MATHEMATICS 335—PROBABILITY—DECEMBER 17–19, 2002

FINAL EXAMINATION SOLUTIONS

- Please complete SIX questions: **ALL of 1, 2, 3, and 4,** and **TWO chosen from 5, 6 and 7.**
- If you submit an extra question, I will flip coins to decide what to grade.
- You must **justify your answers** to receive full credit.
- You may leave answers in terms of binomial coefficients, summations, etc., unless the question specifies that you simplify your answer. (Of course, it may be necessary to simplify or perform other algebraic manipulations on an expression in order to answer another part of the same question.)

1. (18 points) An urn contains n balls, labelled $1, 2, \ldots, n$, respectively. We remove k balls from the urn (without replacement).
   (a) (4 points) Let X be the label of the highest numbered ball removed. Find $P(X = m)$ as a function of n, m, and k.

   Solution. The total number of ways to choose k balls is $\binom{n}{k}$. If the largest ball chosen has label m, then the other $k-1$ balls must be chosen from $1, 2, \ldots, m-1$. There are $\binom{m-1}{k-1}$ ways to do so. Hence, the probability that the largest label on a chosen ball is m is $\binom{m-1}{k-1}/\binom{n}{k}$ for each $m \le n$. (Notice that this value is 0 when $m < k$).
   We get as a corollary that $\sum_{m=k}^{n} \binom{m-1}{k-1} = \binom{n}{k}$, since the sum of all the terms of the distribution function of X must be 1. □

   (b) (5 points) Find $E[X]$.

   Solution 1. By the definition of expected value, this will be

$$\sum_{m=k}^{n} m \frac{\binom{m-1}{k-1}}{\binom{n}{k}} = \frac{1}{\binom{n}{k}} \sum_{m=k}^{n} m \frac{(m-1)!}{(k-1)!(m-k)!} = \frac{k}{\binom{n}{k}} \sum_{m=k}^{n} \frac{(m)!}{(k)!(m-k)!}$$
$$= \frac{k}{\binom{n}{k}} \sum_{m=k}^{n} \binom{m}{k} = \frac{k}{\binom{n}{k}} \binom{n+1}{k+1} = \frac{k(n+1)!k!(n-k)!}{n!(k+1)!(n-k)!} = \frac{k(n+1)}{k+1}.$$

   That summation step is a bit tricky; notice that the terms correspond the the number of subsets of $\{1, \ldots, n+1\}$ with $k+1$ elements and maximal element m, so this is the same computation as in part (a), but with n and k both increased by 1. □

   Solution 2. We use a formula we proved for the expected value of a positive-integer valued random variable, namely that

$$E[X] = \sum_{x=1}^{\infty} P(X \ge x) = \sum_{x=1}^{\infty} (1 - P(X < x)) = \sum_{x=1}^{n} \left(1 - \frac{\binom{x-1}{k}}{\binom{n}{k}}\right)$$
$$= n - \frac{1}{\binom{n}{k}} \sum_{x=k+1}^{n} \binom{x-1}{k} = n - \frac{\binom{n}{k+1}}{\binom{n}{k}} = n - \frac{n!k!(n-k)!}{(k+1)!(n-k-1)!n!}$$
$$= n - \frac{n-k}{k+1} = \frac{nk+n-n+k}{k+1} = \frac{k(n+1)}{k+1}.$$

   Note that $P(X < x)$ is as given above, since $X < x$ exactly when all k balls are chosen from those labelled $1, 2, \ldots, x-1$. Also, as in Solution 1, we've used the identity from part (a) to perform the summation. □

(c) (4 points) Let Y be the sum of the numbers on the k balls removed. Are X and Y independent random variables? Justify your answer.

Solution. In general, no. For example, when $n > k$ and $X = k$, the value of Y is forced to be $1+2+\cdots+k$. Thus $P\left(X = m \text{ and } Y > \frac{k(k+1)}{2}\right) = 0 \neq P(X = m)P\left(Y > \frac{k(k+1)}{2}\right)$. However, it's worth noting that when $n = k$, both X and Y are constant ($X = k$ and $Y = \frac{k(k+1)}{2}$), and thus they are actually independent. □

(d) (5 points) Find $E[Y]$.

