3004
Multiple Hazards

F. P. Treasure

May 22, 2002
Multiple Hazards

Two Sources of Failure

Introduction
An individual may be exposed to two (or more) events simultaneously.

Example: a patient with kidney disease may be exposed simultaneously to the events ‘left kidney failure’ and ‘right kidney failure’.

Example: a patient with cancer is exposed simultaneously to ‘death from cancer’ and ‘death from another cause’.

Here we are interested in modelling the time to the first event using survival analysis techniques. The time from the first event to the second event can be modelled using the usual methods (kidney example) or not at all (second example).

Notation
The two events of interest are $A$ and $B$. The corresponding hazards are $h_A$ and $h_B$. The probability that event $A$ is observed at or before time $t$ is $G_A(t)$, (similarly for $B$). We are often particularly interested in, say, $G_A(\infty)$, which is the probability that event $A$ happens first. Corresponding to $G_A(t)$, $G_B(t)$ are (improper) densities $g_A(t)$, $g_B(t)$ with, typically,

$$G_A(t) = \int_0^t g_A(u) \, du$$

We use the usual (unsubscripted) notation: $f$, $F$, $h$ and $H$ to refer to the composite event ‘first of $A$ or $B$’; so

$$F(t) := \mathcal{P}\{\text{first of } A \text{ or } B > t\}$$

for example.

Parametric Estimation

Relationships between hazards and probabilities
The relationship between the various hazards is simple. We have

$$h = h_A + h_B$$

and

$$H = H_A + H_B.$$
The density $g_A$ can be informally defined by:

$$g_A(t) \delta \simeq \mathbb{P}\{\text{event } A \text{ in } [t, t + \delta]\}$$

with small $\delta$. Using conditional probabilities we can write:

$$g_A(t) \delta \simeq \mathbb{P}\{\text{no event in } [0, t]\} \times \mathbb{P}\{\text{event } A \text{ in } [t, t + \delta]\} | \text{no event in } [0, t]\}$$

or, by definition,

$$g_A(t) = F(t) h_A(t).$$

Using the standard relation $F = \exp(-H)$ we obtain

$$G_A(t) = \int_0^t h_A(u) \exp[-H_A(u) + H_B(u)] du.$$  

In practice, this integral cannot be algebraically integrated (but consider the special case $h_B(t) = \lambda h_A(t)$ with $\lambda$ a constant).

**Maximum Likelihood**

If $h_A, h_B$ depend on a vector $z$ of explanatory variables through a parameter vector $\theta$ then an estimate $\hat{\theta}$ of $\theta$ can be obtained using standard likelihood techniques.

For the $i$th individual we observe $x_i$ the time of event or censoring and $v^A_i, v^B_i$ the visibility indicators. We define

$$v^A_i := I\{x_i \text{ corresponds to event } A\}$$

and similarly for event $B$.

If $v^A_i = v^B_i = 0$ then the $i$th individual contributes $F(x_i)$ to the likelihood. If $v^A_i = 1$ then the $i$th individual experienced event $A$ at time $x_i$ and so that individual contributes $g_A(x_i)$ to the likelihood; similarly for event $B$. The functions $F, g_A$ and $g_B$ are expressed in terms of $\theta$ and the likelihood is maximized to obtain $\hat{\theta}$.

**Non-Parametric Estimation**

**Reminder of Kaplan-Meier**

Kaplan-Meier methods are easily extended to multiple events. The basis of the (single event) Kaplan-Meier calculation is to account for the individuals at each event or censoring time $a_j$: there are $r_j$ individuals at risk at time $a_j$ of which $c_j$ are censored, $d_j$ experience the event and $r_{j+1}$ individuals become the risk set for time $a_{j+1}$. The estimator is obtained from:

$$\hat{P}\{\text{event at } a_j | \text{no event before } a_j\} = d_j/r_j.$$
Extension of Kaplan-Meier

If there are two events of interest, A and B then the \( r_j \) individuals at risk at time \( a_j \) divide into \( c_j \) censored, \( d_j^A \) events of type A, \( d_j^B \) events of type B and the remaining \( r_j+1 \) individuals becoming the risk set at time \( a_{j+1} \). The basic formula is now:

\[
\hat{P}\{\text{event A at } a_j | \text{no event before } a_j \} = \frac{d_j^A}{r_j}.
\]

Non-parametric estimation of \( G_A \)

Using conditional probabilities we can write

\[
G_A(a_j) = \mathbb{P}\{\text{event } A \leq a_j\};
\]

\[
= \sum_{k=1}^{j} \mathbb{P}\{\text{event } A = a_k\};
\]

\[
= \sum_{k=1}^{j} \mathbb{P}\{\text{event A at } a_k | \text{no event before } a_k\} \mathbb{P}\{\text{no event before } a_k\}.
\]

From the Kaplan-Meier argument we have

\[
\hat{P}\{\text{event A at } a_k | \text{no event before } a_k\} = \frac{d_k^A}{r_k}
\]

and it is easy to obtain from this

\[
\hat{P}\{\text{no event before } a_k\} \approx \prod_{l=1}^{k-1} \left(1 - \frac{d_l^A + d_l^B}{r_l}\right)
\]

. The estimate for \( G_A(a_j) \) is therefore given by

\[
\hat{G}_A(a_j) = \sum_{k=1}^{j} \frac{d_k^A}{r_k} \prod_{l=1}^{k-1} \left(1 - \frac{d_l^A + d_l^B}{r_l}\right).
\]