6.262: Discrete Stochastic Processes 5/11/11

L25: Putting it all together?

Outline:

- **Martingales**

- **Markov chains (Countable or finite state)**

- **Countable state Markov processes**

- **Random walks**

Martingales

A sequence $\{Z_n;\ n \geq 1\}$ is a martingale if for all $n > 1$,

$$\mathsf{E}\left[Z_n \mid Z_{n-1}, Z_{n-2}, \ldots, Z_1\right] = Z_{n-1}; \quad \mathsf{E}\left[|Z_n|\right] < \infty$$

Lemma: For a martingale, $\{Z_n;\ n \geq 1\}$, and for $n > i \geq 1$,

$$\mathsf{E}\left[Z_n \mid Z_i, Z_{i-1} \ldots, Z_1\right] = Z_i : \qquad \mathsf{E}\left[Z_n\right] = \mathsf{E}\left[Z_i\right]$$

The increments $X_n = Z_n - Z_{n-1}$ satisfy $\mathsf{E}\left[X_n \mid X_{n-1}, \ldots, X_1\right] = 0$ iff $\{Z_n;\ n \geq 1\}$ is a martingale. A good special case is $X_n = U_n Y_n$ where U_n are IID, $\mathsf{p}_U(1) = \mathsf{p}_U(-1) = 1/2$.

Examples: Zero mean RW and product of unit-mean IID rv's.

Submartingales

$\{Z_n; n \geq 1\}$ **is a submartingale if for all** $n \geq 1$**,**

$$\mathsf{E}\left[Z_{n+1} \mid Z_n, \ldots, Z_1\right] \geq Z_n; \quad \mathsf{E}\left[Z_n\right|] < \infty$$

Lemma: If $\{Z_n; n \geq 1\}$ **is a martingale, then for** $n > i > 0$**,**

$$\mathsf{E}\left[Z_n \mid Z_i, \ldots, Z_1\right] \geq Z_i; \quad \mathsf{E}\left[Z_n\right] \geq \mathsf{E}\left[Z_i\right]$$

If $h(x)$ **is convex, then Jensen's inequality says** $\mathsf{E}\left[h(X)\right] \geq h(\mathsf{E}\left[X\right])$**. If** $\{Z_n; n \geq 1\}$ **is a submartingale (including martingale),** h **is convex, and** $\mathsf{E}\left[|h(X)|\right] < \infty$**, then** $\{h(Z_n); n \geq 1\}$ **is a submartingale.**

Stopped (sub)martingales

A stopped process $\{Z_n^*; n \geq 1\}$ **for a possibly defective stopping time** J **on a process** $\{Z_n; n \geq 1\}$ **satisfies** $Z_n^* = Z_n$ **if** $n \leq J$ **and** $Z_n^* = Z_J$ **if** $n > J$**.**

Theorem: The stopped process, $\{Z_n^*; n \geq 1\}$**, for a (sub) martingale with a (possibly defective) stopping rule is a (sub)martingale and satisfies**

$$\mathsf{E}\left[Z_1\right] \leq \mathsf{E}\left[Z_n^*\right] \leq \mathsf{E}\left[Z_n\right] \qquad \textbf{(submartingale)}$$
$$\mathsf{E}\left[Z_1\right] = \mathsf{E}\left[Z_n^*\right] = \mathsf{E}\left[Z_n\right] \qquad \textbf{(martingale)}$$

For the product martingale $Z_n = \exp[rS_n - n\gamma(r)]$**, and a stopping rule** J **such as threshold crossing, we get a stopped martingale. Subject to some small mathematical nits, this leads to Wald's identity,**

$$\mathsf{E}\left[Z_J\right] = \mathsf{E}\left[\exp[rS_J - J\gamma(r)\right] = 1$$

Kolmgorov's submartingale inequality

Thm: Let $\{Z_n;\ n \geq 1\}$ be a non-negative submartingale. Then for any positive integer m and any $a > 0$,

$$\Pr\left\{\max_{1 \leq i \leq m} Z_i \geq a\right\} \leq \frac{\mathsf{E}\left[Z_m\right]}{a}.$$

This is Kolmogorov's strenghening of the Markov inequality.

