Problem Set 10.5 Solutions

Problems

Problem 1 Eight men and seven women, all single, happen randomly to have purchased single seats in the same 15-seat row of a theater.

(a) What is the probability that the first two seats contain a marriageable couple, i.e., a single man next to a single woman?

Solution.

There are $\binom{15}{2} = 105$ possible pairs that could occupy the first two seats, only $8 \cdot 7 = 56$ of which are marriageable couples. So the probability is $56/105 = 8/15$.

(b) What is the expected number of pairs of adjacent seats which contain marriageable couples? (For example, if the sequence of men and women in the seats is

MMMWMMMMWWWWWWM,

then there are four possible couples sitting in the pairs of adjacent seats that start at the 3rd, 4th, 7th and 14th seats, respectively.)

Solution.

Let the random variable S be the number of pairs of seats containing a marriageble couple; we want to find $\mathrm{E}[S]$. Let A_i be the event "there is a marriageable couple in seats i and $i+1$," and S_i be the indicator random variable for A_i; that is, $S_i = 1$ if A_i and $S_i = 0$ otherwise. Then $S = \sum_{i=1}^{14} S_i$. Since S_i is the indicator random variable for A_i, $\mathrm{E}[S_i] = \Pr(A_i)$. Since each pair of seats is like any other, $\Pr(A_i) = \Pr(A_1) = 8/15$, as we computed in the previous part. Because expectation is linear,

$$\mathrm{E}[S] = \sum_{i=1}^{14} \mathrm{E}[S_i] = 14 \cdot \frac{8}{15}.$$

Note that this works even though the S_i's are not independent. That's the beauty of linearity of expectation!

Problem 2 One hundred twenty students take the 6.042 final exam. The mean on the exam is 90 and the lowest score was 30. You have no other information about the students and the exam, e.g., you should not assume that the final is worth 100 points.

(a) State the best possible upper bound on the number of students who scored at least 180.

Solution.

Let R be the score of a student chosen at random. Apply Markov's Bound to $R - 30$:

$$\Pr(R \geq 180) = \Pr(R - 30 \geq 150) \leq \frac{\mathrm{E}[R-30]}{150} = \frac{60}{150} = \frac{2}{5}.$$

So at most $\frac{2}{5} \cdot 120 = 48$ students scored greater than or equal to 180.

□

(b) Give an example set of scores which achieve your bound.

Solution.

Of the 120 students, 48 score 180 and 72 score 30. The mean is $\frac{48 \cdot 180 + 72 \cdot 30}{120} = 90$, as required.

□

(c) If the maximum score on the exam was 100, give the best possible upper bound on the number of students who scored *at most* 50.

Solution.

Apply Markov's Bound to $100 - R$:

$$\Pr(R \leq 50) = \Pr(100 - R \geq 50) \leq \frac{\mathrm{E}[100-R]}{50} = \frac{10}{50} = \frac{1}{5}.$$

So at most $\frac{1}{5} \cdot 120 = 24$ students scored 50 or less.

□

Problem 3 A couple plans to have children until they have a boy. What is the expected number of children that they have, and what is the variance?

Solution.

Let C be the expected number of children up to and including the first boy. We showed $\mathrm{E}[C] = 2$ in Lecture 22. We can compute the variance as follows:

$$\mathrm{Var}[C] = \mathrm{E}\left[C^2\right] - \mathrm{E}[C]^2 = \sum_{k=1}^{\infty} k^2 \cdot \left(\frac{1}{2}\right)^k - 2^2 = \frac{\frac{1}{2} + (\frac{1}{2})^2}{(1-\frac{1}{2})^3} - 4 = 6 - 4 = 2$$

The sum is computed by differentiating the formula for the sum of an infinite geometric sequence.

□

Problem 4 Suppose that n people have their hats returned at random. Let $X_i = 1$ if the ith person gets his or her own hat back and 0 otherwise. Let $S_n = \sum_{i=1}^{n} X_i$, so S_n is the total number of people who get their own hat back. Show that

(a) $\mathrm{E}\left[X_i^2\right] = 1/n$.

Solution.

Note that $X_i^2 = X_i$, which is 1 with probability $1/n$ and 0 otherwise. Thus, $\mathrm{E}\left[X_i^2\right] = \mathrm{E}\left[X_i\right] = 1/n$.

