{S(a_{j-1}) - S(a_j^\star)}{S(a_{j-1})}$$
- the probability of surviving the full interval I_j is
$$\begin{aligned} p_j &= 1 - q_j \\ &= 1 - \frac{S(a_{j-1}) - S(a_j)}{S(a_{j-1})} \\ &= \frac{S(a_j)}{S(a_{j-1})} \end{aligned}$$

The **conditional** likelihood for interval I_j is

$$L_j(\boldsymbol{\theta}) \quad \propto \quad [q_j]^{d_j} \cdot [p_j^\star]^{W_j} \cdot [p_j]^{Y_j - d_j - W_j} \qquad (2.6\ .11)$$

where Y_j is the number of policies at risk of lapsing in I_j, that is policies that are still alive at a_{j-1}.

The overall likelihood function is

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{k} L_j(\boldsymbol{\theta}) \qquad (2.6\ .12)$$

If class intervals are narrow, another possibility is to treat the data as continuous and assume that all lifetimes in interval I_j occur at the interval midpoint.