21. Proportional Hazards Regression Models

Proportional

The proportional hazards model plays an important role in analyzing data with survival outcomes. This chapter provides a summary of different aspects of this very popular model.

The first part gives the definition of the model and shows how to estimate the regression parameters for survival data with or without ties. Hypothesis testing can be built based on these estimates. Formulas to estimate the cumulative hazard function and the survival function are also provided. Modified models for stratified data and data with time-dependent covariates are also discussed.

The second part of the chapter talks about goodness-of-fit and model checking techniques. These include testing for proportionality assumptions, testing for function forms for a particular covariate and testing for overall fitting.

The third part of the chapter extends the model to accommodate more complicated data structures. Several extended models such as models with random effects, nonproportional models, and models for data with multivariate survival outcomes are introduced.

In the last part a real example is given. This serves as an illustration of the implementation of the methods and procedures discussed in this chapter.

The proportional hazards model has played a pivotal role in survival analysis since it was proposed by *Cox* [21.1]. This model has been widely used in many areas, such as biomedical research and engineering, for assessing covariate effects on the time to some events in the presence of right censoring. For example, when testing the reliability of an electrical instrument, the model can be used to investigate the effects of variables such as humidity, temperature, and voltage on the time to breakdown. Since time constraint might not allow us to observe the failure of every experimental unit, for some units we only know that failure did not occur up to the end of study, which is the censoring event.

Let T be the failure time, C be the censoring time, and $\boldsymbol{Z} = \{\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_p\}^T$ be a p-dimensional vector of covariates. Throughout this chapter the covariate vector $\boldsymbol{Z}$ is assumed to be time-independent, although it is straightforward to extend the theory to time-varying covariates. The failure time T might not always be observed due to censoring, and what we actually observe are $X = \min(T, C)$, the smaller of the failure time and the censoring time, and $\Delta = I\,(T \le C)$, the indicator that failure has been observed. The dataset obtained from a failure-time study consists of n independent realizations of the triplet $(X, \boldsymbol{Z}, \Delta)$. It is usually assumed that the censoring is noninformative in that, given $\boldsymbol{Z}$, the failure and the censoring times are independent. Let $P\,(T > t \mid \boldsymbol{Z})$ be the conditional survival function, and the conditional hazard function is defined as

$$\lambda\,(t \mid \boldsymbol{Z}) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P\,(t \le T < t + \Delta t \mid T \ge t, \boldsymbol{Z}) \ ,$$

which is the instantaneous rate of failure at time t, given that failure has not occurred before t and the covariate vector $\boldsymbol{Z}$.

There are many ways to model the relationship between the failure time and the covariates. The proportional hazards model specifies

$$\lambda\,(t \mid \boldsymbol{Z}) = \lambda_0(t) \exp\left(\beta^T \boldsymbol{Z}\right) \ , \tag{21.1}$$

where $\lambda_0(t)$ is an unknown baseline hazard function corresponding to $\boldsymbol{Z} = (0, \cdots, 0)$, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is the vector of regression coefficients.

This method does not assume a parametric distribution for the failure times, but rather assumes that the effects of the different variables on the time to failure are constant over time and are multiplicative on the hazard. The model is called the proportional hazards model since the ratio of hazards of any two experimental units is always a constant:

$$\frac{\lambda\,(t \mid z)}{\lambda\,(t \mid z')} = \frac{\lambda_0(t) \exp\left(\beta^T z\right)}{\lambda_0(t) \exp\left(\beta^T z'\right)} = \exp\left[\beta^T (z - z')\right] \ ,$$

where z and z' are the respective covariate values of the two units. This quantity is often referred to as the hazard ratio or relative risk.

The interpretation of the parameter β is similar to that in other regression models. For example, $\exp(\beta_1)$ is the hazard ratio of two study units whose values of the first covariate differ by 1 and whose values of any other covariate are the same. Usually, the goal is to make inferences about β or a subset of β to see whether a certain covariate has an effect on the survival rate or not. The baseline hazard $\lambda_0(\cdot)$ is treated as a nuisance parameter function. The proportional hazards model is considered a semiparametric model, in the sense that $\lambda_0(\cdot)$ is an infinite-dimensional parameter.

The semiparametric proportional hazards model includes the parametric Weibull model as a special case. To see this, for the Weibull distribution with density $f(t) = \alpha \lambda t^{\alpha-1} \exp(-\lambda t^{\alpha})$ and survival function $S(t) = \exp(-\lambda t^{\alpha})$, parameterize the parameter λ as $\lambda = \lambda' \exp(\beta^T \boldsymbol{Z})$, then the hazard of failure given $\boldsymbol{Z}$ is

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t) \exp\left(\beta^T \boldsymbol{Z}\right) \ ,$$

where $\lambda_0(t) = \alpha \lambda' t^{\alpha-1}$ is a function with two parameters, instead of the unspecified $\lambda_0(\cdot)$ in the case of the proportional hazards model. It can also be shown that the Weibull model is also a special case of the semiparametric accelerated failure-time model. In fact, the Weibull model is the most general parametric model that has both the proportional hazards and the accelerated failure-time properties. See Chapt. 12 of *Klein* and *Moeschberger* [21.2] for a detailed discussion.

