UNIVERSITY OF CALIFORNIA
SANTA CRUZ

BAYESIAN INFERENCE FOR MEAN RESIDUAL LIFE FUNCTIONS IN SURVIVAL ANALYSIS

A project document submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

STATISTICS AND APPLIED MATHEMATICS

by

Valerie A. Poynor

December 2010

This document is approved:

____
Associate Professor Athanasios Kottas, Chair

____
Associate Professor Raquel Prado

Table of Contents

List of Figures

List of Tables

Abstract

Bayesian Inference for Mean Residual Life Functions in Survival Analysis

by

Valerie A. Poynor

In survival analysis interest lies in modeling data that describe the time to a particular event (e.g., failure of a machine or relapse of a patient). Informative functions, namely the hazard function and mean residual life function, can be obtained from the model's distribution function. We focus on the mean residual life function which provides the expected remaining life given that the subject has survived (i.e., is event-free) up to a particular time. This function is of interest in reliability, medical, and actuarial fields. The mean residual life function not only has a simple and practical interpretation, it characterizes the distribution through the Inversion Formula. Thus the mean residual life function can be used in fitting a model to the data. We review the key properties of the mean residual life function and investigate its form for some common distributions. We also study Bayesian nonparametric inference for mean residual life functions obtained from a flexible mixture model for the corresponding survival distribution. In particular, we develop Markov Chain Monte Carlo posterior simulation methods to fit a nonparametric lognormal Dirichlet process mixture model to two experimental groups. To illustrate the practical utility of the nonparametric mixture model, we compare with an exponentiated Weibull model, a parametric survival distribution that allows various shapes for the mean residual life function.

To my family and friends,

for their support, encouragement, and love.

Acknowledgments

I want to thank my mentors and colleagues, who have provided me with the guidance and tools necessary in attaining this degree.

Chapter 1

Introduction

Survival data are data that describe the time to a particular event. This event is usually referred to as the failure of some machine or death of a person. However, survival data can also represent the time until a cancer patient relapses or time until another infection occurs in burn patients. The survival function of a positive random variable X defines the probability of survival beyond time x.

$$S(x) \quad = \quad Pr(X > x) = 1 - F(x)$$

where $F(x)$ is the distribution function. The hazard rate function computes the probability of a failure in the next instant given survival up to time x,

$$h(x) \quad = \quad \lim_{\Delta x \to 0} \frac{Pr[x < X \leq x + \Delta x | X > x]}{\Delta x} \overset{\text{when X is continuous}}{=} \frac{f(x)}{S(x)}$$

where $f(x)$ is the probability density function. The mean residual life (mrl) function computes the expected remaining survival time of a subject given survival up to time x. Suppose that $F(0) = 0$ and $\mu \equiv E(X) = \int_0^\infty S(x)dx < \infty$. Then the mrl function for continuous X is defined as:

$$m(x) \quad = \quad E(X-x|X>x) = \frac{\int_x^\infty (t-x)f(t)dt}{S(x)} = \frac{\int_x^\infty S(t)dt}{S(x)} \tag{1.1}$$

and $m(x) \equiv 0$ whenever $S(x) = 0$. The mrl function is of particular interest because of its easy interpretability and large area of application [9]. Moreover, it characterizes the survival distribution via the Inversion Formula (1.2). Again for continuous X with finite mean, the survival function is defined through the mrl function:

$$S(x) = \frac{m(0)}{m(x)} exp\left[-\int_0^x \frac{1}{m(t)}dt\right]. \tag{1.2}$$

One point of interest is study of the form for the mrl function of various distributions. Along with discussion on some key probabilistic properties and defining characteristics of the mrl function, we also investigate its form under a number of common distributions. We find that the shape of the mrl function is often quite limited to monotonically increasing (INC) or decreasing (DCR) functions, which may be appropriate for some situations (e.g., the data example of Section 3.4), but not suitable for other populations. For instance, biological lifetime data tend to support lower mrl during infancy and elderly age while there is a higher mrl during the middle ages. The shape of this mrl function is unimodal and commonly referred to as upside-down bathtub (UBT) shape. There have been many papers that have investigated the form of the mrl function in relation to the hazard function. A well-known relationship for monotonically increasing (decreasing) hazard functions is that the corresponding mrl function will be monotonically decreasing (increasing); see Finkelstein [6] for a review. Gupta and Akman [10] establish sufficient conditions for the mrl function to be decreasing (increasing) or UBT (BT) given that the hazard is BT (UBT). Xie et.al. [24] look at the specific change points of mrl function and hazard function. These are just a few examples of the literature on the shape of mrl functions.

Another point of interest lies in inference for the mrl function. There is some literature on inference for the mrl function using nonparametric empirical estimators, as well as parametric maximum likelihood estimates, for settings that may include regression covariates and censoring (related references are given in Chapter 3). We are interested in inference of the mrl function under a Bayesian framework. The literature in this area is quite limited. In Chapter 3, we compare the mrl functions of two experimental groups under an exponentiated Weibull model [20], as well as using a nonparametric lognormal Dirichlet Process (DP) mixture model. In addition to making inference on mrl functions for the two groups, we also perform model comparison, which supports the greater flexibility of the nonparametric mixture model. In Chapter 4 we summarize our findings and discuss future areas of study under this framework.

Notation:

X: a non-negative continuous random variable representing survival time

$F(x)$: the distribution function of X

$f(x)$: the probability density function of X

$S(x)$: denotes the survival function

$h(x)$: denotes the hazard function

$m(x)$: denotes the mean residual life (mrl) function

$T \equiv inf\{x : F(x) = 1\} \leq \infty$

Chapter 2

Mean Residual Life Functions: Theory and Properties

In this chapter we review some important properties and characteristics of the mean residual life function and provide the form of the mrl function for several common distributions. We begin with some elementary properties that are well-established in the literature. These properties will either lead to the development of the Inversion Formula (1.2) or become of more interest once the Inversion Formula is provided. We close the first section by stating the Characterization Theorem (e.g., Hall and Wellner [13]). The second section utilizes various forms of the definition of the mrl function along with convenient transformations to study the various shapes of the mrl function for a number of commonly used distributions.

2.1 Probabilistic Properties

2.1.1 Elementary Identities

We start out by showing an elementary relationship between the survival function and the moments of the distribution. Klein and Moeschberger [16] state that for a continuous random variable taking non-negative values and having finite mean, then $\mu \equiv E(X) = \int_0^\infty x f(x) dx = \int_0^\infty S(x) dx$.

$$
\begin{aligned}
E(X) &= \int_0^\infty x f(x) dx \quad \left(\begin{matrix} \text{using integration by parts with:} \\ u = x, du = dx, dv = f(x), v = -S(x) \end{matrix} \right) \\
&= [-xS(x)]_0^\infty - \int_0^\infty -S(x) dx \\
&= \underbrace{-\lim_{x \to \infty} xS(x)}_{\text{goes to 0 (see discussion below)}} + 0S(0) + \int_0^\infty S(x) dx \\
&= \int_0^\infty S(x) dx
\end{aligned}
$$

where the limit as x goes to infinity of $xS(x)$ is 0, since we assume a finite mean $\left(\int_0^\infty t f(t) dt < \infty\right)$ and continuous distribution function. In general, the distribution function need only be right continuous with finite mean for the limit to be 0. Our argument follows: for a right continuous distribution, the survival function is defined as $S(x) = \int_x^\infty f(t) dt \Rightarrow xS(x) = x \int_x^\infty f(t) dt$ (note that the integral above can easily be broken into a sum of integrals for right continuous distributions containing a jump in the density function). Since x and $f(x)$ are both nonnegative, we have $0 \le x \int_x^\infty f(t) dt \le \int_x^\infty t f(t) dt$. Applying the limit to each expression, $\lim_{x\to\infty} 0 \le \lim_{x\to\infty} x \int_x^\infty f(t) dt \le \underbrace{\lim_{x\to\infty} \int_x^\infty t f(t) dt}_{\text{Goes to zero with finite mean}} \Rightarrow 0 \le \lim_{x\to\infty} x \int_x^\infty f(t) dt \le 0$, so by

Squeeze Theorem, $\lim_{x\to\infty} xS(x) = 0$.

The second moment can also be written as a function of the survival function. Assuming the existence of the 2nd moment, we can write,

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2 f(x)dx = \left[-x^2 S(x)\right]_0^\infty - \int_0^\infty -2xS(x)dx \\ &= \underbrace{-\lim_{x\to\infty}(x^2 S(x))}_{\text{goes to 0 (see below)}} + 0S(0) + 2\int_0^\infty xS(x))dx \\ &= 2\int_0^\infty xS(x)dx \end{aligned}$$

Again, assuming the existence of the second moment $\left(\int_0^\infty x^2 f(x)dx < \infty\right)$, for continuous (at least right continuous) distribution function, we can write $x^2 S(x) = x^2 \int_x^\infty f(t)dt \Rightarrow 0 \le x^2 \int_x^\infty f(t)dt \le \int_x^\infty t^2 f(t)dt$. Applying the limit to each expression, $\lim_{x\to\infty} 0 \le \lim_{x\to\infty} x^2 \int_x^\infty f(t)dt \le \underbrace{\lim_{x\to\infty}\int_x^\infty t^2 f(t)dt}_{\text{Goes to zero with finite 2nd moment}} \Rightarrow 0 \le \lim_{x\to\infty} x^2 \int_x^\infty f(t)dt \le 0$, again by Squeeze Theorem, $\lim_{x\to\infty} x^2 S(x) = 0$.

In general, if the r^{th} moment exists for a continuous random variable X we have the following expression:

$$E(X^r) = r\int_0^\infty x^{r-1} S(x)dx \tag{2.1}$$

This expression is of interest for us, because once we establish the Inversion Formula (1.2), we have a way of obtaining the moments (when they exist) from the mrl function. Additionally, we have an expression for the variance in terms of the survival function:

$$Var(X) = E(X^2) - E^2(X) = 2\int_0^\infty xS(x)dx - \left[\int_0^\infty S(x)dx\right]^2$$

We have already defined the mrl as the expectation of the remaining survival time given survival up to time x. Here we derive the expression for the mrl function through the survival function [23] as stated in (1.1),

$$\begin{aligned} m(x) &= E(X-x|X>x) = \int_x^\infty (t-x)dP(X\le t|X>x) \\ &= \int_x^\infty (t-x)d\left(\frac{F(t)-F(x)}{1-F(x)}\right) = \int_x^\infty (t-x)d\left(\frac{-S(t)+S(x)}{S(x)}\right) \\ &= \int_x^\infty (t-x)\left(d\left[\frac{-S(t)}{S(x)}\right]+d[1]\right) = \int_x^\infty (t-x)\left(\frac{-S'(t)dt}{S(x)}\right) \\ &= \frac{(t-x)S(t)|_x^\infty + \int_x^\infty S(t)dt}{S(x)} = \frac{\lim_{t\to\infty}(t-x)S(t)-(x-x)S(x)+\int_x^\infty S(t)dt}{S(x)} \\ &= \frac{\int_x^\infty S(t)dt}{S(x)} \end{aligned}$$

where the first limit in the last step tends to 0 since we assume that the first moment exists, and the second limit tends to 0 since $F(\infty)=1$. It is now easily seen that the first moment is equivalent to the mrl function at $x=0$.

$$m(0) = \frac{\int_0^\infty (t-0)f(t)dt}{S(0)} = \frac{\int_0^\infty tf(t)dt}{1} = \mu. \tag{2.2}$$

2.1.2 Bounds for MRL Functions

Hall and Wellner [13] list a series of inequalities that provide bounds for the mrl function. First, we have that $m(x)+x \overset{(i)}{=} E(X|X>x)$, which leads to $(m(x)+x)S(x) \overset{(ii)}{=} E\left(X\cdot 1_{(X>x)}\right) \overset{(iii)}{=} \mu - E\left(X\cdot 1_{(X\le x)}\right)$. It is also true that $E\left(X\cdot 1_{(X>x)}\right) \overset{(iv)}{\le} TS(x), \overset{(v)}{\le} \mu$, and $E\left(X\cdot 1_{(X>x)}\right) \overset{(vi)}{\le} (E(X^r))^{\frac{1}{r}} S(x)^{1-\frac{1}{r}}$, for $r>1$. Also, $E\left(X\cdot 1_{(X\le x)}\right) \overset{(vii)}{\le} xF(x)$, and $E\left(X\cdot 1_{(X\le x)}\right) \overset{(viii)}{\le} (E(X^r))^{\frac{1}{r}} F(x)^{1-\frac{1}{r}}$, for $r>1$. Proofs for these results are provided in Appendix A.1.

Now we are ready to address the following bounds for the mrl function. If F is non-degenerate with mrl, $m(x)$, mean, μ, and $\nu_r \equiv E(X^r) \le \infty$,

(a) $m(x) \leq (T-x)^+$ for all x, with equality iff $F(x) = F(T^-)$ or 1,

(note T^- indicates that we are approaching T from the left)

(b) $m(x) \leq \dfrac{\mu}{S(x)} - x$ for all x with equality iff $F(x) = 0$

(c) $m(x) < \left(\dfrac{\nu_r}{S(x)}\right)^{\frac{1}{r}} - x$ for all x and any $r > 1$

(d) $m(x) \geq \dfrac{(\mu - x)^+}{S(x)}$ for $x < T$ with equality iff $F(x) = 0$

(e) $m(x) > \dfrac{\mu - F(x)\left(\frac{\nu_r}{F(x)}\right)^{\frac{1}{r}}}{S(x)} - x$ for $x < T$ and any $r > 1$

(f) $m(x) \geq (\mu - x)^+$ for all x, with equality iff $F(x) = 0$ or 1

If F is degenerate at μ, $m(x) = (\mu - x)^+$, for all x. Proofs are given in Appendix A.1.

2.1.3 Properties of MRL (Inversion Formula)

The properties stated below are also provided in Hall and Wellner [13], and are essential for the development of the characterization theorem for mrl functions, which is stated at the end of this section.

(a) $m(x)$ is a nonnegative and right-continuous, and $m(0) = \mu > 0$

(b) $v(x) \equiv m(x) + x$ is non-decreasing

(c) $m(x^-) > 0$ for $x \in (0, T)$; if $T < \infty$, $m(T^-) = 0$, and m is continuous at T $\left(m(t^-) \equiv \lim\limits_{x \to t^-} m(x)\right)$

(d) $S(x) = \dfrac{m(0)}{m(x)} exp\left[-\int_0^x \dfrac{1}{m(t)}dt\right]$, for all $x < T$ (Inversion Formula)

(e) $\int_0^x \dfrac{1}{m(t)}dt \to \infty$ as $x \to T$

Property (d) is known as the Inversion Formula (1.2) and is proved below (see Appendix A.2 for proofs of (a),(b),(c), and (e)).

Proof of Inversion Formula: Define the function: $k(x) \equiv \int_x^\infty S(t)dt = m(x)S(x)$. We have $k'(x) = f(x)m(x) - S(x)m'(x)$, with $m'(x) = \frac{S^2(x)+f(x)\int_0^x S(t)dt}{S^2(x)} = 1 + \frac{f(x)m(x)}{S(x)}$, and thus $k'(x) = -S(x)$. Now consider,

$$\int_0^x \frac{1}{m(t)}dt = -\int_0^x \frac{-S(t)}{S(t)}\frac{1}{m(t)}dt = -\int_0^x \frac{k'(t)}{k(t)}dt = -\left[log(k(x)) - log(k(0))\right]$$
$$= -log\left(\frac{k(x)}{k(0)}\right) = -log\left(\frac{S(x)m(x)}{S(0)m(0)}\right) = -log\left(\frac{S(x)m(x)}{m(0)}\right)$$
$$\Rightarrow exp\left(-\int_0^x \frac{1}{m(t)}dt\right) = exp\left(log\left(\frac{S(x)m(x)}{m(0)}\right)\right)$$
$$\Leftrightarrow exp\left(-\int_0^x \frac{1}{m(t)}dt\right) = \left(\frac{S(x)m(x)}{m(0)}\right)$$
$$\Leftrightarrow S(x) = \frac{m(0)}{m(x)}exp\left(-\int_0^x \frac{1}{m(t)}dt\right)$$

We conclude the review of properties for mrl functions with a key result that provides necessary and sufficient conditions such that a function is the mrl function for a survival distribution, and thus it characterizes mrl functions.

Characterization Theorem: Suppose a function $m(x)$ which maps $\mathbb{R}^+ \to \mathbb{R}^+$ satisfies (a) $m(x)$ is right-continuous and $m(0) > 0$; (b) $v(x) \equiv m(x) + x$ is non-decreasing; (c) if $m(x^-) = 0$ for some $x = x_0$, then $m(x) = 0$ for $x \in [x_0, \infty)$; (d) if $m(x^-) > 0$ for all x, then $\int_0^\infty \frac{1}{m(t)}dt = \infty$. Let $T \equiv inf\{x : m(x^-) = 0\} \leq \infty$, and define $S(x)$ by (2.3) for $x < T$ and $S(x) \equiv 0$ for $x \geq T$. Then $F(x) \equiv 1 - S(x)$ is a distribution function on $\mathbb{R}^+$ with $F(0) = 0$, $T_F = T$, finite mean $\mu_F = m(0)$, and mrl function $m_F(x) = m(x)$.

