MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Fall 2011)

Problem Set 4 Solutions
Due October 5, 2011

1. (a) From the joint PMF, there are six (x, y) coordinate pairs with nonzero probabilities of occurring. These pairs are (1, 1), (1, 3), (2, 1), (2, 3), (4, 1), and (4, 3). The probability of a pair is proportional to the sum of the squares of the coordinates of the pair, $x^2 + y^2$. Because the probability of the entire sample space must equal 1, we have:

$$(1+1)^2c + (1+3)^2c + (2+1)^2c + (2+3)^2c + (4+1)^2c + (4+3)^2c = 1.$$

Solving for c, we get $c = \boxed{\frac{1}{128}}$.

(a) There are three sample points for which $y < x$:

$$\mathbf{P}(Y < X) = \mathbf{P}(\{(2,1)\}) + \mathbf{P}(\{(4,1)\}) + \mathbf{P}(\{(4,3)\}) = \frac{9}{128} + \frac{25}{128} + \frac{49}{128} = \boxed{\frac{83}{128}}.$$

(b) There are two sample points for which $y > x$:

$$\mathbf{P}(Y > X) = \mathbf{P}(\{(1,3)\}) + \mathbf{P}(\{(2,3)\}) = \frac{16}{128} + \frac{25}{128} = \boxed{\frac{41}{128}}.$$

(c) There is only one sample point for which $y = x$:

$$\mathbf{P}(Y = X) = \mathbf{P}(\{(1,1)\}) = \boxed{\frac{4}{128}}.$$

Notice that, using the above two parts,

$$\mathbf{P}(Y < X) + \mathbf{P}(Y > X) + \mathbf{P}(Y = X) = \frac{83}{128} + \frac{41}{128} + \frac{4}{128} = 1$$

as expected.

(d) There are three sample points for which $y = 3$:

$$\mathbf{P}(Y = 3) = \mathbf{P}(\{(1,3)\}) + \mathbf{P}(\{(2,3)\}) + \mathbf{P}(\{(4,3)\}) = \frac{16}{128} + \frac{25}{128} + \frac{49}{128} = \boxed{\frac{90}{128}}.$$

(e) In general, for two discrete random variable X and Y for which a joint PMF is defined, we have

$$p_X(x) = \sum_{y=-\infty}^{\infty} p_{X,Y}(x,y) \qquad \text{and} \qquad p_Y(y) = \sum_{x=-\infty}^{\infty} p_{X,Y}(x,y).$$

In this problem the ranges of X and Y are quite restricted so we can determine the marginal PMFs by enumeration. For example,

$$p_X(2) = \mathbf{P}(\{(2,1)\}) + \mathbf{P}(\{(2,3)\}) = \frac{34}{128}.$$

Overall, we get:

$$p_X(x) = \begin{cases} 20/128, & \text{if } x = 1, \\ 34/128, & \text{if } x = 2, \\ 74/128, & \text{if } x = 4, \\ 0, & \text{otherwise} \end{cases} \qquad \text{and} \qquad p_Y(y) = \begin{cases} 38/128, & \text{if } y = 1, \\ 90/129, & \text{if } y = 3, \\ 0, & \text{otherwise.} \end{cases}$$

(f) In general, the expected value of any discrete random variable X equals

$$\mathbf{E}[X] = \sum_{x=-\infty}^{\infty} x p_X(x).$$

For this problem,

$$\mathbf{E}[X] = 1 \cdot \frac{20}{128} + 2 \cdot \frac{34}{128} + 4 \cdot \frac{74}{128} = \boxed{3}$$

and

$$\mathbf{E}[Y] = 1 \cdot \frac{38}{128} + 3 \cdot \frac{90}{128} = \boxed{\tfrac{77}{32}}.$$

To compute $\mathbf{E}[XY]$, note that $p_{X,Y}(x,y) \neq p_X(x)p_Y(y)$. Therefore, X and Y are not independent and we cannot assume $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$. Thus, we have

$$\begin{aligned}\mathbf{E}[XY] &= \sum_x \sum_y xy p_{X,Y}(x,y) \\ &= 1 \cdot \frac{4}{128} + 2 \cdot \frac{9}{128} + 4 \cdot \frac{25}{128} + 3 \cdot \frac{16}{128} + 6 \cdot \frac{25}{128} + 12 \cdot \frac{49}{128} = \boxed{\tfrac{227}{32}}.\end{aligned}$$

