6.262: Discrete Stochastic Processes 5/4/11

L23: Martingales, plain, sub, and super

Outline:

- Review of Wald and sequential tests
- Wald's identity with zero-mean rv's
- Martingales
- Simple Examples of martingales
- Sub and super martingales

Thm: (Wald) Let $\{X_i; i \geq 1\}$ be IID with a semi-invariant MGF $\gamma(r) = \ln(\mathsf{E}[\exp(rX)])$ that exists for $(r_- < 0 < r_+)$. Let $\{S_n; n \geq 1$ be the RW with $S_n = X_1 + \cdots + X_n$. If J is the trial at which S_n first crosses $\alpha > 0$ or $\beta < 0$,

$$\mathsf{E}[\exp(rS_J - J\gamma(r)] = 1 \quad \textbf{for } r \in (r_-, r_+)$$

Corollary: If $\overline{X} < 0$ and $\gamma(r^*) = 0$ for $0 < r^*$, then

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

Pf: The Wald identity says $\mathsf{E}[\exp(r^* S_J] = 1$, so this follows from the Markov inequality.

This is valid for all lower thresholds and also for no lower threshold, where it is better stated as

$$\Pr\left\{\bigcup_n \{S_n \geq \alpha\}\right\} \leq \exp(-r^*\alpha)$$

This is stronger (for the case of threshold crossing) than the Chernoff bound, which says that $\Pr\{S_n \geq \alpha\} \leq \exp -r^*\alpha$ for all n.

Review of hypothesis testing: View a binary hypothesis as a binary rv H with $\mathsf{p}_H(0) = p_0$ and $\mathsf{p}_H(1) = p_1$.

We observe $\{Y_n; n \geq 1\}$, which, conditional on $H = \ell$ is IID with density $\mathsf{f}_{Y|H}(y|\ell)$. Define the likelihood ratio

$$\Lambda(\vec{y}^n) = \prod_{i=1}^{n} \frac{\mathsf{f}_{Y_i|H}(y_i|0)}{\mathsf{f}_{Y_i|H}(y_i|1)}$$

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

$$\textbf{MAP rule:} \quad \Lambda(\vec{y}^n) \begin{cases} > p_1/p_0 & ; \quad \textbf{select } \hat{h}{=}0 \\ \leq p_1/p_0 & ; \quad \textbf{select } \hat{h}{=}1. \end{cases}$$

Define the log likelihood ratio as

$$LLR = \ln[\Lambda(\vec{y}^n)] = \sum_{i=1}^{n} \ln \frac{\mathsf{f}_{Y_i|H}(y_i|0)}{\mathsf{f}_{Y_i|H}(y_i|1)}$$

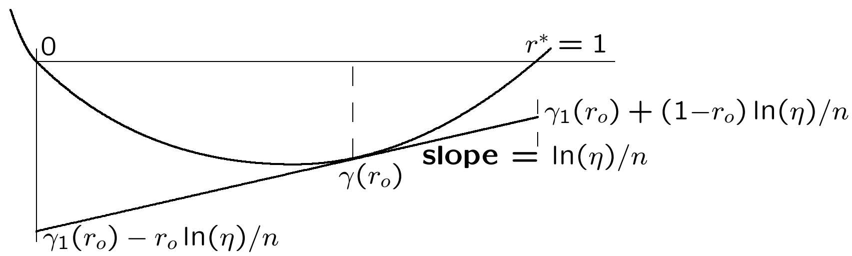
$$s_n = \sum_{i=1}^{n} z_i \qquad \textbf{where } z_i = \ln \frac{\mathsf{f}_{Y_i|H}(y_i|0)}{\mathsf{f}_{Y_i|H}(y_i|1)}$$

Conditional on $H = 1$, $\{S_n;\, n \geq 1\}$ is a RW with $S_n = Z_1 + \cdots Z_n$, where each Z_i is a function of Y_i. The Z_i, given $H = 1$ are then IID.

$$\begin{aligned} \gamma_1(r) &= \ln\left\{\int \mathsf{f}_{Y_i|H}(y_i|1) \exp\left[r \ln \frac{\mathsf{f}_{Y_i|H}(y_i|0)}{\mathsf{f}_{Y_i|H}(y_i|1)}\right] dy\right\} \\ &= \ln\left\{\int \mathsf{f}^{1-r}_{Y_i|H}(y_i|1)\mathsf{f}^{r}_{Y_i|H}(y_i|0)\, dy\right\} \end{aligned}$$

Note that $\gamma_1(1) = 0$, so $r^* = 1$.

