University of Bath

PHD

Modelling techniques for time-to-event data analysis

Davis, Alice

Award date:
2018

Awarding institution:
University of Bath

Link to publication

Modelling techniques for time-to-event data analysis

submitted by

Alice Davis

for the degree of Doctor of Philosophy

of the

University of Bath

Department of Mathematical Sciences

May 2018

Declaration of any previous Submission of the Work

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree.

. . . .

Alice Davis

Declaration of Authorship

I am the author of this thesis, and the work described therein was carried out by myself personally.

. . . .

Alice Davis

Summary

This thesis focusses on the cumulative hazard function as a tool for modelling time-to-event data, as opposed to the more common hazard or survival functions. By focussing on and providing a detailed discussion of the properties of these functions a new framework is explored for building complex models from, the relatively simple, cumulative hazards.

Parametric families are thoroughly explored in this thesis by detailing types of parameters for time-to-event models. The discussion leads to the proposal of combination parametric families, which aim to provide flexible behaviour of the cumulative hazard function.

A common issue in the analysis of time-to-event data is the presence of informative censoring. This thesis explores new models which are useful for dealing with this issue.

Acknowledgements

First and foremost I would like to thank my supervisor Karim Anaya-Izquierdo for his continuous support throughout my PhD. He has been immensely enthusiastic and patient with me, always providing excellent guidance. He has been there during the highs and the lows and I could not imagine a better supervisor.

I would like to thank NHS Blood and Transplant, and especially Dave Collett, for supplying the data used in my thesis.

I would like to thank all my friends who have helped me through all these years and who have made it a truly remarkable experience, thank you all for keeping me (relatively) sane. Special thanks go to all the friends I have made during my PhD, who have made my time in Bath truly the best years of my life. Thanks should also go to all the people who have had to share an office with me over the years. I'd like to thank them for everything from answering dull questions to helping me procrastinate. I wish them all the best in their PhDs.

I would especially like to thank my parents Sue and Andrew, and my brother Rory, who have always supported me, through my PhD and in everything building up to it. I could not have done this without them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Methods of survival analysis are used in the modelling of time-to-event data. This thesis uses the cumulative hazard as a modelling tool for such data. The thesis gives a thorough description of the properties of the set of cumulative hazards and thus exploits some of these properties, such as the composition and the inverse, which have not been used for statistical modelling. These properties are used to construct new univariate and multivariate families of cumulative hazards as well as new interpretable regression models. The thesis also uses these properties to unify known models and families under a single framework.

The thesis describes a class of new frailty mixture models that does not require explicit integration. This new class is based on the concept of Bernstein functions and compositions of cumulative hazards. Another key construct in the thesis is the study of time transformations. The study of these allows for simple methods for simulation and interpretation of the known and proposed models.

This thesis provides a simple framework to construct multi-parametric families of cumulative hazards which provide flexible behaviours. Details of the types of behaviours that can be demonstrated by cumulative hazards are given, and a framework is presented for the combinations of desired behaviours. Illustrations of some of the methodologies are presented with a real data set on the time to death or transplant.

A feature of time-to-event data is the presence of censoring, i.e. missing data. Although this is not one of the main concerns of this thesis, censoring will be explored in the last chapter. The proposed new models are used to study the problem of informative censoring.

1.1 Thesis outline

Chapter 2 introduces some basic properties of cumulative hazard functions. It provides the definition of a cumulative hazard and the set of those which are considered in the thesis. This chapter explores functional transformations of cumulative hazards, discussing whether these transformations remain in or leave our set of cumulative hazards. The transformations discussed are the composition, inverse, product, addition, maximum and integration of cumulative hazards. Throughout this chapter we relate how or if these functional transformations correspond to a time transformation. The properties explored in this chapter will form the basis of the framework for generating parametric models which we will create in chapter 3.

Chapter 3 discusses the form of common parametric families in terms of cumulative hazard functions. The parameters from the literature explored in this chapter are the scale, frailty, power, hazard power, tilt and resilience parameters. From this exploration we note common pairing amongst these parameters, and propose additional parameters to complete the pairings and thus the set of parameters. Based on the properties and transformations of chapter 2 we propose some new parametric families which we term *combination parametric families*. We further generalise these families so that they can be used in chapter 6. We then discuss ways in which to add multiple parameters to a family, whether they be traditional or standard families, or whether they are those we have proposed. The latter part of this chapter explores frailty mixtures. In particular, what they are and how to construct them. The key to this section is relating frailty mixtures to compositions of particular cumulative hazard functions.

Chapter 4 reviews well known regression models for modelling time-to-event data. The beginning of this chapter follows the outline of the first section of chapter 3, by considering the regression model corresponding to the one-dimensional parametric families outlined in the previous chapter. It is noted which are found within the literature and which are candidates for new regression models. The latter part of this chapter then demonstrates how to extend the already stated regression models and how to model the baseline cumulative hazard for a fully parametric model.

Chapter 5 introduces the liver transplantation data set. There we explain the covariates present in this data set, giving initial details such as range, mean or factor levels given for each as applicable. Special attention is given to the United Kingdom End-Stage Liver Disease (UKELD) score as this is a covariate based on another model. The motivation and explanation of this model is given for completeness. An exploratory data analysis is carried out in the latter part of this chapter. This analysis comes in four parts. The first part explores the initial connections between the covariates and the general form of the data set. The second part of this analysis fits traditional survival models to this data set, including the Cox proportional hazards model [18] and the Accelerated Failure Time (AFT) model of Kalbfleisch and Prentice [36]. The third part of the analysis then fits the proposed models from chapter 4 and discusses the fit of these models. The final part of the analysis then explores the presence of informative censoring within the liver transplant data set. This then leads naturally to the next chapter which further discusses informative censoring and the issues thereof.

Chapter 6 specifies what informative censoring is and the issues it raises. Part of the chapter gives details on the type of non-informative censoring via the discussion by Williams and Lagakos [74, 40, 39]. It then describes the ways in which these issues are dealt within the literature, linking these to the bivariate models proposed in previous chapters. We discuss bounds on survival functions proposed by Peterson [51]. We then go on to alternatives to this method and discuss sensitivity analyses of Siannis, Copas and Lu [63] and then the bivariate survival models of Roy and Mukherjee [56]. It is these bivariate models that we relate to the models previously proposed. These models are then used to deal with the informative censoring present within the liver transplantation data set.

1.2 Notation

Here we present the key notations of the thesis. Each chapter relies more heavily on certain notations, thus we have divided the notation for ease of referral.

Notation	**Explanation**
Chapter 2	
H	A Cumulative Hazard (CH)
S	A Survival function
f	A density function
h	A hazard function
$T \sim H$	Random variable T has distribution generated by the CH function H
$T \stackrel{d}{=} Y$	The distributions of T and Y are identical
$H_{ll}(t)$	$= \log(1+t)$, standard log-logistic
$H_G(t)$	$= e^t - 1$, standard Gompertz CH
$H_E^\theta(t)$	$= \theta t$, Exponential CH with scale θ
$H_W^\alpha(t)$	$= t^\alpha$, Weibull CH
$H_\Gamma^\kappa(t)$	$= \log\left(1 + \frac{\Gamma(t,\kappa)}{\Gamma(\kappa)-\Gamma(t,\kappa)}\right)$, Gamma CH, where $\Gamma(t,\kappa) = \int_0^t u^{\kappa-1}e^{-u}du$ is the incomplete gamma function
$H_{LN}(t)$	$= -\log\left(1 - \Phi\left(\log(t)\right)\right)$, Log-Normal CH, where Φ is the standard Normal CDF
$H_r^\theta(t)$	$= \frac{t^\theta}{(1+t)^\theta - t^\theta}$
$H_1 \circ H_2(t)$	Composition of H_1 and H_2
$\vec{\tau}$	A time transformation $T_1 = \vec{\tau}(T_0)$
$\vec{H}$	A Cumulative Hazard functional transformation, $H_1(t) = \vec{H}(H_0(t))$
H^{-1}	Inverse of the CH H
$H^{(2)}(t)$	$= H(H(t))$, iterated composition
$H^{(-2)}(t)$	$= H^{-1} \circ H^{-1}(t)$
$H(t)^2$	$= H(t)H(t)$, product of H with itself
$H^{[1]}(t)$	$= \int_0^t H(s)ds$ an N-CH function
$H^{[-1]}(t)$	$= \int_0^t H^{-1}(s)ds$, complementary N-CH function

$H^{-[1]}(t)$	$\left(H^{[1]}\right)^{-1}(t)$, inverse of an N-CH function
$H^{\psi}(t)$	$= \int_0^t \psi(s)ds$

Chapter 3

$H^{\theta,\alpha}_{AB}(t)$	$= H^{\theta}_{A} \circ H^{\alpha}_{B}(t)$
$H^{\alpha}_{A+B}(t)$	$= \alpha H_A(t) + (1-\alpha)H_B(t)$, the linear combination, $C^{\alpha}_{+}(H_A, H_B)$
$H^{\alpha}_{A\cdot B}(t)$	$= H_A(t)^{\alpha} H_B(t)^{1-\alpha}$, the geometric combination, $C^{\alpha}_{\cdot}(H_A, H_B)$
$H^{\alpha}_{A \overset{\rightarrow}{\circ} B}(t)$	$= \frac{1}{\alpha}(H_A \circ H_B^{-1})(\alpha H_B(t))$, the composition combination, $C^{\alpha}_{\overset{\rightarrow}{\circ}}(H_A, H_B)$
$H^{\alpha}_{A \overset{\leftarrow}{\circ} B}(t)$	$= H_B\left(\frac{1}{\alpha} H_B^{-1} \circ H_A(\alpha t)\right)$, the reverse composition combination, $C^{\alpha}_{\overset{\leftarrow}{\circ}}(H_A, H_B)$
$C^{\alpha}(H_A, H_B)$	A combination model of H_A and H_B
U	Unobserved frailty
$\mathcal{L}_F(t)$	Laplace-Stieltjes transform of a distribution F
$H_{(F)}(t)$	$= -\log \mathcal{L}_F(t)$

Chapter 4

$\mathbf{x}_i$	Vector of explanatory variables for individual i
$\boldsymbol{\beta}$	Vector of regression coefficients
$\boldsymbol{\eta}_i = \boldsymbol{\beta}^T \mathbf{x}_i$	Linear predictor for individual i
Θ	Parameter space
$\psi : \mathbb{R} \to \Theta$	Linking function
$H(t\vert\psi, H_0)$	Regression model with linking function ψ and baseline H_0

Chapter 6

T	Event time
C	Censoring time
Y	Observed time, $= \min(T, C)$
Δ	Censoring indicator, $\Delta = 1$ if censored, 0 otherwise
$f_{X,Y}(x, y)$	Joint distribution of some random variables X and Y
$S_{X,Y}(x, y)$	Joint survival of some random variables X and Y
$f_{X\vert Y}(x\vert y)$	Conditional distribution of X given Y
$Q_X(t)$	Crude survival function of some random variable X
$Q^*_X(x)$	Net survival function of some random variable X

Chapter 2

Mathematical Properties of the Cumulative Hazard Function

2.1 Introduction to cumulative hazards

In the statistical analysis of time-to-event data, the usual approach to modelling is to use the survival or hazard functions rather than the density or cumulative distribution function. The most common hazard models are the *Proportional hazards models* [18], other common alternatives are *Accelerated Failure time models.* While these models are interpreted in terms of the hazard function, they can also be easily written down in terms of the cumulative hazard function. In this chapter we focus on properties of the cumulative hazard function which can potentially be used for modelling.

Let T be the continuous random variable representing the time-to-event of an individual of interest and let P be the probability measure associated to the events of this random variable.

Definition 2.1. The function

$$S_T(t) := P\left(T > t\right) \tag{2.1}$$

is the survival function corresponding to the random variable T.

We now introduce our main assumptions on the survival function.

Assumption 2.1. The survival function has value 1 at time 0, $S_T(0) = 1$.

Assumption 2.2. The survival function tends to a value of 0 as time increases, $\lim_{t\to\infty} S_T(t) = 0$.

Assumption 2.3. The survival function $S_T(t)$ is continuously differentiable on $(0, \infty)$.

Assumption 2.4. The survival function $S_T(t)$ is strictly decreasing on $(0, \infty)$.

There are a number of consequences of these assumptions that need to be discussed.

First, it is clear that assumption 2.1 implies that $P(T \leq 0) = 0$. This is a reasonable assumption which rules out negative times to an event. Assumption 2.1 also implies that $P(T = 0) = 0$, which will rule out models that have a positive probability mass at zero. An example of a model with mass at zero would be one that looks at the lifetimes of babies. The mass at zero would be interpreted as a still birth.

Assumption 2.2 excludes cure models. This means that we do not include models where it is possible for the event of interest to never occur. If we are interested in death due to some disease, this may happen when a patient is actually cured of this disease and thus does not die from it.

Assumptions 2.3 and 2.4 imply that

$$f_T(t) := -\frac{dS_T(t)}{dt},$$

is positive and continuous on $(0, \infty)$. The function $f_T(t)$ is called the probability density function of T.

By the Fundamental Theorem of Calculus, for any $b > t$,

$$\int_t^b f_T(s)ds = S_T(t) - S_T(b).$$

Letting $b \to \infty$ and using assumption 2.2, we obtain

$$\int_t^\infty f_T(s)ds = S_T(t) \tag{2.2}$$

which gives an expression for the survival S_T in terms of the density f_T.

Assumption 2.4, in combination with a further assumption 2.5 to be discussed later, will rule out models for which the density f_T is zero at a point, or where the probability of an event happening in a certain interval is zero.

Finally, assumption 2.3 will rule out models for which the density is discontinuous at a point in time. To see this more clearly consider the following example.

Example 2.2. Let S_T be defined by

$$S_T(t) = \begin{cases} 0 & t \leq 0 \\ e^{-t} & 0 < t < 1 \\ e^{-(2t-1)} & t \geq 1, \end{cases}$$

which can be seen in figure 2.1.

$S_T(t)$ is a continuous function but is not differentiable a $t = 1$. It is easy to show

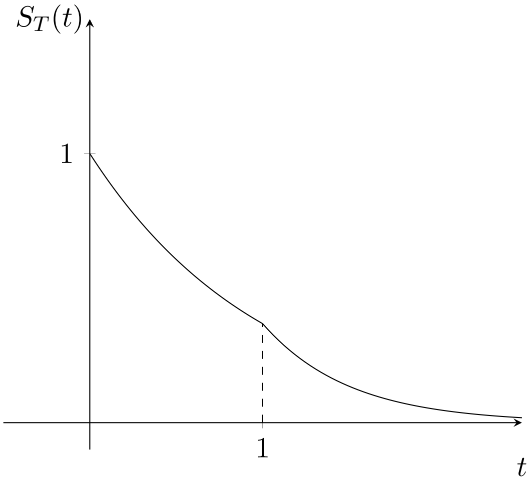

Figure 2.1: Survival function not satisfying assumption 2.3.

that S_T can be expressed as $S_T(t) = \int_t^\infty f_T(s)ds$ where $f_T(t)$ is defined as

$$f_T(t) = \begin{cases} 0 & t < 0 \\ e^{-t} & 0 \leq t < 1 \\ 2e^{-(2t-1)} & t \geq 1, \end{cases}$$

as seen in figure 2.2.

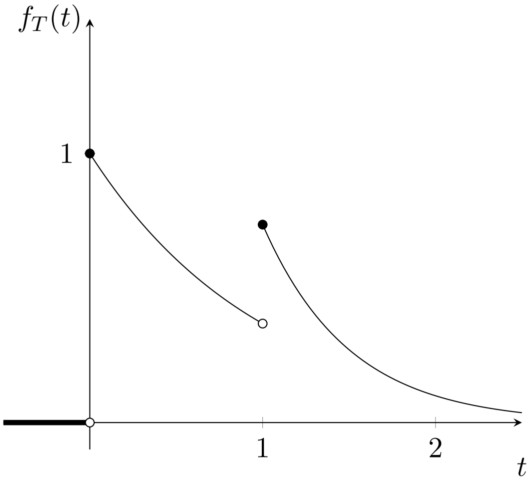

Figure 2.2: Probability distribution function not satisfying assumption 2.3

We note that it is actually irrelevant how we define $f_T(t)$ at the two discontinuities, $t = 0$ and $t = 1$ for the expression 2.2 to hold. The first discontinuity at $t = 0$ is simply a consequence of our assumption 2.1 to have positive times-to-event, thus is not a source of concern. The second discontinuity at $t = 1$ is more concerning and we believe it is unrealistic in many situations. Hence our assumption of continuous differentiability on the survival S_T removes models such as this.

We define $\mathcal{S}$ to be the set of all survival functions that satisfy assumptions 2.1, 2.2, 2.3 and 2.4. Unless otherwise stated we will assume all survival functions are elements of $\mathcal{S}$.

We now define the hazard function, which plays an important role in time-to-event data analysis.

Definition 2.3. The hazard function $h_T(t)$ corresponding to P is defined as

$$h_T(t) := \lim_{\epsilon \searrow 0} \frac{P(t < T \leq t + \epsilon | T > t)}{\epsilon}, \tag{2.3}$$

for all $t \geq 0$.

The hazard function $h_T(t)$ can be written in terms of the survival function and the probability density function. These expressions will now be derived.

$$\begin{aligned}
h_T(t) &= \lim_{\epsilon \searrow 0} \frac{P(t < T \leq t + \epsilon | T > t)}{\epsilon} \\
&= \lim_{\epsilon \searrow 0} \frac{P(t < T \leq t + \epsilon)}{\epsilon P(T > t)} \\
&= \lim_{\epsilon \searrow 0} \frac{S_T(t) - S_T(t + \epsilon)}{\epsilon P(T > t)} \\
&= \frac{-1}{P(T > t)} \lim_{\epsilon \searrow 0} \left(\frac{S_T(t + \epsilon) - S_T(t)}{\epsilon} \right) \\
&\text{which by assumption 2.3} \\
&= -\frac{\frac{d}{dt} S_T(t)}{S_T(t)} = \frac{f_T(t)}{S_T(t)} \qquad (2.4) \\
&= -\frac{d \log S_T(t)}{dt}, \qquad (2.5)
\end{aligned}$$

for $t > 0$.

We note that since both $f_T(t)$ and $S_T(t)$ are continuous on $(0, \infty)$, $h_T(t)$ is continuous on $(0, \infty)$. We have a further assumption that is given in terms of the hazard function:

Assumption 2.5. The hazard function is strictly positive, $h_T(t) > 0$, $\forall t > 0$.

The corresponding hazard function in example 2.2 will not be continuous. The hazard function will be defined in pieces, or piecewise, and will be constant on each piece. The corresponding hazard will be

$$h_T(t) = \begin{cases} 0 & t < 0 \\ 1 & 0 \leq t < 1 \\ 2 & t \geq 1. \end{cases}$$

In general, the probability models with hazard functions such as these are called piecewise constant hazard models [41]. As a consequence of assumption 2.3, we do not pursue these types of models. This assumption could be relaxed so that survival functions, and thus other related functions, are piecewise differentiable. Relaxing this assumption would then allow the inclusion of piecewise constant hazard models. These types of models are not the main focus of this thesis and so this assumption will not be relaxed in this thesis.

Integrating both sides of (2.5) and using the Fundamental Theorem of Calculus, we obtain

$$\begin{aligned}\int_0^t h_T(s)ds &= -\int_0^t \frac{d\log S_T(s)}{ds}ds \\ &= -\left(\log S_T(t) - \log S_T(0)\right) \\ &= -\log S_T(t) \;\text{ by assumption 2.1.}\end{aligned}$$

This implies we can write the survival function uniquely in terms of the hazard function by

$$S_T(t) = e^{-\int_0^t h_T(s)ds}, \quad \forall\, t > 0. \tag{2.6}$$

Our focus will be on the cumulative hazard (CH) function which we define as follows:

Definition 2.4. Let T be a random variable with survival $S_T \in \mathcal{S}$ and corresponding hazard h_T. The cumulative hazard function H_T is defined as

$$H_T(t) = \int_0^t h_T(s)ds, \qquad \forall t > 0.$$

From equation (2.6) we can clearly write the survival function in terms of the cumulative hazard function, namely

$$S_T(t) = e^{-H_T(t)}. \tag{2.7}$$

Note that $H_T(t) = -\log S_T(t)$.

We now notice below some properties of CH functions which are related to, or consequences of, assumptions 2.1 to 2.5.

Property 2.1. The cumulative hazard function is zero at zero, $H_T(0) = 0$ and is otherwise strictly positive for positive t, $H_T(t) > 0$, $\forall t > 0$.

Property 2.2. The cumulative hazard tends to infinity,

$$\lim_{t\to\infty} H_T(t) = \infty.$$

This property will be denoted by $H_T(\infty) = \infty$.

Property 2.3. The cumulative hazard $H_T(t)$ is continuously differentiable on $(0, \infty)$.

Property 2.4. The cumulative hazard function $H_T(t)$ is strictly increasing on $(0, \infty)$.

We define the set $\mathcal{CH}$ as the set of all cumulative hazard functions satisfying properties 2.1 to 2.4. We will assume that any CH function is within this set, unless otherwise stated.

Proposition 2.5. *The survival function $S_T \in \mathcal{S}$ if, and only if, the corresponding cumulative hazard function $H_T \in \mathcal{CH}$.*

Proof. Suppose $S_T \in \mathcal{S}$, then S_T satisfies assumptions 2.1 to 2.5. The combination of all the assumptions imply that $S_T(t) > 0$ for all $t \in (0, \infty)$. Thus as $S_T(0) = 1$ this implies $H_T(0) = 0$, i.e. property 2.1.

Properties 2.2 and 2.3 are direct consequences of assumptions 2.2 and 2.3 respectively.

Assume assumptions 2.4 and 2.5 hold, then

$$\frac{d}{dt}H_t(t) = h_T(t) = -\frac{\frac{d}{dt}S_T(t)}{S_T(t)} > 0$$

since $\frac{d}{dt}S_T(t) < 0$ as S_T is strictly decreasing. Since $h_T(t) > 0$ then property 2.4 holds.

Now assume the properties 2.1 to 2.4 hold. Then $S_T(0) = e^{-H_T(0)} = 1$. Thus assumption 2.1 holds. Now for the second assumption,

$$\lim_{t\to\infty} S_T(t) = e^{-\lim_{t\to\infty} H_T(t)} = 0,$$

thus assumption 2.2 holds. Property 2.3 implies that assumption 2.3 holds. Property 2.4 also implies assumption 2.4. □

Since a CH function uniquely defines a survival function and thus a probability distribution, we can define some notation to denote this. We use

$$T \sim H \tag{2.8}$$

to denote that T follows the probability distribution uniquely defined by the CH function H.

In the next few sections we will see some specific examples of CH functions in $\mathcal{CH}$, have an interpretation of the CH functions and see some properties of the functions in the set $\mathcal{CH}$.

2.1.1 Key cumulative hazard functions

Within the set of CH functions, $\mathcal{CH}$, there are number of important functions we will need to refer back to throughout the course of this thesis. In table 2.1 we highlight which functions we will need to refer back to and define some notation for them.

Note that we have defined a scale parameter via θt rather than the usual t/θ.

Notation	Formula	Explanation
$H_{ll}(t)$	$H_{ll}(t) = \log(1+t)$	Standard log-logistic
$H_G(t)$	$H_G(t) = e^t - 1$	Standard Gompertz
$H_E^\theta(t)$	$H_E^\theta(t) = \theta t$	Exponential with scale θ
$H_W^\alpha(t)$	$H_W^\alpha(t) = t^\alpha$, $\alpha > 0$	Weibull with power parameter α
$H_r^\theta(t)$	$H_r^\theta(t) = \dfrac{t^\theta}{(1+t)^\theta - t^\theta}$	A rational CH with shape parameter θ
$H_\Gamma^\kappa(t)$	$H_\Gamma^\kappa(t) = \log\left(1 + \dfrac{\Gamma(t,\kappa)}{\Gamma(\kappa) - \Gamma(t,\kappa)}\right)$	Gamma with shape parameter κ, $\Gamma(t,\kappa)$ is the incomplete gamma function
$H_{LN}(t)$	$H_{LN}(t) = -\log(1 - \Phi(\log(t)))$	Standard Log-Normal

Table 2.1: Notation for important CH functions.

2.1.2 Baseline transformations

The motivation for the rest of this chapter is to formalise methods for creating new time-to-event models via the transformation of cumulative hazard functions.

As a running example in this chapter, consider a study of individuals who each have a set of recorded covariates. It is assumed that there are some underlying characteristics that all individuals in the study share and that the covariates are believed to explain the differences between the individuals.

The random variable for the time-to-event for a reference individual in the study is denoted by T_0, and we will call it a *baseline time*. A reference individual could be an average or typical individual that the results of the study will make reference or compare to. The cumulative hazard function that defines the distribution of the baseline times is H_0, thus $T_0 \sim H_0$. It is assumed that a particular individual in the study has survival time random variable $T_1 \sim H_1$ and that this relates to the baseline time, T_0, in the sense that there exists some time transformation $\vec{\tau}$ such that

$$T_1 \overset{d}{=} \vec{\tau}(T_0). \tag{2.9}$$

We note that T_0 is a conceptual time random variable which we use to define the actual observable time-to-event T_1 via the transformation $\vec{\tau}$. Also note that this transformation does not imply that T_1 and T_0 are necessarily dependent, only that they are related in a distributional sense. Hence we have used the notation $X \overset{d}{=} Y$ which means that the random variables X and Y have the same probability distribution.

One of the aims of this chapter is to clarify the forms which $\vec{\tau}$ can take. Instead

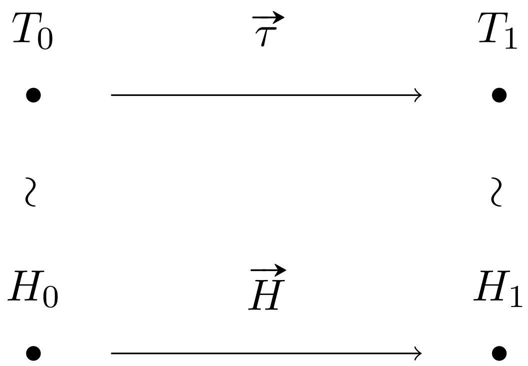

Figure 2.3: Diagram to represent the transformation from baseline time and the corresponding cumulative hazards.

of interpreting the observable time T_1 to be a transformation of the baseline time, we can also think of the baseline cumulative hazard, $H_0(t)$ to be functionally transformed, following the rules of probability, to obtain $H_1(t)$, the cumulative hazard of T_1. The functional transformation will be denoted by $\vec{H}$. This is demonstrated in figure 2.3. Theorem 2.27 will give the details of the transformations $\vec{\tau}$ and $\vec{H}$.

We now see an example of each of the two types of transformations mentioned above. The Cox proportional hazards model [18] states that:

$$H_1(t) = \theta H_0(t),$$

meaning the CH functions are proportional. Thus the transformation $\vec{H}$ is simply multiplying by some positive constant θ. Note that here the functional transformation between CH functions, $\vec{H}$ is the one specified rather than the time transform $\vec{\tau}$.

In the accelerated failure time model [36]

$$T_1 \stackrel{d}{=} T_0/\theta,$$

meaning the survival time of an individual is accelerated or decelerated by a factor of θ of the survival time of a typical individual in the study. If $\theta < 1$ then the effect will be to accelerate the survival time, if $\theta > 1$ then the effect will be to decelerate the time. Note here that we specify the time transformation, $\vec{\tau}$, rather than the functional transformation $\vec{H}$ of CH functions. Both transformations should be equivalent following the rules of probability.

2.2 Interpretation of the cumulative hazard function

The hazard function is a vital tool in time-to-event data analysis. It has a useful and probabilistic interpretation as a conditional probability. From equation (2.3) we can see that the hazard function can be interpreted as an instantaneous time to event rate. However, the cumulative hazard function does not have a probabilistic interpretation.

It is the integral of the hazard function, but it is not an integral over the argument of conditional probabilities where the conditioning event is fixed.

One interpretation of the cumulative hazard can be given in terms of the so called non-informative censoring model. This will be described in detail in chapter 6, so here we only use the necessary details. Let T be the time-to-event random variable and assume that $T \sim H_T$. Let C be the censoring random variable which usually represents the time at which the individual leaves the study, or more generally, the time at which we are no longer able to observe the time-to-event T. We assume that $C \sim H_C$ and the non-informative censoring model assumes that T and C are independent random variables. Let $Y = \min(T, C)$, the actual observed time in the study. The interpretation of the cumulative hazard H_T will be given in terms of the expectation of $H_T(Y)$ where $Y \sim H_Y$. It is easy to show that $H_Y(y) = H_T(y) + H_C(y)$, see proposition 2.52.

Proposition 2.6. *When $T \sim H_T$, T and C are independent and $Y = \min(T, C)$,*

$$E_Y\big(H_T(Y)\big) = P(T < C).$$

Proof.

$$\begin{aligned}
E_Y\big(H_T(Y)\big) &= \int_0^\infty H_T(y) f_Y(y) dy = -\int_0^\infty H_T(y) dS_Y(y) \\
&= -\int_0^\infty H_T(y) \left[\frac{d}{dy} e^{-H_Y(y)}\right] dy \\
&\text{which, by integration by parts,} \\
&= -\left[H_T(y) e^{-H_Y(y)}\right]_0^\infty + \int_0^\infty e^{-H_Y(y)} dH_T(y) \\
&= 0 + \int_0^\infty e^{-H_C(y) - H_T(y)} h_T(y) dy = \int_0^\infty e^{-H_C(y)} f_T(y) dy \\
&= \int_0^\infty f_T(y) S_C(y) dy = \int_0^\infty f_T(y) P(C > y) dy \\
&= \int_0^\infty \int_y^\infty f_T(y) f_C(u) du dy = \int_0^\infty \int_y^\infty f_{T,C}(y, u) du dy \\
&\text{by independence of } T \text{ and } C \\
&= P(T < C).
\end{aligned}$$

□

So then according to proposition 2.6, we can interpret the expected value of the cumulative hazard H_T evaluated at the observed time Y, as the probability of the time-to-event T being observed (rather than being censored) under a non-informative censoring scheme. Note that

$$E_Y\big(H_T(Y)\big) \le E_T\big(H_T(T)\big) = 1$$

since $H_T(T)$ follows an exponential distribution with unit mean. We will use this fact when describing an alternative interpretation of the cumulative hazard function which does not assume any censoring scheme.

Proposition 2.7. *Let $T \sim H_T$, then $H_T(T) \sim H_E^1$.*

Proof.

$$P(H_T(T) > t) = P(T > H_T^{-1}(t)) = \exp(H_T(H_T^{-1}(t))) = \exp(t) = H_E^1(t).$$

□

Another interpretation of the cumulative hazard can be seen in the work of Singpurwalla [64]. The idea of the interpretation is to acknowledge the presence of an underlying random variable $X \geq 0$ which follows an exponential distribution with mean 1.

From equation (2.5) we have that

$$S_T(t) = P(T > t) = e^{-H_T(t)}.$$

This means we can write

$$S_T(t) = e^{-H_T(t)} = P(X > H_T(t)) = S_X(H_T(t)) \tag{2.10}$$

so we can interpret the values of $H_T(t)$ for all $t \geq 0$ as the realised values of the random variable X. More specifically, since, from equation (2.10),

$$P(T \leq t) = P(X \leq H_T(t))$$

we can interpret the time to failure T as the time at which the cumulative hazard $H_T(t)$ crosses a random threshold X, that is,

$$T \stackrel{d}{=} H_T^{-1}(X). \tag{2.11}$$

This in turn gives a way to simulate observations from T for any given choice of H_T. In this way, the hazard potential plays a similar role as the uniform distribution on (0, 1) since $U \stackrel{d}{=} F_X(X) \sim Unif(0,1)$, $\forall F_X$ and $X \sim F_X$ where F_X is continuous.

Definition 2.8. Let $T \geq 0$ be a random variable and let $H_T(t)$ be its corresponding cumulative hazard. Then $X \stackrel{d}{=} H_T(T)$ is defined to be the hazard potential.

Proposition 2.9. *Let $T \sim H_T$ and $X \stackrel{d}{=} H_T(T)$. Then $X \sim H_E^1$.*

The main idea of this interpretation is that the standard exponential distribution of X is completely independent of the context, or setting, of the event of interest. Thus we can interpret X as some unknown resource that has been created at the start time.

In this interpretation, $H_T(t)$ would be the amount of the resource used up by time t and the hazard, $h_T(t) = \frac{d}{dt}H_T(t)$ can be considered as the rate at which the resource is consumed. See Singpurwalla [64] for more details on the hazard potential.

2.3 Analytical properties of the cumulative hazard

The main aim of this chapter is to detail properties of CH functions to find ways of transforming given cumulative hazards, to new ones, and therefore transforming probability distributions to new ones, which will be an essential tool in statistical modelling of time-to-event data.

2.3.1 Composition of cumulative hazards

This section will explore the consequences of composing cumulative hazards, and what properties these types of compositions have. When considering the cumulative hazard H evaluated at time t, we note that t is dimensionless. Thus H is a function from $[0, \infty)$ to $[0, \infty)$. Hence the composition of cumulative hazards is well defined mathematically.

The composition of cumulative hazards can be thought of as the operation that transforms one cumulative hazard function to another. Note that this operation still satisfies all the conditions needed for a function to be a cumulative hazard. We note that all operations involving cumulative hazards can be thought of in the same manner in this chapter.

Proposition 2.10. *The composition of two cumulative hazards is itself a cumulative hazard.*

Proof. Let H_1 and H_2 be cumulative hazards. Since H_1, $H_2 : (0, \infty) \to (0, \infty)$ then $H_1 \circ H_2 : (0, \infty) \to (0, \infty)$. We also have,

$$(H_1 \circ H_2)(0) = H_1(H_2(0)) = H_1(0) = 0,$$

$$\lim_{t\to\infty} H_2(H_1(t)) = H_2(\lim_{t\to\infty} H_1(t)) = H_2(\infty) = \infty$$

as H_1 and H_2 are cumulative hazard functions. Since H_1 and H_2 are strictly increasing then if

$$s < t \iff H_1(s) < H_1(t)$$
$$u < v \iff H_2(u) < H_2(v).$$

Let $u = H_1(s)$ and $v = H_1(t)$, then

$$s < t \iff H_2(H_1(s)) < H_2(H_1(t)).$$

Now for continuous differentiability, we show the derivative is continuous.

$$\frac{d}{dt}H_2(H_1(t)) = h_1(t)h_2(H_1(t))$$

which is the product of continuous functions and so is continuous, thus $H_2(H_1(t))$ is continuously differentiable. □

Example 2.11. Consider composing the log-logistic $H_{ll}(t) = \log(1+t)$ and the Gompertz [44, 33] $H_G(t) = e^t - 1$ in proposition 2.10, then we see that

$$(H_{ll} \circ H_G)(t) = \log(1 + e^t - 1) = t = H_E^1(t),$$

the CH function of the standard exponential distribution.

Example 2.12. Consider two distinct Weibull CHs $H_W^\alpha(t)$ and $H_W^\beta(t)$. The composition of these CH functions yields

$$(H_W^\alpha \circ H_W^\beta)(t) = t^{\alpha\beta},$$

which is another Weibull. Hence, we see that the Weibull family is closed under composition.

Example 2.13. Consider $H_1(t) = \overrightarrow{H}(H_0(t)) = H_G \circ H_0(t)$. Then

$$\begin{aligned} H_1(t) = H_G(H_0(t)) &= e^{H_0(t)} - 1 \\ &= \frac{1 - e^{-H_0(t)}}{e^{-H_0(t)}} = \frac{F_0(t)}{1 - F_0(t)}, \end{aligned}$$

which is the function describing the odds of experiencing the event of interest before time t. This will be relevant later.

The next example refers to a well known survival regression model, the Accelerated Failure Time model. This example details how this model can be constructed using compositions of CH functions, but we don't go into the details of the actual model here. Instead, this model is discussed fully in section 4.1.

Example 2.14. (Accelerated Failure Time model) An important example of the use of the composition transformation is the widely used survival model, the Accelerated Failure Time (AFT) model [36]. Consider a population with an underlying baseline time, $T_0 \sim H_0$, where a particular individual has $T_1 \sim H_1$. In this model the CH functions are related via

$$H_1(t) = H_0(\theta t), \tag{2.12}$$

that is, a transformation of the time scale. As shown in Figure 2.3 we have that $\overrightarrow{H}(H_0(t)) = H_0 \circ H_E^\theta$, that is, composing with the exponential. Then it can be shown that $T_1 \overset{d}{=} \frac{1}{\theta}T_0$ so that $\overrightarrow{\tau}(T_0) = \frac{1}{\theta}T_0$.

In chapter 4 we view models such as the AFT as a linear model. If we take the logs of $T_1 \overset{d}{=} \frac{1}{\theta} T_0$ then we see that $\log T_1 \overset{d}{=} -\log\theta + \log T_0$.

Compositions of CH functions aren't only used in building new CHs, they are also used when generalising distributions and adding new parameters. This is demonstrated in the following example.

Example 2.15. (Scale and Power parameters) The usual accommodation of scale and power parameters over a baseline distribution, corresponds to composing with the Weibull and Exponential CHs. Let $T_0 \sim H_0$ and $T_1 \sim H_1$. Then if

$$\begin{aligned} H_1(t) &= \overrightarrow{H}(H_0(t)) = (H_0 \circ H_E^{\theta} \circ H_W^{\alpha})(t) \\ &= H_0(\theta t^{\alpha}) \end{aligned}$$

we have a scale of θ and a power of α.

The corresponding time transformation will be $T_1 \overset{d}{=} \left(\frac{1}{\theta} T_0\right)^{1/\alpha}$. Note here that a scale of θ where $\alpha > 1$ will result in a deceleration. There is also some trade off in acceleration or deceleration when powering.

Note that in this example the associative property that

$$(H_1 \circ H_2) \circ H_3(t) = H_1 \circ (H_2 \circ H_3(t)).$$

For this to hold we require that

$$\begin{aligned} \text{domain of } H_2 &= \text{codomain of } H_3 \\ \text{domain of } H_1 &= \text{codomain of } H_2. \end{aligned}$$

This is true since $H_i : [0, \infty] \to [0, \infty]$ for $i = 1, 2, 3$.

Proposition 2.16. *Let $H_0 \in \mathcal{CH}$ and let g be a continuously differentiable function. If $H_0 \circ g$ is a cumulative hazard, then g is a CH function.*

Proof. It is assumed that properties 2.1 to 2.4 hold for $H_0 \circ g$. Thus $H_0(g(0)) = 0$, thus we must have that $g(0) = 0$. We also have that H_0 is only defined, or at least positive, on $(0, \infty)$, thus $g(t) > 0$ for all $t \in (0, \infty)$.

If H_0 and g are continuously differentiable, then so must $H_0 \circ g$ be. If $H_0(g(\infty)) = \infty$ and $H_0(\infty) = \infty$, then $g(\infty) = \infty$. If $H_0 \circ g$ is strictly increasing, then

$$\frac{d}{dt} H_0(g(t)) = \frac{dg(t)}{dt} h_0(g(t)) > 0$$

which implies that $\frac{d}{dt} g(t) > 0$, thus g is strictly increasing. Hence g satisfies properties 2.1 to 2.4 and so is a CH function. □

This proposition will be used later in section 2.3.10 as a method to generalise the Accelerated Failure Time model to include time varying covariates. We will then have a way to impose conditions on our covariates so that this model is valid

Iterative compositions of cumulative hazards

We have seen that the composition of CH functions results in a CH function. We now explore the effect of repeatedly composing the same CH function with itself.

We will define notation for the iterated composition. Let H be a CH function, then

$$H^{(2)}(t) := H(H(t)). \tag{2.13}$$

This can be easily generalised, composing H with itself $n > 1$ times is denoted $H^{(n)}$.

Proposition 2.17. *Suppose $H_T(t)$ is a CH function. Iteratively composing with $H_T(t)$ creates a function that gets progressively further from the identity t. I.e., if $H_T^{(n)}(t)$ is the composition of H_T n times, then for all points that are not fixed, $H_T(t) \neq t$, $H_T^{(n+1)}(t) > H_T^{(n)}(t)$ or $H_T^{(n)}(t) < H_T^{(n-1)}(t)$.*

Proof. Suppose t is not a fixed point of H_T, i.e. $H_T(t) \neq t$. Then we have two cases, $H_T(t) > t$ and $H_T(t) < t$.

Case 1: $H_T(t) > t$ Let $s = H_T(t)$. Since H_T is increasing, we have that

$$H_T(s) = H_T^{(2)}(t) > H_T(t) > t.$$

$$\begin{aligned} H_T(t) > t \Rightarrow H_T^{(2)}(t) > H_T(t) \Rightarrow H_T^{(2)}(t) > H_T(t) > t \\ \Rightarrow H_T^{(2)}(t) - t > H_T(t) - t > 0 \Rightarrow |H_T^{(2)}(t) - t| > |H_T(t) - t|, \end{aligned}$$

Thus the result holds for $n = 1$. Suppose it holds for n, thus $u = H_T^{(n+1)}(t) > H_T^{(n)}(t) = v$. Then $H_T(u) > H_T(v)$ since H_T is increasing and so the result holds for $n + 1$. So the result holds by induction when $H_T(t) > t$.

Case 2: $H_T(t) < t$

$$s = H_T(t) < t \Rightarrow H_T(s) = H_T^{(2)}(t) < H_T(t) \Rightarrow H_T^{(2)}(t) < H_T(t) < t,$$

and so using the same approach as above the result holds for this case. □

The best way to illustrate the effects described in proposition 2.17 would be graphically. This can be seen in the next example.

Example 2.18. To illustrate the iterative compositions, see Figure 2.4 where the function $H_{ll}(t) = \log(1 + t)$ is composed with itself multiple times. We see that with each composition, the function gets further from the identity and closer to the t-axis, i.e., with more compositions the function grows more slowly.

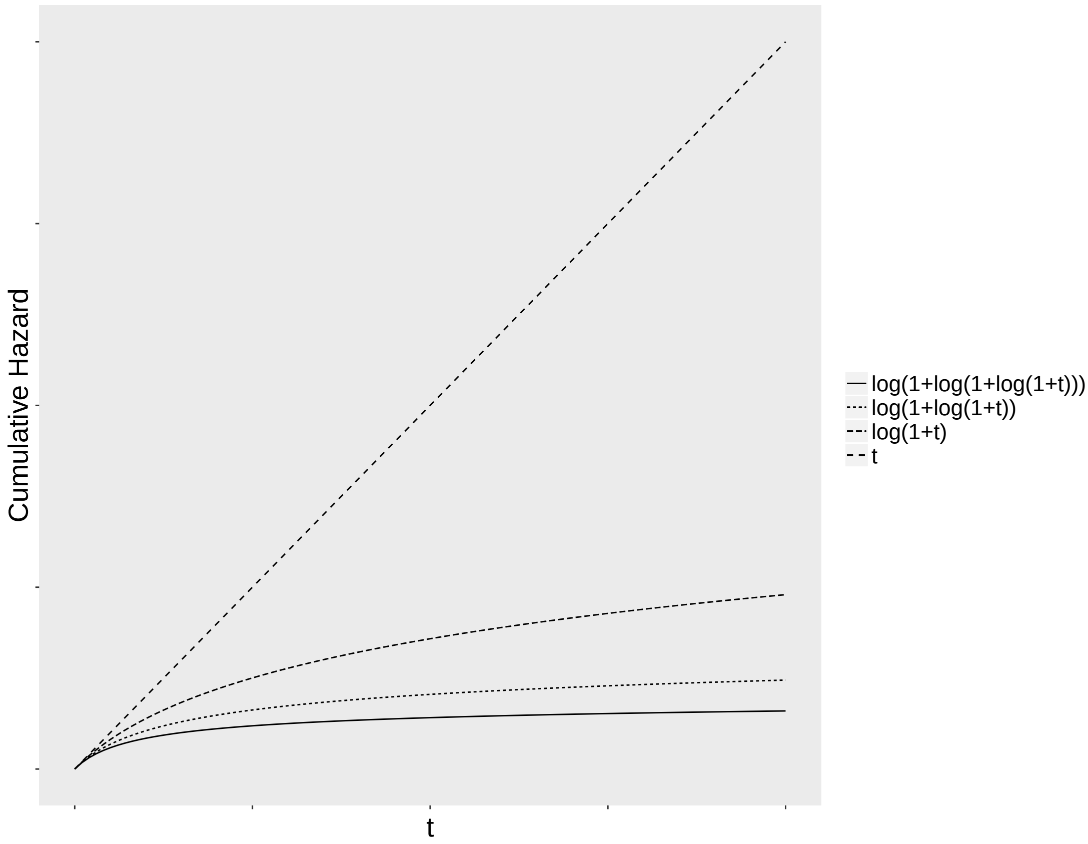

Figure 2.4: The cumulative hazard $H_T(t) = \log(1+t)$ composed iteratively with itself. The more compositions, the function grows more slowly.

Example 2.19. To illustrate how a function that crosses the identity behaves under iterative compositions, see Figure 2.5 where the function t^3 is composed with itself multiple times. We see that with each composition, the function gets further from the identity apart from the point at which it crosses the line $H_T(t) = t$, i.e., a fixed point of $H_T(t)$. We see that before the fixed point, the function is decreased more with every composition, then after the fixed point is increased more.

2.3.2 Inverse of cumulative hazards

The inverse of a CH has been used by Rinne [52], where it is called the hazard quantile. Here we use such inverses in a much more general sense. We refer to H^{-1} as the usual inverse function of H.

Proposition 2.20. *If $H(t)$ is a cumulative hazard function, then so is its inverse $H^{-1}(t)$.*

Proof. If $H : [0,\infty] \to [0,\infty]$ then we must have $H^{-1} : [0,\infty] \to [0,\infty]$.

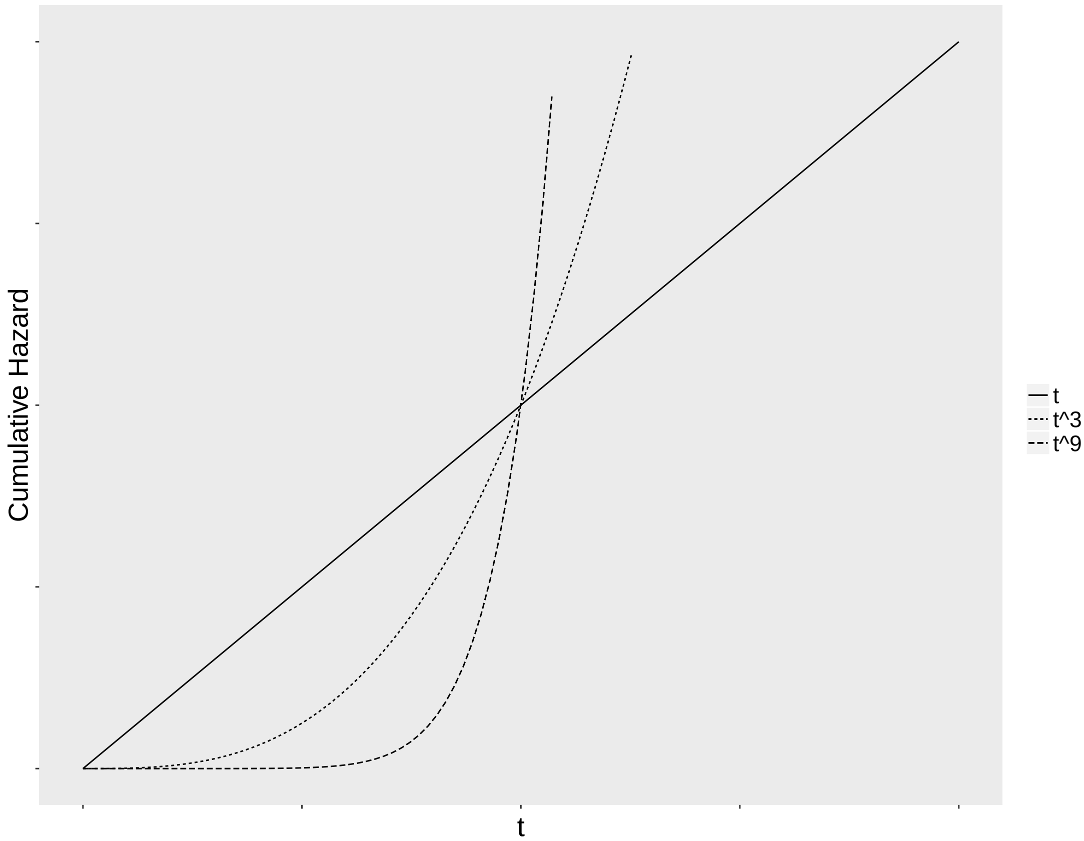

Figure 2.5: The cumulative hazard $H_T(t) = t^3$ composed iteratively with itself.

Since $H(0) = 0$ and $H(\infty) = \infty$, we have

$$H^{-1}(0) = H^{-1}(H(0)) = 0,$$
$$H^{-1}(\infty) = H^{-1}(H(\infty)) = \infty.$$

Finally if $H(t)$ is a cumulative hazard then it must be strictly increasing, thus

$$t < s \iff H(t) < H(s).$$

This implies

$$\begin{aligned} H^{-1}(t) < H^{-1}(s) &\iff H^{-1}(H(t)) < H^{-1}(H(s)) \\ &\iff t < s. \end{aligned}$$

So H^{-1} is increasing too.

By the Inverse function theorem,

$$\frac{d}{dt}H^{-1}(t) = \frac{1}{H'(H^{-1}(t))}.$$

Since H is continuous and bijective, then H^{-1} must be continuous. Hence, the derivative of the inverse is the composition of continuous functions and so is continuously differentiable. Thus H^{-1} satisfies all the conditions stated and hence is a cumulative hazard. □

The next few examples demonstrate that the inverse of a CH function is also a CH function.

Example 2.21. Consider a standard Gompertz cumulative hazard $H_G(t) = e^t - 1$, then $H_G^{-1}(t) = \log(1+t) = H_{ll}(t)$, the standard log-logistic cumulative hazard.

Example 2.22. Consider the inverse of a Weibull CH $H_W^{\alpha}(t) = t^{\alpha}$. Then $\left(H_W^{\alpha}\right)^{-1}(t) = t^{1/\alpha} = H_W^{1/\alpha}$, yet another Weibull.

Example 2.23. Consider the exponential CH $H_E^{\theta}(t) = \theta t$. The inverse is $H_E^{1/\theta}(t) = \frac{1}{\theta}t$. We thus see that the exponential family is also closed under inverses and that $\left(H_E^{\theta}\right)^{-1} = H_E^{1/\theta}$.

Example 2.24. The inverse of the rational CH function $H_r^{\theta}(t) = \frac{t^{\theta}}{(1+t)^{\theta}-t^{\theta}}$ is $H_r^{1/\theta}(t) = \frac{t^{1/\theta}}{(1+t)^{1/\theta}-t^{1/\theta}}$. Hence, this family is also closed under inverses and $\left(H_r^{\theta}\right)^{-1} = H_r^{1/\theta}$.

It is useful to visually compare a CH function with its inverse. In figure 2.6, we see the inverse is the reflection of H about the identity line. We see, for example, that if H is convex, then H^{-1} is concave. This idea of convexity will be revisited later in this chapter, in section 2.3.8. We see that if a CH function accelerates/decelerates in some interval, then the inverse will decelerate/accelerate respectively.

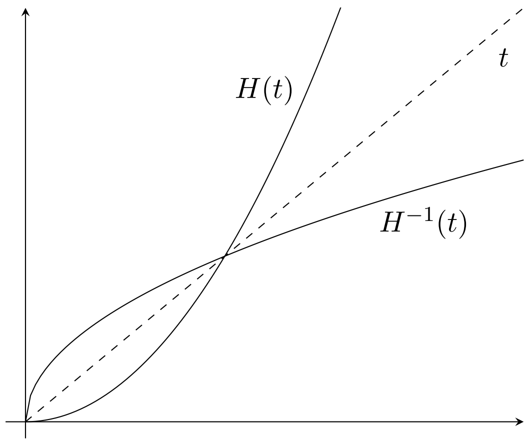

Figure 2.6: Comparison of a CH function H with its inverse H^{-1} to see they are reflections about the identity line (dashed).

We will now see a few more examples of inverses of the functions in our key set of CH functions.

Example 2.25. Consider the Log-Normal CH function, the inverse of this function is

$$H_{LN}^{-1}(t) = \exp\left[\Phi^{-1}\left(1 - e^{-t}\right)\right].$$

Example 2.26. Recall that the CH function of the Gamma distribution is

$$H_\Gamma^\kappa(t) = \log\left(1 + \frac{\Gamma(t,\kappa)}{\Gamma(\kappa) - \Gamma(t,\kappa)}\right).$$

After some rearranging to try to find the inverse, we find we need to solve

$$\frac{\Gamma(t,\kappa)}{\Gamma(\kappa)} = 1 - e^{-u}$$

for t when u is fixed. Since the incomplete gamma function is defined as an integral, we need to invert it numerically in order to find the inverse of H_Γ^κ. Since $1 - e^{-u} \in (0,1)$, finding the solution is equivalent to finding a quantile.

Note here that H^{-1} always exists, but an analytical expression may not be possible to calculate. Thus we may have to calculate the inverse numerically. In general, the computation of H^{-1} can be seen as a computation of a quantile. For details on how to do this see appendix B.

The next section gives an important result which will link the time transformation with both the composition and the inverse.

2.3.3 Linking compositions and inverses with time transformations

We have seen that the set of CH functions is closed under compositions and taking inverses. Given two CH functions in the set, it would be useful to know if and how to get from one to the other using other CH functions in the set. The next Theorem details how to do this.

Theorem 2.27. ***(Closure under composition)*** *Given cumulative hazards H_0 and H_1 with $T_0 \sim H_0$ and $T_1 \sim H_1$. There exists*

1. *a unique cumulative hazard $\overrightarrow{H}$ such that $H_1(t) = \overrightarrow{H} \circ H_0(t)$ which is given by*

$$\overrightarrow{H} = H_1 \circ H_0^{-1},$$

2. *a unique (in distribution) time transformation $\overrightarrow{\tau}$ such that $T_1 \overset{d}{=} \overrightarrow{\tau}(T_0)$ is given by*

$$\overrightarrow{\tau} = H_1^{-1} \circ H_0.$$

Proof. First we will prove 1. Suppose $\overrightarrow{H} = H_1 \circ H_0^{-1}$, then

$$\overrightarrow{H}(H_0(t)) = H_1(H_0^{-1}(H_0(t))) = H_1(t).$$

Now to prove uniqueness. Suppose there exist some other functional transformation,

$\overrightarrow{L}$, such that $H_1 = \overrightarrow{L}(H_0)$. Then,

$$\overrightarrow{L} \circ H_0 = \overrightarrow{H} \circ H_0 \Rightarrow \overrightarrow{L} \circ H_0 \circ H_0^{-1} = \overrightarrow{H} \circ H_0 \circ H_0^{-1} \Rightarrow \overrightarrow{L} \equiv \overrightarrow{H}.$$

Thus $\overrightarrow{H}$ is unique.

Now we aim to prove 2. Let $Y = \overrightarrow{\tau}(T_0) = H_1^{-1}(H_0(T_0))$

$$\begin{aligned} e^{-H_Y(t)} &= P(Y \geq y) = P(H_1^{-1}(H_0(T_0)) \geq y) = P(T_0 \geq H_0^{-1}(H_1(y))) \\ &= e^{-H_0(H_0^{-1}(H_1(y)))} = e^{-H_1(y)}. \end{aligned}$$

Thus $H_Y(y) = H_1(y)$ for all $y > 0$. Hence $T_1 \overset{d}{=} \overrightarrow{\tau}(T_0) = H_1^{-1}(H_0(T_0))$. Let $\overrightarrow{D}$ be another time transformation such that $T_1 \overset{d}{=} \overrightarrow{D}(T_0)$. Then

$$e^{-H_1(t)} = P(T_1 > t) = P(\overrightarrow{D}(T_0) > t)) = e^{-H_0(\overrightarrow{D}^{-1}(t))}.$$

Therefore $H_1 = H_0 \circ \overrightarrow{D}^{-1}$ and thus $\overrightarrow{D} = H_1^{-1} \circ H_0 = \overrightarrow{H}$. So uniqueness follows. □

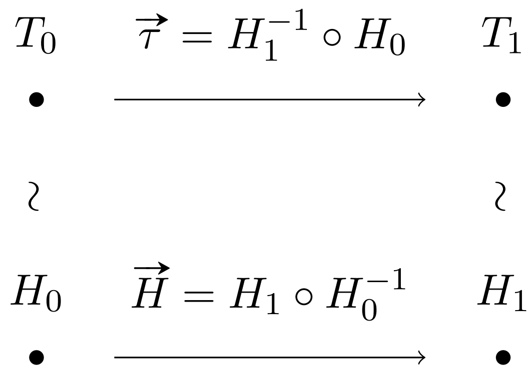

Figure 2.7: Diagram to represent the transformation from baseline time and the corresponding transformation for cumulative hazards.

Theorem 2.27 allows us to complete the diagram in figure 2.3 as seen in figure 2.7. It is important to note that both transformation $\overrightarrow{\tau}$ (time transformation) and $\overrightarrow{H}$ (CH functional transformation) are cumulative hazards themselves since they are compositions of CH functions.

According to Theorem 2.27 we can interpret CH functions as functional transformations of other CHs and also as time transformations. In the latter case, the local shape of the CH determines the acceleration or deceleration of the baseline time.

Theorem 2.27 shows that the set of CH functions is closed under composition. It also shows how to find the unique transformation between any pair of CHs.

