Tom Leighton, Marten van Dijk, and Brooke Cowan

Problem Set 12 Solutions

Due: Friday, December 5, 7pm

Problem 1. [10 points] Here are seven propositions:

$$\begin{array}{rcrcr} x_1 & \vee & x_3 & \vee & \neg x_7 \\ \neg x_5 & \vee & x_6 & \vee & x_7 \\ x_2 & \vee & \neg x_4 & \vee & x_6 \\ \neg x_4 & \vee & x_5 & \vee & \neg x_7 \\ x_3 & \vee & \neg x_5 & \vee & \neg x_8 \\ x_9 & \vee & \neg x_8 & \vee & x_2 \\ \neg x_3 & \vee & x_9 & \vee & x_4 \end{array}$$

Note that:

1. Each proposition is the OR of three terms of the form x_i or the form $\neg x_i$.
2. The variables in the three terms in each proposition are all different.

Suppose that we assign true/false values to the variables $x_1, \ldots, x_9$ independently and with equal probability.

(a) [5 pts] What is the expected number of true propositions?

Solution. Each proposition is true unless all three of its terms are false. Thus, each proposition is true with probability:

$$1 - \left(\frac{1}{2}\right)^3 = \frac{7}{8}$$

Let T_i be an indicator for the event that the i-th proposition is true. Then the number of true propositions is $T_1 + \ldots + T_7$ and the expected number is:

$$\begin{aligned} \mathrm{E}[T_1 + \ldots + T_7] &= \mathrm{E}[T_1] + \ldots + \mathrm{E}[T_7] \\ &= 7/8 + \ldots + 7/8 \\ &= 49/8 = 6\frac{1}{8} \end{aligned}$$

■

(b) [5 pts] Use your answer to prove that there exists an assignment to the variables that makes *all* of the propositions true.

Solution. A random variable can not always be less than its expectation, so there must be some assignment such that:

$$T_1 + \ldots T_7 \geq 6\frac{1}{8}$$

This implies that $T_1 + \ldots + T_7 = 7$ for at least one outcome. This outcome is an assignment to the variables such that all of the propositions are true. ■

Problem 2. [20 points] MIT students sometimes delay laundry for a few days. Assume all random values described below are mutually independent.

(a) [5 pts] A ***busy*** student must complete 3 problem sets before doing laundry. Each problem set requires 1 day with probability 2/3 and 2 days with probability 1/3. Let B be the number of days a busy student delays laundry. What is $\text{E}[B]$?

Example: If the first problem set requires 1 day and the second and third problem sets each require 2 days, then the student delays for $B = 5$ days.

Solution. The expected time to complete a problem set is:

$$1 \cdot \frac{2}{3} + 2 \cdot \frac{1}{3} = \frac{4}{3}$$

Therefore, the expected time to complete all three problem sets is:

$$\begin{aligned}
\text{E}[B] &= \text{E}[\text{pset1}] + \text{E}[\text{pset2}] + \text{E}[\text{pset3}] \\
&= \frac{4}{3} + \frac{4}{3} + \frac{4}{3} \\
&= 4
\end{aligned}$$

■

(b) [5 pts] A ***relaxed*** student rolls a fair, 6-sided die in the morning. If he rolls a 1, then he does his laundry immediately (with zero days of delay). Otherwise, he delays for one day and repeats the experiment the following morning. Let R be the number of days a relaxed student delays laundry. What is $\text{E}[R]$?

Example: If the student rolls a 2 the first morning, a 5 the second morning, and a 1 the third morning, then he delays for $R = 2$ days.

Solution. If we regard doing laundry as a failure, then the mean time to failure is $1/(1/6) = 6$. However, this counts the day laundry is done, so the number of days delay is $6 - 1 = 5$.

