DISCRETE STOCHASTIC PROCESSES

Lecture 20

Chapter 7: Random Walks and Martingales

Random Walks

Three Examples: Simple Random Walk
Hypothesis Testing
G/G/1 Queue

Threshold Crossing: Large Deviation Techniques

Optimization of the Chernov bound

RANDOM WALKS

Definition: The process $\{S_n;\, n \geq 1\}$ is called a **random walk** if $X_1, X_2, \ldots$ are IID random variables and

$$S_n = X_1 + X_2 + \cdots + X_n$$

Each rv S_n is simply a sum of n IID rv's; thus the study of S_n for given n is an old issue. Our interest here is **not** the behavior of S_n for large n, but the process and how it evolves. These differ from renewal processes in that X_k can be negative.

Typical questions

What is the probability that any term in $S_1, S_2, \ldots$ exceeds some fixed number a? (probability an upper boundary is ever crossed)

If S_n exceeds a for at least one n with probability 1, what is the distribution of the first such n? (distribution of first crossing times)

What is the distribution of the overshoot, $S_n - a$, for that n?

Why do we care about such questions?

1) Renewal processes are a special case of random walks (provided we consider S_n as a function of *n,* not *N(t)* as a fctn. of *t* or (arithmetic case) *N(n)* as a function of *n*.)

2) Random walks provide the key to hypothesis testing with a stopping rule.

3) Random walks are the key to G/G/1 queues.

4) Random walks are useful in studying many processes like the stock market.

Example 1 – Simple Random Walk

$P(X_i = 1) = p;\ P(X_i = -1) = 1 - p$

(This will be useful simple example because, for integer boundaries, there is no overshoot.)
Then S_n is number of successes minus number of failures, which is twice number of successes minus n.

$$P(S_n = 2j - n) = \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j}$$

If. $p = 1/2,\ E[S_n] = 0,\ \mathrm{VAR}(S_n) = n$

If $p > 1/\ 2, S_n$ gets big (WP 1) as n gets big.

$\{S_n;\ n \geq 1\}$ can also be modeled as a Markov chain (using all integers as states).
Stopping rule can be used to determine expected first passage times.

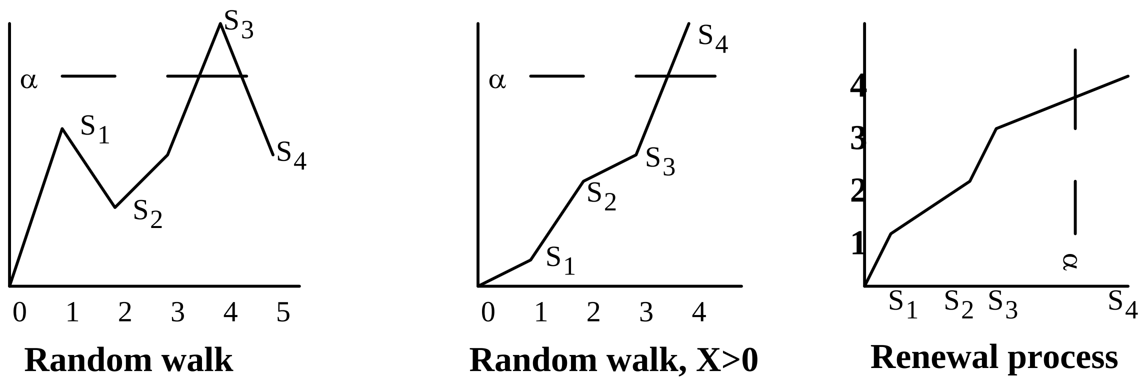
Case II: X_i is an integer valued rv

This can also be modeled as a Markov chain, but the special structure of random walks simplifies the analysis.

Case III: X_i is positive rv

Then S_n is the n^{th} renewal epoch in a renewal process.

When we sketch a sample path of a random walk, n is usually on the horizontal axis, and S_n on vertical; for renewal processes, the axes are conventionally reversed.



Note that the overshoot in a random walk corresponds to the residual life in a renewal process. Analytically, the overshoot for an arbitrary rv is much more complicated to find than the residual life.

