

Problem Set 12 Solutions

Due: Friday, December 3, 7pm

Problem 1. [15 points]

In this problem, we will (hopefully) be making tons of money! Use your knowledge of probability and statistics to keep from going broke!

Suppose the stock market contains N types of stocks, which can be modelled by independent random variables. Suppose furthermore that the behavior of these stocks is modelled by a double-or-nothing coin flip. That is, stock S_i has half probability of doubling its value and half probability of going to 0. The stocks all cost a dollar, and you have N dollars. Say you only keep these stocks for one time-step (that is, at the end of this timestep, all stocks would have doubled in value or gone to 0).

(a) [3 pts] What is your expected amount of money if you spend all your money on one stock? Your variance?

Solution. The stock doubles on a coin flip, so your expected final amount is $.5(2N) + .5(0) = N$. Your variance is calculated as $E[(X - \mu)^2] =$. This, when we take into account the probability distribution of the stock, is

$$1/2(2N - N)^2 + 1/2(0 - N)^2 = N^2$$

■

(b) [3 pts] Suppose instead you diversified your purchases and bought N shares of all different stocks. What is your expected amount of money then? Your variance?

Solution. The amount of money you have in stocks is $X_1 + X_2 + \dots + X_N$, where X_i is a random variable describing how much money you have in stock i . The amount of money you expect to have is

$$E\left[\sum_{i=1}^N X_i\right] = \sum_{i=1}^N E[X_i]$$

But $E[X_i] = 2 * 1/2 + 0 * 1/2 = 1$, so this sum turns out to be N again.

As for variance, recall that $Var(\sum X_i) = \sum Var(X_i)$ if the random variables X_i are independent (which they are in this case). The variance of a single X_i is

$$1/2 * (2 - 1)^2 + 1/2 * (0 - 1)^2 = 1$$

so the variance of your entire portfolio is N .

■

(c) [3 pts] The money that you have invested came from your financially conservative mother. As a result, your goals are much aligned with hers. Given this, which investment strategy should you take?

Solution. Your mother prefers to have less risk, and so would like the stock with less variance. This is the strategy associated with (b). ■

(d) [3 pts] Now instead say that you make money on rolls of dice. Specifically, you play a game where you roll a standard six-sided dice, and get paid an amount (in dollars) equal to the number that comes up. What is your expected payoff? What is the variance?

Solution. The expected payoff is $1/6(1 + 2 + 3 + 4 + 5 + 6) = 3.5$. The variance is

$$1/6(2.5^2 + 1.5^2 + .5^2 + .5^2 + 1.5^2 + 2.5^2) = 35/12$$

■

(e) [3 pts] We change the rules of the game so that your payoff is the cube of the number that comes up. In that case, what is your expected payoff? What is its variance?

Solution. Your expected earnings is $1/6(1 + 8 + 27 + 64 + 125 + 216) = 441/6$. To calculate variance, we can simplify by noting that $Var(X) = E[X^2] - E^2[X]$. So the variance is

$$1/6(1 + 64 + 27^2 + 64^2 + 125^2 + 216^2) - (441/6)^2 = 67171/6 - 194481/36 = 208545/36 \approx 5792$$

■

Problem 2. [10 points] Here are seven propositions:

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

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, \dots, 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 $T_1 + \dots + T_7$ and the expected number is:

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

■

(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:

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

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

Problem 3. [20 points] MIT students sometimes delay laundry for a few days (to the chagrin of their roommates). Assume all random variables 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 $E[B]$?

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:

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

■

(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 $E[R]$?

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:

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

■

(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 $E[U]$?

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:

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

■

(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 $E[D]$?

Solution.

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

■

Problem 4. [10 points] We have two coins: one is a fair coin and the other is a coin that produces heads with probability $3/4$. One of the two coins is picked, and this coin is tossed n times. Explain how to calculate the number of tosses to make us 95% confident which coin was chosen. You do not have to calculate the minimum value of n , though we'd be pleased if you did.

Solution. To guess which coin was picked, set a threshold t between $1/2$ and $3/4$. If the proportion of heads is less than the threshold, guess it was the fair coin; otherwise, guess the biased coin. Let the random variable J be the number of heads in the first n flips. We need to flip the coin enough times so that $\Pr\{J/n \geq t\} \leq 0.05$ if the fair coin was picked, and $\Pr\{J/n \leq t\} \leq 0.05$ if the biased coin was picked. A natural threshold to choose is $5/8$, exactly in the middle of $1/2$ and $3/4$.

