DISCRETE STOCHASTIC PROCESSES

Lecture 21

Chapter 7 - Random Walks

Review: Optimization of the Chernov bound

Wald's Identity

Examples and Applications

Diffusion Model

One Threshold Bound

Kingman Bound for G/G/1 Queue

Application to Hypothesis Testing

Partial Proof

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 the region of convergence is large enough for everything we want to do.

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

Recall **Chernov bound** (Markov inequality applied to $\exp(rS_n)$).

$$P(S_n \ge \alpha) = P\{\exp(rS_n) \ge \exp(r\alpha)\} \le \frac{E[e^{rS_n}]}{e^{r\alpha}} = g(r)^n \exp(-r\alpha);\ r \ge 0$$

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

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

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

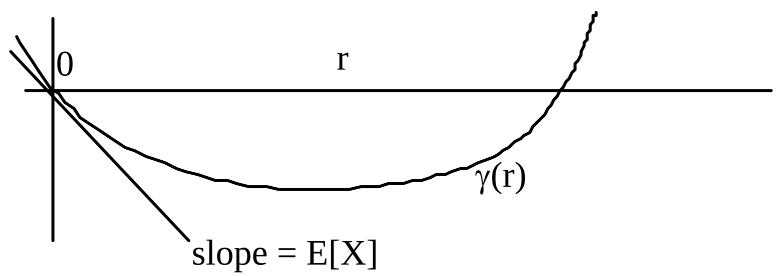
Then $$\boxed{P(S_n \ge \alpha) \le \exp[n\gamma(r) - r\alpha];\ \ r \ge 0}$$

with derivative wrt r of zero at r', where $\gamma'(r') = \alpha/n$.

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

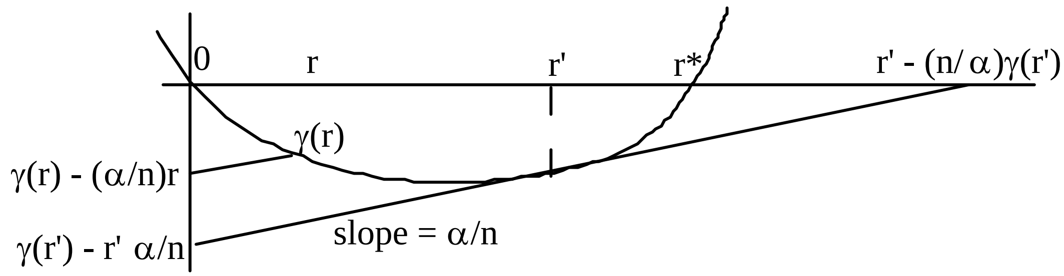
Also $\gamma''(r) = \left[g''(r)g(r) - g'(r)^2\right]/g(r)^2$; $\gamma''(0) = \text{VAR}(X)$.

Also $\gamma''(r) \geq 0$ ($\gamma(r)$ is convex). (Problem 7.8 in Gallager shows this.)



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

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



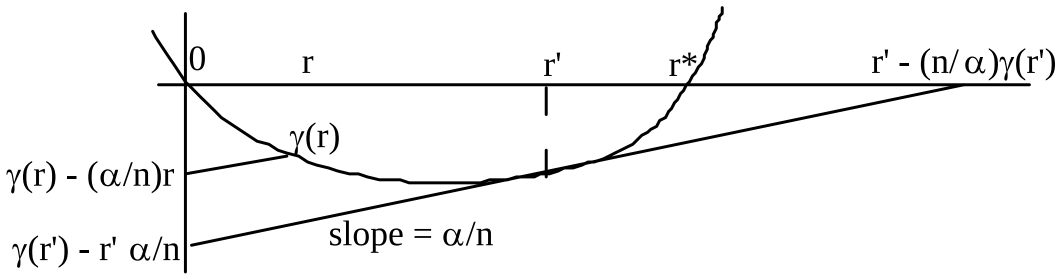
$$\gamma'(r') = \alpha/n$$

$$P(S_n \geq \alpha) \leq \exp\{n[\gamma(r') - r'\alpha/n]\} \leq \exp\{n[\gamma(r) - r\alpha/n]\}, r \geq 0. \qquad (1)$$

Eq. (1) shows how $P(S_n \geq \alpha)$ decays exponentially with n for fixed $\alpha/n = \gamma'(r')$.

