Plan Counting, Renewal, and Point Processes

0. Finish FDR Example
1. The Basic Renewal Process
2. The Poisson Process Revisited
3. Variants and Extensions
4. Point Processes

Reading: G&S: 7.1–7.3, 7.10
Next Time: Poisson and Point Processes and Applications

Homework 7 due Thursday

Left-over Question 1. Does $\mathsf{E}|Z| < \infty$ implie that $\mathsf{E}|Z|1\{|Z| > c\} \to 0$ as $c \to \infty$?

Theorem 2. Optional Sampling
Let X be a submartingale with respect to $(\mathcal{F}_n)$.

1. If T is a stopping time for which $\mathsf{P}\{T \le m\} = 1$ for some $m < \infty$, then $\mathsf{E}X_T^+ < \infty$ and $X_0 \le \mathsf{E}(X_T \mid \mathcal{F}_0)$. (Hence, $\mathsf{E}X_0 \le \mathsf{E}X_T$.)
2. If $T_0 \le T_1 \le \cdots$ are stopping times with $\mathsf{P}\{T_i \le m_i\} = 1$ for $m_i < \infty$, then $Y_n = X_{T_n}$ is a submartingale as well.

Corollary 3. Optional Stopping Theorem Let X be a martingale and T be a stopping time such that

1. $\mathsf{P}\{T < \infty\} = 1$.
2. $\mathsf{E}|X_T| < \infty$
3. $\mathsf{E}X_n 1\{T > n\} \to 0$ as $n \to \infty$.

Then, $\mathsf{E}X_T = \mathsf{E}X_0$.

Corollary 4. Suppose that $X = (X_n)_{n\ge0}$ is a uniformly integrable martingale and T is a stopping time with $\mathsf{P}\{T < \infty\} = 1$, then $\mathsf{E}X_T = \mathsf{E}X_0$.

Corollary 5. If X is a martingale and T is a stopping time with $\mathsf{E}T < \infty$ and assume that $X_{T\wedge n}$ is a uniformly integrable sequence. Then $\mathsf{E}X_0 = \mathsf{E}X_T$.

Application 6. False Discovery Rates

Suppose we have m hypothesis tests and corresponding p-values $P_1, \ldots, P_m$. Let $P_{(0)} \equiv 0$ and $P_{(1)} < \cdots P_{(m)}$ are the p-values in increasing order. Then, Benjamini and Hocbherg (1995) used the following threshold to determine which null hypotheses would be rejected.

$$P^* = \max\left\{P_{(i)} \colon P_{(i)} \leq \alpha \frac{i}{m}\right\} \tag{1}$$

They showed that if we reject all null hypothesis for which $P_i \leq P^*$, we will control the *false discovery rate* at level α. That is,

$$\mathsf{E} \frac{\text{falsely rejected nulls}}{\text{rejected nulls}} \leq \alpha. \tag{2}$$

Let's make this more precise.

Let $\mathcal{H}_0$ and $\mathcal{H}_1$ be a partition of $\{1, \ldots, m\}$. $\mathcal{H}_0$ represents the indices of true null hypotheses; $\mathcal{H}_1$ represents the indices of true alternatives (though which alternative is not specified by the index alone). Let $m_0 = \#(\mathcal{H}_0)$.

For any p-value *threshold*, $0 \leq t \leq 1$, define the following:

$$V(t) = \sum_{i \in \mathcal{H}_0} 1\{P_i \leq t\} \tag{3}$$

$$R(t) = \sum_{i \in \mathcal{H}_0 \cup \mathcal{H}_1} 1\{P_i \leq t\}. \tag{4}$$

$V(t)$ is the number of false rejects at threshold t; $R(t)$ is the total number of rejections. Define the False Discovery Proportion at threshold t by

$$\mathrm{FDP}(t) = \frac{V(t)}{\max(R(t), 1)}. \tag{5}$$

Then, for a threshold T, fixed or random, the False Discovery Rate (FDR) is given by

$$\mathrm{FDR} = \mathsf{E}\,\mathrm{FDP}(T). \tag{6}$$

Benjamini and Hochberg (1995) showed that a threshold equivalent to

$$T_\alpha = \sup\left\{t \in [0,1] \colon \frac{mt}{R(t)} \leq \alpha\right\} \tag{7}$$

satisfies $\mathrm{FDR} \leq \alpha$.

Let's give a martingale proof in three steps. But in this case we need to do two things: take the threshold as our time index, which is thus a continuous index, and reverse time. Both are technical variations on the discrete-time martingales we've seen, but the basic ideas are the same.

Assume that the p-values for which the null is true are independent.

Claim 1. The process $t \mapsto V(t)/t$ is a reversed-time martingale with respect to the filtration

$$\mathcal{F}_t = \sigma\left(1\{P_i \leq s\}, s \in [t,1], i \in \{1, \ldots, m\}\right). \tag{8}$$

Claim 2. T_α is a stopping time (with time reversed) with respect to $(\mathcal{F}_t)$.

