Regression Analysis of Time-to-Event Data

Matthew Pawley

Specialist Reading Course

Supervisor: Dr. Karim Anaya-Izquierdo

January 2021

Contents

CHAPTER 1

Introduction

The focus of this project is to review key parametric regression models in time-to-event analysis and their use in a variety of fields including medicine and engineering.

Chapter 2 briefly outlines the theory of parametric survival analysis and introduces some commonly used models, including accelerated failure time, proportional hazards and proportional odds.

In Chapter 3 we discuss the APGW distribution, a general-purpose parametric distribution proposed in [3] and implement it as a custom distribution in flexsurv, an R package for parametric survival modelling. Efficient algorithms to sample from survival distributions are presented and implemented.

In Chapter 4 we perform some simulation studies. We describe a suite of diagnostic tools for identifying appropriate modelling assumptions, before fitting some regression models. Finally, we investigate how the length of the study observation period affects our ability discriminate between proportional odds and proportional hazards data.

Chapter 5 analyses two real-world datasets from engineering (failure stress of carbon fibres) and medicine (clinical trial). The second dataset requires modifying our approach to deal with continuous covariates.

All the results in this report can be reproduced using the code available at https://github.com/pawleymatthew/RC-Survival-Regression.

CHAPTER 2

Survival Regression Models

In this chapter, we review regression models in survival analysis. The aim of survival regression models is to model the relationship between the survival time and a set of covariates. Regression models are used to answer questions of interest, such as comparing survival between groups, testing/quantifying covariate effects, and making predictions about unseen individuals.

Assume the survival time T is continuous and depends upon p covariates, $X_1, X_2, \ldots, X_p$. For individual $i = 1, \ldots, n$ we observe a set of covariates $\boldsymbol{x}_i = (x_{1i}, \ldots, x_{pi})^T$, an event/right-censoring time t_i, and the censoring indicator

$$\delta_i = \begin{cases} 0 & \text{if event time is observed} \\ 1 & \text{if event time is right-censored.} \end{cases}$$

The following are assumed throughout this report:

(i) Event times are right-censored, with no left-truncation. In principle, the models can be easily generalised to interval-censored and left-truncated data.

(ii) Censoring is random and non-informative. Informally, this means the censoring and event times are independent, and the distribution of survival times provides no information about the distribution of censoring times (and vice versa) [12].

(iii) Covariates are not time-varying. An extensive treatment of survival regression models with time-varying covariates can be found in [14].

A survival regression model specifies the relationship between the covariates and the distribution of T. In particular, the CH function for an individual with covariates $\boldsymbol{x}$ is given by $H(t|\theta, H_0)$, where $\theta = \exp(\boldsymbol{\beta}^T\boldsymbol{x})$ is a parameter which determines the dependence of distribution of T on the covariates, $\boldsymbol{\beta}$ is a vector of regression coefficients, and H_0 is the baseline CH, i.e. the CH for an individual with all covariates equal to zero. It is common to choose $\exp(\boldsymbol{\beta}^T\boldsymbol{x})$ as the 'link' function, but other choices are available are may be preferable in some scenarios [5, 15]. Importantly, our choice of link function ensures that $\theta > 0$, ensuring the CH functions specified in the following sections are valid. We focus our attention on fully-parametric models, meaning we assume a parametric form for the baseline. For brevity, we often write $H(t|\theta)$ as shorthand for $H(t|\theta, H_0)$.

2.1 Accelerated failure time models

The accelerated failure time (AFT) model assumes that the covariates act multiplicatively to accelerate/decelerate the rate of progression to the event [7]. In terms of the CH function, the AFT model is given by

$$H(t|\theta) = H_0(\theta t) \tag{2.1}$$

The parameter θ is called a scale parameter because it stretches the failure distribution along the time axis.

Following the approach of Davis in [5], it is instructive to consider the CH in terms of a composition. Compositions will allow us to define 'pairings' between regression models: two regression models form a pair when the CH compositions are reversals of one another. Moreover, writing the CH in terms of a composition is useful for data simulation purposes (see Section 3.2.1). Rewriting (2.1), we have

$$H(t|\theta) = H_0 \circ H_E^{\theta}(t) \tag{2.2}$$

where $H_E^{\theta}(t) = \theta t$ is the CH of the Exp (θ) distribution.

2.2 Proportional hazards models

The proportional hazards (PH) model assumes that the covariates are related multiplicatively to the hazard [7], so that

$$H(t|\theta) = \theta H_0(t) \tag{2.3}$$

It derives its name from the fact that individuals' hazard ratios are constant in time:

$$\frac{h(t|\theta_1)}{h(t|\theta_2)} = \frac{\theta_1 h_0(t)}{\theta_2 h_0(t)} = \frac{\theta_1}{\theta_2}.$$

PH models are closely related to frailty models (random effects models where the random effect has a multiplicative effect on the hazard) [8], so θ is called a frailty parameter. The PH model can be paired with the AFT model, indicating a close link between the two, since (2.3) is equivalent to

$$H(t|\theta) = H_E^{\theta} \circ H_0(t) \tag{2.4}$$

which is the reverse composition of (2.2).

2.3 Proportional odds models

The proportional odds (PO) model assumes covariates are related multiplicatively to the odds of the event occurring [2]. Thus the PO model is given in terms of the survival function by

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

Using the relation $S(t) = \exp(-H(t))$, this can be rewritten in term of CH functions as

$$H(t|\theta) = \log\left(1 + \theta(e^{H_0(t)} - 1)\right) \tag{2.5}$$

or equivalently

$$H(t|\theta) = H_{ll} \circ H_E^{\theta} \circ H_G \circ H_0(t) \tag{2.6}$$

where $H_{ll}(t) = \log(1 + t)$ and $H_G(t) = e^t - 1$ are the CH functions of the standard log-logistic and standard Gompertz distributions respectively. For reasons discussed in [13, p. 244], θ is called a tilt parameter.

The PO model exhibits interesting asymptotic behaviours, which can be studied by considering the asymptotic expansion of (2.5). In particular, as $t \to 0$, $H(t|\theta) \sim \theta H_0(t)$, so the tilt parameter behaves like a frailty parameter. On the other hand, as $t \to \infty$, $H(t|\theta) \sim H_0(t)$. This is a key property of the PO model: individuals' hazards converge to the baseline hazard in the long run. This may be a desirable property, e.g. if the effects of different treatments between groups is believed to diminish with time [2].

To affect significant change on the survival/hazard scale, the tilt parameter needs to be $\theta \gg 1$ or $\theta \ll 1$. This is illustrated in Figure 2.1. When viewed on the CH scale, increasing/decreasing the odds by a factor of 10 (Figure 2.1c) is not as significant as, say, increasing/decreasing the hazards by a factor of 2 (Figure 2.1b). This is important to bear in mind when interpreting regression coefficients in PO models.

2.4 Proportional Gompertz times models

The theoretical framework developed by Davis in [5, Chapter 3] yields a rich collection of parametric families, each of which could in principle form the basis of a regression model. Some families/models, such as the ones considered so far, are widely studied in the literature and commonly applied in practice. Others are more contrived, in the sense that they are solely born out of the theoretical framework and may have no practical utility for modelling real-world phenomena. One such family is the reverse-tilt family, which gives rise to (adopting Davis' terminology) the proportional Gompertz times (PGT) model.

As the name suggests, the reverse-tilt family is defined simply as a natural pairing to the tilt family. Therefore the CH of the PGT model is given by the reverse composition of (2.6), namely

$$H(t|\theta) = H_0 \circ (H_{ll} \circ H_E^{\theta} \circ H_G)(t) \tag{2.7}$$

or more explicitly

$$H(t|\theta) = H_0\left(\log(1 + \theta(e^t - 1))\right). \tag{2.8}$$

The PGT model is included here for completeness and and will be implemented later to illustrate the flexibility of the computational methods we have developed. However, the PGT model lacks

