BIO 244: Unit 10

Lebesgue-Stieltjes Integrals, Martingales, Counting Processes

This section introduces Lebesgue-Stieltjes integrals, and defines two important stochastic processes: a martingale process and a counting process. It also introduces compensators of counting processes.

Definition: Suppose $G(\cdot)$ is a right-continuous, nondecreasing step function having jumps at $x_1, x_2, \ldots$. Then for any function $f(\cdot)$, we define the integral

$$\int_a^b f(x)\, d\, G(x) \stackrel{\text{def}}{=} \sum_{\substack{j: \\ a<x_j\leq b}} f(x_j)\cdot(G(x_j)-G(x_j-)) = \sum_{\substack{j: \\ a<x_j\leq b}} f(x_j)\cdot \Delta G(x_j),$$

where $\Delta G(x_j) = G(x_j) - G(x_{j-1})$. This is called a Lebesgue-Stieltjes integral. If $G(\cdot)$ is continuous with derivative $g(\cdot)$, then we define $\int_a^b f(x)dG(x)$ to be the Lebesgue integral $\int_a^b f(x)g(x)dx$. Thus, we can define a Lebesgue-Stieltjes integral $\int f(x)dG(x)$ for $G(\cdot)$ either absolutely continuous or a step function.

Let's illustrate what this gives in several examples:

Example 10.1: Survival data; 1-sample problem

$$(U_i, \delta_i) \qquad i = 1, 2, \ldots, n.$$

Define the stochastic process $G(\cdot)$ by:

$$\begin{aligned} G(t) &\stackrel{\text{def}}{=} \sum_{i=1}^{n} 1(U_i \leq t,\ \delta_i = 1) \\ &= \#\text{ of failures observed on or before } t. \end{aligned}$$

As before, let $\tau_1 < \tau_2 < \cdots < \tau_K$ denote the distinct failure times,

$$\begin{aligned} d_j &= \text{\# failures at } \tau_j \\ Y(\tau_j) &= \text{\# at risk at } \tau_j. \end{aligned}$$

Then, for example, if $f(t) = t$, $\displaystyle\int_0^\infty f(t)\, d\, G(t) = \sum_{j=1}^K f(\tau_j)\ \Delta G(\tau_j)$

$$= \sum_{j=1}^K \tau_j \cdot d_j.$$

Or, if $f(t) = \displaystyle\sum_{i=1}^n 1\,(U_i \geq t) = \text{\# at risk at } t\text{=Y(t)},$

$$\begin{aligned} \int_0^\infty f(t)\, d\, G(t) &= \sum_{j=1}^K f(\tau_j)\, \Delta G(\tau_j) \\ &= \sum_{j=1}^K Y(\tau_j) \cdot d_j. \end{aligned}$$

Example 10.2: Survival data; 2-sample problem

$$(U_i, \delta_i, Z_i) \quad i = 1, 2, \ldots, n; \quad Z_i = \begin{cases} 0 & \text{group } 0 \\ 1 & \text{group } 1. \end{cases}$$

Define, for $l = 0, 1$

$$\begin{aligned} N_l(t) &= \sum_{i=1}^n 1\,(U_i \leq t, \delta_i = 1, Z_i = l) && \text{\# failures in group } l \\ Y_l(t) &= \sum_{i=1}^n 1\,(U_i \geq t, Z_i = l) && \text{\# at risk in group } l. \end{aligned}$$

That is, $N_l(t)$ and $Y_l(t)$ are the number of observed failures by time t and number at risk at time t in group l. Using the same notation as when we introduced the logrank test, consider

$$W \stackrel{\text{def}}{=} \int_0^\infty \frac{Y_0(t)}{Y_0(t)+Y_1(t)}\, dN_1(t) \; - \; \int_0^\infty \frac{Y_1(t)}{Y_0(t)+Y_1(t)}\, dN_0(t).$$

Then

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

i.e., W is just the numerator of logrank statistic.

Later we will see that this expression is very useful to study the properties of the logrank test.