Solution 1. Associate a random variable with each ball chosen: $Y = Y_1 + \cdots + Y_k$. Although the Y_i's are not independent (no two can have the same value), they are identically distributed, so $E[Y] = E[Y_1] + \cdots + E[Y_k] = kE[Y_1]$. But then

$$E[Y_1] = \sum_{i=1}^{n} i\left(\frac{1}{n}\right) = \frac{n(n+1)}{2n} = \frac{n+1}{2}, \quad \text{so} \quad E[Y] = \frac{k(n+1)}{2}.$$

□

Solution 2. Associate a random variable Z_i with each ball, so that $Z_1 = 1$ exactly when ball i is chosen. For each i, $P(Z_i = 1) = k/n$, so $E[Z_i] = k/n$. Then $Y = Z_1 + 2Z_2 + \cdots + nZ_n$ implies

$$E[Y] = E[Z_1] + 2E[Z_2] + \cdots + nE[Z_n] = (1 + 2 + \cdots + n)\left(\frac{k}{n}\right) = \frac{k(n+1)}{2}.$$

□

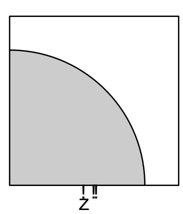
2. (12 points) Let X and Y be independent continuous random variables, each uniform on the interval $(0,1)$. Thus,

$$f_X(x) = \begin{cases} 1 & 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{and similarly for } f_Y.$$

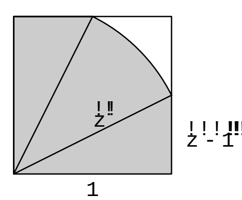
(a) (6 points) Find the density of $X^2 + Y^2$.

Solution 1. Because X and Y are independent, the point (X, Y) is chosen uniformly from the unit square. We can compute the cumulative distribution of $X^2 + Y^2$ geometrically, then differentiate that expression.

$0 \le z \le 1$



$1 \le z \le 2$



When $0 \le z \le 1$, the region satisfying $X^2 + Y^2 \le z$ is a quarter-circle of radius $\sqrt{z}$. When $1 \le z \le 2$, the situation is more complicated: we decompose the region into the union of two right triangles and a circular wedge, as shown above. The triangles each have area $(1/2)\sqrt{z-1}$. The circular wedge has radius $\sqrt{z}$. What is the central angle of the wedge? The angle of each triangle at the origin has cosine $1/\sqrt{z}$, so the remaining angle left for the wedge will be $\pi/2 - 2\arccos(1/\sqrt{z})$. Hence,

$$P(X^2 + Y^2 \le z) = \begin{cases} 0 & z \le 0, \\ \frac{\pi z}{4} & 0 \le z \le 1, \\ \sqrt{z-1} + \frac{z}{2}\left(\frac{\pi}{2} - 2\arccos\left(\frac{1}{\sqrt{z}}\right)\right) & 1 \le z \le 2, \\ 1 & 1 \le z. \end{cases}$$

Differentiation now yields $\quad f_{X^2+Y^2}(z) = \begin{cases} \frac{\pi}{4} & 0 \le z \le 1, \\ \frac{\pi}{4} - \arccos\left(\frac{1}{\sqrt{z}}\right) & 1 \le z \le 2, . \\ 0 & \text{otherwise} \end{cases}$ ☐

Solution 2. First, by the formula we proved about densities of increasing functions of random variables, the density of X^2 is

$$g(x) = \begin{cases} \frac{1}{2\sqrt{x}} & 0 \le \sqrt{x} \le 1 \\ 0 & \text{otherwise,} \end{cases} \qquad \text{and similarly for } Y^2.$$

Because X and Y are independent, so are X^2 and Y^2, and convolving gives the density of $X^2 + Y^2$ as

$$\int_{\max(0,z-1)}^{\min(1,z)} g(x)g(z-x)\,dx = \begin{cases} \int_0^z \frac{1}{4\sqrt{x(z-x)}}\,dx & 0 \le z \le 1, \\ \int_{z-1}^1 \frac{1}{4\sqrt{x(z-x)}}\,dx & 1 \le z \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

