6.262 Discrete Stochastic Processes, Spring 2009
Problem Set 10 — Solutions
due: Wednesday, April 27, 2009

Exercise 7.1

Given the random walk $S_n = X_1 + \ldots + X_n$ with $\mathrm{P}(X_i = 1) = 1 - \mathrm{P}(X_i = -1) = p < 1/2$ and X_i IID, we are interested in $\mathrm{P}(\sup_{n\geq 1} S_n \geq k)$ for any positive integer k. In fact, we're actually interested in the probability of the event that that the random walk reaches or exceeds k (that is, the probability of the event that $S_n \geq k$ for some n), which can be expressed as $\mathrm{P}(\bigcup_{n\geq 1}\{S_n \geq k\})$. Given that the random walk evolves in unit increments, we have that $\bigcup_{n\geq 1}\{S_n \geq k\} = \{\sup_{n\geq 1} S_n \geq k\}$. Can you see why this is not true in general? Still, can you show that $\mathrm{P}\left(\bigcup_{n\geq 1}\{S_n \geq k\}\right) = \mathrm{P}\left(\sup_{n\geq 1} S_n \geq k\right)$? (It's not immediately obvious, but it's not beyond your grasp.)

However, note that for the purpose of this exercise, the event $\{\sup_{n\geq 1} S_n \geq k\}$ has no special additional properties that make it a more desirable object — it merely simplifies the notation.

a The idea of the solution is to say that the moment the random walk reaches 1, it probabilistically restarts, and thus reaching k is the same as reaching 1 exactly k times. However, to show that the walk "probabilistically restarts", do we have to use the memoryless property of the underlying Bernoulli process? Or is this true for any random walk due to its IID increments? It turns out that if we do this argument carefully, we see that the issue is not that of "memorylessness" but is due to an underlying stopping rule, which can be defined for any random walk. Granted, the fact that $\mathrm{P}\left(\sup_{i\geq 1} S_i \geq k\right) = \mathrm{P}\left(\sup_{i\geq 1} S_i \geq 1\right)^k$ will only hold for a random walk that evolves in unit increments (which *significantly* narrows down our choices), but the notion of effectively restarting conditioning on having reached some value is universal to random walks.

For $m = 1, 2, \ldots$, consider the event $A_m = \{S_m = 1 \text{ and } S_n < 1 \text{ for } n < m\}$. Also, let $A = \bigcup_{m\geq 1} A_m$. (Note that A is the event that $S_m = 1$ *for some* m and A_m is the event that $S_m = 1$ *for the first time* at m.) Since in order to exceed $k \geq 1$, the random walk first needs to exceed 1, we have that $A \subset \{\sup_{i\geq 1} S_i \geq k\}$. Thus,

$$\mathrm{P}\left(\sup_{i\geq 1} S_i \geq k\right) = \mathrm{P}\left(\left\{\sup_{i\geq 1} S_i \geq k\right\} \cap A\right) = \sum_{m\geq 1} \mathrm{P}\left(\left\{\sup_{i\geq 1} S_i \geq k\right\} \cap A_m\right)$$

where the last equality follows by the fact that A_m are disjoint. Reformulating the above statement through conditional probabilities yields

$$\mathrm{P}\left(\sup_{i\geq 1} S_i \geq k\right) = \sum_{m\geq 1} \mathrm{P}\left(\left\{\sup_{i\geq 1} S_i \geq k\right\} \mid A_m\right) \mathrm{P}(A_m).$$

Let us take a closer look at $\mathrm{P}\left(\left\{\sup_{i\geq 1} S_i \geq k\right\} \mid A_m\right)$. For $i = 1, 2, \ldots,$ define $X_i' = X_{m+i}$ and similarly let $S_n' = X_1' + \ldots + X_n'$. Now note that

$$\mathrm{P}\left(\left\{\sup_{i\geq 1} S_i \geq k\right\} \mid A_m\right) = \mathrm{P}\left(\left\{\sup_{i\geq 1} S_i' \geq k-1\right\} \mid A_m\right). \quad \text{(why?)}$$