Example 2.28. Following Theorem 2.27, let $T_1 \sim H_G$ and $T_0 \sim H_W^{\alpha}$ then

$$\overrightarrow{\tau}(T_0) = H_1^{-1} \circ H_0(T_0) = H_{ll} \circ H_W^{\alpha}(T_0) = \log(1 + T_0^{\alpha}).$$

From this example we will see that if we time transform a Weibull with a log-logistic CH, the resulting time follows a standard Gompertz.

The next result gives one interpretation of the composition of CHs.

Corollary 2.29. ***(Composition as a time transformation)*** *Let H_0 be a CH and $H_1 = H_0 \circ H$ for some CH function H. Then the corresponding time transformation is given by $T_1 \stackrel{d}{=} H^{-1}(T_0)$.*

Proof. From Theorem 2.27 we have

$$\vec{\tau} = H_1^{-1} \circ H_0 = H^{-1} \circ H_0^{-1} \circ H_0 = H^{-1}.$$

□

Example 2.30. If we let H in corollary 2.29 be H_0^{-1}, then

$$H_1 = H_0 \circ H_0^{-1} = H_E^1$$

and thus $T_1 \sim H_E^1$, the same result as proposition 2.9.

We know that the set of CH functions is closed under composition, but is this action commutative? We see that it is not, as demonstrated by the following argument.

Example 2.31. If we let $T_0 \sim H_0$ and let $T_1 \stackrel{d}{=} \left(H_0^{-1} \circ H^{-1} \circ H_0\right)(T_0)$, for some CH function H. Then using Theorem 2.27, we have that $T_1 \sim H \circ H_0$. Thus if we recall the transformation in corollary 2.29, for $T_1 \sim H_0 \circ H$ we need $T_1 \stackrel{d}{=} H^{-1}(T_0)$. Note that different transformations are needed for $T_1 \sim H \circ H_0$ and $T_1 \sim H_0 \circ H$, and so they do not have the same distributions, i.e. $H_0 \circ H \neq H \circ H_0$. Thus we see that the operation of composition is not commutative.

The next corollary details which time transformation gives $T_0 \sim H$ and $T_1 \sim H^{-1}$.

Corollary 2.32. ***(Inverse as a time transformation)*** *Let H_0 be a CH and let $H_1 = H_0^{-1}$. Then the corresponding time transformation is given by $T_1 \stackrel{d}{=} H_0^{(2)}(T_0)$.*

Proof. From Theorem 2.27 we have $\vec{\tau} = H_1^{-1} \circ H_0 = H_0 \circ H_0 = H_0^{(2)}$. □

This corollary says that if the cumulative hazard of T_1 is the inverse of that of T_0, then T_1 has the same distribution as $H^{(2)}(T_0)$. According to the examples after proposition 2.17, the CH of T_1 will be a local, or perhaps global, deceleration or acceleration of T_0.

Example 2.33. Suppose that $T_0 \sim H_{ll}$ and $T_1 \stackrel{d}{=} H_{ll}^{(2)}(T_0) = \log(1 + \log(1 + T_0))$. Then by corollary 2.32, we have that $T_1 \sim H_{ll}^{-1} = H_G$.

The next example uses a time transformation to define a parametric family of distributions known in the literature.

Example 2.34. The Birnbaum-Saunders distribution is commonly used in reliability theory to model failure times [7] and is usually defined as a time transformation. Suppose T_0 is a standard log-normal. Then if $T_1 \overset{d}{=} \vec{\tau}(T_0)$, where

$$\vec{\tau}(T_0) = \beta \left(\log(T_0^{\alpha/2}) + \sqrt{1 + \log(T_0^{\alpha/2})^2} \right)^2 \tag{2.14}$$

then T_1 follows a Birnbaum-Saunders distribution with shape α and scale β

The corresponding CH of T_1 is given by $H_1(t) = -\log\left(1 - \Phi\left(\frac{1}{\alpha}\left[\frac{\sqrt{t}}{\beta} - \frac{\beta}{\sqrt{t}}\right]\right)\right)$. According to Theorem 2.27, equation (2.14) defines a cumulative hazard given by $\vec{\tau}(t)$.

The next example refers to the most popular model in survival analysis, the Proportional Hazards model, and how it can be constructed using inverses and composition of CH functions. The detail of the model is not needed for this example, but is given in section 4.2. This example simply demonstrates the time transformation used in this model.

Example 2.35. (Proportional Hazards model) Consider a population with baseline time $T_0 \sim H_0$ and a particular individual time $T_1 \sim H_1$. The Proportional Hazards model [18] is such that

$$H_1(t) = \theta H_0(t),$$

i.e. the CH function of the individual is proportional to the baseline CH. Hence we have $H_1 = H_E^{\theta} \circ H_0$. So by Theorem 2.27 we obtain

$$T_1 \overset{d}{=} \vec{\tau}(T_0) = H_0^{-1} \circ H_E^{1/\theta} \circ H_0(T_0).$$

We can view the proportional hazards model as a linear transformation model as in chapter 4, then

$$\begin{aligned} \log H_0(T_1) &\overset{d}{=} -\log(\theta) + \log H_0(T_0) \\ &= -\log(\theta) + \log E \end{aligned}$$

where $E \sim H_E^1$ a standard exponential distribution.

Another important model is the Proportional Odds model [5]. The time transformation used for this model is explained in the next example.

Example 2.36. (Proportional Odds model) Another useful model in the literature is the Proportional Odds model, which will be discussed in more detail in chapter 4. This model is defined by the transformation

$$H_1(t) = \vec{\tau}(H_0(t)) = H_{ll}\left(\theta^{-1} H_{ll}^{-1}\left(H_0(t)\right)\right) = H_{ll} \circ H_E^{1/\theta} \circ H_G \circ H_0(t).$$

This model can be written as $H_G(H_1(t)) = \theta H_G(H_0(t))$ which, according to example 2.13 means that the odds of the event occurring before t are proportional with a factor of θ. If we have that $T_1 \stackrel{d}{=} \vec{\tau}(T_0)$, then we find that

$$\vec{\tau}(T_0) \stackrel{d}{=} H_0^{-1}\left(H_{ll}\left(\theta H_{ll}^{-1}\left(H_0(T_0)\right)\right)\right).$$

The model can be written as a linear model as follows:

$$\begin{aligned}\log\left(H_G\left(H_0(T_1)\right)\right) &\stackrel{d}{=} -\log(\theta) + \log\left(H_G\left(H_0(T_0)\right)\right)\\ &= -\log(\theta) + \log\left(H_G(E)\right),\end{aligned}$$

where $E \sim H_E^1$ the standard exponential.

Theorem 2.27 is related to the concept of the G-Hazard potential of Singpurwalla [64]. From Theorem 2.27 we find that if $Y := H_Y^{-1}(H_T(T))$, then Y has CH function H_Y irrespective of T. In the context of the G-Hazard potential, this means failure occurs when $H_Y^{-1}(H_T(t))$ exceeds a random threshold Y, where Y has CH function H_Y.

2.3.4 Product of cumulative hazards

Before the next few results we need some notation to distinguish between repeated compositions and powers of CH functions. Let H be a CH function and define

$$H(t)^2 := H(t)H(t), \tag{2.15}$$

to be the product of H with itself. We would have that the product of H with itself $n-1$ times is denoted $H(t)^n$.

Proposition 2.37. *Given two cumulative hazard functions $H_1(t)$ and $H_2(t)$, their product $H_1(t)H_2(t)$ is also a cumulative hazard.*

Proof. It is clear that if H_1 and H_2 are CH functions, then $H_1(0)H_2(0) = 0$, $H_1(\infty)H_2(\infty) = \infty$ and $H_1(t)H_2(t) > 0$ for $t > 0$.

Note that

$$\frac{d}{dt}\Big(H_1(t)H_2(t)\Big) = h_1(t)H_2(t) + H_1(t)h_2(t),$$

which is the sum of continuous functions and is therefore continuous itself. The product of CH functions is thus continuously differentiable.

As H_1 and H_2 are strictly increasing, then for

$$s < t, \quad H_1(s) < H_1(t) \text{ and } H_2(s) < H_2(t).$$

Thus

$$H_1(s)H_2(s) < H_1(t)H_2(s) < H_1(t)H_2(t),$$

so $H_1(t)H_2(t)$ is also strictly increasing and thus a CH function. $\square$

This proposition might give a means to producing some new or interesting CH functions. Some possible examples are now explored.

Example 2.38. Consider two Weibull CH functions, $H_W^{\alpha}(t) = t^{\alpha}$ and $H_W^{\beta}(t) = t^{\beta}$. Their product is $H_W^{\alpha+\beta}(t) = t^{\alpha+\beta}$, another Weibull CH. Hence, this family is closed under the product.

So far we have seen that the Weibull family is closed under composition, inverse and product of CH functions.

The next example demonstrates how the product of CH functions can lead to the generation of new distributions.

Example 2.39. Now consider $H_E^1(t) = t$, a standard exponential, and $H_{ll}(t) = \log(1+t)$, a standard log-logistic, then their product $t\log(1+t)$ is also a CH function. This CH function is one that we have not seen before and generates a new distribution.

Time Transformation *Suppose that $T_0 \sim H_0$ and $T_1 \sim H\,H_0$. Theorem 2.27 says that there is some time transformation linking the product with the baseline CH function. If we have that $T_1 \overset{d}{=} \vec{\tau}(T_0)$, by Theorem 2.27 $\vec{\tau}(t) \overset{d}{=} (H\,H_0)^{-1} \circ H_0(t)$. Hence we see that finding the corresponding time transformation may not be analytically tractable.*

Exploring the effect on the survival functions will give a different interpretation.

Proposition 2.40. *Let $S_1(t)$ be the survival function corresponding to the CH function $H_1(t)$. The survival function corresponding to the product of two CH function, $H_1(t)H_2(t)$ is $S(t) = S_1(t)^{H_2(t)}$.*

Proof.

$$S(t) = e^{-H_1(t)H_2(t)} = \left(e^{-H_1(t)}\right)^{H_2(t)} = S_1(t)^{H_2(t)}$$

$\square$

Now we consider multiplying a CH function by some function other than another CH function.

Corollary 2.41. *Multiplying a CH function by a positive, continuously differentiable and non-decreasing function, results in a CH function.*

Proof. Let H be a CH function, and let g be a positive non-decreasing function. It is clear that $H(0)g(0) = 0 \cdot g(0) = 0$ and $H(\infty)g(\infty) = \infty \cdot g(\infty) = \infty$. It is also clear that the product of H and g will be positive. Since g is non-decreasing, we have that the product $H(t)g(t)$ will be increasing as H is. The product of continuously differentiable functions is continuously differentiable. Thus the product satisfies all the conditions to be a CH function. $\square$

Example 2.42. For example, suppose $H(t)$ is a CH function, then $H(t)e^t$ will be a new CH function.

We now relax the non-decreasing condition in g in the previous corollary.

Proposition 2.43. *Let H be a CH function and let $g : (0, \infty) \to (0, \infty)$ be a continuously differentiable function such that*

$$\frac{d}{dt} \log g(t) > -\frac{d}{dt} \log H(t),$$

$g(0) < \infty$ *and* $\lim_{t\to\infty} g(t) \neq 0$*, then $H(t)g(t)$ is a CH function.*

Proof. As g is well defined at 0 and ∞ then $H(t)g(t)$ has the correct limits. The product of continuously differentiable functions is continuously differentiable. Now to prove that $H(t)g(t)$ is increasing. If $H(t)g(t)$ is increasing we require that

$$\frac{d}{dt}\Big(H(t)g(t)\Big) = h(t)g(t) + H(t)\frac{d}{dt}g(t) > 0.$$

This implies that

$$\frac{g'(t)}{g(t)} > -\frac{h(t)}{H(t)}$$

which is equivalent to

$$\frac{d}{dt} \log g(t) > -\frac{d}{dt} \log H(t)$$

which is true by assumption. Thus we have $H(t)g(t)$ is increasing. □

Note that we can find similar constraints for the function $\frac{H(t)}{g(t)}$, obviously requiring g is positive at zero and well defined at infinity.

Proposition 2.43 will be useful when trying to generalise the Proportional Hazards model to include time varying covariates. This proposition gives some conditions on the covariates so that this time varying model is valid. This idea will be explored later in section 2.3.10.

Corollary 2.44. *If $H(t)$ is a cumulative hazard and α is a positive constant, then $H(t)^\alpha$ is a cumulative hazard.*

Proof. This result follows directly from proposition 2.37 for integer valued α. For $\alpha > 0$:

$$H(0)^\alpha = 0^\alpha = 0,$$
$$H(\infty)^\alpha = \infty^\alpha = \infty.$$

Now to show $H(t)^\alpha$ is increasing,

$$\frac{d}{dt}\left(H(t)^\alpha\right) = \alpha h(t)H(t)^{\alpha-1} > 0, \forall\, \alpha > 0.$$

We have that $H(t)^\alpha = H_W^\alpha \circ H(t)$, and thus is the composition of continuously differentiable functions. It is therefore continuously differentiable itself. $\square$

Example 2.45. The most notable example of corollary 2.44 is the CH of the Weibull distribution. If we take $H_E^1(t) = t$, the identity function and CH function of the exponential with rate 1, then raising it to the power of α we get a Weibull CH, i.e. $H_W^\alpha(t) = (H_E^1(t))^\alpha$.

Corollary 2.44 gives rise to some interesting examples of how to manipulate CH functions. Raising a CH to some power will alter its properties as seen in the next example.

Example 2.46. Suppose $H(t) = (\log(1+t))^\alpha$. To see the effect of powering $H(t)$, see Figure 2.8. In this plot we see that if $\alpha > 1$ then as it increases, the function becomes more convex. However, if $\alpha < 1$ then as it decreases, the function becomes more concave.

2.3.5 Addition of cumulative hazards

This section will discuss the addition of cumulative hazards and the properties of these types of transformation.

Proposition 2.47. *The set $\mathcal{CH}$ is a convex cone, i.e., for $H_1, H_2 \in \mathcal{CH}$, $a, b > 0$, $aH_1 + bH_2 \in \mathcal{CH}$.*

Proof. We see that for $a, b > 0$, $aH_1(0) + bH_2(0) = 0$ since $H_i(0) = 0$ for $i = 1, 2$. As we also have that $\lim_{t\to\infty} H_i(t) = \infty$, $\lim_{t\to\infty} aH_1(t) + bH_2(t) = \infty$. Multiplying an increasing function by a positive constant, will result in an increasing function, thus $aH_1(t)$ and $bH_2(t)$ are both increasing. Since the sum of increasing functions will be increasing, we thus have $aH_1(t) + bH_2(t)$ is increasing. The sum of continuously differentiable functions is continuously differentiable. $\square$

Let's see some examples of this proposition.

Example 2.48. Given CH functions $H_1(t) = H_W^\alpha$ and $H_2(t) = H_W^\beta$, then by proposition 2.47 $H(t) = at^\alpha + bt^\beta$ is also a cumulative hazard. We see that the Weibull family is not closed under addition. This family is called the Poly-Weibull distribution as seen in the work of Berger and Sun [6].

Example 2.49. Let $H_1(t) = H_{ll}(t)$ and $H_2(t) = H_G(t)$. Then by proposition 2.47 $H(t) = a\log(1+t) + b\left(e^t - 1\right)$ is a CH function. We note that in contrast to the

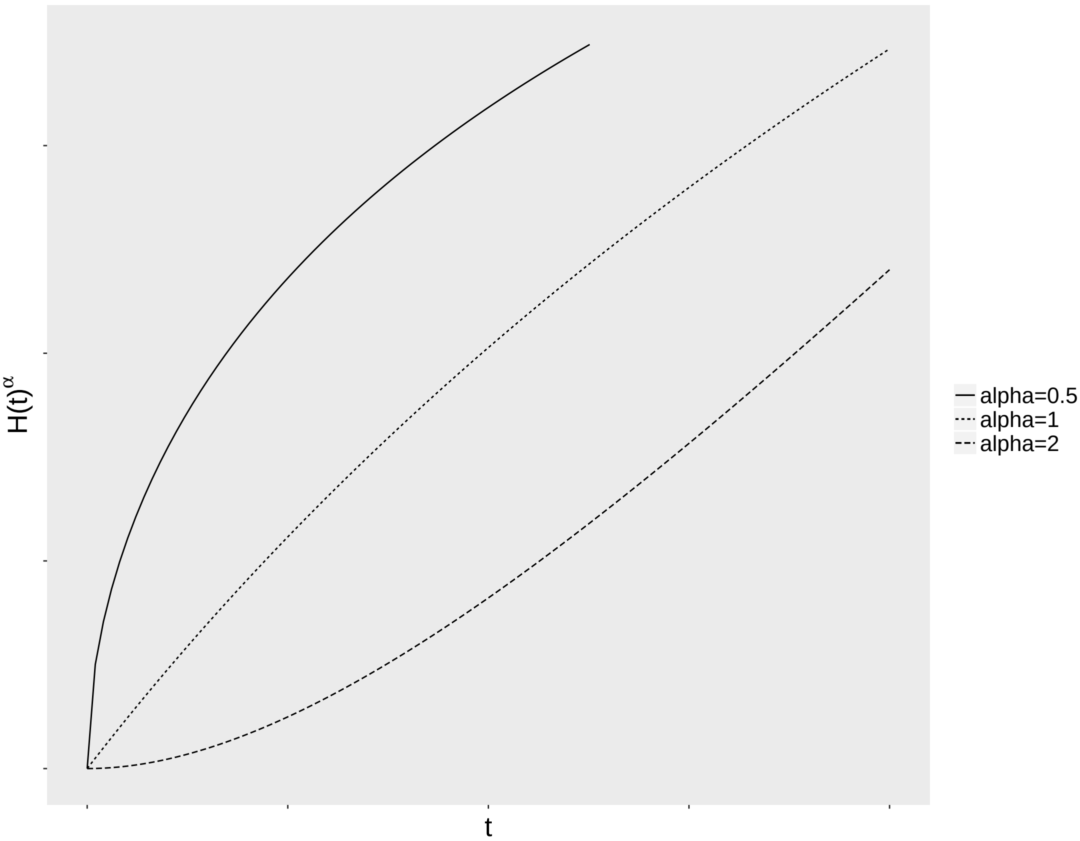

Figure 2.8: The cumulative hazard $H(t) = \log(1+t)$ raised to the powers $\alpha = (0.5, 1, 2)$ to demonstrate the effect of powering a CH function.

previous example, we have the addition of two drastically different, in fact inverse, CH functions and we still preserve the CH function properties.

The following corollary follows directly from Proposition 2.47.

Corollary 2.50. *The sum of two cumulative hazards is a cumulative hazard.*

Proof. To prove this corollary, let $a = b = 1$ in proposition 2.47. □

Example 2.51. (Additive hazards model) The additive hazards model as seen in the work of Breslow and Day [8] is characterised by

$$H_1(t) = H_0(t) + H_E^{\theta}(t)$$

so that the difference in hazards is constant or

$$\frac{H_1(t) - H_0(t)}{t} = \theta.$$

This model is sometimes called the proportional excess hazard model, as seen in the work of Sasieni [59].

The next example explores the form of the time transformation for the sum of CH functions.

Time Transformation *If* $T_0 \sim H_0$ *and* $T_1 \sim H + H_0$*, then if* $T_1 \stackrel{d}{=} \vec{\tau}(T_0)$*, then by Theorem 2.27,* $\vec{\tau}(T) \stackrel{d}{=} (H + H_0)^{-1} \circ H_0(T)$*, which would have to computed numerically in general.*

The following proposition gives a much simpler probabilistic interpretation of the sum of cumulative hazards.

Proposition 2.52. *Given two independent positive random variables* T_0 *and* T_1 *with corresponding CH functions* H_0 *and* H_1*, the survival function of their minimum* $Y = \min(T_0, T_1)$ *is*

$$S_Y(y) = e^{-(H_0(y)+H_1(y))}$$

and we have that $Y \sim H_0 + H_1$*.*

Proof.

$$\begin{aligned} S_Y(y) &= P(Y \geq y) = P(\min(T_0, T_1) \geq y) \\ &= P(T_0 \geq y, T_1 \geq y) = P(T_0 \geq y)P(T_1 \geq y) \\ &= e^{-H_1(y)}e^{-H_2(y)} = e^{-(H_1(y)+H_2(y))}. \end{aligned}$$

Since the CH function of Y is $H_1(y) + H_2(y)$, then $Y \sim H_0 + H_1$. □

This proposition allows for simulating from the distribution whose CH is the sum of two CHs.

As mentioned earlier, the standard model for right censoring is the so called non-informative censoring where T is a potentially observable time to event, and C is an independent unobservable censoring time, so that the actual observed time is $Y = \min(T, C)$. We will see this model again in chapter 6 viewed from a competing risks context.

Corollary 2.53. *Let* H *be a CH function. Let* $g : (0, \infty) \to (0, \infty)$ *be a continuously differentiable function such that* $g(0) = 0$ *and* $\frac{d}{dt}g(t) > -\frac{d}{dt}H(t)$*. Then* $H(t) + g(t)$ *is also a CH function.*

Proof. Since $H(t), g(t) > 0$ for $t > 0$ their sum will be too. And since $g(0) = 0$, $H(0) + g(0) = 0$. Since g is a positive function, then $\lim_{t\to\infty} H(t) + g(t) = \infty + \lim_{t\to\infty} g(t) = \infty$. For $H(t) + g(t)$ to be strictly increasing, we require $\frac{d}{dt}(H(t) + g(t)) > 0$. Since $\frac{d}{dt}g(t) > -\frac{d}{dt}H(t)$, this condition is satisfied. Hence $H + g$ satisfies all the condition to be a CH function. □

Corollary 2.53 gives us a way to expand the set of CH functions easily.

Example 2.54. For example, consider $H(t) = t$ and $g(t) = 1 - e^{-t}$. We can see that $1 + t - e^{-t}$ is a CH function. Note in this example, $g(t)$ is actually the CDF of the exponential distribution with rate parameter 1.

Example 2.55. The Makeham [43] generalisation of the Gompertz distribution adds a constant to the Gompertz hazard function. Thus the corresponding CH is

$$H(t) = \theta t + e^t - 1 = H_E^\theta(t) + H_G(t).$$

Example 2.56. In Marshall and Olkin [44] a parallel system of exponentially distributed random variables have CH function

$$\begin{aligned} H(t) &= -\log\left(e^{-\lambda_1 t} + e^{-\lambda_2 t} + e^{-(\lambda_1+\lambda_2)t}\right) \\ &= \lambda_1 t + \log\left(\frac{e^{\lambda_2 t}}{e^{\lambda_2 t} + e^{\lambda_1 t} + 1}\right), \end{aligned}$$

which is of the form $H(t) + g(t)$.

As well as expanding the set of CH functions, corollary 2.53 gives a method to incorporate time varying covariates into a model through the function $g(t)$. We will explore this in section 2.3.10.

We will now conclude this section by discussing the differences of cumulative hazards. It is clear that if one cumulative hazard is always strictly larger than another, then if their difference is increasing, we have that their difference is a CH function.

Example 2.57. It is well known that $H_{ll}(t) = \log(1+t) < H_E^1(t) = t < H_G(t) = e^t - 1$. Thus, after investigation of the derivatives of the following, we see that

$$\begin{aligned} H_1(t) &= H_G(t) - H_E^1(t), \\ H_2(t) &= H_G(t) - H_{ll}(t), \\ H_3(t) &= H_E^1(t) - H_{ll}(t) \end{aligned}$$

are all CH functions.

2.3.6 Other transformations

Proposition 2.58. *If $H_1(t)$ is a CH function, then so is $H(t) = \frac{1}{H_1(\frac{1}{t})}$.*

Proof. If $t = 0$ then $H_1(\frac{1}{t}) = H_1(\infty) = \infty$, so $H(0) = \frac{1}{\infty} = 0$. If $t = \infty$, $H_1(\frac{1}{t}) = H_1(0) = 0$, so $H(\infty) = \frac{1}{0} = \infty$. We will clearly have that $H(t)$ is a positive function since $H_1(t)$ is. All that remains is to show $H(t)$ is strictly increasing. As $H_1(t)$ is strictly increasing, then $s < t \iff H_1(s) < H_1(t)$. Thus

$$s < t \iff \frac{1}{t} < \frac{1}{s} \iff H_1\left(\frac{1}{t}\right) < H_1\left(\frac{1}{s}\right) \iff \frac{1}{H_1(\frac{1}{s})} < \frac{1}{H_1(\frac{1}{t})}$$

So $H(t)$ is strictly increasing and is therefore a CH function. □

We will now see a couple of examples of CH functions generated using the transformation seen in Proposition 2.58.

Example 2.59. For our first example, consider the Weibull CH $H_1(t) = H_W^{\alpha}(t) = t^{\alpha}$. Then we see that $H(t) = \frac{1}{(1/t)^{\alpha}} = t^{\alpha} = H_W^{\alpha}$.

Example 2.60. For our second example, consider $H_1(t) = H_{ll}(t) = \log(1+t)$. Then $H(t) = \frac{1}{\log(1+\frac{1}{t})}$. In order to see the effect of this transformation, see Figure 2.9.

Example 2.61. Now let $H_1(t) = H_G(t)$, then $H(t) = 1/(e^{1/t} - 1)$. We can see the effect of this transformation in figure 2.10.

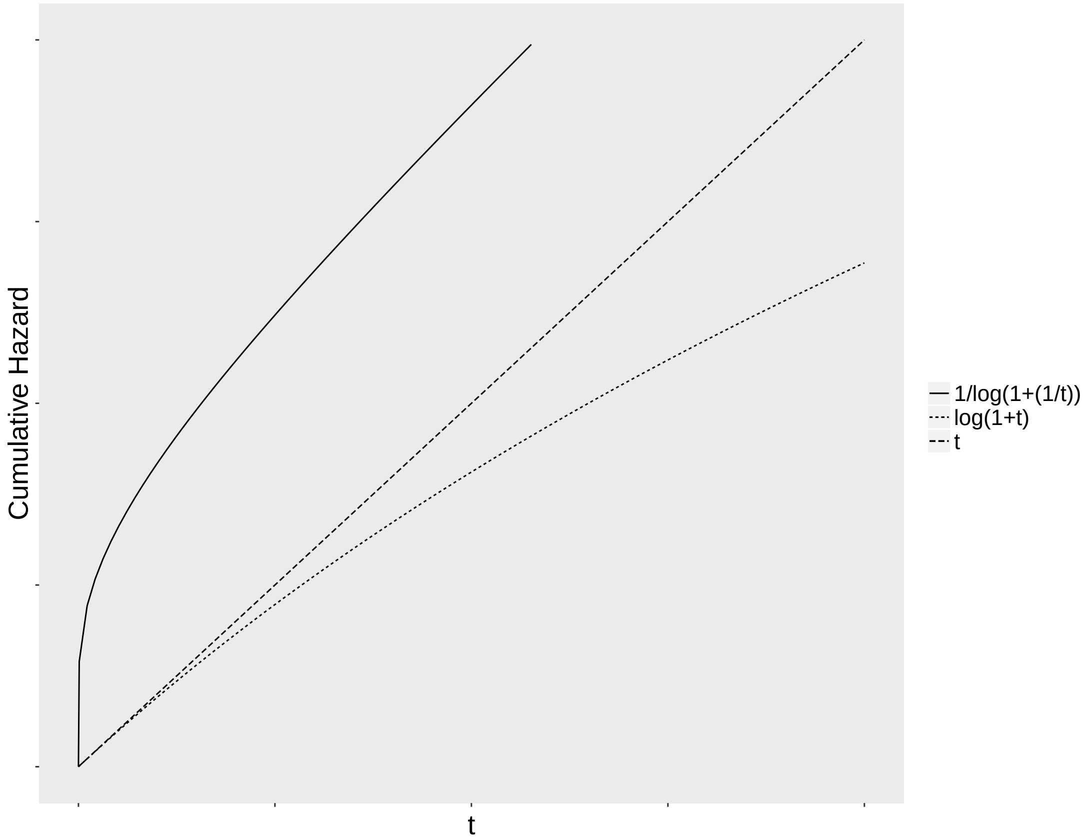

Figure 2.9: The cumulative hazard $H_1(t) = \log(1+t)$ and the cumulative hazard $H(t) = \frac{1}{\log(1+\frac{1}{t})}$.

Example 2.62. Consider $H_1(t) = H_r^2(t) = \frac{t^2}{(1+t)^2 - t^2}$ then $H(t) = (1+t)^2 - 1$.

2.3.7 Maximum of cumulative hazards

In this section we will see an example of some transformation of CH functions that takes us out of our set $\mathcal{CH}$.

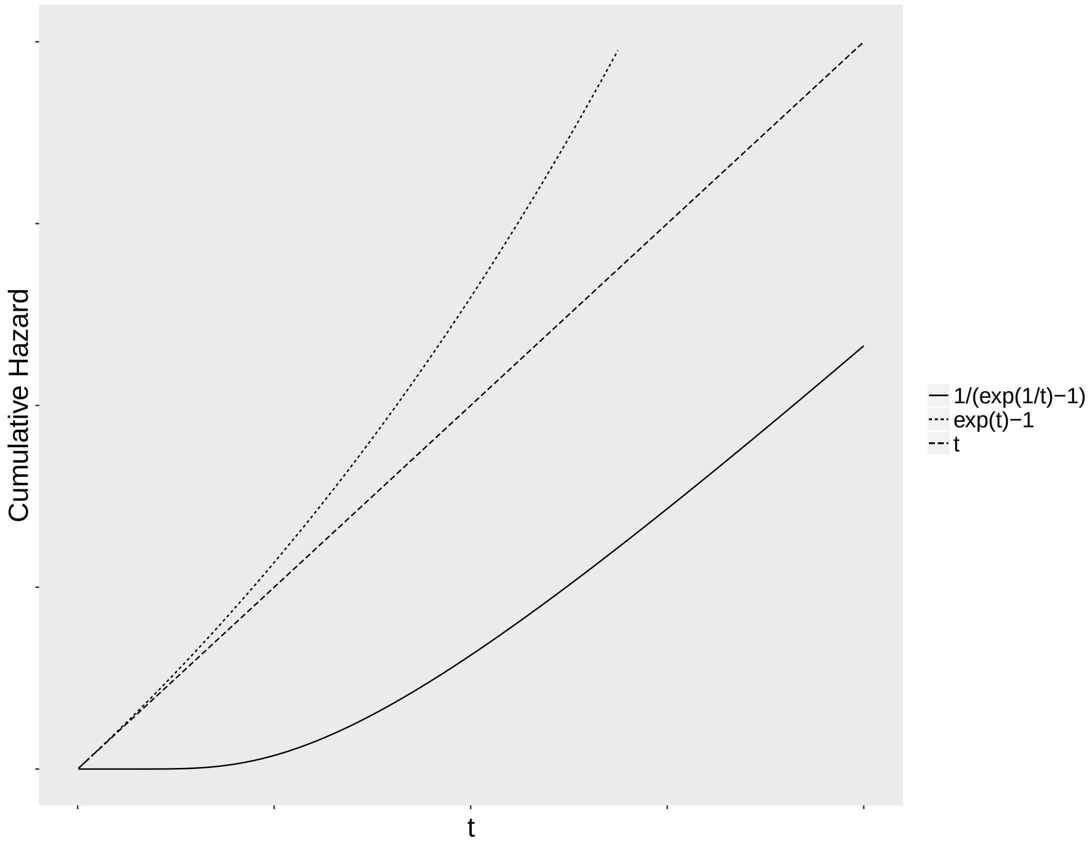

Figure 2.10: The cumulative hazard $H_1(t) = H_G(t)$ and the cumulative hazard $H(t) = \frac{1}{e^{1/t}-1}$.

Proposition 2.63. *Given cumulative hazards $H_1(t)$ and $H_2(t)$, then*

$$H(t) = \max\{H_1(t), H_2(t)\}$$

is a function which satisfies properties 2.1, 2.2 and 2.4.

Proof. First of all

$$\begin{aligned} H(0) &= \max\{H_1(0), H_2(0)\} \\ &= \max\{0, 0\} = 0 \end{aligned}$$

and

$$\begin{aligned} \lim_{t\to\infty} H(t) &= \lim_{t\to\infty} \max\{H_1(t), H_2(t)\} \\ &= \infty \end{aligned}$$

since $\lim_{t\to\infty} H_i(t) = \infty$ for $i = 1, 2$. And finally, we are required to prove $H(t)$ is increasing. Suppose $t > s$,

$$\begin{aligned} H(t) &= \max\{H_1(t), H_2(t)\} \\ &> \max\{H_1(s), H_2(s)\} = H(s) \end{aligned}$$

since H_1 and H_2 are themselves increasing. □

Note here that the cumulative hazard generated by the above proposition will not necessarily be continuously differentiable, i.e. not in $\mathcal{C}^\infty$. To demonstrate this, see figure 2.11 where the cumulative hazard functions t^2 and $\log(1+t)$ are plotted in grey. The maximum of these functions is plotted in black, we see that this curve has a non-differentiable point. This CH function thus violates assumption 2.3 and would yield a density function similar to that seen in figure 2.2. We will not consider these types of CHs further in this thesis.

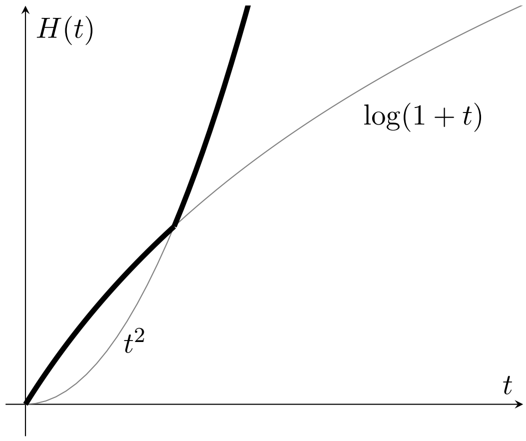

Figure 2.11: Plot of $H(t) = t^2$ and $H(t) = \log(1+t)$ with $\max\{t^2, \log(1+t)\}$ shown in black.

2.3.8 Integration of cumulative hazards

This section gives details on the properties of integrated CH functions. We will see that $\mathcal{CH}$ is closed under integration. The concept of N-functions as seen in Krasnoselaskii and Rutickii [58] will be relevant to this section. The conditions these N-functions must satisfy, continuity, takes the value of zero at zero, is increasing and is even, include those conditions for a function to be a CH function. Thus we define a slightly smaller class of functions, which we call *N-CH functions*.

Definition 2.64. Suppose $H(t)$ is a CH function, then an *N-CH* function is defined as

$$H^{[1]}(t) := \int_0^t H(s)ds.$$

From definition 2.64 we will show that N-CH functions, $H^{[1]}(t)$, are themselves cumulative hazards.

Corollary 2.65. *An N-CH function is a CH function.*

Proof. Let H be a CH function and let $H^{[1]}$ be its corresponding N-CH function. Since $H^{[1]}$ is the integral of a positive increasing function, it is itself increasing and positive. Now we see what $H^{[1]}(0)$ is,

$$H^{[1]}(0) = \int_0^0 H(u)du = 0.$$

Since H is increasing and always positive, its integral over $[0, \infty]$ will be ∞. $H^{[1]}(t)$ is clearly differentiable and since H is a CH, the derivative is continuous. Thus an N-CH function is continuously differentiable. □

To the best of our knowledge, integrated cumulative hazards have not been used in statistical modelling and so may be unfamiliar. We will first demonstrate what an N-CH function is with a few examples and then follow with the properties of these types of functions.

Example 2.66. Let $H(t) = t^\alpha$ be a Weibull CH function, we will see what its corresponding N-CH function is.

$$\left[H_W^\alpha\right]^{[1]}(t) = \int_0^t H(s)ds = \int_0^t s^\alpha ds = \frac{t^{\alpha+1}}{\alpha+1} = \frac{H_W^{\alpha+1}(t)}{\alpha+1} = \frac{tH_W^\alpha(t)}{\alpha+1}.$$

This can be seen either as a scaled Weibull or as a product of a Weibull and an exponential.

From this example we can see that the Weibull family with an added scale parameter is closed under this operation.

Example 2.67. Let H_G be a standard Gompertz CH function, then the corresponding N-CH function is

$$H_G^{[1]}(t) = \int_0^t \left(e^s - 1\right)ds = e^t - t - 1 = H_G(t) - t.$$

Example 2.68. The corresponding N-CH function for the standard log-logistic is

$$H_{ll}^{[1]}(t) = \int_0^t \log(1+s)ds = (1+t)\log(1+t) - t$$

Time Transformation *Let $T_0 \sim H_0$ and $T_1 \sim H_1 = H^{[1]}$, for some H, then the time transformation given by Theorem 2.27 is $\vec{\tau}(T_0) \stackrel{d}{=} \left(H^{-[1]} \circ H_0\right)(T_0)$, where $H^{-[1]}(t) := \left(H^{[1]}\right)^{-1}(t)$, the inverse of the N-CH function $H^{[1]}$. This time transformation will need to be computed numerically in general.*

For the next example of an N-CH function, we need some notation for iteratively integrating a CH function. We can define, for $n \geq 1$

$$H^{[n]}(t) := \int_0^t H^{[n-1]}(s)\, ds,$$

where $H^{[0]}(t) = H(t)$.

Example 2.69. Now let's consider what the iterated N-CH function would be for $n \geq 2$ for the N-CH function in example 2.67. We would see that

$$H_G^{[n]} = e^t - 1 - \sum_{i=1}^{n} \frac{t^i}{i!} = H_G(t) - \sum_{i=1}^{n} \frac{t^i}{i!} \to 0 \quad as \quad n \to \infty.$$

We can see the effect of iteratively integrating these functions in the plot in Figure 2.12. We see that with each iteration, the N-CH function grows more slowly. In this plot we see that as $n \to \infty$, $H_G^{[n]}(t) \to 0$.

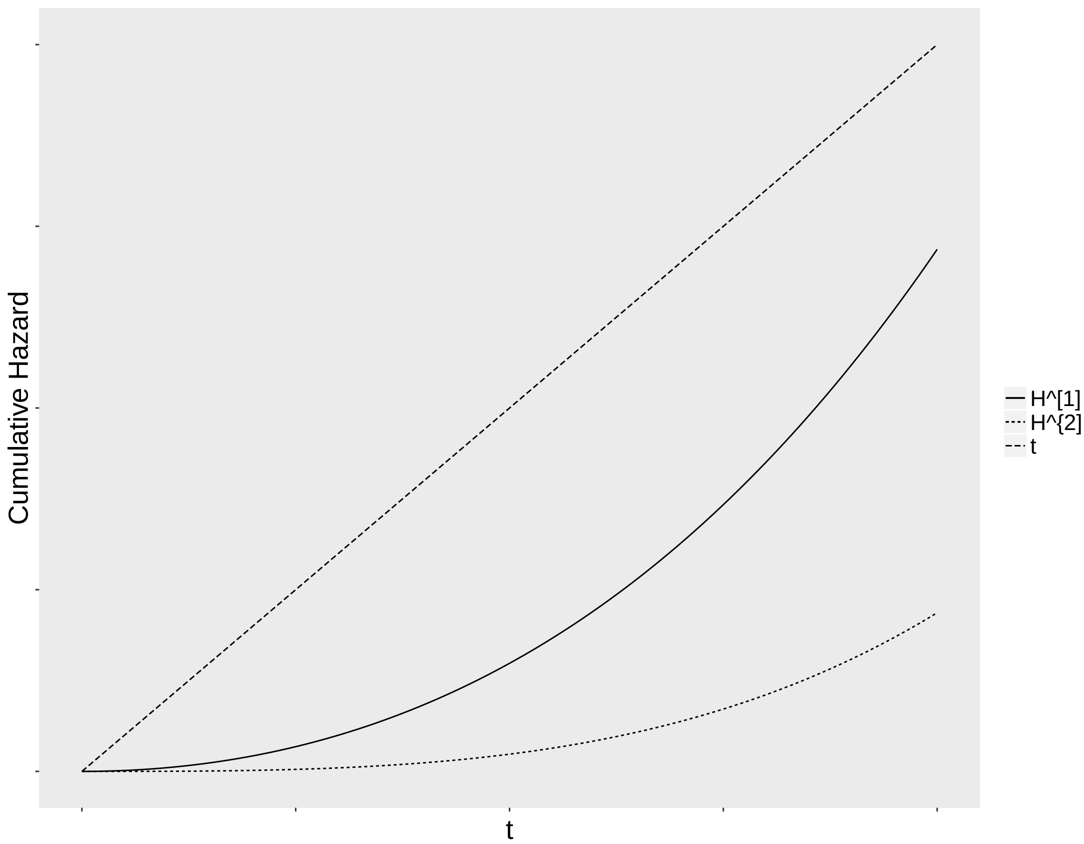

Figure 2.12: The N-CH functions $H^{[1]}(t)$ and $H^{[2]}(t)$ where $H(t) = e^t - 1$ as seen in example 2.67.

We now discuss some properties of N-CH functions.

Corollary 2.70. *$H(t)$ is an N-CH function if, and only if, $\frac{d}{dt}H(t)$ is a CH function.*

We can use this corollary to verify if a given function is an N-CH function. All we need to do is differentiate it.

Example 2.71. To verify that the function $\exp(t^2) - 1$ is an N-CH we differentiate it and check the derivative is a CH function. Its derivative is $2t\exp(t^2)$, which we can see is the product of a CH, $2t$, with a positive non-decreasing function, $\exp(t^2)$, and thus by Corollary 2.41 is a CH function.

Now we note that N-CH functions are closed under sums, products and compositions.

Proposition 2.72. *Given two N-CH functions $H^{[1]}$ and $\tilde{H}^{[1]}$ then*

1. $H^{[1]}(t) + \tilde{H}^{[1]}(t)$*, the sum of N-CH functions,*

2. $H^{[1]}(t)\tilde{H}^{[1]}(t)$*, the product of N-CH functions,*

3. $H^{[1]}(\tilde{H}^{[1]}(t))$*, the composition of N-CH functions,*

are N-CH functions.

Proof. For the first N-CH preserving operation, the sum, is immediate by linearity of the integral.

Consider the derivative of the product of $H^{[1]}(t)$ with $\tilde{H}^{[1]}(t)$

$$\frac{d}{dt}\left(H^{[1]}(t)\tilde{H}^{[1]}(t)\right) = H(t)\tilde{H}^{[1]}(t) + H^{[1]}(t)\tilde{H}(t).$$

This is a CH function since the product and sum of CH functions are CH functions. Hence the product is an N-CH.

Finally consider the derivative of the composition.

$$\frac{d}{dt}H^{[1]}(\tilde{H}^{[1]}(t)) = \tilde{H}(t)H(\tilde{H}^{[1]}(t))$$

which is a CH function. Thus the composition is an N-CH function. □

Now we observe some of the convexity properties of N-CH functions.

Definition 2.73. A twice differentiable CH function H is convex at a point t if and only if $\frac{d^2H}{dt^2}(t) \geq 0$.

Since $\frac{d^2}{dt^2}H^{[1]}(t) = \frac{d}{dt}H(t) = h(t) > 0$ then we have that N-CH functions are convex everywhere. They have increasing hazards since $H(t)$ is the hazard corresponding to $H^{[1]}(t)$.

Example 2.74. Let $\alpha \in [0, 1]$, then we have

$$H^{[1]}(\alpha t) = \int_0^{\alpha t} H(s)ds = \int_0^t H(\alpha u)\alpha du \leq \alpha \int_0^t H(u)du = \alpha H^{[1]}(t)$$

since H is increasing. So we have that

$$H^{[1]}(\alpha t) \leq \alpha H^{[1]}(t).$$

We can thus consider the difference of these functions, and we see that it is a CH function, i.e.

$$H(t) = \alpha H^{[1]}(t) - H^{[1]}(\alpha t)$$

is a CH function for all $\alpha \in [0, 1]$.

We can now define the complement of an N-CH function. This is the N-CH function generated by the inverse of a CH function. Since we know that the inverse is a CH, we know that its integral will be an N-CH function. Thus the purpose of this definition is to define the notation for the complement, and we will learn later why it is termed a *complement*.

Definition 2.75. Let $H^{[1]}(t)$ be an N-CH function. Its corresponding *complementary N-CH* is

$$H^{[-1]}(t) := \int_0^t H^{-1}(s)ds.$$

An equivalent definition is given in the context of *convex conjugate* functions as in the work of Rockafellar [55]. Here we have that

$$H^{[-1]}(t) = \sup_{s \geq 0} \left\{ ts - H^{[1]}(s) \right\}.$$

Thus if the inverse is not analytically tractable, we have another way to compute the complementary N-CH function.

To see this equivalence, first note that

$$\begin{aligned} H^{[-1]}(t) = \int_0^t H^{-1}(s)ds = \int_0^{H^{-1}(t)} vh(v)dv \\ = tH^{-1}(t) - H^{[1]}(H^{-1}(t)) \end{aligned}$$

using the change of variable $s = H(v)$ followed by integration by parts. Then note that the supremum in the alternative definition is attained when

$$\begin{aligned} 0 = \frac{d}{ds}\left(ts - H^{[1]}(s) \right) = t - H(s) \\ \to t = H(s) \to s = H^{-1}(t). \end{aligned}$$

Therefore, the supremum is $H^{[-1]}(t) = tH^{-1}(t) - H^{[1]}(H^{-1}(t))$ which is equivalent to the above.

We will demonstrate some basic properties of the complement. We will first consider what the complement of the complement will be.

Proposition 2.76. *Let $H^{[1]}$ be an N-CH function with complement $H^{[-1]}$, then $H^{[--1]} = H^{[1]}$, i.e. the complement of the complement of $H^{[-1]}$ is $H^{[1]}$.*

Proof.

$$H^{[--1]}(t) = \int_0^t (H^{-1})^{-1}(s)ds = \int_0^t H(s)ds = H^{[1]}(t).$$

□

Corollary 2.77. *If there exist two CH functions $H^{[1]}(t)$ and $\tilde{H}^{[1]}(t)$ such that $H^{[1]}(t) \leq \tilde{H}^{[1]}(t)$ for all $t \geq 0$, then*

$$H^{[-1]}(y) \geq \tilde{H}^{[-1]}(y)$$

for all $y \geq 0$.

Proof. The proof is immediate from the definition of the complementary N-CH function as a convex conjugate. □

Example 2.78. Now let's consider the complementary function to the N-CH function generated by the Gompertz CH, seen in example 2.69. Here $H_G^{-1}(t) = H_{ll}(t) = \log(1+t)$ and

$$H_G^{[-1]}(t) = \int_0^t \log(1+s)ds = (1+t)\log(1+t) - t = (1+t)H_G^{-1}(t) - t.$$

In example 2.67, the CH function is integrated iteratively to find $H^{[2]}(t)$ and $H^{[3]}(t)$. Thus we will iteratively integrate $H_G^{-1}(t)$ to compare.

$$\begin{aligned}
H_G^{[-2]}(t) &= \int_0^t \Big((1+s)\log(1+s) - s\Big)ds = \frac{1}{2}(1+t)^2\log(1+t) - \frac{3}{4}t^2 - \frac{1}{2}t \\
H_G^{[-3]}(t) &= \int_0^t H^{[-2]}(s)ds = \frac{1}{6}(1+t)^3\log(1+t) - \frac{11}{36}t^3 - \frac{5}{12}t^2 - \frac{1}{6}t.
\end{aligned}$$

Here we find that $H_G^{[-n]}(t)$ will have the term $\frac{1}{n!}(1+t)^n\log(1+t)$ in it, and the other terms are actually the first n terms of the series expansion of $\frac{1}{n!}(1+t)^n\log(1+t)$. Hence, as in the previous example, $H_G^{[-n]}(t) \to 0$ as $n \to \infty$. To see an illustration of this, see Figure 2.13.

The next theorem will show that N-CH functions and their complements satisfy Young's inequality for real valued, continuous, strictly increasing functions [28]

$$ab \leq \int_0^a H(x)dx + \int_0^b H^{-1}(x)dx$$

where $H(0) = 0$.

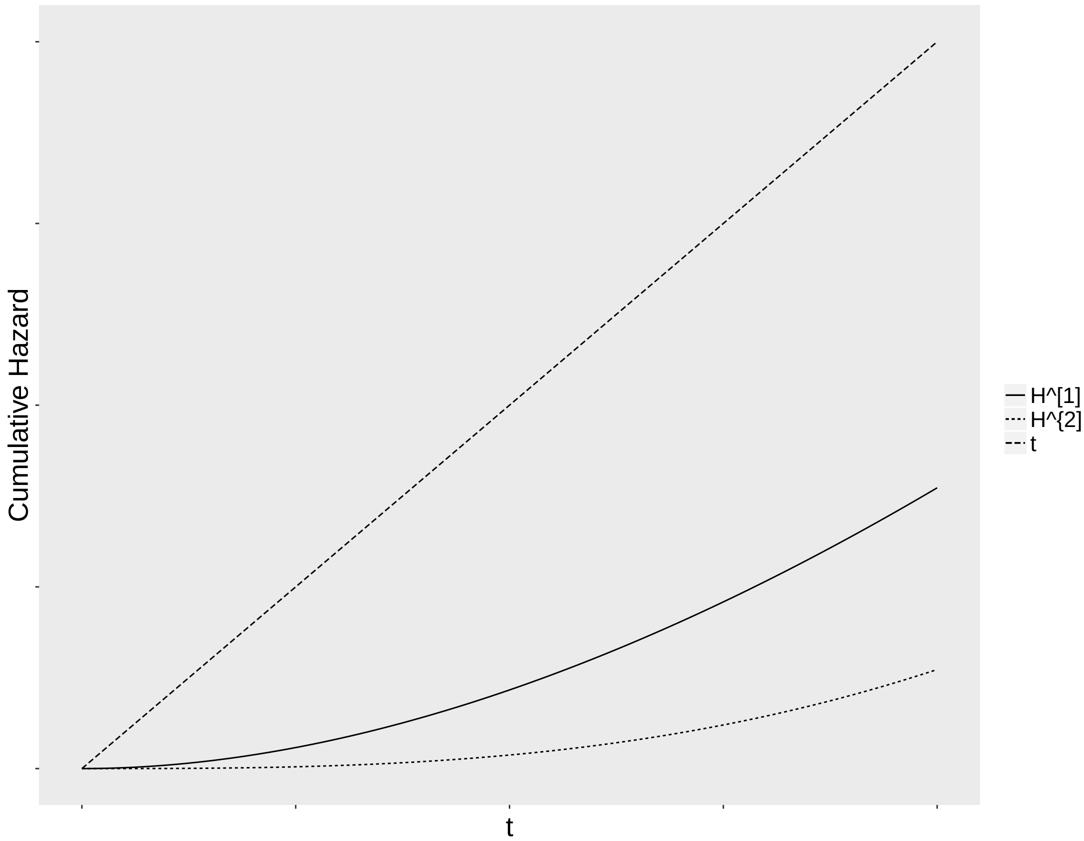

Figure 2.13: The N-CH functions $H^{[1]}(t)$ and $H^{[2]}(t)$ where $H(t) = \log(1+t)$ as seen in example 2.78.

Theorem 2.79. *Let $H^{[1]}$ be an N-CH function with complement $H^{[-1]}$, then for any $t, y \geq 0$*

$$ty \leq H^{[1]}(t) + H^{[-1]}(y). \tag{2.16}$$

This reduces to equality when $y = H(t)$, i.e.,

$$tH(t) = H^{[1]}(t) + H^{[-1]}(H(t)). \tag{2.17}$$

Proof. The inequality (2.16) follows directly from the definition of a complementary N-CH function as a convex conjugate.

$$\begin{aligned} H^{[-1]}(y) &= \sup_{t \geq 0} \left\{ ty - H^{[1]}(t) \right\} \\ &\geq ty - H^{[1]}(t). \end{aligned}$$

The supremum is attained when

$$\frac{d}{dt}\left(ty - H^{[1]}(t)\right) = 0,$$

i.e. when $y = H(t)$. Thus the equation in (2.17) follows. $\square$

Figure 2.14 is a graphical representation of this inequality. The grey region represents $H^{[1]}(t)$ and the white region is $H^{[-1]}(y)$, the complement of the grey region. We see that the sum of these two regions is larger than the rectangle ty, thus demonstrating the inequality in equation (2.16). We also see that if $y = \frac{d}{dt}H^{[1]}(t) = H(t)$, then the sum of the two regions would be the rectangle ty. We note then in particular, Young's inequality implies that $tH(t) \geq H^{[1]}(t)$, $\forall\, t > 0$. This however, is also straightforward from first principles.

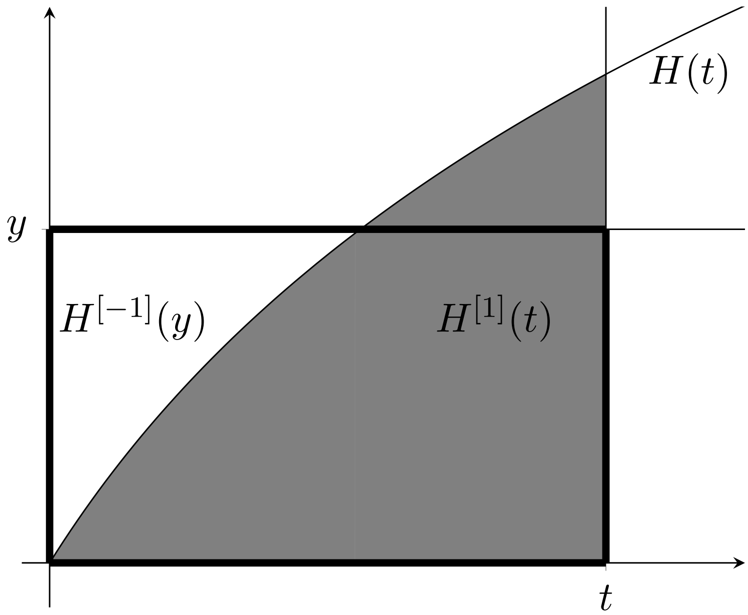

Figure 2.14: Graphical representation of Young's inequality. $H^{[1]}(t)$ is the grey region, $H^{[-1]}(y)$ is the white region in the box region and ty is the lined rectangle.

The next few results deal with integrating functions that are not CH functions. We will see which functions will integrate to be CH functions and see some examples.

Proposition 2.80. *Let $g : (0, \infty) \to (0, \infty)$ be a strictly increasing, continuous function. Then*

$$H(t) = \int_0^t g(s)ds$$

is a CH function.

Proof. The proof is much the same as the proof for corollary 2.65. The difference is g may be bounded above, by $M > 0$ say. Thus as $t \to \infty$, $g(t) \to M$. Hence $\frac{d}{dt}H(t) = g(t) \to M$. Then there must exist some $y > 0$ such that $g(t) > \frac{M}{2}$ for $t \geq y$. Then by the Mean-Value Theorem, there must exist some $c \in (y, t)$ such that

$$\begin{aligned} H(t) - H(y) &= g(c)(t - y) \\ &\geq \frac{M}{2}(t - y). \end{aligned}$$

Then as $t \to \infty$, we have that $H(t) \to \infty$ since M and y are fixed. Note that $\frac{d}{dt}H(t) = g(t)$ is continuous, thus H is continuously differentiable. □

Example 2.81. Let $g(t) = \theta$ for all $t > 0$, be a constant function. Then we have that

$$H(t) = \int_0^t g(s)ds = \theta t = H_E^\theta(t)$$

the exponential with scale θ.

We will see examples of functions which are not constant.

Example 2.82. To see an application of the result in proposition 2.80, let $g(t) = \tilde{F}(t)$, a CDF of a continuous positive random variable. Then

$$H(t) = \int_0^t \tilde{F}(s)ds$$

is a strictly convex CH function. Bassan et al. [4] investigate integrating CDF functions. Suppose $\tilde{F}$ is the CDF of a simple exponential distribution, i.e. $\tilde{F}(t) = 1 - e^{-t}$, then

$$H(t) = t + e^{-t} - 1 = t - (1 - e^{-t}),$$

which is the difference between a CH function and an increasing function. This remains a CH function as the increasing function $1 - e^{-t}$ grows slower than the CH function t.

Now suppose $\tilde{F}$ is the CDF of a simple log-logistic function, i.e. $\tilde{F}(t) = \frac{t}{1+t}$, then

$$H(t) = t - \log(1 + t),$$

which is the difference of two CH functions.

We have seen that integrating an increasing function yields a CH function. The next corollary will clarify the properties needed for the integral of a decreasing function to be a CH function.

Corollary 2.83. *Let $g : (0, \infty) \to (0, \infty)$ be a strictly decreasing, continuous function. Suppose also that g is bounded below by $M > 0$. Then $H(t) = \int_0^t g(s)ds$ is a CH function.*

Proof. The proof is much the same as the proof for corollary 2.65 and proposition 2.80. The difference is g is bounded below by $M \in (0, \infty)$. Thus as $t \to \infty$, $g(t) \to M$. We then have that

$$0 < M \leq g(t)$$

for all t. Thus, since $\int_0^\infty M ds = \infty$, we must have that $\int_0^\infty g(s)ds = \infty$ by the comparison test. □

Example 2.84. Consider the function $g(t) = 1+\tilde{S}(t)$, where $\tilde{S}(t)$ is a survival function. This function is bounded below by $1 > 0$. Thus, for $\tilde{S}(t) = e^{-t}$,

$$\int_0^t 1 + e^{-s} ds = 1 - e^{-t} + t$$

which is the CH function found in example 2.54.

The next proposition gives details on the types of functions that can be integrated to yield a CH function.

Proposition 2.85. *Let $g : (0, \infty) \to (0, \infty)$ be a continuous function that satisfies*

1. $g(0) < \infty$,
2. $g(t) = \Omega\left(\frac{1}{t}\right)$, *i.e.* $\exists\, k > 0, \exists\, t_0$ *sufficiently large so that* $\forall\, t > t_0$ $g(t) \geq \frac{k}{t}$.

