6.262: Discrete Stochastic Processes 5/2/11

L22: Random Walks and thresholds

Outline:

- Review of Chernoff bounds
- Wald's identity with 2 thresholds
- The Kingman bound for G/G/1
- Large deviations for hypothesis tests
- Sequential detection
- Tilted probabilities and proof of Wald's id.

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

$$\Pr\{Z \ge \alpha\} \le g_Z(r) \exp(-r\alpha) = \exp(\gamma_Z(r) - r\alpha)$$

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 $\gamma_{S_n}(r) = n\gamma_X(r)$.

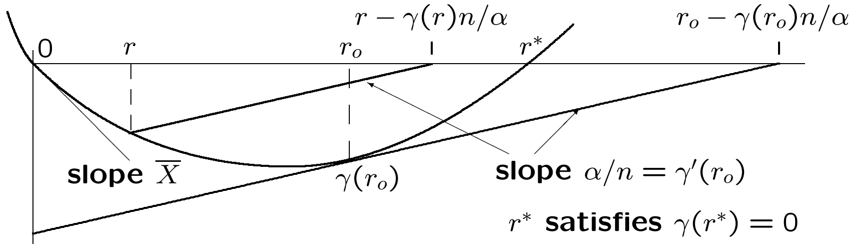
$$\Pr\{S_n \ge na\} \le \min_r \left(\exp[n(\gamma_X(r) - ra)]\right).$$

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

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

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

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



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.

Wald's identity with 2 thresholds

Consider a random walk $\{S_n;\, n \geq 1\}$ with $S_n = X_1 + \cdots + X_n$ and assume that X is not identically zero and has a semi-invariant MGF $\gamma(r)$ for $r \in (r_-, r_+)$ with $r_- < 0 < r_+$. Let $\alpha > 0$ and $\beta < 0$ be two thresholds. Let J be the smallest n for which either $S_n \geq \alpha$ or $S_n \leq \beta$.

Note that J is a stopping trial, i.e., $\mathbb{I}_{J=n}$ is a function of $S_1, \ldots, S_n$ and J is a rv. The fact that J is a rv is proved in Lemma 7.5.1, but is almost obvious.

Wald's identity now says that for any r, $r_- < r < r_+$,

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

If we replace J by a fixed step n, this just says that $\mathsf{E}\left[\exp(rS_n)\right] = \exp(n\gamma(r))$, so this is not totally implausible.

$$\mathsf{E}\left[\exp(rS_J - J\gamma(r))\right] = 1 \qquad \textbf{(Wald's identity)}.$$

Before justifying this, we use it to bound the probability of crossing a threshold.

Corollary: Assume further that $\overline{X} < 0$ **and that** $r^* > 0$ **exists such that** $\gamma(r^*) = 0$**. Then**

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

Wald's id. at r^* **is** $\mathsf{E}\left[\exp(r^*S_J)\right] = 1$**. Since** $\exp(r^*S_J) \geq 0$**,**

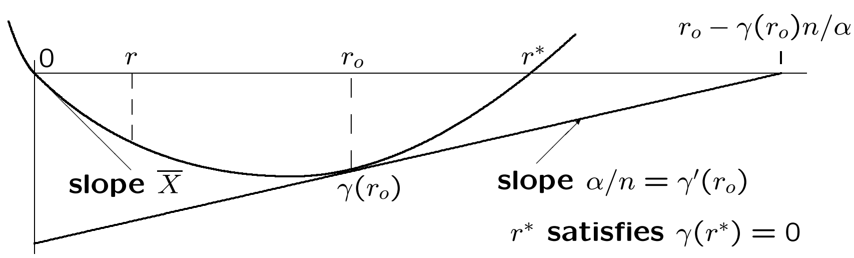
$$\Pr\{S_J \geq \alpha\}\,\mathsf{E}\left[\exp(r^*S_J) \mid S_J \geq \alpha\right] \leq \mathsf{E}\left[\exp(r^*S_J)\right] = 1.$$

For $S_J \geq \alpha$**, we have** $\exp(r^*S_J) \geq \exp(r^*\alpha)$**. Thus**

$$\Pr\{S_J \geq \alpha\}\exp(r^*\alpha) \leq 1.$$

This is valid for all choices of $\beta < 0$**, so it turns out to be valid without a lower threshold, i.e.,** $\Pr\{\bigcup_n\{S_n \geq \alpha\}\} \leq \exp(-r^*\alpha)$**.**

We saw before that $\Pr\{S_n \geq \alpha\} \leq \exp(-\alpha r^*)$ **for all** n**, but this corollary makes the stronger and cleaner statement that** $\Pr\Big\{\bigcup_{n\geq 1}\{S_n \geq \alpha\}\Big\} \leq \exp(-r^*\alpha)$



The Chernoff bound has the advantage of showing that the n **for which the probability of threshold crossing is essentially highest is** $n = \alpha/\gamma'(r^*)$**.**

The Kingman bound for G/G/1

The corollary can be applied to the queueing time W_i for the ith arrival to a G/G/1 system.

We let $U_i = X_i - Y_{i-1}$, i.e., U_i is the difference between the ith interarrival time and the previous service time.

