Edge Corrections for Spatial Point Processes

September 15, 2004

Some Background: The Kaplan Meier Estimator. Let

$$X = \text{lifetime},$$
$$Y = \text{censoring time},$$
$$Z = X \wedge Y = \min[X, y]$$

and

$$\delta = \mathbf{1}_{X \leq Y}.$$

Suppose that X and Y are independent, that $X \sim F$, and that $Y \sim G$, so that

$$F(t) = P[X \leq t],$$
$$G(t) = P[Y \leq t].$$

Suppose also that $F(0) = 0 = G(0)$ and that $0 < F(t), G(t) < 1$ for all $0 < t < \infty$.

The Problem: Given an sample $(\delta_1, Z_1), \cdots, (\delta_n, Z_n)$, estimate F. Ignoring the censoring leads to a biased estimator here, as does only considering uncensored cases. Let

$$0 = t_0 < t_1 < \cdots < t_m < \infty$$

and

$$\Lambda_i = P[X > t_i | Z > t_{i-1}].$$

Then

$$\Lambda_i = P[X > t_i | X > t_{i-1}] = \frac{1 - F(t_i)}{1 - F(t_{i-1})}$$

and

$$\hat{\Lambda}_i = \frac{\#\{j : X_j > t_i\}}{\#\{j : Z_j > t_{i-1}\}}$$

is an unbiased estimator of Λ_i. Here

$$1 - F(t_k) = \Lambda_1 \times \cdots \times \Lambda_k.$$

So,

$$\hat{\Lambda}_1 \times \cdots \times \hat{\Lambda}_k$$

provides and estimator of $1 - F(t_k)$.

Letting $0 < z_1 < \cdots < z_n$ denote the ordered values of $Z_1, \cdots, Z_n$ and taking a limit, the Kaplan-Meier estimator is

$$1 - F(\hat{z_k}) = \prod_{i=1}^{k} \frac{\#\{j : X_j \geq z_k\}}{\#\mathcal{R}_i},$$

where

$$\mathcal{R}_i = \{j : Z_j \geq z_j\}$$

is called the *risk set.*

Euclidean Space: Let R^p be the set of lists $x = (x_1, \cdots, x_p)$ of real numbers, $\|x\|^2 = x_1^2 + \cdots + x_p^2$, and $d(x, y) = \|x - y\|$.

Point Processes: Informally, a point process in $I\!R^p$ is a collection of points, place at random. It can also be described in terms of counts. Let $N_B = N\{B\}$, $B \subseteq I\!R^p$, be jointly distributed, non-negative integer or ∞ valued random variables for which

$$N_B < \infty \ w.p.1$$

for bounded B.

$$N_{A \cup B} = N_A + N_B$$

whenever $A \cap B = \emptyset$, and some other more technical conditions are satisfied.

Interpretation: Here N_B denotes the number of points in the set B. Let $\{x\}$ be the set consisting of x. Then, there is a point at x iff $N_{\{x\}} > 0$. Suppose that there are no multiple points; that is, $N_{\{x\}} \leq 1 \ w.p.1$ for all x. Let

$$\mathcal{N} = \{x : N_{\{x\}} > 0\}.$$

This is the set of points.

Example:*Poisson Processes.* Recall that a random variable X, say, has the Poisson distribution with parameter μ if

$$P[X = k] = \frac{1}{k!}\mu^k e^{-\mu}$$

for $k = 0, 1, 2, \cdots$ in which case $E(X) = \mu$. Thus

$$P[X = 0] = e^{-\mu}.$$

A point process N is called a *Poisson process*, if N_B has a Poisson distribution for every B and $N_{A_1}, \cdots, N_{A_m}$ are independent whenever $A_1, \cdots, A_m$ are mutually exclusive.

In one dimension, it is conventional to let

$$M_t = N_{(0,t]},$$

the number of points in the (time) interval $(0, t]$. Then the points occur at the discontinuities of M_t.

Stationarity: Let $B + x = \{y : y - x \in B\} = \{x + z : z \in B\}$ and suppose that

$$[N_{B_1+x}, \cdots, N_{B_m+x}] =^{\mathcal{D}} [N_{B_1}, \cdots, N_{B_m}]$$

for all $B_1, \cdots, B_m$ and x. Then

$$E[N_B] = \lambda \nu_p(B),$$

where ν_p is p-dimensional volume and λ is a constant, called the *intensity*

Some Quantities of Interest: Let

$$B(x,t) = \{y : d(x,y) \leq t\},$$
$$B_(x,t) = B(x,t) - \{x\}.$$

The Occupied Space Function:

$$p(t) = P[N\_B(x,t) > 0].$$

In view of the assumed stationarity, this does not depend on x

The Interpoint Distribution Function: Let

$$G(t) = P[N_{B_-(0,t)} > 0 | 0 \in \mathcal{N}].$$

The Interpoint Distance Function: Let

$$K(t) = \frac{1}{\lambda} E[N_{B_-(0,t)} | 0 \in \mathcal{N}].$$

Note: Defining the conditional probabilities, in any generality, requires some subtlety in the form Plam distribution.

Examples. If N is a Poisson process in $\mathbb{R}^2$, then

$$E[N_{B(0,t)}] = \lambda \pi t^2$$

and

$$p(t) = 1 - P[N_{B(0,t)} = 0] = 1 - e^{-\lambda \pi t^2}.$$

Also, using the indpendence,

$$G(t) = p(t)$$

and

$$K(t) = \lambda \pi t^2.$$

Estimating p, G, and K From Data. Let $E \subseteq I\!R^p$ be a bounded set with a non-empty interior–for example

$$E = \{x : |x_i| \leq 1, \ i = 1, 2\}$$

in R^2. Suppose that we observe (only) the points that fall in E. Let

$$n = N_E$$

and

$$\mathcal{N}_E = \mathcal{N} \cap E.$$

Estimating p: The Problem. Let

$$W = \{w^1, \cdots, w^m\} \subseteq E$$

be a finite subset of E;

$$\tilde{d}_i = \min_{x \in \mathcal{N}} d(w^1, x)$$

and

$$\tilde{p}(t) = \frac{1}{m} \#\{i : d_i \leq t\}$$

Then

$$P[\tilde{d}_i \leq t] = P[B(w^i, t) > 0] = p(t)$$

and

$$E[\tilde{p}(t)] = \frac{1}{m} \sum_{i=1}^{m} P[\tilde{d}_i \leq t] = p(t).$$

The problem is that we cannot compute $\tilde{d}_i$, because it may require points that are outside of E.

Estimating p: The Solution. Let

$$d_i = \min_{x \in \mathcal{N}_E} d(w^i, x)$$

and

$$r_i = d(w^1, E^c),$$

the distance from w^i to the complement of E. Observe that: If $r_i \geq t$, then $d_i \leq t$ iff $\tilde{d}_i \leq t$, so that

$$P[d_i \leq t] = p(t)$$

for $t \leq d_i$. Let

$$\hat{p}(t) = \frac{\#\{i : d_i \leq t \ \& \ r_i \geq t\}}{\#\{i : r_i \geq t\}}.$$

Then $\hat{p}(t)$ is unbiased. For, letting

$$m_i = \#\{i : r_i \geq t\},$$

$$E[\hat{p}(t)] = \frac{1}{m_i} \sum_{i:r_i \geq t} P[d_i \leq t] = p(t).$$

Some Problems. $p(t)$ is an increasing function of f, $\hat{p}(t)$ need not be.

Also, m_i may be very small for large values of t.