

6.262 Discrete Stochastic Processes, Spring 2009
Problem Set 12 — Solutions
Not due.

Problem 1 (Exercise 7.31)

a First note that

$$\begin{aligned} \mathbb{E}(Y_n \mid Z_{n+m-1}, Z_{n+m-2}, \dots, Z_1) &= \mathbb{E}(Z_{n+m} - Z_m \mid Z_{n+m-1}, Z_{n+m-2}, \dots, Z_1) \\ &= \mathbb{E}(Z_{n+m} \mid Z_{n+m-1}, Z_{n+m-2}, \dots, Z_1) \\ &\quad - \mathbb{E}(Z_m \mid Z_{n+m-1}, Z_{n+m-2}, \dots, Z_1) \end{aligned}$$

by the linearity of the expectation, since, as Z_n is a martingale, both terms are necessarily finite. (Essentially, Z_n being a martingale rules out the fact that

$$\mathbb{E}(Z_{n+m} \mid Z_{n+m-1}, Z_{n+m-2}, \dots, Z_1) = \mathbb{E}(Z_m \mid Z_{n+m-1}, Z_{n+m-2}, \dots, Z_1) = \infty,$$

in which case $\mathbb{E}(Z_{n+m} - Z_m \mid Z_{n+m-1}, Z_{n+m-2}, \dots, Z_1)$ may exist but the difference of the two expectations is undefined.)

Since $\mathbb{E}(Z_{n+m} \mid Z_{n+m-1}, Z_{n+m-2}, \dots, Z_1) = Z_{n+m-1}$ (as Z_n is a martingale) and $\mathbb{E}(Z_m \mid Z_{n+m-1}, Z_{n+m-2}, \dots, Z_1) = Z_m$ (as Z_m is included in the condition on $Z_{n+m-1}, Z_{n+m-2}, \dots, Z_1$), it follows that

$$\mathbb{E}(Y_n \mid Z_{n+m-1}, Z_{n+m-2}, \dots, Z_1) = Z_{n+m-1} - Z_m.$$

b First note that specifying Z_{n+m-1}, \dots, Z_1 entirely specifies Y_{n-1}, \dots, Y_1 but not vice-versa (why?). However, we can see that specifying Z_{n+m-1}, \dots, Z_1 is equivalent to specifying $Y_{n-1}, \dots, Y_1, Z_m, Z_{m-1}, \dots, Z_1$. By Problem 2-a), we therefore have

$$\begin{aligned} \mathbb{E}(Y_n \mid Y_{n-1}, \dots, Y_1) &= \mathbb{E}(\mathbb{E}(Y_n \mid Y_{n-1}, \dots, Y_1, Z_m, Z_{m-1}, \dots, Z_1) \mid Y_{n-1}, \dots, Y_1) \\ &= \mathbb{E}(\mathbb{E}(Y_n \mid Z_{n+m-1}, Z_{n+m-2}, \dots, Y_{m+1}, Z_m, Z_{m-1}, \dots, Z_1) \mid Y_{n-1}, \dots, Y_1) \\ &= \mathbb{E}(Z_{n+m-1} - Z_m \mid Y_{n-1}, \dots, Y_1) \quad \text{by part a)} \\ &= \mathbb{E}(Y_{n-1} \mid Y_{n-1}, \dots, Y_1) \quad \text{by definition} \\ &= Y_{n-1} \end{aligned}$$

c $\mathbb{E}(|Y_n|) = \mathbb{E}(|Z_{n+m} - Z_m|) \leq \mathbb{E}(|Z_{n+m}| + |Z_m|)$. Since Z_n is a martingale, it follows that $\mathbb{E}(|Z_k|) < \infty$ for all k , and so $\mathbb{E}(|Y_n|) \leq \mathbb{E}(|Z_{n+m}|) + \mathbb{E}(|Z_m|) < \infty$.

Problem 2

- a We show the tower property assuming X, Y and Z have densities. The proof adapts easily to the case where the three random variables are discrete by considering the corresponding probability mass functions. The most general case is discussed afterwards.

