Continuous Time Survival Analysis

In continuous time survival analysis we are interested in the distribution of T, which is taken to be a (non-negative) continuous random variable. In general (outside of survival analysis) we are used to characterizing the distribution of a random variable either through the distribution function (CDF), given by

$$F(t) = P(T \leq t),$$

or alternatively by the density function

$$f(t) = \frac{\partial}{\partial t} F(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq T + \Delta t)}{\Delta t}.$$

Note that this also gives the relationship

$$F(t) = \int_0^t f(s) ds.$$

For survival analysis, we add to consideration the survivor function, $S(t)$ and the hazard function $h(t)$. We take the definitions to be

$$S(t) = P(T \geq t),$$

and

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t}.$$

We have seen before that $S(t) = 1 - F(t)$, that $h(t) = \frac{f(t)}{S(t)} = \frac{-S'(t)}{S(t)}$, that $f(t) = -S'(t)$, and that $S(t) = \exp(-\int_0^t h(s) ds)$. As a result, all of these quantities can be interchanged with another; if you know one of them, the rest can be uniquely determined.

Likelihood Construction Under Censoring

Recall that we take $X_i = \min\{T_i, C_i\}$, and set $\delta_i = 1$ if $T_i > C_i$. If we wish to construct the likelihood for these data, we need to consider the joint modelling of T_i and C_i. For an individual that has $\delta_i = 1$ with an observed event time of t_i, we can write down the likelihood contribution to the joint process as

$$L_i = P(T_i \in (t_i, t_i + dt_i), C_i > t_i),$$

where we use the notation $T_i \in (t_i, t_i + dt_i)$ to represent the infinitesimal interval around t_i (e.g. $\lim_{\Delta t \to 0} t < T < t + \Delta t$). If, however, we observe an individual with $\delta_i = 0$ and c_i as the observed censoring time, the likelihood contribution will be

$$L_i = P(T_i > c_i, C_i \in (c_i, c_i + dc_i)).$$

As a result, in general, assuming that x_i is the observed time, we can take

$$L(\theta, \phi) = \prod_{i=1}^{n} L_i = \prod_{i=1}^{n} P(T_i \in (x_i, x_i + dt_i), C_i > x_i)^{\delta_i} P(T_i > x_i, C_i \in (x_i, x_i + dx_i))^{1-\delta_i}.$$

This likelihood expression, in general, requires us to model the joint event-censoring process. This is not typically desirable. Often we are able to make the assumption that $T_i \perp C_i$, and as a result, we can separate these out. If we take $g(t_i; \phi)$ to be the density and $\mathcal{G}(t_i; \phi)$ to be survivor function for the censoring process, then note that

$$\begin{aligned} P(T_i \in (x_i, x_i + dx_i), C_i > x_i) &= [f(x_i; \theta) dx_i] \, \mathcal{G}(x_i; \phi) \\ P(T_i > x_i, C_i \in (x_i, x_i + xt_i)) &= S(x_i; \theta) \, [g(x_i; \phi) dx_i] \end{aligned}$$

Plugging these values in give us the likelihood, under independent censoring, as

$$\begin{aligned} L(\theta, \phi) &= \prod_{i=1}^{n} f(x_i; \theta)^{\delta_i} \mathcal{G}(x_i; \phi)^{\delta_i} S(x_i; \theta)^{1-\delta_i} g(x_i; \phi)^{1-\delta_i} \\ &= \prod_{i=1}^{n} S(x_i; \theta) \left(\frac{f(x_i; \theta)}{S(x_i; \theta)} \right)^{\delta_i} \mathcal{G}(x_i; \phi)^{\delta_i} g(x_i; \phi)^{1-\delta_i} \\ &= \prod_{i=1}^{n} S(x_i; \theta) h(x_i; \theta)^{\delta_i} \mathcal{G}(x_i; \phi)^{\delta_i} g(x_i; \phi)^{1-\delta_i}. \end{aligned}$$

In addition to the independence assumption, we also often assume that censoring is uninformative. That is, the parameter θ and the parameter ϕ are functionally independent – knowing something about one process tells us nothing about the other. If this is the case then we do not need to consider the complete likelihood, since ϕ gives us no added information on θ and we only care about θ, generally. In this scenario, we can consider only the partial likelihood and write

$$L(\theta) \propto \prod_{i=1}^{n} S(x_i; \theta) h(x_i; \theta)^{\delta_i}.$$

If instead of assuming that censoring were random, we took it to be a fixed process, we would derive the same likelihood expression.

