Lecture 8: Wald's identity, Key renewal theorem

Thursday, September 18, 2014  11:00 AM

Reading: Chapter 3 (Gallager Ch.5)
Homework 2: due today

Outline:

3. The Poisson process: Definition, conditional distribution of the arrival times, non-homogeneous Poisson process, compound Poisson random variables and processes

4. Renewal theory: Limit theorems, Wald's identity, key renewal theorem, branching processes, regenerative processes, stationary point processes

**Renewal Theory**

i.i.d. interarrival times

\[ X_1, X_2, X_3, \ldots \]

\[ S_1 = X_1 \]
\[ S_2 = X_1 + X_2 \]
\[ S_3 = X_1 + X_2 + X_3 \]

arrival times

\[ N(t) = \max_n : S_n \leq t \]

arrival count

\[ = 3 \]

Problem: How is \( N(t) \) distributed?

Last time:

- Strong law for renewal processes:

\[ \lim_{t \to \infty} \frac{N(t)}{t} = \frac{1}{\mu} \]

- Central limit theorem for renewal processes:

\[ \Pr \left[ N(t) \leq \frac{t}{\mu} + k \cdot \sqrt{\frac{\mu}{2t}} \right] \xrightarrow{t \to \infty} \Pr [ Z \leq k ] \]

\[ \sim \text{Normal}(0,1). \]
“$N(t)$ tends to Gaussian with mean $\frac{t}{\mu X}$, std $\sqrt{\frac{t}{\mu X} \cdot \frac{\sigma X}{\mu X}}$.”

Today:

Wald’s identity

Elementary renewal theorem: $E[N(t)] \approx \frac{t}{\mu}$

Blackwell’s renewal theorem: $E[N(t+\delta) - N(t)] \approx \frac{\delta}{\mu}$

**WALD’S IDENTITY**

**Goal:** Understand sums

$$\sum_{n=1}^{N} X_n$$

where these are not independent!

**Stopping times:**

**Idea:** Run a process until something interesting happens.

**Examples:**

- Flip coins until you get heads
  
  $$X_1, X_2, X_3, \ldots \in \{0, 1\}$$
  
  $$N = \min\{j \mid X_j = 1\}$$ is a stopping time

- Flip coins until you get 5 heads in a row, or 6 heads total.
• Start at 0

Repeat: With equal probabilities $\frac{1}{2}$, move left/right
Until you hit 10.

$S_0 = 0, S_1 = \pm 1, S_n = \text{position at time } n$

$N = \min \{ n \mid S_n = 10 \}$ is a stopping time

Or,

$x_1, x_2, \ldots \in \{-1, 1\}$ iid.

$N = \min \{ n : \sum_{i=1}^{n} x_i = 10 \}$

? RandomWalkProcess

RandomWalkProcess[p] represents a random walk on a line with
the probability of a positive unit step $p$ and the probability of a negative unit step $1-p$.

RandomWalkProcess[p, q] represents a random walk with the probability of a positive unit step $p$, the probability of a negative unit step $q$, and the probability of a zero step $1-p-q$.

ListLinePlot[
  Append[
    Table[RandomFunction[RandomWalkProcess[1/2], {0, 100}], {3}],
    {{0, 10}, {100, 10}}
  ]
]

20

10

-10

N

N
• Start at vertex u in a graph. Repeat: move to a uniformly random neighboring vertex. Until you get to v.

\[ N = \min \{ t \mid \text{vertex } V_t = v \} \]

is a stopping time (also called a "hitting time")

• For a general renewal process, \( N(t) + 1 \) is a hitting time.

It corresponds to the rule: sample \( X_1, X_2, \ldots \)

until \( S_n = \sum_{i=1}^{n} X_i > t \).

**RenewalProcess**

`RenewalProcess[dist]` represents a renewal process with interarrival times distributed according to `dist`.

```math
ListLinePlot[
  Table[RandomFunction[RenewalProcess[ExponentialDistribution[1]], {10}, {3}],
  InterpolationOrder -> 0
]
```

Note: \( N(t) \) is not a stopping time!

