Survival Analysis: Counting Process and Martingale

Lu Tian and Richard Olshen

Stanford University

Lebesgue-Stieltjes Integrals

- $G(\cdot)$ is a right-continuous step function having jumps at $x_1, x_2, \cdots$ ..

$$\int_a^b f(x)dG(x) = \sum_{a<x_j\leq b} f(x_j)\{G(x_j) - G(x_j^-)) = \sum_{a<x_j\leq b} f(x_j)\Delta\{G(x_j)\}.$$

One-sample problem

- In one-sample problem: observation $(U_i, \delta_i), i = 1, 2, \cdots, n.$
- $N(t) = \sum_{i=1}^{n} I(U_i \leq t, \delta_i = 1)$ # observed failures by time t

$$\int_0^{\infty} f(t)dN(t) = \sum_{j=1}^{K} f(\tau_j)d_j$$

- $f(t) = \{\sum_{i=1}^{n} I(U_i \geq t)\}^{-1}$

$$\int_0^{\infty} f(t)dN(t) = \sum_{j=1}^{K} \frac{d_j}{Y(\tau_j)},$$

The NA estimator for the cumulative hazard function!

Two-sample problem

- In two-sample problem: observation $(U_i, \delta_i, Z_i), i = 1, 2, \cdots, n$.
- $N_l(t) = \sum_{i=1}^{n} I(U_i \le t, \delta_i = 1, Z_i = l)$ # observed failures by time t at group $l, l = 0, 1$.
- $Y_l(t) = \sum_{i=1}^{n} I(U_i \ge t, Z_i = l)$ # at risk by time t at group $l, l = 0, 1$.
- Consider the stochastic integral

$$W = \int_0^\infty \frac{Y_0(t)}{Y(t)} dN_1(t) - \int_0^\infty \frac{Y_1(t)}{Y(t)} dN_0(t)$$

Two-sample problem

- $$\begin{aligned} W &= \sum_{j=1}^{K} \frac{Y_0(\tau_j)}{Y(\tau_j)} \, \Delta N_1(\tau_j) \;-\; \sum_{j=1}^{K} \frac{Y_1(\tau_j)}{Y(\tau_j)} \, \Delta N_0(\tau_j) \\ &= \sum_{j=1}^{K} \left[\frac{Y_0(\tau_j)}{Y(\tau_j)} d_{1j} \;-\; \frac{Y_1(\tau_j)}{Y(\tau_j)} d_{0j} \right] \\ &= \sum_{j=1}^{K} \frac{Y_0(\tau_j) d_{1j} - Y_1(\tau_j)(d_j - d_{0j})}{Y(\tau_j)} \\ &= \sum_{j} \left(d_{1j} - d_j \cdot \frac{Y_1(\tau_j)}{Y(\tau_j)} \right) \;=\; \sum_{j} (O_j - E_j); \end{aligned}$$

Regression problem

- In Cox regression: observations $(U_i, \delta_i, Z_i), i = 1, 2, \cdots, n.$
- $N_i(t) = I(U_i \leq t, \delta_i = 1)$
- The score function related to PL function is

$$S_n(\beta) = \sum_{i=1}^{n} \delta_i \left\{ Z_i - \frac{S^{(1)}(\beta, U_i)}{S^{(0)}(\beta, U_i)} \right\}$$
$$= \sum_{i=1}^{n} \int_0^{\infty} \left\{ Z_i - \frac{S^{(1)}(\beta, s)}{S^{(0)}(\beta, s)} \right\} dN_i(s)$$

σ-algebra

- $\mathcal{F}$ is a non-empty collection of subsets in Ω. $\mathcal{F}$ is called σalgebra if
  - If $E \in \mathcal{F}$ then $E^c \in \mathcal{F}$
  - if $E_i \in \mathcal{F}, i = 1, \cdots, n$ then $\cup_{i=1}^{\infty} E_i \in \mathcal{F}$.

Probability Space

- A probability space $(\Omega, \mathcal{F}, P)$ is an abstract space Ω equipped with a $\sigma-$algebra and a set function P defined on $\mathcal{F}$ such that

  1. If $P\{E\} \geq 0$ when $E \in \mathcal{F}$
  2. $P\{\Omega\} = 1$
  3. $P\{\cup_{i=1}^{\infty} E_i\} = \sum_{i=1}^{\infty} P\{E_i\}$ when $E_i \in \mathcal{F}, i = 1, \cdots, n$ and $E_i \cap E_j = \phi, i \neq j$.