$$P(S_n \geq \alpha) \leq \exp\left\{\alpha\left[\frac{\gamma(r')}{\gamma'(r')} - r'\right]\right\} = \exp\left\{-\alpha\left[r' - \frac{n}{\alpha}\gamma(r')\right]\right\} \qquad (2)$$

Eq. (2) shows how $P(S_n \geq \alpha)$ decays exponentially with α for fixed $\alpha/n = \gamma'(r')$. If n increases while α stays fixed, the slope of the tangent decreases and the right intercept increases, increasing the decay rate with α.



$$\gamma'(r') = \alpha/n \qquad (1)$$

$$P(S_n \geq \alpha) \leq \exp\left\{\alpha\left[\frac{\gamma(r')}{\gamma'(r')} - r'\right]\right\} = \exp\left\{-\alpha\left[r' - \frac{n}{\alpha}\gamma(r')\right]\right\} \qquad (2)$$

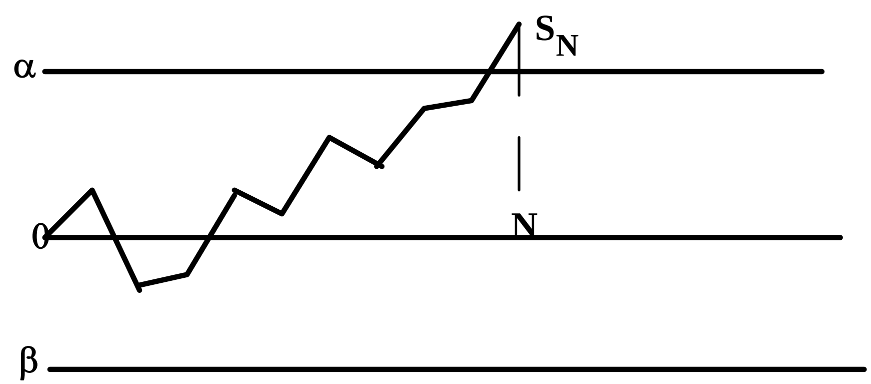
As n decreases, the slope increases and the right intercept decreases until it equals r^*, and as it decreases further, the right intercept increases again. Thus the minimum value of the right intercept as n varies is r^*. Therefore, for $E(X) < 0$,

$$P(S_n \geq \alpha) \leq e^{-r^*\alpha}, \text{ arbitrary } \alpha \geq 0, \ n \geq 1.$$

Part of our interest in the optimized Chernoff bound comes because it is tight in the exponential sense that, for α / n fixed, as $\alpha \to \infty$, $\ln(P(S_n \geq \alpha)) / \ln(\text{bound}) \to 1$

The stronger result that $P(\max_n S_n \geq \alpha) \leq e^{-r^*\alpha}$ will be demonstrated using Wald's Identity

Assume $\alpha > 0, \beta < 0$. Let N be the first time at which the RW reaches or crosses either α or β.



Assuming that X is not identically 0, the RW must cross a threshold eventually WP 1. (If X takes on discrete values, for example, there is some finite string $X_1, \ldots, X_n$ that causes a threshold crossing from any point in the interval with non-zero probability, and this string happens eventually WP1).

Thus N is a rv, and S_N, the value of the walk when the threshold is crossed, is also a rv.

Wald's Identity

Let Xi be iid and $\gamma(r)=\ln\{E[e^{rX}]\}$. Assume $\gamma(r)$ converges in an open interval around 0. Let $\alpha>0$ and $\beta<0$ be two arbitrary numbers and let N be the smallest n for which either $S_n \geq \alpha$ or $S_n \leq \beta$. For any r where $\gamma(r)$ converges

$$E\left\{\exp\left[rS_N - N\gamma(r)\right]\right\} = 1$$

To begin deciphering the algebra, note that

$$E\left\{\exp\left[rS_n - n\gamma(r)\right]\right\} = \frac{E\left\{\exp rS_n\right\}}{\exp[n\gamma(r)]} = \frac{E\left\{e^{rX_1}\cdots e^{rX_n}\right\}}{e^{n\ln E[rX]}} = \frac{(E[e^{rX}])^n}{(E[e^{rX}])^n} = 1$$

A heuristic (but incorrect) way to make this result plausible is to claim that

$$E\left\{\exp\left[rS_N - N\gamma(r)\right] \mid N=n\right\} \overset{??}{=} E\left\{\exp\left[rS_n - n\gamma(r)\right]\right\} = 1$$

(Why is the first equality suspect?)