For non-negative martingales, we can go to the limit $m \to \infty$ since $\mathsf{E}\left[Z_m\right] = \mathsf{E}\left[Z_1\right]$:

$$\Pr\left\{\sup_{n \geq 1} Z_n \geq a\right\} \leq \frac{\mathsf{E}\left[Z_1\right]}{a}$$

Kolmogorov version of Chebyshev: Let $\{Z_n;\ n \geq 1\}$ be a martingale (or submartingale) with $\mathsf{E}\left[Z_n^2\right] < \infty$ for all $n \geq 1$. Then

$$\Pr\left\{\max_{1 \leq n \leq m} |Z_n| \geq b\right\} \leq \frac{\mathsf{E}\left[Z_m^2\right]}{b^2}; \quad m \geq 1,\ b > 0.$$

It is often more useful to maximize only over half the interval and take a union over different intervals, e.g.,

$$\Pr\left\{\bigcup_{j \geq k} \left\{\max_{2^{j-1} < n \leq 2^j} |Z_n| \geq b_j\right\}\right\} \leq \sum_{j=k}^{\infty} \frac{E|Z_{2^j}|^2}{b_j^2}$$

For the zero-mean RW $\{S_n;\ n \geq 1\}$ where $S_n = X_1 + \cdots X_n$ with $\overline{X} = 0$ and $\mathsf{E}\left[X^2\right] = \sigma^2$, $\mathsf{E}\left[Z_{2^j}^2\right] = 2^j \sigma^2$.

$$\Pr\left\{\bigcup_{j \geq k} \left\{\max_{2^{j-1} < n \leq 2^j} |S_n| \geq \left(\frac{3}{2}\right)^j\right\}\right\} \leq \sum_{j=k}^{\infty} \left(\frac{8}{9}\right)^j \sigma^2 = \left(\frac{8}{9}\right)^k 9\sigma^2$$

where $b_j = (3/2)^j$.

SLLN for $\{S_n; \, n \geq 1$ **where** $S_n = X_1 + \cdots X_n$ **and** $\{X_n; \, n \geq 1\}$ **are IID with** $\overline{X} = 0, \mathsf{E}\left[X^2\right] = \sigma^2 < \infty$. **Then**

$$\Pr\left\{\omega : \lim_{n \to \infty} \frac{S_n}{n} = 0\right\} = 1.$$

Proof:

$$\Pr\left\{\bigcup_{j \geq k} \left\{\max_{2^{j-1} < n \leq 2^j} |S_n| \geq \left(\frac{3}{2}\right)^j\right\}\right\} \leq \left(\frac{8}{9}\right)^k 9\sigma^2$$

Lower bounding the left side,

$$\Pr\left\{\bigcup_{j \geq k} \left\{\max_{2^{j-1} < n \leq 2^j} \frac{|S_n|}{n} \geq 2\left(\frac{3}{4}\right)^j\right\}\right\} \leq \left(\frac{8}{9}\right)^k 9\sigma^2$$

Any sample sequence $\{S_n(\omega); \, n \geq 1\}$ **that is not contained in the union on the left must satisfy** $\lim_{n \to \infty} |S_n|/n = 0$, **and thus** $\Pr\{\omega : \lim_n |S_n|/n = 0\} > 1 - (8/9)^k 9\sigma^2$. **Since this is true for all** k, **the theorem is proved. It applies to other martingales also.**

Markov chains (Countable or finite state)

Def: The first passage time T_{ij} **from state** i **to** j **is the smallest** n, **given** $X_0 = i$, **at which** $X_n = j$. T_{ij} **is a possibly defective rv with PMF** $f_{ij}(n)$ **and dist. fcn.** $\mathsf{F}_{ij}(n)$.

$$f_{ij}(n) = \sum_{k \neq j} P_{ik} f_{kj}(n-1); \quad \mathsf{F}_{ij}(n) = \sum_{m=1}^{n} f_{ij}(m); \quad n > 1.$$

Def: State j **is recurrent if** T_{jj} **is non-defective and transient otherwise. If recurrent, it is positive recurrent if** $\mathsf{E}\left[T_{jj}\right] < \infty$ **and null recurrent otherwise.**

For each recurrent j **there is a integer renewal counting process** $\{N_{jj}(t); t > 0\}$ **of visits to** j **starting in** j. **It has interrenewal distribution** $F_{jj}(n)$.

There is a delayed renewal counting process $\{N_{ij}(t); t > 0\}$ **for visits to** j **starting in** i.

Thm: All states in a class are positive recurrent, or all are null recurrent, or all are transient.

Def: A chain is irreducible if all state pairs communicate.

It is called irreducible because the problem of getting from one or another transient class to an 'irreducible class' is largely separable from the analysis of that 'irreducible class' which then becomes the entire chain.