Why?

Stopping at \( N(t) \) would mean that you have to stop
at the last renewal event to occur before time $t$. But you can't know that you've gotten to the last event before $t$ unless you know the future; that the next event will be after $t$.

**Definition:** $N \in \{0, 1, 2, 3, \ldots\}$ is a **stopping time** for the sequence $X_1, X_2, X_3, \ldots$ if the event $\xi N = n$ only depends on $X_1, X_2, \ldots, X_n$.

(That is, if the indicator random variable $\mathbb{1}_{\xi N = n}$ is a function of $X_1, X_2, \ldots, X_n$.

**Remark:** Other definitions are also used, e.g.,

$N$ is a stopping time for the sequence $X_1, X_2, \ldots$ of i.i.d. random variables if the event $\xi N = n$ is independent of $\xi X_m$ for $m > n$.

(Since you can’t know the future when you decide to stop, the stopping time is independent of the future $X_{n+1}, X_{n+2}, \ldots$)

There are also continuous-time versions of the definition, e.g.,

- **Run Brownian motion until you hit 2**

(Brownian motion is a continuous-time and space version of a random walk, $X_{t+\Delta t} = X_t + \mathcal{N}(0, \sqrt{\Delta t})$ or $X_t \pm \sqrt{\Delta t}$)
WienerProcess \( \mu, \sigma \) represents a Wiener process with a drift \( \mu \) and volatility \( \sigma \).

\( \text{WienerProcess[]} \) represents a standard Wiener process with drift 0 and volatility 1.

```math
ListLinePlot[
  Append[
    Table[RandomFunction[WienerProcess[], {0, 10, .01}], {3}],
    {{0, 2}, {10, 2}}
  ]
]
```

**General example:**

Sample \( X_1 \), then \( X_2 \), then \( X_3, \ldots \)

until \( X_1, X_2, \ldots, X_n \) satisfy some condition.

Then stop.

\( N = n \) is a stopping time because it is determined by \( X_1, \ldots, X_n \) (the past), and not by the future \( X_{n+1}, X_{n+2}, \ldots \).

**Wald's identity:**

**Theorem:**

Let \( X_1, X_2, \ldots \) be i.i.d. with \( \mathbb{E}[X_i] < \infty \),

and let \( N \) be a stopping time for the sequence with \( \mathbb{E}[N] < \infty \).

Then...
and let \( N \) be a stopping time for the sequence with \( \mathbb{E}N < \infty \). Then,

\[
\mathbb{E} \left[ \sum_{n=1}^{N} X_n \right] = \mathbb{E}N \mathbb{E} X_1.
\]

**Remark:** This strengthens two identities we have seen before:

- \( \mathbb{E} \left[ \sum_{n=1}^{N} X_n \right] = \sum_{n=1}^{N} \mathbb{E} X_n \) (linearity of expectation)
- \( \mathbb{E} \left[ \sum_{n=1}^{N} X_n \right] = \mathbb{E}[N] \mathbb{E} X_1 \) if the \( X_j \) are i.i.d., \( \forall j \) of \( N \).

We will later show the "martingale stopping theorem," which implies the same conclusion provided \( \mathbb{E}[X_n | X_1, \ldots, X_{n-1}] = m \), without requiring the \( X_n \) be independent or identically distributed.

**Proof:**

The basic problem is that the random variable \( N \) is a limit of the summation. We need to move it into the sum so we can apply identities like \( \mathbb{E}[A B] = \mathbb{E}[A] \mathbb{E}[B] \) for all \( A, B \). We can do that like so:

\[
\sum_{n=1}^{N} X_n = \sum_{n=1}^{\infty} X_n \cdot \mathbb{1}_{\{N \geq n\}}
\]

(The indicator r.v. for \( \{N \geq n\} \)

\( = 1 \) if \( N \geq n \), \( = 0 \) if \( N < n \)

Now \( \mathbb{1}_{\{N \geq n\}} = 1 - \mathbb{1}_{\{N < n\}} \)

\( = 1 - \mathbb{1}_{\{N \leq n-1\}} \).

