Math 302, Assignment 2 Solutions

1. In a small town, there are three bakeries. Each of the bakeries bakes twelve cakes per day. Bakery 1 has two different types of cake, bakery 2 three different types, and bakery 3 four different types. Every bakery bakes equal amounts of cakes of each type.
   You randomly walk into one of the bakeries, and then randomly buy two cakes.
   (a) What is the probability that you will buy two cakes of the same type?
   (b) Suppose you have bought two different types of cake. Given this, what is the probability that you went to bakery 2?

   Solution: (a) Define the events $F_i = \{\text{choose bakery } i\}$, and $E = \{\text{buy different cakes}\}$. Then $\mathbb{P}(F_i) = \frac{1}{3}$, and we compute the conditional probabilities

$$\mathbb{P}(E|F_1) = \frac{\binom{6}{1}^2}{\binom{12}{2}} = \frac{6}{11}$$

$$\mathbb{P}(E|F_2) = \frac{3 \cdot \binom{4}{1}^2}{\binom{12}{2}} = \frac{8}{11}$$

$$\mathbb{P}(E|F_3) = \frac{6 \cdot \binom{3}{1}^2}{\binom{12}{2}} = \frac{9}{11}$$

   By the law of total probability,

$$\mathbb{P}(E) = \frac{1}{3}\left(\frac{6}{11} + \frac{8}{11} + \frac{9}{11}\right) = \frac{23}{33}.$$

   Therefore $\mathbb{P}(\text{buy same type}) = \frac{10}{33}$.
   (b) By Bayes,

$$\mathbb{P}(F_2|E) = \mathbb{P}(E|F_2)\frac{\mathbb{P}(F_2)}{\mathbb{P}(E)} = \frac{8}{23}$$

2. An assembly line produces a large number of products, of which 1% are faulty in average. A quality control test correctly identifies 98% of the faulty products, and 95% of the flawless products. For every product that is identified as faulty, the test is run a second time, independently.
   (a) Suppose that a product was identified as faulty in both tests. What is the probability that it is, indeed, faulty?
   (b) What if the quality control test is only performed once?

   Solution: Let

$$\begin{aligned} F &= \text{the event that a product is faulty,} \\ E_1 &= \text{the event that the first test result is ``faulty''} \\ E_2 &= \text{the event that the second test result is ``faulty''} \end{aligned}$$

   Then, $\mathbb{P}(F) = 1\%$ and, since the second test is independent of the first,

$$\mathbb{P}(E_1 \cap E_2|F) = \mathbb{P}(E_1|F)\mathbb{P}(E_2|F) = 0.98^2$$

$$\mathbb{P}(E_1 \cap E_2|F^c) = \mathbb{P}(E_1|F^c)\mathbb{P}(E_2|F^c) = 0.05^2$$

By Bayes' theorem,

$$\begin{aligned}\mathbb{P}(F|E_1 \cap E_2) &= \frac{\mathbb{P}(E_1 \cap E_2|F)\mathbb{P}(F)}{\mathbb{P}(E_1 \cap E_2|F)\mathbb{P}(F) + \mathbb{P}(E_1 \cap E_2|F^c)\mathbb{P}(F^c)} \\ &= \frac{0.98^2 \cdot 0.01}{0.98^2 \cdot 0.01 + 0.05^2 \cdot 0.99} \approx 80\%.\end{aligned}$$

If that test had only be run once, we would get

$$\mathbb{P}(F|E_1) = \frac{0.98 \cdot 0.01}{0.98 \cdot 0.01 + 0.05 \cdot 0.99} \approx 17\%,$$

so even a very reliable test cannot identify a rare fault with satisfactory accuracy in a single try.

3. Let m be an integer chosen uniformly from $\{1, \ldots, 100\}$. Decide whether the following events are independent:
(a) $E = \{m \text{ is even}\}$ and $F = \{m \text{ is divisible by } 5\}$
(b) $E = \{m \text{ is prime}\}$ and $F = \{\text{at least one of the digits of } m \text{ is a } 2\}$
(c) Can you replace the number 100 by a different number, in such a way that your answer to (a) changes? (E.g., if your answer was "dependent", try to change the number 100 in such a way your answer becomes "independent").

