Chapter 18
Parametric Survival Models

18.1 Homogeneous Models (No Predictors)

The nonparametric estimator of $S(t)$ is a very good descriptive statistic for displaying survival data. For many purposes, however, one may want to make more assumptions to allow the data to be modeled in more detail. By specifying a functional form for $S(t)$ and estimating any unknown parameters in this function, one can

1. easily compute selected quantiles of the survival distribution;
2. estimate (usually by extrapolation) the expected failure time;
3. derive a concise equation and smooth function for estimating $S(t)$, $\Lambda(t)$, and $\lambda(t)$; and
4. estimate $S(t)$ more precisely than $S_{\text{KM}}(t)$ or $S_{\Lambda}(t)$ if the parametric form is correctly specified.

18.1.1 Specific Models

Parametric modeling requires choosing one or more distributions. The Weibull and exponential distributions were discussed in Chapter 18. Other commonly used survival distributions are obtained by transforming T and using a standard distribution. The log transformation is most commonly employed. The *log-normal* distribution specifies that $\log(T)$ has a normal distribution with mean μ and variance σ^2. Stated another way, $\log(T) \sim \mu + \sigma\epsilon$, where ϵ has a standard normal distribution. Then $S(t) = 1 - \Phi((\log(t) - \mu)/\sigma)$, where Φ is the standard normal cumulative distribution function. The *log-logistic* distribution is given by $S(t) = [1 + \exp(-(\log(t) - \mu)/\sigma)]^{-1}$. Here $\log(T) \sim \mu + \sigma\epsilon$ where ϵ follows a logistic distribution $[1+\exp(-u)]^{-1}$. The *log*

F.E. Harrell, Jr., *Regression Modeling Strategies*, Springer Series in Statistics, DOI 10.1007/978-3-319-19425-7_18

extreme value distribution is given by $S(t) = \exp[-\exp((\log(t) - \mu)/\sigma)]$, and $\log(T) \sim \mu + \sigma\epsilon$, where $\epsilon \sim 1 - \exp[-\exp(u)]$.

The generalized gamma and generalized F distributions provide a richer variety of distribution and hazard functions[127,128]. Spline hazard models[286,287,361] are other excellent alternatives.

18.1.2 Estimation

Maximum likelihood (ML) estimation is used to estimate the unknown parameters of $S(t)$. The general method presented in Chapter 9 must be augmented, however, to allow for censored failure times. The basic idea is as follows. Again let T be a random variable representing time until the event, T_i be the (possibly censored) failure time for the ith observation, and Y_i denote the observed failure or censoring time $\min(T_i, C_i)$, where C_i is the censoring time. If Y_i is uncensored, observation i contributes a factor to the likelihood equal to the density function for T evaluated at Y_i, $f(Y_i)$. If Y_i instead represents a censored time so that $T_i = Y_i^+$, it is only known that T_i exceeds Y_i. The contribution to the likelihood function is the probability that $T_i > C_i$ (equal to $\text{Prob}\{T_i > Y_i\}$). This probability is $S(Y_i)$. The joint likelihood over all observations $i = 1, 2, \ldots, n$ is

$$L = \prod_{i:Y_i \text{ uncensored}}^{n} f(Y_i) \prod_{i:Y_i \text{ censored}}^{n} S(Y_i). \tag{18.1}$$

There is one more component to L: the distribution of censoring times if these are not fixed in advance. Recall that we assume that censoring is non-informative, that is, it is independent of the risk of the event. This independence implies that the likelihood component of the censoring distribution simply multiplies L and that the censoring distribution contains little information about the survival distribution. In addition, the censoring distribution may be very difficult to specify. For these reasons we can maximize L separately to estimate parameters of $S(t)$ and ignore the censoring distribution.

Recalling that $f(t) = \lambda(t)S(t)$ and $\Lambda(t) = -\log S(t)$, the log likelihood can be written as

$$\log L = \sum_{i:Y_i \text{ uncensored}}^{n} \log \lambda(Y_i) - \sum_{i=1}^{n} \Lambda(Y_i). \tag{18.2}$$

All observations then contribute an amount to the log likelihood equal to the negative of the cumulative hazard evaluated at the failure/censoring time. In addition, uncensored observations contribute an amount equal to the log of the hazard function evaluated at the time of failure. Once L or $\log L$ is specified, the general ML methods outlined earlier can be used without

change in most situations. The principal difference is that censored observations contribute less information to the statistical inference than uncensored observations. For distributions such as the log-normal that are written only in terms of $S(t)$, it may be easier to write the likelihood in terms of $S(t)$ and $f(t)$.

As an example, we turn to the exponential distribution, for which $\log L$ has a simple form that can be maximized explicitly. Recall that for this distribution $\lambda(t) = \lambda$ and $\Lambda(t) = \lambda t$. Therefore,

$$\log L = \sum_{i:Y_i \text{ uncensored}}^{n} \log \lambda - \sum_{i=1}^{n} \lambda Y_i. \tag{18.3}$$

Letting n_u denote the number of uncensored event times,

$$\log L = n_u \log \lambda - \sum_{i=1}^{n} \lambda Y_i. \tag{18.4}$$

Letting w denote the sum of all failure/censoring times ("person years of exposure"):

$$w = \sum_{i=1}^{n} Y_i, \tag{18.5}$$

the derivatives of $\log L$ are given by

$$\begin{aligned} \frac{\partial \log L}{\partial \lambda} &= n_u/\lambda - w \\ \frac{\partial^2 \log L}{\partial \lambda^2} &= -n_u/\lambda^2. \end{aligned} \tag{18.6}$$

Equating the derivative of $\log L$ to zero implies that the MLE of λ is

$$\hat{\lambda} = n_u/w \tag{18.7}$$

or the number of failures per person-years of exposure. By inserting the MLE of λ into the formula for the second derivative we obtain the observed estimated information, w^2/n_u. The estimated variance of $\hat{\lambda}$ is thus n_u/w^2 and the standard error is $n_u^{1/2}/w$. The precision of the estimate depends primarily on n_u.

Recall that the expected life length μ is $1/\lambda$ for the exponential distribution. The MLE of μ is w/n_u and its estimated variance is w^2/n_u^3. The MLE of $S(t)$, $\hat{S}(t)$, is $\exp(-\hat{\lambda}t)$, and the estimated variance of $\log(\hat{\Lambda}(t))$ is simply $1/n_u$.

As an example, consider the sample listed previously,

$$1 \;\; 3 \;\; 3 \;\; 6^+ \;\; 8^+ \;\; 9 \;\; 10^+.$$

Here $n_u = 4$ and $w = 40$, so the MLE of λ is 0.1 failure per person-period. The estimated standard error is $2/40 = 0.05$. Estimated expected life length is 10 units with a standard error of 5 units. Estimated median failure time is $\log(2)/0.1 = 6.931$. The estimated survival function is $\exp(-0.1t)$, which at $t = 1, 3, 9, 10$ yields $0.90, 0.74, 0.41$, and 0.37, which can be compared to the product limit estimates listed earlier $(0.85, 0.57, 0.29, 0.29)$.

Now consider the Weibull distribution. The log likelihood function is

$$\log L = \sum_{i:Y_i \text{ uncensored}}^{n} \log[\alpha\gamma Y_i^{\gamma-1}] - \sum_{i=1}^{n} \alpha Y_i^{\gamma}. \tag{18.8}$$

Although $\log L$ can be simplified somewhat, it cannot be solved explicitly for α and γ. An iterative method such as the Newton–Raphson method is used to compute the MLEs of α and γ. Once these estimates are obtained, the estimated variance–covariance matrix and other derived quantities such as $\hat{S}(t)$ can be obtained in the usual manner.

For the dataset used in the exponential fit, the Weibull fit follows.

$$\begin{aligned} \hat{\alpha} &= 0.0728 \\ \hat{\gamma} &= 1.164 \\ \hat{S}(t) &= \exp(-0.0728t^{1.164}) \\ \hat{S}^{-1}(0.5) &= [(\log 2)/\hat{\alpha}]^{1/\hat{\gamma}} = 6.935 \text{ (estimated median)}. \end{aligned} \tag{18.9}$$

This fit is very close to the exponential fit since $\hat{\gamma}$ is near 1.0. Note that the two medians are almost equal. The predicted survival probabilities for the Weibull model for $t = 1, 3, 9, 10$ are, respectively, $0.93, 0.77, 0.39, 0.35$.

