6.262: Discrete Stochastic Processes
Lecture - Hypothesis testing & Kolmogorov, 5/6/09

- Threshold tests
- The error curve
- The Neyman-Pearson test
- Sequential hypothesis testing
- Kolmogorov's martingale inequality

Hypothesis testing

Assume a probability space with a binary hypothesis $H = \{\mathbf{H}_0, \mathbf{H}_1\}$, 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 $\mathbf{H}_1$ is chosen.

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

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.

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

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

The max reward problem is just as easy.

$$\frac{r_1\mathbf{P}\{\mathbf{H}_1 \mid \mathbf{y}\}}{r_0\mathbf{P}\{\mathbf{H}_0 \mid \mathbf{y}\}} = \frac{r_1 p_1}{r_0 p_0}\,\frac{f(\mathbf{y} \mid \mathbf{H}_1)}{f(\mathbf{y} \mid \mathbf{H}_0)}$$

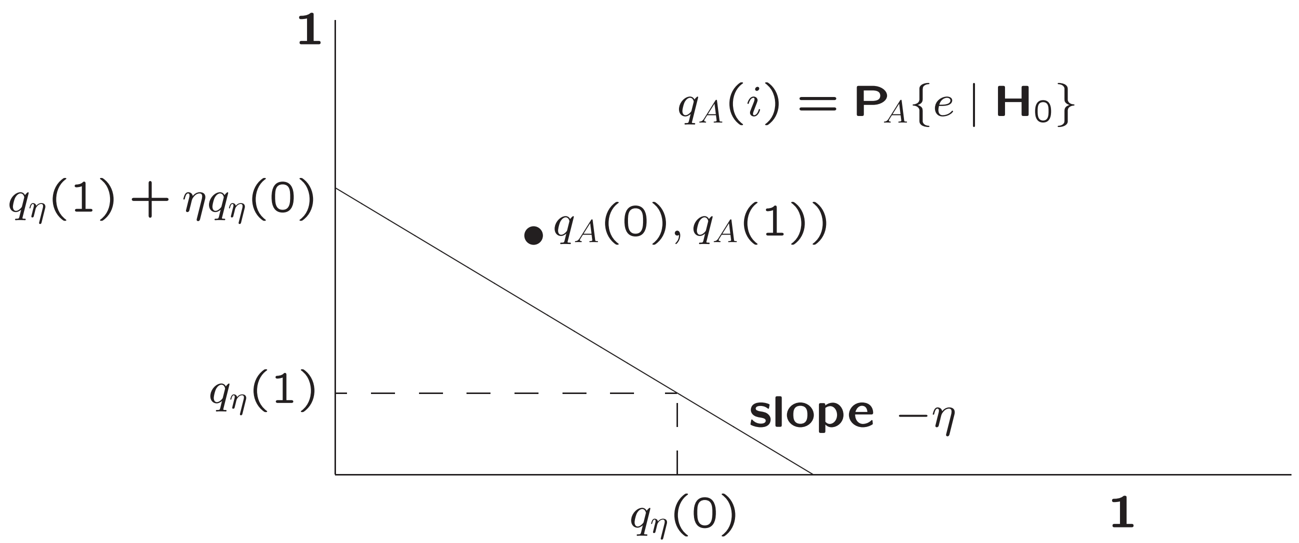
Use a threshold test with $\eta = r_0p_0/r_1p_1$.

Threshold tests are so fundamental that a 'sufficient statistic' is defined as any function of y from which $\Lambda(\mathbf{y})$ **can be calculated.**

Since sufficient statistics are usually one dimensional, this is important in terms of processing.