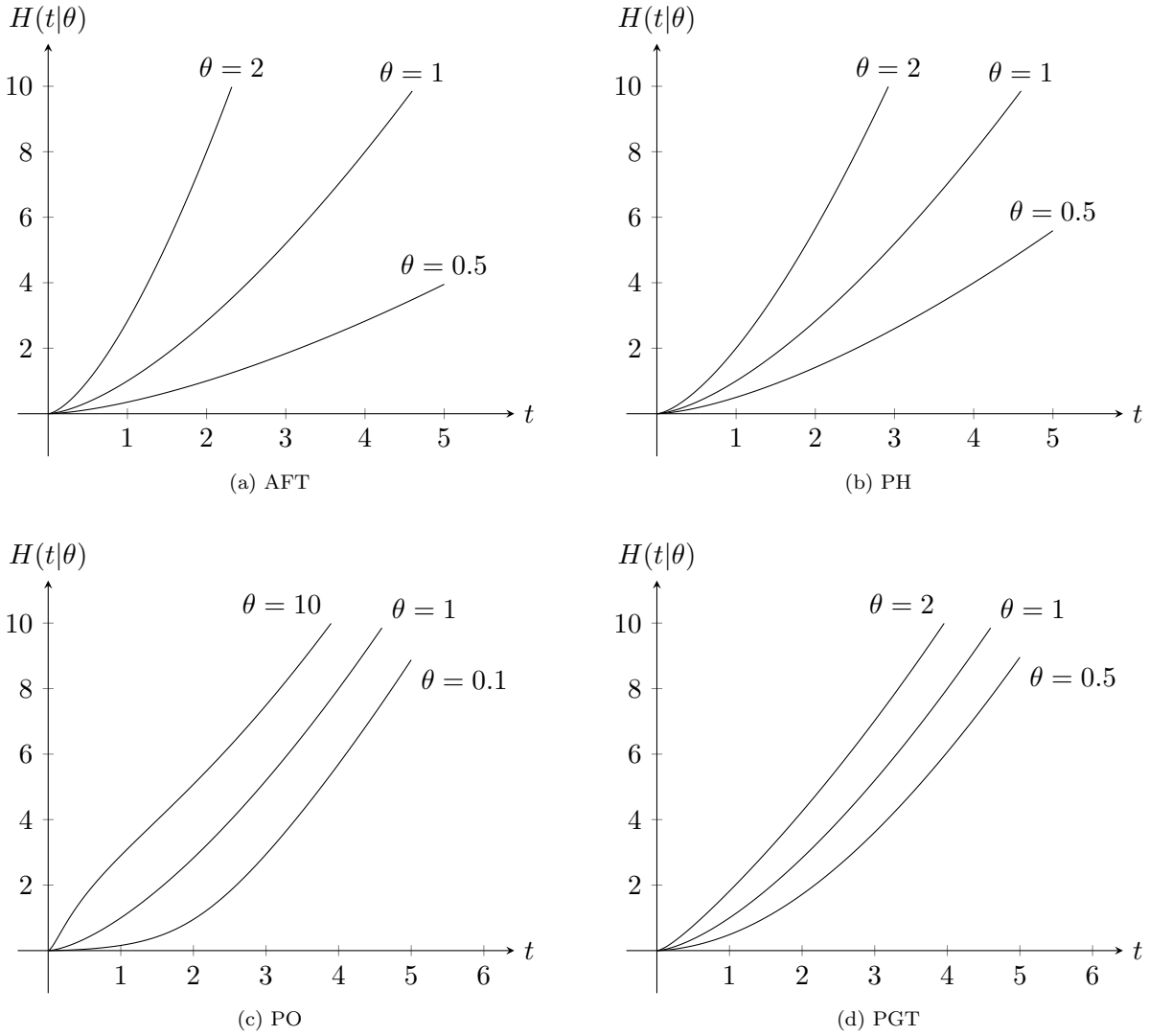


Figure 2.1: The function $H(t|\theta)$ for various values of θ, where θ is taken as an (a) scale, (b) frailty, (c) tilt, and (d) reverse-tilt parameter, with Weibull baseline $H_0(t) = t^{3/2}$.

interpretability and so will rarely be considered or discussed later on when fitting models to simulated and real-world datasets.

For a fully-parametric model, we need not only specify the covariate effects (e.g. AFT/PH/PO/PGT) but also a parametric baseline distribution. The next chapter introduces one possible candidate for the baseline.

CHAPTER 3

The APGW distribution

For fully-parametric models we need to assume a parametric form for the baseline, of which there are infinitely many possible choices and which of those is appropriate will generally depend on the context. In this chapter, we discuss a general-purpose candidate, the adapted power generalised Weibull (APGW) distribution, proposed by Burke, Jones, and Noufaily in [3]. We implement the APGW distribution using flexsurv, an R package, and perform some tests using simulations and a real-world dataset.

3.1 Definition and properties

APGW is a four-parameter distribution defined through its CH function by

$$H(t;\phi,\lambda,\gamma,\kappa) = \lambda\frac{\kappa+1}{\kappa}\left[\left(1+\frac{(\phi t)^\gamma}{\kappa+1}\right)^\kappa - 1\right] \tag{3.1}$$

where $\phi > 0$, $\lambda > 0$ control the horizontal/vertical scaling of the hazard function respectively and $\gamma > 0$, $\kappa > -1$ control the shape.

Having two shape parameters allows the APGW distribution to exhibit a wide range of simple hazard shapes: constant, increasing, decreasing, up-then-down and down-then-up. Burke, Jones, and Noufaily argue that the breadth of distributions encompassed by APGW (e.g. exponential, Weibull, log-logistic, Gompertz) as well as its relative mathematical/computational tractability make it a better choice for a flexible, general parametric model than other two-shape-parameter families, such as the exponentiated Weibull or generalised gamma.

3.2 Sampling from APGW and related distributions

In this section, we describe how to use inversion sampling to generate samples from a distribution with CH function $H(\cdot|\theta, H_0)$, where the baseline distribution H_0 is APGW and θ is a scale/frailty/tilt/reverse-tilt parameter. This will be useful later on (Chapter 4) to check that our regression models are implemented properly, by fitting models to simulated data and checking that they recover the (known) true parameters.

3.2.1 Inversion sampling for survival distributions

Inversion sampling is general approach for simulating data from a given distribution. Suppose we want to simulate a random variable with CDF F, and define the (generalised) inverse of F

as

$$F^{-1}(u) = \min\{t : F(t) \geq u\}, \qquad 0 < u < 1.$$

If $U \sim \text{Uniform}\,(0,1)$, then $T = F^{-1}(U)$ has the required distribution [1].

In survival analysis, distributions are often given through their CH function. The inversion sampling procedure can be easily adapted to simulate from a random variable with a prescribed CH function H [5]. Defining

$$H^{-1}(u) = \min\{t : H(t) \geq u\}, \qquad v > 0.$$

we note that

$$F^{-1}(u) = \min\{t : F(t) \geq u\} = \min\{t : H(t) \geq -\log(1-u)\} = H^{-1}(-\log(1-u))$$

and so if $U \sim \text{Uniform}\,(0,1)$, then $H^{-1}(-\log(1-U))$ has CH function H. Further, we can use the fact that $-\log(1-U) \overset{d}{=} -\log(U)$ and $-\log(U) \sim \text{Exp}\,(1)$. Hence if $V \sim \text{Exp}\,(1)$, then $T = H^{-1}(V)$ has CH function H.

Moreover, if the CH is written in terms of a composition $H = H_1 \circ H_2$ then we can use the fact that if $\widetilde{T} \sim H_1$, then $T = H_2^{-1}(\widetilde{T})$ has the desired distribution [5].

3.2.2 Simulating from the APGW distribution

The inverse of the APGW's CH function ((3.1)) is given by

$$H^{-1}(v) = \frac{1}{\phi}\left((\kappa+1)\left[\left(1 + \frac{\kappa v}{\lambda(\kappa+1)}\right)^{1/\kappa} - 1\right]\right)^{1/\gamma}, \qquad v > 0. \tag{3.2}$$

The procedure (Algorithm 1) for generating APGW samples follows immediately.

Algorithm 1 Sampling from APGW $(\phi, \lambda, \gamma, \kappa)$ distribution

Generate $v \sim \text{Exp}\,(1)$.
Set $t = H^{-1}(v)$.

3.2.3 Simulating scale data with APGW baseline

Suppose now we want to simulate from the distribution characterised by $H(t|\theta, H_0)$ where we fix $H_0 \sim \text{APGW}\,(\phi, \lambda, \gamma, \kappa)$ and θ is a scale parameter. Using (2.1) and (3.1), we have

$$H(t|\theta) = \lambda\frac{\kappa+1}{\kappa}\left[\left(1 + \frac{(\phi\theta t)^\gamma}{\kappa+1}\right)^\kappa - 1\right].$$

This is just the CH function for APGW $(\theta\phi, \lambda, \gamma, \kappa)$. Therefore we can simply apply Algorithm 1.

3.2.4 Simulating frailty data with APGW baseline

If θ is a frailty parameter, then using (2.3) and (3.1) we have

$$H(t|\theta) = \theta\lambda\frac{\kappa+1}{\kappa}\left[\left(1+\frac{(\phi t)^\gamma}{\kappa+1}\right)^\kappa - 1\right].$$

This is the CH function of APGW $(\phi, \theta\lambda, \gamma, \kappa)$, so again we can revert back to Algorithm 1.

3.2.5 Simulating tilt data with APGW baseline

Now assume θ is a tilt parameter. Combining (2.5) and (3.1) and inverting the resulting expression involves unwieldy algebra, but we can avoid this by using the expression for the CH in terms of a composition ((2.6)) and applying the inversion sampling procedure for a composition of CH functions. For brevity, denote

$$\widetilde{H}_\theta(t) = H_{ll} \circ H_E^\theta \circ H_G(t) = \log(1+\theta(e^t-1))$$

so that $H(t|\theta) = \widetilde{H}_\theta \circ H_0(t)$. A useful property of $\widetilde{H}$, which is easy to show, is that $\widetilde{H}_\theta^{-1} = \widetilde{H}_{1/\theta}$. The sampling procedure is given by Algorithm 2.

