Models for Times-to-Event

The "event" of interest is often death or failure; therefore, we often refer to "survival" as the property of interest. The field of study is called "survival analysis". In engineering, it is also called "reliability".

We develop models for the survival time and other random properties of interest, and of the parameters that characterize interesting properties of the survival time, such as its mean and its conditional mean given survival to a fixed time.

The potential data in time-to-event studies are frequently incompletely observed. There are several ways this can happen:

- censoring
  - left, right, or interval
  - fixed, type I or type II
  - random, type I or type II
  - singly or progressively (multiply)
- truncation

This incompleteness of the data does not affect the model, of course, but it does affect how we estimate parameters in the model. In particular, it affects the likelihood. (Most estimators in survival analysis are MLEs.)

Basic Probability Models

Let T be a nonnegative random variable that is associated with the time to the event. The functions of interest that describe this random variable are

- survival function:

  $$S(t) = \Pr(T > t)$$

  Notice that at all points of continuity, $S(t) = 1 - F(t)$, where $F(t)$ is the CDF for T. (Note that the survival function is left-continuous, whereas the CDF is right-continuous. For estimators of the survival function, however, we generally use right-continuous functions—this seems to be a mistake in Harrell.)

- probability density function (or probability mass function):

  $$\begin{aligned} f(t) &= \mathrm{d}F(t)/\mathrm{d}t \\ &= -\mathrm{d}S(t)/\mathrm{d}t \end{aligned}$$

  While all of the functions are related, it is the distribution determined by this PDF that provides the characterization of the survival model.

- hazard rate or hazard function:

  $$\lambda(t) = \lim_{u \to 0} \frac{\Pr(t < T \leq t + u \mid T > t)}{u}$$

  (In most of the literature, the hazard function is denoted as $h(t)$ instead of $\lambda(t)$.)

- cummulative hazard function:

  $$\Lambda(t) = \int_0^t \lambda(v) \mathrm{d}v.$$

Distributions in Survival Analysis

There are several probability distributions that are useful in modeling survival times. Although the random variable has no negative support, even the normal distribution is sometimes used. More commonly, however, are distributions over the positive reals, such as the exponential, the Weibull (and its special case, the exponential), the gamma (also with the exponential as a special case), the log-normal, and the log-logistic.

The characteristics of the distribution, especially the nature of the corresponding hazard function, determine its usefulness.

The Weibull Distribution

Because its corresponding hazard function can assume various useful shapes, the Weibull distribution is one of the most widely used in survival analysis and reliability. Its density is

$$f(t) = \alpha\gamma t^{\alpha-1} \exp(-\alpha t^{\gamma}) \tag{1}$$

(This is Harrell's notation. Most people use α where he uses γ and λ where he uses α; that's why they don't use λ to denote the hazard function, as Harrell does.)

The (baseline) hazard rate is

$$\lambda(t) = \alpha\gamma t^{\gamma-1}, \tag{2}$$

which is monotone (strictly increasing, strictly decreasing, or constant), and the (baseline) survival function is

$$S(t) = \exp(-\alpha t^{\gamma}). \tag{3}$$

Some Interesting Properties and Relationships in the Probability Models

In the following, we'll assume that the underlying distribution is continuous. There are generally simple modifications that can be made for discrete distributions.

- The mean life is

$$\begin{aligned} \mu &= \mathrm{E}(T) \\ &= \int_0^\infty S(v)\mathrm{d}v. \end{aligned}$$

  (Who would've thunk it?? This relationship between the mean and the CDF holds for any nonegative random variable!! Did you learn that in STAT 101?)

- The *expected remaining life* (as a function of t) is:

$$\begin{aligned} \mathrm{E}(T|t) &= \frac{\int_t^\infty (v-t)f(v)\mathrm{d}v}{S(t)} \\ &= \frac{\int_t^\infty S(v)\mathrm{d}v}{S(t)}. \end{aligned}$$

- The hazard function and the survival function:

$$\begin{aligned} \lambda(t) &= \frac{f(t)}{S(t)} \\ &= -\frac{\partial \log(S(t))}{\partial t}. \end{aligned}$$

- The cummulative hazard function and the survival function:

$$\Lambda(t) = -\log(S(t)).$$

Models with Covariates

There are two different approaches we can take to model the effects of covariates, $\underline{x}$, on T. (Here, $\underline{x}$ is a vector, and T is a scalar.)