21.1 Estimating the Regression Coefficients β

The partial likelihood method was introduced by *Cox* [21.3] to estimate the regression parameters β in the proportional hazards model for failure times with possible right censoring. We will first focus on the case when all failure times are distinct. When the failure time follows a continuous distribution, it is very unlikely that two subjects would fail at the same time. In reality, however, the measured time always has a discrete distribution, since it can only take values in a finite set of numbers. Thus tied failure times could happen in a real study, and special attention is needed in this situation.

21.1.1 Partial Likelihood for Data with Distinct Failure Times

Now suppose there is no tie among the failure times. Let $t_1 < \cdots < t_N$ denote the N ordered times of observed failures and let (j) denote the label of the individual that fails at t_j. Let $\mathcal{R}_j$ be the risk set at time t_j, i. e. $\mathcal{R}_j = \{i : X_i \ge t_j\}$.

The partial likelihood for the model (21.1) is defined as

$$\prod_{j=1}^{N} \frac{\exp\left(\beta^T Z_{(j)}\right)}{\sum_{i \in \mathcal{R}_j} \exp\left(\beta^T Z_i\right)} \ , \tag{21.2}$$

and the log partial likelihood is then

$$L(\beta) = \sum_{j=1}^{N} \left\{ \beta^T Z_{(j)} - \log \left[\sum_{i \in \mathcal{R}_j} \exp\left(\beta^T Z_i\right) \right] \right\} .$$

The maximum partial likelihood estimate of β, $\hat{\beta}$, as proposed by *Cox* [21.3], is found by solving the score equation

$$U(\beta) = 0 ,$$

where $U(\beta) = \partial L(\beta)/\partial\beta$.

The information matrix, defined as the negative of the second derivative matrix of the log likelihood, is given by

$$I(\beta) = -\frac{\partial U(\beta)}{\partial \beta} = \sum_{j=1}^{N} \left[\frac{\sum_{i \in \mathcal{R}_j} \exp\left(\beta^T Z_i\right) Z_i^{\otimes 2}}{\sum_{i \in \mathcal{R}_j} \exp\left(\beta^T Z_i\right)} - \left(\frac{\sum_{i \in \mathcal{R}_j} \exp\left(\beta^T Z_i\right) Z_i}{\sum_{i \in \mathcal{R}_j} \exp\left(\beta^T Z_i\right)} \right)^{\otimes 2} \right] ,$$

where $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$ for any vector $\boldsymbol{a}$.

It can be shown that $\hat{\beta}$ is a consistent estimator for β, and $nI^{-1}(\hat{\beta})$ is a consistent estimator for the covariance matrix of $n^{1/2}(\hat{\beta} - \beta)$, where n is the number of all subjects, censored or uncensored. Thus for large samples, $\hat{\beta}$ has an approximately normal distribution with mean β and variance–covariance matrix $I^{-1}(\hat{\beta})$.

21.1.2 Partial Likelihood for Data with Tied Failure Times

In the previous section, we defined the partial likelihood for data with distinct failure times. Now we want to give several alternative partial likelihoods for data with ties between failure times.

Suppose there are N distinct observed failure times $t_1 < \cdots < t_N$, and at each time $t_j (1 \le j \le N)$ there are d_j observed failures. Let $\mathcal{D}_j$ be the set of all individuals who die at time t_j. Let $\mathcal{R}_j$ be the risk set at time t_j, i. e. $\mathcal{R}_j = \{i : X_i \ge t_j\}$.

When there are many ties in the data, the computation of maximum partial likelihood estimates, though still feasible, becomes time-consuming. For this reason, approximations to the partial likelihood function are often used. Two commonly employed approximations are due to Breslow and to Efron.

Breslow [21.4] suggested the following log partial likelihood for data with ties among failure times

$$L_{\mathrm{B}}(\beta) = \sum_{j=1}^{N} \left\{ \beta^T \sum_{l \in D_j} Z_l - d_j \log \left[\sum_{i \in \mathcal{R}_j} \exp\left(\beta^T Z_i\right) \right] \right\} .$$

This approximation works well when there are not many ties. Another approximation of the log partial likelihood is given by *Efron* [21.5]

$$L_{\mathrm{E}}(\beta) = \sum_{j=1}^{N} \left\{ \beta^T \sum_{l \in D_j} Z_l - \sum_{k=1}^{d_j} \log \left[\sum_{i \in \mathcal{R}_j} \exp\left(\beta^T Z_i\right) - (k-1)/d_j \sum_{i \in \mathcal{D}_j} \exp\left(\beta^T Z_i\right) \right] \right\} .$$

Breslow's method is easy to use and is therefore more popular, but Efron's approximation is generally the more accurate of the two. Also both likelihoods reduce to the partial likelihood when there is no tie.