□

(b) $\mathrm{E}\left[X_i X_j\right] = 1/n(n-1)$ for $i \neq j$.

Solution.

$X_i X_j$ is 1 if both X_i and X_j are 1, and 0 otherwise. $\Pr(X_i = 1 \wedge X_j = 1) = \frac{1}{n} \cdot \frac{1}{n-1} = \frac{1}{n(n-1)}$. Thus, $\mathrm{E}\left[X_i X_j\right] = 1/n(n-1)$.

□

(c) $\mathrm{E}\left[S_n^2\right] = 2$. *Hint*: Use (a) and (b).

Solution.

$$\begin{aligned}
\mathrm{E}\left[S_n^2\right] &= \mathrm{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] = \mathrm{E}\left[\sum_i \sum_j X_i X_j\right] = \sum_i \sum_j \mathrm{E}\left[X_i X_j\right] \\
&= \sum_i \mathrm{E}\left[X_i^2\right] + \sum_i \sum_{j \neq i} \mathrm{E}\left[X_i X_j\right] = n \cdot \frac{1}{n} + n(n-1) \cdot \frac{1}{n(n-1)} = 2.
\end{aligned}$$

□

(d) $\mathrm{Var}\left[S_n\right] = 1$.

Solution.

$\mathrm{Var}\left[S_n\right] = \mathrm{E}\left[S_n^2\right] - \mathrm{E}^2\left[S_n\right] = 2 - (n \cdot \frac{1}{n})^2 = 2 - 1 = 1$

□

(e) $\Pr(S_n \geq 11) \leq .01$ for any $n \geq 11$. *Hint*: Use Chebyshev's Inequality.

Solution.

$$\begin{aligned}
\Pr(S_n \geq 11) &= \Pr(S_n - \mathrm{E}\left[S_n\right] \geq 11 - \mathrm{E}\left[S_n\right]) \\
&= \Pr(S_n - \mathrm{E}\left[S_n\right] \geq 10) \\
&\leq \frac{\mathrm{Var}\left[S_n\right]}{\mathrm{Var}\left[S_n\right] + 10^2} \\
&= \frac{1}{101} < .01
\end{aligned}$$

Note that the X_i's are Bernoulli variables but are *not* independent, so S_n does not have a binomial distribution and the estimates from Lecture Notes 21 do not apply.

□

Problem 5 Let X and Y be independent random variables taking on integer values in the range 1 to n uniformly. Compute the following quantities:

(a) $\mathrm{Var}\left[aX+bY\right]$

Solution.

First, we compute $\mathrm{Var}\left[X\right]$, which is also $\mathrm{Var}\left[Y\right]$ since X and Y are identically distributed.

$$\begin{aligned}\mathrm{Var}\left[X\right] = \mathrm{E}\left[X^2\right] - \mathrm{E}\left[X\right]^2 &= \sum_{i=1}^{n} i^2 \cdot \frac{1}{n} - \left(\sum_{i=1}^{n} i \cdot \frac{1}{n}\right)^2 \\ &= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}\end{aligned}$$

We now use this to compute $\mathrm{Var}\left[aX+bY\right]$:

$$\begin{aligned}\mathrm{Var}\left[aX+bY\right] &= \mathrm{Var}\left[aX\right] + \mathrm{Var}\left[bY\right] && \text{because } aX \text{ and } bY \text{ are independent} \\ &= a^2\mathrm{Var}\left[X\right] + b^2\mathrm{Var}\left[Y\right] && \text{Theorem 4.4 in Lecture Notes 24} \\ &= (a^2+b^2)\mathrm{Var}\left[X\right] && \text{because } \mathrm{Var}\left[X\right] = \mathrm{Var}\left[Y\right] \\ &= (a^2+b^2)\frac{n^2-1}{12} && \text{substitution from computation above}\end{aligned}$$

□

(b) $\mathrm{E}\left[\max(X,Y)\right]$

Solution.