Then $H(t) = \int_0^t g(s)ds$ *is a CH function.*

Proof. Since H is the integral of a positive function, it is itself positive. Since $g(t) > 0$ for all $t > 0$ we must also have that H is strictly increasing. It is continuously differentiable since g is continuous. It just remains to show that $H(\infty) = \infty$.

Since $g(t) = \Omega\left(\frac{1}{t}\right)$ then there exists $k > 0$ and t_0 such that $g(t) > k/t$ for all $t > t_0$. Thus

$$H(t) \geq \int_{t_0}^{\infty} g(t)dt \geq \int_{t_0}^{\infty} \frac{k}{t} dt = \infty.$$

Thus H is a CH. □

Note that this proposition does not require that g is monotone. This proposition thus includes function g that are increasing and decreasing.

Note here that H is defined as the integral of some function, then that function, by the Fundamental Theorem of Calculus, must necessarily be the corresponding hazard function. Thus proposition 2.85 gives necessary conditions for a function to be a hazard function.

Example 2.86. Consider $g(t) = \frac{1}{1+t}$. Here we have $g(0) = 1 < \infty$ and for $t > \frac{1}{4}$, $g(t) > \frac{1}{5t}$. We have that

$$H(t) = \int_0^t g(s)ds = \log(1 + t) = H_{ll}(t).$$

Example 2.87. Let $h_0(t)$ be a hazard function and let $g(t) = h_0(t) + \theta$ for some $\theta > 0$. Then

$$\int_0^t g(s)ds = H_0(t) + \theta t$$

which is the additive hazards model of example 2.51

In general, let ψ be a positive function such that $\int_0^t \psi(s)ds$ is a CH function. We will use the following notation

$$H^{\psi}(t) := \int_0^t \psi(s)ds.$$

We note the notation $H^{\psi}(t)$ is motivated by the fact that we will use ψ not as a hazard function, but it will turn out that we need ψ to behave like one. More specifically, we will use ψ as a linking function to introduce time varying covariates in chapter 4. We also note that we can write $H^H(t) = H^{[1]}(t)$ and that if $\psi(t) = \theta \; \forall t > 0$ then $H^{\psi}(t) = H_E^{\theta}(t)$.

Example 2.88. Let ψ be a hazard function and let $T_0 \sim H_0$. Consider the following "inverse" transformation of time $T_1 \sim H^{\psi}(T_0)$, then

$$S_{T_1}(t) = P(T_1 > t) = P\big(T_0 > H^{\psi}(t)\big) = S_0\big(H^{\psi}(t)\big).$$

This implies that $H_1(t) = \left(H_0 \circ H^{\psi}\right)(t)$. In the particular case where ψ is constant and equal to θ, we obtain that $T_1 = \frac{1}{\theta}T_0$ and $H_1(t) = H_0(\theta t)$, the accelerated failure time model of example 2.14. In this sense, the above model generalises the AFT model, the details are deferred until section 2.3.10.

Example 2.89. Let ψ be a hazard function and let H_0 be the baseline CH. We define $H_1(t) = \big(H^{\psi} \circ H_0\big)(t)$ and the corresponding time transformation is given by $T_1 \overset{d}{=} H_0^{-1} \circ (H^{\psi})^{-1} \circ H_0(T_0)$. If $\psi(t) = \theta$ for all $t > 0$ then $H_1(t) = \theta H_0(t)$, the proportional hazards model. In this sense, the above model generalises the proportional hazards model, details of which will be given in section 2.3.10.

2.3.9 Relationship with the Increasing Hazard Rate Average

In reliability, the concept of the Increasing, or Decreasing, Hazard Rate Average has been used [44]. In this section we will explore how this concept relates to our N-CH functions.

Definition 2.90. A continuous distribution of T with CH function $H_T(t)$, is Increasing Hazard Rate Average (IHRA) if

$$\frac{H_T(t)}{t}$$

is increasing.

Note that

$$\frac{H_T(t)}{t} = \frac{1}{t}\int_0^t h_T(u)du$$

and thus $\frac{1}{t}H_T(t)$ is the average of h_T over $(0, t)$, i.e. the hazard rate average. A result from Marshall and Olkin (2007) is that $h_T(t) \geq \frac{1}{t}H_T(t)$. This relates to Young's inequality in Theorem 2.79.

We will now see the relationships between N-CH functions and distributions which are IHRA.

Proposition 2.91. *Let $H^{[1]}(t)$ be an N-CH function. The distribution with CH function $H^{[1]}(t)$ is IHRA.*

Proof. The derivative

$$\begin{aligned}\frac{d}{dt}\frac{H^{[1]}(t)}{t} &= \frac{tH(t) - H^{[1]}(t)}{t^2} \\ &= \frac{H^{[-1]}(H(t))}{t^2},\end{aligned}$$

by Young's inequality equation (2.16), using $H(t) \geq \frac{1}{t}H^{[1]}(t)$. Alternatively, this can be seen directly from the fact that $\left(\frac{1}{t}H(t)\right)' \geq 0$. This is clearly positive since $H^{[-1]}$ is a CH function. Therefore $\frac{H^{[1]}(t)}{t}$ is increasing. □

Proposition 2.92. *The distribution with CH function $H^{[-1]}(t)$ is IHRA.*

Proof. Similar to the above proposition. □

Proposition 2.93. *Let $H^{[1]}$ be an N-CH function, then the following two properties hold:*

1. $\lim_{t\to 0} \dfrac{H^{[1]}(t)}{t} = 0$, *so that $H^{[1]}(t)$ goes to zero at a faster rate than t*
2. $\lim_{t\to \infty} \dfrac{H^{[1]}(t)}{t} = \infty$, *so that $H^{[1]}(t)$ goes to infinity at a faster rate than t*

Proof. Use of L'Hospital's rule shows that

$$\lim_{t\to 0} \frac{H^{[1]}(t)}{t} = \frac{\lim_{t\to 0} \frac{d}{dt}H^{[1]}(t)}{\lim_{t\to 0} \frac{d}{dt}t} = \frac{\lim_{t\to 0} H(t)}{1} = 0\,,$$

so that the first property is true and a similar argument shows the second one. □

Corollary 2.94.

$$\frac{H^{[1]}(t)}{t}$$

is a CH function.

Proof. By Proposition 2.93 we see that $\lim_{t\to 0} \frac{H^{[1]}(t)}{t} = 0$ and $\lim_{t\to\infty} \frac{H^{[1]}(t)}{t} = \infty$. From Proposition 2.91 we have that $\frac{H^{[1]}(t)}{t}$ is increasing. Thus it is a CH function. □

The above corollary essentially says that $H^{[1]}(t)$ tends to 0 faster than t does. To see that $H^{[1]}(t)$ goes to 0 faster than t, recall $H^{[1]}(t)$ in Example 2.67 and refer to the plot of this function in Figure 2.12. Note that around 0, $H^{[1]}(t)$ is much closer to the x-axis than t.

Example 2.95. Consider the N-CH function generated by the Weibull CH, i.e.

$$H^{[1]}(t|\alpha) = \frac{1}{\alpha+1} t^{\alpha+1}$$

for $\alpha > 0$. We see that the CH function $\frac{H^{[1]}(t|\alpha)}{t} = \frac{1}{\alpha+1} t^{\alpha}$. Thus we see that the Weibull family is closed under this type of operation.

Similar properties hold for iterated N-CH functions ($n \geq 2$).

Proposition 2.96. *$H^{[n]}(t)$ has the following properties:*

1. $\lim_{t \to 0} \frac{H^{[n]}(t)}{t^n} = 0\,,$
2. $\lim_{t \to \infty} \frac{H^{[n]}(t)}{t^n} = \infty\,.$

Proof. Consider the first property,

$$\begin{aligned}
\lim_{t \to 0} \frac{H^{[n]}(t)}{t^n} &= \lim_{t \to 0} \frac{H^{[n-1]}(t)}{nt^{n-1}} \text{ by L'Hospital's rule,} \\
&= \ldots \\
&= \lim_{t \to 0} \frac{H(t)}{n!}, \text{ again by L'Hospital's rule,} \\
&= 0.
\end{aligned}$$

The second property is similar. □

We can generalise this proposition to include general CH functions instead of just $\frac{H^{[n]}(t)}{t^n}$.

Corollary 2.97. *Let $H_0^{[1]}(t)$ be an N-CH and $H_1(t)$ and $H_2(t)$ be CH functions. Furthermore if $h_2(t) \in (0, \infty)\, \forall t$ and if $h_1(t) < \infty$ then*

1. $\lim_{t \to 0} \frac{H_0^{[1]}(H_1(t))}{H_2(t)} = 0$
2. $\lim_{t \to \infty} \frac{H_0^{[1]}(H_1(t))}{H_2(t)} = \infty$

Proof. 1.

$$\begin{aligned}
\lim_{t \to 0} \frac{H_0^{[1]}(H_1(t))}{H_2(t)} &= \lim_{t \to 0} \frac{h_1(t) H_0(H_1(t))}{h_2(t)} \text{ by L'Hospital's rule} \\
&= 0 \text{ if } h_2(t) \in (0, \infty) \text{ and } h_1(t) < \infty\, \forall t.
\end{aligned}$$

2. Similar to 1

□

Corollary 2.98. *The function*

$$H(t) = \frac{H_0^{[1]}(H_1(t))}{H_2(t)}, \tag{2.18}$$

where $H_0^{[1]}$, H_1, *and* H_2 *have the same properties as in corollary 2.97, is a CH function if* $\frac{d}{dt}\log H_1(t) > \frac{d}{dt}\log H_2(t)$.

Proof. We have seen that equation (2.18) satisfies $H(0) = 0$ and $H(\infty) = \infty$ in corollary 2.97. We just need to show that $H(t)$ is increasing, we will show that the derivative is positive. We want

$$\frac{d}{dt}\left(\frac{H_0^{[1]}(H_1(t))}{H_2(t)}\right) = \frac{H_2(t)h_1(t)H_0(H_1(t)) - h_2(t)H_0^{[1]}(H_1(t))}{H_2(t)^2} > 0.$$

So we would require that

$$\frac{h_1(t)H_0(H_1(t))}{H_2(t)} > \frac{h_2(t)H_0^{[1]}(H_1(t))}{H_2(t)^2}. \tag{2.19}$$

It can be shown that

$$H_0^{[-1]}(H_0(H_1(t))) = H_1(t)H_0(H_1(t)) - H_0^{[1]}(H_1(t))$$

which is positive since it is a CH function. Thus after some rearranging we find that

$$\frac{h_1(t)H_0(H_1(t))}{H_2(t)} > \frac{H_0^{[1]}(H_1(t))}{H_2(t)}\frac{h_1(t)}{H_1(t)}.$$

Since $\frac{d}{dt}\log H_1(t) > \frac{d}{dt}\log H_2(t)$, we have that

$$\frac{H_0^{[1]}(H_1(t))}{H_2(t)}\frac{h_1(t)}{H_1(t)} > \frac{H_0^{[1]}(H_1(t))}{H_2(t)}\frac{h_2(t)}{H_2(t)},$$

and thus equation (2.19) holds. Thus $H(t)$ is increasing and is thus a CH function. □

This corollary could be used to create rational models with time varying covariates.

2.3.10 Using cumulative hazard transformations to generalise models

In the final section of this chapter we discuss how to use the properties discussed in earlier sections to extend models to include time varying covariates.

This section will build on the properties we have explored throughout this chapter in order to generalise popular models in time-to-event analysis to incorporate time varying covariates. These time varying covariates will be denoted by $x(t)$.

In this section we focus on the so called external time varying covariate defined as follows.

Definition 2.99. Let $x(t)$ be a time varying covariate and let $X(t) = \{x(s), 0 \leq s < t\}$ denote the covariate history up to time t. Then $x(s)$ is external if

$$P\big(s < T \leq s + \epsilon | T > s,\, X(s)\big) = P\big(s < T \leq s + \epsilon | T > s,\, X(t)\big)$$

for all $0 < s \leq t$ and $\epsilon \to 0$.

The idea is that $x(s)$ is associated with the rate of the occurrence of the event of interest over time, but its future up to time $t > s$ is not affected by the occurrence of the event at time s. We note that a time varying covariate can be deterministic in the sense that its path is determined at the beginning of the study. Examples of such a covariate would be the time of day, or the predetermined stress level programme in a reliability study, or the predetermined dosage level programme in a drug study. Another type of external time varying covariate could be a stochastic process that is external to the individuals of the time-to-event study, whose event times are affected by $x(t)$ but the occurrence of an event does not alter the future of $x(t)$. A simple example of this kind of covariate would be the pollution level in a study about asthma.

Given our smoothness restriction we will only consider here the case where the realisations of $x(t)$ are smooth functions of time. In the cases where $X(t)$ is a stochastic process, even assuming that the corresponding realisations are smooth, in practice we only observe the covariate at certain times. To account for this, it has been suggested in the literature, for example in Tseng et al. [69], that $x(t)$, the time varying covariate, is modelled as follows:

$$x_i(t) = \sum_{j=1}^{q} b_{ji} \rho_j(t),$$

where $(\rho_1(t), ..., \rho_q(t))$ is vector of known basis functions and $(b_{1i}, ..., b_{qi})$ is a vector of (usually multivariate normal distributed) random effects.

The idea is that we will assume the function ψ used in section 2.3.8 is a function of $x(t)$, for example, $\psi(t) = e^{\beta x(t)}$, in a simple model with only one time varying covariate and n fixed time covariates. In this case we have that

$$\psi(t) = \exp\{\sum_{j=1}^{q} \beta b_j \rho_j(t)\}.$$

Our results will require that $\psi(t)$ is a hazard function.

Time varying Accelerated Failure Time model

First, recall proposition 2.16 where we explored the composition of the CH function H and the function g. It was found that g must also be a CH function. We use this proposition to generalise the AFT model in the next example.

Example 2.100. (Time varying AFT model 1) Consider the setting of example 2.88 where $H_1(t) = H_0(H^\psi(t))$ for some function $\psi(t)$, where $\psi(t)$ is of the form $\psi(t) = e^{\beta x(t)}$. The corresponding time transformation is given by $T_0 = H^\psi(T_1)$. This proposal is given by Cox and Oakes [21]. Lawless [41] proposes the function H^ψ as a generalised time scale.

By proposition 2.16 a necessary and sufficient condition for $H_1(t)$ to be a CH is for H^ψ to also be a CH function, or equivalently, for ψ to be a hazard function. Proposition 2.85 gives sufficient conditions for ψ to be a hazard function.

In this example, we can let $x_i(t) = \boldsymbol{b}_i^T \boldsymbol{\rho}(t)$, as in Tseng, Hsieh and Wang (2005). This formulation would then require that the model for the time varying covariates is linear in the log scale. In their paper, Tseng et al. propose the basis functions are $\boldsymbol{\rho}(t) = (\log(t), t-1))^T$.

A second proposal for generalising the AFT model to contain time varying covariates in discussed in the next example.

Example 2.101. (Time varying AFT model 2) Consider the baseline $T_0 \sim H_0$ and the time varying AFT model

$$H_1(t) = H_0(g(t)t)$$

where $g(t)$ is a function of $x(t)$.

From proposition 2.16 we see that $g(t)t$ must be a CH function, and from proposition 2.43 we see that $g(t)$ must satisfy $g(0) < \infty$, $g(\infty) > 0$ and $\frac{d}{dt} \log g(t) > -\frac{1}{t}$. Thus if $g(t) = e^{\beta x(t)}$, then we would require $\beta x'(t) > -\frac{1}{t}$.

As in the previous example, we can let $x(t) = \boldsymbol{b}_i^T \boldsymbol{\rho}(t)$. We would propose that $\boldsymbol{\rho}(t) = (\log(t), t)^T$, so that $g(t)$ is generated by the Gompertz and the Weibull depending on which basis function is used.

Time varying Proportional Hazards model

Now recall proposition 2.43 where we explored the product of the CH function H and the function $g : (0, \infty) \to (0, \infty)$. It was found that we must have that $g(0) < \infty$, $g(\infty) \neq 0$ and $\frac{d}{dt} \log g(t) > -\frac{d}{dt} \log H(t)$. We use this proposition to generalise the proportional hazards model in the next example.

Example 2.102. (Time varying Proportional Hazards model 1) Suppose that we have time varying covariates, $x(t)$ such that $g(t) = e^{\beta x(t)}$ satisfies the properties of g in proposition 2.43. Then

$$H_1(t) = e^{\beta x(t)} H_0(t)$$

is a cumulative hazard function which describes the time varying proportional hazards model.

Now we consider the restrictions on $x(t)$ so that this model is valid. In order for this model to be valid we need that $g(t) = e^{\beta x(t)}$ satisfies the properties of proposition 2.43. We need that $g(0) < \infty$, thus we need $x(0) < \infty$. This is reasonable and if we do have a covariate that does not conform to this, some transformation of it will. In order to have that $\lim_{t\to\infty} g(t) > 0$ we need that $\lim_{t\to\infty} x(t) \neq -\infty$. This again is reasonable, and a transformation could be made to make this true. We also require that the CH is strictly increasing, we will therefore need that $\beta x'(t) + \frac{1}{H_0(t)} H_0(t)' > 0$.

In this example we can let $x(t) = \boldsymbol{b}_i^T \boldsymbol{\rho}(t)$. This was seen in the work of Rizopoulos, Verbeke and Lesaffre [54]. Here the basis functions are proposed to be $\boldsymbol{\rho}(t) = (\log(t), t)^T$. This formulation requires that the log cumulative hazard is linear.

In the next example, we will use integration to propose a generalisation of the proportional hazards model.

Example 2.103. (Time varying Proportional Hazard model 2) Let us suppose

$$H_1(t) = \int_0^t e^{\beta x(s)} h_0(s) ds.$$

Note that if $x(t)$ is constant then this reduces to the proportional hazards model.

For this model to be valid, we require that $e^{\beta x(0)} h_0(0) < \infty$ and for $e^{\beta x(s)} h_0(s)$ not to tend to 0 too quickly as in the second condition of proposition 2.85.

There are a few interesting choices for $x(t)$ in this example. If $x(t) = \log H_0(t)$, then $H_1(t) = \frac{1}{\beta+1} H_0(t)^{\beta+1}$. If $x(t) = \frac{1}{\beta} t$, then $H_1(t) = H_0^{[-1]}(H_0(t))$, using Young's equality. Also trivially, if $x(t) = \frac{1}{\beta h_0(t)} H_0(t)$ we will have $H_1(t) = H_0^{[1]}(t)$. These functions would form our basis $\boldsymbol{\rho}(t)$.

Example 2.104. (Time varying Proportional Hazards model 3) Suppose that

$$H_1(t) = H^{\psi}(H_0(t)) = \int_0^{H_0(t)} \psi(s) ds.$$

If $\psi(t) = e^{\beta x(t)}$ as before, we find different conditions on $x(t)$ than in models 1 and 2. We require that H^{ψ} is a CH function, thus that ψ is a hazard function. Thus the conditions on $x(t)$ can be derived from proposition 2.85.

Time varying Additive Hazards model

The additive hazards model can be generalised to include time varying covariates using corollary 2.53. In this corollary the positive function g is added to the CH function H. It is found that the function g must have the property that $g(0) = 0$.

Example 2.105. (Time varying Additive Hazards model) Let H_0 be the baseline cumulative hazard function and let $g(t) = e^{\beta x(t)} - e^{\beta x(0)}$, then

$$H_1(t) = H_0(t) + e^{\beta x(t)} - e^{\beta x(0)}$$

is a time varying version of the additive hazards model.

Note that $g(0) = 0$, so we only require that $\frac{d}{dt} g(t) > -h_0(t)$. Thus we must have $x'(t)e^{\beta x(t)} > -h_0(t)$.

Time varying Accelerated Hazards model

Now recall corollary 2.98 where we explored the form of a rational CH function. The accelerated hazards model is of the form $H_1(t) = \frac{1}{\alpha} H_0(\alpha t)$. We may need a time-varying version of this model, corollary 2.98 could be used to create this model.

Example 2.106. (Time varying Accelerated Hazards model) Suppose that our model is given by

$$H_1(t) = \frac{H_0^{[1]}(g(t)t)}{r(t)},$$

where $g(t)$ and $r(t)$ are functions of the covariates. We require that $g(t)t$ is a CH function and hence must satisfy $\frac{d}{dt} \log g(t) > -\frac{1}{t}$ as in the time varying AFT example. We also require that $r(t)$ is a CH function. We will also require g being finite for all t. A more complicated requirement on these functions is that we need

$$\frac{d}{dt} \log (g(t)t) > \frac{d}{dt} \log r(t).$$

If $g(t) = r(t)$ then this condition would be satisfied.

For this example we could allow $g(t) = e^{\beta x(t)}$ and $r(t) = e^{\alpha z(t)}$. Here, $x(t)$ would contain covariates that might accelerate or decelerate time and $z(t)$ would contain covariates that might act proportionally. We would then require these covariates satisfy the condition that $\beta x'(t) + \frac{1}{t} > \alpha z'(t)$. In this formulation, we could let $x(t) = \boldsymbol{b}_i^T \boldsymbol{\rho}(t)$ and $z(t) = \tilde{\boldsymbol{b}}_i^T \boldsymbol{\rho}(t)$. Good choices of basis functions might be $\boldsymbol{\rho}(t) = (\log(t), t)^T$.

2.4 Summary and future work

This chapter has described the properties of cumulative hazard functions. We have seen that compositions, inverses, products, sums and integrals of cumulative hazards are themselves cumulative hazards. We have also explored how to expand the set of cumulative hazards. The final section of this chapter has given some detail on how the properties previously discussed can be implemented in the construction of time varying regression models.

Cumulative hazard functions are fairly restricted functions. They must be increasing, be zero at zero and infinity at infinity. An advantage of this restriction is that these functions can essentially be classified by their local or global convexity or concavity. A possible goal of future research would be to explore how to characterise, in a simple way, the local convexity or concavity of a cumulative hazard. In terms of convexity, we would pursue the concept of the convex conjugate further, and explore its uses in modelling, possibly in dispersion models.

The next chapter will explore how to use the properties discussed in this chapter to build new models.

Chapter 3

Parametric Families of Cumulative Hazard Functions

3.1 One-dimensional parametric families

In this section we construct one-dimensional parametric families of cumulative hazards using the basic operations (compositions, inverse, etc.) introduced in Chapter 2. In particular, we would like to compose multiple cumulative hazard functions. We will introduce some notation here to simplify the expressions. If we wish to compose H_A^θ followed by H_B^α, then this will be denoted by

$$H_{AB}^{\theta,\alpha}(t) := H_A^\theta \circ H_B^\alpha(t).$$

Note that the parameters are separated by a comma. An example where one of the CH functions is not parametrised is $H_{EG}^\theta(t) = H_E^\theta \circ H_G(t)$.

We note here that a cumulative hazard function maps a time t to $H(t)$, which although it is in $(0, \infty)$, is not time, it is a transformation of the original time scale. Thus to avoid confusion, we would like to view the cumulative hazard functions as functional operators. We will also non-dimensionalise our time variable so that we can focus on the impact of the transformations we discuss in this chapter, i.e. the shapes and distortions of the transformations. In order to non-dimensionalise a variable, one might divide by the mean of that variable, as seen in the work of Shaddick and Zidek [61]. We will divide by one unit of time, i.e. if time is measured in seconds, we divide by one second. This will mean that the range of the variable will be unchanged and when plotting, will appear the same. Hence, when we see $H(t)$ we implicitly mean $H \circ (H_E^1)^{-1}(t)$. Note that H could be any CH function. We will not include this composition in any equation henceforward, as this will result in complicated looking functions.

The underlying baseline will have cumulative hazard $H_0(t)$ and the cumulative hazard of the parametric family, with parameter θ, is denoted $H(t|\theta)$. The random variable associated with the CH $H(t|\theta)$ will be $T_\theta \stackrel{d}{=} \overrightarrow{\tau}(T_0)$ as described in Chapter 2 in Figure 2.7. Most of the terminology for the families in the upcoming sections will follow that used by Marshall and Olkin [44].

We will also be interested in the asymptotic behaviour of a cumulative hazard H when either $t \to 0$ or when $t \to \infty$. The former we will call short term behaviour, and the latter, long term behaviour. This will be denoted by $H(t) \sim g(t)$ when $t \to a$, where $a = 0$ or $a = \infty$, for some function $g(t)$. This notation is shorthand for $\lim_{t\to a} \frac{H(t)}{g(t)} = 1$. In either case, we will say that $H(t)$ behaves like $g(t)$ when $t \to a$.

3.1.1 Families with a scale parameter

The first parametric family we will describe is the addition of a scale parameter to the underlying baseline distribution.

Definition 3.1. Suppose that the cumulative hazard of a parametric family is

$$H(t|\theta) = H_0(\theta t), \qquad \theta > 0.$$

Then the parameter θ is called a scale parameter.

This can be written in terms of a composition. The family generated by a scale parameter is equivalent to

$$H(t|\theta) = H_0 \circ H_E^\theta(t), \tag{3.1}$$

where H_E^θ is the cumulative hazard of an exponential distribution with rate θ.

Example 3.2. Letting $H_0(t) = H_G$, the standard Gompertz CH, in equation (3.1), it is easy to see that $H(t|\theta) = H_G \circ H_E^\theta(t) = H_{GE}^\theta(t) = e^{\theta t} - 1$.

Example 3.3. Suppose $H_0(t) = H_{ll}(t) = \log(1+t)$ is the CH function of the underlying baseline distribution, the standard log-logistic. Then, if we add a scale parameter, we have that

$$H(t|\theta) = H_{ll} \circ H_E^\theta(t) = H_{llE}^\theta(t) = \log(1 + \theta t).$$

The next paragraph will detail what time transformation is generated by the scale parameter.

Time Transformation ***(for scale parameter families)*** *If T_0 is the random variable corresponding to the distribution generated with H_0, then $T_\theta \stackrel{d}{=} T_0/\theta$ is the random variable corresponding to the distribution generated by $H(t|\theta)$.*

The incorporation of a scale parameter into a distribution is the basis of the Accelerated Failure Time model, as discussed in section 4.1. In this model the scale parameter θ accelerates or decelerates the progression of time to event, as shown above, depending on whether θ is less than or greater than 1.

Change of Baseline *Consider a reparametrisation of the family defined by* $\alpha = \frac{\theta}{\theta_1}$ *for some* $\theta_1 > 0$. *Let* $H_1(t) = H_0(\theta_1 t)$. *Then we have that*

$$H(t|\alpha) = H_0(\theta\alpha t) = H_1(\alpha t)$$

so that H_1 *now plays the role of the baseline in the new parametrisation. Therefore, any member of the scale parameter family can be made to be the baseline.*

3.1.2 Families with a frailty parameter

Another important parametric family is that of including a frailty parameter.

Definition 3.4. Suppose that the cumulative hazard of a parametric family is

$$H(t|\theta) = \theta H_0(t), \qquad \theta > 0.$$

Then the parameter θ is called a frailty parameter.

This family is very important as it describes the family of Proportional Hazards models, which are the most well known models within survival analysis. When θ is considered to be a random variable such models are often called *frailty models* [29] and so θ is termed a frailty parameter [44]. This type of parameter is sometimes called a proportional hazards parameter, but we will continue with the frailty terminology here.

The addition of a frailty parameter results in the survival function of the underlying baseline distribution being raised by the power θ, i.e.

$$S(t|\theta) = S_0(t)^\theta.$$

The addition of a frailty parameter can be written in terms of a composition of cumulative hazards. The family generated by a frailty parameter is equivalent to

$$H(t|\theta) = H_E^\theta \circ H_0(t).$$

Note that this is the reverse composition to equation (3.1), thus there is a direct connection between scale and frailty parameters. There will be connections of the same type for other parameters we will discuss in this chapter.

Example 3.5. Let $H_0(t) = H_G(t)$ be the baseline CH, the family generated by the frailty parameter is given by

$$H(t|\theta) = H_E^\theta \circ H_G(t) = H_{EG}^\theta(t) = \theta\left(e^t - 1\right).$$

Example 3.6. Let $H_0(t) = H_{ll}(t)$ be the baseline CH, the family generated by the frailty parameter is given by

$$H(t|\theta) = H_E^\theta \circ H_{ll}(t) = H_{Ell}^\theta(t) = \theta \log(1+t).$$

We can find the corresponding time transform for the frailty parameter using theorem 2.27.

Time Transformation ***(for frailty parameter families)*** *Suppose* $T_0 \sim H_0$ *and* $T_\theta \sim \theta H_0$. *Using theorem 2.27 we have that* $T_\theta \overset{d}{=} \overrightarrow{\tau}(T_0)$ *where* $\overrightarrow{\tau}(T_0) = H_0^{-1}(\frac{1}{\theta}H_0(T_0))$. *This implies that the transformed time* $H_0(T_\theta)$ *is an accelerated, or decelerated, version of the baseline transformed time* $H_0(T_0)$, *which always follows a standard exponential distribution by Proposition 2.9 in Chapter 2.*

Change of Baseline *Consider a reparameterisation of the family defined by* $\alpha = \frac{\theta}{\theta_1}$ *for some* $\theta_1 > 0$ *and let* $H_1(t) = \theta_1(t)$. *Then we have that*

$$H(t|\theta) = \alpha\theta_1 H_0(t) = \alpha H_1(t)$$

so that H_1 *now plays the role of the baseline in the new parametrisation. Thus any member of the frailty parameter family can be made to be the baseline.*

3.1.3 Families with a power parameter

Definition 3.7. Suppose that the cumulative hazard of a parametric family is

$$H(t|\theta) = H_0(t^\theta), \qquad \theta > 0.$$

Then the parameter θ is called a power parameter.

In terms of a composition, the power parameter is

$$H(t|\theta) = H_0 \circ H_W^\theta(t), \tag{3.2}$$

that is, composing with the Weibull. Note that the power parameter is equivalent to a scale parameter on the log scale

Example 3.8. Consider $H_0(t) = H_{ll}(t) = \log(1+t)$. We find the addition of a power parameter yields the CH function $H(t|\theta) = H_{ll} \circ H_W^\theta(t) = H_{llW}^\theta = \log(1+t^\theta)$.

Example 3.9. Suppose $T_0 \sim H_G$. Adding a power parameter yields the cumulative hazard $H(t|\theta) = H_{GW}^\theta(t) = \exp(t^\theta) - 1$.

The time transformation that generates the addition of a power parameter is given next.

Time Transformation ***(for power parameter families)*** *If* $T_0 \sim H_0$, *then if* $T_\theta \overset{d}{=} T_0^{1/\theta}$, *then* $H(t|\theta) = H_0(t^\theta)$. *Thus the baseline time is accelerated when raising to a* *power larger that one ($\theta < 1$), or decelerated when raising to a power less than one ($\theta > 1$).*

Change of Baseline *Consider the reparametrisation $\alpha = \frac{\theta}{\theta_1}$ for some $\theta_1 > 0$ and let $H_1(t) = H_0(t^{\theta_1})$. Then we have that*

$$H(t|\theta) = H_0(t^{\alpha\theta_1}) = H_1(t^{\alpha})$$

so that H_1 now plays the role of the baseline in the new parametrisation. Therefore any member of the power parameter family can be made to be the baseline.

3.1.4 Families with a hazard power parameter

The hazard power parameter, as it is termed in Marshal and Olkin [44], is given by powering the cumulative hazard, not the hazard function.

Definition 3.10. Suppose the cumulative hazard function of a parametric family is

$$H(t|\theta) = H_0(t)^{\theta}, \qquad \theta > 0.$$

Then the parameter θ is called a hazard power parameter

Note that $H(t|\theta) = H_0(t)^{\theta}$ can be seen as

$$\log H(t|\theta) = \theta \log H_0(t).$$

Thus the hazard power parameter can be interpreted as the proportionality parameter for the proportional log cumulative hazard model.

The hazard power parameter can be written as a composition of cumulative hazards,

$$H(t|\theta) = H_W^{\theta} \circ H_0(t).$$

We again note that the hazard power parameter is the opposite composition to the power parameter in equation (3.2). We have that there is a pairing between the power and hazard power parameters.

The next example will demonstrate that the Weibull family is closed under the addition of a hazard power parameter.

Example 3.11. Consider the addition of a power parameter, θ to the Weibull CH with parameter β, i.e.

$$H(t|\theta, \beta) = \left(H_W^{\beta}(t)\right)^{\theta} = t^{\theta\beta}.$$

We see that this is another Weibull with parameter $\theta\beta$, i.e. $H(t|\theta, \beta) = H_W^{\theta\beta}(t)$.

We note that the addition of a hazard power parameter to the Exponential distribution will result in a Weibull distribution.

Example 3.12. We can introduce a hazard power parameter into the log-logistic distribution that has CH function

$$H(t|\theta) = [\log(1+t)]^{\theta} = H_W^{\theta} \circ H_{ll}(t) = H_{Wll}^{\theta}(t).$$

Note that this transformation was seen in chapter 2 when we showed that powering a CH function is itself a CH function.

Time Transformation ***(for hazard power parameter families)*** *The corresponding time transformation for the hazard power parameter is given by* $T_\theta \stackrel{d}{=} H_0^{-1} \circ H_W^{1/\theta} \circ H_0(T_0)$*. This relationship can be simplified to* $H_0(T_\theta) \stackrel{d}{=} H_0(T_0)^{1/\theta}$*. Here we see again the acceleration (deceleration) of the transformed baseline time* $H_0(T_0)$*.*

Change of Baseline *Consider a reparametrisation of the family defined by* $\alpha = \frac{\theta}{\theta_1}$ *for some* $\theta_1 > 0$ *and let* $H_1(t) = H_0(t)^{\theta_1}$*. Then*

$$H(t|\theta) = (H_0(t))^{\theta_1 \alpha} = H_1(t)^{\alpha}$$

so that H_1 *now plays the role of the baseline in the new parametrisation. Thus any member of the hazard power family can be made to be the baseline.*

3.1.5 Families related to a tilt parameter

A tilt parameter is best described in terms of the survival function, $S(t)$, see Marshall and Olkin section 7.F [44], but since our focus is on the cumulative hazard, we will define a tilt parameter in terms of the cumulative hazard first.

Definition 3.13. Suppose that the cumulative hazard function of a parametric family is

$$H(t|\theta) = \log\left(1 + \theta(e^{H_0(t)} - 1)\right), \qquad \theta > 0. \tag{3.3}$$

Then the parameter θ is called a tilt parameter.

This is a sequential composition that involves the log-logistic, exponential, Gompertz and baseline CH, i.e.

$$\begin{aligned} H(t|\theta) &= H_{ll} \circ (\theta H_G(H_0(t))) = H_{ll} \circ H_E^{\theta} \circ H_G \circ H_0(t) \\ &= H_{llEG}^{\theta} \circ H_0(t), \end{aligned} \tag{3.4}$$

where, notationally, $H_{ll} \circ H_E^{\theta} \circ H_G(t) = H_{llEG}^{\theta}(t)$.

Example 3.14. Suppose $H_0(t) = H_{ll}(t)$, the standard log-logistic. The addition of a tilt parameter simply gives $H(t|\theta) = H_{llE}^{\theta}(t) = \log(1+\theta t)$, using the fact that H_G and H_{ll} are inverses of each other. We saw previously this is the same as adding a scale parameter, hence for the log-logistic distribution, tilt and scale parameters have the same interpretation.

In the literature, the addition of the tilt parameter is generally viewed in terms of the survival function, S, specifically

$$S(t|\theta) = \frac{\theta S_0(t)}{1 - (1-\theta)S_0(t)}.$$

Time Transformation ***(for tilt parameter families)*** *The time transformation generated by the addition of a tilt parameter is given by*

$$T_\theta \stackrel{d}{=} H_0^{-1} \circ H_{ll}\left(\frac{1}{\theta} H_G \circ H_0(T_0)\right) = H_0^{-1} \circ H_{llEG}^{1/\theta} \circ H_0(T_0)$$

where $T_0 \sim H_0$ and $T_\theta \sim H_{llEG}^{\theta} \circ H_0$.

It is interesting to note that $\left(H_{llEG}^{\theta}\right)^{-1}(t) = H_{llEG}^{1/\theta}(t)$, hence the simplified formula above. We also note that $H_{llEG}^{1}(t) = H_{E}^{1}(t)$, the standard exponential.

The addition of a tilt parameter to a distribution is the basis for the proportional odds model. The odds of the event are $(1 - S(t))/S(t)$. The model is proportional because

$$\frac{1 - S(t|\theta)}{S(t|\theta)} = \theta \frac{1 - S_0(t)}{S_0(t)}.$$

It is interesting to note the long term and short term effects of the tilt parameter family of CHs. To see this we look at the asymptotic behaviours of the CH function as $t \to 0$ and $t \to \infty$. The asymptotic expansion of H_{llEG}^{θ} when $t \to 0$ has leading term θt. In this case we write $H_{llEG}^{\theta}(t) \sim \theta t$ when $t \to 0$. Thus we see that the tilt parameter acts similarly to a frailty parameter when $t \to 0$. Similarly, for large t, $H_{llEG}^{\theta} \sim t$ for all $\theta > 0$, i.e. so that the CH (3.3) behaves like $H_0(t)$ in the long term.

Each of the parameters we have mentioned so far can be written in terms of compositions. In doing this we have noticed there are pairings amongst the parameters, scale/frailty and power/hazard power, where the order of compositions has been reversed. We can perform this operation with the composition for the tilt parameter (3.4) to define a different family, namely

$$H(t|\theta) = H_0 \circ H_{llEG}^{\theta}(t) = H_0\left(\log\left(1 + \theta(e^t - 1)\right)\right). \tag{3.5}$$

The parameter θ will now be the natural pairing to the tilt parameter and we will call it a reverse-tilt parameter.

Example 3.15. Let $H_0(t) = H_{llEG}^{\theta_0}(t)$ for some θ_0. Then we have

$$H(t|\theta, \theta_0) = H_{llEG}^{\theta} \circ H_{llEG}^{\theta_0} = \log\left(1 + \theta\theta_0\left(e^t - 1\right)\right) = H_{llEG}^{\theta\theta_0}(t),$$

so that the family of CHs $\left\{H_{llEG}^{\theta} : \theta > 0\right\}$ is closed under composition.

Time Transformation ***(for reverse tilt parameter families)*** *The time transfor-*

mation for the reverse-tilt parameter is

$$T_\theta = H_{llEG}^{1/\theta}(T_0).$$

Change of Baseline *Consider the reparamatrisation* $\alpha = \frac{\theta}{\theta_1}$ *for some* $\theta_1 > 0$ *and let* $H_1(t) = H_0 \circ H_{llEG}^{\theta_1}(t)$*. Then we have that*

$$\begin{aligned} H(t|\theta) &= H_0 \circ H_{llEG}^{\theta_1 \alpha}(t) = H_0 \circ H_{llEG}^{\theta_1} \circ H_{llEG}^{\alpha}(t) \\ &= H_1 \circ H_{llEG}^{\alpha}(t), \end{aligned}$$

using the closure property shown in example 3.15. Then H_1 *now plays the role of the baseline in the new parametrisation and any member of the reverse tilt parameter family can be made to be the baseline. Clearly this is also true for the tilt parameter family.*

In figure 3.1, we plot the function $H_{llEG}^{\theta}(t)$ for various values of θ. We see that for values of $\theta > 1$ the function $H_{llEG}^{1/\theta}(t)$ is below the identity line, thus we have deceleration of the baseline time. For values of $\theta < 1$ we have the function $H_{llEG}^{1/\theta}$ is above the identity line and thus have acceleration of the baseline time T_0. We can clearly see the long term behaviour from the plot as all the CH curves become parallel as t increases. The short term behaviour is less obvious from the plot alone, but from the discussion earlier we see it acts like θt close to $t = 0$.

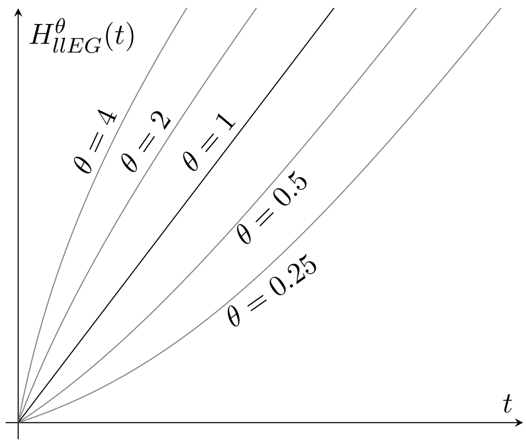

Figure 3.1: The CH function H_{llEG}^{θ} for $\theta = 0.25$, 0.5, 1, 2, 4.

Example 3.16. If we let $H_0(t) = H_G(t)$ in the reverse-tilt family, then $H(t|\theta) = H_{EG}^{\theta}(t) = \theta H_G(t)$. Thus, for the Gompertz distribution, frailty and reverse-tilt parameters have the same interpretation.

3.1.6 Families related to a resilience parameter

There are some parametric families that are slightly more difficult to express in terms of the cumulative hazard function. Letting $F_0(t)$ denote the cumulative distribution function for the underlying baseline time, then we can define a resilience parameter. Families with a resilience parameter are often called exponential distributions.

Definition 3.17. Suppose that the cumulative distribution function of a parametric family is

$$F(t|\theta) = F_0(t)^\theta, \qquad \theta > 0.$$

Then the parameter θ is called a resilience parameter.

In terms of the cumulative hazard, the addition of a resilience parameter is given by

$$\begin{aligned} H(t|\theta) &= \log\left(\frac{1}{1 - F_0(t)^\theta}\right) \\ &= \log\left(\frac{1}{1 - \left(1 - e^{-H_0(t)}\right)^\theta}\right). \end{aligned}$$

It can be shown that

$$H(t|\theta) = H_{ll} \circ H_r^\theta \circ H_G \circ H_0(t) = H_{llrG}^\theta(t) \circ H_0(t) \tag{3.6}$$

where

$$H_r^\theta(t) = \frac{t^\theta}{(1+t)^\theta - t^\theta}$$

is a cumulative hazard function which is a rational function, hence the r subscript.

Example 3.18. In this example we will add a resilience parameter to the exponential distribution. The CDF of the exponential distribution with CH function H_E^1 is $F_E^1(t) = 1 - \exp(-t)$. Thus the exponential distribution with a resilience parameter has cdf

$$F(t|\theta) = [1 - \exp(-t)]^\theta.$$

This distribution is called the Verhulst distribution in Marshall and Olkin [44] and has CH function

$$H(t|\theta) = \log\left(\frac{1}{1 - [1 - \exp(-t)]^\theta}\right).$$

This distribution is also called the exponentiated exponential distribution within the literature [26].

Recalling that adding a frailty parameter is equivalent to powering a survival function, i.e. $S(t|\beta) = S_0(t)^\beta$, and that $F(t) = 1 - S(t)$, we see that there is some duality between resilience and frailty parameters. This is further explored in the following proposition.

Proposition 3.19. *If $T_0 \sim \theta H_0$ and $T_\theta \stackrel{d}{=} 1/T_0$, then θ is a resilience parameter for T_θ.*

This proposition says that if θ is a frailty parameter for T_0, then if $T_\theta \stackrel{d}{=} 1/T_0$, θ is a resilience parameter for T_θ. We do not expect this type of relationship between parameters to be unique to the frailty and resilience parameters, but do not explore this further.

Time Transformation ***(for resilience parameter families)*** *The time transformation for the resilience parameter is given by* $T_\theta = H_0^{-1} \circ H_{llrG}^{1/\theta} \circ H_0(T_0)$.

It is interesting to note that $\left(H_{llrG}^{\theta}\right)^{-1}(t) = H_{llrG}^{1/\theta}(t)$. This CH function can be seen in figure 3.2 for different values of θ where we can see that it looks similar to H_{llEG}^{θ} in figure 3.1. Here however, values of $\theta < 1$ now mean $H_{llrG}^{1/\theta}$ is decelerating and values of $\theta > 1$ correspond to accelerating $H_0(T_0)$.

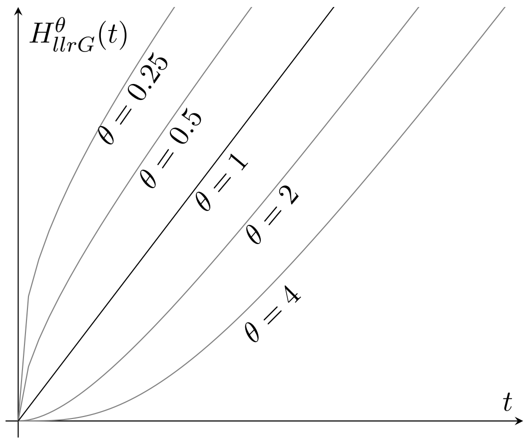

Figure 3.2: The function H_{llrG}^{θ} for $\theta = 0.25$, 0.5, 1, 2, 4.

Recalling the definition of the resilience parameter in terms of the CDF, we note that the family can be written as

$$\log F(t|\theta) = \theta \log F_0(t).$$

Thus the family with a resilience parameter can be viewed as a proportional log CDF model.

We investigate the behaviour of the resilience parameter family around $t = 0$ and infinity. We find that around $t = 0$ the cumulative hazard acts like t^θ, i.e. the resilience parameter acts like a hazard power parameter. Exploring the behaviour at infinity, we find H_{llrG}^{θ} acts like t, so that the CH of the family behaves like $H_0(t)$.

If we reverse the composition in equation (3.6), then we define a different family,

$$H(t|\theta) = H_0 \circ H_{llrG}^{\theta}(t). \tag{3.7}$$

The parameter θ is the pairing to the resilience parameter and will be called the reverse-resilience parameter.

Example 3.20. Let $H_0(t) = H_{llrG}^{\theta_0}$ for some $\theta_0 > 0$. Then we have

$$H(t|\theta_0, \theta) = H_{llrG}^{\theta_0} \circ H_{llrG}^{\theta}(t) = H_{ll} \circ H_r^{\theta_0} \circ H_r^{\theta} \circ H_G(t).$$

It is easy to show that $H_r^{\theta_0} \circ H_r^{\theta} = H_r^{\theta_0 \theta}$. Thus $H(t|\theta_0, \theta) = H_{llrG}^{\theta_0\theta}(t)$, so the family of CHs $\left\{H_{llrG}^{\theta} : \theta > 0\right\}$ is closed under composition.

Time Transformation *(for reverse resilience parameter families)* *The time transformation for the reverse-resilience parameter is* $T_\theta = H_{llrG}^{1/\theta}(T_0)$. *As seen in figure 3.2, we will have acceleration of the baseline time for values of* $\theta < 1$ *and deceleration for* $\theta > 1$.

Change of Baseline *Consider the reparametrisation* $\alpha = \frac{\theta}{\theta_1}$ *for some* $\theta_1 > 0$ *and let* $H_1(t) = H_0 \circ H_{llrG}^{\theta_1}(t)$. *Then we have that*

$$\begin{aligned} H(t|\theta) &= H_0 \circ H_{llrG}^{\theta_1 \alpha}(t) = H_0 \circ H_{llrG}^{\theta_1} \circ H_{llrG}^{\alpha}(t) \\ &= H_1 \circ H_{llrG}^{\alpha}(t), \end{aligned}$$

using the closure property shown in example 3.20. Then H_1 *now plays the role of the baseline in the new parametrisation. Thus any member of the reverse resilience parameter family can be made to be the baseline. It is easy to show that this is also true for the resilience parameter family.*

3.1.7 Families of alternative parameter pairings

In previous sections we have emphasised the existence of parameter pairings. We have also proposed pairings for other parameters in the literature. Now we propose a new type of parameter and its corresponding pairing.

Consider the family

$$H(t|\theta) = H_{GEll}^{\theta} \circ H_0(t). \tag{3.8}$$

In this family there is proportionality between what we call the logistic cumulative hazard functions, i.e.

$$\log(1 + H(t|\theta)) = H_{ll} \circ H(t|\theta) = H_E^{\theta} \circ H_{ll} \circ H_0(t) = \theta \log(1 + H_0(t)).$$

We will call this parameter the proportional logistic hazards parameter.

Time Transformation *(for proportional logistic hazard parameter families)* *The time transformation for the logistic hazards parameter is given by* $T_\theta = H_0^{-1} \circ H_{GEll}^{1/\theta} \circ H_0(T_0)$. *This is equivalent to* $H_0(T_\theta) = H_{GEll}^{1/\theta} \circ H_0(T_0)$. *Figure 3.3 demonstrates the form* $H_{GEll}^{\theta}(t)$ *takes for various values of* θ. *We see* $H_{llEG}^{1/\theta}$ *is below the identity for* $\theta > 1$ *and thus would result in deceleration and acceleration of* $H_0(T_0)$ *for* $\theta < 1$.

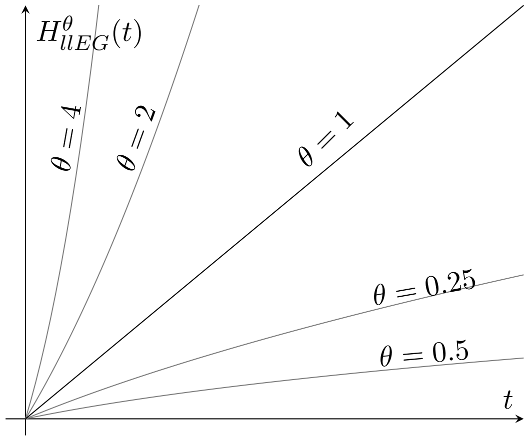

Figure 3.3: The function H^{θ}_{GEll} for $\theta = 0.25$, 0.5, 1, 2, 4.

Note that, we have again, $\left(H^{\theta}_{GEll}\right)^{-1}(t) = H^{1/\theta}_{GEll}(t)$.

We can investigate the asymptotic behaviour when $t \to 0$ and $t \to \infty$. It is easy to show that, and is confirmed in figure 3.3, that when $t \to 0$, $H^{\theta}_{GEll}(t) \sim \theta t$ so the reverse tilt parameter is similar to a frailty parameter in the short term. It can be shown that $H_{GEll} \sim t^{\theta}$ when $t \to \infty$, thus the reverse-tilt parameter acts like a power hazard parameter in the long term.

The reason for the introduction of this family is due to the fact that a distribution exists in the literature that can be written as a specific example of this family. This distribution and its relation to the proposed family is discussed in the next example.

Example 3.21. The Generalised Power Weibull family is proposed by Nikulin and Haghighi [27]. The CH function of this distribution is given by

$$H(t|\alpha, \theta) = (1 + t^{\alpha})^{\theta} - 1 = H^{\theta}_{GEll} \circ H^{\alpha}_{W}.$$

We see that this is the family with a proportional logistic parameter θ where $H_0 = H^{\alpha}_{W}$, the Weibull with a power parameter α. This family has also been studied by Jones and Noufaily [34].

The corresponding pairing of the parameter of the family in (3.8) is generated by the reverse composition,

$$H(t|\theta) = H_0 \circ H^{\theta}_{GEll}(t),$$

which we will term the reverse logistic hazards family.

Example 3.22. Let $H_0(t) = H^{\theta_0}_{GEll}(t)$ for some $\theta_0 > 0$. Then we have

$$H(t|\theta_0, \theta) = H^{\theta_0}_{GEll} \circ H^{\theta}_{GEll}(t) = (1 + t)^{\theta\theta_0} - 1 = H^{\theta\theta_0}_{GEll}(t)$$

so that the family of CHs $\left\{H^{\theta}_{GEll} : \theta > 0\right\}$ is closed under compositions

Time Transformation ***(for the reverse logistic hazard parameter families)*** *The time transformation for the reverse logistic hazards family is given by* $T_\theta = H^{1/\theta}_{GEll}(T_0)$.

Change of Baseline *Consider the reparametrisation* $\alpha = \frac{\theta}{\theta_1}$ *for some* $\theta_1 > 0$ *and let* $H_1(t) = H_0 \circ H^{\theta_1}_{GEll}(t)$. *Then*

$$\begin{aligned} H(t|\theta) &= H_0 \circ H^{\theta_1 \alpha}_{GEll}(t) = H_0 \circ H^{\theta_1}_{GEll} \circ H^{\alpha}_{GEll}(t) \\ &= H_1 \circ H^{\alpha}_{GEll}(t), \end{aligned}$$

using the closure property shown in example 3.22. Then H_1 *now plays the role of the baseline in the new parametrisation. Thus any member of the reverse logistic hazard parameter family can be made to be the baseline. It is clear that this is also true for the logistic parameter family.*

Table 3.1 summarises all the information from section 3.1. We see that there are many parameters that can be interpreted as a proportionality parameter for some functions. Taking the logarithm of these models allows us to linearise them. This will be seen again in chapter 4.

The last three pairs of families described in table 3.1 involve a composition of cumulative hazards of the form $H_1^{-1} \circ H^\theta \circ H_1$, where H_1 is some standard distribution and H^θ defines a family such that $\left(H^\theta\right)^{-1} = H^{1/\theta}$. So far, for H_1 we have considered either the standard log-logistic or the standard Gompertz, and for H^θ we have considered the exponential, Weibull and rational families. In principle, we can consider different choices to the ones we have considered so far. For example, one could consider the pair of families defined by $H^\theta_{llWG} = H_{ll} \circ H^\theta_W \circ H_G = \log\left(1 + (e^t - 1)^\theta\right)$. Inspection of the plots of the CH for different values of θ reveal that the corresponding cumulative hazards cross at a fixed point. This makes subsequent analysis and interpretation more difficult, thus we do not consider any other possible families further.

3.1.8 Combination parametric families

In this section we introduce parametric families that not only include a given baseline CH H_A, but also another given CH H_B.

Recalling that the sum of cumulative hazard functions is itself a cumulative hazard from proposition 2.47, we can define a new model.

Definition 3.23. (Linear Combination Family) Given two CH functions H_A and H_B,

$$H^\alpha_{A+B}(t) = \alpha H_A(t) + (1-\alpha) H_B(t), \qquad \alpha \in [0,1] \tag{3.9}$$

is a candidate for generating a parametric family. We call this the *linear combination* family with combination parameter α. We see that $\alpha = 0$ recovers $H_B(t)$ and $\alpha = 1$ recovers $H_A(t)$.

Family	Cumulative Hazard	Time transformation	Parameter interpretation
Scale	$H(t\|\theta) = H_0(\theta t)$	$T_\theta = \frac{1}{\theta}T_0$	Proportional times
Frailty	$H(t\|\theta) = \theta H_0(t)$	$T_\theta = H_0^{-1} \circ H_E^{1/\theta} \circ H_0(T_0)$ $H_0(T_\theta) = \frac{1}{\theta}H_0(T_0)$	Proportional hazards parameter
Power	$H(t\|\theta) = H_0(t^\theta)$	$T_\theta = T_0^{1/\theta}$ $\log(T_\theta) = \frac{1}{\theta}\log(T_0)$	Proportional log times
Hazard Power	$H(t\|\theta) = H_0(t)^\theta$	$T_\theta = H_0^{-1} \circ H_W^{1/\theta} \circ H_0(T_0)$ $H_0(T_\theta) = H_0(T_0)^{1/\theta}$	Proportional log hazards parameter $\log H(t\|\theta) = \theta \log H_0(t)$
Tilt	$H(t\|\theta) = H_{llEG}^\theta \circ H_0(t)$	$T_\theta = H_0^{-1} \circ H_{llEG}^{1/\theta} \circ H_0(T_0)$ $H_G(H_0(T_\theta)) = \frac{1}{\theta}H_G(H_0(T_0))$	Proportional odds parameter $H_G(H(t\|\theta)) = \theta H_G(H_0(t))$
Reverse-tilt	$H(t\|\theta) = H_0 \circ H_{llEG}^\theta(t)$	$T_\theta = H_{llEG}^{1/\theta}(T_0)$ $H_G(T_\theta) = \frac{1}{\theta}H_G(T_0)$	Proportional Gompertz times
Resilience	$H(t\|\theta) = H_{llrG}^\theta \circ H_0(t)$	$T_\theta = H_0^{-1} \circ H_{llrG}^{1/\theta} \circ H_0(T_0)$ $\log\left(1 - e^{-H_0(T_\theta)}\right) = \frac{1}{\theta}\log\left(1 - e^{-H_0(T_0)}\right)$	Proportional log CDF parameter $\log\left(1 - e^{-H(t\|\theta)}\right) = \theta \log\left(1 - e^{-H_0(t)}\right)$
Reverse-resilience	$H(t\|\theta) = H_0 \circ H_{llrG}^\theta(t)$	$T_\theta = H_{llrG}^{1/\theta}(T_0)$ $\log\left(1 - e^{-T_\theta}\right) = \frac{1}{\theta}\log\left(1 - e^{-T_0}\right)$	Proportional transformed times
Proportional logistic	$H(t\|\theta) = H_{GEll}^\theta \circ H_0(t)$	$T_\theta = H_0^{-1} \circ G_{GEll}^{1/\theta} \circ H_0(T_0)$ $H_{ll}(H_0(T_\theta)) = \frac{1}{\theta}H_{ll}(H_0(T_0))$	Proportional logistic parameter $H_{ll}(H(t\|\theta)) = \theta H_{ll}(H_0(t))$
Reverse-proportional logistic	$H(t\|\theta) = H_0 \circ H_{GEll}^\theta(t)$	$T_\theta = H_{GEll}^{1/\theta}(T_0)$ $H_{ll}(T_\theta) = \frac{1}{\theta}H_{ll}(T_0)$	Proportional logistic times

Table 3.1: Collation of all the information for the families mentioned in section 3.1.

Example 3.24. If we let $H_B(t) = H_{ll}(t)$ and let $H_A(t) = H_G(t)$, its inverse, then the linear combination family is given by

$$H^{\alpha}_{ll+G}(t) = (1-\alpha)\log(1+t) + \alpha\left(e^t - 1\right).$$

We can see that this family will be fairly flexible in the sense that it contains a range of CH behaviours from concavity ($\alpha = 0$) to convexity ($\alpha = 1$).