Example 10.3: 1-sample problem

$$(U_i, \delta_i) \qquad i = 1, 2, \ldots, n.$$

Consider

$$\hat{\Lambda}(t) \; = \; \int_0^t \frac{dN(u)}{Y(u)},$$

where

$$N(t) \;=\; \sum_{i=1}^{n} 1\,(U_i \leq t, \delta_i = 1)$$

and

$$Y(t) \;=\; \sum_{i=1}^{n} 1\,(U_i \geq t).$$

Then it is easily shown, using the notation used for the Kaplan-Meier estimator, that

$$\hat{\Lambda}(t) \;=\; \sum_{\tau_j \leq t} \frac{d_j}{Y(\tau_j)} \;=\; \text{Nelson-Aalen estimator.}$$

Note 1: The Lebesgue-Stieltjes integrals in these examples involve random quantities and hence are called stochastic integrals.

Note 2: The examples illustrate that some of the statistics we considered in Units 5 and 6 can be written as stochastic integrals. In subsequent units, we will consider theorems that will enable us to determine the properties of such stochastic integrals, and thus be able to prove useful results for the statistics considered earlier.

Let $X(\cdot) = \{X(t) : \;\; t \geq 0\}$ be a stochastic process, and let $\mathcal{F}_t$ denote the σ–algebra generated by the random variables $(X(u) : \;\; 0 \leq u \leq t)$. The increasing family of σ-algebras $(\mathcal{F}_t : \;\; t \geq 0)$ is called a filtration, and we say that $X(\cdot)$ is adapted to $(\mathcal{F}_t \; : \; t \geq 0)$, since once $\mathcal{F}_t$ is known, $X(t)$ is known (or: $X(t)$ is $\mathcal{F}_t$-measurable).

For any $t \geq 0$ and $s \geq 0$, define

$$E[X(t+s) \mid \mathcal{F}_t] \;=\; E[X(t+s) \mid X(u), 0 \leq u \leq t].$$

The values $(X(u) \; : \; 0 \leq u \leq t)$ are called the history of $X(\cdot)$ from 0 to t.

For more about conditional expectations we refer to the Appendix.

Definition: Suppose $X(\cdot)$ is a right-continuous stochastic process with left-hand limits that is adapted to $(\mathcal{F}_t)$. $X(\cdot)$ is a martingale if

$$\text{(a)} \quad E\,|X(t)| < \infty \quad \forall t, \quad \text{and}$$

$$\text{(b)} \quad E[X(t+s) \mid \mathcal{F}_t] \overset{\text{a.s.}}{=} X(t) \quad \forall t \geq 0, s \geq 0.$$

The idea is that given the information at time t ($\mathcal{F}_t$), the expected value of $X(t+s)$ is equal to $X(t)$. Again, for more about conditional expectations see the Appendix.

Also, $X(\cdot)$ is called a sub-martingale if "=" in (b) is replaced by "$\geq$", and called a super-martingale if "=" in (b) is replaced by "$\leq$".

Note 1: As before, by saying that $X(\cdot)$ has left hand limits, we mean that with probability 1, the limit of X(u) as u $\uparrow$ t exists for every t. I.e., there exists set A with probability 0 (not depending on t) such that on $\Omega \setminus A$, for every t, the limit of X(u) as u $\uparrow$ t exists.

Note 2: Martingale processes are commonly used in economics to model fluctuations in financial processes. Here the response is some continuous random quantity such as the value of a commodity. In failure time applications, the martingales we will deal with usually consist of the arithmetic difference of a counting process and a so-called "compensator" process.

Example 10.4: Let us consider a random walk process, which is a simple example of a discrete time martingale. Suppose that $Y_1, Y_2, \ldots$ are i.i.d. random variables satisfying

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

$$\Longrightarrow \qquad E(Y_i) = 0, \quad Var(Y_i) = 1.$$

Define

$$X(0) = 0, \quad \text{and}$$
$$X(n) = \sum_{j=1}^{n} Y_j \qquad n = 1, 2, \ldots$$

It follows that $E(X(n)) = 0$ and $Var(X(n)) = n$. Also, knowing

$$(X(u) \ : \ 0 \le u \le n) = (X(1), \ldots, X(n)),$$

is equivalent to knowing $(Y_1, \ldots, Y_n)$. Hence, $\mathcal{F}_n = \sigma(X(1), \ldots, X(n)) = \sigma(Y_1, \ldots, Y_n)$.