In **one approach**, we first transform T by its log, and model $\log(T)$ as we would in ordinary regression modelling, writing

$$\log(T) \approx g(\underline{x}, \underline{\theta}). \tag{4}$$

Often, of course, $\underline{x}$ and $\underline{\theta}$ are vectors of the same dimension, and we use a linear function:

$$\log(T) = \mu + \underline{\theta}^{\mathrm{T}}\underline{x} + \sigma W. \tag{5}$$

(Here, W is a random variable respresenting a component of an "error" that we have scaled by a positive constant, σ.) This model, of course, introduces the effects of the covariates in the survival function, hazard function, and other functions describing the distribution of T.

In the **other approach**, we model the effects of the covariates directly on the function that describe the distribution of T. We usually do this directly on the hazard function, but of course this affects all of the other functions, such as the survival function, the cummulative hazard, and so on. We use models conditioned on $\underline{x}$; that is, we use models for $\lambda(t|x)$, $S(t|\underline{x})$, and so on, often in term of a baseline function for $\underline{x} = \underline{0}$.

There are two obvious ways of doing this. One is to incorporate some function of the covariates as a multiplier,

$$\lambda(t|\underline{x}) = \lambda_0(t)g_m(\underline{x}, \underline{\theta}), \tag{6}$$

(these are called "multiplicative hazard rate models"), and the other is to model an additive effect,

$$\lambda(t|\underline{x}) = \lambda_0(t) + g_a(\underline{x}, \underline{\theta}). \tag{7}$$

The Linear Model for the Log and the Accelerated Failure-Time Model

The transformation $\log(T)$ leads to a nonconstant scaling of the survival time. The effect in the survival function in the linear model, for example, is

$$\begin{aligned} \Pr(T > t \mid \underline{x}) &= \Pr(\log(T) > \log(t) \mid \underline{x}) \\ &= \Pr(\mu + \sigma W > \log(t) - \underline{\theta}^{\mathrm{T}}\underline{x} \mid \underline{x}) \\ &= \Pr(\mathrm{e}^{\mu+\sigma W} > t\mathrm{e}^{-\underline{\theta}^{\mathrm{T}}\underline{x}} \mid \underline{x}) \\ &= S_0(t\mathrm{e}^{-\underline{\theta}^{\mathrm{T}}\underline{x}}), \end{aligned}$$

where S_0 is the survival function when $\underline{x} = \underline{0}$; that is, the survival function with time transformed as $T = \mathrm{e}^{\mu+\sigma W}$.

Because the linear model for $\log(T)$ effectively leads to a scaling of the survival or failure time, the model is sometimes called an "accelerated failure-time model". (The transformation could, of course, lead to a "deceleration" in the failure times.) Although the scaling factor shown here is most common, we could in general write the survival function in an accelerated failure-time model as

$$S_\alpha(t) = S_0(\alpha t). \tag{8}$$

All of the other functions we use in survival analysis, f, λ, and so on, for $t|\underline{x}$ could be written in terms of a baseline function with a scaled argument.

Notes

The model in equation (4) could be written with an explicit additive error term, similar to, but more general than, equation (5):

$$\log(T) = g(\underline{x}, \underline{\theta}) + E, \tag{9}$$

where E is some random variable with mean 0. This more general formulation also leads to an accelerated failure-time model. In this case, time is transfromed as $T = \mathrm{e}^{E}$, and the multiplier in equation (8) is $\alpha = \mathrm{e}^{-g(\underline{x}, \underline{\theta})}$.

The relationship between the distributional family of T and that of E in equation (9) or of W in equation (5) is of interest.

