

5 Modeling Survival Data with Parametric Regression Models

5.1 The Accelerated Failure Time Model

Before talking about parametric regression models for survival data, let us introduce the *accelerated failure time* (AFT) Model. Denote by $S_1(t)$ and $S_2(t)$ the survival functions of two populations. The AFT models says that there is a constant $c > 0$ such that

$$S_1(t) = S_2(ct) \quad \text{for all } t \geq 0. \quad (5.1)$$

This model implies that the aging rate of population 1 is c times as much as that of population 2. (For example, if $S_1(t)$ is the survival function for the dog population and $S_2(t)$ is the survival function for the human population, then the conventional wisdom that a year for a dog is equivalent to 7 years for a human implies $c = 7$, and $S_1(t) = S_2(7t)$. So the probability that a dog can survive 10 years or beyond is the same as the probability that a human subject can survive 70 years or beyond)

Let μ_i be the mean survival time for population i and let φ_i be the population quantiles such that $S_i(t)(\varphi_i) = \theta$ for some $\theta \in (0, 1)$. Then

$$\begin{aligned} \mu_2 &= \int_0^\infty S_2(t)dt \\ &= c \int_0^\infty S_2(cu)du \quad (t = cu) \\ &= c \int_0^\infty S_1(u)du \\ &= c\mu_1 \end{aligned}$$

and

$$S_2(\varphi_2) = \theta = S_1(\varphi_1) = S_2(c\varphi_1).$$

Assume that $S_2(t)$ is a strictly decreasing function. Then we have

$$\varphi_2 = c\varphi_1.$$

This simple argument tells us that under the accelerated failure time model (5.1), the expected survival time, median survival time of population 2 all are c times as much as those of population 1.

Suppose we have a sample of size n from a target population. For subject i ($i = 1, 2, \dots, n$), we have observed values of covariates $z_{i1}, z_{i2}, \dots, z_{ip}$ and possibly censored survival time T_i . The procedure `Proc Lifereg` in `SAS` fits models to data specified by the following equations

$$\log(T_i) = \beta_0 + \beta_1 z_{i1} + \dots + \beta_p z_{ip} + \sigma \varepsilon_i, \quad (5.2)$$

where β_0, \dots, β_p are the regression coefficients of interest, σ is a scale parameter and ε_i are the random disturbance terms, usually assumed to be independent and identically distributed with some density function $f(\varepsilon)$. The reason why we take logarithm of T_i is obvious considering the fact that the survival times are always positive (with probability 1).

Equation (5.2) is very similar to a linear regression model for the log-transformed response variable $Y_i = \log(T_i)$. In a linear regression, the random error term e_i is usually assumed to be *i.i.d.* from $N(0, \sigma^2)$ so that e_i can be written as $e_i = \sigma \varepsilon_i$, where ε_i are *i.i.d.* from $N(0, 1)$ in this case.

At this moment, let us see how the regression coefficients in model (5.2) can be interpreted in general. We will investigate their interpretation more closely later when we consider more specific models (*i.e.*, with different distributional assumptions for ε_i). For this purpose, let us consider β_k ($k = 1, \dots, p$). Holding other covariate values fixed, let us increase covariate z_k by one unit from z_k to $z_k + 1$ and denote by T_1 and T_2 the corresponding survival times for the two populations with covariate values z_k and $z_k + 1$ (with other covariate values fixed). Then T_1 and T_2 can be expressed as

$$\begin{aligned} T_1 &= e^{\beta_0 + \beta_1 z_1 + \dots + \beta_k z_k + \dots + \beta_p z_p} e^{\sigma \varepsilon_1} = c_1 e^{\sigma \varepsilon_1} \\ T_2 &= e^{\beta_0 + \beta_1 z_1 + \dots + \beta_k (z_k + 1) + \dots + \beta_p z_p} e^{\sigma \varepsilon_2} = c_2 e^{\sigma \varepsilon_2} \end{aligned}$$

where c_2 and c_1 are two constants related by $c_2 = c_1 * e^{\beta_k}$. The corresponding survival functions

are

$$S_1(t) = P[T_1 \geq t] = P[c_1 e^{\sigma \varepsilon_1} \geq t] = P[e^{\sigma \varepsilon_1} \geq c_1^{-1} t],$$

$$S_2(t) = P[T_2 \geq t] = P[c_2 e^{\sigma \varepsilon_2} \geq t] = P[e^{\sigma \varepsilon_2} \geq c_2^{-1} t]$$

Since ε_1 and ε_2 have the same distribution, and $c_2 = c_1 * e^{\beta_k}$, we have

$$S_2(e^{\beta_k} t) = P[e^{\sigma \varepsilon_2} \geq c_2^{-1} e^{\beta_k} t] = P[e^{\sigma \varepsilon_2} \geq c_1^{-1} e^{-\beta_k} e^{\beta_k} t] = P[e^{\sigma \varepsilon_2} \geq t] = P[e^{\sigma \varepsilon} \geq t] = S_1(t).$$