21.2 Estimating the Hazard and Survival Functions

The cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)\,\mathrm{d}u$ can be estimated by *Breslow* [21.6]

$$\hat{\Lambda}_0(t) = \sum_{j: t_j \le t} \frac{\delta_j}{\sum_{i \in R_j} \exp(\hat{\beta}^T Z_i)} ,$$

where $\delta_j = I(T_j \le C_J)$. Note that $\hat{\Lambda}_0$ is a right-continuous step function with jumps at the observed failure times, and it is often referred to as the Breslow estimator. In the case of tied events, each of the subjects in a tie contributes its own term to the sum, and this term is the same for all subjects who failed at the specific time. This estimator can also be derived through a profile likelihood approach (*Johansen* [21.7], *Klein* and *Moeschberger* [21.2]). The baseline survival function $S_0(t) = \exp[-\Lambda_0(t)]$ can thus be estimated by $\hat{S}_0(t) = \exp[-\hat{\Lambda}_0(t)]$. The estimated survival function of an individual with covariate value z is given by

$$\hat{S}(t \mid z) = \exp\left[-\hat{\Lambda}_0(t) \mathrm{e}^{\hat{\beta}^T z} \right] .$$

21.3 Hypothesis Testing

Without loss of generality, assume that we are interested in hypothesis testing involving only the first q components of the regression parameter β. Write $\beta = \left(\beta_1^T, \beta_2^T\right)^T$, where β_1 is of dimension q and β_2 is of dimension $(p-q)$. For testing the null hypothesis $\beta_1 = \beta_{01}$ against the alternative $\beta_1 \neq \beta_{01}$ for any fixed β_{01} in the presence of the unknown parameters β_2, there are three types of tests: the likelihood ratio test, the Wald test, and the score test. This type of test with $\beta_{01} = 0$ is often used in model selection procedures, testing whether a given model can be improved by including a certain additional covariate or covariate combinations.

21.3.1 Likelihood Ratio Test

The test statistic for the likelihood ratio test is given by

$$TS_{\text{LR}} = 2\left[\log L(\hat{\beta}) - \log L(\tilde{\beta})\right] ,$$

where $\tilde{\beta} = (\beta_{01}, \tilde{\beta}_2^T)^T$ and $\tilde{\beta}_2$ maximizes $L(\beta)$ when β_1 is fixed at β_{01}. Under the null hypothesis, the asymptotic distribution of TS_{LR} is χ_q^2.

21.3.2 Wald Test

Let $\hat{\beta} = \left(\hat{\beta}_1^T, \hat{\beta}_2^T\right)$ denote the usual maximum partial likelihood estimate of the full parameter vector $\boldsymbol{\beta} = \left(\beta_1^T, \beta_2^T\right)$, and partition the inverse of the information matrix as

$$\mathbf{I}^{-1}(\beta) = \begin{pmatrix} \mathbf{I}^{11}(\beta) & \mathbf{I}^{12}(\beta) \\ \mathbf{I}^{21}(\beta) & \mathbf{I}^{22}(\beta) \end{pmatrix},$$

where $\mathbf{I}^{11}(\beta)$ is a $q \times q$ matrix. The test statistic for the Wald test is given by

$$TS_{\text{wald}} = (\hat{\beta}_1 - \beta_{01})^T \times \mathbf{I}^{11}(\hat{\beta})^{-1}(\hat{\beta}_1 - \beta_{01}) .$$

Under the null hypothesis, the asymptotic distribution of TS_{wald} is χ_q^2.

21.3.3 Score Test

Let $S_1(\beta)$ denote the vector of the first q components of the score function $S(\beta)$. The test statistic for the score test is

$$TS_{\text{score}} = S_1(\tilde{\beta})^T \times \mathbf{I}^{11}(\tilde{\beta}) S_1(\tilde{\beta}) ,$$

where $\tilde{\beta}$ and $\mathbf{I}_{11}(\beta)$ are defined as before. Again, the large sample distribution of the test statistic under the null hypothesis is χ_q^2.

21.4 Stratified Cox Model

The proportional hazards model can be stratified to account for heterogeneity in the baseline hazards. To achieve this, the subjects are divided into several groups with distinct baseline hazard functions and a common vector of regression coefficients β, and proportional hazards are assumed within each stratum. For a subject with covariate $\boldsymbol{Z}$ in the k-th stratum, let the hazard at time t be

$$\lambda(t \mid Z) = \lambda_k(t) \exp\left(\beta^T Z\right) .$$

Within each stratum a partial likelihood function can be defined as in (21.2), and the partial likelihood for the stratified Cox model is defined as the sum of the partial likelihood functions for all strata.

21.5 Time-Dependent Covariates

The Cox model can be extended to include time-dependent covariates.