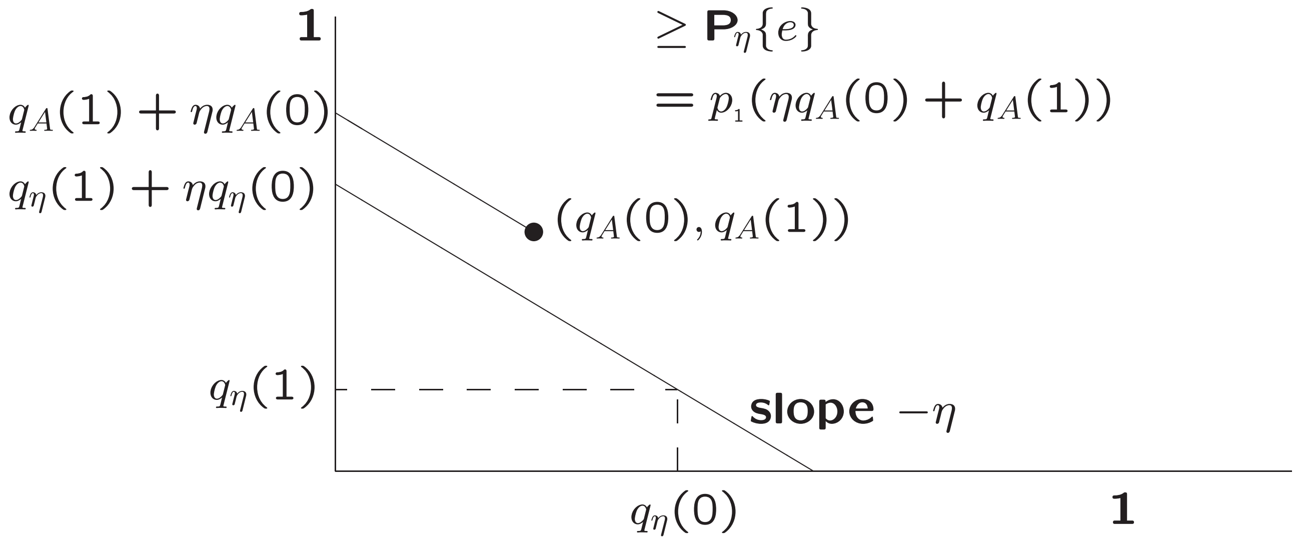
Denote $\mathsf{P}_A\{e \mid \mathsf{H}_i\}$ as $q_A(i)$, $i = 0, 1$. Every test A corresponds to two error probabilities $q_A(0)$ and $q_A(1)$. For a threshold test at η, replace A with η.

Thm: On a graph plotting $(q_A(1), q_A(0))$ for different A, all points are NE of a straight line of slope $-\eta$ through the point $(q_\eta(1), q_\eta(0))$.

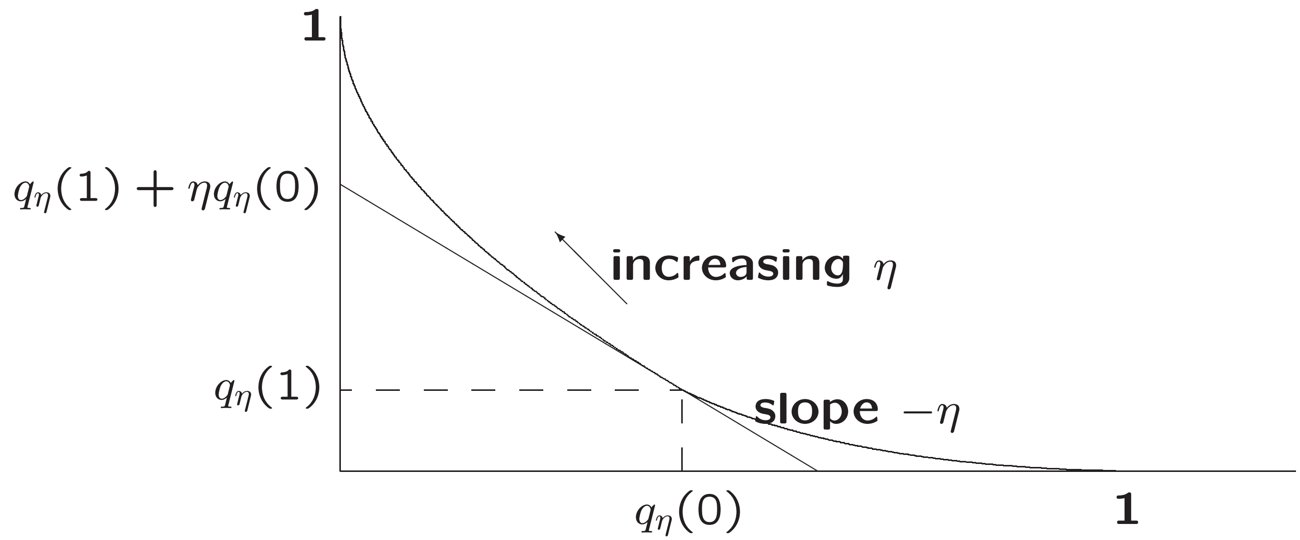


Proof:

$$
\begin{aligned}
\mathbf{P}_A\{e\} &= p_0 q_A(0) + p_1 q_A(1) \\
&= p_1(\eta q_A(0) + q_A(1)) \\
&\geq \mathbf{P}_\eta\{e\} \\
&= p_1(\eta q_A(0) + q_A(1))
\end{aligned}
$$