2.2 MRL Functions for Specific Distributions

2.2.1 Linear MRL

Oakes and Dasu [21] focus on linear mrl functions discussed in Hall and Wellner [13]. The key result is that if the mrl function is linear, $m(x) = Ax+B$ $(A > -1, B > 0)$, then by use of the Inversion Formula (1.2), the survival function has the form:

$$S(x) = \left[\frac{B}{Ax+B}\right]_+^{\frac{1}{A}+1} \tag{2.3}$$

We show the arrival to this survival form when $A \neq 0$ below:

$$S(x) = \left(\frac{B}{Ax+B}\right) exp\left[-\int_0^x \frac{1}{At+B}dt\right] = \left(\frac{B}{Ax+B}\right) exp\left[-\frac{1}{A}ln(At+B)\right]_0^x$$
$$= \left(\frac{B}{Ax+B}\right) \frac{exp\left[ln(Ax+B)^{-\frac{1}{A}}\right]}{exp\left[ln(B)^{-\frac{1}{A}}\right]} = \left(\frac{B}{Ax+B}\right)\left(\frac{B}{Ax+B}\right)^{\frac{1}{A}} = \left(\frac{B}{Ax+B}\right)_+^{\frac{1}{A}+1}$$

where the positive part is necessary to satisfy the nonnegative property of the survival function. For $A > 0$ the survival function is a Pareto distribution. The form of the survival function of the Pareto distribution for random variable Z is,

$$S(z) = \left(\frac{\beta}{z}\right)^{\alpha} \text{ for } \beta > 0 \ \text{ (scale)}, \ \alpha > 0 \ \text{ (shape), and } z \in [\beta, +\infty]$$

If we consider the transformation $Z = AX + B$ where $B = \beta$ and $\frac{1}{A} + 1 = \alpha$, then we have $Z \sim Pareto(\alpha, \beta)$. To clarify, we know that $\beta > 0$ is satisfied from $B = \beta$ with $B > 0$. We also know that the shape parameter satisfies $\alpha > 1$ from $\frac{1}{A} + 1 = \alpha$ and $\frac{1}{A} > 0$. Note that the first moment only exists for the Pareto distribution when $\alpha > 1$ therefore the mean of survival function exists for linear mrl with $A, B > 0$. The support is given by $z \in [\beta, +\infty]$ since $z = Ax+B$ and $Ax \geq 0$. Finally, since $Z \geq \beta > 0 \Rightarrow \frac{\beta}{z} > 0$

the survival function is always positive, so no precautions need be made with taking only the positive part of the function.

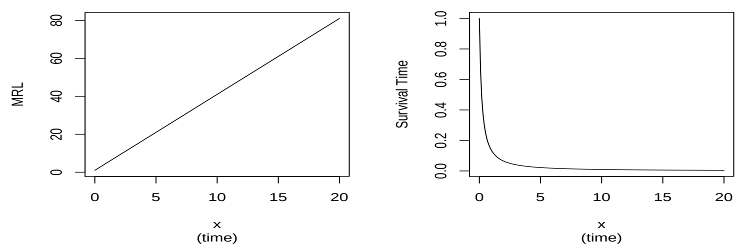

Figure 2.1: (left) Linear mrl for X with $A = 4$ (slope) and $B = 1$ (intercept). (right) Corresponding survival function of X.

For $-1 < A < 0$ the survival function is a rescaled beta distribution. The pdf of a rescaled beta distribution is given by

$$f(z; a, b, p, q) = \frac{(z-a)^{p-1}(b-z)^{q-1}}{\mathbf{B}(p,q)(b-a)^{p+q+1}}$$

where $a \leq z \leq b, \quad p, q > 0$, and $\mathbf{B}(.,.)$ is the beta function defined as $\mathbf{B}(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}\,dt$. Start with the form of the survival function from the linear mrl to obtain the pdf. The pdf will reveal what type of reparameterization yields the form of the rescaled beta. Start with $S(x) = \left[\frac{B}{Ax+B}\right]_+^{\frac{1}{A}+1}$, note: that the positive part is obtained when $-Ax \leq B \rightarrow x \leq -B/A$. Then $F(x) = 1 - \left[\frac{B}{Ax+B}\right]^{\frac{1}{A}+1}$. Thus we have,

$$\begin{aligned} f(x) &= -\left(\frac{1}{A}+1\right)\left[\frac{B}{Ax+B}\right]^{\frac{1}{A}}\left[\frac{AB}{(Ay+B)^2}\right] = -\frac{\left(\frac{1}{A}+1\right)AB^{\frac{1}{A}+1}}{(Ax+B)^{\left(\frac{1}{A}+1\right)+1}} \\ &= -\frac{A\,(Ax+B)^{-\left(\frac{1}{A}+1\right)-1}}{\left(\frac{1}{A}+1\right)^{-1}B^{-\left(\frac{1}{A}+1\right)}}, \text{ Let } Z = -AX \Rightarrow \frac{dx}{dz} = -\frac{1}{A} \\ \Rightarrow f(z) &= \frac{+\frac{A}{A}\,(B-z)^{-\left(\frac{1}{A}+1\right)-1}}{\left(\frac{1}{A}+1\right)^{-1}B^{-\left(\frac{1}{A}+1\right)}} \text{ (with } \overset{q=-\left(\frac{1}{A}+1\right))}{=} \frac{(B-y)^{q-1}}{\left(-\frac{1}{q}\right)B^{-q}} \end{aligned}$$

Now we can see that it is necessary for $B = b, a = 0, p = 1$. When $p = 1 \Rightarrow \mathbf{B}(p = 1, q) = \int_0^1 (1-t)^{q-1} dt = -\left(\frac{1}{q}\right)$ we have, $f(z) = \frac{(z-0)^{1-1}(b-z)^{q-1}}{\mathbf{B}(1,q)(b-0)^{q+1-1}} \quad 0 \leq z \leq b$. Then the cdf and survival functions are given by,

$$\begin{aligned} F(z) &= \int_0^z \frac{(t-0)^{1-1}(b-t)^{q-1}}{\mathbf{B}(1,q)(b-0)^{q+1-1}} dt = \frac{\int_0^z (b-t)^{q-1}\, dt}{\mathbf{B}(1,q) b^q} \\ &= \frac{-\frac{1}{q}\left[(b-z)^q - b^q\right]}{-\left(\frac{1}{q}\right) b^q} = \frac{\left[(b-z)^q - b^q\right]}{b^q} = \left(\frac{b-z}{b}\right)^q - 1 \\ \Rightarrow S(z) &= \left(\frac{b-z}{b}\right)^q = \left(\frac{b}{b-z}\right)^{-q} \end{aligned}$$

which is precisely the transformed survival function.

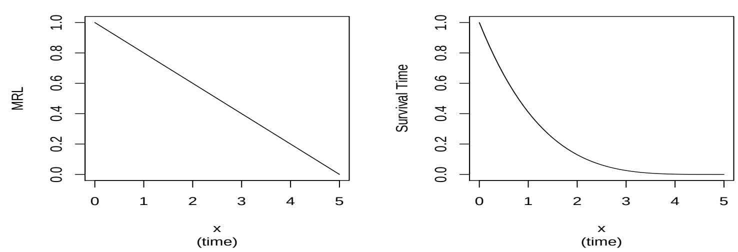

Figure 2.2: (left) Linear mrl for X with $A = -0.2$ (slope) and $B = 1$ (intercept). (right) Corresponding survival function of X.

For $A = 0$, the survival function is exponential: $S(x) = \left(\frac{B}{B}\right) exp\left[-\int_0^x \frac{1}{B} dt\right] = e^{-\frac{1}{B}x}$

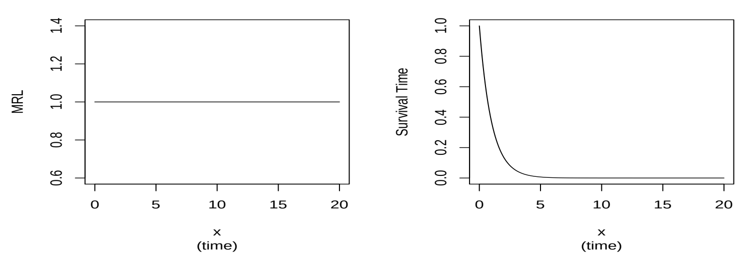

Figure 2.3: (left) Linear mrl for X with $A = 0$ (slope) and $B = 1$ (intercept). (right) Corresponding survival function of X.

Figures 2.1, 2.2, and 2.3 above show both the linear mrl function and the resulting survival function using a value of the slope parameter (A) from each of the three domains previously mentioned. The intercept was set as $B = 1$ for all three forms to make the role of the slope parameter more obvious. Note that aside from the exponential survival function, where no transformation is necessary, both the mrl and survival functions are of the original survival times X rather than the transformed survival times that are expected to follow well-defined distributions (Pareto and rescaled beta). Turning now to the domains of the survival functions, recall from above that for $A \leq 0$ X can take on values from 0 to infinity, however, for $-1 < A < 0$ X goes from 0 to $-B/A$. In our example, the domain for the survival function when $A = -0.2$ and $B = 1$ is $[0, 5]$. We can see that by increasing the magnitude of A the domain can get very narrow.

2.2.2 The Form of the Mean Residual Life for Some Common Distributions

In this section we summarize our investigation of the forms of mrl functions for a number of common distributions. In the previous section, we discussed the distributions having a linear mean residual life function namely the exponential, Pareto, and rescaled beta. These distributions share the convenient feature that they yield a closed form for the mrl function. On the other hand, the linearity of the mrl is too limiting to be of much practical use. There are a number of distributions having more flexible mrl functions, such as increasing and decreasing curvatures as well as BT or UBT shapes.

The difficulty for these distributions lies in obtaining a closed form of the mrl. Recall from (1.1) that the mrl is defined as

$$m(x) \quad = \quad \frac{\int_x^\infty x(t-x)f(t)dt}{S(x)} = \frac{\int_x^\infty S(t)dt}{S(x)}$$

Alternatively, the mrl can be written as,

$$m(x) \quad = \quad \frac{\int_x^\infty x(t-x)f(t)dt}{S(x)} = \frac{\int_x^\infty tf(t)dt}{S(x) - x\int_x^\infty tf(t)dt}S(x) = \frac{\int_x^\infty tf(t)dt}{S(x)} - x \quad (2.4)$$

$$m(x) \quad = \quad \frac{\int_x^\infty S(t)dt}{S(x)} = \frac{\int_0^\infty S(t)dt - \int_0^x S(t)}{S(x)} \overset{(2.2)}{=} \frac{\mu - \int_0^x S(t)}{S(x)} \quad (2.5)$$

Govilt and Aggarwal [8] derive (2.4) by starting with $\int_x^\infty f(t)dt$ and applying integration by parts and solving for $\int_x^\infty S(t)dt$ to obtain $\int_x^\infty S(t)dt = \int_x^\infty tf(t)dt - xS(x)$. Dividing both sides by S(x) results in the survival distribution form of the mrl function. This derivation requires that $xS(x) \to 0$ as $x \to \infty$. As stated in Section 2.1.1, this limit converges to 0 as long as the distribution function is right continuous and has finite mean. The distributions that we discuss meet these requirements. The mrl can also be obtained, perhaps more directly, from the first equality stated in Section 2.1.2 (i) by subtracting x from both sides.

The distributions discussed here have no known closed form for their associated mrl making them difficult to explore. However, through the use of (2.4) or (2.5) and/or simple transformations of X, we are able to obtain forms of the mrl functions comprised of well-known integrals. Although these forms are far from an ideal closed form, they are easy to evaluate with most statistical programming software.

Gamma Distribution

The survival function of the gamma distribution has no closed form, therefore we will work with (2.4) to obtain the mrl. The pdf of the gamma distribution with shape parameter α and scale parameter λ is given by

$$f(x) = \frac{x^{\alpha-1} exp\left[-\frac{x}{\lambda}\right]}{\lambda^{\alpha}\Gamma(\alpha)} \text{ with } \Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}e^{-t}dt$$

The numerator in (2.4) is simplified as follows:

$$\int_x^{\infty} tf(t)dt = \int_x^{\infty} t\left(\frac{x^{\alpha-1} exp\left[-\frac{x}{\lambda}\right]}{\lambda^{\alpha}\Gamma(\alpha)}\right)dt = \frac{1}{\Gamma(\alpha)}\int_x^{\infty}\left(\frac{t}{\lambda}\right)^{\alpha} exp\left[-\frac{t}{\lambda}\right]dt$$

Under the integration by parts with the follows substitutions: $u = \left(\frac{t}{\lambda}\right)^{\alpha}$ and $dv = exp\left[-\frac{t}{\lambda}\right]dt$, then $du = \frac{\alpha}{\lambda}\left(\frac{t}{\lambda}\right)^{\alpha-1}dt$ and $v = -\lambda exp\left[-\frac{t}{\lambda}\right]$. The numerator is equivalent to,

$$\frac{1}{\Gamma(\alpha)}\left(-\left(\frac{1}{\lambda}\right)^{\alpha-1} t^{\alpha} exp\left[-\frac{t}{\lambda}\right]|_x^{\infty(*)}\right) + \lambda\alpha\int_x^{\infty}\frac{1}{\Gamma(\alpha)}t^{\alpha-1}\left(\frac{1}{\lambda}\right)^{\alpha} exp\left[-\frac{t}{\lambda}\right]$$
$$= \frac{1}{\Gamma(\alpha)}\left(\frac{x}{\lambda}\right)^{\alpha-1} x^{\alpha} exp\left[-\frac{x}{\lambda}\right] + \lambda\alpha\underbrace{\int_x^{\infty} f_{\mathbf{X}}(t)dt}_{S_{\mathbf{X}}(x)}$$

Returning this expression to the numerator in (2.4), the mrl function is given by,

$$m(x) = \frac{x^{\alpha} exp\left[-\frac{x}{\lambda}\right]}{\lambda^{\alpha-1}\Gamma(\alpha)S_{\mathbf{X}}(x)} + \lambda\alpha - x \tag{2.6}$$

where by repeated use of L'Hopital's Rule (*) goes to 0.

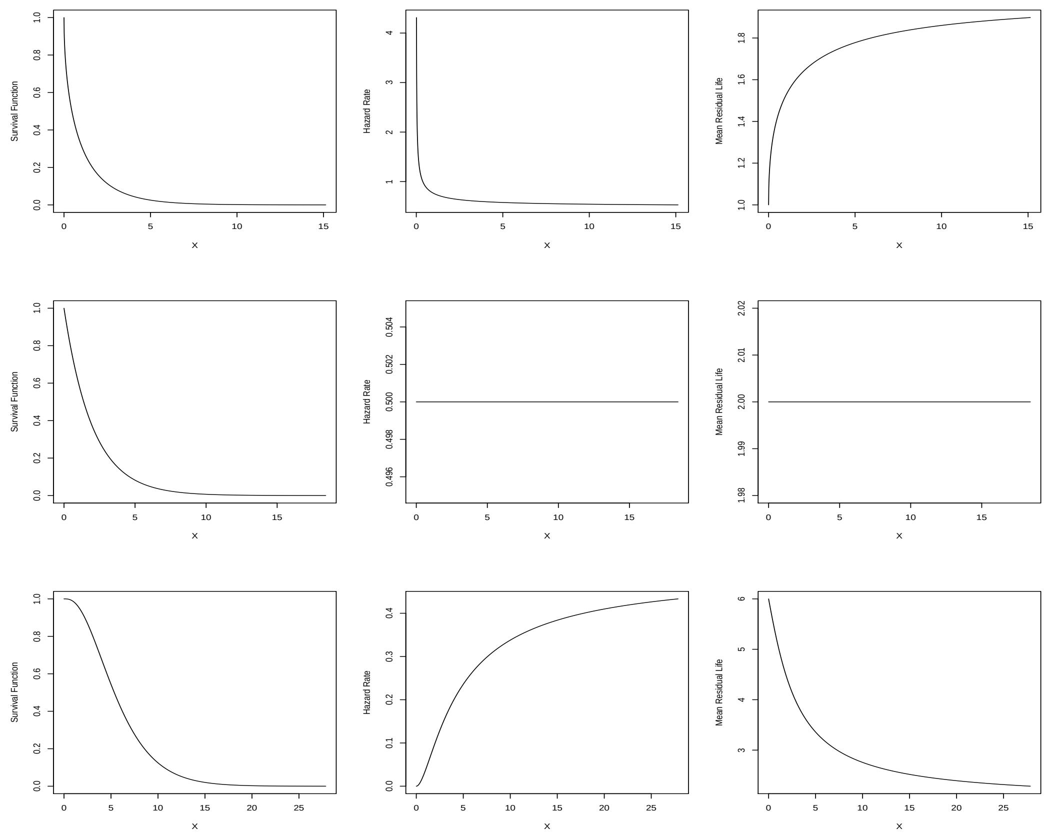

Figure 2.4: (Top) Gamma distribution with shape 0.5 and scale 2. (Middle) Gamma distribution with shape 1 and scale 2. (Bottom) Gamma distribution with shape 3 and scale 2.

