3005
Counting Processes

F. P. Treasure

March 12, 2003

Counting Processes

This is an account aimed at applied statisticians, not mathematical probabilists.

Notation by example

$f(t-)$ is shorthand for $\lim_{\delta \downarrow 0} f(t-\delta)$.
$[a, b[$ is short for the interval $a \leq x < b$.
$\mathcal{I}\{A\} = 1$ if A is true; $= 0$ otherwise.
$\mathcal{P}\{\ldots\}$ is probability, $\mathcal{E}\{\ldots\}$ is expectation and $\mathcal{V}\{\ldots\}$ is variance.
Both $a := B$ and $B =: a$ mean that we are defining a to mean B.

Introduction

Definition

The random variable $N(t)$ represents a counting process on $[0, \infty[$ if

1. $N(t)$ is a non-negative integer;
2. $N(s) \leq N(t)$ for $s < t$;
3. $\mathrm{d}N(t) = N(t) - N(t-)$ is either 0 or 1;
4. $\mathcal{E}N(t) < \infty$.

Normally we take $N(0) = 0$.

History

The *history* $\mathcal{H}_t$ (properly *filtration*) of a counting process is all that is known at time t. In particular the history includes the values of random variables known up to and including time t. $\mathcal{H}_{t-}$ represents what is known up to but not including time t.

Intensity Function

The probability (conditional on the history) of $\mathrm{d}N(t) = 1$ at any time can be written in terms of an *intensity* $\lambda(t)$:

$$\mathcal{P}\{N(t+\delta) - N(t-) = 1 | \mathcal{H}_{t-}\} \simeq \lambda(t)\delta.$$

or, equivalently:

$$\mathcal{P}\{\mathrm{d}N(t) = 1 | \mathcal{H}_{t-}\} = \mathrm{d}\Lambda(t)$$

where

$$\Lambda(t) = \int_0^t \lambda$$

is the *integrated intensity*.

Predictability

$\Lambda(t)$ is required to be *predictable* with respect to $\mathcal{H}_t$: that is, $\Lambda(t)$ is known given $\mathcal{H}_{t-}$. In practice, this means that $\Lambda(t)$ has to be continuous.

Expectations and Martingales

The probability can be converted into an expectation:

$$\mathcal{E}\{\mathrm{d}N(t)|\mathcal{H}_{t-}\} = \mathrm{d}\Lambda(t)$$

and using the predictability of Λ:

$$\mathcal{E}\{\mathrm{d}N(t) - \mathrm{d}\Lambda(t)|\mathcal{H}_{t-}\} = 0.$$

We now define $M(t)$ by

$$M(t) = N(t) - \Lambda(t)$$

and the expectation becomes

$$\mathcal{E}\{\mathrm{d}M(t)|\mathcal{H}_{t-}\} = 0.$$

showing, as far as we are concerned, that $M(t)$ is a *Martingale*.

Compensation

The counting process $N(t)$ can now be written as a *Doob-Meyer decomposition*:

$$N(t) = \Lambda(t) + M(t)$$

where $\Lambda(t)$ is the *compensator* of the process.

Survival Analysis

A single individual

A survival analysis is a time-to-event analysis where an individual can only experience one event. The time-to-event is a random variable T. The counting process $N(t)$ represents whether or not the event has happened by or at t:

$$N(t) = \mathcal{I}\{T \leq t\}.$$

The intensity $\lambda(t)$ is equal to the hazard $h(t)$ when the individual is at risk of the event and equal to zero when the event has happened. We express this by writing the intensity as

$$\lambda(t) = Y(t)h(t)$$

where $Y(t)$ is an indicator for being at risk:

$$Y(t) = \mathcal{I}\{T \geq t\}.$$

Note that $Y(t) = 1 - N(t-)$ for an uncensored individual.

Several individuals

The ith of a set of n individuals has a counting process $N_i(t)$, a compensator $\Lambda_i(t)$, a Martingale process $M_i(t)$, and a risk indicator $Y_i(t)$.

These can all be summed over the n individuals to give $N_+(t)$, $\Lambda_+(t)$, $M_+(t)$ and $Y_+(t)$. $N_+(t)$ is a counting process, $\Lambda_+(t)$ its compensator and $M_+(t)$ its Martingale:

$$N_+(t) = \Lambda_+(t) + M_+(t). \tag{1}$$

The compensator can be written in terms of the individual hazards $h_i(t)$:

$$\Lambda_+(t) = \int_0^t \sum_{i=1}^n Y_i(u) h_i(u) \mathrm{d}u. \tag{2}$$

Estimating the Integrated Hazard: Nelson-Aalen

The Compensator and the Integrated Hazard

If all individuals are exposed to the same hazard then the expression for the overall compensator reduces to:

$$\Lambda_+(t) = \int_0^t Y_+(u) h(u) \mathrm{d}u = \int_0^t Y_+(u) \mathrm{d}H(u).$$

where H is the integrated hazard.

