18.310 Homework 2 Solutions

Instructions: Remember to submit a separate PDF for each question.Do not forget to include a list of your collaborators or to state that you worked on your own.

1. The following theorem and its proof are mathematically incorrect.

   Let n be any positive integer and recall that $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$. For a set $X \subseteq \mathbb{Z}_n$, let $2X = \{2x \pmod{n} : x \in X\}$.

   Theorem 1. *Suppose that X and Y are drawn independently and uniformly at random among all 2^n subsets of $\mathbb{Z}_n$. Then $\mathbb{P}(2X \cap Y = \emptyset) = \frac{3^n}{4^n}$.*

   Proof. We know that a uniformly random set X can be generated by independently deciding to include i in X with probability $\frac{1}{2}$, for each $i \in \mathbb{Z}_n$. Thus, we obtain that

   $$\mathbb{P}(2X \cap Y = \emptyset) = \prod_{i=0}^{n-1} \mathbb{P}(\text{if } i \in X \text{ then } (2i \pmod{n}) \notin Y) = \left(\frac{3}{4}\right)^n.$$

   □

   - **To be handed in in recitation on 9/12/2013.** Show that the theorem is false by explicitly calculating $\mathbb{P}(2X \cap Y = \emptyset)$ for $n = 2$. Does your counterexample generalize to $n = 3$? (Also, think about what step of the current proof is incorrect prior to recitation.)
   - **Writing assignment.** (To be submitted with the rest of the problem set on 9/18/2013.) Correct the statement of the theorem above so that it is true for every n. Provide a well-written proof. Pay attention to notation, and also to the issues that made the "proof" above wrong.

Solution. The correct statement is the following.

Theorem 2. *Suppose that X and Y are drawn independently and uniformly at random among all 2^n subsets of $\mathbb{Z}_n$. Then*

$$\mathbb{P}(2X \cap Y = \emptyset) = \begin{cases} \left(\frac{3}{4}\right)^n & \text{if } n \text{ is odd} \\ \left(\frac{5}{8}\right)^{n/2} & \text{if } n \text{ is even.} \end{cases}$$

Proof. We know that a uniformly random set X can be generated by independently deciding to include i in X with probability $\frac{1}{2}$, for each $i \in \mathbb{Z}_n$. Similarly for Y. Thus we can see the process of generating X and Y as coming from $2n$ *independent* coin tosses, n for X and n for Y.

Now define event A_i for $i \in \mathbb{Z}_n$ to be the event that if i is in X then $2i \bmod n$ is not in Y. The probability we are looking, $\mathbb{P}(2X \cap Y = \emptyset)$, can thus be expressed as

$$\mathbb{P}\left(\bigwedge_{i=0}^{n-1} A_i\right).$$

It is easy to see that $\mathbb{P}(A_i) = \frac{3}{4}$ for any i.

In the case in which n is odd, we have that all n events $(A_i)_{i=0,\cdots,n-1}$ are independent since each depends on the result of two coin tosses (one for whether i is in X, the other whether $2i$ is in Y) and overall these correspond to distinct coins since $2i \bmod n$ is never equal to $2j \bmod n$ unless $i = j$. Thus, we have that

$$\mathbb{P}\left(\bigwedge_{i=0}^{n-1} A_i\right) = \prod_{i=0}^{n-1} \mathbb{P}(A_i) = \left(\frac{3}{4}\right)^n.$$

However, when n is even, the events A_i and $A_{i+n/2}$ are not independent as they both involve whether $2i \bmod n$ are not in Y. Instead, we define the event B_i to be $A_i \wedge A_{i+n/2}$ for $0 \le i < \frac{n}{2}$. Observe that B_i is the event that if either i or $i + \frac{n}{2}$ (or both) is in X then $2i \bmod n$ is not in Y; therefore $\mathbb{P}(B_i) = 1 - \mathbb{P}(2i \bmod n \text{ is in } Y)\mathbb{P}(\text{either } i \text{ or } i + n/2 \text{ is not in } X) = 1 - \frac{1}{2}\frac{3}{4} = \frac{5}{8}$. We can now write

$$\mathbb{P}\left(\bigwedge_{i=0}^{n-1} A_i\right) = \mathbb{P}\left(\bigwedge_{i=0}^{\frac{n}{2}-1} B_i\right).$$