In Figure 2.4, the survival function (left), hazard rate function (center), and mrl function (right) for three different values of the shape parameter. When the shape parameter is < 1 (we use 0.5, see top row), the hazard rate function is monotone decreasing and the mrl function is monotone increasing. For shape parameter $= 1$ (middle row), the hazard and mrl functions are constant at the rate (1/scale) $= 1/2$ and scale $= 2$, respectively. For shape parameter > 1 (we use 3, see bottom row), the hazard rate function in monotone decreasing and the mrl function is monotone increasing. The scale parameter does not play a role in the shape of the hazard or mrl function, so was kept constant.

Gompertz Distribution

The Gompertz distribution with shape and scale parameters $\alpha, \lambda > 0$ respectively has survival function

$$\begin{aligned} S(x) &= exp\left[\frac{\lambda}{\alpha}\left(1 - e^{\alpha x}\right)\right] \\ \Rightarrow \int_x^\infty S(t)dt &= \int_x^\infty exp\left[\frac{\lambda}{\alpha}\left(1 - e^{\alpha t}\right)\right] dt = e^{\lambda/\alpha} \int_x^\infty exp\left[-\frac{\lambda}{\alpha} e^{\alpha t}\right] dt \end{aligned}$$

If we let $z(t) = z = \frac{\lambda}{\alpha} e^{\alpha t}$, then $t = \frac{1}{\alpha} ln\left[\frac{\lambda}{\alpha} z\right] \Rightarrow dt = \frac{1}{\alpha}\left(\frac{1}{z}\right) dz$. Substituting back into the survival function provides,

$$\begin{aligned} S(x) &= e^{\lambda/\alpha}\left(\frac{1}{\alpha}\right) \int_{z(x)}^\infty z^{-1} e^{-z} dz = e^{\lambda/\alpha}\left(\frac{1}{\alpha}\right) \Gamma_{inc}(0, z(x)) \\ & \quad \text{where } \Gamma_{inc}(a, x) = \int_x^\infty t^{a-1} e^{-t} dt \text{ where } x, a \geq 0 \\ \Rightarrow m(x) &= \frac{e^{\lambda/\alpha}\left(\frac{1}{\alpha}\right) \Gamma_{inc}(0, z(x))}{exp\left[\frac{\lambda}{\alpha}\left(1 - e^{\alpha x}\right)\right]} = e^{z(x)}\left(\frac{1}{\alpha}\right) \Gamma_{inc}(0, z(x)) \qquad (2.7) \\ & \quad \left(\text{where } \; z(x) = \frac{\lambda}{\alpha} e^{\alpha x}\right) \end{aligned}$$

The Gompertz distribution has only monotone increasing hazard rate function and decreasing mrl function. In Figure 2.5, the survival function (left), hazard rate function (middle), and mrl function (right) are shown under a shape parameter value of 3 and scale parameter value of 0.5.

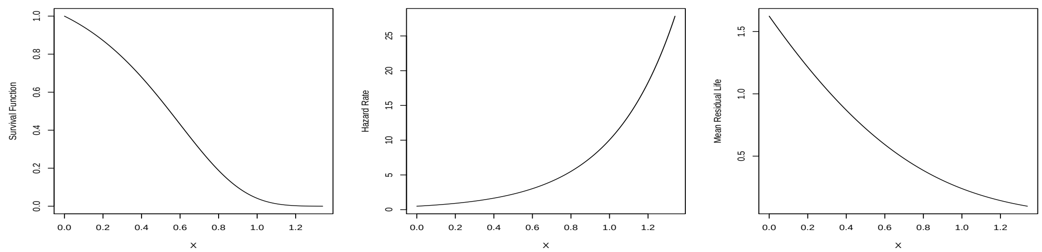

Figure 2.5: Gompertz distribution with shape parameter 3 and scale parameter 0.5

Log-logistic Distribution

The Survival Function for the log-logistic distribution with shape and scale parameters $\alpha, \lambda > 0$ respectively is given by,

$$S(x) = \left[1 + \left(\frac{x}{\lambda}\right)^{\alpha}\right]^{-1}$$

The mean of the log-logistic distribution is only finite when the shape parameter is greater than 1, thus the mrl is only defined when $\alpha > 1$. The mrl for the log-logistic distribution is easily obtained from by simplifying (1.1) as is done by Gupta, Akman, and Lvin [11]. The numerator in (1.1) is defined as $\int_x^{\infty} S(t)dt = \int_x^{\infty}\left[1 + \left(\frac{x}{\lambda}\right)^{\alpha}\right]^{-1}$. Let $z(t) = z = \frac{\left(\frac{t}{\lambda}\right)^{\alpha}}{1+\left(\frac{t}{\lambda}\right)^{\alpha}}$. Then $t = \lambda\left(\frac{z}{1-z}\right)^{\frac{1}{\alpha}}$ and $dt = \frac{\lambda}{\alpha}\left(\frac{z}{1-z}\right)^{\frac{1}{\alpha}-1}\left(\frac{1}{(1-z)^2}\right)dz$. Applying the transformation, the integral becomes,

$$\begin{aligned}
&= \int_{z(x)}^{\lim_{t\to\infty} z(t)} \left[1 + \frac{z}{1-z}\right]^{-1}\left(\frac{\lambda}{\alpha}\right)\left(\frac{z}{a-z}\right)^{\frac{1}{\alpha}-1}(1-z)^{-2}dz \\
&= \left(\frac{\lambda}{\alpha}\right)\int_{z(x)}^{1}(1-z)^{\left(1-\frac{1}{\alpha}\right)-1} z^{\frac{1}{\alpha}-1}dz \\
&= \left(\frac{\lambda}{\alpha}\right)\Gamma\left(1-\frac{1}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)\underbrace{\int_{z(x)}^{1}\frac{\Gamma\left(1-\frac{1}{\alpha}+\frac{1}{\alpha}\right)}{\Gamma\left(1-\frac{1}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)}(1-z)^{\left(1-\frac{1}{\alpha}\right)-1} z^{\frac{1}{\alpha}-1}dz}_{\text{survival function of a beta}}
\end{aligned}$$

$$m(x) = \frac{\left(\frac{\lambda}{\alpha}\right)\Gamma\left(1-\frac{1}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)S_{\mathbf{Z}}\left(z(x);\text{shape}=1-\frac{1}{\alpha},\text{scale}=\frac{1}{\alpha}\right)}{S_{\mathbf{X}}(x)}$$

$$= \left(\frac{\lambda}{\alpha}\right)\Gamma\left(1-\frac{1}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)S_{\mathbf{Z}}\left(z(x);1-\frac{1}{\alpha},\frac{1}{\alpha}\right)\left(1+\left(\frac{x}{\lambda}\right)^{\alpha}\right) \quad (2.8)$$

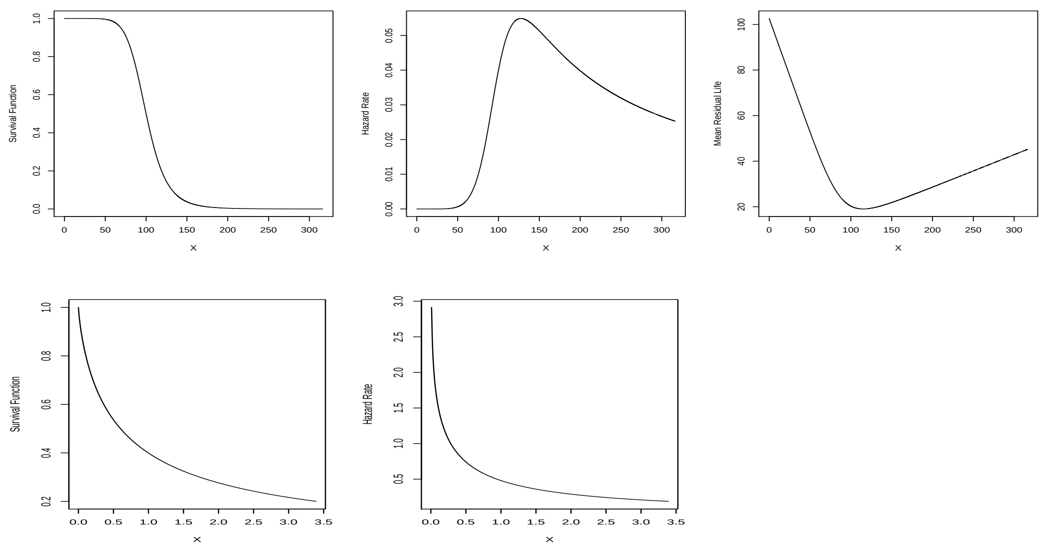

Figure 2.6: Loglogistic Distribution with shape parameter 8 and scale parameter 100. (bottom) Loglogistic Distribution with shape parameter 0.8 and scale parameter 0.6

The loglogistic distribution provides an UBT shape for the hazard rate function with corresponding BT mrl function when the shape parameter is greater than 1, see the top row in Figure 2.6. However, this is the only shape that the distribution offers for the mrl. When the shape parameter is less than or equal to 1 (bottom of Figure 2.6), the hazard rate function is decreasing, but the mrl function is undefined.

Log-Normal Distribution

The log-normal distribution falls in with those distributions having no closed form for the survival function, so (2.4) will be used to obtain the mrl function. The pdf of a lognormal is given by,

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2}\left(\frac{ln(x)-\mu}{\sigma}\right)^2\right]$$

The cdf is given by $F(x) = \int_0^x \frac{1}{t\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2}\left(\frac{ln(t)-\mu}{\sigma}\right)^2\right] dt = \Phi\left(\frac{ln(x)-\mu}{\sigma}\right)$, so the survival function is $S(x) = 1 - \Phi\left(\frac{ln(x)-\mu}{\sigma}\right)$. Working from (2.5) the numerator is $\int_x^\infty tf(t)dt = \frac{1}{\sqrt{2\pi}}\int_x^\infty \frac{1}{t\sigma} exp\left[-\frac{1}{2}\left(\frac{ln(t)-\mu}{\sigma}\right)^2\right] dt$. Let $z(t) = z = \frac{ln(t)-\mu}{\sigma}$, then $t = exp\left[z\sigma+\mu\right]$ and $dt = \sigma exp\left[z\sigma+\mu\right]dz$. The numerator becomes,

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi}}\int_{z(x)}^{\infty} exp\left[-\frac{1}{2}z^2 + z\sigma + \mu\right] dz \\
&= \frac{1}{\sqrt{2\pi}} e^{\left(\mu+\frac{\sigma^2}{2}\right)} \int_{z(x)}^{\infty} exp\left[-\frac{1}{2}(z-\sigma)^2\right] dz \\
&= e^{\left(\mu+\frac{\sigma^2}{2}\right)}\left[1-\Phi\left(\frac{ln(x)-(\mu+\sigma^2)}{\sigma}\right)\right] \\
m(x) &= \frac{e^{\left(\mu+\frac{\sigma^2}{2}\right)}\left[1-\Phi\left(\frac{ln(x)-(\mu+\sigma^2)}{\sigma}\right)\right]}{1-\Phi\left(\frac{ln(x)-\mu}{\sigma}\right)} - x \qquad (2.9)
\end{aligned}$$

Contrary to what we have seen thus far, the scale parameter determines the shape of the hazard and mrl functions. In Figure 2.7, we provide the survival (left), hazard (center), and mrl (left) functions under three different values of σ and constant $\mu = 1$, in the lognormal distribution. When $\sigma < 1$ (top), the hazard rate function is increasing and the mrl function is decreasing. When $\sigma = 1$ (middle), the hazard rate has an UBT shape and the corresponding mrl function has a BT shape. For $\sigma > 1$ (bottom), the hazard rate function is decreasing, and the mrl function is increasing.

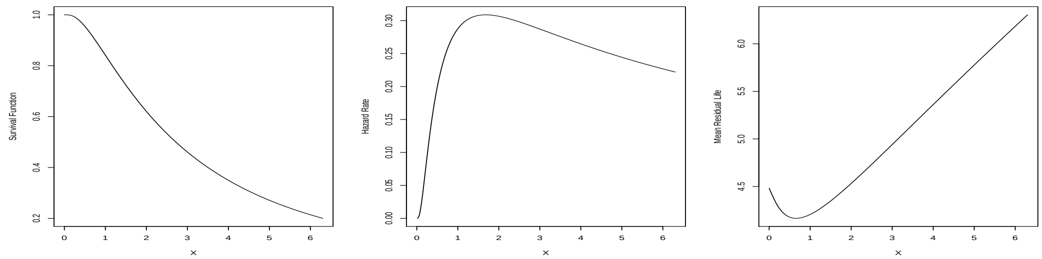

Figure 2.7: Log-normal distribution with location parameter, $\mu = 1$, and scale parameter $\sigma = 1$.

Truncated Normal Distribution

Once again the lack of a closed form for the survival function of the normal distribution requires use of (2.4) to obtain the mrl function. We will extend the results of Govilt and Aggarwal [8] for the standard normal distribution to the general normal distribution with a lower truncation at 0 to fit the non-negative criteria of survival times. Let X follow a truncated normal distribution with mean μ and variance σ^2 and let Y follow a normal distribution with the same mean μ and variance σ^2. Then the cdf of Y is given by,

$$F_{\mathbf{Y}}(y) \quad = \quad \frac{1}{\sqrt{\sigma^2 2\pi}} \int_0^y exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right] dt = \Phi\left(\frac{y-\mu}{\sigma}\right)$$

The density and survival functions are then given by $f_{\mathbf{Y}}(y) = \frac{1}{\sqrt{\sigma^2 2\pi}} exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]$ and $S_{\mathbf{Y}}(y) = 1 - \Phi\left(\frac{y-\mu}{\sigma}\right)$, respectively. The density function of X can be expressed in terms of the normal distribution as $f_{\mathbf{X}}(x) = \frac{f_{\mathbf{Y}}(x)}{1-F_{\mathbf{Y}}(0)} = \frac{f_{\mathbf{Y}}(x)}{c}$ where $c = 1 - F_{\mathbf{Y}}(0)$. The cdf of X can also be written in terms of the normal distribution: $F_{\mathbf{X}}(x) = \int_0^x f_{\mathbf{X}}(t)dt = \frac{1}{c}\int_0^x f_{\mathbf{Y}}(t)dt = \frac{1}{c}F_{\mathbf{Y}}(x) - \frac{1}{c}\underbrace{F_{\mathbf{Y}}(0)}_{1-c} = 1 - \frac{1}{c}(1 - F_{\mathbf{Y}}(x)) = 1 - \frac{1}{c}S_{\mathbf{Y}}(x)$. The survival function of X follows as $S_{\mathbf{X}}(x) = \frac{1}{c}S_{\mathbf{Y}}(x)$. We can write the numerator in (2.5) as,

$$\int_x^\infty tf_{\mathbf{X}}(t)dt \quad = \quad \frac{1}{c}\int_x^\infty tf_{\mathbf{Y}}(t)dt = \int_x^\infty \frac{t}{c\sqrt{\sigma^2 2\pi}} exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right]$$

Let $z(t) = z = \frac{t-\mu}{\sigma}$, then $t = z\sigma + \mu$ and $dt = \sigma dz$. Applying the above transformation, the integral becomes,

$$= \quad \frac{\sigma}{c\sigma\sqrt{2\pi}}\int_{z(x)}^\infty (z\sigma+\mu)e^{-\frac{z^2}{2}}dz = \frac{\sigma}{c\sqrt{2\pi}}\int_{z(x)}^\infty ze^{-\frac{z^2}{2}}dz + \frac{\mu}{c}\underbrace{\frac{1}{\sqrt{2\pi}}\int_{z(x)}^\infty e^{-\frac{z^2}{2}}dz}_{S_{\mathbf{Y}}(z(x))}$$

$$= \quad -\frac{\sigma}{c\sqrt{2\pi}}e^{-\frac{z^2}{2}}|_{z(x)}^\infty + \frac{\mu}{c}S_{\mathbf{Y}}(z(x)) = \frac{\sigma}{c\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}^2\right)} + \frac{\mu}{c}S_{\mathbf{Y}}(z(x))$$

$$m_{\mathbf{X}}(x) \quad = \quad \frac{\frac{\sigma}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}^2\right)} + \mu S_{\mathbf{Y}}(z(x))}{S_{\mathbf{Y}}(x)} - x \tag{2.10}$$

The shape of the hazard rate and mrl functions are especially limited under the truncated normal distribution. The hazard rate function is increasing, and the mrl is decreasing for all values of the parameters (Figure 2.8).

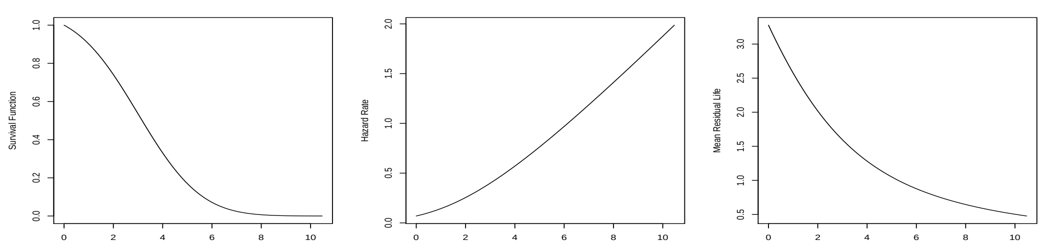

Figure 2.8: Truncated normal distribution with mean, $\mu = 3$, and variance, $\sigma^2 = 4$.

