6.262: Discrete Stochastic Processes 4/27/11

L21: Hypothesis testing and Random Walks

Outline:

- Random walks
- Detection, decisions, & Hypothesis testing
- Threshold tests and the error curve
- Thresholds for random walks and Chernoff

Random walks

Def: Let $\{X_i;\ i \geq 1\}$ be a sequence of IID rv's, and let $S_n = X_1 + X_2 + \cdots + X_n$ for $n \geq 1$. The integer-time stochastic process $\{S_n;\ n \geq 1\}$ is called a random walk, or, specifically, the random walk based on $\{X_i;\ i \geq 1\}$.

Our focus will be on threshold-crossing problems. For example, if X is binary with $\mathsf{p}_X(1) = 1,\ \mathsf{p}_X(-1) = q = 1-p$, then

$$\Pr\left\{ \bigcup_{n=1}^{\infty} \{S_n \geq k\} \right\} = \left(\frac{p}{1-p} \right)^k \qquad \text{if } p \leq 1/2.$$

Detection, decisions, & Hypothesis testing

The model here contains a discrete, usually binary, rv H called the hypothesis rv. The sample values of H, say 0 and 1, are called the alternative hypotheses and have marginal probabilities, called a priori probabilities $p_0 = \Pr\{H = 0\}$ and $p_1 = \Pr\{H = 1\}$.

Among arbitrarily many other rv's, there is a sequence $\vec{Y}^{(m)} = (Y_1, Y_2, \dots, Y_m)$ of rv's called the observation. We usually assume that $Y_1, Y_2, \dots$, are IID conditional on $H = 0$ and IID conditional on $H = 1$. Thus, if the Y_n are continuous,

$$f_{\vec{Y}^{(m)}|H}(\vec{y} \mid \ell) = \prod_{n=1}^{m} f_{Y|H}(y_n \mid \ell).$$

Assume that, on the basis of observing a sample value $\vec{y}$ of $\vec{Y}$, we must make a decision about H, i.e., choose $H = 0$ or $H = 1$, i.e., detect whether or not H is 1.

Decisions in probability theory, as in real life, are not necessarily correct, so we need a criterion for making a choice.

We might maximize the probability of choosing correctly, for example, or, given a cost for the wrong choice, might minimize the expected cost.

Note that the probability experiment here includes not only the experiment of gathering data (i.e., measuring the sample value $\vec{y}$ of $\vec{Y}$) but also the sample value of the hypothesis.

From Bayes', recognizing that $\mathsf{f}(\vec{y}) = p_0\mathsf{f}(\vec{y}|0) + p_1\mathsf{f}(\vec{y}|1)$

$$\Pr\{H{=}\ell \mid \vec{y}\} = \frac{p_\ell f_{\vec{Y}|H}(\vec{y} \mid \ell)}{p_0 f_{\vec{Y}|H}(\vec{y} \mid 0) + p_1 f_{\vec{Y}|H}(\vec{y} \mid 1)}.$$

Comparing $\Pr\{H{=}0 \mid \vec{y}\}$ and $\Pr\{H{=}1 \mid \vec{y}\}$,

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

The probability that $H = \ell$ is the correct hypothesis, given the observation, is $\Pr\left\{H{=}\ell \mid \vec{Y}\right\}$. Thus we maximize the a posteriori probability of choosing correctly by choosing the maximum over ℓ of $\Pr\left\{H{=}\ell \mid \vec{Y}\right\}$.

This is called the MAP rule (maximum a posteriori probability). It requires knowing p_0 and p_1.

The MAP rule (and other decision rules) are clearer if we define the likelihood ratio,

$$\Lambda(\vec{y}) = \frac{f_{\vec{Y}|H}(\vec{y} \mid 0)}{f_{\vec{Y}|H}(\vec{y} \mid 1)}.$$

The MAP rule is then

$$\Lambda(\vec{y}) \begin{cases} > p_1/p_0 & ; \qquad \text{select } \hat{h}{=}0 \\ \leq p_1/p_0 & ; \qquad \text{select } \hat{h}{=}1. \end{cases}$$

Many decision rules, including the most common and the most sensible, are rules that compare $\Lambda(\vec{y})$ to a fixed threshold, say η, independent of $\vec{y}$. Such decision rules vary only in the way that η is chosen.