For the fair coin, J has an $(n, 1/2)$ -binomial distribution, so we need to choose n so that

$$\Pr\left\{J > \left(\frac{5}{8}\right)n\right\} \leq 0.05$$

which is equivalent to

$$CDF_J\left(\frac{5}{8}n\right) \geq 0.95 \quad (1)$$

For the biased coin, J has an $(n, 3/4)$ -binomial distribution, so we need to choose n so that

$$\Pr\left\{J \leq \left(\frac{5}{8}\right)n\right\} \leq 0.05$$

which is equivalent to

$$CDF_J\left(\frac{5}{8}n\right) \leq 0.95 \quad (2)$$

We can now search for the minimum n that satisfies both (1) and (2), using one of the several ways we know to calculate or approximate the binomial cumulative distribution function.

■

Problem 5. [13 points] Each 6.042 final exam (out of 100 points) will be graded according to a rigorous procedure:

- With probability $\frac{4}{7}$ the exam is graded by a TA; with probability $\frac{2}{7}$ it is graded by a lecturer; and with probability $\frac{1}{7}$, it is accidentally dropped behind the radiator and arbitrarily given a score of 84.

- *TAs* score an exam by scoring each problem individually and then taking the sum.
 - There are ten true/false questions worth 2 points each. For each, full credit is given with probability $3/4$, and no credit is given with probability $1/4$.
 - There are four questions worth 15 points each. For each, the score is determined by rolling two fair dice, summing the results, and adding 3.
 - The single 20 point question is awarded either 12 or 18 points with equal probability.
- *Lecturers* score an exam by rolling a fair die twice, multiplying the results, and then adding a “general impression” score.
 - With probability $\frac{4}{10}$, the general impression score is 40.
 - With probability $\frac{3}{10}$, the general impression score is 50.
 - With probability $\frac{3}{10}$, the general impression score is 60.

Assume all random choices during the grading process are independent.

(a) [5 pts] What is the expected score on an exam graded by a TA?

Solution. Let the random variable T denote the score a TA would give. By linearity of expectation, the expected sum of the problemscores is the sum of the expected problem scores. Therefore, we have:

$$\begin{aligned}
 E[T] &= 10 \cdot E[\text{T/F score}] + 4 \cdot E[\text{15pt prob score}] + E[\text{20pt prob score}] \\
 &= 10 \cdot \left(\frac{3}{4} \cdot 2 + \frac{1}{4} \cdot 0 \right) + 4 \cdot \left(2 \cdot \frac{7}{2} + 3 \right) + \left(\frac{1}{2} \cdot 12 + \frac{1}{2} \cdot 18 \right) \\
 &= 10 \cdot \frac{3}{2} + 4 \cdot 10 + 15 \\
 &= 70
 \end{aligned}$$

■

(b) [5 pts] What is the expected score on an exam graded by a lecturer?

Solution. Now we find the expected value of L , the score a lecturer would give. Employing linearity again, we have:

$$\begin{aligned}
 E[L] &= E[\text{product of dice}] + E[\text{general impression}] \\
 &= \left(\frac{7}{2} \right)^2 + \left(\frac{4}{10} \cdot 40 + \frac{3}{10} \cdot 50 + \frac{3}{10} \cdot 60 \right) \\
 &= \frac{49}{4} + 49 \\
 &= 61\frac{1}{4}
 \end{aligned}$$

■

(c) [3 pts] What is the expected score on a 6.042 final exam?

Solution. Let X equal the true exam score. The total expectation law implies:

$$\begin{aligned} \mathbb{E}[X] &= \frac{4}{7} \cdot \mathbb{E}[T] + \frac{2}{7} \cdot \mathbb{E}[L] + \frac{1}{7} \cdot 84 \\ &= \frac{4}{7} \cdot 70 + \frac{2}{7} \cdot \left(\frac{49}{4} + 49 \right) + \frac{1}{7} \cdot 84 \\ &= 40 + \frac{7}{2} + 14 + 12 \\ &= 69\frac{1}{2} \end{aligned}$$

■

Problem 6. [32 points]

Suppose n balls are thrown randomly into n boxes, so each ball lands in each box with uniform probability. Also, suppose the outcome of each throw is independent of all the other throws.

