

Chapter 13

Time to Event Regression

We now introduce covariates onto our time to event analysis via the *accelerated failure time (AFT)* model and the *proportional hazards (PH)* model. Throughout, our data take the form $D = (y, \delta)$ with $y = (y_1, \dots, y_n)'$, where $y_i = \min(T_i, C_i)$, and $\delta = (\delta_1, \dots, \delta_n)'$, where $\delta_i = I_{[0, \infty)}(C_i - T_i) \equiv I(T_i \leq C_i)$. The censoring times are assumed to be noninformative and to be independent of event times. In addition, we now have x_i , an r vector of predictor information, associated with individual i .

13.1 Accelerated Failure Time Models

The AFT model is a log-linear regression model for event times $T_1, \dots, T_n$. The AFT model is

$$\log(T_i) = x_i' \beta + \sigma \varepsilon_i; \quad \varepsilon_i \stackrel{iid}{\sim} F_\varepsilon(\cdot) \quad (1)$$

where $\sigma = 1/\sqrt{\tau}$ is a scale parameter and F_ε is a known cdf defined on the real line with corresponding density and hazard, f_ε and h_ε . When model (1) holds, write

$$T_i \stackrel{iid}{\sim} AFT(F_\varepsilon, \beta, \tau | x_i).$$

When $F_\varepsilon = \Phi$, the cdf of a $N(0, 1)$, model (1) reduces to the linear regression model of Chapter 9 applied to the log-transformed event times. Of course, Chapter 9 did not deal with censored data. The other distributions used for F_ε include the logistic and extreme value (complementary log-log) distributions, cf. Table 13.1. These are the same three distributions used in Section 8.1 to define binomial regression models. We typically include an “intercept” parameter in $x_i' \beta$ by setting $x_{i1} \equiv 1$ and standardize the other variables. Defining $e_1 = (1, 0, \dots, 0)'$, with standardized variables $\beta_1 = e_1' \beta$ is the regression parameter for a “standard” individual in the study.

We require the survivor function for an individual with event time T and covariate information x , that is, the survival function when $T \sim AFT(F_\varepsilon, \beta, \tau | x)$.

Table 13.1: Common baseline distributions for AFT models.

Baseline Distribution	$f_\varepsilon(v)$	$F_\varepsilon(v)$	$W \equiv e^\varepsilon$
Normal	$\frac{1}{\sqrt{2\pi}} e^{-0.5v^2}$	$\Phi(v)$	Log-normal
Logistic	$\frac{e^v}{(1+e^v)^2}$	$\frac{e^v}{1+e^v}$	Log-Logistic
Extreme Value	$e^v \exp[-e^v]$	$1 - \exp[-e^v]$	Weib(1,1)

Model (1) gives

$$\begin{aligned}
S(t|x, \beta, \tau) &= \Pr(T > t | x, \beta, \tau) \\
&= \Pr(\log(T) > \log(t) | x, \beta, \tau) \\
&= \Pr[(\log(T) - x'\beta)\sqrt{\tau} > (\log(t) - x'\beta)\sqrt{\tau} | x, \beta, \tau] \\
&= \Pr[\varepsilon > (\log(t) - x'\beta)\sqrt{\tau} | x, \beta, \tau] \\
&= 1 - F_\varepsilon[(\log(t) - x'\beta)\sqrt{\tau}].
\end{aligned} \tag{2}$$

The corresponding density and hazard functions derived from this are:

$$f(t|x, \beta, \tau) = \frac{\sqrt{\tau}}{t} f_\varepsilon[(\log(t) - x'\beta)\sqrt{\tau}] \tag{3}$$

$$h(t|x, \beta, \tau) = \frac{\sqrt{\tau}}{t} h_\varepsilon[(\log(t) - x'\beta)\sqrt{\tau}]. \tag{4}$$

These functions depend on F_ε . If ε is standard normal,

$$f(t|x, \beta, \tau) = \frac{1}{\sqrt{2\pi}} \frac{\sqrt{\tau}}{t} \exp\left\{-\frac{\tau}{2}[\log(t) - x'\beta]^2\right\},$$

so similar to Example 12.2.3, we see that $T \sim LN(x'\beta, 1/\tau)$. Using the extreme value cdf for F_ε , it is not difficult to show that $T \sim \text{Weib}(\sqrt{\tau}, e^{-x'\beta\sqrt{\tau}})$, which is an exponential distribution whenever $\tau = 1$. To see this, substitute the extreme value cdf for F_ε into (2) and simplify to recognize the Weibull survival function from Example 12.2.2.

The likelihood function based on a sample is constructed by substituting (2) and (3) into (12.1.2). Of course as indicated by Exercise 12.2(b), we could just as well use the distribution of $\log(T)$.

We have previously noted that the median of the $LN(\mu, 1/\tau)$ distribution was e^μ . The same argument gives the median of a $LN(x'\beta, 1/\tau)$ distribution as $e^{x'\beta}$. In an arbitrary AFT model, let $\tilde{t}_\varepsilon$ satisfy

$$0.5 = F_\varepsilon(\tilde{t}_\varepsilon).$$

Now, for $T \sim AFT(F_\varepsilon, \beta, \tau|x)$, the median $\tilde{t}$ satisfies

$$0.5 = S(\tilde{t}|x, \beta, \tau).$$

From (2),

$$0.5 = 1 - F_\varepsilon[(\log(\tilde{t}) - x'\beta)\sqrt{\tau}],$$

so we must have

$$\tilde{t}_\varepsilon = (\log(\tilde{t}) - x'\beta)\sqrt{\tau}$$

or

$$\tilde{t} = \exp(x'\beta + \tilde{t}_\varepsilon/\sqrt{\tau}).$$

We are often interested in $\tilde{t}$ as it depends on x , so write $\tilde{t}(x)$.

If $\tilde{t}_\varepsilon = 0$, we get $\tilde{t}(x) = \exp(x'\beta)$. The standard normal and logistic distributions both have $\tilde{t}_\varepsilon = 0$. Often the extreme value distribution is redefined so that it also has $\tilde{t}_\varepsilon = 0$. This is accomplished by redefining the extreme value cdf as

$$F_\varepsilon(v) \equiv 1 - \exp[-\log(2)e^v].$$

The relative median comparing an individual with covariate x_1 to another individual with covariate x_2 is

$$RM = e^{(x_1 - x_2)'\beta}. \quad (5)$$

Of course we are often interested in estimating survival probabilities (2), and hazard functions (4) for selected choices of x , as well as relative hazards

$$RH = \frac{h(t|x_1, \beta, \tau)}{h(t|x_2, \beta, \tau)} = \frac{h_\varepsilon[(\log(t) - x_1'\beta)\sqrt{\tau}]}{h_\varepsilon[(\log(t) - x_2'\beta)\sqrt{\tau}]}. \quad (6)$$

EXAMPLE 13.1.1. *Cancer of the larynx: log-normal regression analysis.* Data on 90 males with cancer of the larynx were collected by Kardaun (1983) and analyzed in Klein and Moeschberger (2003). The data consist of times in months from diagnosis to death or censoring. Predictor variables include the **stage** of the disease at diagnosis (one through four), the year of diagnosis *Yr*, and the age at diagnosis *Age*. The **stage** variable is recoded into indicators S_i for stages 2, 3, and 4, making stage one the baseline stage. We also standardize *Age* and *Yr* before using them. Our log-normal AFT model is

$$\log(T) = \beta_1 + \beta_2 S_2 + \beta_3 S_3 + \beta_4 S_4 + \beta_5 \text{Age} + \beta_6 \text{Yr} + \sigma\varepsilon,$$

with $\varepsilon \sim N(0, 1)$.

We are interested in the relative medians for individuals in stages 2, 3 or 4 relative to stage 1 for people with the same age and year of diagnosis. These are e^{β_i} , $i = 2, 3, 4$. The relative median that compares two individuals who are one (sample) standard deviation (10.8 years) apart in age, but with the same stage and year of diagnosis, is e^{β_5} . The relative median that compares two individuals of the same stage and age but who differ by one standard deviation (2.2 years) in year of diagnosis is e^{β_6} .

We handle censored observations in WinBUGS as in Chapter 12. With NA used to indicate missing data, transform (y, δ) into (t, c) where

$$t = \begin{cases} y & \text{if } \delta = 1 \\ NA & \text{if } \delta = 0 \end{cases} \quad c = \begin{cases} 0 & \text{if } \delta = 1 \\ y & \text{if } \delta = 0 \end{cases}$$

Our WinBUGS code for the log-normal survival regression model includes a subset of the data listed at the end.

```
model{
  for(i in 1:106){
    sAge[i] <- (age[i]-mean(age[ ]))/sd(age[ ])
    sYr[i] <- (Yr[i]-mean(Yr[ ]))/sd(Yr[ ])
    t[i] ~ dlnorm(mu[i], tau) I(c[i] , )
    mu[i] <- beta[1] + beta[2]*equals(stage[i],2)
              + beta[3]*equals(stage[i],3)
              + beta[4]*equals(stage[i],4)
              + beta[5]*sAge[i] + beta[6]*sYr[i]
    # 5 month survival probabilities using covariates in the data
    S[i] <- 1- phi((log(5) - mu[i])*sqrt(tau))
    # Medians corresponding to the covariates in the data
    med[i] <- exp(mu[i])
  }
  for (i in 1:6){
    beta[i] ~ dnorm(0,0.000001)
  }
  # Get all possible RMs for each variable
  rm[i] <- exp(beta[i])
  prob[i] <- step(beta[i])
}
tau ~ dgamma(0.001,0.001)
sigma <- sqrt(1/tau)
}
list(beta=c(0,0,0,0,0,0),tau = 1)
stage[ ] t[ ] age[ ] Yr[ ] c[ ]
1 0.6 77 76 0
1 1.3 53 71 0
1 2.4 45 71 0
1 NA 57 78 2.5
1 3.2 58 74 0
1 NA 51 77 3.2
The last 84 data lines have been removed to save space
1 NA 50 71 0 # 16 lines of augmented data start here
1 NA 50 77 0
1 NA 70 71 0
1 NA 70 77 0
2 NA 50 71 0
2 NA 50 77 0
```

```

2 NA 70 71 0
2 NA 70 77 0
3 NA 50 71 0
3 NA 50 77 0
3 NA 70 71 0
3 NA 70 77 0
4 NA 50 71 0
4 NA 50 77 0
4 NA 70 71 0
4 NA 70 77 0
END

```

The last 16 data lines contain no actual data since the time of death is *NA* and the censoring time is 0 in each. These indicate predictor vectors at which we want to make inferences. We picked a sequence of age, year and stage values for the predictors so that allows us to assess the practical import of changing these variates on the median time to death and 5 month survival prospects. We could also compute predictive densities for the time to death of 16 new larynx cancer patients with these predictor profiles.

Posterior results on regression coefficients, relative survival medians, the error standard deviation are presented in Table 13.2. There is very little evidence that Stage 2 leads to shorter survival than Stage 1, but there is substantial evidence that Stages 3 and 4 are progressively worse than Stage 1. (Compare β_j to 0 or e^{β_j} to 1, $j = 2, 3, 4$.) We are 95% sure that the median time to death under Stage 1 is between $3 \doteq 1/0.36$ and $22 \doteq 1/0.046$ times longer than it is under Stage 4. There is some indication that older patients have worse survival prospects ($\beta_5 < 0$) and that treatments are improving over time since later year of diagnosis has $\beta_6 > 0$. But the posterior probability intervals for β_5 and β_6 both contain 0, so the evidence for these indications is relatively weak. More directly, we can compute $\Pr(\beta_5 < 0 | D) = 0.90$, and $\Pr(\beta_6 > 0 | D) = 0.74$, which indicate a reasonable amount of evidence for both assertions.

If the probability intervals for β_5 and β_6 not only contained 0 but were very narrow, we would remove Age and Year from the analysis. However, Table 13.3 indicates that these variables have interesting effects on our analysis. The table provides information on the median survival times for 50 and for 70 year-olds who were diagnosed in 1971 and in 1977. For example, Stage 2 individuals of age = 50 and diagnosed in 1971 have an estimated median time to death of about 9 months, while if they were diagnosed in 1977, their estimated time to death would be about 12 months. The 97.5 percentiles differ by nearly a year. Comparing 70 year-olds, these numbers are about 6 and 8 months respectively, interestingly lower, although the same relative change. Age and Year display interesting but modest effects that we cannot be sure are real. Considering the comparable Stage 4 results, the effects of Year and Age are so small, scientifically they seem practically meaningless, although perhaps less so to patients. The only clear effects are that Stages 3 and 4 give worse mean survivals than Stage 1. **It would be nice to see whether stages 3 and 4 are different from**

Table 13.2: Larynx cancer: log-normal regression.