Example: For maximum likelihood, the threshold is 1 (this is MAP for $p_0 = p_1$, but it is also used in other ways).

Back to random walks: Note that the logarithm of the threshold ratio is given by

$$\ln \Lambda(\vec{y}^{(m)}) = \sum_{n=1}^{m} \Lambda(y_n); \qquad \Lambda(y_n) = \ln\left(\frac{f_{Y|H}(y_n|0)}{f_{Y|H}(y_n|1)}\right)$$

Note that $\Lambda(y_n)$ is a real-valued function of y_n, and is the same function for each n. Thus, since $Y_1, Y_2, \dots,$ are IID rv's conditional on $H = 0$ (or $H = 1$), $\Lambda(Y_1), \Lambda(Y_2)$, are also IID conditional on $H = 0$ (or $H = 1$).

It follows that $\ln \Lambda(\vec{y}^{(m)})$, conditional on $H = 0$ (or $H = 1$) is a sum of m IID rv's and $\{\ln \Lambda(\vec{y}^{(m)});\, m \geq 1\}$ is a random walk conditional on $H = 0$ (or $H = 1$). The two random walks contain the same sequence of sample values but different probability measures.

Later we look at sequential detection, where observations are made until a treshold is passed.

Threshold tests and the error curve

A general hypothesis testing rule (a test) consists of mapping each sample sequence $\vec{y}$ into either 0 or 1. Thus a test can be viewed as the set A of sample sequences mapped into hypothesis 1. The error probability, given $H = 0$ or $H = 1$, using test A, is given by

$$q_0(A) = \Pr\{Y \in A \mid H = 0\}\,; \qquad q_1(A) = \Pr\{Y \in A^{\mathsf{c}} \mid H = 1\}$$

With a priori probabilities p_0, p_1 and $\eta = p_1/p_0$,

$$\Pr\{\mathbf{e}(A)\} = p_0 q_0(A) + p_1 q_1(A) = p_0[q_0(A) + \eta q_1(A)]$$

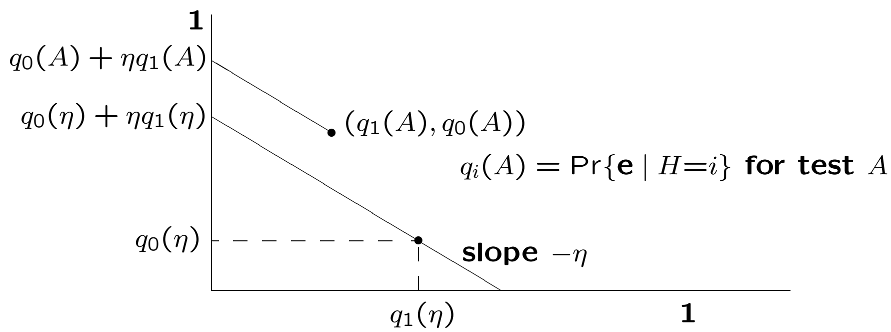
For the threshold test based on η,

$$\Pr\{\mathbf{e}(\eta)\} = p_0 q_0(\eta) + p_1 q_1(\eta) = p_0[q_0(\eta) + \eta q_1(\eta)]$$

