SECTION 13

# Estimation from Censored Data

Let $P$ be a a nonatomic probability distribution on $[0,\infty)$. The cumulative hazard function $\beta$ is defined by

$$\beta(t) = \int \frac{\{0 \le x \le t\}}{P[x,\infty)} P(dx).$$

It uniquely determines $P$. Let $T_2, T_2, \ldots$ be independent observations from $P$ and $\{c_i\}$ be a deterministic sequence of nonnegative numbers representing censoring times. Suppose the data consist of the variables

$$T_i \wedge c_i \qquad \text{and} \qquad \{T_i \le c_i\} \qquad \text{for } i = 1, \ldots, n.$$

That is, we observe $T_i$ if it is less than or equal to $c_i$; otherwise we learn only that $T_i$ was censored at time $c_i$. We always know whether $T_i$ was censored or not.

If the $\{c_i\}$ behave reasonably, we can still estimate the true $\beta$ despite the censoring. One possibility is to use the Nelson estimator:

$$\widehat{\beta}_n(t) = \frac{1}{n} \sum_{i \le n} \frac{\{T_i \le c_i \wedge t\}}{L_n(T_i)},$$

where

$$L_n(t) = \frac{1}{n} \sum_{i \le n} \{T_i \wedge c_i \ge t\}.$$

It has become common practice to analyze $\widehat{\beta}_n$ by means of the theory of stochastic integration with respect to continuous-time martingales. This section will present an alternative analysis using the Functional Central Limit Theorem from Section 10. Stochastic integration will be reduced to a convenient, but avoidable, means for calculating limiting variances and covariances.

**Heuristics.** Write $G(t)$ for $\mathbb{P}\{T_i \geq t\}$ and define

$$\Gamma_n(t) = \frac{1}{n}\sum_{i\leq n}\{c_i \geq t\}.$$

Essentially we need to justify replacement of $L_n$ by its expected value,

$$\mathbb{P}L_n(t) = \frac{1}{n}\sum_{i\leq n}\mathbb{P}\{T_i \geq t\}\{c_i \geq t\} = G(t)\Gamma_n(t).$$

That would approximate $\widehat{\beta}_n$ by an average of independent processes, which should be close to its expected value:

$$\begin{aligned}
\widehat{\beta}_n(t) &\approx \frac{1}{n}\sum_{i\leq n}\frac{\{T_i \leq c_i \wedge t\}}{G(T_i)\Gamma_n(T_i)} \\
&\approx \frac{1}{n}\sum_{i\leq n}\mathbb{P}\frac{\{T_i \leq t\}\{T_i \leq c_i\}}{G(T_i)\Gamma_n(T_i)} \\
&= \mathbb{P}\left(\frac{\{T_1 \leq t\}}{G(T_1)\Gamma_n(T_1)}\frac{1}{n}\sum_{i\leq n}\{T_1 \leq c_i\}\right) \\
&= \beta(t).
\end{aligned}$$

A more precise analysis will lead to a functional central limit theorem for the standardized processes $\sqrt{n}(\widehat{\beta}_n - \beta)$ over an interval $[0,\tau]$, if we assume that:

(i) the limit $\Gamma(t) = \lim_{n\to\infty}\Gamma_n(t)$ exists for each $t$;

(ii) the value $\tau$ is such that $G(\tau) > 0$ and $\Gamma(\tau) > 0$.

The argument will depend upon a limit theorem for a process indexed by pairs $(t,m)$, where $0 \leq t \leq \tau$ and $m$ belongs to the class $\mathcal{M}$ of all nonnegative increasing functions on $[0,\tau]$. Treating $\beta$ as a measure on $[0,\tau]$, define

$$\begin{aligned}
\beta(t,m) &= \int\{0 \leq x \leq t\}m(x)\beta(dx), \\
f_i(\omega,t,m) &= \{T_i \leq t \wedge c_i\}m(T_i) - \beta(t \wedge T_i \wedge c_i, m).
\end{aligned}$$

Such a centering for $f_i$ is suggested by martingale theory, as will be explained soon. We will be able to establish a functional central limit theorem for

$$\begin{aligned}
X_n(t,m) &= \frac{1}{\sqrt{n}}\sum_{i\leq n} f_i(\omega,t,m) \\
&= \sqrt{n}\left(\left(\frac{1}{n}\sum_{i\leq n}\{T_i \leq t \wedge c_i\}m(T_i)\right) - \beta(t, mL_n)\right).
\end{aligned}$$

Putting $m$ equal to $1/L_n$ we get the standardized Nelson estimator:

$$X_n(t, 1/L_n) = \sqrt{n}\big(\widehat{\beta}_n(t) - \beta(t)\big).$$

The limit theorem for $X_n$ will justify the approximation

$$X_n(t, 1/L_n) \approx X_n(t, 1/G\Gamma_n).$$

It will also give the limiting distribution for the approximating process.

