Competing Risks

To motivate competing risks, consider how the hazard rate can be expressed in terms of competing risks. Assume that there are $K, k = 1, 2, \ldots r$ possible events that an observation is at risk of experiencing. Accounting for the K events in the hazard rate gives

$$h_k(t \mid \mathbf{x}) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T \leq t + \Delta t \mid T \geq t, \mathbf{x})}{\Delta t}, \tag{1}$$

which differs from the hazard rate for a single-state process because the rate is subscripted for each of the k events that could occur. As such, models used for competing risks problems typically model the type-specific hazard rate.

Latent Survivor Time Approach to Competing Risks

A commonly applied approach to the competing risks problem assumes that there are $K, k = 1, 2, \ldots r$ specific outcomes or destination states and that there is assumed to exist a *potential* or latent failure time associate with each outcome.

For K outcomes, there are theoretically $T_k = (T_1, T_2, \ldots, T_r$ duration times but only the shortest duration time is actually observed; that is, $T_k = \min\{T_1, T_2, \ldots, T_r\} = T_C$, where T_C is the duration time associated with observed event or "cause" of failure.

The basic result that drives the model is that if there are k specific outcome states, then the overall survivor function can be partitioned into marginal survivor functions, each corresponding to one of the k destination states.

Suppose there are n observations at risk of failing in one of k ways. The individual contribution of the ith individual failing by event k is given by

$$\mathcal{L}_i = f_k(t_i \mid X_{ik}, \beta_k) \prod_{k \neq r} S_r(t_i \mid X_{ir}, \beta_r), \tag{2}$$

where the subscript k denotes the kth event and r in the product term implies that the product is taken over the survivor times for all states except k. The likelihood function for the full sample is then given by

$$\mathcal{L} = \prod_{i=1}^{n} f_k(t_i \mid X_{ik}, \beta_k) \prod_{k=1}^{r} S_k(t_i \mid X_{ik}, \beta_k). \tag{3}$$

But since only T_C is observed (i.e. only one failure among the K possible outcomes is observed), the overall likelihood can be partitioned in terms of the number of observations failing by each of the K outcomes:

$$\mathcal{L} = \prod_{k=1}^{r} \prod_{i=1}^{n_k} f_k(t_i \mid X_{ik}, \beta_k) S_k(t_i \mid X_{ik}, \beta_j). \tag{4}$$

This partitioning is easier to see if we define a censoring indicator such that

$$\delta_{ik} = \begin{cases} 1 & \text{if } i \text{ failed due to } k \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

When $\delta_i = 1$, the observation is observed failing due to risk k (and hence we only observe T_C; when $\delta_i = 0$, the observation is right-censored. Incorporating δ_{ik} into the likelihood function, the likelihood of the sampled duration times may be expressed as

$$\mathcal{L} = \prod_{k=1}^{r} \prod_{i=1}^{n} f_k(t_i \mid X_{ik}, \beta_k)^{\delta_{ik}} S_k(t_i \mid X_{ik}, \beta_j)^{1-\delta_{ik}}. \tag{6}$$

The overall likelihood function factors into k sub-contributions, where failures due to risks other than k are treated as right-censored. In turn, $k-$specific hazards can be generated from equation (6).

Implementation of this model is straightforward because it simply requires that K models be estimated where all events other than k are treated as randomly censored.

Multinomial Logit

We've discussed the discrete-time model for single-spell processes. There is a discrete-time analog to competing risks. It is given by the multinomial logit model.

The multinominal logit (hereafter MNL) model is a series of "linked" logit models. If there are k possible events (destinations, outcomes, states) that an observation is at risk of experiencing, the MNL model estimates $k-1$ logit models to obtain parameter estimates on the type-specific or destination-specific hazards The $k-1$ logits produced by the MNL model are interpretable as logit models. Note that the hazard probability for the MNL model is

$$\lambda_(ik) = \frac{\exp \beta_k' \mathbf{x}}{\Sigma_k^K \exp(\beta_k' \mathbf{x_i})}. \tag{7}$$

If there are $k = 3$ possible outcomes, then the type specific hazard probabilities from the multinomial logit model are given by

$$\lambda(y_i = 1 \mid \mathbf{x}_i) = \lambda_{i1} = \frac{1}{1 + \exp(\beta_2' \mathbf{x_i}) + \exp(\beta_3' \mathbf{x_i})}, \tag{8}$$

for the baseline category ($k = 1$),

$$\lambda(y_i = 2 \mid \mathbf{x}_i) = \lambda_{i2} = \frac{\exp(\beta_2' \mathbf{x_i})}{1 + \exp \beta_2' \mathbf{x_i} + \exp \beta_3' \mathbf{x_i}}, \tag{9}$$

for $k = 2$, and

$$\lambda(y_i = 3 \mid \mathbf{x}_i) = \lambda_{i3} = \frac{\exp(\beta_3' \mathbf{x_i})}{1 + \exp \beta_2' \mathbf{x_i} + \exp \beta_3' \mathbf{x_i}} \tag{10}$$

for $k = 3$ (where y_i denotes the outcome variable). The form of (7) is clearly similar to the logit model. The issues pertinent to the logit version of the event history model are relevant here. The interpretation of the model is akin to what you've learned about the standard MNL model.