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

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

Thus, $X(\cdot)$ is a martingale.

<u>Note 1:</u> We could have made $X(\cdot)$ a continuous-time process by taking

$$X(t) \stackrel{\text{def}}{=} \sum_{j=1}^{[t]} Y_j, \quad \text{where } [t] = \text{ greatest integer } \le t.$$

Note 2: This example illustrates that, given the history of a martingale process up to time t, the expected value at some future time is $X(t)$.

Note 3: It follows from the definition of a martingales (taking $t = 0$) that $E(X(s)) = E(X(0))$ when $X(\cdot)$ is a martingale. That is, the mean function of a martingale process is constant. In most of our applications, X(0)=0, and hence our martingales will have a mean function that is identically zero.

Note 4: Martingales have uncorrelated increments (Exercise).

Definition: A stochastic process $N(\cdot) = (N(t) \ : \ t \geq 0)$ is called a **counting process** if

- $N(0) = 0$
- $N(t) < \infty, \quad \text{all } t$
- With probability 1, $N(t)$ is a right-continuous step function with jumps of size +1.

Example 10.5: 1-sample survival data (assume no ties). Observations: $(U_i, \delta_i) \quad i = 1, 2, \ldots, n$. Define the process $N(\cdot)$ by

$$N(t) \stackrel{\text{def}}{=} \sum_{i=1}^{n} 1 \, (U_i \leq t, \ \delta_i = 1)$$

$$= \# \text{ observed failures in } [0, t].$$

Then $N(\cdot)$ is a counting process.

Example 10.6: Now fix i and define the process $N_i(\cdot)$ by

$$N_i(t) \stackrel{\text{def}}{=} 1\,(U_i \le t,\ \delta_i = 1)$$
$$= \begin{cases} 0 & \text{no failure by } t \\ 1 & \text{failure observed on/before } t. \end{cases}$$

Then $N_i(\cdot)$ is a counting process.

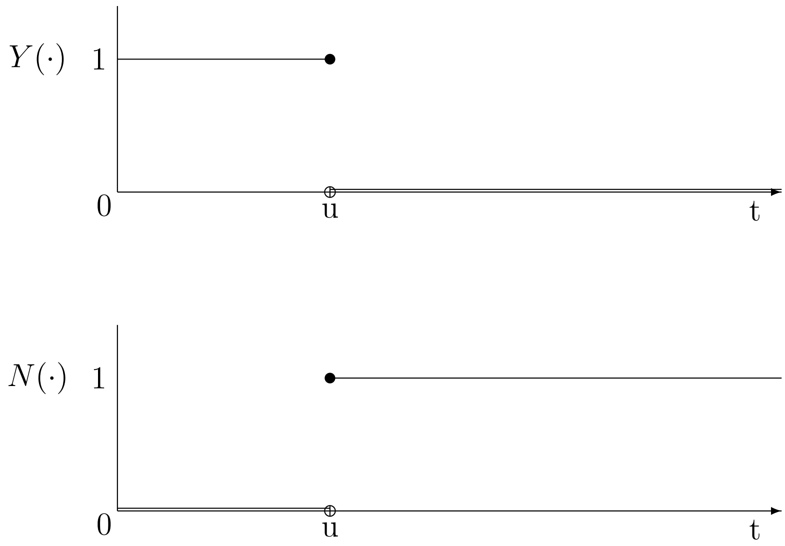
Let's focus on a single subject, and for simplicity drop the subscript i. Suppose T denotes survival time and denote the c.d.f. and hazard functions for T by $F(\cdot)$ and $\lambda(\cdot)$. As before, we denote the observation for this subject by (U, δ) and assume that censoring is noninformative.