Application to Hypothesis Testing

Assume we are trying to distinguish between 2 possible hypotheses, H_0 and H_1;
$P(H_j) = p_j$; $p_0 + p_1 = 1$.

We make n measurements $Y_1, Y_2, \ldots, Y_n$. Conditional on H_0; $Y_1, \ldots, Y_n$ are IID with density $f(y|H_0)$, and conditional on H_1; $Y_1, \ldots, Y_n$ are IID with density $f(y|H_1)$.

$$f_{\vec{Y}|H}\left(\vec{y}\middle|H_j\right) = \prod_{i=1}^{n} f\left(y_i\middle|H_j\right); \quad P\left(H_j\middle|\vec{Y} = \vec{y}\right) = \frac{p_j f_{\vec{Y}|H}\left(\vec{y}\middle|H_j\right)}{f_{\vec{Y}}(\vec{y})}$$

$$\frac{P\left(H_1\middle|\vec{Y} = \vec{y}\right)}{P\left(H_0\middle|\vec{Y} = \vec{y}\right)} = \frac{p_1 \prod_{i=1}^{n} f\left(y_i\middle|H_1\right)}{p_0 \prod_{i=1}^{n} f\left(y_i\middle|H_0\right)}$$

By taking the ratio, we avoid dealing with the unconditional density of the observation $Y_1, \ldots, Y_n$.

If we take the logarithm of both sides, we get

$$\ln\left[\frac{P\left(H_1\middle|\vec{Y} = \vec{y}\right)}{P\left(H_0\middle|\vec{Y} = \vec{y}\right)}\right] = \ln\left[\frac{p_1}{p_0}\right] + \sum_{i=1}^{n} \ln\left[\frac{f\left(y_i\middle|H_1\right)}{f\left(y_i\middle|H_0\right)}\right]$$

The terms in the sum on the right side are called **log likelihood ratios**,

$$Z_i = \ln\left[\frac{f\left(Y_i\middle|H_1\right)}{f\left(Y_i\middle|H_0\right)}\right]; \quad z_i = \ln\left[\frac{f\left(y_i\middle|H_1\right)}{f\left(y_i\middle|H_0\right)}\right]$$

For a given hypothesis, $\{Z_i \;\; i \geq 1\}$ are IID rv's.

$P\left(H_0 \middle| \vec{Y} = \vec{y}\right)$ is the *a posteriori* probability that H_0 is correct.

The MAP (maximum *a posteriori* probability) rule:

choose the hypothesis that maximizes the probability of being correct.

$$\ln\left[p_1 / \; p_0\right] + \sum_{i=1}^{n} Z_i \begin{cases} > 0 \;; & \text{choose } \; H_1 \\ \leq 0 \;; & \text{choose } \; H_0 \end{cases}$$

This says that the classical MAP rule adds the sample values of *n* IID rv (IID under either hypothesis) and chooses accordingly. The *a priori* probabilities, p_0 and p_1 are factored in.

Wald said "This is crazy; you know the probability of error when you make this decision; if it is too large, make more observations."

This means we want to look at the RW $\{S_n; \, n \geq 1\}$ where $S_n = Z_1 + \ldots + Z_n$, and use a stopping rule in our decision.

More concretely, let LR be the log likelihood ratio:

$$LR = \ln\left[\frac{P\left(H_1 \middle| \vec{Y} = \vec{y}\right)}{P\left(H_0 \middle| \vec{Y} = \vec{y}\right)}\right] = \ln\left[\frac{p_1}{p_0}\right] + \sum_{i=1}^{n} \ln\left[\frac{f\left(y_i \middle| H_1\right)}{f\left(y_i \middle| H_0\right)}\right] = \ln\left[\frac{p_1}{p_0}\right] + \sum_{i=1}^{n} Z_i$$

Then (after some minor algebra)

$$P\left(H_1 \middle| \vec{Y} = \vec{y}\right) = \frac{e^{LR}}{e^{LR}+1},\ P\left(H_0 \middle| \vec{Y} = \vec{y}\right) = \frac{1}{e^{LR}+1}$$