Estimation

The decomposition (1) expressed in differentials becomes

$$\mathrm{d}N_+(t) = Y_+(t)\mathrm{d}H(t) + \mathrm{d}M_+(t).$$

$N_+(t)$ and $Y_+(t)$ are the data. Conditional on the history $\mathcal{H}_{t-}$, $\mathrm{d}M_+(t)$ has zero expectation. So an estimate of H can be obtained by setting $\mathrm{d}M_+(t)$ equal to zero:

$$\mathrm{d}\hat{H}(t) = \frac{\mathrm{d}N_+(t)}{Y_+(t)} \tag{3}$$

where the RHS is taken to be zero if $Y_+(t)$ is zero.

Definition of the Nelson-Aalen estimator

Let the n distinct, uncensored, event times from a set of n individuals be $a_1, \ldots, a_j, \ldots, a_n$ with $a_{j-1} < a_j$. The estimated integrated hazard from (3) is

$$\hat{H}(t) = \int_0^t \frac{\mathrm{d}N_+(u)}{Y_+(u)}.$$

The numerator of the integrand, $\mathrm{d}N_+(t)$ is zero unless $t = a_j$ for some j. $Y_+(a_j)$ is the number in the risk set at a_j, conventionally written as r_j.

The *Nelson-Aalen estimator* is therefore:

$$\hat{H}(t) = \sum_{j:a_j \le t} \frac{1}{r_j}.$$

Censored data

Censored individuals are not an problem for the Nelson-Aalen estimator.

If an individual is censored between two failure times a_{j-1} and a_j then that individual is counted in the risk sets up to and including the set at a_{j-1} but not in any subsequent ones.

If an individual is censored at a failure time a_j then that individual *is* included in the risk set for a_j but not any later ones.

Exercises

Exercise 1: Show that:

$$\sum_i \hat{H}(X_i) = \sum_i N_i(X_i).$$

(Interpretation in words: show that the sum of the estimated integrated hazards at time of event or censoring is equal to the number of observed events.)

Exercise 2: (Harder) Show that the expected sum over the integrated hazards (actual not estimated) at the time of event or censoring is equal to the expected number of observed events.

Hint: start with a single individual with fixed censoring time c. The event is only observed if the event time T is less than or equal to c. You need to show for that individual that $\mathcal{E}H(X) = \mathcal{P}\{T \le c\}$ (where $X = \min(T, c)$) and then generalize.

Nelson-Aalen: variance and confidence limits

Variance and standard error

Locally, $\mathrm{d}N(t)$ can be treated as a Poisson variable with mean and variance $\mathrm{d}\Lambda(t)$. The estimated variance of $\mathrm{d}N(t)$ is therefore equal to $\mathrm{d}\hat{\Lambda}(t)$ which is itself just $\mathrm{d}N(t)$.

The estimated variance of $\mathrm{d}\hat{H}(t)$ is therefore given by:

$$\hat{\mathcal{V}}\{\mathrm{d}\hat{H}(t)\} = \hat{\mathcal{V}}\left\{\frac{\mathrm{d}N(t)}{Y_+(t)}\right\} = \frac{\mathrm{d}N(t)}{[Y_+(t)]^2}.$$

Integration is a sum of independent increments, so:

$$\hat{\mathcal{V}}\{\hat{H}(t)\} = \int_0^t \frac{\mathrm{d}N(t)}{[Y_+(t)]^2} = \sum_{j:a_j \le t} \frac{1}{{r_j}^2} =: [s(t)]^2$$

where $s(t)$ is the *standard error* (estimated standard deviation).

Confidence interval and transformation

A $1-\alpha$ confidence interval for $H(t)$ can be constructed:

$$\left[\hat{H}(t) - \Phi^{-1}(1-\alpha/2)s(t), \hat{H}(t) + \Phi^{-1}(1-\alpha/2)s(t)\right]$$

where Φ is the standard Normal distribution function.

It is better, however, to find a confidence interval for $\log H(t)$ and transform back.

Exercise 3: Use the useful approximation (*law of propagation of errors*) for finding the variance of a function of a random variable:

$$\mathcal{V}\{u(X)\} \simeq [u'(\mathcal{E}X)]^2 \mathcal{V}X$$

to obtain the estimate of variance:

$$\hat{\mathcal{V}}\{\log \hat{H}(t)\} = [s(t)/\hat{H}(t)]^2$$

(with $s(t)$ defined as above) and the confidence interval for $H(t)$:

$$\left[\hat{H}(t) exp\{-\Phi^{-1}(1-\alpha/2)s(t)/\hat{H}(t)\}, \hat{H}(t) exp\{+\Phi^{-1}(1-\alpha/2)s(t)/\hat{H}(t)\}\right].$$

Nelson-Aalen: handling ties

Ties in the data

Although mathematically $\mathrm{d}N(a_j)$ cannot equal 2 (or more), in practice we do not measure time with infinite precision and so we do see ties in the data.

These ties cause peculiar difficulties with the Nelson-Aalen method. To avoid complicated notation, we demonstrate by example.

