Contents

1 Introduction

Definition 1.1 (**Random Censorship Model**). *In the random censorship model, the ordered pairs* (T_{ij}, U_{ij}), $j = 1, \ldots, n_i, i = 1, \ldots, r$, *are* $n(= \sum_{i=1}^{r} n_i)$ *independent finite failure and censoring time random variables that satisfy the equation* $\frac{dF(z)}{1-F(z-)} = \frac{-dP\{T \geq z; U \geq T\}}{P\{T \geq z; U \geq T\}}$ *for all* z *such that* $P\{T \geq z; U \geq T\} > 0$ *for each* $n = 1, 2, \ldots$. *The observable data are:*

$$X_{ij} = min(T_{ij}, U_{ij})$$

$$\equiv T_{ij} \wedge U_{ij}$$

and

$$\delta_{ij} = I_{(X_{ij} = T_{ij})}$$

The notation for the underlying distributions in the random censorship model will be

$$S_i(t) = P\{T_{ij} > t\},$$

$$F_i(t) = 1 - S_i(t),$$

$$C_{ij}(t) = P\{U_{ij} > t\},$$

$$L_{ij}(t) = 1 - C_{ij}(t),$$

and

$$\pi_{ij}(t) = P\{X_{ij} \geq t\}.$$

The following stochastic processes are introduced:

$$\overline{N_i}(t) = \sum_{j=1}^{n_i} N_{ij}(t) = \sum_{j=1}^{n_i} I_{(X_{ij} \leq t, \delta_{ij} = 1)},$$

$$N_{ij}^U(t) = I_{(X_{ij} \leq t, \delta_{ij} = 0)},$$

$$\overline{Y_i}(t) = \sum_{j=1}^{n_i} Y_{ij}(t) = \sum_{j=1}^{n_i} I_{(X_{ij} \geq t)},$$

$$M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(s) d\Lambda s,$$

$$M_i(t) = \sum_{j=1}^{n_i} M_{ij}(t) = \overline{N_i}(t) - \int_0^t \overline{Y_i}(s) d\Lambda s,$$

where,

$$\Lambda_i(t) = \int_0^t [1 - F_i(s-)]^{-1} dF_i(s).$$

All martingale properties depend on a specification of the way information accrues over time or, in other words, a filtration. Until specified otherwise, we will use the filtration $\{\mathcal{F}_t : t \geq 0\}$ is given by:

$$\mathcal{F}_t = \sigma\{N_{ij}(s), N_{ij}^U(s) : 0 \leq s \leq t, j = 1, \ldots, n_i, i = 1, \ldots, r\}.$$

This filtration specifies at time t, which items have failed or have been censored up to and including that time and, in this setting the assumption that for each $t \geq 0$, given $\mathcal{F}_{t-}$, $\{\Delta N_{ij}(t) : j = 1, \ldots, n_i, i = 1, \ldots, r\}$ are all independent $0, 1$ random variables for the counting processes $\{N_{ij}(t) : j = 1, \ldots, n_i, i = 1, \ldots, r\}$.

2 Non-Parametric Estimation of the Survival Distribution

In this section, we examine finite sample properties of estimators of the survival distribution in a single homogeneous sample $(r = 1)$, and suppress the subscript i.

We first examine methods for estimating the cumulative hazard function, $\Lambda(t)$. From **Theorem 3.1** and **Lemma 3.1** $M_j(t) = N_j(t) - \int_0^t I_{(X_j \geq u)} d\Lambda(u)$ is a martingale for each j with respect to $\{\mathcal{F}_t : t \geq 0\}$. In turn, $M(t) = \overline{N}(t) - \int_0^t \overline{Y}(s) d\Lambda(s)$ is a martingale, where $\overline{N}(t) = \sum_{j=1}^n N_j(t)$ and $\overline{Y}(t) = \sum_{j=1}^n Y_j(t)$. Since, the process given at time t by

$$\frac{I_{(\overline{Y}(t)>0)}}{\overline{Y}(t)} = \begin{cases} 1/\overline{Y}(t), & \overline{Y}(t) > 0 \\ 0, & \overline{Y}(t) = 0 \end{cases}$$

is a left-continuous adapted process with right-hand limits, then $\{\mathcal{M}(t) : t \geq 0\}$ given by

$$\begin{aligned} \mathcal{M}(t) &= \int_0^t \frac{I_{(\overline{Y}(s)>0)}}{\overline{Y}(s)} dM(s) \qquad \textbf{(Theorem 3.2)} \\ &= \int_0^t \frac{I_{(\overline{Y}(s)>0)}}{\overline{Y}(s)} [d\overline{N}(s) - \overline{Y}(s) d\Lambda(s)] \qquad \text{(by definition of } M(t)\text{)} \\ &= \int_0^t \frac{d\overline{N}(s)}{\overline{Y}(s)} - \int_0^t I_{(\overline{Y}(s)>0)} d\Lambda(s) \qquad \text{(since } I_{(\overline{Y}(s)>0)} \cdot d\overline{N}(s) = d\overline{N}(s) \text{ almost surely)} \end{aligned}$$

is a martingale. It follows since $\mathcal{M}(0) = 0$, that

$$E\left\{\int_0^t \frac{d\overline{N}(s)}{\overline{Y}(s)}\right\} = E\left\{\int_0^t I_{(\overline{Y}(s)>0))} d\Lambda(s)\right\} \qquad \text{(by the martingale property)} \tag{1}$$

Let $\Lambda^*(t) = \int_0^t I_{(\overline{Y}(s)>0)} d\Lambda(s)$. Then, if $T = \inf\{t : \overline{Y}(t) = 0\}$, $\Lambda^*(t) = \int_0^{t \wedge T} d\Lambda(s) = \Lambda(t \wedge T)$. By Eq. 1, we might expect that $\hat{\Lambda}(t) = \int_0^t \frac{d\overline{N}(s)}{\overline{Y}(s)}$ would be a good "estimator" for $\Lambda^*(t) = \Lambda(t \wedge T)$, but that it would not be possible to obtain an unbiased estimator of $\Lambda(t)$ without making parametric assumptions.

