6.262: Discrete Stochastic Processes
Lecture - Putting it all together?, 5/6/09

- Markov chains

- Markov processes

- Random walks

Markov chains (finite or countable)

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. $F_{ij}(n)$.

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.

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

A chain is irreducible if all states communicate with each other.

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.

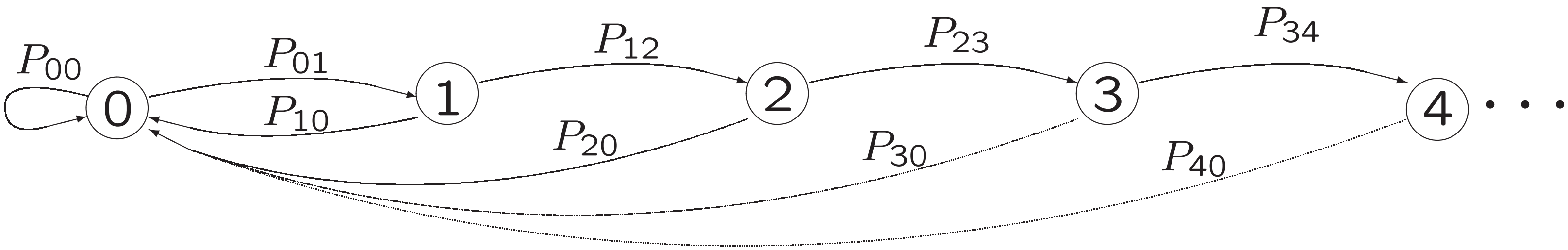
An irreducible class is positive recurrent, null recurrent, or transient.

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.

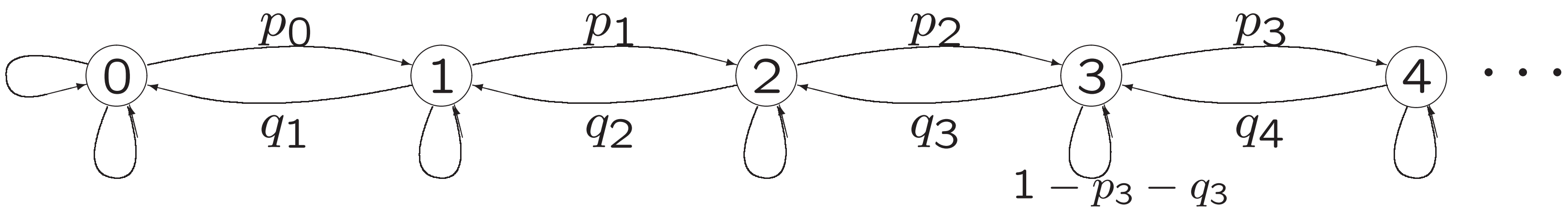


Markov model of age of renewal process

$$\pi_n = \pi_0 P_{01} P_{12} \cdots P_{n-1,n} = \mathbf{P}\{W > n\}$$

$$1 = \sum_i \pi_i = \pi_0 \sum_{i=0}^{\infty} \mathbf{P}\{W > n\} = \pi_0 \overline{W}$$

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 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.

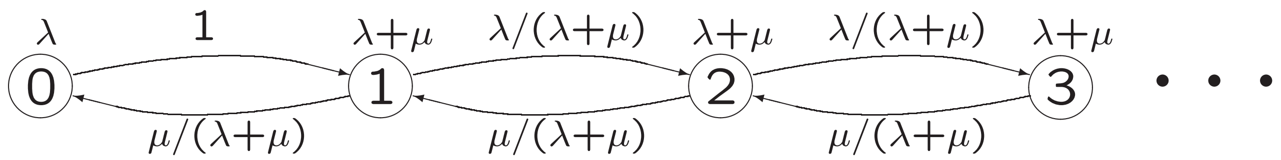
$$\mathbf{P}\{U_n \leq \tau \mid X_n = j, \textbf{past}\} = 1 - \exp(\tau\nu_j)$$

$$\begin{tikzcd}
 & U_1 & & U_2 & & U_3 & & U_4 & \\
X_0 \arrow[ur] \arrow[rr] & & X_1 \arrow[ur] \arrow[rr] & & X_2 \arrow[ur] \arrow[rr] & & X_3 \arrow[ur] \arrow[rr] & & {}
\end{tikzcd}$$