(g) The variance of a random variable X can be computed as $\mathbf{E}[X^2] - \mathbf{E}[X]^2$ or as $\mathbf{E}[(X - \mathbf{E}[X])^2]$. We use the second approach here because X and Y take on such limited ranges. We have

$$\operatorname{var}(X) = (1-3)^2 \frac{20}{128} + (2-3)^2 \frac{34}{128} + (4-3)^2 \frac{74}{128} = \boxed{\tfrac{47}{32}}$$

and

$$\operatorname{var}(Y) = (1 - \frac{7}{3})^2 \frac{38}{128} + (3 - \frac{7}{3})^2 \frac{90}{128} = \boxed{\tfrac{855}{1024}}.$$

X and Y are not independent, so we cannot assume $\operatorname{var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y)$. The variance of $X+Y$ will be computed using $\operatorname{var}(X+Y) = \mathbf{E}[(X+Y)^2] - (\mathbf{E}[X+Y])^2$. Therefore, we have

$$\mathbf{E}[(X+Y)^2] = 4 \cdot \frac{4}{128} + 9 \cdot \frac{9}{128} + 25 \cdot \frac{25}{128} + 16 \cdot \frac{16}{128} + 25 \cdot \frac{25}{128} + 49 \cdot \frac{49}{128} = \frac{1001}{32}.$$

$$(\mathbf{E}[X+Y])^2 = (\mathbf{E}[X] + \mathbf{E}[Y])^2 = \left(3 + \frac{77}{32}\right)^2 = 29.22.$$

Therefore,

$$\operatorname{var}(X+Y) = \frac{1001}{32} - 29.22 = \boxed{\tfrac{2103}{1024}}.$$

(h) There are four (x, y) coordinate pairs in A : (1,1), (2,1), (4,1), and (4,3). Therefore, $\mathbf{P}(A) = \frac{1}{128}(4 + 9 + 25 + 49) = \frac{87}{128}$. To find $\mathbf{E}[X \mid A]$ and $\operatorname{var}(X \mid A)$, $p_{X|A}(x)$ must be calculated. We have

$$p_{X|A}(x) = \begin{cases} 4/87, & \text{if } x = 1, \\ 9/87, & \text{if } x = 2, \\ 74/87, & \text{if } x = 4, \\ 0, & \text{otherwise,} \end{cases}$$

$$
\begin{aligned}
\mathbf{E}[X \mid A] &= 1 \cdot \frac{4}{87} + 2 \cdot \frac{9}{87} + 4 \cdot \frac{74}{87} = \boxed{\tfrac{106}{29}}, \\
\mathbf{E}[X^2 \mid A] &= 1^2 \cdot \frac{4}{87} + 2^2 \cdot \frac{9}{87} + 4^2 \cdot \frac{74}{87} = \frac{408}{29}, \\
\mathrm{var}(X \mid A) &= \mathbf{E}[X^2 \mid A] - (\mathbf{E}[X \mid A])^2 = \frac{408}{29} - \left(\frac{106}{29}\right)^2 = \boxed{\tfrac{596}{841}},
\end{aligned}
$$

2. (a) We know that I_A is a random variable that maps a 1 to the real number line if ω occurs within an event A and maps a 0 to the real number line if ω occurs outside of event A. A similar argument holds for event B. Thus we have,

$$I_A(\omega) = \begin{cases} 1, & \text{with probability } \mathbf{P}(A) \\ 0, & \text{with probability } 1 - \mathbf{P}(A) \end{cases}$$

$$I_B(\omega) = \begin{cases} 1, & \text{with probability } \mathbf{P}(B) \\ 0, & \text{with probability } 1 - \mathbf{P}(B) \end{cases}$$

If the random variables, A and B, are independent, we have $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. The indicator random variables, I_A and I_B, are independent if, $p_{I_A,I_B}(x, y) = p_{I_A}(x)p_{I_B}(y)$
We know that the intersection of A and B yields.

$$
\begin{aligned}
p_{I_A,I_B}(1,1) &= p_{I_A}(1)p_{I_B}(1) \\
&= \mathbf{P}(A)\mathbf{P}(B) \\
&= \mathbf{P}(A \cap B)
\end{aligned}
$$