Node	mean	sd	2.5%	med	97.5%
β_1	2.324	0.318	1.745	2.308	2.989
β_2	-0.250	0.508	-1.256	-0.251	0.746
β_3	-0.973	0.413	-1.804	-0.967	-0.182
β_4	-2.018	0.528	-3.081	-2.004	-1.021
β_5	-0.221	0.167	-0.557	-0.219	0.101
β_6	0.121	0.182	-0.227	0.116	0.491
e^{β_2} St2/St1	0.886	0.486	0.285	0.778	2.109
e^{β_3} St3/St1	0.411	0.174	0.165	0.380	0.833
e^{β_4} St4/St1	0.152	0.083	0.046	0.135	0.360
e^{β_5} (Age +1sd)/Age	0.813	0.136	0.573	0.803	1.106
e^{β_6} (Yr +1sd)/Yr	1.148	0.215	0.797	1.123	1.634
σ	1.396	0.1648	1.118	1.38	1.762

Table 13.3: Larynx cancer: log-normal regression. Median survival times and 5 month survival probabilities.

	Node	mean	sd	2.5%	med	97.5%
Stage 2 Medians	Age 50 Yr 71	10.57	6.911	3.081	8.88	28.2
	Age 50 Yr 77	14.51	9.912	4.525	11.99	39.32
	Age 70 Yr 71	6.787	3.752	2.265	5.944	16.31
	Age 70 Yr 77	9.261	5.142	3.406	8.03	22.5
Stage 4 Medians	Age 50 Yr 71	1.83	1.18	0.482	1.545	4.93
	Age 50 Yr 77	2.421	1.395	0.802	2.102	5.99
	Age 70 Yr 71	1.177	0.660	0.356	1.031	2.827
	Age 70 Yr 77	1.548	0.720	0.6173	1.4	3.353
Stage 1 5 Month Surv.	Age 50 Yr 71	0.718	0.095	0.512	0.727	0.879
	Age 50 Yr 77	0.782	0.097	0.564	0.794	0.935
	Age 70 Yr 71	0.618	0.091	0.429	0.621	0.786
	Age 70 Yr 77	0.694	0.104	0.475	0.701	0.874
Stage 3 5 Month Surv.	Age 50 Yr 71	0.461	0.119	0.238	0.460	0.697
	Age 50 Yr 77	0.546	0.122	0.306	0.547	0.777
	Age 70 Yr 71	0.352	0.100	0.174	0.348	0.560
	Age 70 Yr 77	0.435	0.110	0.232	0.431	0.656

stage 2. ron Finally, 5 month survival probabilities seem better for 50 year-olds diagnosed in 1977 in stage 1 than for 70 year-olds diagnosed in 1971 at stage 3; the posterior probability intervals do not overlap.

We earlier noted that if $T \sim AFT(F_\varepsilon, \beta, \tau|x)$, using the extreme value cdf for F_ε leads to $T \sim \text{Weib}(\sqrt{\tau}, e^{-x'\beta\sqrt{\tau}})$. To simplify interpretations of regression coefficients, we use the modified extreme value distribution that has median 0, i.e.,

$$F_\varepsilon(v) \equiv 1 - \exp[-\log(2)e^v],$$

which leads to

$$T \sim \text{Weib}(\sqrt{\tau}, \log(2)e^{-x'\beta\sqrt{\tau}}).$$

We now use this form of Weibull regression to reanalyze the larynx cancer data.

EXAMPLE 13.1.2. *Cancer of the larynx: Weibull regression analysis.* Using the same reference priors as in the log-normal analysis, the main change in the WinBUGS code given earlier is that

```
t[i] ~ dlnorm(mu[i], tau) I(c[i] , )
```

is replaced by the code

```
t[i] ~ dweib(alpha, lambda[i]) I(c[i] , )
lambda[i] <- log(2)*exp(-mu[i]*sqrt(tau))
alpha <- sqrt(tau)
```

If one just copies this code, `alpha` will get assigned 106 times. It would be better to move the assignment outside the “for” loop. In addition, the log-normal 5 month survival probability,

```
S[i] <- 1- phi((log(5) - mu[i] )*sqrt(tau))
```

needs to be replaced with the Weibull 5 month survival probability

```
S[i] <- exp(-log(2)*exp(( log(5) - mu[i])*sqrt(tau)))
```

Comparing results between Tables 13.2 and 13.3 and Tables 13.5 and 13.5, we see that the two analyses are substantially the same, with some relatively minor modifications. Using the log-normal model, relative medians comparing 50 year olds to 70 year olds were estimated to be 12 and 8 in Stage 2, and 2.1 and 1.4 in Stage 4, while under the Weibull model these estimates were 11.4 and 7.7 in Stage 2, and 2.5 and 1.7 in Stage 4. Estimated relative medians comparing Stages 3 and 4 to Stage 1 under the *LN* were 0.38 and 0.14 while they were 0.52 and 0.19 under the Weibull. For the Weibull regression, the posterior probability that β_5 is negative is 0.91 and the posterior probability that β_6 is positive is 0.68. These probabilities were 0.91 and 0.74 under the *LN* model. Interestingly, we obtained the DIC statistics for both models and got 297.3 and 297.7 for *LN* and Weibull regression, respectively. One final note on

Table 13.4: Larynx cancer: Weibull regression analysis.

Node	mean	sd	2.5%	med	97.5%
β_1	2.22	0.29	1.71	2.20	2.86
β_2	-0.14	0.49	-1.07	-0.15	0.86
β_3	-0.66	0.38	-1.45	-0.65	0.05
β_4	-1.69	0.46	-2.64	-1.68	-0.83
β_5	-0.21	0.16	-0.54	-0.21	0.09
β_6	0.09	0.18	-0.24	0.08	0.45
Relative Medians:					
e^{β_2} St2/St1	0.99	0.54	0.35	0.87	2.37
e^{β_3} St3/St1	0.55	0.21	0.24	0.52	1.04
e^{β_4} St4/St1	0.20	0.10	0.07	0.19	0.43
e^{β_5} (Age +1sd)/Age	0.82	0.13	0.58	0.81	1.109
e^{β_6} (Year +1sd)/Year	1.1	0.2	0.78	1.08	1.58
σ	1.007	0.14	0.77	0.99	1.32

Table 13.5: Larynx cancer: Weibull regression analysis.

Node	mean	sd	2.5%	med	97.5%
Medians					
Stage 2					
Age 50 Yr 77	13.64	8.674	4.862	11.42	35.72
Age 70 Yr 77	8.674	5.258	3.429	7.71	22.3
Medians:					
Stage 4					
Age 50 Yr 77	2.749	1.359	1.10	2.45	6.19
Age 70 Yr 77	1.773	0.66	0.87	1.7	3.39
5 Month Surv:					
Stage 1					
Age 50 Yr 77	0.76	0.09	0.55	0.77	0.90
Age 70 Yr 77	0.67	0.10	0.45	0.68	0.84
5 Month Surv:					
Stage 3					
Age 50 Yr 77	0.59	0.12	0.33	0.60	0.81
Age 70 Yr 77	0.47	0.13	0.22	0.47	0.71

this analysis is that the estimate of σ is virtually one, which indicates that the exponential model may well suffice.

EXERCISE 13.13. Fit the exponential regression model to the larynx cancer data. This is accomplished by modifying the Weibull code so that σ is identically one. Obtain the DIC statistic and decide if this model would be sufficient based on that criterion. Compare inferences with those in Table 13.?.

Before reanalyzing the data with F_ε as the logistic distribution, we discuss the logistic distribution in more detail. Suppose ε has the logistic distribution of Table 13.1. Write

$$\varepsilon \sim \text{Logis}(0, 1).$$

For positive σ and real μ consider

$$W = \sigma\varepsilon + \mu.$$

Write the distribution of W as

$$W \sim \text{Logis}(\mu, 1/\sigma) = \text{Logis}(\mu, \tau^*)$$

where

$$\tau^* \equiv 1/\sigma.$$

Using arguments similar to those in (2), it is easy to see that W has cdf

$$F_W(w) = F_\varepsilon[(w - \mu)\tau^*] = 1 - \frac{1}{1 + \exp[(w - \mu)\tau^*]}.$$

If

$$\log(T) \sim \text{Logis}(\mu, \tau^*),$$

define T to have a log-logistic distribution and write

$$T \sim \text{Log-Logis}(\mu, \tau^*).$$

EXAMPLE 13.1.3. *Cancer of the larynx: log-logistic regression analysis.* Now consider the $T_i \stackrel{\text{ind}}{\sim} AFT(F_\varepsilon, \beta, \tau | x_i)$ model with the logistic error distribution of Table 13.1 for the larynx cancer data. Using (2),

$$S(t | x, \beta, \tau) = 1 - F_\varepsilon[(\log(t) - x'\beta)\sqrt{\tau}] = \frac{1}{1 + \exp[(\log(t) - x'\beta)\sqrt{\tau}]}$$

and

$$T_i \stackrel{\text{ind}}{\sim} \text{Log-Logis}(x'_i\beta, \sqrt{\tau}).$$

Equivalently,

$$\log(T_i) \stackrel{\text{ind}}{\sim} \text{Logis}(x'_i\beta, \sqrt{\tau}).$$

We use the same reference priors as in the log-normal analysis. The main change to the WinBUGS code is that the analysis is performed on the log scale, using the logistic distribution rather than the log-logistic. As usual in WinBUGS, transform (y, δ) into (t, c) but now use $(\log(t), \log(c))$ in the WinBUGS analysis. When t or c is zero, replace it with a large negative number, here, -1000 . When $t = \text{NA}$, define $\log(t) = \text{NA}$. As discussed before, if t and c remain in the data file, they must be used.

```
junk1 <- t[1]
junk2 <- c[1]
stage[] t[] logt[] age[] Yr[] c[] logc[]
1      0.6 -0.5108 77    76    0  -1000
1      1.3  0.2623 53    71    0  -1000
1      2.4  0.8754 45    71    0  -1000
1      NA   NA     57    78    2.5 0.9162
1      3.2  1.1631 58    74    0  -1000
1      NA   NA     51    77    3.2  1.1631
more data lines and augmented data
END
```

The other key change from the log-normal code is that

```
t[i] ~ dlnorm(mu[i], tau) I(c[i] , )
```

is replaced by

```
logt[i] ~ dlogis(mu[i], taustar)I(logc[i],)
taustar <- sqrt(tau)
```

As with `alpha` in the Weibull example, `taustar` should be assigned outside the “for” loop. In addition, the log-normal 5 month survival probability,

```
S[i] <- 1- phi((log(5) - mu[i]) * sqrt(tau))
```

needs to be replaced with the log-logistic 5 month survival probability

```
S[i] <- 1/(1+ exp((log(5)-mu[i])*sqrt(tau)))
```

To create the variables `logt` and `logc` we copied the original data from WinBUGS into Excel. There we divided it into separate columns using the “Text to Columns” tool located under the Data menu in the Excel toolbar. Then we used the `ln` function in Excel to create a new column of log transformed survival times (if t_i is missing then set its log to `NA`) and a column of log transformed censoring times (if $c_i = 0$ then set the log equal to -1000 or some other large negative value). Then we copied that data from Excel back into WinBUGS (in WinBUGS go to Paste Special under the Edit menu and select Unicode Text to do this). As always, you will need to type `END` with a hard return after the last data line. Then you are set to run the code with the transformed data.

The results of the log-logistic analysis looked superficially similar to those of the log-normal and Weibull. The log-logistic model has $DIC = 231.2$, which

is considerably smaller than for the other models that both have $DIC = 297$. The log-logistic model may deserve further scrutiny. We leave as an exercise to discuss the analysis of the larynx cancer data using log-logistic regression and to compare it with the other models.

EXERCISE 13.14. (a) Analyze the larynx cancer data using the log-logistic regression model. Compare results with the previous analyses. (b) Pick one of the three models used to analyze these data, and investigate whether adding interaction terms will improve the selected model. If yes, modify the model interpretation accordingly.

13.1.1 Abortion Data

In the remainder of this section, our discussion revolves around the data introduced in Section 1.4 on time to natural abortion in dairy cattle. In this subsection, we focus on inferential procedures using an informative prior. The next subsection details the elicitation of the informative prior. The last three subsections examine case deletion diagnostics, model selection, and sensitivity analysis.

Our 45 observations were retrieved from a larger set of 375 observations collected by Drs. Mark Thurmond and Sharon Hietala of the University of California, Davis. While all 375 cows were at risk of abortion, only 45 actually aborted. Our analysis is conditional on the event of a known abortion. Most of the remaining cows had full term pregnancies. Some cows were culled from the herd, but there is no way to tell if they would have aborted or gone full term. We make the assumption that culled animals would have come to term.