Weibull Distribution

The Weibull distribution is closely related to the gamma distribution. Since the

the mrl defines the distribution it makes sense that we see a relationship between the mrl functions of the the two distributions. The survival function of the Weibull distribution with shape parameter $\alpha > 0$ and scale parameter $\lambda > 0$ is given by $S(x) = exp\left[-\left(\frac{x}{\lambda}\right)^{\alpha}\right]$. Then the numerator in (1.1) becomes $\int_x^{\infty} S(t)dt = \int_x^{\infty} exp\left[-\left(\frac{t}{\lambda}\right)^{\alpha}\right]$. Let $z(t) = z = t^{\alpha}$, then $t = z^{1/\alpha}$ and $dt = \frac{1}{\alpha} z^{\frac{1}{\alpha}-1} dz$. Applying the transformation, the integral becomes,

$$= \frac{1}{\alpha} \int_{z(x)}^{\infty} z^{\frac{1}{\alpha}-1} e^{-\frac{z}{\lambda^{\alpha}}} dz = \frac{1}{\alpha} (\lambda^{\alpha})^{\frac{1}{\alpha}} \Gamma\left(\frac{1}{\alpha}\right) \int_{z(x)}^{\infty} \frac{z^{\frac{1}{\alpha}-1} e^{-\frac{z}{\lambda^{\alpha}}}}{(\lambda^{\alpha})^{\frac{1}{\alpha}} \Gamma\left(\frac{1}{\alpha}\right)}$$

where the last integral is exactly the survival function $S_{\mathbf{Z}}(z(x))$ with $Z \sim \Gamma\left(\text{shape} = \frac{1}{\alpha}, \text{scale} = \lambda^{\alpha}\right)$. Then the mrl is given by,

$$m(x) = \frac{\left(\frac{\lambda}{\alpha}\right) \Gamma\left(\frac{1}{\alpha}\right) S_{\mathbf{Z}}(z(x))}{S_{\mathbf{X}}(x)} \tag{2.11}$$

In Figure 2.9, the survival (left), hazard rate (center), and mrl (right) functions are shown for three different values of the shape parameter. Note the scale parameter does not play a role in determining the shapes of the hazard rate and mrl functions. When the shape parameter is less than 1 (top), the hazard rate function is decreasing, and the mrl function is increasing. For shape parameter equal to 1 (middle), the hazard rate and mrl functions are constant at rate (1/scale) and the scale parameter values, respectively. For shape parameter greater than 1 (bottom), the hazard rate function in increasing with decreasing corresponding mrl function.

Table 2.1, provides a summary of the possible shapes of the hazard rate and mrl functions for the distributions discussed in this section. The table shows how restricted these commonly used distribution are in modeling the mrl function. The gamma and

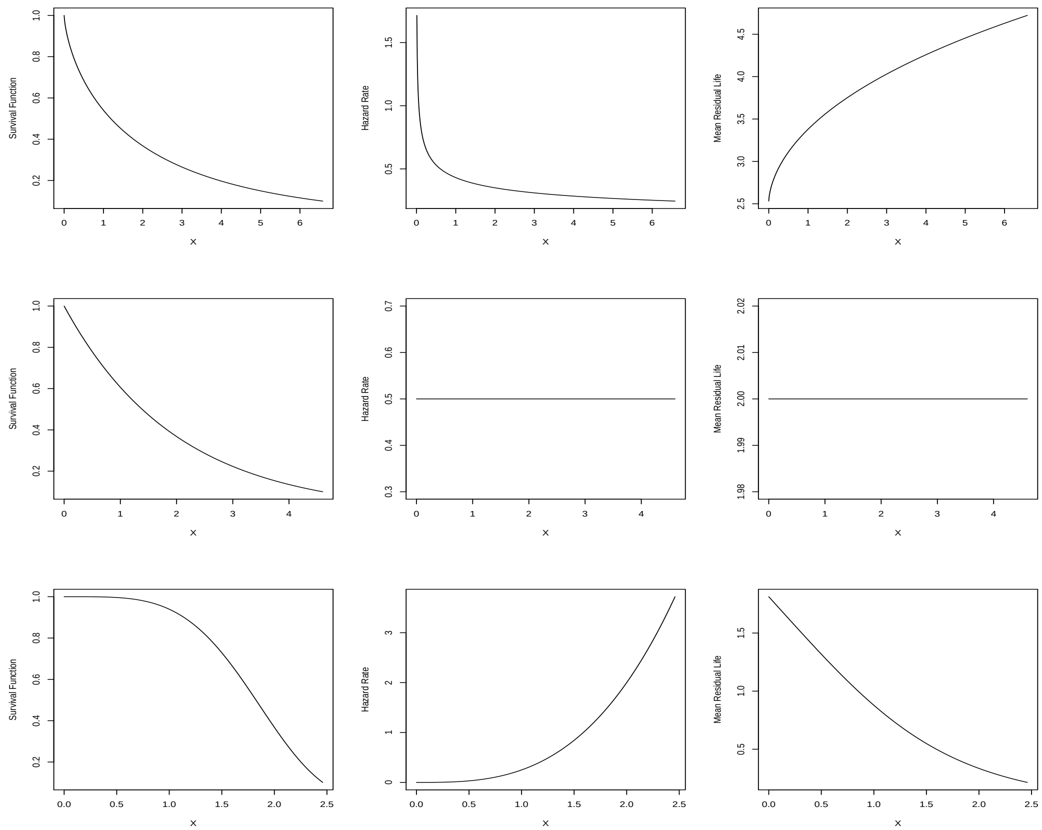

Figure 2.9: (Top) Weibull distribution with shape 0.7 and scale 2. (Middle) Weibull distribution with shape 1 and scale 2. (Bottom) Weibull distribution with shape 4 and scale 2.

Weibull are more versatile they offer three potential shapes for the mrl function, but none of the three shapes consider change points in the mrl function. The loglogistic offers three shapes of the mrl function as well, one being a BT shape, but the UBT shaped mrl function is the more appropriate shape in modeling natural age data. In fact, none of the distribution in Table 2.1 offer an UBT shaped mrl function. Generally speaking, the distributions are restrictive in modeling mrl functions.

Distribution	Hazard Rate	Mean Residual Life
Gamma(α, λ) shape parameter $\alpha > 0$ scale parameter $\lambda > 0$	$\alpha < 1$ DCR $\alpha = 1$ constant $(1/\lambda)$ $\alpha > 1$ INC	$\alpha < 1$ INC $\alpha = 1$ constant(λ) $\alpha > 1$ DCR
Gompertz(α, λ) shape parameter $\alpha > 0$ scale parameter $\lambda > 0$	$\forall \alpha$ INC	$\forall \alpha$ DCR
Loglogistic(α, λ) shape parameter $\alpha > 0$ scale parameter $\lambda > 0$	$\alpha \leq 1$ DCR $\alpha > 1$ UBT	$\alpha \leq 1$ undefined $\alpha > 1$ BT
Lognormal(μ, σ) mean $\mu \in \mathbf{R}$ variance $\sigma^2 > 0$	UBT	BT
Truncated normal(μ, σ) mean $\mu \in \mathbf{R}$ variance $\sigma^2 > 0$	INC	DCR
Weibull(α, λ) shape parameter $\alpha > 0$ scale parameter $\lambda > 0$	$\alpha < 1$ DCR $\alpha = 1$ constant $(1/\lambda)$ $\alpha > 1$ INC	$\alpha < 1$ INC $\alpha = 1$ constant(λ) $\alpha > 1$ DCR

Table 2.1: Summary Table

Chapter 3

Bayesian Inference for MRL Functions

3.1 Literature Review

Interest in estimating the mrl function has been around for many years in the classical survival analysis literature. There has been much development in this area, from the nonparametric empirical estimator for completely observed survival times to semiparametric estimates under regression settings and censored survival times.

The most basic estimator, being the empirical estimate, was first studied in Yang [25]. Yang defines the empirical estimate by $\hat{m}_n(x) = \frac{\int_x^\infty S_n(t)dt}{S_n(x)} \delta_{[0, X_{(n)}]}(x)$ where $S_n(x)$ is the empirical survival function and $X_{(n)}$ is the maximum observed survival time. It is shown that under this fixed finite interval, the estimator is asymptotically unbiased, is uniformly strong consistent, and as n goes to infinity it converges in distribution to a Gaussian process. Hall and Wellner [12] extend Yang's empirical estimator by defining it for values on the entire real line. Furthermore, they provide nonparametric confidence bands for the estimate via transformations of the limiting process of the estimator into Brownian motion. Kochar et. al. [17] modify the empirical estimate for monotonic mrl functions by taking $m_n^*(x) = \delta_{[0,X_{(n)}]}(x) inf_{y \leq x} \hat{m}_n(y)$ for monotonically decreasing mrl and $m_n^{**}(x) = \delta_{[0,X_{(n)}]}(x) sup_{y \leq x} \hat{m}_n(y)$ for monotonically increasing mrl. They also prove consistency of the estimator. Abdous and Berred [1] use a local linear fitting technique to find a smooth estimate assuming only that the smoothing kernel is symmetric. A nonparametric hypothesis testing procedure for comparing mrl functions from two independent groups was introduced by Berger et. al. [2]. A practical benefit of this procedure is that mrl estimates of the two groups were allowed to cross, a pattern that is likely to arise in applications.

Classical estimation for the mrl function began to have a semiparametric regression flavor when Oakes and Dasu [21] extended the class of distribution having linear mrl functions [13], to a family having proportional mrl functions, $m_1(x) = \psi m_2(x)$ for $\psi > 0$. Maguluri and Zhang [19] further extended the proportional mrl model to a regression setting, $m(x|z) = exp(\psi z) m_0(x)$, where z is a vector of covariates, ψ is of vector of regression coefficients, and $m_0(x)$ is a baseline mrl function. They provide estimators for the vector of covariate effects ψ. One estimator is based off of maximum likelihood methods of the exponential regression model, while a second arises from the proportional hazards version of the model. The two estimators are compared using simulations and they find the estimator under the exponential regression model performs superior to the estimate arising from the proportional hazards model. Both estimates are consistent and asymptotically normal. Chen and Cheng [3] also extend the proportional mrl model to include inference for the regression parameters with censored data. The method is developed through counting process theory.

In contrast to the classical literature, there has been very little work on modeling and inference for mrl functions under the Bayesian framework. Lahiri and Park [18] present nonparametric Bayes and empirical Bayes estimators under a Dirichlet process [5] prior for the probability distribution. They show that the Bayes estimator becomes a weighted average of the prior guess for the mrl function and the empirical mrl function of the data. Johnson [15] discusses a Bayesian method for estimation of the mrl function under interval and right censored data, also using a Dirichlet process prior for the corresponding survival function.

In this chapter, we develop Bayesian inference methods for mrl functions based on both a parametric and a more general nonparametric mixture model for the survival distribution (Sections 3.2 and 3.3, respectively). The two approaches are compared using a data set from the literature (Section 3.4), which illustrates the flexibility of the nonparametric model setting. Possible extensions to more general modeling for mrl functions are discussed in Chapter 4.

3.2 Exponentiated Weibull Model

The exponentiated Weibull Model [20] has been considered as a flexible parametric model with regard to the shapes of its mrl function. Specifically, the mrl function may take on a number of various forms, namely monotone increasing, monotone decreasing, constant, UBT, or BT. The survival function has a closed form hence so does the hazard rate function; however, the mrl function still requires numerical methods to obtain. The distribution, density, hazard, and mrl function for the exponentiated Weibull model are given by the following expressions:

$$\begin{aligned}
F(x;\alpha,\theta,\sigma) &= \left[1-exp\left(-\left(\frac{x}{\sigma}\right)^{\alpha}\right)\right]^{\theta}, \quad x>0,\alpha,\theta,\sigma>0 \qquad (3.1)\\
f(x;\alpha,\theta,\sigma) &= \frac{\alpha\theta}{\sigma}\left[1-exp\left(-\left(\frac{x}{\sigma}\right)^{\alpha}\right)\right]^{\theta-1} exp\left(-\left(\frac{x}{\sigma}\right)^{\alpha}\right)\left(\frac{x}{\sigma}\right)^{\alpha-1}\\
h(x;\alpha,\theta,\sigma) &= \frac{\alpha\theta\left[1-exp\left(-\left(\frac{x}{\sigma}\right)^{\alpha}\right)\right]^{\theta-1} exp\left(-\left(\frac{x}{\sigma}\right)^{\alpha}\right)\left(\frac{x}{\sigma}\right)^{\alpha-1}}{\sigma\left[1-\left[1-exp\left(-\left(\frac{x}{\sigma}\right)^{\alpha}\right)\right]^{\theta}\right]}\\
m(x;\alpha,\theta,\sigma) &= \frac{\int_x^\infty\left[1-\left[1-exp\left(-\left(\frac{t}{\sigma}\right)^{\alpha}\right)\right]^{\theta}\right]dt}{1-\left[1-exp\left(-\left(\frac{x}{\sigma}\right)^{\alpha}\right)\right]^{\theta}}
\end{aligned}$$

where α and θ are shape parameters and σ is a scale parameter. Note that σ, being a scale, will not play a role in determining the form of the hazard and mrl functions. Table 3.1 provides the parameter sets that result in each distinct shape for the mrl function.

α	θ	$\alpha\theta$	form of mrl function
1	1	1	exponential distribution $\rightarrow$ constant mrl
–	1	–	weibull distribution $\rightarrow$ monotone (inc, dcr or constant) mrl
< 1	$\neq 1$	< 1	increasing
> 1	$\neq 1$	> 1	decreasing
> 1	< 1	< 1	UBT
< 1	> 1	> 1	BT

Table 3.1: Forms of MRL for Exponentiated Weibull Distribution

Mudholkar and Strivastava [20] provide a similar table as Table 3.1 for the hazard rate function for specific domains of α and θ. Xie et. al. [24] look at the role of the product of the shape parameters on the form of the hazard rate. Gupta and Akman [10] prove that if the hazard rate function is BT and $h(0) > 1/m(0)$, then the mrl is UBT, while $h(0) \leq 1/m(0)$ implies decreasing mrl function. Similarly if the hazard rate function is UBT and $h(0) > 1/m(0)$, then the mrl is BT, while $h(0) \geq 1/m(0)$ implies increasing function. Combining the aforementioned results, we can improve the table in [20] to specify the exact shape of the mrl function for particular values of α and θ in conjunction with the value of the product of the parameters.

We use the exponentiated Weibull distribution under a Bayesian framework to model survival data with a parametric approach. Exponential priors for both shape parameters as well as the scale parameter provide a natural choice given that the parameters take on values in $\mathbb{R}^+$, as well as a convenient option for prior specification. Assuming prior independence we have

$$p(\alpha, \theta, \sigma) \sim Exp(\alpha; a_\alpha) Exp(\theta; a_\theta) Exp(\sigma; a_\sigma) \tag{3.2}$$

where a_α, a_θ, and a_σ are the means of the respective exponential distributions. Prior specification can be obtained using three prior quantiles (Q_1, Q_2, Q_3) that describe prior guesses of the survival population percentiles (P_1, P_2, P_3). First, we create a system of equations (3.3) to solve for α, θ, and σ. Then, the resulting values may be used as the prior means a_α, a_θ, and a_σ, respectively.

$$P_j = \left[1 - exp\left(-\left(\frac{Q_j}{\sigma}\right)^\alpha\right)\right]^\theta \text{ for } j = 1, 2, 3 \tag{3.3}$$

The posterior conditional distributions are not conjugate, so we use a Metropolis-Hastings algorithm for MCMC posterior simulation. Due to the strong correlation amongst the parameters, we use a trivariate normal distribution for the proposal distribution. For numerical stability, we used the proposal on the log-scale. The mean of the proposal is given by the log of the current posterior sample of α, θ, and σ. The covariance matrix is first specified as $diag(1,1,1) * c$ where c is a constant or vector of constants that improve mixing. We run the model for 1000 iterations, and update the covariance matrix with the covariance of the log of the 1000 posterior samples. We repeat this process until the covariance matrix is virtually unchanged. We will call the resulting covariance matrix $\boldsymbol{H}$. Let x_i for $i = 1, ..., n$ be the observed survival times, $\boldsymbol{\phi} = (\alpha, \theta, \sigma)'$, and B be the number of iterations of the MCMC. We obtain posterior samples of α, θ, and σ by

$$
\begin{array}{rcl}
\text{Initialize} & & \boldsymbol{\phi}^{(1)} = (a_\alpha, a_\theta, a_\sigma)' \\
\text{for } b = 1, ..., B+1 & & \\
log(\boldsymbol{\phi}^*) & \overset{\text{draw}}{\sim} & MVN_3\left(log\left(\boldsymbol{\phi}^{(b)}\right), \boldsymbol{H} * c'\right) \\
\eta & \overset{\text{draw}}{\sim} & Unif(0,1) \\
\text{if } \eta & < & min\left\{ \frac{\overbrace{\left(\prod_{i=1}^{n} f(x_i; \boldsymbol{\phi}^*)\right)}^{\text{likelihood}} \times \overbrace{p(\boldsymbol{\phi}^*; a_\alpha, a_\theta, a_\sigma)}^{\text{prior}} \times \alpha^*\theta^*\sigma^*}{\left(\prod_{i=1}^{n} f(x_i; \boldsymbol{\phi}^{(b)})\right) \times p(\boldsymbol{\phi}^{(b)}; a_\alpha, a_\theta, a_\sigma) \times \alpha^{(b)}\theta^{(b)}\sigma^{(b)}}, 1 \right\} \\
 & & \text{let } \boldsymbol{\phi}^{(b+1)} = \boldsymbol{\phi}^* \\
\text{else} & & \boldsymbol{\phi}^{(b+1)} = \boldsymbol{\phi}^{(b)}
\end{array}
$$

Finally, burn-in and thinning may be applied to obtain uncorrelated posterior samples of α, θ, and σ. Once we have the desired posterior samples, we can compute point and interval estimates for the density, survival, and hazard functions by evaluating their expressions in (3.1) over a grid of values x_0 for survival time. Regarding the mrl function, we can compute it using the expression in (2.5), that is,

$$m(x_0; \alpha, \theta, \sigma) = \frac{\int_0^\infty S(t|\alpha, \theta, \sigma)dt - \int_0^{x_0} S(t|\alpha, \theta, \sigma)dt}{S(x_0|\alpha, \theta, \sigma)}$$

This form of the mrl function was chosen to minimize numerical instability. Since we are essentially truncating the survival function by evaluating over a grid, the numerical computation is not as reliable when we integrate using the form of (1.1) at upper grid values.