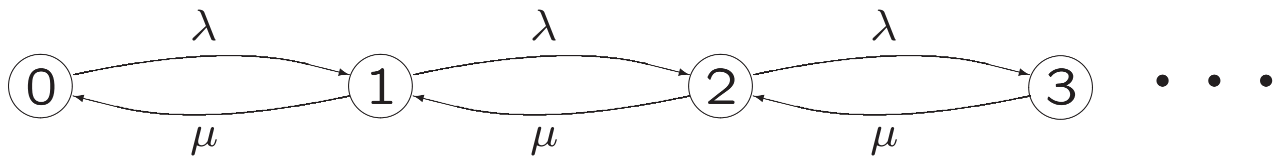
A Markov process is specified by the embedded transition probabilities P_{ij} the rates ν_i.

Transition rates are given by $q_{ij} = \nu_i P_{ij}$.

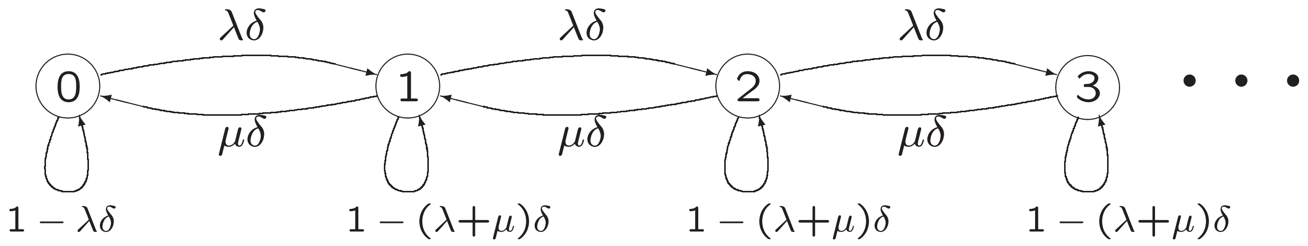
Three ways to represent a Markov process.



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.

Review: An irreducible countable state Markov chain is positive recurrent iff the steady state equations,

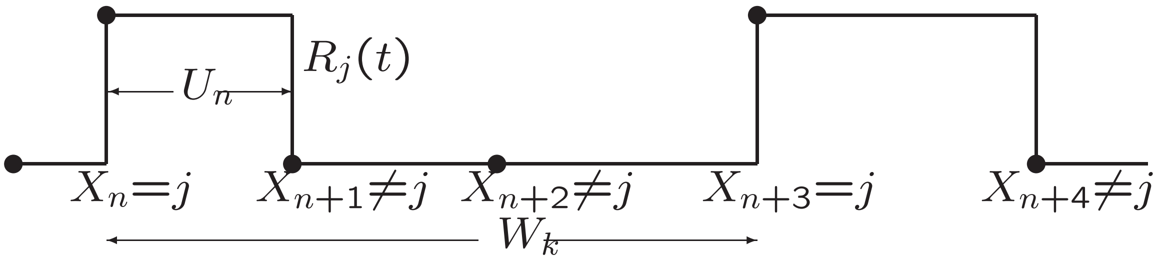
$$\pi_j = \sum_i \pi_i P_{ij} \text{ for all } j; \;\; \pi_j \geq 0 \text{ for all } j; \;\; \sum_j \pi_j = 1$$

have a solution. If there is a solution, it is unique and $\pi_i > 0$ for all i. Also, the number of visits $N_{ij}(n)$ in the first n transitions to j given $X_0 = i$ satisfies

$$\lim_{n\to\infty} N_{ij}(n) = \pi_j \qquad \textbf{W.P.1}$$

The fraction of time in state j (if $\sum_i \pi_i/\nu_i < \infty$) is

$$p_j = \frac{\pi_j/\nu_j}{\sum_i \pi_i/\nu_i}$$



From the (delayed) renewal reward theorem,

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

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

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

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

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

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.