The data involve four covariates. In Section 1.4 we only used infection status to distinguish two groups. The variable $x_2 = IS$ is an indicator of infection status wherein infected with *Neospora caninum* corresponds to $x_2 = 1$ and no infection corresponds to zero. Days open, $x_3 = DO$, is the number of days between most recent previous birth and conception. Calving ease, $x_4 = CE$, is coded $x_4 = 1$ for cows with difficulties in their most recent previous birth and zero otherwise. Prostaglandin, $x_5 = PR$, has $x_5 = 1$ corresponding to prostaglandin being given during the open days. Along with an intercept $x_1 \equiv 1$, the scientists believed there may be an interaction $x_6 = DO * CE$, so there are 6 regression coefficients in our full model,

$$\log(T_i) = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_6 x_{i6} + \sigma \varepsilon_i.$$

The scientists believed that infection status would be the most important variable for predicting time to abortion, with abortion occurring later in infected animals. They expected days open to be the second most important variable, anticipating that increasing DO would slightly increase the time to abortion. Difficulty with calving and the application of prostaglandin were expected to slightly decrease the time to abortion. The primary goal of this study is to characterize the effect of neospora on the time to abortion. This is part of an

ongoing investigation of neospora and other infectious abortifacients with the ultimate goal of reducing fetal wastage, cf. Paré, Thurmond, and Hietala (1997).

Figure 13.??? presents eight predictive survival curves for various choices of $x = (1, x_2, x_3, x_4, x_5, x_6)'$ using our elicited prior and a logistic error distribution. Also presented are median survival times for each set of covariates. These curves represent cows with $PR = 0$. We also constructed similar plots for $PR = 1$. Superimposing the plots shows no visual differences between them, however, predicted median survival times are a few days longer for $PR = 0$ than for $PR = 1$. These differences are of no practical importance. For example, in plot (b), for $x' = (1, 1, 45, 1, 0, 45)$, the estimated median time to abortion is 114 days. The 90% posterior interval is (92, 141), so clearly a couple of days difference in estimated median survival times is of no importance. Similarly, in plot (a), for $x' = (1, 0, 45, 0, 0, 0)$, the median time to abortion 63 days with 90% PI of (54, 74). Note also that when $CE = 1$, the curves depend on the infection status but do not seem to depend much on the days open. On the other hand, when $CE = 0$, fetal survival depends on both the infection status and the number of days open.

We show the summary estimates based on posterior and maximum likelihood methods in Table 13.6. The posterior means are remarkably similar to the maximum likelihood estimates. Standard deviations are either about the same or smaller for the posterior. The larger differences are due to having little data on some conditions with significant prior information. Statistical import or lack thereof is clear from looking at the various probability intervals for all the factors. Everything but prostaglandin seems to matter as we have already seen from looking at the survivor curves.

The predictive distribution can be thought of as the predictive probability of survival for a particular individual with covariate x , or it can be thought of as an estimate of the proportion of individuals in the population of individuals with covariate x that will survive past t units of time. With the latter interpretation, we can get approximate posterior distributions for $h(\beta, \sigma) = S(t|x, \beta, \sigma) = S_0((\log(t) - x'\beta)/\sigma)$ and, therefore, obtain posterior probability bands for these proportions for every t .

13.1.2 Prior Elicitation for AFTs

Our approach is based on Bedrick, Christensen and Johnson (2000) who use conditional median priors. In these the prior specification is made on a collection of median responses; each corresponding to a particular potential covariate combination. This joint prior induces a distribution on the regression coefficients.

We require a joint prior $p(\beta, \tau)$. The standard diffuse prior in this setting is $p(\beta, \tau) \propto 1/\tau$, cf. Ibrahim and Laud (1991). We used a proper approximation to this prior in the previous section for all three analyses of the larynx cancer data. A standard alternative to this is to let β be normal given τ and the marginal of τ be gamma, cf. West (1985) and Racine-Poon et al. (1986), which is the usual conjugate prior when the baseline is normal or log-normal (the

Table 13.6: Log logistic regression estimates for the cow abortion study.

Variable	Posterior Summaries			
	Mean	Std. Dev.	5%	95%
Intercept	3.985	0.146	3.742	4.223
<i>IS</i>	0.388	0.086	0.247	0.525
<i>DO</i>	0.004	0.001	0.002	0.006
<i>CE</i>	0.377	0.201	0.045	0.711
<i>PR</i>	−0.020	0.101	−0.182	0.146
<i>DO * CE</i>	−0.004	0.002	−0.007	−0.001
Scale	0.261	0.032	0.213	0.314
Variable	Maximum Likelihood		Flat Prior	
	Estimate	S.E.	Mean	Std. Dev.
Intercept	3.928	0.152	3.929	0.600
<i>IS</i>	0.342	0.128	0.340	0.134
<i>DO</i>	0.004	0.001	0.004	0.001
<i>CE</i>	0.349	0.902	0.294	1.028
<i>PR</i>	−0.019	0.289	−0.035	0.315
<i>DO * CE</i>	−0.004	0.006	−0.004	0.007
Scale	0.245	0.031	0.269	0.033

likelihood functions have exactly the same form). Normal-gamma priors are also convenient in large sample situations where the posterior distribution of β is approximately normal. However, as we keep reminding you, it is difficult to specify real prior information directly for esoteric regression parameters.

Our informative priors of choice satisfy $p(\beta, \tau) = p(\beta)p(\tau)$. We thus discuss the marginals for β and τ separately.

Specifying the marginal prior for β

For AFT models with a baseline survival distribution that has median 0 and with covariates x , the median is $m = \exp(x'\beta)$. We elicit expert information about a collection of medians and use that information to induce a prior on β . As we have done several times before, we select a collection of $r \times 1$ covariate vectors, $\tilde{x}_i = (1, \tilde{x}_{i2}, \tilde{x}_{i3}, \dots, \tilde{x}_{ir})'$. We then define the corresponding median times to event $\tilde{m}_i = \exp(\tilde{x}_i'\beta)$, and elicit expert opinion about these quantities. The induced distribution on β is obtained from the specified prior distribution on $\tilde{m} \equiv (\tilde{m}_1, \dots, \tilde{m}_r)'$. Defining $\tilde{X} = (\tilde{x}_1, \dots, \tilde{x}_r)'$, we can write $\tilde{m} \equiv \exp(\tilde{X}\beta)$. Since we assume $\tilde{X}$ is nonsingular, we obtain $\beta = \tilde{X}^{-1} \log(\tilde{m})$. We also assume that the $\tilde{x}_i$ s are sufficiently “far apart” that it is reasonable to assume that knowledge about the $\tilde{m}_i$ s can be regarded independently. Thus, if we have

specified $p_{\tilde{m}}(\cdot) = \prod_{i=1}^r p_{\tilde{m}_i}(\cdot)$ as our prior on the vector of medians, then

$$p(\beta) \propto \prod_{i=1}^r p_{\tilde{m}_i}(e^{\tilde{x}'_i \beta}) e^{\tilde{x}'_i \beta}, \quad (1)$$

since the Jacobian of the transformation is

$$\begin{aligned} \left| \frac{d}{d\beta} \exp(\tilde{X}\beta) \right| &= \left| \left\{ \frac{d}{d\beta} \exp(\tilde{x}'_i \beta) : i = 1, \dots, r \right\} \right| \\ &= \left| \{ \tilde{x}'_i \exp(\tilde{x}'_i \beta) : i = 1, \dots, r \} \right| \\ &= |\text{Diag}\{e^{\tilde{x}'_i \beta}\} \tilde{X}| \\ &= \prod_{i=1}^r e^{\tilde{x}'_i \beta} |\tilde{X}| \\ &\propto \prod_{i=1}^r e^{\tilde{x}'_i \beta}. \end{aligned}$$

The point is that $\tilde{m}$ is on the scale of the data and is more natural to think about. Moreover, we need only specify a single prior on $\tilde{m}$ and the same prior can be used to induce a prior on β for any median zero baseline distribution S_0 , which is a nice feature since we only need to make a single elicitation regardless of which or how many such AFT models we will consider.

Since the median survival time is non-negative, and since the normal distribution is particularly easy to work with, independent log-normal distributions are natural and convenient. Thus, one possibility for $p_{\tilde{m}_i}$ is to specify

$$\tilde{m}_i \sim LN(\log(\tilde{m}_{0i}), \tilde{\sigma}_{0i}^2). \quad (2)$$

The median of the prior is $\tilde{m}_{0i}$, which serves as the expert's prior estimate of $\tilde{m}_i$, and $\tilde{\sigma}_{0i}$ is a subjectively chosen scale factor for the distribution. It is straightforward to ask an experimenter to select $\tilde{m}_{0i}$. To select a value for the scale factor, the experimenter can be asked to reflect on the 95th or the 5th (or both) percentiles of their prior for the median survival time. Suppose the experimenter is 95% certain that the median survival time is below c_i years. Then there is a unique value of $\tilde{\sigma}_{0i}$ that satisfies this additional constraint and thus the prior distribution is completely determined, namely we set

$$\frac{\log(c_i) - \log(\tilde{m}_{0i})}{\tilde{\sigma}_{0i}} = 1.645$$

and solve for $\tilde{\sigma}_{0i}$.

WinBUGS code for inducing a prior on β based on independent log-normal priors for the $\tilde{m}_i$ s is:

```
model{
  for(i in 1:r){
    mtilde[i] ~ dlnorm(mu0[i],tautilde0[i]) # mtilde[i] ~ LN
```

```

    mu0[i] <- log(mtilde0[i]) # Prior guess is mtilde0[i]
    tautilde0[i] <- pow(1/sigmatilde0[i],2)
    lmtilde[i] <- log(mtilde[i])
    beta[i] <- inprod(G[i,1:r] , lmtilde[1:r])
  }
}
list(r=, mtilde0 = c(), sigmatilde0 = c())
# The matrix G = solve(Xtilde) must be obtained in R
G[,1]  G[,2]  G[,3]  ...  G[,r]
list the columns of G here
END

```

EXERCISE 13.15. *Log normal prior.* (a) Let $r = 2$. Obtain explicit log-normal priors for $\tilde{m}_1$ and $\tilde{m}_2$ if the best guesses for the two medians are 10 and 20, and if the values of the c s are 20 and 30, respectively. (b) Now assume that the single predictor variable is age and that it has been standardized. Assume the average age in the data is 50 and that the standard deviation is 5. Then write the code to induce the prior on β . Assume that you have selected an average age of 50 for the first predictor and an age of 60, two standard deviations above the average, for the second. Then

$$\tilde{X} = \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}$$

(c) Now re-write the code for the log-normal so that you don't have to invert $\tilde{X}$. Run the code and show the induced priors on the β_i s.

EXAMPLE 13.1.4. *BCJ priors for β .* Bedrick, Christensen and Johnson (2000) considered Log-Extreme-Value (LEV) distributions as priors for regression coefficients in the Weibull model with the AFT parametrization, and Log-Logistic (LL) distributions as priors for regression coefficients in the Log-Logistic model. They also used a combination of *LN* and *LEV* priors for the cow abortion data in BCJ (2000) to better match elicited prior information.

Suppose we are working with baseline distribution S_0 . Then we can let

$$\frac{\log(\tilde{m}_i) - \log(\tilde{m}_{0i})}{\sigma_{0i}} \stackrel{iid}{\sim} S_0. \quad (3)$$

Note that (2) is a special case of (3) with $S_0 = N(0,1)$. It is straightforward to obtain the analytical form of the induced prior for β like we did for the *LN* distribution, but since we will be inducing the prior in WinBUGS, we don't do that here. The actual specification of the prior is almost identical to the *LN* case. We always have the solution $\beta = \tilde{X}^{-1} \log(\tilde{m})$. So this part of the code remains the same for all models.