Therefore this event is completely determined by \( X_1, \ldots, X_{n-1} \).

It is independent of \( X_n \).

Thus,

\[
\mathbb{E} \left[ \sum_{n=1}^{N} X_n \right] = \sum_{n=1}^{\infty} \mathbb{E}[X_n \cdot \mathbb{1}_{\{N \geq n\}}]
\]

\[= \sum_{n=1}^{\infty} \mathbb{E}[X_n] \cdot \mathbb{E}[\mathbb{1}_{\{N \geq n\}}]
\]

\[= \sum_{n=1}^{\infty} \mathbb{E}[X_n] \cdot \mathbb{P}[N \geq n] \]
\[
\begin{align*}
E[X] &= \sum_{n=1}^{\infty} P[N = n] \\
&= E[X] \cdot E[N]. \\
&= E[X] \cdot E[N].
\end{align*}
\]

(by the "integrating the tail" method)

Remark: I cheated in the above proof when I wrote
\[
E \left[ \sum_{n=1}^{\infty} X_n \cdot 1_{N = n} \right] = \sum_{n=1}^{\infty} E[X_n \cdot 1_{N = n}].
\]

Why is this allowed?

\textbf{Theorem: (Fubini's Theorem)}

If \( Y_1, Y_2, \ldots \) are random variables with \( \sum_{n=1}^{\infty} E[|Y_n|] < \infty \)
then
\[
E \left[ \sum_{n=1}^{\infty} Y_n \right] = \sum_{n=1}^{\infty} E[Y_n].
\]

This theorem applies, with \( Y_n = X_n 1_{N = n} \), because
\[
\sum_{n=1}^{\infty} E[|X_n| 1_{N = n}] = E[|X|] \sum_{n=1}^{\infty} P[N = n] = E[|X|] \cdot E[N] < \infty.
\]

\textbf{Proof of Fubini's Theorem:}

Since \( \sum_{n=1}^{\infty} E[|Y_n|] < \infty \), for any \( \varepsilon > 0 \) we can find a \( k \) st.
\[
\sum_{n>k} E[|Y_n|] < \varepsilon.
\]

Hence \( E[\sum_{n=1}^{\infty} Y_n] = E[\sum_{n=k}^{\infty} Y_n + \sum_{n=k}^{\infty} Y_n] \)
\[
= \sum_{n=k}^{\infty} E[Y_n] + E[\sum_{n=k}^{\infty} Y_n] \]
\[
\Rightarrow |E[\sum_{n=1}^{\infty} Y_n] - \sum_{n=1}^{k} E[Y_n]| = \left| E[\sum_{n=k}^{\infty} Y_n] \right| \leq E[\sum_{n=k}^{\infty} |Y_n|] \leq \sum_{n=k}^{\infty} E[|Y_n|] < \varepsilon.
\]

Thus \( \lim_{k \to \infty} \sum_{n=1}^{k} E[Y_n] = E[\sum_{n=1}^{\infty} Y_n]. \)

\textbf{Example 1:} \( X_n = \{ 0 \} \) w/ prob. 1/2.
\[ N = \min\{n \mid X_1 + \cdots + X_n = 10\} \]

What is \( \mathbb{E}[N] \)?

Answer:
\[ \mathbb{E}\left[ \sum_{n=1}^{\infty} X_n \right] = \mathbb{E}[X_1] \cdot \mathbb{E}[N] \]
\[ 10 \cdot \frac{1}{2} = \mathbb{E}[N] = 20. \checkmark \]

**Example 2:** Random walk on the line

\[ \cdots -2 -1 0 1 2 \cdots \]

Starting from 0, what is \( \mathbb{E}[\text{# of steps to reach } 1] \)?