First we compute the probability that the maximum of X and Y is equal to i:

$$\begin{aligned}\Pr(\max(X,Y)=i) &= \Pr(X \le i \wedge Y \le i) - \Pr(X < i \wedge Y < i) \\ &= \Pr(X \le i)\Pr(Y \le i) - \Pr(X < i)\Pr(Y < i) \\ &= \frac{i}{n} \cdot \frac{i}{n} - \frac{i-1}{n} \cdot \frac{i-1}{n} \\ &= \frac{2i-1}{n^2}\end{aligned}$$

We now compute the expectation using these probabilities:

$$\begin{aligned}\mathrm{E}\left[\max(X,Y)\right] &= \sum_{i=i}^{n} i \cdot \Pr(\max(X,Y)=i) = \sum_{i=i}^{n} i \cdot \frac{2i-1}{n^2} \\ &= \frac{1}{n^2} \cdot \sum_{i=1}^{n} (2i^2 - i) = \frac{1}{n^2} \cdot \left(2 \cdot \frac{n(2n+1)(n+1)}{6} - \frac{n(n+1)}{2}\right) \\ &= \frac{(n+1)(4n-1)}{6n}\end{aligned}$$

Alternatively, we could also compute it as follows:

$$\begin{aligned}
\mathrm{E}[\max(X,Y)] &= \sum_{i=0}^{\infty} \Pr(\max(X,Y) > i) = \sum_{i=0}^{n-1} (1 - \Pr(\max(X,Y) \leq i)) \\
&= n - \sum_{i=0}^{n-1} \Pr(X \leq i \wedge Y \leq i) = n - \sum_{i=0}^{n-1} \Pr(X \leq i)\Pr(Y \leq i) \\
&= n - \sum_{i=0}^{n-1} \left(\frac{i}{n}\right)^2 = n - \frac{1}{n^2} \cdot \frac{n(n-1)(2n-1)}{6} \\
&= \frac{(n+1)(4n-1)}{6n}
\end{aligned}$$

□

(c) $\mathrm{E}[\min(X,Y)]$

Solution.

Note that $\max(X,Y) + \min(X,Y) = X + Y$. Thus,

$$\begin{aligned}
\mathrm{E}[\min(X,Y)] &= \mathrm{E}[X] + \mathrm{E}[Y] - \mathrm{E}[\max(X,Y)] \\
&= \frac{n+1}{2} + \frac{n+1}{2} - \frac{(n+1)(4n-1)}{6n} \\
&= \frac{(n+1)(2n+1)}{6n}
\end{aligned}$$

□

(d) $\mathrm{E}[|X-Y|]$

Solution.

We could compute this directly. However, note that $|X-Y| = \max(X,Y) - \min(X,Y)$. Thus,

$$\begin{aligned}
\mathrm{E}[|X-Y|] &= \mathrm{E}[\max(X,Y) - \min(X,Y)] \\
&= \mathrm{E}[\max(X,Y)] - \mathrm{E}[\min(X,Y)] \\
&= \frac{(n+1)(4n-1)}{6n} - \frac{(n+1)(2n+1)}{6n} \\
&= \frac{n^2-1}{3n}
\end{aligned}$$

□

(e) $\mathrm{Var}[|X-Y|]$.

Solution.

We start with one of the standard formulas for variance:

$$\mathrm{Var}\left[|X-Y|\right] = \mathrm{E}\left[|X-Y|^2\right] - \mathrm{E}\left[|X-Y|\right]^2 = \mathrm{E}\left[(X-Y)^2\right] - \mathrm{E}\left[|X-Y|\right]^2$$

We could compute $\mathrm{E}\left[(X-Y)^2\right]$ directly, but note that $\mathrm{Var}\left[X-Y\right] = \mathrm{E}\left[(X-Y)^2\right] - \mathrm{E}\left[X-Y\right]^2$, and $\mathrm{E}\left[X-Y\right] = 0$. We can compute $\mathrm{Var}\left[X-Y\right]$ using the result in part (a). Thus,

$$\begin{aligned}
\mathrm{Var}\left[|X-Y|\right] &= \mathrm{Var}\left[X-Y\right] - \mathrm{E}\left[|X-Y|\right]^2 \\
&= (1^2 + (-1)^2) \cdot \frac{n^2-1}{12} - \left(\frac{n^2-1}{3n}\right)^2 \\
&= \frac{(n^2-1)(n^2+2)}{18n^2}
\end{aligned}$$

□

Problem 6 Suppose you are playing the game "Hearts" with three of your friends. In Hearts, all the cards are dealt to the players, in this case the four of you will each have 13 cards.