Note that these are the *a posteriori probabilities of error*. For example, if you want to the probability of error to be $\leq$ 0.1, you can decide H_1 is true if $e^{LR} \geq 9$ and that H_0 is true if $e^{LR} \leq 1/9$ (i.e.,$LR \geq 2.2$ or $LR \leq -2.2$.) This gives the following random walk formulation of the decision problem: Conduct the set of trials until either

$$\sum_{i=1}^{N} Z_i \geq 2.2 - \ln[\frac{p_1}{p_0}] \text{ or } \sum_{i=1}^{N} Z_i \leq \text{ -}2.2 - \ln[\frac{p_1}{p_0}]$$

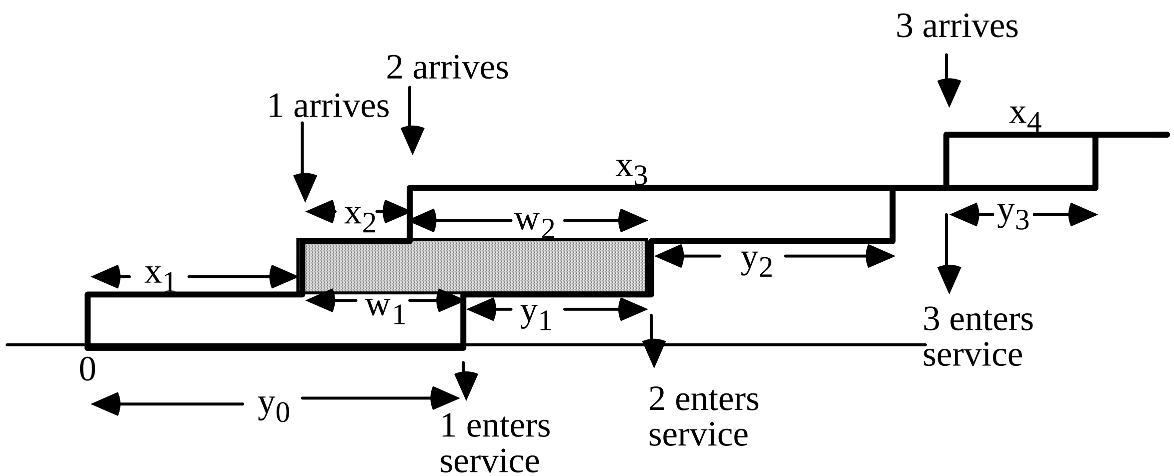
In the former case, conclude H_1 is true. In the latter case, conclude H_0 is true. Note that N is now random, and the number of trials cannot be specified in advance.

This is a Great Idea!

This is vastly superior in many cases to choosing a fixed number of trials, n, in advance. If the evidence is compelling, the threshold is reached quickly on average, and you can terminate the trials early. If it is less conclusive, you continue until you are sufficiently sure you are right, i.e., until you meet your probability of error specification. The smaller the probability of error you allow, the further apart the thresholds become for the random walk, and the longer it will take on average to cross one of them.

Application to G/G/1 queue

Start analysis with beginning of busy period (arrival of customer 0).