Now all our events B_i's for $i = 0, \cdots, \frac{n}{2} - 1$ are independent and thus we obtain:

$$\mathbb{P}\left(\bigwedge_{i=0}^{n-1} A_i\right) = \mathbb{P}\left(\bigwedge_{i=0}^{\frac{n}{2}-1} B_i\right) = \prod_{i=0}^{\frac{n}{2}-1} \mathbb{P}(B_i) = \left(\frac{5}{8}\right)^{n/2}.$$

□

2. The classroom that we are in has six blackboard frames. In some of the lectures, the instructor enjoys showing his (lack of) drawing skills and draws a pigeon on one or several board frames. Show that over the course of a semester with 36 lectures, there exist two lectures and three board frames such that these three frames either all had no pigeons drawn on them in both lectures, or all had at least one pigeon drawn on them in both lectures.

   Solution 1. In a lecture, there are 2^5 ways in which the instructor can draw or not draw a pigeon in the first 5 frames. Since $2^5 = 32 < 36$, there are two lectures in which the professor leaves the first 5 blackboards the same way. Suppose this way involves leaving 3 of those frames empty. In this case these three frames all had no pigeons drawn in both lectures. If there are no 3 empty frames, there must be 3 frames that were drawn on. In this case these 3 frames had pigeons in both lectures.

Solution 2. In any given lecture, there are at least two groups of three boards that either all have pigeons, or none of them do. Indeed, we have either (i) 3 boards with pigeons and 3 boards without (and these are the two groups), or (ii) at least 4 boards all with pigeons or all without pigeons and in this latter case, we can choose any two subsets of size 3 of these (at least) 4 boards. So in 36 lectures, we have at least 72 of these groups of three. These groups of three can be any of $2\binom{6}{3} = 40$ in a given lecture. Since $40 < 72$ there are two among these 72 that occupy the same boards, and they correspond to two such lectures.

3. A random variable $Y : \Omega \to \mathbb{Z}$ is distributed according to the Poisson distribution with parameter $\lambda \geq 0$ if for all $i \geq 0$:
$$\mathbb{P}(Y = i) = \mathrm{e}^{-\lambda}\frac{\lambda^i}{i!}.$$
   - Verify that $\sum_{i=0}^{\infty} \mathbb{P}(Y = i) = 1$.
   - Show that $\mathbb{E}[Y] = \mathrm{Var}(Y) = \lambda$.
   - Suppose that each random variable $X_1, X_2, \ldots, X_n$ follows the Poisson distribution with parameter λ_i. Assume that all X_i are independent and let $X = \sum_{i=1}^{n} X_i$. Show that, for $\mu \geq \mathbb{E}[X]$ and for all $\delta > 0$:
$$\mathbb{P}(X > (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}.$$

     Hint: Compare this theorem with part (i) of Theorem 1 in the lecture notes on Chernoff Bound. Try to follow the proof of Theorem 1 closely.

Solution:

(a) This is just stating that
$$\sum_{i\geq 0} \frac{\lambda^i}{i!} = e^{\lambda},$$
which is true by Taylor's theorem.

(b) This is just a calculation:
$$\mathbb{E}[Y] = e^{-\lambda}\sum_{i\geq 1} i\frac{\lambda^i}{i!} = \lambda e^{-\lambda}\sum_{i\geq 0}\frac{\lambda^i}{i!} = \lambda,$$
and
$$\mathbb{E}[X^2] = e^{-\lambda}\sum_{i\geq 1} i^2\frac{\lambda^i}{i!} = \lambda e^{-\lambda}\sum_{i\geq 1} i\frac{\lambda^i}{i!} + \lambda e^{-\lambda}\sum_{i\geq 0}\frac{\lambda^i}{i!} = \lambda^2 + \lambda,$$
and since $\mathrm{Var}[Y] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, the result follows.