(a) [5 pts] Let X_i be an indicator random variable whose value is 1 if box i is empty and 0 otherwise. Write a simple closed form expression for the probability distribution of X_i . Are X_1, X_2, \dots, X_n independent random variables?

Solution. Box i is empty iff all n balls land in other boxes. The probability that a ball will land in another box is $(n-1)/n = 1 - (1/n)$, and since the balls are thrown independently, we have

$$\Pr(X_i = 1) = \left(1 - \frac{1}{n}\right)^n. \quad (3)$$

The X_i 's are not independent. For example,

$$\Pr(X_1 = X_2 = \dots = X_n = 1) = 0 < \prod_{i=1}^n \Pr(X_i = 1).$$

■

(b) [5 pts] Find a constant, c , such that the expected number of empty boxes is asymptotically equal (\sim) to cn .

Solution. The number of empty boxes is the sum of the X_i 's. So the expected number of empty boxes is the sum of the expectations of the X_i 's. By (3), we now have

$$\mathbb{E}(\text{number of empty boxes}) = n \mathbb{E}(X_1) = n \left(1 - \frac{1}{n}\right)^n \sim n \cdot \frac{1}{e}$$

That is,

$$c = \frac{1}{e}$$

■

(c) [5 pts] Show that

$$\Pr(\text{at least } k \text{ balls fall in the first box}) \leq \binom{n}{k} \left(\frac{1}{n}\right)^k.$$

Solution. Let S be a set of k of the n balls, and let E_S be the event that each of these k balls falls in the first box. Since the probability that a ball lands in this box is $1/n$, and the throws are independent, we have

$$\Pr(E_S) = \left(\frac{1}{n}\right)^k. \quad (4)$$

The event that *at least* k balls land in the first box is the union of all the events E_S . There are $\binom{n}{k}$ subsets, S , of k balls, so by the Union Bound,

$$\Pr(\text{at least } k \text{ balls fall in the first box}) \leq \binom{n}{k} \cdot \Pr(E_S).$$

Using the value for $\Pr(E_S)$ from (4) in the preceding inequality yields the required bound.

■

(d) [7 pts] Let R be the maximum of the numbers of balls that land in each of the boxes. Conclude from the previous parts that

$$\Pr\{R \geq k\} \leq \frac{n}{k!}.$$

Solution. Note that $R \geq k$ exactly when some box has at least k balls. Since the bound on the probability of at least k balls in the first box applies just as well to any box, we can apply the Union Bound to having at least k balls in at least one of the n boxes:

$$\Pr(R \geq k) \leq n \Pr(\text{at least } k \text{ balls fall in the first box}).$$

So from the previous problem part, we have

$$\begin{aligned} \Pr(R \geq k) &\leq n \binom{n}{k} \left(\frac{1}{n}\right)^k \\ &= n \left(\frac{n(n-1) \cdots (n-k+1)}{k! n^k} \right) \\ &= \frac{n}{k!} \left(\frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} \right) \\ &\leq \frac{n}{k!} \end{aligned}$$

■

(e) [10 pts] Conclude that

$$\lim_{n \rightarrow \infty} \Pr \{R \geq n^\epsilon\} = 0$$

for all $\epsilon > 0$.

Solution. Using Stirling's formula, and the upper bound from the previous part, we have

$$\Pr \{R \geq k\} \leq \frac{n}{k!} \sim \frac{n}{\sqrt{2\pi k}(k/e)^k} \leq \frac{n}{(k/e)^k} = \frac{ne^k}{k^k} = \frac{e^{k+\ln n}}{e^{k \ln k}}.$$

Now let $k = n^\epsilon$. Then the exponent of e in the numerator above is $n^\epsilon + \ln n$, and the exponent of e in the denominator is $n^\epsilon \ln n^\epsilon$. Since

$$n^\epsilon + \ln n = o(n^\epsilon \ln n^\epsilon),$$

we conclude

$$\Pr \{R \geq n^\epsilon\} \leq \frac{e^{n^\epsilon + \ln n}}{e^{n^\epsilon \ln n^\epsilon}} \rightarrow 0$$

as n approaches ∞ . ■