Random Variables and Stochastic Process

- $(\Omega, \mathcal{F}, P)$ is a probability space. A random variable $X : \Omega \to R$ is a real-valued function satisfying $\{\omega : X(\omega) \in (-\infty, b]\} \in \mathcal{F}$ for any b.

- A stochastic process is a family of random variables $X = \{X(t) : t \in \Gamma\}$ indexed by a set Γ all defined on the same probability space $(\Omega, \mathcal{F}, P)$. Often times $\Gamma = [0, \infty)$ in survival analysis.

Filtration and Adaptability

- A family of σ-algebra $\{\mathcal{F}_t : t \geq 0\}$ is called a filtration of a probability space $(\Omega, \mathcal{F}, P)$ if $\mathcal{F}_t \subset \mathcal{F}$ and $\mathcal{F}_s \subset \mathcal{F}_t$ for $s \leq t$.
- A filtration $\{\mathcal{F}_t; t \geq 0\}$ is right-continuous if for any t $\mathcal{F}_{t^+} = \mathcal{F}_t$ where $\mathcal{F}_{t^+} = \cap_{\epsilon > 0} \mathcal{F}_{t+\epsilon}$.
- A stochastic process $\{X(t) : t \geq 0\}$ is said to be adapted to $\{\mathcal{F}_t, t \geq 0\}$ if $X(t)$ is $\mathcal{F}_t$ measurable for each t, i.e., $\{\omega : X(t; \omega) \in (-\infty, b]\} \in \mathcal{F}_t$.

History of a stochastic process

- Let $X = \{X(t) : t \geq 0\}$ be a stochastic process, and let $\mathcal{F}_t = \sigma\{X(s) : 0 \leq s \leq t\}$, the smallest σ-algebra making all of the random variables $X(s), 0 \leq s \leq t$ measurable. The filtration $\{\mathcal{F}_t : t \geq 0\}$ is called the history of X.

Conditional Expectation

- Suppose that Y is random variable on a probability space $(\Omega, \mathcal{F}, P)$ and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Let X be a random variable satisfying

  1. X is $\mathcal{G}$-measurable
  2. $\int_B Y dP = \int_B X dP$ for all sets $B \in \mathcal{G}$.

  The random variable X is the called the conditional expectation of Y given $\mathcal{G}$ and denoted by $E(Y|\mathcal{G})$.

Conditional Expectation

- If $\mathcal{F}_t = \{\phi, \Omega\}$, then $E(X|\mathcal{F}_t) = E(X)$
- $E\{E(X|\mathcal{F}_t)\} = E(X)$
- If $\mathcal{F}_s \subset \mathcal{F}_t$, then $E\{E(X|\mathcal{F}_s)|\mathcal{F}_t\} = E(X|\mathcal{F}_s)$
- If Y is $\mathcal{F}_t$-measurable, then $E(XY|\mathcal{F}_t) = YE(X|\mathcal{F}_t)$
- If X and Y are independent, then $E(Y|\sigma(X)) = E(Y)$
- $E(aX + bY|\mathcal{F}_t) = aE(X|\mathcal{F}_t) + bE(Y|\mathcal{F}_t)$
- If $g(\cdot)$ is a convex function, then $E\{g(X)|\mathcal{F}_t\} \geq g\{E(X|\mathcal{F}_t)\}$.
- $E(Y|\sigma(X_t, t \in \Gamma)) = E(Y|X_t, t \in \Gamma)$

Example of conditional expectation

- The probability space $(\Omega, \mathcal{F}, P)$
  1. $\Omega = [0, 1)$
  2. $\mathcal{F}$ is the Borel algebra generated by all the open interval $(a, b) \subset [0, 1]$
  3. The probability measure: the lebesgue measure $P((a, b)) = b - a$.
- Define the sigma fields
  1. $\mathcal{F}_1 = \sigma([0, 1/2), [1/2, 1))$
  2. $\mathcal{F}_2 = \sigma([0, 1/4), [1/4, 2/4), [2/4, 3/4), [3/4, 1))$
  3. $\mathcal{F}_k = \sigma([i/2^k, (i+1)/2^k), i = 0, 1, \cdots, 2^k - 1)$