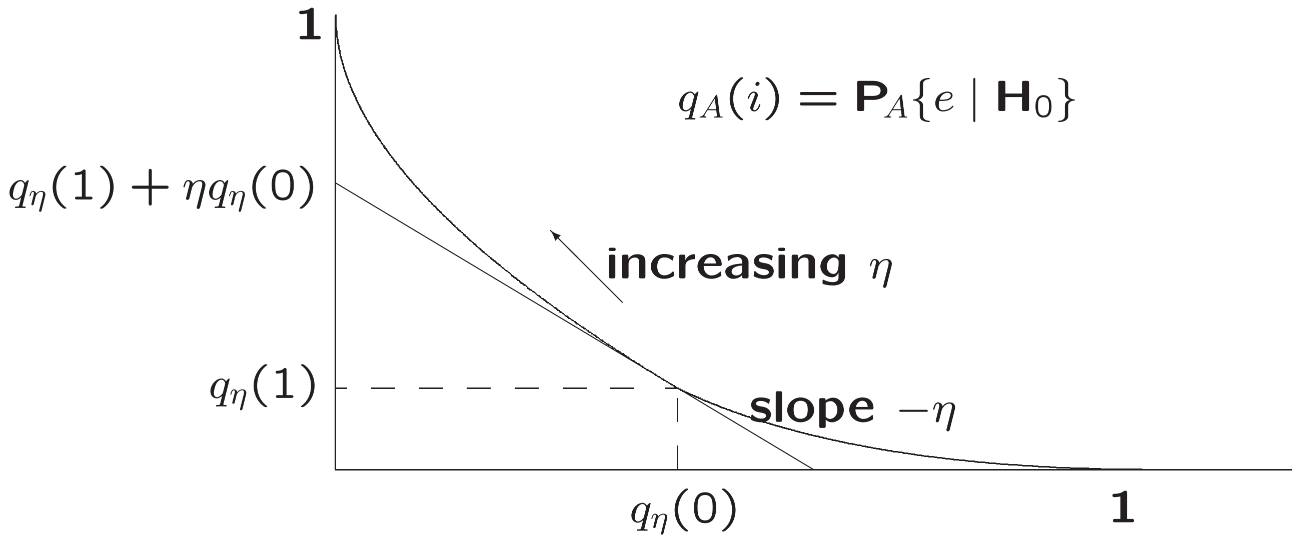
The theorem is valid for all η, $0 < \eta < \infty$. Thus for all tests A $(q_A(0), q_A(1))$ lies above the convex hull of the straight lines of slopes $-\eta$ through $q_\eta(1) + \eta q_\eta(0)$.



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

As η increases, $q_\eta(0)$ decreases and $q_\eta(1)$ increases.

The straight line for each η is tangent to the parametric function $(q_\eta(0), q_\eta(1))$ of η.

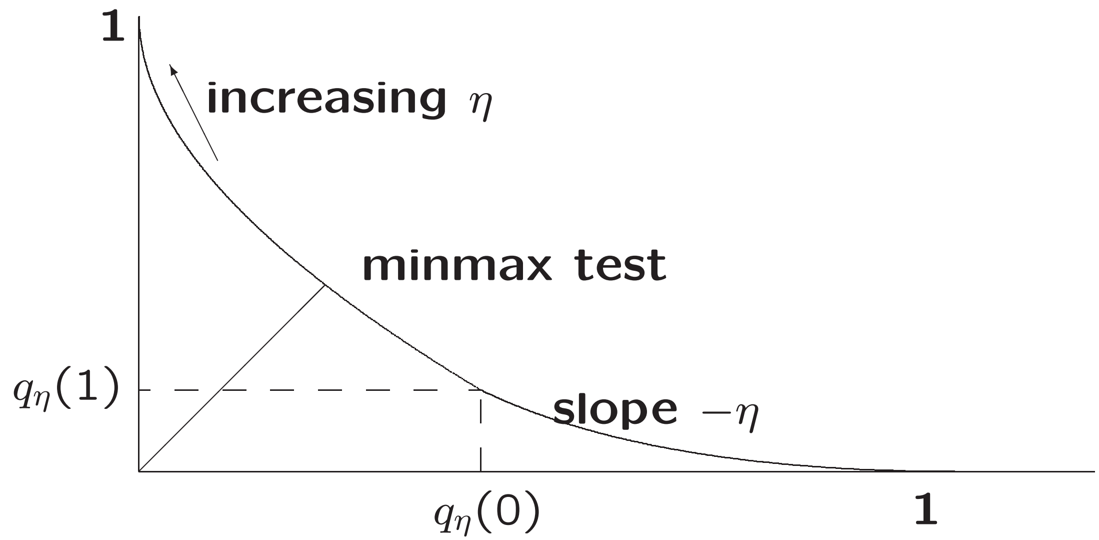


Since the tangents all lie below the curve, the curve is convex.