First, we select $\tilde{m}_{0i}$ in exactly the same way as before. Then, if we let w_α be the $1 - \alpha$ percentile of the distribution S_0 and if $\Pr(\tilde{m}_i < c) = 1 - \alpha$, we

must have

$$w_\alpha = \frac{\log(c) - \log(\tilde{m}_{0i})}{\sigma_{0i}}.$$

We again solve for $\tilde{\sigma}_{0i}$. For example, if $\alpha = 0.05$, and we have selected the *LEV* baseline, we find

$$0.05 = S_0(w_{0.05}) = \exp[-\log(2)e^{w_{0.05}}] \Leftrightarrow w_{0.05} = \log\{\log(20)/\log(2)\} = 1.46.$$

WinBUGS code for inducing a prior on β based on an *LL* specification for the $\tilde{m}_i$ s is:

```
model{
  for(i in 1:r){
    w[i] ~ dlogist(mu0[i],tautilde0[i])
    mtilde[i] <- exp(w[i]) # So mtilde[i] ~ Log-Logistic
    mu0[i] <- log(mtilde0[i])
    tautilde0[i] <- 1/sigmatilde0[i]
    lmtilde[i] <- log(mtilde[i])
    beta[i] <- inprod(G[i,1:r] , lmtilde[1:r])
  }
}
```

WinBUGS code for inducing a prior on β based on an *LEV* (Weibull) specification for the $\tilde{m}_i$ s is:

```
model{
  for(i in 1:r){
    mtilde[i] <- dweib(alphatilde0[i],thetatilde0[i])
    thetatilde0[i] <- log(2)*exp(-mu0[i]*sqrt(tautilde0[i]))
    alpha[i] <- sqrt(tautilde0[i])
    mu0[i] <- log(mtilde0[i])
    tautilde0[i] <- 1/sigmatilde0[i]
    lmtilde[i] <- log(mtilde[i])
    beta[i] <- inprod(G[i,1:r] , lmtilde[1:r])
  }
}
```

Someone else should check these. I've checked them twice but... Is this line correct `tautilde0[i] <- 1/sigmatilde0[i]`. Should it be squared? Also, what is `alphatilde0` in the *LEV* code?

EXERCISE 13.16. BCJ PRIORS. (a) Repeat parts (a) and (b) of Exercise 13.15 only now for the *LL* and for the *LEV* baseline distributions.

Example 13.1.5. For the cow abortion problem we designed a thought experiment as discussed in BCJ, and elicited prior information from Dr. Thurmond, a veterinary epidemiologist. Specifically, we asked for his beliefs about the median time to abortion at six sets of covariates. We asked him to specify his best

Table 13.7: Specification of the prior for the cow abortion study.

i	Design for Prior						Prior Percentiles			
	Int.	IS	DO	CE	PR	$DO * CE$	5%	$\tilde{m}_{0i}$	95%	$\tilde{\sigma}_{0i}$
1	1	0	95	0	0	0	70	80	95	0.1045
2	1	0	45	1	1	45	60	75	90	0.1108
3	1	0	45	1	0	45	65	75	90	0.1108
4	1	1	130	1	1	130	80	110	130	0.1141
5	1	1	130	0	1	0	80	120	140	0.1053
6	1	1	95	0	0	0	85	125	150	0.1246

guess for the median time to abortion (the median of his prior distribution) and extreme values for his distributions, which we took to be the fifth and ninety fifth percentiles. The covariate combinations were chosen to be in the range of Dr. Thurmond's experience and were chosen so that median times to abortion for the various covariate combinations could be regarded independently. The six covariate combinations need to be linearly independent so that $\tilde{X}^{-1}$ exists. The information that we extracted is given in Table 12.5.

For example, the row with $i = 1$ indicates that for uninfected cows ($IS = 0$), with 95 days open ($DO = 95$), no prior history of difficult births ($CE = 0$), and no prostaglandin given ($PR = 0$), Dr. Thurmond's best guess for the median time to abortion was 80 days with only a 5% chance that the median time to abortion would be less than 70 days and a 95% chance that the median time to abortion would be less than 95 days. Similarly, $i = 4$ corresponds to infected cows with 130 days open, a history of difficult births, and prostaglandin administered, his best guess for the median time to abortion was 110 days with only a 5% chance that the median time to abortion would be less than 80 days and a 95% chance that the median time to abortion would be less than 130 days. Note the strong influence of infection on the a priori median times to abortion. Dr. Thurmond tended to be more certain about situations with smaller median times as one might expect, e.g., percentiles tend to be closer when the median is smaller.

The medians and 95th percentiles of the prior were used to determine a corresponding Weibull or log-normal distribution with the same median and 95th percentile. The log-normal distribution was used for cases 1, 2, and 3 and the Weibull for the others. The choice between Weibull and log-normal was based on how well the 5th percentiles matched the elicited percentiles (the log-normal being, for these values, more right skewed and the Weibull more left skewed). Dr. Thurmond was generally more confident about the 95th percentiles than the 5th percentiles, which is why this method was used. The distributions were plotted and we reconfirmed that they appropriately represented the expert's prior opinions.

Wes, do you have the figure in the next paragraph? Also, what

are the locations? I don't have this data set so I can't produce the figure. Figure **** presents plots of the prior distributions for the $\tilde{m}_i$'s and the corresponding posteriors for the median times to abortion under each of the six covariate combinations $\tilde{x}_i$. It is interesting that the prior and posterior are nearly identical for location 2. This is because there are only a few animals among the 45 that were given prostaglandin. There is little information in the data for this location and the prior information dominates. At the other locations, there is generally more information in the posterior than the prior. The priors and posteriors for locations 2, 3 and 5 have approximately the same centers while the centers for the posteriors at locations 1, 4 and 6 are lower than for the corresponding priors.

Partial prior information

As previously discussed in Sections 8.4.4 and 9.4.4, we may not have the resources to specify a full prior for r medians. The issue is resolved here in more or less the same way that we approached it before. We would only specify independent *LN* (or *LL* or *LEV*) distributions on say $p < r$ medians, $e^{x_i'\beta}$, $i = 1, \dots, p$. We would standardize all continuous covariates and attempt to specify prior information corresponding to a $p \times r$ matrix $\tilde{X} = (\tilde{X}_1, 0)$, where $\tilde{X}_1$ is $p \times p$ and nonsingular. Partition $\beta = (\beta_1^*, \beta_2^*)'$ and define the corresponding vector of medians as $\tilde{m}_1^* = e^{\tilde{X}_1\beta_1^*}$. We then obtain $\beta_1^* = \tilde{X}_1^{-1} \log(\tilde{m}_1^*)$. The prior on β_1^* is induced by the specification for $\tilde{m}_1^*$. We finally specify an uninformative *LN*(0,0.001) prior for the components of β_2^* .

EXERCISE 13.17. (a) For the problem posed in Exercise 13.15, write WinBUGS code to induce a partial prior on β where $p = 1, r = 2$ and with the first row of $\tilde{X}$ from Exercise 13.15 being used to specify a prior on the corresponding $\tilde{m}_1^* \equiv \tilde{m}_1$ from that problem. Then repeat using the second row of $\tilde{X}$ from Exercise 13.15 and let $\tilde{m}_1^* \equiv \tilde{m}_2$ from that problem. (b) Now suppose that there is a second covariate in Exercise 13.15, let's say a dichotomous variate indicating sex ($M = 1, F = 0$). Let

$$\tilde{X} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \end{pmatrix}$$

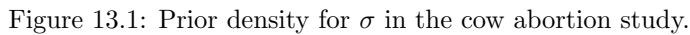
and assume that the prior information for the two medians in that exercise were actually specified for women. Using this partial prior information, write code to induce a prior on β .

Uncertainty about τ

The prior on τ is elicited the same way it was in Example 5.2.3 using log transformed data from the Diasorin study. We stay on the scale of the data by considering a percentile of the survival distribution itself, like the 90th percentile. Information about this parameter gives us information about τ and/or $\sigma = 1/\sqrt{\tau}$. We focus on σ here but the prior for τ is similarly obtained.

$$0.5 = \Pr[\tilde{m}_1 \exp(\sigma w_{0.25}) \leq 60 \mid \tilde{m}_1 = 80] = \\ \Pr\left[\frac{\log[\tilde{m}_1 e^{(\sigma w_{0.25})}] - \log[\tilde{m}_1]}{w_{0.25}} \leq \frac{\log(60/80)}{w_{0.25}} \mid \tilde{m}_1 = 80\right] = \Pr[\sigma \leq 0.262],$$

We used the logistic percentile in evaluating the prior for σ because we will use the log logistic to model the data. The prior on σ , therefore, depends on which parametric model is assumed for the data. This follows since σ has a different meanings in different AFT models. Similarly, Dr. Thurmond was 95% certain that the value of the quartile was at least 35, giving another percentile of the distribution of σ . These percentiles approximately match a $\text{Gamma}(2.25, 7.25)$ distribution for σ , which is plotted in Figure 13.*.



13.1.3 Case Deletion Diagnostics for AFTs

Johnson (1996) introduced case deletion diagnostics for estimating survival curves in the Bayesian log-normal survival model. The goal is to detect cases that noticeably affect inferences, as discussed in Section 9.4. For example, it is important to know if the predictive survivor function changes radically upon deleting one case. It is also relevant to examine the effect of the prior on posterior inferences and, with a BCJ prior, it is possible to look at the effect of deleting a “prior observation” corresponding to $\tilde{m}_i$.

First, we give an expression for the joint posterior with a case deleted. Denote the complete data as $D \equiv (y, \delta)$, the data with the i th case deleted as $D_{(i)}$, and the i th case as $d_i = (y_i, \delta_i)$. The likelihood based on all the data except d_i is $L(\beta, \sigma | D_{(i)}) \equiv L(\beta, \sigma | D) / L(\beta, \sigma | d_i)$, so

$$\begin{aligned} p(\beta, \sigma | D_{(i)}) &= \frac{L(\beta, \sigma | D_{(i)})p(\beta, \sigma)}{\int L(\beta, \sigma | D_{(i)})p(\beta, \sigma)d\beta d\sigma} \\ &= \frac{p(\beta, \sigma | D)/L(\beta, \sigma | d_i)}{\int p(\beta, \sigma | D)/L(\beta, \sigma | d_i)d\beta d\sigma}. \end{aligned} \quad (1)$$

A standard tool for measuring the effect of case deletion on prediction (estimation) is the Kullback-Leibler divergence (KLD) between predictive (posterior) densities based on full and case deleted data (cf. Johnson and Geisser 1983, 1985; Johnson 1996). The KLD is a measure of the discrepancy between two densities. If the predictive density based on the full data is appreciably different from the predictive density based on case deletion, we would like to take a further look at the particular case. A similar discussion was presented in Section 9.6 on case deletion diagnostics for linear regression. Johnson (1996) gives results for the log-normal survival model where integrations are carried forth via Gibbs sampling. We introduce a variation on these ideas that can be implemented using WinBUGS in conjunction with R.

It is simpler to consider the effect of case deletion on the joint posterior itself, so we present this first. Define the diagnostic

$$KLD_i = \int p(\beta, \sigma | D) \log \left(\frac{p(\beta, \sigma | D)}{p(\beta, \sigma | D_{(i)})} \right) d\beta d\sigma$$

This simplifies using (1) and some algebra to

$$\begin{aligned} KLD_i &= \log \left(\int \frac{1}{L(\beta, \sigma | d_i)} p(\beta, \sigma | D) d\beta d\sigma \right) - \\ &\quad \int \log \left(\frac{1}{L(\beta, \sigma | d_i)} \right) p(\beta, \sigma | D) d\beta d\sigma \\ &= \log \left(E \left[\frac{1}{L(\beta, \sigma | d_i)} \right] \right) - E \left[\log \left(\frac{1}{L(\beta, \sigma | d_i)} \right) \right] \geq 0 \end{aligned}$$

where both expectations are with respect to the full posterior and nonnegativity is guaranteed by Jensen’s inequality.

Now suppose we have a sample from the joint posterior, say $\{\beta^j, \sigma^j : j = 1, \dots, m\}$, which for us would have been obtained using WinBUGS after we threw out the burn-in. Then a numerical approximation is

$$KLD_i \doteq \log \left(\frac{1}{m} \sum_{j=1}^m [1/L(\beta^j, \sigma^j | d_i)] \right) - \frac{1}{m} \sum_{j=1}^m \log [1/L(\beta^j, \sigma^j | d_i)], \quad (2)$$

which can be computed by taking the MC sample from the joint posterior into R and doing the arithmetic there. This would be done for all cases in the data, censored and uncensored. We remind the reader that

$$L(\beta, \sigma | d_i) = \frac{1}{\sigma y_i} p_0[(y_i - x'_i \beta)/\sigma]^{\delta_i} S_0[(y_i - x'_i \beta)/\sigma]^{1-\delta_i},$$

where $d_i = (y_i, \delta_i)$ is the observed data for case i .

Since there is no natural scale for KLD_i we normalize all values by dividing by the largest one. The case that has the largest value should be scrutinized as discussed in Section 9.6. If it is discovered that there is an error associated with the case, it can either be fixed, if possible, or deleted. Otherwise, the model should be run with this case deleted and a determination made whether there is any substantial impact on inferences after removal. If not, proceed with the analysis based on the full data. If there is a substantial impact, it should be reported how inferences would change if it were deleted. If there are other cases that have relatively large KLD values, they should be explored in the same way.