Some examples of use of Wald's identity

Taking the first derivative of Wald's identity yields

$$\frac{d}{dr} E\left\{\exp\left[rS_N - N\gamma(r)\right]\right\} =$$

$$E\left\{\left[S_N - N\gamma'(r)\right]\exp\left[rS_N - N\gamma(r)\right]\right\} = 0$$

Evaluating at $r = 0$,

$$E\left[S_N - N\gamma'(0)\right] = 0 \text{ or } \boxed{E\left[S_N\right] = E[N]E[X]} \quad \textbf{(Wald's equality)}$$

If $E[X] = 0$, this says that $E\left[S_N\right] = 0$, but does not help in finding $E[N]$.
Taking the second derivative at $r = 0$,

$$\frac{d}{dr}E\left\{\left[S_N - N\gamma'(r)\right]\exp\left[rS_N - N\gamma(r)\right]\right\}\Big|_{r=0} = E\left\{\left[S_N - N\gamma'(0)\right]^2 - N\gamma''(0)\right\} = 0$$

For $E[X] = 0$, $\gamma'(0) = 0$, $\gamma''(0) = E[X^2] = \text{var}(X)$, and this becomes

$$\boxed{\text{For } E[X] = 0, \quad \text{VAR}\left[S_N\right] = E\left[(S_N)^2\right] = E[N]\ \text{VAR}[X]}$$

This will allow us a very simple way to derive the result that diffusion to a boundary takes, on average, time proportional to the square of the distance. You derived this by more laborious techniques using first passage times in Problem J of Problem Set #8.

In a simple RW ($X = 0, \ +1$ or -1) with $E[X] = 0$ and α, β integers, we must have $S_N = \alpha$ or $S_N = \beta$.

Let $q_\alpha = P(S_N = \alpha)$. Since there is no overshoot, $\alpha q_\alpha + \beta(1 - q_\alpha) = 0$.

From the Wald equality for E[X] = 0, $\mathrm{VAR}[S_N] = E\left[(S_N)^2\right] = E[N]\,\mathrm{VAR}[X]$, we have

$$\alpha^2 q_\alpha + \beta^2 (1 - q_\alpha) = E[N]\mathrm{VAR}(X)$$

Note that $\alpha^2 q_\alpha = -\alpha\beta(1 - q_\alpha)$ and $\beta^2 (1 - q_\alpha) = -\alpha\beta q_\alpha$, so

$$E[N] = -\beta\alpha \,/\, \mathrm{VAR}(X)$$

This says that when $E[X] = 0$, and when $\beta = -\alpha$, the expected time to cross the threshold is quadratic in α, and inversely proportional to the variance.

This is not surprising since the RW has no drift, and relies on the variance of S_n to cross the threshold.

One Threshold Bound

When $E[X] < 0$ and we are interested in crossing a threshold at α, let β be negative and choose r in Wald's identity as the positive root r^* of $\gamma(r)$.

Then $\gamma(r^*) = 0$ and $E\{\exp[r^* S_N - N\gamma(r^*)]\} = E\left[e^{r^* S_N}\right] = 1$.

Writing this out by separating the case $S_N \geq \alpha$ from $S_N \leq \beta$, and letting $q_\alpha = P(S_N \geq \alpha)$,

$$1 = E\left[e^{r^* S_N} \mid S_N \geq \alpha\right] q_\alpha + E\left[e^{r^* S_N} \mid S_N \leq \beta\right](1 - q_\alpha)$$

Approximating bound: $1 \geq E\left[e^{r^* S_N} \mid S_N \geq \alpha\right] q_\alpha \geq \exp(r^* \alpha) q_\alpha$ so

$$P(S_N \geq \alpha) = q_\alpha \leq \exp(-r^* \alpha)$$

This is true for all $\beta < 0$, and as

$\beta \to -\infty$, it implies that $P\left\{\left(\max_n S_n\right) \geq \alpha\right\} \leq e^{-r^* \alpha}$.

$$1 = E\left[e^{r^*S_N} \mid S_N \geq \alpha\right]q_\alpha + E\left[e^{r^*S_N} \mid S_N \leq \beta\right](1 - q_\alpha)$$

In the limit as $\beta \to -\infty, \quad 1 = q_\alpha\, E\left[e^{r^*S_N} \mid S_N \geq a\right]$

$$q_\alpha = e^{-r^*\alpha}\left\{E\left[e^{r^*(S_N - \alpha)} \mid S_N \geq \alpha\right]\right\}^{-1}$$

If X is non-arithmetic, the final term above approaches a constant as $\alpha \to \infty$. If $\gamma(r)$ is a rational function, this term can be evaluated by solving a Wiener Hopf equation.

