

1 Outline

Suppose:

- $(\Omega, \mathcal{A}, \mu)$ is a probability space.
- $n \in \mathbb{N}$ is an natural number (positive integer).
- $T_1, T_2, \dots, T_n : \Omega \longrightarrow [0, \infty]$ are independent identically distributed extended \mathbb{R} -valued random variables.
- $U_1, U_2, \dots, U_n : \Omega \longrightarrow [0, \infty]$ are independent identically distributed extended \mathbb{R} -valued random variables.
- For each $i = 1, 2, \dots, n$, let $X_i := \min\{T_i, U_i\}$, and $C_i := I_{\{T_i \leq U_i\}}$.

For each subject $i = 1, 2, \dots, n$, the random variable T_i is interpreted to be the “survival time” of subject i , while U_i is interpreted to be the “censoring time” of subject i .

We wish to make inference about the (common) *survival function*

$$S(t) := P(T > t) = \mu\left(\left\{\omega \in \Omega \mid T(\omega) > t\right\}\right)$$

of T_1, T_2, \dots, T_n . However, in survival analysis, the inference about $S(t)$ is made based on the *right-censored survival time data* $\{X_i, C_i\}$, $i = 1, 2, \dots, n$ (rather than on the T_i ’s directly).

The *hazard function*:

$$\lambda(t) := \lim_{h \rightarrow 0^+} \frac{1}{h} \cdot P\left(t \leq T < t + h \mid t \leq T\right)$$

The *cumulative hazard function*:

$$\Lambda(t) := \int_0^t \lambda(t) dt$$

The *Nelson-Aalen estimator* for the cumulative hazard function $\Lambda(t)$:

$$\hat{\Lambda}(\omega, t) := \sum_{\substack{T_i(\omega) \leq t \\ C_i(\omega) = 1}} \frac{1}{Y(\omega, T_i(\omega))},$$

where

$$Y_i(\omega, t) := \begin{cases} 1, & t - h < T_i(\omega) \leq U_i(\omega), \text{ for each } h > 0 \\ 0, & \text{otherwise} \end{cases}$$

and

$$Y(\omega, t) := \sum_{i=1}^n Y_i(\omega, t)$$

The aggregated counting process for subject i :

$$N_i(\omega, t) := I_{\{X_i(\omega) \leq t\}}$$

The aggregated counting process:

$$N(\omega, t) := \sum_{i=1}^n N_i(\omega, t) = \sum_{i=1}^n I_{\{X_i(\omega) \leq t\}}$$

The aggregated intensity process:

$$\alpha(\omega, t) := \lim_{h \rightarrow 0^+} \frac{1}{h} \cdot P\left(N(\omega, t+h) - N(\omega, t) = 1 \mid \mathcal{F}_t\right) = \lim_{h \rightarrow 0^+} \frac{1}{h} \cdot E\left[N(\omega, t+h) - N(\omega, t) \mid \mathcal{F}_t\right]$$

The aggregated cumulative intensity process:

$$A(\omega, t) := \int_0^t \alpha(\omega, t) dt$$

Then, the process

$$M(\omega, t) := N(\omega, t) - A(\omega, t) = N(\omega, t) - \int_0^t \alpha(\omega, t) dt$$

is a martingale process. In particular, $M(\cdot, t)$ satisfies

$$E\left[M(\cdot, t+h) - M(\cdot, t) \mid \mathcal{F}_t\right](\omega) = 0$$

References

- [1] ALIPRANTIS, C. D., AND BURKINSHAW, O. *Principles of Real Analysis*, third ed. Academic Press, 1998.
- [2] BILLINGSLEY, P. *Probability and Measure*, third ed. John Wiley & Sons, 1995.
- [3] BILLINGSLEY, P. *Convergence of Probability Measures*, second ed. John Wiley & Sons, 1999.
- [4] FERGUSON, T. S. *A Course in Large Sample Theory*, first ed. Texts in Statistical Science. CRC Press, 1996.
- [5] SHAO, J. *Mathematical Statistics*, second ed. Springer, 2003.