An irreducible class can be positive recurrent, null recurrent, or transient.

Thm: For an irreducible Markov chain, if the 'steady state' equations

$$\pi_j = \sum_i \pi_i P_{ij} \text{ and } \pi_j \geq 0 \text{ for all j;} \sum_j \pi_j = 1$$

has a solution, then the solution is unique, $\pi_j = 1/\overline{T}_{jj} > 0$ for all j, and the states are positive recurrent. Also if the states are positive recurrent, then the steady state equations have a solution.

This is an infinite set of equations, so not necessarily computer solvable.

The counting processes under positive recurrence must satisfy

$$\lim_{n \to \infty} \frac{N_{ij}(n)}{n} = \pi_j \quad \textbf{WP1}$$
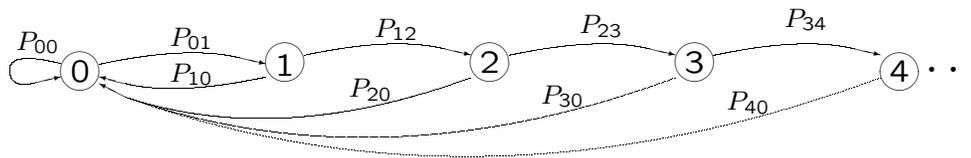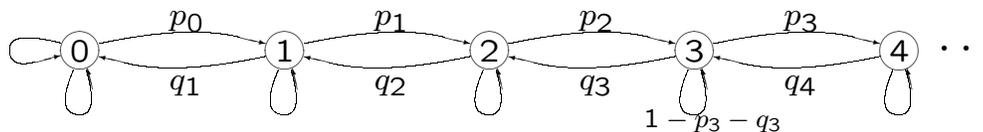
Markov model of age of renewal process

$$\pi_n = \pi_0 P_{01} P_{12} \cdots P_{n-1,n} = \Pr\{Z > n\}$$

$$1 = \sum_i \pi_i = \pi_0 \sum_{i=0}^{\infty} \Pr\{Z > n\} = \pi_0 \overline{Z}$$

This is a nice chain for examples about null-recurrence.

A birth-death Markov chain

This is another very useful model for examples about recurrence and for models of sampled-time queueing systems.
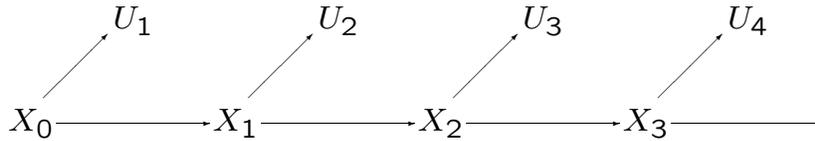
Note that the steady state equations are almost trivial.

$$\pi_i p_i = \pi_{i+1} q_{i+1}; \qquad \frac{\pi_{i+1}}{\pi_i} = \rho_i$$
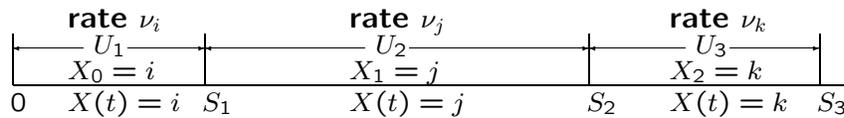
where $\rho_i = p_i/q_{i+1}$.

Markov processes

A Markov process $\{X(t); t \geq 0\}$ is a combination of a countable state Markov chain $\{X_n; n \geq 1\}$ along with an exponential holding time U_n for each state.

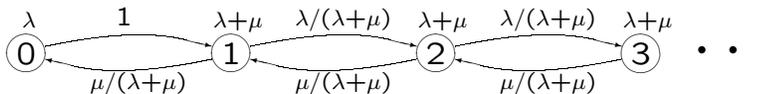$$\Pr\{U_n \leq \tau \mid X_n = j, \mathbf{past}\} = 1 - \exp(\tau \nu_j)$$

A Markov process is specified by the embedded transition probabilities P_{ij} and the rates ν_i.
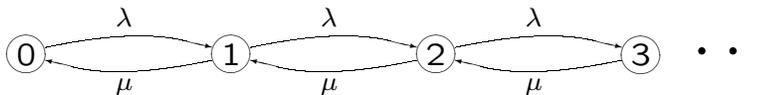

Def: Transition rates are given by $q_{ij} = \nu_i P_{ij}$.
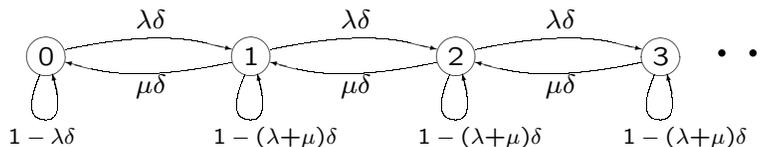
Three ways to represent a Markov process.