- \( X_n \in \{\pm 1, 13\} \) uniform, i.i.d.
- \( N = \min\{n \mid \sum_{i=1}^{n} X_i = 13\} \)

\[ \mathbb{E}[N] \cdot \mathbb{E}[X_n] = \mathbb{E}\left[ \sum_{n=1}^{N} X_n \right] \]
\[ 0 \cdot 1 = \mathbb{E}[N] \]

This is impossible unless \( \mathbb{E}[N] = \infty. \checkmark \)

**Example 3:** \( X_n \in \{1, 2, 3, \ldots, 10\} \) uniform, i.i.d. \( (\mathbb{E}[X] = 5.5) \)

\[ N = \min\{n \mid X_n = 5.5\} \]

Q: What is \( \mathbb{E}\left[ \sum_{n=1}^{N} X_n \right] \)?

Answer:
\[ \mathbb{P}[N = n] = \mathbb{P}[X_1 \neq 5.5, X_2 \neq 5.5, \ldots, X_{n-1} \neq 5.5, X_n = 5.5] \]
\[ = \left(\frac{9}{10}\right)^{n-1} \cdot \frac{1}{10} \quad \text{for } n \geq 1 \]

\[ \Rightarrow \mathbb{E}[N] = 10 \]

Wald's identity \( \Rightarrow \mathbb{E}\left[ \sum_{n=1}^{N} X_n \right] = \mathbb{E}[N] \cdot \mathbb{E}[X] = 55. \)

Remark: Note that \( X_N = 5.5 \) and \( X_n \) conditioned on \( n < N \) is biased, \( \mathbb{E}[X_n] \mid n < N \neq 5.5 \), since \( X_n \neq 5.5 \).

Nonetheless, Wald's identity lets us treat the sum.
as if the \( X_n \) were independent of \( N \), and not biased!

Also,

\[
E \left[ \sum_{n=1}^{\min\{m: X_m = \delta_3\}} X_n \right] = SS, \text{ still! Nothing special about } \sigma.
\]

**Elementary renewal theorem**

**Corollary:** In a renewal process, since \( N(t)+1 \) is a stopping time (corresponding to the rule: stop once \( S_n > t \)),

\[
E[S_{N(t)+1}] = E[X] \cdot E[N(t)+1]
\]

\[
\Rightarrow E[N(t)] = \frac{E[S_{N(t)+1}]}{\mu_X} - 1.
\]

**Theorem:** *(Elementary renewal theorem)*

For a renewal process, with \( \mu = E[X_i] \),

\[
\lim_{t \to \infty} \frac{E[N(t)]}{t} = \frac{1}{\mu}.
\]

**Remark:** We have seen already that, w\%/ prob. 1,

\[
\lim_{t \to \infty} \frac{N(t)}{t} = \frac{1}{\mu}.
\]

So this is not surprising. But in general, (*) does not imply (*). For example, if

\( U \sim \text{Uniform}(0, 1) \)

and \( Y_t = t \times [\frac{1}{2} \leq u < \frac{1}{t}] = \begin{cases} t, & \text{if } U \leq \frac{1}{t} \\ 0, & \text{if } U > \frac{1}{t} \end{cases} \)

then \( \lim_{t \to \infty} Y_t = 0 \) even though \( E[Y_t] = 1 \) for all \( t > 0 \).

**Proof of the elementary renewal theorem:**

By our corollary to Wald's identity,

\[
\frac{E[N(t)]}{t} = \frac{E[S_{N(t)+1}]}{t \cdot \mu} - \frac{1}{t}
\]

Since \( S_{N(t)+1} > t \), this is

\[
> \frac{1}{\mu} - \frac{1}{t} \longrightarrow \frac{1}{\mu} \text{ as } t \to \infty.
\]
\[
\frac{1}{\mu} - t \xrightarrow{t \to \infty} \frac{1}{\mu}
\]

Showing that \( \frac{1}{\mu} \) is also an upper bound is slightly harder. The problem is that we don't have any upper bound on \( S_{N(t)+1} \). To get an upper bound, let us define the \"truncated\" renewal process, with interarrival times
\[
X_n = \min \{X_n, M\},
\]
for some constant \( M \). Then
\[
N(t) \leq N(t) \quad \text{and} \quad S_{N(t)+1} \leq t + X_{N(t)+1} \leq t + M,
\]
so
\[
\frac{1}{t} E[N(t)] \leq \frac{1}{t} E[N(t)]
\]
\[
= \frac{1}{t} \cdot E[S_{N(t)+1}] - \frac{1}{t}
\]
\[
\leq \frac{t + M}{t} \cdot \frac{1}{\mu} - \frac{1}{t} \quad \text{since} \quad S_{N(t)+1} \leq t + M
\]