Define the processes $N(\cdot)$ and $Y(\cdot)$ by

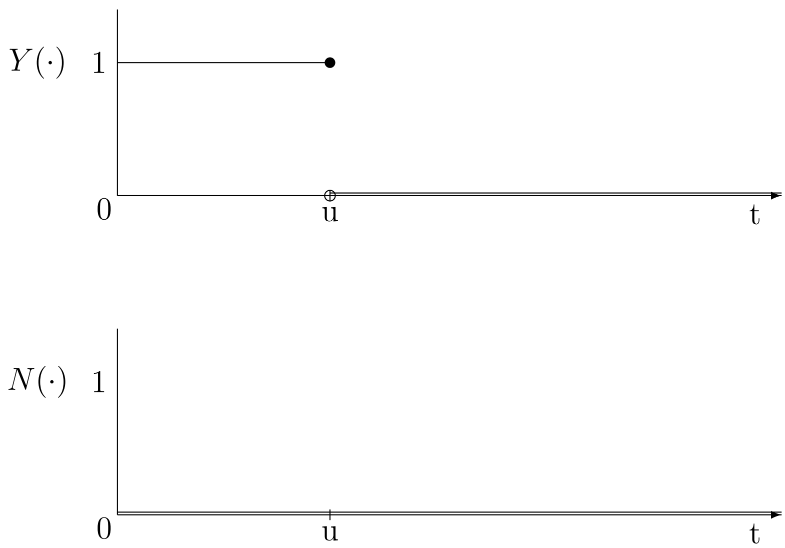
$$N(t) \stackrel{\text{def}}{=} 1\,(U \le t, \delta = 1)$$
$$Y(t) \stackrel{\text{def}}{=} 1\,(U \ge t) \quad = \text{‘at risk’ indicator.}$$

Note that N(t) is just an indicator that a failure is observed by (that is, on or before) time t, and Y(t) is an indicator of whether the subject is at risk at time t. What do these look like?

Observe failure at U=u



Observe censoring at U=u



Note 1: Suppose that $N(\cdot)$ is any counting process adapted to some filtration $\{\mathcal{F}_t : \quad t \geq 0\}$. This means that once $\mathcal{F}_t$ is known, $N(t)$ is known, or more accurately, that $N(t)$ is $\mathcal{F}_t$-measurable. Then since $N(\cdot)$ is nondecreasing, N(t+s) $\geq$ N(t) for every s and t and hence $E[N(t+s) \mid \mathcal{F}_t] \geq N(t)$. Thus every counting process is a submartingale.

Note 2: Notice that the counting process N(t)=1(U$\leq$ t, δ=1) has to do with observing a failure on or before time t. This is not the same as failing on or before t (without regard to censoring status), which is governed by the hazard function $\lambda(\cdot)$. To see the connection between these, note that

$$\begin{aligned}
\lambda(t) \cdot \Delta t &= \frac{f(t)}{S(t-)}\, \Delta t \\
&\approx P\left[t \leq T < t + \Delta t \mid T \geq t\right] \\
&= P\left[t \leq T < t + \Delta t \mid T \geq t \text{ and } C \geq t\right] \qquad \text{(because of noninformative censoring)} \\
&= P\left[t \leq T < t + \Delta t \mid U \geq t\right] \\
&= P\left[t \leq T < t + \Delta t \mid Y(t) = 1\right] \\
&= P\left[N(t + \Delta t-) - N(t-) = 1 \mid Y(t) = 1\right] \\
&= E\left[N(t + \Delta t-) - N(t-) \mid Y(t) = 1\right] \qquad \text{(since } N(t+\Delta t-) - N(t-) = 0 \text{ or } 1\text{).}
\end{aligned}$$

Thus, conditional on $Y(t) = 1$, $\lambda(t) \cdot \Delta t$ is the expected change in the counting process $N(t)$ at time t.

(continuation of Note 2)

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

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

Then

$$\begin{aligned} E(N(t)) &= P(N(t)=1) \\ &= \int_0^t \lambda(u)P(U \geq u)du \qquad \longleftarrow \\ &= \int_0^t \lambda(u)E(Y(u))du \\ &= E\left(\int_0^t \lambda(u)Y(u)du\right) \\ &= E(A(t)). \end{aligned}$$

if $U = \min\,(T, C)$ and $C \sim G(\cdot)$, $P(N(t)=1) = P(T \leq t \text{ and } C \geq T)$ $= \int_0^t \int_u^\infty f(u)g(c)dc\,du$ $= \int_0^t f(u)(1-G(u))du$ $= \int_0^t \lambda(u)(1-F(u))(1-G(u))du$ $= \int_0^t \lambda(u)P(U \geq u)du$

Thus, $\quad M(t) \stackrel{\text{def}}{=} N(t) - A(t) \quad$ has mean 0.

