UC BERKELEY, CS 174: COMBINATORICS AND DISCRETE PROBABILITY (FALL 2010)

Solutions to Problem Set 2

1. **(MU 2.4; Jensen's Inequality)** Prove that $\mathbb{E}[X^k] \geq \mathbb{E}[X]^k$ for any even integer $k \geq 1$.

   By Jensen's inequality, $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$ for any convex function f. If f is twice differentiable and its second derivative is non-negative, then f is convex. For $f(x) = x^k$, the second derivative is $f''(x) = k(k-1)x^{k-2}$ which is non-negative if $x \geq 0$.

2. **(MU 2.7)** Let X and Y be independent geometric random variables, where X has parameter p and Y has parameter q.

   (a) What is the probability that $X = Y$?

   $$\begin{aligned} \mathbb{P}[X = Y] &= \sum_x (1-p)^{x-1} p (1-q)^{x-1} q \\ &= \sum_x [(1-p)(1-q)]^{x-1} pq \end{aligned}$$

   Recall that from page 31, for geometric random variables, we have the identity

   $$\mathbb{P}[X \geq i] = \sum_{n=i}^{\infty} (1-p)^{n-1} p = (1-p)^{i-1}. \tag{1}$$

   So, we obtain

   $$\mathbb{P}[X = Y] = \frac{pq}{p+q-pq}$$

   (b) What is $\mathbb{E}[\max(X,Y)]$? We know from problem MU 2.9 that $\mathbb{E}[\max(X,Y)] = \mathbb{E}[X] + \mathbb{E}[Y] - \mathbb{E}[\min(X,Y)]$. From below, in part (c), we know that $\min(X,Y)$ is a geometric random variable mean $p+q-pq$. Therefore, $\mathbb{E}[\min(X,Y)] = \frac{1}{p+q-pq}$, and we get

   $$\mathbb{E}[\max(X,Y)] = \frac{1}{p} + \frac{1}{q} - \frac{1}{p+q-pq}.$$

   (c) What is $\mathbb{P}[\min(X,Y) = k]$? We split this event into two disjoint events.

   $$\begin{aligned} \mathbb{P}[\min(X,Y) = k] &= \mathbb{P}[X = k, Y \geq k] + \mathbb{P}[X > k, Y = k] \\ &= \mathbb{P}[X = k]\mathbb{P}[Y \geq k] + \mathbb{P}[X > k]\mathbb{P}[Y = k] \end{aligned}$$

   Recall the identity in Eqn 1. So we have $\mathbb{P}[X > k] = \mathbb{P}[X \geq k] - \mathbb{P}[x = k] = (1-p)^{k-1}(1-p)$. Finally, we get

   $$\begin{aligned} \mathbb{P}[\min(X,Y) = k] &= (1-p)^{k-1} p (1-q)^{k-1} + (1-p)^{k-1}(1-p)(1-q)^{k-1} q \\ &= [(1-p)(1-q)]^{k-1}(p + (1-p)q) \\ &= [(1-p)(1-q)]^{k-1}(p + q - pq) \end{aligned}$$

(d) What is $\mathbb{E}[X|X \leq Y]$?

$$\begin{aligned}\mathbb{E}[X|X \leq Y] &= \sum_{x \geq 1} x\mathbb{P}[X = x|x \leq Y] \\ &= \sum_{x} x\frac{\mathbb{P}[X = x \cap x \leq Y]}{\mathbb{P}[X \leq Y]}\end{aligned}$$

First, let's consider the denominator.

$$\begin{aligned}\mathbb{P}[X \leq Y] &= \sum_{z \geq 1} \mathbb{P}[X = z \cap z \leq Y] \\ &= \sum_{z} \mathbb{P}[X = z]\mathbb{P}[z \leq Y] \\ &= \sum_{z} (1-p)^{z-1}p(1-q)^{z-1} \\ &= \sum_{z} [(1-p)(1-q)]^{z-1}p \\ &= p\sum_{z} [(1-p-q+pq)]^{z-1} \\ &= \frac{p}{p+q-pq}\end{aligned}$$

The last step above is again by the identity in Eqn 1. Now we can compute the whole equation.

$$\begin{aligned}\mathbb{E}[X|X \leq Y] &= \frac{p+q-pq}{p}\sum_{x} x\mathbb{P}[X = x]\mathbb{P}[x \leq Y] \\ &= \frac{p+q-pq}{p}\sum_{x} x(1-p)^{x-1}p(1-q)^{x-1} \\ &= (p+q-pq)\sum_{x} x(1-p-q+pq)^{x-1}\end{aligned}$$

This is equal to the expectation of a geometric random variable with mean $p + q - pq$. Therefore

$$\mathbb{E}[X|X \leq Y] = \frac{1}{p+q-pq}.$$

3. **(MU 2.9; Linearity of expectation)**

(a) Suppose that we roll twice a fair k-sided die with the numbers 1 through k on the die's faces, obtaining values X_1 and X_2. What is $\mathbb{E}[\max(X_1, X_2)]$? What is $\mathbb{E}[\min(X_1, X_2)]$?