Therefore, we have accelerated failure time model between populations 1 (covariate value= z_k) and 2 (covariate value= $z_k + 1$) with $c = e^{\beta_k}$. So if we increase the covariate value of z_k by one unit while holding other covariate values unchanged, the corresponding average survival time μ_2 and μ_1 will be related by

$$\mu_2 = e^{\beta_k} \mu_1.$$

If β_k is small, then

$$\frac{\mu_2 - \mu_1}{\mu_1} = e^{\beta_k} - 1 \approx \beta_k.$$

Similarly we have for the population quantiles φ_i

$$\frac{\varphi_2 - \varphi_1}{\varphi_1} = e^{\beta_k} - 1 \approx \beta_k.$$

Therefore, when β_k is small, it can be interpreted as the percentage increase if $\beta_k > 0$ or percentage decrease if $\beta_k < 0$ in the average survival time and/or median survival time when we increase the covariate value of z_k by one unit. Thus the greater value of the covariate with positive β_k is more beneficial in improving survival time for the target population. This interpretation of β_k is very similar to that in a linear regression model.

5.2 Some Popular AFT Models

We can assume different distributions for the disturbance term ε_i in model (5.2). For example, we can assume $\varepsilon_i \stackrel{i.i.d.}{\sim} N(0, 1)$. This assumption is equivalent to assuming that T_i has log-normal

distribution (of course, conditional on the covariates z 's). In this section, we will introduce some popular parametric models for T_i (equivalently for ε_i). The following table gives some of these distributions:

Distribution of ε	Distribution of T	Syntax in Proc Lifereg
extreme values (2 par.)	Weibull	<code>dist = weibull</code>
extreme values (1 par.)	exponential	<code>dist = exponential</code>
log-gamma	gamma	<code>dist = gamma</code>
logistic	log-logistic	<code>dist = llogistic</code>
normal	log-normal	<code>dist = lnormal</code>

In Proc Lifereg of SAS, all models are named for the distribution of T rather than the distribution of ε . Although these above models fitted by Proc Lifereg all are AFT models (so the regression coefficients have a unified interpretation), different distributions assume different shapes for the hazard function.

The exponential model

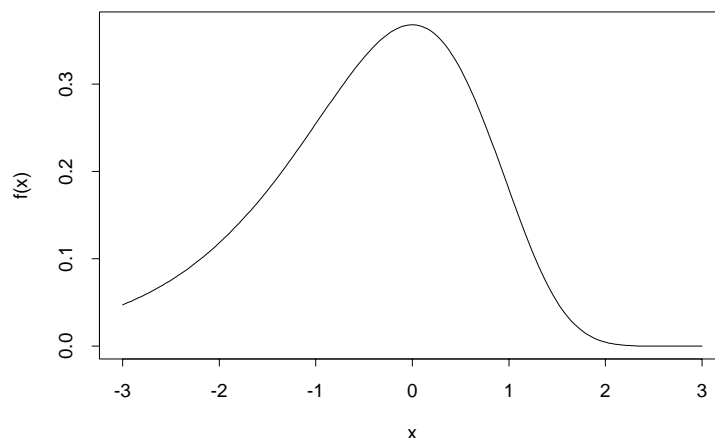
The simplest model is the exponential model where T at $z = 0$ (usually referred to as the baseline) has exponential distribution with constant hazard $\exp(-\beta_0)$. This is equivalent to assuming that $\sigma = 1$ and ε has a standard *extreme value* distribution

$$f(\varepsilon) = e^{\varepsilon - e^{\varepsilon}},$$

which has the density function shown in Figure 5.1. (So e^{ε} has the standard exponential distribution with constant hazard 1.)

From this specification, it is easy to see that the distribution of T at any covariate vector z is exponential with constant hazard (independent of t)

$$\lambda(t|z) = e^{-\beta_0 - \beta_1 z_1 - \dots - \beta_p z_p}.$$

Figure 5.1: *The density function of the standard extreme value distribution*

So automatically, we get *proportional hazards models*. For a given set of covariates (z_1, z_2, \dots, z_p) , the corresponding survival function is

$$S(t|z) = e^{-\lambda(t|z)t},$$

where $\lambda(t|z) = e^{-\beta_0 - \beta_1 z_1 - \dots - \beta_p z_p}$. Let $\beta_j^* = -\beta_j$. Then equivalently

$$\lambda(t|z) = e^{\beta_0^* + \beta_1^* z_1 + \dots + \beta_p^* z_p}.$$

Therefore, if we increase the value of covariate z_k ($k = 1, \dots, p$) by one unit from z_k to $z_k + 1$ while holding other covariate values fixed, then the ratio of the corresponding hazards is equal to

$$\frac{\lambda(t|z_k + 1)}{\lambda(t|z_k)} = e^{\beta_k^*}.$$

Thus $e^{\beta_k^*}$ can be interpreted as the hazard ratio corresponding to one unit increase in the covariate z_k , or equivalently, β_k^* can be interpreted as the increase in **log**-hazard as the value of covariate z_k increases by one unit (while other covariate values being held constant).