Sometimes a formal test can be made to assess the fit of the proposed parametric survival distribution. For the data just analyzed, a formal test of exponentiality versus a Weibull alternative is obtained by testing $H_0 : \gamma = 1$ in the Weibull model. A score test yielded $\chi^2 = 0.14$ with 1 d.f., $p = 0.7$, showing little evidence for non-exponentiality (note that the sample size is too small for this test to have any power).

18.1.3 Assessment of Model Fit

The fit of the hypothesized survival distribution can often be checked easily using graphical methods. Nonparametric estimates of $S(t)$ and $\Lambda(t)$ are primary tools for this purpose. For example, the Weibull distribution $S(t) = \exp(-\alpha t^{\gamma})$ can be rewritten by taking logarithms twice:

$$\log[-\log S(t)] = \log \Lambda(t) = \log \alpha + \gamma(\log t). \tag{18.10}$$

The fit of a Weibull model can be assessed by plotting $\log \hat{\Lambda}(t)$ versus $\log t$ and checking whether the curve is approximately linear. Also, the plotted curve provides approximate estimates of α (the antilog of the intercept) and γ (the slope). Since an exponential distribution is a special case of a Weibull distribution when $\gamma = 1$, exponentially distributed data will tend to have a graph that is linear with a slope of 1.

For any assumed distribution $S(t)$, a graphical assessment of goodness of fit can be made by plotting $S^{-1}[S_{\Lambda}(t)]$ or $S^{-1}[S_{\mathrm{KM}}(t)]$ against t and checking for linearity. For log distributions, S specifies the distribution of $\log(T)$, so we plot against $\log t$. For a log-normal distribution we thus plot $\Phi^{-1}[S_{\Lambda}(t)]$ against $\log t$, where Φ^{-1} is the inverse of the standard normal cumulative distribution function. For a log-logistic distribution we plot $\mathrm{logit}[S_{\Lambda}(t)]$ versus $\log t$. For an extreme value distribution we use $\log - \log$ plots as with the Weibull distribution. Parametric model fits can also be checked by plotting the fitted $\hat{S}(t)$ and $S_{\Lambda}(t)$ against t on the same graph.

18.2 Parametric Proportional Hazards Models

In this section we present one way to generalize the survival model to a survival regression model. In other words, we allow the sample to be heterogeneous by adding predictor variables $X = \{X_1, X_2, \ldots, X_k\}$. As with other regression models, X can represent a mixture of binary, polytomous, continuous, spline-expanded, and even ordinal predictors (if the categories are scored to satisfy the linearity assumption). Before discussing ways in which the regression part of a survival model might be specified, first recall how regression effects have been modeled in other settings. In multiple linear regression, the regression effect $X\beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$ can be thought of as an increment in the expected value of the response Y. In binary logistic regression, $X\beta$ specifies the log odds that $Y = 1$, or $\exp(X\beta)$ multiplies the odds that $Y = 1$.

18.2.1 Model

The most widely used survival regression specification is to allow the hazard function $\lambda(t)$ to be multiplied by $\exp(X\beta)$. The survival model is thus generalized from a hazard function $\lambda(t)$ for the failure time T to a hazard function $\lambda(t)\exp(X\beta)$ for the failure time given the predictors X:

$$\lambda(t|X) = \lambda(t)\exp(X\beta). \tag{18.11}$$

This regression formulation is called the *proportional hazards (PH)* model. The $\lambda(t)$ part of $\lambda(t|X)$ is sometimes called an *underlying hazard function* or a *hazard function for a standard subject*, which is a subject with $X\beta = 0$. Any parametric hazard function can be used for $\lambda(t)$, and as we show later, $\lambda(t)$ can be left completely unspecified without sacrificing the ability to estimate β, by the use of Cox's semi-parametric PH model.[132] Depending on whether the underlying hazard function $\lambda(t)$ has a constant scale parameter, $X\beta$ may or may not include an intercept β_0. The term $\exp(X\beta)$ can be called a *relative hazard function* and in many cases it is the function of primary interest as it describes the (relative) effects of the predictors.

The PH model can also be written in terms of the cumulative hazard and survival functions:

$$\begin{aligned} \Lambda(t|X) &= \Lambda(t)\exp(X\beta) \\ S(t|X) &= \exp[-\Lambda(t)\exp(X\beta)] = \exp[-\Lambda(t)]^{\exp(X\beta)}. \end{aligned} \tag{18.12}$$

$\Lambda(t)$ is an "underlying" cumulative hazard function. $S(t|X)$, the probability of surviving past time t given the values of the predictors X, can also be written as

$$S(t|X) = S(t)^{\exp(X\beta)}, \tag{18.13}$$

where $S(t)$ is the "underlying" survival distribution, $\exp(-\Lambda(t))$. The effect of the predictors is to multiply the hazard and cumulative hazard functions by a factor $\exp(X\beta)$, or equivalently to raise the survival function to a power equal to $\exp(X\beta)$.

18.2.2 Model Assumptions and Interpretation of Parameters

In the general regression notation of Section 2.2, the log hazard or log cumulative hazard can be used as the property of the response T evaluated at time t that allows distributional and regression parts to be isolated and checked. The PH model can be linearized with respect to $X\beta$ using the following identities.

$$\begin{aligned} \log\lambda(t|X) &= \log\lambda(t) + X\beta \\ \log\Lambda(t|X) &= \log\Lambda(t) + X\beta. \end{aligned} \tag{18.14}$$

No matter which of the three model statements are used, there are certain assumptions in a parametric PH survival model. These assumptions are listed below.

1. The true form of the underlying functions (λ, Λ, and S) should be specified correctly.

2. The relationship between the predictors and log hazard or log cumulative hazard should be linear in its simplest form. In the absence of interaction terms, the predictors should also operate additively.
3. The way in which the predictors affect the distribution of the response should be by multiplying the hazard or cumulative hazard by $\exp(X\beta)$ or equivalently by adding $X\beta$ to the log hazard or log cumulative hazard at each t. The effect of the predictors is assumed to be the same at all values of t since $\log \lambda(t)$ can be separated from $X\beta$. In other words, the PH assumption implies no t by predictor interaction.

The regression coefficient for X_j, β_j, is the increase in log hazard or log cumulative hazard at any fixed point in time if X_j is increased by one unit and all other predictors are held constant. This can be written formally as

$$\beta_j = \log \lambda(t|X_1, X_2, \ldots, X_j + 1, X_{j+1}, \ldots, X_k) - \log \lambda(t|X_1, \ldots, X_j, \ldots, X_k), \tag{18.15}$$

which is equivalent to the log of the ratio of the hazards at time t. The regression coefficient can just as easily be written in terms of a ratio of hazards at time t. The ratio of hazards at $X_j + d$ versus X_j, all other factors held constant, is $\exp(\beta_j d)$. Thus the effect of increasing X_j by d is to increase the hazard of the event by a factor of $\exp(\beta_j d)$ at all points in time, assuming X_j is linearly related to $\log \lambda(t)$. In general, the ratio of hazards for an individual with predictor variable values X^* compared to an individual with predictors X is

$$\begin{aligned} X^* : X \text{ hazard ratio} &= [\lambda(t)\exp(X^*\beta)]/[\lambda(t)\exp(X\beta)] \\ &= \exp(X^*\beta)/\exp(X\beta) = \exp[(X^* - X)\beta]. \end{aligned} \tag{18.16}$$

If there is only one predictor X_1 and that predictor is binary, the PH model can be written

$$\begin{aligned} \lambda(t|X_1 = 0) &= \lambda(t) \\ \lambda(t|X_1 = 1) &= \lambda(t)\exp(\beta_1). \end{aligned} \tag{18.17}$$

Here $\exp(\beta_1)$ is the $X_1 = 1 : X_1 = 0$ hazard ratio. This simple case has no regression assumption but assumes PH and a form for $\lambda(t)$. If the single predictor X_1 is continuous, the model becomes

$$\lambda(t|X_1) = \lambda(t)\exp(\beta_1 X_1). \tag{18.18}$$

Without further modification (such as taking a transformation of the predictor), the model assumes a straight line in the log hazard or that for all t, an increase in X by one unit increases the hazard by a factor of $\exp(\beta_1)$.

As in logistic regression, much more general regression specifications can be made, including interaction effects. Unlike logistic regression, however, a model containing, say age, sex, and age $\times$ sex interaction is not equivalent to

fitting two separate models. This is because even though males and females are allowed to have unequal age slopes, both sexes are assumed to have the

Table 18.1 Mortality differences and ratios when hazard ratio is 0.5

Subject	5-Year Survival		Difference	Mortality Ratio (T/C)
	C	T		
1	0.98	0.99	0.01	0.01/0.02 = 0.5
2	0.80	0.89	0.09	0.11/0.2 = 0.55
3	0.25	0.50	0.25	0.5/0.75 = 0.67

underlying hazard function proportional to $\lambda(t)$ (i.e., the PH assumption holds for sex in addition to age).