Algorithm 2 Simulating tilt data with APGW baseline

Generate $v \sim \text{Exp}(1)$.
Set $\widetilde{t} = \widetilde{H}_{1/\theta}(v)$.
Set $t = H_0^{-1}(\widetilde{t})$.

3.2.6 Simulating reverse-tilt data with APGW baseline

Finally, we suppose θ is a reverse-tilt parameter. Using (2.7) and the notation from Section 3.2.5, $H(t|\theta) = H_0 \circ \widetilde{H}_\theta(t)$. The sampling procedure follows immediately.

Algorithm 3 Simulating reverse-tilt data with APGW baseline

Generate $\widetilde{t} \sim$ APGW using Algorithm 1.
Set $t = \widetilde{H}_{1/\theta}(\widetilde{t})$.

The R code for these sampler is shown in Listing 1. Using the composition form of the CH functions and exploiting the fact that the APGW distribution is closed under the addition of a scale/frailty parameter, we avoid the need to write out several long, complicated expressions. This makes the code concise and reduces the likelihood of typos.

3.3 Survival regression models with APGW baseline

In this section, we implement the survival regression models discussed in Chapter 2 with an APGW baseline using flexsurv. flexsurv is an R package for parametric survival modelling. A

Listing 1 Simulating scale/frailty/tilt/reverse-tilt APGW data.

```
invHapgw <- function(x, phi, lambda, gamma, kappa, log=FALSE){
  (1/phi)*((kappa+1)*((1+kappa*x/(lambda*(kappa+1)))^(1/kappa)-1))^(1/gamma)
}
H_lleg <- function(x, theta){log(1+theta*(exp(x)-1))}

rapgw <- function(n, phi, lambda, gamma, kappa){
  v <- rexp(n=n, rate=1)
  x <- invHapgw(v, phi, lambda, gamma, kappa, log=FALSE)
  return(x)
}
rapgw_scale <- function(n, phi, lambda, gamma, kappa, theta){
  rapgw(n, theta*phi, lambda, gamma, kappa)
}
rapgw_frailty <- function(n, phi, lambda, gamma, kappa, theta){
  rapgw(n, phi, theta*lambda, gamma, kappa)
}
rapgw_tilt <- function(n, phi, lambda, gamma, kappa, theta){
  v <- rexp(n=n, rate=1)
  x <- H_lleg(v, 1/theta)
  x <- invHapgw(x, phi, lambda, gamma, kappa, log=FALSE)
  return(x)
}
rapgw_revtilt <- function(n, phi, lambda, gamma, kappa, theta){
  x <- rapgw(n, phi, lambda, gamma, kappa)
  x <- H_lleg(x, 1/theta)
  return(x)
}
```

comprehensive overview of flexsurv's functionality and principles can be found in [9]. For our purposes, the key features of the package are the ability to implement custom distributions and fit parametric models with flexible covariate effects.

3.3.1 Implementing APGW as a custom distribution in flexsurv

Specify a custom survival distribution in flexsurv involves specifying several pieces of information about the distribution.

First, we provide the hazard function as a vectorised R function. If available analytically, we should also supply the CH function; otherwise it will be estimated using numerical integration and the model fitting process will be slower and less stable.

Next, we assign one of the parameters to be the 'location' parameter. In flexsurv the location parameter is the primary parameter of interest which will depend on covariates by default; all other parameters are called 'ancillary'. The natural candidates for the location parameter of APGW would be ϕ or λ. Which one we choose isn't particularly important, since covariate effects can respecified later on via flexsurv's model fitting function `flexsurvreg`.

Next, flexsurv requires a list of functions mapping the parameters' ranges to $\mathbb{R}$. In our case, ϕ, λ and γ are all positive, so log is one such suitable mapping. Since $\kappa > -1$ it requires a slightly

different mapping (see line 7 of Listing 2).

Finally, we specify plausible initial values for parameters as a function of the data. Suitably chosen values will aid convergence of the optimiser, but again we can override these choices later on in `flexsurvreg` if necessary. For APGW, we choose initial shape parameters $\gamma = \kappa = 1$, which corresponds to an exponential distribution with rate $\lambda\phi$. Then we choose $\phi = 1$ and λ as the median event/censoring time divided by $\log 2$. The initialisation of λ is inspired by the expression for the median of the exponential distribution. Strictly speaking, we should calculate the median of the observed event times, but it wasn't clear from the package documentation how to incorporate censoring information in the initialisation. Listing 2 shows how these pieces of information are passed to flexsurv.

Listing 2 Implementing APGW as a custom flexsurv distribution.

```
1  hapgw <- function(x, phi, lambda, gamma, kappa, log=FALSE){
2    lambda*gamma*phi^gamma*(x)^{gamma-1}*(1+((phi*x)^gamma)/(kappa+1))^(kappa-1)
3  }
4  Hapgw <- function(x, phi, lambda, gamma, kappa, log=FALSE){
5    lambda*(kappa+1)/kappa*((1+(phi*x)^gamma/(kappa+1))^kappa-1)
6  }
7  kappa_trans <- function(x){log(x+1)}
8  kappa_invtrans <- function(x){exp(x)-1}
9
10 custom.apgw <- list(
11   name="apgw",
12   pars=c("phi","lambda","gamma","kappa"),
13   location="lambda",
14   transforms=c(log, log, log, kappa_trans),
15   inv.transforms=c(exp, exp, exp, kappa_invtrans),
16   inits=function(t){c(1, median(t)/log(2), 1, 1)})
```

Maximum likelihood estimation in flexsurv is done using `optim`, R's general-purpose optimisation function. To aid convergence, we can optionally supply analytical derivatives when defining the custom distribution. No major convergence issues arose during the project so we didn't do this, instead allowing R to estimate the gradients numerically, but it would be a useful future addition.

3.3.2 Testing the APGW sampler and flexsurv implementation

Now we run some simulations to verify our algorithms/implementations are working correctly. First, we simulate 5000 uncensored event times from the APGW distribution (with $\phi = 1$, $\lambda = 5$, $\gamma = 2$, $\kappa = -0.2$) using the `rapgw` function in Listing 1. The sample histogram in Figure 3.1 closely matches the true density function, implying the sampler is generating samples from the correct distribution. To test our flexsurv APGW implementation is working correctly, we fit an APGW model to the simulated event time data. We fix $\phi = 1$ and estimate the remaining three parameters, since ϕ and λ tend to be correlated and cannot be reliably estimated together [3]. The parameter estimates in Table 3.1 are consistent with the true values.

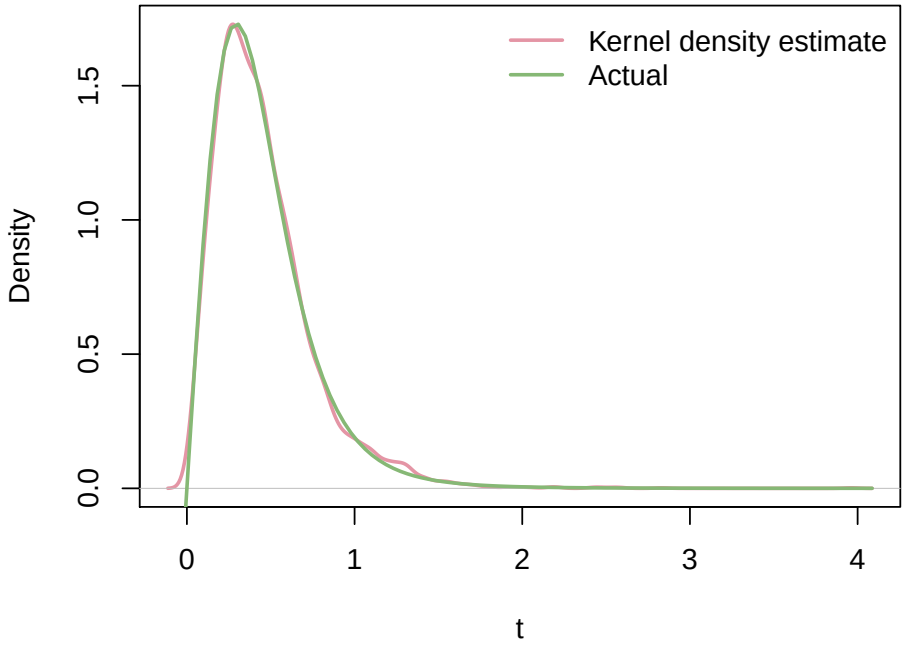


Figure 3.1: Histogram of 5000 APGW samples alongside the APGW density function.

Table 3.1: Comparison of the true and estimated APGW parameters.

Parameter	**Actual**	**Estimate (95% CI)**
λ	5.00	5.04 (4.68, 5.43)
γ	2.00	2.02 (1.96, 2.08)
κ	-0.20	-0.22 (-0.29, -0.13)

3.3.3 Fitting the APGW distribution to the `bc` dataset

Now we test the flexsurv APGW implementation works when we include a covariate. This time, instead of simulating data, we use a real-world dataset. The `bc` dataset is included in the flexsurv package and provides the following information about 686 breast cancer patients [9]:

`censrec`: the censoring indicator, i.e. 1 if the patient died and 0 if their survival time is censored. In total there are 387 (56.4%) censored patients.