Note: Another SAS procedure Proc Phreg fits a proportional hazards model to the data and outputs the regression coefficient estimates in **log**-hazard form (i.e., in β_k^*). Therefore, if an

exponential model fits the data well (`Proc Phreg` will also fit the data well in this case.) then the regression coefficient estimates in outputs from `Proc Lifereg` using `dist=exponential` and `Proc Phreg` should be just opposite to each other (opposite sign but almost the same absolute value). We should be able to shift back and forth between these two models.

Example (Autologous and Allogeneic Bone Marrow Transplants for Hodgkin's and Non-Hodgkin's Lymphoma, pages 11-12 of the textbook): Data on 43 bone marrow transplant patients were collected. Patients had either Hodgkin's disease or Non-Hodgkin's Lymphoma, and were given either an allogeneic (Allo) transplant (from a HLA match sibling donor) or autogeneic (Auto) transplant (their own marrow was cleansed and returned to them after a high dose of chemotherapy). Other covariates are Karnofsky score (a subjective measure of how well the patient is doing, ranging from 0-100) and waiting time (in months) from diagnosis to transplant. It is of substantial interest to see the difference in leukemia-free survival (in days) between those patients given an Allo or Auto transplant, after adjusting for patients' disease status, Karnofsky score and waiting time. The data were given in Table 1.5 of the textbook. We used the following SAS program to fit an exponential model to the data

```
title "Exponential fit";
proc lifereg data=bone;
  model time*status(0) = allo hodgkins kscore wtime / dist=exponential;
run;
```

where `allo=1` for allogeneic transplant and `allo=0` for autologous transplant, `hodgkins=1` for Hodgkin's disease and `hodgkins=0` for Non-Hodgkin's Lymphoma, `kscore` is the Karnofsky score and `wtime` is the waiting time. We got the following output:

```
Exponential fit 1
17:35 Tuesday, March 1, 2005
```

The LIFEREG Procedure

Model Information

Data Set	WORK.BONE
Dependent Variable	Log(time)
Censoring Variable	status
Censoring Value(s)	0

Number of Observations	43
Noncensored Values	26
Right Censored Values	17
Left Censored Values	0
Interval Censored Values	0
Name of Distribution	Exponential
Log Likelihood	-62.49090652

Algorithm converged.

Type III Analysis of Effects

Effect	DF	Wald	
		Chi-Square	Pr > ChiSq
allo	1	0.0837	0.7723
hodgkins	1	6.2467	0.0124
kscore	1	64.8976	<.0001
wtime	1	1.6610	0.1975

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	0.6834	0.6408	-0.5726	1.9394	1.14	0.2862
allo	1	0.1333	0.4607	-0.7697	1.0363	0.08	0.7723
hodgkins	1	-1.3185	0.5275	-2.3524	-0.2845	6.25	0.0124
kscore	1	0.0758	0.0094	0.0574	0.0942	64.90	<.0001
wtime	1	0.0093	0.0072	-0.0049	0.0235	1.66	0.1975
Scale	0	1.0000	0.0000	1.0000	1.0000		
Weibull Shape	0	1.0000	0.0000	1.0000	1.0000		

Lagrange Multiplier Statistics

Parameter	Chi-Square	Pr > ChiSq
Scale	1.9089	0.1671

According to this model, allogeneic transplant is slightly better than autologous transplant after adjusting for disease status, Karnofsky score and waiting time. Hodgkins patients did worse than Non-Hodgkins patients (the average disease-free survival time for Hodgkins patients is only $\exp(-1.3185) = 0.27$ of that of the Non-Hodgkins patients). The patients with higher Karnofsky scores have better survival (with one point higher of Karnofsky score, the patients' average survival time will increase by about 7%). Waiting time has no effect on the disease-free survival.