18.2.3 Hazard Ratio, Risk Ratio, and Risk Difference

Other ways of modeling predictors can also be specified besides a multiplicative effect on the hazard. For example, one could postulate that the effect of a predictor is to add to the hazard of failure instead of to multiply it by a factor. The effect of a predictor could also be described in terms of a mortality ratio (relative risk), risk difference, odds ratio, or increase in expected failure time. However, just as an odds ratio is a natural way to describe an effect on a binary response, a hazard ratio is often a natural way to describe an effect on survival time. One reason is that a hazard ratio *can* be constant.

Table 18.1 provides treated (T) to control (C) survival (mortality) differences and mortality ratios for three hypothetical types of subjects. We suppose that subjects 1, 2, and 3 have increasingly worse prognostic factors. For example, the age at baseline of the subjects might be 30, 50, and 70 years, respectively. We assume that the treatment affects the hazard by a constant multiple of 0.5 (i.e., PH is in effect and the constant hazard ratio is 0.5). Note that $S_T = S_C^{0.5}$. Notice that the mortality difference and ratio depend on the survival of the control subject. A control subject having "good" predictor values will leave little room for an improved prognosis from the treatment.

The hazard ratio is a basis for describing the mechanism of an effect. In the above example, it is reasonable that the treatment affect each subject by lowering her hazard of death by a factor of 2, even though less sick subjects have a low mortality difference. Hazard ratios also lead to good statistical tests for differences in survival patterns and to predictive models. Once the model is developed, however, survival differences may better capture the impact of a risk factor. Absolute survival differences rather than relative differences

(hazard ratios) also relate more closely to statistical power. For example, even if the effect of a treatment is to halve the hazard rate, a population where the control survival is 0.99 will require a much larger sample than will a population where the control survival is 0.3.

Figure 18.1 depicts the relationship between survival $S(t)$ of a control subject at any time t, relative reduction in hazard (h), and difference in survival $S(t) - S(t)^h$. This figure demonstrates that absolute clinical benefit

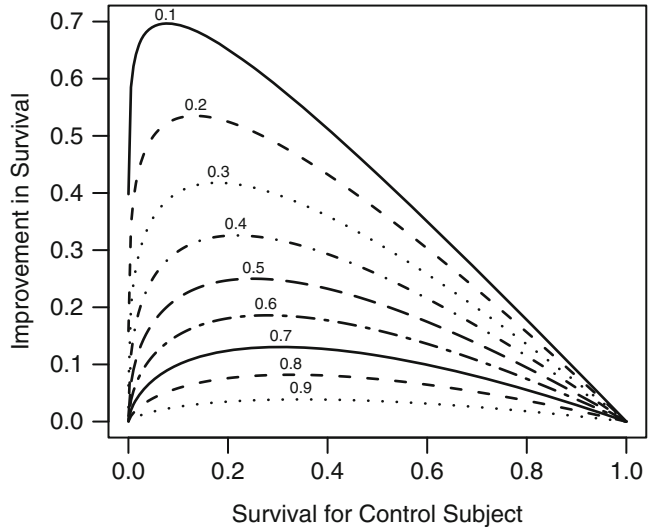

Fig. 18.1 Absolute clinical benefit as a function of survival in a control subject and the relative benefit (hazard ratio). The hazard ratios are given for each curve.

is primarily a function of the baseline risk of a subject. Clinical benefit will also be a function of factors that interact with treatment, that is, factors that modify the relative benefit of treatment. Once a model is developed for estimating $S(t|X)$, this model can be used to estimate absolute benefit as a function of baseline risk factors as well as factors that interact with a treatment. Let X_1 be a binary treatment indicator and let $A = \{X_2, \ldots, X_p\}$ be the other factors (which for convenience we assume do not interact with X_1). Then the estimate of $S(t|X_1 = 0, A) - S(t|X_1 = 1, A)$ can be plotted against $S(t|X_1 = 0)$ or against levels of variables in A to display absolute benefit versus overall risk or specific subject characteristics. 1

18.2.4 Specific Models

Let $X\beta$ denote the linear combination of predictors excluding an intercept term. Using the PH formulation, an exponential survival regression model[218] can be stated as

$$\begin{aligned}\lambda(t|X) &= \lambda \exp(X\beta)\\ S(t|X) &= \exp[-\lambda t \exp(X\beta)] = \exp(-\lambda t)^{\exp(X\beta)}.\end{aligned} \tag{18.19}$$

The parameter λ can be thought of as the antilog of an intercept term since the model could be written $\lambda(t|X) = \exp[(\log \lambda) + X\beta]$. The effect of X on the expected or median failure time is as follows.

$$\begin{aligned}E\{T|X\} &= 1/[\lambda \exp(X\beta)]\\ T_{0.5}|X &= (\log 2)/[\lambda \exp(X\beta)].\end{aligned} \tag{18.20}$$

The exponential regression model can be written in another form that is more numerically stable by replacing the λ parameter with an intercept term in $X\beta$, specifically $\lambda = \exp(\beta_0)$. After redefining $X\beta$ to include β_0, λ can be dropped in all the above formulas.

The Weibull regression model is defined by one of the following functions (assuming that $X\beta$ does not contain an intercept).

$$\begin{aligned}\lambda(t|X) &= \alpha\gamma t^{\gamma-1} \exp(X\beta)\\ \Lambda(t|X) &= \alpha t^{\gamma} \exp(X\beta)\\ S(t|X) &= \exp[-\alpha t^{\gamma} \exp(X\beta)]\\ &= [\exp(-\alpha t^{\gamma})]^{\exp(X\beta)}.\end{aligned} \tag{18.21}$$

Note that the parameter α in the homogeneous Weibull model has been replaced with $\alpha \exp(X\beta)$. The median survival time is given by

$$T_{0.5}|X = \{\log 2/[\alpha \exp(X\beta)]\}^{1/\gamma}. \tag{18.22}$$

As with the exponential model, the parameter α could be dropped (and replaced with $\exp(\beta_0)$) if an intercept β_0 is added to $X\beta$.

For numerical reasons it is sometimes advantageous to write the Weibull PH model as

$$S(t|X) = \exp(-\Lambda(t|X)), \tag{18.23}$$

where

$$\Lambda(t|X) = \exp(\gamma \log t + X\beta). \tag{18.24}$$

18.2.5 Estimation

The parameters in λ and β are estimated by maximizing a log likelihood function constructed in the same manner as described in Section 18.1. The only difference is the insertion of $\exp(X_i\beta)$ in the likelihood function:

$$\log L = \sum_{i:Y_i \text{ uncensored}}^{n} \log[\lambda(Y_i)\exp(X_i\beta)] - \sum_{i=1}^{n} \Lambda(Y_i)\exp(X_i\beta). \quad (18.25)$$

Once $\hat{\beta}$, the MLE of β, is computed along with the large-sample standard error estimates, hazard ratio estimates and their confidence intervals can readily be computed. Letting s denote the estimated standard error of $\hat{\beta}_j$, a $1-\alpha$ confidence interval for the $X_j+1 : X_j$ hazard ratio is given by $\exp[\hat{\beta}_j \pm zs]$, where z is the $1-\alpha/2$ critical value for the standard normal distribution.

Once the parameters of the underlying hazard function are estimated, the MLE of $\lambda(t)$, $\hat{\lambda}(t)$, can be derived. The MLE of $\lambda(t|X)$, the hazard as a function of t and X, is given by

$$\hat{\lambda}(t|X) = \hat{\lambda}(t)\exp(X\hat{\beta}). \quad (18.26)$$

The MLE of $\Lambda(t)$, $\hat{\Lambda}(t)$, can be derived from the integral of $\hat{\lambda}(t)$ with respect to t. Then the MLE of $S(t|X)$ can be derived:

$$\hat{S}(t|X) = \exp[-\hat{\Lambda}(t)\exp(X\hat{\beta})]. \quad (18.27)$$

For the Weibull model, we denote the MLEs of the hazard parameters α and γ by $\hat{\alpha}$ and $\hat{\gamma}$. The MLE of $\lambda(t|X)$, $\Lambda(t|X)$, and $S(t|X)$ for this model are

$$\begin{aligned} \hat{\lambda}(t|X) &= \hat{\alpha}\hat{\gamma}t^{\hat{\gamma}-1}\exp(X\hat{\beta}) \\ \hat{\Lambda}(t|X) &= \hat{\alpha}t^{\hat{\gamma}}\exp(X\hat{\beta}) \\ \hat{S}(t|X) &= \exp[-\hat{\Lambda}(t|X)]. \end{aligned} \quad (18.28)$$

Confidence intervals for $S(t|X)$ are best derived using general matrix notation to obtain an estimate s of the standard error of $\log[\hat{\lambda}(t|X)]$ from the estimated information matrix of all hazard and regression parameters. A confidence interval for $\hat{S}$ will be of the form

$$\hat{S}(t|X)^{\exp(\pm zs)}. \quad (18.29)$$

The MLEs of β and of the hazard shape parameters lead directly to MLEs of the expected and median life length. For the Weibull model the MLE of the median life length given X is

$$\hat{T}_{0.5}|X = \{\log 2/[\hat{\alpha}\exp(X\hat{\beta})]\}^{1/\hat{\gamma}}. \quad (18.30)$$

For the exponential model, the MLE of the expected life length for a subject having predictor values X is given by

$$\hat{\mathrm{E}}(T|X) = [\hat{\lambda}\exp(X\hat{\beta})]^{-1}, \quad (18.31)$$

where $\hat{\lambda}$ is the MLE of λ.