In proposition 2.37 and corollary 2.44 we saw that products of CH functions are CH functions themselves, and that raising a CH function to a power resulted in a CH function also. With this is mind, the following parametric family is proposed.

Definition 3.25. (Geometric Combination Family) Given two CH functions H_B and H_B

$$H^{\alpha}_{A\cdot B}(t) = H_A(t)^{\alpha} H_B(t)^{1-\alpha}, \qquad \alpha \in [0,1], \tag{3.10}$$

is a candidate for generating a parametric model. We will call this the *geometric combination* family with combination parameter α. As with the linear combination we see that letting $\alpha = 0$ recovers H_B and letting $\alpha = 1$ recovers H_A.

Example 3.26. We can again let $H_B(t) = H_{ll}(t)$ and $H_A(t) = H_G(t)$, the geometric combination will be

$$H^{\alpha}_{ll\cdot G}(t) = \log(1+t)^{1-\alpha}\left(e^t - 1\right)^{\alpha}.$$

Now, we recall from proposition 2.10 that composing CH functions results in a CH function, and from proposition 2.20 that the inverse of a CH function is a CH function. Then the following is a candidate for generating a new 1-dimensional family.

Definition 3.27. (Composition Combination Family) Given two CH functions H_B and H_A,

$$H^{\alpha}_{A \vec{\circ} B}(t) = \frac{(H_A \circ H_B^{-1})(\alpha H_B(t))}{\alpha}, \qquad \alpha > 0. \tag{3.11}$$

We will call this the *composition combination* family with combination parameter α. We also note that the parameter α can be larger than 1. We observed above that letting $\alpha = 0$ in (3.9), (3.10) recovers H_B, and letting $\alpha = 1$ recovers H_A. This result is not immediate in the case of (3.11) and further conditions need to be met. If $\alpha = 1$ we recover H_A, and if $h_B(0) = h_A(0) \in (0,\infty)$ then H_B is recovered when $\alpha \to 0$. We will only assume $h_B(0)$, $h_A(0) \in (0,\infty)$, then w.l.o.g. $h_B(0) = h_A(0)$ as the cumulative hazards can be rescaled.

Example 3.28. An important example of a composition combination is when we take $H_B(t) = H^1_E(t) = t$. The cumulative hazard of this family can then be seen to be

$$H^{\alpha}_{A \vec{\circ} E}(t) = H_E^{1/\alpha} \circ H_A \circ H^{\alpha}_E(t) = \frac{1}{\alpha} H_A(\alpha t). \tag{3.12}$$

This family is the basis for the Accelerated Hazards regression model [14] that will be discussed again in chapter 4.

Example 3.29. The Burr Type XII distribution [41] is defined by choosing $H_A = H_{ll}$ in 3.12. Then the family has CH of the form

$$H^{\alpha}_{ll \vec{\circ} E}(t) = \frac{1}{\alpha} H_{ll}(\alpha t) = \frac{1}{\alpha} \log\left(1 + \alpha t\right), \quad \alpha > 0.$$

Example 3.30. If we let $H_B(t) = H_{ll}(t) = \log(1+t)$ and let $H_A(t) = H_{ll}^{-1}(t) = H_G(t)$ in the composition combination (3.11), i.e.

$$H^{\alpha}_{G \vec{\circ} ll}(t) = \frac{(H_G \circ H_G)(\alpha H_{ll}(t))}{\alpha} = \frac{e^{(1+t)^{\alpha}-1} - 1}{\alpha}.$$

This could be written as $H(t|\alpha) = \frac{1}{\alpha} H^{\alpha}_{GGEll}(t)$. If now we let $H_A(t) = H_{ll}(t)$ and $H_B(t) = H_G(t)$, then

$$H^{\alpha}_{ll \vec{\circ} G}(t) = \frac{\log(1 + \log(1 + \alpha(e^t - 1)))}{\alpha}.$$

This can be written as $H(t|\alpha) = \frac{1}{\alpha} H^{\alpha}_{llllEG}(t)$.

Next we will show there exists a common structure in the three combination families described above.

It is easy to see that our set $\mathcal{CH}$ of cumulative hazards endowed with the composition operation forms a group. In this group, the identity element is $H^1_E(t) = t \in \mathcal{CH}$ and the inverse element of $H \in \mathcal{CH}$ is the functional inverse $H^{-1} \in \mathcal{CH}$. Note that the identity element has mean 1 in the units specified for t.

We could extend the set $\mathcal{CH}$ to a larger set of continuously differentiable functions, the details of which we do not give here, and we endow this set with the product operation. Then the group identity element is the function identical to one, and the reciprocal $\frac{1}{H}$ is the inverse element of H.

If we rather endow the set with the addition operation, then we obtain a group with identity element the function identical to zero, and the inverse of H is $-H$.

Then the elements of the linear, geometric and combination families, respectively, can be written as follows

$$H^{\alpha}_{A+B}(t) = [\alpha H_A(t)] + [(-\alpha H_B(t)) + H_B(t)] \tag{3.13}$$

$$H^{\alpha}_{A \cdot B}(t) = [H_A(t)^{\alpha}] \cdot \left[\left(\frac{1}{H_B(t)^{\alpha}} \right) \cdot H_B(t) \right] \tag{3.14}$$

$$H^{\alpha}_{A \vec{\circ} B}(t) = \left[\frac{1}{\alpha} H_A(t) \right] \circ \left[\left(\frac{1}{\alpha} H_B \right)^{-1} \circ H_B(t) \right] \tag{3.15}$$

which reveals a common algebraic structure.

Reparametrisation of combination families

Consider the reparametrisation $\theta = \frac{\alpha}{\alpha_0}$ for some $\alpha_0 \in (0,1)$ and let

$$H_0(t) = \alpha_0 H_A(t) + (1-\alpha_0) H_B(t)$$

for the linear combination family,

$$H_0(t) = H_A(t)^{\alpha_0} H_B(t)^{1-\alpha_0}$$

for the geometric combination family and

$$H_0(t) = \frac{1}{\alpha_0} H_A \circ H_B^{-1}\left(\alpha_0 H_B(t)\right)$$

for the composition combination family. Then for the linear combination family we have that

$$\begin{aligned} H^{\alpha}_{A+B}(t) &= \theta\alpha_0 H_A(t) + (1-\theta\alpha_0) H_B(t) \\ &= \theta\alpha_0 H_A(t) + \theta(1-\alpha_0) H_B(t) + (1-\theta) H_B(t) \\ &= \theta H_0(t) + (1-\theta) H_B(t) \end{aligned}$$

where $\theta \in (0, \frac{1}{\alpha_0})$.

For the geometric combination we have that

$$\begin{aligned} H^{\alpha}_{A \cdot B}(t) &= H_A(t)^{\theta\alpha_0} H_B(t)^{1-\theta\alpha_0} = H_A(t)^{\theta\alpha_0} H_B(t)^{\theta(1-\alpha_0)} H_B(t)^{1-\theta} \\ &= H_0(t)^{\theta} H_B(t)^{1-\theta} \end{aligned}$$

where $\theta \in (0, \frac{1}{\alpha_0})$.

For the composition combination family we have

$$\begin{aligned} H^{\alpha}_{A \vec{\circ} B}(t) &= H_E^{\frac{1}{\theta\alpha_0}} \circ H_A \circ H_B^{-1} \circ \left(H_E^{\theta\alpha_0}\right) \circ H_B(t) \\ &= H_E^{\frac{1}{\theta\alpha_0}} \circ H_A \circ H_B^{-1} \circ \left(H_E^{\theta} \, [\circ] \, H_E^{\alpha_0}\right) \circ H_B(t) \\ &= H_E^{\frac{1}{\theta\alpha_0}} \circ H_A \circ H_B^{-1} \circ \Big(H_E^{\theta} \circ \left[H_B \circ H_A^{-1} \circ H_E^{\alpha_0} \circ H_E^{\frac{1}{\alpha_0}} \circ H_A \circ H_B^{-1}\right] \\ &\quad \circ H_E^{\alpha_0}\Big) \circ H_B(t) \\ &= \left(H_E^{\frac{1}{\theta\alpha_0}}\right) \circ H_A \circ H_B^{-1} \circ H_E^{\theta} \circ H_B \circ H_A^{-1} \circ H_E^{\alpha_0} \circ H_0(t) \\ &= \left(H_E^{\frac{1}{\theta}} \circ H_0 \circ H_0^{-1} \circ H_E^{\frac{1}{\alpha_0}}\right) \circ H_A \circ H_B^{-1} \circ H_E^{\theta} \circ H_B \circ H_A^{-1} \circ H_E^{\alpha_0} \circ H_0(t) \\ &= H_E^{\frac{1}{\theta}} \circ H_0 \circ \bar{H}^{-1} \circ H_E^{\theta} \circ \bar{H}(t), \end{aligned}$$

another composition combination. Here $\bar{H} = H_B \circ H_A^{-1} \circ H_E^{\alpha_0} \circ H_0$ is also a CH.

Thus, by the reparametrisation $\theta = \alpha/\alpha_0$ we can express any of the combination families above in terms of any one of H_A or H_B and an arbitrary member of the family H_0.

Time Transformation ***(for Combination Families)*** *This transformation involves the inverse of $H(t|\alpha)$ according to the formula given in Figure 2.7. Clearly, for both the linear and geometric combination families there is no general analytic inverse. For the composition combination, the inverse is explicit and the time transformation is given by*

$$T_\alpha = H_B^{-1}\left(\frac{1}{\alpha}H_B \circ H_A^{-1}(\alpha H_A(T_A))\right)$$

where $T_A \sim H_A$ and T_α follows the composition combination family distribution. We note that, because α appears twice in the above formula, there will not be any proportionality interpretation as in the previous parametric families in this chapter. This will lead to a non-linear model in the next chapter.

Example 3.31. If we let $H_A = H_0$ with $h_0(0) = 1$ and $H_B = H_G$ in the composition combination family, then the corresponding CH is of the form

$$H^{\alpha}_{0 \overset{\to}{\circ} G}(t) = H_E^{1/\alpha} \circ H_0 \circ H^{\alpha}_{llEG}(t), \quad \alpha > 0.$$

We note that this is almost identical to the cumulative hazard of the reverse tilt parameter family given in (3.5), only with an exponential $H_E^{1/\alpha}$ at the very left. This extra composition does not change the qualitative behaviour but it changes the family members near the limit $\alpha \to 0$. In the reverse tilt families, the limit as $\alpha \to 0$ is degenerate, but the composition combination has a proper CH function in the limit $\alpha \to 0$, the Gompertz CH.

One price to pay for this attribute is that the parameter α loses its proportionality interpretation in table 3.1.

In the same way as before, we propose pairs of parametric families by reversing the order of compositions. We reverse the order of the composition combination to obtain the following family.

Definition 3.32. (Reverse Composition Combination Family) Given two cumulative hazards H_A and H_B where $h_A(0) = h_B(0) \in (0, \infty)$, the parametric family with members

$$H^{\alpha}_{A \overset{\leftarrow}{\circ} B}(t) = H_B\left(\frac{1}{\alpha}H_B^{-1} \circ H_A(\alpha t)\right), \quad \alpha > 0 \tag{3.16}$$

is called the reverse composition combination family. We note that the usual limits hold since when $\alpha = 1$ we recover H_A and when $\alpha = 0$ we recover H_B.

We also note that there is no need to define the reversed family for the linear and geometric combination families since the sum and product are symmetric, unlike the composition.

Example 3.33. If we let $H_A = H_0$ with $h_0(0) = 1$ and $H_B = H_{ll}$ in the reverse composition combination family, then the corresponding CHs are of the form

$$H^{\alpha}_{0 \overleftarrow{\circ} ll}(t)) = H^{1/\alpha}_{llEG} \circ H_0 \circ H^{\alpha}_E(t), \quad \alpha > 0,$$

which is similar to the CHs of the tilt parameter family, except for the innermost composition with an exponential. This extra composition modifies the tilt family members near the limit $\alpha \to 0$, where in this case, the standard log-logistic is included in the limit.

Time Transformation ***(for reverse composition combination)*** *This is given by*

$$T_\alpha = H^{1/\alpha}_E \circ H^{-1}_A \circ H_B \circ H^{\alpha}_E \circ H^{-1}_B \circ H_A(T_A)$$

for $T_A \sim H_A$ and T_α from the reverse composition combination.

In appendix A the behaviours of the likelihoods of the combination families are explored.

3.1.9 Linear-Composition combination families

The following proposed families will be of particular interest in section 6.4.3 where their properties will be further discussed. The following families will combine the techniques discussed earlier in building families.

First of all we propose a general model which combines summing, composing and inverting CHs to create new ones. The CH function of this model follows.

Definition 3.34. (Linear-Composition Combination) Given two CH functions H_A and H_B and a parametrised CH function H_α,

$$H(t|\alpha) = H^{-1}_\alpha \left(H_\alpha \left(H_A(t)\right) + H_\alpha \left(H_B(t)\right)\right), \tag{3.17}$$

is called the linear-composition combination with parameter α. We would like that some particular value of $\alpha = \tilde{\alpha}$ recovers the case $H(t|\tilde{\alpha}) = H_A(t) + H_B(t)$. Later in chapter 6, this combination will be used to model multiple times-to-event. In this context, it will be useful to recover the sum of $H_A(t)$ and $H_B(t)$ since this will correspond to the case of independent random variables.

Example 3.35. Consider the linear-composition combination (3.17) where $H_\alpha(t) = t^\alpha$. This incorporates summing and powering CH functions. The corresponding CH is given by

$$H(t|\alpha) = \left((H_A(t))^\alpha + (H_B(t))^\alpha\right)^{1/\alpha}, \tag{3.18}$$

where we can see that $H(t|1) = H_A(t) + H_B(t)$, so that $\tilde{\alpha} = 1$.

Example 3.36. Let $H_\alpha(t) = H^\alpha_{llEG}(t)$, then the corresponding linear-composition combination has CH

$$H(t|\alpha) = H^{1/\alpha}_{llEG}\left(H^\alpha_{llEG}(H_A(t)) + H^\alpha_{llEG}(H_B(t))\right).$$

Since $H^1_{llEG}(t) = H^1_E$ it is clear that

$$H(t|1) = H_A(t) + H_B(t).$$

Similar families can be constructed in the same way using H^α_{llrG} or H^α_{GEll} since $H^1_{llrG} = H^1_{GEll} = H^1_E$.

Example 3.37. Consider the CH family (3.12) $H_\alpha(t) = \frac{1}{\alpha}\tilde{H}(\alpha t)$, where $\tilde{h}(0) \in (0, \infty)$. Then the corresponding linear-composition combination has CH of the form

$$H(t|\alpha) = \frac{1}{\alpha}\tilde{H}^{-1}\left(\tilde{H}(\alpha H_A(t)) + \tilde{H}(\alpha H_B(t))\right).$$

It is easy to show that $H(t|0) = H_A(t) + H_B(t)$ using L'Hospital's rule, so that $\tilde{\alpha} = 0$.

All the previous combination families can be seen as flexible, since H_A and H_B can be very different types of cumulative hazards. For example, H_A could be concave, e.g. log-logistic, and H_B could be convex, e.g. Weibull with power parameter $\theta > 1$, or the Gompertz. Then, the parameter α would determine how convex or how concave the actual family, or perhaps baseline family, should be. This means that although our approach would be fully parametric, it would not be too restrictive and the parameter α would determine how flexible this family would actually be.

3.1.10 Equivariance of combination families

In this subsection we explore the equivariance properties of combination families to the addition of different types of parameters such as scale, frailty, etc. We will let $C^\alpha(H_A, H_B)$ denote a general combination model. We will also let

$$\begin{aligned}
C^\alpha_+(H_A, H_B) &= H^\alpha_{A+B}(t) = \alpha H_A(t) + (1-\alpha)H_B(t) \\
C^\alpha_\cdot(H_A, H_B) &= H^\alpha_{A\cdot B}(t) = H_A(t)^\alpha H_B(t)^{1-\alpha} \\
C^\alpha_{\overset{\rightarrow}{\circ}}(H_A, H_B) &= H^\alpha_{A\overset{\rightarrow}{\circ}B}(t) = \frac{(H_A \circ H_B^{-1})(\alpha H_B(t))}{\alpha} \\
C^\alpha_{\overset{\leftarrow}{\circ}}(H_A, H_B) &= H^\alpha_{A\overset{\leftarrow}{\circ}B}(t) = H_B\left(\frac{1}{\alpha}H_B^{-1} \circ H_A(\alpha t)\right).
\end{aligned}$$

Proposition 3.38. ***(Equivariance to scale and frailty parameters)*** *All four combination models are equivariant to the addition of a scale or a frailty parameter. That* *is*

$$C^{\alpha}\left(H_A \circ H_E^{\theta}(t), H_B \circ H_E^{\theta}(t)\right) = C^{\alpha}\left(H_A, H_B\right) \circ H_E^{\theta}(t), \quad \theta > 0,$$
$$C^{\alpha}\left(H_E^{\theta} \circ H_A(t), H_E^{\theta} \circ H_B(t)\right) = H_E^{\theta} \circ C^{\alpha}\left(H_A(t), H_B(t)\right), \quad \theta > 0,$$

where C^{α} is the function that takes two CHs and returns the combined CH.

Proof. The proof for the composition and reverse composition combination are the least intuitive so will be given here. The proofs for the other combinations are much more straightforward so won't be included.

First, let $C^{\alpha}_{\vec{\circ}}$ denote the composition combination. Suppose H_A and H_B both have the same scale parameter θ, then $H_A(t|\theta) = H_A(\theta t)$ and $H_B(t|\theta) = H_B(\theta t)$ and $H_A^{-1}(t|\theta) = \frac{1}{\theta} H_A^{-1}(t)$. Hence

$$\begin{aligned} C^{\alpha}_{\vec{\circ}}\left(H_A \circ H_E^{\theta}(t), H_B \circ H_E^{\theta}(t)\right) &= \frac{1}{\alpha} H_A(\theta \frac{1}{\theta} H_B^{-1}(\alpha H_B(\theta t))) \\ &= \frac{1}{\alpha} H_A(H_B^{-1}(\alpha H_B(\theta t))) = C^{\alpha}_{\vec{\circ}}\left(H_A, H_B\right) \circ H_E^{\theta}(t). \end{aligned}$$

Now suppose H_A and H_B both have the same frailty parameter θ, then $H_A(t|\theta) = \theta H_A(t)$ and $H_B(t|\theta) = \theta H_B(t)$ and $H_A^{-1}(t|\theta) = H_A^{-1}(\frac{1}{\theta} t)$. Hence

$$\begin{aligned} C^{\alpha}_{\vec{\circ}}\left(H_E^{\theta} \circ H_A(t), H_E^{\theta} \circ H_B(t)\right) &= \frac{\theta}{\alpha} H_A(H_B^{-1}(\frac{\alpha}{\theta} \theta H_B(t))) \\ &= \frac{\theta}{\alpha} H_A(H_B^{-1}(\alpha H_B(t))) = H_E^{\theta} \circ C^{\alpha}_{\vec{\circ}}\left(H_A(t), H_B(t)\right). \end{aligned}$$

Now let $C^{\alpha}_{\overleftarrow{\circ}}$ denote the reverse combination. Suppose H_A and H_B both have the same scale parameter. Then

$$\begin{aligned} C^{\alpha}_{\overleftarrow{\circ}}\left(H_A \circ H_E^{\theta}(t), H_B \circ H_E^{\theta}(t)\right) &= H_B\left(\frac{\theta}{\alpha} \frac{1}{\theta} H_B^{-1} \circ H_A(\theta \alpha t)\right) \\ &= H_B\left(\frac{1}{\alpha} H_B^{-1} \circ H_A(\theta \alpha t)\right) = C^{\alpha}_{\overleftarrow{\circ}}\left(H_A, H_B\right) \circ H_E^{\theta}(t). \end{aligned}$$

Now suppose H_A and H_B both have a frailty parameter. Then

$$\begin{aligned} C^{\alpha}_{\overleftarrow{\circ}}\left(H_E^{\theta} \circ H_A(t), H_E^{\theta} \circ H_B(t)\right) &= \theta H_B\left(\frac{1}{\alpha} H_B^{-1}\left(\frac{1}{\theta} \theta H_A(\alpha t)\right)\right) \\ &= \theta H_B\left(\frac{1}{\alpha} H_B^{-1}\left(H_A(\alpha t)\right)\right) \\ &= H_E^{\theta} \circ C^{\alpha}_{\overleftarrow{\circ}}\left(H_A(t), H_B(t)\right). \end{aligned}$$

□

Proposition 3.39. ***(Equivariance to tilt, resilience or proportional logistic hazard parameters)*** *The reversed composition combination families are equivariant to the addition of a tilt, resilience or proportional logistic hazards parameter. That is*

$$\begin{aligned}
C^{\alpha}\left(H_{llEG}^{\theta} \circ H_A(t), H_{llEG}^{\theta} \circ H_B(t)\right) &= H_{llEG}^{\theta} \circ C^{\alpha}\left(H_A(t), H_B(t)\right), \quad \theta > 0, \\
C^{\alpha}\left(H_{llrG}^{\theta} \circ H_A(t), H_{llrG}^{\theta} \circ H_B(t)\right) &= H_{llrG}^{\theta} \circ C^{\alpha}\left(H_A(t), H_B(t)\right), \quad \theta > 0, \\
C^{\alpha}\left(H_{GEll}^{\theta} \circ H_A(t), H_{GEll}^{\theta} \circ H_B(t)\right) &= H_{GEll}^{\theta} \circ C^{\alpha}\left(H_A(t), H_B(t)\right), \quad \theta > 0.
\end{aligned}$$

Proof. This proposition will be proved for the tilt parameter as the others are similar. Recall the reverse composition can be expressed as

$$C_{\overset{\leftarrow}{\circ}}^{\alpha}\left(H_A(t), H_B(t)\right) = H_B \circ H_E^{1/\alpha} \circ H_B^{-1} \circ H_A \circ H_E^{\alpha}(t).$$

Also recall that $\left(H_{llEG}^{\theta}\right)^{-1} = H_{llEG}^{1/\theta}$. Then,

$$\begin{aligned}
C_{\overset{\leftarrow}{\circ}}^{\alpha}&\left(H_{llEG}^{\theta} \circ H_A(t), H_{llEG}^{\theta} \circ H_B(t)\right) \\
&= H_{llEG}^{\theta} \circ H_B \circ H_E^{1/\alpha} \circ H_B^{-1} \circ H_{llEG}^{1/\theta} \circ H_{llEG}^{\theta} \circ H_A \circ H_E^{\alpha}(t) \\
&= H_{llEG}^{\theta} \circ C_{\overset{\leftarrow}{\circ}}^{\alpha}\left(H_A(t), H_B(t)\right).
\end{aligned}$$

The other proofs are similar since $\left(H_{llrG}^{\theta}\right)^{-1} = H_{llrG}^{1/\theta}$ and $\left(H_{GEll}^{\theta}\right)^{-1} = H_{GEll}^{1/\theta}$. □

Proposition 3.40. ***(Equivariance to reversed tilt, reversed resilience and reversed proportional logistic hazards parameters)*** *The linear, geometric and composition families are equivariant to the addition of a power, reversed tilt, reversed resilience or a reversed proportional logistic hazard parameter. That is*

$$\begin{aligned}
C^{\alpha}\left(H_A \circ H_W^{\theta}(t), H_B \circ H_W^{\theta}(t)\right) &= C^{\alpha}\left(H_A, H_B\right) \circ H_W^{\theta}(t), \quad \theta > 0, \\
C^{\alpha}\left(H_A \circ H_{llEG}^{\theta}(t), H_B \circ H_{llEG}^{\theta}(t)\right) &= C^{\alpha}\left(H_A, H_B\right) \circ H_{llEG}^{\theta}(t), \quad \theta > 0, \\
C^{\alpha}\left(H_A \circ H_{llrG}^{\theta}(t), H_B \circ H_{llrG}^{\theta}(t)\right) &= C^{\alpha}\left(H_A, H_B\right) \circ H_{llrG}^{\theta}(t), \quad \theta > 0, \\
C^{\alpha}\left(H_A \circ H_{GEll}^{\theta}(t), H_B \circ H_{GEll}^{\theta}(t)\right) &= C^{\alpha}\left(H_A, H_B\right) \circ H_{GEll}^{\theta}(t), \quad \theta > 0.
\end{aligned}$$

Proof. Again, the proof will only be given for the composition combination and for the reversed tilt parameter. Recall the composition combination is $C_{\overset{\rightarrow}{\circ}}^{\alpha}\left(H_A(t), H_B(t)\right) = H_E^{1/\alpha} \circ H_A \circ H_B^{-1} \circ H_E^{\alpha} \circ H_B(t)$. Then

$$\begin{aligned}
C_{\overset{\rightarrow}{\circ}}^{\alpha}&\left(H_A \circ H_{llEG}^{\theta}(t), H_B \circ H_{llEG}^{\theta}(t)\right) \\
&= H_E^{1/\alpha} \circ H_A \circ H_{llEG}^{\theta} \circ H_{llEG}^{1/\theta} \circ H_B^{-1} \circ H_E^{\alpha} \circ H_B \circ H_{llEG}^{\theta}(t) \\
&= H_E^{1/\alpha} \circ H_A \circ H_B^{-1} \circ H_E^{\alpha} \circ H_B \circ H_{llEG}^{\theta}(t) \\
&= C_{\overset{\rightarrow}{\circ}}^{\alpha}\left(H_A, H_B\right) \circ H_{llEG}^{\theta}(t).
\end{aligned}$$

The other proofs are similar since $\left(H_W^{\theta}\right)^{-1} = H_W^{1/\theta}$, $\left(H_{llrG}^{\theta}\right)^{-1} = H_{llrG}^{1/\theta}$ and $\left(H_{GEll}^{\theta}\right)^{-1} = H_{GEll}^{1/\theta}$. □

Proposition 3.41. ***(Equivariance to hazard power parameters)*** *The geometric and reversed composition families are equivariant to the addition of a hazard power parameter. That is*

$$C^{\alpha}\left(H_W^{\theta} \circ H_A(t), H_W^{\theta} \circ H_B(t)\right) = H_W^{\theta} \circ C^{\alpha}\left(H_A(t), H_B(t)\right), \quad \theta > 0.$$

Proof. The proof for the geometric is simple, thus only the proof for the reversed composition, $C^{\alpha}_{\overleftarrow{\circ}}\left(H_A(t), H_B(t)\right) = H_B \circ H_E^{1/\alpha} \circ H_B^{-1} \circ H_A \circ H_E^{\alpha}(t)$, will be given.

$$\begin{aligned}
&C^{\alpha}_{\overleftarrow{\circ}}\left(H_W^{\theta} \circ H_A(t), H_W^{\theta} \circ H_B(t)\right) \\
&\qquad = H_W^{\theta} \circ H_B \circ H_E^{1/\alpha} \circ H_B^{-1} \circ H_W^{1/\theta} \circ H_W^{\theta} \circ H_A \circ H_E^{\alpha}(t) \\
&\qquad = H_W^{\theta} \circ C^{\alpha}_{\overleftarrow{\circ}}\left(H_A(t), H_A(t)\right).
\end{aligned}$$

□

Proposition 3.42. *Recall the Accelerated Hazards model parameter as in* (3.12)*,*

$$H^{\theta}_{A\vec{\circ}E}(t) = \frac{1}{\theta} H_A(\theta t).$$

The four combination families are equivariant to the accelerated hazards parameter, that is

$$C^{\alpha}\left(H_E^{1/\theta} \circ H_A \circ H_E^{\theta}(t), H_E^{1/\theta} \circ H_B \circ H_E^{\theta}(t)\right) = H_E^{1/\theta} \circ C^{\alpha}\left(H_A, H_B\right) \circ H_E^{\theta}(t).$$

Proof. This will only be shown for the reversed composition.

$$\begin{aligned}
&C^{\alpha}\left(H_E^{1/\theta} \circ H_A \circ H_E^{\theta}(t), H_E^{1/\theta} \circ H_B \circ H_E^{\theta}(t)\right) \\
&\quad = H_E^{1/\theta} \circ H_B \circ H_E^{\theta} \circ H_E^{1/\alpha} \circ H_E^{1/\theta} \circ H_B^{-1} \circ H_E^{\theta} \circ H_E^{1/\theta} \circ H_A \circ H_E^{\theta} \circ H_E^{\alpha}(t) \\
&\quad = H_E^{1/\theta} \circ H_B \circ H_E^{\theta} \circ H_E^{1/\alpha} \circ H_E^{1/\theta} \circ H_B^{-1} \circ H_A \circ H_E^{\theta} \circ H_E^{\alpha}(t).
\end{aligned}$$

Then using $H_E^{\alpha} \circ H_E^{\theta}(t) = H_E^{\theta} \circ H_E^{\alpha}(t)$, we have the above is equal to

$$\begin{aligned}
&= H_E^{1/\theta} \circ H_B \circ H_E^{\theta} \circ H_E^{1/\alpha} \circ H_E^{1/\theta} \circ H_B^{-1} \circ H_A \circ H_E^{\theta} \circ H_E^{\alpha}(t) \\
&= H_E^{1/\theta} \circ H_B \circ H_E^{1/\alpha} \circ H_B^{-1} \circ H_A \circ H_E^{\alpha} \circ H_E^{\theta}(t) \\
&= H_E^{1/\theta} \circ C^{\alpha}\left(H_A, H_B\right) \circ H_E^{\theta}(t).
\end{aligned}$$

□

3.2 Multi-parameter families

There are many considerations we can take into account when combining two or more parameters in a family of cumulative hazards. A sensible and ideal consideration is that each parameter plays a separate and different role. This is a modelling approach to avoid any possible confounding between parameters. At the same time, flexibility of the possible behaviours of the family of cumulative hazards is important.

There are two key criteria for cumulative hazards that we consider,

1. behaviours such as convexity, concavity, linearity or some combinations of those,

2. short term, $t \to 0$, and long term, $t \to \infty$, behaviours.

We will combine these two criteria in order to achieve flexibility.

When considering the first criterion we will explore combining archetypal families whose CH functions are convex, concave or linear. The typical convex families include the Gompertz and the Weibull, where the power parameter is greater than one. Concave families include the log-logistic and the Weibull for power parameters less than one. The only example of a linear family is the Exponential. This is related to combining behaviours such as Increasing hazard rate (IFR), decreasing hazard rate (DFR) and the exponential hazard.

In the previous section we have explored the long and short term behaviours of families. We have seen that short term behaviours include linear, θt, and Weibull, t^θ for any $\theta > 0$, and long term behaviours include Weibull, t^θ for $\theta > 0$ as well as e^t and $\log(t)$. We expect that the combinations of these types of behaviours could be quite fruitful and would provide a flexible family.

3.2.1 Combining one-dimensional families

We can start by combining scale and power parameters to create a distribution with two parameters. The order in which we include the parameters now plays a minor role. For example, for some baseline distribution with CH H_0 we could have

$$H(t|\theta,\alpha) = H_0 \circ H_E^\theta \circ H_W^\alpha(t) = H_0 \circ H_{EW}^{\theta,\alpha}(t) = H_0(\theta t^\alpha)$$

or

$$H(t|\theta,\alpha) = H_0 \circ H_w^\alpha \circ H_E^\theta(t) = H_0 \circ H_{WE}^{\alpha,\theta}(t) = H_0\left((\theta t)^\alpha\right).$$

From this it is easy to see these correspond to different parametrisations of the same two-dimensional family of distributions.

We will see an example of a distribution with scale and power parameters, there are of course many other examples.

Example 3.43. Let $H_0(t) = H_E^1(t)$. We have that in the first parametrisation, $H(t|\theta,\alpha) = \theta t^\alpha$ and θ acts as a frailty parameter. In the second parametrisation, $H(t|\theta,\alpha) = (\theta t)^\alpha$ and θ does act as a scale parameter.

Example 3.44. Let $H_0(t) = H_{ll}(t)$. The first parametrisation gives that $H(t|\theta,\alpha) = \log(1+\theta t^\alpha)$ and the second yields $H(t|\theta,\alpha) = \log(1+\theta^\alpha t^\alpha)$. Using the first parametrisation, in the short term $H(t|\theta,\alpha) \sim \theta t^\alpha$, $\forall \alpha > 0$ so that it behaves like a Weibull with frailty parameter θ.

It is well known that scale and power parameters are a simple reparametrisation of location and scale parameters in the log time scale. This clearly shows scale and power are, in principle, well identified and play two different roles.

We can now look at the reverse compositions of the above to see how this effects the baseline distribution. The reverse compositions of the above results in the addition of frailty and hazard power parameters. We can consider the parametrisations

$$H(t|\theta,\alpha) = H_E^\theta \circ H_W^\alpha \circ H_0(t) = H_{EW}^{\theta,\alpha} \circ H_0(t) = \theta H_0(t)^\alpha \tag{3.19}$$

or

$$H(t|\theta,\alpha) = H_W^\alpha \circ H_E^\theta \circ H_0(t) = H_{WE}^{\alpha,\theta} \circ H_0(t) = \theta^\alpha H_0(t)^\alpha.$$

An example of this combination of parameters can be seen next.

Example 3.45. Let $H_0(t) = H_G(t)$, a Gompertz CH, then the first parametrisation 3.19 for frailty and hazard power combination is

$$H(t|\theta,\alpha) = H_{EW}^{\theta,\alpha} \circ H_G(t) = \theta\left(e^t - 1\right)^\alpha.$$

We can understand the role of the parameters by looking at the asymptotic behaviour of this CH as $t \to 0$ and $t \to \infty$. In the short term we find that $H(t|\theta,\alpha)$ behaves like θt^α, like a Weibull with frailty θ and power parameter α. In the long term we find that $H(t|\theta,\alpha)$ behaves like $\theta e^{\alpha t}$, so that it behaves like the Gompertz with frailty θ and scale α.

We see that in the long term, this family will be convex, regardless of the values of the parameters. However, in the short term, the shape of this family is determined by the value of α. If $\alpha > 1$ it will be convex, if $\alpha < 1$ it will be concave and it will be linear for $\alpha = 1$. Thus we see that the flexibility of this family is achieved by the parameter α which controls the type of short term behaviour combined with long term convexity.

Example 3.46. In example 3.21 we saw the combination of the proportional logistic parameter with a Weibull baseline with a power parameter, i.e. $H(t|\theta,\alpha) = H_{GEll}^\theta \circ H_W^\alpha$ the Generalised Power Weibull. We can understand the role of each parameter again by looking at the short and long term behaviour. In the short term $H(t|\theta, \alpha)$ behaves like a Weibull θt^α with frailty θ and power α. In the long term $H(t|\theta, \alpha)$ acts like a different Weibull, namely $t^{\theta\alpha}$, with power parameter $\theta\alpha$. A simple reparametrisation, where we introduce a new parameter $\lambda = \theta\alpha$, simplifies the interpretation, specifically, as seen in teh work of Jones and Noufaily [34]

$$H(t|\lambda, \alpha) = H_{GEll}^{\lambda/\alpha} \circ H_W^\alpha(t).$$

Now the family behaves like a Weibull with power α and frailty $\frac{\lambda}{\alpha}$ in the short term and like a Weibull with power λ in the long term.

This family is very flexible as both the short and long term behaviours can include concavity, convexity and linearity depending on the values of the parameters α and λ.

Example 3.47. A similar family to the previous example arises as the Exponentiated Weibull, for example Jones and Noufaily [34]. This family combines the resilience with Weibull baseline, namely, $H_{llrG}^\theta \circ H_W^\lambda$. If $\theta = \alpha/\lambda$ then in the short term the family behaves like t^α, so that α is a power parameter. In the long term this family behaves like t^λ so that λ is now the power parameter. Similar to the previous example, this family is very flexible as it can combine concavity, convexity or linearity in both the short term and the long term.

Example 3.48. Yang and Prentice [76] propose a model where

$$H(t|\lambda, \theta) = H_E^\lambda \circ H_{llEG}^{\theta/\lambda} \circ H_0(t).$$

We first note that if $\lambda = 1$ then θ is a tilt parameter and if $\theta = \lambda$ then λ is a frailty parameter. In the short term the CH function acts like $\theta H_0(t)$, so that θ is a frailty parameter. In the long term it behaves like $\lambda H_0(t)$, so that the frailty parameter is now λ. Thus it can be viewed as one Exponential distribution in the short term and another in the long term. The parameter θ is thus called the short term hazard ratio and the λ is the long term hazard ratio. In the particular case of $H_0(t) = t$, a standard exponential, we can see that this family smoothly combines two potentially different linear behaviours in the short term compared to the long term. Note that Yang and Prentice propose a proportional odds model with a frailty parameter.

Now we look at an equivariance property of the Yang and Prentice family.

Corollary 3.49. *The reverse composition is equivariant to the combination of parameters described by the Yang and Prentice model in example 3.48. That is,*

$$C_{\overleftarrow{\circ}}^{\alpha}\left(H_E^\lambda \circ H_{llEG}^{\theta/\lambda} \circ H_A(t), H_E^\lambda \circ H_{llEG}^{\theta/\lambda} \circ H_B(t)\right) = H_E^\lambda \circ H_{llEG}^{\theta/\lambda} \circ C_{\overleftarrow{\circ}}^{\alpha}\left(H_A(t), H_B(t)\right).$$

Proof. The Yang and Prentice model incorporates a scale and a tilt parameter. We see that the only combination that is equivariant to both parameters, as seen in propositions 3.38 and 3.39, is the reverse composition. □

Example 3.50. Chen and Jewell [13] propose a model where

$$H(t|\lambda,\theta) = H_E^{\lambda/\theta} \circ H_0 \circ H_E^{\theta}(t).$$

This model combines frailty and scale parameters. Note that, for equivariance, we could have that H_0 is any of our four combination families since these are all equivariant to both scale and frailty, as discussed in Proposition 3.38.

There are of course many ways in which parameters can be combined in order to create new flexible models. Our framework provides a principled way for how we might actually do this and allows us to provide some interpretation of the parameters. In the next section we discuss how to create other flexible parametric models.

3.2.2 Multi dimensional combination families

Another way of combining one-dimensional families is to turn the double appearance of the combination parameters into two different parameters.

First we aim to generalise the linear combination in equation (3.9) to include an extra parameter. This results in the family given by

$$H(t|\alpha,\beta) = \alpha H_A(t) + \beta H_B(t), \tag{3.20}$$

for $\alpha, \beta > 0$.

A further generalisation of the model in (3.20), would be to add some other function of the CH functions H_A and H_B. This type of generalisation can be seen in the literature. Consider the following multi-dimensional family inspired by the bivariate models proposed by Murthy, Xie and Jiang [49],

$$H(t|\alpha,\beta,\nu) = \alpha H_A(t) + \beta H_B(t) + \nu\phi(H_A(t), H_B(t)) \tag{3.21}$$

where $\phi(H_A(t), H_B(t))$ is a CH function and $\alpha, \beta, \nu > 0$. An example of $\phi(\cdot,\cdot)$ is $\phi(H_A(t), H_B(t)) = H_A(t)H_B(t)$. Note that the 2-dimensional parametric family in (3.20) can be recovered from (3.21) where we let $\nu = 0$.

Another way to generalise the families we proposed in the earlier section is to have H_A and H_B be parametrised, i.e. we have $H_A(t|\theta)$ and $H_B(t|\gamma)$. Then our model, the linear-composition combination (3.17), can be extended to

$$H(t|\alpha,\theta,\gamma) = H_\alpha^{-1}\left(H_\alpha(H_A(t|\theta)) + H_\alpha(H_B(t|\gamma))\right). \tag{3.22}$$

This model will be used later in chapter 6 where we model the failure and censoring times in a joint model.

Consider the following equation which aims to generalise the composition combination family in equation (3.11) and the reverse composition in (3.16),

$$H(t|\alpha,\beta) = \frac{1}{\alpha} H_A \circ H_B^{-1}(\beta H_B(t)) \tag{3.23}$$

$$H(t|\alpha,\beta) = H_B\left(\frac{1}{\alpha} H_B^{-1}\left(H_A(\beta t)\right)\right) \tag{3.24}$$

where $\alpha > 0$, $\beta > 0$.

The following examples show the flexibility of this family.

Example 3.51. If we let $H_A = H_0$ and $H_B = H_G$ in (3.23), then we obtain

$$H(t|\alpha,\beta) = H_E^{1/\alpha} \circ H_0 \circ H_{llEG}^{\beta}(t),$$

a family that when $\alpha = 1$ we recover the reverse tilt family, and when $\beta = 1$ we recover the frailty parameter family. With respect to equivariance, this family could have the linear, geometric or composition combinations as H_0 since these are all equivariant to the addition of both the reverse tilt and the frailty parameters. Conversely if we let $H_B = H_{ll}$, then

$$H(t|\alpha,\beta) = H_E^{1/\alpha} \circ H_0 \circ H_{GEll}^{\beta}(t),$$

then when $\alpha = 1$ the reverse logistic hazard parameter family is recovered and when $\beta = 1$ the frailty parameter family is recovered. Again, we could use the linear, geometric and composition combinations as H_0.

If we let $H_A = H_0$ and $H_B = H_{ll}$ in (3.24), then,

$$H(t|\alpha,\beta) = H_{llEG}^{1/\alpha} \circ H_0 \circ H_E^{\beta}(t)$$

then when $\beta = 1$ we recover the tilt parameter family, and when $\alpha = 1$ we recover a scale parameter family. A reverse composition combination would be useful as a choice of H_0 since this family is equivariant to the addition of both the tilt and scale parameters. Conversely, if we let $H_B = H_G$ then,

$$H(t|\alpha,\beta) = H_{GEll}^{1/\alpha} \circ H_0 \circ H_E^{\beta}(t)$$

and when $\alpha = 1$ we again recover a scale parameter family and when $\beta = 1$ we recover a logistic hazard parameter family. Again, a good choice of H_0 would be the reverse composition combination.

Example 3.52. If we let $H_B(t) = H_E^1(t)$ and $H_A(t) = H_0(t)$, then the two families in (3.23) and (3.24) are identical, with CH of the form

$$H(t|\alpha,\beta) = \frac{1}{\alpha} H_0(\beta t).$$

This is the Chen and Jewell [13] family in example 3.52 which we will revisit in the next chapter. Thus, when $\alpha = 1$ we recover the scale parameter family. When $\beta = 1$ we recover the frailty parameter family. Finally, if $\alpha = \beta$, we recover the composition combination family with $H_B(t) = t$, which is the Accelerated hazards CH of example 3.28

An alternative way of constructing a two dimensional family is as follows

$$H(t|\alpha, \beta) = \beta H_B \left(\frac{1}{\alpha} H_B^{-1}(H_A(t)) \right). \tag{3.25}$$

Example 3.53. If we let $H_B = H_{ll}$ and $H_A = H_0$ in (3.25) then when $\alpha = 1$ we recover a frailty parameter family with baseline H_A, and when $\beta = 1$ we recover a tilt parameter family with the same baseline H_A.

Furthermore, if $\alpha = \beta$, we recover a composition combination with the property that when $\alpha = 1$ we obtain H_A and when $\alpha = 0$ we obtain $H_B^{-1}(H_A(t))$. This requires the condition that $h_B(\infty) \in (0, \infty)$ so that we can take $h_B(\infty) = 1$. Examples of distributions that satisfy this condition are the Exponential, the H_{llEG}^{θ} family, and any other such that $H(t)$ that acts like t as $t \to \infty$.

The previous two-dimensional combination families have the characteristic that each parameter played a different role, i.e. frailty, scale, tilt, etc., when the other one was fixed. A different situation arises when a combination parameter controls the role of the other parameter, i.e. if it is a scale, frailty or another parameter.

Example 3.54. Consider the two-dimensional parameter family of CHs defined by

$$H(t|\alpha, \beta) = \frac{1}{\alpha} H_{llEG}^{\beta}(\alpha H_0(t)) = \beta C_{\circlearrowleft}^{\alpha} \left(\frac{1}{\beta} H_{llEG}^{\beta}, H_E^{1} \right) \circ H_0(t), \tag{3.26}$$

where we note the inclusion of the denominator β inside the combination. This is needed since $h_{llEG}^{\beta}(0) = \beta$, so $\frac{1}{\beta} h_{llEG}^{\beta}(0) = 1$ as required by the definition of the composition combination. Then we clearly have that when $\alpha = 0$ we recover $H(t|\alpha, \beta) = \beta H_0(t)$ meaning β is a frailty parameter, and when $\alpha = 1$, $H(t|\alpha, \beta) = H_{llEG}^{\beta} \circ H_0(t)$ meaning β is a tilt parameter. Thus, β has a different role depending on the value of α. We contrast the above family with that of example 3.53 where the same parameter has a distinct role and cannot be both a frailty or tilt. Nevertheless, both families might be useful to discriminate between a frailty and a tilt parameter family.

The family (3.26) was introduced by Royston and Parmar [57], although they only consider two cases when $\alpha = 0$ or $\alpha = 1$ and did not let α be a free parameter.

This family will only be equivariant for the reversed composition families for H_0.

The idea in the previous example can be used to define many other two-dimensional families. For example

$$H(t|\alpha,\beta) = \frac{1}{\alpha} H^{\beta}_{GEll}(\alpha H_0(t)) = \beta C_{\overrightarrow{\circ}}^{\alpha} \left(\frac{1}{\beta} H^{\beta}_{GEll}, H^{1}_{E}\right) \circ H_0(t),$$

so that when $\alpha = 0$ we get a frailty family with parameter β and when $\alpha = 1$ we get a proportional logistic hazard family with parameter β. This family will be equivariant only to reversed composition combination for H_0.

Analogously, we can define

$$H(t|\alpha,\beta) = H_0\left(\beta C_{\overrightarrow{\circ}}^{\alpha} \left(\frac{1}{\beta} H^{\beta}_{llEG}(t), H^{1}_{E}(t)\right)\right)$$

so that when $\alpha = 0$ we get a scale family with parameter β and when $\alpha = 1$ we get a reverse tilt family with parameter β. This family will be equivariant to linear, geometric and composition combinations for H_0.

3.3 Frailty mixtures

We will consider survival models which are mixtures, more specifically where we can include heterogeneity by mixing over an unobserved random variable, usually called frailty. In order to detail the form of these distributions, we will need some preliminary mathematical results. These results can be seen in Schilling, Song and Vondracek [60] in more detail.

3.3.1 Preliminary Results

Throughout this section, the positive random variable U, which might be discrete or continuous, will denote the unobserved frailty which represents unobserved heterogeneity in a population. For simplicity in the expressions, we will use cumulative distribution functions and Stieltjes integrals. We first define the Laplace-Stieltjes transform of the distribution of the frailty.

Definition 3.55. (Laplace-Stieltjes transform) The Laplace-Stieltjes transform of a cumulative distribution function F on $[0,\infty)$ is defined by

$$\mathcal{L}_F(t) = \int_0^\infty e^{-tu}\, dF(u) \tag{3.27}$$

for $t > 0$.

Example 3.56. Let U be a frailty random variable following standard exponential, i.e. $F_E(t) = 1 - e^{-t}$. The Laplace-Stieltjes transform of $F_E(t)$ is

$$\begin{aligned}\mathcal{L}_E(t) &= \int_0^\infty e^{-tu}\, dF_E(u) = \int_0^\infty e^{-tu} e^{-u} du \\ &= \frac{1}{1+t}, \quad t > 0.\end{aligned}$$

Example 3.57. Let U be a positive frailty random variable following a Gamma distribution with shape parameter $1/\alpha$ and scale parameter α. This ensures the mean is equal to one, as is usual for frailty distributions. The frailty variance is equal to α. The corresponding Laplace-Stieltjes transform is given by

$$\begin{aligned}\mathcal{L}_\Gamma^\alpha(t) &= \frac{1}{\alpha^{1/\alpha}\Gamma(\frac{1}{\alpha})} \int_0^\infty u^{\frac{1}{\alpha}-1} e^{-u(\frac{1}{\alpha}+t)} du \\ &= \left(\frac{1}{1+\alpha t}\right)^{1/\alpha}, \quad t > 0.\end{aligned}$$

Example 3.58. Let U be a discrete frailty random variable taking values $\{1, 2, 3, ...\}$, following a zero truncated Geometric distribution with probability mass function

$$dF(u) = \frac{\alpha^{u-1}}{(1+\alpha)^u}, \quad \alpha > 0.$$

The untruncated distribution counts the numbers of successes until one failure occurs. The probability of success is $\frac{\alpha}{1+\alpha}$. The Laplace-Stieltjes transform is given by

$$\begin{aligned}\mathcal{L}_{ZTG}^\alpha(t) &= \int_1^\infty e^{-tu} dF(u) = \sum_{u=1}^\infty e^{-tu} \frac{\alpha^{u-1}}{(1+\alpha)^u} \\ &= \alpha^{-1} \sum_{u=1}^\infty \left(\frac{e^{-t}\alpha}{1+\alpha}\right)^u = \alpha^{-1}\left(\frac{1}{1-\frac{e^{-t}\alpha}{1+\alpha}} - 1\right) \\ &= \left[1 + (1+\alpha)(e^t - 1)\right]^{-1}, \quad \forall t > 0.\end{aligned}$$

Example 3.59. Let V be a discrete frailty random variable taking values $\{1, 2, 3, ...\}$ following a zero truncated Geometric distribution with probability mass function

$$dF_V(v) = p^{v-1}(1-p), \quad p \in (0, 1).$$

The untruncated distribution counts the numbers of successes until one failure occurs. The probability of success is p. It is easy to show that $E(V) = \frac{1}{1-p}$ and that $Var(V) = \frac{p}{(1-p)^2}$. Then the modified random variable $U = (1-p)V$ has mean 1 and variance p. The probability mass function of U is given by $dF_V\left(\frac{u}{1-p}\right)$. The corresponding

Laplace-Stieltjes transform is given by

$$\begin{aligned}
\mathcal{L}^p_{ZTG}(t) &= \int_0^\infty e^{-tu} dF_V\left(\frac{u}{1-p}\right) = \int_0^\infty e^{-t(1-p)v} dF_V(v) \\
&= \sum_{u=1}^{\infty} e^{-t(1-p)v} p^{v-1}(1-p) = \frac{1-p}{p}\left(\frac{1}{1-e^{-t(1-p)}p} - 1\right) \\
&= \left(1 + \frac{1}{1-p}\left(e^{t(1-p)} - 1\right)\right)^{-1}, \quad \forall t > 0.
\end{aligned}$$

We now define what it means for a function to be completely monotone, followed by the definition of a Bernstein function.

Definition 3.60. (Completely monotone) A function $S : (0, \infty) \to \mathbb{R}$ is completely monotone if it is of the class $C^\infty$and

$$(-1)^n \frac{d^n S(t)}{dt^n} \geq 0 \qquad \forall\, n \in \mathbb{N} \cup \{0\} \quad \text{and} \quad t > 0.$$

It is easy to verify that the sum and product of completely monotone functions, are themselves completely monotone as seen in corollary 1.6 in Schilling, Song and Vondračeks book [60].

Definition 3.61. (Bernstein functions) A function $H : (0, \infty) \to \mathbb{R}$ is Bernstein if, and only if, $\frac{dH}{dt}$ is completely monotone.

The following proposition relates completely monotone functions with the Laplace-Stieltjes transform.

Proposition 3.62. *The function $S : (0, \infty) \to \mathbb{R}$ is completely monotone if, and only if, we can write*

$$S(t) = \mathcal{L}_F(t) = \int_0^\infty e^{-tu}\, dF(u) \tag{3.28}$$

where F is unique, positive, bounded and non-decreasing.

The above result says that a function S is completely monotone if it can be written as the Laplace transform of some other function. Note that the function $S(t)$ in proposition 3.62 can be interpreted as the survival function of an exponential mixture distribution [31].

The following proposition relates Bernstein functions with completely monotone functions.

Proposition 3.63. *A function $H : (0, \infty) \to \mathbb{R}$ is Bernstein if, and only if,* $\exp(-u\, H)$ *is completely monotone for every $u > 0$.*

Complete monotonicity is a property of survival functions, hence the use of "S" above, while being a Bernstein function is a property of cumulative hazards.

3.3.2 Univariate mixtures

After the section of preliminary results, we are now able to give details on how to construct univariate frailty mixtures. We first see how cumulative hazards and Laplace transforms are related.

Proposition 3.64. *Let $\mathcal{L}_F(t)$ be a Laplace transform of some probability distribution of a positive random, discrete or continuous, variable, such that $\mathcal{L}_F(t)$ is defined for all $t > 0$. Then*

$$H_{(F)}(t) = -\log \mathcal{L}_F(t)$$

is a cumulative hazard function.

Proof. From proposition 3.62 and from Jewell (1982) we recognise $S(t) = \mathcal{L}_F(t)$ as a survival function of an exponential mixture. Thus $H_{(F)}(t) = -\log S(t)$ is a cumulative hazard. □

Note that the cumulative hazard $H_{(F)}$ is not the cumulative hazard corresponding to the CDF F, that is $e^{-H_{(F)}(t)} \neq 1 - F(t)$. Hence the bracket notation is used to indicate the CH generated by the Laplace transform of F.

Example 3.65. As seen in example 3.56 the Laplace-Stieltjes transform of the standard exponential distribution function is given by $\mathcal{L}_E(t) = \frac{1}{1+t}$. Thus

$$H_{(E)}(t) = -\log \mathcal{L}_{F_E}(t) = \log(1+t),$$

the CH of the standard log-logistic. So in this example we have that $H_{(E)} = H_{ll}$.

Example 3.66. As seen in example 3.57, the Laplace-Stieltjes transform of the Gamma distribution with mean 1 and variance α is given by $\mathcal{L}_\Gamma^\alpha(t) = \left(\frac{1}{1+\alpha t}\right)^{1/\alpha}$. Thus

$$H_{(\Gamma)}^\alpha(t) = \frac{1}{\alpha} \log\left(1 + \alpha t\right).$$

In this example we have that $H_{(\Gamma)}^\alpha(t) = H_E^{1/\alpha} \circ H_{ll} \circ H_E^\alpha(t) = H_{EllE}^{1/\alpha,\alpha}(t)$, which is the Burr Type XII family from example 3.29.

Example 3.67. According to example 3.59 the Laplace-Stieljes transform of the zero truncated Geometric distribution is given by

$$\mathcal{L}_{ZTG}^p(t) = \left[1 + \frac{1}{1-p}\left(e^{t(1-p)} - 1\right)\right]^{-1}.$$

Then we have that

$$\begin{aligned} H_{(ZTG)}^p(t) &= -\log \mathcal{L}_{ZTG}^p(t) = \log\left[1 + \frac{1}{1-p}\left(e^{t(1-p)} - 1\right)\right] \\ &= H_{llEG}^{1/(1-p)}\left((1-p)t\right), \quad p \in (0,1). \end{aligned}$$

Corollary 3.68. *A cumulative hazard of the form $H_{(F)}(t) = -\log \mathcal{L}_F(t)$, where $\mathcal{L}_F$ is the Laplace-Stieltjes transform of a distribution F, is Bernstein.*

Proof. Let $H_{(F)}(t) = -\log \mathcal{L}_F(t)$. Then $\exp(-uH_{(F)}(t)) = \mathcal{L}_F(t)^u$. Since products of completely monotone functions are completely monotone, then for all $u > 0$, $\exp(-uH_{(F)}(t))$ is completely monotone. For the proofs of these statement see Schilling, Song and Vondraček [60]. Thus by proposition 3.63 $H_{(F)}$ is a Bernstein function. □

Example 3.69. From example 3.66, we have that $H^{\alpha}_{(\Gamma)}(t) = \frac{1}{\alpha}\log(1+\alpha t)$ is Bernstein for all $\alpha > 0$.

Example 3.70. The degenerate distribution D_μ at a constant μ is of interest. The corresponding Laplace transform is given by $\mathcal{L}_{D_\mu}(t) = e^{-\mu t}$ and the associated CH is given by

$$H_{(D_\mu)}(t) = -\log \mathcal{L}_{D_\mu}(t) = \mu t.$$

In this case we have $H_{(D_\mu)} = H^{\mu}_{E}$.

The following result gives a simple way to construct mixtures using only compositions of CHs and without integrating.

Proposition 3.71. ***(Frailty mixtures)*** *Let H_0 be a baseline CH function and U be a continuous, positive random variable with cumulative distribution function F. Then $T|U = u \sim u\,H_0$ where $T \sim H_{(F)} \circ H_0$ if, and only if, $H_{(F)}$ is a Bernstein CH function.*

Proof. We have that $T|U = u \sim uH_0$. The marginal survival function of T is given by

$$S_T(t) = P(T > t) = \int_0^\infty P(T > t|U = u)\, dF(u) = \int_0^\infty \exp(-uH_0(t))\, dF(u)$$

Then we can write $S_T(t) = \mathcal{L}_F(H_0(t))$, where $\mathcal{L}_F(s)$ is the Laplace-Stieltjes transform of F. Therefore $H_T(t) = -\log \mathcal{L}_F(H_0(t)) = H_{(F)}(H_0(t))$. By Proposition 3.62, we know the function $\mathcal{L}_F$ is completely monotone.

Since $\mathcal{L}_F(t)$ is completely monotone, then by corollary 3.68 and proposition 3.64 $H_{(F)}(t) = -\log \mathcal{L}_F(t)$ is a Bernstein CH function.

Conversely, suppose the CH function $H_{(F)}$ is a Bernstein function. Proposition 3.63 gives that the function $\exp(-H_{(F)}(t))$ is completely monotone. Thus by proposition 3.62 $\exp(-H_{(F)}(t))$ can be represented by

$$\exp(-H_{(F)}(t)) = \mathcal{L}_F(t) = \int_0^\infty \exp(-tu)\, dF(u)$$

for some non-decreasing function F defined on $[0, \infty)$. Since $H_{(F)}(0) = 0$, we have

$$L_F(0) = \int_0^\infty dF(u) = 1,$$

so that $F(u)$ is bounded by one and is therefore a cumulative distribution function of some U.