Let $\boldsymbol{Z}(t)$ be a covariate vector measured at time t. Again we assume that the censoring is noninformative in that the failure time T and the censoring time C are conditionally independent, given the history of the covariate vector $\boldsymbol{Z}^*(X)$, where $\boldsymbol{Z}^*(t) = \{Z(u) : 0 \leq u < t\}$ for any $0 \leq t \leq X$. The dataset $\left\{\left[X_i, Z_i^*(X_i), \Delta_i\right] : i = 1, \cdots, n\right\}$ is an i.i.d. sample of $\left\{\left[X, Z^*(X), \Delta\right]\right\}$. Using similar notations as in Sect. 21.1, the hazard function for T is defined as

$$\begin{aligned}\lambda\left[t \mid \boldsymbol{Z}^*(t)\right] &= \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P\Big[t \leq T < t + \Delta t \mid T \geq t, \boldsymbol{Z}^*(t)\Big] \\ &= \lambda_0(t) \exp\left[\beta^T z(t)\right] .\end{aligned}$$

The partial likelihood is then defined as

$$\prod_{j=1}^{N} \frac{\exp\left[\beta^T Z_{(j)}(X_{(j)})\right]}{\sum_{i\in\mathcal{R}_j} \exp\left[\beta^T Z_i(X_{(j)})\right]}\,,$$

where, like in Sect. 21.2, (j) denotes the label of the subject with the j-th earliest observed failure time, and R_j denotes the corresponding risk set.

Though the extension of the model and partial likelihood is simple, the validity of the underlying assumption and the computation of the partial likelihood are quite complicated in practice. Extra care is needed in dealing with time-dependent covariates.

Kalbfleisch and *Prentice* [21.8] pointed out that there are two types of time-dependent covariates: external covariates, whose value do not depend on the failure process, and internal covariates, which usually carry information about the failure process. For example, when studying how long an experimental unit remains functioning, the air humidity rate is essentially external to the units work duration. But a patient's daily blood pressure is an internal time-dependent covariate, since it carries information about the health status and hence the failure time of the patient. Though the definition of the hazard function and the construction of the partial likelihood apply to both type of time-dependent covariates, it is not possible to estimate the conditional survival function when there are internal covariates.

Another problem concerns the measurement of time-dependent covariates. The calculation of partial likelihood requires that the values of any unit's time-dependent covariates be available at all failure times when it is still at risk. This cannot be achieved in general, since we can never know the failure times in advance and the information on the covariates are usually collected at predetermined time points.

One way to deal with incomplete history of time-dependent covariates is imputation, and there are several possible ways to impute the intermittent values. An ad hoc approach commonly used in practice, referred to as the LVCF method, is to impute the missing covariate at a certain time point with the nearest previous observation of the same unit. Other nonparametric approaches, like the smoothing methods, can be used to estimate the unobserved part of the time-dependent data.

If one can make the assumption that the time-dependent covariates follow certain models, for example, linear mixed effects models, several strategies can be applied, including regression calibration methods [21.9–11], joint likelihood methods, [21.12–14], and conditional score methods, [21.15].

One also needs to be aware of the possible informative censoring due to the fact that the time-dependent covariates are truncated by failure [21.16]. *Allison* [21.17] suggested avoiding using observations after the failure time, unless one is dealing with external time-dependent covariates. Failure to take into account these problems can lead to biased parameter estimates.

21.6 Goodness-of-Fit and Model Checking

21.6.1 Tests of Proportionality

A key assumption of the Cox proportional hazard model is the proportionality of the hazards. Note that the hazard function for an individual depends on the covariate values and the value of the baseline hazard. For any two individuals, it is easy to see from (21.1) that the ratio of the hazards over time will be constant. The validity of this assumption needs to be checked. Various methods can be used for this purpose.

One graphical approach to check the proportionality assumption is the log–log survival plot. We first divide all study units into several groups according to the covariate values, and then estimate the survival function within each group using the *Kaplan–Meier* method, [21.18]. If the proportional hazard assumption is satisfied, plotting the transformed survival functions, $\log\left[-\log S(t)\right]$, would result in parallel curves. This method works well with categorical covariates without many levels. For continuous predictors, one may first divide them into a few categories based on quantiles or other grouping criterion. For categorical covariates with many levels, one may want to combine them into fewer groups.

Other graphic approaches includes *Andersen* plots [21.19], *Arjas* plots [21.20], and the use of plots based on the score residuals or Schoenfeld residuals. The details of these approaches can be found in *Klein* and *Moeschberger* [21.2], Chapt. 11.

Alternatively, one can check the proportionality assumption by adding a time-dependent covariate to the model. The time-dependent covariate can be defined as the interaction of a time-independent covariate and a function of time. If the time-dependent covariate is significant in the Cox model, it could be concluded that the effect of the time-independent covariate varies with time, and the proportional hazards assumption is violated.