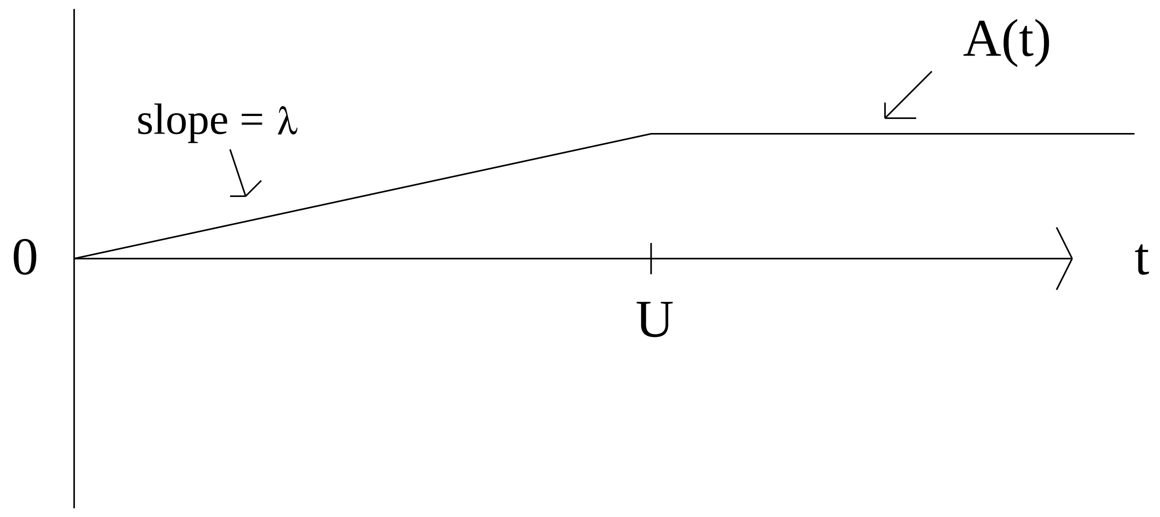
Note: $A(\cdot)$ is called the compensator process corresponding to $N(\cdot)$. It represents the cumulative risk of being observed to fail by time t, whereas $N(t)$ is the indicator of whether we observe a failure by time t. As we will see later, the difference between these two processes, $M(\cdot)$, is a zero-mean martingale.