The following theorem summarizes some properties of $\hat{\Lambda}$, an estimator first proposed by Nelson (1969).

Theorem 2.1. *Let $t \geq 0$ be such that $\Lambda(t) < \infty$. Then*

a. $E[\hat{\Lambda}(t) - \Lambda^*(t)] = 0$,

b. $E[\hat{\Lambda}(t) - \Lambda(t)] = -\int_0^t [\prod_{j=1}^n \{1 - \pi_j(s)\}] d\Lambda(s)$

b'. if $\pi_j(s) = \pi(s)$ for all j then

$$E[\hat{\Lambda}(t) - \Lambda(t)] = \int_0^t \{1 - \pi(s)\}^n d\Lambda(s) \geq -\{1 - \pi(t)\}^n \Lambda(t),$$

c. $\sigma_*^2(t) = E[\sqrt{n}\{\hat{\Lambda}(t) - \Lambda^*(t)\}]^2 = E\left[n \int_0^t \frac{I_{(\overline{Y}(s)>0)}}{\overline{Y}(s)} \{1 - \Delta\Lambda(s)\} d\Lambda(s)\right]$

Proof. From equation (1), we conclude that $E[\hat{\Lambda}(t)] = E[\Lambda^*(t)]$, since $\hat{\Lambda}(t) = \int_0^t \frac{d\overline{N}(s)}{\overline{Y}(s)}$ and $\Lambda^*(t) = \int_0^t I_{(\overline{Y}(s)>0)} d\Lambda(s)$, thus establishing (a).

Now, $E[\hat{\Lambda}(t) - \Lambda^*(t)] = 0$ implies

$$\begin{aligned}
E[\hat{\Lambda}(t) - \Lambda(t)] &= E[\Lambda^*(t) - \Lambda(t)] \\
&= E\left[\int_0^t I_{(\overline{Y}(s)>0)} d\Lambda(s) - \int_0^t d\Lambda(s)\right] \\
&= E\left[\int_0^t (I_{(\overline{Y}(s)>0)} - 1) d\Lambda(s)\right] \\
&= E\left[-\int_0^t I_{(\overline{Y}(s)=0)} d\Lambda(s)\right] = -E\left[\int_0^t I_{(\overline{Y}(s)=0)} d\Lambda(s)\right] \\
&= -\int_0^t P(\overline{Y}(s) = 0) d\Lambda(s) \\
&= -\int_0^t P\left(\sum_{i=1}^n \overline{Y_i}(s) = 0\right) d\Lambda(s) \\
&= -\int_0^t P\left(\sum_{i=1}^n I_{(X_i \geq t)} = 0\right) d\Lambda(s)
\end{aligned}$$

and since, $\{\sum_{i=1}^n I_{(X_i \geq t)} = 0\} \equiv \{X_1 \leq t, X_2 \leq t, \ldots, X_n \leq t\}$ we have,

$$E[\hat{\Lambda}(t) - \Lambda(t)] = -\int_0^t \prod_{i=1}^n \{1 - \pi_i(s)\} d\Lambda(s),$$

thus establishing (b).

Note that, if $\pi_j(s) = \pi(s)$ for all j then the above expression changes to,

$$E[\hat{\Lambda}(t) - \Lambda(t)] = -\int_0^t \{1 - \pi(s)\}^n d\Lambda(s) \geq -\{1 - \pi(t)\}^n \Lambda(t)$$

since $\pi(s)$ is non-decreasing, thus establishing (b').

To prove (c), the equation

$$\langle M, M \rangle(t) = \int_0^t \overline{Y}(s) \{1 - \Delta\Lambda(s)\} d\Lambda(s) \tag{2}$$

follows from **Theorem 3.3 (2)**, for our case we replace $M = M_j$ and $A = A_j = \int_0^t \overline{Y}(s) d\Lambda(s)$ and substitute $\Delta A(t) = \overline{Y}(t)\Delta\Lambda(t)$ into the theorem thus yielding the above equation. Therefore,

$$\begin{aligned}
E[\hat{\Lambda}(t) - \Lambda^*(t)]^2 &= E\left\{\int_0^t \frac{I_{(\overline{Y}(s)>0)}}{\overline{Y}(s)} dM(s)\right\}^2 \\
&= E\left\{\int_0^t \frac{I_{(\overline{Y}(s)>0)}}{\overline{Y}^2(s)} d\langle M, M\rangle(s)\right\} \qquad \text{(from \textbf{Corollary 3.1} and \textbf{Theorem 3.4})} \\
&= E\left\{\int_0^t \frac{I_{(\overline{Y}(s)>0)}}{\overline{Y}(s)} \{1 - \Delta\Lambda(s)\} d\Lambda(s)\right\} \qquad \text{(using equation 2)}
\end{aligned}$$