It is possible to make inferences about functions of the parameters without running the code again with the case deleted. For example, suppose we are interested in making inferences about $h(\beta, \sigma)$ using the case deleted posterior. From (1), we obtain

$$\begin{aligned} \int h(\beta, \sigma) p(\beta, \sigma | D_{(i)}) d\beta d\sigma &= \int h(\beta, \sigma) \frac{p(\beta, \sigma | D)/L(\beta, \sigma | d_i)}{\int p(\beta, \sigma | D)/L(\beta, \sigma | d_i) d\beta d\sigma} d\beta d\sigma \\ &= \frac{\int h(\beta, \sigma) \left\{ \frac{1}{L(\beta, \sigma | d_i)} \right\} p(\beta, \sigma | D) d\beta d\sigma}{\int \left\{ \frac{1}{L(\beta, \sigma | d_i)} \right\} p(\beta, \sigma | D) d\beta d\sigma} \\ &\doteq \frac{\sum_j h(\beta^j, \sigma^j)/L(\beta^j, \sigma^j | d_i)}{\sum_j 1/L(\beta^j, \sigma^j | d_i)}. \end{aligned} \quad (3)$$

Inferences about h would be made based on taking the MC sample from WinBUGS into R to performing the arithmetic implicit in (3).

EXERCISE 13.18. *Case deletion diagnostics.* For the log-normal regression analysis of the larynx cancer data, take the Monte Carlo output from WinBUGS into R and obtain the values of KLD_i for all subjects in the study. Find and remove the most influential case and then re-run the code. Comment on the impact of deleting that case on all inferences.

EXAMPLE 13.1.7*. *Checking predictive influence.*

We proceed to develop two slightly more complicated measures of the effect of case deletion on prediction. The first measure is simply a *KLD* for assessing the effect of case deletion of a predictive density corresponding to a predictor x . This is defined as

$$KLD P_i^x = \int f(t|x, D) \log \left(\frac{f(t|x, D)}{f(t|x, D_{(i)})} \right) dt, \quad (4)$$

where $f(t|x, D)$ and $f(t|x, D_{(i)})$ are predictive densities for a future observation with predictor information x that are based on full and case deleted data, respectively.

We also propose a measure that directly examines the effect of case deletion on the right tail of survival curves. Ferguson (1973) uses a loss function for cumulative distribution functions of $\int [F(t) - \hat{F}(t)]^2 f(t) dt$. To emphasize the right tails of the survival function, we transform to logs and consider a similar measure, namely

$$DS_i^x = \int f(t|x, D) \{ \log[S(t|x, D)/S(t|x, D_{(i)})] \}^2 dt. \quad (5)$$

DS_i^x will be large if case i has a substantial effect on the right tail of the predictive density.

These measures will again be calculated in R using the Monte Carlo sample $\{(\beta^j, \sigma^j) : j = 1, \dots, m\}$ from the joint posterior. We first need to get numerical approximations to the two predictive densities. These are computed as

$$f(t|x, D) \doteq \frac{1}{m} \sum_{j=1}^m h_t(\beta^j, \sigma^j) \quad \text{where} \quad h_t(\beta, \sigma) = \frac{1}{\sigma t} p_0\{(\log(t) - x'\beta)/\sigma\},$$

and $f(t|x, D_{(i)}) = \int h_t(\beta, \sigma) p(\beta, \sigma | D_{(i)}) d\beta d\sigma$ is approximated using (3) with this h . We obtain these two predictive densities over a grid of equally spaced points $\{t_1, t_2, \dots, t_k\}$. You could decide on the range of the grid by obtaining a numerical approximation to the full data predictive density for the specified x in WinBUGS.

We then get discrete approximations to both predictive densities as

$$\hat{f}(t|x, D) \equiv \frac{f(t|x, D)}{\sum_{j=1}^k f(t_j|x, D)}, \quad t \in \{t_1, \dots, t_k\},$$

with a similar definition of $\hat{f}(t|x, D_{(i)})$. Define the corresponding approximations to the predictive survivor functions to be $\hat{S}(t|x, D)$ and $\hat{S}(t|x, D_{(i)})$, which are obtained by numerically integrating the functions $\hat{f}(t|x, D)$ and $\hat{f}(t|x, D_{(i)})$ over the grid points. We then obtain the approximations

$$KLD P_i^x \doteq \sum_{j=1}^k \hat{f}(t_j|x, D) \log \left(\frac{\hat{f}(t_j|x, D)}{\hat{f}(t_j|x, D_{(i)})} \right),$$

and

$$DS_i^x \doteq \sum_{j=1}^k \hat{f}(t_j | x, D) \left\{ \log \left(\frac{\hat{S}(t_j | x, D)}{\hat{S}(t_j | x, D_{(i)})} \right) \right\}^2.$$

We can consider the collective effect of deleting case i over a future sample of individuals with covariates, say, $x_{f\ell}$, $\ell = 1, \dots, s$,

$$KLDP_i = \sum_{\ell=1}^s KLDP_i^{x_{f\ell}}, \quad DS_i = \sum_{\ell=1}^s DS_i^{x_{f\ell}}.$$

Johnson and Geisser (1983) let the collection of $x_{f\ell}$ s be the matrix X of predictor vectors associated with the observed data. Since there is no natural scale for $KLDP_i$, $KLDP_i$, or DS_i , we normalize all values by dividing by the largest one.

For the cow abortion study, we considered the eight locations used in Figure **** as our choices of x_{fi} 's. Case 10 was most influential with normalized values of 1 on both predictive measures. This was followed by case 26 with $KLDP_{26} = 0.958$ and $DS_{26} = 0.817$. The third most influential value based on $KLDP$ was case 2 with $KLDP_2 = 0.472$, but for this case $DS_2 = 0.000$, indicating that case 2 had some impact on the predictive densities but virtually no impact on the tails of the corresponding survival functions. Finally, case 38 was third most influential with respect to the DS measure with $DS_{38} = 0.762$, while $KLDP_{38} = 0.135$. Thus, case 38 had a relatively large impact on estimation of the tails of the survival distribution but had little impact on the predictive densities.

We note that cases 10, 26, and 38 had by far the highest observed survival times of 259, 254, 251 days, respectively. So these pregnancies nearly went full term. The next highest survival time was 150 days.

A full analysis would delete some or all of the influential observations and evaluate whether this leads to changes of practical importance. This is left to the following exercise.

EXERCISE 13.19. *Predictive influence.* **ADAM? I can't get to this now, but it seems fine to put an exercise on this**

13.1.4 Model Selection for AFTs

We now allow the Bayesian paradigm to indicate which among the various choices of parametric baseline survivor functions is most appropriate for the data. Specifically, we select from among the LL (M_1), LN (M_2) and Weibull (LEV) (M_3) regression models. The DIC statistic is a candidate model selection instrument that we have used in this chapter. Alternatively, the LPML statistic, which can be used to construct pseudo Bayes factors, can be used for this purpose just as they were used for linear regression. We can also calculate Bayes Factors, which have been discussed in detail in Chapters 4 and 8. The reader may wish to refer to Chapter 4 for basic details of all of these methods

and Chapters 8 and 9 for previous applications of them; the basic techniques applied there will apply here as well. We use BF's for AFT model selection.

Let $f(y|M)$ be the marginal density for the observed data under model M , which is obtained by integrating the corresponding likelihood function with respect to the prior on (β, σ) for that model. Then recall that the Bayes factor for comparing models M_j and $M_{j'}$ is

$$BF_{jj'} = \frac{f(y|M_j)}{f(y|M_{j'})}.$$

EXAMPLE 13.1.8. *Cow abortion data.* The Bayes factors for the 3 pairwise comparisons under the BCJ prior are $BF_{12} = 2.6$, $BF_{13} = 4922$, and thus (with some round-off error) $BF_{23} = BF_{13}/BF_{12} = 4922/2.6 = 1880$. There seems to be good reason not to use Weibull regression. Log logistic regression appears to be the best among these candidates, but the log-normal does not seem too bad. In particular, if the prior odds for the log logistic versus log-normal models are 1, the posterior odds in favor of the log logistic are 2.6.

13.1.5 Sensitivity Analysis

Where are the figures? Numbers here aren't right; I DIDN'T TOUCH THIS SUBSECTION - WJ. Me either; I don't even have the data - Adam.

We redid the analysis of the Cow Abortion data using the proper approximation to the prior $p(\beta, \tau) \propto 1/\tau$ **is this right**. Survival curves similar to Figure *** are given as Figure***. Curves for $PR = 1$ are almost visually indistinguishable from those shown for $PR = 0$. The curves for $IS = 0$, $CE = 0$ are almost indistinguishable from those using our informative prior. The curves for $IS = 1$, $CE = 0$ are slightly lower than with our prior, indicating slightly shorter survivals. Predicted median survivals dropped by just under 10 days. Posterior 90% intervals for these are about 30 days long, so a change of 10 is not drastic.

In plot (b) the tails are much higher with the flat prior and the curves are closer together, indicating less difference due to infection status. The flat prior median survival of 98 days is nearly outside the informative prior 90% interval of (92, 141). What is really going on is that there is little information in the data corresponding to these conditions, so the informative prior is exerting considerable weight. As mentioned, the 90% informative posterior interval for $IS = 1, DO = 45, CE = 1, PR = 0$ is (92, 141). For the flat prior, this is (27, 311). For $IS = 0$ and the flat prior, the interval is (19, 227). These wide intervals are indicative of little information.

The differences between curves for the flat and informative priors in plots (d) are larger than those for plots (a) and (c), but not nearly as extreme as those for plot (b). In the data, there are many observations with $CE = 0$ as in plots (a) and (c), so the prior has relatively little effect on these plots. In the

Table 13.8: Flat Prior Results

Variable	Posterior Summaries			
	Estimate	Std. Dev.	0.05	0.95
Intercept	3.99	0.15	3.74	4.22
<i>IS</i>	0.39	0.09	0.25	0.52
<i>DO</i>	0.0036	0.0013	0.0015	0.0058
<i>CE</i>	0.38	0.20	0.05	0.71
<i>PR</i>	-0.02	0.10	-0.18	0.15
<i>DO * CE</i>	-0.004	0.002	-0.007	-0.001
Scale	0.26	0.03	0.21	0.31
Variable	Maximum Likelihood			
	Estimate	S.E.		
Intercept	3.93	0.15		
<i>IS</i>	0.34	0.13		
<i>DO</i>	0.0038	0.0013		
<i>CE</i>	0.35	0.90		
<i>PR</i>	-0.019	0.289		
<i>DO * CE</i>	-0.004	0.057		
Scale	0.25	0.03		

data there are only three observations with $CE = 1$, while the informative prior has three $\tilde{m}_i$ s with $CE = 1$, hence the larger differences in plots (b) and (d).

Posterior means and standard deviations for the regression coefficients using the flat prior are reported in Table 12.7.

13.2 Proportional Hazards Modeling

As previously mentioned, parametric models are restricted. Let's face it, we only really considered three parametric families of models for survival data so far, the LN, LL and LEV distributions. While it's certainly better to have three than only one, they all have constraints in terms of their shape. They are all unimodal and thus preclude the possibility of modeling situations where there might be clusters of individuals who survive longer than others, resulting in multimodal or exceptionally fat tailed survival distributions. We could consider mixtures of parametric models in order to achieve greater skewness and allow for multimodality, but we would be required to guess the number of modes and or distributions that are needed. If this were solely a book on survival analysis, we would certainly pursue that route, but instead, we continue in a more traditional approach by considering a broad family of models that allows for greater flexibility, namely the Cox (1972) proportional hazards (PH) model. Having said that, we must also mention that this model is also constrained. But not in the ways we have so far described. We will get into that in due time.

In this section, we develop the PH model, then discuss a discrete but still flexible approximation to it that is amenable to WinBUGS coding as per the discussion in their help file. We illustrate with some examples.

13.2.1 The PH Model

PH regression is all about flexible modeling of survival as a function of predictor information x . This is accomplished through modeling the hazard function as

$$h(t|x) = e^{x'\beta} h_0(t)$$

where $h_0(\cdot)$ is a completely arbitrary “baseline” hazard function. Thus, as $x'\beta$ becomes larger, the greater the hazard of event. Since the baseline hazard may be close to zero over some range of time, it may be that having a larger $x'\beta$ isn't such a big deal over that period. Of course since $h_0(\cdot)$ is a hazard function, it can't always be small. Unlike AFT and other regression models we have encountered so far, the vector x in the PH model does not contain a one in the first slot; there is no intercept term in the Cox model. If there were, it would be confounded with the baseline hazard function.