An M/M/1 queue using $[P]$ and ν
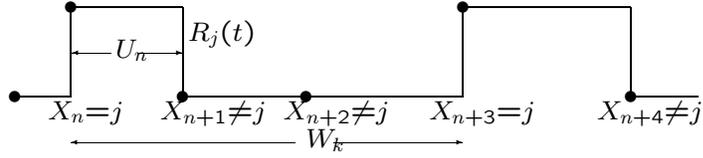

The same M/M/1 queue using $[q]$.

The same M/M/1 queue in sampled time.

From the (delayed) renewal reward theorem,

$$p_j = \lim_{t \to \infty} \frac{\int_0^t R_j(\tau)d\tau}{t} = \frac{\overline{U}(j)}{\overline{W}(j)} = \frac{1}{\nu_j \overline{W}(j)} \qquad \text{W.P.1.}$$

We can also assign unit reward for each transition in the renewal interval from state j to j. Let $M_i(t)$ be number of transitions in $(0, t]$ given that $X(0) = i$.

$$\overline{M}_i = \lim_{t \to \infty} \frac{M_i(t)}{t} = \frac{1}{\pi_j \overline{W}_j} \qquad \text{W.P.1.}$$

$$p_j = \frac{\pi_j}{\nu_j} \overline{M}_i$$

$$p_j = \frac{1}{\nu_j \overline{W}(j)}; \qquad \overline{M} = \frac{1}{\pi_j \overline{W}_j}; \qquad p_j = \frac{\pi_j}{\nu_j} \overline{M}$$

If $0 < \overline{M} < \infty$, then each $p_j > 0$ and $\sum_j p_j = 1$.

$$\overline{M} = \frac{1}{\sum_i \pi_i/\nu_i}; \qquad p_j = \frac{\pi_j/\nu_j}{\sum_i \pi_i/\nu_i}$$

Similarly, since $\sum_i \pi_i = 1$,

$$\overline{M} = \sum_i p_i \nu_i; \qquad \pi_j = \frac{p_j \nu_j}{\sum_i p_i \nu_i}$$

A sampled time MP exists if $\nu_i \leq A$ for some A and all i. The steady state probabilities are the time average probabilities, $\{p_i\}$, which satisfy

$$\nu_j p_j = \sum_i \nu_i p_j; \qquad p_j \geq 0; \qquad \sum_i p_i = 1$$

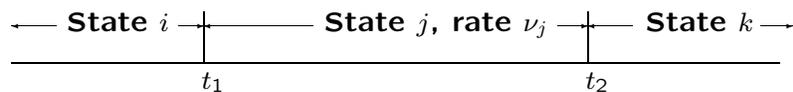The strange cases occur when $\overline{M}$ is 0 or infinite.

$\overline{M} = 0$ for the rattled server and discouraged customer queue. Here the embedded chain is positive recurrent, but all p_j are zero and the sampled-time chain is null recurrent.

The case $\overline{M} = \infty$ is not possible for a positive recurrent embedded chain, but is possible when the equations $\nu_j p_j = \sum_i \nu_i p_i$ have a solution. These processes are irregular, and allow an infinite number of transitions in finite time.

Reversibility for MP's

For any (embedded) Markov chain, the backward transition probabilities are defined as

$$\pi_i P_{ij}^* = \pi_j P_{ji}$$

$$\longleftarrow \text{State } i \quad \big| \quad \text{State } j, \text{ rate } \nu_j \longrightarrow \big| \longleftarrow \text{State } k \longrightarrow$$
$$ t_1 t_2$$

Moving right, after entering state j, the exit rate is ν_j, i.e., we exit in each δ with probability $\nu_j \delta$. The some holds moving left.

Thus $\{\pi_i\}$ and $\{\nu_i\}$ are the same going left as going right

Note that the probability of having a (right) transition from state j to k in $(t, t+\delta)$ is $p_j q_{jk} \delta$. Similarly, for the left going process, if q^*_{kj} is the process transition rate, the probability of having the same transition is $p_k q^*_{kj}$. Thus

$$p_j q_{jk} = p_k q^*_{kj}$$

By fiddling equations, $q^*_{kj} = \nu_k P^*_{kj}$.