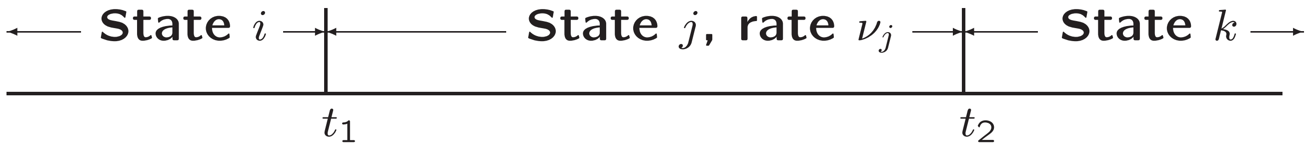
$\overline{M} = 0$ for the rattled server and discouraged customer queue. Here the embedded chain is positive recurrent, but all p_j are zero and the sampled-time chain is null recurrent.

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

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



State i

State j, rate ν_j

State k

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.

- **Wald's identity and threshold crossings**
- **Waiting time in a G/G/1 queue**
- **Large deviations - binary hypothesis testing**

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

$$\mathsf{E}[\exp(rS_N - N\gamma(r)] = 1 \qquad \text{for } r \in (r_-, r_+)$$

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.

$$\mathsf{E}[S_N] = \mathsf{E}[N]\,\overline{X}$$

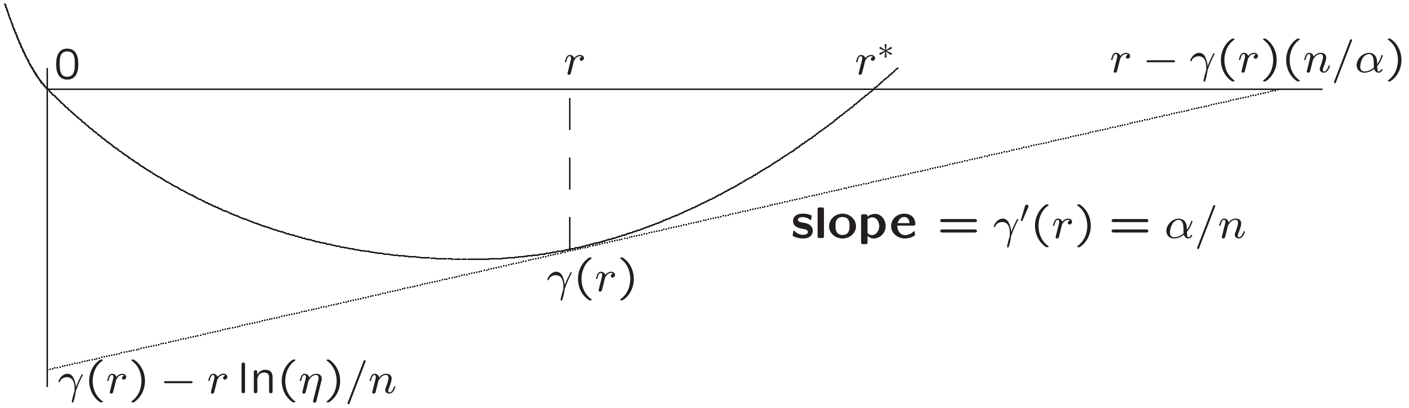
$$\mathsf{E}\left[S_N^2\right] = \mathsf{E}[N]\,\sigma_X^2 \qquad \text{if } \overline{X} = 0$$

Large deviations: Suppose E$[X] < 0$ and we want to upperbound P$\{S_N \geq \alpha, N \geq n\}$.

$$\begin{aligned} 1 &\geq \mathbf{E}\left[\exp[rS_N - N\gamma(r)] \mid N{\geq}n, S_N{\geq}\alpha\right] \mathbf{P}\{N \geq n,\ S_N{\geq}\alpha\} \\ &\geq \exp[r\alpha - n\gamma(r)]\mathbf{P}\{N{\geq}n, S_N{\geq}\alpha\} \end{aligned}$$

for any r such that $r > 0$ and $\gamma(r) \leq 0$, ($r \leq r^*$).