Solution: (a) $\mathbb{P}(E)\mathbb{P}(F) = (1/2)(1/5) = 1/10 = \mathbb{P}(E \cap F)$, so E and F are independent.
(b) $\mathbb{P}(E) = 25/100$, $\mathbb{P}(F) = 19/100$, and $\mathbb{P}(E \cap F) = 3/100$. The events E and F are not independent since $\mathbb{P}(E \cap F) \neq \mathbb{P}(E)\mathbb{P}(F)$.
(c) Already replacing it by 101 makes the events dependent: $\frac{50}{101} \cdot \frac{20}{101} \neq \frac{10}{101}$. The fact that E and F of (a) were dependent was a pure coincidence, and had nothing to do with number theoretic properties of divisibility by 2 and 5. Whether one could change the 100 in part (b) to make the events mentioned there independent would be a much harder question!

4. Let X be a discrete random variable with values in $\mathbb{N} = \{1, 2, \ldots\}$. Prove that X is geometric with parameter $p = \mathbb{P}(X = 1)$ if and only if the *memoryless property*

$$\mathbb{P}(X = n + m \,|\, X > n) = \mathbb{P}(X = m)$$

holds.
Hint: To show that the memoryless property implies that X is geometric, you need to prove that the p.m.f. of X has to be $\mathbb{P}(X = k) = p(1-p)^{k-1}$. For this, use $\mathbb{P}(X = k) = \mathbb{P}(X = k+1|X > 1)$ repeatedly.

Solution: We first show that a geometric RV has the memoryless property: We learned that $\mathbb{P}(X = m) = p(1-p)^{m-1}$ and that $\mathbb{P}(X > m) = (1-p)^m$, therefore by the definition of conditional probability we obtain

$$\begin{aligned}\mathbb{P}(X = n + m \,|\, X > n) &= \frac{\mathbb{P}(X = n + m)}{\mathbb{P}(X > n)} = \frac{p(1-p)^{n+m-1}}{(1-p)^n} \\ &= p(1-p)^{m-1} = \mathbb{P}(X = m)\end{aligned}$$

Now we show that the memoryless property implies that $\mathbb{P}(X = k) = p(1-p)^{k-1}$ with $p = \mathbb{P}(X = 1)$. Using the law of total probability and the hint, for $k > 1$ we have

$$\begin{aligned} P(X = k) &= P(X = k|X > 1)P(X > 1) + P(X = k|X = 1)P(X = 1) \\ &= P(X = k|X > 1)(1 - P(X = 1)) + 0 \\ &= P(X = k - 1)(1 - P(X = 1)) = \cdots = P(X = 1)(1 - P(X = 1))^{k-1}. \end{aligned}$$

Therefore, X is Geom(p) with $p = P(X = 1)$.

5. In a card game, 13 cards are given to you out of a deck of 52. This game is being played 50 times. Identify (with names and parameters) the following random variables:
(a) The number of games in which all cards you receive have the same suit.
(b) The first time where the number of aces you receive is at least 1.
(c) The number of games in which you receive exactly three aces.
(d) The third time in which you received no aces.

Solution: a) $\mathrm{Bin}(50, 4/\binom{52}{13})$

b) $\mathrm{Geom}(1 - \binom{48}{13}/\binom{52}{13})$

c) $\mathrm{Bin}(50, \binom{48}{10}\binom{4}{3}/\binom{52}{13})$

d) $\mathrm{NegBin}(3, \binom{48}{13}/\binom{52}{13})$

Note that we have been cheating a little bit in b) and d): The geometric/negative binomial random variables would only be applicable if the game is played continuously (rather than "only" 50 times). Since 50 is a rather large number, it might still be a good approximation to use Geom/NegBin. The question was phrased ambiguously to not give the answers away.

6. You have one million \$, but for some reason want to earn an additional \$50. Your strategy is to play roulette at a casino, and always bet \$1 on black, until you own \$1,000,050. What is the probability that you will be successful?
Hint: The probability of black in a single game is $p = \frac{18}{38}$. Prove a recursion relation for the probability P_n of finishing at \$1,000,050, starting with \$$n$.

Solution:

Let $M = 1,000,050\$$ be your goal. For all $0 \leq n \leq M$ let P_n denote the probability that you do indeed end up with M dollars, starting with n dollars. Clearly $P_0 = 0$ and $P_M = 1$. Applying the law of total probability (conditioning on the first flip) we obtain the recursion

$$P_n = pP_{n+1} + (1-p)P_{n-1}$$

for all $1 \leq n \leq M - 1$, with $p = \frac{18}{38}$. That is,

$$(1-p)(P_n - P_{n-1}) = p(P_{n+1} - P_n).$$

Let $a = (1-p)/p$, then clearly

$$P_{n+1} - P_n = a^n(P_1 - P_0).$$

Using $P_0 = 0$ we obtain that

$$P_n = \sum_{k=0}^{n-1}(P_{k+1} - P_k) = (P_1 - P_0)\sum_{k=0}^{n-1} a^k. \tag{0.1}$$