**Some martingale theory.** The machinery of stochastic integration with respect to martingales provides a very neat way of calculating variances and covariances for the $f_i$ processes. We could avoid stochastic integration altogether by direct, brute force calculation; but then the happy cancellations arranged by the martingales would appear most mysterious and fortuitous.

The basic fact, not altogether trivial (Dellacherie 1972, Section V.5), is that both

$$Z_i(t) = \{T_i \leq t\} - \beta(t \wedge T_i) \qquad \text{and} \qquad Z_i(t)^2 - \beta(t \wedge T_i)$$

are continuous parameter martingales in $t$. That is, both the simple jump process $\{T_i \leq t\}$ and the submartingale $Z_i^2$ have compensator $\beta(t \wedge T_i)$. The $f_i$ process is expressible as a stochastic integral with respect to $Z_i$:

$$f_i(\omega, t, m) = \int \{0 \leq x \leq t \wedge c_i\} m(x) Z_i(dx).$$

It follows that, for fixed $m$, the process $f_i$ is also a martingale in $t$. In particular, $\mathbb{P} f_i(\omega, t, m) = \mathbb{P} f_i(\omega, 0, m) = 0$ for every $t$.

From now on let us omit the $\omega$ from the notation.

Stochastic integration theory tells us how to calculate compensators for new processes derived from the martingales $\{Z_i\}$. In particular, for fixed $t_1$, $t_2$, $m_1$, and $m_2$, the product $f_i(t \wedge t_1, m_1) f_i(t \wedge t_2, m_2)$ has compensator

$$A_i(t) = \int \{0 \leq x \leq t \wedge t_1 \wedge t_2 \wedge T_i \wedge c_i\} m_1(x) m_2(x) \beta(dx);$$

the difference $f_i(t \wedge t_1, m_1) f_i(t \wedge t_2, m_2) - A_i(t)$ is a martingale in $t$. This implies that

$$\mathbb{P} f_i(t \wedge t_1, m_1) f_i(t \wedge t_2, m_2) = \mathbb{P} A_i(t) \qquad \text{for each } t.$$

Put $t = \max(t_1, t_2)$, then average over $i$. Because each $T_i$ has the same distribution, we get

$$\begin{aligned} \mathbb{P} X_n(t_1, m_1) X_n(t_2, m_2) &= \frac{1}{n} \sum_{i \leq n} \mathbb{P} f_i(t_1, m_1) f_i(t_2, m_2) \\ &= \mathbb{P} \int \{0 \leq x \leq t_1 \wedge t_2\} L_n(x) m_1(x) m_2(x) \beta(dx) \\ &= \beta(t_1 \wedge t_2, G\Gamma_n m_1 m_2). \end{aligned} \tag{13.1}$$

The calculations needed to derive this result directly would be comparable to the calculations needed to establish the martingale property for $Z_i$.

**Manageability.** For each positive constant $K$ let $\mathcal{M}(K)$ denote the class of all those $m$ in $\mathcal{M}$ for which $m(\tau) \leq K$. To establish manageability of the $\{f_i(t, m)\}$ processes, as $t$ ranges over $[0, \tau]$ (or even over the whole of $\mathbb{R}^+$) and $m$ ranges over $\mathcal{M}(K)$, it suffices to consider separately the three contributions to $f_i$.

Let us show that the indicator functions $\{T_i \leq t \wedge c_i\}$ define a set with pseudodimension one. Suppose the $(i, j)$-projection could surround some point in $\mathbb{R}^2$. Suppose $T_i \leq T_j$. We would need to be able to find $t_1$ and $t_2$ such that both pairs of inequalities,

$$T_i \leq t_1 \wedge c_i \quad \text{and} \quad T_j \leq t_1 \wedge c_j,$$
$$T_i > t_2 \wedge c_i \quad \text{and} \quad T_j \leq t_2 \wedge c_j,$$

were satisfied. The first pair would imply $T_i \leq c_i$ and $T_j \leq c_j$, and then the second pair would lead to a contradiction, $t_2 \geq T_j \geq T_i > t_2$, which would establish the assertion about pseudodimension.

For the factors $\{m(T_i)\}$ with $m$ ranging over $\mathcal{M}(K)$, we can appeal to the result from Example 6.3 if we show that no 2-dimensional projection of the convex cone generated by $\mathcal{M}(K)$ can surround the point $(K, K)$. This is trivial. For if $T_i \leq T_j$ then no $r \in \mathbb{R}^+$ and $m \in \mathcal{M}(K)$ can achieve the pair of inequalities $rm(T_i) > K$ and $rm(T_j) < K$.

The argument for the third contribution to $f_i$ is similar. For each $t \leq \tau$ and $m \in \mathcal{M}(K)$, the process $\beta(t \wedge T_i \wedge c_i, m)$ is less than $K' = K\beta(\tau)$. If, for example, $T_i \wedge c_i \leq T_j \wedge c_j$ then it is impossible to find an $r \in \mathbb{R}^+$, an $m \in \mathcal{M}(K)$, and a $t \in [0, \tau]$ such that $r\beta(t \wedge T_i \wedge c_i, m) > K'$ and $r\beta(t \wedge T_j \wedge c_j, m) < K'$.