$$\mathbf{P}\{N \geq n, S_N \geq \alpha\} \leq \exp[-r\alpha + n\gamma(r)]$$



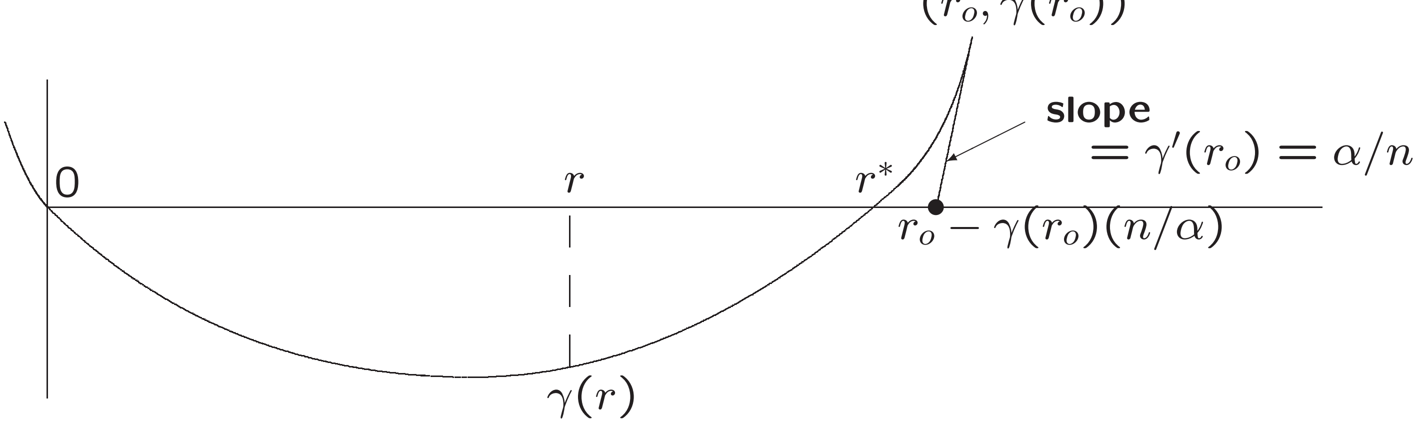
If $\alpha/n < \gamma'(r^*)$, this is the same bound as the exponential bound on P$\{S_n \geq \alpha\}$.

To upperbound $\mathbf{P}\{S_N \geq \alpha, N \leq n\}$,

$$\begin{aligned} 1 &\geq \mathsf{E}[\exp[rS_N - N\gamma(r)] \mid N{\leq}n, S_N{\leq}\alpha]\,\mathsf{P}\{N \leq n,\ S_N{\geq}\alpha\} \\ &\geq \exp[r\alpha - n\gamma(r)]\mathsf{P}\{N{\leq}n, S_N{\geq}\alpha\} \end{aligned}$$

for any r such that $r > 0$ and $\gamma(r) \geq 0$, ($r \geq r^*$).

$$\mathsf{P}\{N \leq n, S_N \geq \alpha\} \leq \exp[-r\alpha + n\gamma(r)]$$



If $\alpha/n \geq \gamma'(r^*)$, this is the same bound as the exponential bound on $\mathbf{P}\{S_n \geq \alpha\}$.

The bound on $\mathsf{P}\{S_N \ge \alpha, N \ge n\}$ is useful for $n \ge n^* = \alpha/\gamma'(r^*)$.

The bound on $\mathsf{P}\{S_N \le \alpha, N \le n\}$ is useful for $n \le \alpha/\gamma'(r^*)$.

The overall bound is

$$\mathsf{P}\{S_N \ge \alpha\} \le \exp(-r^*\alpha)$$

This is also the bound for $\mathsf{P}\{S_N \ge \alpha, N \ge n^*\}$, for $\mathsf{P}\{S_N \ge \alpha, N \le n^*\}$, and $\mathsf{P}\{S_{n^*} \ge \alpha\}$.