Def: A MP is reversible if $q^*_{ij} = q_{ji}$ for all i, j

Assuming positive recurrence and $\sum_i \pi_i / \nu_i < \infty$, the MP process is reversible if and only if the embedded chain is.

The guessing theorem: Suppose a MP is irreducible and $\{p_i\}$ is a set of probabilities and satisfies $p_i q_{ij} = p_j q_{ji}$ for all i, j and satisfies $\sum_i p_i \nu_i < \infty$.

Then $p_i > 0$ for all i, p_i is the steady state time average probability of state i, the process is reversible, and the embedded chain is positive recurrent.

Useful application: All birth/death processes are reversible (if $\sum_j p_j \nu_j < \infty$)

Random Walks

Def: A random walk is a sequence $\{S_n; n \geq 1\}$ of successive sums $S_n = X_1 + \cdots + X_n$ of IID rv's X_i.

We are interested in exponential bounds on S_n for large n (Chernoff bounds)

We are also interested in threshold crossings: for 2 thresholds, $\alpha > 0$ and $\beta < 0$, what is the stopping time when S_n first crosses α or β, what is the probability of crossing each, and what is the distribution of the overshoot.

Let a rv Z have an MGF $g_Z(r)$ for $0 \leq r < r_+$ and mean $\overline{Z} < 0$. By the Chernoff bound, for any $\alpha > 0$ and any $r \in (0, r_+)$,

$$\Pr\{Z \geq \alpha\} \leq g_Z(r)\exp(-r\alpha) = \exp(\gamma_Z(r) - r\alpha)$$

where $\gamma_Z(r) = \ln g_Z(r)$. If Z is a sum $S_n = X_1 + \cdots + X_n$, of IID rv's, then $\gamma_{S_n}(r) = n\gamma_X(r)$.

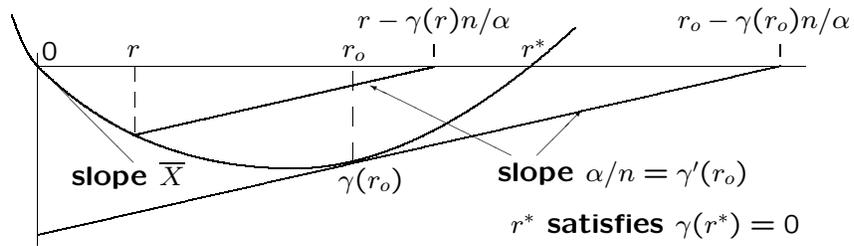$$\Pr\{S_n \geq na\} \leq \min_r \left(\exp[n(\gamma_X(r) - ra)]\right).$$

This is exponential in n for fixed a (i.e., $\gamma'(r) = a$). We are now interested in threshold crossings, i.e., $\Pr\{\bigcup_n (S_n \geq \alpha)\}$. As a preliminary step, we study how $\Pr\{S_n \geq \alpha\}$ varies with n for fixed α.

$$\Pr\{S_n \geq \alpha\} \leq \min_r \left(\exp[n\gamma_X(r) - r\alpha]\right).$$

Here the minimizing r varies with n (i.e., $\gamma'(r) = \alpha/n$).

$$\Pr\{S_n \geq \alpha\} \leq \min_{0 < r < r_+} \exp\left(-\alpha\left[r - \frac{n}{\alpha}\gamma_X(r)\right]\right)$$

$r - \gamma(r)n/\alpha$ $r_o - \gamma(r_o)n/\alpha$

0   r r_o r^*

slope $\overline{X}$ $\gamma(r_o)$ **slope** $\alpha/n = \gamma'(r_o)$

r^* **satisfies** $\gamma(r^*) = 0$

When n is very large, the slope $\frac{\alpha}{n} = \gamma'_X(r_0)$ is close to 0 and the horizontal intercept (the negative exponent) is very large. As n decreases, the intercept decreases to r^* and then increases again.

Thus $\Pr\{\bigcup_n\{S_n \geq \alpha\}\} \approx \exp(-\alpha r^*)$, where the nature of the approximation will be explained in terms of the Wald identity.

Thm: (Wald) Let $\gamma(r) = \ln(\mathsf{E}\left[\exp(rX)\right])$ exist over (r_-, r_+) containing 0. Let J be trial at which S_n first exceeds $\alpha > 0$ or $\beta < 0$. Then

$$\mathsf{E}\left[\exp(rS_J - J\gamma(r)\right] = 1 \quad \text{for } r \in (r_-, r_+)$$

More generally theorem holds if stopping time is a rv under both the given probability and the tilted probability.