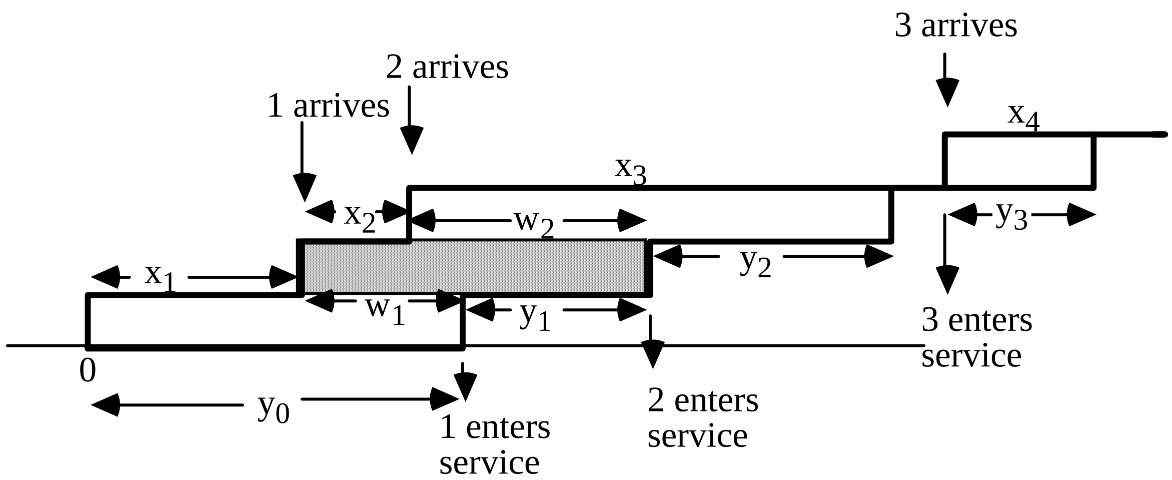


$\{X_i;\, i \ge 1\}$ are interarrival intervals (not the increments of a random walk here).

$\{Y_i;\, i \ge 0\}$ are service periods. $\{W_i,\ i \ge 1\}$ are waiting times in queue, prior to service.

If customer 0 is still in service when 1 arrives $(Y_0 > X_1)$ then 1 waits from X_1 until Y_0. Otherwise, 1 does not wait.

$$W_1 = \max\left[Y_0 - X_1, 0\right]$$

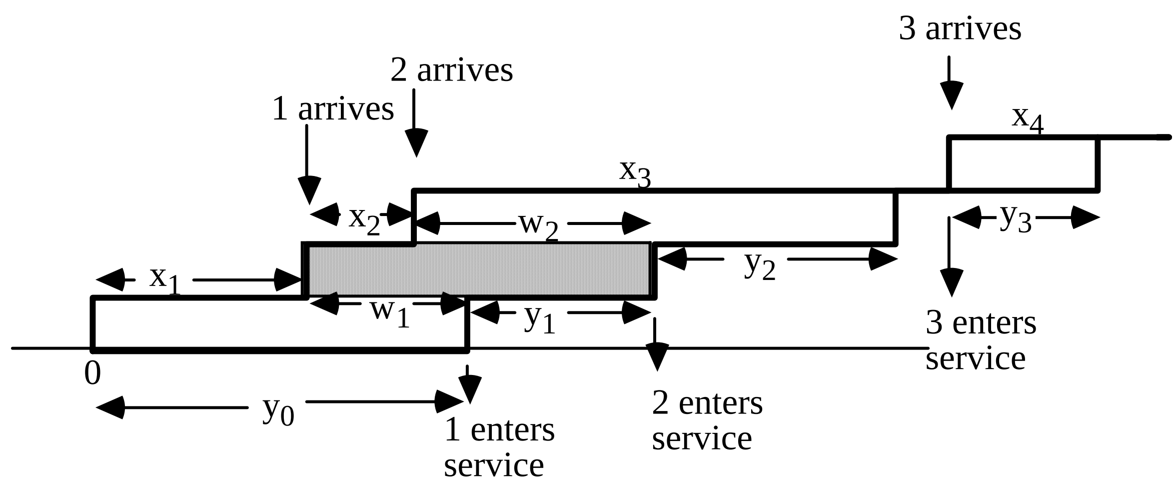


$$W_1 = \max\left[Y_0 - X_1, 0\right]$$

Customer 2 arrives at $X_1 + X_2$, and 1 departs at $Y_0 + Y_1$ or at $X_1 + Y_1$, whichever is larger.
Simpler expression: customer 1 departs at $X_1 + W_1 + Y_1$.

If $X_1 + W_1 + Y_1 > X_1 + X_2$, then customer 2 waits, and otherwise doesn't wait.

$$W_2 = \max\left[W_1 + Y_1 - X_2, 0\right] = \max[\underbrace{(Y_1 - X_2) + (Y_0 - X_1)}_{\text{neither cust. 1 nor cust. 2 entered empty system}}, \underbrace{(Y_1 - X_2)}_{\text{cust. 1 but not cust. 2 entered empty system}}, \underbrace{(\ 0\ )}_{\text{cust. 2 enters empty system}}]$$



$$W_2 = \max\left[W_1 + Y_1 - X_2, 0\right] = \max[\ \underbrace{(Y_1 - X_2) + (Y_0 - X_1)}_{\text{neither cust. 1 nor cust. 2 entered empty system}},\quad \underbrace{(Y_1 - X_2)}_{\text{cust. 1 but not cust. 2 entered empty system}},\quad \underbrace{(\ 0\ )}_{\text{cust. 2 enters empty system}}\ ]$$

$U_n = Y_{n-1} - X_n > 0$ is the wait customer #n would experience due to # (n-1)'s service time alone.