21.6.2 Test of the Functional Form of a Continuous Covariate

Next we want to examine the functional form of a given covariate on the survival and see whether any transformation of the original covariate is needed. One method is based on martingale residuals.

Let $\hat{\beta}$ be the estimated regression coefficient and $\hat{\Lambda}_0(\cdot)$ be the Breslow estimator of the cumulative baseline hazard function, the martingale residual for the i-th subject is defined as

$$\hat{M}_i = \delta_i - \hat{\Lambda}_0(X_i)\exp\left(\hat{\beta}^T Z_i\right) = \delta_i - \hat{\Lambda}_i(X_i)\,.$$

It can be shown that when sample size is large, expectation of $\hat{M}_i$ is close to 0 and, for $i \neq j$, $\mathrm{cov}(\hat{M}_i, \hat{M}_j)$ is also close to 0. We also have $\sum \hat{M}_i = 0$.

When the functional form of a specific covariate needs to be examined, we first fit a Cox model with all covariates excluding the covariate to be investigated, and then plot the martingale residuals against the excluded covariate. If the plot shows an approximately linear trend, no transformation is needed and the untransformed covariate can be included in the model with the other covariates. If, however, there appears to be a certain pattern, a proper transformation of the original covariate might be needed.

Nonparametric methods can also be used to explored the nature of covariate effects on survival. Local likelihood or penalized likelihood methods with different smoothing methods have been proposed to estimate the functional form of a single covariate or a linear combination of covariates in the survival model [21.21–25].

21.6.3 Test for the Influence of Individual Observation

The influence of individual observation may be studied by the use of score residuals. We can fit the Cox model with and without the i-th observation of the data sample, and obtain the estimates $\hat{\beta}$ and $\hat{\beta}_{(i)}$, respectively. If $\hat{\beta} - \hat{\beta}_{(i)}$ is close to zero the individual observation has little influence on the estimate. We can plot this difference against the observation number to identify those influential observations.

21.6.4 Test for the Overall Fit

The overall fit of the model can be assessed using the Cox–Snell residuals, which are defined as

$$r_i = \hat{\Lambda}_0(X_i)\exp\left(\hat{\beta}^T Z_i\right), \qquad i = 1, \cdots, n\,,$$

where $\hat{\Lambda}$ is the Breslow estimator for the cumulative baseline hazard function and $\hat{\beta}$ are the estimated regression coefficients. It can be shown that, when there is no censoring and the true values of the parameters are known, $\Lambda_0(T_i)\exp\left(\beta^T Z_i\right)$ follows an exponential distribution with unit rate. Thus we can treat $\{r_i, i = 1, \ldots, n\}$ as a possibly right-censored sample of failure times from the unit exponential distribution, which has a constant hazard rate $\lambda_r(r) = 1$ and a cumulative hazard function $\Lambda_r(r) = r$. The failure indicator δ_i for X_i can also serve as the failure indicator for r_i. The Nelson–Aalen estimator can be used to estimate Λ_r:

$$\hat{\Lambda}_r(r) = \sum_{i:r_i \le r} \frac{\delta_i}{\#\{r_j : r_j \ge r_i\}}\,,$$

where # counts the number of elements in a set. If the model is correct, the function $\hat{\Lambda}_r$ would be close to the straight line $\Lambda_r(r) = r$. Thus, plotting $\hat{\Lambda}_r(r_i)$ against r_i will provide an assessment of the departure from the model assumptions.

21.6.5 Test of Time-Varying Coefficients

Schoenfeld residuals [21.26] can be used to test for time-varying coefficients in the Cox model. Let $Y_i(t) = I(X_i \ge t)$, be the indicator of whether unit i is still at risk at time t. For a subject who fails at time t_i, the vector of Schoenfeld residuals is defined as

$$Z_i(t_i) - \bar{Z}(t_i)\,,$$

where $\bar{Z}(t)$ is defined as

$$\bar{Z}(t) = \frac{\sum_{j=1}^{n} Y_j(t) Z_j \exp\left(\hat{\beta}^T Z_j\right)}{\sum_{j=1}^{n} Y_j(t)\exp\left(\hat{\beta}^T Z_j\right)}\,,$$

which is a weighted average of the covariate $\mathbf{Z}$ over all individuals at risk at time t.

Under the proportional hazards assumption, the Schoenfeld residuals should be independent of time. Therefore, a clear pattern of the Schoenfeld residuals over time implies a departure from the model assumption.

For the j-th covariate, a time-varying coefficient can be expressed as

$$\beta_j(t) = b_{0j} + b_{1j} f_j(t) ,$$

where $f_j(t)$ is a known function of time. *Grambsch* and *Therneau* [21.27] showed that the scaled Schoenfeld residuals from a Cox model have a mean of approximately $b_{1j} f_j(t)$ for the j-th covariate at time t. So the plot of the scaled Schoenfeld residuals against the event times can be used to check whether the coefficient of f_j is zero.

