Applied Survival Analysis

Regression Modeling of Time to Event Data

DAVID W. HOSMER, Jr.

Department of Biostatistics and Epidemiology
University of Massachusetts
Amherst, Massachusetts

STANLEY LEMESHOW

Department of Biostatistics and Epidemiology
University of Massachusetts
Amherst, Massachusetts

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

APPENDIX 1

The Delta Method

A problem faced by statisticians when developing an estimator of a parameter is deriving an expression for an estimator of the variance of the estimator. Both estimators are needed for confidence interval estimation and/or hypothesis testing.

A procedure commonly called the *delta method* has been used by statisticians for many years to obtain an estimator of the variance when the estimator is not a simple sum of observations. The basic idea is to use a method from calculus called a *Taylor series expansion* to derive a linear function that approximates the more complicated function. We refer the reader to any introductory calculus text for a discussion of the Taylor series expansion.

To apply the delta method, the function must be one that can be approximated by a Taylor series and, in general, this means that it is a "smooth" function, with no "corners." Consider such a function of a random variable X denoted as $f(X)$. To apply the delta method we use the first two terms of a Taylor series expansion about the mean of the variable to approximate the value of the function as

$$f(X) \cong f(\mu) + (X - \mu) f'(\mu) \tag{A.1}$$

where

$$f'(\mu) = \frac{\partial f(X)}{\partial X}\Big|_{X=\mu}$$

is the derivative of the function with respect to X evaluated at the mean of X. It follows from (A.1) that the variance of the function is approximately

$$\begin{aligned}\mathrm{Var}[f(X)] &\cong \mathrm{Var}(X-\mu)\times[f'(\mu)]^2 \\ &\cong \sigma^2 \times [f'(\mu)]^2, \end{aligned} \tag{A.2}$$

where σ^2 is the variance of X. The delta method estimator of the variance of the function is obtained when we use the estimators of μ and σ^2 in (A.2) as follows:

$$\hat{\mathrm{Var}}[f(x)] \cong \hat{\sigma}^2 \times [f'(\hat{\mu})]^2. \tag{A.3}$$

As an example, consider the function $\ln(X)$. The expansion from (A.1) is

$$\ln(X) \cong \ln(\mu) + (X-\mu)\frac{1}{\mu}. \tag{A.4}$$

The delta method estimator of the variance from (A.3) is

$$\hat{\mathrm{Var}}[\ln(X)] \cong \hat{\sigma}^2 \frac{1}{\hat{\mu}^2},$$

where $\hat{\sigma}^2$ and $\hat{\mu}$ denote estimators of σ^2 and μ.

As a second example, we provide the details for the development of the delta method estimator of the variance of the log of the Kaplan–Meier estimator of the survivorship function shown in (2.3) and that of the Kaplan–Meier estimator itself in (2.5). The estimator as shown in (2.1) is

$$\hat{S}(t) = \prod_{t_{(i)} \le t} \frac{n_i - d_i}{n_i},$$

and its log is

$$\begin{aligned}\ln[\hat{S}(t)] &= \sum_{t_{(i)} \le t} \ln\left(\frac{n_i - d_i}{n_i}\right) \\ &= \sum_{t_{(i)} \le t} \ln(\hat{p}_i), \end{aligned}$$

where $\hat{p}_i = (n_i - d_i)/n_i$. The first key assumption in the development of the variance estimator is that the observations of survival among the n_i subjects at risk are independent Bernoulli with constant probability, p_i. Under this assumption the estimator of the constant probability is $\hat{p}_i$ with variance estimator $\hat{p}_i(1-\hat{p}_i)/n_i$. The Taylor series expansion for the log function in (A.4) yields

$$\ln(\hat{p}_i) \cong \ln(p_i) + (\hat{p}_i - p_i)\frac{1}{\hat{p}_i}$$

and from (A.3) the delta method variance estimator is

$$\hat{\mathrm{Var}}[\ln(\hat{p}_i)] \cong \frac{1}{\hat{p}_i^2}\frac{\hat{p}_i(1-\hat{p}_i)}{n_i}$$
$$\cong \frac{d_i}{n_i(n_i - d_i)}.$$

The second key assumption is that observations in different risk sets are independent. Thus the delta method estimator of the variance of the log of the Kaplan–Meier estimator is, as shown in (2.3),

$$\hat{\mathrm{Var}}\{\ln[\hat{S}(t)]\} \cong \sum_{t_{(i)} \le t} \hat{\mathrm{Var}}[\ln(\hat{p}_i)]$$
$$\cong \sum_{t_{(i)} \le t} \frac{d_i}{n_i(n_i - d_i)}.$$