$$
\begin{aligned}
\mathbb{E}[\max(X_1, X_2)] &= \sum_{x_1} \sum_{x_2} \max(x_1, x_2)(1/k)(1/k) \\
&= 1/k^2 \sum_{x_1} \sum_{x_2 \leq x_1} x_1 + \sum_{x_2 > x_1} x_2
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[\min(X_1, X_2)] &= \sum_{x_1} \sum_{x_2} \min(x_1, x_2)(1/k)(1/k) \\
&= 1/k^2 \sum_{x_1} \sum_{x_2 \leq x_1} x_2 + \sum_{x_2 > x_1} x_1
\end{aligned}
$$

(b) Show from your calculation in part (a) that

$$
\mathbb{E}[\max(X_1, X_2)] + \mathbb{E}[\min(X_1, X_2)] = \mathbb{E}[X_1] + \mathbb{E}[X_2]. \tag{2}
$$

$$
\begin{aligned}
\mathbb{E}[\max(X_1, X_2)] + \mathbb{E}[\min(X_1, X_2)] &= 1/k^2 \sum_{x_1} \sum_{x_2 \leq x_1} x_1 + \sum_{x_2 > x_1} x_2 + 1/k^2 \sum_{x_1} \sum_{x_2 \leq x_1} x_2 + \sum_{x_2 > x_1} x_1 \\
&= 1/k^2 \sum_{x_1} \left(\sum_{x_2} x_2 + \sum_{x_2} x_1 \right) \\
&= 1/k^2 \sum_{x_1} \sum_{x_2} x_2 + x_1 \\
&= \mathbb{E}[x_1] + \mathbb{E}[x_2]
\end{aligned}
$$

(c) Explain why Eqn. (2) must be true by using the linearity of expectations instead of a direct computation.

By using linearity of expectation twice, we get

$$
\begin{aligned}
\mathbb{E}[\max(X_1, X_2)] + \mathbb{E}[\min(X_1, X_2)] &= \mathbb{E}[\max(X_1, X_2) + \min(X_1, X_2)] \\
&= \mathbb{E}[X_1 + X_2] \\
&= \mathbb{E}[X_1] + \mathbb{E}[X_2].
\end{aligned}
$$

4. **(MU 2.15; Coupon Collector)** For a coin that comes up heads independently with probability p on each flip, what is the expected number of flips until the kth heads?

Let X be the number of flips until the kth heads. Let X_i be the number of coin flips for the next heads.

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[\sum_{i=1}^{k} X_i] \\
&= \sum_{i=1}^{k} \mathbb{E}[X_i] \quad \text{(by linearity of expectation)} \\
&= \sum_{i=1}^{k} 1/p \quad \text{(since } X_i \sim \text{geom(p))} \\
&= k/p
\end{aligned}$$

5. **(MU 2.18; Induction)** The following approach is often called *reservoir sampling*. Suppose we have a sequence of items passing by one at a time. We want to maintain a sample of one item with the property that it is uniformly distributed over all the items that we have seen at each step. Moreover, we want to accomplish this without knowing the total number of items in advance or storing all of the items that we see.

   Consider the following algorithm, which stores just one item in memory at all times. When the first item appears, it is stored in the memory. When the kth item appears, it replaces the item in memory with probability $1/k$. Explain why this algorithm solves the problem.

   Let $b_1, b_2, ...b_n$ be the the values of the items observed at time b_t. We will prove this by induction. Let M_t be a random variable that takes the value of the item in memory at time t. We need to show that at time t, $\mathbb{P}[M_t = b_i] = 1/t$ for all $1 \leq i \leq t$.

   The base case is when $t = 1$, which is trivially true since $M_t = b_1$ with probability 1. Assume that at time t, $\mathbb{P}[M_t = b_i] = 1/t$ for all $1 \leq i \leq t$. Now we prove that this property holds for time $t+1$. At time $t+1$, we set $M_{t+1} = b_{t+1}$ with probability $1/(t+1)$. Therefore, $\mathbb{P}[M_{t+1} = b_{t+1}] = 1/(t+1)$. For $1 \leq i \leq t$,

$$\begin{aligned}
\mathbb{P}[M_{t+1} = a_i &= \mathbb{P}[\text{no swap at time } t \text{ and} M_t = b_i] \\
&= \mathbb{P}[\text{no swap at time } t]\mathbb{P}[M_t = b_i] \\
&= \frac{t}{t+1}\frac{1}{t} \\
&= \frac{1}{t+1}
\end{aligned}$$

6. **(MU 2.24)** We roll a standard die over and over. What is the expected number of rolls until the first pair of consecutive sixes appears?