3.3 Nonparametric Lognormal Mixture Model

3.3.1 Model Formulation

When the data exhibits unusual distributional features such as multi-modality or skewness, parametric models tend to fail to capture these important features. A way to go about this issue is to use a mixture model that combines a number of distributions that we will refer to as components of the model. The question then becomes how many components should be used and how should they be combined together? These concerns can be addressed by bringing in a nonparametric aspect to the model, in particular, to the weights of each component and to the number of components.

We use a Dirichlet process (DP) prior for the mixing distribution resulting in a DP nonparametric mixture model, $f(x; G) = \int k(x; \boldsymbol{\theta})dG(\boldsymbol{\theta})$, for the density of the survival distribution. Specifically, we take a lognormal distribution for the kernel, thus, $k(x;\boldsymbol{\theta}=(\mu,\sigma^2))=\frac{1}{x\sqrt{2\pi\sigma^2}}exp\left[-\frac{1}{2}\left(\frac{ln(x)-\mu}{\sigma}\right)^2\right]$, and assign a $DP(\alpha, G_0)$ prior to G. The DP is a stochastic process with random sample paths that are distributions [5]. Thus a realization from the DP provides a random cdf sample path. The G_0 parameter is the baseline or centering distribution, while the α parameter is a precision parameter; the larger the value of α the closer the DP sample path is to the centering distribution. We use the stick-breaking (SB) constructive definition of the DP defined by Sethuraman [22], which states that a sample $G(\cdot)$ from $DP(\alpha, G_0)$ is almost surely of the form $\sum_{l=1}^{\infty} w_l \delta_{\theta_l}(\cdot)$ where $\delta_{\theta_l}(\cdot)$ is a point mass at θ_l. The θ_l, for all $l \in \{1,2,...\}$, are iid samples from the baseline distribution, G_0, and the w_l are the corresponding weights constructed sampling iid latent variables $v_r \sim$Beta$(1,\alpha)$, for all $r \in \{1,2,...\}$, then $w_1 = v_1$ and $w_l = v_r \prod_{r=1}^{l-1}(1-v_r)$, for all $l \in \{2,3,...\}$.

We will use the truncated version of the SB constructive definition of the DP, $G_N(\cdot) = \sum_{l=1}^{N} p_l \delta_{\theta_l}(\cdot)$, where $\theta_l \overset{iid}{\sim} G_0$ for $l = 1,...,N$, and $p_1 = v_1$ and $p_l = v_r \prod_{r=1}^{l-1}(1-v_r)$, for $l = 2,3,...N$, where $v_r \overset{iid}{\sim} Beta(1,\alpha)$ for $r = 1,...,N-1$. Thus the model for the observed survival times x_i, $i = 1,...,n$, becomes

$$\begin{aligned} f(x_i|G) &\overset{ind}{\sim} \int LN(x_i;\mu,\sigma^2)dG(\mu,\sigma^2) \\ &= \sum_{l=1}^{N} p_l LN(x_i;\mu_l,\sigma_l^2) \end{aligned} \tag{3.4}$$

where p_l for $l = 1,...,N$ are the weights obtained via the DP SB construction, described above, corresponding to the component $\theta_l = (\mu_l, \sigma_l^2)$ and N is the total number of components in the mixture model. Technically, since the number of components is predetermined there is no nonparametric element to the number of components.

However, N is generally chosen to overestimate the true number of components, so that the number of components suggested by the data is captured by the model. In fact, many of the components will just be assigned a probability that is virtually zero. The number of components for the finite sum DP approximation can be found using $E\left(\sum_{l=1}^{N} p_l\right) = 1 - \left(\frac{\alpha}{\alpha+1}\right)^N$, in particular, solving for N in $\left(\frac{\alpha}{\alpha+1}\right)^N = \epsilon$ for small $\epsilon > 0$.

The lognormal kernel for the DP mixture is chosen to provide the appropriate support on $\mathbb{R}^+$ and, at the same time, a useful connection with the more standard normal DP mixture model. Consider the transformation $Y = log(X)$ and note that,

$$Pr(X \leq x; G) = Pr(X \leq exp(y); G)$$
$$= \int_0^{exp(y)} f(x; G)dx = \int_0^{exp(y)} \sum_{l=1}^{N} p_l LN(x; \mu_l, \sigma_l^2)dx$$
$$= \sum_{l=1}^{N} p_l \int_0^{exp(y)} LN(x; \mu_l, \sigma_l^2)dx = \sum_{l=1}^{N} p_l \Phi\left(\frac{log(exp(y)) - \mu_l}{\sigma_l}\right)$$
$$= \sum_{l=1}^{N} p_l \Phi\left(\frac{y - \mu_l}{\sigma_l}\right)$$

Therefore, modeling X with a lognormal kernel is essentially equivalent to modeling Y with a normal kernel. Hence we can obtain inference for the lognormal mixture on the original scale by fitting the following normal DP mixture model to the transformed responses $y_i = log(x_i)$ for $i = 1, .., n$:

$$y_i|G \overset{ind}{\sim} \int N(y_i; \mu, \sigma^2)dG(\mu, \sigma^2) = \sum_{l=1}^{N} p_l N(y_i; \mu_l, \sigma_l^2) \text{ for } i = 1 : n \quad (3.5)$$

In our data example (Section 3.4) we use the following conditionally conjugate priors

$$
\begin{aligned}
(\mu_l, \sigma_l^2)|\lambda, \tau^2, \rho &\sim G_0(\mu_l, \sigma_l^2; \lambda, \tau^2, \rho) = N(\mu_l; \lambda, \tau^2)\Gamma^{-1}(\sigma_l^2; a, \rho(\text{scale})) \qquad (3.6)\\
\lambda &\sim N(\lambda; a_\lambda, b_\lambda)\\
\tau^2 &\sim \Gamma^{-1}(\tau^2; a_\tau, b_\tau(\text{scale}))\\
\rho &\sim \Gamma(\rho; a_\rho, b_\rho(\text{rate}))\\
\alpha &\sim \Gamma(\alpha; a_\alpha, b_\alpha(\text{rate})).
\end{aligned}
$$

When it comes to prior specification often there is not much prior knowledge on the behavior of the population of of interest, but typically the researcher will have at least somewhat of an idea of the range of the population. We would want to set our priors to have a prior predictive distribution that encompasses this range. One way to do so is to imagine one dispersed normal distribution that is centered at the midrange with 2 standard deviations either way representing the prior range. We can then divide the range by 4 and square that value to get the prior variance of the data. Using the formulation below, we can divide the prior variance amongst the three additive components:

$$
\begin{aligned}
\left(\frac{range(Y)}{4}\right)^2 = Var(Y) &= Var(E(Y|\mu, \sigma^2)) + E(Var(Y|\mu, \sigma^2)) \qquad (3.7)\\
&= Var(\mu) + E(\sigma^2)\\
&= Var(E(\mu|\lambda, \tau^2)) + E(Var(\mu|\lambda, \tau^2)) + E(E(\sigma^2|a, \rho))\\
&= Var(\lambda) + E(\tau^2) + E(\rho/(a-1))\\
&= b_\lambda + \frac{b_\tau}{a_\tau - 1} + \left(\frac{1}{a-1}\right)\left(\frac{a_\rho}{b_\rho}\right)
\end{aligned}
$$

Using a shape parameter of 2 for b_τ and a would provide infinite prior variance for their respective inverse gamma distributions. The variance for a gamma distribution with rate b_ρ and shape a_ρ is given by a_ρ/b_ρ^2 thus a larger shape parameter relative to the square of the rate would give a larger prior variance.

Regarding the prior for α, we consider the relationship between the number of distinct components and the value of α. In general, the number of distinct components is large for large α and small for small α. If the data set is moderately large, E(#distinct components$|\alpha) \approx \alpha log\left(\frac{\alpha+n}{\alpha}\right)$ can be used to suggest an appropriate α value.

3.3.2 Posterior Inference

Posterior samples of the unknown parameters can easily be obtained using the block Gibbs sampler for DP mixtures [14]. Observe that before we introduce $\theta_l = (\mu_l, \sigma_l^2)$, the first two levels of the model are,

$$\begin{aligned} y_i|\zeta_i &\overset{ind}{\sim} N(y_i;\zeta_i),\ i=1,...,n \\ \zeta_i|\boldsymbol{p},\boldsymbol{\theta} &\overset{iid}{\sim} G_N,\ i=1,...,n \end{aligned}$$

where $\boldsymbol{p} = (p_1, ..., p_N)$ are the weights corresponding to the weights, $\boldsymbol{\theta} = (\theta_1, ..., \theta_N)$. By marginalizing over the ζ_i we obtain the finite mixture model in (3.5). Now we can augment the model with configuration variables $\boldsymbol{L} = (L_1, ..., L_n)$ such that $L_i = l$ iff $\zeta_i = \theta_l$. Then the full hierarchical model is given by,

$$\begin{aligned} y_i|\boldsymbol{\mu},\boldsymbol{\sigma}^2, L_i &\overset{ind}{\sim} N(y_i;\mu_{L_i},\sigma_{L_i}^2) \\ L_i|\mathbf{p} &\overset{iid}{\sim} \sum_{l=1}^{N} p_l\delta_l(L_i) \\ \mathbf{p}|\alpha &\sim f(\mathbf{p};\alpha)\ \ (SB) \\ (\mu_l,\sigma_l^2)|\lambda,\tau^2,\rho &\sim N(\mu_l;\lambda,\tau^2)\Gamma^{-1}(\sigma_l^2;a,\rho(\text{scale})) \\ \lambda &\sim N(\lambda;a_\lambda,b_\lambda) \\ \tau^2 &\sim \Gamma^{-1}(\tau^2;a_\tau,b_\tau(\text{scale})) \\ \rho &\sim \Gamma(\rho;a_\rho,b_\rho(\text{rate})) \\ \alpha &\sim \Gamma(\alpha;a_\alpha,b_\alpha(\text{rate})) \end{aligned}$$

where $f(\boldsymbol{p}|\alpha) = \alpha^{N-1} p_N^{\alpha-1}(1-p_1)^{-1}(1-(p_1+p_2))^{-1} \times ... \times (1-\sum_{l=1}^{N-2} p_l)^{-1}$ is a special case of the generalized Dirichlet distribution as is Connor and Mosimann [4]. Let n^* be the number of distinct components of $\boldsymbol{L}$ where $\{L_j^* : j = 1, ..., n^*\}$ are the distinct components. Let Ψ represent the vector of the most recent iteration of all other parameters. Let $b = 1, ..., B$ be the number of iterations in the MCMC. Then the samples from the joint posterior distribution are obtained by

for $b = 1, ..., B + 1$

- Posterior conditional distribution for θ_l for $l = 1, ..., N$:
  If l IS NOT already a component: $l \notin \{L_j^{*(b)} : j = 1, ..., n^{*(b)}\}$

$$\begin{aligned} \mu_l^{(b+1)}|data, \Psi &\overset{\text{draw}}{\sim} N(\lambda^{(b)}, \tau^{2(b)}) \\ \sigma_i^{2(b+1)}|data, \Psi &\overset{\text{draw}}{\sim} \Gamma^{-1}(a, \rho^{(b)}) \end{aligned}$$

If l IS a component: $l \in \{L_j^{*(b)} : j = 1, ..., n^{*(b)}\}$

$$\begin{aligned} \mu_l^{(b+1)}|data, \Psi \quad & \left(\propto N(\mu_l; \lambda, \tau^2) \prod_{\{i:L_i=l\}} N(y_i; \mu_l, \sigma_l^2) \right) \\ \overset{\text{draw}}{\sim} \quad & N(m_{\mu_l}, s_{\mu_l}^2) \\ & s_{\mu_l}^2 = \frac{1}{\frac{1}{\tau^{2(b)}} + \frac{n_j^{(b)}}{\sigma_l^{2(b)}}}, \quad n_j^{(b)} = \{\# : l = L_i^{(b)}, i = 1, ..., n\} \\ & m_{\mu_l} = \left(\frac{\sum_{\{i:L_i^{(b)}=l\}} y_i}{\sigma_l^{2(b)}} + \frac{\lambda^{(b)}}{\tau^{2(b)}} \right) s_{\mu_l}^2 \\ \sigma_i^{2(b+1)}|data, \Psi \quad & \left(\propto \Gamma^{-1}(\sigma_l^2; a, \rho) \prod_{\{i:L_i=l\}} N(y_i; \mu_l, \sigma_l^2) \right) \\ \overset{\text{draw}}{\sim} \quad & \Gamma^{-1} \left(\frac{n_r^{(b)}}{2} + a, \frac{1}{2} \sum_{\{i:L_i^{(b)}=l\}} (y_i - \mu_l^{(b+1)})^2 + \rho^{(b)} \right) \end{aligned}$$

- Update for $\boldsymbol{p}$:

$$\boldsymbol{p}^{(b+1)}|data, \Psi \qquad \left(\propto f(\boldsymbol{p}|\alpha)\prod_{l=1}^{N} p_l^{M_l}\ M_l = |\{i : L_i = l\}|\ , l = 1, ..., N\right)$$

$$\overset{\text{draw}}{\sim} \quad \text{Generalized Dirichlet Distribution}$$

$$\text{for } l = 1, ..., N \text{ draw latent variable:}$$

$$V_l^{*(b+1)} \overset{ind}{\sim} Beta(1 + M_l^{(b)}, \alpha^{(b)} + \sum_{r=l+1}^{N} M_r^{(b)})$$

$$\begin{aligned}
\Rightarrow \quad p_1^{(b+1)} &= V_1^{*(b+1)} \\
p_l^{(b+1)} &= V_l^{*(b+1)} \prod_{r=1}^{l-1}(1 - V_r^{*(b+1)})\ (l = 2, ..., N-1) \\
p_N^{(b+1)} &= 1 - \sum_{l=1}^{n-1} p_l^{(b+1)}
\end{aligned}$$

- Update for L_i for $i = 1, .., n$:

$$L_i^{(b+1)}|data, \Psi \quad \overset{\text{draw}}{\sim} \quad \sum_{l=1}^{N} \tilde{p}_{li}\delta_{(l)}(\cdot)$$

$$\tilde{p}_{li} = \frac{p_l^{(b+1)} N(y_i; \mu_l^{(b+1)}, \sigma_l^{2(b+1)})}{\sum_{l=1}^{N} p_l^{(b+1)} N(y_i; \mu_l^{(b+1)}, \sigma_l^{2(b+1)})},\ l = 1, ..., N$$

- The posterior conditional distribution for λ:

$$\lambda^{(b+1)}|data, \Psi \qquad \left(\propto N(\lambda; a_\lambda, b_\lambda)\prod_{l=1}^{N} N(\mu_l; \lambda, \tau^2)\right)$$

$$\overset{\text{draw}}{\sim} \quad N(m_\lambda, s_\lambda^2)$$

$$m_\lambda = \left(\frac{a_\lambda}{b_\lambda} + \frac{\sum_{l=1}^{N} \mu_l^{(b+1)}}{\tau^{2(b)}}\right) s_\lambda^2,\ s_\lambda^2 = \frac{1}{\frac{1}{b_\lambda} + \frac{N}{\tau^{2(b)}}}$$

- The posterior conditional distribution for τ^2:

$$\tau^{2(b+1)}|data, \Psi \qquad \left(\propto \Gamma^{-1}(\tau^2; a_\tau, b_\tau)\prod_{l=1}^{N} N(\mu_l; \lambda, \tau^2)\right)$$

$$\overset{\text{draw}}{\sim} \quad \Gamma^{-1}\left(\frac{N}{2} + a_\tau, \frac{1}{2}\sum_{l=1}^{N}(\mu_l^{(b+1)} - \lambda^{(b+1)})^2 + b_\tau\right)$$