Recall that we showed that $\{U_i;\, i \geq 1\}$ is a modification of a random walk. The text shows that it is a random walk looking backward.

Letting $\gamma(r)$ be the semi-invariant MGF of each U_i, then the Kingman bound (the corollary to the Wald idenity for the G/G/1 queue) says that for all $n \geq 1$,

$$\Pr\{W_n \geq \alpha\} \leq \Pr\{W \geq \alpha\} \leq \exp(-r^*\alpha)\,; \qquad \text{for all } \alpha > 0.$$

Large deviations for hypothesis tests

Let $\vec{Y} = (Y_1, \dots, Y_n)$ be IID conditional on H_0 and also IID condtional on H_1. Then

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

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

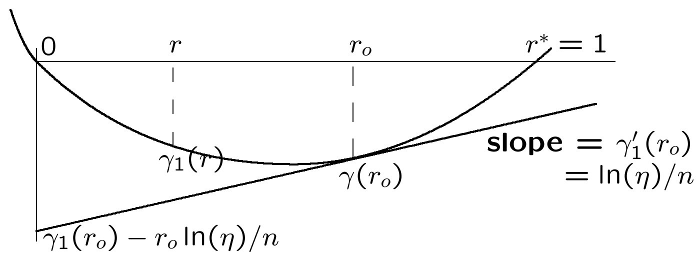
A threshold test compares $\sum_{i=1}^{n} z_i$ with $\ln(\eta) = \ln(p_1/p_0)$.

Conditional on H_1, make error if $\sum_i Z_i^1 > \ln(\eta)$ where Z_i^1, $1 \leq i \leq n$, are IID conditional on H_1.

Exponential bound for $\sum_i Z_i^1$

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

At $r = 1$, this is $\ln(\int f(y\mid\mathbf{H}_0)\,dy) = 0$.



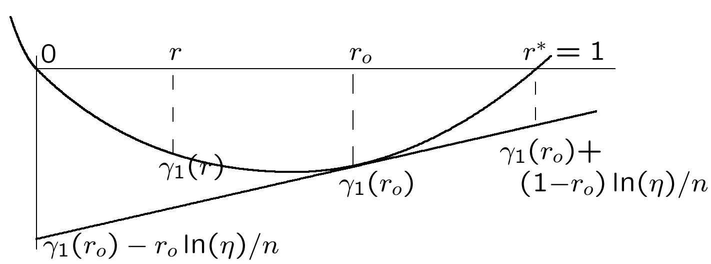
$$q_1(\eta) \le \exp n\,[\gamma_1(r_0) - r_0\ln(\eta)/n]$$

where $q_\ell(\eta) = \Pr\{e \mid H = \ell\}$

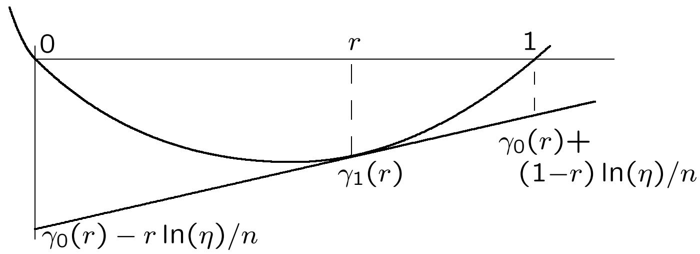
Exponential bound for $\sum_i Z_i^0$

$$\gamma_0(s) = \ln\left\{\int f(y \mid 0)\exp\left[s\ln\frac{f(y\mid\mathbf{H}_0)}{f(y\mid\mathbf{H}_1)}\right]\,dy\right\}$$
$$= \ln\left\{\int f^{-s}(y\mid\mathbf{H}_1)f^{1+s}(y\mid\mathbf{H}_0)\,dy\right\}$$

At $s = -1$, this is $\ln(\int f(y\mid\mathbf{H}_1)\,dy) = 0$. **Note:** $\gamma_0(s) = \gamma_1(r{-}1)$.



$$q_0(\eta) \le \exp n\,[\gamma_1(r_o) + (1{-}r_o)\ln(\eta)/n]$$



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 the essential characteristic here, but the bound for MAP is obtimized at r such that $\ln(\eta)/n - \gamma_0'(r)$

Sequential detection

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

We have two coupled random walks, one based on $\mathbf{H}_0$ and one on $\mathbf{H}_1$.

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

Thus crossing α is a rare event given the random walk with $\mathbf{H}_0$ and crossing β is rare given $\mathbf{H}_1$.

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

This is not surprising; for the simple RW with $p_1 = 1/2$, $\sum_i Z_i = \alpha$ means that

$$\ln[\Pr\{e \mid \mathbf{H}_1\} / \Pr\{e \mid \mathbf{H}_0\} = \alpha$$

Also, $\Pr\{e \mid \mathbf{H}_1\} \le e^{\beta}$.