This says that, given a crossing of α, the trial at which it occurs is close to n^*, with exponentially smaller probabilities elsewhere.

These bounds are all exponentially tight.

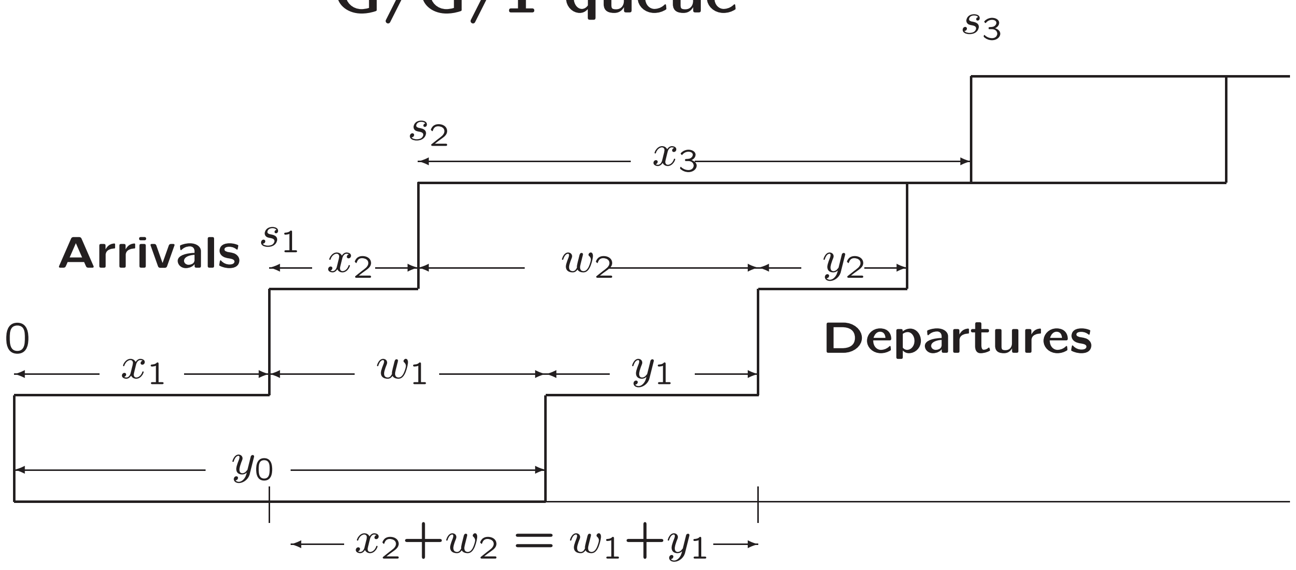
If any ε is added to any such exponenent, the upper bound becomes a lower bound at sufficiently large α with fixed α/n.

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.

G/G/1 queue



$$U_i = Y_{i-1} - X_i; \qquad \{U_1, U_2, \dots\} \text{ are IID}$$

$$\begin{aligned} W_i &= \max[U_i + W_{i-1},\ 0] \\ &= \max[(U_i{+}U_{i-1}{+}\cdots + U_1),\ (U_i{+}U_{i-1}),\ U_i,\ 0] \\ &= \max[\textbf{first i terms of a backward RW}]. \end{aligned}$$

P $\{W_i \geq \alpha\}$ **is the probability of a threshold crossing by trial** i**. If E** $[U] < 0$**,** W_i **converges in distribution to** W**. P** $\{W \geq \alpha\} \leq \exp(-r^*\alpha)$**.**

Hypothesis testing

Assume a probability space with a binary hypothesis $H = \{\mathsf{H}_0, \mathsf{H}_1\}$, PMF $\{p_0, p_1\}$, and a random vector $\mathbf{Y}$ described by a density $f(\mathbf{Y}|H)$.

Given a sample value of $\mathbf{Y}$, we choose a sample value of H to satisfy some criterion.

The rule for choosing is called a statistical test (test for short).