Then we can write

$$e^{-H_{(F)}(H_0(t))} = \int_0^\infty e^{-H_0(t)u} dF(u) = S_T(t)$$

which implies $T \sim H_{(F)} \circ H_0$ for some variable T such that $T|U = u \sim uH_0$. $\square$

The usefulness of the above result is in the necessary condition, where the choice of a Bernstein cumulative hazard corresponds to the choice of the distribution of the frailty. Thus mixtures can be constructed by a simple composition. Note that we use the notation $H_{(F)} \circ H_0$ to emphasise the mixing distribution is F.

Example 3.72. Let $H^{\alpha}_{(\Gamma)}(t) = \frac{1}{\alpha}\log(1 + \alpha t)$, the Burr Type XII family, which is a Bernstein function. By proposition 3.71, the composition

$$H^{\alpha}_{(F)} \circ H_0(t) = \frac{1}{\alpha} \log\left(1 + \alpha H_0(t)\right)$$

is the marginal CH function of T in the mixture defined by $T|U = u \sim uH_0$. Here U follows a Gamma distribution with mean 1 and variance α. This can be seen in Hougaard (1984) in example 1 where $\delta = 1/\theta = \alpha$.

Example 3.73. It is easy to verify that $H^{\nu}_{GEll}(t) = (1 + t)^{\nu} - 1$ is Bernstein for $\nu \in (0, 1)$, see section 16.2 of Schilling, Song and Vondraček [60]. Then clearly, the CH function

$$\begin{aligned} H^{\nu,\alpha}(t) &= H_E^{\frac{1-\nu}{\nu\alpha}} \circ H^{\nu}_{GEll} \circ H_E^{\frac{\alpha}{1-\nu}}(t) = H_{EGEllE}^{\frac{1-\nu}{\nu\alpha},\nu,\frac{\alpha}{1-\nu}} \\ &= \frac{1-\nu}{\nu\alpha}\left[\left(1 + \frac{\alpha t}{1-\nu}\right)^{\nu} - 1\right] \end{aligned}$$

is also Bernstein. This modification is necessary to have a frailty distribution with mean 1 and variance α. This can be verified using the fact that $\frac{d}{dt}H_{(F)}(t)|_{t=0} = E(U)$ and $-\frac{d^2}{dt^2}H_{(F)}(t)|_{t=0} =Var(U)$. Then by proposition 3.71, the composition $H^{\nu,\alpha} \circ H_0$ is the marginal CH of T in the mixture defined by $T|U = u \sim uH_0$. Here U follows the Power Variance Frailty (PVF) distribution with mean 1 and variance α, see Hougaard [30]. Note we must restrict to the case where all moments exist.

We note two special cases. First, when $\alpha = 0$, using L'Hospital's rule, it is easy to show that $\lim_{\alpha\to 0} H^{\nu,\alpha}(t) = t = H^1_E(t)$. By example 3.70, we have that it corresponds to a degenerate distribution at the value of 1. Secondly, the case $\nu = 0$ is addressed. Again, using L'Hospital's rule, it can be shown that

$$\lim_{\nu\to 0} H^{\nu,\alpha}(t) = \frac{1}{\alpha}\log(1 + \alpha t) = H^{\alpha}_{(\Gamma)}(t),$$

the Bernstein CH corresponding to a Gamma frailty with mean 1 and variance α.

Finally, one particular example is given when $H_0 = H_E^{\frac{1-\nu}{\alpha}} \circ H_G$, a Gompertz with a frailty parameter. Then

$$H^{\nu,\alpha} \circ H_0 = H_E^{\frac{1-\nu}{\nu\alpha}} \circ H_G \circ H_E^{\nu} = H_{EGE}^{\frac{1-\nu}{\nu\alpha},\nu}$$

another Gompertz with extra scale and frailty parameters.

Example 3.74. From example 3.67 we have that

$$H_{(ZTG)}^{p}(t) = H_{llEG}^{1/(1-p)}\left((1-p)t\right) = \log\left[1 + \frac{1}{1-p}\left(e^{t(1-p)} - 1\right)\right]$$

is a Bernstein CH function. By proposition 3.71 the composition $H_{llEG}^{p}(H_0(t))$ is the marginal CH of T in the mixture defined by $T|U = u \sim uH_0(t)$ where U follows the zero truncated Geometric distribution with mean 1 and variance p. Clearly, if $Var(U) = p = 0$, then

$$H_{(ZTG)}^{p}(H_0(t)) = H_{llEG}^{1}(H_0(t)) = H_0(t).$$

It is interesting to note that there is a different way to obtain a convenient Bernstein CH when the mixing distribution is zero truncated Geometric. If we do not modify the original frailty random variable V in example 3.59, it is easy to show that the generated Bernstein CH is given by

$$\tilde{H}_{(ZTG)}^{p}(t) = H_{llEG}^{1/(1-p)}(t).$$

We can then use the formulae

$$E(U) = \frac{d}{dt}\tilde{H}_{(ZTG)}^{p}(t)|_{t=0}$$
$$Var(U) = -\frac{d^2}{dt^2}\tilde{H}_{(ZTG)}^{p}|_{t=0}$$

to modify the Bernstein CH to have mean one. We have that $E(U) = ... = \frac{1}{1-p}$. It is clear that if we define $\tilde{\tilde{H}}_{(ZTG)}^{p}(t) = (1-p)H_{llEG}^{1/(1-p)}(t)$ then $\frac{d}{dt}\tilde{\tilde{H}}_{(ZTG)}^{p}(t)|_{t=0} = 1$. Finally, we have that $-\frac{d^2}{dt^2}\tilde{\tilde{H}}_{(ZTG)}^{p}(t) = ... = \frac{p}{1-p}$. Then the corresponding new mixture is given by

$$\tilde{\tilde{H}}_{(ZTG)}^{p}(t)\left(H_0(t)\right) = (1-p)H_{llEG}^{1/(1-p)}\left(H_0(t)\right) = (1-p)\log\left(1 + \frac{1}{1-p}\left(e^{H_0(t)} - 1\right)\right)$$

which is different from $H_{(ZTG)}^{p}\left(H_0(t)\right) = H_{llEG}^{1/(1-p)}\left((1-p)H_0(t)\right)$ obtained above. Specifically, $H_{(ZTG)}^{p}(t)$ is a composition combination family with $H_B = H_G$ and $H_A = H_E^1$. We have that $\tilde{\tilde{H}}_{(ZTG)}^{p}(t)$ is a reversed composition combination family with $H_A = H_E^1$ and $H_B = H_{ll}$. When $Var(V) = 0$ then $p = 0$, thus we clearly have that $\tilde{\tilde{H}}_{(ZTG)}^{p}(t) =$

$H_0(t)$.

Example 3.75. (Proportional Frailty model) An important special case is when there is no heterogeneity. This means the distribution of U is degenerate and is concentrated at a value, say $\mu > 0$ as in example 3.70. In this case we have $H_{(D_\mu)}(t) = H_U^\mu(t) = \mu t$ which corresponds to an exponential distribution with rate μ. Then the frailty mixture $H_{(D_\mu)} \circ H_0(t) = \mu H_0(t)$ is a simple frailty parameter family.

Example 3.76. Let $H_{(F)}^\alpha(t) = t^\alpha$, which is Bernstein for $\alpha \in (0,1)$. Then by proposition 3.71, the composition $H_{(F)}^\alpha \circ H_0(t)$ is the marginal CH of T in the mixture $T|U = u \sim uH_0$ where U follows a distribution in the family of positive stable distributions with parameter α. The members of this family do not have finite mean and variance. For more details see Hougaard [30] or Aalen [1].

Example 3.77. Section 16.2 of the book by Schilling et al [60] provides a list of Bernstein functions, some of which are also cumulative hazard functions. One particularly simple one is

$$H_{(F)}(t|\alpha) = \frac{t}{\sqrt{1+\alpha t}}, \quad \alpha > 0.$$

It is easy to show that $E(U) = 1$ and $Var(U) = \alpha$ so that the composition in proposition 3.71, given by

$$H_{(F)} \circ H_0(t) = \frac{H_0(t)}{\sqrt{1+\alpha H_0(t)}},$$

defines a simple new frailty mixture.

Example 3.78. The family with a resilience parameter can be written as

$$H(t|\theta) = H_{llrG}^\theta \circ H_0(t) = H_{ll}\left(H_{rG}^\theta(H_0(t))\right).$$

Thus, since H_{ll} is Bernstein then by proposition 3.71, the family with a resilience parameter is a frailty mixture where $T|U = u \sim H_{rG}^\theta$ and $U \sim H_E^1$.

We will now see an example that relates tilt and frailty parameter families in a mixture setting.

Example 3.79. Consider the tilt parameter family, i.e. the proportional odds, with CHs of the form

$$H_{llEG}^\theta(t) = H_{ll}\left(\theta H_G(t)\right).$$

Then since H_{ll} corresponds to a standard exponential frailty mixing distribution, then we interpret the tilt parameter family as a frailty mixture where $T|U = u \sim u(\theta H_G)$ and $U \sim H_E^1$. That is, a mixture of a frailty parameter family with a standard Gompertz baseline and a standard exponential mixing distribution.

This relationship was discussed in the work of Clayton and Cuzick and Murphy et al. [15, 48].

3.3.3 Bivariate mixtures

By extending proposition 3.71 we can propose a method for modelling the joint distribution of two failure times T_1 and T_2. Here, these models are termed shared frailty models.

If we let the joint survival time be

$$S_{T_1,T_2|U}(t_1,t_2|u) = \exp\left\{-\left(H_{T_1}(t_1) + H_{T_2}(t_2)\right)u\right\}$$

then the joint marginal is given by

$$\begin{aligned} S_{T_1,T_2}(t_1,t_2) &= \int_0^\infty \exp\left\{-\left(H_{T_1}(t_1) + H_{T_2}(t_2)\right)u\right\} f_U(u)du \\ &= \mathcal{L}_F\left(H_{T_1}(t_1) + H_{T_2}(t_2)\right). \end{aligned}$$

Then by proposition 3.71, there must be some Bernstein CH function, $H_{(F)}(u) = -\log(\mathcal{L}_F(u))$ such that

$$S_{T_1,T_2}(t_1,t_2) = \exp\left\{-H_{(F)}\left(H_{T_1}(t_1) + H_{T_2}(t_2)\right)\right\}. \tag{3.29}$$

In chapter 6 we will propose new bivariate survival models. These models can often be viewed as frailty models. The next example will show how a commonly referenced model in the literature can be seen as a frailty model. In later chapters we will view it as a model for informative censoring.

Example 3.80. An important example of a bivariate survival model is that proposed by Clayton [16] where the joint survival function is given by

$$S_{T_1,T_2}(t_1,t_2|\theta) = (1 + \alpha(H_A(t_1) + H_B(t_2)))^{-1/\alpha}$$

where $H_A(\cdot)$ and $H_B(\cdot)$ are non-decreasing functions with $H_A(0) = H_B(0) = 0$. Note that it is not required here that $H_A(\infty)$ and $H_B(\infty) = \infty$. From this survival function, we find the corresponding CH function is

$$\begin{aligned} H_{T_1,T_2}(t_1,t_2) &= \frac{1}{\alpha}\log(1 + \alpha(H_A(t_1) + H_B(t_2)) \\ &= H^{\alpha}_{(\Gamma)}(H_A(t_1) + H_B(t_2)) \end{aligned}$$

Thus the joint model proposed by Clayton (1978) is a gamma frailty mixture. In this model, H_A and H_B are only the marginals in the independent case, $\alpha = 0$.

Example 3.81. Suppose that the random variable U follows a positive stable distribution with parameter $\alpha < 1$ as in the work of Hougaard [30]. Then the joint survival function will be

$$S_{T_1,T_2}(t_1, t_2) = \exp\left\{-\left(H_{T_1}(t_1) + H_{T_2}(t_2)\right)^{\alpha}\right\}.$$

If we let $H_{T_i}(t_i) = H_W^{\beta_i}(t_i)$ for $i = 1, 2$, then we can derive a version of the multivariate Weibull distribution [22].

3.4 Summary and future work

This chapter has defined ways to introduce parameters into a single distribution to create a family. We have seen the scale, frailty, power, hazard power and resilience parameters which are common to the literature. We have also proposed reverse-tilt and reverse-resilience parameters along with the proportional logistic hazard and reverse proportional logistic hazards parameters. These proposed parameters have come about due to our unique framework and the way in which we view our models via the cumulative hazard.

We have also looked at the asymptotic behaviour of the families generated by these parameters, specifically around zero and infinity. We mentioned that these parameters could have different interpretations in the limits, thus we could view these families as ways to join different families between zero and infinity. We may assume a certain type of behaviour at zero and something else around infinity, and may want to find a way to join these behaviours smoothly. This is a concept we would like to explore further.

As an example, consider the proportional logistic hazard parameter given by

$$H(t|\theta) = H_{GEll}^{\theta} \circ H_0(t).$$

Around zero this family acts like $\theta H_0(t)$, thus θ behaves like a frailty parameter. As $t \to \infty$, this family acts like $H_0(t)^{\theta}$, thus θ behaves like a hazard power parameter. Another way to join these behaviours smoothly would be with a very simple spline-like model given by

$$H_s^{\theta}(H_0(t)) = \begin{cases} \theta H_0(t) & t \in [0, 1) \\ \theta H_0(t) + (H_0(t) - H_0(1))^{\theta} & t \geq 1 \end{cases}$$

for $\theta > 0$. This would ensure that the cumulative hazard and the hazard functions are continuous and ensure the chosen behaviours in the limits. In the future, we would hope to develop this spline theory and provide a framework for the joining of whichever behaviours are necessary.

Chapter 4

Survival Regression Models

In this chapter we will review different approaches to survival regression modelling. Many models in the literature focus on the hazard function, as defined in equation (2.3). Hence we will describe the models, first in terms of the hazard function, then put them in terms of the cumulative hazard, to align them with our framework. We will follow the order of section 3.1 when presenting models within the literature and also to propose some new models.

We assume to have structured data of the form of n individuals and that the time to the event of interest, T, is continuous and depends upon p explanatory variables X_1, X_2, ..., X_p. Let $\mathbf{x}_i = (x_{1i}, ..., x_{pi})^T$ denote a vector of the values of the explanatory variable for individual i. We will let $\boldsymbol{\eta}(\mathbf{x}) = \boldsymbol{\beta}^T\mathbf{x}$ denote the linear predictor where $\boldsymbol{\beta}$ is a vector of regression parameters, and for individual i it will be denoted $\boldsymbol{\eta}(\mathbf{x}_i)$ or simply $\boldsymbol{\eta}_i$. We will consider both covariates that do not change over time as well as time varying covariates.

Throughout this chapter $\psi : \mathbb{R} \to \Theta$ is a function linking the linear predictor $\boldsymbol{\eta}$ with a parameter θ which determines the dependence of the distribution of T on the covariates. ψ is of course a function of the covariates, thus should be denoted by $\psi(\boldsymbol{\eta}(\mathbf{x}))$, but will be denoted more succinctly by $\psi(\boldsymbol{\eta})$. Note that ψ is not a link function in the GLM sense but can be thought of as a time-to-event equivalent.

The specification of ψ is an important part of the modelling process, since the interpretation of the regression parameters depends on ψ. The most common choice of ψ is $\psi(\eta) = e^{\eta}$ but other choices can be made. For example, Taulbee [67] suggest the use of alternative linking functions such as $\psi(\eta) = 1 + \eta$ and $\psi(\eta) = (1 + \eta)^{-1}$.

For practical purposes, we impose the restriction $\psi(0) = 1$, which is equivalent to not allowing the linear predictor $\boldsymbol{\eta}$ to have an intercept term. An intercept term can be included by adding an extra parameter into the baseline cumulative hazard. This will be discussed in the last section of this chapter.

We will describe regression models by the cumulative hazard function of the time to event T and it will depend on the covariates $\mathbf{x}$, and the regression parameter $\boldsymbol{\beta}$ via the linear predictor $\boldsymbol{\eta} = \boldsymbol{\beta}^T\mathbf{x}$; the linking function, ψ, and the baseline cumulative hazard function, H_0. In cases where the covariates can affect the distribution of T in more than one way, we will use more than one linking function, ψ_1, ψ_2, ..., on the same or different linear predictors. This will be clarified through specific examples. As opposed to previous chapters, where H_0 was considered to be a fixed CH, in this chapter we will consider H_0 as a parameter of the CH of T.

We will use the following notation for the cumulative hazard of T

$$H(t|\psi(\boldsymbol{\eta}), H_0). \tag{4.1}$$

Here $H(t|\psi, H_0)$ is a cumulative hazard for all $\psi > 0$ and H_0 such that $H(t|1, H_0) = H_0$. This last restriction is common and defines an interpretation of the baseline CH H_0 as the cumulative hazard of T in the case where the covariates are equal to zero. Hence we require $\psi(0) = 1$. As usual, the baseline H_0 can be interpreted as the CH for a typical individual. For example, consider age as a covariate. We may not think that age zero is a "typical" age, we may think that fifty should be our baseline, or typical age. Thus we would define this covariate as $x = \text{age} - 50$. We discuss the specification of H_0 briefly at the end of this chapter.

4.1 Accelerated Failure Time models

The accelerated failure time (AFT) model assumes that the covariates affect the rate at which an individual progresses to the event of interest multiplicatively, i.e. it is accelerated or decelerated [36]. So the covariate effect is multiplicative with respect to the survival time and is a scale parameter. This model can be specified via the hazard function. Let $h_i(t)$ be the hazard function for the ith individual, then the AFT model is such that

$$h_i(t|\psi(\boldsymbol{\eta}_i), h_0) = \psi(\boldsymbol{\eta}_i) h_0(\psi(\boldsymbol{\eta}_i)t)$$

where $h_0(t)$ is the baseline hazard as before and $\boldsymbol{\eta}$ is the linear predictor of the model. Details of this model can be found in Lawless, Meeker and Escobar, and Wei [41, 46, 72]. In terms of the cumulative hazard function, this model can be written as

$$H_i(t|\psi(\boldsymbol{\eta}_i), H_0) = H_0(\psi(\boldsymbol{\eta}_i)t). \tag{4.2}$$

The survival function for the ith individual is $S_i(t|\psi(\boldsymbol{\eta}_i), S_0) = S_0(\psi(\boldsymbol{\eta}_i)t)$, where $S_0(t)$ is the baseline survivor function.

We now view the AFT model in terms of the random variables as a time transformation. This model is equivalent to

$$T_i = \frac{1}{\psi(\boldsymbol{\eta}_i)} T_0. \tag{4.3}$$

This is the same as the time transformation 3.1.1 for the scale parameter. Here we see that the times are proportional, and thus $1/\psi(\boldsymbol{\eta}_i)$ is the proportionality constant. In the literature, the linking function is usually $\psi(\boldsymbol{\eta}) = e^{\boldsymbol{\beta}^T \mathbf{x}}$. This allows the regression model in (4.3) to be written as a linear model.

The log-linear form of the AFT model shows that this class of models is related to the general linear model. Suppose that the random variable T_i represents the lifetime of the ith individual, then the log-linear model says that

$$\begin{aligned}\log(T_i) &= -\boldsymbol{\beta}^T \mathbf{x}_i + \log(T_0) \\ &= -\boldsymbol{\beta}^T \mathbf{x}_i + \log(H_0^{-1}(E)),\end{aligned}$$

where $E \sim H_E^1$. Changing the distribution of T_0 gives a different distribution on T_i. Or, given E, the distribution of the error term is specified through H_0^{-1}.

The AFT model is usually fully parametric since a distribution is generally specified for T_0. There are alternative methods including that of Buckley and James [9] that are semiparametric, where the baseline survivor function is estimated nonparametrically. Jin et al. [32] estimate the parameters of a semi-parametric AFT model using rank based estimation, as do Martinussen and Scheike [45].

We now suppose that the explanatory variables, $\mathbf{x}(t)$, are time dependent. Cox and Oakes [21] suggest a time varying extension of the AFT model. An individual with covariates $\mathbf{x}_i(t)$ at time $t^{(x_i)}$ evolves relative to the time $t^{(0)}$, the time at which $\mathbf{x}_i(t) = \mathbf{0}$, via

$$t^{(0)} = \int_0^{t^{(\mathbf{x}_i)}} \psi\left(\boldsymbol{\beta}^T \mathbf{x}_i(s)\right) ds = \Psi(t^{(\mathbf{x}_i)s}),$$

where ψ is the linking function. Note that here $t^{(x_i)}$ refers to the time at which the ith individuals covariates are $\mathbf{x}_i(t)$. This implies that the time transformation is

$$T_i = \Psi^{-1}(T_0).$$

We see from this formulation, the lifetime of an individual at time t depends on the history of the covariates up to time t. Further, if $\mathbf{x}(t) = \mathbf{x}$ for all t, that is the covariates do not change over time, then we recover the AFT model. Recall that this was seen in section 2.3.10 in example 2.100

In some situations the proportional time assumption (4.3) may be violated, but may hold when the overall population is divided into subgroups, or factors. If this is the case, a stratified AFT model can be fitted. This model allows for the baseline hazard to vary across the levels of the factor variable. Therefore the factor can be adjusted for, without having to estimate its effect in the model. Suppose there are g subgroups.

The cumulative hazard for the ith individual in strata $j = 1, ..., g$ is

$$H_{ij}(t|\boldsymbol{\beta}, H_0) = \prod_{k=1}^{g} H_{0k}(\exp(\boldsymbol{\beta}^T \mathbf{x}_{ij})t)^{I(k=j)}.$$

Here I is the indicator function. We can stratify any model as it only affects the specification of the baseline. The stratification procedure will be discussed in the baseline modelling section.

4.2 Cox Proportional Hazards models

In very simplistic terms, the proportional hazards model says that the covariates are related multiplicatively to the hazard function. This model was proposed by David Cox in 1972 [18] where he wished to extend the work of Kaplan and Meier [37] to include regression arguments into life table analysis. We will now discuss the form that this hazards model takes.

For individual i, the hazard function at time t can be written as

$$h_i(t|\psi(\boldsymbol{\eta}_i), h_0) = \psi(\boldsymbol{\eta}_i)h_0(t).$$

Here, $h_0(t)$ is the *baseline hazard* function, as it is common to all individuals in the study. $\psi(\cdot)$ must be a positive function in order for h_i to be well defined, and so it is often written as $\psi(\boldsymbol{\eta}_i) = \exp(\boldsymbol{\beta}^T \mathbf{x}_i)$.

The Proportional hazards model can be written in terms of the cumulative hazard function,

$$H_i(t|\boldsymbol{\beta}, H_0) = \exp(\boldsymbol{\beta}^T \mathbf{x}_i)H_0(t). \tag{4.4}$$

Let us investigate the interpretation of the β parameters. Consider the log hazard ratio between individuals i and j:

$$\log \left\{ \frac{h_i(t|\boldsymbol{\beta}, h_0)}{h_j(t|\boldsymbol{\beta}, h_0)} \right\} = \beta_1(x_{1i} - x_{1j}) + ... + \beta_p(x_{pi} - x_{pj}).$$

Thus, for $l = 1, ..., p$, β_l is the log hazard ratio of two individuals whose lth covariates x_{li} and x_{lj} differ by 1, with all other covariates equal. Note that the same interpretation is valid for the ratio of cumulative hazards rather than ratios of hazards.

The proportional hazards model assumes that the hazards $h_0(t)$ and $h_i(t|\boldsymbol{\beta}, h_0)$ are proportional, which implies that the hazard ratio between any pair of individuals is constant. This means that if these hazards cross, or intersect, at any time, then this assumption is violated.

As discussed by Cox [19], in the semi-parametric version of the model, and using a partial likelihood, we can estimate the parametric part of the model, $\boldsymbol{\beta}$, without knowledge of $h_0(t)$. Then once we have estimates of the regression coefficients, we can estimate the non-parametric part of the model, $h_0(t)$.

Time dependent Cox model

There are many cases when the model that must be fitted is not a proportional hazards model, one case is when the covariates are time dependent. A Cox model with time dependent variables or coefficients is not proportional as the hazard ratio is not constant. This model is discussed at length by Therneau and Grambsch [68]. The exponential form of the linking function is used. The hazard for the ith individual is

$$h_i(t|\boldsymbol{\beta}, h_0) = \exp\left\{\sum_{j=1}^{p} \boldsymbol{\beta}_j \mathbf{x}_{ji}(t)\right\} h_0(t)$$

where $h_0(t)$ is the hazard for the individual whose values of the covariates are 0 for all t [17]. The hazard ratio for the rth and sth individual is

$$\frac{h_r(t|\boldsymbol{\beta}, h_0)}{h_s(t|\boldsymbol{\beta}, h_0)} = \exp\{\beta_1(x_{r1}(t) - x_{s1}(t)) + ... + \beta_p(x_{rp}(t) - x_{sp}(t))\}.$$

So β_j is the log hazard ratio for two individuals whose values of the jth variable differ by 1 but the other $p-1$ variables are equal. We see that the baseline hazard and the coefficients of the covariates may be harder to interpret. Here the time varying covariates must satisfy the same conditions as in example 2.102.

Stratified Cox model

Occasionally the proportional hazards assumption is violated, but holds within subgroups or factors of the population. If this is the case, a stratified Cox model can be fitted. This model allows for the baseline hazard to vary across the levels of the factor variable.

Suppose there are g subgroups. As seen by Collett [17], the hazard for the ith individual in strata $j = 1, ..., g$ is

$$h_{ij}(t|\boldsymbol{\beta}, h_0) = h_{0j}(t)\exp(\boldsymbol{\beta}^T\mathbf{x}_{ij}).$$

In terms of the cumulative hazard, this can be written more succinctly as

$$H_{ij}(t|\boldsymbol{\beta}, H_0) = \exp(\boldsymbol{\beta}^T\mathbf{x}_{ij})\prod_{k=1}^{g} H_{0k}(t)^{I(k=j)}. \tag{4.5}$$

Note that the coefficients $\boldsymbol{\beta}$ are the same in each stratum but the baseline hazards are different.

It is not only the AFT or the proportional hazards model that can be stratified. The stratified model in (4.5) can be modified for each model for stratification.

4.3 Power Accelerated Failure Time models

The Power Accelerated Failure Time (PAFT) model is motivated by the power parameter as seen in section 3.1. This model is proposed by Burke and MacKenzie [10]. This model is given by

$$H_i(t|\psi(\boldsymbol{\eta}_i), H_0) = H_0\left(t^{\psi(\boldsymbol{\eta}_i)}\right).$$

The time transformation for this model is given by

$$T_i = T_0^{1/\psi(\boldsymbol{\eta}_i)}.$$

We see this is similar to the AFT model, in that the baseline time is accelerated in some way. However, this model accelerates time via powering the baseline time, rather than multiplying by a constant. This type of acceleration thus justifies the name of this model.

Taking logs of the time transformation, we have

$$\log T_i = \frac{1}{\psi(\boldsymbol{\eta}_i)} \log T_0.$$

Choosing $\psi(\eta) = (1+\eta)^{-1}$ is convenient as this generates a model that is linear in the linear predictor $\boldsymbol{\eta}$.

4.4 Proportional Log Hazards models

The Proportional Log Hazards model is motivated by the hazard power parameter, as described in section 3.1. It has CH given by

$$H_i(t|\psi(\boldsymbol{\eta}_i), H_0) = H_0(t)^{\psi(\boldsymbol{\eta}_i)}.$$

Taking logs of this model gives

$$\log H_i(t|\psi(\boldsymbol{\eta}_i), H_0) = \frac{1}{\psi(\boldsymbol{\eta}_i)} \log H_0(t).$$

A convenient choice of ψ would be $\psi(\eta) = (1+\eta)^{-1}$, as this gives linearity with respect to $\boldsymbol{\eta}$. In this case we can write

$$\frac{\log\left[H_i(t|\boldsymbol{\beta}, H_0)/H_j(t|\boldsymbol{\beta}, H_0)\right]}{\log(H_0(t))} = \beta_1(x_{1i} - x_{1j}) + ... + \beta_p(x_{pi} - x_{pj}),$$

giving an interpretation of the regression parameters.

4.5 Proportional Odds models

The main assumption of the Proportional Hazards model is that the covariates are related multiplicatively to the hazard function. The Proportional Odds model proposed by Bennett is similar, in that it assumes covariates are related multiplicatively to the odds of the event happening before time t [5]. Thus the model is specified as follows,

$$\left(\frac{e^{-H_i(t|\psi(\boldsymbol{\eta}_i),H_0)}}{1-e^{-H_i(t|\psi(\boldsymbol{\eta}_i),H_0)}}\right)^{-1} = \psi(\boldsymbol{\eta}_i)\left(\frac{e^{-H_0(t)}}{1-e^{-H_0(t)}}\right)^{-1},$$

where $H_0(t)$ is the baseline cumulative hazard function. Often $\psi(\boldsymbol{\eta}_i) = e^{\boldsymbol{\beta}^T\mathbf{x}_i}$ is used. For this linking function, the parameter $\boldsymbol{\beta}$ is interpreted in terms of the log-odds ratios. If we take the log of the odds for $\psi(\eta) = e^{\eta}$, we find

$$\log\left[\frac{1-e^{-H_i(t|\boldsymbol{\beta},h_0)}}{e^{-H_i(t|\boldsymbol{\beta},h_0)}} \Big/ \frac{1-e^{-H_j(t|\boldsymbol{\beta},h_0)}}{e^{-H_j(t|\boldsymbol{\beta},h_0)}}\right] = \beta_1(x_{1i}-x_{1j}) + ... + \beta_p(x_{pi}-x_{pj}).$$

Thus for $k = 1, ..., p$, β_k is the log odds ratio of two individuals whose kth covariates x_{ki} and x_{kj} differ by 1, with all other covariates equal.

In terms of a CH functional transformation, we find the proportional odds model can be written as

$$H_i(t|\psi(\boldsymbol{\eta}_i), H_0) = H_{ll}\left(\psi(\boldsymbol{\eta}_i)H_G(H_0(t))\right) = H_{llEG}^{\psi(\boldsymbol{\eta}_i)} \circ H_0(t). \quad (4.6)$$

For the proportional hazards model the hazard ratio of individuals with different covariates is constant over time, but in the proportional odds model the same hazard ratio will converge over time. In some cases, except when $H_0 = H_{ll}$, $H_i = H_{ll}(\psi(\boldsymbol{\eta}_i)t)$, so the model becomes an AFT. Recall that at zero, $H_{llEG}^{\theta}(t)$ acts like θt and around infinity acts like t. Thus the CH $H_{llEG}^{\psi(\boldsymbol{\eta}_i)} \circ H_0(t)$ will act like $\psi(\boldsymbol{\eta}_i)H_0(t)$ around zero and H_0 at infinity. The corresponding hazard function will behave like $\psi(\boldsymbol{\eta}_i)h_0(t)$ at zero and $h_0(t)$ at infinity. Thus the hazard ratio converges to 1 as $t \to \infty$.

In Bennett (1983) the proportional odds model is considered semi-parametric and the baseline distribution is estimated. This is achieved by transforming the failure times be log-logistic. The parameters $\boldsymbol{\beta}^T$ are then estimated via maximum likelihood estimation.

4.5.1 Short term Accelerated Failure Time models

The Short term Accelerated Failure Time (StAFT) is motivated by the reverse-tilt parameter. This model is given by

$$H_i(t|\psi(\boldsymbol{\eta}_i), H_0) = H_0 \circ H_{llEG}^{\psi(\boldsymbol{\eta}_i)}(t).$$

The time transformation for this model can be written as

$$T_i = H_{llEG}^{1/\psi(\boldsymbol{\eta}_i)}(T_0).$$

To justify the name of this model we look at the asymptotic behaviour of $H_{llEG}^{1/\psi(\boldsymbol{\eta}_i)}(T_0)$. Around zero $H_{llEG}^{1/\psi(\boldsymbol{\eta}_i)}(T_0)$ acts like $\frac{1}{\psi(\boldsymbol{\eta}_i)}T_0$. Thus in the short term this model can be interpreted as an AFT model. Around infinity $H_{llEG}^{1/\psi(\boldsymbol{\eta}_i)}(T_0)$ acts like T_0.

Rearranging the time transformation and taking logs we find

$$\log\left(e^{T_i} - 1\right) = \log\left(\frac{1}{\psi(\boldsymbol{\eta}_i)}\right) + \log\left(e^{T_0} - 1\right).$$

A convenient choice of $\psi(\boldsymbol{\eta}_i)$ is $\psi(\boldsymbol{\eta}_i) = e^{\boldsymbol{\eta}_i}$, as then we see that this model is linearisable in $\boldsymbol{\eta}_i$. This model does not appear to be proposed in the literature.

4.6 Proportional Log CDF models

The Proportional Log CDF model is motivated by the resilience parameter and is given by

$$H_i(t|\psi(\boldsymbol{\eta}_i), H_0) = H_{llrG}^{\psi(\boldsymbol{\eta}_i)} \circ H_0(t).$$

We find that $\log\left(1 - e^{-H(t|\psi(\boldsymbol{\eta}_i), H_0)}\right) = \psi(\boldsymbol{\eta}_i)\log\left(1 - e^{-H_0(t)}\right)$ which implies

$$\log\left(F_i(t|\psi(\boldsymbol{\eta}_i), H_0)\right) = \psi(\boldsymbol{\eta}_i)\log\left(F_0(t)\right),$$

where F_0 and F_i are CDFs. This equation justifies the name of Proportional log CDF model since $\psi(\boldsymbol{\eta}_i)$ is the proportionality constant for the log CDFs.

In the particular case when $\psi(\eta) = e^{\eta}$ we see

$$\log\left[\frac{\log(F_i(t|\boldsymbol{\beta}, H_0))}{\log(F_j(t|\boldsymbol{\beta}, H_0))}\right] = \beta_1(x_{1i} - x_{1j}) + ... + \beta_p(x_{pi} - x_{pj}).$$

Thus for $k = 1, ..., p$, β_p is the log-log probability ratio of two individuals whose kth covariates x_{ki} and x_{kj} differ by 1 where all other covariates are equal.

Investigating the behaviours of this family at zero and infinity, we find that it acts like $H_0(t)^{\psi(\boldsymbol{\eta}_i)}$ at zero and H_0 at infinity. Thus both the cumulative hazard ratio and the hazard ratio for individuals with different covariates tend to one as $t \to \infty$. The ratio of cumulative hazards tends to $\frac{\psi(\boldsymbol{\eta}_i)}{\psi(\boldsymbol{\eta}_j)}H_0(t)^{\psi(\boldsymbol{\eta}_i) - \psi(\boldsymbol{\eta}_j)}$ as $t \to 0$.

4.6.1 Short term power AFT models

When considering tilt parameter, we proposed a model based on the reverse-tilt parameter. We will do the same now for the reverse-resilience parameter. This model is given by

$$H_i(t|\psi(\boldsymbol{\eta}_i), H_0) = H_0 \circ H_{llrG}^{\psi(\boldsymbol{\eta}_i)}(t).$$

The time transformation is given by

$$T_i = H_{llrG}^{1/\psi(\boldsymbol{\eta}_i)}(T_0).$$

Around zero $H_{llrG}^{1/\psi(\boldsymbol{\eta}_i)}(T_0)$ acts like $T_0^{1/\psi(\boldsymbol{\eta}_i)}$ and around infinity acts like T_0. Thus in the short term this model acts like a power AFT model, thus justifying the name. This model does not appear to be proposed in the literature.

4.7 Proportional logistic hazards models

In section 3.1 we introduced the proportional logistic parameter. Here, we propose a regression model motivated by this parameter given by,

$$H_i(t|\psi(\boldsymbol{\eta}_i), H_0) = H_{GEll}^{\psi(\boldsymbol{\eta}_i)} \circ H_0(t).$$

This model is called the proportional logistic hazards model since

$$H_{ll}\left(H(t|\psi(\boldsymbol{\eta}_i), H_0)\right) = \psi(\boldsymbol{\eta}_i) H_{ll}\left(H_0(t)\right).$$

Thus $\psi(\boldsymbol{\eta}_i)$ is the proportionality constant for the CHs transformed with the log-logistic CH.

For the particular case when $\psi(\eta) = e^{\eta}$, we can see the interpretation of the parameters. We have that

$$\log\left[\frac{\log(1 + H_i(t|\boldsymbol{\beta}, H_0))}{\log(1 + H_j(t|\boldsymbol{\beta}, H_0))}\right] = \beta_1(x_{1i} - x_{1j}) + ... + \beta_p(x_{pi} - x_{pj}).$$

Thus we have that for $k = 1, ..., p$, β_k is the log of the log logistic hazard ratio for two individuals whose kth covariates x_{ki} and x_{kj} differ by 1 with all other covariates equal.

Investigating the behaviours of this family at zero and infinity, we find that it acts like $\psi(\boldsymbol{\eta}_i)H_0(t)$ at zero and $H_0^{\psi(\boldsymbol{\eta}_i)}$ at infinity. Thus the ratio of hazards tends to $\psi(\boldsymbol{\eta}_i)/\psi(\boldsymbol{\eta}_j)$ as $t \to 0$ and to $\frac{\psi(\boldsymbol{\eta}_i)}{\psi(\boldsymbol{\eta}_j)} H_0(t)^{\psi(\boldsymbol{\eta}_i) - \psi(\boldsymbol{\eta}_j)}$ as $t \to \infty$. This model does not appear to be proposed in the literature.

4.8 Accelerated Hazards models

Chen and Wang [14] proposed a new model called the Accelerated Hazards (AH) model, that is not restricted by the condition that either the hazards or survival functions cross over at some time point. It is as follows,

$$h_i(t|\psi(\boldsymbol{\eta}_i), h_0) = h_0\left(\psi(\boldsymbol{\eta}_i)t\right). \tag{4.7}$$

Here $\psi(\boldsymbol{\eta})$ alters the time scale of $h_0(t)$, if $\psi(\boldsymbol{\eta}) > 1$ the effect is to accelerate and is $\psi(\boldsymbol{\eta}) < 1$ it is to decelerate. Note that this acceleration/deceleration is on the hazard scale, not the time or CH scale.

In terms of the cumulative hazard function this model is

$$H_i(t|\psi(\boldsymbol{\eta}_i), H_0) = \frac{1}{\psi(\boldsymbol{\eta}_i)} H_0\left(\psi(\boldsymbol{\eta}_i)t\right).$$

Again, a common form for $\psi(\eta)$ is e^{η}.

Chen and Wang discuss that the proportional hazards model and the AFT model both require that $h_i(t|\psi, h_0)$ and $h_0(t)$ are different at time $t = 0$. Under these models, if $h_i(0|\psi, h_0) = h_0(0)$, then it must be true that $h_i(t|\psi, h_0) = h_0(t)$ for all $t \geq 0$. This is unlikely to be true in settings such as a clinical trial. In a trial investigating the effect of some treatment over a placebo, one would not expect an immediate effect, unless the treatment is highly effective, but the effect will develop over time.

The AH model does not require that the hazards be different at time $t = 0$, instead it is necessary that they are equal. This model therefore should be used when the basic form of the hazards is the same, only the time scale is different

It can be seen that this model is inappropriate when the baseline hazard is exponential as this is constant over time. Thus in this case one wouldn't be able to estimate the regression coefficients. It can also be shown that when the underlying distribution is Weibull, the AH, AFT and proportional hazards models coincide as discussed by Chen and Wang [14]. The AH model also allows the inclusion of time-dependent covariates, via the extension

$$h_i(t|\psi(\boldsymbol{\eta}_i), h_0) = h_0(\psi(\boldsymbol{\eta}_i(t))t).$$

The time transformation corresponding to the Accelerated Hazards model is

$$T_i = H_0\left(\frac{1}{\psi(\boldsymbol{\eta}_i)} H_0^{-1}(\psi(\boldsymbol{\eta}_i)T_0)\right).$$

4.9 Excess Risk models

An excess risk model is used to compare the mortalities between two populations. Sasieni [59] and Crowther [24] show that this model is given by

$$h_i(t|\boldsymbol{\beta}, h_i^*, \lambda_0) = h_i^*(t) + \lambda_0 \exp(\boldsymbol{\beta}^T \mathbf{x}_i),$$

where $h_i^*(t)$ is the expected mortality rate at time t and λ_0 is the baseline excess hazard function. Here h_i is termed the total mortality hazard rate. The parameters $\boldsymbol{\beta}$ are interpreted as the log excess hazard ratios.

A similar model is the Additive Hazards model mentioned by Lin and Ying [42]. In this semi-parametric model the hazard function has the form

$$h_i(t|\boldsymbol{\beta}, h_0) = h_0(t) + \boldsymbol{\beta}^T \mathbf{x}_i(t) \tag{4.8}$$

where $h_0(t)$ is the baseline hazard as before, $\mathbf{x}_i(t)$ is a vector of (possibly) time-varying covariates and $\boldsymbol{\beta}$ is a vector of regression coefficients. Note that $h_0(t)$ is the hazard corresponding to when $\mathbf{x}_i(t) = \mathbf{0}$. This model is similar to that of Aalen et al. [2].

The proportional hazards model estimates hazard ratios, whereas the additive model estimates the difference in the hazards for certain value of $\boldsymbol{\beta}^T \mathbf{x}_i(t)$. If this difference is of interest, the additive hazard is of more use than the proportional hazards model. We may also prefer the additive model when the proportional hazards assumption is violated, that is when the hazard of each individual $h_i(t|\boldsymbol{\beta}, h_0)$ is not proportional to some baseline hazard $h_0(t)$.

Due to the fact that hazard functions must be non-negative, we can immediately see that in order for $h_i(t|\boldsymbol{\beta}, h_0)$ in (4.8) to be a hazard function, we would require $h_0(t) + \boldsymbol{\beta}^T \mathbf{x}_i(t) \geq 0$. This would mean we would need to place specific constraints on $\boldsymbol{\beta}^T \mathbf{x}_i(t)$, which may not always be realistic given certain sets of data. One way to ensure that the right hand side of (4.8) is always positive is to replace $\boldsymbol{\beta}^T \mathbf{x}_i(t)$ with $e^{\boldsymbol{\beta}^T \mathbf{x}_i(t)}$. Note that this would change the interpretation of $h_0(t)$, it would then be the hazard corresponding to when $\boldsymbol{\beta}^T \mathbf{x}_i(t) = -\infty$. These concerns were raised by Lin and Ying [42] along with a detailed description of how to carry out the inferences in both cases.

4.10 Extended regression models

The survival regression models in the previous sections have one linear predictor. In the proportional hazards model, this parameter is interpreted as the hazard ratio and as the odds ratio in the proportional odds model. This section explores models with more parameters which extend the models in the previous section. These extend the previous models by having them as submodels. This section will include extended regression models within the literature by authors such as Yang and Prentice [75] and Chen and Jewell [13], but will also include some new models we propose. Related ideas also appear in the work of Burke and MacKenzie [11].

4.10.1 Proportional hazard-Proportional odds (PH-PO) models

Yang and Prentice (2005) develop a model that generalises the Proportional hazards and Proportional Odds models [75] so that their model can accommodate crossing survival curves and include meaningful parameters. Their model focuses on the two-sample case. They suppose they have two groups, a control and a treatment group, with corresponding hazards $h_C(t)$ and $h_T(t)$. Their model is then of the form

$$h_T(t|\theta_1, \theta_2, h_C) = \frac{\theta_1 \theta_2 h_C(t)}{\theta_1 + (\theta_2 - \theta_1)e^{-H_C(t)}},$$

where θ_1 and θ_2 are positive. Yang and Prentice claim that θ_1 and θ_2 can be interpreted as the short-term and long-term hazard ratios respectively. Observe that if we let $\theta_1 = \theta_2$ in equation (4.9), we recover the proportional hazards model. Note also that letting $\theta_2 = 1$ gives us the proportional odds model.

After a reparametrisation, $\psi(\eta_1) = \theta_2$ and $\psi(\eta_2) = \theta_1/\theta_2$, we can write this model in terms of the CH functions in our framework

$$H_i(t|\psi(\eta_{1i}), \psi(\eta_{2i}), H_C) = H_E^{\psi(\eta_{1i})} \circ H_{llEG}^{\psi(\eta_{2i})} \circ H_C(t). \tag{4.9}$$

Here we have $\eta_1 = \boldsymbol{\beta}^T \mathbf{x}$ and $\eta_2 = \boldsymbol{\gamma}^T \mathbf{x}$ and ψ is the same function. We note that if $\boldsymbol{\beta} = \mathbf{0}$ then we recover the proportional odds model, and the proportional hazards model is recovered if $\boldsymbol{\gamma} = \mathbf{0}$. We also note that $\mathbf{x}$ is the collection of all covariates, thus the η_1 and η_2 could included some of the same covariates, but could also have no covariates in common. This form of the model is a regression extension of the original two-sample model, as discussed at the end of the Yang and Prentice (2005) paper.

4.10.2 Proportional hazards-AFT-Accelerated Hazards (PH-AFT-AH) models

Chen and Jewell present a general family of semiparametric hazards models in their paper [13]. This model generalises the Proportional Hazards, Accelerated Failure Time and Accelerated Hazards models. It achieves this by combining these models so that given certain values of the parameters, their model reduces to one of the three main models.

Their generalised model is given by the hazard function

$$h(t|\boldsymbol{\lambda}, \boldsymbol{\theta}, h_0) = e^{\boldsymbol{\lambda}^T \mathbf{x}} h_0(e^{\boldsymbol{\theta}^T \mathbf{x}} t).$$

The cumulative hazard of Chen and Jewell's model is

$$H_i(t|\eta_{1i}, \eta_{2i}, H_0) = e^{\eta_{1i} - \eta_{2i}} H_0(e^{\eta_{2i}} t). \tag{4.10}$$

Here H_0 is the baseline CH. We can see that if $\eta_2 = 0$, their model reduces to the proportional hazards model. If $\eta_1 = \eta_2$ then they recover the accelerated failure time model, and the accelerated hazards model if $\eta_1 = 0$. Note that if $\eta_1 = \boldsymbol{\lambda}^T \mathbf{x}$ and $\eta_2 = \boldsymbol{\theta}^T \mathbf{x}$ then, the covariates in the linear predictors can be disjoint or there can be some covariates, depending on the zero values of $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$.

This model was proposed as semi-parametric since the baseline hazard is not specified and the parameters λ and η are estimated. For the two-sample case, where there is a treatment and a control group, H_0 would be the CH function for the control group and H_i the CH for the treatment group. It is noted in Chen and Jewell (2001) that this model is not identifiable if the baseline hazard is from a Weibull distribution.

4.10.3 Proportional hazards-Proportional odds (PH-PO) spline models

Royston and Parmar [57] proposed a flexible parametric model based on the proportional hazards and the proportional odds models. Their model is presented in terms of the survival function and is given by

$$g\left[S(t|\boldsymbol{\beta}, S_0)\right] = g\left[S_0(t)\right] + \boldsymbol{\beta}\mathbf{x}, \tag{4.11}$$

where g is a link function. They choose a parametrised link $g(x|\alpha) = \log\left(\frac{x^{-\alpha}-1}{\alpha}\right)$. In our framework, this model is

$$\begin{aligned} H(t|\alpha, \psi(\eta), H_0) &= \frac{1}{\alpha} H_{llEG}^{\psi(\eta)}(\alpha H_0(t)) \\ &= \psi(\eta) C_{\overset{\rightarrow}{\circ}}^{\alpha} \left(\frac{1}{\psi(\eta)} H_{llEG}^{\psi(\eta)}, H_E^1 \right) \circ H_0(t), \end{aligned} \tag{4.12}$$

as seen in equation (3.26). We see here that the baseline cumulative hazard is given by $H(t|0, \psi(\mathbf{0}), H_0) = H_0(t)$. They choose to only model the cases $\alpha = 0$, the proportional hazards model, and $\alpha = 1$, the proportional odds model. This is due to the interpretation of η, or lack thereof, if α was not either of these values. For each model they then model the log baseline CH, or the baseline log odds, respectively, with a cubic spline, thus resulting in a PH spline or PO spline model.

Other models can be generated by using H_{GEll}^{θ}, for example, instead of H_{llEG}^{θ} in equation (4.12).

4.10.4 Proportional odds-AFT (PO-AFT) models

We will now extend the models in the first few sections to two dimensional families, using similar techniques to Yang and Prentice [75] and Chen and Jewell [13]. We will also aim to generalise the one dimensional models we propose in section 3.2.2.

The first proposed model is inspired by the model formulation proposed by Yang and Prentice. Their model contains the proportional hazards and the proportional odds models as submodels. Our model differs from theirs as we propose a model with the proportional odds and the accelerated failure time models as submodels. This model is based on that proposed in equation (3.24), given by

$$H(t|\alpha,\beta) = H_B\left(\frac{1}{\alpha}H_B^{-1}\left(H_A(\beta t)\right)\right)$$

Thus we consider the following equation,

$$H(t|\eta_1,\eta_2,H_0) = H_{llEG}^{\psi(\eta_1)} \circ H_0 \circ H_E^{\psi(\eta_2)}(t), \tag{4.13}$$

obtained by letting $H_B = H_{ll}$ and $H_A = H_0$. If $\eta_1 = \mathbf{0}$ we recover the accelerated failure time model. Letting $\eta_2 = \mathbf{0}$ we recover the proportional odds model. We could see this model as an accelerated proportional odds model. Other models can be generated by using H_{GEll}^{θ} instead of H_{llEG}^{θ}.

4.10.5 Proportional hazards-Short term Accelerated Failure Time (PH-StAFT) models

The following regression model is based on the family given in (3.23)

$$H(t|\alpha,\beta) = \frac{1}{\alpha}H_A \circ H_B^{-1}(\beta H_B(t)).$$

We propose

$$H(t|\eta_1,\eta_2,H_0) = H_E^{\psi(\eta_1)} \circ H_0 \circ H_{llEG}^{\psi(\eta_2)}(t),$$

which is obtained by letting $H_A = H_0$ and $H_B = H_{ll}$. Then if $\eta_2 = \mathbf{0}$, we recover the proportional hazards model, and if $\eta_1 = \mathbf{0}$ we recover the short term accelerated failure time model. Other models can be generated by using H_{GEll}^{θ} instead of H_{llEG}^{θ}.

4.10.6 Accelerated Failure Time-Short term Accelerated Failure Time (AFT-StAFT) models

The concept of the PH-PO spline model proposed by Royston and Parmar can be extended to other models. We may wish to combine the AFT and StAFT models in a way such that varying one parameter implies the other parameters come from one of these models. This would be achieved by the model,

$$H(t|\alpha,\eta) = H_0\left(\psi(\eta)C_{\circlearrowleft}^{\alpha}\left(\frac{1}{\psi(\eta)}H_{llEG}^{\psi(\eta)}(t), H_E^{1}(t)\right)\right), \tag{4.14}$$

which is the reverse composition of (4.12). Here $\alpha = 0$ yields an AFT model with parameter $\psi(\eta)$ and $\alpha = 1$ recovers an StAFT model with parameter $\psi(\eta)$. This family will be equivariant to linear, geometric and composition combinations for H_0.

Other models can be obtained by using H_{GEll}^{θ} instead of H_{llEG}^{θ} in equation (4.14).

4.11 Modelling of the baseline hazard

The specification of the baseline cumulative hazard H_0 in the model (4.1), $H(t|\eta, H_0)$, can be done in three different ways.

1. The baseline can be completely specified. For example, we can specify a standard CH, such as the standard Gompertz or the standard log-logistic.

2. The baseline can be specified up to some unknown parameters. More specifically, we specify H_0 to be a member of a parametric family of cumulative hazards $\mathcal{H}_0^{\boldsymbol{\alpha}} = \{H(t|\boldsymbol{\alpha}); \boldsymbol{\alpha} \in \mathcal{A}\}$, where $\boldsymbol{\alpha}$ might be a vector.

3. The baseline can be completely unspecified. We do not restrict H_0 to be of any specific parametric form. The model is then called semi-parametric.

In the second case, the full set of unknown parameters is given by $(\boldsymbol{\beta}, \boldsymbol{\alpha})$, where $\boldsymbol{\beta}$ is the vector of regression parameters and $\boldsymbol{\alpha}$ relate to the baseline. Specification of these parameters thus completely specifies the model (4.1). This modelling technique is usually called fully parametric regression as seen in Crowther (2014) and references therein [24].

In order to capture different behaviours of the baseline hazard, ie. concavity, convexity, linearity or combination thereof, it is usual to use a flexible parametric family which can capture different behaviours for different values of the parameter $\boldsymbol{\alpha}$. The earliest approach is to use piecewise exponential models [25], so that the baseline CH is a piecewise linear increasing function. This requires the specification of the joining time points, or knots, and the scale parameter for each exponential piece. This falls outside our set $\mathcal{CH}$ of CHs, which requires smoothness, since the piecewise linear CHs are not differentiable at the knots.

A different approach was used by Royston and Parmar [57] where they use smooth cubic splines, in the log time scale, to model the logarithm of either the cumulative hazard or the odds of survival. The splines require the specification of a number of knots and parameters, giving the flexibility to this approach.

The parametric model we use for H_0 starts with a combination family $C^{\alpha}(H_A, H_B)$, for a given choice of standard CHs H_A and H_B. Clearly we may want to add some extra parameters. For example, scale and power parameters, as well as a tilt parameter, can be added so that the resulting parametric family for the baseline is of the form

$$\mathcal{H}_0^{\boldsymbol{\alpha}} = \left\{ H_{llEG}^{\alpha_1} \circ C^{\alpha_0}(H_A, H_B) \circ H_{WE}^{\alpha_2, \alpha_3}(t); \alpha_1 > 0,\, i = 0, ..., 3 \right\}.$$

Our model will thus be of the form

$$H(t|\psi(\eta), \mathcal{H}_0^{\boldsymbol{\alpha}}). \tag{4.15}$$

We note that in the special case where one of the added baseline parameters, say α_j, is of the same type as the parameter that includes the covariates, that is $\psi(\eta)$, then we have that H in (4.15) depends on $\psi(\eta)$ and α_j only through $\alpha_j\psi(\eta)$. To see this, consider the example of a proportional odds model, so that $\psi(\eta)$ is a tilt parameter, where we want to add a tilt parameter to the baseline. We then have that (4.15) can be written as

$$H_{llEG}^{\psi(\eta)} \circ \left[H_{llEG}^{\alpha_j} \circ H_0(t)\right] = H_{llEG}^{\psi(\eta)\alpha_j} \circ H_0(t).$$

Other parameters, scale, frailty, etc., have the same property.

In the case where $\psi(\eta) = e^{\eta}$ we can readily see that $\log \alpha_j$ plays the role of the intercept in the linear predictor, i.e. $\psi(\eta)\alpha_j = e^{\log \alpha_j + \eta}$, and is not redundant since we imposed $\psi(0) = 1$ earlier. In the case where $\psi(\eta) = 1 + \eta$ we have that $\psi(\eta)\alpha_j = \alpha_j + \alpha_j\eta$ and the effect of introducing α_j to the baseline is that of adding an intercept and rescaling all the regression parameters by the same factor α_j. For the choice $\psi(\eta) = (1+\eta)^{-1}$, the effect is the same but with $1/\alpha_j$ instead of α_j.

Now we illustrate why the equivariance properties of section 3.1.8 are useful in a regression context. Consider the following example of a proportional odds regression model. We can write this model as

$$H(t|\psi(\eta), \mathcal{H}_0^{\boldsymbol{\alpha}}) = H_{llEG}^{\psi(\eta)} \circ C^{\alpha}(H_A(t), H_B(t))$$

and assume the combination C^{α} is equivariant to the addition of a tilt parameter. Then, we have

$$H(t|\psi(\eta,)\mathcal{H}_0^{\boldsymbol{\alpha}}) = C^{\alpha}\left(H_{llEG}^{\psi(\eta)} \circ H_A(t), H_{llEG}^{\psi(\eta)} \circ H_B(t)\right)$$

meaning this regression model can be interpreted as the combination of two simpler PO models, $H_{llEG}^{\psi(\eta)} \circ H_A(t)$ and $H_{llEG}^{\psi(\eta)} \circ H_B(t)$ where the baseline is fully specified. Clearly the same equivariance property holds if we add parameters of which their effect is equivariant under C^{α}.

4.12 Summary and future work

This chapter reviewed different regression models for time-to-event data including the Accelerated Failure Time, Proportional hazards and Proportional Odds models and some extended regression models. This uses the parametric families discussed in chapter 3 to construct new models including those based on the combination families. The combination families are found to be useful in the very last section in this chapter which discussed the flexible modelling of the baseline cumulative hazard function. This way of modelling the baseline will be illustrated in the following chapter when fitting models to the liver transplantation data.

An area of future work is to explore how to perform automatic flexible estimation of the baseline. This estimation would let the data decide which and how many parameters need to be added to the baseline to have a flexible estimate.

Chapter 5

Liver Transplantation Data

In this chapter we will give details of the data set we use to illustrate our methodologies throughout this thesis. The data are on the use of liver transplantation to treat end stage liver disease. The data consists of measurements on patients from the liver transplant registry. These patients are suffering from some liver disease, where transplant is the preferred treatment. For each patient in the data set, some patient characteristics are recorded at entry to the registry. Also recorded are their *survival times*. Here this indicates the number of days on the registry until either death, transplant, or in a few cases, the time from registration until the date of data extraction. The data was provided by NHS Blood and Transplant to investigate the presence of informative censoring and possible methods to overcome these issues.

The first section of this chapter will give further details of the data set. In the following section we will perform an initial analysis, in order to give us a better understanding of the data set. This will help us to uncover what issues are inherent with this data set which we will need to attempt to solve in later chapters.