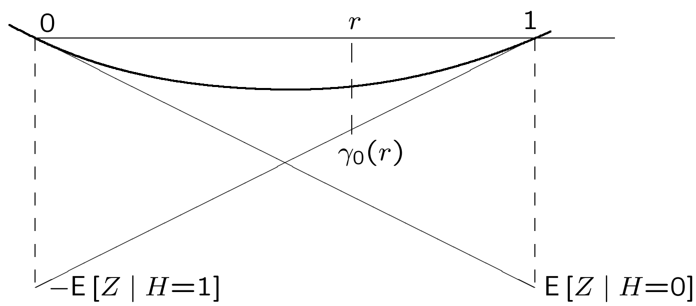
The coupling between errors given $\mathbf{H}_1$ and errors given $\mathbf{H}_0$ is weaker here than for fixed n.

Increasing α lowers $\Pr\{e \mid \mathbf{H}_0\}$ exponentially and increases $\mathsf{E}[J \mid \mathbf{H}_1] \approx \alpha/\mathsf{E}[Z \mid \mathbf{H}_1]$ (from Wald's equality since $\alpha \approx \mathsf{E}[S_J \mid H = 1]$). Thus

$$\Pr\{e \mid H{=}0\} \sim \exp(-\mathsf{E}[J \mid H{=}1]\,\mathsf{E}[Z \mid H{=}1])$$

In other words, $\Pr\{e \mid H{=}0\}$ is essentially exponential in the expected number of trials given $H{=}1$. The exponent is $\mathsf{E}[Z \mid H{=}1]$, illustrated below.

Similarly, $\Pr\{e \mid H{=}1\} \sim \exp(\mathsf{E}[J \mid H{=}0]\,\mathsf{E}[Z \mid H{=}0])$.



Tilted probabilities

Let $\{X_n;\, n \geq 1\}$ be a sequence of IID discete rv's with a MGF at some given r. Given the PMF of X, define a tilted PMF (for X) as

$$\mathsf{q}_{X,r}(x) = \mathsf{p}_X(x)\exp[rx - \gamma(r)].$$

Summing over x, $\sum q_{X,r}(x) = g_X(r)e^{-\gamma_X(r)} = 1$. We view $\mathsf{q}_{X,r}(x)$ as the PMF on X in a new probability space with this given relationship to the old space.

We can then use all the laws of probability in this new measure. In this new measure, $\{X_n;\, n \geq 1\}$ are taken to be IID. The mean of X in this new space is

$$\begin{aligned}
\mathsf{E}_r[X] &= \sum_x x\mathsf{q}_{X,r}(x) = \sum_x x\mathsf{p}_X(x)\exp[rx - \gamma(r)] \\
&= \frac{1}{g_X(r)}\sum_x \frac{d}{dr}\mathsf{p}_X(x)\exp[rx] \\
&= \frac{g'_X(r)}{g_X(r)} = \gamma'(r).
\end{aligned}$$

The joint tilted PMF for $\vec{X}^n = (X_1, \dots, X_n)$ is then

$$\mathsf{q}_{\vec{X}^n,r}(x_1, \dots, x_n) = \mathsf{p}_{\vec{X}^n}(x_1, \dots, x_n) \exp(\sum_{i=1}^{n} [rx_i - \gamma(r)].$$

Let $A(s_n)$ be the set of n-tuples such that $x_1 + \cdots x_n = s_n$. Then (in the original space) $\mathsf{p}_{S_n}(s_n) = \Pr\{S_n = s_n\} = \Pr\{A(s_n)\}$. Also, for each $\vec{x}^n \in A(s_n)$,

$$\begin{aligned} \mathsf{q}_{\vec{X}^n,r}(x_1, \dots, x_n) &= \mathsf{p}_{\vec{X}^n}(x_1, \dots, x_n) \exp(rs_n - n\gamma(r)] \\ \mathsf{q}_{S_n,r}(s_n) &= \mathsf{p}_{S_n}(s_n) \exp[rs_n - n\gamma(r)], \end{aligned}$$

where we have summed over $A(s_n)$. This is the key to much of large deviation theory. For $r > 0$, it tilts the probability measure on S_n toward large values, and the laws of large numbers can be used on this tilted measure.

Proof of Wald's identity

The stopping time J for the 2 threshold RW is a rv (from Lemma 7.5.1) and it is also a rv for the tilted probability measure. Let $\mathcal{T}_n = \{\vec{x}_n : s_n \notin (\beta, \alpha); s_i \in (\beta, \alpha); 1 \le i < n\}$.

That is, $\mathcal{T}_n$ is the set of n tuples for which stopping occurs on trial n. Letting $\mathsf{q}_{J,r(n)}$ be the PMF of J in the tilted probability measure,

$$\begin{aligned} \mathsf{q}_{J,r}(n) &= \sum_{\vec{x}^n \in \mathcal{T}_n} \mathsf{q}_{\vec{X}^n,r}(\vec{x}^n) = \sum_{\vec{x}^n \in \mathcal{T}_n} \mathsf{p}_{\vec{X}^n}(\vec{x}^n) \exp[rs_n - n\gamma(r)] \\ &= \mathsf{E}\left[\exp[rS_n - n\gamma(r) \mid J{=}n\right] \; \Pr\{J = n\}\,. \end{aligned}$$

Summing over n completes the proof.

6.262 Discrete Stochastic Processes
Spring 2011