Letting \( t \to \infty \), we get
\[
\limsup_{t \to \infty} \frac{1}{t} E[N(t)] \leq \frac{1}{\mu} = \frac{1}{E[\min \{X_1, M\}]
\]
This holds for every \( M \). Letting \( M \to \infty \),
\[
E[\min \{X, M\}] = \int_0^\infty xf(x)dx + \int_M^\infty f(x)dx
\]
\[
\Rightarrow E[\min \{X, M\}] \to \infty \quad \text{since} \quad E[X] < \infty
\]
\[
\Rightarrow E[\min \{X, M\}] \xrightarrow{M \to \infty} E[X], \quad \text{proving the theorem.} \quad \Box
\]

**Blackwell's Theorem**

"The expected renewal rate approaches steady state as \( t \to \infty \)."

\[
\lim_{t \to \infty} E[N(t+t) - N(t)] = \frac{E}{\mu X}.
\]

More precisely, call the distribution \( F_X \) of interarrival times,
“lattice with period d” if \( X \) can only take on values that are multiples of \( d \) (and there is no larger \( d' \) that also works).

**Theorem:** *(Blackwell's Theorem)*

If the distribution of interarrival times is not lattice, then for any \( \delta > 0 \),
\[
\lim_{t \to \infty} \frac{\mathbb{E}[N(t+t) - N(t)]}{\mathbb{E}[X]} = \delta.
\]

(There is another statement for lattice interarrival times; see the text.)

**Proof idea:**
\[
\mathbb{E}[N(t)] = \sum_{n=1}^{\infty} \mathbb{P}[N(t) = n]
\]
\[
= \sum_{n=1}^{\infty} \mathbb{P}[S_n \leq t]
\]
\[
= \mathbb{P}[X_1 \leq t] + \sum_{n=2}^{\infty} \int_{0}^{t} \text{d}x \cdot f_X(x) \cdot \sum_{n=1}^{\infty} \mathbb{P}[S_{n-1} \leq t-x]
\]
\[
= \mathbb{P}[X < t] + \int_{0}^{t} \text{d}x \cdot \mathbb{E}[N(t-x)] \cdot f_X(x).
\]

Now, assume that \( X \) is a convex combination of exponentially distributed random variables. (This is not general!)

For example, say that
\[
X \sim \begin{cases} 
\text{Exp}(\lambda = 1) & \text{w/prob.} \frac{1}{2} \\
\text{Exp}(\lambda = 3) & \text{w/prob.} \frac{1}{2}
\end{cases}
\]
\[
\Rightarrow f_X(x) = \frac{1}{2} e^{-x} + \frac{1}{2} \cdot 3 e^{-3x}.
\]

Applying the Laplace transform
\[
L_X(s) = \int_{0}^{\infty} e^{-sx} f(x) 
\]
\[
= \frac{1}{2} \cdot \frac{1}{s+1} + \frac{3}{2} \cdot \frac{1}{s+3}
\]

Here we have used the first of the following two integral calculations:
Facts: \( L(e^{-ct}) = \frac{1}{s+c} \)

\( L(x^p(s)) = \frac{1}{sy^{p+1}}, \quad p = 0, 1, 2, \ldots \)

Let \( m(t) = \mathbb{E}[N(t)] \), satisfying

\[
m(t) = \mathbb{P}[X \leq t] + \int_0^t m(t-x) f_X(x) \, dx.
\]

Taking the Laplace transform of both sides,

\[
L_m(s) = \int_0^\infty e^{-st} F_X(t) \, dt + \int_0^\infty e^{-st} \int_0^t m(t-x) f_X(x) \, dx = \frac{L_f(s)}{s} \cdot \frac{L_m(s)}{s} \cdot \frac{1}{1-L_f(s)}
\]

\[
\Rightarrow L_m(s) = \frac{L_f(s)}{s(1-L_f(s))}
\]