By the lemma below, these reduce to

$$f_{X^2+Y^2}(z) = \begin{cases} \frac{1}{4}\left(\arcsin\left(\frac{2z}{z} - 1\right) - \arcsin\left(\frac{2(0)}{z} - 1\right)\right) = \frac{1}{4}(2\arcsin(1)) = \frac{\pi}{4} & 0 \le z \le 1, \\ \frac{1}{4}\left(\arcsin\left(\frac{2(1)}{z} - 1\right) - \arcsin\left(\frac{2(z-1)}{1} - 1\right)\right) = \frac{1}{2}\arcsin\left(\frac{2}{z} - 1\right) & 1 \le z \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

☐

Lemma. $\displaystyle \int \frac{1}{4\sqrt{x(z-x)}}\,dx = \frac{1}{4}\arcsin\left(\frac{2x}{x} - 1\right) + C.$

Proof. Of course, we could just differentiate to verify the given formula... but really: how does one antidifferentiate this expression? Complete the square in the denominator, substitute, and recognize the resulting form. Here we go.

$$\int \frac{1}{4\sqrt{x(z-x)}}\,dx = \frac{1}{4}\int \frac{dx}{\sqrt{\frac{z^2}{4} - \left(\frac{z^2}{4} - zx + x^2\right)}} = \frac{1}{4}\int \frac{dx}{\sqrt{\frac{z^2}{4} - \left(x - \frac{z}{2}\right)^2}}$$

$$= \frac{1}{4}\int \frac{2\,dx/z}{\sqrt{1 - \left(\frac{2x}{z} - 1\right)^2}} = \frac{1}{4}\int \frac{du}{\sqrt{1-u^2}} = \frac{1}{4}\arcsin(u) + C = \frac{1}{4}\arcsin\left(\frac{2x}{z} - 1\right) + C.$$

☐

Remark. Are the two solutions actually giving the same answer? It's clear that they do, except for the case $1 \le z \le 2$. Let $\alpha = \arccos(1/\sqrt{z})$; then $\cos\alpha = 1/\sqrt{z}$. We can compute

$$\sin\left(\frac{\pi}{2} - 2\alpha\right) = \cos(2\alpha) = 2\cos^2\alpha - 1 = \frac{2}{z} - 1.$$

But then

$$\frac{\pi}{2} - 2\alpha = \arcsin\left(\frac{2}{z} - 1\right),$$

and dividing both sides by 2 shows that our two answers are equivalent. (As is always the case with inverse trig functions, one should be worried about issues of unique definition—there are potentially infinitely many values of each angle that work. Those work out well here, and the reasoning above does suffice.)

(b) (6 points) Find $E[X^2 + Y^2]$.

Solution. $E[X^2 + Y^2] = E[X^2] + E[Y^2] = \displaystyle\int_0^1 x^2\,dx + \int_0^1 y^2\,dy = \frac{2}{3}.$ ☐

3. (15 points) A bank accepts rolls of pennies and gives 50 cents credit for each roll to a customer without counting the pennies. Assume that the rolls customers bring in contain 49 pennies 30 percent of the time, 50 pennies 60 percent of the time, and 51 pennies 10 percent of the time. (You may assume that the contents of different rolls are independent of each other.)

   (a) (3 points) Find the expected value and variance of the amount of money the bank *loses* on a typical roll.

   Solution. The expected value of the loss per roll, in cents, is $(0.3)(1) + (0.6)(0) + (0.1)(-1) = 0.2$ (the loss can be negative!) The variance per roll is $(0.3)(1^2)+(0.6)(0^2)+(0.1)(-1^2) - (0.2)^2 = 0.36$. $\square$

   (b) (6 points) Estimate the probability that the bank loses more than 25 cents when 100 rolls are brought in.

   Solution. In 100 rolls, the bank expects to lose $100(0.2) = 20$ cents, with a standard deviation of $\sqrt{100(0.36)} = 6$ cents. Hence the probability that the bank loses more than 25 cents is the probability that we are at most $5/6 \approx 0.8333$ standard deviations above the mean. By the CLT and the table, that's about $1 - 0.7967 = 0.2033$. $\square$

   (c) (6 points) How many rolls does the bank need to collect to have a 99 percent chance of a net loss?