Wn = the wait of customer #n is at least zero (and is zero iff cust. #n starts a busy period.)

It is at least $U_n = Y_{n-1} - X_n$ if $U_n > 0$ (and if $U_n > 0$, Wn = Un > 0 iff cust.# (n-1) starts a busy period.)

It is at least $U_n + U_{n-1}$ if $U_n > 0$ and $U_{n-1} > 0$ (and if $U_n > 0$,, $U_{n-1} > 0$, Wn = Un + Un-1 iff cust # (n-2) starts a busy period.)

In general, customer $n+1$ arrives at $X_1+\ldots+X_{n+1}$ and customer n departs at $X_1+\ldots+X_n+W_n+Y_n$.

$$W_{n+1} = \max[W_n + Y_n - X_{n+1}, 0]$$

Define $U_{n+1} = Y_n - X_{n+1}$, so $\quad W_{n+1} = \max\left[W_n + U_{n+1}, 0\right].$

This is like a random gambling walk with a bankruptcy law.

$$\begin{aligned} W_n &= \max\left[W_{n-1}+U_n, 0\right] = \max\left[W_{n-2}+U_{n-1}+U_n, U_n, 0\right] \\ &= \max\left[W_{n-3}+U_{n-2}+U_{n-1}+U_n, U_{n-1}+U_n, U_n, 0\right] \\ &= \max\left[U_1+\ldots+U_n, U_2+\ldots+U_n, \ldots, U_{n-1}+U_n, U_n, 0\right] \end{aligned}$$

Define: $\quad Z_1^n = U_n, Z_2^n = U_n + U_{n-1}, \ldots, Z_i^n = U_n + \ldots + U_{n-i+1}$

Then $Z_1^n, Z_2^n, \ldots, Z_n^n$ are n terms in a random walk, and

$$W_n = \max\left[0, Z_1^n, Z_2^n, \ldots, Z_n^n\right]$$

$P(W_n \ge a) = P$ (walk exceeds $a > 0$ in sometime in the first n steps).
As $n \to \infty$, $P(W_\infty \ge a) = P$(walk ever exceeds $a > 0$); the distribution of steady state wait in the queue

RANDOM WALKS & THRESHOLD CROSSING
LARGE DEVIATION TECHNIQUES

The usual stopping rule is stop when the RW crosses either a large positive threshold a or a large negative threshold b.

For the examples, we were interested in low probability events (waiting too long in a queue, making errors, etc.).

We focus on **large deviation techniques** – approximating or bounding very small probabilities. These are based on **moment generating functions**.

$$g(r) = E\left[\exp(rX)\right] = \int \exp(rx)\, dF_X(x)$$

The "region of convergence" of $g(r)$ is the set of r for which this integral converges. Assume for now that the region of convergence is large enough for everything we want to do.

$$g_{S_n}(r) = E\left[\exp\left(rS_n\right)\right] = E\left[\exp\left(rX_1 + \ldots + rX_n\right)\right]$$
$$= E\left[\exp\left(rX_1\right)\right]E\left[\exp\left(rX_2\right)\right]\ldots E\left[\exp\left(rX_n\right)\right] = g(r)^n$$

Markov bound for nonnegative random variable X: $P(X \geq a) \leq E(X)/a$.
Recall **Chernov bound** (Markov inequality applied to $\exp\left(rS_n\right)$).

$$P(S_n \geq \alpha) = P\left\{\exp(rS_n) \geq \exp(r\alpha)\right\} \leq g(r)^n \exp(-r\alpha); \ r \geq 0$$

Since Chernov bound is valid for any $r \geq 0$, we want to choose the r that minimizes $g(r)^n \exp(-r\alpha)$, but we postpone that to the next page.