21.6.6 Test for a Common Coefficient Across Different Groups

To test whether the effect of a covariate on failure is identical across different groups of study units, a stratified analysis can be utilized. First, a stratified Cox model is fitted with different regression coefficients for different groups. Then another stratified Cox model is fitted with the same coefficient for all groups. The difference between the log likelihoods, which has an approximate χ^2 distribution, can be used to test for any heterogeneity in the covariate effect.

21.7 Extension of the Cox Model

21.7.1 Cox Model with Random Effects

Similarly to the case of a linear model, random effects can be added to the proportional hazards model to handle clustered or heterogeneous survival data [21.28]. Let

$$\lambda(t \mid \boldsymbol{Z}) = \lambda_0(t) \exp\left(\beta^T \boldsymbol{Z} + b^T \boldsymbol{W}\right) ,$$

where b is the random effect of the covariate vector $\boldsymbol{W}$ on survival. This model allows for a multivariate random effect with known distribution. Maximum likelihood estimates of the regression parameters, the variance components and the baseline hazard function can be obtained via the expectation maximization (EM) algorithm.

21.7.2 Nonproportional Models

Several models can be applied when one suspects that the proportionality assumption does not hold for a certain dataset. The most common among them are frailty models and cure-rate models.

Frailty models can be used to account for individual randomness in an experiment. An unobserved random variable W is added to a Cox proportional hazards model, which is assumed to follow a known distribution. The effect of W is multiplicative on the hazard: given W, the hazard rate is given by

$$\lambda(t \mid \boldsymbol{W}, \boldsymbol{Z}) = \lambda_0(t) W \exp\left(\beta^T \boldsymbol{Z}\right) .$$

With a common choice for the distribution of W, Gamma$(1, \theta)$, the marginal hazard rate given the covariate vector $\boldsymbol{Z}$ is

$$\lambda(t \mid \boldsymbol{Z}) = \frac{\lambda_0(t) \exp\left(\beta^T \boldsymbol{Z}\right)}{1 + \theta \exp\left(\beta^T \boldsymbol{Z}\right) \Lambda_0(t)} .$$

Another commonly used model for nonproportional hazards is the cure-rate model. This assumes that there are two sub-populations: cured subjects and uncured subjects. Suppose the proportion of the cured subjects is π and that of the uncured subjects is $1-\pi$. The survival probability at t given covariates $\boldsymbol{Z}$ is then given by

$$S(t \mid \boldsymbol{Z}) = \pi S_1(\boldsymbol{Z}) + (1-\pi) S_2(t \mid \boldsymbol{Z}) ,$$

where $S_1(\boldsymbol{Z})$ is the probability of being cured and $S_2(t \mid \boldsymbol{Z})$ represents the survival function of the uncured,

Table 21.1 Data table for the example

Time to breakdown (s)				
45 kV	**40 kV**	**35 kV**	**30 kV**	**25 kV**
1	1	30	50	521
1	1	33	134	2517
1	2	41	187	4056
2	3	87	882	12553
2	12	93	1448	40290
3	25	98	1468	50560+
9	46	116	2290	52900+
13	56	258	2932	67270+
47	68	461	4138	83990
50	109	1182	15750	85500+
55	323	1350	29180+	85700+
71	417	1495	86100+	86420+

+ represents censoring

which is modeled as in a proportional hazards model. The cure-rate model can be used to fit failure-time data when the right tail of the survival function looks like a plateau.

21.7.3 Multivariate Failure Time Data

Sometimes an experimental unit might experience multiple failures, which could be of the same nature (e.g., same event recurring over time), or of different nature (e.g., distinct types of problems). Also, in some situations there are clustering of study units such that failure times within the same cluster are expected to be correlated. For example, pieces of equipment in the same factory might behave similarly to those in another factory.

Various methods have been developed for these kinds of multivariate failure-time data. See, for example, [21.29–33].

21.8 Example

We now use a dataset from the book by *Wayne* [21.34] as an illustration. The data came from an experiment on testing the fatigue limit for two steel specimens in two forms under different stress ratios. The data are shown in Table 21.1.

The fitting result is shown in Table 21.2. We can see that voltage has a negative effect on the failure rate. The higher the voltage is, the shorter the time to break down will be. One unit increase in voltage results in an hazard ratio of 1.27 (95% confidence interval = 1.19–1.36).