   Solution. We want to find the smallest value of n for which

$$P\left(S_n > 0\right) = P\left(\frac{S_n - (0.2)n}{\sqrt{(0.36)n}} \geq \frac{-(0.2)n}{\sqrt{(0.36)n}}\right) \geq 0.99,$$

   Or, at least, the smallest value of n for which that inequality holds when the exact probability is replaced by its normal approximation. By the table, the 99 percent chance mark is at 2.33 standard deviations from the mean; that is,

$$\int_{-\infty}^{2.33} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \int_{-2.33}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \doteq 0.9901.$$

   Thus, we solve

$$\frac{-(0.2)n}{\sqrt{(0.36)n}} \leq -2.33,$$

   obtaining $n \geq ((3)(2.33))^2 = 48.8601$. The bank has just about a 99 percent chance of losing money when it accepts 49 or more rolls.

   $\square$

4. (15 points) Let $X_1, X_2, \ldots$ be independent and identically distributed Poisson(1) random variables. Thus, for each integer $i \geq 1$ and each integer $k \geq 0$,

$$P(X_i = k) = \frac{e^{-1}}{k!}.$$

   Let $S_n = X_1 + \ldots X_n$.

   (a) (7 points) Prove that $\lim_{n\to\infty} P(S_n > 2n) = 0$.

   Solution. We know that each X_i has expected value 1 and variance 1, so $E[S_n] = Var[S_n] = n$ (in fact, S_n is Poisson(n)). By elementary manipulations and Chebyshev's inequality,

$$\begin{aligned} P(S_n > 2n) &= P(S_n - n > n) \leq P(|S_n - E[S_n]| > n) \\ &\leq \frac{\mathrm{Var}[S_n]}{n^2} = \frac{n}{n^2} = \frac{1}{n} \to 0. \end{aligned}$$

   Since $P(S_n > 2n) \geq 0$ (because it's a probability), we can conclude that, in fact, $\lim_{n\to\infty} P(S_n > 2n) = 0$. $\square$

(b) (8 points) Prove that $\lim_{n\to\infty} P(S_n > n + \sqrt{n}) = \int_1^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$.

Solution. By elementary manipulations and the Central Limit Theorem,

$$P(S_n > n + \sqrt{n}) = P\left(S_n - n > \sqrt{n}\right) = P\left(\frac{S_n - E[S_n]}{\sqrt{n}} > 1\right)$$
$$= P\left(\frac{S_n - E[S_n]}{\mathrm{Var}[S_n]} > 1\right) \to \int_1^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx.$$

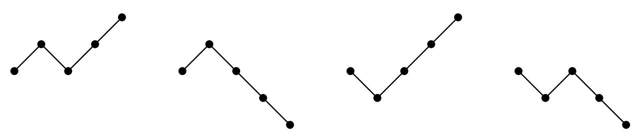
□

5. (20 points) [CORRECTED VERSION OF PROBLEM] Let $X_1, X_2, \ldots$ be independent and identically distributed random variables, each satisfying $P(X_i = 1) = P(X_i = -1) = 1/2$. Let $S_n = X_1 + \cdots + X_n$.

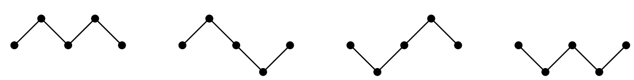
Let r_{2n} be the probability that *exactly one* return to 0 occurs *before* time $2n$, AND $S_{2n} \neq 0$, and let s_{2n} be the probability that $S_{2n} = 0$ and there is *at least one* return to 0 *before* time $2n$.

(a) (5 points) Find r_2, s_2, r_4, and s_4.

Solution. First, r_2 and s_2 are both 0, since there is no even time between 0 and 2 at which a return can take place. Here are the paths for r_4 (and thus $r_4 = 4/16 = 1/4$):



Here are the paths for s_4 (and thus $s_4 = 4/16 = 1/4$):



□

(b) (15 points) Prove that $r_{2n} = s_{2n}$ for $n \geq 2$.