Example of conditional expectation

- Random Variable $X : (0,1) \to R^1 : X(\omega) = f(\omega)$
- $X_k = E(X|\mathcal{F}_k)$
- X_k is $\mathcal{F}_k$ measurable
  - $X_k(\omega)$ is a step function: $X_k(\omega) = a_i, \omega \in [i/2^k, (i+1)/2^k]$
- $\int_B X_k dP = \int_B X dP$
  - For $i = 0, 1, \cdots, 2^k - 1$,

$$a_i = 2^k \int_{i/2^k}^{(i+1)/2^k} f(\omega) d\omega$$

Martingale

- Let $X(\cdot) = \{X(t), t \geq 0\}$ be a right-continuous a stochastic process with left-hand limit and $\mathcal{F}_t$ be a filtration on a common probability space. $X(\cdot)$ is a martingale if
  1. X is adapted to $\{\mathcal{F}_t : t \geq 0\}$.
  2. $E|X(t)| < \infty$ for any t
  3. $E[X(t+s)|\mathcal{F}_t] = X(t)$ for any $t, s \geq 0$.
- $X(\cdot)$ is called a sub-martingale if $=$ is replaced by $\geq$ and super-martingale if $=$ is replaced by $\leq$.

Martingale

- In the survival analysis: $\mathcal{F}_t = \sigma\{X(s) : 0 \leq u \leq t\}$.
- For $t, s \geq 0$ $E[X(t+s)|\mathcal{F}_t] = E[X(t+s)|X(u), 0 \leq u \leq t]$, where $(X(u), 0 \leq u \leq t)$ is the history of the process $X(\cdot)$ from 0 to t.

Example of martingale

- $Y_1, Y_2, \ldots$ are i.i.d. random variables satisfying

$$Y_i = \begin{cases} +1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2 \end{cases}$$

- Define

$$\begin{aligned} X(0) &= 0, \quad \text{and} \\ X(n) &= \sum_{j=1}^{n} Y_j \quad n = 1, 2, \cdots \end{aligned}$$

- $(X(u) \; : \; 0 \le u \le n) = (X(1), \ldots, X(n)) \sim (Y_1, \ldots, Y_n)$

$$\mathcal{F}_n = \sigma(X(1), \ldots, X(n)) = \sigma(Y_1, \ldots, Y_n)$$

Example of martingale

- Clearly, $E \mid X(n) \mid < \infty$ for each n.
-

$$
\begin{aligned}
E[X(n+k) \mid \mathcal{F}_n] &= E\left[\sum_{j=1}^{n+k} Y_j \mid X(1), \ldots, X(n)\right] \\
&= E\left[\sum_{j=1}^{n+k} Y_j \mid Y_1, \ldots, Y_n\right] \\
&= Y_1 + \cdots + Y_n \; + \; E\left[Y_{n+1} + \cdots + Y_{n+k} \mid Y_1, \ldots, Y_n\right] \\
&= Y_1 + \cdots + Y_n \\
&= X(n)
\end{aligned}
$$

Two important properties of martingale

- Let X be a martingale with respect to a filtration $\{\mathcal{F}_t : t \geq 0\}$ then $E\{X(t)|\mathcal{F}_{t^-}\} = X(t^-)$, where $\mathcal{F}_{t^-} = \cup_{s<t}\mathcal{F}_s$.
- Let X be a martingale with respect to a filtration $\{\mathcal{F}_t : t \geq 0\}$ then $E\{dX(t)|\mathcal{F}_{t^-}\} = 0$

$$E\{X(t) - X(t-\epsilon)|\mathcal{F}_{t-\epsilon}\} = X(t-\epsilon) - X(t-\epsilon) = 0.$$

Counting Process

- A stochastic process $N(\cdot) = (N(t) \ : \ t \geq 0)$ is called a **counting process** if
  1. $N(0) = 0$
  2. $N(t) < \infty, \quad$ all t
  3. With probability 1, $N(t)$ is a right-continuous step function with jumps of size +1.
- $N_i(t)$, $N(t)$ and $N_l(t)$ are all counting process.

Compensator Process

- the stochastic process $A(\cdot)$, defined by

$$A(t) \stackrel{\text{def}}{=} \int_0^t Y(u)\ h(u)du.$$

- $E(N(t)) = E(A(t))$
- $M(t) = N(t) - A(t)$ is a mean zero process

Example

- $T \sim EXP(\lambda) \Rightarrow h(t) = \lambda$
- $A(t) = \int_0^t h(u)Y(u)du = \lambda \int_0^t Y(u)du = \lambda \min(t, U)$
- $M(t) = I(U \leq t)\delta - \lambda \min(t, U)$
- In general $M(t) = I(U \leq t)\delta - \int_0^t Y(u)h(u)du$
- Indeed, $M(t)$ is a martingale under independent censoring assumption!