Recalling formula (12.1.1), we must also have

$$h(t|x) = \frac{f(t|x)}{S(t|x)} = -\frac{d}{dt} \log[S(t|x)],$$

which implies, by the Fundamental Theorem of Calculus that

$$\log[S(t|x)] = -\int_0^t h(s|x)ds + c \Leftrightarrow S(t|x) = \exp\left[-\int_0^t h(s|x)ds + c\right].$$

Since the survivor function at $t = 0$ must be 1, we must have $c = 0$. Define the cumulative hazard function to be $H(t|x) = \int_0^t h(s|x)ds = e^{x'\beta} \int_0^t h_0(s)ds \equiv e^{x'\beta} H_0(t)$, where $H_0(\cdot)$ is termed the “baseline” cumulative hazard function. It follows that

$$S(t|x) = \exp[-e^{x'\beta} H_0(t)] = \{\exp[-H_0(t)]\}^{e^{x'\beta}} \equiv [S_0(t)]^{e^{x'\beta}}, \quad (1)$$

where we again term S_0 the “baseline” survivor function. The density function is thus

$$f(t|x) = h(t|x)S(t|x) = e^{x'\beta} h_0(t)[S_0(t)]^{e^{x'\beta}}. \quad (2)$$

A key feature of the PH model is that, with two distinct sets of covariates, call them $\tilde{x}_1$ and $\tilde{x}_2$, the hazard ratio (HR) is

$$HR = \frac{h(t|\tilde{x}_1)}{h(t|\tilde{x}_2)} = \exp[(\tilde{x}_1 - \tilde{x}_2)'\beta],$$

which is completely free of the time t . Thus, hazard function in the numerator is equal to this constant times the denominator hazard function, which is the reason for calling this a proportional hazards model. Moreover, if $\tilde{x}_1'\beta > \tilde{x}_2'\beta$,

then we must also have $S(t|\tilde{x}_1) < S(t|\tilde{x}_2)$ for all t , since a probability to a greater power must be less than one to a smaller power. Hence, a constraint of the PH model is that it forces survival curves for distinct individuals to never cross. The problem that arises with this is that there will be situations where, say, there are two treatments for cancer, and one is more toxic than the other early in a study so that survival prospects are worse under that treatment in the beginning, but also where those that survive the toxicity may perhaps do better with this same treatment later on, leading to crossing survival curves. Having said this, there are many situations where this will not be an issue. Moreover, when the problem exists, it can often be remedied through the use of what is called a time dependent covariate. We discuss this later.

While the hazard ratio is a simple function of β , median times to event and relative medians are not simple. It is beyond the scope of this book to pursue medians further. We instead focus on survival curve estimates and relative hazards.

Consider standard survival data $\{(y_i, \delta_i)\}_{i=1}^n$. Using (12.1.2), the likelihood function is

$$L(\beta, h_0(\cdot)) = \prod_{i=1}^n \{e^{x_i' \beta} h_0(y_i)\}^{\delta_i} [S_0(y_i)]^{e^{x_i' \beta}}. \quad (3)$$

A key feature of this likelihood that we have not seen before is that h_0 is infinite dimensional, and thus very flexible, since it is an entire function that cannot be described with a denumerable number of parameters. This model is actually termed semi-parametric since we have a combination of the infinite dimensional h_0 , alongside of our usual parameter vector β . So the model is only partially parametric or semi-parametric. In the next subsection, we approximate the PH model by discretizing h_0 , which gives a “richly parametric” model.

EXERCISE 13.20. Show that the Weibull model is a special case of a PH model. Give the explicit form for the baseline hazard h_0 , the baseline cumulative hazard H_0 , and the baseline survivor function S_0 in this case.

EXERCISE 13.21. (a) In R, plot the hazard ratio comparing the $LN(0, 1)$ to the $LN(2, 2)$. Are the hazards proportional? (b) Repeat part (a) for $LL(0, 1)$ versus $LL(2, 2)$ distributions.

EXERCISE 13.22. Argue why $\lim_{t \rightarrow \infty} H_0(t) = \infty$.

13.2.2 The Approximate PH Model

The idea of the discrete approximation to h_0 has been around for a long time. The plan is to break up the time interval $(0, \infty)$ into a finite number of subintervals. We thus partition $(0, \infty) = \bigcup_{i=1}^k [t_i, t_{i+1})$, where $0 < t_1 < \dots < t_{k-1} < t_k < t_{k+1} = \infty$. Define the piecewise constant function

$$\tilde{h}_0(t) = \sum_{i=1}^k \lambda_{0i} I_{[t_i, t_{i+1})}(t),$$

which simply means that $\tilde{h}_0(t) = \lambda_{0i}$ if $t \in [t_i, t_{i+1})$, where the λ_{0i} s are unknown parameters. We subsequently let the first k partition points be the distinct event times in the data, and define the baseline hazard to be zero on the sets $(0, t_1)$ and (t_k, ∞) since there is no information in the data to suggest how survival curves would drop over them. So we've gone from an infinite dimensional parameter h_0 to a $k - 1$ dimensional parameter vector $\lambda_0 = (\lambda_{01}, \dots, \lambda_{0(k-1)})$. We obtain the approximate (discretized) likelihood by substituting $\tilde{h}_0$ for h_0 in (3), and define

$$L(\beta, \lambda_0) = L(\beta, \tilde{h}_0). \quad (4)$$

If k is very large, we now have a flexible “richly parametric” model that is an approximation to the semi-parametric Cox model.

From here, we proceed to develop a standard trick in survival analysis that relies on Poisson modeling of “expanded data” that will generate the likelihood (4). We use counting process notation. See Klein and Moeschberger (2003) for a brief introduction to counting processes and their relevance to survival analysis.

We “expand” the data corresponding to the i th actual data point (y_i, δ_i) . Define the processes $N[i, j] = I(y_i \leq t_j)$ and $Y[i, j] = I(y_i \geq t_j)$. Then define $dN[i, j] = Y[i, j] * N[i, j] * \delta_i$. So $dN[i, j]$ is zero for all intervals except the interval that includes y_i if $\delta_i = 1$, and is otherwise always zero, and $Y[i, j]$ is one for all intervals up to and including the interval that includes the time y_i , and is zero for all intervals afterwards. Also define

$$\begin{aligned} dH[i, j] &= H(t_{j+1} | x_i) - H(t_j | x_i) \\ &= e^{x_i' \beta} [H_0(t_{j+1}) - H_0(t_j)] \\ &\equiv e^{x_i' \beta} dH_0[j] = e^{x_i' \beta} \lambda_{0j} [t_{j+1} - t_j]. \end{aligned} \quad (5)$$

To make life simpler, we need to have $y_i \in \{t_1, \dots, t_k\}$ if $\delta_i = 1$. So all the event times are represented as endpoints of our grid. Let $t_{i*} \in \{t_1, \dots, t_k\}$ be the largest element among the t_j s that is less than y_i . Then

$$\begin{aligned} H_0(y_i) &= H_0(t_1) + [H_0(t_2) - H_0(t_1)] + \dots + [H_0(y_i) - H_0(t_{i*})] \\ &= \sum_{j: t_j < y_i} dH_0[j], \end{aligned}$$

and thus

$$\begin{aligned} H(y_i | x_i) &= e^{x_i' \beta} H_0(y_i) \\ &= e^{x_i' \beta} \sum_{j: t_j < y_i} dH_0[j] \\ &= \sum_{j: t_j < y_i} dH[i, j]. \end{aligned}$$

The contribution to the approximate likelihood from the data for subject i is therefore

$$L_i(\beta, \tilde{h}_0) = \left\{ \prod_{j: t_j < y_i} \{e^{x_i' \beta} \tilde{h}_0(y_i)\}^{dN[i, j]} \right\} \exp \left\{ -e^{x_i' \beta} \sum_{j: t_j < y_i} dH_0[j] \right\}.$$

Defining $Idt[i, j] = dH[i, j]Y[i, j]$ and using (5), this reduces to

$$\begin{aligned} L_i(\beta, \tilde{h}_0) &\propto \prod_{j: t_j < y_i} \{dH[i, j]\}^{dN[i, j]} \exp\{-dH[i, j]\} \\ &= \prod_j \{Idt[i, j]\}^{dN[i, j]} \exp\{-Idt[i, j]\}. \end{aligned} \quad (6)$$

Now this looks suspiciously like a Poisson likelihood where the values taken are either 0 or 1. So if we model $dN[i, j] \sim \text{Poisson}(Idt[i, j])$, $j = 1, \dots, n$, we get the right likelihood contribution for case i . Many values of $Idt[i, j]$ will be 0, but WinBUGS has no problem with this; these will not contribute to the likelihood. The product of terms in (6) thus gives the full approximate PH model likelihood (4), namely

$$L(\beta, \tilde{h}_0) = \prod_i L_i(\beta, \tilde{h}_0).$$

The selection of t_j s to be the distinct event times in the data is consistent with the illustration of the Cox model in the WinBUGS manual. It is well-known in survival analysis that, with nonparametric approaches, there is no information in the data about the survival function in between the event times. So while we could pick times in addition to the event times, it would be only the prior that contributes information about how survival curves drop in between them. So we also recommend just picking the event times in the absence of additional information.

13.2.3 Priors

The standard type of uninformed prior involves placing independent diffuse Gamma priors on the λ_{0j} s, and diffuse normal priors on the components of β . This is always an option. However, we proceed to discuss how to place informative priors on the regression coefficients.

Informative and Partially Informative Priors for β

In order to obtain an informative prior for β , as always, we need to get onto a scale that leads to expert opinion. Since researchers familiar with survival methods are used to thinking about relative hazards, we start here. A relative hazard of 1 means no effect, and a relative hazard of 100 or 1/100 would be quite something to write home about. Often, an HR of 1.5, or 2/3 is almost not worth mentioning, while an HR of 2 or 1/2 is. We ask our expert for their best guess for the HR associated with particular predictors in the model, holding all other predictors fixed at values that the expert can assess. If there are no interactions in the model, the hazard ratio is of the form $HR_i = e^{\beta_i}$ when comparing two subjects who differ on covariate i by 1 unit. If the best estimate is HR_{0i} , and a best guess for the 95th percentile is u , we obtain $\beta_{0i} = \log(HR_{0i})$ and

$$0.95 = \Pr[e^{\beta_i} \leq u] = \Phi\left(\frac{\log(u) - \log(HR_{0i})}{\sigma_{0i}}\right),$$

where we have assumed $\beta_i \sim N(\beta_{0i}, \sigma_{0i}^2)$. Solving, we get $1.645 = [\log(u) - \log(HR_{0i})]/\sigma_{0i} \Leftrightarrow \sigma_{0i} = [\log(u) - \log(HR_{0i})]/1.645$.

This type of specification should be relatively easy for binary predictors. For example, a researcher may believe that the risk of men drivers being involved in an alcohol related accident is triple that for women, for typical values of other predictors like age. In this case, $\beta_{0i} = \log(3)$. However, if we are asking about the relative hazard corresponding to age, what does that mean? It is the hazard ratio comparing an individual that is, say, one year older than the other. It may be easier to think about an HR that corresponds to a larger difference in time. So we may ask the expert to think about $e^{10\beta_i}$. Then our best estimate is $HR_{0i} = e^{10\beta_{0i}}$, so $\beta_{0i} = \log(HR_{0i})/10$. Similarly, $\sigma_{0i} = [\log(u)/10 - \log(HR_{0i})/10]/1.645$. This is equivalent to saying that $10\beta_i \sim N(\log(HR_{0i}), \{[\log(u) - \log(HR_{0i})]/1.645\}^2)$.

With interaction terms in the model, we have to work a little harder. Suppose our predictors are Sex ($F = 1, M = 0$) and Race (Black = 1, White = 0). So with an interaction term, we have

$$h(t | Sex, Race, Sex \times Race) = \exp(\beta_1 Sex + \beta_2 Race + \beta_3 Sex \times Race) h_0(t).$$

We let our baseline individual be a white male (WM). Their hazard function is simply h_0 . Then consider the three HRs corresponding to comparisons between a (i) WF and a WM, (ii) a BM and a WM, and (iii) a BF and a WM. The three hazard ratios are (i) e^{β_1} , (ii) e^{β_2} , and (iii) $e^{\beta_1 + \beta_2 + \beta_3}$. We ask our expert to ponder all three of these and to provide their best estimates of each and an upper or lower value. We illustrate with upper values. So let HR_{0i} s denote the best guesses and u_i s denote the upper 95% values. We let $\beta_i \sim N(\beta_{0i}, \sigma_{0i}^2)$ independently for $i = 1, 2, 3$. The first two are easy; we have $\beta_{0i} = \log(HR_{0i})$, $i = 1, 2$. And the corresponding variances are selected just as above when there was no interaction, the only difference being that the elicitation here for the HR corresponding to sex depends on race, and the elicitation for the HR corresponding to race depends on the sex, whereas in the absence of an interaction term, it was not necessary to be specific about the race and sex.