First note that

$$E(X | Y = y, Z = z) = \int x f_{X|Y,Z}(x | y, z) dx = \int x \frac{f_{X,Y,Z}(x, y, z)}{f_{Y,Z}(y, z)} dx.$$

Hence,

$$\begin{aligned} E(E(X | Y, Z)) &= \int E(X | Y = y, Z = z) f_{Z|Y}(z | y) dz = \int \int x \frac{f_{X,Y,Z}(x, y, z)}{f_{Y,Z}(y, z)} \frac{f_{Y,Z}(y, z)}{f_Y(y)} dx dz \\ &= \int \int x \frac{f_{X,Y,Z}(x, y, z)}{f_Y(y)} dx dz = \int \int x \frac{f_{X,Y,Z}(x, y, z)}{f_Y(y)} dz dx \quad (\star) \text{ see note below} \\ &= \int \int x \frac{f_{X,Y}(x, y)}{f_Y(y)} dx = \int \int x f_{X|Y}(x | y) dx = E(X | Y = y) \end{aligned}$$

This implies the property $E(Z_n | Z_{n-1}, \dots, Z_1) = Z_{n-1} \forall n$ is stronger than the property $E(Z_n | Z_{n-1}) = Z_{n-1} \forall n$. In particular, if Z_n satisfies $E(Z_n) < \infty$ and $E(Z_n | Z_{n-1}) = Z_{n-1} \forall n$, it need not follow that $\{Z_n\}_{n \geq 1}$ is a martingale. Can you think of a good example?

Note: The interchange of integrals in (\star) holds when $E(|X| | Y = y) < \infty$. However, the tower property is true in general, and is much easier to prove in its general form. In probabilistic lingo, a σ -field is a collection of events that satisfies some closure properties (for instance, we want an intersection of events to be an event). Let \mathcal{G} be the σ -field that contains all the events of interest in our problem. Then, $E(X)$ is really shorthand for $E(X | \mathcal{G})$, i.e. the expectation of the random variable X taking into account all possible events. Suppose we want to restrict the set of possible events (for instance, look at events consistent with $Y = y$). Then, denoting the corresponding restricted σ -field by \mathcal{F} , we have $\mathcal{F} \subset \mathcal{G}$ (i.e. the restricted events are contained in the collection of all events of interest). Then, the more general statement of the tower property is the following: if $\mathcal{F} \subset \mathcal{G}$, then $E(X | \mathcal{F}) = E(E(X | \mathcal{G}) | \mathcal{F})$ w.p.1 (e.g. if we condition on $Y = y$, this works for “most” – in the sense of w.p.1 – values of y). It turns out that the proof of this fact does not involve any interchange of integrals.

- b For some fixed sequence $1 \leq n_1 \leq n_2 \leq \dots$, let $W_k = Z_{n_k}$. Since $E(|Z_n|) < \infty$ for all n , it follows that $E(|W_k|) < \infty$ for all k . Furthermore,

$$\begin{aligned} E(W_k | W_{k-1}, \dots, W_1) &= E(Z_{n_k} | Z_{n_{k-1}}, \dots, Z_{n_1}) \quad \text{by definition} \\ &= Z_{n_{k-1}} \quad \text{since } Z_n \text{ is a martingale} \\ &= W_{k-1} \end{aligned}$$

for all $k \geq 1$. It follows that W_k is a martingale.

Problem 3 To be skipped. The desired counterexample did not work out.

Problem 4

Let G_n denote the number of green rabbits in the hat after the n^{th} trial and set $\tilde{G}_n = G_n/(n+2)$. Note that starting with exactly 2 rabbits, there are $n + 2$ rabbits in the hat after the n^{th} trial. Conditioned on knowing G_{n-1} , the probability of drawing a green rabbit is therefore $G_{n-1}/(n+1)$, in which case there will be $G_{n-1} + 1$ rabbits in the hat after the n^{th} trial, and the probability of drawing a white rabbit is $1 - \tilde{G}_{n-1}/(n+1)$, in which case there will be G_{n-1} rabbits in the hat after the n^{th} trial. Thus,

$$E(G_n \mid G_{n-1}, G_{n-2}, \dots, G_1) = G_{n-1} + \frac{G_{n-1}}{n+1} = \frac{n+2}{n+1} G_{n-1}.$$

But, since the process has the Markov property (all that matters for the outcome of trial n is the makeup of the hat's population at time $n - 1$), $E(G_n \mid G_{n-1}, G_{n-2}, \dots, G_1) = E(G_n \mid G_{n-1})$, so

$$E(G_n \mid G_{n-1}, G_{n-2}, \dots, G_1) = \frac{n+2}{n+1} G_{n-1}.$$

Now, we're interested in $E(\tilde{G}_n \mid \tilde{G}_{n-1}, \tilde{G}_{n-2}, \dots, \tilde{G}_1)$, but $E(\tilde{G}_n \mid \tilde{G}_{n-1}, \tilde{G}_{n-2}, \dots, \tilde{G}_1) = E(\tilde{G}_n \mid G_{n-1}, G_{n-2}, \dots, G_1)$ (why?). Furthermore, $E(\tilde{G}_n \mid \tilde{G}_{n-1}, \tilde{G}_{n-2}, \dots, \tilde{G}_1) = E(G_n \mid \tilde{G}_{n-1}, \tilde{G}_{n-2}, \dots, \tilde{G}_1)/(n+2)$. Thus,

$$E(\tilde{G}_n \mid \tilde{G}_{n-1}, \dots, \tilde{G}_1) = \frac{1}{n+2} E(G_n \mid \tilde{G}_{n-1}, \dots, \tilde{G}_1) = \frac{G_{n-1}}{n+1} = \tilde{G}_{n-1}.$$

Finally, for all n , the maximum number of green rabbits in the hat is finite, so $E(|\tilde{G}_n|) < \infty$. It follows that \tilde{G}_n is a **martingale**.