Substitute in our result \( L_f(s) = \frac{1}{2s+1} + \frac{3}{2} \cdot \frac{1}{s+3} \)

\[
L_f = \frac{1}{2} \cdot \frac{1}{s+1} + \frac{3}{2} \cdot \frac{1}{s+3} \;
\]

\[
\frac{L_f}{s(1-L_f)} \quad // \text{Simplify}
\]

\[
\% - \left( \frac{3}{2s^2} + \frac{1}{4s} - \frac{1/4}{s+2} \right) \quad // \text{Simplify}
\]

\[
\frac{3+2s}{s^2(2+s)}
\]

\[
0
\]

Hence \( L_m(s) = \frac{3/2}{s^2} + \frac{1/4}{s} - \frac{1/4}{s+2} \),

implying (by the above facts) that

\[
m(t) = \mathbb{E}[N(t)] = \frac{3}{2}t + \frac{1}{4} - \frac{1}{4} e^{-2t}. \]

Note that \( \mathbb{E}[X] = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{3} = \frac{2}{3} \), so

\[
\mathbb{E}[N(t)] = \frac{5}{6}t + \frac{1}{6} - \frac{1}{6} e^{-2t}
\]
The elementary renewal theorem $E[N(t)]\downarrow \nu_\infty$, and Blackwell's theorem, $E[N(t+\delta)-N(t)] \to \delta/\nu_\infty$, both follow (for this special case). 

**Example:** $X = \left\{ \begin{array}{ll} 1 & \text{w/ prob. } \frac{1}{2} \\ \pi & \text{w/ prob. } \frac{1}{2} \end{array} \right.$ a non-lattice distr

```
dist = EmpiricalDistribution[{{1/2, 1/2} -> {1, \pi}});

ListLinePlot[
  RandomFunction[RenewalProcess[dist], {100}],
  InterpolationOrder -> 0]
```

**A sample process**
Now consider $E[N(t)]$. We can estimate $E[N(t)]$ by taking the average of 100 samples of $N(t)$. 

50 samples of $N(t)$ vs. $t$: 

```math
samples = Table[
    RandomFunction[RenewalProcess[dist], {100}],
    {50}];
ListLinePlot[samples, InterpolationOrder -> 0]
```
data = RandomFunction[RenewalProcess[dist], {100}, 100];
Plot[Mean[data[t]], {t, 0, 10}]
Plot[Mean[data[t]], {t, 90, 100}]

• for t small,
  \[ \mathbb{E}[N(t)] \text{ is very jagged} \]

• But for large t,
  \[ \mathbb{E}[N(t)] \text{ smooths out, as Blackwell's theorem says} \]

\[
\text{``slope'' } \frac{\mathbb{E}[N(t+\delta) - N(t)]}{\delta} \rightarrow \frac{1}{\lambda} \text{ as } t \rightarrow \infty
\]

For a lattice distribution of interarrival times, \( \mathbb{E}[N(t)] \)
never smooths out. Here, for example, \( X_\delta = \begin{cases} 1 \text{ with prob } \frac{1}{2} \\ 2 \text{ with prob } \frac{1}{2} \end{cases} \)
\textit{dist} = \texttt{EmpiricalDistribution} \left[ \left\{ \frac{1}{2}, \frac{1}{2} \right\} \rightarrow \{1, 2\} \right];

data = \texttt{RandomFunction[RenewalProcess[dist], \{100\}, 100]};

\texttt{Plot[Mean[data[t]], \{t, 0, 10\}]}
\texttt{Plot[Mean[data[t]], \{t, 90, 100\}]}

\[ \mathbb{E}[N(t)] \]

\( t \rightarrow \)

\[ \mathbb{E}[N(t)] \]

\( t \rightarrow \)

\(* \text{ even for large } t, \)

\( \mathbb{E}[N(t)] \text{ stays jagged with the same period, } t \)