Non-Parametric Discrete Time Estimation

Recall that the discrete-time hazard function, denoted $h(k)$, represents the probability of an event occurring at a given time k, supposing that it had not yet occurred. That is, $h(k) = P(T = k|T \geq k)$. Also note that, in the discrete time setting, we are envisioning any events that occur in $[k, k+1)$ to be defined as "happening at k". If we let d_k denote the number of observed events happening at time k, and we let r_k denote the number of individuals who were at risk at the start of k, then an obvious estimator for the hazard function is

$$\widehat{h}(k) = \frac{d_k}{r_k}.$$

This estimator is quite similar to the approach we took for transition probabilities, and indeed, can be derived in a similar way.

Note that this estimator implicitly takes into account censoring of individuals. The reason is that $r_{k+1} \neq r_k - d_k$, and the estimator recognizes this. Instead, if c_k individuals are censored at time k (meaning they are, for instance, lost to follow-up in the study) then $r_{k+1} = r_k - d_k - c_k$. In this way, the denominator of this estimator accounts for censoring. Note that we are making a subtle assumption about censoring in this formulation, and it is one that we will continue to make. We are assuming that at each stage we first observe whether or not an event occurs, and then we observe whether or not someone is censored. In this sense, if $T_i = C_i$, then we actually take $C_i = T_i^+$, some time just after T_i, and count their event as being recorded. Put differently, you cannot be censored and have an event observed during the same time interval.

After considering the hazard function, it also makes sense to try to estimate the survivor function. One estimator that seems reasonable at face value would be to take

$$\widehat{S}(k) = \frac{r_k}{n},$$

where n is the total population. This represents the proportion of the total individuals still at risk at time k, which is a reasonable sounding estimator. The problem is that the individuals who are no longer at risk $(n - r_k)$ may have been those who experience the event $(\sum_{j=1}^{k} d_j)$, which is what we care about, or they may have been those that were censored $(\sum_{j=1}^{k} c_j)$. As a result, under right censoring this estimator will not be valid for the quantity of interest, as it will **underestimate** the true probability of surviving beyond a certain point.

We saw previously that

$$S(k) = P(T > k) = \prod_{j=1}^{k} P(T \geq j|T > j) = \prod_{j=1}^{k}(1 - h(j)).$$

This gives motivation for an estimator for the survivor function, based on the discrete hazard function. Namely, we can take

$$\widehat{S}(k) = \prod_{j=1}^{k}(1 - \widehat{h}(j)).$$

This ends up being a valid estimator for the survivor function, even in the event of censoring.

If we want to take a measure of central tendency, or to provide answers to questions of the "average" survival time, sample means are not valid under censoring. If right censoring is present then the sample mean will underestimate the true average survival time, since the denominator will be inflated. Getting an accurate estimate, without parametric assumptions, of the expected value under censoring is not possible. Instead, we will typically turn to the median, which can be estimated from the survivor function. If we take m to be a value such that $\widehat{S}(m) > 0.5 > \widehat{S}(m+1)$, then we can take

$$\text{median} = m + \left[\frac{\widehat{S}(m) - 0.5}{\widehat{S}(m) - \widehat{S}(m+1)}\right].$$

This gives a non-parametric estimator for the measure of central tendency.

Stochastic Processes and Partial Likelihood

Just as we did with transition models, it can be helpful to think of survival data framed as a stochastic process. We take $Y(s)$ to represent the state of the process at time s, where $Y(s) = 0$ means that the event has not yet happened, and $Y(s) = 1$ means that the event has occurred. In this sense we are concerned only with the probabilities of moving from 0 to either 0 or 1 (as 1 acts as an absorbing state). Using this notation we can define the hazard function as

$$h(s) = P(Y(s) = 1|Y(s-1) = 0).$$

Alongside the event process, we can also define a censoring process. Let $Z(s) = 1$ if the individual is still under observation at time s and $Z(s) = 0$ otherwise. We will also borrow the history notation from transition models, where $\mathcal{H}^Y(s)$ is the history vector for all occurrences of Y up to (not including) s, and $\mathcal{H}^Z(s)$ is the same for Z. Using this notation, we can write down the likelihood contribution for a single individual as

$$L = \prod_{s=1}^{\infty} P\left\{Y(s), Z(s)|\mathcal{H}^Y(s), \mathcal{H}^Z(s)\right\} = \prod_{s=1}^{\infty} P\left\{Y(s)|Z(s), \mathcal{H}^{Y,Z}(s)\right\} P\left\{Z(s)|\mathcal{H}^{Y,Z}(s)\right\}.$$

There are two commonly made assumptions that are made to work with this quantity more easily.