We proceed by determining a prior on β_3 . Since $HR_3 = HR_1 HR_2 e^{\beta_3}$, our prior specification for HR_3 naturally depends on what we thought about HR_1 and HR_2 . Thus, our best estimate of β_3 is taken to be $\beta_{03} = \log[HR_{03}/(HR_{01} HR_{02})]$. Moreover, using the independence of the β_i s,

$$\begin{aligned} 0.95 &= \Pr(HR_3 \leq u_3 | \beta_{01}, \beta_{02}) \\ &= \Pr \left[e^{\beta_3} \leq \frac{u_3}{HR_{01} HR_{02}} \right] \\ &= \Pr \left[\frac{\beta_3 - \beta_{03}}{\sigma_{03}} \leq \frac{\log[u_3/(HR_{01} HR_{02})] - \log[HR_{03}/(HR_{01} HR_{02})]}{\sigma_{03}} \right] \\ &= \Phi[\log(u_3/HR_{03})/\sigma_{03}]. \end{aligned}$$

We get $\sigma_{03} = \log(u_3/HR_{03})/1.645$.

We can obtain a partial prior specification by placing informative priors on only a few of the β_i s. The remainder of the β_j s will get diffuse normal prior

distributions. If there is an interaction in the model, we could elicit information about fewer relative hazards. For example, in the previous illustration we could take the informative prior on β_1 and β_2 , and place a diffuse $N(0, b)$ prior on β_3 .

EXERCISE 13.23. Suppose you have data on time to death from diagnosis with cancer and that you will fit the PH model to those data. Suppose you have the two dichotomous predictors Sex and Race as discussed above. (a) Assume that there is no interaction for the moment. Then construct an informative prior on the two regression coefficients that reflects your beliefs that the RH comparing females to males is 0.5 with a 5% lower bound of 0.1, and that the RH for comparing blacks to whites is 1.0 with a 5% lower bound of 0.8. (b) Now assume an interaction, and assume that your first elicitation in (a) corresponded to the comparison of WF to WM and that the second comparison corresponded to comparison of BF to BM. Add to this the comparison of a BM to a WM where the best estimate of the RH is 2 with an upper 95% bound of 3. Construct the prior on β_3 . (c) Write code to induce the prior on β in part (b). Then modify the prior to be only partially informative, using only the first two specifications there. Obtain the induced prior on the third RH. (d) Given the prior guesses in part (b), what would your best guess be for the RH comparing a BF to a WF?

Prior for the Baseline Hazard $\tilde{h}_0$

The baseline hazard is not the easiest object to think about. However, in the beginning of this chapter, we talked about a hazard that reflected a shape that might govern human mortality from birth to death. Selecting the actual values of the hazard function, on the other hand, is not so easy without a lot of experience in estimating these objects. Here, we are estimating a very flexible baseline hazard function, and the idea was to not be overly constrained by any particular shape. So we settle for uninformed priors here.

In the beginning of this section, we alluded to, say $\lambda_{0i} \stackrel{iid}{\sim} \text{Gamma}(rc, c)$, where the prior mean is r , which the expert would select, and then we would let $c = 0.001$ to allow much uncertainty for our prior guess. A value of $r = 1000$ would correspond to a belief that events were happening at a high rate across all intervals, while a value of $1/1000$ would indicate a belief that events were somewhat rare. An uninformed not too blatant prior guess might be $1/10$. The WinBUGS folks have used just such a guess in their illustration of the PH model.

One problem with assuming independent Gamma priors for the components of λ_0 is that you would expect knowledge of adjacent λ_{0j} s to be correlated, and indeed this would be desirable since in modeling such correlation you would expect greater smoothness in the overall plot of $\hat{h}_0(t) = \sum_{i=1}^{k-1} \hat{\lambda}_{0j} I_{[t_i, t_{i+1})}(t)$. We can easily correlate all of the λ_{0j} by modeling them with LN distributions and including a common random effect. One suggestion is

$$\log(\lambda_{0j}) | \rho, \sigma_w \stackrel{iid}{\sim} N(\log(\rho), d), \quad \log(\rho) \sim N(\log(r), e).$$

Then, conditional on ρ , the median of λ_{0j} is ρ . The prior parameter d determines how different the baseline hazard bits are across intervals. If d were selected

to be very small, the implication would be that the hazard was constant at ρ . If the value of d is selected to be large, then much heterogeneity is allowed in the hazard bits across intervals. We would select e to be large, the same as we did with the independence prior. If we were even more adventurous, we would place a prior on d as well and let the data decide how much the hazard function deviates from a constant.

Perhaps a more realistic prior would have additional structure that would allow for more correlation among terms that are closer to one another in time. Perhaps even standard autoregressive structure (see Chapter 10) would be reasonable. We do not pursue this further here.

13.2.4 WinBUGS Code for PH Model Approximation

We have written our model using more or less the same style and notation that is used in WinBUGS to illustrate the PH model. Our code is essentially their code, in a slightly revised notation. The first part of the code sets up the “expanded data” as discussed above. It then proceeds to describe the likelihood contributions (6) based on the assumption of Poisson data, ultimately resulting in the approximate likelihood (4). We then include code for constructing survivor functions and relative hazards plus the prior specifications. The reader is warned that with a large n and k , this code will run considerably slower than previous code we have run. In particular, we ran this code with $n = 876$ and $k = 129$, and it took on the order of one second per iteration. In other words, about 10 minutes for every 600 iterations.

EXAMPLE 13.2.1. *Leukemia data revisited.* To illustrate the WinBUGS code, we return to the leukemia study from Example 12.2.1. There are two groups of patients, no censored observations, and time until death from diagnosis with leukemia was recorded for $n = 33$ patients. The two groups are $AG+$ and $AG-$. The code below is in regression form using the indicator variable $AGgroup[i]$, which takes the value 1 if subject i is $AG+$ and 0 if $AG-$.

```
model{ # Set up data
  for(i in 1:n){
    for(j in 1:(k-1)){
      #Y[i,j] keeps track of whether subject i is at risk at t[j]
      #Y[i,j]=1 if y[i] >= t[j], and 0 else
      Y[i,j] <- step(y[i] - (t[j] - eps))

      # Event process jump=1 if t[j] <= y[i] < t[j+1]
      # and delta[i]=1. Jump=0 otherwise.
      dN[i, j] <- Y[i,j] * step(t[j+1]-(y[i]+eps))*delta[i]
    }
  }

  # Construct the likelihood function. Each individual in
  # the sample contributes to the likelihood over each of
```

```

# the death intervals
for(i in 1:n){
  mu[i] <- b*AGgroup[i]
  for(j in 1:(k-1)){ # Likelihood
    dN[i,j] ~ dpois(Idt[i,j])
    Idt[i,j] <- (t[j+1]-t[j])* Y[i,j]*exp(mu[i])*lam0[j]
  }
}
for(j in 1:(k-1)){ # prior on baseline hazard intervals
  lam0[j] ~ dgamma(mmu,c)
}
mmu <- r*c # prior mean hazard guess is r
# Inferences for posterior survival functions
for(j in 1:(k-1)){
  dH0[j] <- (t[j+1]-t[j])* lam0[j] # (H_0(t[j+1])-H_0(t[j]))
# Survivor function at t[j] = exp(-H_0(t[j]))^mu[j]
# Survivor for AG+ and AG-
  S.pos[j] <- pow(exp(-sum(dH0[1 : j])), exp(.5*b))
  S.neg[j] <- pow(exp(-sum(dH0[1 : j])), exp(-.5*b))
  S.diff[j] <- S.pos[j] - S.neg[j]
}
# Inferences for the HRs and prior on b
HR <- exp(b) # relative hazard
b ~ dnorm(0,0.000001)
}
# Data
list(n=33, k=22, eps=1.0E-10, r=.1, c=.001,
y=c(65,156,100,134,16,108,121,4,39,143,56,26,22,1,1,5,65,
56,65,17,7,16,22,3,4,2,3,8,4,3,30,4,43),
AGgroup=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),
delta = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),
t=c(1,2,3,4,5,7,8,16,17,22,26,30,39,43,56,65,100,
108,121,134,143,156))
# Initial Values
list(b=0, lam0=c(.1,.1,.1,.1,.1,.1,.1,.1,.1,.1,.1,.1,
.1,.1,.1,.1,.1,.1,.1,.1,.1,.1))

```

Table 12.8 gives inferences for the regression coefficient β , relative hazard, e^β , survivor functions for $AG+$ and $AG-$ groups at 2, 16, 26, 100 and 143 weeks, and their differences at the same times. All the 95% posterior intervals for differences in survival exclude 0 indicating statistically different probabilities of survival at each of the times considered. Patients are clearly better off if they are $AG+$ at diagnosis. The relative hazard comparing $AG-$ to $AG+$ is estimated to be $1/0.33 = 3$, and we are 95% sure that the instantaneous failure rate for an

Table 13.9: Output from the PH analysis of the leukemia data. **Need to verify these numbers.**

Node	mean	sd	2.5%	med	97.5%
HR	0.35	0.15	0.14	0.32	0.72
$AG- S(2)$	0.87	0.07	0.71	0.88	0.97
$S(16)$	0.48	0.11	0.28	0.48	0.69
$S(26)$	0.30	0.10	0.13	0.30	0.51
$S(100)$	0.06	0.05	0.005	0.05	0.20
$S(143)$	0.01	0.02	0.0	0.003	0.07
$AG+ S(2)$	0.95	0.03	0.88	0.96	0.99
$S(16)$	0.78	0.08	0.61	0.79	0.90
$S(26)$	0.66	0.10	0.46	0.67	0.83
$S(100)$	0.38	0.11	0.18	0.37	0.60
$S(143)$	0.16	0.08	0.04	0.15	0.36
$\Delta S(2)$	0.08	0.05	0.01	0.07	0.21
$\Delta S(16)$	0.30	0.11	0.09	0.30	0.50
$\Delta S(26)$	0.36	0.12	0.11	0.37	0.59
$\Delta S(100)$	0.32	0.12	0.09	0.31	0.55
$\Delta S(143)$	0.15	0.08	0.03	0.14	0.34

$AG-$ patient is between $1/0.73 = 1.37$ and $1/0.14 = 7.14$ times greater than it is for an $AG+$ patient. Recall from Example 12.2.3 that, under the exponential model, the estimated relative median was 3.42 with a 95% PI (1.78, 6.63). But this is also the inference for the RH comparing $AG-$ to $AG+$ under that model so we have obtained very similar results here with the more flexible PH model.

Figure 12.5 plots estimated survival functions from both the PH regression and an exponential AFT analysis using independent $\text{Gamma}(0.0001, 0.0001)$ priors for θ_1 and θ_2 . From the Cox analysis, the median time to death is about 26 days for $AG+$ patients and is under 5 days for $AG-$ patients. From the exponential regression analysis, these parameters are estimated to be 44 days and 13 days for $AG+$ and $AG-$ patients, respectively.

EXAMPLE 13.2.2. *Kidney data.* These data were obtained from Klein and Moeschberger (2003). Background for the data is discussed in their Section 1.7. Briefly, there are 863 times (in days) to death or censoring after kidney transplantation. The surgeries were performed at the Ohio State University Transplant Center between 1982 and 1992. We use a proportional hazards regression model to investigate differences in survival between the sexes (coded as 1 = male, 2 = female) and races (coded as 1 = white, 2 = black), controlling for age in years at the time of transplant.

There were 140 deaths (16%) over the course of follow-up (about 9.5 years). The number of deaths during the study for the four groups are as follows: among

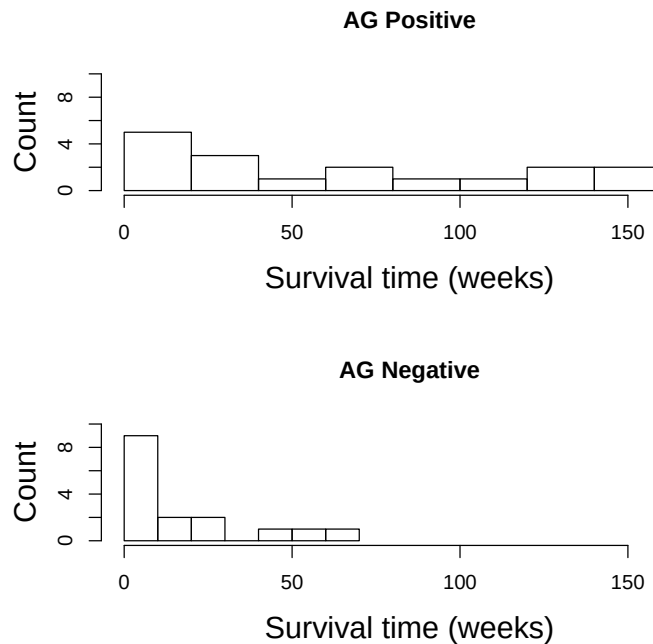


Figure 13.2: Estimated survival curves for *AG+* and *AG-* leukemia patients. **These are based on Adam's numbers. Wes and Adam need to agree on output. Currently we have different estimates.**

white males, 73 out of 432 (17%); among black males, 14 out of 92 (15%); among white females, 39 out of 280 (14%); among black females, 14 out of 59 (24%). The median (sd) age at transplant was 43.5 (13.2) years for white males, 47 (13.1) years for black males, 41 (14.2) years for white females, and 44 (12.2) years for black females.