□

Suppose $\pi_j(s) = \pi(s)$ for all j and s. If $\pi(t) > 0$, Theorem 2.1 indicates that $\hat{\Lambda}(t)$ is an unbiased estimator of $\Lambda(t)$, with bias converging to zero at an exponential rate as $n \to \infty$. For the second moment,

$$\begin{aligned}
\sigma_*^2(t) &= E[n\{\hat{\Lambda}(t) - \Lambda^*(t)\}^2] \\
&= E\int_0^t \frac{n}{\overline{Y}(s)} I_{(\overline{Y}(s)>0)} \{1 - \Delta\Lambda(s)\} d\Lambda(s) \qquad \text{(from above theorem)}
\end{aligned}$$

which for large n should approach

$$\sigma^2(t) \equiv \int_0^t \{\pi(s)\}^{-1}\{1 - \Delta\Lambda(s)\} d\Lambda(s)$$

since $E[n\{\hat{\Lambda}(t) - \Lambda(t)\}^2] = E[n\{\Lambda^*(t) - \Lambda(t)\}^2]$ and $E[n\{\Lambda^*(t) - \Lambda(t)\}^2]$ converges to zero when $\pi(t) > 0$, since

$$\sqrt{n}\{\hat{\Lambda}(t) - \Lambda^*(t)\} = \frac{1}{\sqrt{n}} \int_0^t n \frac{I_{(\overline{Y}(s)>0)}}{\overline{Y}(s)} dM_j(s) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \int_0^t \frac{n}{\overline{Y}(s)} dM_j(s),$$

where $\{M_j\}$ is independent and identically distributed collection, and since $n\{\overline{Y}(s)\}^{-1}$ converges to $\{\pi(s)\}^{-1}$, we might expect that $\sqrt{n}\{\hat{\Lambda}(t) - \Lambda(t)\}$ is approximately distributed as $N(0, \sigma^2(t))$ for large n.

The precision of $\hat{\Lambda}$ at time t can be measured either by its variance, $E\{\hat{\Lambda}(t) - E\hat{\Lambda}(t)\}^2$ or since it is biased, by its mean square error, $E[\{\hat{\Lambda}(t) - \Lambda(t)\}^2]$. Since the squared bias satisfies

$$[E\{\hat{\Lambda}(t) - \Lambda(t)\}^2 \leq \{1 - \pi(t)\}^{2n}\{\Lambda(t)\}^2 \qquad \text{(from \textbf{Theorem 2.1(b')})}$$

these will be nearly equal, and the variance of $\hat{\Lambda}(t)$ can be safely used except when $\Lambda(t)$ is large and n is small, or when $\pi(t) = 0$. The variance is given by

$$\begin{aligned}
var[\hat{\Lambda}(t)] &= E\{\hat{\Lambda}(t) - E\hat{\Lambda}(t)\}^2 = E\{\hat{\Lambda}(t) - E\Lambda^*(t)\}^2 \qquad \text{(since } E\hat{\Lambda}(t) = E\Lambda^*(t)\text{)} \\
&= E\{\hat{\Lambda}(t) - \Lambda^*(t) + \Lambda^*(t) - E\Lambda^*(t)\}^2 \\
&= n^{-1}\sigma_*^2(t) + 2E[\{\hat{\Lambda}(t) - \Lambda^*(t)\}\{\Lambda^*(t) - E\Lambda^*(t)\}] + E\{\Lambda^*(t) - E\Lambda^*(t)\}^2 \\
&\approx \frac{1}{n}\sigma_*^2(t),
\end{aligned}$$

for even relatively small values of n. In estimating the variance of $\hat{\Lambda}(t)$, it is therefore sufficient to find a good estimator of $n^{-1}\sigma_*^2(t)$.

Theorem 2.2. *Let $t \geq 0$ be such that $\Lambda(t) < \infty$. Define*

$$\frac{1}{n}\hat{\sigma}^2(t) = \int_0^t \frac{I_{(\overline{Y}(s)>0)}}{\overline{Y}^2(s)} \left\{1 - \frac{\Delta\overline{N}(s) - 1}{\overline{Y}(s) - 1}\right\} d\overline{N}(s),$$

where $0/0 \equiv 0$ as usual. Then

$$E\left\{\frac{1}{n}\hat{\sigma}^2(t) - \frac{1}{n}\sigma_*^2(t)\right\} = \int_0^t P\{\overline{Y}(s) = 1\}\Delta\Lambda(s)d\Lambda(s).$$

Proof.

$$\begin{aligned}
E\left\{\frac{1}{n}\hat{\sigma}^2(t) - \frac{1}{n}\sigma_*^2(t)\right\} &= E\int_0^t \frac{I_{(\overline{Y}(s)>0)}}{\overline{Y}^2(s)}\{d\overline{N}(s) - \overline{Y}(s)d\Lambda(s)\} \\
&\quad - E\int_0^t \frac{I_{(\overline{Y}(s)>1)}}{\overline{Y}^2(s)\{\overline{Y}(s)-1\}}[\{\Delta\overline{N}(s) - 1\}d\overline{N}(s) - \overline{Y}(s)\{\overline{Y}(s)-1\}\Delta\Lambda(s)d\Lambda(s)] \\
&\quad + E\int_0^t I_{(\overline{Y}(s)=1)}\Delta\Lambda(s)d\Lambda(s).
\end{aligned} \tag{3}$$