Martingale

- $M(t) = N(t) - \int_0^t Y(s)h(s)ds$ and $\mathcal{F}_t = \sigma\{N(u), N^C(u), u \in [0,t]\}$. The key is to show that

$$E(M(t+s) - M(t)|\mathcal{F}_t) = 0 \;\; a.s.$$

- Need to show that

$$E[N(t+s) - N(t)|\mathcal{F}_t] = E\left[\int_t^{t+s} h(u)Y(u)du|\mathcal{F}_t\right]$$

- Consider the value of $Y(t)$

$$\begin{aligned} & E\left[N(t+s) - N(t)|\mathcal{F}_t\right] \\ = & \left\{ \begin{array}{ll} 0 & \text{if } Y(t) = 0 \\ E\left[N(t+s) - N(t)|Y(t) = 1\right] & \text{if } Y(t) = 1. \end{array} \right. \end{aligned}$$

Martingale

- Since N(t+s)-N(t) is zero or one,

$$
\begin{aligned}
&E[N(t+s)-N(t)|Y(t)=1] \\
&\quad = P[N(t+s)-N(t)=1|Y(t)=1] \\
&\quad = P[t<U\le t+s, \delta=1|U\ge t] \\
&\quad = \frac{P[t<T\le t+s, C>T]}{P[U\ge t]}.
\end{aligned}
$$

Martingale

- Consider the value of $Y(t)$

$$E\left[\int_t^{t+s} h(u)Y(u)du|\mathcal{F}_t\right]$$
$$= \begin{cases} 0 & \text{if } Y(t)=0 \\ E\left[\int_t^{t+s} h(u)Y(u)du|Y(t)=1\right] & \text{if } Y(t)=1. \end{cases}$$

- When $Y(t)=1$

$$E\left[\int_t^{t+s} h(u)Y(u)du|Y(t)=1\right] = \int_t^{t+s} h(u)E\left[Y(u)|Y(t)=1\right]du$$
$$= \int_t^{t+s} h(u)P[U \geq u|U \geq t]du$$
$$= \frac{\int_t^{t+s} h(u)P(U \geq u)du}{P(U \geq t)} = \frac{P[t < T \leq t+s, C > T]}{P[U \geq t]}.$$

Martingale

- In the last step, we argue that

$$\frac{\int_t^{t+s} h(u)P(U \geq u)du}{P(U \geq t)} = \frac{P[t < T \leq t+s, C > T]}{P[U \geq t]}$$

- It holds if

$$h(u) = \lim_{\epsilon \to 0} \frac{P(u \leq T \leq u + \epsilon | T \geq u, C \geq u)}{\epsilon},$$

which is the noninformative censoring assumption!

Predictable process

- A stochastic process $X(\cdot)$ is called predictable with respect to a filtration $\mathcal{F}_t$ if for very t, $X(t)$ is measurable with respect to $\mathcal{F}_{t-}$
- Any left-continuous adapted process is predictable. $N(t)$ is not predictable since it is right-continuous but $Y(t) = I(U \geq t)$ is a predictable process.

Doob-Meyer decomposition

- Suppose that $X(\cdot)$ is a non-negative submartinagle adapted to $\{\mathcal{F}_t, t \geq 0\}$. Then there exists a right-continuous and non-decreasing predictable process $A(\cdot)$ such that $E(A(t)) < \infty$ for all t and

$$M(\cdot) = X(\cdot) - A(\cdot)$$

  is a martinagle. If $A(0) = 0$ a.s., then $A(\cdot)$ is unique!

- The process $A(\cdot)$ is the called the compensator for $X(\cdot)$.
- $\int_0^t h(u)Y(u)du$ is the unique compensator for the counting process $N(t) = I(U \leq t, \delta = 1)$.

Property of Martingale associated with counting process

- If $M(\cdot)$ is a martinagle, then $M^2(\cdot)$ is a submartingale (why?)
- By Doob-Meyer decomposition, there is a unique predicable process $< M, M >$ such that

$$M^2(\cdot) - < M, M > (\cdot)$$

  is a martingale.
- $E(M^2(t)) = E\{< M, M > (t)\}$
- If $M(\cdot) = N(\cdot) - A(\cdot)$, then $< M, M > (\cdot) = A(\cdot)$
- We can calculate the variance of $N(\cdot) - A(\cdot)$!

$$\text{var}\{N(t) - A(t)\} = E\{A(t)\} = E\int_0^t Y(u)h(u)du = \int_0^t P(U > u)h(u)du.$$