For fixed n, a threshold rule says choose $\hat{H} = 0$ if $S_n \geq \ln \eta$. Thus, given $H = 1$, an error occurs if $S_n \geq \ln \eta$. From the Chernoff bound,



$$\Pr\{e \mid H{=}1\} \leq \exp(n\gamma_1(r_o) - r_o \ln \eta)$$

Given $H = 0$, a similar argument shows that

$$\Pr\{e \mid H{=}0\} \leq \exp(n\gamma_1(r_o) + (1 - r_o) \ln \eta)$$

A better strategy is sequential decisions. For the same pair of RW's, continue trials until either $S_n \geq \alpha$ or $S_n \leq \beta$ where $\alpha > 0$ and $\beta < 0$.

Given $H = 1$, $\{S_n; n \geq 1\}$ is a random walk. Choose some $\alpha > 0$ and $\beta < 0$ and let J be a stopping time, stopping when first $S_n \geq \alpha$ or $S_n \leq \beta$.

If $S_J \geq \alpha$, decide $\hat{H} = 0$ and if $S_J \leq \beta$, decide $\hat{H} = 1$. Conditional on $H = 1$, an error is made if $S_J \geq \alpha$. Then

$$\Pr\{e \mid H{=}1\} = \Pr\{S_J \geq \alpha \mid H{=}1\} \leq \exp[-\alpha r^*]$$

where r^* is the root of $\gamma(r) = \ln \mathsf{E}\,[\exp(rZ) \mid H = 1]$, i.e., $r^* = 1$.

$$\begin{aligned}\gamma(r) &= \ln \int_y f_{Y|H}(y|1) \exp\left[r \ln\left(\frac{f_{Y|H}(y|0)}{f_{Y|H}(y|1)}\right)\right] \\ &= \ln \int_y [f_{Y|H}(y|1)]^{1-r} [f_{Y|H}(y|0)]^r \, dy\end{aligned}$$

Choose apriori's $p_0 = p_1$. Then at the end of trial n

$$\frac{\Pr\{H{=}0 \mid S_n\}}{\Pr\{H{=}1 \mid S_n\}} = \exp(S_n); \qquad \frac{1 - \Pr\{H{=}1 \mid S_n\}}{\Pr\{H{=}1 \mid S_n\}} = \exp(S_n)$$

$$\Pr\{H{=}1 \mid S_n\} = \frac{\exp(-S_n)}{1 + \exp(-S_n)}$$

This is the probability of error if a decision $\hat{h} = 0$ is made at the end of trial n. Thus deciding $\hat{h} = 0$ on crossing α guarantees that $\Pr\{e \mid H{=}1\} \le \exp -\alpha$.

As we saw last time, the cost of choosing α to be large is many trials under $H = 0$. In particular, the stopping time J satisfies

$$\mathsf{E}[J \mid H{=}0] = \frac{\mathsf{E}[S_J \mid H{=}0]}{\mathsf{E}[Z \mid H{=}0]} \approx \frac{\alpha + \mathsf{E}[\textbf{overshoot} \mid H{=}0]}{\mathsf{E}[Z \mid H{=}0]}$$

Wald's identity with zero-mean rv's

If we take the first 2 derivatives of Wald's identity at $r = 0$, we get Wald's equality and a useful result for zero-mean rv's.

$$\frac{d}{dr}\mathsf{E}[\exp(rS_J - J\gamma(r)] = \mathsf{E}\left[[S_J - J\gamma'(r)]\exp(rS_J - J\gamma(r))\right]$$

$$\frac{d}{dr}\mathsf{E}[\exp(rS_J - J\gamma(r)]\Big|_{r=0} = \mathsf{E}\left[S_J - J\overline{X}\right] = 0; \qquad \textbf{(Wald eq.)}$$

$$\frac{d^2}{dr^2}\mathsf{E}[\exp(rS_J - J\gamma(r)]\Big|_{r=0} = \mathsf{E}\left[S_J^2 - \sigma_X^2 \overline{J}\right] = 0; \qquad \textbf{if } \overline{X} = 0$$

For zero-mean simple RW with threshold at $\alpha > 0$ and $\beta < 0$, we have $\overline{J} = -\beta\alpha$