The first term on the right-hand side of Eq. (3) is zero since it is the expectation at time t of a zero mean martingale. The second term will be zero for the same reason if

$$\int [\{\Delta\overline{N}(s) - 1\}d\overline{N}(s) - \overline{Y}(s)\{\overline{Y}(s)-1\}\Delta\Lambda(s)d\Lambda(s)] \tag{4}$$

is also a zero mean martingale, where an integral without limits denotes a process whose value at time t is the integral over the interval $(0, t]$. Now,

$$\begin{aligned}
&[\{\Delta\overline{N}(s) - 1\}d\overline{N}(s) - \textstyle\int \overline{Y}(s)\{\overline{Y}(s)-1\}\Delta\Lambda(s)d\Lambda(s)] \\
&= \int\{\Delta\overline{N}(s) - 1\}d\overline{N}(s) - \int \Delta\overline{N}(s)\overline{Y}(s)d\Lambda(s) + \int \overline{Y}(s)\Delta\Lambda(s)d\Lambda(s) \\
&\quad + \int \overline{Y}(s)\Delta\Lambda(s)\{d\overline{N}(s) - \overline{Y}(s)d\Lambda(s)\} \qquad \text{(since } \Delta\overline{N}(s) = d\overline{N}(s) \text{ at jump points and 0 everywhere else)} \\
&= \int\{\Delta\overline{N}(s) - \overline{Y}(s)\Delta\Lambda(s)\}\{d\overline{N}(s) - \overline{Y}(s)d\Lambda(s) - \int \overline{Y}(s)\{1 - \Delta\Lambda(s)\}d\Lambda(s) \\
&\quad + \int\{2\overline{Y}(s)\Delta\Lambda(s) - 1\}\{d\overline{N}(s) - \overline{Y}(s)d\Lambda(s)\} \\
&= \int \Delta M(s)dM(s) - \langle M, M\rangle(s) + \int\{2\overline{Y}(s)\Delta\Lambda(s) - 1\}dM(s),
\end{aligned}$$

where $M(s) = \overline{N}(s) - \int \overline{Y}(s)d\Lambda(s)$. Thus, to establish expression (4) is a zero mean martingale, it is sufficient to show that $\int \Delta M(s)dM(s) - \langle M, M\rangle(s)$ is a zero mean martingale which will follow from **Theorem 3.3** if $\int \Delta M_i(s)dM_j(s)$ is a zero mean martingale whenever $i \neq j$.Also, $\int\{2\overline{Y}(s)\Delta\Lambda(s) - 1\}dM(s)$ is a bounded predictable process.

We first show that $\int \Delta N_i(s) dM_j(s)$ is a zero mean martingale when $i \neq j$. For $s < t$,
$E\left\{\int_s^t \Delta N_i(v) dM_j(v) \mid \mathcal{F}_s\right\}$

$$= E\left[\sum_{s<v\leq t} \Delta N_i(v)(\Delta N_j - \Delta A_j)(v) \mid \mathcal{F}_s\right] \text{ (since } dM_j = dN_j - dA_j = \Delta N_j - \Delta A_j \text{ at the jump points)}$$

$$= E\left[\sum_{s<v\leq t} E\{\Delta N_i(v)\Delta N_j(v) \mid \mathcal{F}_{v-}\} \mid \mathcal{F}_s\right] \qquad \text{(since } E\{\Delta N_i \Delta N_j \mid \mathcal{F}_{v-}\} = \Delta N_i \Delta N_j)$$

$$- E\left[\sum_{s<v\leq t} \Delta A_j(v) E\{\Delta N_i(v) \mid \mathcal{F}_{v-}\} \mid \mathcal{F}_s\right]$$

$$= E\left[\sum_{s<v\leq t} E\{\Delta N_i(v) \mid \mathcal{F}_{v-}\} E\{\Delta M_j(v) \mid \mathcal{F}_{v-}\} \mid \mathcal{F}_s\right]$$

where the second equality follows from the predictability of A_j and the third follows because of the assumption that for each $t \geq 0$, given $\mathcal{F}_{t-}$, $\{\Delta N_j(t) : j = 1, \ldots, n\}$ are all independent $0, 1$ random variables for the counting processes $\{N_j(t) : j = 1, \ldots, n\}$. Since ΔA_i is a bounded predictable process, $\int \Delta M_i dM_j = \int \Delta N_i dM_j - \int \Delta A_i dM_j$ is a zero mean martingale whenever $i \neq j$. $\square$

When Λ is a continuous function of t, **Theorem 2.2** indicates that $\hat{\sigma}^2(t)$ provides an unbiased estimator of $\sigma_*^2(t)$. In fact, the theorem suggests that bias arises from situations in which $\Delta\Lambda(s) > 0$ and $\overline{Y}(s) = 1$. the situations here is closely related to problems encountered in estimating a binomial parameter since, when $\Delta\Lambda(s) > 0$,

$$\Delta\Lambda(s) = P\{T = s \mid T \geq s\}.$$

Suppose $\Delta\Lambda(t_0) \equiv p > 0$ for some $t_0 < t$. The "binomial parameter" p has the usual estimator $\hat{p} \equiv \Delta\overline{N}(t_0)/\overline{Y}(t_0)$ with conditional variance, given $\overline{Y}(t_0) \equiv n, \tilde{\sigma}^2 \equiv p(1-p)/n$. When $n > 1$,

$$\begin{aligned}\hat{\tilde{\sigma}}^2 &\equiv \frac{n}{n-1} \frac{\hat{p}(1-\hat{p})}{n} \\ &= \frac{1}{\overline{Y}^2(t_0)} \left\{1 - \frac{\Delta\overline{N}(t_0) - 1}{\overline{Y}(t_0) - 1}\right\} \Delta\overline{N}(t_0)\end{aligned}$$

is an unbiased estimator of $\tilde{\sigma}^2$, while no unbiased estimator exists when $n = 1$. If we use $\hat{\tilde{\sigma}}^2 \equiv \hat{p} = \Delta\overline{N}(t_0)/\overline{Y}(t_0)$ when $n = 1$, a bias of $p^2 = \Delta\Lambda d\Lambda$ will arise in that special case. In summary, then, the classical setting of estimating a binomial parameter motivates $\hat{\Lambda}$ as an "unbiased estimator" of $\Lambda^*(t)$, the form of $\sigma_*^2(t)$ and of its estimator $\hat{\sigma}^2$, and the bias of $\hat{\sigma}^2$ as an estimator of $\sigma_*^2(t)$.