The survival function estimates for five groups of voltage are displayed in Fig. 21.1. The graph shown in Fig. 21.2 of the graph of log[-log(survival)] versus the log of survival time results in parallel straight lines, so the proportional hazard assumption is satisfied. The

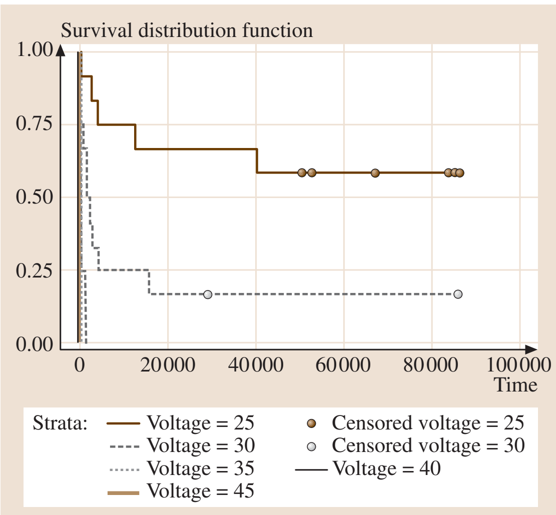

Fig. 21.1 Estimated survival function

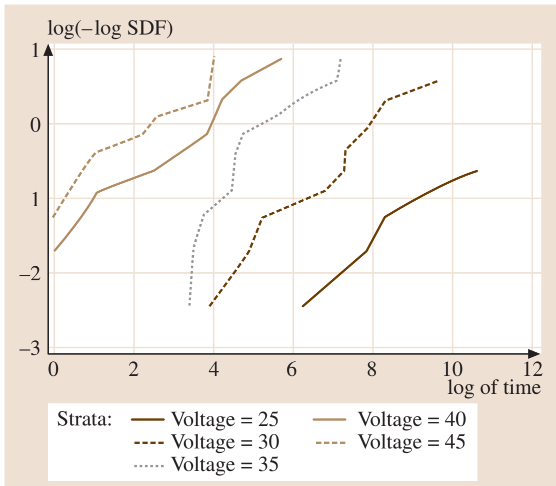

Fig. 21.2 Checking the proportional hazard assumption

Table 21.2 Model fitting result

	coef	exp(coef)	se(coef)	*z*	*p*
Voltage	0.241	1.27	0.0339	7.12	1.1e-012
	exp(coef)	exp(-coef)	lower 0.95	upper 0.95	
Voltage	1.27	0.786	1.19	1.36	
Rsquare = 0.68 (max possible = 0.997)					
Likelihood ratio test = 68.3				on 1 df,	*p* = 1.11e-016
Wald test = 50.7				on 1 df,	*p* = 1.07e-012
Score (logrank) test = 68.1				on 1 df,	*p* = 1.11e-016

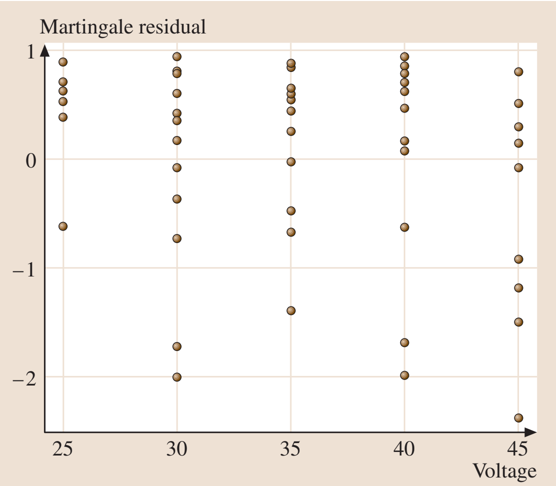

Fig. 21.3 Martingale residual plot

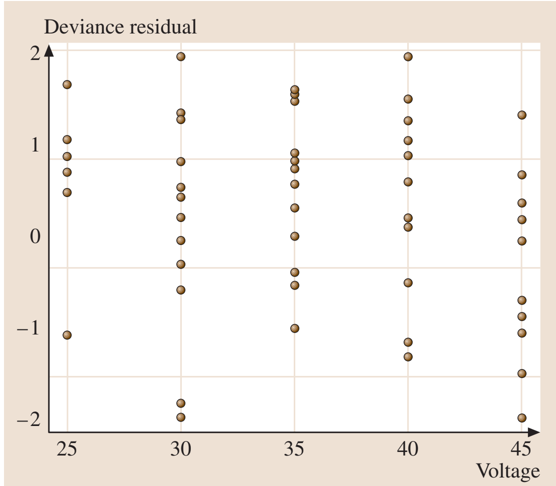

Fig. 21.4 Deviance residual plot

martingale residual plot is shown in Fig. 21.3 and the deviance residual plot is shown in Fig. 21.4. There is no indication of a lack of fit of the model to individual observations.