We also have,

$$
\begin{aligned}
p_{I_A,I_B}(1,1) &= \mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B) = p_{I_A}(1)p_{I_B}(1) \\
p_{I_A,I_B}(0,1) &= \mathbf{P}(A^c \cap B) = \mathbf{P}(A^c)\mathbf{P}(B) = p_{I_A}(0)p_{I_B}(1) \\
p_{I_A,I_B}(1,0) &= \mathbf{P}(A \cap B^c) = \mathbf{P}(A)\mathbf{P}(B^c) = p_{I_A}(1)p_{I_B}(0) \\
p_{I_A,I_B}(0,0) &= \mathbf{P}(A^c \cap B^c) = \mathbf{P}(A^c)\mathbf{P}(B^c) = p_{I_A}(0)p_{I_B}(0)
\end{aligned}
$$

(b) If $X = I_A$, we know that

$$\mathbf{E}[X] = \mathbf{E}[I_A] = 1 \cdot \mathbf{P}(A) + 0 \cdot (1 - \mathbf{P}(A)) = \mathbf{P}(A)$$

3. (a) An easy way to derive $p_{X,Y,Z}(x, y, z)$ is in sequential terms as $p_X(x) \cdot p_{Y,Z|X}(y, z|x)$. Note $p_X(x)$ is geometric with parameter p. Conditioned on X even, $(Y, Z) = (0, 0)$ with probability 1. Conditioned on X odd, $p_{Y,Z|X}(y, z) = \frac{1}{4}$ for $(y, z) \in \{(0, 0), (0, 2), (2, 0), (2, 2)\}$.

$$p_{X,Y,Z}(x, y, z) = \begin{cases} \frac{1}{4}p(1-p)^{x-1}, & \text{if } x \text{ is odd and } (y, z) \in \{(0, 0), (0, 2), (2, 0), (2, 2)\} \\ p(1-p)^{x-1}, & \text{if } x \text{ is even and } (y, z) = (0, 0) \\ 0, & \text{otherwise.} \end{cases}$$

(b) (i) No. Notice that even though conditional on X (i.e. given a realization, x, of random variable X), the random variables Y and Z are independent (that's why they look "regular"), Y and Z are not independent. Given Y, the distribution over Z changes (i.e. if Y is 2, Z is equally likely to be 0 or 2; however if Y is 0, Z is more likely to be 0).

(ii) Yes. Given $Z = 2$, if we are further given $X = x$, Y is equally likely to take on the value 0 or 2.

(iii) No. Given $Z = 0$, if we are further given $X = x$, then if x is even, Y must be 0, whereas if x is odd, Y is equally likely to take on 0 or 2.

(iv) Yes. Given $Z = 2$, if we are further given $X = x$, $Z = 2$ still holds (i.e. with probability 1)! Double conditioning has no effect.

(c)

$$
\begin{aligned}
\mathbf{E}[X|Y=2] &= \sum_x x p_{X|Y}(x|2) \\
&= \sum_{x \text{ odd}} x p_{X|Y}(x|2) + \sum_{x \text{ even}} x p_{X|Y}(x|2) \\
&= \sum_{x \text{ odd}} x p_{X|Y}(x|2) \\
&= \sum_{x \text{ odd}} x \frac{p_{X,Y}(x,2)}{p_Y(2)} \\
&= \sum_{x \text{ odd}} x \frac{2 \cdot \frac{1}{4} p(1-p)^{x-1}}{p_Y(2)} \\
&= \frac{2p/4}{p_Y(2)} \sum_{x \text{ odd}} x(1-p)^{x-1} \\
\text{Note } p_Y(2) &= \sum_{x \text{ odd}} 2 \cdot \frac{1}{4} p(1-p)^{x-1} = \frac{p}{2}(1-(1-p)^2) \\
\text{Thus, } \mathbf{E}[X|Y=2] &= (2-p)p \sum_{x \text{ odd}} x(1-p)^{x-1}
\end{aligned}
$$