Now note that \tilde{G}_n takes values in the set $[0, 1]$ for all n . Thus, $E(|\tilde{G}_n|^2) \leq 1$ for all n . By the Martingale Convergence Theorem, \tilde{G}_n **converges** to some random variable (which need not be a constant) G w.p.1. However, the theorem does not tell us what that G is. To establish the actual limit, first note that for any $k \in \{1, \dots, n+1\}$, the probability of drawing first $k - 1$ green rabbits and then $n - k$ white rabbits (for a total of k green rabbits and $n + 1 - k$ white rabbits following the n^{th} trial) equals

$$\frac{1}{2} \frac{2}{3} \cdots \frac{k-1}{k} \frac{1}{k+1} \cdots \frac{n-k}{n+1} = \frac{(k-1)!(n-k)!}{(n+1)!}.$$

Now change the order in which we draw the $k - 1$ green rabbits and notice that every drawing sequence of green and white rabbits such that $k - 1$ are green has the above probability. (We simply shuffle the numerator and the denominator accordingly.) There are $\binom{n}{k-1}$ such sequences, and thus,

$$P(G_n = k) = \frac{(k-1)!(n-k)!}{(n+1)!} \binom{n}{k-1} = \frac{1}{n+1},$$

for every value of $k \in \{1, \dots, n+1\}$. Now, for every interval of integer width $(m_1, n_1) \subset [0, 1]$, we therefore have $P(G_n \in (m_1, m_2)) = m_2 - m_1$ (i.e. the width of the interval) for every n . Thus, $\lim_{n \rightarrow \infty} P(G_n \in (m_1, m_2)) = m_2 - m_1$ for every pair of integers $0 \leq m_1 \leq m_2 \leq 1$. It follows that \tilde{G}_n converges *in distribution* to a random variable uniform on $[0, 1]$. But we know that \tilde{G}_n converges w.p.1. Therefore, \tilde{G}_n converges w.p.1 to a random variable uniform on $[0, 1]$. (Technically, we've

only shown that G has a uniform law over intervals of integer width, but it's not too difficult to extend the result to rationals and then irrationals. We did a similar thing in the recitation at the beginning of the semester.)

Note 1: Here's an example of how we could have done this problem *wrong*. Denoting by W_n the analogous quantity (fraction of white rabbits after n^{th} trial), we have that $W_n + G_n = 1$ for all n . Thus, $W + G = 1$ always. But, by the symmetry of the problem, we have that W and G are identically distributed. So far, everything is true, but it does not give us sufficient grounds to conclude that W and G are in fact uniform on the unit interval. It merely tells us that W and G have probability distributions that are symmetric about the mean.

Note 2: Abstracted of rabbits, this problem is referred to as Polya's Urn and its generalizations have consumed a non-trivial amount of mathematical bandwidth. One common generalization is the following: start with an urn containing i_r red balls and i_b black balls. At each time n , one draws a ball, notes its color, puts it back into the urn and adds several more balls according to the following rule: if the drawn ball is red, add a red balls and b black balls; if the drawn ball is black, add c red balls and d black balls (it is common to impose $a + b = c + d$). The question has been attacked through various methods in combinatorics and probability. Here is a small sampling of various fundamentally different approaches to the problem:

- G. Pólya, *Sur quelques points de la théorie des probabilités*, Ann. Inst. Henri Poincaré 1 (1931), 117-161.
- S. Janson, *Functional limit theorem for multitype branching processes and generalized Pólya urns.*, Stochastic Processes and Applications, 110 (2004), 177-245.
- P. Flajolet, J. Gabarró, H. Pekari, *Analytic urns*, The Annals of Probability, 33 (3) (2005), 1200-1233.
- N. Pouyanne *An algebraic approach to Pólya processes*, (2008). Ann. Inst. Henri Poincaré, Vol. 44, No. 2, 293-323.
- H. M. Mahmoud, *Pólya urn models*, CRC Press, (2008)

Problem 5

To show Y_n is a martingale with respect to X_1, \dots, X_n , we need to show two things: that $E(|Y_n|) < \infty$ and that $E(Y_n | X_{n-1}, \dots, X_1) = Y_{n-1}$ for all n . First, by Jensen's inequality, $E(|Y_n|)^2 \leq E(|Y_n|^2)$. But, $E(|Y_n|^2) = E(E(Y | X_n, \dots, X_1)^2) \leq E(E(Y^2 | X_n, \dots, X_1))$ also by Jensen's inequality. By Problem 2 part a), $E(E(Y^2 | X_n, \dots, X_1)) = E(Y^2) < \infty$ (this is why we require Y to have a finite second moment).