The quantity $P(S_N \geq \alpha) = q_\alpha$ is different from the previous $P(S_n \geq \alpha)$ for each n.

$$P(S_N \geq \alpha) = P\left(\left(\sup_n S_n\right) \geq \alpha\right)$$

Remember, all this was derived for the negatively drifting walk. The quantity q_α is the probability that this walk will exit through the positive threshold at all. The quantity $1 - q_\alpha$ is the probability that the walk will exit through the negative threshold when β is finite and the probability that the walk will never exit when β is minus infinity.

Kingman bound

Let Xi and Yi be the interarrival and service times of a G/G/1 queue. Let Ui = Yi-1 – Xi. For the G/G/1 queue, let W be the steady- state waiting time in queue. If $\gamma(r)$ has a root at $r^* > 0$, then

$$P(W \geq \alpha) \leq \exp(-r^* \alpha).$$

Hypothesis Testing

Let Y_i be IID conditional on H_0 or H_1. Assume that *a priori*, H_0 and H_1 are equally likely.

$$\frac{P(H_1 \mid \vec{Y})}{P(H_0 \mid \vec{Y})} = \frac{P(H_1)}{P(H_0)} \frac{\prod_{i=1}^{n} f(Y_i \mid H_1)}{\prod_{i=1}^{n} f(Y_i \mid H_0)}$$

$$S_n = \ln \frac{P(H_1 | \vec{Y})}{P(H_0 | \vec{Y})} = \sum_{i=1}^{n} Z_i; \quad \text{where} \quad Z_i = \ln \frac{f(Y_i | H_1)}{f(Y_i | H_0)}$$

Condition on H_0 in what follows.

Let $S_n = Z_1 + \ldots + Z_n$ (under H_0). Use sequential test: stop when S_n crosses either $\alpha > 0$ or $\beta < 0$. If $S_N \geq \alpha$, declare H_1 and if $S_N \leq \beta$, declare H_0. q_a is then the probability of error (given H_0).

$$g(r) = E\left(e^{rZ}\right) = \int f(y|H_0)\exp\left\{r\ln\frac{f(y|H_1)}{f(y|H_0)}\right\}dy$$

$$= \int\left[f(y|H_0)\right]^{1-r}\left[f(y|H_1)\right]^{r}dy$$

Note that $g(1) = 1$ so $r^* = 1$.

Therefore, for hypothesis testing, $q_\alpha = P(\text{conclude } 'H_1' \mid H_0) = P(S_N \geq \alpha) \leq e^{-\alpha}$.

This is not surprising, since, ignoring the overshoot, H_1 is declared when

$$\ln\left[P(H_1|\vec{y}) \,/\, P(H_0|\vec{y})\right] = \alpha \quad \text{i.e.,} \quad P(H_0|\vec{y}) = e^{-\alpha}P(H_1|\vec{y})$$

Assume the problem is symmetrical so that the natural choice, $\beta = -\alpha$, is made.

The expected number of trials, $E[N|H_0]$, is estimated using Wald's equality $\overline{S_N} = \bar{N}\bar{X}$:

$$E[S_N \mid H_0] \approx \alpha q_\alpha - \alpha(1 - q_\alpha) = -\alpha(1 - 2q_\alpha) \approx -\alpha$$

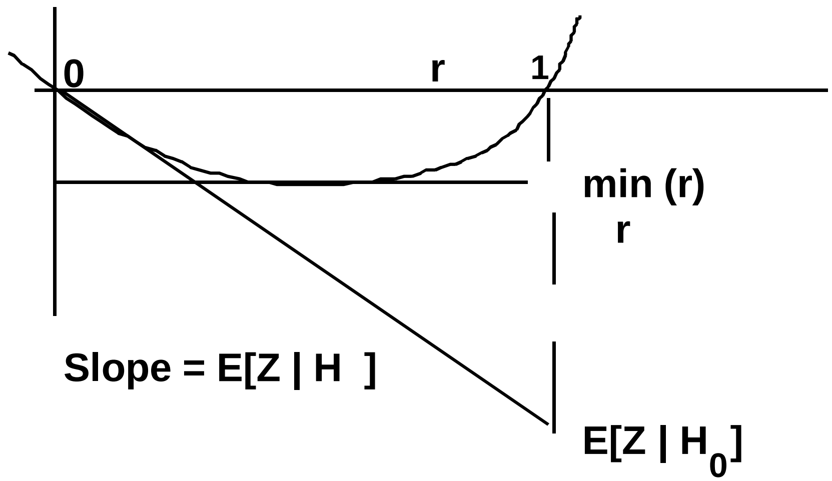
$$E[N|H_0] = E[S_N|H_0] / E[Z|H_0] \approx -\alpha / E[Z|H_0]$$

