NON PARAMETRIC ESTIMATION OF SURVIVAL AND INTEGRATED HAZARD I (PART 1)

Lorenza Menghetti

22 May 2006

1 Introduction

In preceding chapters we saw that many censored survival data statistics can be written as $\sum_i \int H_i dM_i$. Such statistics are martingales, a structure yielding formulas for first and second moments. In our presentation, we will use these results to explore some finite-sample properties of estimators of survival and hazard functions. Without martingale methods, even the computation of first and second moments for these tests and estimators can be tedious.
We will use the random censorship model, this notion has been presented earlier. In many applications, especially in clinical research, this is the form of censoring commonly encoutered. Next remember Thm 1.3.2 from preceding chapter.

- **Theorem 1.3.2**
  Let T and U be failure and censoring variables, and let $X = \min(T, U)$, $\delta = I_{\{T \leq U\}}$, $N(t) = I_{\{X \leq t, \delta = 1\}}$ and $N^U(t) = I_{\{X \leq t, \delta = 0\}}$.
  Define $\mathcal{F}_t = \sigma\{N(u), N^U(u) : 0 \leq u \leq t\}$.
  The process M given by

  $$M(t) = N(t) - \int_0^t I_{\{X \geq u\}} d\Lambda(u)$$

  is a martingale with respect to $\mathcal{F}_t$ if and only if

  $$\frac{dF(z)}{1 - F(z-)} = \frac{-dP\{T \geq z; U \geq T\}}{P\{T \geq z; U \geq z\}} \qquad (3.7)$$

  for all z such that $P\{T \geq z, U \geq z\} > 0$.

- **Definition 3.1.1. (Random Censorship Model)**
  In the random censorship model, the ordered pairs (T_j, U_j), $(j = 1, 2, \ldots, n)$ are n independent finite failure and censoring time random variables that satisfy Eq. (3.7) in Thm 1.3.2, for each $n = 1, 2, \ldots$.
  The observable data are

  $$X_j = min(T_j, U_j) \equiv T_j \wedge U_j$$

  and

  $$\delta = I_{\{X_j = T_j\}}.$$

The notation for the underlying distributions in the random censorship model will be

$$S(t) = P\{T_j > t\},$$
$$F(t) = 1 - S(t),$$
$$C_j(t) = P\{U_j > t\},$$
$$L_j(t) = 1 - C_j(t),$$

and

$$\pi_j(t) = P\{X_j \geq t\}.$$

The following stochastic processes have been introduced in previous chapters:

$$\overline{N}(t) \equiv \sum_{j=1}^{n} N_j(t) \equiv \sum_{j=1}^{n} I_{\{X_j \leq t, \delta_j = 1\}}$$

$$N_j^U(t) = I_{\{X_j \leq t, \delta_j = 0\}}$$

$$\overline{Y}(t) \equiv \sum_{j=1}^{n} Y_j(t) \equiv \sum_{j=1}^{n} I_{\{X_j \geq t\}}$$

$$M_j(t) = N_j(t) - \int_0^t Y_j(s) d\Lambda(s)$$

$$M(t) = \sum_{j=1}^{n} M_j(t) = \overline{N}(t) - \int_0^t \overline{Y}(s) d\Lambda(s)$$

where

$$\Lambda(t) = \int_0^t \frac{1}{1 - F(s-)} dF(s).$$

All martingale properties depend on a specification of the way information accrues over time, i.e. a filtration. Until specified otherwise the filtration $\{\mathcal{F}_t : t \geq 0\}$ we will use will be given by

$$\mathcal{F}_t = \sigma\{N_j(s), N_j^U(s) : 0 \leq s \leq t, j = 1, \ldots, n\}.$$

2 Nelson estimator

2.1 Bias of Nelson estimator

We examine methods for estimating the cumulative hazard function $\Lambda(t)$ of single homogeneous sample.
From Theorem 1.3.2 ,

$$M_j(t) = N_j(t) - \int_0^t Y_j(s) d\Lambda(s)$$

is a martingale for each j with respect to $\{\mathcal{F}_t : t \geq 0\}$. In turn

$$M_j(t) = \overline{N}_j(t) - \int_0^t \overline{Y}_j(s) d\Lambda(s)$$

is a martingale.

Since the process give at time t by

$$\frac{I_{\{\overline{Y}(t)>0\}}}{\overline{Y}(t)} = \begin{cases} \frac{1}{\overline{Y}(t)} & \text{if } \overline{Y}(t) > 0 \\ 0 & \text{if } \overline{Y}(t) = 0 \end{cases}$$

is a left-continuous adapted process with right-hand limits, $\{\mathcal{M}(t) : t \geq 0\}$ given by

$$\mathcal{M}(t) = \int_0^t \frac{I_{\{\overline{Y}(t)>0\}}}{\overline{Y}(t)} dM(s) = \int_0^t \frac{d\overline{N}(s)}{\overline{Y}(s)} - \int_0^t I_{\{\overline{Y}(s)>0\}} d\Lambda(s)$$

is a martingale. It follow, since $\mathcal{M}(0) = 0$, that

$$E \int_0^t \frac{d\overline{N}(s)}{\overline{Y}(s)} = E \int_0^t I_{\{\overline{Y}(s)>0\}} d\Lambda(s). \tag{1}$$

Let $\Lambda^*(t) = \int_0^t I_{\{\overline{Y}(s)>0} d\Lambda(s)\}$. Then, if $T = \inf\{t : \overline{Y}(t) = 0\}$, $\Lambda^*(t) = \int_0^{t \wedge T} d\Lambda(s) = \Lambda(t \wedge T)$.
By (1), we might expect that

$$\hat{\Lambda}(t) \equiv \int_0^t \frac{d\overline{N}(s)}{\overline{Y}(s)}$$

would be a good "estimator" for $\Lambda^*(t) = \Lambda(t \wedge T)$, but that it would not be possible to obtain an unbiased estimator of $\Lambda(t)$ without making parametric assumptions.