First method

A natural way to deal with, say, 2 individuals having events at a_j is to write that bit of the summation thus:

$$\cdots + \frac{1}{r_{j+1}} + \frac{2}{r_j} + \frac{1}{r_{j+1}} \cdots. \tag{4}$$

A difficulty with this approach is that the estimate $\hat{H}(t)$ for $t > a_j$ is not the same as would be obtained by substituting two distinct event times $a_j - \Delta, a_j + \Delta$ for the two tied times and letting $\Delta \downarrow 0$.

Second method

The inconsistency in the limit of the first method makes some statisticians prefer to replace (4) by

$$\cdots + \frac{1}{r_{j+1}} + \frac{1}{r_j} + \frac{1}{r_j - 1} + \frac{1}{r_{j+1}} \cdots . \tag{5}$$

so although the two events are apparently simultaneous we treat one as in fact happening before the other.

Exercise 4: Taking there to be d_j tied events at each time a_j, convert (5) to a proper formula (i.e. a summation over j).

Exercise 5: Why do we not have the same problem with ties when using the Kaplan-Meier estimator?

Nelson-Aalen and Kaplan-Meier

We write (temporarily) the Nelson-Aalen estimator of the integrated hazard as $\hat{H}_{\mathrm{NA}}(t)$ and the Kaplan-Meier estimator of the survivor function as $\hat{S}_{\mathrm{KM}}(t)$.

We can obtain the 'Nelson-Aalen' estimator of the survivor function and the 'Kaplan-Meier' estimate of the integrated hazard by

$$\hat{S}_{\mathrm{NA}}(t) = \exp\{-\hat{H}_{\mathrm{NA}}(t)\}$$

and

$$\hat{H}_{\mathrm{KM}}(t) = -\log\{\hat{S}_{\mathrm{KM}}(t)\}$$

respectively.

For reasonably sized risk sets $\hat{H}_{\mathrm{NA}}(t)$ and $\hat{H}_{\mathrm{KM}}(t)$ are close to each other: the difference is small relative to the width of the confidence interval. (Likewise $\hat{S}_{\mathrm{NA}}(t)$ and $\hat{S}_{\mathrm{KM}}(t)$).

Exercise 6: Convince yourself, for large risk sets, that $\hat{H}_{\mathrm{NA}}(t)$ is nearly equal to $\hat{H}_{\mathrm{KM}}(t)$.

Nelson-Aalen and proportional hazards

Reminder of proportional hazards

A *proportional hazards* model relates the hazard $h_i(t)$ seen by the ith individual to a vector of explanatory variables $z^{(i)}$ by

$$h_i(t) = \phi(z^{(i)}, \beta)h_0(t)$$

where β is a vector of parameters and h_0 is the baseline hazard function. The function ϕ is usually of the form $\phi(z^{(i)}, \beta) = \exp(\beta^{\mathrm{T}} z^{(i)})$.

The essence of proportional hazards modelling is that β can be estimated without needing to estimate $h_0(t)$.

Estimation of the baseline integrated hazard

There are occasions when an estimate of $H_0(t) = \int_0^t h_0$ would be useful. We can use a method very similar to that used to derive the Nelson-Aalen estimator. Here, again, is equation (2) for the compensator:

$$\Lambda_+(t) = \int_0^t \sum_{i=1}^n Y_i(u) h_i(u) \mathrm{d}u.$$

which in the proportional hazards formulation becomes:

$$\Lambda_+(t) = \int_0^t \sum_{i=1}^n Y_i(u) \phi(z^{(i)}, \beta) h_0(u) \mathrm{d}u.$$

As in the derivation of the Nelson-Aalen estimator, we estimate $\mathrm{d}\Lambda_+(t)$ by $\mathrm{d}N_+(t)$, so we have

$$\mathrm{d}\hat{\Lambda}_+(t) = \sum_{i=1}^n Y_i(u) \phi(z^{(i)}, \hat{\beta}) \mathrm{d}\hat{H}_0(t) = \mathrm{d}N_+(t)$$

where I have replaced β by its estimate $\hat{\beta}$.

Rearranging the RH equality and integrating, we obtain

$$\hat{H}_0(t) = \int_0^t \frac{\mathrm{d}N_+(u)}{\sum_{i=1}^n Y_i(u) \phi(z^{(i)}, \hat{\beta})}$$

as an estimator of the baseline integrated hazard.

References

In (exponentially) increasing order of difficulty:

Crowder M.J. (2001) *Classical Competing Risks.* Chapman & Hall / CRC [Chapter 8 for neat introduction]

Thernau T.M. and Grambsch P.M. (2000) *Modeling Survival Data: Extending the Cox Model.* Springer-Verlag

Fleming T.R. and Harrington D.P. (1991) *Counting Processes and Survival Analysis.* Wiley

Andersen P.K., Borgan Ø., Gill R.D. and Keiding N. (1993) *Statistical Models Based on Counting Processes.* Springer-Verlag