The conventional (non-sequential test) is to take n trials and then choose H_1 if $S_n > 0$.
What n to choose to keep P(error) $|H_0) = e^{-\alpha}$?

By the Chernoff bound,

$$P(S_n \ge \alpha) \le \exp\{\gamma(r)n - \alpha r\} \Rightarrow P(e|H_0) = P(S_n \ge 0) \le \exp[n \min_r \gamma(r)].$$

To make $P(e|H_0) \approx e^{-\alpha}$, $n \approx -\alpha / [\min_r \gamma(r)]$.



Note that the number of trials for the sequential test is several times smaller than for the fixed test (by a ratio of $|E[Z|H_0]| / |\min_r \gamma(r)|$).

Semi-Proof of Wald's identity, $E\{\exp[rS_N - N\gamma(r)]\} = 1$

First, to get used to the algebra, let N just be some constant n. (This is a simple but valid special case of a stopping rule.)

$$E\left\{e^{rS_n - n\gamma(r)}\right\} = E\left\{\frac{e^{r\sum_{k=1}^{n} X_k}}{e^{n\gamma(r)}}\right\} = \frac{\left[E\left\{e^{rX}\right\}\right]^n}{g^n(r)} = 1.$$

Now let N be the number of terms until the first threshold crossing.

For each integer $n > 0$, define $I_n = 1$ if $N \geq n$, $I_n = 0$ if $N < n$. Then $N = I_1 + I_2 + \ldots$ and $S_N = X_1 I_1 + X_2 I_2 + \ldots$.

I_n is determined by $X_1, \ldots, X_{n-1}$, so N is a stopping rule for $\{X_i;\, i \geq 1\}$ and I_n is independent of $X_n, X_{n+1}, \ldots$.

$$E\left\{\exp\left[rS_N - N\gamma(r)\right]\right\} = E\left\{\exp\left[\sum_n \left[rX_n I_n - I_n\gamma(r)\right]\right]\right\} = E\left\{\prod_n \exp\left[\left[rX_n - \gamma(r)\right]I_n\right]\right\}$$

$$= E_{I_1,X_1}\left\{\exp\left[\left[rX_1 - \gamma(r)\right]I_1\right]E_{2|1}\left\{\exp\left[\left[rX_2 - \gamma(r)\right]I_2 \middle| I_1, X_1\right]\right.\right.$$

$$\left.\left.E_{3|2,1}\left\{\exp\left[\left[rX_3 - \gamma(r)\right]I_3 \middle| I_1, X_1, I_2, X_2\right]\right\}\right\}\cdots\right\}$$

Note that $E_{X_m}\left\{\exp\left[\left[rX_m - \gamma(r)\right]I_m\right] \mid X_1, I_1, \dots, X_{m-1}, I_{m-1}, I_m = 0\right\} = E\left\{e^0\right\} = 1$ and

$E_{X_m}\left\{\exp\left[\left[rX_m - \gamma(r)\right]I_m\right] \mid X_1, I_1, \dots, X_{m-1}, I_{m-1}, I_m = 1\right\} = E\left\{\exp\left[rX_m - \ln\left(g(r)\right)\right]\right\} = 1$,
because X_m is independent of all other X_k's and of $I_1, I_2, \dots, I_m$.

Thus $E_{I_m,X_m}\left\{\exp\left[\left[rX_m - \gamma(r)\right]I_m\right] \mid X_1, I_1, \dots, X_{m-1}, I_{m-1}\right\} = 1$.

Thus all terms in the product are 1.

This argument is incomplete in that it does not justify passing to the limit as the number of terms goes to ∞. That is why the proof in the text is so long. We see later why the interchange of expection and limit needs to be done carefully here.