The estimator of the variance of the Kaplan–Meier estimator comes from a second application of the delta method. In this application the function is

$$f(X) = \exp(X),$$

e.g., $\hat{S}(t) = \exp\{\ln[\hat{S}(t)]\}$. It follows from (A.1) that the series expansion is

$$\exp(X) \cong \exp(\mu) + (X - \mu)\exp(\mu)$$

and, from (A.2) the approximate variance is

$$\hat{\mathrm{Var}}[\exp(X)] \cong \sigma^2[\exp(\mu)]^2. \tag{A.5}$$

Application of the approximation in (A.5) yields the Greenwood estimator in (2.5), namely

$$\hat{\mathrm{Var}}[\hat{S}(t)] \cong [\hat{S}(t)]^2 \sum_{t_{(i)} \le t} \frac{d_i}{n_i(n_i - d_i)}.$$

The confidence interval estimator for the Kaplan–Meier estimator discussed in Chapter 2 is based on the log-log survivorship function, that is, $\ln\{-\ln[\hat{S}(t)]\}$. The variance estimator of this function requires a second application of the expansion of the log function. In this case $X = \ln[\hat{S}(t)]$ and application of the approximation to the variance of the log of a random variable yields the estimator shown in (2.6), namely,

$$\hat{\mathrm{Var}}\{\ln[-\ln(\hat{S}(t))]\} \cong \frac{1}{[\ln(\hat{S}(t))]^2} \sum_{t_{(i)} \le t} \frac{d_i}{n_i(n_i - d_i)}.$$

The results presented in this appendix provide a brief introduction to the use of the delta method within the specific context of deriving an estimator of the variance of the Kaplan–Meier estimator, or functions of it. The technique is quite general and has been used in a variety of settings [see Agresti (1990) for applications in general categorical data models].

APPENDIX 2

An Introduction to the Counting Process Approach to Survival Analysis

We refer to the counting process approach to the analysis of survival time throughout the text. This method has been the source of many new developments in the field since it was first used by Aalen [(1975) and (1978)]. Two recent texts, Fleming and Harrington (1991) and Andersen, Borgan, Gill and Keiding (1993), document the mathematical details of this powerful method in a thorough manner. Fleming and Harrington (1991) focus primarily on the analysis of survival time while Andersen et al. (1993) consider analysis of survival time as well as other, more general statistical problems. We encourage readers of this text to see Andersen et al. (1993, Chapter 1) for an excellent overview of the types of statistical problems that can be formulated as counting processes. Nontechnical mathematical introductions to the approach are provided by Fleming and Harrington (1991, Chapter 0) and Andersen et al. (1993, Section II.1). The purpose of this appendix is to introduce a few of the key ideas and constructs used in the counting process approach to the analysis of survival time. For this reason many of the more technical mathematical assumptions and details will not be discussed.

For the time being, suppose we follow a single subject from time of enrollment, $t = 0$, in a study of a particular cancer until death from this cancer. Furthermore, we assume it is a 5-year study and that this subject is enrolled on the first day of the study. Thus the maximum length of follow-up for this subject is 60 months. We denote the survival time random variable as X. A common approach to modeling the possibility of right censoring is to assume that there is a second random variable,

independent of X, that records the time until observation terminates due to anything other than the event of interest, for example, death from another cause or loss to follow-up for reasons unrelated to any study factor. We denote this random variable as Z. The actual observed time random variable is $T = \min(X, Z)$ and the available data for a subject consists of T and an indicator variable C whose value is 1 if $T = X$ and 0 if $T = Z$. Thus the variable T records follow-up time and C is the censoring indicator variable.

Three functions of time central to the counting process approach are: the counting process

$$N(t) = I(T \le t, C = 1),$$

the at risk process

$$Y(t) = I(T \ge t)$$

and the intensity process

$$\lambda(t)dt = Y(t)h(t)dt\ ,$$

where

$$h(t)dt = \Pr(t \le T < t + dt, C = 1 | T \ge t)$$

is the hazard function for survival time. The function $I(\cdot)$ is the indicator function whose value is 1 if the argument is true and 0 otherwise.

The counting process records, in our example, whether death from cancer occurs at time t. The function "counts" this by jumping from a value of 0 to a value of 1. For example, suppose our hypothetical subject died from cancer after being in the study 32 months, $(T = 32, C = 1)$. The counting process function for this subject is equal to zero until 32 months. At exactly 32 months the function jumps to a value of 1. The function is equal to 1 for the remaining 28 months of follow-up time. If the subject's follow-up time is right censored, $C = 0$, then a death is not counted and the counting process is equal to 0 for all values of t. For example, if our hypothetical subject was removed from the study at 32 months for reasons unrelated to the cancer, $(T = 32, C = 0)$, the counting process for this subject is equal to zero over the 60 months of follow-up.

The at risk process indicates whether the subject is still being followed, at risk for death, at time t. This function jumps from a value of 1 to a value of 0 when follow-up ends due to death or censoring. For a hypothetical subject with follow-up time of 32 months, the at risk process is equal to 1 from the beginning of follow-up until 32 months. The function jumps/drops to a value of zero just after 32 months as the subject is no longer at risk for the remaining 28 months of the study.