A test is identified by the set A of sample values of $\mathbf{Y}$ for which H_1 is chosen.

$$\begin{aligned}
\mathsf{P}_A\{e \mid \mathsf{H}_0\} &= \mathsf{P}_A\{\mathbf{Y} \in A \mid \mathsf{H}_0\} \\
\mathsf{P}_A\{e \mid \mathsf{H}_1\} &= \mathsf{P}_A\{\mathbf{Y} \in A^c \mid \mathsf{H}_1\} \\
\mathsf{P}_A\{e\} &= p_0\mathsf{P}_A\{\mathbf{Y} \in A \mid \mathsf{H}_0\} + p_1\mathsf{P}_A\{\mathbf{Y} \in A^c \mid \mathsf{H}_1\}
\end{aligned}$$

Minimizing error probability is easy. For any given y, choose max of $\mathsf{P}\{\mathsf{H}_0 \mid \mathbf{y}\}$ and $\mathsf{P}\{\mathsf{H}_1 \mid \mathbf{y}\}$

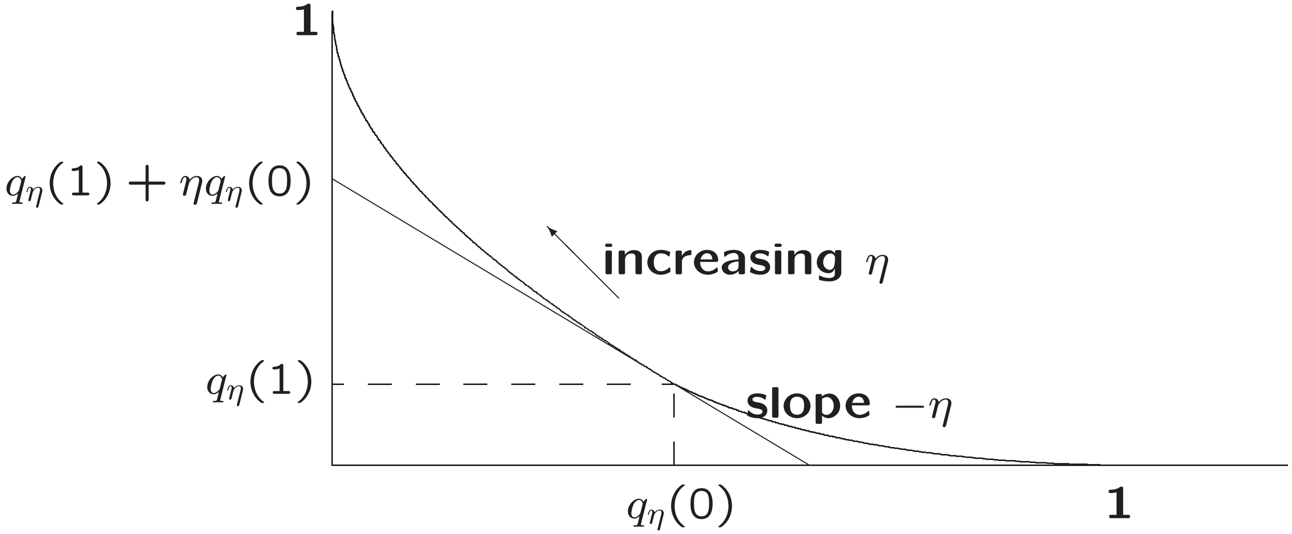
$$\frac{\mathsf{P}\{\mathsf{H}_1 \mid \mathbf{y}\}}{\mathsf{P}\{\mathsf{H}_0 \mid \mathbf{y}\}} = \frac{p_1}{p_0}\,\frac{f(\mathbf{y} \mid \mathsf{H}_1)}{f(\mathbf{y} \mid \mathsf{H}_0)}$$

$$\text{Let } \Lambda(\mathbf{y}) = \frac{f(\mathbf{y} \mid \mathsf{H}_1)}{f(\mathbf{y} \mid \mathsf{H}_0)}; \qquad \eta = \frac{p_0}{p_1}$$

This is the MAP test and also a threshold test T_η.