`rectime`: the event/censoring time, in years.

`group`: the prognostic group, either `Good`, `Medium` or `Poor`.

The aim is to fit a separate survival curve to each prognostic group. First, we use fit a Weibull model, where the `scale` parameter depends on the `group` covariate, using the in-built functionality of flexsurv. Then we fit the same model using the APGW custom distribution, by fixing $\lambda = 1$ and $\kappa = 1$. If the two models agree, then we can be confident the APGW implementation is working correctly. The results (Table 3.2) for the two models are not immediately comparable since the models use different parametrisations and link functions (see details in [9]). However, they agree once we have accounted for these differences, i.e. the shape parameters are reciprocals of one another, and the covariate effects agree up to sign differences.

Table 3.2: AFT-Weibull regression model parameter estimates on the `bc` dataset using the APGW custom flexsurv distribution and the in-built Weibull distribution.

Parameter	**APGW (flexsurv custom)**	**Weibull (flexsurv built-in)**
-	$\phi = 0.0875$	`scale` $= 11.4229$ ($\Rightarrow$ `scale`$^{-1} = 0.0875$)
-	$\gamma = 1.3796$	`shape` $= 1.3797$
`groupMedium`	0.6136	-0.6136
`groupPoor`	1.2123	-1.2122

3.3.4 Regression models with APGW baseline in flexsurv

Given survival data $\{(t_i, \delta_i, \boldsymbol{x}_i)\}_{i=1}^n$, our goal is to fit an AFT/PH/PO/PGT regression model with APGW baseline, so that each individual has a survival curve defined through the CH function by

$$H_i(t) = H(t|\theta_i, \phi, \lambda, \gamma, \kappa) \tag{3.3}$$

where ϕ, λ ,γ, κ are the APGW baseline parameters common to all individuals and θ_i is the 'proportionality' parameter for individual i, whose interpretation depends on what model we are using. This requires implementing four five-parameter custom distributions in flexsurv, where now θ is the location parameter and is log-linear in the covariates. An example of how this works is shown in Listing 3 for the reverse-tilt family (PGT model). As discussed in Section 3.2.3 and Section 3.2.4, the AFT and PH models are non-identifiable when the baseline is APGW. We resolve this by fixing $\phi = 1$ and $\lambda = 1$ respectively, leaving three unknown baseline parameters.

In the next chapter we will fit these models to simulated data with a single categorical covariate and discuss how to decide which of the modelling assumptions are appropriate.

Listing 3 Implementing the reverse-tilt APGW family as a custom flexsurv distribution.

```
hapgwrevtilt <- function(x, phi, lambda, gamma, kappa, theta, log=F){
  ...
}
Hapgwrevtilt <- function(x, phi, lambda, gamma, kappa, theta, log=F){
  ...
}
custom.apgwrevtilt <- list(
  name="apgwrevtilt",
  pars=c("phi","lambda","gamma","kappa","theta"),
  location="theta",
  transforms=c(log,log,log,kappa_trans,log),
  inv.transforms=c(exp,exp,exp,kappa_invtrans,exp),
  inits=function(t) c(1, median(t)/log(2), 1, 1, 1)
)
```

CHAPTER 4

Simulation Studies

In this chapter, we simulate a dataset that has a single categorical covariate and is known to satisfy the PH assumption with APGW baseline. We explore the use of diagnostic tools to identify valid/invalid modelling assumptions. We then fit and interpret the regression models developed in Chapter 3.

4.1 Methodology for simulating survival data in groups

For simplicity, we consider survival data with a single categorical covariate, so that the observations form distinct groups. This type of data is common in practice [16]. We have already seen an example of this in the `bc` dataset, where there were three groups corresponding to the patients' prognoses. In Chapter 5 we see another example from engineering, where the groups correspond to the lengths of carbon fibres. Note that the term 'grouped survival data' is best avoided here, since that may also refer to data with interval-censored event times.

Suppose we have K groups and that each group $1 \leq k \leq K$ has n_k individuals, so that the total number of individuals is $n = \sum_{k=1}^{K} n_k$. The event times of individuals in group $k = 1, \ldots, K$ are distributed according to CH function $H_k(\cdot) = H(\cdot|\theta_k, H_0)$, where H_0 is the CH function of the baseline distribution $T_0 \sim \text{APGW}(\phi, \lambda, \gamma, \kappa)$ and θ_k is the group k proportionality constant (relative to the baseline). We arbitrarily assign group $k = 1$ as our reference level, with proportionality constant θ_1. The proportionality constants for groups $k = 2, \ldots, K$ are then defined through a vector of regression parameters $\boldsymbol{\beta} = (\beta_2 \ldots, \beta_K)^T \in \mathbb{R}^{K-1}$ by $\theta_k = \theta_1 \exp(\beta_k)$.

The censoring mechanism has two elements. First, censoring times are generated from an exponential distribution. To ensure that censoring is non-informative, the rate parameter is common to all individuals. Second, there is administrative censoring at time $t_{\max}$, representing the end of the study observation period. The implementation of this methodology is shown in Listing 4 and can be used simulate survival data satisfying any of the AFT, PH, PO or PGT assumptions by replacing `rapgw_<<family>>` in line 6 with one of the functions from Listing 1.

There are many ways we could've encoded the categorical variable and parameterised the model. The notation adopted here, where the first level of a factor variable as the reference level and other levels are interpreted relative to this reference level via regression parameters, is consistent with the default contrast coding system in R. This is convenient for interpreting R output. Moreover, it is often a natural choice in practice, e.g. in a clinical trial treatment groups are compared with the placebo group.

Listing 4 Simulating survival data in groups

```
sim_df <- function(K, nk, t_max, c_rate, phi, lambda, gamma, kappa, theta_1,
    beta, family){
  # setup groups and compute theta values
  group <- rep(1:K, nk)
  theta <- ifelse(group==1, theta_1, theta_1*exp(beta[group-1]))
  # simulate event times
  t <- mapply(rapgw_<<family>>, phi, lambda, gamma, kappa, theta, MoreArgs=list(
    n=1))
  # censoring
  c <- rexp(n=sum(nk), rate=c_rate)
  c <- pmin(c, t_max)
  delta <- 1*(t<c)
  t <- pmin(t,c)
  # create dataframe
  group <- as.factor(group)
  df <- data.frame(group, t, delta)
  return(df)
}
```

`K`: number of groups
`nk`: group sizes, a K-vector
`t_max`: administrative censoring time
`c_rate`: exponential rate parameter for censoring
`phi,lambda,gamma,kappa`: APGW baseline parameters
`theta_1`: proportionality constant for group 1
`beta`: regression parameters, a $(K-1)$-vector
`family`: interpretation of the proportionality constants, a string (e.g. `"scale"`)
`rapgw_<<family>>`: an R function that simulates from the relevant APGW family

4.2 Simulation study: frailty data

In this section, we simulate data where the groups have proportional hazards. Then we investigate the use of diagnostic tools to deduce (from the data) that PH is the correct modelling assumption. Finally, we fit and compare some regression models. We will also investigate how an shortening the study observation period impairs our ability to discriminate between modelling assumptions.

4.2.1 Data simulation

We simulate frailty data with three groups of size 300 and baseline distribution $T_0 \sim \text{APGW}(\phi = 1, \lambda = 1, \gamma = 4, \kappa = -0.2)$. The proportionality constants are defined by $\theta_1 = 2$, $\beta_2 = 0.05$, $\beta_3 = 0.6$. The censoring distribution has exponential rate parameter $c_{\text{rate}} = 0.1$ and administrative censoring time $t_{\max} = 1.5$. Figure 4.1 shows the Kaplan-Meier estimates of the survival curves for each group. We observe that the survival curves for groups 1 and 2 are very similar because β_2 is close to zero.

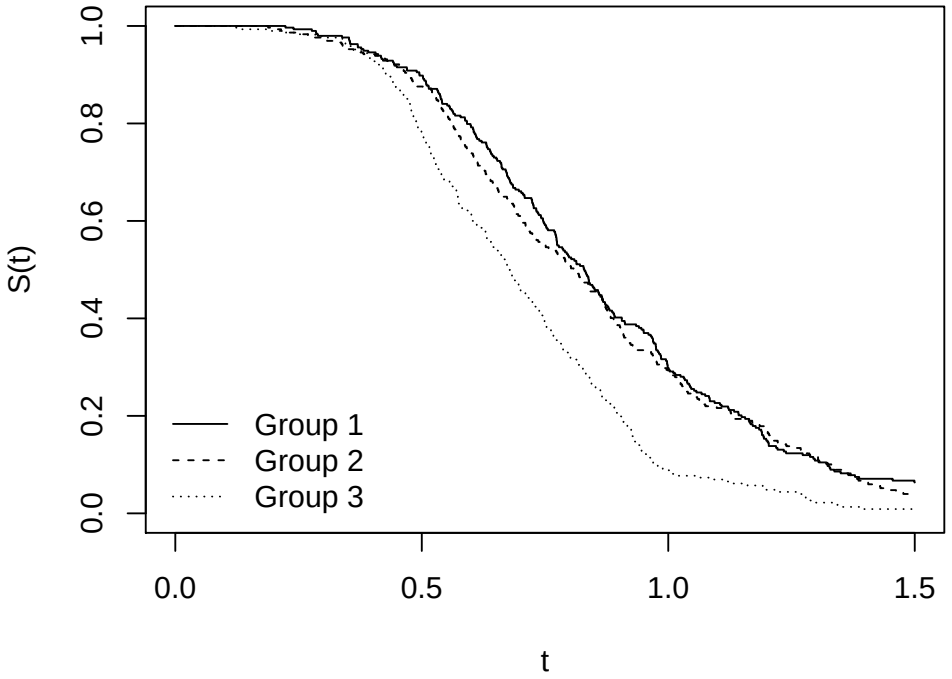


Figure 4.1: Kaplan-Meier estimates of the survival curves.