(d) If $X = 5$, then Y and Z are uniformly distributed on the set S specified in the problem statement, so $Y + Z$ takes the values 0 and 4 with probability $\frac{1}{4}$, and takes the value 2 with probability $\frac{1}{2}$. This PMF is symmetric about 2, so the mean value of $Y + Z$ is evidently 2. Hence the variance is

$$(0-2)^2 \frac{1}{4} + (4-2)^2 \frac{1}{4} = 2.$$

4. (a) Let L_i be the event that Joe played the lottery on week i, and let W_i be the event that he won on week i. We are asked to find

$$\mathbf{P}(L_i \mid W_i^c) = \frac{\mathbf{P}(W_i^c \mid L_i)\mathbf{P}(L_i)}{\mathbf{P}(W_i^c \mid L_i)\mathbf{P}(L_i) + \mathbf{P}(W_i^c \mid L_i^c)\mathbf{P}(L_i^c)} = \frac{(1-q)p}{(1-q)p + 1 \cdot (1-p)} = \frac{p - pq}{1 - pq}.$$

(b) Conditioned on $X = x$, the random variable Y is binomial. That is, for $0 \leq x \leq n$:

$$p_{Y|X}(y \mid x) = \begin{cases} \begin{pmatrix} x \\ y \end{pmatrix} q^y(1-q)^{(x-y)}, & 0 \leq y \leq x; \\ 0, & \text{otherwise.} \end{cases}$$

(c) Realizing that X has a binomial PMF, we have

$$\begin{aligned} p_{X,Y}(x,y) &= p_{Y|X}(y \mid x)p_X(x) \\ &= \begin{cases} \begin{pmatrix} x \\ y \end{pmatrix} q^y(1-q)^{(x-y)} \begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{(n-x)}, & 0 \leq y \leq x \leq n; \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

(d) Using the result from (c), we could compute

$$p_Y(y) = \sum_{x=y}^{n} p_{X,Y}(x,y) \ ,$$

but the algebra is messy. An easier method is to realize that Y is just the sum of n independent Bernoulli random variables, each having a probability pq of being 1. Therefore Y has a binomial PMF:

$$p_Y(y) = \begin{cases} \begin{pmatrix} n \\ y \end{pmatrix} (pq)^y(1-pq)^{(n-y)}, & 0 \leq y \leq n; \\ 0, & \text{otherwise.} \end{cases}$$

(e)

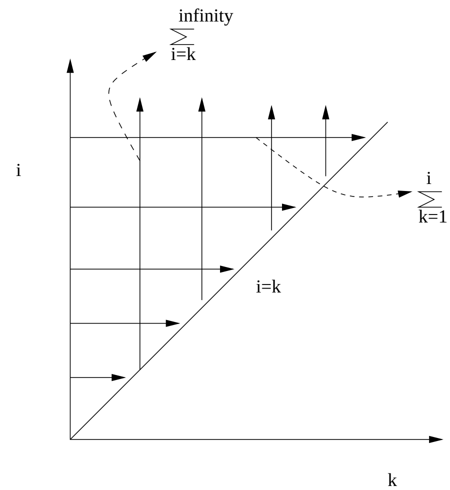
$$\begin{aligned} p_{X|Y}(x \mid y) &= \frac{p_{X,Y}(x,y)}{p_Y(y)} \\ &= \begin{cases} \dfrac{\begin{pmatrix} x \\ y \end{pmatrix} q^y(1-q)^{(x-y)} \begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{(n-x)}}{\begin{pmatrix} n \\ y \end{pmatrix} (pq)^y(1-pq)^{(n-y)}}, & 0 \leq y \leq x \leq n; \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

(f) Given $Y = y$, we know that Joe played y weeks with certainty. For each of the remaining $n - y$ weeks that Joe did not win there are $x - y$ weeks where he played. Each of these events occured with probability $\mathbf{P}(L_i \mid W_i^c)$ (the answer from part (a)). Using this logic we see that that X conditioned on Y is binomial:

$$p_{X|Y}(x \mid y) = \begin{cases} \begin{pmatrix} n-y \\ x-y \end{pmatrix} \left(\frac{p-pq}{1-pq}\right)^{x-y} \left(1 - \frac{p-pq}{1-pq}\right)^{n-x}, & 0 \leq y \leq x \leq n; \\ 0, & \text{otherwise.} \end{cases}$$

After some algebraic manipulation, the answer to (e) can be shown to be equal to the one above.