- The posterior conditional distribution for ρ:

$$\rho^{(b+1)}|data, \Psi \qquad \left(\propto \Gamma(\rho; a_\rho, b_\rho)\prod_{l=1}^{N}\Gamma^{-1}(\sigma_l^2; a, \rho)\right)$$

$$\overset{\text{draw}}{\sim} \quad \Gamma\left(a_\rho + aN, \sum_{l=1}^{N}\frac{1}{\sigma_l^{2(b+1)}} + b_\rho^{(b+1)}\right)$$

- The posterior conditional distribution for α:

$$\alpha^{(b+1)}|data, \Psi \qquad (\propto \Gamma(\alpha; a_\alpha, b_\alpha)f(\mathbf{p}|\alpha))$$

$$\overset{\text{draw}}{\sim} \quad \Gamma\left(N + a_\alpha - 1, -\sum_{s=1}^{N-1} log(1 - V_s^{*(b+1)}) + b_\alpha\right)$$

Once the posterior samples are obtained we can compute point and interval estimates for the density function over a grid of values $\boldsymbol{x}_0$ (on the original scale) by,

$$f(x_0|G_N) = \int\int LN(x_0; \mu_{L_0}, \sigma_{L_0}^2)\left(\sum_{l=1}^{N} p_l\delta_l(L_0)\right)$$

$$\times p(\mu, \sigma^2, \boldsymbol{p}, \boldsymbol{L}, \lambda, \tau^2, \rho, \alpha|data)dL_0 d\mu d\sigma^2 d\boldsymbol{p} d\boldsymbol{L} d\lambda d\tau^2 d\rho d\alpha$$

and integrating over all possible values of a new L_0

$$= \int\left(\sum_{l=i}^{N} p_l LN(x_0; \mu_l, \sigma_l^2)\right)p(\mu, \sigma^2, \boldsymbol{p}, \boldsymbol{L}, \lambda, \tau^2, \rho, \alpha|data)d\mu d\sigma^2 d\boldsymbol{p} d\boldsymbol{L} d\lambda d\tau^2 d\rho d\alpha$$

Moreover, the survival function at grid point x_0 is given by,

$$S(x_0|G_N) =$$

$$\int\left(\sum_{l=i}^{N} p_l \Phi\left(\frac{log(x_0) - \mu_l}{\sigma_l}\right)\right)p(\mu, \sigma^2, \boldsymbol{p}, \boldsymbol{L}, \lambda, \tau^2, \rho, \alpha|data)d\mu d\sigma^2 d\boldsymbol{p} d\boldsymbol{L} d\lambda d\tau^2 d\rho d\alpha$$

Notice that both the density and survival functions are approximated by a summation of the mixture components. This is not a problem for obtaining the hazard rate function at x_0, which is still given directly from its definition $h(x_0|G_N) = \frac{f(x_0|G_N)}{S(x_0|G_N)}$. Obtaining the mrl function must be done by numerical integration approximation for the integral over the survival distribution. The survival function is monotone decreasing so the trapezoid technique is an appropriate technique that is also quite simple. Of course point and interval estimates for each of the aforementioned functions are desirable, so if we consider a grid of possible new values $\boldsymbol{x}_0 = x_{0,1}, ..., x_{0,K}$

$$
\begin{aligned}
&\text{for } b = 2, ..., B+1 \\
&\quad \text{for } k = 1, ..., K \\
&\qquad f(x_{0,k}|G_N)^{(b)} = \sum_{l=i}^{N} p_l LN(x_{0,k}; \mu_l^{(b)}, \sigma_l^{2(b)}) \\
&\qquad S(x_{0,k}|G_N)^{(b)} = \sum_{l=i}^{N} p_l \Phi\left(\frac{log(x_{0,k}) - \mu_l^{(b)}}{\sigma_l^{(b)}}\right) \\
&\qquad h(x_{0,k}|G_N)^{(b)} = \frac{f(x_{0,k}|G_N)^{(b)}}{S(x_{0,k}|G_N)^{(b)}} \\
&\text{for } k = 1, ..., K-1 \\
&\qquad m(x_{0,k}|G_N)^{(b)} = \frac{\frac{1}{2}\sum_{h=k}^{K-1}\left((x_{0,(h+1)} - x_{0,h})(S(x_{0,(h+1)}|G_N)^{(b)} + S(x_{0,(h)}|G_N)^{(b)})\right)}{S(x_{0,k}^{(b)}|G_N)}
\end{aligned}
$$

Just as in the exponentiated Weibull model we can save the 2.5% and 97.5% quantiles along with the mean at each grid point for each function to obtain the desired point and interval estimates.

3.4 Example

We use the data set considered in Berger et. al. [2] to illustrate posterior inference under both the exponentiated Weibull model and the nonparametric DP lognormal mixture model. The data set consists of survival times of rats in two experimental groups. The first group (Ad libitum group) is comprised of 90 rats who were allowed to eat freely as they desired. The second group (Restricted group) is comprised of 106 rats that were placed on a restricted diet. Our interest lies in studying the form of the mrl function under each condition, and moreover whether the mrl functions are significantly different from one another. We are also interested in how each model performs in comparison to one another.

3.4.1 Results

Under the exponentiated Weibull model (3.1), we used the 10%, 50%, and 90% quantiles of the data with formula (3.3) to approximate appropriate priors for each group. The restricted group had respective quantile values of ($Q_1 = 1.55$, $Q_2 = 2.84$, $Q_3 = 3.34$). If we set $\alpha = 2$, $\theta = 5$, and $\sigma = 2$, then the corresponding quantiles are given as $Q_1^{'} = 1.99$, $Q_2^{'} = 2.85$, and $Q_3^{'} = 4.07$ which we considered to be reasonably close to the observed quantiles. Therefore, we set hyper-parameters in (3.2) to be $a_\alpha = 2$, $a_\theta = 5$, and $a_\sigma = 2$. Following the same methodology for the ad libitum group, we set the hyper-parameters to $a_\alpha = 4$, $a_\theta = 1$, and $a_\sigma = 2$. Point and interval estimates of the density function are plotted in the top row of Figure 3.1.

Prior selection under the nonparametric lognormal DP mixture model (3.4) was decided using the approximation for the variance of Y in (3.7). Note in (3.7) we use the transformed random variable Y, but since the location and scale parameters of Y and X are equivalent, the formulation remains the same under the original random variable X. The range of the restricted group on the log scale is (4.65396,7.26892). This gives us a spread of about 2.7 so the prior variance is about 0.46. We distribute the variance evenly across the terms, and set the shape parameters $a = a_\tau = 2$ so that the

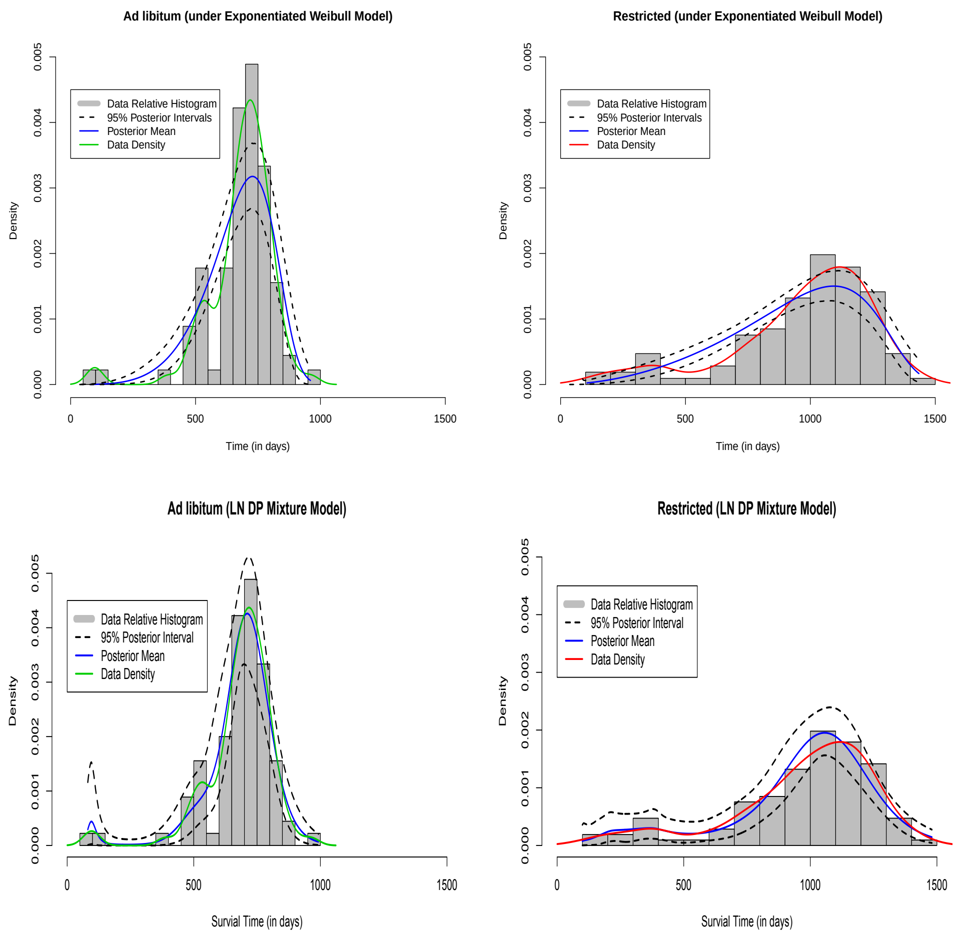

Figure 3.1: Relative frequency histogram and densities of lifetime (in days) of the two experimental groups (Ad libitum is left and Restricted is right) along with posterior mean and 95% interval estimates for the density functions under the exponentiated Weibull model (top) and LN DP mixture model (bottom).

the corresponding prior have infinite variance. We set $a_\rho = 20$. This leaves us with $b_\lambda = 0.15$. $b_\tau = 0.15$ and $b_\rho = 133.3$. The prior mean for λ was set at the prior mean of the group $a_\lambda = 6.8$. For the ad libitum group, we followed the same approach. The range on the log scale is (4.488636 ,6.870053) so the spread is about 2.5 leading to an approximated prior variance of 0.39. Dividing the variance evenly amongst the terms keeping $a = a_\tau = 2$ and decreasing $a_\rho = 19$, we get that $b_\lambda = 0.13$, $b_\tau = 0.13$ and $b_\rho = 146.2$. We set a_λ once again to the mean of the data, 6.5. For both groups, we set $a_\alpha = 2$ and $b_\alpha = 4$ which leads to a prior expected number of distinct components to be about 3. Finally, we set the number of mixture components to $N = 20$. Posterior estimates for the densities for the two groups under the DP mixture model are shown in the bottom row of Figure 3.1.

In Figure 3.1 we note that the parametric model has some trouble capturing some of the characteristics of the data. In the ad libitum group (upper left) a minor mode is suggested just below the 200^{th} day. The unimodality of the exponentiated Weibull distribution makes it impossible for the parametric model to capture this shape. We note that the model tries to by reaching the tail of the estimated density out to these values, but this is at a cost of underestimating the density where most of the data exist, and overestimating the density where there is no data at all. There are many regions where the data and the density of the data (green) do not even fall within the interval estimates (black dashed). If we compare to how well the nonparametric model (lower left) performs we see quite a bit of improvement. The extra structure at the lower survival times is now being captured without the consequences of modeling poorly in

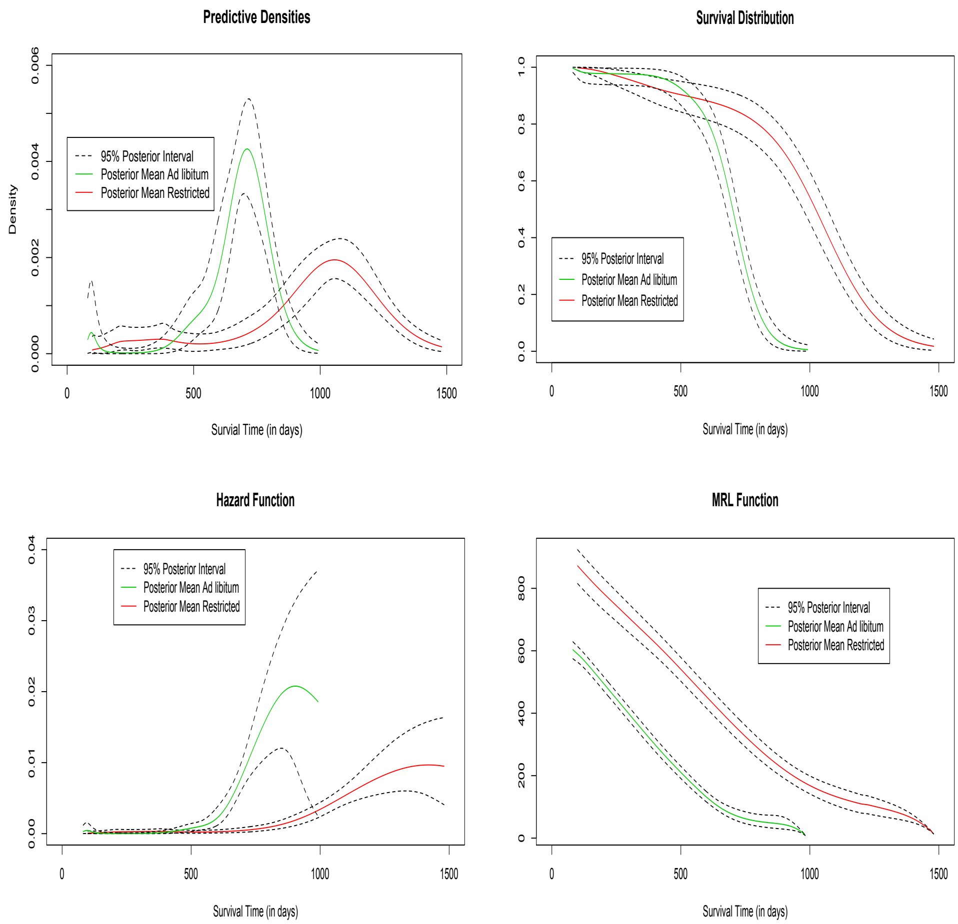

Figure 3.2: Point and interval estimates of lifetime (in years) for the density (top left), survival (top right), hazard rate (lower left), and ml (lower right) functions of the two experimental groups under the LN DP mixture model.

other regions of the data. The data density remains within the interval estimates over the entire range of the data. We see similar results for the restricted group, which has a large left skew with a slight mode in the far tail. The exponentiated Weibull model (upper right) is able to model some of the skewness, but again runs into trouble by smoothing over obvious peaks and valleys. Again there are a number of regions in which the density of the data (red) is not contained in the interval estimates of the model. The lognormal DP mixture model (lower right) is able to capture the peaks and valleys that the exponentiated Weibull model could not. There is a slight discrepancy from the point estimate (blue) and the density of the data (red) around 1250 days. Nonetheless, the data density remains within the interval estimates of the model.

By comparing the densities under the two models, there is clear evidence that the nonparametric lognormal DP mixture model is superior to the exponentiated Weibull model. Therefore, we will use the results under the nonparametric lognormal DP mixture model to compare the mrl functions under the two groups. In Figure 3.2, we plot point and interval estimates of the posterior density functions (upper left), survival functions (upper right), hazard functions (lower left), and mrl functions (lower right) for both the ad libitum (green) and restricted (red) groups. Looking at the estimated densities we can see that the majority of the ad libitum group have lower survival times compared to the restricted group. The survival function estimates show that after about 700 days the survival curve of the restricted group is significantly higher than the ad libitum survival curve. The hazard function shows that the probability of death in the next instant is much higher for the ad libitum group past 500 days. The mrl functions are monotonically decreasing and do not cross. This leads us to conclude that the remaining life expectancy of a rat in the restricted group is higher than the remaining life expectancy of a rat in the ad libitum group at any given time within the range of the data.

3.4.2 Model Comparison

We use the minimum posterior predictive loss approach by Gelfand and Gosh [7] to compare the exponentiated Weibull model to the nonparametric DP lognormal mixture model. Under this criterion the goal is to minimize, within the collection of models under consideration, the expectation of a specified loss function under the posterior predictive distribution of replicate responses $\boldsymbol{x}_{rep}$ given the observed data $\boldsymbol{x}_{obs}$. Here, we use the square error loss function so that the general criterion is given by

$$D_k(m) = \sum_{i=1}^{n} var(x_{i,rep}|\boldsymbol{x}_{obs}, m) + \frac{k}{k+1} \sum_{i=1}^{n} (E(x_{i,rep}|\boldsymbol{x}_{obs}, m) - x_{i,obs})^2$$

where $x_{i,rep}$ is a replicate of the i^{th} observation, $x_{i,obs}$, under the posterior predictive distribution of the m^{th} model. The first term is representative of a penalty measure $P(m)$, and the second term is a goodness-of-fit measure $G(m)$. The value of k is specified as the relative regret for departure from $x_{i,rep}$. Note that as k tends to infinity, the criterion becomes the sum of the penalty $P(m)$ and goodness-of-fit $G(m)$ measures.