$$\Lambda(\mathbf{y}) \begin{cases} > \eta & ; \quad \text{choose } \mathsf{H}_1 \\ < \eta & ; \quad \text{choose } \mathsf{H}_0 \\ = \eta & ; \quad \text{don't care, choose either} \end{cases}$$

$$\begin{aligned} \mathsf{P}_\eta\{e \mid \mathsf{H}_0\} &= \mathsf{P}\{\Lambda(\mathbf{Y}) > \eta \mid \mathsf{H}_0\} \\ \mathsf{P}_\eta\{e \mid \mathsf{H}_1\} &= \mathsf{P}\{\Lambda(\mathbf{Y}) \le \eta \mid \mathsf{H}_1\} \end{aligned}$$



$$\begin{aligned} q_\eta(0) &= \mathsf{P}_\eta\{e \mid \mathsf{H}_0\} &= \mathsf{P}\{\Lambda(\mathbf{Y}) > \eta \mid \mathsf{H}_0\} \\ q_\eta(1) &= \mathsf{P}_\eta\{e \mid \mathsf{H}_1\} &= \mathsf{P}\{\Lambda(\mathbf{Y}) \le \eta \mid \mathsf{H}_1\} \end{aligned}$$

As $\eta \nearrow$, $q_\eta(0) \searrow$ and $q_\eta(1) \nearrow$.

The parametric function $(q_\eta(0), q_\eta(1))$ of η is called the error curve.

It is convex and all tests lie NE of it.

Large deviations for hypothesis tests

Let Y be IID conditional on H_0 and also IID condtional on H_1. Then

$$\ln(\Lambda(\mathbf{y})) = \ln \frac{f(\mathbf{y} \mid \mathbf{H}_1)}{f(\mathbf{y} \mid \mathbf{H}_0)} = \sum_{i=1}^{n} \ln \frac{f(y_i \mid \mathbf{H}_1)}{f(y_i \mid \mathbf{H}_0)}$$

$$z_i = \ln \frac{f(y_i \mid \mathbf{H}_1)}{f(y_i \mid \mathbf{H}_0)}$$

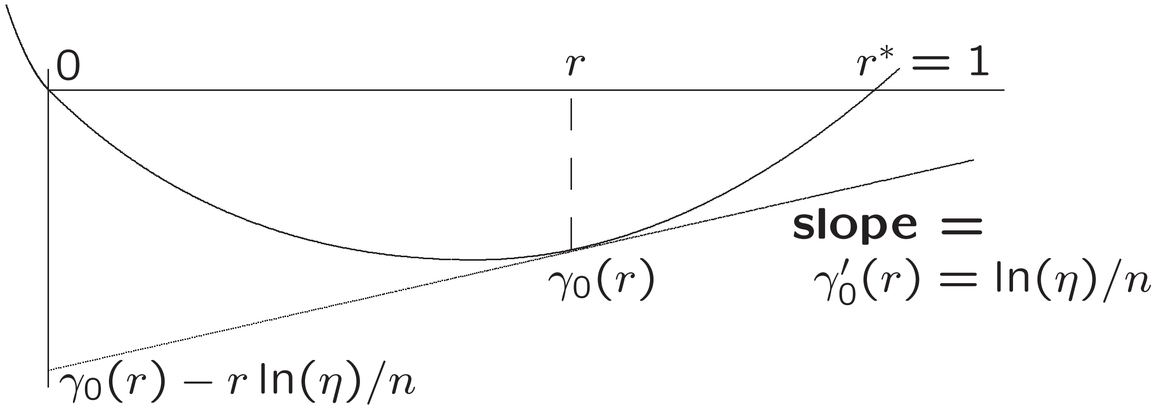
A threshold test compares $\sum_{i=1}^{n} z_i$ with $\ln(\eta)$.

Conditional on H_0, make error if $\sum_i Z_i^0 > \ln(\eta)$ where Z_i^0, $1 \le i \le n$ are IID conditional on H_0.