Alternatively, we could derive the answer as follows:

$$\begin{aligned}
\mathrm{E}[R] &= \sum_{k=0}^{\infty} \Pr\{R > k\} \\
&= \frac{5}{6} + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^3 + \dots \\
&= \frac{5}{6} \cdot \left(1 + \frac{5}{6} + \left(\frac{5}{6}\right)^2 + \dots\right) \\
&= \frac{5}{6} \cdot \frac{1}{1 - 5/6} \\
&= 5
\end{aligned}$$

■

(c) [5 pts] Before doing laundry, an ***unlucky*** student must recover from illness for a number of days equal to the product of the numbers rolled on two fair, 6-sided dice. Let U be the expected number of days an unlucky student delays laundry. What is $\mathrm{E}[U]$?

Example: If the rolls are 5 and 3, then the student delays for $U = 15$ days.

Solution. Let D_1 and D_2 be the two die rolls. Recall that a die roll has expectation 7/2. Thus:

$$\begin{aligned}
\mathrm{E}[U] &= \mathrm{E}[D_1 \cdot D_2] \\
&= \mathrm{E}[D_1] \cdot \mathrm{E}[D_2] \\
&= \frac{7}{2} \cdot \frac{7}{2} \\
&= \frac{49}{4}
\end{aligned}$$

■

(d) [5 pts] A student is ***busy*** with probability 1/2, ***relaxed*** with probability 1/3, and ***unlucky*** with probability 1/6. Let D be the number of days the student delays laundry. What is $\mathrm{E}[D]$?

Solution.

$$\mathrm{E}[D] = \frac{1}{2}\mathrm{E}[B] + \frac{1}{3}\mathrm{E}[R] + \frac{1}{6}\mathrm{E}[U]$$

■

Problem 3. [20 points] A gambler plays 120 hands of draw poker, 60 hands of black jack,and 20 hands of stud poker per day. He wins a hand of draw poker withprobability 1/6, a hand of black jack with probability 1/2, and a hand ofstud poker with probability 1/5. Assume the outcomes of the card games are mutually independent.

(a) [5 pts] What is the expected number of hands the gambler wins in a day?

Solution. $120(1/6) + 60(1/2) + 20(1/5) = 54$. ■

(b) [5 pts] What is the variance in the number of hands won per day?

Solution. The variance can also be calculated using linearity of variance. For an individual hand the variance is $p(1-p)$ where p is the probability of winning. Therefore the variance is

$$120 \cdot \frac{1}{6} \cdot \frac{5}{6} + 60 \cdot \frac{1}{2} \cdot \frac{1}{2} + 20 \cdot \frac{1}{5} \cdot \frac{4}{5} = \frac{50}{3} + 15 + \frac{16}{5} = \frac{250}{15} + \frac{225}{15} + \frac{48}{15} = \frac{523}{15} = 34\,\frac{13}{15}.$$

■

(c) [5 pts] What would the Markov bound be on the probability that the gambler will win 108 hands on a given day?

Solution. The expected number of games won is 54, so by Markov, $\Pr\{R \geq 108\} \leq 54/108 = 1/2$. ■

(d) [5 pts] What would the Chebyshev bound be on the probability that the gambler will win 108 hands on a given day?

Solution.

$$\Pr\{R - 54 \geq 54\} \leq \Pr\{|R - 54| \geq 54\} \leq \frac{V}{54^2} = \frac{523/15}{54^2} \approx \frac{1}{85}.$$

■

Problem 4. [10 points] Prove that for any random variable, R, and constant, b,

(a) [5 pts] if R and S are independent, then so are $(R + b)$ and S.

Solution.

$$\begin{aligned} \Pr\{(R+b) = r \wedge S = s\} &= \Pr\{R = r - b \wedge S = s\} \\ &= \Pr\{R = r - b\} \cdot \Pr\{S = s\} \qquad \text{(independence of } R \text{ and } S\text{)} \\ &= \Pr\{(R+b) = r\} \cdot \Pr\{S = s\} \end{aligned}$$

for all r, s, which implies $(R + b)$ and S are independent. ■

(b) [5 pts] Prove that $\text{Var}[R] = 0$ iff R is a constant with probability 1.