Solution. Consider s_{2n} first. Let $2k$ be the time of the walk's first return to 0; we must have $2 \leq 2k \leq 2k - 2$. Also, the walk from time $2k$ to time $2n$ consists of $2n - 2k$ steps that sum to 0, and that are independent of the first $2k$ steps. Thus

$$s_{2n} = \sum_{k=1}^{n-1} f_{2k} u_{2n-2k}.$$

Now, for r_{2n}. Let $2k$ be the time of the single return to 0; hence, $2 \leq 2k \leq 2n - 2$. The walk after time $2k$ must not return to 0 at or before time $2n$; hence, it is a segment of random walk of length $2n - 2k$ that starts at 0, but does not return to 0. These steps are independent of the first k steps, so we can write

$$r_{2n} = \sum_{k=1}^{n-1} f_{2k}(1 - (f_2 + f_4 + \cdots + f_{2n-2k})) = \sum_{k=1}^{n-1} f_{2k} u_{2n-2k},$$

where the second equality follows from the given identity (which was on your homework). □

You may find it helpful to recall that we proved that when $u_{2n} = P(S_{2n} = 0)$ and f_{2n} is the probability that the first return to 0 occurs at time $2n$ (thus $u_0 = 1$ and $f_0 = 0$), then

$$f_2 + f_4 + \cdots + f_{2n} = 1 - u_{2n} \quad \text{for } n \geq 1.$$

6. (20 points) Consider a circle with M points equally spaced around it, labelled $0, 1, \ldots, M-1$ clockwise around the circle. A caterpillar starts at point 0. At each time, the caterpillar walks to the next point clockwise with probability $1/2$ and to the next point counterclockwise with probability $1/2$.

   (a) (6 points) Find the probability that the caterpillar is at 0 at time t, in terms of M and t. (Your answer may be left as a summation.)

   Solution. First, we pretend that the caterpillar is walking not on a circle, but on a line. Any path on the circle can be "lifted" to a path on the line: turn clockwise steps into up steps, and counterclockwise steps into down steps. The probability that at time t a caterpillar walking on a line is at position k is $\frac{1}{2^t}\binom{t}{\frac{t-k}{2}}$, as long as $(t-k)/2$ is an integer, and zero otherwise. (We showed this in class; the key is that there must be exactly k more +1's than -1's in the sequence).
   However, the caterpillar is on a circle. If the caterpillar's position on the corresponding walk on the line is any multiple of M, then his position on the circle is 0. Thus, we must sum:

$$P(\text{caterpillar at 0 at time } t) = \sum_k \frac{1}{2^t}\binom{t}{\frac{t-k}{2}},$$

   where the sum is taken over all values of k such that $M|k$ and $\frac{t-k}{2}$ is an integer. □

   (b) (14 points) [CORRECTED VERSION OF PROBLEM] Show that the probability that the caterpillar visits all the points of the circle before his first return to 0 is $1/(M-1)$.

   Solution. At the first step, the caterpillar goes either to 1 (with probability 1/2) or $M-1$ (with probability 1/2).
   We first consider the case where the caterpillar steps to 1. In order to visit all the states before returning to 0, it is both necessary and sufficient that the caterpillar reach $M-1$ before returning to 0. Necessary, because $M-1$ is one of the states we are requiring that he visit; sufficient, because, in order to go from 1 to $M-1$ without hitting 0, the caterpillar must walk through $2, 3, \ldots, M-2$—all the remaining states of the circle. Hence, we can consider the caterpillar's sojourn from 1 until the first visit to either 0 (failure) or $M-1$ (success) to be a gambler's ruin situation, with fair odds, maximum fortune $M-1$, and starting with fortune 1. By the result cited below (which we proved in lecture), the caterpillar has probability $1/(M-1)$ of hitting $M-1$ befpre hitting 0.
   If the caterpillar goes to $M-1$ in his first step, the same reasoning goes through, with the points 1 and $M-1$ switched throughout. Thus, the total probability that the caterpillar visits all states before returning to 0 is

$$\frac{1}{2}\left(\frac{1}{M-1}\right) + \frac{1}{2}\left(\frac{1}{M-1}\right) = \frac{1}{M-1}.$$

   □

   You may find it helpful to recall that in a "fair" gambler's ruin situation (i.e. one in which the individual bets are equally likely to be won by each player) with maximum fortune M, the probability that a player starting with k dollars ultimately wins is k/M for $0 \le k \le M$.