The Weibull model

The only difference between the Weibull model and the exponential model is that the scale parameter σ is estimated rather than being set to be one. In this case, the distribution of $\sigma\varepsilon$ is an extreme value distribution with scale parameter σ . The survival function of T at covariate value $z = (1, z_1, \dots, z_p)^T$ can be shown to be

$$S(t|z) = \exp \left\{ - \left[t e^{-z^T \beta} \right]^{\frac{1}{\sigma}} \right\},$$

where $\beta = (\beta_0, \dots, \beta_p)^T$ is the vector of regression coefficients. Equivalently, in terms of (log)-hazard function

$$\log \lambda(t|z) = \left(\frac{1}{\sigma} - 1 \right) \log t - \log \sigma - z^T (\beta / \sigma).$$

Let $\alpha = 1/\sigma$, $\beta_0^* = -\log \sigma - \beta_0/\sigma$, and $\beta_j^* = -\beta_j/\sigma$ for $j = 1, \dots, p$ (again, pay close attention to the negative sign). Then we have

$$\log \lambda(t|z) = (\alpha - 1) \log t + \beta_0^* + z_1 \beta_1^* + \dots + z_p \beta_p^*.$$

Thus we also get a proportional hazards model and the coefficient β_k^* ($k = 1, \dots, p$) also has the interpretation that it is the increase in **log**-hazard when the value of covariate z_k increases by one unit while other covariate values being held unchanged. The function

$$\lambda_0(t) = t^{\alpha-1} e^{\beta_0^*} = \alpha t^{\alpha-1} e^{-\beta_0/\sigma} = \alpha t^{\alpha-1} e^{-\alpha \beta_0}$$

is the baseline hazard (*i.e.*, when $z = 0$).

Note: If the Weibull model is a reasonable model for your data and you use `Proc Lifereg` and `Proc Phreg` to fit the data, then the regression coefficient estimates not only have opposite signs (except possibly for the intercept) but also have different magnitude (depending on whether $\sigma > 1$ or $\sigma < 1$).

Since $\beta_k^* = -\beta_k/\sigma$ for $k = 1, \dots, p$. Testing $H_0 : \beta_k^* = 0$ is equivalent to testing $H_0 : \beta_k = 0$.

If we are interested in calculating standard error for the estimate of β_k^* and constructing a confidence interval for β_k^* , we can use delta method for this purpose.

Example (Bone marrow transplant data revisited) We used the following SAS program to fit a Weibull model to the bone marrow transplant data

```
title "Weibull fit";
proc lifereg data=bone;
  model time*status(0) = allo hodgkins kscore wtime / dist=weibull;
run;
```

and got the following output:

```

                                Weibull fit                                2
                                17:35 Tuesday, March 1, 2005

The LIFEREG Procedure

Model Information

Data Set                WORK.BONE
Dependent Variable      Log(time)
Censoring Variable      status
Censoring Value(s)     0
Number of Observations  43
Noncensored Values      26
Right Censored Values   17
Left Censored Values    0
Interval Censored Values 0
Name of Distribution     Weibull
Log Likelihood          -61.21034611
```

Algorithm converged.

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
allo	1	0.1351	0.7132
hodgkins	1	4.5212	0.0335
kscore	1	43.2179	<.0001
wtime	1	1.2210	0.2692

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi- Square	Pr > ChiSq
Intercept	1	0.4258	0.8463	-1.2329	2.0845	0.25	0.6148
allo	1	0.2080	0.5659	-0.9012	1.3172	0.14	0.7132
hodgkins	1	-1.3746	0.6465	-2.6417	-0.1075	4.52	0.0335
kscore	1	0.0793	0.0121	0.0557	0.1029	43.22	<.0001
wtime	1	0.0104	0.0094	-0.0081	0.0289	1.22	0.2692

Scale	1	1.2733	0.2044	0.9297	1.7440
Weibull Shape	1	0.7854	0.1260	0.5734	1.0757

If we think that the Weibull model is a reasonable one, then the likelihood ratio test statistic is $2*(-61.21034611 - (-62.49090652)) = 2.56$ and the p-value = 0.1096, not a strong evidence against the exponential model. From this model, we see similar results in the transplant methods.

The log-normal model

The log-normal model simply assumes that $\varepsilon \sim N(0, 1)$. Let $\lambda_0(t)$ be the hazard function of T when $\beta = 0$ ($\beta_0 = \beta_1 = \dots = \beta_p = 0$). Then it can be shown that $\lambda_0(t)$ has the following functional form

$$\lambda_0(t) = \frac{\phi\left(\frac{\log(t)}{\sigma}\right)}{\left[1 - \Phi\left(\frac{\log(t)}{\sigma}\right)\right] \sigma t},$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ is the probability density function and $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}}e^{-u^2/2}du$ is the cumulative distribution function of the standard normal distribution. Then the log-hazard function of T at any covariate value z can be expressed as

$$\log\lambda(t|z) = \log\lambda_0(te^{-z^T\beta}) - z^T\beta. \quad (5.3)$$

Obviously we no longer have a proportional hazards model.

Note: The function $\lambda_0(t)$ is not the baseline hazard function. If such a function is desired, it can be obtained from equation (5.3) by setting $z = 0$.