Exponential bound for $\sum_i Z_i^0$

$$\begin{aligned}\gamma_0(r) &= \ln\left\{\int f(y \mid \mathbf{H}_0)\exp\left[r\ln\frac{f(y \mid \mathbf{H}_1)}{f(y \mid \mathbf{H}_0)}\right]\,dy\right\}\\ &= \ln\left\{\int f^{1-r}(y \mid \mathbf{H}_0)f^{r}(y \mid \mathbf{H}_1)\,dy\right\}\end{aligned}$$

At $r = 1$**, this is** $\ln(\int f(y \mid \mathbf{H}_1)\,dy) = 0$**.**

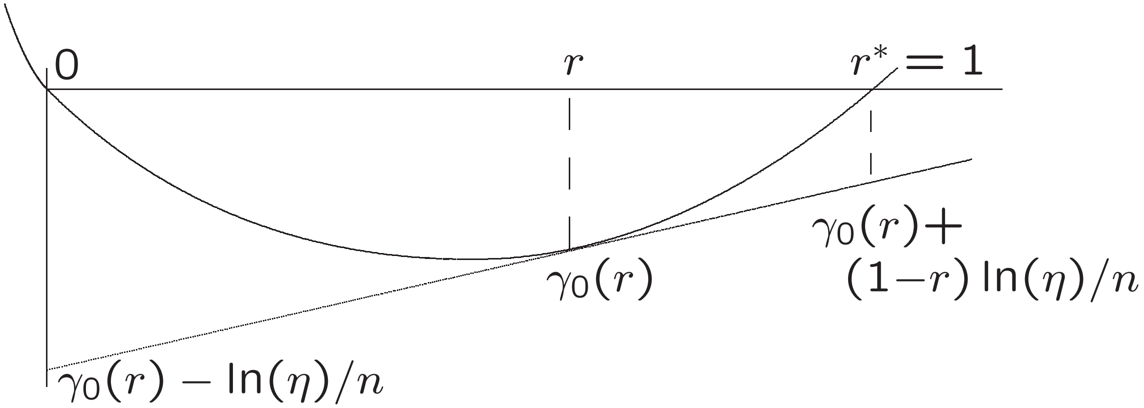


$$q_0(\eta) \le \exp n\left[\gamma_0(r_0) - r_0\ln(\eta)/n\right]$$

Exponential bound for $\sum_i Z_i^1$

$$\begin{aligned}\gamma_1(s) &= \ln\left\{\int f(y \mid \mathbf{H}_1)\exp\left[s\ln\frac{f(y \mid \mathbf{H}_1)}{f(y \mid \mathbf{H}_0)}\right]dy\right\}\\ &= \ln\left\{\int f^{-s}(y \mid \mathbf{H}_0)f^{1+s}(y \mid \mathbf{H}_1)\,dy\right\}\end{aligned}$$

At $s=-1$**, this is** $\ln(\int f(y \mid \mathbf{H}_0)\,dy)=0$. **Note:** $\gamma_1(s)=\gamma_0(r{-}1)$.



$$q_1(\eta) \le \exp n\left[\gamma_0(r)+(1{-}r)\ln(\eta)/n\right]$$

For hypothesis testing with a variable number of trials, use two thresholds, $\alpha > 0, \beta < 0$.

Given H_0, an error occurs if $\sum_i z_i \ge \alpha$.

$$\mathbf{P}\{e \mid \mathbf{H}_0\} \le e^{-\alpha}; \qquad \mathbf{E}[N \mid \mathbf{H}_0] \approx \frac{\beta}{\gamma_0'(0)}$$

$$\mathbf{P}\{e \mid \mathbf{H}_1\} \le e^{\beta}; \qquad \mathbf{E}[N \mid \mathbf{H}_1] \approx \frac{\alpha}{\gamma_0'(1)}$$

Choose $n^* = \beta/\gamma_0'(0) = \alpha/\gamma_0'(1)$. Then

$$\mathbf{P}\{e \mid \mathbf{H}_0\} \le \exp[-n^*\gamma_0'(1)]; \qquad \mathbf{P}\{e \mid \mathbf{H}_1\} \le \exp[-n^*\gamma_0'(0)]$$