Claim 3. The optional stopping theorem implies that $\mathsf{E}\,\mathrm{FDP}(T_\alpha) = (m_0/m)\alpha$.

Proof of Claim 1. What do we need to show? What do we know (or need to assume)? What are the first steps?

Proof of Claim 2. What do we need to show? (It's not exactly the same as in the discrete case but what's the natural analogue?) What do we know (or need to assume)? How does $mt/R(t)$ change with t?

Proof of Claim 3. Thinking about the process $mt/R(t)$, we see that

$$R(T_\alpha) = \frac{m}{\alpha} T_\alpha. \tag{9}$$

Hence,

$$\mathrm{FDP}(T_\alpha) = \frac{\alpha}{m} \frac{V(T_\alpha)}{T_\alpha}. \tag{10}$$

Note that the stopped process is bounded by m/α. Hmmmm...

The optional stopping theorem gives

$$\mathsf{E}\frac{V(T_\alpha)}{T_\alpha} = \mathsf{E}\frac{V(1)}{1} = m_0, \tag{11}$$

and thus

$$\mathrm{FDR}_{T_\alpha} = \frac{\alpha}{m}\mathsf{E}\frac{V(T_\alpha)}{T_\alpha} = \frac{m_0}{m}\alpha.$$

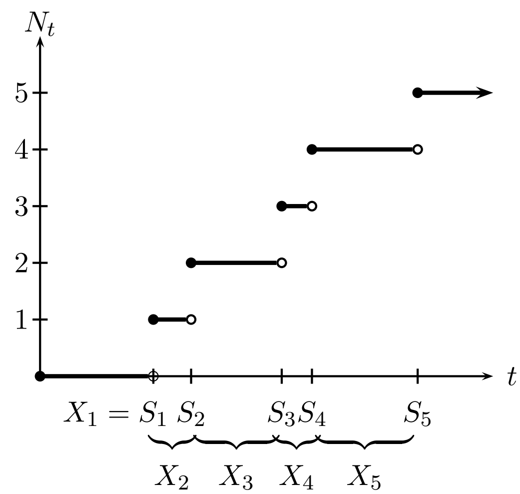
Sweet.

Definition 7. A *counting process* $N = (N_t)_{t \ge 0}$ satisfies

1. N_t is non-negative integer-valued.
2. N_t is non-decreasing.

The *increment* $N_t - N_s$ for $s \le t$ represents the number of "events" that have occurred in the interval $(s, t]$.

Figure 8.



The Basic Renewal Process 9.

Let $X_1, X_2, \ldots$ be a sequence of IID, non-negative random variables with common CDF F. Assume, to avoid odd boundary cases, that $F(0) = 0$. Let $\mu = \mathsf{E}X_1$.

We will interpret X_i as the time between the $i-1$st and ith event or "renewal".

Write $S_0 = 0$ and $S_n = \sum_{i=1}^n X_i$. Then, S_n is the time of the nth event.

Finally, define

$$N_t = \sup\{n \in \mathbb{Z}_\oplus \colon S_n \le t\}. \tag{12}$$

This counting process is called a *renewal process.*

Question 10. Can an infinite number of events occur in a finite time interval? Why or why not?

A Representation 11. Consider the following claim:

$$N_t = \sum_{k=1}^{\infty} 1\{S_k \le t\}\,. \tag{13}$$

Is this true? Why?

Definition 12. Let $m(t) = \mathsf{E}N_t$. This is called the *renewal function*. By the above representation, we have

$$m(t) = \sum_{k=1}^{\infty} \mathsf{P}\{S_k \le t\} = \sum_{k=1}^{\infty} F_k(t), \tag{14}$$

where F_k is the CDF of S_k.

Question 13. What is the distribution of S_k in terms of the distribution of S_{k-1}? (Start with $k = 1$.)

Derivation 14. The Distribution of N_t.

We have the following logical relation:

$$N_t \ge n \iff S_n \le t. \tag{15}$$

From this, we have

$$\begin{aligned}
\mathsf{P}\{N_t = n\} &= \mathsf{P}\{N_t \ge n\} - \mathsf{P}\{N_t \ge n+1\} &\quad (16)\\
&= \mathsf{P}\{S_n \le t\} - \mathsf{P}\{S_{n+1} \le t\} &\quad (17)\\
&= F_n(t) - F_{n+1}(t). &\quad (18)
\end{aligned}$$

Notation 15. For our purposes here, we will use $\star$ for the convolution operator given by

$$(G \star H)(t) = \int_0^t G(t-s)\, dH(s), \tag{19}$$

whenever the integral exists, for real-valued functions G, H on $[0, \infty)$. You can show that this a commutative and associative operator.

Equation 16. A function u is a solution to a *renewal-type* equation if it satisfies

$$u = G + u \star F, \tag{20}$$

for bounded G on $[0, \infty)$.