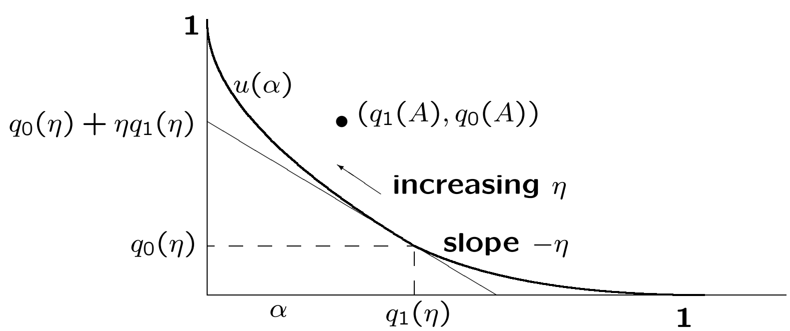
$$q_0(\eta) + \eta q_1(\eta) \leq q_0(A) + \eta q_1(A); \qquad \textbf{by MAP}$$

$$q_0(\eta) + \eta q_1(\eta) \le q_0(A) + \eta q_1(A); \qquad \textbf{by MAP}$$

Note that the point $q_0(A), q_1(A)$ does not depend on p_0; the a priori probabilities were simply used to prove the above inequality.



For every A and every η, $(q_0(A), q_1(A))$ lies NorthEast of the line of slope $-\eta$ through $(q_0(\eta), q_1(\eta))$. Thus $(q_0(A), q_1(A))$ is NE of the upper envelope of these straight lines.



If the vertical axis of the error curve is inverted, it is called a receiver operating curve (ROC) which is a staple of radar system design.

The Neyman-Pearson test is a test that chooses A to minimize $q_1(A)$ for a given constraint on $q_0(A)$. Typically this is a threshold test, but sometimes, especially if Y is discrete, it is a randomized threshold test.

Thresholds for random walks and Chernoff bounds

The Chernoff bound says that for any real b and any r such that $g_Z(r) = \mathsf{E}\left[e^{rZ}\right]$ exists,

$$\begin{aligned} \Pr\{Z \ge b\} &\le g_Z(r)\exp(-rb); && \textbf{for } b > \overline{Z},\ r > 0 \\ \Pr\{Z \le b\} &\le g_Z(r)\exp(-rb); && \textbf{for } b < \overline{Z},\ r < 0 \end{aligned}$$

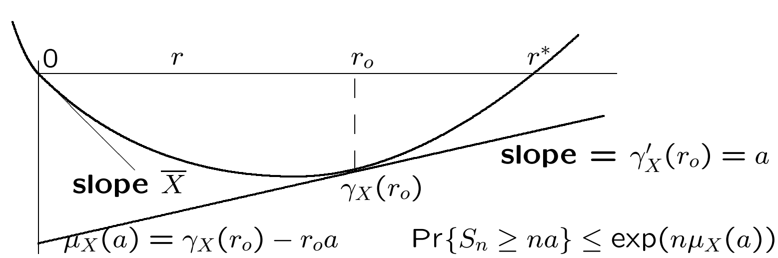
This is most useful when applied to a sum, $S_n = X_1 + \cdots X_n$ of IID rv's. If $g_X(r) = \mathsf{E}\left[e^{rX}\right]$ exists, then

$$\mathsf{E}\left[e^{rS_n}\right] = \mathsf{E}\left[\prod_{i=1}^{n} e^{rX_i}\right] = g_X^n(r)$$

$$\begin{aligned} \Pr\{S_n \ge na\} &\le g_X^n(r)\exp(-rna); && \textbf{for } a > \overline{X},\ r > 0 \\ \Pr\{S_n \le na\} &\le g_X^n(r)\exp(-rna); && \textbf{for } a < \overline{X},\ r < 0 \end{aligned}$$

This is easier to interpret and work with if expressed in terms of the semi-invariant MGF, $\gamma_X(r) = \ln g_X(r)$. Then $g_X^n(r) = e^{n\gamma_X(r)}$ and

$$\begin{aligned} \Pr\{S_n \ge na\} &\le \exp(n[\gamma_X(r) - ra]); && \textbf{for } a > \overline{X},\ r > 0 \\ \Pr\{S_n \le na\} &\le \exp(n[\gamma_X(r) - ra]); && \textbf{for } a < \overline{X},\ r < 0 \end{aligned}$$



The Chernoff bound, optimized over r, is essentially exponentially tight; i.e., $\Pr\{S_n \ge na\} \ge \exp(n(\mu_X(a) - \epsilon))$ for large enough n.