(a) What is the expectation and variance of the number of hearts in your hand?

Solution.

Let H be the number of hearts in your hand, and let X_i be the indicator random variable for the event that the ith card in your hand is a heart. Then $H = \sum_{i=1}^{13} X_i$. So

$$\mathrm{E}\left[H\right] = \sum_{i=1}^{13} \mathrm{E}\left[X_i\right] = \sum_{i=1}^{13} \Pr(i\text{th card is a heart}) = \sum_{i=1}^{13} \frac{1}{4} = \frac{13}{4}.$$

To compute variance, we first compute $\mathrm{E}\left[H^2\right]$, following the solution to Problem 4c.

$$\mathrm{E}\left[H^2\right] = \mathrm{E}\left[\left(\sum_{i=1}^{13} X_i\right)^2\right] = \sum_i \mathrm{E}\left[X_i\right] + \sum_i \sum_{j \neq i} \mathrm{E}\left[X_i X_j\right]$$

Note that $X_i^2 = X_i$, so $\mathrm{E}\left[X_i^2\right] = 1/4$. When $i \neq j$, $X_i X_j = 1$ if cards i and j are both hearts, and 0 otherwise. The probability that both cards are hearts is $\binom{13}{2}/\binom{52}{2}$, so $\mathrm{E}\left[X_i X_j\right] = \binom{13}{2}/\binom{52}{2} = 3/51$. In the summation above, there are 13 $\mathrm{E}\left[X_i^2\right]$ terms, and $13 \cdot 12 = 156$ terms of the form $\mathrm{E}\left[X_i X_j\right]$ with $i \neq j$. Thus,

$$\mathrm{Var}\left[H\right] = \mathrm{E}\left[H^2\right] - \mathrm{E}\left[H\right]^2 = 13 \cdot \frac{1}{4} + 156 \cdot \frac{3}{51} - (\frac{13}{4})^2 \approx 1.864.$$

□

(b) What is the expectation and variance of the number of suits in your hand?

Solution.

Let N denote the number of suits in a hand. Let X_s be the indicator random variable for the event that there is a spade in the hand. Define X_h, X_d, X_c analogously for the three remaining suits: hearts, diamonds, and clubs. Then $N = X_s + X_h + X_d + X_c$.

The expectation of each of the indicator variables is the probability that a hand contains the corresponding suit. Thus, for $i \in \{s, h, d, c\}$

$$\mathrm{E}[X_i] = 1 - \frac{\binom{52-13}{13}}{\binom{52}{13}} \approx 0.9872,$$

because $\binom{52-13}{13}$ is the number of hands that are missing the particular suit, and $\binom{52}{13}$ is the total number of hands.

$$\mathrm{E}[N] = \mathrm{E}[X_s + X_h + X_d + X_c] = \mathrm{E}[X_s] + \mathrm{E}[X_h] + \mathrm{E}[X_d] + \mathrm{E}[X_c] \approx 4 \cdot 0.9872 = 3.9488.$$

To compute variance, we compute $\mathrm{E}\left[N^2\right]$ directly, and then $\mathrm{Var}[N] = \mathrm{E}\left[N^2\right] - \mathrm{E}[N]^2$. We first calculate $\Pr(N = i)$ for $i = 1, 2, 3, 4$.

$$\Pr(N = 1) = \frac{4}{\binom{52}{13}} \approx 6.299 \times 10^{-12},$$

because there are four possible hands of all one suit out of $\binom{52}{13}$ possible hands of 13 cards.

$$\Pr(N = 2) = \frac{\binom{4}{2}\left[\binom{26}{13} - 2\right]}{\binom{52}{13}} \approx 9.827 \times 10^{-5},$$

because there are $\binom{4}{2}$ ways to choose which two suits. There are $\binom{26}{13}$ ways to choose 13 cards from the 26 possible cards of those two suits. Two of those $\binom{26}{13}$ hands actually contain only one suit.

$$\Pr(N = 3) = \frac{\binom{4}{3}\left[\binom{39}{13} - \binom{3}{2}\binom{26}{13} + 3\right]}{\binom{52}{13}} \approx 0.05097,$$

because there are $\binom{4}{3}$ ways to choose the three suits, and $\binom{39}{13}$ ways to choose 13 cards from the 39 cards of those three suits. But that includes hands with only one or two suits. So, using Inclusion-Exclusion, we subtract the $\binom{3}{2}\binom{26}{13}$ ways to choose two of the three suits and then choose 13 cards from the 26 cards of those suits. This subtracts out the three hands with only one suit twice, so we add these back.