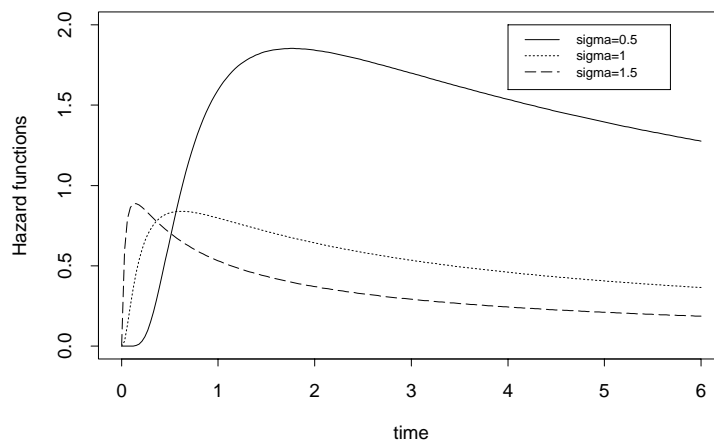
Some typical patterns that the hazard function $\lambda_0(t)$ assumes are presented in Figure 5.2.

The inverted U-shaped of the log-normal hazard is often appropriate for repeated events such as a residential move (*i.e.*, the interest is the time to next move). Immediately after a move, the hazard for another move is likely to be low, then increases with time, and eventually begins to decline since people tend to not move as they get older.

The survival function $S(t|z)$ at any covariate value z can be expressed as

$$\Phi^{-1}[S(t|z)] = \beta_0^* + \beta_1^*z_1 + \dots + \beta_p^*z_p - \alpha\log(t), \quad (5.4)$$

Figure 5.2: Typical hazard functions for a log-normal model



or equivalently

$$S(t|z) = \Phi[\beta_0^* + \beta_1^* z_1 + \cdots + \beta_p^* z_p - \alpha \log(t)],$$

where $\alpha = 1/\sigma$ and $\beta_j^* = \beta_j/\sigma$ for $j = 0, 1, \dots, p$. This is a *probit regression model* with intercept depending on t .

Note: Equation (5.4) indicates that the coefficients β_j^* can be estimated using Proc Logistic or Proc Genmod by specifying probit link function. Specifically, pick a time point of interest, say t_0 . Then dichotomize each subject based on his/her survival status at t_0 (in this case $\alpha \log(t_0)$ is absorbed into the intercept). Of course, there are some limitations using this approach. First, there should not be censoring prior to time t_0 . Second, the scale parameter σ is not estimable. Third, we will lose efficiency since we did not use all the information on the exact timing of the events. Since normal distribution and the logistic distribution we will introduce soon behave similarly to each other, the parameters β_k^* here have similar interpretation as the parameters in log-logistic model.

Example (Bone marrow transplant data revisited) If we want to fit a log-normal for the bone marrow transplant data, we use the following tt SAS program:

```
title "Log-normal fit";
```

```
proc lifereg data=bone;
  model time*status(0) = allo hodgkins kscore wtime / dist=lnormal;
run;
```

The following output is from the above program:

```

                                Log-normal fit
                                17:35 Tuesday, March 1, 2005
                                3

The LIFEREG Procedure

Model Information

Data Set                WORK.BONE
Dependent Variable      Log(time)
Censoring Variable      status
Censoring Value(s)      0
Number of Observations  43
Noncensored Values      26
Right Censored Values   17
Left Censored Values    0
Interval Censored Values 0
Name of Distribution     Lognormal
Log Likelihood          -60.93151115
```

Algorithm converged.

```

Type III Analysis of Effects

Effect      DF      Wald
            Chi-Square  Pr > ChiSq
allo        1         0.3556      0.5509
hodgkins    1         3.6277      0.0568
kscore      1        21.5358     <.0001
wtime       1         1.5309      0.2160
```

```

Analysis of Parameter Estimates

Parameter      DF      Estimate      Standard
              Error      95% Confidence
              Limits      Chi-Square
              Pr > ChiSq
Intercept      1      0.5064      1.1563
allo           1      0.3520      0.5902
hodgkins       1     -1.3137      0.6898
kscore         1      0.0663      0.0143
wtime          1      0.0142      0.0115
Scale          1      1.6296      0.2441
```

The results from this model are quite similar to the results from other models.