4.2.2 PH diagnostics

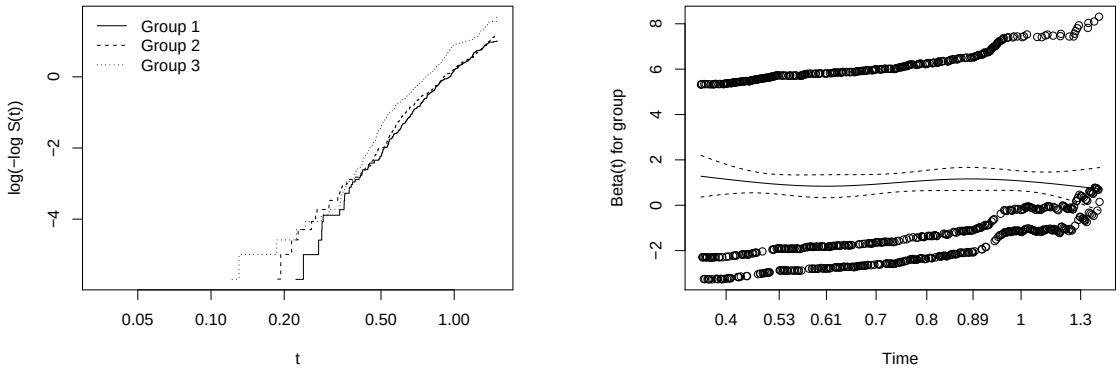
Figure 4.1 is useful for getting a feel for the data, but isn't particularly helpful for ascertaining whether a particular proportionality assumption is reasonable or not. For example, determining whether the hazards are proportional involves judging whether the survival curves are related by a power law, which is not something we can intuitively judge. The idea behind graphical diagnostics is to apply transformations to the data so that it is easier to judge whether a particular relationship holds or not, e.g. the assumption holds if the transformed curves are (approximately) parallel. Of course, reading graphical diagnostics involves some subjectivity and different observers may reach different conclusions, especially if the KM estimates are not very smooth due to small samples or heavy censoring.

There are two popular ways to check the validity of PH assumption [**kl**]:

(i) A graphical approach involves plotting $\log(-\log \hat{S})$, known as the log-log survival curve. Under the PH assumption,

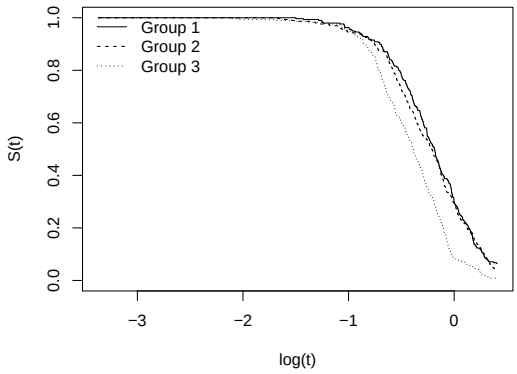
$$\begin{aligned} & H(t|\theta) = \theta H_0(t) \\ \implies & -\log S(t|\theta) = -\theta \log(S_0(t)) \\ \implies & \log(-\log S(t|\theta)) = \log(-\log(S_0(t))) + \log(\theta) \end{aligned}$$

and therefore the log-log survival curves are parallel. Hence, for our grouped data we should check whether the log-log survival curves for each group are approximately parallel. A log transformation of the time axis often produces a tidier plot and has no effect on the pattern of vertically shifted curves.

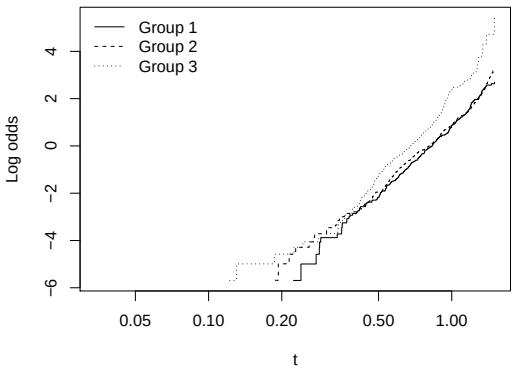


(a) Log-log survival curves.

(b) Schoenfeld residuals.



(c) Log-time survival curves.



(d) Log odds curves.

Figure 4.2: Diagnostic plots to check the AFT, PH and PO assumptions for the simulated (PH) data.

(ii) Perform a statistical test based on the Schoenfeld residuals. Suppose individual i has an observed event time t_i. Then the Schoenfeld residual for individual i and covariate k is given by the covariate value x_{ik} minus the expected value for the covariate for the risk set at t_i. Under the PH assumption, the Schoenfeld residuals should uncorrelated with survival time. We can test this by plotting the residuals and by performing a statistical test on the correlation.

Figure 4.2a shows that the log-log survival curves are approximately parallel, suggesting PH is reasonable. Likewise the fitted curve in Figure 4.2b is fairly flat and straight, suggesting the (scaled) Schoenfeld residuals are uncorrelated with time. The p-value for the statistical test (using the `cox.zph` function) is 0.84, so we do not reject PH.

4.2.3 AFT diagnostics

Similarly, we will consider two approaches for checking the AFT assumption:

(i) Under AFT, the KM estimates of the group-wise survival curves plotted against $\log t$ should be horizontally shifted copies of each other.

(ii) Under AFT, the distributions of $\log T$ for each group are the same up to a location shift, meaning the central moments should be equal. In particular, we can test whether the

Table 4.1: Parameter estimates and goodness-of-fit measures for the fitted regression models. (* indicates the parameter was fixed when fitting the model.)

Model	ϕ	λ	γ	κ	θ	β_2	β_3	**Log-likelihood**	**AIC**
AFT	-	1.00*	3.91	0.16	1.17	0.03	0.21	-194.15	398.30
PH	1.00*	-	3.77	-0.09	1.82	0.07	0.60	-192.93	395.86
PO	1.00*	0.38	4.17	0.64	5.65	0.15	0.89	-195.91	403.82
PGT	1.00*	0.49	4.70	0.33	1.70	0.05	0.29	-196.63	405.26

variances of the log survival times for each group are equal using Bartlett's test (for details see [11]). The test statistic, which is a function of the groups' sample variances, follows a χ^2_K distribution under the null (data is AFT).

The PH and AFT assumptions coincide if and only if the baseline is Weibull [4]. Our data is PH and the APGW shape parameters are $\gamma = 4$ and $\kappa = -0.2$ so the baseline is non-Weibull. This means that, theoretically, the AFT assumption doesn't hold.

In Figure 4.2c it could be argued that the Group 3 log-time survival curve is slightly steeper than the Group 1/2 curves, but overall the plot is probably inconclusive. The sample variances of the log event times for each group (ignoring censored observations) are 0.134, 0.162 and 0.135, which gives a Bartlett test statistic equal to 3.180 and a p-value of 0.204, which is insignificant. Therefore neither diagnostics lead us to rule out an AFT model.

4.2.4 PO diagnostics

Under the PO assumption, the odds functions are proportional and therefore the log-odds curves plotted against t (or $\log t$) should be parallel.

In this case, the diagnostic plot (Figure 4.2d) seems quite conclusive. The log-odds curves do not appear to be parallel: the vertical distance between the Group 3 curve and the curves for Groups 1/2 increases with time, implying the odds are non-proportional. We note that the non-parallelism is most apparent in the right-hand tail. If we were prevented from seeing this region of the graph, perhaps due to a short study observation period, the diagnostic would be less conclusive. This observation will be investigated further in Section 4.2.6.

4.2.5 Fitting regression models

The diagnostics suggest PH and AFT are reasonable modelling assumptions, so we now fit PH and AFT regression models. For comparison, we also fit PO and PGT models. Table 4.1 shows the parameter estimates and some goodness-of-fit measures for each model.

The best fitting model (largest log-likelihood, smallest AIC) is the PH model, as expected. Moreover, the estimates of the baseline parameters and regression coefficients are reasonably close to the true values. Figure 4.3 shows that the CH function of the fitted model agrees closely with the non-parametric estimate of the CH function for each group.