(c) Let $r > 0$ be arbitrary. We start with the inequality
$$e^{r(1+\delta)\mu}\mathbb{1}_{\{X>(1+\delta)\mu\}} \leq e^{rX}.$$
By the monotonicity of expectation, we get
$$e^{r(1+\delta)\mu}\mathbb{P}(X > (1+\delta)\mu) \leq \mathbb{E}(e^{rX}).$$

Using the independence of the variables X_i we can calculate

$$\mathbb{E}(e^{rX}) = \prod_{i=1}^{n} \mathbb{E}(e^{rX_i}) = e^{n\lambda(e^r-1)}.$$

where we have used that

$$\mathbb{E}(e^{rX_i}) = e^{-\lambda} \sum_{i\geq 0} \frac{e^{ri}\lambda^i}{i!} = e^{\lambda(e^r-1)}.$$

Plugging this calculation into our inequality, and using that $\mu \geq \mathbb{E}(X) = n\lambda$, we get

$$\mathbb{P}(X > (1+\delta)\mu) \leq \exp[n\lambda(e^r-1) - r(1+\delta)\mu] \leq \exp[\mu(e^r - 1 - r(1+\delta))].$$

We optimize the value of $e^r - 1 - r(1+\delta)$ to obtain $r = \log(1+\delta)$, which in turn gives us

$$\mathbb{P}(X > (1+\delta)\mu) \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu.$$

4. You may have heard recently some story about former MIT students (and other groups) winning a fair amount of money at the Massachusetts lottery game Cash WinFall (if not, just google 'Cash WinFall MIT students'). Let's analyze the game (or some simplification of it).

   In Cash WinFall, a customer can buy a ticket for \$2 which let him/her choose 6 numbers between 1 and 46 hoping to match the 6 (distinct) numbers being randomly selected at the next drawing. If the 6 numbers on the ticket match the 6 numbers that are drawn, he/she wins the jackpot, which is at least \$500,000. The customer also wins prize money if 5, 4 or 3 of the numbers are matched, see the second column in the table below for the prize money in each case.

Match	Prize money	Example (from 2/8/2010) of prize money when rolldown
6 out of 6	jackpot	-
5 out of 6	\$4,000	\$22,096
4 out of 6	\$150	\$807
3 out of 6	\$5	\$26

   The MIT students and the other groups exploited the fact that if the jackpot reaches \$2,000,000 and the jackpot is not won then part of the jackpot money is used to considerably increase the prize money for matching 5, 4 or 3 of the numbers; see the third column in the table above. Notice that the increase is more than 5-fold. Such a drawing is known as a *rolldown drawing*. The precise increase for a rolldown drawing is based on formulas that are not (quite) revealed to the public (and depends on the amount of the jackpot, etc.), but the increase is always very significant and of the order of magnitude shown in the 3rd column above.

   (a) For $i = 6, 5, 4, 3$, what is the probability p_i that one ticket matches precisely i of the 6 numbers that are randomly drawn? Give a formula and also numerically compute these probabilities.

Solution. We have that p_i is

$$p_i = \frac{\binom{6}{i}\binom{40}{6-i}}{\binom{46}{6}},$$

as we need to choose i numbers among the 6 winning ones, and $6-i$ numbers among the remaining 40. The values are:

$$\begin{aligned}
p_6 &= \frac{1}{9366819} = 0.000000106\cdots \\
p_5 &= \frac{240}{9366819} = 0.000025622\cdots \\
p_4 &= \frac{11700}{9366819} = 0.001249090\cdots \\
p_3 &= \frac{197600}{9366819} = 0.021095742\cdots
\end{aligned}$$

(b) Let A be the event that one wins any amount of prize money when buying a single ticket. What is $\mathbb{P}(A)$?