The main issue this data set provides is that the censoring is informative to the event process. If a patient is very ill they are likely to die, but they are also more likely to receive a transplant. Therefore, we see that there is dependence between death, the event of interest, and transplant, the censoring. This is termed *informative censoring*. The final section of this chapter will demonstrate that the censoring is indeed informative.

5.1 Liver transplantation data set

This section will provide the details of the liver transplantation data set. We will explain what covariates are present and give any necessary description of what these covariates are.

Table 5.1 gives the details of the variables in the data set, stating their names and giving a description of what the variable is. If a variable is categorical then the number of patients in each category is given. If a variable is continuous, then its range and mean are given. There are 4449 patients in this data set, but not all of them have observations for each of the covariates. Details of the number of missing observations for each covariate are also given in table 5.1. Table 5.2 gives the details of each disease category.

In the data set there are some patients with `status` and `txcens` both equal to 0. This means that they have neither died nor received a transplant at the date of data extraction, July 2009, carried out by NHS Blood and Transplant. These individuals are removed from the analysis. There are no details of any possible death after transplant, i.e. there are no patients where `status` and `txcens` both equal 1. We thus have that time until death is censored by transplant, or time until transplant is censored by death.

There are 122 patients who have a survival time of 0. This would imply they spent no time on the transplant registry, and on the day they received a transplant or died, their information was added to the data set. It is standard practice to remove such patients from the analysis. This is consistent with our assumption that $H(0) = 0$, see property 2.1 in chapter 2.

Variable Name	Description
`patient`	Patient number ($n = 4449$)
`survtime`	Survival time from listing to either death or transplant
`status`	Indicator of whether the patient is alive or dead at the survival time: 0 = alive ($n = 3590$), 1 = dead ($n = 859$)
`txcens`	Indicator of whether the patient has had a transplant at the survival time: 0 = no transplant ($n = 1016$), 1 = transplant ($n = 3433$)
`disease`	Primary liver disease at time of registration
`recip_age`	Age of patient at registration (Range: 17-78, Mean: 52)
`UKELD`	UK End stage Liver Disease score (1930 missing) (Range: 38-83.2, Mean: 55.4)

Table 5.1: The variable names and descriptions in the Liver Transplantation data set. The range and mean are given for the continuous variables, along with details of how many missing observations there are. For categorical variables, the numbers of observations in each category are given.

5.1.1 United Kingdom End-Stage Liver Disease Score

Transplants are often assigned to the sickest patients, meaning those with the greatest need get priority. In order to assign transplants fairly, a model or score is required to identify who needs the transplant most and who might benefit the most. In the United States, the Mayo End-Stage Liver Disease (MELD) score has been used since February 2002 [73].

The MELD score is based on three common measurements, serum bilirubin, creatinine and international normalized ratio (INR) for prothrombin time. It is calculated

Disease Group	Primary liver disease
1	Primary biliary cirrhosis (PBC) ($n = 581$)
2	Primary sclerosing cholangitis (PSC) ($n = 420$)
3	Alcoholic liver disease (ALD) ($n = 1161$)
4	Auto-immune and cryptogenic disease (AID) ($n = 526$)
5	Hepatitis C cirrhosis (HCV) ($n = 701$)
6	Hepatitis B cirrhosis (HBV) ($n = 163$)
7	Cancer ($n = 215$)
8	Metabolic liver disease ($n = 199$)
9	Other liver diseases ($n = 453$)
10	Acute hepatic failure ($n = 30$)

Table 5.2: Primary liver disease groups with details of how many patients with each disease given.

using the formula

$$\begin{aligned} \text{MELD} =& 9.57 \ln(\text{creatinine mg/dl}) + 3.78 \ln(\text{bilirubin mg/dl}) \\ &+ 11.2 \ln(\text{INR}) + 6.43. \end{aligned} \tag{5.1}$$

The MELD score has proved useful but was not specifically developed to select which patients need liver transplants. Thus the UK End-Stage Liver Disease (UKELD) score was developed to predict transplant list mortality on a cohort of patients on the UK liver transplant waiting list [3]. The UKELD score is given by the formula

$$\begin{aligned} \text{UKELD} =& [5.395 \ln(\text{INR}) + 1.485 \ln(\text{creatinine}) + 3.130 \ln(\text{bilirubin}) \\ &- 81.565 \ln(\text{sodium})] + 435. \end{aligned} \tag{5.2}$$

We see that the UKELD score in (5.2) uses all the same measurements as the MELD score in (5.1) with the addition of serum sodium. In a general sense, UKELD score is a measure of how ill a patient is and is thus expected to be key to modelling death due to liver disease censored by transplantation. We also note that this covariate is the output of a model itself and is subject to measurement error.

5.2 Exploratory data analysis

In this section we carry out an exploratory data analysis. Firstly we present some plots and statistics to demonstrate the characteristics of the data set. We will then fit some proportional hazards models to the data assuming non-informative censoring. We will then see how there is a need to incorporate informative censoring into the model.

First, we investigate the dependence between the covariates and transplantation. In figures 5.1, 5.2, 5.3 and 5.4 we see graphs comparing the proportions or distributions of patients who did or did not have a transplant for the variables disease, UKELD score, age at registration and for patient survival times.

In figure 5.1 we compare the proportion of transplants received for each disease category. We see that on average the proportion of transplants over each disease category is around 0.8, and although there is slight variation between the diseases, there does not appear to be a very strong relationship between transplantation and disease category.

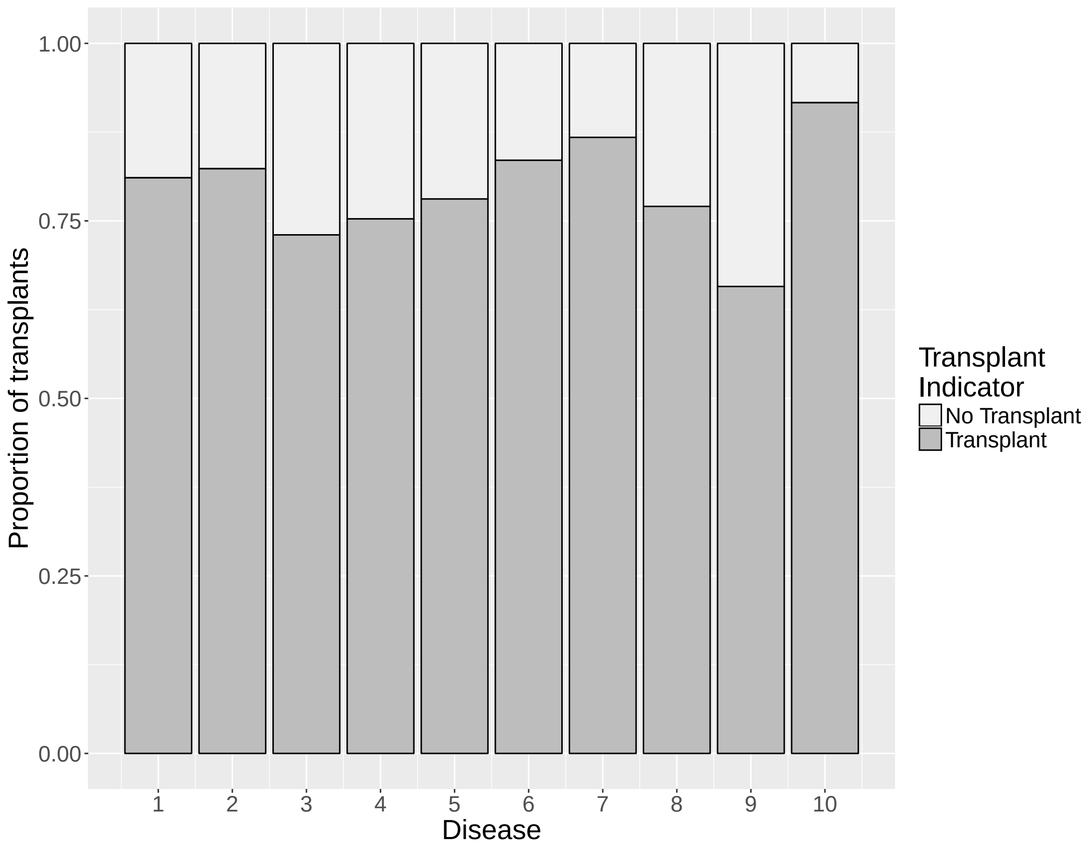

Figure 5.1: Proportion of transplants per disease

In figure 5.2 we compare the kernel density estimates of UKELD score at registration for the patients who have, or have not, had a transplant. We see that the corresponding kernel density estimates differ, so there may be some relationship between transplant and UKELD score. This does not seem surprising, since UKELD score is a measure of patient health. If this measure indicates that the patient is very unwell, we would expect them to be more likely to receive a liver transplant.

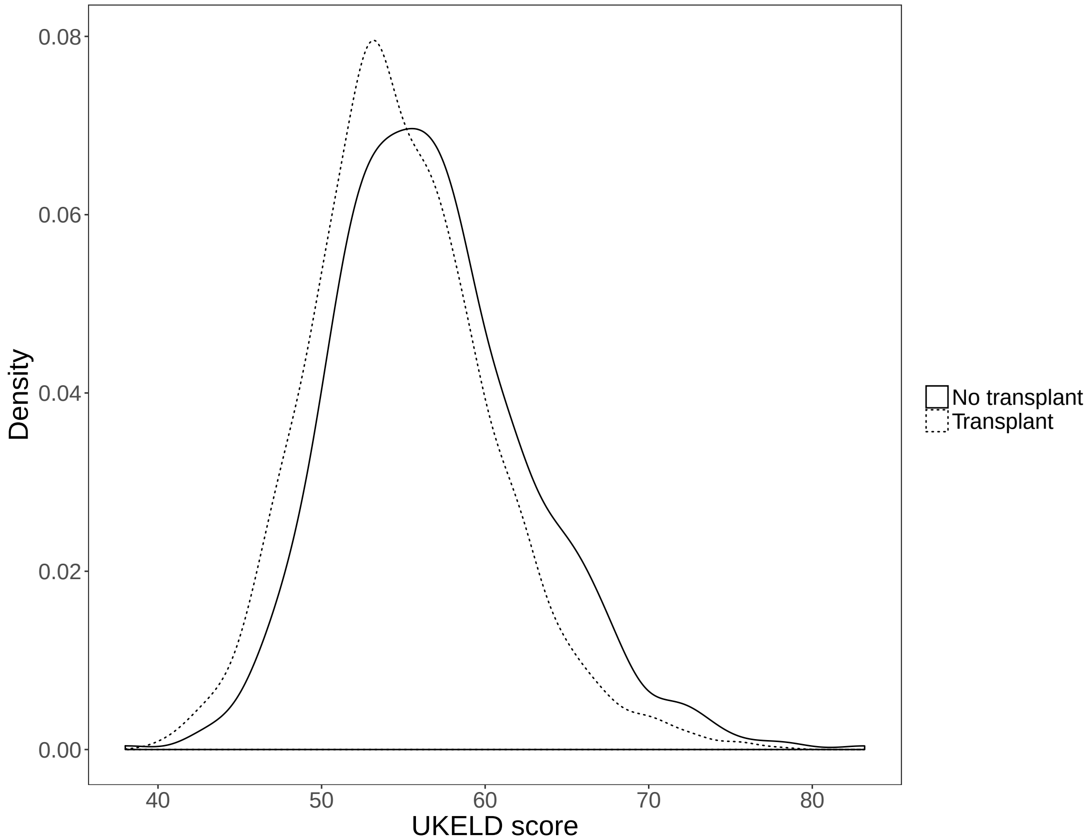

Figure 5.2: Density estimate of UKELD score at registration

In figure 5.3 we investigate whether there is some dependence between recipient age and transplant indicator. We see that the kernel density estimates of both groups are similarly shaped. We might infer that there is likely no relationship between patient age and whether they might receive a transplant.

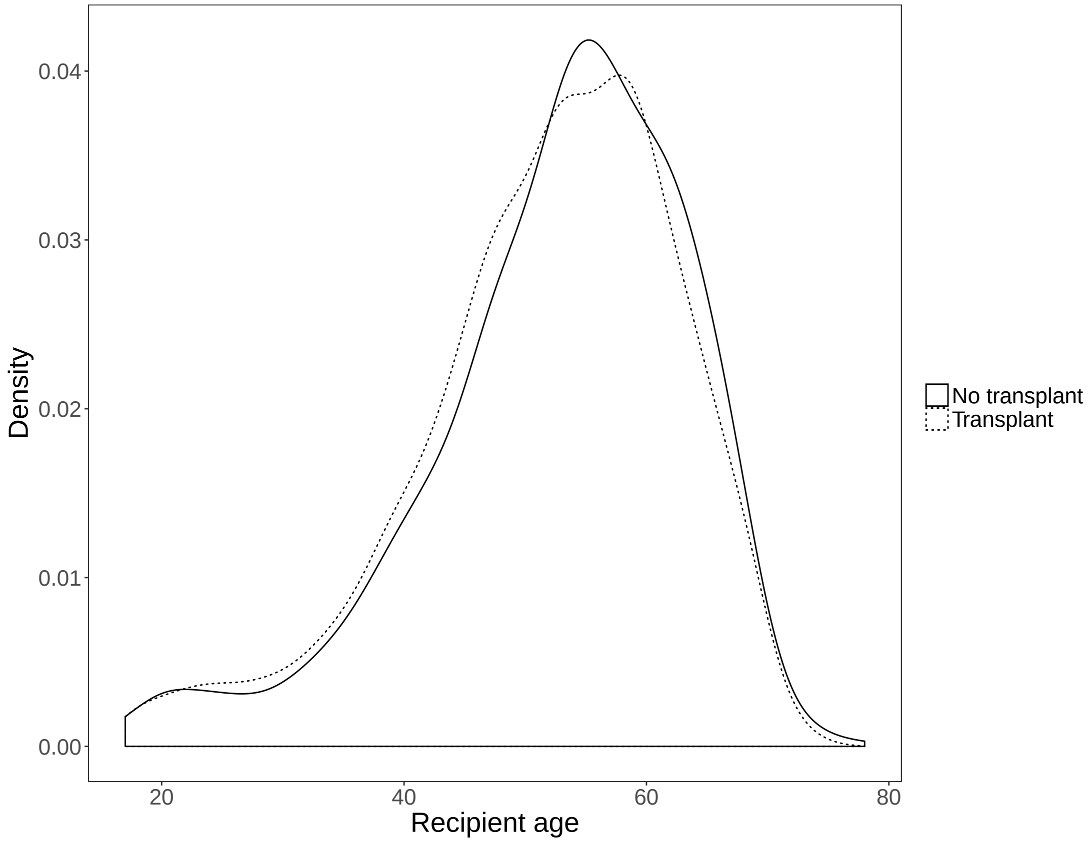

Figure 5.3: Density estimate of recipient ages at registration

The final plot figure 5.4 shows the kernel density estimates of the survival times for each group. We see that these may be differently distributed. We see that transplants are more likely to occur at shorter survival times, just like the death times. However, we might suppose that more transplants occur closer to time of registration than deaths. We may use this plot to justify that there is little dependence between the survival time, and whether a transplant was received or not. We aim to show that this is not true, and further investigation is needed.

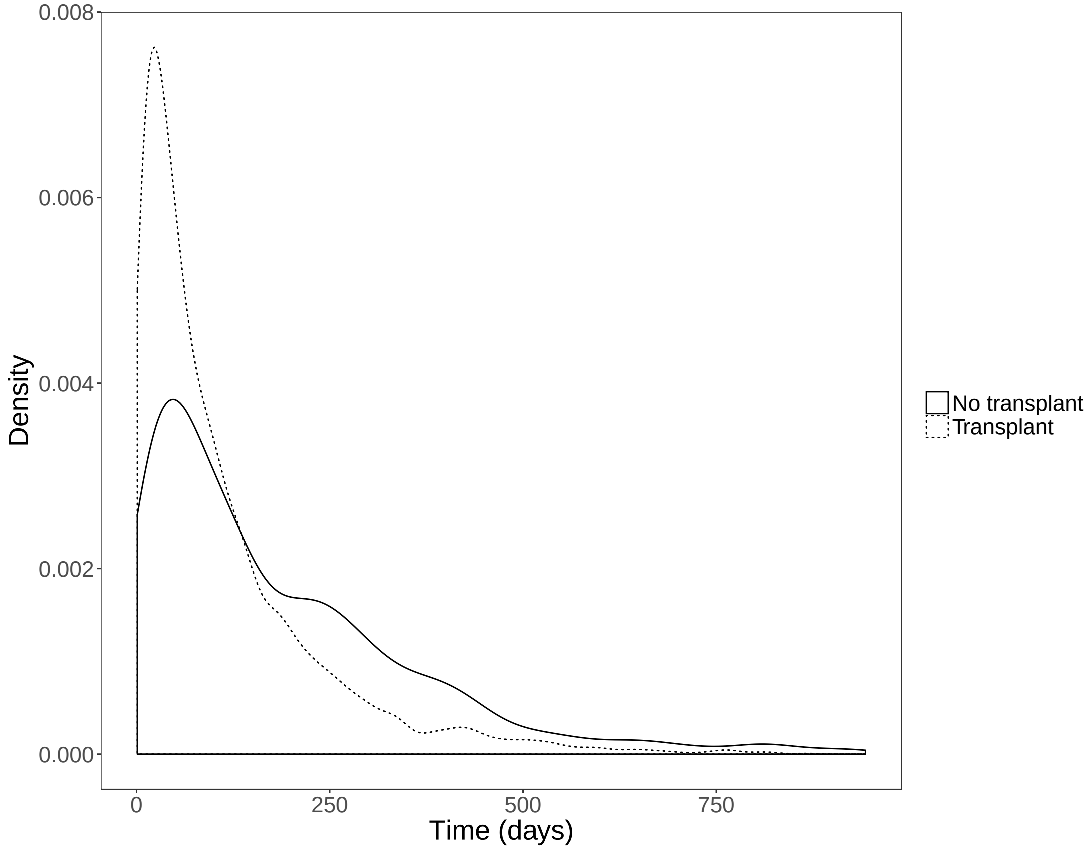

Figure 5.4: Density estimate of patient survival times

Figure 5.5 shows the Kaplan-Meier estimates of the cumulative hazard function for the time to death and time to transplant, i.e. censoring. In the next chapter these will be referred to as the crude survival functions. The cumulative hazards are clearly separated, where the median survival time for time to death is 446 days, whereas for time to transplant is 97 days. This indicates that patients who are transplanted tend to spend less time on the transplant list than those who die.

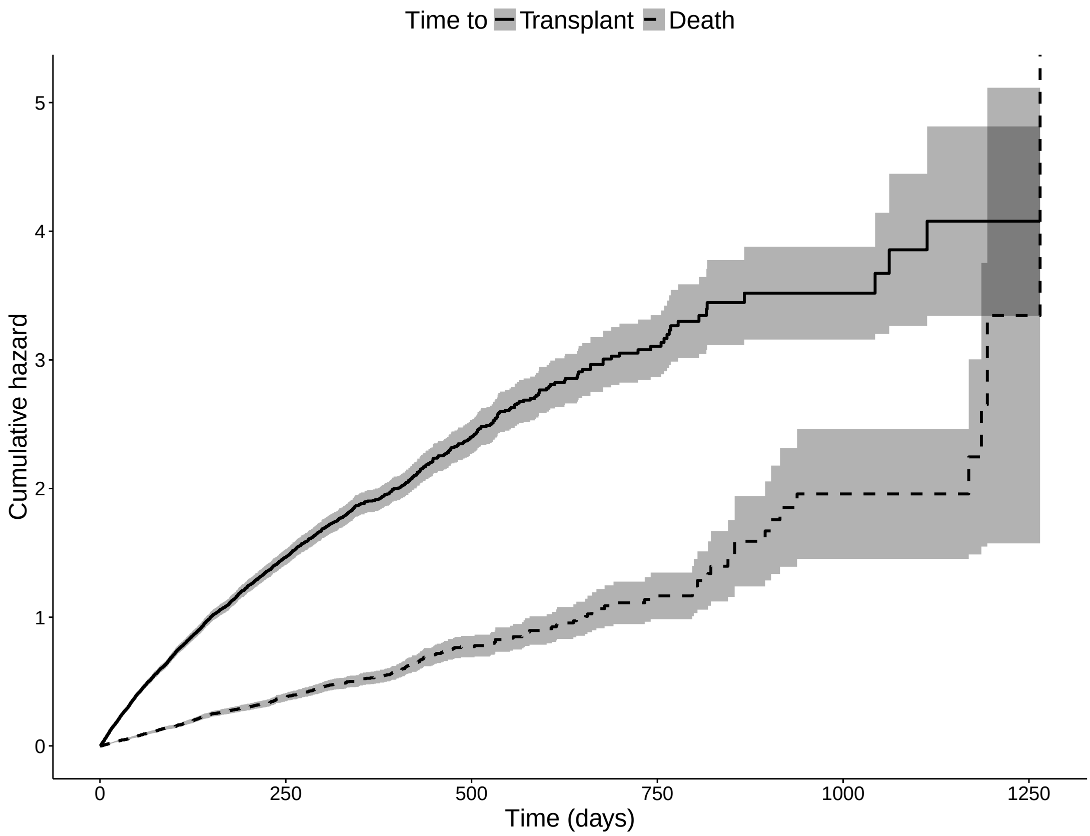

Figure 5.5: Kaplan-Meier estimates of the cumulative hazard functions for time to death censored by transplant and for time to transplant censored by death.

5.2.1 Fitting known regression models

Semi-parametric model

As part of an initial analysis a semi-parametric Cox proportional hazards model was fitted to the data for time to death and for time to transplant. The Cox proportional hazards model is given by

$$H_i(t) = \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i) H_0(t),$$

where $\boldsymbol{\beta}$ is the vector of parameters, $\boldsymbol{x}_i$ is the vector of covariates for individual i and $H_0(t)$ is the baseline cumulative hazard function.

Parameter	Hazard Ratio $\exp(\beta_i)$	Standard Error	p-value
Disease (PBC)			
Disease (PSC)	0.73	0.23	0.18
Disease (ALD)	1.17	0.15	0.32
Disease (AID)	1.30	0.18	0.15
Disease (HCV)	1.34	0.18	0.11
Disease (HBV)	1.16	0.31	0.62
Disease (Cancer)	1.66	0.29	0.08
Disease (Metabolic)	2.10	0.22	< 0.001
Disease (Other)	2.01	0.18	< 0.001
Disease (Acute)	0.84	0.73	0.81
Age	1.02	0.004	< 0.001
UKELD score	1.15	0.01	< 0.001

Table 5.3: Results from the Cox model for time to death, assuming non-informative censoring due to transplant. PBC disease is the reference factor.

Parameter	Hazard Ratio	Standard Error	p-value
Disease (PBC)			
Disease (PSC)	1.19	0.10	0.09
Disease (ALD)	0.99	0.08	0.95
Disease (AID)	1.02	0.10	0.87
Disease (HCV)	1.24	0.09	0.02
Disease (HBV)	1.50	0.15	0.01
Disease (Cancer)	2.33	0.12	< 0.001
Disease (Metabolic)	1.27	0.13	0.07
Disease (Other)	0.80	0.11	0.04
Disease (Acute)	1.75	0.28	0.05
Age	1.00	0.002	0.82
UKELD score	1.03	0.005	< 0.001

Table 5.4: Results from the Cox model for time to transplant, assuming non-informative censoring due to death. PBC disease is the reference factor.

Tables 5.3 and 5.4 contain the results of fitting a Cox proportional hazards model to the liver transplantation data for time to death and time to transplant, respectively. The baseline hazard gives the hazard of death, or transplant, for a patient with Primary biliary cirrhosis, with age 0 and a UKELD score of 0. Here the hazard ratio is reported instead of the parameter estimates. For example, in table 5.3 the ratio of hazards where the UKELD score differs by one, with all other covariates equal, is 1.15. The p-values relate to the null hypothesis that the coefficients are equal to zero.

We see that for the model for time to death, age and UKELD score are both significant, as is disease category. However, for the model for time to transplant, we do not have that age is significant. Removal of age from the model ensured all other covariates were significant.

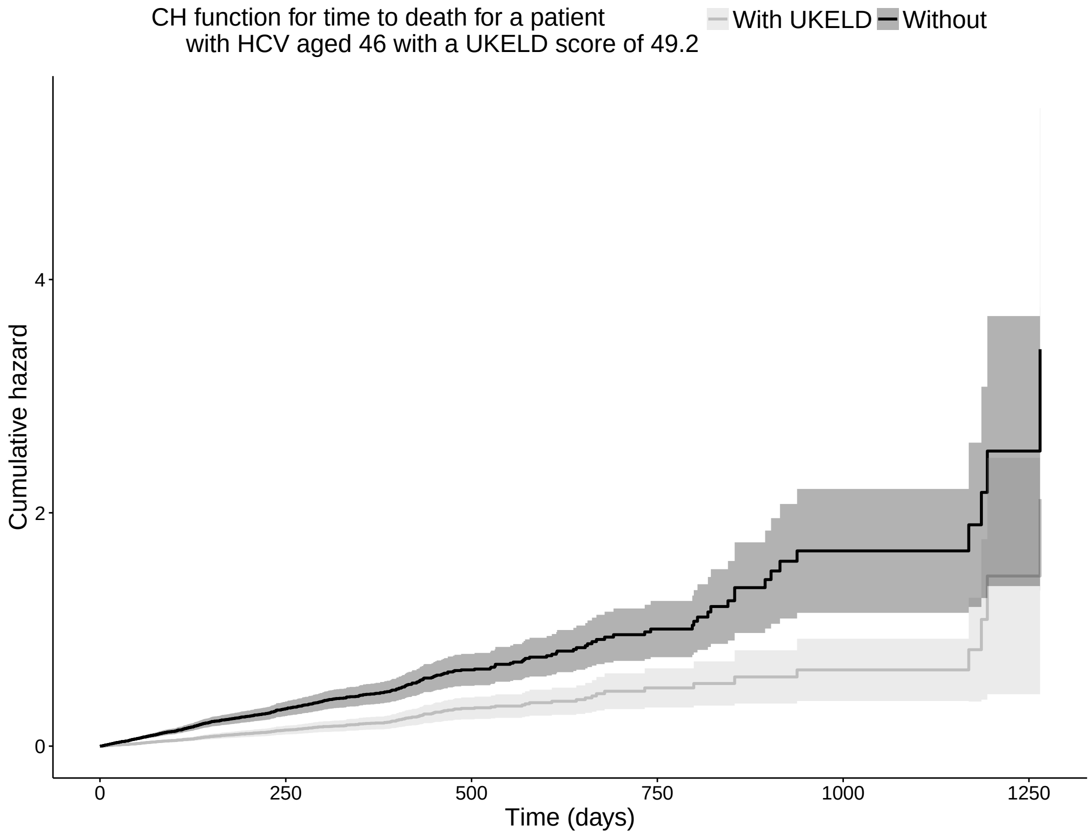

Figure 5.6: Cumulative hazard function for a patient aged 46 with disease HCV and a UKELD score of 49.2 in a Cox model fitted with and without UKELD score.

An important issue to note is that there are a lot of missing values for UKELD score, as highlighted in table 5.1. Our analysis only deals with complete cases, thus a model with UKELD score included has fewer observations than a model without UKELD score. Figure 5.6 shows two cumulative hazard curves, one from a model with UKELD included and all other covariates, and the other with it excluded. Here we see that the inclusion of UKELD score in the model drastically changes the cumulative hazard function, leading to a more conservative estimate of survival. We cannot rule out that the UKELD score could be missing-not-at-random (MNAR), meaning there is some reason why particular patients have missing UKELD score. It may be that patients who are healthier have their score missing, which would mean the survival estimate would be more conservative than necessary as suggested by figure 5.6. It could also be that patients who are sicker have missing UKELD and the survival curves should in fact be even more conservative than they are now. The issue of MNAR covariates however, is not the focus of this thesis, instead we investigate informative censoring.

Fully parametric model

As the models we propose are fully parametric, part of our initial analysis will include the use of such models in the literature. We have chosen to focus on the accelerated failure time model. The AFT model is given by

$$H_i(t) = H_0(\exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)t),$$

where $\boldsymbol{\beta}$ is the vector of parameters, $\boldsymbol{x}_i$ is the vector of covariates for individual i and $H_0(t)$ is the baseline cumulative hazard function.

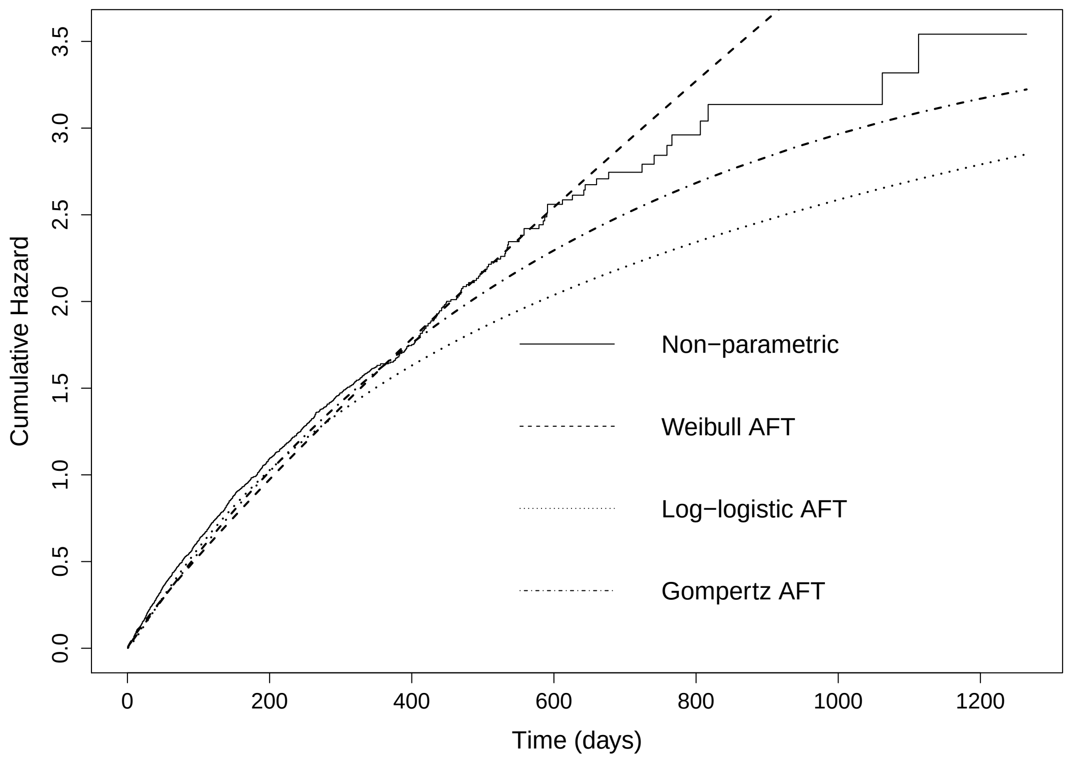

Figure 5.7: Non-parametric estimate of the cumulative hazard function (solid line) compared to the Weibull AFT (dashed line), Log-logistic AFT (dotted line) and the Gompertz AFT (dot-dashed line) estimated cumulative hazards for a patient with mean age (52), mean UKELD score (55) with ALD.

Fitting an AFT model with a Weibull baseline hazard is equivalent to fitting a proportional hazards model. The results of this model are given in table 5.5, where we see all covariates are significant except recipient age. We see the cumulative hazard of the Weibull AFT in figure 5.7 for an average individual, i.e. a patient with mean age and UKELD score and ALD, the most common disease in this data set. In this figure the Weibull AFT CH is compared with the estimated cumulative hazard for the log-logistic AFT model and the Gompertz AFT model for the same individual and the non-parametric estimate of the cumulative hazard. Note that comparing the CH functions of the average individual is comparing the baseline CH in our context, after re-centring the covariates appropriately. The log-logistic AFT model appears to reflect the shape of the non-parametric estimate, whereas the Weibull AFT model seems to capture the early characteristics of the non-parametric model. We see that the CH for the average individual from the Gompertz AFT model more accurately reflects the shape and characteristics of the non-parametric estimate.

Parameter	Estimate	Standard Error	p-value
Intercept	7.27	0.36	< 0.001
Disease (PSC)	-0.21	0.12	0.08
Disease (ALD)	0.01	0.10	0.96
Disease (AID)	-0.02	0.11	0.84
Disease (HCV)	-0.26	0.11	0.01
Disease (HBV)	-0.46	0.17	0.01
Disease (Cancer)	-1.00	0.13	< 0.001
Disease (Metabolic)	-0.26	0.15	0.09
Disease (Other)	0.28	0.13	0.03
Disease (Acute)	-0.52	0.32	0.10
Age	-0.001	0.003	0.70
UKELD score	-0.03	0.01	< 0.001
Log(scale)	0.13	0.02	< 0.001

Table 5.5: Results from the AFT model with Weibull baseline, for time to death with censoring due to transplant. PBC disease is the reference factor.

5.2.2 Fitting combination models

In chapter 4 we discussed how to use the combination models in chapter 3 as regression models. Recall the models we propose are termed the linear, geometric and composition combinations. In this section we use these models as the baseline hazard in the parametric proportional hazards, or the accelerated failure time models. We expect this will provide some flexibility for the baseline, if we use $H_1(t)$ is convex and $H_2(t)$ is concave.

As an example, we consider the linear combination.

$$H(t|\alpha, \beta, \nu) = (1 - \alpha)H_1(t^\beta) + \alpha H_2(t^\nu).$$

We tried various combinations of H_1 and H_2, but restricted ourselves to consider the cases where one was convex and one was concave. The parameters of the model were estimated in R by maximum likelihood estimation. We have that the parameter θ is the exponential of the linear predictor, so

$$\log(\theta) = \beta_0 + \beta_1 \texttt{disease} + \beta_2 \texttt{recip\_age} + \beta_3 \texttt{UKELD}.$$

For the AFT model with linear combination baseline, we found the combination of the Gompertz and the log-logistic worked best. This model was easily fitted using the `flexsurv` package. This package requires the cumulative hazard and the hazard of the model to be defined and outputs non-parametric estimates of the cumulative hazard and the model parameters. This package thus provides a method for model checking and allows for the easy fitting of our more complex models. The combination models could be fitted using other packages such as `timereg`, we however, find the `flexsurv` package to be the most convenient.

The estimates of the cumulative hazard are compared to the non-parametric estimate in figure 5.8. The confidence interval for the coefficient of `recip_age` contained one and thus was removed for the model. We see that this model stays within the confidence limits of the non-parametric estimate of the CH function in figure 5.8.

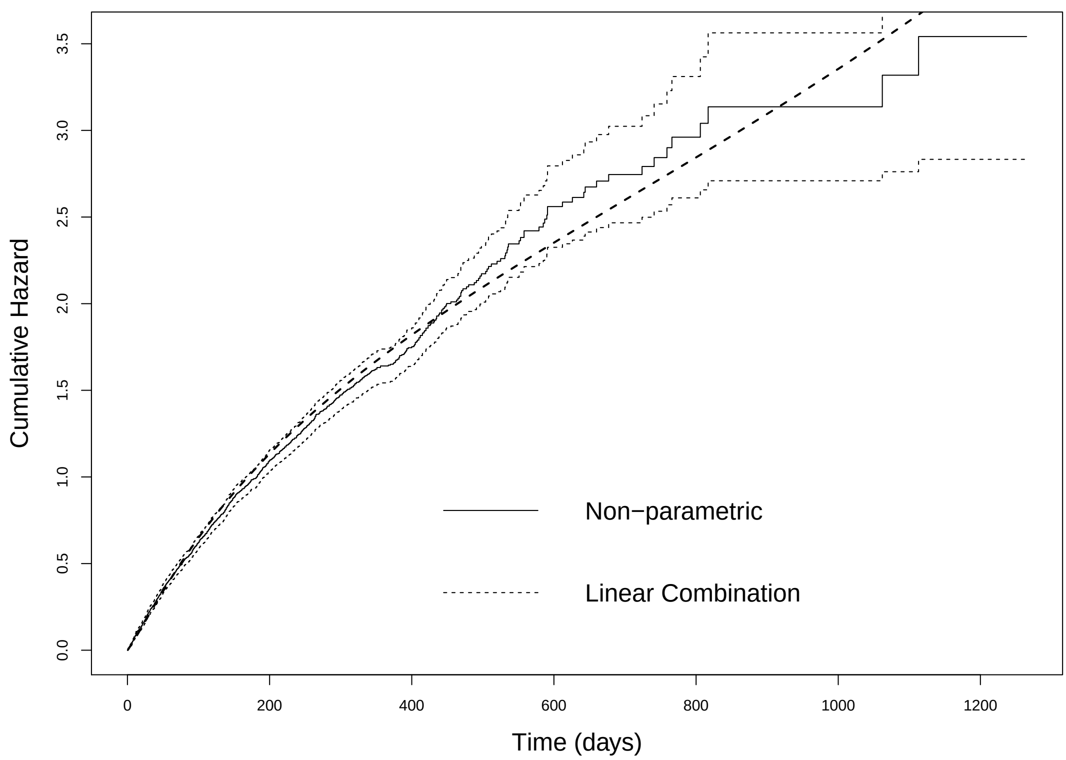

Figure 5.8: The non-parametric estimate of the CH function compared with the CH for the parametric AFT model with a linear combination baseline.

5.2.3 Issues with informative censoring

In this data set, informative censoring would occur when the unknown mechanisms causing death or transplant are dependent in some way. As detailed in chapter 6, it is impossible to determine whether the censoring is informative or not from the data.

If we perform a simple logistic regression on the censoring indicator using the UKELD score, recipient age and disease category as covariates, we see that they are all significant in table 5.6. This implies that censoring, or transplantation, depends on each of these covariates, and thus we could suspect that the censoring is informative since we have seen from the Cox regressions that the survival time also depends on all of these covariates too. This method for investigating the independent censoring assumption was discussed by Collett [17].

Parameter	Estimate	Standard Error	p-value
Disease (PBC)	0		
Disease (PSC)	0.39	0.22	0.07
Disease (ALD)	-0.18	0.16	0.26
Disease (AID)	-0.19	0.19	0.32
Disease (HCV)	-0.06	0.18	0.76
Disease (HBV)	0.36	0.32	0.26
Disease (Cancer)	0.82	0.30	0.01
Disease (Metabolic)	-0.19	0.24	0.43
Disease (Other)	-0.74	0.20	< 0.001
Disease (Acute)	1.14	0.78	0.15
Age	-0.01	0.004	0.001
UKELD score	-0.06	0.01	< 0.001

Table 5.6: Results from the logistic regression on the censoring indicator.

Chapter 6

Cumulative Hazards Models for Informative Censoring

6.1 Introduction

A key element of survival data is the presence of censoring, which occurs when the event of interest is not fully observed. There are many types of censoring and it can occur in many ways. For example, if a trial has a fixed end point, the event of interest may not have occurred for all trial participants by the end of the study, and so those participants who have not experienced the event would be censored. If the event is death by some disease, a patient may die from some other competing disease, and thus they would be censored. Another common cause for censoring is loss-to-follow-up, where the patient can no longer be contacted and so there is no way of knowing if they have experienced the event or not.

The problem in practice is that the only available information is if the individual is censored or not and not the reasons for why it was censored. For this reason, most methods for analysing time-to-event data conveniently assume that the censoring mechanism is in no way related to the event time mechanism. This is called *non-informative censoring*. This assumes that the censoring and event times are independent and that the corresponding distributions are functionally independent. This assumption is not always appropriate and can sometimes have serious effects on the inference.

In this chapter we focus on *right censoring* where it is known that the lifetime is longer than reported, but it isn't known how much longer. The most common approach to dealing with this type of censoring is to model the observed time as the minimum of the event time and the censoring time, i.e. whichever occurs first. If our potential event time is T and the censoring time is C, then we observe

$$Y = \min(T, C),$$

and the censoring indicator

$$\Delta = \begin{cases} 0 & T \leq C, \text{ event time observed} \\ 1 & T > C, \text{ event time censored.} \end{cases}$$

In a non-informative context, we assume both that the random variables T and C are independent and the parameters θ and γ of the distributions for T and C, $f_T(t|\theta)$ and $f_C(c|\gamma)$ respectively, are functionally independent. I.e., there does not exist some function $q(\cdot)$ such that $\gamma = q(\theta)$. In this chapter we review what is meant by informative censoring via non-informative censoring.

The goal of analysing the survival data of the form (Y, Δ) is to make inference about the joint distribution of (T, C), and in most problems, only the marginal distribution of T. However, given (Y, Δ), there are an infinite number of joint distributions (T, C) that could correspond to the distribution of (Y, Δ) [74]. For the same joint distribution of (Y, Δ), there will be some compatible joint distributions where T and C are independent, and others where they will be dependent. Thus given only data (Y, Δ) for different individuals, it is impossible to learn whether the event and censoring times are independent or not.

The latter part of this chapter aims to propose a sensitivity analysis to learn about the marginal distributions of T and C and overcome the above non-identifiability issue.

6.2 Specification of informative censoring

In this section we discuss different types of right censoring, the assumptions required for such models and the impact on the analyses. We aim to specify what is meant by informative censoring. This is achieved by discussing what is meant by non-informative censoring, which is the complement of informative censoring. Non-informative censoring is a common, and often unjustified, assumption in time-to-event data analysis. Thus determining the class of non-informative censoring models gives insight into when these models are actually justified. When they are not, we look to informative censoring models.

The following defines what is meant by non-informative censoring, which Crowder refers to as the Makeham assumption [23, 35]. We follow a direct distributional approach, but there are other approaches, such as that based on counting processes as in Kalbfleisch and Prentice [36].

Definition 6.1. Given a survival time random variable T and a censoring random variable C, the joint survival model characterised by the joint survival $S_{T,C}$ is a non-informative censoring model if

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} P(t < T \le t + \epsilon, T \le C | T > t) = h_T(t)$$
$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} P(c < C \le c + \epsilon, C < T | C > c) = h_C(c).$$

This definition states the hazard functions for the event and censoring mechanisms in a system where both events and censoring can occur.

In a general model for right censoring, the random variable Y is defined such that

$$Y \le T,$$

where Δ is defined as

$$\Delta = \begin{cases} 0 & Y = T, \text{ event time observed} \\ 1 & Y < T, \text{ event time censored.} \end{cases}$$

This general model does not require any latent censoring variable, and in some circumstances, is a more intuitive representation of the system under study.

Under the constraint that an individual must either be censored or experience the event, we have that

$$\begin{aligned} 1 = P(Y = T) + P(Y < T) &= \int_0^\infty \Big(P(Y = y, Y = T) + P(Y = y, Y < T) \Big) dy \\ &= \int_0^\infty P\left(Y = T | T = y\right) P\left(T = y\right) dy + \int_0^\infty P\left(Y < T, Y = y | T > y\right) P\left(T > y\right) dy. \end{aligned}$$

Here we have informally used probabilities with the event $\{Y = y\}$ instead of a more formal measure theoretic quantity. Within this chapter we will write $f_T(t)dt = P(T = t)$, meaning that in a neighbourhood around t of length dt the probability that $\{T = t\}$ is the density evaluated at that value multiplied by the length. A more formal way to write the above equation is

$$\int_0^\infty P(Y = T | T = y) f_T(y) dy + \int_0^\infty f_Y(y | T > Y) P(T > Y) dy = 1$$

where $f_T(y)$ is the marginal density of T evaluated at y and $f_Y(y|T > Y)$ is the conditional density of the observed time, Y, when censoring occurs. The same equation can be written as

$$\begin{aligned} 1 &= \int_0^\infty a(y) f_T(y) dy + \int_0^\infty S_T(y) dB(y) \\ &= \int_0^\infty \left[a(y) + B(y)\right] f_T(y) dy \end{aligned}$$

where

$$a(y) = P(Y = T|T = y) = P(T = Y)\frac{f_T(y|T = Y)}{f_T(y)}, \text{ and} \tag{6.1}$$

$$dB(y) = P(Y < T, Y = y|T > y)dy = \frac{f_Y(y|T > Y)}{S_T(y)}P(T > Y)dy \tag{6.2}$$

as in the works of Williams and Lagakos [74, 40, 39]. In the above, $a(y)$ is the probability that the potentially observable time $\{T = y\}$ is actually uncensored, and $dB(y)$ is the conditional probability that the event time is censored at a time y given that the individual has survived up to time y.

To see the last equivalence in the above, $\int_0^\infty S_T(y)dB(y) = \int_0^\infty B(y)f_T(y)dy$, we use integration by parts, i.e.

$$\begin{aligned}\int_0^\infty S_T(y)dB(y) &= [S_T(y)B(y)]_{y=0}^{y=\infty} + \int_0^\infty f_T(y)B(y)dy \\ &= \int_0^\infty B(y)f_T(y)dy\end{aligned}$$

since here $B(y) = \int_0^y dB(s)$ and so $B(0) = 0$ and also $B(\infty) < \infty$.

Since any observed time is either censored or not, then the contribution to the likelihood for an uncensored observation is $P(Y = y, Y = T) = a(y)f_T(y)$ and is $P(Y = y, Y < T) = dB(y)S_T(y)$ for a censored observation. The likelihood for independent and identically distributed data $(y_i,\ \delta_i)$, then becomes

$$\prod_i \left(a(y_i)f_T(y_i)\right)^{1-\delta_i} \left(S_T(y_i)dB(y_i)\right)^{\delta_i}. \tag{6.3}$$

As discussed by Williams and Lagakos [74], the condition

$$a(y) + B(y) = 1, \text{ for all } y \text{ such that } f_T(y) > 0 \tag{6.4}$$

characterises all models for which the likelihood is of the form

$$\prod_i f_T(y_i)^{1-\delta_i} S_T(y_i)^{\delta_i}. \tag{6.5}$$

This can be seen from the following arguments. Clearly from (6.1), $a(y) = P(Y = T|T = y)$ tends to zero as $y \to \infty$. We have that $B(y) = \int_0^y dB(s)$ is an increasing function with $B(0) = 0$. Then the equation (6.4), $a(y) + B(y) = 1$ implies that $B(y)$ tends to one when $y \to 0$, so that $B(y)$ is a CDF given by $B(y) = 1 - a(y) = P(Y < T|T = y)$. This can be re-written as follows

$$B(y) = P(Y < T|T = y) = P(Y < T, Y < y|T = y)$$

since $\{Y < T, T = y\} \subset \{Y < y\}$. Then we have $dB(y) = P(Y < T, Y = y|T = y)dy$

which is almost the same as (6.2) except that here the condition event is $\{T = y\}$ instead of $\{T > y\}$. Then $dB(y)$ and $a(y)$ are, respectively, the conditional probabilities of the event time being censored at y and not censored given the event time T is y.

The likelihood (6.3) can be re-written as

$$\prod_i f_T(y_i)^{1-\delta_i} S_T(y_i)^{\delta_i} \times \prod_i a(y_i)^{1-\delta_i} dB(y_i)^{\delta_i}. \tag{6.6}$$

The first factor of the likelihood involves only the marginal distribution of the event time T and the second factor only involves the conditional distribution of the censoring process, conditioned on realisations of the event times. Therefore, using the likelihood principle in Cox and Hinkley [20], only the first factor in the likelihood (6.6) can be used for inferences about the distribution of interest F_T and the second factor can be safely ignored.

Condition (6.4) is called the constant-sum property and is equivalent to non-informative censoring as described in definition 6.1 and by Kalbfleisch and MacKay [35].

Now consider the following different assumption

$$P\left(T = t | Y < T, Y = y\right) = \frac{f_T(t)}{S_T(y)}, \forall\, 0 < y < t. \tag{6.7}$$

Lagakos [39] states that the likelihood for this model can be written also as in (6.5). To see this, note first that

$$\{T = t\} = (\cup_{y \le t} \{T = t, Y = y, Y < T\}) \cup \{T = t, Y = T\},$$

then

$$\begin{aligned} f_T(t)dt = P(T = t) &= \int_0^t \big(P(T = t|Y = y, Y < T)\big) P(Y = y, Y < T) dy \\ &\quad + P(Y = T|T = t)P(T = t) \\ &= \int_0^t \big(P(T = t|Y = y, Y < T)\big) S_T(y) dB(y) + a(t) f_T(t) dt \\ &= \int_0^t \Big(\frac{f_T(t)dt}{S_T(y)}\Big) S_T(y) dB(y) + a(t) f_T(t) dt \text{ by (6.7)} \\ &= \left(\int_0^t dB(y) + a(t)\right) f_T(t) dt \\ &= \left[a(t) + B(t)\right] f_T(t) dt. \end{aligned}$$

Thus for these to be equivalent, we require $a(t) + B(t) = 1$ for all $t > 0$, i.e. a constant-sum model. Thus models satisfying (6.7) are non-informative censoring models.

Models satisfying (6.7) are referred to as *non-prognostic* censoring models. This property states that the information provided by the censoring at time y is only that the true survival time is larger than y. Williams and Lagakos [74] give an example of a model that is constant-sum but does not satisfy the non-prognostic censoring condition. This implies that the set of non-prognostic censoring models is a proper subset of non-informative censoring models.

The usual model for right censoring includes an interpretation of the censoring random variable, with the observed survival time being

$$Y = \min(T, C),$$

and the censoring indicator

$$\Delta = \begin{cases} 0 & T \leq C, \text{ event time observed} \\ 1 & T > C, \text{ event time censored.} \end{cases}$$

If we assume that T and C are independent, then using (6.2) we have

$$\begin{aligned} dB(y) &= P(Y < T, Y = y | T > y) = P(C < T, C = y | T > y) \\ &= \frac{P(C < T, C = y, T > y)}{S_T(y)} = \frac{P(C < T, C = y)}{S_T(y)} \\ &= \frac{P(C = y, T > y)}{S_T(y)} = \frac{P(C = y)P(T > y)}{S_T(y)} \\ &= P(C = y) = f_C(y)dy. \end{aligned}$$

Also by (6.1) we have that

$$\begin{aligned} a(y) &= P(Y = T | T = y) = P(T \leq C | T = y) \\ &= \frac{P(T \leq C, T = y)}{P(T = y)} = \frac{P(C \geq y, T = y)}{P(T = y)} \\ &= \frac{P(T = y)P(C \geq y)}{P(T = y)} = P(C \geq y) = S_C(y). \end{aligned}$$

This implies $a(y) + B(y) = 1$ and thus an independent censoring model is a constant-sum model.

Under the assumption that T and C are independent, we then find that the likelihood function is given by

$$\prod_i \left(f_T(y_i)S_C(y_i)\right)^{1-\delta_i} \left(f_C(y_i)S_T(y_i)\right)^{\delta_i}.$$

If we further have that the distributions f_T and f_C are functionally independent, i.e. they have no common parameters, then inferences about the distribution of T can be based only on the likelihood (6.5). This setting is called independent censoring. As non-prognostic censoring and independent censoring models have the same form of likelihood, one might expect these models to be equivalent. In fact, independent censoring models are non-prognostic censoring models, but not vice versa, i.e. independent censoring models form a proper subset of non-prognostic censoring models [39]. In figure 6.1 we can see a visual representation of the types of censoring and which ones are subsets of each other. We see that independent censoring is a subset of non-prognostic censoring, which is a subset of constant-sum models which is complement to informative censoring.

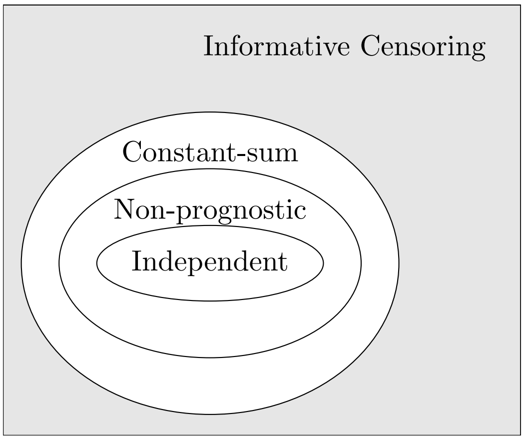

Figure 6.1: Diagram to represent the types of non-informative censoring, constant-sum, non-prognostic and independent, and their compliment informative censoring.

Each type of censoring model discussed in this section has their own unique worth in the analysis of survival data. Independent censoring models are commonly used in clinical trials and other settings where the failure and censoring times can be assumed independent. Non-prognostic censoring models are useful when the consideration of potential censoring times is not feasible. Non-informative censoring models characterised by the constant sum property are the most general of all these models and are equivalent to the Makeham assumption stated in definition 6.1. This type of model would be used when there is not enough evidence to suggest either independent or non-prognostic censoring. When the non-informative censoring assumption cannot be considered valid, an informative censoring model is required. It is difficult to determine from the data alone whether informative or non-informative censoring is appropriate. This identifiability issue is discussed in the next section.

6.2.1 Censoring as a missing data problem

Survival data is a particular type of missing data, in that the censoring mechanism is a type of missingness mechanism [12]. In the context of right censoring we know the range of the censored, or missing, value. In terms of missingness, the censoring can take one of these forms:

1. Censored completely at random (CCAR) which mimics missing completely at random (MCAR). This is where the censoring and event processes are completely independent. This is equivalent to independent censoring.

2. Censored at random (CAR) which mimics missing at random (MAR). In this mechanism, the censoring is independent of the failure conditional on some covariates. This is equivalent to non-informative censoring.

3. Censored not at random (CNAR) which mimics missing not at random (MNAR). Here the censoring and failure times are dependent. This is informative censoring.

6.3 Identifiability problem of informative censoring

There have been two main approaches to dealing with informative censoring. The first approach is to characterise the joint distribution of the event times and the censoring times. The second approach is to propose a model for the functions a and B, as in Lagakos and Williams [40]. It is the problems faced with the former approach that this chapter focuses on, and ways in which to overcome those said problems.

We have introduced the idea that given the data (Y, Δ), it is impossible to learn whether the T and C are independent or not, thus there is an identifiability issue. In this section this issue will be explained mathematically. However, first we will discuss the form of the joint distribution of (Y, Δ), $f_{Y,\Delta}(y, \delta)$.

6.3.1 The joint distribution of Y and Δ

In order to model the joint distribution of T and C we need to understand the forms the distribution can take and see how functions like the survival and cumulative hazard generalise to two variables.

We define the CDF of the joint distribution of T and C as

$$F_{T,C}(t, c) = P(T \leq t,\, C \leq c) = \int_0^t \int_0^c f_{T,C}(s, b)dbds.$$

We define the joint survival function as

$$S_{T,C}(t, c) = P(T > t,\, C > c) = \int_t^\infty \int_c^\infty f_{T,C}(s, b)dbds.$$

Note that the sum of the joint survival and CDF does not equal 1, i.e. $S_{T,C}(t, c) + F_{T,C}(t, c) \neq 1$ as seen in [77]. Thus $S_{T,C}$ and $F_{T,C}$ are not the only probabilities needed when discussing a bivariate distribution. This point is illustrated graphically in Figure 6.2 as seen by Yang and Nachlas [77]. We see that the probabilities $P(T \leq t,\, C > c)$ and $P(T > t,\, C \leq c)$ are required for the full characterisation of this distribution.

A key equation to note is the probability of being in the rectangle $[t_1 \leq T \leq t_2,\, c_1 \leq C \leq c_2]$. This equation is

$$P(t_1 \leq T \leq t_2,\, c_1 \leq C \leq c_2) = F_{T,C}(t_2, c_2) + F_{T,C}(t_1, c_1) - F_{T,C}(t_1, c_2) - F_{T,C}(t_2, c_1). \qquad (6.8)$$

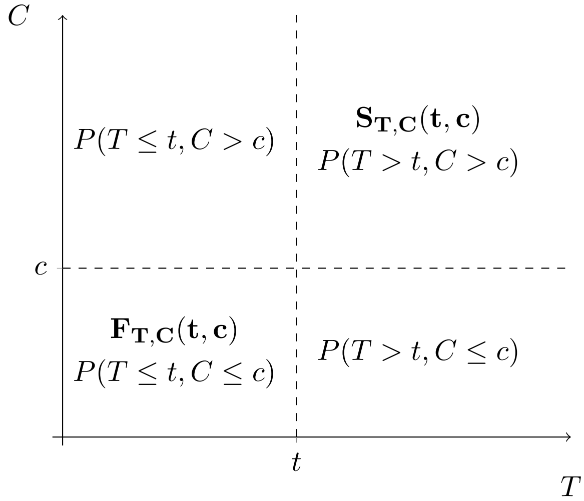

Figure 6.2: Graphical representation of bivariate probabilities.

A particularly useful case of equation (6.8) enables us to relate the joint survival function to the joint CDF and the marginal CDFs,

$$\begin{aligned} S_{T,C}(t,c) &= P(t \leq T < \infty,\, c \leq C < \infty) \\ &= 1 - F_C(c) - F_T(t) + F_{T,C}(t,c). \end{aligned} \tag{6.9}$$

From equation (6.9) we can deduce that

$$\frac{\partial^2}{\partial t \partial c} S_{T,C}(t,c) = f_{T,C}(t,c),$$

and thus a joint survival function will characterise a joint distribution.

Using the joint distribution induced by the joint survival, we are able to define the joint distribution of the observed time Y and the censoring indicator Δ. To calculate this distribution we use the fact that

$$f_{Y,\Delta}(y,\delta) = -\frac{\partial}{\partial y} P(Y > y, \Delta = \delta). \tag{6.10}$$

Since Δ can only take the values of 0 and 1, then the joint survival $S_{Y,\Delta}$ will be

$$\begin{aligned} P(Y > y, \Delta = \delta) &= \begin{cases} P(Y > y, \Delta = \delta), & \delta = 0 \\ P(Y > y, \Delta = \delta), & \delta = 1 \end{cases} \\ &= \begin{cases} P(T > y, T \leq C), & \delta = 0 \\ P(C > y, T > C), & \delta = 1. \end{cases} \end{aligned} \tag{6.11}$$

Thus we need to calculate the following probabilities

$$Q_T(y) = P(T > y, T \leq C) \text{ and } Q_C(y) = P(C > y, T > C). \tag{6.12}$$

These probabilities are called the crude survival functions of T and C evaluated at y. In the same vein, the quantities in the left hand side of definition 6.1 are called the crude hazard functions. A crude survival function is the probability of the event given that the event could be caused by some other mechanism, it is sometimes called the cause-specific survival function or the sub-survival function [23].