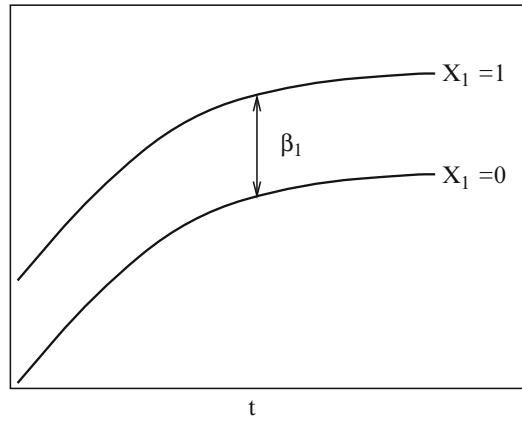

Fig. 18.2 PH model with one binary predictor. Y-axis is $\log \lambda(t)$ or $\log \Lambda(t)$. For $\log \Lambda(t)$, the curves must be non-decreasing. For $\log \lambda(t)$, they may be any shape.

18.2.6 Assessment of Model Fit

Three assumptions of the parametric PH model were listed in Section 18.2.2. We now lay out in more detail what relationships need to be satisfied. We first assume a PH model with a single binary predictor X_1. For a general underlying hazard function $\lambda(t)$, all assumptions of the model are displayed in Figure 18.2. In this case, the assumptions are PH and a shape for $\lambda(t)$.

If $\lambda(t)$ is Weibull, the two curves will be linear if $\log t$ is plotted instead of t on the x-axis. Note also that if there is no association between X and survival ($\beta_1 = 0$), estimates of the two curves will be close and will intertwine due to random variability. In this case, PH is not an issue.

If the single predictor is continuous, the relationships in Figures 18.3 and 18.4 must hold. Here linearity is assumed (unless otherwise specified) besides PH and the form of $\lambda(t)$. In Figure 18.3, the curves must be parallel for any choices of times t_1 and t_2 as well as each individual curve being linear. Also, the difference between ordinates needs to conform to the assumed distribution. This difference is $\log[\lambda(t_2)/\lambda(t_1)]$ or $\log[\Lambda(t_2)/\Lambda(t_1)]$.

Figure 18.4 highlights the PH assumption. The relationship between the two curves must hold for any two values c and d of X_1. The shape of the function for a given value of X_1 must conform to the assumed $\lambda(t)$. For a Weibull model, the functions should each be linear in $\log t$.

When there are multiple predictors, the PH assumption can be displayed in a way similar to Figures 18.2 and 18.4 but with the population additionally cross-classified by levels of the other predictors besides X_1. If there is one binary predictor X_1 and one continuous predictor X_2, the relationship in

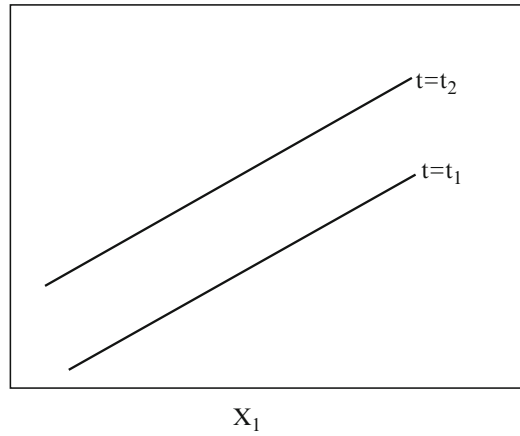

Fig. 18.3 PH model with one continuous predictor. Y-axis is $\log \lambda(t)$ or $\log \Lambda(t)$; for $\log \Lambda(t)$, drawn for $t_2 > t_1$. The slope of each line is β_1.

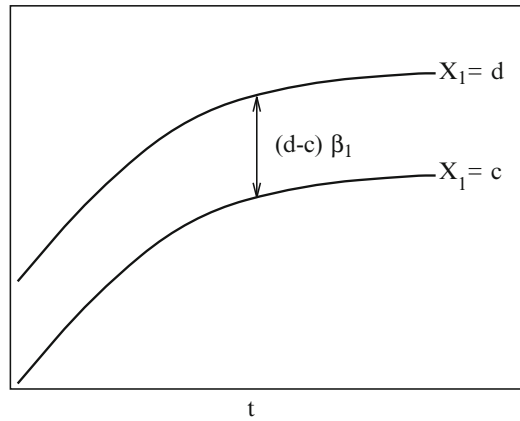

Fig. 18.4 PH model with one continuous predictor. Y-axis is $\log \lambda(t)$ or $\log \Lambda(t)$. For $\log \lambda$, the functions need not be monotonic.

Figure 18.5 must hold at each time t if linearity is assumed for X_2 and there is no interaction between X_1 and X_2. Methods for verifying the regression assumptions (e.g., splines and residuals) and the PH assumption are covered in detail under the Cox PH model in Chapter 20.

The method for verifying the assumed shape of $S(t)$ in Section 18.1.3 is also useful when there are a limited number of categorical predictors. To validate a Weibull PH model one can stratify on X and plot $\log \Lambda_{\mathrm{KM}}(t|X \text{ stratum})$ against $\log t$. This graph simultaneously assesses PH in addition to shape assumptions—all curves should be parallel as well as straight. Straight but nonparallel (non-PH) curves indicate that a series of Weibull models with differing γ parameters will fit.

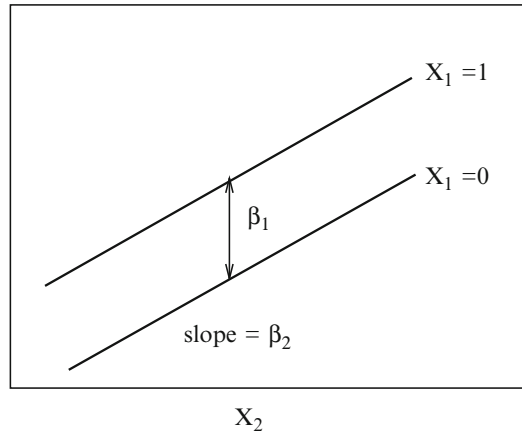

Fig. 18.5 Regression assumptions, linear additive PH or AFT model with two predictors. For PH, Y-axis is $\log \lambda(t)$ or $\log \Lambda(t)$ for a fixed t. For AFT, Y-axis is $\log(T)$.

18.3 Accelerated Failure Time Models

18.3.1 Model

Besides modeling the effect of predictors by a multiplicative effect on the hazard function, other regression effects can be specified. The *accelerated failure time (AFT) model* is commonly used; it specifies that the predictors act multiplicatively on the failure time or additively on the log failure time. The effect of a predictor is to alter the rate at which a subject proceeds along the time axis (i.e., to accelerate the time to failure [331, pp. 33–35]). The model is

2

$$S(t|X) = \psi((\log(t) - X\beta)/\sigma), \tag{18.32}$$

where ψ is any standardized survival distribution function. The parameter σ is called the *scale parameter*. The model can also be stated as $(\log(T) - X\beta)/\sigma \sim \psi$ or $\log(T) = X\beta + \sigma\epsilon$, where ϵ is a random variable from the distribution ψ. Sometimes the untransformed T is used in place of $\log(T)$. When the log form is used, the models are said to be log-normal, log-logistic, and so on.

The exponential and Weibull are the only two distributions that can describe either a PH or an AFT model.