1. **Conditionally Independent Censoring:** Given the event history, the probability of an event occurring is independent of the censoring process. That is

$$P(Y(s)|\mathcal{H}^{Y,Z}(s), Z(s)) = P(Y(s)|\mathcal{H}^Y(s)).$$

2. **Non-Informative Censoring:** The process of censoring provides no information to (or shares no parameters with) the event process. That is $P(Z(s)|\mathcal{H}^{Y,Z}(s))$ and $P(Y(s)|\mathcal{H}^Y(s))$ are functionally independent of one another.

Under these two assumptions, we can simplify these terms and consider only the **partial likelihood** for each individual, given by

$$L = \prod_{s=1}^{\infty} P(Y(s)|\mathcal{H}^Y(s)).$$

We used conditionally independent censoring to simplify the first term, and non-informative censoring to drop the second (since no common parameters are in it). We do not lose any efficiency by dropping this component *if* non-informative censoring holds; if the assumption is violated, this partial likelihood is still valid, it is simply not efficient. If conditionally independent censoring is violated, however, the expression is no longer valid.

Note that the term $P(Y(s)|\mathcal{H}^Y(s))$ can be decomposed based on the hazard function. If the individual observed $Y(s) = 1$ while $Y(s-1) = 0$, then this quantity is captured exactly by $h(s)$. If $Y(s) = 0$ while $Y(s-1) = 0$, then the contribution will be $1 - h(s)$. If we fix their censoring time to be C_i, then we can exploit this to write

$$L = \prod_{s=1}^{C_i} (1 - h(s))^{1-Y(s)} h(s)^{Y(s)}.$$

In practice, with a dataset of n individuals, that is going to give

$$L(\theta) = \prod_{i=1}^{n} \prod_{s=1}^{C_i} (1 - h_i(s;\theta))^{1-Y_i(s)} h_i(s;\theta)^{Y_i(s)},$$

where the form of the hazard function depends on some unknown parameter θ. Note that this looks **exactly** like the likelihood expression for a binomial random variable, with probabilities given by $h_i(s;\theta)$ – this is going to be our key to estimating transition survivor distribution with covariates!

In particular, if we specify a standard logistic regression model (say) for $\text{logit}(h(s))$ then this model can be fit using standard statistical software. We can take the estimated probabilities as $\widehat{h}(s)$, and use these to build $\widehat{S}(s)$.

Confidence Intervals for Survivor Function

While the confidence intervals output from the GLM will suffice for the hazard function, attempting to estimate a confidence interval for $\widehat{S}(t)$ is more challenging. Doing so relies on the multivariate delta method, and we typically use a log transformation. We will ignore the specific details, but in broad strokes:

1. Consider $\log \widehat{S}(t)$, since this will be given by $\sum_{j=1}^{t} \log(1 - \widehat{h}(j))$.

2. Note that

$$\widehat{\text{var}} \begin{pmatrix} \log \widehat{S}(1) \\ \vdots \\ \log \widehat{S}(C) \end{pmatrix} \approx \widehat{G}\widehat{\text{var}}(\widehat{\alpha})\widehat{G}',$$

where $\widehat{\text{var}}(\widehat{\alpha})$ is the estimated covariance matrix from the GLM, and

$$\widehat{G} = \begin{bmatrix} -\widehat{h}(1) & 0 & 0 & \cdots & 0 \\ -\widehat{h}(1) & -\widehat{h}(2) & 0 & \cdots & 0 \\ -\widehat{h}(1) & -\widehat{h}(2) & -\widehat{h}(3) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\widehat{h}(1) & -\widehat{h}(2) & -\widehat{h}(3) & \cdots & -\widehat{h}(C) \end{bmatrix}.$$

3. Using this approximation, compute confidence intervals for $\log \widehat{S}(t)$ based on a normal approximation.

4. Exponentiate the confidence intervals for the survivor function.