For the exponentiated Weibull model (m_1), obtaining $E(x_{i,rep}|\boldsymbol{x}_{obs}, m)$ and $var(x_{i,rep}|\boldsymbol{x}_{obs}, m)$ is straightforward. The posterior predictive distribution is given by $p(x_{i,rep}|\boldsymbol{x}_{obs}) = \int EW(x_{i,rep}|\alpha, \theta, \sigma)p(\alpha, \theta, \sigma|data)d\alpha d\theta d\sigma$ and can thus be sampled by taking the posterior samples $(\alpha_b, \theta_b, \sigma_b)$, for $b = 1, ..., B$, and drawing $x_{i,rep,b}$ from the exponentiated Weibull distribution given each posterior parameter vector. Next, we compute the mean and variance of the B replicates. Important to note is that the mean and variance for one experimental group is going to be the same for each observation in that group. We find the $E(x_{i,rep}|\boldsymbol{x}_{obs}, m_1)$ and $var(x_{i,rep}|\boldsymbol{x}_{obs}, m_1)$ for the ad libitum group to be 671.2 and 17433.0, respectively, and for the the restricted group to be 949.5 and 74691.7, respectively. Thus the ad libitum group has $G(m_1)^a = \sum_{i=1}^{90}(671.2 - x_{i,obs})^2 = 1615787$ and $P(m_1)^a = 90 * (17433.0) = 1568967$. The restricted group has $G(m_1)^r = \sum_{i=1}^{106}(949.5 - x_{i,obs})^2 = 8542725$ and $P(m_1)^r = 106 * (74691.7) = 7917319$.

Obtaining the criterion under the nonparametric DP lognormal mixture model (m_2) takes a little more care. Recall that $x_i|G \overset{ind}{\sim} \int LN(x_i; \mu, \sigma^2)dG(\mu, \sigma^2)$ for $i =, ..., n$. In order to obtain replicates for each x_i, we need to know the l^{th} component from which the observed x_i came from according to the model. Thus we need to sample $x_{i,rep}|\boldsymbol{x}_{obs}, m_2 \sim \int LN(x_{i,rep}|\mu_{l_i}, \sigma^2_{l_i})p(\mu_{l_i}, \sigma^2_{l_i}|data)d\mu_{l_i}d\sigma^2_{l_i}$, for $i = 1, ..., n$, where the subscript l_i is the i^{th} value of the posterior sample of L and μ_{l_i} and $\sigma^2_{l_i}$ are the l_i^{th} posterior samples of μ and σ. Essentially a single $x_{i,rep}$ is sampled from the lognormal distribution at each posterior iteration $b = 1, ..., B$ integrating out all possible values of μ_{l_i} and $\sigma^2_{l_i}$. After obtaining B $x_{i,rep}$'s, we compute the mean $(E(x_{i,rep}|\boldsymbol{x}_{obs}, m_2)$) and variance $(var(x_{i,rep}|\boldsymbol{x}_{obs}, m_2))$ at each i^{th} replicate. Now the penalty and goodness-of-fit terms can be computed via the definition of the criterion. For the ad libitum group we obtained $G(m_2)^a = 393819.1$ and $P(m_2)^a = 1569166$, and for the restricted group $G(m_2)^r = 1561413$ and $P(m_2)^r = 5595397$.

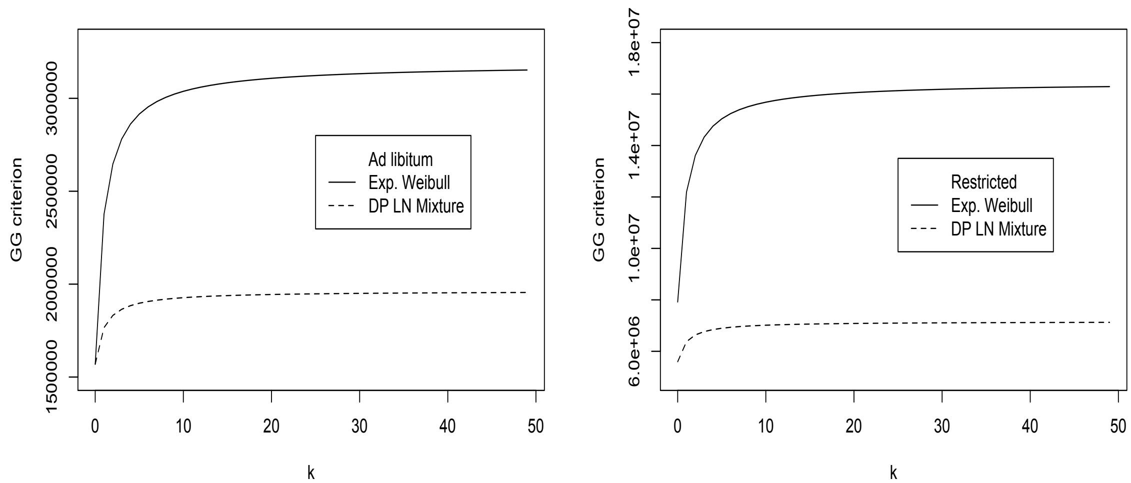

Figure 3.3: Values of the posterior predictive loss criterion for comparison between the parametric exponentiated Weibull model (solid lines) and nonparametric lognormal DP mixture model (dashed lines).

Figure 3.3 is a plot of the criterion values over a grid of k values. For both groups the nonparametric lognormal DP mixture model performs significantly better than the exponentiated Weibull model. The results of the formal model comparison support our earlier argument that the nonparametric lognormal DP mixture model is indeed a better model for these data compared to the exponentiated Weibull model.

Chapter 4

Discussion and Conclusion

We began this document by presenting some basic properties and essential characteristics of the mrl function, showing in particular that the survival distribution is completely defined by the mrl function via the Inversion Formula (2.3). We next presented an easy-to-work with (yet limiting) class of distributions that correspond to a linear mrl function. We provided methods for obtaining the mrl function of several common distributions allowing us to study the various shapes of the mrl function. We find that the form of the mrl function for these distributions is again limited. Knowledge of the form of the mrl function would need to be available in order to select a proper model for mrl inference. The exponentiated Weibull model shows more promise in inference for the mrl function. The mrl function corresponding to the exponentiated Weibull distribution is able to take on several forms, namely constant, linear, increasing, decreasing, BT, and UBT. Another benefit of the exponentiated Weibull distribution is that it has a closed form for its survival function. This helps lower numerical error in estimating the mrl function. Also, the extension to censored data follows very naturally. The likelihood under the exponentiated Weibull model (3.1) with observed survival times, $x_1, ..., x_r$, and censored survival times, $x_{r+1}, ..., x_n$, would be given by

$$f(\boldsymbol{x}|\alpha,\theta,\sigma) = \prod_{i=1}^{r} f(x_r|\alpha,\theta,\sigma) \times \prod_{j=r+1}^{n} \left(F(b_{x_j}|\alpha,\theta,\sigma) - F(a_{x_j}|\alpha,\theta,\sigma)\right)$$

where a_{x_j} is the minimum known survival time and b_{x_j} is the maximum known survival time. In the case of right censoring, $F(b_{x_j}|\alpha,\theta,\sigma) = 1$, and for left censoring, $F(a_{x_j}|\alpha,\theta,\sigma) = 0$.

We fit the exponentiated Weibull model to a data set with two experimental groups consisting of fully observed survival times. The model was able to capture some of the skewness observed in the data, but the unimodality of its density proved to be restrictive. We also fit a nonparametric lognormal DP mixture model to the two groups. The posterior inference results captured the shape of the data much better than the exponentiated Weibull model. The drawback in working with the nonparametric lognormal DP mixture model (3.4) is that the survival function is not available in closed form, rather it is approximated over a grid of survival times by a weighted sum of the survival values of each component of the model at each grid point. Hence the mrl function is also approximated by an operation involving summations. The extension to censoring is also available. For a censored observation x_i the contribution to the likelihood would be given by

$$\sum_{l=1}^{N} p_l \left(\Phi\left(\frac{log(b_{x_i})-\mu_l}{\sigma_l}\right) - \Phi\left(\frac{log(a_{x_i})-\mu_l}{\sigma_l}\right)\right)$$

where a_{x_i} is the minimum and b_{x_i} is the maximum known survival times. For right censoring, $\Phi\left(\frac{log(b_{x_i})-\mu_l}{\sigma_l}\right) = 1$, and for left censoring, $\Phi\left(\frac{log(a_{x_i})-\mu_l}{\sigma_l}\right) = 0$. Of course, under this setting we no longer have conditional conjugacy for all of the parameters, so more general MCMC methods will be needed.

A practically important future research direction will seek to address the question of how to model the mrl function directly under a Bayesian framework. In application, there is often more interest in the mrl function over the survival function, hence it would be practically useful to have a prior for the mrl function directly. To the best of our knowledge, there is no methodology that has been established for this approach of inference for the mrl function. The support of the nonparametric prior would have to be functions that satisfy the properties stated in the Characterization Theorem of the mrl function. Under such a prior, inference for the entire distribution would be obtainable by using the Inversion Formula.

Bibliography

[1] Belkacem Abdous and Alexandre Berred. Mean residual life estimation. *Statistical Planning and Inference*, 132:3–19, 2005.

[2] Roger L. Berger, Dennis D. Boos, and Frank M. Guess. Tests and confidence sets for comparing two mean residual life functions. *Biometrics*, 44:103–115, March 1988.

[3] Y. Q. Chen and S. Cheng. Semiparametric regression analysis of mean residual life with censored data. *Biometrika*, 92(1):19–29, 2005.

[4] R.J. Conner and J.E. Mosemann. Concepts of independence for proportions with a generalization of the dirichlet distribution. *Journal of the American Statistical Association*, 64:194–206, 1969.

[5] Thomas S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973.

[6] M. S. Finkelstein. On the shape of the mean residual lifetime function. *Applied Stochastic Models in Business and Industry*, 18(2):135–146, 2002.

[7] Alan E. Gelfand and Sujit K. Ghosh. Model choice: A minimum posterior predictive loss approach. *Biometrika*, 85(1):1–11, 1998.

[8] K.K. Govilt and K.K. Aggarwal. Mean residual life function of normal, gamma and lognormal densities. *Reliability Engineering*, 5:47–51, 1983.

[9] Frank Guess and Frank Proschan. Mean residual life: theory and applications. Technical Report 85-178, North Carolina State University and Florida State University, Tallahassee,Florida, June 1985.

[10] Rameh C. Gupta and H. Olcay Akman. Mean residual life functions for certain types of non-monotonic ageing. *Commun. Statist.-Stochastic Models*, 11(1):219–225, 1995.

[11] Ramesh C. Gupta, Olcay Akman, and Sergey Lvin. A study of log-logistic model in survival analysis. *Biometrical Journal*, 41(4):431–443, 1999.

[12] W. J. Hall and Jon A. Wellner. Estimation of mean residual life. University of Rochester, January 1979.

[13] W. J. Hall and Jon A. Wellner. Mean residual life. In M. Csörgö, D.A. Dawson, J.N.K. Rao, and A.K.Md.E. Saleh, editors, *Statistics and Related Topics*. North-Holland Publishing Company, 1981.

[14] Hemant Ishwaran and Lancelot F. James. Gibbs sampling methods for stick-breaking priors. *American Statistical Association*, 96(453):161–173, 2001.

[15] Wesley O. Johnson. Survival analysis for interval data. In P. Grambsch and S. Geisser, editors, *Institute for Mathematics and Its Applications: Statistics in the Health Sciences: Diagnosis and Prediction*, volume 114, pages 75–90. Springer-Verlag, 1999.

[16] John P. Klein and Melvin L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Statistics for Biology and Health. Springer, 1997.

[17] Subhash C. Kochar, Hari Mukerjee, and Francisco J. Samaniego. Estimation of a monotone mean residual life. *The Annals of Statistics*, 28(3):905–921, 2000.

[18] Parthasarthi Lahiri and Dong Ho Park. Nonparametric Bayes and empirical Bayes estimators of mean residual life. *Journal of Statistical Planning and Inference*, 29:125–136, 1991.

[19] Gangaji Maguluri and Cun-Hui Zhang. Estimation in the mean residual life regression model. *Journal of the Royal Statistical Society Series B*, 56(3):477–489, 1994.

[20] Govind S. Mudholkar and Deo Kumar Strivasta. Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions of Reliability*, 42(2):299–302, June 1993.

[21] David Oakes and Tamraparni Dasu. A note on mean residual life. *Biometrika*, 77(2):409–410, 1990.

[22] J. Sethuraman. A constructive definition of dirichlet priors. *Statistica Sinica*, 4:639–650, 2001.

[23] Peter J. Smith. Analysis of failure and survival data. In C. Chatfield, Jim Lindsey, Martin Tanner, and J. Zidek, editors, *Texts in Statistical Science Series*. Chapman and Hall/CRC, 2002.

[24] M. Xie, T.N. Goh, and Y. Tang. On changing points of mean residual life and failure rate function fro some generalized Weibull distributions. *Reliability Engineering and System Safety*, 84:293–299, 2004.

[25] Grace L. Yang. Estimation of a biometric function. *The Annals of Statistics*, 6(1):112–116, January 1978.

Appendix A

Proofs

A.1 Equalities and Bounds of MRL

Below we provide the proofs for the equalities and bounds, respectively, of the mrl function stated in Section 2.1.2.

(i) $m(x) + x = \frac{\int_x^\infty (t-x)f(t)dt}{S(x)} + x = \left[\int_x^\infty tf(t)dt - x\int_x^\infty f(t)dt + xS(x)\right]/S(x)$
$= \left[\int_x^\infty tf(t)dt - x\int_x^\infty f(t)dt + x\left(1 - \int_0^x f(t)dt\right)\right]/S(x)$
$= \left[\int_x^\infty tf(t)dt - x\int_0^\infty f(t)dt + x\right]/S(x) = \left[\int_x^\infty tf(t)dt - x + x\right]/S(x)$
$= \left[\int_x^\infty tf(t)dt\right]/S(x) = E(X|X > x)$.

(ii) From (i) we have $(m(x) + x)\,S(x) = E(X|X > x)S(x) = \frac{\int_x^\infty tf(t)dt}{S(x)}S(x) = \int_x^\infty tf(t)dt = E\left(X \cdot 1_{(X>x)}\right)$.

(iii) $E\left(X \cdot 1_{(X>x)}\right) = \int_x^\infty tf(t)dt$ (for X > x, and = 0 o.w.) $= \int_0^\infty tf(t)dt - \int_0^x tf(t)dt = \mu - E\left(X \cdot 1_{(X\le x)}\right)$

(iv) $E\left(X \cdot 1_{(X>x)}\right) = \int_x^\infty tf(t)dt$ (for X > x, and = 0 o.w.) $\overset{\text{since } t\le T}{\le} T\int_x^\infty f(t)dt = TS(x)$.

(v) $E\left(X \cdot 1_{(X>x)}\right) = \int_x^\infty tf(t)dt$ (for X > x, and = 0 o.w.) $\le \int_0^\infty tf(t)dt = \mu$.

(vi) For this proof we make use of Holder's inequality: for r.v. X and Y, $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$, $E(XY) \leq [E(X^p)]^{\frac{1}{p}} [E(Y^q)]^{\frac{1}{q}}$. Using the following substitutions: $p = r$, $q = (1 - \frac{1}{r})^{-1}$, $Y = 1_{(X>x)} \Rightarrow E\left(X \cdot 1_{(X>x)}\right) \leq [E(X^r)]^{\frac{1}{r}} \left[E\left((1_{(X>x)})^{(1-\frac{1}{r})^{-1}}\right)\right]^{(1-\frac{1}{r})}$. This leaves us to show that $S(x)^{(1-\frac{1}{r})} = E\left((1_{(X>x)})^{(1-\frac{1}{r})^{-1}}\right)^{(1-\frac{1}{r})}$. So, $S(x)^{(1-\frac{1}{r})} = \left[\int_x^\infty f(t)dt\right]^{1-\frac{1}{r}} = \left[\int_0^\infty 1_{(X>x)} f(t)dt\right]^{1-\frac{1}{r}} = \left[E(1_{(X>x)})\right]^{1-\frac{1}{r}} = \left[E((1_{(X>x)})^{(1-\frac{1}{r})^{-1}})\right]^{1-\frac{1}{r}}$.

(vii) $E\left(X \cdot 1_{(X \leq x)}\right) = \int_0^x t f(t)dt \overset{\text{since } t \leq x}{\leq} x \int_0^x f(t)dt = xF(x)$.

(viii) Using the substitutions as in the proof for (vi), we need to show that: $F(x)^{(1-\frac{1}{r})} = \left[E\left((1_{(X \leq x)})^{(1-\frac{1}{r})^{-1}}\right)\right]^{1-\frac{1}{r}}$. So, $F(x)^{1-\frac{1}{r}} = \left[\int_0^x f(t)dt\right]^{1-\frac{1}{r}} = \left[\int_0^\infty 1_{(X \leq x)} f(t)dt\right]^{1-\frac{1}{r}} = \left[E\left(1_{(X \leq x)}\right)\right]^{1-\frac{1}{r}} = \left[E\left((1_{(X \leq x)})^{(1-\frac{1}{r})^{-1}}\right)\right]^{1-\frac{1}{r}}$.

Turning to the proofs for the bounds of the mrl function, we have the following derivations.