These probabilities can be calculated using

$$\begin{aligned}
Q_T(y) = P(T > y, T \leq C) &= \int_y^\infty \int_y^c f_{T,C}(t,c)dtdc \\
&= \int_y^\infty \int_y^c \frac{\partial^2}{\partial t \partial c} S_{T,C}(t,c)dtdc \\
&= \int_y^\infty \left[\frac{\partial}{\partial c} S_{T,C}(c,c) - \frac{\partial}{\partial c} S_{T,C}(y,c) \right] dc \\
&= \int_y^\infty \frac{\partial}{\partial c} S_{T,C}(c,c)dc - \left[S_{T,C}(y,c) \right]_{c=y}^{c=\infty} \\
&= S_{T,C}(y,y) + \int_y^\infty \frac{\partial}{\partial c} S_{T,C}(c,c)dc
\end{aligned}$$

and using a similar calculation we find

$$Q_C(y) = P(C > y, T > C) = S_{T,C}(y,y) + \int_y^\infty \frac{\partial}{\partial t} S_{T,C}(t,t)dt.$$

Thus using equation (6.10), we have that the joint distribution of Y and Δ is found by calculating

$$f_{Y,\Delta}(y,\delta) = \begin{cases} \frac{\partial}{\partial c} S_{T,C}(t,c)|_{t=y,c=y} - \frac{d}{dy} S_{T,C}(y,y), & \delta = 0 \\ \frac{\partial}{\partial t} S_{T,C}(t,c)|_{t=y,c=y} - \frac{d}{dy} S_{T,C}(y,y), & \delta = 1. \end{cases} \tag{6.13}$$

Now that we have outlines how the joint distribution of (Y, Δ) can be found, we will continue by explaining how either independent or dependent (T, C) can give the same joint distribution of (Y, Δ).

6.3.2 Identifiability issue

Tsiatis [70] proved that there is an identifiability issue when it comes to determining the dependence structure. This lack of identifiability is explained in terms of the crude and net survival functions. Crude and net survivals arise in the context of competing risks. In the setting of informative censoring, the censoring and event times are competing. As previously discussed crude survival of some event is the probability of this event occurring where other events could also occur, i.e. censoring. The net survival is the probability of the event in a hypothetical world where only this type of event can occur, i.e. there is no censoring.

The joint distribution of the event time T and the censoring time C uniquely defines the crude survival functions, $Q_T(t)$ and $Q_C(c)$, for T and C, see Tsiatis [70]. In our context, the crude survival functions are found by

$$Q_T(t) = -\int_t^\infty \frac{\partial}{\partial s} S_{T,C}(s,b)|_{s=s,b=s} ds \tag{6.14}$$

$$Q_C(c) = -\int_c^\infty \frac{\partial}{\partial b} S_{T,C}(s,b)|_{s=b,b=b} db, \tag{6.15}$$

and we note that these expression are equal to those as in (6.12). Similarly, the net survival functions are found by

$$Q_T^*(t) = S_{T,C}(t,0) = S_T(t)$$
$$Q_C^*(c) = S_{T,C}(0,c) = S_C(c).$$

Thus the net survival functions correspond to the marginal survival functions.

The following theorem demonstrates the identifiability issue in terms of the crude and net survivals.

Theorem 6.2. *Given any set of crude survival functions $Q_T(t)$ and $Q_C(c)$, there exist some net survival functions $\tilde{Q}_T^*(t)$ and $\tilde{Q}_C^*(c)$ corresponding to the case where T and C are assumed independent, and the corresponding crude survivals $\tilde{Q}_T(t)$ and $\tilde{Q}_C(c)$are such that $\tilde{Q}_T(t) = Q_T(t)$ and $\tilde{Q}_C(c) = Q_C(c)$.*

The proof of this theorem is given by Tsiatis [70]. Theorem 6.2 thus says that given some crude survivals, we cannot determine whether T and C were independent or not.

Theorem 6.2 thus states that the identifiability issue of making inferences about the dependence parameter arises from the fact that the likelihood obtained from a joint survival with dependence can be equivalent to a likelihood obtained from the joint survival with independence. This is demonstrated by the fact that the likelihood is written in terms of the crude survivals, as seen from (6.12) and (6.13). An example of this issue can be seen in the work of Tsiatis [70].

To combat this problem, we attempt to follow the procedure outlined by Siannis, Copas and Lu [63] and perform a sensitivity analysis on the dependence parameter. In these analyses we set the dependence parameter to be some value, or within some interval, and then perform likelihood inferences on the parameters of the event time T and the censoring time C. This will be discussed in detail later in section 6.4.2.

6.4 Approaches to informative censoring

In the literature there exist many approaches to how one should deal with the issue of informative censoring. A few of these approaches will be discussed in this section.

A natural response to the lack of identifiability is to attempt to place bounds on the joint and marginal survival functions. The more assumptions that are made, and the more information that becomes available allows the bounds to become tighter. These bounds allow one to learn about the joint behaviour of T and C without having to attempt to model it. Other approaches discussed in this section actually attempt to model said behaviour.

A popular approach is one that aims to model the conditional distribution of the censoring given the event time. Knowing this distribution would then enable us to ascertain the dependence structure and thus the joint distribution of T and C. This approach is discussed in section 6.4.2.

If we know the joint survival, we can define the joint distribution of T and C. Then we are able to define the joint distribution of $Y = \min(T, C)$ and $\Delta = I(T > C)$. Thus we would like to propose the form of the joint survival function. Suppose that the joint survival takes the form

$$S_{T,C}(t, c|\alpha) = \exp\left\{\varphi_\alpha(H_T(t), H_C(c))\right\}, \tag{6.16}$$

for some function $\varphi_\alpha(\cdot, \cdot)$. We would like that a certain value of α will give the independent case, i.e. if $\alpha = 1$, say, then

$$\begin{aligned} S_{T,C}(t, c|\alpha = 1) &= \exp\left\{\varphi_1(H_T(t), H_C(c))\right\} \\ &= \exp\left\{-(H_T(t) + H_C(c))\right\}. \end{aligned} \tag{6.17}$$

In section 6.4.3 we will detail the possible functions φ_α, either from the literature or propose new functions. We will provide three new different ways in which to specify the functions φ_α.

6.4.1 Bounds on the marginal survival function

We have previously discussed the identifiability issues and the fact that we are unable to determine whether there is some dependence between T and C. In this context, it may be of use to place bounds on the marginal survival function of the event time, and the censoring time.

Peterson [51] obtains sharp bounds on the marginal survival functions based on the crude survival functions. He finds that

$$Q_T(y) + Q_C(y) \leq S_T(y) \leq Q_T(y) + Q_C(0), \tag{6.18}$$

$$Q_T(y) + Q_C(y) \leq S_C(y) \leq Q_C(y) + Q_T(0). \tag{6.19}$$

He also shows that one can bound the joint survival function as follows

$$(Q_T + Q_C)(\max(t, c)) \leq S_{T,C}(t, c) \leq Q_T(t) + Q_C(c).$$

These bounds however, can be very wide.

Tighter bounds do exist in the literature, but they require additional information. For example, Slud and Rubinstein [65] develop bounds based on the function

$$\rho(t) = \lim_{\epsilon \to 0} \frac{P\left(t < T < t + \epsilon | T > t, C \leq t\right)}{P\left(t < T < t + \epsilon | T > t, C > t\right)}.$$

The bounds on the survival function require the investigator to bound the function $\rho(t)$, which may be possible with large enough samples.

Klein and Moeschberger [38] propose that the marginal survival function of $Y = \min(T, C)$ is given by

$$S_Y(y) = \left\{ \left[\frac{1}{S_T(y)} \right]^{\theta-1} + \left[\frac{1}{S_C(y)} \right]^{\theta-1} \right\}^{-1/(\theta-1)}.$$

The parameter θ is related to the concordance, which is a measure of how agreeable T and C are. The bounds on the marginal survival depend on the investigator being able to provide some bounds on the parameter θ. This would be possible given a sufficiently large sample. These bounds would be plugged into the survival function for Y and thus would provide the bounds on the survival function.

6.4.2 Approach based on the conditional distribution

In order to model T and C jointly, one could focus their attention on the form of the conditional distributions $f_{C|T}$ and $f_{T|C}$. If we could ascertain these distributions, we would know the dependence structure of the random variables T and C. The problem with this method is that, given the data, it will be impossible to determine whether there really should be any dependence between the two variables, as described in the identifiability issue shown before.

Siannis, Copas and Lu [63] propose a sensitivity analysis on a parameter that describes the level of dependence between T and C. The authors don't estimate this parameter, but set it to be a certain value, or within a certain range and learn about the sensitivity of the inferences on the joint distribution [62, 66].

They propose that the conditional distribution of C given T take the same form as the marginal distribution of C, but the parameter of this conditional distribution may depend on T. They assume the marginals of C and T depend on some parameter γ and θ respectively, i.e. $f_C(c|\gamma)$ and $f_T(t|\theta)$. They propose

$$f_{C|T}(c|t) = f_C\left(c|\gamma + \nu i_\gamma^{-\frac{1}{2}} B(t, \theta)\right),$$

where $B(t, \theta)$ is a bias function measuring the pattern of the dependence and $i_\gamma = \text{Var}\left(\frac{\partial}{\partial \gamma} \log f_C(c|\gamma)\right)$. Here if $\nu = 0$ then T and C are independent. This parameter, ν, is meant to measure the size of the dependence between T and C. The choice of the bias function $B(t, \theta)$ reflects the beliefs held about the dependence structure between

T and C. For example, if it is believed that censoring is more likely to occur for larger values of t, then an increasing bias function would be used. Thus choosing this function is key to constructing this model so that it accurately reflects the beliefs about the true model.

Using this conditional distribution and the distribution of T the joint distribution of T and C can then be found. Siannis et al. approximate the joint distribution as follows

$$\begin{aligned} f_{T,C}(t,c) &= f_T(t|\theta) f_C(c|\gamma + \nu i_\gamma^{-\frac{1}{2}} B(t,\theta)) \\ &\approx f_T(t|\theta) f_C(c|\gamma) \left[1 + \nu i_\gamma^{-\frac{1}{2}} s_C(c,\gamma) B(t,\theta) \right], \end{aligned}$$

where $s_C(c,\gamma)$ is the score function for the distribution of C. This then allows the formulation of the likelihood in which the independent case appears as a factor in the likelihood. They then propose a local sensitivity analysis on ν instead of conducting a full analysis with an aim to estimate ν. They fix ν to be within a range of values and use maximum likelihood to estimate θ and γ. It is then investigated whether the value of ν chosen had much impact on the estimates of the other parameters.

In section 6.5 we will perform a sensitivity analysis on the dependence parameter of models we fit to the liver transplant data set which aim to deal with informative censoring.

6.4.3 Approaches based on the cumulative hazard

Roy and Mukherjee [56] extend univariate survival distributions to a multivariate model. They propose a method for the joint modelling of k survival times. Here, we propose to adapt their methodology to jointly model the event and censoring times, hence $k = 2$. Their model proposes that the joint survival function of T and C is

$$S_{T,C}(t,c) = \exp\left(- \left[H_T(t)^\alpha + H_C(c)^\alpha \right]^{1/\alpha} \right), \tag{6.20}$$

where $\alpha \geq 1$. Notice here, that if $\alpha = 1$ then we recover the independent case (6.17). Recall that this model was first introduced in section 3.1.8 as a univariate model that depended on two CH functions H_A and H_B that were both functions of t, in equation (3.18). Here we have extended this to a bivariate model, by replacing one of the CHs with a CH for another variable, c.

In their paper, Roy and Mukherjee show that if the variables all have increasing hazard functions then so does the joint distribution. The same is true if each variable is of Increasing Hazard Rate Average (IHRA) class as seen in section 2.3.9. We note that these properties do rely on the condition that $\alpha \geq 1$.

Thanks to our investigation of the analytical properties of cumulative hazard functions in Chapter 2, we know that powering and summing CH functions preserves the

CH function properties. Thus it is easy to identify that the equation (6.20) will in fact define a joint survival function.

Noting the form of the joint survival function (6.20), we see that the cumulative hazard is a function of other, known cumulative hazards, namely the Weibull CH, H_T and H_C. Thus, we aim to generalise this model to give us the flexibility we desire. We propose that the joint survival function of T and C be of the form

$$S_{T,C}(t,c) = \exp\left\{-H_\alpha^{-1}\left[H_\alpha(H_T(t)) + H_\alpha(H_C(c))\right]\right\}, \qquad (6.21)$$

where H_α is a cumulative hazard function with parameter α.

The survival function in (6.21) is a generalisation of (6.20) since the Weibull CH has been replaced with some more general CH H_α. Thus letting $H_\alpha(x) = x^\alpha$ we recover (6.20). Note that this was similar to the model proposed in section 3.1.8 in equation (3.17), the linear-composition combination. It is a bivariate extension of the model in (3.17), just as (6.20) was a bivariate extension of (3.17). In this model the choice of H_α will reflect the beliefs of the dependence structure of T and C. This function is thus key to the use of this model. In order to recover the independent case, for some $\alpha = \tilde{\alpha}$, we require that $H_{\tilde{\alpha}} = H_E^1$, i.e.

$$H_{\tilde{\alpha}}^{-1}\left[H_{\tilde{\alpha}}(H_T(t)) + H_{\tilde{\alpha}}(H_C(c))\right] = H_T(t) + H_C(c).$$

Here H_α plays a similar role to that of $\nu i_\gamma^{-\frac{1}{2}} B(t,\theta)$ in the Siannis et al. approach.

Our approach is similar to the dependence modelling via Archimedean copula [71]. In these models the joint survival of T and C is given by

$$S_{T,C}(t,c) = \phi^{-1}\left(\phi(S_T(t)) + \phi(S_C(c))\right),$$

where $\phi : [0,1] \to [0,\infty)$ is a convex, decreasing function with $\phi(1) = 0$ and $\lim_{x\to 0}\phi(x) = \infty$.

Proposition 6.3. *Our model proposed in* (6.21) *is equivalent to the Archimedean copula model where* $\phi(x) = H_\alpha(-\log(x))$, *when* H_α *is convex.*

Proof. If $\phi(x) = H_\alpha(-\log(x))$ and H_α is convex, then

$$\begin{aligned} S_{T,C}(t,c) &= \phi^{-1}\left(\phi(S_T(t)) + \phi(S_C(c))\right) \\ &= \exp\left\{-H_\alpha^{-1}\left[H_\alpha(-\log(S_T(t))) + H_\alpha(-\log(S_C(c)))\right]\right\} \\ &= \exp\left\{-H_\alpha^{-1}\left[H_\alpha(H_T(t)) + H_\alpha(H_C(c))\right]\right\}, \end{aligned}$$

which is the model given in (6.21). □

The Archimedean copula model requires that ϕ must be convex. This does not imply H_α is convex. Our model is equivalent to the Archimedean copula model.

Letting $H_\alpha(x) = e^{\alpha x} - 1$ in equation (6.21), we recover the joint distribution in Klein and Moeschberger [38] which was an adaptation of a model proposed by Clayton [16]. This model is of the form

$$S_{T,C}(t,c) = \exp\left\{-\frac{1}{\alpha}\log\left(e^{\alpha H_T(t)} + e^{\alpha H_C(c)} - 1\right)\right\}. \tag{6.22}$$

Note that as $\alpha \to 0$ we recover the independent case. Notice that this CH was seen in example 3.37, there $\tilde{H} = H_G$. Note that since $\log\left(e^{\alpha H_T(t)} + e^{\alpha H_C(c)} - 1\right) = \log\left(1 + [e^{\alpha H_T(t)} - 1] + [e^{\alpha H_C(c)} - 1]\right)$ then the argument inside the logarithm will always be greater than one, and thus the whole expression will always be positive.

We see that particular choices of H_α, thus far the Weibull and the Gompertz, allow us to reproduce bivariate survival models from within the literature. Further to these examples, the archimedean model generated by $H_\alpha(x) = \log(1 + \alpha x)$ can be seen in Nelsen's book in copulas [50]. Another example from this book is the Ali-Mikhail-Haq family, which is generated by $H_\alpha(t) = H^\alpha_{llEG}(t)$. The choice of H_α can introduce specific short and long term behaviours that we require. For example, choosing $H_\alpha(x) = e^{\alpha x} - 1$, we have that in the short term this family acts like the independent case. We also have that choosing $H_\alpha = H^\alpha_{llEG}$, then in the long term this family would act like the independent case. Using the Weibull H_α, we ensure Weibull type long and short term behaviours.

The event time and the censoring time can be viewed as competing risks. If we were to have further risks, we may need a multivariate survival model, i.e. we may need a model for more than two competing risks. This extension would be quite simple to construct thanks to the simple construction of our model (6.21). If we have p competing risks, T_1, ... T_p, our multivariate model may be given by the joint survival

$$S_{T_1,\dots,T_p}(t_1,\dots,t_p) = \exp\left\{-H_\alpha^{-1}\left[H_\alpha(H_{T_1}(t_1)) + \dots + H_\alpha(H_{T_p}(t_p))\right]\right\}.$$

Here there are further restrictions required on H_α, but these will not be explored as this is not the focus of this thesis and we will continue with only two competing risks, the event time and the censoring time.

Another way to create a bivariate survival model is to introduce a common frailty as seen in section 3.3.3. Then

$$H_{T,C}(t,c) = H^\alpha_{(F)}(H_T(t) + H_C(c))$$

for some mixing CDF F. For this model we require that $H^\alpha_{(F)}$ must be a Bernstein function. In this model we would want that for some $\alpha = \tilde{\alpha}$, $H_{(F)} = H_{(D_\mu)} = H^1_E$ so that we can recover the independent case. For example, we could let $H_{(F)}(t|\alpha) = \frac{t}{\sqrt{1+\alpha t}}$, as in example 3.77. The parameter α is usually, but not always, the variance of the mixing distribution.

More generally we could generate a bivariate survival model such as

$$H_{T,C}(t,c) = H_\alpha(H_T(t) + H_C(c)),$$

for some H_α such that, for some $\alpha = \tilde{\alpha}$, $H_\alpha = H_E^1$. Thus we can recover the independent case. We see that this would be a generalisation of the bivariate frailty mixture, as H_α would not need to be Bernstein. An example of this would be $H_\alpha = H_{GEll}^\alpha$ for $\alpha \geq 1$.

With our new bivariate survival models, we can calculate the joint distribution of the observed time and the censoring indicator, i.e. we can calculate $f_{Y,\Delta}(y,\delta)$. Then we can fit this model to data and analyse the fit and perform a sensitivity analysis on the parameter that controls the dependence, α. The next section aims to do this with the liver transplant data set described in chapter 5.

6.5 Informative censoring in the liver transplant data set

In chapter 5 the presence of informative censoring in the liver transplant data set was discussed. Patients who are more ill, are in more need of a transplant, and thus more likely to receive one. However, patients who are more ill are also more likely to die. Thus we see that death and censoring due to transplant are dependent.

We aim to fit the models discussed in the previous section, described in equation (6.21), with $H_\alpha(x) = e^{\alpha x} - 1$, followed by $H_\alpha(x) = x^\alpha$, to the liver transplant data and perform a sensitivity analysis on α. Note that when $H_\alpha(x) = e^{\alpha x} - 1$, the independent case occurs when $\alpha \to 0$ and thus is on the edge of the parameter space, whereas for $H_\alpha(x) = x^\alpha$ the independent case corresponds to $\alpha = 1$ which is not on the edge of the parameter space. These choices of H_α provide a range of short and long term behaviours. The choice of the Gompertz H_α ensures an independent type short term behaviour, and the Weibull ensures a Weibull type short term behaviour.

For simplicity, we assume that the cumulative hazard functions for death and for transplant are

$$H_T(t|\theta) = \theta t,$$
$$H_C(c|\gamma) = \gamma c.$$

Simple forms for H_T and H_C are desirable since anything flexible with lots of parameters will exacerbate the identifiability issues we are already dealing with. However, give the form of the cumulative hazards of H_T and H_C in figure 5.5, we see that perhaps a better choice would be perhaps Weibull or log-logistic.

First we will perform a sensitivity analysis without covariates, and then will repeat the analysis with the covariates in the proportional hazards models discussed in section 5.2.

In a sensitivity analysis, we fix a value of α and then find the maximum likelihood

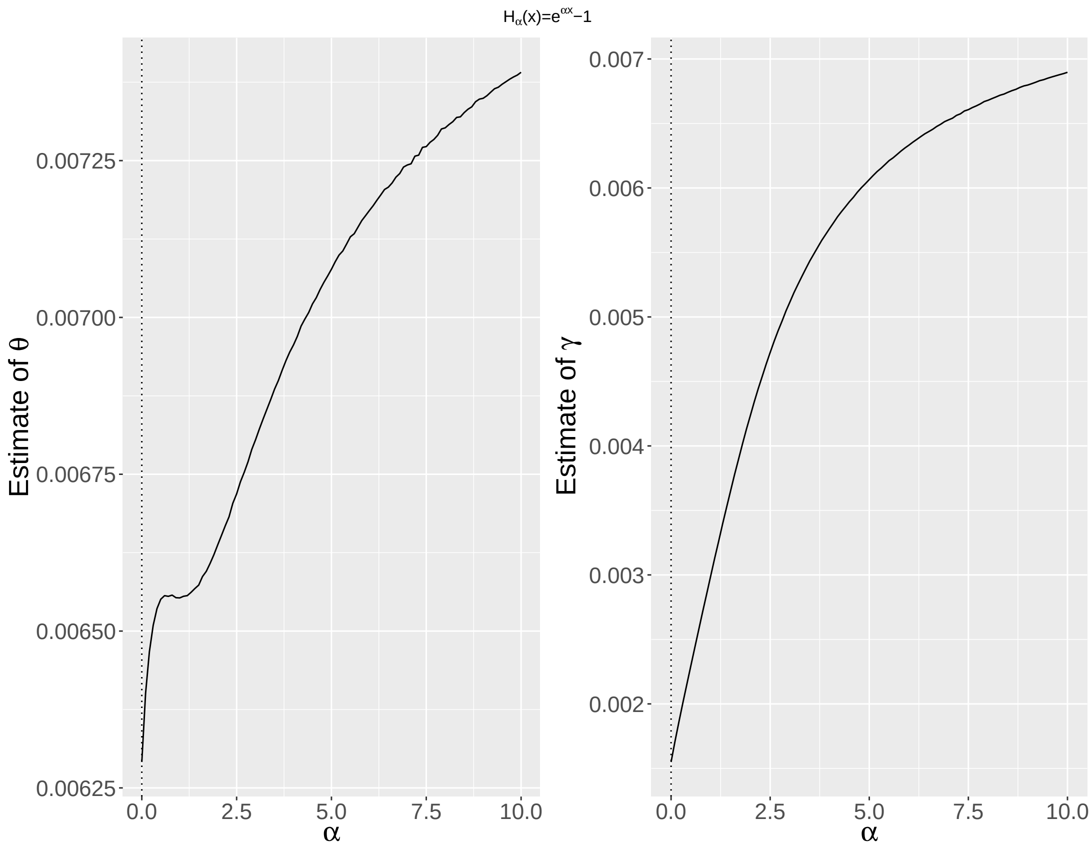

Figure 6.3: Estimates of θ and γ, the parameters of the CH functions for T and C, plotted against the value of α used in the analysis. Here $H_\alpha(x) = e^{\alpha x} - 1$. The dotted line corresponds to the independence case.

estimates of θ and γ. We then repeat this process for a different value of α. In figure 6.3 we see the estimates of θ and γ from H_T and H_C plotted against α. It appears that the estimates are fairly dependent on the value of α, but if we look closely at the scale of the range, we see the range is very small being less than 0.005. We thus might conclude that the estimates of both parameters are not overly dependent on the value of α. The dotted line refers to the independent case, thus a local sensitivity analysis would focus around this line.

We can make similar conclusions about the sensitivity analysis performed with $H_\alpha(x) = x^\alpha$. The estimates of θ and γ are plotted against α in figure 6.4. We again see that the estimates do not vary too much with α, the ranges being less than 0.01. We see around the dotted line, which refers to the independent case, there is little variability in α.

Although we have seen the parameters themselves do not vary much given the value of α, it would be interesting to see what effect this has on the net survival functions of T and C, seen in figures 6.5 and 6.6. Here the net survivals are plotted for two values of $\hat{\theta}_\alpha$ and $\hat{\gamma}_\alpha$, the estimates of θ and γ corresponding to particular α values. This is to investigate whether the choice of α effects the estimate of the survival. In figure 6.5

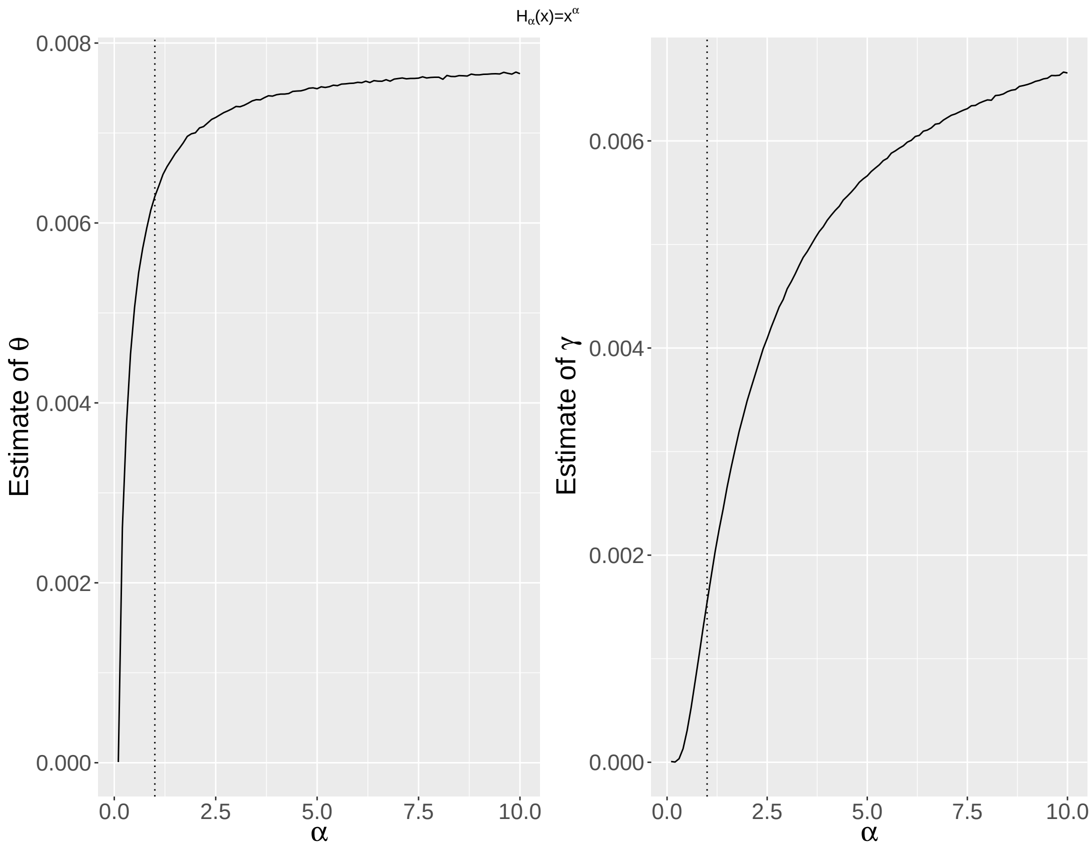

Figure 6.4: Estimates of θ and γ, the parameters of the CH functions for T and C, plotted against the value of α used in the analysis. Here $H_\alpha(x) = x^\alpha$. The dotted line corresponds to the independent case.

we see that the net survival of T does not change very much given the value of α, it is thus not sensitive to this choice. However, in figure 6.6 we see the opposite. The net survival of C is very dependent on the choice of α. In these figures we have chosen α to be 0 and 4.9. $\alpha = 0$ was chosen because this corresponds to the independent case. $\alpha = 4.9$ was chosen because this corresponds to a dependent case. Note that these values were chosen and not estimated from the data

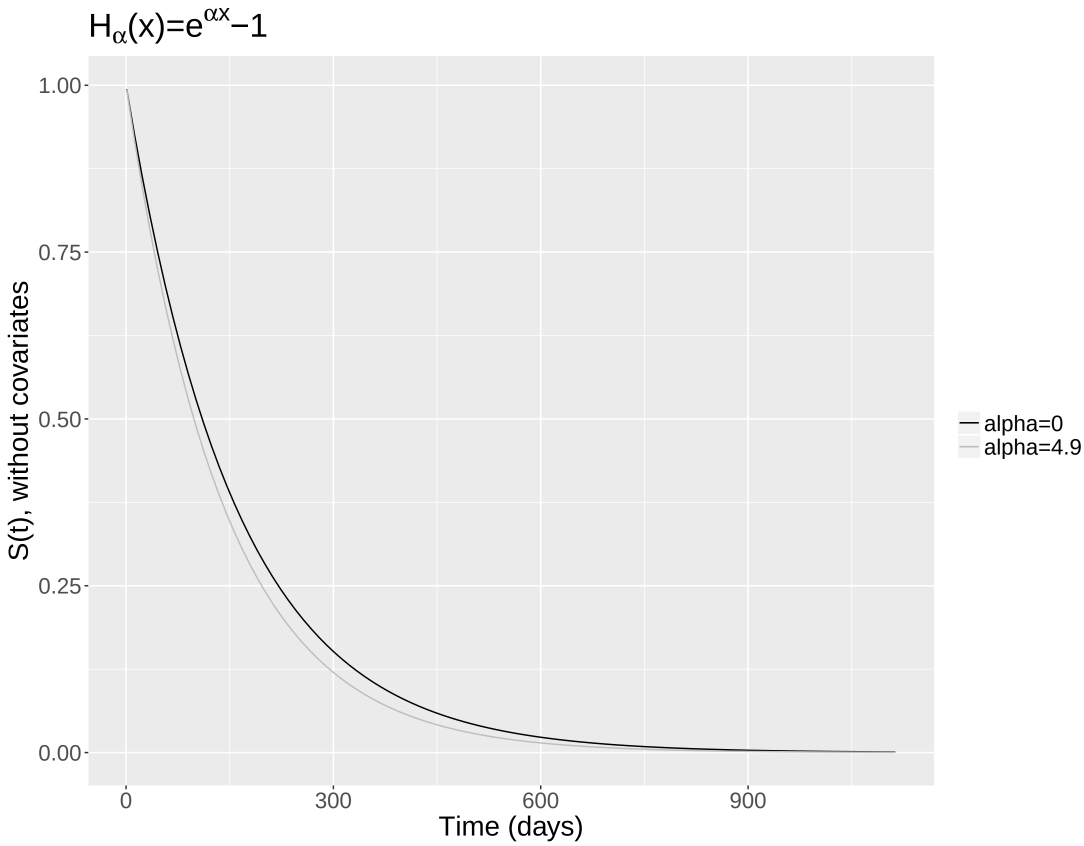

Figure 6.5: The estimated net survival function $S_T(t) = e^{-\hat{\theta}_\alpha t}$, where θ is estimated from the joint model described in equation (6.22) for fixed α. Here $\alpha = 0$ is the black line (and independent case) and $\alpha = 4.9$ is the grey line (an example of a dependent case). There were no covariates in this model.

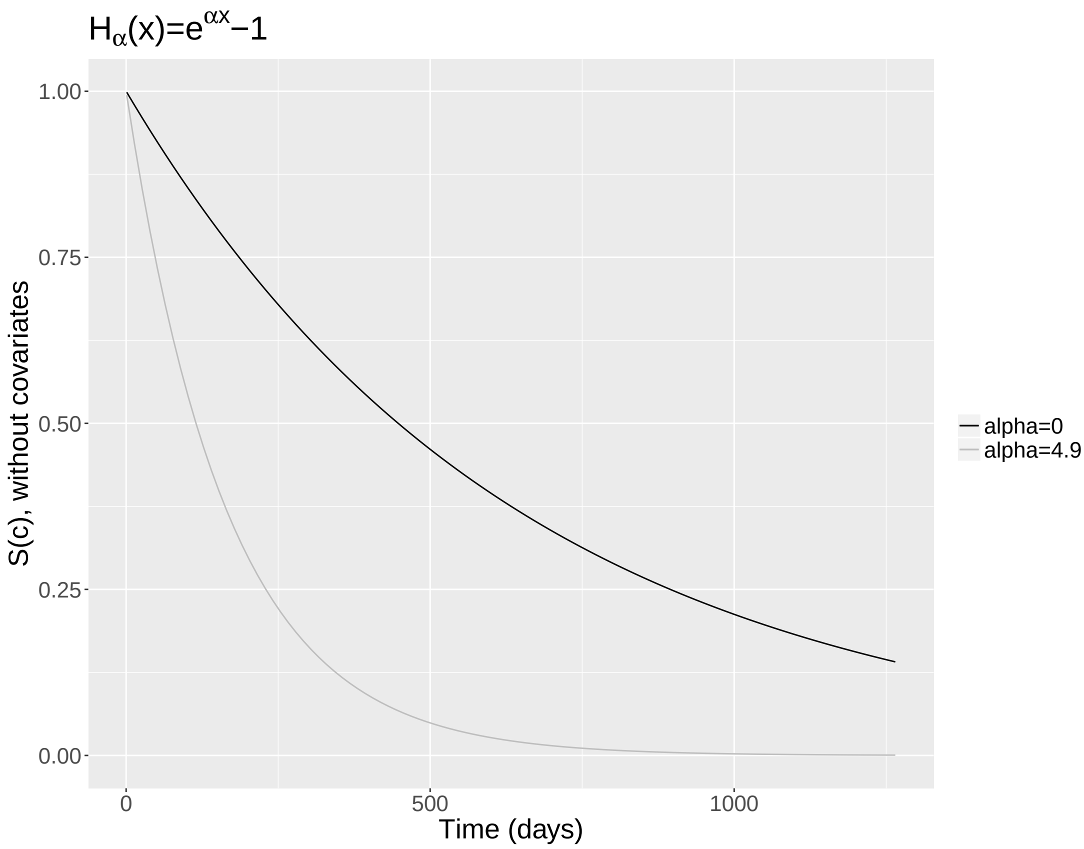

Figure 6.6: The estimated net survival function $S_C(c) = e^{-\hat{\gamma}_\alpha c}$, where γ is estimated from the joint model described in equation (6.22) for fixed α. Here $\alpha = 0$ is the black line (and independent case) and $\alpha = 4.9$ is the grey line (an example of a dependent case). There were no covariates in this model.

To investigate further, and to allow a better comparison, we include covariates into the model. We include the covariates that were deemed significant in the Cox models in section 5.2, i.e. UKELD score, patient age and disease category for the event T and only UKELD score and disease category for the censoring C. These were included in the model by allowing

$$\log(\theta) = \theta_0 + \theta_1 \texttt{UKELD} + \theta_2 \texttt{recip\_age} + \theta_3 \texttt{disease},$$
$$\log(\gamma) = \gamma_0 + \gamma_1 \texttt{UKELD} + \gamma_2 \texttt{disease}.$$

The analysis was then performed using the model in (6.22) with $H_\alpha(x) = e^{\alpha x} - 1$ again, for $\alpha \in [0, 5]$. Then the net survivals are plotted in figures 6.7 and 6.8. These appear to be very similar to those in figures 6.5 and 6.6 where no covariates were included in the model. We do however see that the differences between the the two curves in each plot has become larger. Thus now, there is a greater dependence on α, however slight. Again, we see that the dependence is more pronounced for the censoring than for the event times.

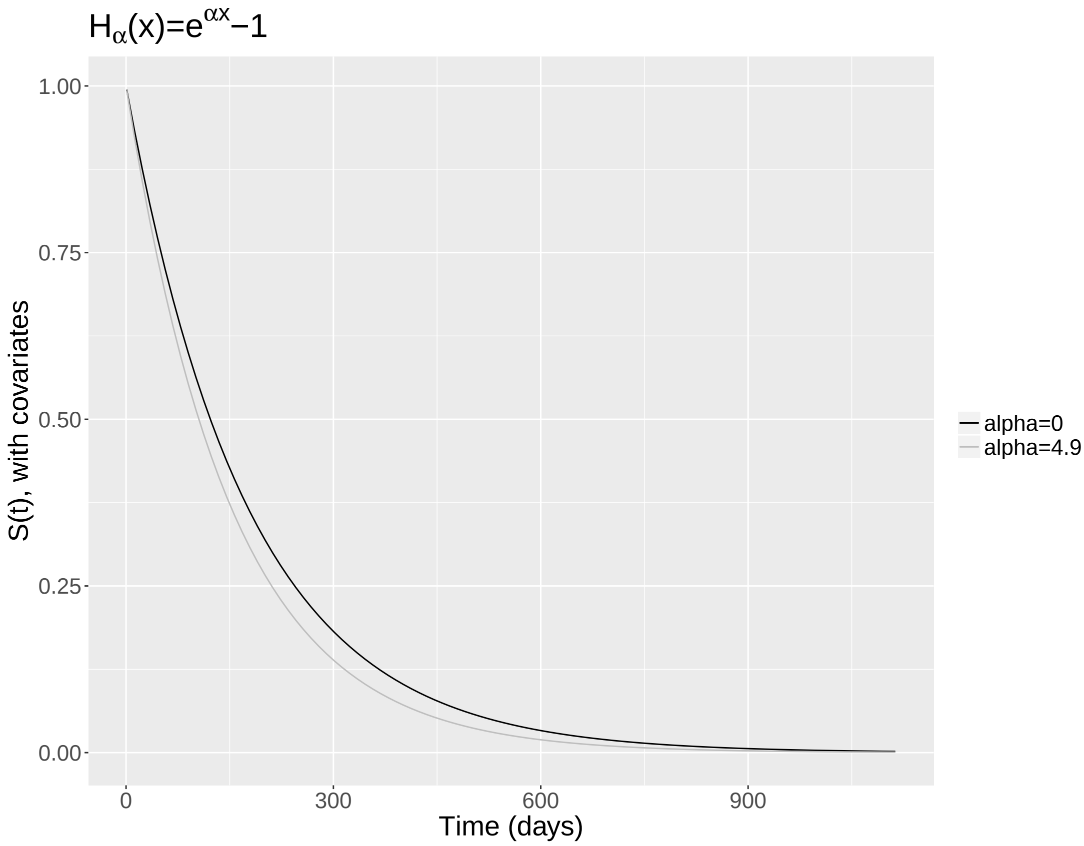

Figure 6.7: The estimated net survival function $S_T(t) = e^{-\hat{\theta}_\alpha t}$, where $\hat{\theta}_\alpha$ is an estimate from the joint model described in equation (6.22) for fixed α. In this plot, $\alpha = 0$ is the black line (and independent case) and $\alpha = 4.9$ is the grey line (an example of a dependent case). The covariates included in this model were UKELD score (49.2), patient age (46) and disease category (5 =HCV).

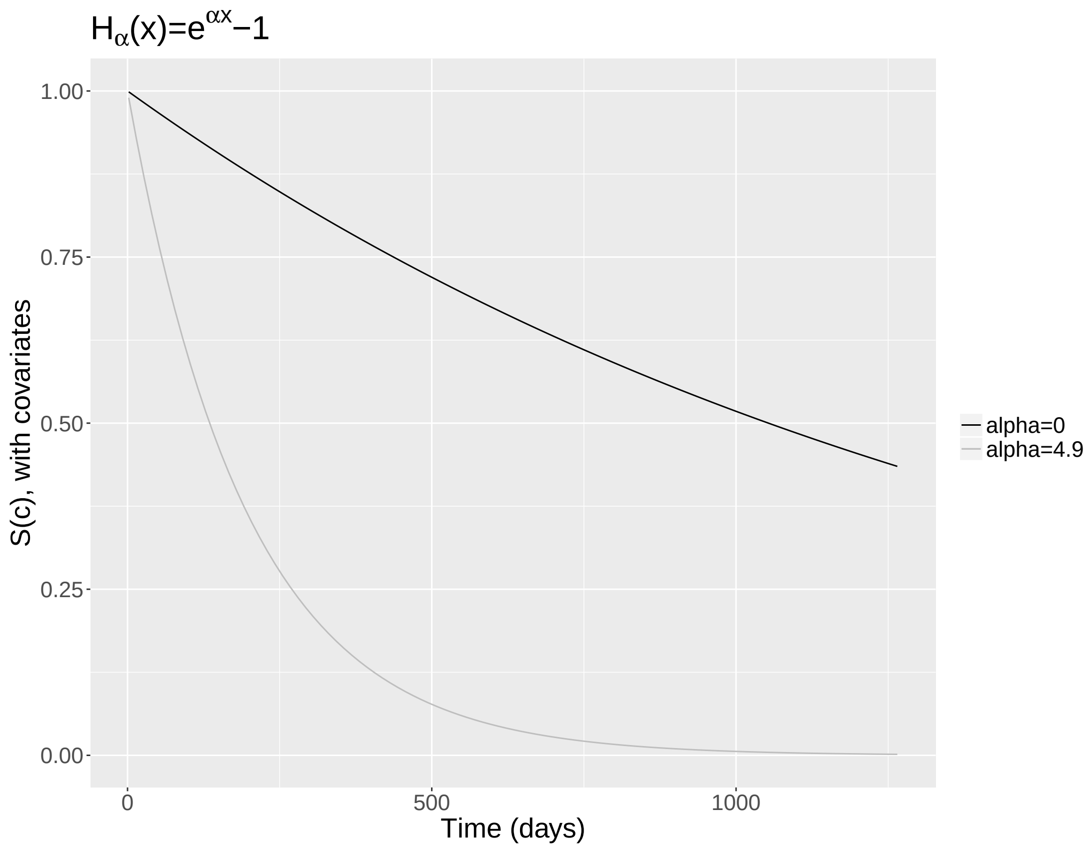

Figure 6.8: The estimated net survival function $S_C(c) = e^{-\hat{\gamma}_\alpha c}$, where $\hat{\gamma}_\alpha$ is an estimate from the joint model described in equation (6.22) for fixed α. In this plot, $\alpha = 0$ is the black line (and independent case) and $\alpha = 4.9$ is the grey line (an example of a dependent case). The covariates included in this model were UKELD score (49.2) and disease category (5 =HCV).

We then repeat the above analysis with $H_\alpha(x) = x^\alpha$ with and without covariates. Figures 6.9 and 6.10 show the net survivals for this model without covariates. We include three values of α. We chose $\alpha = 1$ as this is the independent case for this model. We then chose two other values of α, one smaller than 1 and another larger than 1. We see that the survival for T is less sensitive to α than the survival for C.

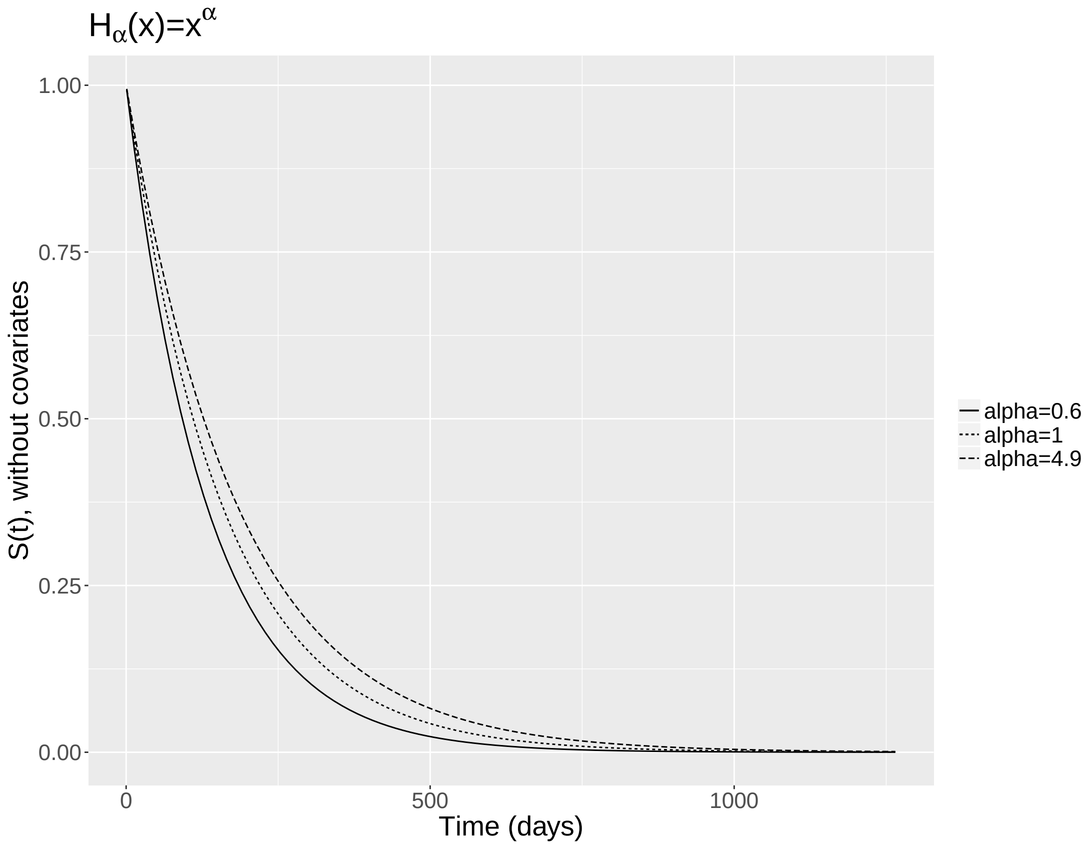

Figure 6.9: The estimated net survival function $S_T(t) = e^{-\hat{\theta}_\alpha t}$, where $\hat{\theta}_\alpha$ is an estimate from the joint model described in equation (6.22) for fixed α. In this plot, $\alpha = 1$ is the dotted line (and independent case), $\alpha = 0.6$ is the solid line (an example of a dependent case) and $\alpha = 4.9$ is the dashed line (another example of a dependent case). There were no covariates in this model.

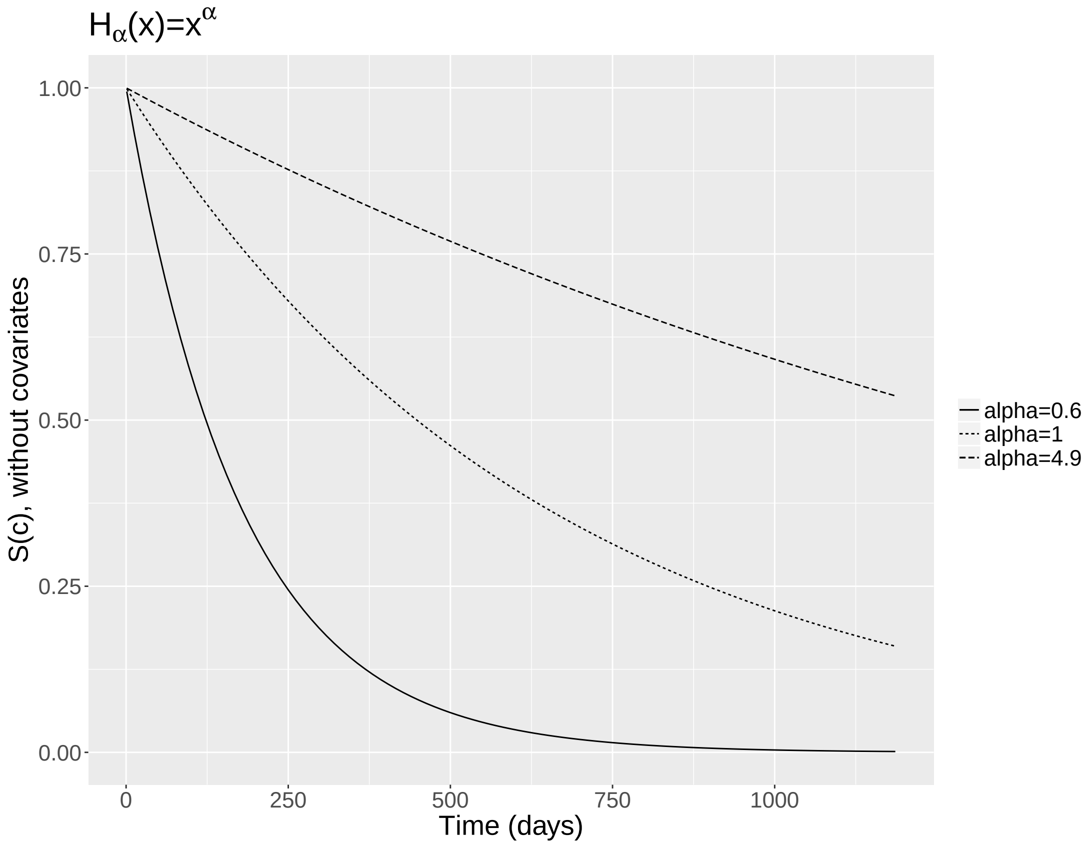

Figure 6.10: The estimated net survival function $S_C(c) = e^{-\hat{\gamma}_\alpha c}$, where $\hat{\gamma}_\alpha$ is an estimate from the joint model described in equation (6.22) for fixed α. In this plot, $\alpha = 1$ is the dotted line (and independent case), $\alpha = 0.6$ is the solid line (an example of a dependent case) and $\alpha = 4.9$ is the dashed line (another example of a dependent case). There were no covariates in this model.

In figures 6.11 and 6.12 we see the net survivals of the model with covariates. We see that the net survival of T in figure 6.11 is not very sensitive to values of $\alpha \geq 1$, but is for values close to 0. Here, values of α close to 0 imply the survival function is given by a constant, thus the analysis is unstable for small values of α. In figure 6.12 there is a very great dependence on the value of α. We see that very small values of α yield a survival function close to being constant, which makes very little sense. Other values of α give very similar results to the analysis performed with $H_\alpha(x) = e^{\alpha x} - 1$.

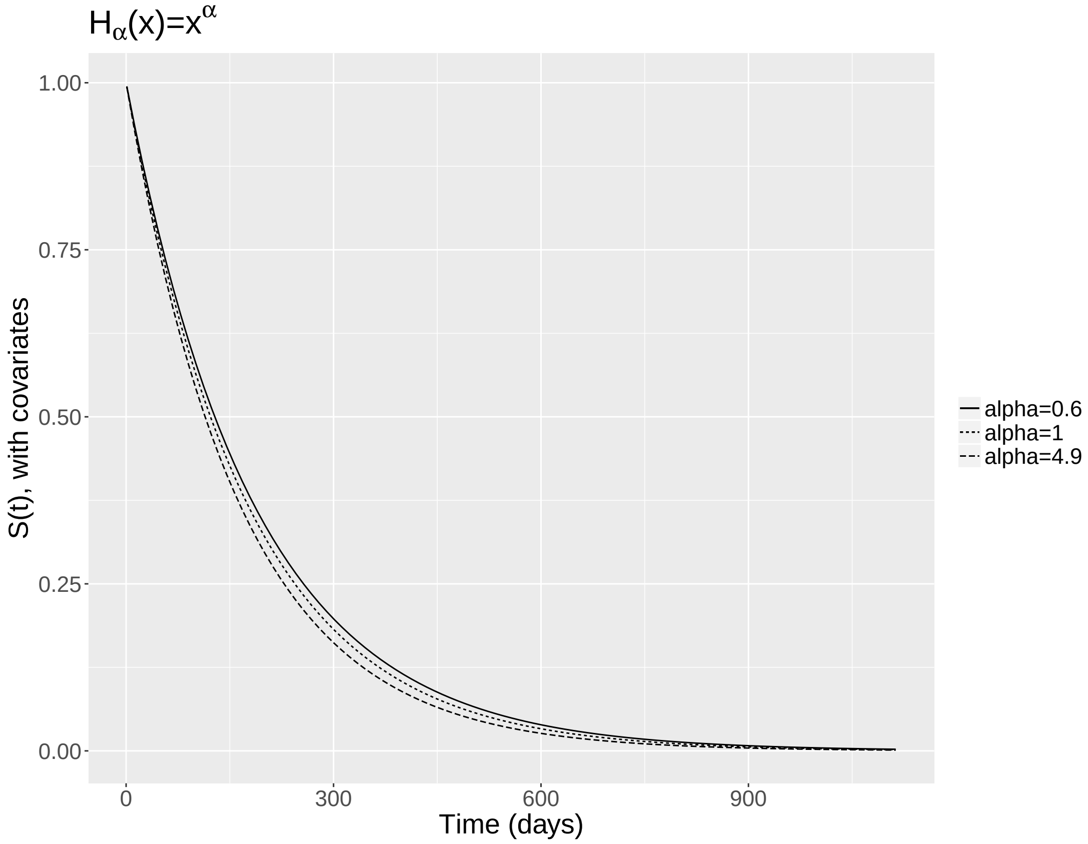

Figure 6.11: The estimated net survival function $S_T(t) = e^{-\hat{\theta}_\alpha t}$, where $\hat{\theta}_\alpha$ is estimated from the joint model described in equation (6.20) for fixed α. In this plot $\alpha = 1$ is the dotted line (and independent case), $\alpha = 0.6$ is the solid line (an example of a dependent case) and $\alpha = 4.9$ is the dashed line (another example of a dependent case). The covariates included in this model were UKELD score, patient age and disease category.

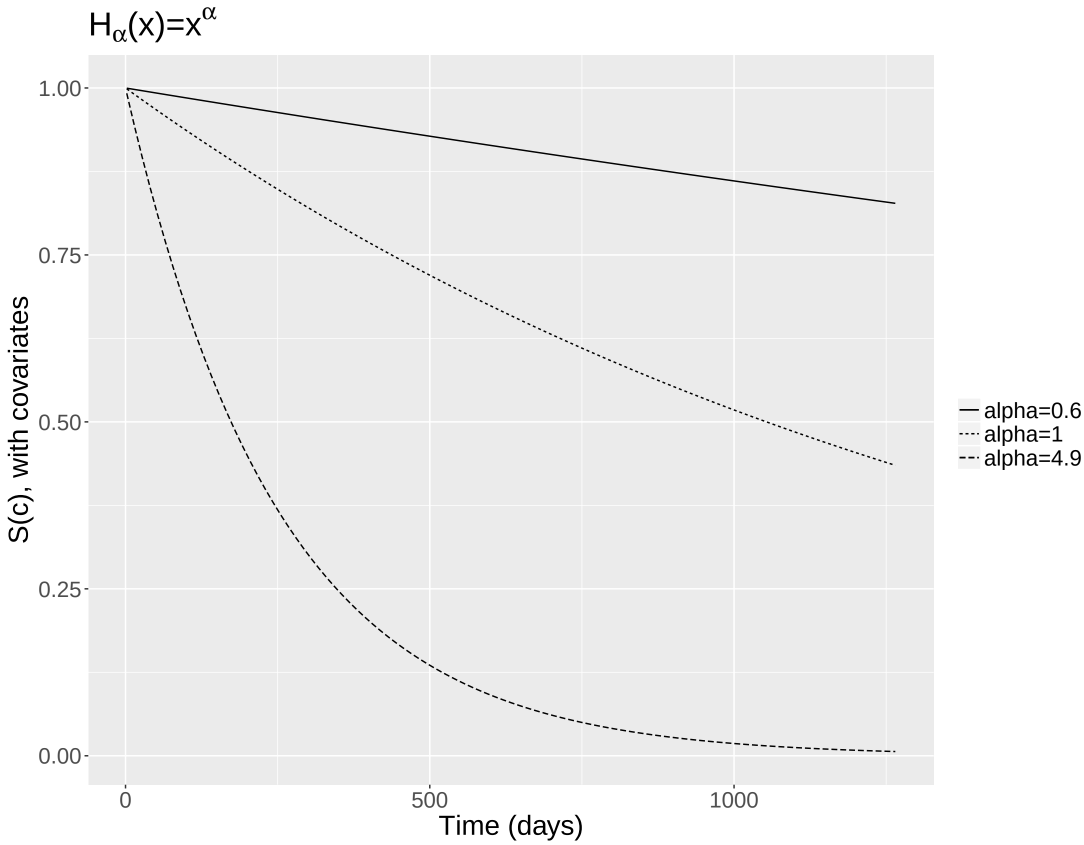

Figure 6.12: The estimated net survival function $S_C(c) = e^{-\hat{\gamma}_\alpha c}$, where $\hat{\gamma}_\alpha$ is an estimate from the joint model described in equation (6.20) for fixed α. In this plot, $\alpha = 1$ is the dotted line (and independent case), $\alpha = 0.6$ is the solid line (an example of a dependent case) and $\alpha = 4.9$ is the dashed line (another example of a dependent case). The covariates included in this model were UKELD score and disease category.

From these analyses we can conclude that in this form of model the censoring mechanism depends to some degree on the dependence between T and C, but the event time mechanism is influenced less by this dependence. Since we are usually interested in the event time mechanism more, time to death usually being of more interest, we may decide that the sensitivity of C is not too important. This conclusion is supported by the conclusions found in the sensitivity analysis from Staplin [66].

As well as determining if the estimates of the model parameters are sensitive to the choice of dependence parameter, it is also useful to know if the model actually fits the data. To investigate this we first separate the data into the deaths and the transplants, i.e. into events and censored observations. We then plot a Kaplan-Meier estimate of the survival for each group and compare to the net survival estimated by the model. Here, the Kaplan-Meier estimate would estimate the net survivals. We see this for the model with $H_\alpha(x) = e^{\alpha x} - 1$, with and without covariates in figures 6.14 and 6.13. We see that the net survival estimated from the models are very similar to the Kaplan-Meier estimates. Similar results can be seen for the model with $H_\alpha(x) = x^\alpha$.