3

18.3.2 Model Assumptions and Interpretation of Parameters

The $\log \lambda$ or $\log \Lambda$ transformation of the PH model has the following equivalent for AFT models.

$$\psi^{-1}[S(t|X)] = (\log(t) - X\beta)/\sigma. \tag{18.33}$$

Letting as before ϵ denote a random variable from the distribution S, the model is also

$$\log(T) = X\beta + \sigma\epsilon. \tag{18.34}$$

So the property of the response T of interest for regression modeling is $\log(T)$. In the absence of censoring, we could check the model by plotting an X against $\log T$ and checking that the residuals $\log(T) - X\hat{\beta}$ are distributed as ψ to within a scale factor.

The assumptions of the AFT model are thus the following.

1. The true form of ψ (the distributional family) is correctly specified.
2. In the absence of nonlinear and interaction terms, each X_j affects $\log(T)$ or $\psi^{-1}[S(t|X)]$ linearly.
3. Implicit in these assumptions is that σ is a constant independent of X.

A one-unit change in X_j is then most simply understood as a β_j change in the log of the failure time. The one-unit change in X_j increases the failure time by a factor of $\exp(\beta_j)$.

The median survival time is obtained by solving $\psi((\log(t) - X\beta)/\sigma) = 0.5$ giving

$$T_{0.5}|X = \exp[X\beta + \sigma\psi^{-1}(0.5)] \tag{18.35}$$

18.3.3 Specific Models

Common choices for the distribution function ψ in Equation 18.32 are the extreme value distribution $\psi(u) = \exp(-\exp(u))$, the logistic distribution $\psi(u) = [1 + \exp(u)]^{-1}$, and the normal distribution $\psi(u) = 1 - \Phi(u)$. The AFT model equivalent of the Weibull model is obtained by using the extreme value distribution, negating β, and replacing γ with $1/\sigma$ in Equation 18.24:

$$\begin{aligned} S(t|X) &= \exp[-\exp((\log(t) - X\beta)/\sigma)] \\ T_{0.5}|X &= [\log(2)]^{\sigma}\exp(X\beta). \end{aligned} \tag{18.36}$$

The exponential model is obtained by restricting $\sigma = 1$ in the extreme value distribution.

The log-normal regression model is

$$S(t|X) = 1 - \Phi((\log(t) - X\beta)/\sigma), \tag{18.37}$$

and the log-logistic model is

$$S(t|X) = [1 + \exp((\log(t) - X\beta)/\sigma)]^{-1}. \tag{18.38}$$

The t distribution allows for more flexibility by varying the degrees of freedom. Figure 18.6 depicts possible hazard functions for the log t distribution for varying σ and degrees of freedom. However, this distribution does not have a late increasing hazard phase typical of human survival.

```
require(rms)
```

```
haz ← survreg.auxinfo$t$hazard
times ← c(seq(0, .25, length=100), seq(.26, 2, length=150))
high ← c(6, 1.5, 1.5, 1.75)
low  ← c(0, 0, 0, .25)
dfs  ← c(1, 2, 3, 5, 7, 15, 500)
cols ← rep(1, 7)
ltys ← 1:7
i ← 0
for(scale in c(.25, .6, 1, 2)) {
  i ← i + 1
  plot(0, 0, xlim=c(0,2), ylim=c(low[i], high[i]),
    xlab=expression(t), ylab=expression(lambda(t)), type="n")
  col ← 1.09
  j ← 0
  for(df in dfs) {
    j ← j+1
    ## Divide by t to get hazard for log t distribution
    lines(times,
          haz(log(times), 0, c(log(scale), df))/times,
          col=cols[j], lty=ltys[j])
    if(i==1) text(1.7, .23 + haz(log(1.7), 0,
         c(log(scale),df))/1.7, format(df))
  }
  title(paste("Scale:", format(scale)))
}   # Figure 18.6
```

All three of these parametric survival models have median survival time $T_{0.5}|X = \exp(X\beta)$.

18.3.4 Estimation

Maximum likelihood estimation is used much the same as in Section 18.2.5. Care must be taken in the choice of initial values; iterative methods are especially prone to problems in choosing the initial $\hat{\sigma}$. Estimation works better if σ is parameterized as $\exp(\delta)$. Once β and σ ($\exp(\delta)$) are estimated, MLEs of secondary parameters such as survival probabilities and medians can readily be obtained:

$$\begin{aligned}\hat{S}(t|X) &= \psi((\log(t) - X\hat{\beta})/\hat{\sigma}) \\ \hat{T}_{0.5}|X &= \exp[X\hat{\beta} + \hat{\sigma}\psi^{-1}(0.5)].\end{aligned} \tag{18.39}$$

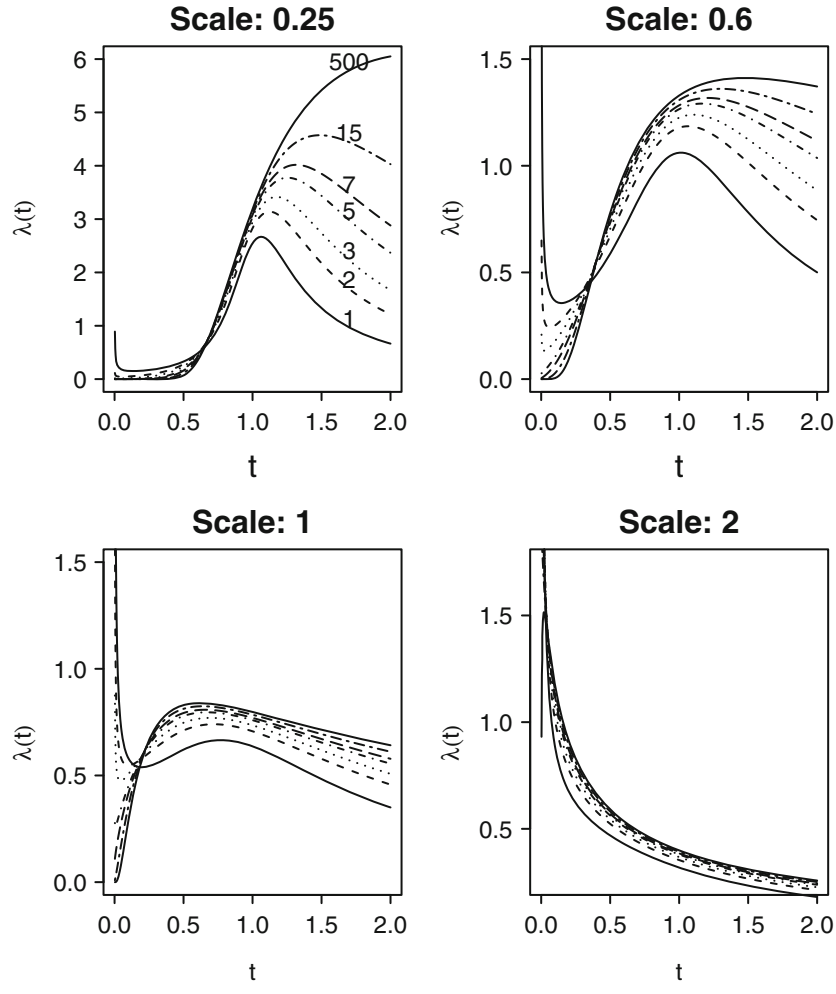

Fig. 18.6 $\log(T)$ distribution for $\sigma = 0.25, 0.6, 1, 2$ and for degrees of freedom $1, 2, 3, 5, 7, 15, 500$ (almost log-normal). The top left plot has degrees of freedom written in the plot.

For normal and logistic distributions, $\hat{T}_{0.5}|X = \exp(X\hat{\beta})$. The MLE of the effect on $\log(T)$ of increasing X_j by d units is $\hat{\beta}_j d$ if X_j is linear and additive.

The delta (statistical differential) method can be used to compute an estimate of the variance of $f = [\log(t) - X\hat{\beta}]/\hat{\sigma}$. Let $(\hat{\beta}, \hat{\delta})$ denote the estimated parameters, and let $\hat{V}$ denote the estimated covariance matrix for these parameter estimates. Let F denote the vector of derivatives of f with respect to $(\beta_0, \beta_1, \ldots, \beta_p, \delta)$; that is, $F = [-1, -X_1, -X_2, \ldots, -X_p, -(\log(t) - X\hat{\beta})]/\hat{\sigma}$. The variance of f is then approximately

$$\text{Var}(f) = F\hat{V}F'. \tag{18.40}$$

Letting s be the square root of the variance estimate and $z_{1-\alpha/2}$ be the normal critical value, a $1-\alpha$ confidence limit for $S(t|X)$ is

$$\psi((\log(t) - X\hat{\beta})/\hat{\sigma} \pm z_{1-\alpha/2} \times s). \tag{18.41}$$

18.3.5 Residuals

For an AFT model, standardized residuals are simply

$$r = (\log(T) - X\hat{\beta})/\sigma. \tag{18.42}$$

4 When T is right-censored, r is right-censored. Censoring must be taken into account, for example, by displaying Kaplan–Meier estimates based on groups of residuals rather than showing individual residuals. The residuals can be used to check for lack of fit as described in the next section. Note that examining individual uncensored residuals is not appropriate, as their distribution is conditional on $T_i < C_i$, where C_i is the censoring time.