The Chernov bound is easier to work with if we define the **semi-invariant MGF**,

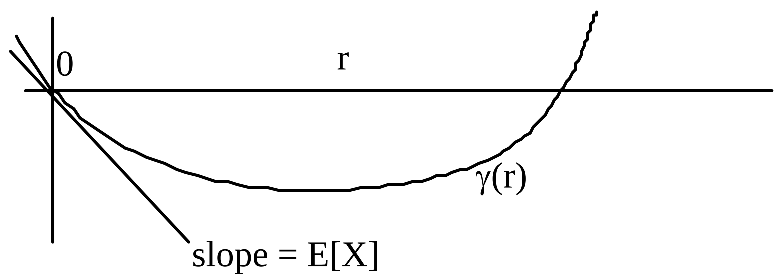
$$\boxed{\gamma(r) = \ln\left(g(r)\right)} \text{ so that } g(r)^n = e^{n\gamma(r)}.$$

Then $\boxed{P(S_n \geq \alpha) \leq \exp\left[n\gamma(r) - r\alpha\right]; \ r \geq 0}$.

Note that $\gamma'(r) = g'(r)/g(r)$ and $\gamma'(0) = E[X]$.

Also $\gamma''(r) = \left[g''(r)g(r) - g'(r)^2\right]/g(r)^2$; $\gamma''(0) = \mathrm{VAR}(X)$.

Also $\gamma''(r) \geq 0$ ($\gamma(r)$ is convex). (Problem 7.7 in Gallager shows this.)

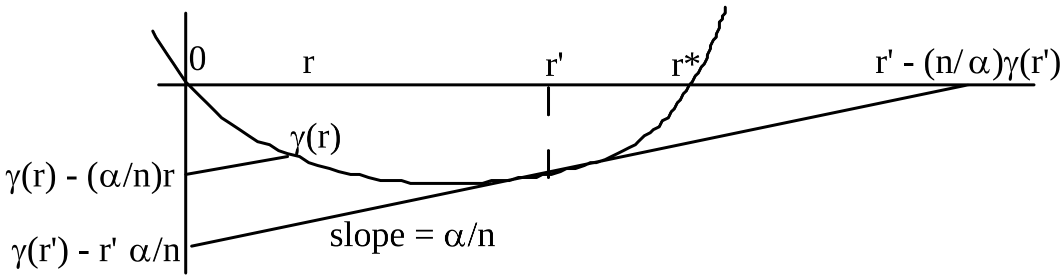


Optimization of the Chernov bound when $E[X] < 0$

$$P(S_n \geq \alpha) \leq \exp[n\gamma(r) - r\alpha]; \quad r \geq 0$$

$$P(S_n \geq \alpha) \leq \exp\{-n[r\alpha/n - \gamma(r)]\} = \exp\{-\alpha[r - (n/\alpha)\gamma(r)]\} \quad \text{True for all } r \geq 0.$$

$$P(S_n \geq \alpha) \leq \exp\{-n[r\alpha / n - \gamma(r)]\} = \exp\{-\alpha[r - (n/\alpha)\gamma(r)]\}$$ True for all $r \geq 0$.



For given $\alpha / n > 0 > E[X]$, r' above minimizes the bound. Analytically, r' is the r such that $d\gamma(r)/dr = \alpha / n$.

First form shows that, for fixed $\alpha / n > 0 > E[X]$, and increasing n, that $P(S_n \geq \alpha)$ approaches 0 exponentially in n. Exponent is the negative of the y intercept above.

Second form shows that for all n, and fixed $\alpha > 0 > E[X]$, that $P(S_n \geq a) \leq \exp(-r^* \alpha)$. Note $r' - (n/\alpha)\gamma(r') \geq r^*$. Thus, $P(S_n \geq \alpha) \leq e^{-r^*\alpha}$ is a less than optimally tight bound for each n.

Part of our interest in the optimized Chernoff bound comes because it is tight in the exponential sense that,

For α / n fixed, as $\alpha \to \infty$, $\ln(P(S_n \geq \alpha)) / \ln(\text{bound}) \to 1$