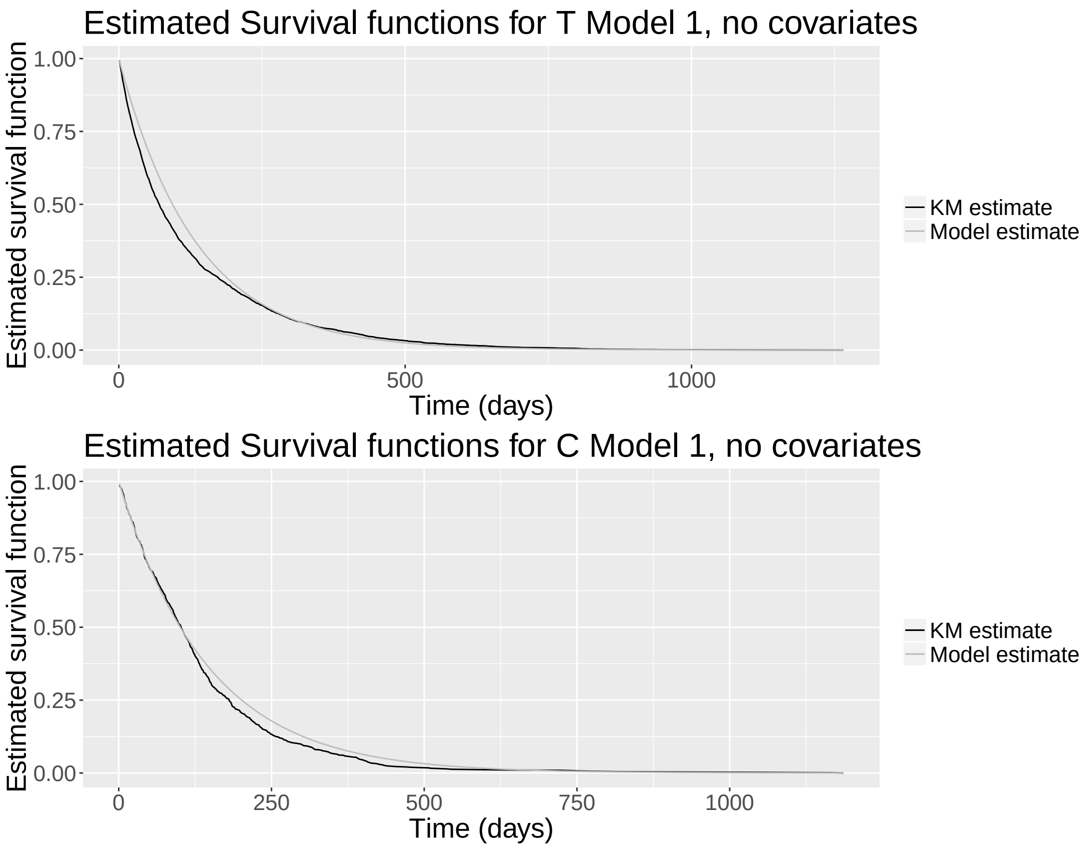

Figure 6.13: The estimated net survival functions for the model without covariates with $H_\alpha(x) = e^{\alpha x} - 1$. Here $\alpha = 9.9$.

We will also compare the marginal survival of $Y = \min(T, C)$ from both models with the usual Kaplan-Meier estimate, as seen in figures 6.15 and 6.16. We compare the Kaplan-Meier estimate with the model fitted using covariates and the one without,

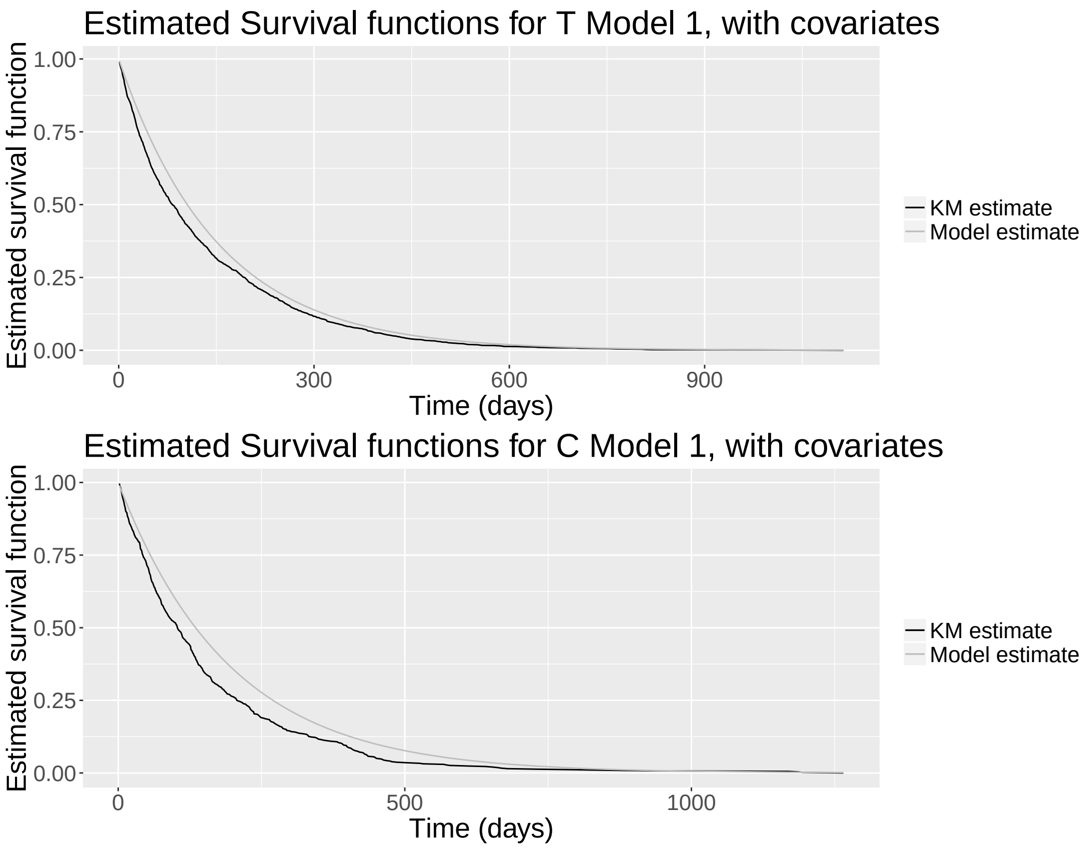

Figure 6.14: The estimated net survival functions for the model with covariates with $H_\alpha(x) = e^{\alpha x} - 1$. Here $\alpha = 4.9$.

for each model with $H_\alpha(x) = e^{\alpha x} - 1$ in 6.15 and $H_\alpha(x) = x^\alpha$ in 6.16. We see that for smaller survival times, less than 200 days, there exists some value of α for each model that predicts survival well. For larger survival times our models give more conservative estimates of survival. In terms of prediction, a conservative estimate is preferred.

6.6 Summary and future work

At the end of section 6.4.3 we suggested how our models might be extended to the case where we have more than two competing risks. We would like to explore the fitting of these models and compare them to those that already exist in the literature, for example compare to those of Roy and Mukherjee [56]. We would also like to study more closely the type of copulas generated by our multivariate models.

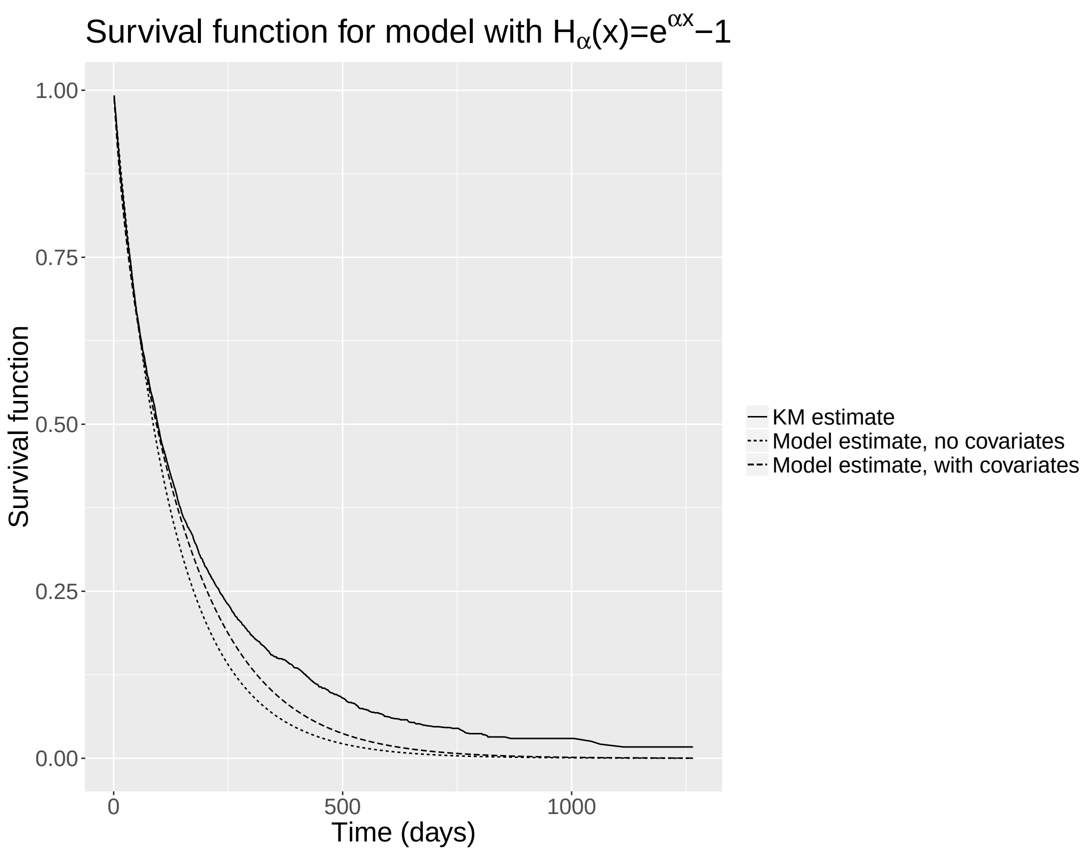

Figure 6.15: Estimated survival functions for the model with $H_\alpha(x) = e^{\alpha x} - 1$. Model with no covariates has $\alpha = 0.1$ (dotted line), model with covariates $\alpha = 4.9$ (dashed line).

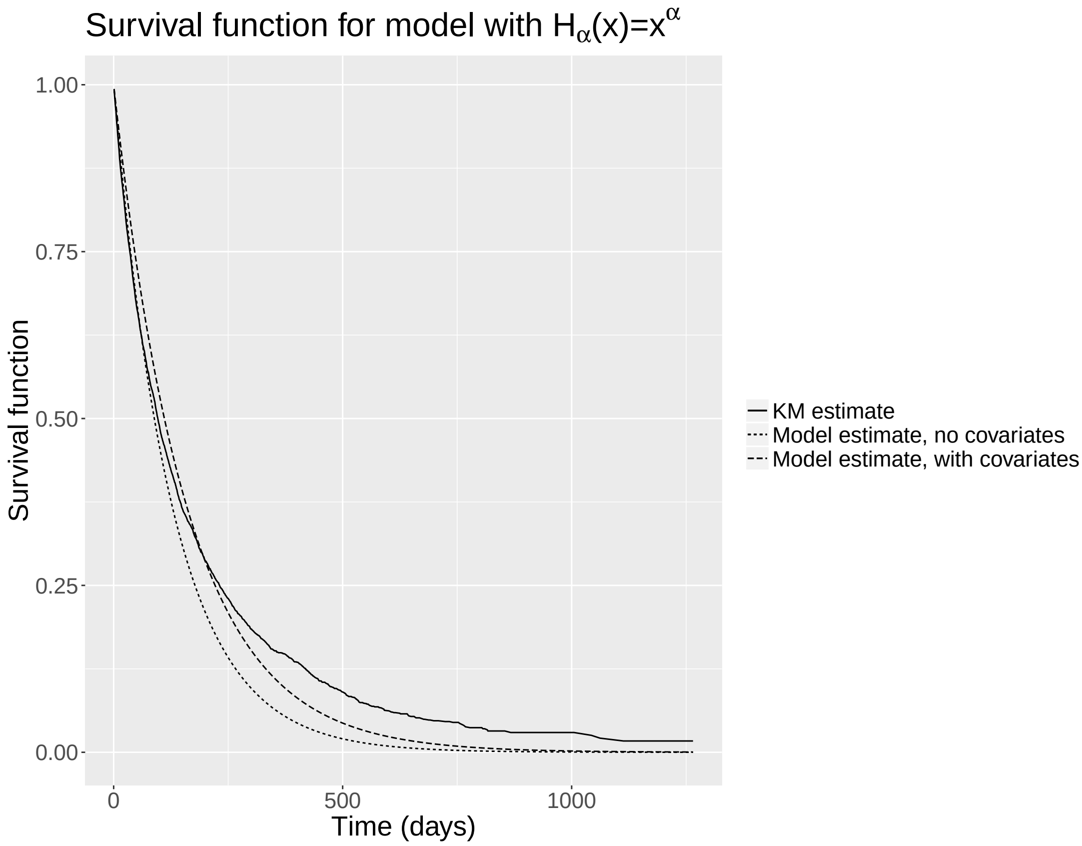

Figure 6.16: Estimated survival functions for the model with $H_\alpha(x) = x^\alpha$. Model with no covariates has $\alpha = 0.1$ (dotted line), model with covariates $\alpha = 10$ (dashed line).

Chapter 7

Discussion and Future Work

The aim of this thesis was to provide a framework for using the cumulative hazard function as a modelling tool for time-to-event data. This thesis has demonstrated how the study of this function can simplify the form of a parametric model, thus making the construction of new models very straightforward. This thesis has also demonstrated the use of cumulative hazard functions as a tool for modelling multivariate time-to-event models.

7.1 Discussion

Chapter 2 discussed a collection of properties of cumulative hazard functions. More specifically, this chapter gave details on the types of operations that could be performed on a cumulative hazard, so that one would stay within the set of smooth cumulative hazards. It was found that the set of cumulative hazards was closed under composing, inverting, taking the product, summing and integrating. Of particular interest were compositions and inverses of cumulative hazards, which allow one to get from one cumulative hazard in the set to another, but also to describe time transformations.

Chapter 3 first explores the construction of one-dimensional parametric families. In this exploration a common structure was found between the types of parameter. It was found that pairings arise between parameters in the literature, such as frailty and scale, power and hazard power. This then lead to the construction of new types of parameters, generated as a pair to already known parameters, such as the tilt or resilience which we have termed the reverse-tilt and reverse-resilience. Further to this, the properties discussed in chapter 2, addition, taking the product and composition, allowed for the construction of so called *combination families*. These were found to be very flexible families and were later used for modelling purposes. The one-dimensional parametric families were then extended to multi-dimensional parametric families. Methods to combine one-dimensional families were explored and various families, some of which were new, were described. The latter part of this chapter discusses the use of composition of cumulative hazard functions in the construction of frailty mixture models.

Chapter 4 reviewed different regression models for modelling time-to-event data. The chapter gives details on many of the well known models in the literature, including the Accelerated Failure Time, Proportional hazards and Proportional Odds models. This chapter also proposes regression models generated by the parametric families discussed in chapter 3. The latter part of the chapter then discussed the extended regression models in the literature and used the methods of combination of models in order to propose new extended regression models. The very last section in this chapter then discussed the baseline cumulative hazard function. The key of this section was to introduce the idea that the combination families of the previous chapter can be used to model a flexible baseline cumulative hazard. This way of modelling the baseline is then used in the following chapter when fitting models to the liver transplantation data.

Chapter 5 introduced and gave an exploratory analysis of the liver transplantation data set. A collection of models discussed and proposed in the previous chapters are fitted to the data and compared. It was found that using the linear combination of the Gompertz and the log-logistic to flexibly model the baseline cumulative hazard provided some promising results. The latter part of the analysis explored the presence of informative censoring in this data set, where it was found that there was significant evidence that this was present.

Chapter 6 aimed to clarify what is meant by informative and non-informative censoring. It explained the different sub-classes of non-informative censoring. This chapter also explained the identifiability issues when determining the dependence structure between events and censoring and approaches to dealing with these issues and to modelling. These approaches included bounding the joint and marginal survival functions, specifying the conditional distribution and performing a sensitivity analysis and finally specifying the form of the joint distribution of the event and censoring times. The approach that this chapter expands on is that of specifying the joint distribution and this method is applied to the liver transplantation data. It was found that using this joint model method for this data set results in a a good fit for the marginal distribution of the event times, and a less good fit for the marginal distribution of the censoring times. It was concluded that a good fit for the marginal of the event times is of more interest than that of the censoring times, thus this joint model is useful for dealing with informative censoring present in this data set.

7.2 Future work

In chapter 2 convexity of cumulative hazards was briefly explored. Further investigation is suspected to provide even greater insight into the form of the set of cumulative hazards and give another way to classify cumulative hazards. Another possible area of future work based on chapter 2 is that of the generalised models suggested in section 2.3.10. These models were simple extensions of already well known models, using properties such as composition and integration of cumulative hazards. Further exploration of these models is needed in order to clarify how to fit such models and how they compare to the generalised models proposed in the literature.

In chapter 3 there was a key focus on the long and short term behaviours of certain parametric families. As previously discussed, a possible avenue for further work would be to explore spline models in order to smoothly combine the types of behaviours which are deemed useful or necessary to explore. Another possible area of future work motivated by this chapter would be to experiment with the fitting of frailty models to the liver transplantation data set.

In chapter 4 many new regression models are proposed, and methods for constructing others are put forward. This methodology could be used further in order to explore these possible models, and then these could be used in fitting.

Chapter 6 proposed models for dealing with informative censoring. In this chapter only a selection of these models were used to explore the informative censoring in the liver transplantation data set. Thus an area of further work would be to explore the rest of the models in this chapter. It would be interesting to also extend these models to the case of multiple competing risks.

Appendix A

Concavity of the likelihoods

In this appendix we will explore the behaviours of the likelihoods of the one-dimensional combination parametric families in equations (3.9), the linear combination, and (3.10), the geometric combination in chapter 3.

It is useful to know whether the log-likelihoods corresponding to the cumulative hazard functions (3.9) and (3.10) are concave in the parameter α. If the log-likelihood is concave, then the search for the maximum likelihood estimate of α will be made easier, and many other results of maximum likelihood estimation follow. To see if the log-likelihood is concave, we find its second derivative and see if it is negative.

We will look at the concavity for one sample. If we have an independent sample of size n, i.e. $t_1, ..., t_n$, and the log-likelihood is concave for just one sample, it will be for the sample of size n. This is due to the independence of the sample which will mean the log-likelihood will be the sum of n identical terms. The concavity is ensured as the sum of concave functions is a concave function.

We will also look at the concavity of a censored sample as a simple extension of the families we have proposed. Here we use the density for independent censoring, given by

$$f(t) = h(t)^{1-\delta} e^{-H(t)}.$$

Here δ is the indicator of censoring, i.e. when $\delta = 0$ this indicates that sample was not censored, and when $\delta = 1$ this indicates the sample was censored.

Proposition A.1. *The censored log-likelihood corresponding to the CH function of the linear combination in equation* (3.9) *is concave.*

Proof. The censored log-likelihood for the linear combination is

$$L(\alpha) = \log f(t;\alpha) = (1-\delta)\log\left(\alpha h_A(t) + (1-\alpha)h_B(t)\right) \\ - \left(\alpha H_A(t) + (1-\alpha)H_B(t)\right).$$

The second derivative is

$$\frac{d^2L}{d\alpha^2} = -(1-\delta)\left(\frac{h_A(t) - h_B(t)}{\alpha h_A(t) + (1-\alpha)h_B(t)}\right)^2 \le 0,$$

thus the censored log-likelihood corresponding to the linear combination in (3.9) is concave. □

Corollary A.2. *The log-likelihood corresponding to the CH function in the linear combination where $\delta = 0$ in proposition A.1,*

$$H(t|\alpha) = \alpha H_A(t) + (1-\alpha)H_B(t), \qquad \alpha \in [0,1],$$

is (strictly) concave.

Proof. The log-likelihood for the linear combination is given by

$$L(\alpha) = \log f(t;\alpha) = \log\left(\alpha h_A(t) + (1-\alpha)h_B(t)\right) - \left(\alpha H_A(t) + (1-\alpha)H_B(t)\right).$$

The second derivative is

$$\frac{d^2L}{d\alpha^2} = -\left(\frac{h_A(t) - h_B(t)}{\alpha h_A(t) + (1-\alpha)h_B(t)}\right)^2 < 0,$$

thus the log-likelihood corresponding to the linear combination (3.9) is (strictly) concave. □

Proposition A.3. *The censored log-likelihood corresponding to the CH function of the geometric combination* (3.10) *is concave.*

Proof. To prove this it is again better to view the CH function as

$$H(t|\alpha) = e^{\alpha r_A(t) + (1-\alpha)r_B(t)},$$

where $r_A(t) = \log H_A(t)$ and $r_B(t) = \log H_B(t)$. Then the log-likelihood will be

$$\begin{aligned} L(\alpha) = (1-\delta)\log(\alpha r'_A(t) + (1-\alpha)r'_B(t)) \\ + (1-\delta)(\alpha r_A(t) + (1-\alpha)r_B(t)) - e^{\alpha r_A(t) + (1-\alpha)r_B(t)}. \end{aligned}$$

Thus the second derivative is

$$\frac{d^2L}{d\alpha^2} = -(1-\delta)\left(\frac{r'_A(t) - r'_B(t)}{\alpha r'_A(t) + (1-\alpha)r'_B(t)}\right)^2 - \left(r_A(t) - r_B(t)\right)^2 e^{\alpha r_A(t) + (1-\alpha)r_B(t)} < 0,$$

and hence the censored log-likelihood is concave. □

Corollary A.4. *The log-likelihood corresponding to the CH function of the geometric combination where $\delta = 0$ in proposition A.3,*

$$H(t|\alpha) = H_A(t)^\alpha H_B(t)^{1-\alpha}, \qquad \alpha \in [0,1],$$

is concave.

Proof. To prove this it is better to view the CH function as

$$H(t|\alpha) = e^{\alpha r_A(t) + (1-\alpha) r_B(t)},$$

where $r_A(t) = \log H_A(t)$ and $r_B(t) = \log H_B(t)$. Then the log-likelihood will be

$$L(\alpha) = \log(\alpha r'_A(t) + (1-\alpha) r'_B(t)) + \alpha r_A(t) + (1-\alpha) r_B(t) - e^{\alpha r_A(t) + (1-\alpha) r_B(t)}.$$

Thus the second derivative is

$$\frac{d^2 L}{d\alpha^2} = -\left(\frac{r'_A(t) - r'_B(t)}{\alpha r'_A(t) + (1-\alpha) r'_B(t)}\right)^2 - (r_A(t) - r_B(t))^2 \, e^{\alpha r_A(t) + (1-\alpha) r_B(t)} < 0,$$

and hence the log-likelihood is concave. □

The concavity of the log-likelihoods for the distributions of families generated by the linear and geometric combinations have both been confirmed, the same cannot be said of the family generated by the composition and reverse composition combinations and the linear-composition combination. Since these families are generated using compositions, the concavity of the log-likelihood would entirely depend on the cumulative hazard functions used in the composition.

For the log-likelihood to be concave in two variables α and β, we need

$$\frac{d^2 L(\alpha,\beta)}{d\alpha^2} \leq 0, \frac{d^2 L(\alpha,\beta)}{d\beta^2} \leq 0 \quad \text{and} \quad \frac{d^2 L(\alpha,\beta)}{d\alpha^2}\frac{d^2 L(\alpha,\beta)}{d\beta^2} - \left(\frac{d^2 L(\alpha,\beta)}{d\alpha d\beta}\right)^2 \geq 0.$$

For the two-dimensional linear combination, we have that the log-likelihood and the censored log-likelihood are concave.

Appendix B

Simulation for Survival Models

There are many ways to simulate from a distribution. This appendix will explore how to simulate from distributions with different forms of cumulative hazard functions, detailed in previous chapters.

A well known method that will be frequently used in this appendix is called Inversion sampling [47]. Inversion sampling is a simple algorithm which only requires that we know the inverse of the CDF of the density we wish to sample from. Suppose we wish to simulate a random variable X with CDF $F(t)$. Then the random variable $F(T)$ is uniformly distributed on $[0, 1]$, as seen in proposition 2.9. If we can find the inverse of the CDF F^{-1}, then given $U \sim Unif(0, 1)$, $T = F^{-1}(U)$ is the required random variable. So if we know the inverse of the CDF, then, since we can easily simulate from a $Unif(0, 1)$, we can simulate T without rejecting samples or loss of information [53].

For survival analysis, we are more likely to know the survival function rather than the CDF. Thus we can invert the survival function instead of the CDF, if this is more attainable. This is because $F(T) = 1 - S(T)$ and thus $S(T)$ will also be uniformly distributed on $[0, 1]$. Suppose that $T \sim H$ for some CH function H. Then we have that $P(T > t) = e^{-H(t)}$, which is a number between 0 and 1. Thus for some $U \sim Unif(0, 1)$, $T = H^{-1}(-\log(U))$ is a random variable with CH function H. Hence the problem of simulating from such a distribution, is reduced to knowledge of H^{-1}, together with the fact that $-\log(U)$ is Exponentially distributed with mean 1.

There are a number of numerical operations needed throughout this appendix. Firstly, in order to perform inversion sampling, we obviously need to be able to invert a CH function. As well as being able to compose CH functions, we will need to be able to integrate CH functions. This integration may need to be performed numerically, and thus may not be exact. Finally, we will also need to be able to simulate from a Uniform distribution on the interval from 0 to 1. This will be an essential part of inversion sampling.

B.1 Addition of cumulative hazard functions

Consider a random variable $T \sim H_1 + H_2$, for some CH functions H_1 and H_2. We can immediately see that inverting $H_1(T) + H_2(T)$ could be difficult, and would depend entirely on the forms of both CH functions. Hence, inversion sampling may not seem like an appropriate method of sampling for such a distribution. A useful property that can be taken advantage of for sampling is outlined in the following proposition, which is a restatement of proposition 2.52.

Proposition B.1. *Given two independent positive random variables $T_1 \sim H_1$ and $T_2 \sim H_2$, the survival function of their minimum $M = \min(T_1, T_2)$ is*

$$S(m) = e^{-(H_1(m)+H_2(m))}.$$

Proof. See proposition 2.52. □

Hence, the problem of inverting $H_1 + H_2$ has been reduced to inverting H_1 and H_2 separately. Hence, the procedure for sampling from $T \sim H_1 + H_2$ is as follows:

1. Sample U_1 and U_2 from $Unif(0, 1)$
2. Calculate $T_1 = H_1^{-1}(-\log(U_1))$ and $T_2 = H_2^{-1}(-\log(U_2))$
3. $T = \min(T_1, T_2)$.

Example B.2. Let $H_1(t) = H_G(t)$ and $H_2(t) = H_W^2(t)$. We simulated 10000 samples from $T \sim H_1 + H_2$ and compared the results with the exact distribution. A plot can be seen in figure B.1 of this comparison. We see that even at a sample of 10000, we are already approaching the exact distribution. This method was quick and with large samples would be very accurate.

B.2 Composition of cumulative hazard functions

Suppose that $T \sim H_1 \circ H_2$, then the survival function of T is

$$S(t) = e^{-H_1(H_2(t))}.$$

If we know how to invert both H_1 and H_2 then we can use inversion sampling to generate T. Suppose we only know how to invert H_2, but we know that sampling from $\tilde{T} \sim H_1$ is simple, then we should

1. generate $\tilde{T}$ from H_1
2. calculate $T = H_2^{-1}(\tilde{T})$.

The proof of the procedure for sampling from the composition when we know how to sample from $\tilde{T} \sim H_2$ is demonstrated in the following proposition.

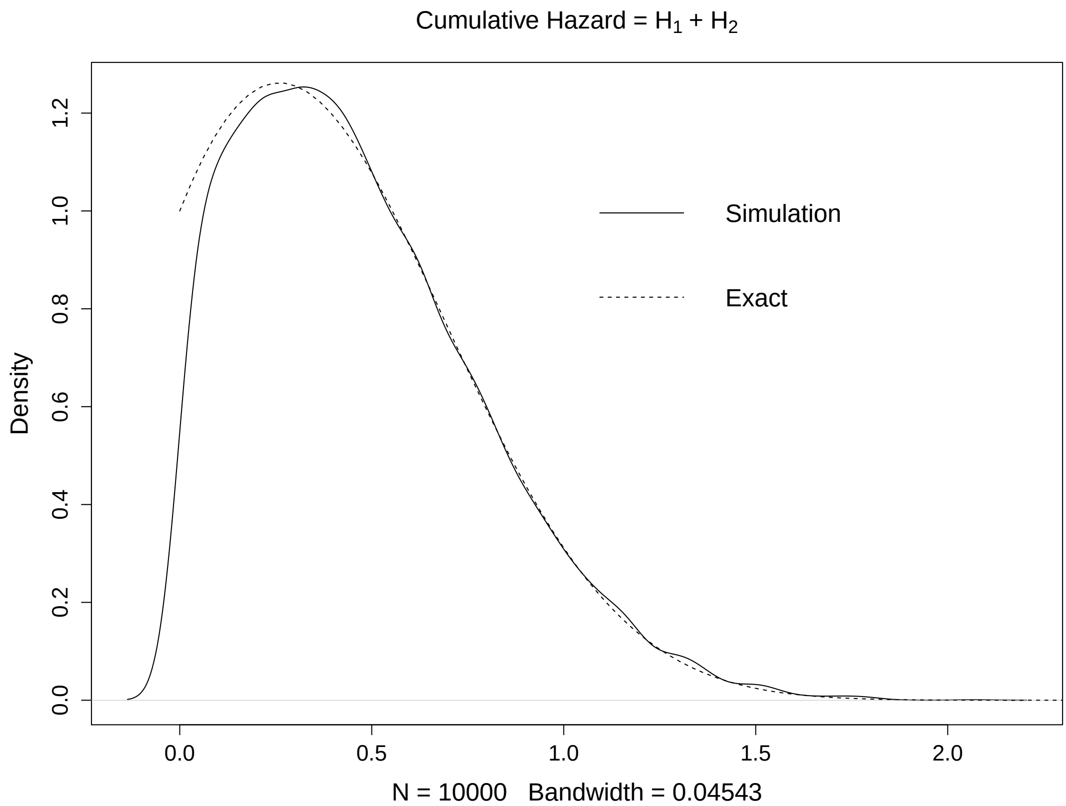

Figure B.1: Simulating from $T \sim H_1 + H_2$ where $H_1(t) = H_G(t)$ and $H_2(t) = H_W^2(t)$ using the minimum method, and comparing with the exact distribution.

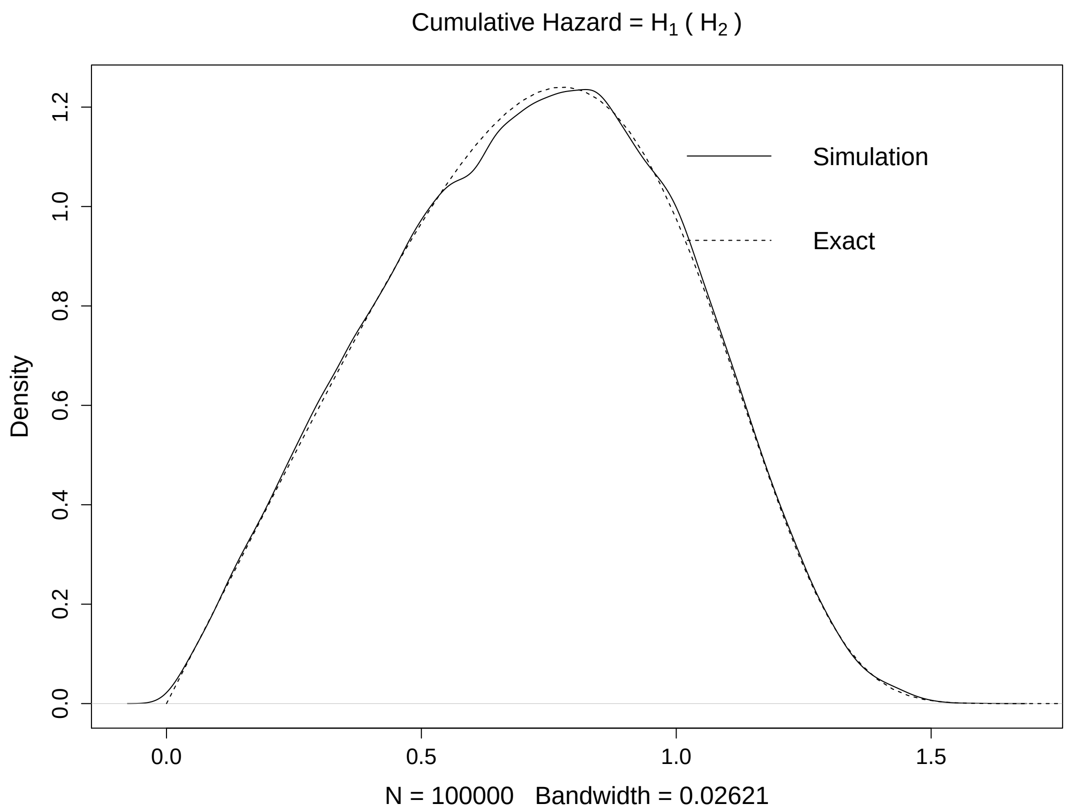

Figure B.2: Simulating from $T \sim H_1 \circ H_2$ where $H_1(t) = H_G(t)$ and $H_2(t) = H_W^2(t)$, and comparing with the exact distribution.

Proposition B.3. *Given two independent positive random variables* $T_1 \sim H_1$ *and* $T_2 \sim H_2$*, the survival function of* $Y = H_2^{-1}(T_1)$ *is*

$$S(y) = e^{-H_1(H_2(y))}.$$

Proof.

$$\begin{aligned} S(y) &= P(Y \geq y) = P(H_2^{-1}(T_1) \geq y) \\ &= P(T_1 \geq H_2(y)) = S_1(H_2(y)) \\ &= e^{-H_1(H_2(y))} \end{aligned}$$

□

Example B.4. Let $H_1(t) = H_G(t)$ and $H_2(t) = H_W^2(t)$ We simulated 100000 samples from the distribution $T \sim H_1 \circ H_2$ using the above method. In figure B.2 we see the comparison of the simulation with the exact distribution. We see that even with this small sample, we are getting close to the exact distribution. As $\tilde{T} \sim H_1$ we expect this to be accurate for small samples.

B.3 Integral of cumulative hazard functions

This section explores methods to simulate from a distribution whose CH function is an N-CH function, as seen in section 2.3.8. Hence $T \sim H^{[1]}$, then the survival function for this distribution will be

$$S(t) = e^{-H^{[1]}(t)}.$$

We see that in order to use inversion sampling we will need to solve

$$-\log(U) = \int_0^T H(s)ds = e^{-\int_0^t H(s)ds}.$$

Newton's method

If H is sufficiently nice, i.e. convex or concave, we can use Newton's method to find T. Thus

1. Generate U from $Unif(0,1)$.

2. Choose t_0

3. Iterate $t_{n+1} = t_n - \frac{H^{[1]}(t_n)+\log(U)}{H(t_n)}$, for $n \in \mathbb{N}$, until convergence is reached.

Since $H^{[1]}$ is increasing and convex, Newton's method should converge fairly quickly. This method requires knowledge of the functional form of at least one of H or $H^{[1]}$. If we know H, we must be able to integrate it, at least numerically, for Newton's method. If we know $H^{[1]}$ we must be able to differentiate in order to calculate $H(t_n)$.

Example B.5. Let $H(t) = \log(1 + t) = H_{ll}(t)$, we simulated 1000 samples for this distribution and compared the results with the exact distribution. A plot can be seen in figure B.3 of this comparison. We see that even at a small sample of 1000, we are already approaching the exact distribution. This method was very quick, and increasing the tolerance of the convergence would give even better, if slower, results.

Initial Value Problem method

If the previous method is too computationally demanding, we can re-pose the problem as an Initial Value Problem (IVP). Note that solving $-\log(U) = \int_0^T H(s)ds$, is the same as solving

$$\frac{dT}{dx} = \frac{1}{H(T)}$$

where $T(x)$ is defined by $x = \int_0^{T(x)} H(s)ds$. We note that $T(1) = 0$, the initial condition, means that we may not have a unique solution for T since $\frac{1}{H(T)}$ is not continuous at $T = 0$. We thus must know some other initial condition, t_0 say, so that $\int_0^{t_0} H(s)ds = x_0$, i.e. $T(x_0) = t_0$.

Thus the method for simulating is

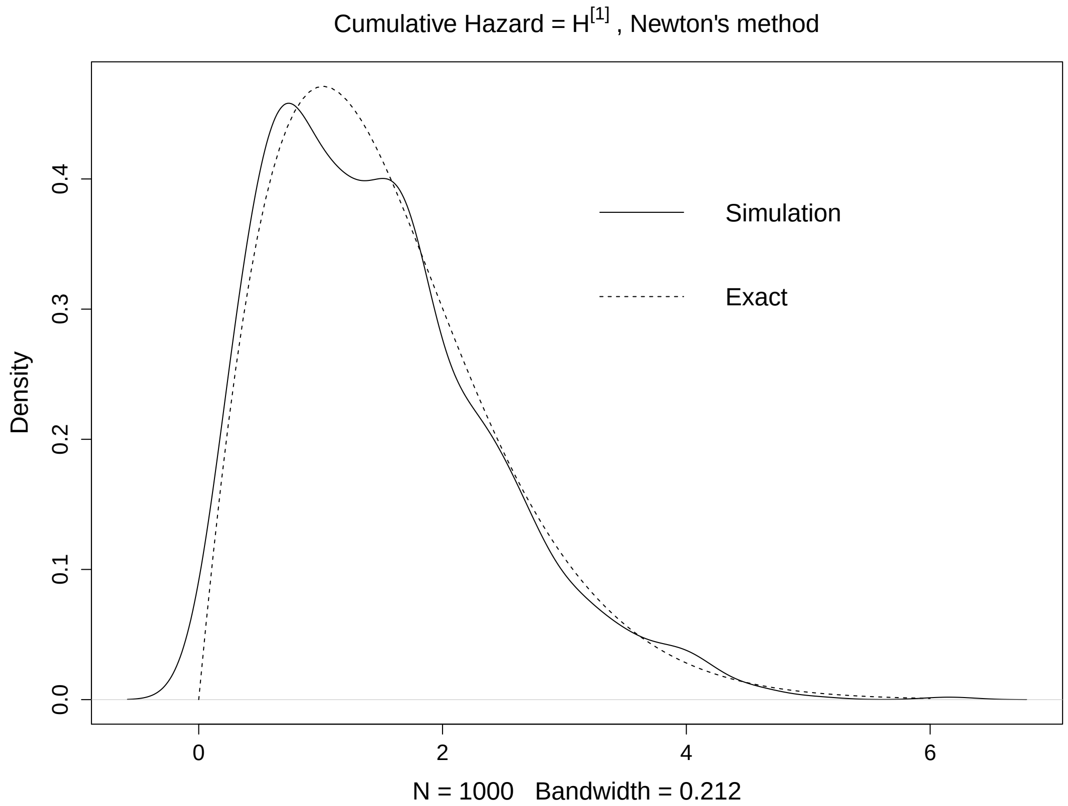

Figure B.3: Simulating from $T \sim H^{[1]}$ where $H(t) = H_{ll}(t)$ using Newton's method, and comparing with the exact distribution.

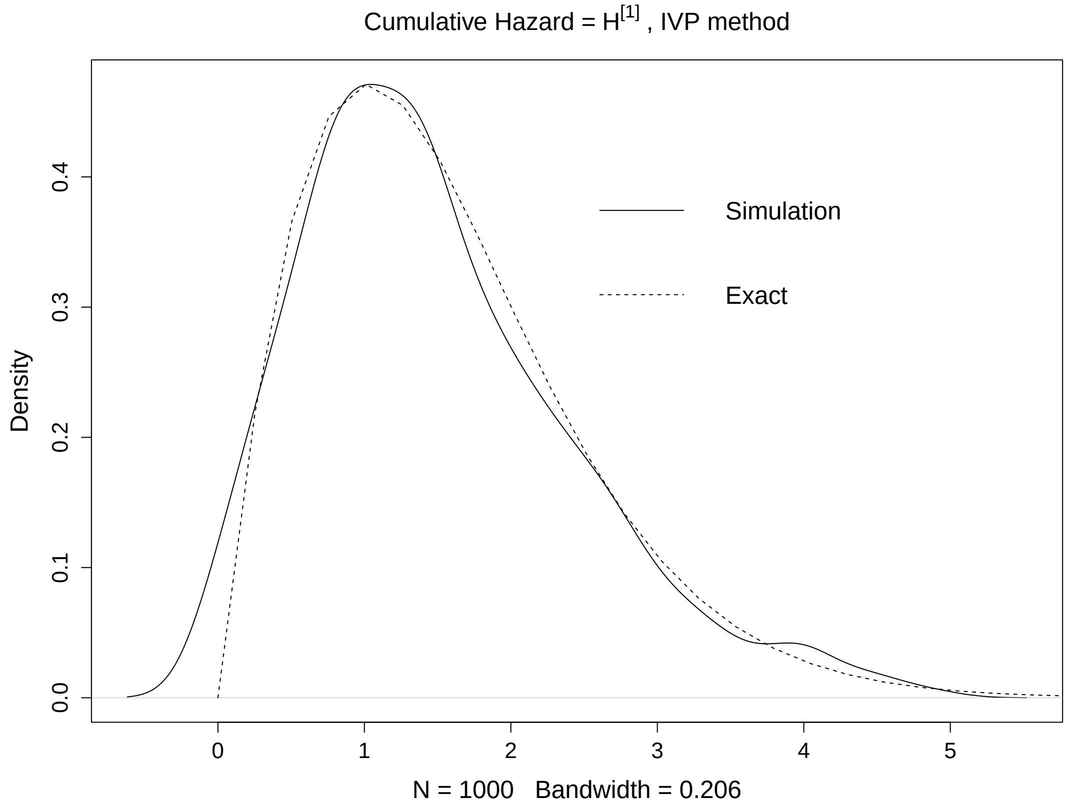

Figure B.4: Simulating from $T \sim H^{[1]}$ where $H(t) = H_{ll}(t)$ using the IVP method, and comparing with the exact distribution. Here the initial condition is $H^{[1]}(t_0) = 0.0001$.

1. Choose initial condition (x_0, t_0)
2. Generate U from $Unif(0, 1)$
3. Solve the ODE

$$\frac{dT(x)}{dx} = \frac{1}{H(T(x))}$$

where $-\log(U) = x$ for $T(x)$.

Example B.6. We simulated 1000 samples from $T \sim H^{[1]}$ for $H(t) = H_{ll}(t)$ using the IVP method, and compared the results with the exact distribution. In this case we knew $H^{[1]}$ and had to invert to find t_0. A plot can be seen in figure B.4 of this comparison. We see that even at a small sample of 1000, we are already approaching the exact distribution. This method was quite fast, knowing the initial condition completely would speed up the computation even more, but more accurate simulations will slow down the computation.

Sum method

When approximating an integral, it is common to use a summation based method. We therefore, may have that $H^{[1]}(t) = \sum_{i=1}^{n} H_i(t)$. Thus the problem of simulating from a distribution whose CH function is an integral is reduced to the problem of simulating from a distribution whose CH function is a sum. Note that this problem is different to that as discussed in section B.1 since the sum in this setting is not the sum of independent terms.

Our procedure will be to approximate the integral with a sum and then find the value of t such that $-\log(u) = H^{[1]}(t)$ using a grid search algorithm. Hence we will

1. Simulate $U \sim Unif(0,1)$
2. Solve

$$-\log(U) = \int_0^t H(s)ds \approx \frac{t}{3n}\left[4H\left(\frac{t}{n}\right) + 2H\left(\frac{2t}{n}\right) + \ldots + 2H\left(\frac{(n-1)t}{n}\right) + 4H\left(\frac{(n-1)t}{n}\right) + H(t)\right].$$

   for t using a grid search.

Example B.7. Let $T \sim H^{[1]}$ where $H(t) = H_{ll}(t)$. Then we approximated the integral $H^{[1]}$ by a sum using Simpson's rule. We simulated 10000 samples from this distribution where $T \sim H^{[1]}$ and compared the results with the exact distribution. A plot can be seen in figure B.5 of this comparison. We see that with a sample of 10000, we are very close to the exact distribution.

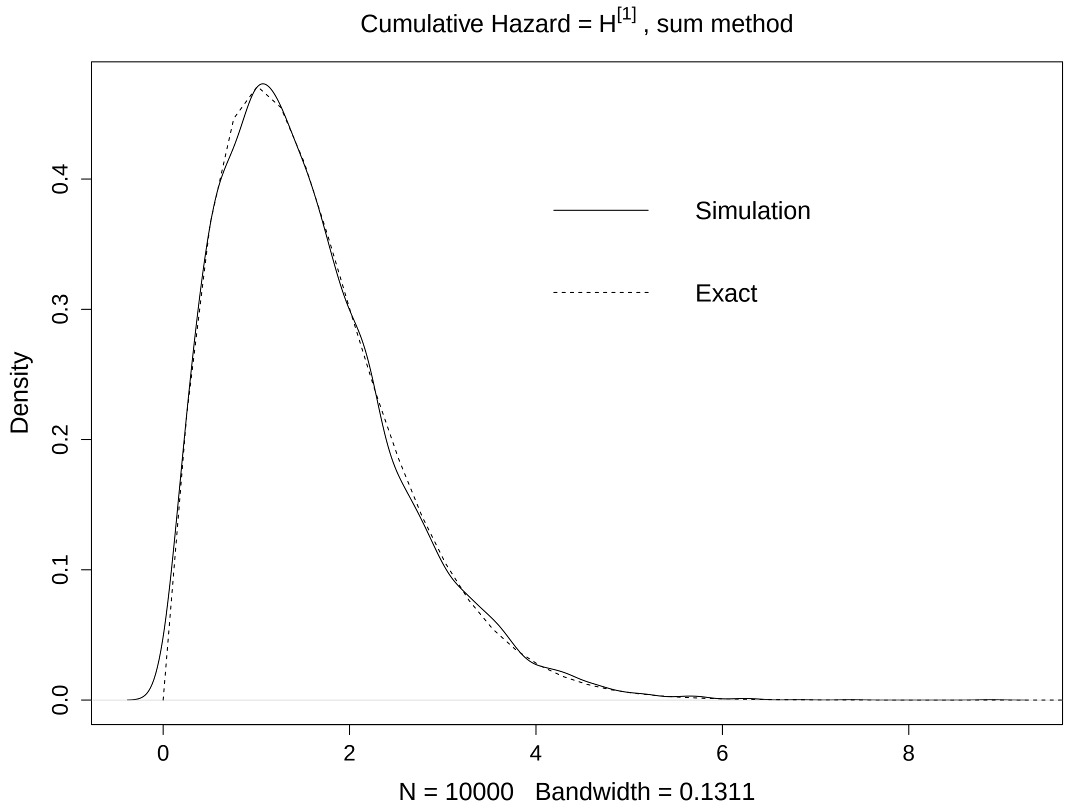

Figure B.5: Simulating from $T \sim H^{[1]}$ where $H(t) = H_{ll}(t)$ using the sum method for approximating the integral, and comparing with the exact distribution.

Bibliography

[1] O. Aalen. Heterogeneity in survival analysis. *Statistics in Medicine*, 7:1121–1137, 1988.

[2] O. Aalen, O. Borgan, and H. Gjessing. *Survival and Event History Analysis: A Process Point of View*. Springer Science & Business Media, 2008.

[3] K. Barber, S. Madden, J. Allen, D. Collett, J. Neuberger, A. Gimson, et al. Elective liver transplant list mortality: development of a United Kingdom end-stage liver disease score. *Transplantation*, 92:469–476, 2011.

[4] B. Bassan, M. Denuit, and M. Scarsini. Variability orders and mean differences. *Statistics & Probability Letters*, 45:121–130, 1999.

[5] S. Bennett. Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2:273–277, 1983.

[6] J. O. Berger and D. Sun. Bayesian analysis for the poly-weibull distribution. *Journal of the American Statistical Association*, 88:1412–1418, 1993.

[7] Z.W. Birnbaum and S.C. Saunders. A new family of life distributions. *Journal of Applied Probability*, 6:319–327, 1969.

[8] N. E. Breslow and N. E. Day. *Statistical Methods in Cancer Research*, volume 1. International agency for research on cancer Lyon, 1980.

[9] J. Buckley and I. James. Linear regression with censored data. *Biometrika*, 66:429–436, 1979.

[10] K. Burke and G. MacKenzie. Multi-parameter regression survival modeling: an alternative to proportional hazards. *Biometrics*, 73:678–686, 2017.

[11] K. Burke and G. MacKenzie. Multi-parameter regression survival modeling: an alternative to proportional hazards. *Biometrics*, 73:678–686, 2017.

[12] J. Carpenter and M. Kenward. *Multiple Imputation and its Application*. John Wiley & Sons, 2012.

[13] Y. Q. Chen and N. P. Jewell. On a general class of semiparametric hazards regression models. *Biometrika*, 88:687–702, 2001.

[14] Y. Q. Chen and M.-C. Wang. Analysis of accelerated hazards models. *Journal of the American Statistical Association*, 95:608–618, 2000.

[15] D. Clayton and J. Cuzick. The semiparametric pareto model for regression analysis of survival times. *Papers on Semiparametric Models*, pages 19–31, 1986.

[16] D. G. Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65:141–151, 1978.

[17] D. Collett. *Modelling Survival Data in Medical Research.* Text in Statistical Science Series. CRC Press, second edition, 2003.

[18] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 187–220, 1972.

[19] D. R. Cox. Partial likelihood. *Biometrika*, 62:269–276, 1975.

[20] D. R. Cox and D. V. Hinkley. *Theoretical Statistics.* CRC Press, 1979.

[21] D. R. Cox and D. Oakes. *Analysis of Survival Data.* CRC Press, 1984.

[22] M. Crowder. On repeated measures analysis with misspecified covariance structure. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63:55–62, 2001.

[23] M. J. Crowder. *Multivariate Survival Analysis and Competing Risks.* CRC Press, 2012.

[24] M. J. Crowther. *Development and application of methodology for the parametric analysis of complex survival and joint longitudinal-survival data in biomedical research.* PhD thesis, Department of Health Sciences, University of Leicester, 2014.

[25] M. Friedman. Piecewise exponential models for survival data with covariates. *The Annals of Statistics*, pages 101–113, 1982.

[26] R. D. Gupta and D. Kundu. Exponentiated exponential family: an alternative to gamma and weibull distributions. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 43:117–130, 2001.

[27] F. Haghighi and M. Nikulin. A chi-squared test for power generalized weibull family for the head-and-neck cancer censored data. *Journal of Mathematical Sciences*, 133:1333–52, 2007.

[28] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, 1952.

[29] P. Hougaard. Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika*, 71:75–84, 1984.

[30] P. Hougaard. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73:387–396, 1986.

[31] N. P. Jewell. Mixtures of exponential distributions. *The Annals of Statistics*, pages 479–484, 1982.

[32] Z. Jin, D. Y. Lin, L. J. Wei, and Z. Ying. Rank-based inference for the accelerated failure time model. *Biometrika*, 90:341–353, 2003.

[33] N. L. Johnson, S. Kotz, and N. Balakrishnan. Continuous univariate distributions, vol. 2 of Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, 1995.

[34] M. C. Jones and A. Noufaily. Log-location-scale-log-concave distributions for survival and reliability analysis. *Electronic Journal of Statistics*, 9:2732–2750, 2015.

[35] J. D. Kalbfleisch and R. J. MacKay. On constant-sum models for censored survival data. *Biometrika*, 66:87–90, 1979.

[36] J. D. Kalbfleisch and R. L. Prentice. The Statistical Analysis of Failure Time Data. *Hoboken, NJ*, 1980.

[37] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.

[38] J. P. Klein and M. L. Moeschberger. Bounds on net survival probabilities for dependent competing risks. *Biometrics*, pages 529–538, 1988.

[39] S. W. Lagakos. General right censoring and its impact on the analysis of survival data. *Biometrics*, pages 139–156, 1979.

[40] S. W. Lagakos and J. S. Williams. Models for censored survival analysis: A cone class of variable-sum models. *Biometrika*, 65:181–189, 1978.

[41] J. F. Lawless. *Statistical Models and Methods for Lifetime Data*, volume 362. John Wiley & Sons, 2011.

[42] DY Lin and Z. Ying. Semiparametric analysis of the additive risk model. *Biometrika*, 81:61–71, 1994.

[43] W. M. Makeham. On the law of mortality and construction of annuity tables. *Journal of the Institute of Actuaries*, 8:301–310, 1860.

[44] A. W. Marshall and I. Olkin. *Life Distributions.* Springer, 2007.

[45] T. Martinussen and T. H. Scheike. *Dynamic Regression Models for Survival Data.* Springer Science & Business Media, 2007.

[46] W. Q. Meeker and L. A. Escobar. *Statistical Methods for Reliability Data.* John Wiley & Sons, 2014.

[47] B. J. T. Morgan. *Elements of Simulation*, volume 4. Chapman and Hall, 1984.

[48] S. A. Murphy, A. J. Rossini, and A. W. van der Vaart. Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*, 92:968–976, 1997.

[49] D. N. P. Murthy, M. Xie, and R. Jiang. *Weibull Models*, volume 505. John Wiley & Sons, 2004.

[50] R. B. Nelsen. *An Introduction to Copulas.* Springer Science & Business Media, 2007.

[51] A. V. Peterson. Bounds for a joint distribution function with fixed sub-distribution functions: application to competing risks. *Proceedings of the National Academy of Sciences*, 73:11–13, 1976.

[52] Horst Rinne. *The Weibull distribution: a handbook.* Chapman and Hall/CRC, 2008.

[53] B. D. Ripley. *Stochastic Simulation*, volume 316. John Wiley & Sons, 1987.

[54] D. Rizopoulos, G. Verbeke, and E. Lesaffre. Fully exponential laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71:637–654, 2009.

[55] R. T. Rockafellar. Convex Analysis. Number 28 in Princeton Mathematical Series. *Princeton University Press, Princeton*, 50:55–57, 1970.

[56] D. Roy and S. P. Mukherjee. Multivariate extensions of univariate life distributions. *Journal of Multivariate Analysis*, 67:72–79, 1998.

[57] P. Royston and M. K. B. Parmar. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, 21:2175–2197, 2002.

[58] B. Rutickii, M. A. Kranosel'skii-Ya. Convex functions and Orlicz spaces. *Noordhoff, Gröningen*, 1961.

[59] P. D. Sasieni. Proportional excess hazards. *Biometrika*, 83:127–141, 1996.

[60] R.L. Schilling, R. Song, and Z. Vondraček. *Bernstein Functions: Theory and Applications*. De Gruyter studies in mathematics. De Gruyter, 2010.

[61] G. Shaddick and J. V. Zidek. A case study in preferential sampling: Long term monitoring of air pollution in the uk. *Spatial Statistics*, 9:51–65, 2014.

[62] F. Siannis. Applications of a parametric model for informative censoring. *Biometrics*, 60:704–714, 2004.

[63] F. Siannis, J. Copas, and G. Lu. Sensitivity analysis for informative censoring in parametric survival models. *Biostatistics*, 6:77–91, 2005.

[64] N. D. Singpurwalla. The hazard potential: introduction and overview. *Journal of the American Statistical Association*, 101:1705–1717, 2006.

[65] E. V. Slud and L. V. Rubinstein. Dependent competing risks and summary survival curves. *Biometrika*, 70:643–649, 1983.

[66] N. D. Staplin, A. C. Kimber, D. Collett, and P. Roderick. Dependent censoring in piecewise exponential survival models. *Statistical Methods in Medical Research*, 24 3:325–41, 2015.

[67] J. D. Taulbee. A general model for the hazard rate with covariables. *Biometrics*, pages 439–450, 1979.

[68] T. M. Therneau and P. M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Springer Science & Business Media, 2013.

[69] Y. Tseng, F. Hsieh, and J. Wang. Joint modelling of accelerated failure time and longitudinal data. *Biometrika*, 92:587–603, 2005.

[70] A. Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72:20–22, 1975.

[71] A. Wang. Properties of the marginal survival functions for dependent censored data under an assumed archimedean copula. *Journal of Multivariate Analysis*, 129:57–68, 2014.

[72] L. Wei. The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. *Statistics in Medicine*, 11:1871–1879, 1992.

[73] R.l Wiesner, E. Edwards, R. Freeman, A. Harper, R. Kim, P. Kamath, W. Kremers, J. Lake, T. Howard, R. M. Merion, et al. Model for end-stage liver disease (meld) and allocation of donor livers. *Gastroenterology*, 124:91–96, 2003.

[74] J. S. Williams and S. W. Lagakos. Models for censored survival analysis: constant-sum and variable-sum models. *Biometrika*, 64:215–224, 1977.

[75] S. Yang and R. Prentice. Semiparametric analysis of short-term and long-term hazard ratios with two-sample survival data. *Biometrika*, 92:1–17, 2005.

[76] S. Yang and R. L. Prentice. Assessing potentially time-dependent treatment effect from clinical trials and observational studies for survival data, with applications to the women's health initiative combined hormone therapy trial. *Statistics in Medicine*, 34:1801–1817, 2015.

[77] S. C. Yang and J. A. Nachlas. Bivariate reliability and availability modeling. *IEEE Transactions on Reliability*, 50:26–35, 2001.