The Log-Logistic Model

The log-logistic model assumes that the disturbance term ε has a standard logistic distribution

$$f(\varepsilon) = \frac{e^\varepsilon}{(1 + e^\varepsilon)^2}.$$

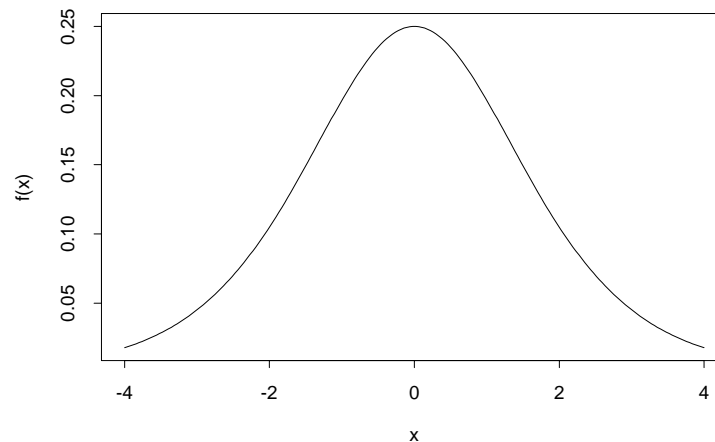
The density is plotted in Figure 5.3. Graphically, it looks like the standard normal density except that the standard normal density puts more mass around its mean value (0).

The hazard function of T at any covariate value z has a closed form:

$$\lambda(t|z) = \frac{\alpha t^{\alpha-1} e^{-z^T \beta / \sigma}}{1 + t^\alpha e^{-z^T \beta / \sigma}},$$

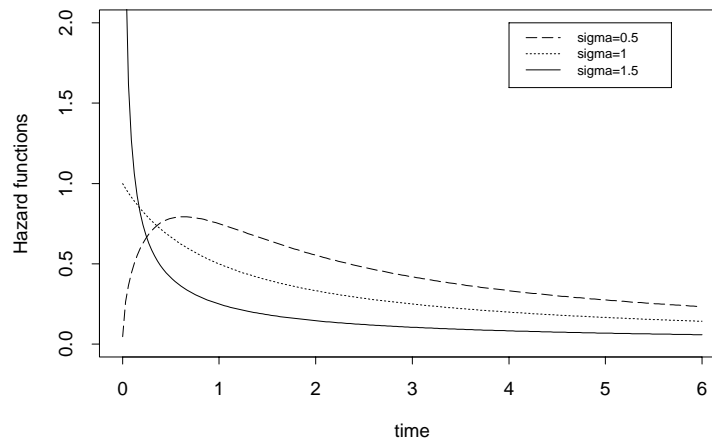
where $\alpha = 1/\sigma$.

Figure 5.3: *Density function of standard logistic distribution*



Since the logistic distribution looks similar to the normal distribution, it is expected that the hazard function of log-logistic distribution would also look like that of log-normal distribution, *i.e.*, would have an inverted U-shaped hazard. However, this is the case only for $\sigma < 1$. The hazard function of T when $\beta = 0$ for some values of σ 's is presented in Figure 5.4.

Figure 5.4: Hazard functions for the log-logistic distribution



The random variable T has a very simple survival function at covariate value z

$$S(t|z) = \frac{1}{1 + (te^{-z^T\beta})^{1/\sigma}}.$$

Some simple algebra then shows that

$$\log \left[\frac{S(t|z)}{1 - S(t|z)} \right] = \beta_0^* + \beta_1^* z_1 + \cdots + \beta_p^* z_p - \alpha \log(t), \quad (5.5)$$

where $\beta_j^* = \beta_j/\sigma$ for $j = 0, 1, \dots, p$. This is nothing but a *logistic regression model* with the intercept depending on t . Since $S(t|z)$ is the probability of surviving to time t for any given time t , the ratio $S(t|z)/(1 - S(t|z))$ is often called the *odds* of surviving to time t . Therefore, with one unit increase in z_k while other covariates being held fixed, the *odds ratio* is given by

$$\frac{S(t|z_k + 1)/(1 - S(t|z_k + 1))}{S(t|z_k)/(1 - S(t|z_k))} = e^{\beta_k^*} \text{ for all } t \geq 0,$$

which is a constant over time. Therefore, we have a *proportional odds models*. Hence β_k^* can be interpreted as the log odds ratio (for surviving) with one unit increase in z_k and $-\beta_k^*$ is the log odds ratio of dying before time t with one unit increase in z_k . At the times when the event of failure is rare (such as the early phase of a study), $-\beta_k^*$ can also be approximately interpreted as the log relative risk of dying. The log-logistic model is the *only* one that is both an AFT model and a proportional odds model.

Obviously, ε has the following cumulative distribution function

$$F(u) = \frac{e^u}{1 + e^u}, \quad u \in (-\infty, \infty),$$

whose inverse function

$$\text{logit}(\pi) = \log \left[\frac{\pi}{1 - \pi} \right], \quad \pi \in (0, 1),$$

is often called the logit function.

Note: As in the case of log-normal distribution, equation (5.5) indicates that the regression coefficients β_k^* can be estimated using Proc `logistic` or Proc `Genmod`. See the notes for log-normal model for the procedure and the limitations of such approach.