(a) $m(x) \leq (T - x)^+$ for all x, with equality iff $F(x) = F(T^-)$ or 1:

INEQUALITY: **Case 1** ($x < T$): $m(x) \leq (T - x)^+ \Leftrightarrow m(x) + x \leq T \overset{(i)}{\Leftrightarrow} E\left(X \cdot 1_{(X>x)}\right) \leq T$ The last inequality is always true since $P(x > T) = 0$. **Case 2** ($x \geq T$ or $x = T^-$) in the case where T is infinite: $m(x) \leq (T - x)^+ \Rightarrow m(x) \leq 0$ since the mrl is defined on R^+ $\Rightarrow m(x) = 0$. **EQUALITY**: **Forward Direction** Let $m(x) = (T - x)^+$ **Case 1** ($x < T$): then, $m(x) = (T - x)$, but we have $m(x) = \int_x^T (t - x) f(t)dt = [(t - x)F(t)]_x^T - \int_x^T F(t)dt = (T - x)F(T^-) - (x - x)F(x) - \int_x^T F(t)dt = (T - x) - \int_x^T F(t)dt \neq (T - x)$. Since $\int_x^T F(t)dt > 0$ when $x < T$ we do not have equality when $x < T$. **Case 2** ($x \geq T$) or $x = T^-$ in the case where T is infinite: $/m(x) \equiv m(T) = \int_T^T (t - T) f(t)dt = 0 = (T - T)^+ \equiv (T - x)^+$ and $F(x) = F(T) = 1$ the same argument holds for T^-.. Hence, when we have quality $F(x) = F(T^-)$ or 1 **Backward Direction** Let $F(x) = F(T^-)$ or 1: $\Rightarrow S(x) = 0 \Rightarrow m(x) = 0 = (T - T)^+$. This completes the if and only if argument.

(b) $m(x) \leq \frac{\mu}{S(x)} - x$ for all x with equality iff $F(x) = 0$:

INEQUALITY: $m(x) \leq \frac{\mu}{S(x)} - x \Leftrightarrow (m(x) + x)S(x) \leq \mu \overset{(ii)}{\Leftrightarrow} E\left(X \cdot 1_{(X>x)}\right) \leq \mu \Leftrightarrow \int_x^\infty t f(t)dt \leq \int_0^\infty t f(t)dt$. The last inequality is always true since $\int_0^x t f(t)dt \geq 0$. **EQUALITY**: **Forward Direction**. Suppose $m(x) = \frac{\mu}{S(x)}$, from the inequality we have $\Leftrightarrow \int_x^\infty t f(t)dt = \int_0^\infty t f(t)dt$ thus $x \equiv 0$ so that $F(x) = 0$. **Backward Direction**.

Suppose F(x) = 0. Then this implies that $x \equiv 0$ so that $\int_x^\infty tf(t)dt = \int_0^\infty tf(t)dt$ and thus $m(x) = \frac{\mu}{S(x)} - x$.

(c) $m(x) < \left(\frac{\nu_r}{S(x)}\right)^{\frac{1}{r}} - x$ for all y and any $r > 1$:
$m(x) < \left(\frac{\nu_r}{S(x)}\right)^{1/r} - x \Leftrightarrow [m(x)+x] < \left(\frac{\nu_r}{S(x)}\right)^{1/r} \Leftrightarrow [m(x)+x]\,S(x) < \left(\frac{\nu_r}{S(x)}\right)^{1/r} S(x) \overset{(ii)}{\Leftrightarrow}$ $E\left(X1_{(X>x)}\right) < \left(\frac{\nu_r}{S(x)}\right)^{1/r} S(x)$. From (vi) we have that $E\left(X1_{(X>x)}\right) \leq \left(\frac{\nu_r}{S(x)}\right)^{1/r} S(x)$. Equality for Holder's Theorem is present when for all r, $\nu_r < \infty$, there exists constants c_1 and c_2 not both zero such that $c_1 X^r = c_2 \left(1_{(X>x)}\right)^{(1-1/r)^{-1}}$ for all values of x and any $r > 1$ for $X \leq x \Rightarrow \left(1_{(X>x)}\right)^{(1-1/r)^{-1}} = 0 \Rightarrow c_1 X^r = 0$. Since F is nonnegative and non degenerate $E(X^r) > 0$, then $c_1 = 0$ and c_2 can be any nonzero constant and the equality holds. For $X > x \Rightarrow \left(1_{(X>x)}\right)^{(1-1/r)^{-1}} = 1 \Rightarrow c_1 X^r = c_2$ but since $c_1 = 0$, then that leaves $c_2 = 0$ and the equality does not hold for all x. Therefore we only have a strict inequality for (c).

(d) $m(x) \geq \frac{(\mu-x)^+}{S(x)}$ for $x < T$ with equality iff $F(x) = 0$:
Note that $x < T$ so that $S(x) > 0$.
INEQUALITY: **Case 1** $(x > \mu)$: $\Rightarrow (\mu - x)^+ = 0 \Rightarrow m(x) \geq 0$ which is true since the mrl function is nonnegative by definition. **Case 2** $(x \leq \mu)$: $\Rightarrow (\mu - x)^+ = \mu - x, \Rightarrow m(x) \geq \frac{\mu - x}{S(x)} \Leftrightarrow m(x)S(x) + x \geq \mu \Leftrightarrow S(x)\,(m(x)+x) - xS(x) + x \geq \mu \overset{(iii)}{\Leftrightarrow}$ $\mu - E\left(X \cdot 1_{(X\leq x)}\right) - xS(x) + x \geq \mu \Leftrightarrow -E\left(X \cdot 1_{(X\leq x)}\right) + x(1 - S(x)) \geq 0 \Leftrightarrow xF(x) \geq E\left(X \cdot 1_{(X\leq x)}\right)$ which is true by (vii).
EQUALITY: **Forward Direction** Suppose $m(x) = \frac{(\mu-x)^+}{S(x)}$. **Case 1** $(x > \mu)$: $\Rightarrow (\mu - x)^+ = 0 \Rightarrow m(x) = 0$, but $m(x)$ is zero iff $x \geq T$ (or degenerate at μ), but here $x < T$. Therefore $m(x) \neq (\mu - x)^+/S(x)$ when $x \geq \mu$. **Case 2** $(x \leq \mu) \Rightarrow m(x) = \frac{\mu - x}{S(x)} \Leftrightarrow m(x)S(x) + x = \mu \overset{\text{from (d) INEQ. Case 2}}{\Longleftrightarrow} xF(x) = E\left(X \cdot 1_{(X\leq x)}\right) \Leftrightarrow xF(x) \overset{\text{by partial fractions}}{=} xF(x) - \int_0^x F(t)dt \Leftrightarrow \int_0^x F(t)dt = 0$. Since F(x) is nonnegative this is only true when $F(x) = 0$. **Backward Direction** Suppose $F(x) = 0 \Rightarrow S(x) = 1 \Rightarrow m(x) = \int_x^\infty (t-x)f(t)dt = \int_x tf(t)dt - x\int_x^\infty f(t)dt = \mu - xS(x) = \mu - x$. Thus, we have equality when $F(x) = 0$. Note that $\mu \geq x$ since $m(x) \geq 0$.

(e) $m(x) > \frac{\mu - F(x)\left(\frac{\nu_r}{F(x)}\right)^{\frac{1}{r}}}{S(x)} - x$ for $x < T$ and any $r > 1$:
$m(x) > \frac{\mu - F(x)\left(\frac{\nu_r}{F(x)}\right)^{\frac{1}{r}}}{S(x)} - x \Leftrightarrow (m(x)+x)S(x) > \mu - F(x)\left(\frac{\nu_r}{F(x)}\right)^{1/r} \overset{(ii)}{\Leftrightarrow} \mu - E\left(X \cdot 1_{(X\leq x)}\right) >$ $\mu - F(x)\left(\frac{\nu_r}{F(x)}\right)^{1/r} \overset{(iii)}{\Leftrightarrow} E\left(X \cdot 1_{(X\leq x)}\right) > F(x)\left(\frac{\nu_r}{F(x)}\right)^{1/r}$. From (viii) we know that this is true, to show that we only have a strict inequality here, we proceed as in (c) with

showing that there does not exist two constants c_1, c_2 that are not both nonzero such that $c_1 X^r = c_2 \left(1_{(X \leq x)}\right)^{(1-1/r)^{-1}}$ for $X > x \Rightarrow \left(1_{(X \leq x)}\right)^{(1-1/r)^{-1}} = 0 \Rightarrow c_1 X^r = 0 \Rightarrow$ $c_1 = 0, c_2 \neq 0$ for $X \leq x \Rightarrow \left(1_{(X \leq x)}\right)^{(1-1/r)^{-1}} = 1 \Rightarrow c_1 X^r = c_2$ but $c_1 = 0 \Rightarrow c_2 = 0$. Thus the equality for (e) does not hold.

(f) $m(x) \geq (\mu - x)^+$ for all x, with equality iff $F(x) = 0$ or 1:
INEQUALITY: **Case 1** $(x > \mu)$: $\Rightarrow (\mu - x)^+ = 0 \Rightarrow m(x) \geq 0$ is true by definition of mrl function. **Case 2**: $(x \leq \mu) \Rightarrow (\mu - x)^+ = \mu - x \overset{\text{from } (d)}{\Rightarrow} m(x) \geq (\mu - x)/S(x) \overset{S(x) \leq 1}{\geq}$ $\mu - x$. Therefore, the inequality holds.
EQUALITY: **Forward Direction** Suppose $m(x) = (\mu - x)^+$. **Case 1** $(x > \mu)$ $\Rightarrow (\mu - x) = 0 \Rightarrow m(x) = 0$ which is only true for $x \geq T$ or $T^- \Rightarrow F(x) = 1$ or $F(T^-)$. **Case 2** $(x \leq \mu)$: $\Rightarrow (\mu - x)^+ = \mu - x \Rightarrow m(x) = \mu - x \Leftrightarrow m(x) + x = \mu \overset{(iii)}{\Leftrightarrow} E(X|X > x) = \mu$, which is true only when $F(x) = 0$. **Backward Direction**: **Case 1:** Suppose $F(x) = 0 \Rightarrow x < \mu \Rightarrow m(x) = \mu - x \Leftrightarrow m(x) + x = \mu \overset{(i)}{\Leftrightarrow} E(X|X > x) = \mu$ which is true for x such that $F(x) = 0$. **Case 2**: Suppose $F(x) = F(T^-)$ or $1 \Rightarrow \mu < x \Rightarrow (\mu - x)^+ = 0$ Also, since $S(x) = 0 \Rightarrow m(T) = 0$. Therefore we have equality.
DEGENERATE: **EQUALITY**: If F is degenerate at μ, $m(x) = (\mu - x)^+$. Suppose that F is degenerate $\Rightarrow X = \mu \Rightarrow (\mu - x)^+ = 0$ Also, $X = T \Rightarrow S(x) = 0 \Rightarrow m(x) = 0$. Therefore we have the equality.

A.2 properties of MRL

Below we provide the proofs for the properties of the mrl function stated in Section 2.1.3.

(a) m is a nonnegative and right-continuous, and $m(0) = \mu > 0$:
NON-NEGATIVE: Since $0 \leq F(x) \leq 1 \Rightarrow 0 \leq 1 - S(x) \leq 1 \Rightarrow 0 \leq S(x) \leq 1$. Therefore, $S(x)$ is non-negative. Now consider when $x \geq T$, then $S(x) \equiv 0$, so $m(x) \equiv 0$. For $x < T \Rightarrow S(x) > 0$ thus $\int_x^\infty S(t)dt > 0$. Hence $m(x) = \frac{\int_x^\infty S(t)dt}{S(x)} \geq 0$.
RIGHT-CONTINUITY: We know that $F(x)$ is right-continuous (ie. $\lim_{h \to 0^+} F(x+h) = F(x)$). Now, $\lim_{h \to 0^+} S(x+h) = \lim_{h \to 0^+} (1 - F(x+h)) = 1 - lim_{h \to 0^+} F(x+h) = 1 - F(x) = S(x)$. Hence $S(x)$ is right-continuous as well. If $S(x)$ is right-continuous, then its integral must also be right-continuous (i.e., the limit, $lim_{h \to 0^+} \left[\int_{x+h}^\infty S(t)dt\right] = \int_x^\infty S(t)dt$). Finally, $lim_{h \to 0^+} m(x+h) = lim_{h \to 0^+} \left[\frac{\int_{x+h}^\infty S(t)dt}{S(x+h)}\right] = \frac{\int_x^\infty S(t)dt}{S(x)} = m(x)$, thus $m(x)$ is right-continuous.

FIRST MOMENT STRICTLY POSITIVE: From equation (2.1) we have established that $\mu = m(0)$. Further, $m(0) = \frac{\int_0^\infty S(t)dt}{S(0)} = \int_0^\infty S(t)dt$, which must be greater than 0 because $S(t)$ is nonnegative for all $0 \le t < \infty$ and $S(t+\epsilon) - S(t-\epsilon) > 0$ for at least one value of t and $\epsilon > 0$ in the domain. Therefore, $m(0) \equiv \mu > 0$.

(b) $v(x) \equiv m(x) + x$ is non-decreasing:
Let $h > 0$. **Case 1** ($x + h < T$): $\Rightarrow v(x+h) - v(x) = m(x+h) + (x+h) - m(x) - x = m(x+h) - m(x) + h = \frac{\int_{x+h}^\infty S(t)dt}{S(x+h)} - \frac{\int_x^\infty S(t)dt}{S(x)} + h$. Since $S(x)$ is monotone decreasing then $S(x+h) \le S(x)$ so the former expression is $\ge \frac{\int_{x+h}^\infty S(t)dt}{S(x)} - \frac{\int_x^\infty S(t)dt}{S(x)} + h = -\frac{\int_x^{x+h} S(t)dt}{S(x)} + h$ we need to show that this expression is nonnegative. Assume that it is, $\Leftrightarrow h \ge \frac{\int_x^{x+h} S(t)dt}{S(x)} \Leftrightarrow \int_x^{x+h} S(t)dt \le hS(x)$ this is true since the survival function is non-increasing. Hence, $v(x+h) - v(x) \ge 0 \Rightarrow v(x)$ is non-decreasing. **Case 2** ($x < T \le x + h$): $\Rightarrow v(x+h) - v(x) \overset{\text{from Case 1}}{=} \frac{\int_{x+h}^\infty S(t)dt}{S(x+h)} - \frac{\int_x^\infty S(t)dt}{S(x)} + h$, but the first integral is 0 since $x + h > T$. Thus, the expression becomes $-\frac{\int_x^\infty S(t)dt}{S(x)} + h = -\frac{\int_x^T S(t)dt}{S(x)} + h$. Again we need to show that this expression is nonnegative. Assuming that it is $\Leftrightarrow \int_x^{x+h} S(t)dt \le hS(t)$, which is true since the survival function in non-increasing. Therefore, $v(x+h) - v(x) \ge 0 \Rightarrow v(x)$ is non-decreasing. **Case 3** ($T \le x < x + h$): $\Rightarrow v(x+h) - v(x) = m(x+h) + (x+h) - m(x) - x$, but since $T \le x < x + h \Rightarrow m(x+h) = m(x) = 0$. Thus, $v(x+h) - v(x) = h > 0 \Rightarrow v(x)$ is non-decreasing.

(c) $m(x^-) > 0$ for $x \in (0, T)$; if$T < \infty m(T^-) = 0$ and m is continuous at T:
Part 1: Let $x \in (0, T)$, then $m(x^-) = \frac{\int_{x^-}^T S(t)dt}{S(x^-)}$. Since $S(x^-) < S(T) \le 1 \Rightarrow \frac{\int_{x^-}^T S(t)dt}{S(x^-)} > \int_{x^-}^T S(t)dt$ which is > 0. Therefore, $m(x^-) > 0$.
Part 2: Let $x < T < \infty \Rightarrow v(x) \overset{\text{from (b)}}{\le} v(T) = m(T) + T = T \Rightarrow v(x) = m(x) + x \le T \Leftrightarrow m(x) \le T - x \Rightarrow lim_{x \to T^-} m(x) \le lim_{x \to T^-}(T - x) = T - T^- = 0 \Rightarrow m(T^-) = m(T) = 0$ proving that $m(x)$ is continuous at T.

(e) $\int_0^x \frac{1}{m(t)} dt \to \infty$ as $x \to T$:
Using (e) $\lim_{x \to T} \int_0^x \frac{-k'(t)}{k(t)} dt = -\lim_{x \to T} [log(k(x)) - log(k(0))] = -\lim_{x \to T} log\left[\frac{k(x)}{(k(0)}\right] = log\left[\frac{lim_{x \to T} k(x)}{lim_{x \to T} k(0)}\right]$. Since the limit of the numerator can be found by, $lim_{x \to T} k(x) = lim_{x \to T} S(x) lim_{x \to T} m(x) = 0$, and the denominator is $k(0) = \mu$ which is strictly positive from (a), the limit inside the log function is 0 with convergence from the right. $\Rightarrow lim_{x \to 0^+ log(x)} = -\infty$, hence $\lim_{x \to T} \int_0^x \frac{1}{m(t)} dt = -(-\infty) = \infty$.