Of course,

$$\Pr(N = 4) = 1 - (\Pr(N = 1) + \Pr(N = 2) + \Pr(N = 3)) \approx 0.9489.$$

Thus,

$$\begin{aligned} \mathrm{E}\left[N^2\right] &= 1 \cdot \Pr(N = 1) + 4 \cdot \Pr(N = 2) + 9 \cdot \Pr(N = 3) + 16 \cdot \Pr(N = 4) \\ &\approx 0 + 0.0004 + 0.4587 + 15.1824 = 15.6415 \end{aligned}$$

This allows us to complete the calculation of the variance:

$$\mathrm{Var}[N] \approx 15.6415 - (3.9488)^2 \approx 0.0485.$$

□

Problem 7 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.

(a) Does the Weak Law of Large Numbers allow us to *predict* what limit, if any, is approached by the expected proportion of heads that turn up as n approaches infinity? Briefly explain.

Solution.

The Weak Law of Large Numbers tells us that the proportion of heads will approach 1/2 if the fair coin was picked, and it will approach 3/4 if the other coin was picked. But it does not tell us anything about which of these two numbers it will approach, as we have no information about which coin is picked.

□

(b) How many tosses suffice to make us 95% confident which coin was chosen? Explain.

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 H_n be the number of heads in the first n flips. We need to flip the coin enough times so that $\Pr(H_n/n > t) \leq 0.05$ if the fair coin was picked, and $\Pr(H_n/n < t) \leq 0.05$ if the biased coin was picked. A natural threshold to choose is 5/8, exactly in the middle of 1/2 and 3/4.

H_n is the sum of independent Bernoulli variables, which each have variance 1/4 for the fair coin and 3/16 for the biased coin. Using Chebyshev's Inequality for the fair coin,

$$\begin{aligned}
\Pr\left(\frac{H_n}{n} > \frac{5}{8}\right) &= \Pr\left(\frac{H_n}{n} - \frac{1}{2} > \frac{5}{8} - \frac{1}{2}\right) = \Pr\left(H_n - \frac{n}{2} > \frac{n}{8}\right) \\
&= \Pr\left(H_n - \mathrm{E}[H_n] > \frac{n}{8}\right) \leq \Pr\left(|H_n - \mathrm{E}[H_n]| > \frac{n}{8}\right) \\
&\leq \frac{\mathrm{Var}[H_n]}{(n/8)^2} = \frac{n/4}{n^2/64} = \frac{16}{n}
\end{aligned}$$

For the biased coin, we have

$$\begin{aligned}
\Pr\left(\frac{H_n}{n} < \frac{5}{8}\right) &= \Pr\left(\frac{3}{4} - \frac{H_n}{n} > \frac{3}{4} - \frac{5}{8}\right) = \Pr\left(\frac{3n}{4} - H_n > \frac{n}{8}\right) \\
&= \Pr\left(\mathrm{E}[H_n] - H_n > \frac{n}{8}\right) \leq \Pr\left(|H_n - \mathrm{E}[H_n]| > \frac{n}{8}\right) \\
&\leq \frac{\mathrm{Var}[H_n]}{(n/8)^2} = \frac{3n/16}{n^2/64} = \frac{12}{n}
\end{aligned}$$

We are 95% confident if these are at most 0.05, which is satisfied if $n \geq 320$.

Because the variance of the biased coin is less that of the fair coin, we can do slightly better if make our threshold a bit bigger, to about 0.634, which gives 95% confidence with 279 coin flips.

Because H_n has a binomial distribution, we can get a much better bound using the estimates from Lecture 21, giving 95% confidence when $n > 42$.

□

Problem 8

(a) Let X be a random variable whose value is an observation drawn uniformly at random from the $\{i \in \mathbb{Z} \mid -n \leq i \leq n\}$. Let $Y = X^2$. Show that $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$. Are X and Y independent?

Solution.