Consider the case where T follows a Weibull distribution. Its survival function (equation (3)) is

$$S(t) = \exp(-\alpha t^{\gamma}).$$

Let $Y = \log(T)$, and make the substitution of variables in the CDF to get its survival function:

$$S(y) = \exp(-\alpha \mathrm{e}^{\gamma y}). \tag{10}$$

Now if we set $\sigma = 1/\gamma$ and $\mu = -\sigma \log(\alpha)$, we have

$$S(y) = \exp(-\mathrm{e}^{(y-\mu)/\sigma}),$$

or with $w = (y - \mu)/\sigma$,

$$S(y) = \exp(-\mathrm{e}^{w}),$$

which is the survival function of an extreme value random variable (that is, it is 1 minus the CDF of an extreme value random variable). With $W = (\log(T) - \mu)/\sigma$, we can write

$$\log(T) = \mu + \sigma W,$$

which is equation (5) with $\underline{x} = \underline{0}$. (Of course we could have introduced $\underline{\theta}^{\mathrm{T}}\underline{x}$ along with the μ above. How did we know to make the substitutions above? We recognized equation (10) as being of the general form of a survival function (or CDF) of an extreme value random variable.)

The usefulness of a model depends on its fidelity to observational data and on its simplicity. The linear model in $\log(T)$ is limited by the fact that there are not many well-known, simple pairs of distributions for the variable of interest T and the variable E in equation (9). One other simple pair is the log-logistic (for T) and the logistic (for W or E/σ).

In practical applications, it is usually easier to think of reasonable probability models for T, and we often do this in terms of the behavior of the hazard function.

Proportional Hazards Models

For the multiplicative hazard rate models we have the interesting relationship for two values of the covariates, $\underline{x}_1$ and $\underline{x}_2$:

$$\begin{aligned}\frac{\lambda(t|\underline{x}_1)}{\lambda(t|\underline{x}_2)} &= \frac{\lambda_0(t)g_m(\underline{x}_1,\underline{\theta})}{\lambda_0(t)g_m(\underline{x}_2,\underline{\theta})} \\ &= \frac{g_m(\underline{x}_1,\underline{\theta})}{g_m(\underline{x}_2,\underline{\theta})}.\end{aligned}$$

Multiplicative hazard rate models are therefore often called proportional hazards or PH models.

Using the relationship

$$S(t) = \exp\left(-\int_0^x \lambda(v)\mathrm{d}v\right)$$

that holds for any continuous distribution from the fact that $\lambda(t) = -\partial \log(S(t))/\partial t$, we have

$$S(t|\underline{x}) = S_0(t)^{g_m(\underline{x},\underline{\theta})}$$

for any multiplicative hazard rate model with a continuous distribution. Moreover, this property of the survival function characterizes a multiplicative hazard rate model. (For discrete distributions, we must start with slightly different definitions of the basic functions.)

For simplicity, the arguments of the function g_m are usually combined into a single dot product: $\underline{\theta}^{\mathrm{T}}\underline{x}$.

Furthermore, the function g_m is often taken as the exponential: $\exp(\underline{\theta}^{\mathrm{T}}\underline{x})$.

Weibull Models

In the accelerated failure-time model with a scaling factor $g(\underline{x}, \underline{\theta})$ that depends on the covariates, the Weibull yields

$$\begin{aligned} S(t|\underline{x}) &= S_0(tg(\underline{x}, \underline{\theta})) \\ &= \exp(-\alpha(tg(\underline{x}, \underline{\theta}))^\gamma) \\ &= (\exp(-\alpha t^\gamma))^{g(\underline{x}, \underline{\theta})^\gamma} \\ &= (S_0(t))^{g(\underline{x}, \underline{\theta})^\gamma} \end{aligned} \tag{11}$$

This is exactly the expression we have for the multiplicative hazard rate model (see above). This form can only occur if the survival function is of the form $\exp(a(t-c)^b)$, for some constants a, b, and c; hence, the equivalence of a multiplicative hazard rate model and an accelerated failure-time model is a characterization of the Weibull distribution (that is, only the Weibull has this property). (Notice that the requirement from equations (11) is that the survival function have the property that for some b, $S_0(rt) = (S_0(t))^{r^b}, \ \forall r$.)