5. (a) The picture below illustrates the double sum needed to prove the statement of this problem:



We first note that

$$\mathbf{P}(X \geq k) = \sum_{i=k}^{\infty} p_X(i)$$

and proceed as follows:

$$\sum_{k=1}^{\infty} \mathbf{P}(X \geq k) = \sum_{k=1}^{\infty} \sum_{i=k}^{\infty} p_X(i) = \sum_{i=1}^{\infty} \sum_{k=1}^{i} p_X(i) = \sum_{i=1}^{\infty} i\, p_X(i) = \mathbf{E}[X].$$

(b) We first compute

$$\mathbf{P}(Y \geq k) = \begin{cases} 1 & k \leq a \\ \frac{b-k+1}{b-a+1} & a+1 \leq k \leq b \\ 0 & k \geq b+1 \end{cases}$$

So

$$\begin{aligned} \sum_{k=1}^{\infty} \mathbf{P}(Y \geq k) &= \sum_{k=1}^{a} 1 + \sum_{k=a+1}^{b} \frac{b-k+1}{b-a+1} \\ &= a + \frac{1}{b-a+1} \sum_{k=1}^{b-a} k \\ &= a + \frac{1}{b-a+1} \frac{(b-a+1)(b-a)}{2} \\ &= a + \frac{b-a}{2} \\ &= \frac{b+a}{2} \end{aligned}$$

Therefore $\mathbf{E}[Y] = \frac{b+a}{2}$.

G1†. (a) i. Die A beats die B when die A rolls a 4. This occurs with probability 4/6.

†Required for 6.431; optional for 6.041

ii. Die B beats die C when die C rolls a 2. This occurs with probability 4/6.

iii. Die C beats die D when die C rolls a 2 and die D rolls a 1, which occurs with probability $(4/6)(3/6) = 1/3$, or if die C rolls a 6, which occurs with probability 2/6. Therefore, the total probability that C beats D is 4/6.

iv. Die D beats die A when die D rolls a 1 and die A rolls a 0, which occurs with probability $(3/6)(2/6) = 1/6$, or if die D rolls a 5, which occurs with probability 3/6. Therefore, the total probability that D beats A is 4/6.

(b) The strategy to ensure a 2/3 probability of winning is to choose the die "below" my die. In other words, if I choose A then you choose D, if I choose B then you choose A, etc.

G2†. (a) There are only two possible values of W_n: it is 2^n if heads was flipped all n times, and otherwise it is 0 (i.e., if there was at least one tails out of the n flips). The pmf of W_n is

$$p_{W_n}(w) = \begin{cases} \frac{1}{2^n}, & \text{if } w = 2^n, \\ 1 - \frac{1}{2^n}, & \text{if } w = 0, \\ 0, & \text{otherwise.} \end{cases}$$

(b) From the pmf, we see that the expected wealth after n tosses is

$$\mathbf{E}[W_n] = 2^n \cdot \frac{1}{2^n} = 1.$$

(c)

$$\lim_{n\to\infty} \mathbf{P}(W_n = 0) = \lim_{n\to\infty} 1 - \frac{1}{2^n} = 1.$$

(d) As n increases, the probability of getting heads repeatedly with no tails becomes exponentially small and reaches 0 in the limit. The expected wealth at any n, however, is a constant 1. This seems counterintuitive because the expected value in this case is inconsistent with the result you are likely to get as you play the game, which would be to lose all of your money. The reason for this is the exponential increase in rewards, which balances the exponential decrease in the likeliness of winning and getting those rewards. The result is a positive constant expected value of wealth.

(e)

$$\begin{aligned} \text{var}(W_n) &= \mathbf{E}\Big[(W_n - \mathbf{E}[W_n])^2\Big] \\ &= \mathbf{E}[W_n^2] - (\mathbf{E}[W_n])^2 \\ &= \mathbf{E}[W_n^2] - 1 \\ &= \frac{1}{2^n} \cdot (2^n)^2 - 1 \\ &= 2^n - 1 \\ \sigma_{W_n} &= \sqrt{\text{var}(W_n)} \\ &= \sqrt{2^n - 1} \\ &\approx 2^{n/2} \end{aligned}$$

†Required for 6.431; optional for 6.041