Example (Bone marrow transplant data revisited) We fit a log-logistic model to the data using the following SAS program:

```
title "Log-logistic fit";
proc lifereg data=bone;
  model time*status(0) = allo hodgkins kscore wtime / dist=llogistic;
run;
```

and got the following output:

```

                                Log-logistic fit                                4
                                                                17:35 Tuesday, March 1, 2005

                                The LIFEREG Procedure

                                Model Information

Data Set                                WORK.BONE
Dependent Variable                      Log(time)
Censoring Variable                       status
Censoring Value(s)                       0
Number of Observations                   43
Noncensored Values                       26
Right Censored Values                    17
Left Censored Values                     0
Interval Censored Values                  0
Name of Distribution                      LLogistic
Log Likelihood                           -61.15252729
```

Algorithm converged.

Type III Analysis of Effects

Effect	DF	Wald	
		Chi-Square	Pr > ChiSq
allo	1	0.3518	0.5531
hodgkins	1	4.6895	0.0303
kscore	1	24.0413	<.0001
wtime	1	1.4635	0.2264

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	0.5715	1.0620	-1.5101	2.6530	0.29	0.5905
allo	1	0.3419	0.5764	-0.7878	1.4715	0.35	0.5531
hodgkins	1	-1.4494	0.6693	-2.7613	-0.1376	4.69	0.0303
kscore	1	0.0669	0.0137	0.0402	0.0937	24.04	<.0001
wtime	1	0.0131	0.0108	-0.0081	0.0343	1.46	0.2264
Scale	1	0.9382	0.1567	0.6763	1.3016		

We got the same conclusion that two transplants are not significantly different after adjusting for other covariates. The effects of other covariates are similar too.

The gamma model

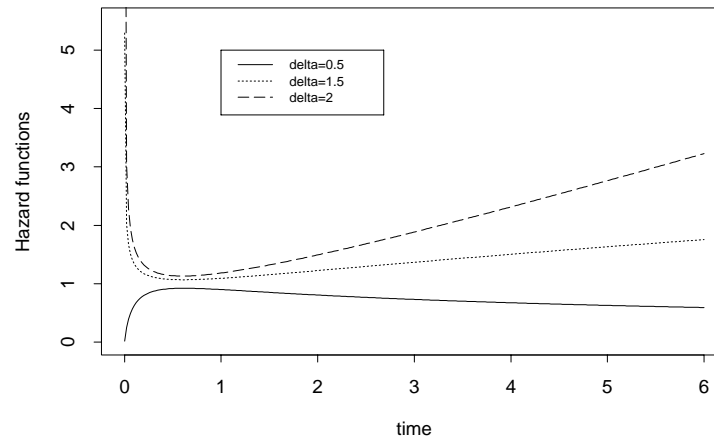
The procedure `Proc Lifereg` in SAS actually fits a generalized gamma model (not a standard gamma model) to the data by assuming $T_0 = e^\varepsilon$ to have the following density function

$$f(t) = |\delta|(t^\delta/\delta^2)^{1/\delta^2} \exp(-t^\delta/\delta^2)/(t\Gamma(1/\delta^2)),$$

where δ is called **shape** with label `Gamma shape parameter` in the output of `Proc Lifereg` when we specify `dist=gamma`. The hazard function of this gamma distribution does not have a closed form and is presented graphically in Figure 5.5 for some δ 's. **Note:** They are **not** the hazard functions of the standard gamma distribution.

Clearly from this plot, the hazard function is an inverted U-shaped function of time when $\delta < 1$ and takes a U-shape when $\delta > 1$. This feature makes gamma model very appropriate to model the survival times, especially for human. In practice, the hazard function is determined jointly by the scale parameter σ and the shape parameter δ and we need to examine the resulting

Figure 5.5: Hazard functions for the gamma distribution



hazard function case by case.

For a given set of covariates (z_1, z_2, \dots, z_p) , let $c = e^{\beta_0 + z_1\beta_1 + \dots + z_k\beta_k} = e^{z^T\beta}$. Then $\log(T) = z^T\beta + \sigma\epsilon$ implies $T = e^{z^T\beta} * [e^\epsilon]^\sigma = cT_0^\sigma$. Hence the survival function for this population is

$$\begin{aligned}
 S(t|z) &= P[T \geq t] = P[cT_0^\sigma \geq t] \\
 &= P[T_0 \geq (c^{-1}t)^{1/\sigma}] = P[T_0 \geq b(t)] \\
 &= \int_{b(t)}^{\infty} |\delta| (x^\delta/\delta^2)^{1/\delta^2} \exp(-x^\delta/\delta^2) / (x\Gamma(1/\delta^2)) dx \\
 &\stackrel{x^\delta/\delta^2=y}{=} \begin{cases} \int_{\alpha b(t)^\delta}^{\infty} \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)} dy & \text{if } \delta > 0 \\ \int_0^{\alpha b(t)^\delta} \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)} dy & \text{if } \delta < 0, \end{cases}
 \end{aligned}$$

where $\alpha = 1/\delta^2$. This final integration can be calculated using built-in functions in any popular software.