Second, note that

$$E(Y_n | X_{n-1}, \dots, X_1) = E(E(Y | X_n, \dots, X_1) | X_{n-1}, \dots, X_1) = E(Y | X_{n-1}, \dots, X_1) = Y_{n-1}$$

where the second equality follows by Problem 2 part a) and the third follows from the definition of Y_n . The martingale in this problem is commonly referred to as *Doob's martingale*, named after Joseph L. Doob.

Problem 6

- a First note that $X_n = Z_n - Z_{n-1}$. Furthermore, $E(Z_n | Z_{n-1}, \dots, Z_1) = E(Z_n | X_{n-1}, \dots, X_1)$ (why?). Thus, for $n \geq 2$,

$$\begin{aligned} E(X_n | X_{n-1}, \dots, X_1) &= E(Z_n - Z_{n-1} | X_{n-1}, \dots, X_1) \\ &= E(Z_n | X_{n-1}, \dots, X_1) - E(Z_{n-1} | X_{n-1}, \dots, X_1) \\ &= E(Z_n | Z_{n-1}, \dots, Z_1) - E(Z_{n-1} | Z_{n-1}, \dots, Z_1). \end{aligned}$$

If Z_n is a martingale, then $E(Z_n | Z_{n-1}, \dots, Z_1) - E(Z_{n-1} | Z_{n-1}, \dots, Z_1) = Z_{n-1} - Z_{n-1} = 0$. Conversely, since $E(Z_{n-1} | Z_{n-1}, \dots, Z_1) = Z_{n-1}$, that $E(Z_n | Z_{n-1}, \dots, Z_1) - E(Z_{n-1} | Z_{n-1}, \dots, Z_1) = 0$ implies that Z_n is a martingale.

- b By a), if Z_n is a martingale, then for $n \geq 2$, $E(X_n | X_{n-1}, \dots, X_1) = 0$. By Problem 2-a), it follows that

$$E(X_n) = E(E(X_n | X_{n-1}, \dots, X_1)) = E(0) = 0.$$

Thus, for all $n \neq m$, $\text{cov}(X_n, X_m) = E(X_n X_m) - E(X_n)E(X_m) = E(X_n X_m)$. Let $n < m$. Then,

$$E(X_n X_m) = E(E(X_n X_m | X_{m-1}, \dots, X_1)) = E(X_n E(X_m | X_{m-1}, \dots, X_n)) = E(X_n \times 0) = 0.$$

It follows that $\text{cov}(X_n, X_m) = 0$.

- c Since $\text{var}(Z_n) = E((\sum_{i=1}^n X_i)^2) - (E(\sum_{i=1}^n X_i))^2$ and since $E(\sum_{i=1}^n X_i) = 0$ when Z_n is a martingale, by b), we have that $\text{var}(Z_n) = E((\sum_{i=1}^n X_i)^2) = E(\sum_{i=1}^n X_i^2) + \sum_{i \neq j} E(X_i X_j)$. But, also by b), $E(X_i X_j) = 0$ for $i \neq j$. It follows that $\text{var}(Z_n) = E(\sum_{i=1}^n X_i^2) = \sum_{i=1}^n \text{var}(X_i)$, again since X_i have zero mean by part b).
- d Since $\text{var}(Z_n) = \sum_{i=1}^n \text{var}(X_i)$ by part c), and since $\text{var}(X_i) \geq 0$ (for any random variable, by Jensen's inequality... right?), it follows that $\text{var}(Z_1) \leq \text{var}(Z_2) \leq \text{var}(Z_3) \dots$. Since $\text{var}(Z_n)$ is non-decreasing, it tends to some limit in \mathbb{R} . The limit of a non-decreasing sequence can be finite (think of $x_n = 1 - 1/n$) or infinite (think of $z_n = n$). What is it here? Well, it depends. Suppose $\text{var}(X_n) = 1/n^2$. What does $\text{var}(Z_n)$ tend to in the limit of large n ? How about for $\text{var}(X_n) = n$? (Of course, we'd have to show that martingales with such variances exist. But, in both cases, it is easy exercise to construct the corresponding increments X_n as, by part a), all we need is that $E(X_n | X_{n-1}, \dots, X_1) = 0$. *Hint*: Make them independent and zero-mean, but not identically distributed.)