Stratified Cox

Another variant of the competing risks model we consider has considerable similarity to the Cox conditional risk set model discussed previously. This model assumes observations are perpetually at risk. So in terms of policy adoption processes, it is assumed that states are "at risk" of adopting one of m type policies.

Imagine there are $k = 5$ possible (unordered) events that could occur to an observation. Because there are no restrictions on when or in what order any of these 5 events an observation can experience, we must assume that *for each* of the 5 events, the observation is at risk for all them.

The data structure will consist of multiple records per observation with each observation having at minimum, 5 records of data—one for each possible event.

If one assumes that the covariate effects are common to each event type, but the baseline hazard for each risk is allowed to vary across risks, then a stratified Cox model can be used to estimate the parameters of interest.

The stratified Cox model, in this context, would estimate a single set of parameters for each of the event types. But if one stratifies on the different kinds of events ($k = 5$), then one can back out of the stratified Cox model a unique baseline hazard function for each of the k risks.

Split Population

The split population model is unique in that it does not assume that eventually every observation will experience the event. Instead, the model splits the population into two groups—one that will experience the event and one that will not.

It is an assumption of all duration models discussed to this point that if t is sufficiently large, then the probability of an event occurrence will approach one.

Schmidt and Witte (1984, 1988, 1989) developed the so-called "split population" model.

Note that there are a wide variety of split-population models out there!

Basic Schmidt and Witte model. Assume there is some unobserved variable indicating whether or not an observation will eventually experience an event. Letting Z denote this variable, then let

$$\Pr(Z = 1) = \delta$$

and

$$\Pr(Z = 0) = 1 - \delta,$$

where δ is a parameter indicating the estimated probability that observations experience an event. In the case of criminal recidivism, δ is the estimated probability an individual will return to crime. It should be clear that if $\delta = 1$, then all observations are assumed to eventually experience an event.

Next, we define a cumulative distribution function for the duration times of individuals who eventually experience the event and denote this as

$$F(t \mid Z = 1), \tag{11}$$

with corresponding density

$$f(t \mid Z = 1).$$

The distribution function in (11) is conditional on whether or not an observation eventually fails. As Schmidt and Witte (1989, 148) note, this distribution function is irrelevant for individuals for whom $Z = 0$.

Since we do not directly observe Z, Schmidt and Witte (1989) define an indicator R, which is a dummy variable, to denote whether or not an observation experienced an event. For those who experience the event, the probability density is given by

$$\Pr(Z = 1) f(t \mid Z = 1) = \delta f(t \mid Z = 1), \tag{12}$$

and for those who do not experience an event, i.e., for whom $Z = 0$, Schmidt and Witte (1989, 148) define the probability that an event is never observed as:

$$\begin{aligned} \Pr(R = 0) &= \Pr(Z = 0) + \Pr(Z = 1) \Pr(t > T \mid Z = 1) \\ &= 1 - \delta + \delta[1 - F(T \mid Z = 1)]. \end{aligned} \tag{13}$$

Under this model, the log-likelihood function for the full data is obtained by combining (12) and (13) to obtain

$$\begin{aligned} \log L = \sum_{i=1}^{N} & R_i[\log \delta + \log f(t_i \mid Z_i = 1)] \\ & + (1 - R_i) \log[1 - \delta + \delta(1 - F(T_i \mid Z_i = 1))]. \end{aligned} \tag{14}$$

The appeal of the log-likelihood expressed in this way is that observations that never experience an event by the end of the observation period contribute information only to the second part of the function. As such, the log-likelihood "splits" the two populations, hence the name of the model.

This result is conceptually similar to the log-likelihoods derived earlier for other models. The basic difference here is that one is explicitly accounting for the fact that some observations may *never* experience an event, no matter how long t is.

The log-likelihood in (14) may be parameterized in terms of any of the log-linear parametric models discussed in this class. The only additional complication is the presence of the δ parameter in the log-likelihood. Schmidt and Witte (1989, 151) propose augmenting the model by estimating δ as a logit model. Practically, this is accomplished by estimating

$$\Pr(R_i = 1) = \frac{1}{1 + \exp^{\beta'\mathbf{x}}}, \tag{15}$$

which is a logit model where $\beta'\mathbf{x}$ are the covariates and parameters. The model in (15) will give an estimate of δ. If δ is not significantly different from 1, then a model not accounting for the "splitting" is obtained. Thus, if one estimates a "splitting Weibull" model and δ is not different from 1, then the model reduces to a standard Weibull duration model.

Important: this is the Schmidt and Witte version. Other variants exist! Many (most?) are conceptually similar but statistically different. Currently (still) Limdep is the only package with canned modules to estimate this model directly.