For $n = M$ the boundary condition gives

$$1 = P_M = (P_1 - P_0)\sum_{k=0}^{M-1} a^k,$$

so

$$P_1 - P_0 = \frac{1}{\sum_{k=0}^{M-1} a^k}. \tag{0.2}$$

Therefore (0.3) and (0.4) imply that

$$P_n = \frac{\sum_{k=0}^{n-1} a^k}{\sum_{k=0}^{M-1} a^k} = \frac{a^n - 1}{a^M - 1}.$$

Thus we have

$$P_N = \frac{a^N - 1}{a^M - 1}.$$

Plugging in numbers, $a = \frac{10}{9}$ and

$$P_{1{,}000{,}000} = \frac{(\frac{10}{9})^{1{,}000{,}000} - 1}{(\frac{10}{9})^{1{,}000{,}050} - 1} = (\tfrac{10}{9})^{-50}\frac{1 - (\frac{10}{9})^{-1{,}000{,}000}}{1 - (\frac{10}{9})^{-1{,}000{,}050}} \approx (\tfrac{10}{9})^{-50} = 0.005.$$

Even though you started with 1 million \$, your chances of making it to 1,000,050 dollars are less than 1% (and instead, hitting 0\$ and loosing all your money with this strategy has probability $99,5\%$).

7. Let X take values $\{1, 2, 3, 4, 5\}$, and p.m.f. given by

Table 1: The p.m.f. of X

k	1	2	3	4	5
$\mathbb{P}(X = k)$	1/7	1/14	3/14	2/7	2/7

(a) Calculate $\mathbb{P}(X \le 3)$

(b) Calculate $\mathbb{P}(X < 3)$

(c) Calculate $\mathbb{P}(X < 4.12 | X > 1.6)$

(d) Calculate $\mathbb{E}\,X$

(e) Calculate $\mathbb{E}|X - 2|$

Solution:

a) $\mathbb{P}(X \le 3) = 1/7 + 1/14 + 3/14 = 3/7$.

b) $\mathbb{P}(X < 3) = 1/7 + 1/14 = 3/14$.

c)

$$\begin{aligned}\mathbb{P}(X < 4.12 \,|\, X > 1.6) &= \frac{\mathbb{P}(1.6 < X < 4.12)}{\mathbb{P}(X > 1.6)} = \frac{\mathbb{P}(2 \le X \le 4)}{\mathbb{P}(X \ge 2)} \\ &= \frac{1/14 + 3/14 + 2/7}{1/14 + 3/14 + 2/7 + 2/7} = \frac{2}{3}.\end{aligned}$$

d) $\mathbb{E}(X) = 1 \cdot (1/7) + 2 \cdot (1/14) + 3 \cdot (3/14) + 4 \cdot (2/7) + 5 \cdot (2/7) = \frac{49}{14} = 3.5.$

e) $\mathbb{E}(|X-2|) = |1-2| \cdot (1/7) + |2-2| \cdot (1/14) + |3-2| \cdot (3/14) + |4-2| \cdot (2/7) + |5-2| \cdot (2/7) = 1/7 + 3/14 + 4/7 + 6/7 = \frac{25}{7}.$

8. Consider the following lottery: There are a total of 10 tickets, of which 5 are “win” and 5 are “lose”. You draw tickets until you draw the first “win”. Drawing one ticket costs \$2, 2 tickets \$4, 3 tickets \$8, and so on. A winning ticket pays out \$8.

   (a) Let X be the number of tickets you draw in the lottery (i.e. the number of tickets until the first win, including the winning ticket). Calculate the p.m.f. of X.

   (b) Calculate the expectation $\mathbb{E}\, X$.

   (c) Calculate the variance $\sigma^2(X)$.

   (d) What are your expected winnings in this game?

   Solution:

   a) The values that X can take are $\{1, 2, 3, 4, 5, 6\}$, and

$$\mathbb{P}(X = k) = \frac{5}{10}\frac{4}{9}\frac{3}{8}\cdots\frac{5-(k-2)}{10-(k-2)}\frac{5}{N-(k-1)} = \frac{\binom{10-k}{5-1}}{\binom{10}{5}}$$

   b) We have

$$\mathbb{E}\, X = \sum_{k=1}^{6} k \frac{\binom{10-k}{5-1}}{\binom{10}{5}} = \frac{11}{6}$$

   c) We have

$$\sigma^2(X) = \sum_{k=1}^{6} (k - \tfrac{11}{6})^2 \frac{\binom{10-k}{5-1}}{\binom{10}{5}} = \frac{275}{252}$$

   d) If the game finishes with k balls, then you gain \$8 and pay \$$2^k$. Therefore, the expected gain is

$$\mathbb{E}(8 - 2^X) = \sum_{k=1}^{6} (8 - 2^k) \frac{\binom{10-k}{5-1}}{\binom{10}{5}} = \frac{185}{63}$$

9. Prove the following claims. Here, X, Y are discrete random variables on the same sample space, and $a, b \in \mathbb{R}$.