Cox and Snell[134] proposed a type of general residuals that also work for censored data. Using their method on the cumulative probability scale results in the probability integral transformation. If the probability of failure before time t given X is $S(t|X)$, $F(T|X) = 1 - S(T|X)$ has a uniform $[0, 1]$ distribution, where T is a subject's actual failure time. When T is right-censored, so is $1 - S(T|X)$. Substituting $\hat{S}$ for S results in an approximate uniform $[0, 1]$ distribution for any value of X. One minus the Kaplan–Meier estimate of $1 - \hat{S}(T|X)$ (using combined data for all X) is compared against a 45° line to check for goodness of fit. A more stringent assessment is obtained by repeating this process while stratifying on X.

18.3.6 Assessment of Model Fit

For a single binary predictor, all assumptions of the AFT model are depicted in Figure 18.7. That figure also shows the assumptions for any two values of a single continuous predictor that behaves linearly. For a single continuous predictor, the relationships in Figure 18.8 must hold for any two follow-up times. The regression assumptions are isolated in Figure 18.5.

To verify the fit of a log-logistic model with age as the only predictor, one could stratify by quartiles of age and check for linearity and parallelism of the four logit $S_\Lambda(t)$ or $S_{\text{KM}}(t)$ curves over increasing t as in Figure 18.7, which stresses the distributional assumption (no T by X interaction and linearity vs. $\log(t)$). To stress the linear regression assumption while checking for absence of time interactions (part of the distributional assumptions), one could make

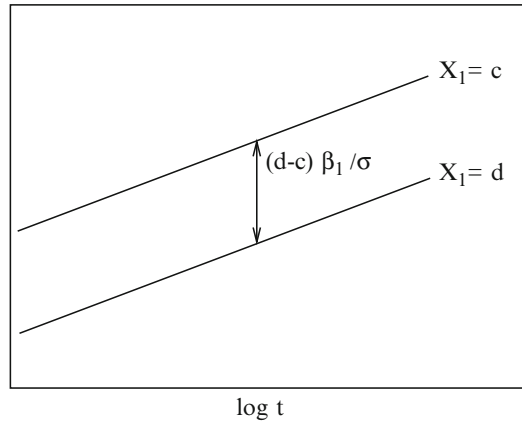

Fig. 18.7 AFT model with one predictor. Y-axis is $\psi^{-1}[S(t|X)] = (\log(t) - X\beta)/\sigma$. Drawn for $d > c$. The slope of the lines is σ^{-1}.

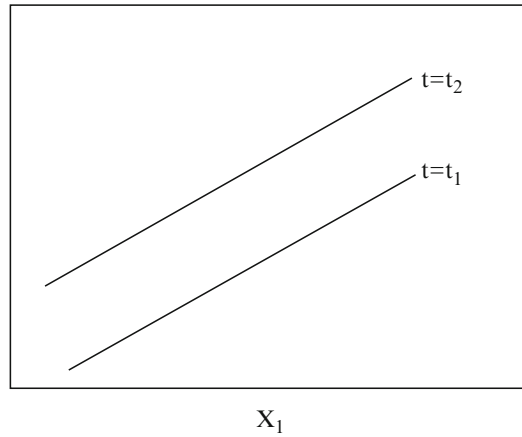

Fig. 18.8 AFT model with one continuous predictor. Y-axis is $\psi^{-1}[S(t|X)] = (\log(t) - X\beta)/\sigma$. Drawn for $t_2 > t_1$. The slope of each line is β_1/σ and the difference between the lines is $\log(t_2/t_1)/\sigma$.

a plot like Figure 18.8. For each decile of age, the logit transformation of the 1-, 3-, and 5-year survival estimates for that decile would be plotted against the mean age in the decile. This checks for linearity and constancy of the age effect over time. Regression splines will be a more effective method for checking linearity and determining transformations. This is demonstrated in Chapter 20 with the Cox model, but identical methods apply here.

As an example, consider data from Kalbfleisch and Prentice [331, pp. 1–2], who present data from Pike[508] on the time from exposure to the carcinogen DMBA to mortality from vaginal cancer in rats. The rats are divided into two groups on the basis of a pre-treatment regime. Survival times in days (with censored times marked $^+$) are found in Table 18.2.

Table 18.2 Rat vaginal cancer data from Pike[508]

Group 1	143	164	188	188	190	192	206	209	213	216
	220	227	230	234	246	265	304	216^+	244^+	
Group 2	142	156	163	198	205	232	232	233	233	233
	233	239	240	261	280	280	296	296	323	204^+
	344^+									

```
getHdata(kprats)
kprats$group ← factor(kprats$group, 0:1, c('Group 1', 'Group 2'))
dd ← datadist(kprats); options(datadist="dd")

S ← with(kprats, Surv(t, death))
f ← npsurv(S ~ group, type="fleming", data=kprats)
survplot(f, n.risk=TRUE, conf='none',   # Figure 18.9
         label.curves=list(keys='lines'), levels.only=TRUE)
title(sub="Nonparametric estimates", adj=0, cex=.7)

# Check fits of Weibull, log-logistic, log-normal
xl ← c(4.8, 5.9)
survplot(f, loglog=TRUE, logt=TRUE, conf="none", xlim=xl,
         label.curves=list(keys='lines'), levels.only=TRUE)
title(sub="Weibull (extreme value)", adj=0, cex=.7)
survplot(f, fun=function(y)log(y/(1-y)), ylab="logit S(t)",
         logt=TRUE, conf="none", xlim=xl,
         label.curves=list(keys='lines'), levels.only=TRUE)
title(sub="Log-logistic", adj=0, cex=.7)
survplot(f, fun=qnorm, ylab="Inverse Normal S(t)",
         logt=TRUE, conf="none",
         xlim=xl,cex.label=.7,
         label.curves=list(keys='lines'), levels.only=TRUE)
title(sub="Log-normal", adj=0, cex=.7)
```

The top left plot in Figure 18.9 displays nonparametric survival estimates for the two groups, with the number of rats "at risk" at each 30-day mark written above the x-axis. The remaining three plots are for checking assumptions of three models. None of the parametric models presented will completely allow for such a long period with no deaths. Neither will any allow for the early crossing of survival curves. Log-normal and log-logistic models yield very similar results due to the similarity in shapes between $\Phi(z)$ and $[1+\exp(-z)]^{-1}$ for non-extreme z. All three transformations show good parallelism after the early crossing. The log-logistic and log-normal transformations are slightly more linear. The fitted models are:

```
fw ← psm(S ~ group, data=kprats, dist='weibull')
fl ← psm(S ~ group, data=kprats, dist='loglogistic',
         y=TRUE)
fn ← psm(S ~ group, data=kprats, dist='lognormal')
latex(fw, fi='')
```

