SECTION 13

Estimation from Censored Data

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