To see that $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$, note that $\mathrm{E}[X] = 0$ and $\mathrm{E}[XY] = \mathrm{E}\left[X^3\right] = 0$ since X and X^3 are both symmetric; that is, $\Pr(X = i) = \Pr(X = -i)$ and $\Pr(X^3 = i) = \Pr(X^3 = -i)$ for all i.

However X and Y are *not* independent, since Y is determined by X. There are many cases that demonstrate this, for example, $\Pr(Y = 0) = 1/(2n+1)$, but $\Pr(Y = 0 \mid X = 0) = 1$.

□

(b) Show that in general for any two random variables X and Y that if $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$ then $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Solution.

$$
\begin{aligned}
\mathrm{Var}[X+Y] &= \mathrm{E}\left[(X+Y)^2\right] - (\mathrm{E}[X+Y])^2 \\
&= \mathrm{E}\left[X^2 + 2XY + Y^2\right] - (\mathrm{E}[X] + \mathrm{E}[Y])^2 \\
&= \mathrm{E}\left[X^2\right] + 2\mathrm{E}[XY] + \mathrm{E}\left[Y^2\right] - \mathrm{E}[X]^2 - 2\mathrm{E}[X]\,\mathrm{E}[Y] - \mathrm{E}[Y]^2 \\
&= \left(\mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2\right) + \left(\mathrm{E}\left[Y^2\right] - \mathrm{E}[Y]^2\right) + 2\left(\mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]\right) \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y].
\end{aligned}
$$

□

Problem 9 Consider the following 3-stage experiment:

In the first stage, I throw a 6-sided die. The number of spots that I get determines how many instances of the second stage to perform.

In the second stage, I throw a 6-sided die. This stage is repeated once for each spot I got in the first stage. The total number of spots that I get in this stage determines how many instances of the third stage to perform.

In the third stage, I throw a 6-sided die. This stage is repeated once for each spot I got in all instances of the second stage.

For example, if the outcome of the first stage is 3, then stage 2 gets played 3 times. Suppose the outcomes of the 3 instances of stage 2 are $1, 3, 4$. Then stage 3 is played $1 + 3 + 4 = 8$ times, and the score is the sum of 8 instances of stage 3.

Assume that all throws are mutually independent.

My score is the total number of spots on all the instances of the third stage.

What is my expected score?

(Analyze this as carefully as you can, using Wald's theorem. Be careful about uses of independence.)

Solution.

Let X be the random variable whose value is the roll in the first stage. Let $Y_1, Y_2, \ldots, Y_X$ be the random variable whose values are the rolls in the second stage. Define $Y = Y_1 + \cdots + Y_X$. Let $Z_1, Z_2, \ldots, Z_Y$ be the random variable whose values are the rolls in the third stage. Define $Z = Z_1 + \cdots + Z_Y$. We want to find $\mathrm{E}[Z]$.

All the rolls are mutually independent, so we can use Wald's theorem to get $\mathrm{E}[Y] = \mathrm{E}[X]\,\mathrm{E}[Y_i] = (7/2)^2$. Since Y is determined by rolls that are independent of the rolls in the third stage, Y and Z_i are independent. So we apply Wald's theorem again to get $\mathrm{E}[Z] = \mathrm{E}[Y]\,\mathrm{E}[Z_i] = (7/2)^3 = 42.875$.

□

Problem 10 Suppose that someone is infected with an unknown incurable contagious disease. Every day he encounters n healthy people and infects each of them with probability p. Encounters are mutually independent. These people in turn meet n people and infect them with probability p and so on.

(a) Write a recurrence for the expected number of sick people on the $(k+1)$-st day in terms of the expected number of sick people on the kth day. Assume that each day every sick person meets n *different* persons.

Solution.

Let the random variable S_k be the number of sick people at the end of the kth day. We want an expression for S_{k+1} in terms of S_k. The base case is $S_0 = 1$.