The intensity process may be viewed as being like an "expected number of deaths" at time t. This follows from the fact that the function is of the form " $n \times p$" (i.e., the expected number of events in a binomial distribution). The at risk process corresponds to "n" and the hazard function to "p".

The process of following the hypothetical subject from time zero to time t may be thought of as an accumulation of many conditional independent steps, much like the argument used to construct the Kaplan–Meier estimator in Chapter 2. The total expected number of deaths up to time t is obtained from the intensity process in the same manner as the cumulative hazard is obtained from the hazard function, namely by integrating the intensity process over time to obtain

$$\begin{aligned}\Lambda(t) &= \int_0^t \lambda(u)\, du \\ &= \int_0^t Y(u)h(u)\, du,\end{aligned}$$

and this function is called the cumulative intensity process.

Thus one may think of the counting process as the total number of observed events and the cumulative intensity process as the total number of expected events up to time t. The difference between these two quantities is a residual-like quantity called the *counting process martingale*, namely

$$M(t) = N(t) - \Lambda(t).$$

This function is the basis for the martingale residuals that play a central role in model evaluation methods; see Chapter 6. Another way to express the relationship between the counting, intensity and martingale processes is via a linear-like model

$$N(t) = \Lambda(t) + M(t).$$

When expressed in this way we see that the counting process, the observed part of the model, is the sum of a systematic component, the cumulative intensity process and a residual, the martingale process. In our hypothetical example death or censoring can occur only one time and at the actual follow-up time, T, the value of the martingale is

$$M(T) = \begin{cases} 1 - \Lambda(T) \text{ if } C = 1, \\ 0 - \Lambda(T) \text{ if } C = 0. \end{cases}$$

It is well beyond the scope of this appendix to explain what makes a process a martingale and what gives $M(t)$ this quality. We refer the interested reader to the texts cited above for these technical details. It suffices for the purposes of this appendix and text to think of $M(t)$ as being similar to a residual.

Now suppose we have observations of follow-up time and censoring indicator variable on n subjects in our hypothetical cancer study. We assume that observations of time are independent and identically distributed. We denote the actual observed times and right censoring indicator variables in the usual way as (t_i, c_i), $i = 1,2,\ldots,n$. In this setting, a basic result from counting process theory is that the estimator of the cumulative intensity process for the ith subject at time t is

$$\hat{\Lambda}_i(t) = Y_i(t)\hat{H}(t),$$

where

$$Y_i(t) = I(t_i \geq t),$$

$$\hat{H}(t) = \sum_{t_j \leq t} \frac{c_j}{n_j}$$

is the Nelson–Aalen estimator of the cumulative hazard at t and

$$n_j = \sum_{i=1}^{n} Y_i(t_j)$$

is the number at risk at time t_j. The estimator of the martingale residual for the ith subject at his/her follow-up time is

$$\begin{aligned}\hat{M}(t_i) &= c_i - \hat{\Lambda}(t_i) \\ &= c_i - Y_i(t_i)\hat{H}(t_i) \\ &= c_i - \hat{H}(t_i)\end{aligned}$$

since $Y_i(t_i) = I(t_i \geq t_i) = 1$. We denote this martingale residual as $\hat{M}_i$. We note that, like residuals from most regression models, $\Sigma\hat{M}_i = 0$.

Assume that we have, in addition to follow-up time and censoring indicator variables, observations on p fixed (not time-varying) covariates. Assume that we fit a proportional hazards regression model. The estimator of the cumulative intensity process for the ith subject at time t is

$$\hat{\Lambda}\left(t, \mathbf{x}_i, \hat{\boldsymbol{\beta}}\right) = -Y_i(t)e^{\mathbf{x}_i'\hat{\boldsymbol{\beta}}} \ln\left[\hat{S}_0(t)\right].$$

Thus the value of the martingale residual for the ith subject at his/her follow-up time is

$$\hat{M}_i = c_i - e^{\mathbf{x}_i'\hat{\boldsymbol{\beta}}} \ln\left[\hat{S}_0(t_i)\right].$$

The estimated martingale residuals are the basis for many of the diagnostic methods for assessing various aspects of the fitted model; see Chapters 5 and 6.

One, if not the, major theoretical benefit derived from formulating a survival analysis as a counting process is that a number of theorems from martingale theory may be used to prove many of the distributional results cited in this text. For example, this theory may be used to prove that the maximum partial likelihood estimators of the coefficients in a proportional hazards model are asymptotically normally distributed with a covariance matrix that may be estimated by the observed information matrix (Chapter 3). A second example involves the proof that the Kaplan–Meier estimator and functions of it are asymptotically normally distributed (Chapter 2). The list of applications of this theory in survival analysis is quite long. The central theme in all of the applications involves proving that a particular scaled and centered estimator, such as the Kaplan–Meier estimator, $\sqrt{n}\left[\hat{S}(t) - S(t)\right]$, is a martingale.

In summary we feel that it is important for anyone using the regression methods for the analysis of survival time described in this text to

have at least a superficial knowledge of the basics of the counting process paradigm.