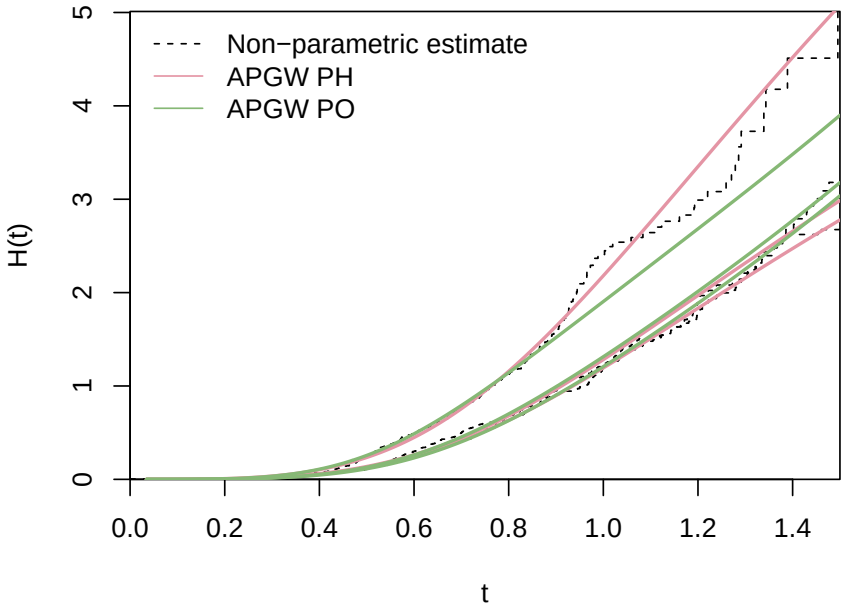


Figure 4.3: The CH functions of the fitted PH and PO models compared with the estimate from the data.

As well as parameter estimates, flexsurv provides asymptotic confidence intervals. These are calculate from the observed information matrix. The 95% confidence interval β_2 is $(-0.10, 0.24)$, which means the 95% confidence interval for the hazard ratio between groups 2 and 1 is $(\exp(-0.10), \exp(0.24)) = (0.91, 1.27)$. Since 1 is contained in the interval, we cannot reject the hypothesis that Group 1 and Group 2 have the same survival curves. Indeed, we noted earlier that the KM estimates for the survival curves in Figure 4.1 are virtually identical, so this conclusion is unsurprising.

The AFT model is the second best in terms of fit, and the PO and PGT models are the worst. These findings are in accordance with what we observed in the diagnostic plots: the data violates the PO assumption and therefore we cannot expect a PO model to fit the data.

The PO model is constrained by the fact that the hazards of the groups must converge in the long run. We can see this occurring in Figure 4.3. Within this constraint the model behaves sensibly: convergence of the hazards is achieved by underestimating the Group 3 hazard, sacrificing a good fit for a minority while maintaining a good fit for a large number of individuals in Group 1 and Group 2.

4.2.6 Effect of varying the censoring mechanism

A PO model will generally fit PH data poorly in the long run, because under PO the hazards converge whereas under PH they are proportional at all times. Conversely, from Section 2.3 we know that as $t \to 0$ a tilt parameter behaves like a frailty parameter. From these facts, we claim that a PO model may be able to achieve a good fit for PH data provided the observation period is short. To test the claim, we perform the following experiment:

(i) For a given value of $t_{\max}$, simulate 20 PH datasets. The parameters for the simulation are

the same as those we used previously.

(ii) Fit a PH and PO model to each dataset.

(iii) For each model, compute the AIC to measure the goodness-of-fit.

(iv) Compute the mean AIC values achieved by the PH and PO models.

(v) Repeat for values of $t_{\max}$ from 0.5 to 2.5 in increments of 0.2.

Figure 4.4 shows the difference in goodness-of-fit achieved by the models increases as the observation period becomes longer, supporting the claim. This finding has an implication for data collection: if we fail to run a sufficiently long study, we may be unable to discern appropriate modelling assumptions, potentially leading to flawed analyses and conclusions.

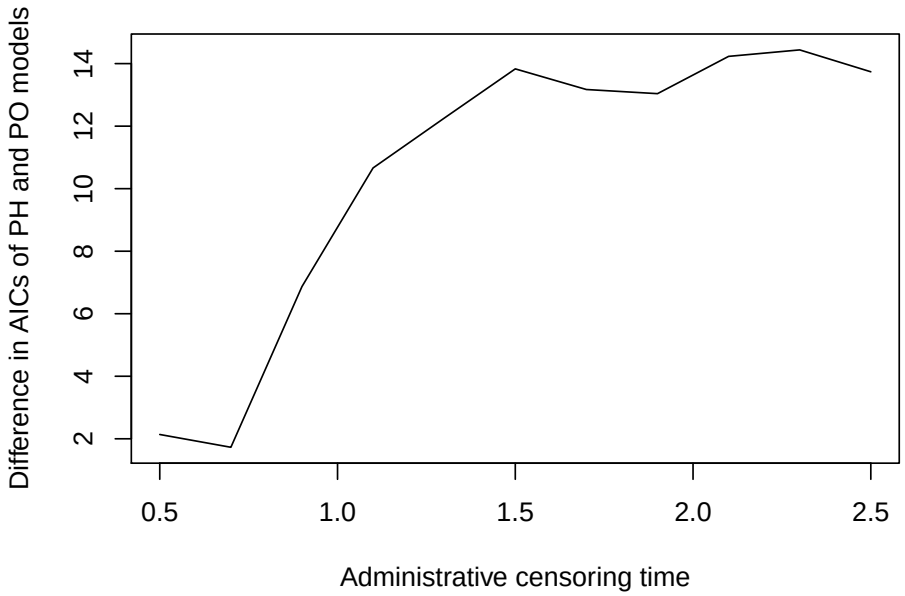


Figure 4.4: Difference in AIC achieved by PH and PO models as the administrative censoring time is varied. Results averaged over 20 simulations.

CHAPTER 5

Real Data Experiments

In this chapter, we repeat the process of performing diagnostics and fitting models using real-world datasets from engineering and medicine. The latter will require modifying our approach to deal with continuous covariates.

5.1 Carbon fibres strength data

The data considered in this section is sourced from [4, Table 4.1]. In experiments conducted by the University of Surrey, the failure stresses (in GPa) of 244 carbon fibres of varying lengths (1mm, 10mm, 20mm and 50mm) were measured. The goal is to understand the relationship between fibre length and failure stress. This dataset fits nicely with our framework. We have a single covariate, fibre length, which takes four distinct values and therefore defines four separate groups. The target variable, failure stress, is non-negative and so can viewed as a 'time'. All fibres were tested to breaking point so there is no censoring.

Figure 5.1 shows the empirical survival curves for each fibre length. From this we observe that shorter fibres tend to withstand greater stress before failure, but diagnostic tools are required to determine the nature of this relationship more precisely.

The diagnostic plots discussed in Chapter 4 are replicated for the fibres data in Figure 5.2. Figure 5.2a and Figure 5.2b show that the PH assumption does not hold. The log-log survival curves are not parallel (e.g. the 10mm curve is very curved, while the others are reasonably straight) and the fitted curve for the Schoenfeld residuals decreases with time. Figure 5.2c suggests AFT is a reasonable modelling assumption, as the log-time survival curves are approximately location shifts of one another. Moreover, the result of the Bartlett test for equal variances is insignificant [4], with a p-value of 0.87. Finally, we reject PO because the log-odds curves in Figure 5.2d are non-parallel (the 1mm and 10mm curves are concave and convex respectively).

Table 5.1: Parameter estimates of the AFT model for the fibres data.

Parameter	**Estimate**	**95% CI**
γ	6.99	(5.87, 8.32)
κ	0.28	(0.05, 0.57)
θ	0.23	(0.22, 0.24)
β_{10}	0.33	(0.26, 0.40)
β_{20}	0.53	(0.46, 0.61)
β_{50}	0.63	(0.56, 0.70)

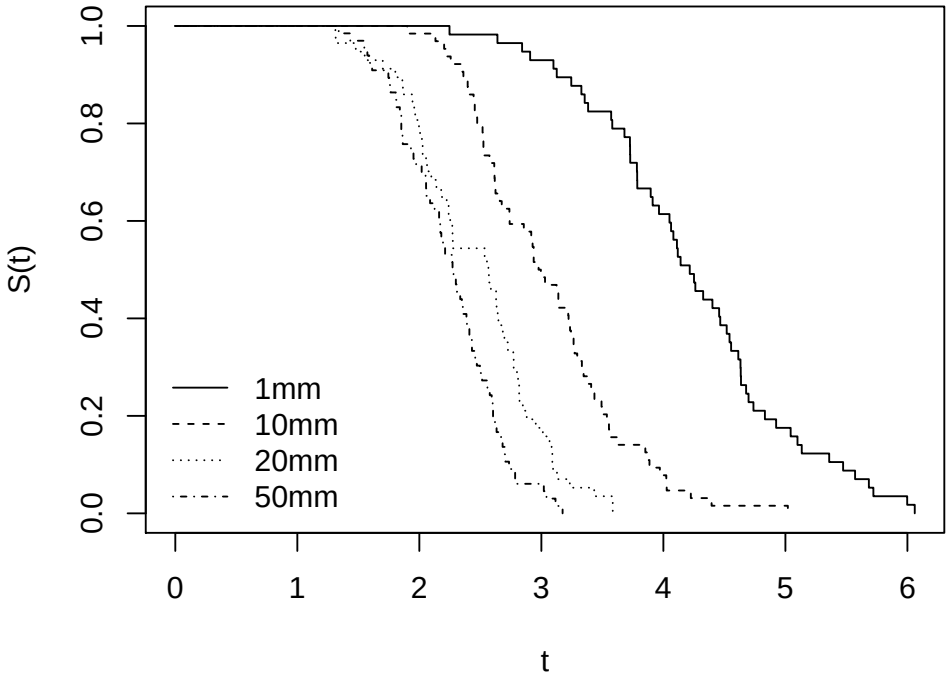


Figure 5.1: KM estimates for the survival curves of the fibres dataset.