Let the sick people after day k be numbered $1, \ldots, S_k$, and let the random variable $H_{i,k}$ be the number of (healthy) people infected on day $k+1$ by the ith sick person from day k. We get $S_{k+1} = S_k + \sum_{i=1}^{S_k} H_{i,k}$. $H_{i,k}$ has a binomial distribution with parameters n and p, so $\mathrm{E}[H_{i,k}] = np$, provided $i \leq S_k$. That is, $\Pr(H_{i,k} = x | S_k \geq i) = \Pr(H = x)$, where H is the binomial random variable with parameters n and p. Thus,

$$\begin{aligned}
\mathrm{E}[S_{k+1}] &= \mathrm{E}[S_k] + \mathrm{E}\left[\sum_{i=1}^{S_k} H_{i,k}\right] && \text{by linearity of expectation} \\
&= \mathrm{E}[S_k] + \mathrm{E}[S_k] \cdot \mathrm{E}[H] && \text{by Wald's Theorem} \\
&= \mathrm{E}[S_k](1 + np) && \text{since } \mathrm{E}[H] = np
\end{aligned}$$

If we let $e_k = \mathrm{E}[S_k]$ then $e_0 = 1$ and $e_{k+1} = e_k(1 + np)$.

□

(b) Solve the recurrence of part (a).

Solution.

$e_k = (1 + np)^k$.

□

(c) Now suppose that any sick person recovers from the disease each day with probability r. Find the expected number of sick people on the kth day.

Solution.

Let $B_{i,k}$ be the indicator variable for whether the ith sick person from day k gets better. Then $\mathrm{E}[B_{i,k}] = r$ and $S_{k+1} = S_k + \sum_{i=1}^{S_k}(H_{i,k} - B_{i,k})$. Let B be the Bernoulli random variable with parameter r. Then, as before,

$$\begin{aligned}
\mathrm{E}[S_{k+1}] &= \mathrm{E}[S_k] + \mathrm{E}\left[\sum_{i=1}^{S_k} H_{i,k}\right] - \mathrm{E}\left[\sum_{i=1}^{S_k} B_{i,k}\right] && \text{by linearity of expectation} \\
&= \mathrm{E}[S_k] + \mathrm{E}[S_k] \cdot \mathrm{E}[H] - \mathrm{E}[S_k] \cdot \mathrm{E}[B] && \text{by Wald's Theorem} \\
&= \mathrm{E}[S_k](1 + np - r) && \text{since } \mathrm{E}[H] = np \text{ and } \mathrm{E}[B] = r
\end{aligned}$$

and with this new recurrence, $e_{k+1} = e_k(1 + np - r)$, so $e_k = (1 + np - r)^k$.

□

Problem 11 In this problem, we analyze probabilistically the first stage of bubble sort performed on a list of n integers $\{i_1, \ldots i_n\}$. We assume that all the integers are distinct.

On the first step we compare integers i_1 and i_2. If $i_1 > i_2$ we swap them otherwise we leave them as they are.

On step k ($2 \leq k \leq n-1$), we compare the new i_k with i_{k+1}. If $i_k > i_{k+1}$ we swap else we leave them as they are.

We assume that the order of the integers is chosen uniformly from all possible orderings.

(a) If a number x is randomly (uniformly) chosen from a set of n distinct numbers, what is the probability that x is the largest number in the set?

What is the probability that on the kth step ($2 \leq k \leq n-1$), i_{k+1} is the biggest integer to appear in the list so far?

Solution.

Pr(choosing largest of n numbers) $= 1/n$.

Pr(i_{k+1} in k step is biggest so far) $= 1/(k+1)$.

□

(b) Let C_k be the indicator random variable for the event "$i_{k+1} > i_k$". What is $\mathrm{E}[C_k]$?

Solution.

We can prove by a simple induction that i_k is always the biggest of the first k elements at the beginning of the kth step.

$$\mathrm{E}[C_k] = \Pr(i_{k+1} > i_k) = \Pr(i_{k+1} \text{ is biggest so far}) = \frac{1}{k+1}.$$

□

(c) Let random variable $C = \sum_{k=1}^{n} C_k$. Give an asymptotic bound for $\mathrm{E}[C]$ as well as an interpretation of what $\mathrm{E}[C]$ is.

Solution.

Since we swap iff $i_k > i_{k+1}$, C_k is also the indicator variable for *not* swapping on the kth step. There are $n-1$ steps in a stage of bubble sort, so $n-1-C$ is the total number of swaps in a stage, and $\mathrm{E}[C]$ is the expected number of non-swapping steps. We can compute it as follows:

$$\mathrm{E}[C] = \sum_k \mathrm{E}[C_k] = \sum_{k=2}^{n} \frac{1}{k} \sim \ln n.$$

$\square$