Solution. *Proof.* By the definition of variance,

$$\text{Var}[R] = 0 \text{ iff } \text{E}\left[(R - \text{E}[R])^2\right] = 0.$$

The inner expression on the right, $(R - \text{E}[R])^2$, is always nonnegative because of the square. As a result, $\text{E}[(R - \text{E}[R])^2] = 0$ if and only if $\Pr\{(R - \text{E}[R])^2 \neq 0\}$ is zero, which is the same as saying that $\Pr\{(R - \text{E}[R])^2 = 0\}$ is one. That is,

$$\text{Var}[R] = 0 \text{ iff } \Pr\left\{(R - \text{E}[R])^2 = 0\right\} = 1.$$

But the $(R - \text{E}[R])^2 = 0$ and $R = \text{E}[R]$ are different descriptions of the same event. Therefore,

$$\text{Var}[R] = 0 \text{ iff } \Pr\{R = \text{E}[R]\} = 1.$$

So R equals the constant $\text{E}[R]$ with probability 1. □

■

Problem 5. [15 points] A man has a set of n keys, one of which fits the door to his apartment. He tries the keys until he finds the correct one. Give the expectation and variance for the number of trials until success if he tries the keys at random (possibly repeating a key tried earlier)

Solution. This is a mean time to failure problem if finding a key is taken to be a "failure". The probability of failure on the ith try, given "success" on the previous tries, is $1/n$, so if T is the number of tries to find the right key, then $\text{E}[T] = 1/(1/n) = n$.

The variance of T can be computed directly from the formula,

$$\text{Var}[T] = \text{E}\left[T^2\right] - \text{E}^2[T],$$

using the fact that $\Pr\{T = k\} = pq^{k-1}$. By definition we have

$$\text{E}\left[T^2\right] = \sum_{k=1}^{n} k^2 pq^{k-1}$$

To evaluate this sum, we use the following trick. We know that

$$\sum_{k=1}^{\infty} kq^k = q \sum_{k=1}^{\infty} kq^{k-1} = \frac{q}{(1-q)^2}$$

Moreover,

$$\begin{aligned}
\sum_{k=1}^{\infty} k^2 q^{k-1} &= \sum_{k=1}^{\infty} \frac{d}{dq} kq^k \\
&= \frac{d}{dq} \sum_{k=1}^{\infty} kq^k \\
&= \frac{d}{dq} \frac{q}{(1-q)^2} \\
&= \frac{1}{(1-q)^2} + \frac{2q}{(1-q)^3} = \frac{1}{p^2} + \frac{2q}{p^3}.
\end{aligned}$$

Finally, the expectation $\mathrm{E}[T^2]$ is

$$\mathrm{E}\left[T^2\right] = p\left(\frac{1}{p^2} + \frac{2q}{p^3}\right) = \frac{1}{p} + \frac{2q}{p^2},$$

and the variance is

$$\mathrm{Var}[T] = \frac{1}{p} + \frac{2q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2} = n(n-1).$$

■

Problem 6. [10 points] Ike Harmon wants to get married, but he isn't sure that he's met his soulmate yet. He decides on the following strategy. He will marry the first woman he meets. Then, if he meets someone he finds more suitable, he will divorce his current wife and marry the newcomer.

Suppose Ike meets a total of n women throughout his life, and for any two women, Ike prefers one to the other. Ike's preference relation is a total order, that is, it is transitive. Due to the randomness in everyday life, Ike unfortunately meets these n women in random order.

Prove that the expected number of times Ike has to marry is $\sim \ln n$.

Hint: Let M_i be the indicator variable for the event that Ike marries the ith woman.