Solution. It is the probability that we win some prize, i.e.

$$p_6 + p_5 + p_4 + p_3 = \frac{209541}{9366819} = 0.022370561\cdots$$

(c) Let the random variable X be the prize money for a single ticket, assuming (i) that the jackpot amount is \$1,900,000 and (ii) that the drawing is not a rolldown drawing. What is $\mathbb{E}(X)$? Compute its numerical value. (Should you play?)

Solution. $\mathbb{E}(X) = 1900000p_6 + 4000p_5 + 150p_4 + 5p_3 = 0.59817\cdots$. (As this is less than the price of the winning, you shouldn't be playing if you are rational...)

(d) Assume that we have a rolldown drawing (i.e. no one wins the jackpot which happens to be over \$2,000,000). Suppose furthermore that the prize money for matching 5, 4 or 3 numbers are as in the 3rd column in the table. Let Y be the prize money for a single ticket under these assumptions.
What is $\mathbb{E}(Y)$ and $\mathrm{Var}(Y)$? Compute their values.

Solution. We have

$$\mathbb{E}(Y) = 22096p_5 + 807p_4 + 26p_3 = 2.12265658\cdots$$

Also,

$$\mathbb{E}(Y^2) = 22096^2p_5 + 807^2p_4 + 26^2p_3 = 13337.416\cdots$$

and thus

$$\mathrm{Var}(Y) = E(Y^2) - E(Y)^2 = 13332.910451\cdots$$

The standard deviation is $\sqrt{\mathrm{Var}(Y)} = 115.46$ (much larger than the expected earnings).

(e) If you purchase only one ticket, you have a large probability of not recovering your bet. Now suppose you purchase 1,000,000 tickets[1], each randomly drawn. Let Z be the total prize money received.

What is $\mathbb{E}(Z)$? What is $\text{Var}(Z)$? Use Chebyshev's inequality to compute an upper bound on the probability that $Z < 2,000,000$ (i.e. that you are losing money).

Solution. We have $\mathbb{E}(Z) = 1000000\mathbb{E}(Y) = 2122656.58$, while $\text{Var}(Z) = 1000000\text{Var}(Y) = 13332910451$. (Not asked: Here the standard deviation is 115468, much more comparable to the expected profit of $\mathbb{E}(Z) - 2000000$.)

Chebyshev's inequality says that

$$\mathbb{P}(|Z - E(Z)| \geq 2122656.58 - 2000000) \leq \frac{\text{Var}(Z)}{122656.58^2} = 0.886\cdots.$$

This implies that the probability that $Z < 2000000$ is less than 0.886.

(f) Now use the Chernoff-Hoeffding bound to compute a better upper bound on the probability that $Z < 2,000,000$. How much better is your result?

(g) If the jackpot goes over \$2,000,000, a rolldown might not happen since some ticket might win the jackpot. Suppose that, for a given drawing, the total number of (distinct) tickets sold[2] is $1,000,000$. Let B be the event that someone wins the jackpot. What is $\mathbb{P}(B)$? To evaluate this numerically, it is convenient to use the approximation[3] $1 - x \sim e^{-x}$.

Solution. We have

$$\mathbb{P}(B) = 1 - \mathbb{P}(\neg B) = 1 - \left(1 - \frac{1}{\binom{46}{6}}\right)^{1000000} \sim 1 - e^{-1000000/\binom{46}{6}} = 1 - e^{-0.10675983} = 0.1012585.$$

[1]The MIT students purchased up to 700,000 tickets for one drawing...

[2]In 2004-2005, the number of tickets sold in anticipation of a rolldown drawing was never more than 950,000 and typically less than 600,000 while in 2007, the number of tickets sold in a rolldown drawing was typically between 1,200,000 and 1,400,000.

[3]One has $1 - x \leq e^{-x}$ for *all* x, and the approximation $1 - x \sim e^{-x}$ is very good for x close to 0.