3 Appendix

Theorem 3.1. *Let T be an absolutely continuous failure time random variable and U is a censoring time random variable with an arbitrary distribution. Let $X = min(T, U)$, $\delta = I_{(T \leq U)}$ and let λ denote the hazard function for T. Define*

$$N(t) = I_{(X \leq t, \delta = 1)},$$

$$N^U(t) = I_{(X \leq t, \delta = 0)},$$

$$\mathcal{F}_t = \sigma\{N(s), N^U(s) : 0 \leq s \leq t\}.$$

Then the process M given by

$$M(t) = N(t) - \int_0^t I_{(X \geq u)} d\Lambda(u)$$

is an $\mathcal{F}_t$ martingale if and only if

$$\frac{dF(z)}{1 - F(z-)} = \frac{-dP(T \geq z, U \geq T)}{P(T \geq z, U \geq z)}, \qquad \text{for all } z \text{ such that } P(T \geq z, U \geq z) > 0. \tag{5}$$

Proof. In order to show that $M(t)$ is a martingale, first we need to show that $E|M(t)| < \infty$ for all $t \geq 0$. Note that

$$\begin{aligned}
E\{|M(t)|\} &\leq E\{N(t)\} + E\left\{\int_0^t I_{(X \geq u)} \lambda(u) du\right\} \\
&\leq 1 + \int_0^t P(X \geq u) \lambda(u) du \\
&\leq 1 + \int_0^t P(T \geq u) \lambda(u) du \qquad \text{since } P(X \geq u) = P(T \geq u, U \geq u) \leq P(T \geq u) \\
&= 1 + 1 - S(t) \leq 2
\end{aligned}$$

Thus it only remains to show that $E[M(t+s) \mid \mathcal{F}_t] = M(s)$ almost surely for all $s \geq 0, t \geq 0$ iff Condition 5 holds. Now,

$$\begin{aligned}
E\{M(t+s) \mid \mathcal{F}_t\} &= E\left\{N(t+s) - \int_0^{t+s} I_{(X \geq u)} \lambda(u) du \mid \mathcal{F}_t\right\} \\
&= E\left\{N(t+s) - N(t) + N(t) - \int_0^t I_{(X \geq u)} \lambda(u) du - \int_t^{t+s} I_{(X \geq u)} \lambda(u) du \mid \mathcal{F}_t\right\} \\
&= N(t) - \int_0^t I_{(X \geq u)} \lambda(u) du + E\{N(t+s) - N(t) \mid \mathcal{F}_t\} - E\left\{\int_t^{t+s} I_{(X \geq u)} \lambda(u) du \mid \mathcal{F}_t\right\}
\end{aligned}$$

The first two terms on the right-hand side are $M(t)$, so it remains to show that the last two terms are equal to zero almost surely. Now,

$$\begin{aligned}
E\{N(t+s) - N(t) \mid \mathcal{F}_t\} &= E\{I_{(X \leq t+s, \delta=1)} - I_{(X \leq t, \delta=1)} \mid \{N(u), N^U(u) : 0 \leq u \leq t\}\} \\
&= E\{I_{(t < X \leq t+s, \delta=1)} \mid \{N(u), N^U(u) : 0 \leq u \leq t\}\}
\end{aligned}$$

If either N or N^U has jumped at or before time t, then $I_{(t<X\leq t+s,\delta=1)} = 0$, so the conditional expectation must be 0 on the set $\{X \leq t\}$. On the $\mathcal{F}_t$ set $\{N(u) = N^U(u) = 0 \text{ for any } u \in [0,t]\} = \{X > t\}$.

Now, given a random variable Y on $(\Omega, \mathcal{F}, P)$ and let $G \subseteq \mathcal{F}$. Then if X satisifes:

- X is $G-$measureable,
- $\int_B Y dP = \int_B X dp$ for all $B \in G$,

then $X = E[Y \mid G]$.

So, $N(t) = I_{(X\leq t,\delta=1)}$ indicates whether failure has occured by time t and $N^U(t) = I_{(X\leq t,\delta=0)}$ indicates whether censoring has occurred by time t.

Thus, knowing $\mathcal{F}_t$ is equivalent to knowing whether $\{X \leq t\}$ or $\{X > t\}$, since if $X > t$, then $N(t) = N^U(t) = 0$ and if $X \leq t$, then $N(t) = 1$ or $N^U(t) = 1$. Therefore, given $\mathcal{F}_t$. we can partition the probability space into two sets, $\{X \leq t\}$ and $\{X > t\}$.