We do not present the entire WinBUGS program because some of it (setting up the data and defining the prior on the baseline hazard) is the same code used in the leukemia analysis. The baseline group was selected to be white males with standardized age 0 (42 years). The code starts with the likelihood.

```
model{
# code for setting up the data and the prior on lam0 goes here
for(i in 1:n){
  mu[i] <- b[1]*(sex[i]-1) + b[2]*(race[i]-1)+
           b[3]*(age[i]- mean(age[ ]))/sd(age[ ])
  for(j in 1:(k-1)){
    dN[i,j] ~ dpois(Idt[i, j]) # Likelihood
    Idt[i,j] <- (t[j+1]-t[j])*Y[i,j]*exp(mu[i])*lam0[j]
  }
}
```

```

}
# Inferences for posterior survival functions
for (j in 1:(k-1)){
  dH0[j] <- (t[j+1]-t[j])* lam0[j] # (H_0(t[j+1])-H_0(t[j]))
# Survivor function at t[j] = exp(-H_0(t[j]))^mu[j]
# Survivor function for M and F among whites of average age
S.wm[j] <- pow(exp(-sum(dH0[1:j])),1);
S.wf[j] <- pow(exp(-sum(dH0[1:j])),exp(b[1]))
S.wdiff[j] <- S.wm[j] - S.wf[j] # Diff in surv probs
# Survivor function for M and F among blacks of average age
S.bm[j] <- pow(exp(-sum(dH0[1:j])),exp(b[2]))
S.bf[j] <- pow(exp(-sum(dH0[1:j])),exp(b[1]+b[2]))
S.bdiff[j] <- S.bm[j] - S.bf[j]
# Survivor function for WF aged 2 sd above the mean
S.age.2.wf[j] <- pow(exp(-sum(dH0[1:j])),exp(b[1]+2*b[3]))
}
# Inferences for the HRs and prior on the bs
for(j in 1:3){
  HR[j] <- exp(b[j]) # relative hazards
  b[j] ~ dnorm(0.0,0.000001)
}
# Have to use the ID column in the data file
junk <- I[1]
}
# Initial Values
list(b=c(0,0,0),lam0=c(0.1,...,0.1)) # 128 inits for lam0
# Data
list(n=863, k=129, eps=1.0E-10, r= 0.1, c=0.001,
t =c(2,3,7,10,17,21,26,28,37,40,43,44,45,50,52,56,
57,59,62,68,69,78,79,88,91,97,98,104,106,119,
the remaining times go here))
I[ ]    y[ ]    delta[ ] sex[ ] race[ ] age[ ]
1      1      0      1      1      46
2      5      0      1      1      51
3      7      1      1      1      55
the remaining data lines go here
END

```

Partial output is given in Table 12.9. At the aver age of 42 years, the estimated survival probability never dips below 70% for any of the 4 groups over the course of the study (Figure 12.*). Notice from Table 12.9 that we can barely tell the difference in survival prospects between white females and black females of average age (42 years) at diagnosis [**true up to the highest time point in table 12.9. but going beyond that some differences occur between the races, but not between the sexes. see survival plots.**]. Probability intervals for the differences have zero squarely in the middle. The same was

Table 13.10: Output for PH analysis of the kidney transplant data

Node	mean	sd	2.5%	med	97.5%
e^{β_1}	1.05	0.18	0.73	1.04	1.43
e^{β_2}	1.15	0.24	0.73	1.13	1.66
e^{β_3}	2.01	0.19	1.66	2.01	2.41
β_1	0.033	0.17	-0.31	0.037	0.36
β_2	0.12	0.21	-0.31	0.12	0.51
β_3	0.70	0.10	0.51	0.70	0.88
	Ave Age	+ 2sd	WF		
$S(4)$	0.978	0.001	0.954	0.980	0.993
$S(45)$	0.80	0.04	0.72	0.81	0.88
$S(91)$	0.60	0.07	0.46	0.60	0.73
$S(110)$	0.50	0.08	0.35	0.50	0.64
$S(128)$	0.34	0.08	0.19	0.33	0.50
	Ave Age		WF		
$S(4)$	0.995	0.002	0.989	0.995	0.998
$S(45)$	0.95	0.01	0.93	0.95	0.96
$S(91)$	0.88	0.02	0.84	0.88	0.91
$S(110)$	0.84	0.02	0.79	0.84	0.88
$S(128)$	0.76	0.03	0.69	0.76	0.82
	Ave Age		BF		
$S(4)$	0.994	0.003	0.987	0.994	0.998
$S(45)$	0.94	0.01	0.91	0.94	0.96
$S(91)$	0.87	0.03	0.81	0.87	0.92
$S(110)$	0.82	0.04	0.75	0.82	0.89
$S(128)$	0.73	0.05	0.62	0.73	0.83

true when comparing White Males to White Females so we do not present those values due to the redundancy. If we look at the regression coefficients for Sex and Race, we see that zero is squarely in the middle of those intervals. Since we have a fairly large sample size we conclude that there is no Sex or Race effect **[at the later time points we should be careful]**. There is however a clear effect of age, as can be seen by looking at the HR for age and by comparing survival prospects among women who are two standard deviations above the mean in age (70 years) with those that are of average age (see Figure 12.*) Clearly, being older at the time the transplant takes place reduces prospects of survival relative to younger transplant patients. Since 4 of the 5 PIs for the older patient do not overlap with their counterparts for the younger patients, there is clear statistical import.

We ran the code with all three predictors and the DIC was 2242, and with Sex and Race removed it was 2237, indicating that the simpler model is an improvement. Inferences for RH and survival functions were practically identical

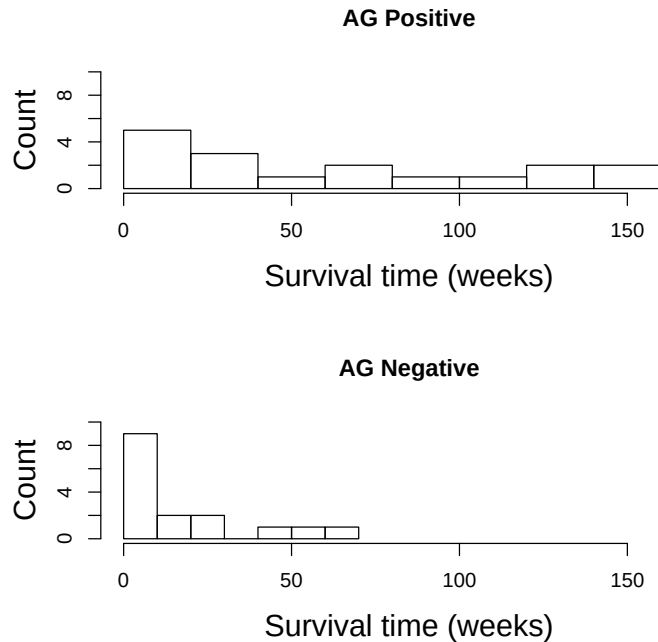


Figure 13.3: Estimated survival curves from a PH analysis of the kidney data.

to those in Table 12.9 so we don't reproduce them here.

Figures 12.* and 12.* give estimated survival plots that tell virtually the same story as Table 12.9. Figure 12.* gives a smoothed estimate of the baseline hazard function as well as the actual estimate, which is a step function. This is the hazard function corresponding to individuals with $x \equiv 0$, which here is white males of average age.

EXERCISE 13.24. *Cow Abortion Data.* Take the two sample version of the Cow Abortion Data from Exercise ***, and analyze the data using the PH model. Make inferences for the RH and for survivor probabilities in the two groups. **ADAM can you add a bit of flavor to this.**

EXERCISE 13.25. *Ovarian Cancer Data.* Analyze the Ovarian Cancer Data taken from Collett, ****. Compare a PH analysis with the best AFT model analysis.

EXERCISE 13.26. *Kidney data.* Re-analyze the Kidney Data to check if there are any interactions among the three predictors Age, Race and Sex. Just look at one interaction at a time, and don't run the code for more than 2000 iterations since it will take a while to run the code each time anyway. Obtain

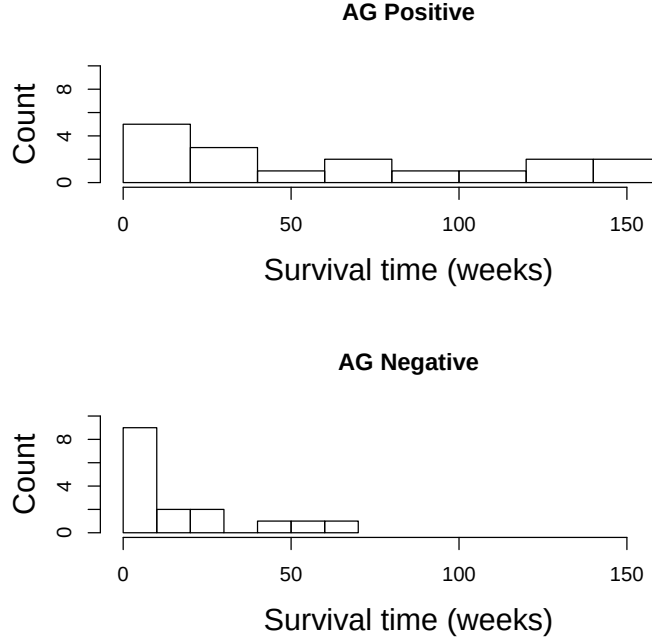


Figure 13.4: Estimated survival curves for white females with average age at transplant and 2 sds above average age.

the DIC statistics in each case based on 1000 iterations. Be sure to check convergence based on history and quantile plots. If any of the interactions seems worthwhile to pursue, give some interpretations of what they mean using appropriate relative hazards and survival probabilities.

Earlier in this section, we alluded to the situation where survival curves might cross. Suppose you have two treatments and that survival is better under the first treatment early on in the study, while survival is better under the second treatment later on. Let's assume an indicator variable to distinguish the first treatment, $\text{Trt}=1$, from the second, $\text{Trt}=0$. Then what we have is an interaction between treatment and time. So an appropriate model will allow for an effect of Trt up to a particular point in time, say t_c , and for a different effect after that. We let $z(t)$ take the value 0 for $t < t_c$, and the value 1 for $t \geq t_c$. Then we write the model

$$h(t | \text{Trt}, z(t)) = \exp(\beta_1 \text{Trt} + \beta_2 \text{Trt} \times z(t)) h_0(t).$$

Individuals are randomly assigned to one of the two treatment groups at the beginning of the study. So, under this model, the RH comparing an individual under $\text{Trt}=1$ to one under $\text{Trt}=0$ is e^{β_1} for $t < t_c$, and $e^{\beta_1 + \beta_2}$ for $t \geq t_c$.

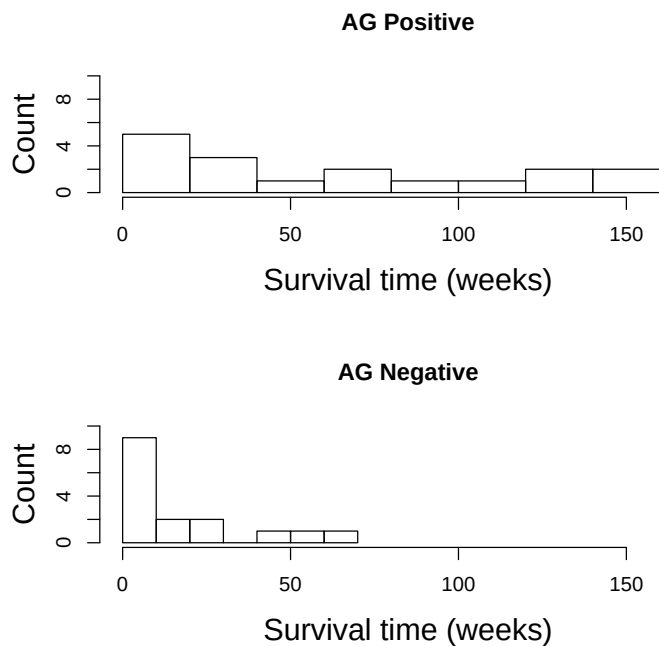


Figure 13.5: Empirical (top) and LOWESS smoothed (bottom) estimates of the hazard function for 42 year old white males.

The variable $z(t)$ is called a (fixed) time dependent covariate (TDC). There are also random TDCs. For example, the conditions under which an individual is under during a study that are relevant to event time may change, for example eating and or smoking habits. The further study of time dependent covariates is beyond the scope of this book.