The following theorem summarizes some properties of $\hat{\Lambda}$, an estimator first proposed by Nelson (1969).

- **Theorem 3.2.1**
  Sei $t \geq 0$ be such that $\Lambda(t) < \infty$. Then

  1. $E\{\hat{\Lambda}(t) - \Lambda^*(t)\} = 0$,
  2. $E\{\hat{\Lambda}(t) - \Lambda(t)\} = -\int_0^t [\prod_{j=1}^n \{1 - \pi_j(s)\}] d\Lambda(s)$,
  3. if $\pi_j(s) = \pi(s)$ for all j, then

     $$E\{\hat{\Lambda}(t) - \Lambda(t)\} = -\int_0^t \{1 - \pi_j(s)\}^n d\Lambda/(s) \geq -\{1 - \pi_j(s)\}^n \Lambda(t),$$

  4. $$\sigma_*^2(t) = E[\sqrt{n}\{\hat{\Lambda}(t) - \Lambda^*(t)\}]^2 = E\left[n \int_0^t \frac{I_{\{\overline{Y}(t)>0\}}}{\overline{Y}(t)} \{1 - \Delta\Lambda(s)\} d\Lambda(s)\right].$$

2.2 Variance of Nelson estimator

Suppose $\pi_j(s) = \pi(s)$ for all j and s. If $\pi(t) > 0$, Thm 3.2.1 indicate that $\hat{\Lambda}(t)$ is asymptotically unbiased estimator of $\Lambda(t)$, with bias converging to zero at as exponential rate as $n \to 0$. For the second moment $\sigma_*^2(t)$, for large n, should approach

$$\sigma^2(t) \equiv \int_0^t \frac{1 - \Delta\Lambda(s)}{\pi(s)} d\Lambda(s).$$

Since $En\{\Lambda^*(t) - \Lambda(t)\}^2$ converges to zero when $\pi(t) > 0$, since

$$\sqrt{n}\{\hat{\Lambda}(t) - \Lambda^*(t)\} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \int_0^t \frac{n}{\overline{Y}(s)} dM_j(s),$$

where $\{M_j\}$ is an independent and identically distributed collection, and since $n\{\overline{Y}(s)\}^{-1}$ converges to $\{\pi(s)\}^{-1}$, we might expect that $\sqrt{n}\{\hat{\Lambda}(t) - \Lambda(t)\}$ is approximately distributed as $N(0, \sigma^2(t))$ for large n.

The precision of $\hat{\Lambda}(t)$ at time t can be measured either by its variance, $E\{\hat{\Lambda}(t) - E\hat{\Lambda}(t)\}^2$ or, since it is biased, by its mean quared error, $E[\{\hat{\Lambda}(t) - \Lambda(t)\}^2]$
. Since the squared bias satisfies

$$[E\{\hat{\Lambda}(t) - \Lambda(t)\}]^2 \leq \{1 - \pi(t)\}^{2n}\{\Lambda(t)\}^2,$$

these will be nearly equal, and the variance of $\hat{\Lambda}(t)$ can be safely used exept when $\Lambda(t)$ is large and n is small, or when $\pi(t)$ is zero.
The variance is given by

$$var\hat{\Lambda}(t) = n^{-1}\sigma_*^2(t) + 2E[\{\hat{\Lambda}(t) - \Lambda^*(t)\}\{\hat{\Lambda}(t) - E\Lambda^*(t)\}] + E\{\Lambda^*(t) - E\Lambda^*(t)\}^2 \approx \frac{1}{n}\sigma_*^2(t),$$

for even relatively small values of n.
In estimating the variance of $\hat{\Lambda}(t)$, it is therefore sufficient to find a good estimator of $n^{-1}\sigma_*^2$.

- **Theorem 3.2.2**
  Let $t \geq 0$ be such that $\Lambda(t) < \infty$. Define

$$\frac{1}{n}\hat{\sigma}^2(t) = \int_0^t \frac{I_{\{\overline{Y}(s) > 0\}}}{\overline{Y}^2(s)} \left\{1 - \frac{\Delta\overline{N}(s) - 1}{\overline{Y}(s) - 1}\right\} d\overline{N}(s),$$

where $\frac{0}{0} \equiv 0$ as usual. Then

$$E\left\{\frac{1}{n}\hat{\sigma}^2(t) - \frac{1}{n}\sigma_*^2(t)\right\} = \int_0^t P\{\overline{Y}(s) = 1\}\Delta\Lambda(s) d\Lambda(s).$$