7. (20 points) Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be two sequences of independent and identically distributed random variables, each of which is equally likely to be 1 or -1. Thus, for each integer $i \ge 1$ and each integer $j \ge 1$,

$$P(X_1 = 1) = P(X_i = -1) = P(Y_j = 1) = P(Y_j = -1) = 1/2.$$

   We also assume that the two sequences are independent of each other.

   Let $S_n = X_1 + \cdots + X_n$ and let $T_n = Y_1 + \cdots + Y_n$. The pair (S_n, T_n) can be viewed as performing a random walk in the plane as n increases. (We take $S_0 = 0$ and $T_0 = 0$, so the walk starts at the origin.)

(a) (3 points) Determine all possible values of the pair (S_1, T_1) and the probability that each occurs.

Solution. The possible values are $(1, 1)$, $(1, -1)$, $(-1, 1)$, and $(-1, -1)$. Each occurs with probability $1/4$, since, by independence,

$$P((S_1, T_1) = (i, j)) = P(S_1 = i \text{ and } T_1 = j) = P(X_1 = i)P(Y_1 = j).$$

□

(b) (3 points) Show that this random walk can only return to the origin at even times.

Solution. The pair (S_k, T_k) is equal to $(0, 0)$ when it is true that $S_k = 0$ *and* $T_k = 0$. Since both of those can only happen when k is even (as we discussed in class: the number of positive steps must equal the number of negative steps, so the total number of steps must be even), the ordered pair can only be $(0, 0)$ when k is even. □

(c) (6 points) Let $w_{2n} = P((S_{2n}, T_{2n}) = (0, 0))$. Show that $\quad w_{2n} \sim \dfrac{1}{\pi n}$.

Solution. We proved in class that $u_{2n} \sim \dfrac{1}{\sqrt{\pi n}}$ as $n \to \infty$. Because they depend on disjoint sets of independent underlying steps, S_{2n} and T_{2n} are themselves independent. Thus,

$$P\left((S_{2n}, T_{2n} = (0, 0)\right) = P(S_{2n} = 0)P(T_{2n} = 0) \sim \frac{1}{\sqrt{\pi n}}\left(\frac{1}{\sqrt{\pi n}}\right) = \frac{1}{\pi n}.$$

□

(d) (8 points) Is the probability that this walk eventually returns to the origin 1, or is it less than 1? Justify your answer.

Solution 1. Define u^*_{2n}, $U^*(z)$, f^*_{2n}, and $F^*(z)$ to be this walk's version of the quantities and generating functions analogous to those for 1-dimensional random walk (which we wrote without the stars). Exactly as before, it will be true that

$$F^*(z) = \frac{U^*(z) - 1}{U^*(z)},$$

so that the probability of eventual return, which is $F^*(1)$, will be equal to 1 if and only if $U^*(1)$ diverges.

However, in part (c), we showed that $u^*_{2n} \sim \frac{1}{\pi n}$, and $\sum_{n=1}^{\infty} \frac{1}{\pi n}$ diverges. By the comparison test, so must $\sum_{n=1}^{\infty} u^*_{2n} = U^*(1)$. Thus, $F^*(1) = 1$, and this walk returns to the origin with probability 1. □

Solution 2. This walk is actually isomorphic to the two-dimensional random walk we studied earlier (geometrically, via a rotation through $\pi/4$ radians and a magnification by $\sqrt{2}$). Even though we step in both dimensions at every time (instead of picking one dimension, then picking a direction in that dimension, as we did earlier), our unit steps are still all the same length and consist of perpendicular pairs: NE/SW and NW/SE instead of left/right and up/down.

Thus, the probability that we return to the origin is the same as it was for the 2D RW we looked at earlier—which is 1. □