Solution. Note that the number of times Ike marries is precisely $M = M_1 + \ldots + M_n$. Since expectation is a linear operator, we can compute $\mathrm{E}[M]$ by finding $E(M_i)$ for all i and summing them up. Note also that since M_i is an "indicator" variable, we only have to find $\Pr\{M_i = 1\}$. In a random permutation, this happens with probability $1/i$. Why? Because all permutations of the first i women that Ike meets are equally likely, and the most suitable for Ike occurs as the last woman of the permutation in $1/i$ of the cases. Thus

$$\begin{aligned} \mathrm{E}[M] &= \sum_i \Pr\{M_i = 1\} \\ &= \sum_{i=1}^{n} 1/i \\ &= H_n \sim \ln n, \end{aligned}$$

where H_n is the nth Harmonic number. ■

Problem 7. [15 points] Let X, X_1, …, X_n be independent identically distributed random variables. Define the random variable $S = \sum_{i=1}^{n} X_i$. Define the function $M_Y(s) = \mathrm{E}\left[e^{sY}\right]$ for random variables Y.

(a) [5 pts] Prove $\Pr\{S \geq na\} \leq (M_X(s)e^{-sa})^n$ for $s > 0$. (Hint: Prove $M_S(s) = (M_X(s))^n$ and use the Markov bound.)

Solution. By linear independence,

$$M_S(s) = \mathrm{E}\left[e^{sS}\right] = \mathrm{E}\left[\prod_{i=1}^{n} e^{sX_i}\right] = \prod_{i=1}^{n} \mathrm{E}\left[e^{sX_i}\right] = \mathrm{E}\left[e^{sX}\right]^n = M_X(s)^n.$$

By the Markov bound,

$$\Pr\{S \geq na\} = \Pr\left\{e^{sS} \geq e^{nsa}\right\} \leq \mathrm{E}\left[M_S(s)\right] e^{-nsa} = (M_X(s)e^{-sa})^n.$$

■

(b) [5 pts] Let X be a binary variable with $\Pr\{X = 0\} = p$ and $\Pr\{X = 1\} = 1 - p$. Let $1 - p < a < 1$. Show that $M_X(s)e^{-sa}$ is minimized for $s = \ln \frac{ap}{(1-a)(1-p)}$ giving

$$\min_{s>0} M_X(s)e^{-sa} = \left(\frac{p}{1-a}\right)^{1-a} \left(\frac{1-p}{a}\right)^a.$$

By using part (a), this yields the bound

$$\Pr\{S \geq na\} \leq \left[\left(\frac{p}{1-a}\right)^{1-a} \left(\frac{1-p}{a}\right)^a\right]^n.$$

Solution. We compute

$$M_X(s)e^{-sa} = (p + (1-p)e^s)e^{-sa} = pe^{-sa} + (1-p)e^{s(1-a)}.$$

Differentiating gives

$$-pae^{-sa} + (1-p)(1-a)e^{s(1-a)},$$

which is equal to 0 if $-pa + (1-p)(1-a)e^s = 0$, that is, $s = \ln \frac{ap}{(1-a)(1-p)}$. Nptice that $s > 0$ for $a > 1 - p$. Substituting this into $M_X(s)e^{-sa} = (p + (1-p)e^s)e^{-sa}$ gives

$$\left(p + (1-p)\frac{ap}{(1-a)(1-p)}\right)\left(\frac{ap}{(1-a)(1-p)}\right)^{-a}$$

and reordering terms gives

$$\frac{p}{1-a}\left(\frac{ap}{(1-a)(1-p)}\right)^{-a}$$

which is equal to the desired expression. ■

(c) [5 pts] Compute the bound of part (b) for $p = 0.05$ and $a = 0.99$ and compare this bound to the Chernoff bound as presented in the lecture notes.

Solution. The bound of part (b) gives 0.97557^n. Since $\mathrm{E}[S] = (1-p)n$ and $a > 1 - p$, the Chernoff bound gives

$$\Pr\{S \geq na\} = \Pr\left\{S \geq \frac{a}{1-p}\mathrm{E}[S]\right\} \leq e^{-(\frac{a}{1-p}\ln\frac{a}{1-p}+1-\frac{a}{1-p})(1-p)n} = e^{-(a\ln\frac{a}{1-p}+1-p-a)n}.$$

For the specified p and a this bound is equal to 0.99917^n. ■