If we have truncation, then the likelihood necessarily depends on conditional probabilities. Suppose that we cannot observe any event that occurs before u. As a concrete example, imagine the study of survival among those individuals who were exposed to radiation from the 1945 atomic bombings of Hiroshima and Nagasaki. It is unlikely that these individuals could be studied starting in 1945, at $T = 0$, and instead seems more likely that research would have began starting some years later. If you imagine a study beginning in 1950, then we cannot observe anyone in the study who had a survival time less than 5 years. In the case of truncation, we must work on conditional inference. Here we may observe

$$L_i(\theta) = f(t_i | T > u; \theta) = \frac{f(t_i; \theta)}{S(u; \theta)}.$$

Example of Parametric Continuous Survival Analysis

Note that we have fairly strong requirements on the distribution of T, in general. We assume that $T \geq 0$, and in real applications generally assume that T is skewed, with a fairly long tail. That is because T represents the survival time, which we know must be positive, and in general can be the full lifetime of interest. The natural choice – and indeed, a choice you've probably seen in previous courses when discussing, for instance, light bulbs – is the exponential distribution. Recall that, if we take $T \sim \text{Exponential}(\rho)$, we can write

$$f(t) = \frac{1}{\rho} \exp\left(-\frac{t}{\rho}\right)$$
$$F(t) = 1 - \exp\left(-\frac{t}{\rho}\right).$$

If we wanted to find the survival function and hazard function, we could use the aforementioned relationships, and we could get that

$$S(t) = 1 - F(t) = \exp\left(-\frac{t}{\rho}\right).$$

Then, we see that

$$h(t) = \frac{-S'(t)}{S(t)} = \frac{-\left(-\frac{1}{\rho}\exp\left\{-\frac{1}{\rho}\right\}\right)}{\exp\left(-\frac{1}{\rho}\right)} = \frac{1}{\rho}.$$

This means that the exponential distribution represents a model which has a **constant hazard function**. This is a property you already know about the exponential distribution, just framed differently: a constant hazard (roughly) means that any instant there is an equal probability of the event occurring, no matter how long survival has happened thus far. This is actually a fairly restrictive assumption, and we will see models that relax this is a natural way afterwards, but an exponential distribution is fairly easy to work with, and meets the minimum requirements for a distribution for use in survival analysis.

Taking the likelihood result from above, and defining $\theta = \rho$, we get that

$$L(\theta) \propto \prod_{i=1}^{n} \left(\frac{1}{\theta}\right)^{\delta_i} \exp\left(-\frac{x_i}{\theta}\right)$$
$$\implies \ell(\theta) = \sum_{i=1}^{n} \log\left(\frac{1}{\theta}\right)^{\delta_i} - \frac{x_i}{\theta}$$
$$= -\log\theta \sum_{i=1}^{n} \delta_i - \frac{1}{\theta}\sum_{i=1}^{n} x_i.$$

This leads us to the Score equations given by

$$\frac{\partial}{\partial\theta}\ell(\theta) = -\frac{1}{\theta}\sum_{i=1}^{n}\delta_i + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i.$$

As a result, the MLE for θ is given by

$$\widehat{\theta} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} \delta_i},$$

which gives the MLE of the hazard to be $\frac{1}{\widehat{\theta}}$. This is the number of observed events divided by the sum of the observed times.

Location-Scale Families

As discussed above, families which are appropriate for time to event data are fairly restricted. What's more, distributions that are constrained to meet those criteria, while plentiful, are very often challenging to work with. It turns out, however, that there is often a relationship between suitable distributions and location-scale families.

In general, a location-scale family is defined through closure under addition and multiplicative scaling by constants. That is, Y belongs to a location scale family if $Y \stackrel{d}{=} \mu_Y + \sigma_Y X$ where X is a random variable with 0 mean and variance equal to 1, where Y and X to belong to the same family. Some common location-scale families include:

- Normal Distribution;
- Continuous uniform distribution;
- Logistic distribution;
- t-distribution;
- Extreme Value Distribution.

Several of the popular distributions for T are such that if we use a log-transformation, then $Y = \log T$ will follow a location-scale family. For instance, consider the exponential distribution. Take $f(t) = \frac{1}{\rho}\exp\left(-\frac{t}{\rho}\right)$. If we take $Y = \log T$, then a simple change of variable transformation shows us that $f(y) = f(e^y)e^y = \exp\left(y - \log\rho - e^{y-\log\rho}\right)$. This is an **extreme value** distribution, with parameter $\log\rho$ and scale parameter $\sigma = 1$.

The **Weibull** distribution is a two-parameter generalization of the exponential distribution, with density function given by

$$f(t) = \frac{\kappa}{\rho}\left(\frac{t}{\rho}\right)^{\kappa-1}\exp\left\{-\left(\frac{t}{\rho}\right)^{\kappa}\right\}.$$

Here if $\kappa = 1$, then we have the exponential distribution. In general, a similar process shows us that $Y = \log T$ where $T \sim \text{Weibull}(\kappa, \rho)$ will be expressible as $\mu + \sigma W$, where $W \sim \text{EV}(0, 1)$, $\mu = \log\rho$, and $\sigma = \kappa^{-1}$.

Distribution of T	Distribution of Y	μ	σ
Exponential(ρ)	Extreme Value	$\log\rho$	1
Weibull(κ, ρ)	Extreme Value	$\log\rho$	κ^{-1}
Log-Logistic(κ, ρ)	Logistic	$\log\rho$	κ^{-1}
Log-Normal(μ, σ)	Normal	μ	σ

This means that if we assume that T follows one of the parametric distributions that was listed, then considering the transformed random variable given by $Y = \log T$, we are able to model Y as $\mu + \sigma W$, where W is a mean-zero error term (with unit variance), μ is the mean of Y and σ is the variance. This framework naturally lends to regression-type models for survival data, when such a parametric assumption is appropriate.