Now if failure or censoring has already happened by time t, then $N(t+s) = N(t) \implies E\{N(t+s) - N(t) \mid X \leq t\} = 0$, but for $\{X > t\}$, we have the conditional expectation of $I_{(t<X\leq t+s,\delta=1)}$ must be a constant since it is $\{X > t\}$ measurable. Thus,

$$\begin{aligned} E[I_{\{t<X\leq t+s,\delta=1\}} \mid X > t] &= P(t < X \leq t+s, \delta = 1 \mid X > t) \\ &= \frac{P(t < X \leq t+s, \delta = 1)}{P(X > t)} \end{aligned}$$

$$\begin{aligned} \implies E[N(t+s) - N(t) \mid \mathcal{F}_t] &= \begin{cases} 0, & X \leq t \\ \frac{P(t<X\leq t+s,\delta=1)}{P(X>t)}, & X > t \end{cases} && (6) \\ &= I_{\{X>t\}} \frac{P(t < X \leq t+s, \delta = 1)}{P(X > t)} && (7) \end{aligned}$$

Now, for $\{X \leq t\}, \int_t^{t+s} I_{\{X\geq u\}}\lambda(u)du = 0 \implies E\left[\int_t^{t+s} I_{\{X\geq u\}}\lambda(u)du \mid \mathcal{F}_t\right] = 0$ and for $\{X > t\}$,

$$E\left[\int_t^{t+s} I_{\{X\geq u\}}\lambda(u)du \mid X > t\right] = \frac{\int_t^{t+s} P(X \geq u)\lambda(u)du}{P(X > t)} \tag{8}$$

Therefore, $E[N(t+s) - N(t) \mid \mathcal{F}_t] = E\left[\int_t^{t+s} I_{\{X\geq u\}}\lambda(u)du \mid \mathcal{F}_t\right]$ a.s., is equivalent to,

$$I_{\{X>t\}} \frac{P(t < X \leq t+s, \delta = 1)}{P(X > t)} = I_{\{X>t\}} \frac{\int_t^{t+s} P(X \geq u)\lambda(u)du}{P(X > t)}, \text{ a.s.} \tag{9}$$

If $P(X > t) = 0, I_{\{X>t\}} = 0$ a.s. and if $P(X > t) > 0$,

$$
\begin{aligned}
P(t < X \leq t+s, \delta = 1) &= P(t < T \leq t+s, T \leq U) \\
&= \int_{z=t}^{t+s} -dP(T \geq z, U \geq T) \\
&= \int_{z=t}^{t+s} -P(X \geq z)\frac{dP(T \geq z, U \geq T)}{P(T \geq z, U \geq z)} \\
&= \int_{z=t}^{t+s} P(X \geq z)\lambda^{\#}(z)dz, \qquad \text{since } \frac{-dP(T \geq z, U \geq T)}{P(T \geq z, U \geq z)} = \lambda^{\#}(z)
\end{aligned}
$$

Therefore, $\int_t^{t+s} P(X \geq z)\lambda^{\#}(z)dz = \int_t^{t+s} P(X \geq z)\lambda(z)dz$ a.s. if eqn.(5) holds which as we have shown is equivalent to $E[M(t+s) \mid \mathcal{F}_t] = M(t)$ a.s. □

Lemma 3.1. *Suppose we have n independent and identically distributed pairs $(T_i, U_i), i = 1, \ldots, n$, where each pair satisfies equation (5). Let $X_i = min(T_i, U_i)$ and $\delta_i = I_{(X_i = T_i)}$ and let $N_i(t) = I_{(X_i \leq t, \delta_i = 1)}$ and $N_i^U(t) = I_{(X_i \leq t, \delta_i = 0)}$. Let λ denote the hazard function for t_i.*

The process $M_i(t) = N_i(t) - \int_0^t I_{(X_i \geq u)}\lambda_i(u)du$ is a martingale w.r.t the richer σ-algebras $\{\mathcal{F}_t^n : t \geq 0\}$, where $\mathcal{F}_t^n = \sigma[\{N_i(s), N_i^U(s) : 0 \leq s \leq t\}, i = 1, \ldots, n]$. It immediately follows that $\{\overline{N_i}(t) - \int_0^t \overline{Y_i}(s)d\Lambda s\}$ is a martingale w.r.t $\{\mathcal{F}_t^n : t \geq 0\}$ where $\overline{N_i}(t) = \sum_{j=1}^{n_i} N_{ij}(t)$ and $\overline{Y_i}(t) = \sum_{j=1}^{n_i} Y_{ij}(t)$.

Lemma 3.2. *Let $\{\mathcal{F}_t : t > 0\}$ be a filtration and X a left-continuous real- valued process adapted to $\{\mathcal{F}_t : t > 0\}$. Then X is predictable.*

Proof. let I_A denote the indicator of any interval $A \subset \mathcal{R}+$. We show that a left-continuous process X is a limit of predictable processes, and hence must be predictable. Let

$$X^n(t, \omega) = X(0, \omega)I_{[0]}(t) + \sum_{k=0}^{\infty} X(k/n, \omega)I_{(\frac{k}{n}, \frac{k+1}{n}]}(t)$$

Each term in the infinite sum of X^n is predictable since each is the limit of sums of simple predictable processes. hence, X^n is a predictable process. Because X is a.s. left continuous,

$$X(t) = \lim_{n \to \infty} X^n(t)$$

except on a set of probability zero. □

Theorem 3.2. *Let N be a counting process with $E[N(t)] < \infty$ for any t. Let $\{\mathcal{F}_t : t > 0\}$ be a right-continuous filtration such that*

1. *$M = N - A$ is an $\mathcal{F}_t$-martingale, where $A = \{A(t) : t > 0\}$ is an increasing $\mathcal{F}_t$-predictable process with $A(0) = 0$;*

2. *H is a bounded, $\mathcal{F}_t$-predictable process.*

Then the process L given by

$$L(t) = \int_0^t H(u)dM(u)$$

is an $\mathcal{F}_t$-martingale.

Proof. Let $\mathcal{S}$ denote the class of measurable rectangles: $\begin{cases}[0]\times A, & A\in\mathcal{F}_0\\ (a,b]\times A, & A\in\mathcal{F}_a\end{cases}$ along with the empty set ϕ. It is easy to show that $\mathcal{S}$ is closed under finite intersection since $\mathcal{F}_0$ and $\mathcal{F}_a$ are both $\sigma-$algebras.