In particular, the renewal function satisfies

$$m = F + m \star F. \tag{21}$$

There are a couple ways to see this. First, we can plug in $\sum_k F_k$ into the equation. Check. Second, we can condition on the first arrival. The function $\mathsf{E}(N_t \mid X_1 \text{ near } s)$ takes the following form:

$$\mathsf{E}(N_t \mid X_1 \text{ near } s) = \begin{cases} 1 + m(t-s) & \text{if } s \le t \\ 0 & \text{if } s > t. \end{cases} \tag{22}$$

Why?

It follows that

$$\begin{aligned} m(t) &= \mathsf{E} N_t & (23)\\ &= \mathsf{E}\mathsf{E}(N_t \mid X_1) & (24)\\ &= \int_0^\infty \mathsf{E}(N_t \mid X_1 \text{ near } s)\, ds & (25)\\ &= \int_0^t (1 + m(t-s))\, dF(s) & (26)\\ &= F(t) + (m \star F)(t). & (27) \end{aligned}$$

More generally, a solution to the renewal-type equation (20) is given by

$$u = G + G \star m. \tag{28}$$

To see this, convolve both sides of the equation with F:

$$\begin{aligned} u \star F &= G \star F + G \star m \star F & (29)\\ &= G \star F + G \star (m - F) & (30)\\ &= G \star m & (31)\\ &= u - G, & (32) \end{aligned}$$

hence,

$$u = G + u \star F. \tag{33}$$

We can also show that if G is bounded on compact sets, then u is a unique solution and is also bounded on compact sets.

Definition 17. A real-valued random variable is said to be *lattice* if it takes values in a set $\{ka\colon k \in \mathbb{Z}\}$. The *span* (or period) of the lattice is the maximal such a.

Limit Theorems 18. Renewal processes behave asymptotically as you might expect. (ha ha) As $t \to \infty$, the following hold

1. $\dfrac{N_t}{t} \to \dfrac{1}{\mu}$ almost surely.
2. $\dfrac{m(t)}{t} \to \dfrac{1}{\mu}$.
3. $\dfrac{m(t+h) - m(t)}{h} \to \dfrac{1}{\mu}$, for all h if X_1 is not lattice and for all h that are integer multiples of the span if X_1 is lattice.

In addition, N_t/t, properly centered and scaled, is asymptotically Normal. See the book (section 10.2) for proofs.

Process 19. The *Poisson Process* is a renewal process with interarrival distribution F equal to an Exponential$\langle\lambda\rangle$. More precisely, this is called a (homogeneous) *Poisson process* with rate λ.

Recall that from this setup, we showed that the Poisson process has two other important properties.

First, *independent increments*. The random variables $N_t - N_s$ and $N_v - N_u$ are independent whenever $(s, t]$ and $(u, v]$ are disjoint. (The same goes for multiple such intervals.)

Second, *stationary increments*, the distribution of $N_t - N_s$ depends only on $t - s$.

We can thus characterize this process in two other equivalent ways.

Alternate Definition 1 20.
Let $N = (N_t)_{t\geq 0}$ be a process with the following properties:

1. $N_0 = 0$
2. For $0 \leq s < t$, $N_t - N_s$ has a Poisson$\langle\lambda(t - s)\rangle$ distribution.
3. For $0 \leq s_1 < t_1 < s_2 < t_2 < \cdots < s_m < t_m$, the random variables $N_{t_i} - N_{s_i}$ are independent.

Then, N is a Poisson process with rate λ.

Alternate Definition 2 21.
Let $N = (N_t)_{t\geq 0}$ be a process with the following properties:

1. $N_0 = 0$
2. The process has stationary and independent increments.
3. $\mathsf{P}\{N_h = 1\} = \lambda h + o(h)$.
4. $\mathsf{P}\{N_h \geq 2\} = o(h)$.

Then, N is a Poisson process with rate λ.

Theorem 22. The two alternate definitions are equivalent.

Basic idea of the proof: Poisson approximation to the binomial.

Note that from these alternative definitions, we can derive the interarrival distribution by

$$\mathsf{P}\{X_1 > t\} = \mathsf{P}\{N_t = 0\} = e^{-\lambda t}. \tag{34}$$

Useful Fact 23. ... which you derived in homework.

Given $N_t = n$, the set of n arrival times $\{S_1, \dots, S_n\}$ have the same distribution as the order statistics of n independent Uniform$\langle 0, t\rangle$ random variables.

Process 24. By generalizing the alternate definitions, we get a whole new class of interesting processes.

Let $N = (N_t)_{t \geq 0}$ be a process with the following properties:

1. $N_0 = 0$
2. The process has independent increments.
3. $\mathsf{P}\{N_{t+h} - N_h \geq 2\} = o(h)$.
4. $\mathsf{P}\{N_{t+h} - N_h = 1\} = \lambda(t)h + o(h)$.

Then, N is called an inhomogeneous Poisson process with rate function $\lambda(t)$.

We can show that $N_{t+s} - N_t$ has a Poisson distribution with mean $m(t+s) - m(t)$, where $m(t) = \int_0^t \lambda(s)\, ds$.

This leads us to the idea of Point Processes which we'll take up next time.