Example (Bone marrow transplant data revisited) We fit a gamma model to the bone marrow transplant data using the following SAS program

```

title "Gamma fit";
proc lifereg data=bone;
  model time*status(0) = allo hodgkins kscore wtime / dist=gamma;
run;

```

and got the following results:

Gamma fit

5

17:35 Tuesday, March 1, 2005

The LIFEREG Procedure

Model Information

Data Set	WORK.BONE
Dependent Variable	Log(time)
Censoring Variable	status
Censoring Value(s)	0
Number of Observations	43
Noncensored Values	26
Right Censored Values	17
Left Censored Values	0
Interval Censored Values	0
Name of Distribution	Gamma
Log Likelihood	-60.91369551

Algorithm converged.

Type III Analysis of Effects

Effect	DF	Wald	
		Chi-Square	Pr > ChiSq
allo	1	0.4046	0.5247
hodgkins	1	3.7275	0.0535
kscore	1	11.3847	0.0007
wtime	1	1.2282	0.2678

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	0.4299	1.1407	-1.8057	2.6656	0.14	0.7062
allo	1	0.3734	0.5870	-0.7772	1.5240	0.40	0.5247
hodgkins	1	-1.3098	0.6784	-2.6394	0.0199	3.73	0.0535
kscore	1	0.0696	0.0206	0.0292	0.1100	11.38	0.0007
wtime	1	0.0133	0.0120	-0.0102	0.0367	1.23	0.2678
Scale	1	1.5771	0.3805	0.9829	2.5307		
Shape	1	0.2047	1.0278	-1.8098	2.2192		

Again, there is no significant different in the transplant methods. Some special cases:

1. $\delta = 1 \Rightarrow T|z$ has the Weibull distribution.
2. $\delta = 0 \Rightarrow T|z$ has the log-normal distribution. We need the following approximation in

order to show this:

$$\Gamma(x) \approx \sqrt{2\pi x} x^{-\frac{1}{2}} e^{-x} \quad \text{as } x \rightarrow \infty.$$

3. $\delta = \sigma \Rightarrow T|z$ has the standard gamma distribution with the following density:

$$f(t|z) = \frac{t^{K-1} e^{-\frac{t}{\gamma}}}{\gamma^K \Gamma(K)},$$

where $K = 1/\delta^2$ is the shape parameter and $\gamma = \delta^2 e^{z^T \beta}$ is the scale parameter in the standard gamma distribution. However, `Proc Lifereg` will not fit a standard gamma distribution to data. We have to use grid search to fit a standard gamma distribution. Specifically, use the output from a generalized gamma distribution to get an idea about the true value of $\sigma = \delta$. Then form a grid. For each grid point of σ (or δ), say, $\sigma = 1.2$, fit a gamma distribution with the specification `dist=gamma noshape1 shape1=1.2 noscale scale=1.2;`. Select the model that gives the largest log-likelihood.

Categorical Variables and Class statement in Proc Lifereg

If a covariate z_k is categorical, then we can use `class` statement in `Proc Lifereg` to tell the procedure that z_k is categorical and just enter z_k in the model statement in a usual way. Then the output of `Proc Lifereg` will provide a χ^2 statistic and a p -value to test the null hypothesis that the survival time is not associated with z_k . Then it reports the estimates, standard errors, and p -values, etc., for the contrasts between each level and the highest level of the category. However, we need to create appropriate variables for the interaction.

5.3 Goodness-of-fit Using Likelihood Ratio Test

The following fact on the models we described in this chapter allows us to perform likelihood ratio tests in order to pick a right model:

1. generalized gamma $(\delta, \sigma) \supset$ standard gamma $(\delta = \sigma) \supset$ exponential distribution $(\delta = \sigma = 1)$.

2. generalized gamma $(\delta, \sigma) \supset$ Weibull $(\delta = 1, \sigma) \supset$ exponential distribution $(\delta = \sigma = 1)$.
3. generalized gamma $(\delta, \sigma) \supset$ log-normal $(\delta = 0, \sigma)$.

We can use the above nested models to conduct likelihood ratio test for the bone marrow transplant data:

Maximum likelihood values for different models

Model	Maximum Likelihood
Gamma	-60.91
Log-logistic	-61.15
Log-normal	-60.93
Weibull	-61.21
Exponential	-62.49

Assuming the Gamma model is a reasonable model for the data, the LRT indicates that the log-normal model is equally good and the Weibull model is also acceptable. Since the log-normal model and the Weibull model have the same number of parameters, we might want to take the log-normal model as the final model based on the larger maximum log-likelihood value.