Furthermore, note that for any i, $S_i' = X_1' + \ldots + X_i' = X_{m+1} + \ldots + X_{m+1}'$ is independent of $X_1, \ldots, X_m$ and therefore of the event A_m. (Does this ring a bell? Can you se why it was important to define A_m as the event that the walk reaches 1 *for the first time* at m? It tells us that A_m depends only on $X_1, \ldots, X_m$, and therefore not on X_{m+1} and onwards.) It therefore follows that

$$\mathrm{P}\left(\left\{\sup_{i\geq 1} S_i \geq k\right\} \mid A_m\right) = \mathrm{P}\left(\sup_{i\geq 1} S_i' \geq k-1\right) = \mathrm{P}\left(\left\{\sup_{i\geq 1} S_i \geq k-1\right\}\right),$$

where the last equality follows by the fact that S_i' has the same distribution as S_i (why?). Since this holds for any $m \in \mathbb{Z}_+$, it follows that

$$\begin{aligned}
\sum_{m\geq 1} \mathrm{P}\left(\left\{\sup_{i\geq 1} S_i \geq k\right\} \mid A_m\right) \mathrm{P}(A_m) &= \sum_{m\geq 1} \mathrm{P}\left(\sup_{i\geq 1} S_i \geq k-1\right) \mathrm{P}(A_m) \\
&= \mathrm{P}\left(\sup_{i\geq 1} S_i \geq k-1\right) \sum_{m\geq 1} \mathrm{P}(A_m) \\
&= \mathrm{P}\left(\sup_{i\geq 1} S_i \geq k-1\right) \mathrm{P}(A) \\
&= \mathrm{P}\left(\sup_{i\geq 1} S_i \geq k-1\right) \mathrm{P}\left(\sup_{i\geq 1} S_i \geq 1\right).
\end{aligned}$$

The above recursive equation yields

$$\mathrm{P}\left(\sup_{i\geq 1} S_i \geq k\right) = \mathrm{P}\left(\sup_{i\geq 1} S_i \geq 1\right)^k$$

and the result follows.

Note that we did not use the condition that $p < 1/2$.

b First note that $S_1 = X_1$ and therefore S_1 takes values in the set $\{-1, 1\}$ with probabilities $1-p$ and p, respectively. Conditioning yields

$$\mathrm{P}\left(\sup_{i\geq 1} S_i \geq 1\right) = \mathrm{P}\left(\left\{\sup_{i\geq 1} S_i \geq 1\right\} \mid S_1 = 1\right) p + \mathrm{P}\left(\left\{\sup_{i\geq 1} S_i \geq 1\right\} \mid S_1 = 1\right)(1-p).$$

Clearly, $\mathrm{P}\left(\left\{\sup_{i\geq 1} S_i \geq 1\right\} \mid S_1 = 1\right) = 1$. Furthermore, extending argument of a),

$$\mathrm{P}\left(\left\{\sup_{i\geq 1} S_i \geq 1\right\} \mid S_1 = 1\right) = \mathrm{P}\left(\sup_{i\geq 1} S_i \geq 1\right)^2.$$

The desired quadratic equation therefore becomes

$$x = p + x^2(1-p) \quad \Longleftrightarrow \quad x^2 - \frac{x}{1-p} + \frac{p}{1-p} = 0,$$

c Let $Q(x) = x^2 - \frac{x}{1-p} + \frac{p}{1-p}$ and notice that for $Q(1) = 0$ for any real p. Furthermore, for $p \neq 1$, $Q(p/(1-p)) = 0$ as well. For $p = 1/2$, the solutions naturally match. (Recall a betting example in class where gaining a \$1 for heads nad losing \$1 for tails, we would eventually be ahead by \$1 w.p.1. for a fair coin.) For $p > 1/2$, we have shown in the past that S_n drifts off to $+\infty$ (specifically, $S_n \to \infty$ w.p.1) and $\mathrm{P}\left(\sup_{i\geq 1} S_i \geq k\right) = 1$, so the larger root is the correct one. For $p < 1/2$, reversing the argument for $p > 1/2$, we obtain that the walk reaches -1 w.p.1, from which it follows that the walk eventually reaches $-k$ for any positive integer k with probability 1. Supose we still have $\mathrm{P}\left(\sup_{i\geq 1} S_i \geq k\right) = 1$. It follows that the walk eventually reaches k for any integer k with probability 1. Therefore, the walk crosses the origin infinitely often w.p.1, which is a contradiction since 0 is a transient state. It follows that for $p < 1/2$, we have $\mathrm{P}\left(\sup_{i\geq 1} S_i \geq k\right) = p/(1-p)$.