The proof simply sums the probabilities of the stopping nodes under both the probability measure and the tilted probability measure.

$$\mathsf{E}\left[S_J\right] = \mathsf{E}\left[J\right]\overline{X}$$

$$\mathsf{E}\left[S_J^2\right] = \mathsf{E}\left[J\right]\sigma_X^2 \qquad \text{if } \overline{X} = 0$$

These bounds are all exponentially tight.

If any ϵ is added to any such exponent, the upper bound becomes a lower bound at sufficiently large α with fixed α/n.

The slack in the bounds come partly from the overshoot and partly from the lower threshold.

The lower threshold is unimportant if both thresholds are far from 0.

The overshoot is similar to residual life for renewal processes. It doesn't exist for simple random walks and is easy to calculate if the positive part of the density of X is exponential.

Wald's identity for 2 thresholds

Let $\{X_i; i \geq 1\}$ be IID with $\gamma(r) = \ln(\mathsf{E}\,[\exp(rX)])$ for $r \in (r_- < 0 < r_+)$. Let $\{S_n; n \geq 1$ be the RW with $S_n = X_1 + \cdots + X_n$. If J is the trial at which S_n first crosses $\alpha > 0$ or $\beta < 0$, then

$$\mathsf{E}\,[\exp(rS_J - J\gamma(r)] = 1 \quad \text{for } r \in (r_-, r_+)$$

Corollary: If $\overline{X} < 0$ and $r^* > 0$ satisfies $\gamma(r^*) = 0$, then

$$\Pr\{S_J \geq \alpha\} \leq \exp(-\alpha r^*)$$

Review of hypothesis testing: View a binary hypothesis as a binary rv H with $\mathsf{p}_H(0) = p_0$ and $\mathsf{p}_H(1) = p_1$.

We observe $\{Y_n; n \geq 1\}$, which, conditional on $H = \ell$ is IID with density $\mathsf{f}_{Y|H}(Y|\ell)$. Define the log likelihood ratio

$$LLR(\vec{Y}^n) = \sum_{i=1}^{n} \ln \frac{\mathsf{f}_{Y_i|H}(Y_i|0)}{\mathsf{f}_{Y_i|H}(y_i|1)}$$

$$\ln \frac{\Pr\{H=0 \mid \vec{y}^n\}}{\Pr\{H=1 \mid \vec{y}^n\}} = \ln \frac{p_0 f_{\vec{Y}^n|H}(\vec{y}^n \mid 0)}{p_1 f_{\vec{Y}^n|H}(\vec{y}^n \mid 1)} = \ln \frac{p_0}{p_1} + LLR(\vec{y}^n).$$

MAP rule: $\quad LLR(\vec{y}^n) \begin{cases} > \ln p_1/p_0 & ; \qquad \textbf{select } \hat{h}=0 \\ \leq \ln p_1/p_0 & ; \qquad \textbf{select } \hat{h}=1. \end{cases}$

$$S_n = \sum_{i=1}^{n} Z_i \qquad \textbf{where } Z_i = \ln \frac{\mathsf{f}_{Y_i|H}(Y_i|0)}{\mathsf{f}_{Y_i|H}(Y_i|1)}$$

Conditional on $H = 1$, $\{S_n; n \geq 1\}$ is a RW; $S_n = Z_1 + \cdots Z_n$ with the density $\mathsf{f}_{Y|H}(y|1)$ where Z is a function of Y. $\gamma_1(r) = \ln \mathsf{E}\left[e^r Z \mid H = 1\right]$ and $r^* = 1$ for $\gamma_1(r)$.

Conditional on $H = 0$, $\{S_n; n \geq 1\}$ is a RW; $S_n = Z_1 + \cdots Z_n$ with the density $\mathsf{f}_{Y|H}(y|0)$ where Z is the same function of Y.

Under $H = 1$, the RW has negative drift. With thresholds at $\alpha > 0$ and $\beta < 0$, $\Pr\{S_J \geq \alpha\} \leq e^{-\alpha}$. This is the probability of error given $H = 1$ and is essentially the probability of error conditional on any sample path crossing α.

Under $H = 0$, the RW has positive drift and $\Pr\{S_J \leq \beta\} \leq e^{\beta}$. This also essentially holds for each sample path.

6.262 Discrete Stochastic Processes
Spring 2011