(a) $\mathbb{E}\,(aX+b) = a\mathbb{E}\,X + b$

(b) $\sigma^2(aX+b) = a^2\sigma^2(X)$

(c) $\sigma^2(X) = \mathbb{E}(X^2) - (\mathbb{E}\,X)^2$

Solution:

(a) According to the theorem stated in the lecture about the expectation of a function of a random variable.

$$\begin{aligned}\mathbb{E}\,(aX+b) &= \sum_k (ak+b)\mathbb{P}(X=k) = a\sum_k k\mathbb{P}(X=k) + b\sum_k \mathbb{P}(X=k)\\ &= a\mathbb{E}\,X + b,\end{aligned}$$

where we used the linearity of the sum, the definition of $\mathbb{E}\,X$, and the normalization property of the p.m.f.

(b) According to (a), the expectation of $aX+b$ is $a\mathbb{E}\,X+b$. Therefore,

$$\begin{aligned}\sigma^2(aX+b) &= \mathbb{E}\big(aX+b-(a\mathbb{E}\,X+b)\big)^2\\ &= \mathbb{E}\big(a(X-\mathbb{E}\,X)\big)^2 = a^2\mathbb{E}\big(X-\mathbb{E}\,X\big)^2 = a^2\sigma^2(X).\end{aligned}$$

(c) We compute

$$\begin{aligned}\sigma^2(X) &= \mathbb{E}\big(X-\mathbb{E}(X)\big)^2 = \mathbb{E}\big(X^2 - 2X\mathbb{E}(X) + (\mathbb{E}\,X)^2\big)\\ &= \mathbb{E}(X^2) - 2(\mathbb{E}\,X)^2 + (\mathbb{E}\,X)^2 = \mathbb{E}(X^2) - (\mathbb{E}\,X)^2\end{aligned}$$

In the second line, we have used $\mathbb{E}(X+Y) = \mathbb{E}\,X + \mathbb{E}\,Y$, but since all expressions are functions of X, we could have also used the theorem from the lecture and linearity of the sum.

10. In a town, there are on average 2.3 children in a family and a randomly chosen child has on average 1.6 siblings. Determine the variance of the number of children in a randomly chosen family.

Solution: Let X be the number of children in a randomly chosen family and let Y be the number of siblings of a randomly chosen child. Let n be the maximal number of children in a family, and assume that there are a_i families with exactly i children for each $i = 0, 1, \ldots, n$. Then the total number of families is $F = \sum_{i=0}^n a_i$ and the total number of children is $C = \sum_{i=0}^n ia_i$. Thus we have

$$P(X=i) = \frac{a_i}{F} \quad \text{and} \quad \mathbb{P}(Y=i-1) = \frac{ia_i}{C}$$

for all $i = 0, \ldots, n$. The definition of mean and our condition give

$$\mathbb{E}(X) = \sum_{i=0}^n i\mathbb{P}(X=i) = \sum_{i=1}^n \frac{ia_i}{F} = \frac{C}{F} = 2.3, \tag{0.3}$$

and similarly

$$\mathbb{E}(Y) = \sum_{i=0}^n (i-1)\mathbb{P}(Y=i-1) = -1 + \sum_{i=0}^n i\mathbb{P}(Y=i-1) = -1 + \sum_{i=0}^n \frac{i^2a_i}{C} = 1.6,$$

so

$$\sum_{i=0}^{n} \frac{i^2 a_i}{C} = 2.6. \tag{0.4}$$

Using (0.3) and (0.4) the second moment of X is

$$\mathbb{E}(X^2) = \sum_{i=0}^{n} i^2 \mathbb{P}(X = i) = \sum_{i=0}^{n} \frac{i^2 a_i}{F} = \frac{C}{F} \sum_{i=0}^{n} \frac{i^2 a_i}{C} = 2.3 \cdot 2.6.$$

Thus

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = 2.3 \cdot 2.6 - 2.3^2 = 2.3 \cdot 0.3 = 0.69.$$