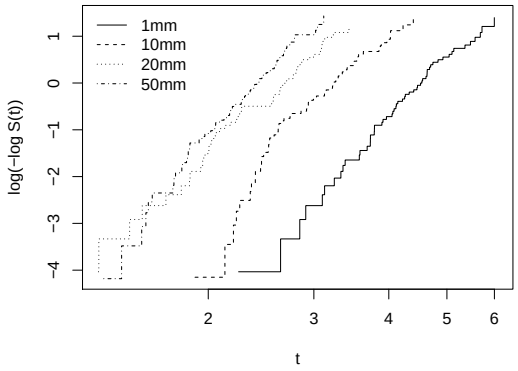
The next step is to fit an AFT regression model using APGW as the baseline. We take 1mm as the reference level and β_{10}, β_{20}, β_{50} are regression parameters defined in the same way as in Section 4.1. The estimates and confidence intervals are shown in Table 5.1.

The $\hat{\beta}$ values increase with length, meaning longer fibres are more likely fail at lower stresses. Since we are using an AFT model, the parameters can be interpreted in terms of an acceleration/deceleration of the progression towards the event of interest. For example, we have $\hat{\beta}_{10} = 0.33$, and so $\hat{\theta}_{10}/\hat{\theta}_1 = \exp(\hat{\beta}_{10}) = 1.39$. This means that the progression of a 10mm to failure is accelerated (here we imagine the stress steadily increasing) by a factor of 1.39 relative to a 1mm fibre. We can similarly compute acceleration factors of $\exp(\hat{\beta}_{20}) = 1.70$ and $\exp(\hat{\beta}_{50}) = 1.88$ for 20mm and 50mm fibres respectively.

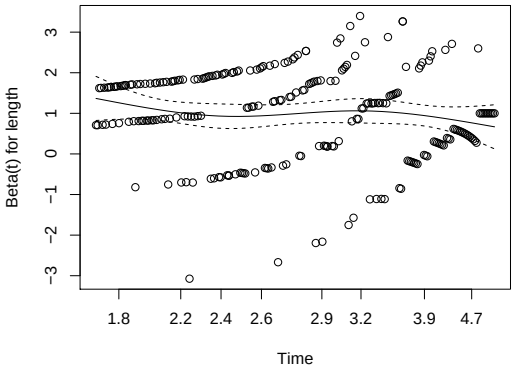
5.2 Primary biliary cirrhosis (PBC) data

The `pbc` dataset is commonly used in survival analysis and is built into the survival package. It comes from a clinical trial which was testing a treatment for primary biliary cirrhosis (PBC) of the liver. The time (in days) until death or censoring was recorded for 418 patients, along with 17 covariates, including the treatment variable, basic information about the patient (age, sex, etc.) and a variety of clinical measurements. Full details of the covariates can be found in [6], where the dataset was originally studied. We restrict our analysis to the five important covariates identified by Fleming and Harrington:

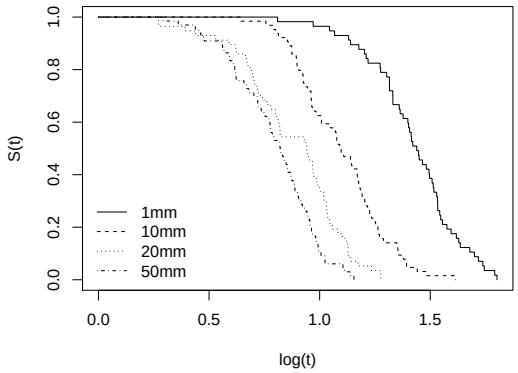
`age`: age in years (continuous).
`edema`: presence of edema (categorical).



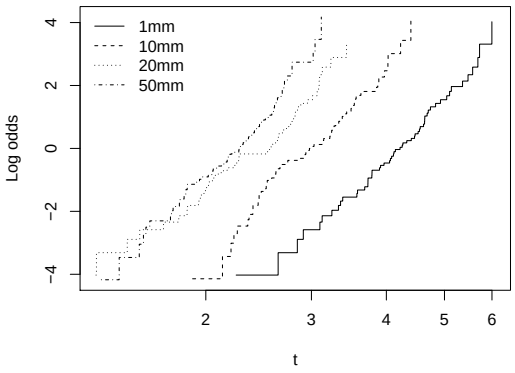
(a) Log-log survival curves.



(b) Schoenfeld residuals.



(c) Log-time survival curves.



(d) Log odds curves.

Figure 5.2: Diagnostic plots to check the AFT, PH and PO assumptions for the simulated (PH) data.

`albumin`: albumin in g/dl (continuous).

`bili`: serum bilirubin in mg/dl (continuous).

`protime`: prothrombin time in seconds (continuous).

Before the analysis, we perform the same pre-processing steps as Martinussen and Scheike, so that we compare our results. This involves converting the `time` variables to years, taking log transforms of the `albumin`, `bili` and `protime` variables, and centring continuous variables around their means.

The `pbc` dataset has continuous covariates, so we do not have a group structure like in the previous examples and our graphical diagnostic tools cannot be readily applied in this case. Martinussen and Scheike employ sophisticated diagnostics based on comparing the observed score process with score processes simulated using the modelling assumption being examined. Our approach is very simple and crude: simply convert the continuous variables into categorical ones and apply the same tools we used earlier. We now illustrate this approach by checking the PO assumption for log(albumin) and log(protime).

Figure 5.3 shows the histograms of the centred log(albumin) and log(protime) variables. These plots inform our decision as to how we might group patients. For log(albumin) we define two groups using breakpoint of -0.05, i.e. if log(albumin) < -0.05 then the patient is put in the

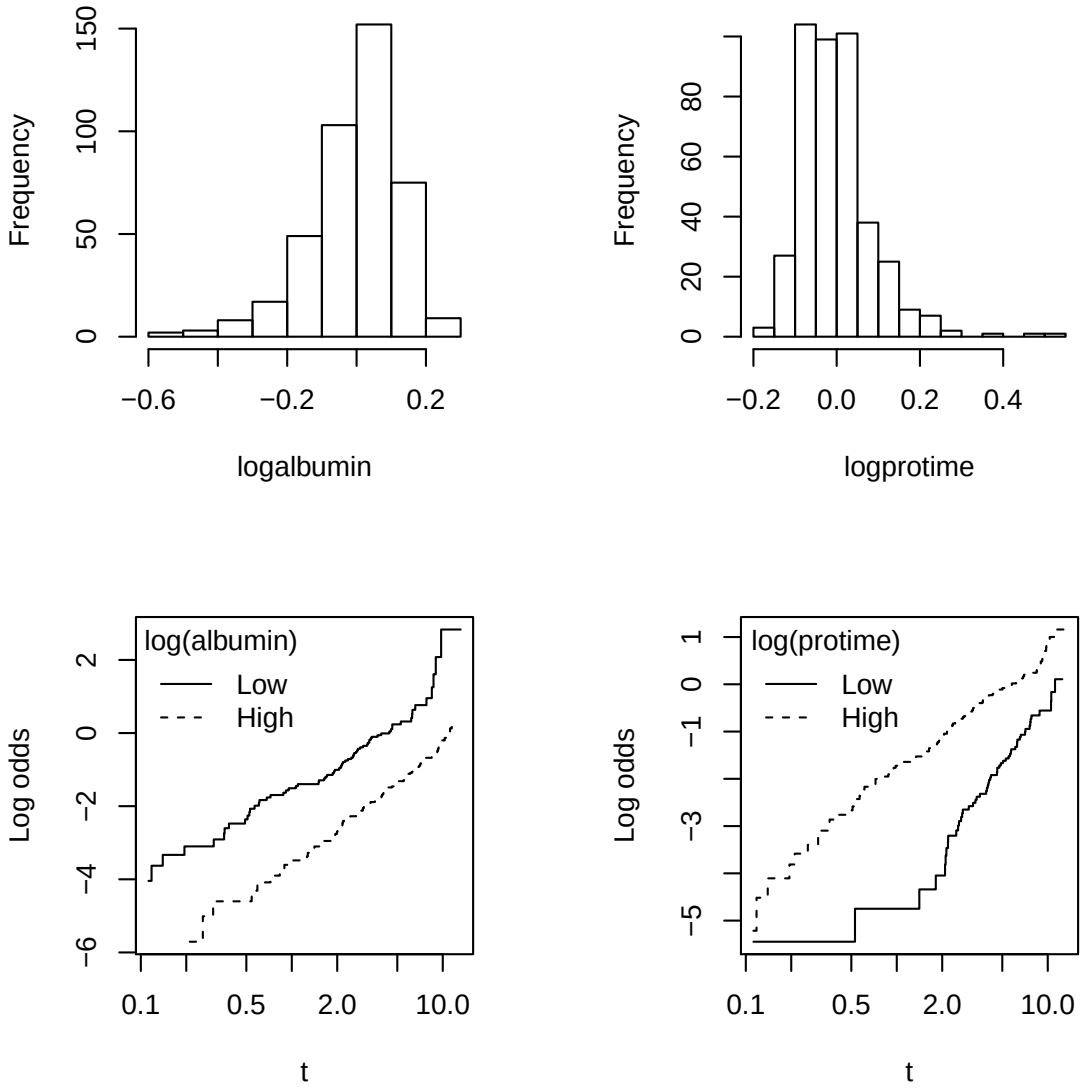


Figure 5.3: Histograms and log-odds curves for log(albumin) and log(protime).