Martingales

A sequence $\{Z_n;\, n \geq 1\}$ of rv's is a martingale if $\mathsf{E}[|Z_n|] < \infty$ for all $n \geq 1$ and

$$\mathsf{E}[Z_n \mid Z_{n-1}, Z_{n-2}, \dots, Z_1] = Z_{n-1} \qquad (1)$$

The condition $\mathsf{E}[|Z_n|] < \infty$ is almost a mathematical fine point, and we mostly ignore it here. The condition (1) appears to be a very weak condition, but it leads to surprising applications. In times of doubt, write (1) as

$$\mathsf{E}[Z_n \mid Z_{n-1}{=}z_{n-1}, \dots, Z_1{=}z_1] \;=\; z_{n-1}$$

for all sample values $z_{n-1}, z_{n-2}, \dots, z_1$

Lemma: For a martingale, $\{Z_n;\, n \geq 1\}$, and for $n > i \geq 1$,

$$\mathsf{E}[Z_n \mid Z_i, Z_{i-1} \dots, Z_1] = Z_i$$

Pf: To start, we show that $\mathsf{E}[Z_3 \mid Z_1] = Z_1$. Recall the meaning of $\mathsf{E}[X] = \mathsf{E}[\mathsf{E}[X|Y]]$. Then

$$\mathsf{E}[Z_3 \mid Z_1] = \mathsf{E}[\mathsf{E}[Z_3 \mid Z_2, Z_1] \mid Z_1]$$

$$\mathsf{E}[Z_3 \mid Z_1] = \mathsf{E}[\mathsf{E}[Z_3 \mid Z_2, Z_1] \mid Z_1] = \mathsf{E}[Z_2|Z_1] = Z_1$$

In the same way,

$$\begin{aligned} \mathsf{E}\Big[Z_{i+2} \mid Z_i, \dots, Z_1\Big] &= \mathsf{E}\Big[\mathsf{E}\Big[Z_{i+2} \mid Z_{i+1}, \dots, Z_1\Big] \mid Z_i, \dots, Z_1\Big] \\ &= \mathsf{E}\Big[Z_{i+1} \mid Z_i, \dots, Z_1\Big] = Z_i \end{aligned}$$

After more of the same, $\mathsf{E}[Z_n \mid Z_i, \dots, Z_1] = Z_i$.

The most important special case is $\mathsf{E}[Z_n \mid Z_1] = Z_1$, and thus $\mathsf{E}[Z_n] = \mathsf{E}[Z_1]$.

Simple Examples of martingales

1) Zero-mean random walk: Let $Z_n = X_1 + \cdots X_n$ where $\{X_i;\ i \geq 1\}$ are IID and zero mean.

$$\begin{aligned} \mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1] &= \mathsf{E}[X_n + Z_{n-1} \mid Z_{n-1}, \dots, Z_1] \\ &= \mathsf{E}[X_n] + Z_{n-1} = Z_{n-1}. \end{aligned}$$

2) Sums of 'arbitrary' dependent rv's: Suppose $\{X_i;\ i \geq 1\}$ satisfy $\mathsf{E}[X_i \mid X_{i-1}, X_{i-2}, \dots, X_1] = 0$. Then $\{Z_n;\ n \geq 1\}$ where $Z_n = X_1 + \cdots + X_n$ is a martingale.

This can be taken as an alternate definition of a martingale. We can either start with the sums Z_n or with the differences between successive sums.

3) Let $X_i = U_i Y_i$ where $\{U_i;\ i \geq 1\}$ are IID, equiprobable ± 1. The Y_i are non-negative and independent of the U_i but otherwise arbitrary. Then

$$\mathsf{E}[X_n \mid X_{n-1}, \dots, X_1] = 0$$

Thus $\{Z_n;\ n \geq 1\}$ where $Z_n = X_1 + \cdots X_n$ is a martingale.

4) Product form martingales. Suppose $\{X_i;\ i \geq 1\}$ is a sequence of IID unit-mean rv's. Then $\{Z_n;\ n \geq 1\}$ where $Z_n = X_1 X_2 \cdots X_n$ is a martingale.

$$\begin{aligned} \mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1] &= \mathsf{E}[X_n Z_{n-1} \mid Z_{n-1}, \dots, Z_1] \\ &= \mathsf{E}[X_n]\, \mathsf{E}[Z_{n-1} \mid Z_{n-1}, \dots, Z_1] \\ &= \mathsf{E}[Z_{n-1} \mid Z_{n-1}] = Z_{n-1}. \end{aligned}$$

5) Special case of product form martingale: let X_i be IID and equiprobably 2 or 0.

$$\Pr\{Z_n = 2^n\} = 2^{-n}; \quad \Pr\{Z_n = 0\} = 1 - 2^{-n}; \quad \mathsf{E}[Z_n] = 1.$$

Thus $\lim_n Z_n = 0$ **WP1 but** $\mathsf{E}[Z_n] = 1$ **for all** n

6) Recall the branching process where X_n is the number of elements in gen n and $X_{n+1} = \sum_{i=1}^{X_n} Y_{i,n}$ where the $Y_{i,n}$ are IID.