   For a given sequence of coin flips ending in 66, we can write it as $**..*6***..*6*...**..*66$, where $*$ is any roll that is not a 6. Let X_i be the number of flips until obtaining a 6. Notice that X_i is geometric with parameter 1/6. Let N be the number of 6's that are observed before observing 66. Notice that N is also geometric with parameter 1/6. Now we can use conditional expectation to write

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\left[\mathbb{E}[X|N]\right] \\
&= \mathbb{E}\left[\mathbb{E}[\sum_{i=1}^{N} X_i + 1|N]\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{N} \mathbb{E}[X_i] + 1\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{N} 7\right] \\
&= 7 \cdot \mathbb{E}[N] \\
&= 7 \cdot 6 \\
&= 42
\end{aligned}$$

7. **(MU 2.27)** Consider the following distribution on the integers $x \geq 1$: $\mathbb{P}[X = x] = (6/\pi^2)x^{-2}$. This is a valid distribution, since $\sum_{k=1}^{\infty} k^{-2} = \pi^2/6$. What is its expectation?

   We begin by simply writing out the definition of expectation

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=1}^{\infty} x\mathbb{P}[X = x] \\
&= \sum_{x=1}^{\infty} \frac{6}{\pi^2}\frac{1}{x}
\end{aligned}$$

   This sum diverges since $\sum \frac{1}{x}$ diverges, so the expectation is not finite.

8. **(MU 2.32)** You need a new staff assistant, and you have n people to interview. You want to hire the best candidate for the position. When you interview a candidate, you can give them a score, with the highest score being the best and no ties being possible. You interview the candidates one by one. Because of your company's hiring practices after you interview the kth candidate, you either offer the candidate the job before the next interview or you forever lose the chance to hire that candidate. We suppose the candidates are interviewed in a random order, chosen uniformly at random from all $n!$ possible orderings.

   We consider the following strategy. First, interview m candidates but reject them all; these candidates give you an idea of how strong the field is. After the mth candidate, hire the first candidate you interview who is better than all of the previous candidates you interviewed.

   (a) Let E be the event that we hire the best assistant, and let E_i be the event the ith candidate is the best and we hire them. Determine $\mathbb{P}[E_i]$, and show that

$$\mathbb{P}[E] = \frac{m}{n} \sum_{j=m+1}^{n} \frac{1}{j-1}.$$

Notice that the E_i are disjoint events, therefore $\mathbb{P}[E] = \sum_{i=1}^{n} \mathbb{P}[E_i]$. For $i <= m$, $\mathbb{P}[E_i] = 0$, since none of the first m candidates are selected. Now, we see that for $i > m$ two independents events make up E_i.

$$\begin{aligned}
\mathbb{P}[E_i] &= \mathbb{P}[i\text{th candidate is the best}] \cdot \mathbb{P}[\text{the } i\text{th candidate is chosen}] \\
&= \frac{1}{n} \cdot \mathbb{P}[\text{best of the } i-1 \text{ candidates is in the first } m \text{ candidates}] \\
&= \frac{1}{n} \frac{m}{i-1}
\end{aligned}$$

Now, putting this all together, we get

$$\mathbb{P}[E] = \sum_{i=m+1}^{n} \mathbb{P}[E_i] = \frac{m}{n} \sum_{i=m+1}^{n} \frac{1}{i-1}$$

(b) Bound $\sum_{j=m+1}^{n} \frac{1}{j-1}$ to obtain

$$\frac{m}{n}(\ln n - \ln m) \leq \mathbb{P}[E] \leq \frac{m}{n}(\ln(n-1) - \ln(m-1)).$$

Using Lemma 2.10 from the book, we get the solution

$$\mathbb{P}[E] \geq \frac{m}{n} \int_{m+1}^{n+1} \frac{1}{x-1} dx = \ln(x-1)|_{m+1}^{n+1} = \frac{m}{n}\left(\ln(n) - \ln(m)\right)$$

and

$$\mathbb{P}[E] \leq \frac{m}{n} \int_{m}^{n} \frac{1}{x-1} dx = \ln(x-1)|_{m}^{n} = \frac{m}{n}\left(\ln(n-1) - \ln(m-1)\right)$$

(c) Show that $m(\ln n - \ln m)/n$ is maximized when $m = n/e$, and explain why this means $\mathbb{P}[E] \geq 1/e$ for this choice of m.

How should we find the best m? Since the bound from above is concave, we can take the derivative, set it equal to zero, and solve for m. This yields the m that maximizes $Pr[E]$. We have

$$\frac{d}{dm} \frac{m}{n}\left(\ln(n) - \ln(m)\right) = \frac{\ln(n)}{n} - \frac{\ln(m)}{n} + \frac{1}{n} = 0.$$

Then we get $\ln(m) = \ln(n) - 1$, which is

$$m = e^{\ln(n)-1} = e^{\ln(n)}e^{-1} = ne^{-1} = \frac{n}{e}.$$

Substituting this m back into the bound from part (b), we get

$$\mathbb{P}[E] \geq \frac{1}{e}\left(\ln n - \ln \frac{n}{e}\right) = 1/e.$$