Let $\mathcal{H}$ denote the vector space of bounded, measurable and adapted processes H such that $\int HdM$ is a martingale with respect to $\{\mathcal{F}_t : t>0\}$. $\mathcal{H}$ obviously contains the constant functions, and also contains I_B where $B\subset\mathcal{R}^+\times\Omega$ and belongs to the class $\mathcal{S}$. Now let $H_n, n=1,2,\dots,$ be an increasing sequence of mappings from $\mathcal{R}^+\times\Omega$ to $\mathcal{R}$ in $\mathcal{H}$ such that $\sup_n H_n \equiv H$ is bounded on a set of probability one. We must show that $\int HdM$ is a martingale.

For any t,

$$E\left|\int_0^t H(u)dM(u)\right|<\infty$$

Each process $\int H_n dM$ is a martingale and consequently an adapted process. Since, $\int_0^t H(u)dM(u)$ is the pointwise limit of the $\mathcal{F}_t$-measurable variables $\int_0^t H_n(u)dM(u)$, hence, $\int_0^t H(u)dM(u)$ is itself $\mathcal{F}_t$-measurable. Finally,

$$\begin{aligned}E\left\{\int_0^{t+s} H(u)dM(u)\mid\mathcal{F}_t\right\} &= E\left\{\int_0^{t+s}\lim_{n\to\infty} H_n(u)dM(u)\mid\mathcal{F}_t\right\}\\ &= E\left\{\lim_{n\to\infty}\int_0^{t+s} H_n(u)dM(u)\mid\mathcal{F}_t\right\}\\ &= \lim_{n\to\infty}\int_0^t H_n(u)dM(u)\\ &= \int_0^t H(u)dM(u),\end{aligned}$$

where all changes of limits with integrals or conditional expectations follow from the Monotone Convergence Theorem. □

Theorem 3.3. *Let $N_j : j=1,\dots,n$ be a collection of counting processes on a stochastic basis $(\Omega,\mathcal{F},\{\mathcal{F}_t, t\geq 0\},P)$, A_j the $\mathcal{F}_t$-compensator of N_j, and $M_j = N_j - A_j$. Then,*

1. *$\int\Delta M_j dM_j - \int(1-A_j)dA_j$ is an $\mathcal{F}_t$-martingale.*
2. *$\langle M_j, M_j\rangle = \int(1-\Delta A_j)dA_j$; that is, $\int(1-\Delta A_j)dA_j$ is the unique predictable right-continuous increasing process which is a.s. zero at $t=0$ and finite at any t, such that $M_j^2-\int(1-\Delta A_j)dA_j$ is a local martingale.*
3. *If $N_1, N_2,\dots,N_n$ is a multivariate counting process, $\langle M_i, M_j\rangle = -\int\Delta A_i dA_j$ when $i\neq j$.*
4. *Suppose $E[N_i(t)]<\infty$ for any t,i. Then the local martingales mentioned above are martingales.*

Proof. We only provide the proof the statements (1) and (2). Using the integration by parts formula fro Lebegue-Stieljes integrals,

$$M_j^2(t) = 2\int_0^t M_j(s-)dM_j(s)+\int_0^t\Delta M_j(s)dM_j(s)$$

In turn,

$$\int_0^t \Delta M_j(s) dM_j(s) = \sum_{s\le t} \Delta N_j(s)\{\Delta N_j(s) - \Delta A_j(s)\} - \int_0^t \Delta A_j(s) dM_j(s)$$
$$= \sum_{s\le t} \Delta N_j(s) - \int_0^t \Delta A_j(s) dA_j(s) - 2\int_0^t A_j(s) dM_j(s),$$

so,

$$\int \Delta M_j dM_j - \int (1 - A_j) dA_j = M_j - 2\int_0^t \Delta A_j dM_j.$$

Since, ΔA_j is bounded and predictable and M_j is a local martingale of locally bounded variation, hence, $\int_0^t \Delta A_j dM_j$ is a local square integrable martingale, thus proving (1).

Since, M_j is a compensated local martingale, we can conclude that

$$2\int M_j(s-) dM_j(s) + M_j - 2\int \Delta A_j(s) dM_j(s)$$

is a local square integrable martingale. The predictable process $\int (1 - A_j) dA_j$ is increasing and consequently the Doob-Meyer Decomposition yields (2). □

Lemma 3.3. *Let $\{N_j : j = 1, \ldots, n\}$ be a collection of counting processes such that for each $t \ge 0$, given $\mathcal{F}_{t-}$, $\{\Delta N_1(t), \ldots, \Delta N_n(t)\}$ are independent 0, 1 random variables. Set $M_j = N_j - A_j$, where A_j is the compensator for N_j. Then for any $i \ne j$ and $t \ge 0$,*

$$\langle M_i, M_j\rangle(t) = 0 \quad a.s.$$

Proof. By integration by parts,

$$M_1(t)M_2(t) = M_1(0)M_2(0) + \int_0^t M_1(s-) dM_2(s) + \int_0^t M_2(s-) dM_1(s) + \int_0^t \Delta M_1(s) dM_2(s).$$