This curve is called the error curve for the detection problem.

The threshold tests $(q_A(0), q_A(1)$ **lie on the curve, all other tests** $(q_A(0), q_A(1)$ **lie above the curve.**

It is hard to imagine reasons for using non-threshold tests.



The Neyman Pearson test A minimizes $q_A(1)$ for some required value of $q_A(0)$.

Thus it chooses η to satisfy $q_A(1) = q_\eta(1)$ and uses the threshold test T_η.

The minmax test minimizes the larger of $q_A(1)$ and $q_A(0)$

Large deviations for hypothesis tests

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

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

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

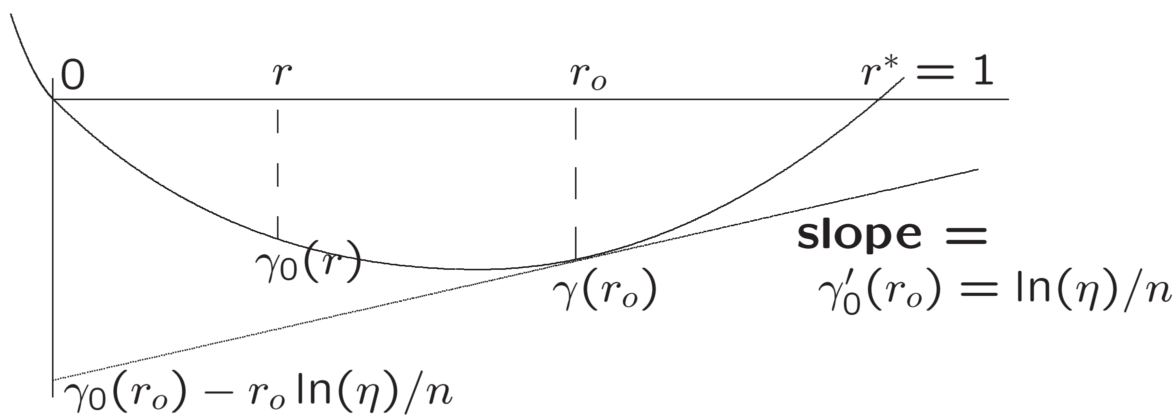
A threshold test compares $\sum_{i=1}^{n} z_i$ with $\ln(\eta) = \ln(p_0/p_1)$.