d We have $g_X(r) = (1-p)e^{-r} + pe^r$. Letting $g_X(r) = 1$ and setting $x = e^{-r}$, we obtain that $1 = (1-p)x + p/x$. This is equivalent to $x^2 - x/(1-p) - p/(1-p)$ for $x \neq 0$, which is the quadratic equation solved in c). The roots are therefore $e^{r'} = 1$ and $e^{r''} = p/(1-p)$, that is, $r' = 0$ and $r'' = \log(p/(1-p))$, which is strictly greater than 0 for $0 < p < 1$.

Why is this interesting? A sub-optimal version of the Chernoff bound tells us that for any n, $\mathrm{P}(S_n \geq a) \leq e^{-r^*a}$ where r^* is such that $g_X(r^*) = 1$. Therefore, for any n, $\mathrm{P}(S_n \geq 1) \leq \frac{p}{1-p}$. However, we have shown something stronger: we have that $\mathrm{P}(\sup_{n\geq 1} S_n \geq 1) = \frac{p}{1-p}$. Our answer is in fact more closely related to the bound obtained from Wald's Identity. Namely, set the thresholds as $\beta = -\infty$ and $\alpha = 1$. Then we have that the event $\{S_N \geq 1\}$ is equal to the event that S_n reaches 1 for some n, i.e. $\{\sup_{n\geq 1} S_n \geq 1\}$ in our case. Since from Wald's Identity, $\mathrm{P}(S_N \geq a) \leq e^{-r^*a}$ for the same r^* as previously, it follows that $\mathrm{P}(\sup_{n\geq 1} S_n \geq 1) \leq \frac{p}{1-p}$. We have in fact shown that the bound obtained by setting the thresholds as specified is tight (i.e. there exists a random walk that meets it with equality).

Exercise 7.2

a From Eqn. 7.5 in the notes, we have $W_n = \max\{U_1 + \ldots + U_n, U_2 + \ldots + U_n, \ldots, U_{n+1} + U_n, U_n, 0\}$. Thus, $W_6 = \max\{0, Z_1^6, Z_2^6, \ldots, Z_6^6\} = \max\{0, -2, 1, 0, 2, 0, 1\} = 2$.

b For $n = 1, \ldots, 6$, we have $W_n = \max\{0, Z_1^n, Z_2^n, \ldots, Z_n^n\}$, where $Z_i^n = U_n + \ldots + U_{n-i+1}$. This yields, $W_1 = 1$, $W_2 = 0$, $W_3 = 2$, $W_4 = 1$, $W_5 = 4$.

c Customers 0 (trivially) and 2 start a busy period. Interestingly, from a) we notice that letting $i = 4$ maximizes Z_i^6. Is that a coincidence? Well, if the maximum of Z_i^6 occurs at some $i = i^*$, then for all $i > i^*$, $Z_i^6 \leq Z_{i^*}^6$. From the way Z_i^6 is defined, this implies that for all $i > i^*$, $U_{6-i} + \ldots + U_{6-i^*} \leq 0$, and thus,