'Low' group, otherwise he is assigned to the 'High' group. Similarly we define two groups for log(protime) using 0 as the breakpoint. Here the breakpoints were set fairly arbitrarily but one can imagine there are more sophisticated approaches doing this, e.g. based on quartiles. There is also a trade-off involved in deciding the number of groups: a more granular grouping (e.g. 'Low', 'Medium', 'High') would certainly be preferable, but using a higher number of groups necessarily reduces the size of the groups, meaning the KM estimates are less smooth and the diagnostic plots will be harder to read. Based on these newly-defined categorical variables, we can plot the log-odds functions (Figure 5.3). The log(protime) curves are clearly not parallel, so a PO model provides a poor description the effect of the log(protime) covariate. On the other hand, an optimistic observer might conclude that the PO assumption is reasonable for log(albumin). These findings agree with the conclusions of Martinussen and Scheike's more sophisticated techniques.

Table 5.2: PO model parameter estimates for `pbc`. The values in the first column are the estimates using our fully-parametric method with APGW baseline and $\lambda = 1$ fixed. The second column shows the regression coefficients estimated using semi-parametric modelling in [14, p. 306]

Parameter	**Parametric (APGW baseline)**	**Semi-parametric**
ϕ	0.071	-
γ	1.69	-
κ	7.55	-
θ	0.076	-
β(`age`)	0.055	0.051
β(`edema`)	1.40	1.28
β(`logBili`)	1.17	1.12
β(`logAlbumin`)	-3.09	-2.95
β(`logProtime`)	3.92	3.94

In spite of the fact that the PO assumption is not valid for all the covariates, we now fit a PO regression model with APGW baseline. Here we encounter some convergence issues when fitting the model, due to poor initial guesses for the parameters causing an error in optim. To find better choices for the initialisation we fit a Weibull model, which uses analytical derivatives and so converges without issues, and then use the given Weibull model as our initial guess for the APGW model. However, even with a better initialisation the optimiser didn't converge completely to the maximum. This could probably be resolved by tuning the `control` arguments in `flexsurvreg`, but we didn't try this and just accepted slightly inaccurate parameter estimates. The confidence intervals given by flexsurv are based on the Hessian at the maximum, the convergence issues mean they are wrong too and so we ignore them.

The parameter estimates are shown in Table 5.2. The estimates of the regression coefficients agree quite strongly with those obtained by Martinussen and Scheike using semi-parametric methods. We find that, for example, an increase in the prothrombin time (protime) by exp(1) seconds increases the event odds by a factor of $\exp(3.92) = 50.4$. From a clinical point of view this makes sense, because a long prothrombin time signifies issues with blood clotting [10]. Conversely, the log(albumin) coefficient is negative because a low albumin levels is indicative of liver disease [17].

CHAPTER 6

Conclusion

6.1 Conclusion

Parametric survival modelling offers a framework to analyse data from a variety of fields, such as engineering and medicine. While the PH model is extremely prevalent in practice, there are several alternatives available, and we can use simple graphical diagnostics to decide which might be most appropriate. Each model has an intuitive interpretation, so it is straightforward for practitioners to explain the results of their statistical analyses. flexsurv provides a very flexible framework for fitting parametric models: we can specify any distribution as the baseline (so long as we provide a valid hazard/density function) and the user has a lot of control over the relationship between the covariates and the distributional parameters.

6.2 Future work

Here are a few ideas for future work and topics that would be interesting to look into:

(i) The computational aspect of the project was frustrating at times due to a lack of unification between the various R packages for survival analysis. For example, the plotting method in flexsurv is much less versatile than the plotting methods in survival or rms. Producing plots such as the one in Figure 4.3 is impossible using flexsurv alone, we have to use survival too. Developing flexsurv's functionality in areas like this would be helpful.

(ii) Resolve the minor convergence issues (e.g. when fitting the PO model to the pbc dataset) by supplying the analytical derivatives to flexsurv.

(iii) Consider a wider variety of regression models. Davis [5] reviews an extensive list of regression models. Of particular interest would be the so-called 'extended regression models', which combine the assumptions mentioned in this report to generate new models, e.g. PH-PO or PO-AFT. We could also experiment with different baseline distributions.

(iv) Consider data from a wider range of fields, e.g. the European forestry data, which we had hoped to analyse but didn't get time.

(v) Learn how to incorporate time-dependent covariates in regression models and how to implement these in R using the timereg package. This topic is treated in [14].

(vi) Learn about (or develop) diagnostics for continuous covariates, such as the diagnostic based on score processes in [14]. The method used to discretise the continuous variables in the pbc dataset was very crude and could definitely be improved upon.

Bibliography

[1] Søren Asmussen and Peter Glynn. *Stochastic Simulation: Algorithms and Analysis*. Vol. 57. Stochastic Modelling and Applied Probability. Springer, 2007. ISBN: 978-0-387-69033-9.

[2] Steve Bennett. "Analysis of survival data by the proportional odds model". In: *Statistics in Medicine* 2.2 (1983), pp. 273–277. DOI: 10.1002/sim.4780020223.

[3] Kevin Burke, Chris Jones, and Angela Noufaily. "A flexible parametric modelling framework for survival analysis". In: *Applied Statistics Series C* 69.2 (2020), pp. 429–457. DOI: 10.1111/rssc.12398.

[4] Martin Crowder et al. *Statistical Analysis of Reliability Data*. 1st ed. Chapman and Hall/CRC, 1991. ISBN: 9780412594809.

[5] Alice Davis. "Modelling techniques for time-to-event data analysis". Doctoral dissertation. University of Bath, 2018.

[6] Thomas Fleming and David Harrington. *Counting Processes and Survival Analysis*. Wiley Series in Probability and Statistics. Wiley, New York, 2005. ISBN: 9780471522188.

[7] Timothy Hanson and Wesley Johnson. "A Bayesian Semiparametric AFT Model for Interval-Censored Data". In: *Journal of Computational and Graphical Statistics* 13.2 (2012), pp. 341–361. DOI: 10.1198/1061860043489.

[8] Philip Hougaard. "Frailty models for survival data". In: *Lifetime Data Analysis* 1 (1995), pp. 255–273. DOI: 10.1007/BF00985760.

[9] Christopher Jackson. "flexsurv: a platform for parametric survival modeling in R". In: *Journal of Statistical Software* 70.8 (2016), pp. 1–33. DOI: 10.18637/jss.v070.i08.

[10] Arif Kamal, Ayalew Tefferi, and Rajuv Pruthi. "How to Interpret and Pursue an Abnormal Prothrombin Time, Activated Partial Thromboplastin Time, and Bleeding Time in Adults". In: *Mayo Clinic Proceedings* 82.7 (2007), pp. 864–873. DOI: 10.4065/82.7.864.

[11] John Kennedy and Adam Neville. *Basic Statistical Methods for Engineers and Scientists*. 3rd ed. Harper & Row, 1986. ISBN: 9780060436339.

[12] Damjan Krstajic. *How real is the random censorship model in medical studies?* Serbia: Research Centre for Cheminformatics, 2017.

[13] Albert Marshall and Ingram Olkin. *Life Distributions*. Springer Series in Statistics. Springer, 2007. ISBN: 978-0-387-20333-1.

[14] Torben Martinussen and Thomas Scheike. *Dynamic Regression Models for Survival Data*. 1st ed. Statistics for Biology and Health. 2006. ISBN: 978-0-387-20274-7.

[15] John Taulbee. "A general model for the hazard rate with covariables". In: *Biometrics* 35.2 (1979), pp. 439–450. DOI: 10.2307/2530346.

[16] Nora Villanueva, Marta Sestelo, and Luìs Meira-Machado. “A method for determining groups in multiple survival curves”. In: *Statistics in Medicine* 38.5 (2019), pp. 866–877. DOI: 10.1002/sim.8016.

[17] Florence Wong. “Drug Insight: the role of albumin in the management of chronic liver disease”. In: *Nature Clinical Practice Gastroenterology & Hepatology volume* 4 (2007), pp. 43–51. DOI: 10.1038/ncpgasthep0680.