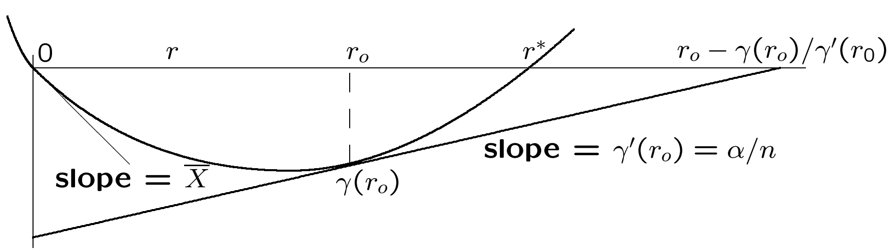
In looking at threshold problems, we want to find the probability that Pr$\{S_n \geq \alpha\}$ **for any** n. **Thus we want a bound that focuses on variable** n **for a fixed** α, **i.e., on when the threshold is crossed if it is crossed.**

We want a bound of the form Pr$\{S_n \geq \alpha\} \leq \exp \alpha f(n)$

Start with the bound Pr$\{S_n \geq na\} \leq \exp(n[\gamma_X(r_0) - r_0 a])$, **with** $\alpha = an$ **and** r_0 **such that** $\gamma_X'(r_0) = \alpha/n$. **Substituting** $\alpha/\gamma_X'(r_0)$ **for** n,

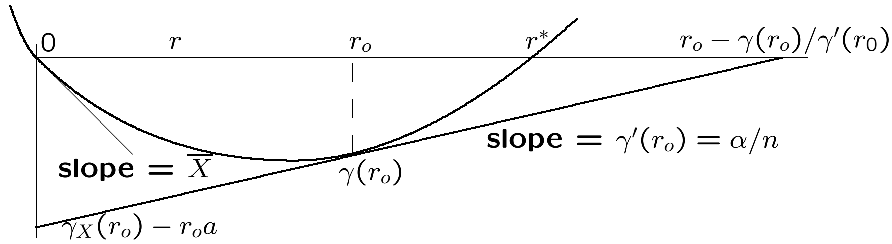
$$\Pr\{S_n \geq \alpha\} \leq \exp\left(\alpha\left[\frac{\gamma_X(r_0)}{\gamma_X'(r_0)} - r_0\right]\right)$$

$$\Pr\{S_n \geq \alpha\} \leq \exp\left(\alpha\left[\frac{\gamma_X(r_0)}{\gamma_X'(r_0)} - r_0\right]\right)$$



When n **is very large, the slope** $\gamma_X'(r_0)$ **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 remains to be explained.**



Example: $\mathsf{p}_X(1) = p,\ \mathsf{p}_X(-1) = 1{-}p;\ p < 1/2$. **Then** $g_X(r) = pe^r + (1{-}p)e^{-r}; \quad \gamma_X(r) = \ln[pe^r + (1{-}p)e^{-r}]$

Since $\gamma_X(r^*) = 0$**, we have** $pe^{r^*} + (1{-}p)e^{-r^*} = 1$**. Letting** $z = e^{r^*}$**, this is** $pz + (1{-}p)/z = 1$ **so** z **is either 1 or** $(1-p)/p$**. Thus** $r^* = \ln(1{-}p)/p$ **and**

$$\Pr\left\{\bigcup_n S_n \ge \alpha\right\} \approx \exp(-\alpha r^*) = \left(\frac{1{-}p}{p}\right)^{-\alpha}$$

which is exact for α **integer. The bound for individual** n **is the exponent in the Gaussian approximation.**