The first term on the right hand side of the equation is zero a.s., and since, $\{M_i(s-) : s \ge 0\}$ is an adapted left-continuous process with right-hand limits, the next two are local square integrable martingales. By the almost sure finiteness of the counting process, we can write the last term as finite sum,

$$\int_0^t \Delta M_1(s) dM_2(s) = \sum_{0<s\le t} \Delta M_1(s)\Delta M_2(s)$$

It remains to show this last term is a local martingale. Let $\{\tau_n\}$ be a localizing sequence for M_1 and M_2, and $0 \le u \le t$. Then

$$E\left\{\int_u^t \Delta M_1(s\wedge\tau_n) dM_2(s\wedge\tau_n) \mid \mathcal{F}_u\right\}$$
$$= \sum_{u<s\le t} E[E\{\Delta M_1(s\wedge\tau_n)\Delta M_2(s\wedge\tau_n) \mid \mathcal{F}_{s-}\} \mid \mathcal{F}_u] = 0$$

□

Corollary 3.1. *If the local square integrable martingale M is a martingale with $M(0) = 0$ a.s.,* *then for any* $t \geq 0$,

$$EM^2(t) = E\langle M, M\rangle(t).$$

Proof. Suppose $EM^2(t) < \infty$. Then, by the Doob-Meyer Decomposition, $M^2 - \langle M, M\rangle$ is a martingale on $[0,t]$ and $EM^2(t) = E\langle M, M\rangle(t)$.

If $EM^2(t) = \infty, E\langle M, M\rangle(t) = \infty$ as well since

$$EM^2(t) \leq \lim_{n\to\infty} EM^2(t \wedge \tau_n) = E\langle M, M\rangle(t),$$

where $\{\tau_n\}$ is a localizing sequence for M. □

Theorem 3.4. *Assume that on a stochastic basis* $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$:

1. H_i *is locally bounded* $\mathcal{F}_t$*-predictable process.*
2. N_i *is a counting process.*

Then for the local martingales $M_i = N_i - A_i$,

$$\left\langle \int H_1 dM_1, \int H_2 dM_2 \right\rangle = \int H_1 H_2 d\langle M_1, M_2\rangle; \tag{10}$$

that is, the process

$$\int H_1 dM_1 \int H_2 dM_2 - \int H_1 H_2 d\langle M_1, M_2\rangle$$

is a local martingale over $[0, \infty)$.

Proof. Since H_i is locally bounded $\mathcal{F}_t$-predictable process, N_i is a counting process and M_i are local martingales hence, each $\int H_i dM_i$ is a local square integrable martingale on $[0, \infty)$. This implies that there exists a unique predictable right-continuous increasing process

$$\left\langle \int H_i dM_i, \int H_i dM_i \right\rangle$$

and is given almost surely by

$$\left\langle \int H_i dM_i, \int H_i dM_i \right\rangle(t) = \lim_{n\to\infty} \left\langle \int_0^{\cdot\wedge\tau_n^i} H_i dM_i, \int_0^{\cdot\wedge\tau_n^i} H_i dM_i \right\rangle(t), \tag{11}$$

where $\{\tau_n^i\}$ is a localizing sequence for $\int H_i dM_i$. The localizing sequences $\{\tau_n^i\}$ can be taken to be

$$\tau_n^i = n \wedge \sup\{t : N_i(t) < n\} \wedge \tau_n^{0i} \wedge \tau_n^{1i},$$

where $\{\tau_n^{0i}\}$ is such that $\sup_{0\leq t} A_i(t \wedge \tau_n^{0i}) \leq n$ a.s., and $\{\tau_n^{1i}\}$ is a sequence rendering H_i locally bounded. Since $H_i(. \wedge \tau_n^i)$ and $N_i(. \wedge \tau_n^i)$ are bounded and $M_i(. \wedge \tau_n^i$ is a square integrable martingale, it follows that

$$\begin{aligned}\left\langle \int_0^{\cdot\wedge\tau_n^i} H_i dM_i, \int_0^{\cdot\wedge\tau_n^i} H_i dM_i \right\rangle(t) &= \int_0^t H_i^2(u \wedge \tau_n^i) d\langle M_i, M_i\rangle(u \wedge \tau_n^i) \\ &= \int_0^{t\wedge\tau_n^i} H_i^2(u) d\langle M_i, M_i\rangle(u) \text{ a.s.}\end{aligned}$$

this last equation and Eq.11 establishes Eq.10 for the case when both $H_1 = H_2$ and $M_1 = M_2$. When $i \neq j$, since $\int H_i dM_i, \int H_j dM_j$ are right-continuous integrable local square martingale on $[0, \infty)$ thus there uniquely exists predictable right-continuous process $\langle \int H_i dM_i, \int H_j dM_j \rangle$ and is given almost surely by

$$\lim_{n \to \infty} \left\langle \int_0^{\cdot \wedge \tau_n} H_i dM_i, \int_0^{\cdot \wedge \tau_n} H_j dM_j \right\rangle (t) \tag{12}$$

where $\{\tau_n^i\}$ is a localizing sequence for $\int H_i dM_i$, and $\tau_n \equiv \tau_n^i \wedge \tau_n^j$. Each of τ_n^i and τ_n^j can be taken as in the first part of the proof and then it follows

$$\begin{aligned} \left\langle \int_0^{\cdot \wedge \tau_n} H_i dM_i, \int_0^{\cdot \wedge \tau_n} H_j dM_j \right\rangle (t) &= \int_0^t H_i(u \wedge \tau_n) H_j(u \wedge \tau_n) d\langle M_i, M_j \rangle (u \wedge \tau_n) \\ &= \int_0^{t \wedge \tau_n} H_i(u) H_j(u) d\langle M_i, M_j \rangle (u \wedge \tau_n) \text{ a.s.} \end{aligned}$$

For $i \neq j$, the result now follows from Eq.12 and the last equation. $\square$