**Functional Central Limit Theorem.** It is a simple matter to check the five conditions of the Functional Central Limit Theorem from Section 10 for the triangular array of processes

$$f_{ni}(t, m) = \frac{1}{\sqrt{n}} f_i(t, m) \qquad \text{for } i = 1, \ldots, n, \quad t \in [0, \tau], \quad m \in \mathcal{M}(K),$$

for some constant $K$ to be specified. These processes have constant envelopes,

$$F_{ni} = K(1 + \beta(\tau))/\sqrt{n},$$

which clearly satisfy conditions (iii) and (iv) of the theorem. The extra $1/\sqrt{n}$ factor does not affect the manageability. Taking the limit in (13.1) we get

$$H\big((t_1, m_1), t_2, m_2)\big) = \beta(t_1 \wedge t_2, G\Gamma m_1 m_2).$$

For simplicity suppose $t_1 \leq t_2$. Then, because $f_{ni}$ has zero expected value, (13.1) also gives

$$\begin{aligned}
&\rho_n\big((t_1, m_1), (t_2, m_2)\big)^2 \\
&\quad = \mathbb{P}|X_n(t_1, m_1) - X_n(t_2, m_2)|^2 \\
&\quad = \beta(t_1, G\Gamma_n m_1^2) + \beta(t_2, G\Gamma_n m_2^2) - 2\beta(t_1, G\Gamma_n m_1 m_2) \\
&\quad = \int \{0 \leq x \leq t_1\} G\Gamma_n (m_1 - m_2)^2 \beta(dx) + \int \{t_1 \leq x \leq t_2\} G\Gamma_n m_2^2 \beta(dx) \\
&\quad \leq \int \{0 \leq x \leq t_1\} (m_1 - m_2)^2 \beta(dx) + \int \{t_1 \leq x \leq t_2\} m_2^2 \beta(dx).
\end{aligned}$$

A similar calculation with $\Gamma_n$ replaced by $\Gamma$ gives

$$\begin{aligned}
&\rho\big((t_1, m_1), (t_2, m_2)\big)^2 \\
&\quad = \int \{0 \leq x \leq t_1\} G\Gamma (m_1 - m_2)^2 \beta(dx) + \int \{t_1 \leq x \leq t_2\} G\Gamma m_2^2 \beta(dx),
\end{aligned}$$

which is greater than the positive constant factor $G(\tau)\Gamma(\tau)$ times the upper bound just obtained for $\rho_n\big((t_1, m_1), (t_2, m_2)\big)^2$. The second part of condition (v) of the Functional Central Limit Theorem follows.

The processes $\{X_n(t, m)\}$, for $0 \le t \le \tau$ and $m \in \mathcal{M}(K)$, converge in distribution to a Gaussian process $X(t, m)$ with $\rho$-continuous paths, zero means, and covariance kernel $H$.

**Asymptotics for $\widehat{\beta}_n$.** We now have all the results needed to make the heuristic argument precise. A straightforward application of Theorem 8.2 shows that

$$\sup_t |L_n(t) - G(t)\Gamma_n(t)| \to 0 \qquad \text{almost surely.}$$

If we choose the constant $K$ so that $G(\tau)\Gamma(\tau) > 1/K$, then, with probability tending to one, both $1/L_n$ and $1/G\Gamma_n$ belong to $\mathcal{M}(K)$ and

$$\sup_{0 \le t \le \tau} \rho\big((t, 1/L_n), (t, 1/G\Gamma_n)\big) \to 0 \qquad \text{in probability.}$$

From stochastic equicontinuity of $\{X_n\}$ we then deduce that

$$\begin{aligned}\sqrt{n}(\widehat{\beta}_n(t) - \beta(t)) &= X_n(t, 1/L_n) \\ &= X_n(t, 1/G\Gamma_n) + o_p(1) \qquad \text{uniformly in } 0 \le t \le \tau \\ &\rightsquigarrow X(t, 1/G\Gamma).\end{aligned}$$

The limit is a Gaussian process on $[0, \tau]$ with zero means and covariance kernel $\beta(t_1 \wedge t_2, 1/G\Gamma)$. It is a Brownian motion with a stretched out time scale.

REMARKS. As suggested by Meier (1975), deterministic censoring times $\{c_i\}$ allow more flexibility than the frequently made assumption that the $\{c_i\}$ are independent and identically distributed random variables. A conditioning argument would reduce the case of random $\{c_i\}$ to the deterministic case, anyway.

The method introduced in this section may seem like a throwback to the original proof by Breslow and Crowley (1974). However, the use of processes indexed by $\mathcal{M}(K)$ does eliminate much irksome calculation. More complicated forms of multivariate censoring might be handled by similar methods. For a comparison with the stochastic integral approach see Chapter 7 of Shorack and Wellner (1986).

I am grateful to Hani Doss for explanations that helped me understand the role of martingale methods.