Example: Suppose $T \sim Exp(\lambda)$

$$\Longrightarrow \quad \lambda(t) = \lambda.$$

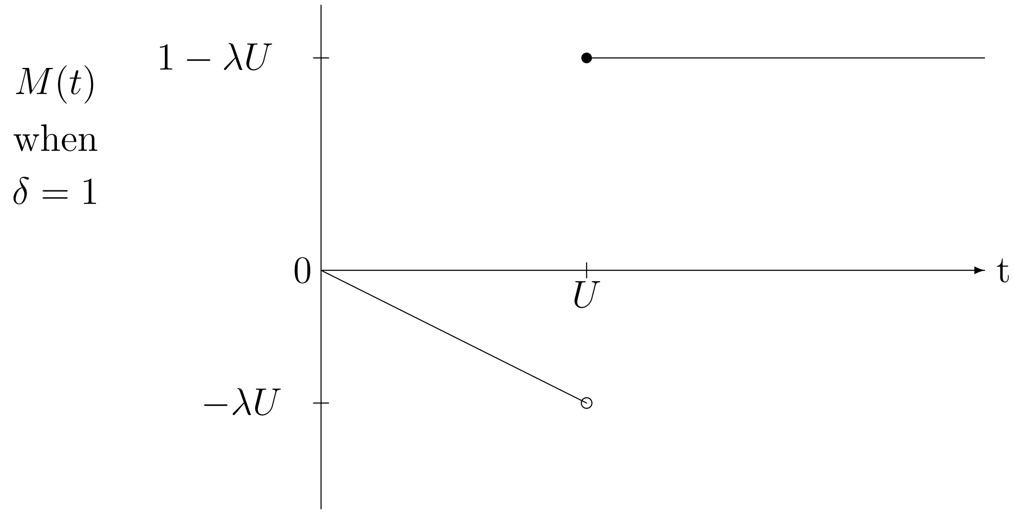
Then

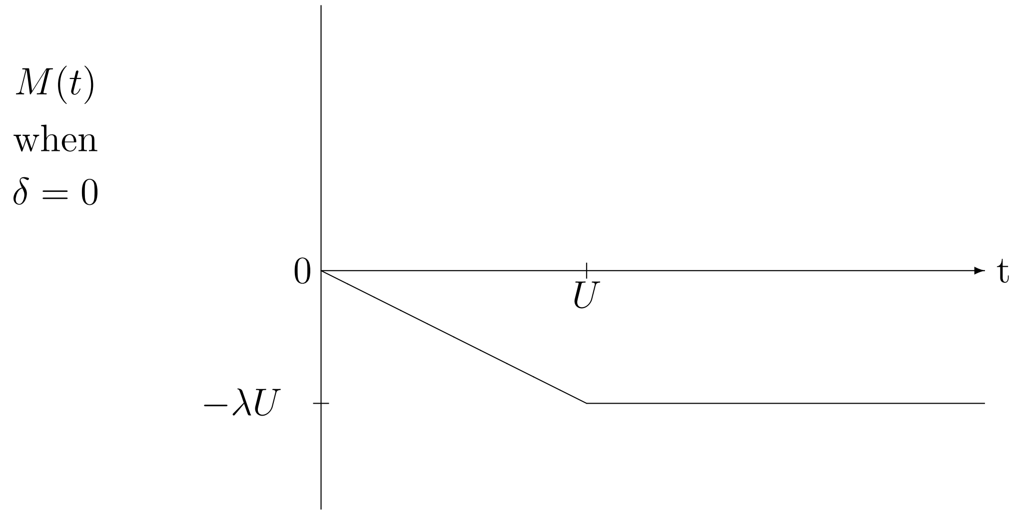
$$\begin{aligned} A(t) &= \int_0^t \lambda(u)Y(u)du = \lambda \int_0^t Y(u)du \\ &= \lambda \cdot \int_0^t 1_{\{u \le U\}}\, du \\ &= \lambda \cdot \min(t, U). \end{aligned}$$



Note that $N(t) = 1_{\{U \le t, \delta = 1\}}$. Thus, $N(t) = 0$ for $t < U$ and for $t \ge U$, $N(t) = 1$ if $\delta = 1$ or $N(t) = 0$ if $\delta = 0$.

Thus, we can combine $N(\cdot)$ and $A(\cdot)$ to get $M(\cdot)$, as shown in the following picture.





For t fixed, $M(t)$ could be

- < 0 (if $t < U$ and sometimes even if $t > U$)
- > 0 (sometimes if $t \geq U$).

For every fixed t, $E(M(t)) = 0$.

Exercises

1. With $Y_l(\cdot)$ and $N_l(\cdot)$ $\quad (l = 0, 1)$ $\quad$ defined as in example 2 on page 3, re-express the following in the old notation:

$$W^* = \int_0^\infty Y_1(u) dN_0(u) \; - \; \int_0^\infty Y_0(u) dN_1(u)$$

2. Consider the 2-sample problem with (U_i, δ_i, Z_i), $\quad i = 1, 2, \ldots, n$ $\quad$ defined in the usual way. Define

$$\begin{aligned} Y_i(t) &= 1\,(U_i \geq t) \\ N_i(t) &= 1\,(U_i \leq t, \delta_i = 1) \end{aligned}$$

Simplify

$$\sum_{i=1}^{n} \int_0^\infty \left[Z_i \; - \; \frac{\sum_{l=1}^{n} Y_l(s) Z_l}{\sum_{l=1}^{n} Y_l(s)} \right] dN_i(s)$$

3. Suppose $\quad Y_1, Y_2, \ldots, Y_n, \ldots \quad$ are *iid* with

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

Define $\quad X(0) = 0 \quad$ and, for $t > 0$, $\quad X(t) = \sum_{j=1}^{[t]} Y_j$.

Define $\mathcal{F}_t$ the filtration generated by X.

Show that $X(\cdot)$ is a martingale with respect to $\mathcal{F}_t$.