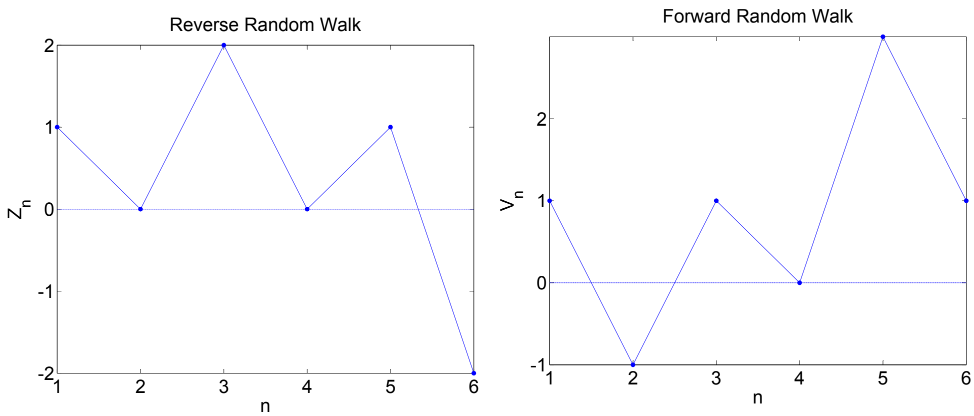
$$W_{6-i^*} = \max\{U_1 + \ldots + U_{6-i^*}, U_2 + \ldots + U_{6-i^*}, \ldots, U_{6-i^*}, 0\} = 0,$$

i.e. customer $6 - i^*$ comes in on an empty system.

d Letting $V_n = U_1 + \ldots + U_n$, Eqn. 7.5 then yields

$$W_n = \max\{V_n, V_n - V_1, V_n - V_2, \ldots, V_n - V_{n-1}, 0\}.$$

Given the sample path $(u_1, \ldots, u_6) = (1, -2, 2, -1, 3, -2)$, the two random walks of this problem are the following:



Exercise 7.5

Given an observation y, chosing H_1 incurs an expected cost of $C_0\mathrm{P}(H_0 \mid y)$, while chosing H_0 incurs an expected cost of $C_1\mathrm{P}(H_1 \mid y)$. Conditioned on $Y = y$, the expected cost is therefore minimized by choosing H_1 if $C_0\mathrm{P}(H_0 \mid y) < C_1\mathrm{P}(H_1 \mid y)$, and choosing H_0 otherwise. (Why does it not matter which hypothesis we choose when $C_0\mathrm{P}(H_0 \mid y) = C_1\mathrm{P}(H_1 \mid y)$?) Taking logs,

$$\log \frac{C_1\mathrm{P}(H_1 \mid y)}{C_0\mathrm{P}(H_0 \mid y)} \begin{cases} > 0, & \text{choose } H_1 \\ \leq 0, & \text{choose } H_0 \end{cases}.$$

But, for $y = (y_1, \ldots, y_n)$,

$$\log \frac{\mathrm{P}(H_1 \mid y)}{\mathrm{P}(H_0 \mid y)} = \log \frac{p_1}{p_0} + \sum_{i=1}^{n} z_i \quad \text{where} \quad z_i = \log \frac{f(y \mid H_1)}{f(y \mid H_0)}$$

where the above expansion is simply a reformulation through conditional probabilities (Eqn. 7.13 in the text). The threshold test can therefore be expressed as

$$\sum_{i=1}^{n} z_i \begin{cases} > \log((p_0 C_0/(p_1 C_1)), & \text{choose } H_1 \\ \leq \log((p_0 C_0/(p_1 C_1)), & \text{choose } H_0 \end{cases}.$$

Obviously, the above answer is not unique as from the point of view of minimizing the expected cost, it does not matter which hypothesis we choose when $C_0\mathrm{P}(H_0 \mid y) = C_1\mathrm{P}(H_1 \mid y)$.

Exercise 7.8

a First note that $g_{X-c}(r) = \mathrm{E}(e^{(X-c)r}) = e^{-rc}\mathrm{E}(e^{Xr}) = e^{-rc}g_X(r)$, which is defined for all r in the region of convergence of g_X. Taking derivatives,

$$g''_{X-c}(r) = \frac{d^2}{dx^2}\mathrm{E}(e^{(X-c)r}) = \mathrm{E}(\frac{d^2}{dX^2}e^{(X-c)r}) = \mathrm{E}((X-c)^2e^{(X-c)r}),$$

where taking the derivative inside the expectation integral is justified by Leibniz's integral rule from Calculus (why?). Since $(x-c)^2e^{(x-c)r} > 0$ for all $x \in \mathbb{R}$, it follows that $g''_{X-c}(r) \geq 0$.