Let $Z_n = X_n/\overline{Y}^n$, i.e., $\{Z_n;\, n \geq 1\}$ is a scaled down branching process.

$$\mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1] = \mathsf{E}\left[\frac{X_n}{\overline{Y}^n} \mid X_{n-1}, \dots, X_1\right] = \frac{\overline{Y} X_{n-1}}{\overline{Y}^n} = Z_{n-1}.$$

Thus this is a martingale.

Submartingales and supermartingales

These are sequences $\{Z_n; n \geq 1\}$ with $\mathsf{E}[|Z_n|] < \infty$ like martingales, but with inequalities instead of equalities. For all $n \geq 1$,

$$\begin{aligned} \mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1] &\geq Z_{n-1} && \textbf{submartingale} \\ \mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1] &\leq Z_{n-1} && \textbf{supermartingale} \end{aligned}$$

We refer only to submartingales in what follows, since the supermartingale case results from replacing Z_n with $-Z_n$.

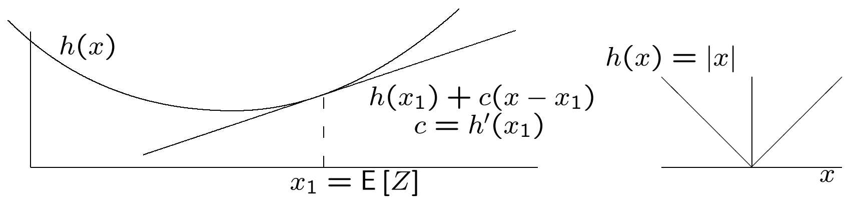
For submartingales,

$$\mathsf{E}[Z_n \mid Z_i, \dots, Z_1] \geq Z_i \qquad \textbf{for all } n > i > 0$$

$$\mathsf{E}[Z_n] \geq \mathsf{E}[Z_i] \qquad \textbf{for all } n > i > 0$$

Convex functions

A function $h(x)$, $\mathbb{R} \to \mathbb{R}$, is convex if each tangent to the curve lies on or below the curve. The condition $h''(x) \geq 0$ is sufficient but not necessary.



Lemma (Jensen's inequality): If h is convex and Z is a rv with finite expectation, then

$$h(\mathsf{E}[Z]) \leq \mathsf{E}[h(Z)]$$

Jensen's inequality can be used to prove the following theorem. See Section 7.7 for a proof.

If $\{Z_n;\, n \geq 1\}$ is a martingale or submartingale, if h is convex, and if $\mathsf{E}[|h(Z_n)|] < \infty$ for all n, then $\{h(Z_n);\, n \geq 1\}$ is a submartingale.

For example, if $\{Z_n;\, n \geq 1\}$ is a martingale, then essentially $\{|Z_n|;\, n \geq 1\}$, $\{Z_n^2;\, n \geq 1\}$ and $\{e^{rZ_n};\, n \geq 1$ are submartingales.

Stopped martingales

The definition of a stopping time for a stochastic process $\{Z_n;\, n \geq 1\}$ applies to any process. That is, J must be a rv and $\{J = n\}$ must be specified by $\{Z_1, \dots, Z_n\}$.

This can be extended to possibly defective stopping times if J is possibly defective (consider a random walk with a single threshold).

A stopped process $\{Z_n^*;\, n \geq 1\}$ for a possibly defective stopping time J on a process $\{Z_n;\, n \geq 1\}$ satisfies $Z_n^* = Z_n$ if $n \leq J$ and $Z_n^* = Z_J$ if $n > J$.

For example, a given gambling strategy, where Z_n is the net worth at time n, could be modified to stop when Z_n reaches some given value. Then Z_n^* would remain at that value forever after, while Z_n follows the original strategy.

Theorem: If $\{Z_n;\, n \geq 1\}$ is a martingale (submartingale) and J is a possibly defective stopping rule for it, then the stopped process $\{Z_n^*;\, n \geq 1\}$ is a martingale (submartingale).

Pf: Obvious??? The intuition here is that before stopping occurs, $Z_n^* = Z_n$, so Z_n^* satisfies the martingale (subm.) condition. Afterwards, Z_n^* is constant, so it again satisfies the martingale (subm) condition. Section 7.8 does this carefully.

MIT OpenCourseWare
http://ocw.mit.edu

6.262 Discrete Stochastic Processes
Spring 2011

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.