4. Let $(U_i, \delta_i), \quad i = 1, 2, \ldots, n$ denote a set of censored survival data.

Define

$$\begin{aligned} N_i(t) &= 1\,(U_i \le t, \delta_i = 1) \qquad t \ge 0 \\ Y_i(t) &= 1\,(U_i \ge t) \qquad t \ge 0 \end{aligned}$$

Then with $N(\cdot) \overset{\text{def}}{=} \sum_{i=1}^{n} N_i(\cdot)$

and $A(\cdot) \overset{\text{def}}{=} \sum_{i=1}^{n} \left\{ \int_0^{\cdot} \lambda(u) Y_i(u) du \right\}$,

let $M(\cdot) = N(\cdot) - A(\cdot)$.

Suppose $\lambda(t) = \lambda$, as in the example on page 13, and that we observe

$n = 8$ and $(U_i, \delta_i) = (4.7, 0),\ (11.6, 0),\ (2.1, 1),\ (5.2, 1),\ (3.4, 0),\ (5.9, 1),$ $(17.3, 0)$, and $(12.1, 1)$.

Plot $N(\cdot), A(\cdot)$, and $M(\cdot)$ for these data, assuming $\lambda = .06$.

5. Suppose that $N(\cdot)$ denotes a homogeneous Poisson Process with parameter $\lambda > 0$. Thus, $N(t)$ represents the number of events that occur by time t. Define the process $M(\cdot)$ by $M(t) = N(t) - \lambda t$. Is $M(\cdot)$ a martingale? If so, with respect to which filtration? Justify your answer. You may assume the memoryless property of Poisson processes; e.g.,

$$P(\text{event in } (t, t+s] | \text{history of process up to time } t) = 1 - e^{-\lambda s}$$

and that for any given $t, N(t)$ has a Poisson distribution.

6. Suppose that $\mathcal{F}_t$ is a filtration, and that $X(\cdot)$ is adapted to that filtration. Show that for all $s \le t$, $X(s)$ is $\mathcal{F}_t$-measurable. Interpretation: the complete history of X until t is known from the information at t.

7. Suppose that $\mathcal{A}$ is a σ-algebra, and $\mathcal{A}_1$ and $\mathcal{A}_2$ are σ-algebra's with $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \mathcal{A}$. Suppose that X is a random variable that is measurable with respect to $\mathcal{A}$ (you can think: just some kind of random variable). Show that $E\left[E\left[X|\mathcal{A}_2\right]|\mathcal{A}_1\right] = E\left[X|\mathcal{A}_1\right]$ (this is called the tower property of conditional expectations).

Appendix with Unit 10

This appendix briefly introduces conditional expectations and filtrations. For more about conditional expectations, see Billingsley, Probability and Measure, Chapter 34. For more about filtrations, see Andersen et al.

1. Recall that a *scalar* random variable X on $(\Omega, \mathcal{F}, P)$ with values in $\mathbb{R}$ is called ***measurable*** if $\{\omega : X(\omega) \leq x\} \in \mathcal{F}$ for every x in $\mathbb{R}$. Equivalently, $X^{-1}((\infty, x]) \in \mathcal{F}$ for every x in $\mathbb{R}$. Recall that we mentioned this is equivalent to $X^{-1}(B) \in \mathcal{F}$ for every Borel-set B (with Borel sets the σ-algebra generated by the open sets in $\mathbb{R}$).

2. A random variable on $(\Omega, \mathcal{F}, P)$ with values in a topological space (e.g., a metric space), is called ***measurable*** if $X^{-1}(B) \in \mathcal{F}$ for every Borel-set B (with Borel sets the σ-algebra generated by the open sets).

3. Suppose that X is integrable and $\mathcal{F}$-measurable, and $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$. Then there exists a random variable $E[X|\mathcal{G}]$ with:

   (a) $E[X|\mathcal{G}]$ is $\mathcal{G}$-measurable and integrable

   (b) $\int_G E[X|\mathcal{G}]dP = \int_G XdP$ for every $G \in \mathcal{G}$.

   Such variable is called the ***conditional expecation*** of X with respect to $\mathcal{G}$. It is almost surely unique.