References

21.1 D. R. Cox: Regression models, life-tables (with discussion), J. R. Stat. Soc. B **34**, 187–220 (1972)

21.2 J. P. Klein, M. L. Moeschberger: *Survival Analysis: Techniques for Censored and Truncated Data* (Springer, Berlin Heidelberg New York 1997)

21.3 D. R. Cox: Partial likelihood, Biometrika **62**, 269–276 (1975)

21.4 N. E. Breslow: Covariance analysis of censored survival data, Biometrics **30**, 89–99 (1974)

21.5 B. Efron: The efficiency of Cox's likelihood function for censored data, J. Am. Stat. Assoc. **72**, 557–565 (1977)

21.6 N. E. Breslow: Contribution to the discussion of the paper by D. R. Cox, J. R. Stat. Soc. B **34**, 187–220 (1972)

21.7 S. Johansen: An extension of Cox's regression model, Int. Stat. Rev. **51**, 258–262 (1983)

21.8 J. D. Kalbfleisch, R. L. Prentice: *The Statistical Analysis of Failure Time Data* (Wiley, New York 1980)

21.9 Y. Pawitan, S. Self: Modeling disease marker processes in AIDS, J. Am. Stat. Assoc. **83**, 719–726 (1993)

21.10 U. G. Dafni, A. A. Tsiatis: Evaluating surrogate markers of clinical outcome when measured with error, Biometrics **54**, 1445–1462 (1998)

21.11 A. A. Tsiatis, V. Degruttola, M. S. Wulfsohn: Modeling the relationship of survival to longitudinal data measured with error: applications to survival, CD4 counts in patients with AIDS, J. Am. Stat. Assoc. **90**, 27–37 (1995)

21.12 C. J. Faucett, D. C. Thomas: Simultaneously modeling censored survival data, repeatedly measured covariates: a Gibbs sampling approach, Stat. Med. **15**, 1663–1685 (1996)

21.13 M. S. Wulfsohn, A. A. Tsiatis: A joint model for survival, longitudinal data measured with error, Biometrics **53**, 330–339 (1997)

21.14 R. Henderson, P. Diggle, A. Dobson: Joint modelling of longitudinal measurements, event time data, Biostat. **4**, 465–480 (2000)

21.15 A. A. Tsiatis, M. Davidian: A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error, Biometrika **88**, 447–458 (2001)

21.16 M. C. Wu, R. J. Carroll: Estimation, comparison of changes in the presence of informative right censoring by modeling the censoring process, Biometrics **44**, 175–188 (1988)

21.17 P. D. Allison: *Survival Analysis Using the SAS System: A Practical Guide* (SAS Institute, Cary 1995)

21.18 E. L. Kaplan, P. Meier: Nonparametric estimation from incomplete observations, J. Am. Stat. Assoc. **53**, 457–481 (1958)

21.19 P. K. Andersen: Testing goodness of fit of Cox's regression and life model, Biometrics **38**, 67–77 (1982)

21.20 E. A. Arjas: Graphical method for assessing goodness of fit in Cox's proportional hazards model, J. Am. Stat. Assoc. **83**, 204–212 (1988)

21.21 R. Tibshirani, T. Hastie: Local likelihood estimation, J. Am. Stat. Assoc. **82**, 559–567 (1987)

21.22 L. A. Sleeper, D. P. Harrington: Regression splines in the Cox model with application to covariate effects in liver disease, J. Am. Stat. Assoc. **85**, 941–949 (1990)

21.23 R. Gentleman, J. Crowley: Local full likelihood estimation for the proportional hazards model, Biometrics **47**, 1283–1296 (1991)

21.24 J. Fan, I. Gijbels, M. King: Local likelihood, local partial likelihood in hazard regression, Ann. Stat. **25**, 1661–1690 (1997)

21.25 W. Wang: Proportional hazards regression with unknown link function, time-dependent covariates, Stat. Sin. **14**, 885–905 (2004)

21.26 D. Schoenfeld: Partial residuals for the proportional hazards regression model, Biometrika **69**, 239–241 (1982)

21.27 P. M. Grambsch, T. M. Therneau: Proportional hazards tests, diagnostics based on weighted residuals, Biometrika **81**, 515–526 (1994)

21.28 D. J. Sargent: A general framework for random effects survival analysis in the Cox proportional hazards setting, Biometrics **54**, 1486–1497 (1998)

21.29 R. L. Prentice, B. J. Williams, A. V. Peterson: On the regression analysis of multivariate failure time data, Biometrika **68**, 373–379 (1981)

21.30 L. J. Wei, D. Y. Lin, L. Weissfeld: Regression analysis of multivariate incomplete failure time data by modeling marginal distribution, J. Am. Stat. Assoc. **84**, 1065–1073 (1989)

21.31 C. F. Spiekerman, D. Y. Lin: Marginal regression models for multivariate failure time data, J. Am. Stat. Assoc. **93**, 1164–1175 (1998)

21.32 D. Y. Lin, L. J. Wei, I. Yang, Z. Ying: Semiparametric regression for the mean, rate functions of recurrent events, J. R. Stat. Soc. B **62**, 711–730 (2000)

21.33 P. K. Andersen, R. D. Gill: Cox's regression model counting process: a large sample study, Ann. Stat. **10**, 1100–1120 (1982)

21.34 N. Wayne: *Accelerated Testing: Statistical Models, Test Plans, And Data Analysis* (Wiley, New York 1990)