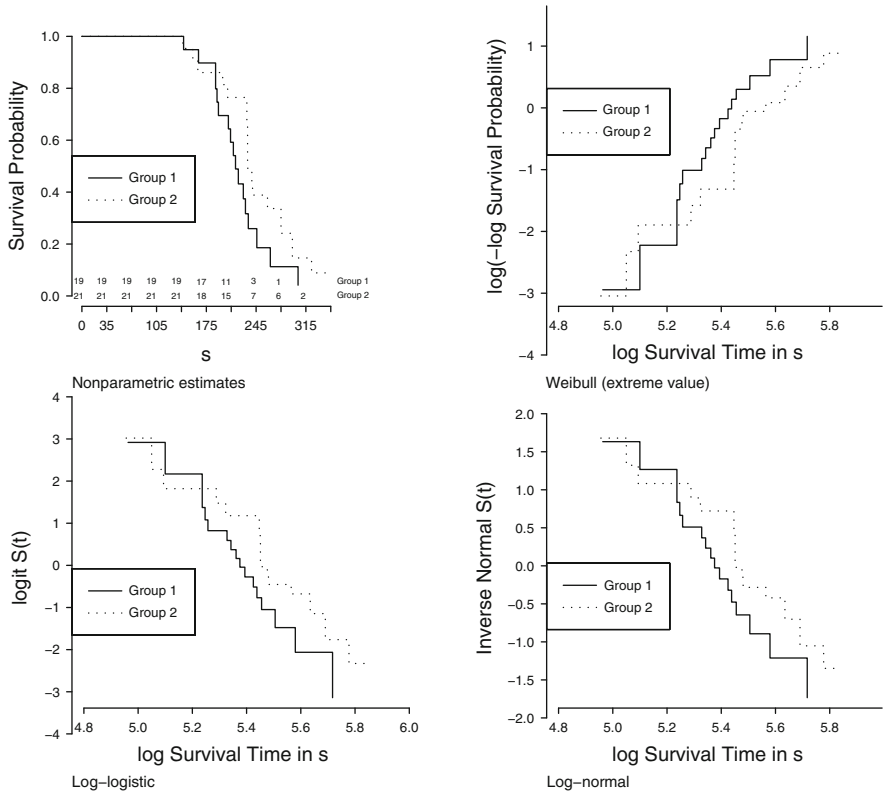

Fig. 18.9 Altschuler–Nelson–Fleming–Harrington nonparametric survival estimates for rats treated with DMBA,[508] along with various transformations of the estimates for checking distributional assumptions of three parametric survival models.

$$\text{Prob}\{T \geq t\} = \exp[-\exp(\frac{\log(t) - X\beta}{0.1832976})] \quad \text{where}$$

$$X\hat{\beta} =$$
$$5.450859$$
$$+0.131983[\text{Group 2}]$$

and $[c] = 1$ if subject is in group c, 0 otherwise.

```
latex(fl, fi='')
```

Table 18.3 Group effects from three survival models

Model	Group 2:1 Failure Time Ratio	Median Survival Time	
		Group 1	Group 2
Extreme Value (Weibull)	1.14	217	248
Log-logistic	1.11	217	241
Log-normal	1.10	217	238

$$\text{Prob}\{T \geq t\} = [1 + \exp(\frac{\log(t) - X\beta}{0.1159753})]^{-1} \quad \text{where}$$

$$X\hat{\beta} = 5.375675 + 0.1051005[\text{Group 2}]$$

and $[c] = 1$ if subject is in group c, 0 otherwise.

```
latex(fn, fi='')
```

$$\text{Prob}\{T \geq t\} = 1 - \Phi(\frac{\log(t) - X\beta}{0.2100184}) \quad \text{where}$$

$$X\hat{\beta} = 5.375328 + 0.0930606[\text{Group 2}]$$

and $[c] = 1$ if subject is in group c, 0 otherwise.

The estimated failure time ratios and median failure times for the two groups are given in Table 18.3. For example, the effect of going from Group 1 to Group 2 is to increase log failure time by 0.132 for the extreme value model, giving a Group 2:1 failure time ratio of $\exp(0.132) = 1.14$. This ratio is also the ratio of median survival times. We choose the log-logistic model for its simpler form. The fitted survival curves are plotted with the nonparametric estimates in Figure 18.10. Excellent agreement is seen, except for 150 to 180 days for Group 2. The standard error of the regression coefficient for group in the log-logistic model is 0.0636 giving a Wald χ^2 for group differences of $(.105/.0636)^2 = 2.73, P = 0.1$.

```
survplot(f, conf.int=FALSE,    # Figure 18.10
         levels.only=TRUE, label.curves=list(keys='lines'))
survplot(fl, add=TRUE, label.curves=FALSE, conf.int=FALSE)
```

The Weibull PH form of the fitted extreme value model, using Equation 18.24, is

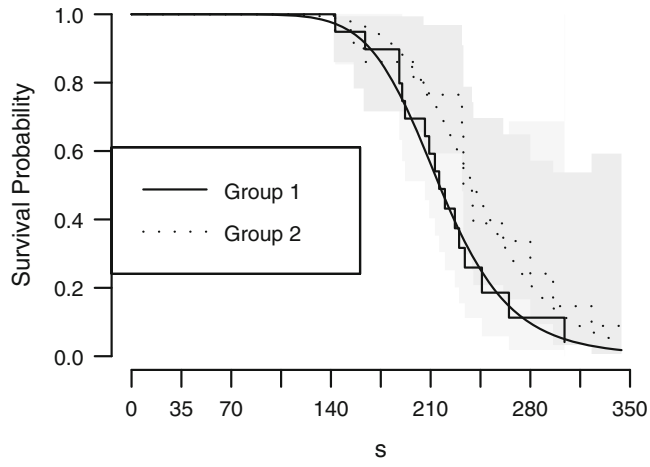

Fig. 18.10 Agreement between fitted log-logistic model and nonparametric survival estimates for rat vaginal cancer data.

$$\text{Prob}\{T \geq t\} = \exp\{-t^{5.456}\exp(X\hat{\beta})\} \quad \text{where}$$

$$X\hat{\beta} = \\ -29.74 \\ -0.72[\text{Group 2}]$$

and $[c] = 1$ if subject is in group c, 0 otherwise.

A sensitive graphical verification of the distributional assumptions of the AFT model is obtained by plotting the estimated survival distribution of standardized residuals (Equation 18.3.5), censored identically to the way T is censored. This distribution is plotted along with the theoretical distribution ψ. The assessment may be made more stringent by stratifying the residuals by important subject characteristics and plotting separate survival function estimates; they should all have the same standardized distribution (e.g., same σ).

```
r ← resid(fl, 'cens')
survplot(npsurv(r ~ group, data=kprats),
         conf='none', xlab='Residual',
         label.curves=list(keys='lines'), levels.only=TRUE)
survplot(npsurv(r ~ 1), conf='none', add=TRUE, col='red')
lines(r, lwd=1, col='blue')    # Figure 18.11
```

As an example, Figure 18.11 shows the Kaplan–Meier estimate of the distribution of residuals, Kaplan–Meier estimates stratified by group, and the assumed log-logistic distribution.

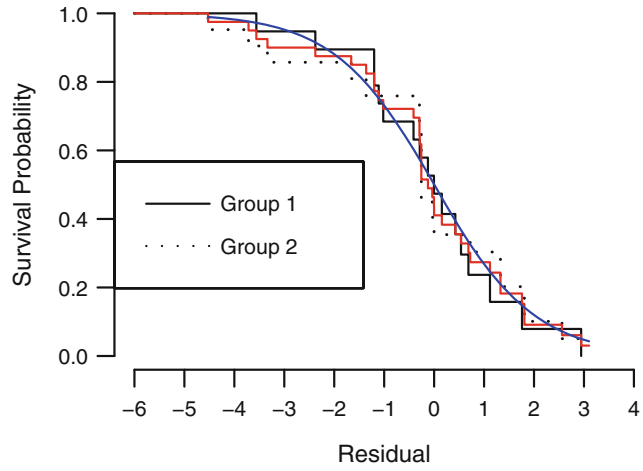

Fig. 18.11 Kaplan–Meier estimates of distribution of standardized censored residuals from the log-logistic model, along with the assumed standard log-logistic distribution (dashed curve). The step functions in red is the estimated distribution of all residuals, and the step functions in black are the estimated distributions of residuals stratified by group, as indicated. The blue curve is the assumed log-logistic distribution.

Section 19.2 has a more in-depth example of this approach.

18.3.7 Validating the Fitted Model

AFT models may be validated for both calibration and discrimination accuracy using the same methods that are presented for the Cox model in Section 20.11. The methods discussed there for checking calibration are based on choosing a single follow-up time. Checking the distributional assumptions of the parametric model is also a check of calibration accuracy in a sense. Another indirect calibration assessment may be obtained from a set of Cox–Snell residuals (Section 18.3.5) or by using ordinary residuals as just described. A higher resolution indirect calibration assessment based on plotting individual uncensored failure times is available when the theoretical censoring times for those observations are known. Let C denote a subject's censoring time and F the cumulative distribution of a failure time T. The expected value of $F(T|X)$ is 0.5 when T is an actual failure time random variable. The expected value for an event time that is observed *because it is uncensored* is the expected value of $F(T|T \leq C, X) = 0.5F(C|X)$. A smooth plot (using, say, `loess`) of $F(T|X) - 0.5F(C|X)$ against $X\hat{\beta}$ should be a flat line through $y = 0$ if the model is well calibrated. A smooth plot of $2F(T|X)/F(C|X)$ against $X\hat{\beta}$ (or anything else) should be a flat line through $y = 1$. This method assumes that the model is calibrated well enough that we can substitute $1 - \hat{S}(C|X)$ for $F(C|X)$.