b It suffices to note that

$$g''_{X-c}(r) = \frac{d^2}{dx^2}\mathrm{E}(e^{(X-c)r}) = \frac{d^2}{dx^2}(e^{-rc}g_X(r)) = e^{-rc}(g''_X(r) - 2cg'_X(r) + c^2g_X(r)).$$

c By part a, for any c, we have that $g_{X-c}(r) \geq 0$, for all r in the ROC of g_X. Fixing any such r, it follows that $e^{-rc}(g''_X(r) - 2cg'_X(r) + c^2g_X(r)) \geq 0$ for all c. This is equivalent to requiring that $g''_X(r) - 2cg'_X(r) + c^2g_X(r) \geq 0$ for all c, which is in turn equivalent to stating that the resulting quadratic equation in c has at most one real root (why?). Finally, this is equivalent to having the discriminant of the quadratic equation be non-positive, that is, $4c^2(g'_X(r))^2 - 4c^2g''_X(r)g_X(r) \leq 0$. For $c \neq 0$, this is equivalent to $(g'_X(r))^2 - g''_X(r)g_X(r) \leq 0$ and the result follows.

To show the second claim, recall that $g_X(r) > 0$ for all r in the ROC. Taking logs, $\gamma_X = \log g_X$ and thus $\gamma''_X(r) = \frac{g''_X(r)g_X(r)-(g'_X(r))^2}{g_X^2} \geq 0$, where the inequality follows immediately from the previous claim. Note that we have just shown that γ_X is convex!

d It suffices here to show that if $\mathrm{P}(C = c) \neq 1$, then $g''_{X-c}(r) > 0$ for all c and all r in the ROC. From there, it will follow that the discriminant needs to be strictly negative (why?) and thus $\gamma''_X(r) > 0$ (i.e. γ_X is strictly convex). To show the first claim, note that if for some $x \neq c$, we have $\mathrm{P}(X = x) > 0$, then $g''_{X-c}(r) = \mathrm{E}((X-c)^2e^{(X-c)r}) > P(X = x)(x-c)^2e^{(X-c)r}) > 0$. Suppose instead X has some density f. The point is to show that $\int_{\mathbb{R}}(x-c)^2e^{(x-c)r}f(x)dx > 0$. While the statement appears obvious (and should be taken as such in an exam-type situation), it is good practice to be a little more careful. Since $\int_{\mathbb{R}-\{c\}} f(x)dx > 0$ and $f \geq 0$, there must exist some $x_0 \in \mathbb{R} - \{c\}$ such that $f(x_0) > 0$. By the continuity of f, it follows that there exists some interval $(x_0 - \delta, x_0 + \delta)$ such that for all x in that interval $f(x) \geq f(x_0)/2$ (go back to our definition of continuity and let $\epsilon = f(x_0)/2$). Since $(x-c)^2e^{(x-c)r}f(x) \geq 0$ for all x, we have

$$\int_{\mathbb{R}}(x-c)^2e^{(x-c)r}f(x)dx \geq \int_{(x_0-\delta,x_0+\delta)}(x-c)^2e^{(x-c)r}f(x)dx$$

and, in turn,

$$\int_{(x_0-\delta,x_0+\delta)}(x-c)^2e^{(x-c)r}f(x)dx \geq \int_{(x_0-\delta,x_0+\delta)}(x-c)^2e^{(x-c)r}\frac{f(x_0)}{2}dx > 0$$

where the last inequality follows from the fact that $(x-c)^2e^{(x-c)r} > 0$ if $x \neq c$ and $f(x_0)/2 > 0$ by construction. Thus, $\int_{\mathbb{R}}(x-c)^2e^{(x-c)r}f(x)dx > 0$ and we're done.

(That wasn't so bad, wasn't it?)

As a question of deeper interest, the above proof is actually incomplete. It is possible to construct random variables on $\mathbb{R}$ that do not have a density, but cannot be expressed through some density + collection of point masses (i.e. points c for which $P(X = c) > 0$). Such random variables are not considered in this course — recall that we consider only random variables that are "continuous", "discrete" or "mixed". The probability of encountering a random variable that does not subscribe to the previous three types in a practical setting is zero. However, it is good to keep in mind that probability theory is a deep and interesting subject, and there's always something more than meets the eye (perhaps much more so than in any other mathematical discipline of engineering interest).