4. ***Example*** of conditional expectation: suppose that B_1, B_2, $\ldots$ is a countable partition of Ω generating $\mathcal{G}$, with $P(B_i) > 0$ for every i. Recall that a partition covers the entire space, and its members are disjoint. Then
   $$E[X|\mathcal{G}](\omega) = \frac{\int_{B_i} XdP}{P(B_i)}$$
   for $\omega \in B_i$. Intuition: "$E[X|B_i]$". Proof: $\mathcal{G}$ consists of sets of the form $\cup_{\text{some } i} B_i$. Thus, $E[X|\mathcal{G}]^{-1}(x)$ has to be a union of B_i, for every x, so that $E[X|\mathcal{G}]$ has to be constant on each B_i. By definition of conditional expectation,
   $$\int_{B_i} E[X|\mathcal{G}]dP = \int_{B_i} XdP.$$
   The left hand side has to be equal to the value of $E[X|\mathcal{G}]$ on B_i times $P(B_i)$. That concludes the proof.

5. ***Example*** of conditional expectation: If (X, Y) is continuous, then $E[Y|\sigma(X)] = \int y f_{Y|X}(y) dy$.

6. ***Theorem***: for constants a and b, $E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$. Proof: measurability follows since sums of measurable random variables are measurable. So, we still have to check:
$$\int_G (aE[X|\mathcal{G}] + bE[Y|\mathcal{G}])\, dP = \int_G (aX + bY)\, dP.$$
But the left hand side equals
$$\begin{aligned}\int_G (aE[X|\mathcal{G}] + bE[Y|\mathcal{G}])\, dP &= a\int_G E[X|\mathcal{G}]dP + b\int_G E[Y|\mathcal{G}]dP \\ &= a\int_G XdP + b\int_G YdP,\end{aligned}$$
because of the definition of conditional expectation given $\mathcal{G}$. That concludes the proof.

7. ***Theorem***: If X is $\mathcal{G}$-measurable and Y and XY are integrable, then $E[XY|\mathcal{G}] = XE[Y|\mathcal{G}]$. Why does this make sense? Proof for simple functions $X = \sum_{i=1}^n c_i I_{G_i}$ $(G_i \in \mathcal{G})$: first: for $X = I_G$. To check:
$$E[1_G Y|\mathcal{G}] = 1_G E[Y|\mathcal{G}],$$
or, for every $\tilde{G} \in \mathcal{G}$,
$$\int_{\tilde{G}} 1_G E[Y|\mathcal{G}]dP = \int_{\tilde{G}} 1_G Y dP.$$
But the left hand side equals
$$\begin{aligned}\int_{\tilde{G}} 1_G E[Y|\mathcal{G}]dP &= \int_{\tilde{G}\cap G} E[Y|\mathcal{G}]dP \\ &= \int_{\tilde{G}\cap G} YdP \\ &= \int_{\tilde{G}} 1_G Y dP.\end{aligned}$$
Next, for $X = \sum_{i=1}^n c_i I_{G_i}$:
$$\begin{aligned}E[(\sum_{i=1}^n c_i I_{G_i})Y \mid \mathcal{G}] &= \sum_{i=1}^n c_i E[1_{G_i} Y|\mathcal{G}] \\ &= \sum_{i=1}^n c_i 1_{G_i} E[Y|\mathcal{G}] = XE[Y|\mathcal{G}],\end{aligned}$$

where in the first line we use the previous theorem and in the second line we use the statement for 1_{G_i}. This concludes the proof for simple X.

8. ***Note***: conditioning on a σ-algebra generated by random variables is the same as conditioning on the random variables generating the σ-algebra.

9. ***Note***: $E[Y|\sigma(X)]$ is a function of X.

10. A ***filtration*** is an increasing family of σ-algebra's. The idea is that information, and hence the information contained in the σ-algebra, increases over time.

11. We say that a process X is ***adapted*** to a filtration $\mathcal{F}_t$ if $X(t)$ is $\mathcal{F}_t$-measurable. The idea is that once the information on $\mathcal{F}_t$ is available, the value of $X(t)$ is known. Notice that since $\mathcal{F}_t$ is increasing in t, this also means that the value of $X(s)$ for $s < t$ is known (or, $X(s)$ is $\mathcal{F}_t$-measurable).