18.4 Buckley–James Regression Model

Buckley and James[81] developed a method for estimating regression coefficients using least squares after imputing censored residuals. Their method does not assume a distribution for survival time or the residuals, but is aimed at estimating expected survival time or expected log survival time given predictor variables. This method has been generalized to allow for smooth nonlinear effects and interactions in the S `bj` function in the `rms` package, written by Stare and Harrell[585].

18.5 Design Formulations

Various designs can be formulated with survival regression models just as with other regression models. By constructing the proper dummy variables, ANOVA and ANOCOVA models can easily be specified for testing differences in survival time between multiple treatments. Interactions and complex nonlinear effects may also be modeled.

18.6 Test Statistics

As discussed previously, likelihood ratio, score, and Wald statistics can be derived from the maximum likelihood analysis, and the choice of test statistic depends on the circumstance and on computational convenience.

18.7 Quantifying Predictive Ability

See Section 20.10 for a generalized measure of concordance between predicted and observed survival time (or probability of survival) for right-censored data.

18.8 Time-Dependent Covariates

Time-dependent covariates (predictors) requires special likelihood functions and add significant complexity to analyses in exchange for greater versatility and enhanced predictive discrimination[604]. Nicolaie et al.[477] and D'Agostino et al.[145] provide useful static covariate approaches to modeling time-dependent predictors using landmark analysis.

18.9 R Functions

Therneau's `survreg` function (part of his `survival` package) can fit regression models in the AFT family with left–, right–, or interval–censoring. The time variable can be untransformed or log-transformed (the default). Distributions supported are extreme value (Weibull and exponential), normal, logistic, and Student-t. The version of `survreg` in `rms` that fits parametric survival models in the same framework as `lrm`, `ols`, and `cph` is called `psm`. `psm` works with `print`, `coef`, `formula`, `specs`, `summary`, `anova`, `predict`, `Predict`, `fastbw`, `latex`, `nomogram`, `validate`, `calibrate`, `survest`, and `survplot` functions for obtaining and plotting predicted survival probabilities. The `dist` argument to `psm` can be `"exponential"`, `"extreme"`, `"gaussian"`, `"logistic"`, `"loglogistic"`, `"lognormal"`, `"t"`, or `"weibull"`. To fit a model with no covariables, use the command

```
psm(Surv(d.time, event) ~ 1)
```

To restate a Weibull or exponential model in PH form, use the `pphsm` function. An example of how many of the functions are used is found below.

```
units(d.time) ← "Year"
f ← psm(Surv(d.time,cdeath) ~ lsp(age,65)*sex)
# default is Weibull
anova(f)
summary(f)          # summarize effects with delta log T
latex(f)            # typeset math. form of fitted model
survest(f, times=1) # 1y survival est. for all subjects
survest(f, expand.grid(sex="female", age=30:80), times=1:2)
# 1y, 2y survival estimates vs. age, for females
survest(f, data.frame(sex="female",age=50))
# survival curve for an individual subject
survplot(f, sex=NA, age=50, n.risk=T)
# survival curves for each sex, adjusting age to 50
f.ph ← pphsm(f)     # convert from AFT to PH
summary(f.ph)       # summarize with hazard ratios
                    # instead of changes in log(T)
```

Special functions work with objects created by `psm` to create S functions that contain the analytic form for predicted survival probabilities (`Survival`), hazard functions (`Hazard`), quantiles of survival time (`Quantile`), and mean or expected survival time (`Mean`). Once the S functions are constructed, they can be used in a variety of contexts. The `survplot` and `survest` functions have a special argument for `psm` fits: `what`. The default is `what="survival"` to estimate or plot survival probabilities. Specifying `what="hazard"` will plot hazard functions. `Predict` also has a special argument for `psm` fits: `time`. Specifying a single value for `time` results in survival probability for that time being plotted instead of $X\hat{\beta}$. Examples of many of the functions appear below, with the output of the `survplot` command shown in Figure 18.12.

```
med   ← Quantile(fl)
meant ← Mean(fl)
```

```
haz   ← Hazard(fl)
surv  ← Survival(fl)
latex(surv, file='', type='Sinput')
```

```
surv ← function (times = NULL, lp = NULL,
                 parms = -2.15437773933124)
{
    1/(1 + exp((logb(times) - lp)/exp(parms)))
}
```

```
# Plot estimated hazard functions and add median
# survival times to graph
survplot(fl, group, what="hazard")   # Figure 18.12
# Compute median survival time
m ← med(lp=predict(fl,
          data.frame(group=levels(kprats$group))))
m
```

```
      1       2
216.0857 240.0328
```

```
med(lp=range(fl$linear.predictors))
```

```
[1] 216.0857 240.0328
```

```
m ← format(m, digits=3)
text(68, .02, paste("Group 1 median: ", m[1],"\n",
                    "Group 2 median: ", m[2], sep=""))
# Compute survival probability at 210 days
xbeta ← predict(fl,
                data.frame(group=c("Group 1","Group 2")))
surv(210, xbeta)
```

```
       1        2
0.5612718 0.7599776
```

The S object called `survreg.distributions` in Therneau's `survival` package and the object `survreg.auxinfo` in the `rms` package have detailed information for extreme-value, logistic, normal, and t distributions. For each distribution, components include the deviance function, an algorithm for obtaining starting parameter estimates, a LaTeX representation of the survival function, and S functions defining the survival, hazard, quantile functions, and basic survival inverse function (which could have been used in Figure 18.9). See Figure 18.6 for examples. `rms`'s `val.surv` function is useful for indirect external validation of parametric models using Cox–Snell residuals and other approaches of Section 18.3.7. The `plot` method for an object created by `val.surv` makes it easy to stratify all computations by a variable of interest to more stringently validate the fit with respect to that variable.

`rms`'s `bj` function fits the Buckley–James model for right-censored responses.

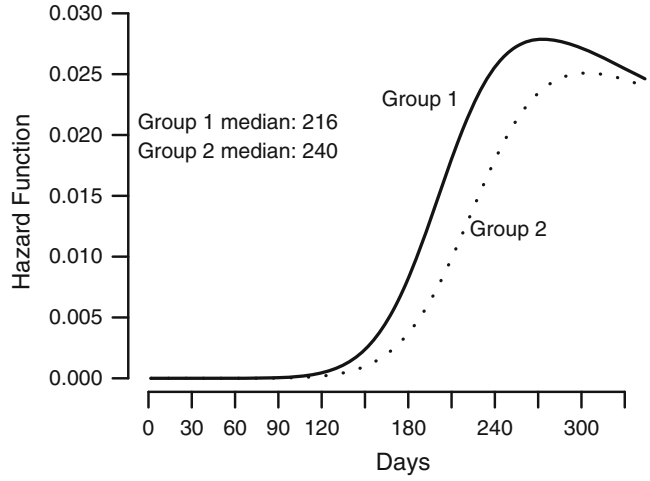

Fig. 18.12 Estimated hazard functions for log-logistic fit to rat vaginal cancer data, along with median survival times.

Kooperberg et al.'s adaptive linear spline log-hazard model[360,361,594] has been implemented in the S function `hare`. Their procedure searches for second-order interactions involving predictors (and linear splines of them) and linear splines in follow-up time (allowing for non-proportional hazards). `hare` is also used to estimate calibration curves for parametric survival models (`rms` function `calibrate`) as it is for Cox models.

18.10 Further Reading

1. Wellek[657] developed a test statistic for a specified maximum survival difference after relating this difference to a hazard ratio.
2. Hougaard[308] compared accelerated failure time models with proportional hazard models.
3. Gore et al.[226] discuss how an AFT model (the log-logistic model) gives rise to varying hazard ratios.
4. See Hillis[293] for other types of residuals and plots that use them.
5. See Gore et al.[226] and Lawless[382] for other methods of checking assumptions for AFT models. Lawless is an excellent text for in-depth discussion of parametric survival modeling. Kwong and Hutton[369] present other methods of choosing parametric survival models, and discuss the robustness of estimates when fitting an incorrectly chosen accelerated failure time model.

18.11 Problems

1. For the failure times (in days)

$$1 \quad 3 \quad 3^+ \quad 6^+ \quad 7^+$$

compute MLEs of the following parameters of an exponential distribution by hand: λ, μ, $T_{0.5}$, and $S(3 \text{ days})$. Compute 0.95 confidence limits for λ and $S(3)$, basing the latter on $\log[\Lambda(t)]$.

2. For the same data in Problem 1, compute MLEs of parameters of a Weibull distribution. Also compute the MLEs of $S(3)$ and $T_{0.5}$.