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 $\sum_i Z_i^0$

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

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

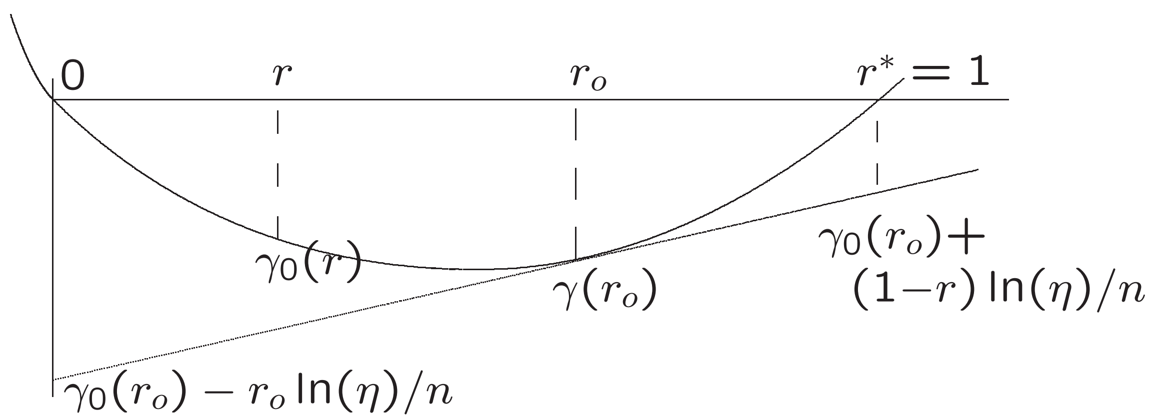


$$q_0(\eta) \le \exp n\left[\gamma_0(r_0) - r_0\ln(\eta)/n\right]$$

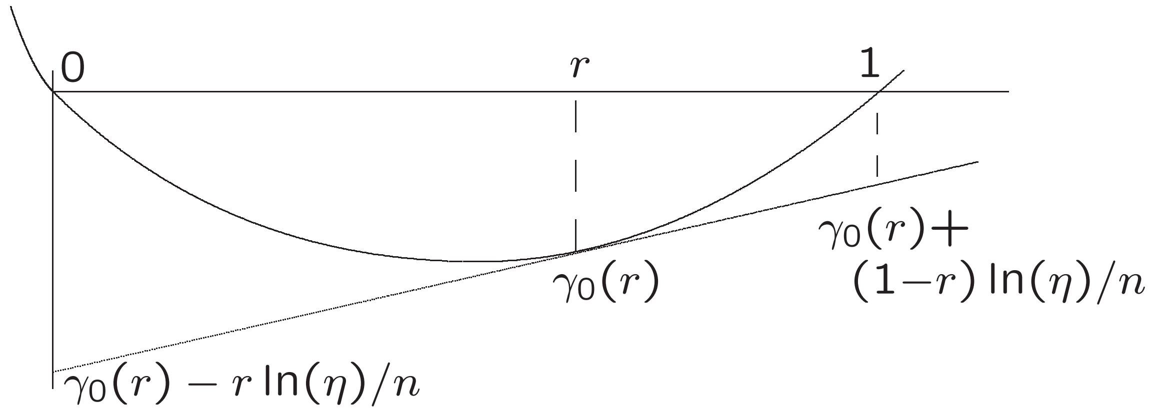
Exponential bound for $\sum_i Z_i^1$

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

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



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



These are the exponents for the two kinds of errors. This can be viewed as a large deviation form of Neyman Pearson. Choose one exponent and the other is given by the inverted see-saw above.

The a priori probabilities are usually not important here.

This large deviation hypothesis testing problem screams for a variable number of trials.

We have two coupled random walks, one based on H_0 and one on H_1.

We use two thresholds, $\alpha > 0$ and $\beta < 0$. Note that $\mathsf{E}[Z \mid \mathsf{H}_0] < 0$ and $\mathsf{E}[Z \mid \mathsf{H}_1] > 0$.

Thus crossing α is a rare event given the random walk with H_0 and crossing β is rare given H_1.

Since $r^* = 1$ for the H_0 walk, $\mathsf{P}\{e \mid \mathsf{H}_0\} \le e^{-\alpha}$.

This is not surprising since $\sum_i Z_i = \alpha$ means $\ln[\mathsf{P}\{e \mid \mathsf{H}_1\} / \mathsf{P}\{e \mid \mathsf{H}_0\} = \alpha$ for $p_0 = 1/2$.

Also, $\mathsf{P}\{e \mid \mathsf{H}_1\} \le e^{\beta}$.

The coupling between errors given H_1 and errors given H_0 is weaker here than for fixed n.

Increasing α lowers $\mathbf{P}\{e \mid \mathsf{H}_0\}$ exponentially and increases $\mathbf{E}[N \mid \mathsf{H}_1] \approx \alpha^{-1}\mathbf{E}[Z \mid \mathsf{H}_1]$.

Decreasing β lowers $\mathbf{P}\{e \mid \mathsf{H}_1\}$ exponentially and increases $\mathbf{E}[N \mid \mathsf{H}_0] \approx \beta^{-1}\mathbf{E}[Z \mid \mathsf{H}_0]$.

Kolmgorov martingale inequality

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

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

If we replace the max with Z_m, this is the lowly but useful Markov inequality.

Proof: Let N be the stopping time defined as the smallest $n \leq m$ such that $Z_n \geq a$.

N is minimum $n \leq m$ such that $Z_n \geq a$.

If $Z_n < a$ for all $n \leq m$, then $N = m$. Thus the process must stop by time m, and $Z_N \geq a$ iff $Z_n \geq a$ for some $n \leq m$. Thus

$$P\left(\max_{1 \leq n \leq m} Z_n \geq a\right) = \mathbf{P}\left\{Z_N \geq a\right\} \leq \frac{\mathbf{E}\left[Z_N\right]}{a}.$$

Since the process must be stopped by time m, we have $Z_N = Z_m^*$.

$\mathbf{E}\left[Z_m^*\right] \leq \mathbf{E}\left[Z_m\right]$, so the right hand side above is less than or equal to $\mathbf{E}\left[Z_m\right]/a$, completing the proof.