

Math 461 B/C, Spring 2009
Midterm Exam 2 Solutions and Comments

1. Assume A and B are independent events with $P(A) = 0.2$ and $P(B) = 0.3$. Let C denote the event that **none** of the events A and B occurs, and let D be the event that **exactly one** of the events A and B occurs.

- (a) Find $P(C)$.

5 pts

Solution. The event C (“none of A and B occurs”) is the intersection of the complements of A and B . Thus,

$$\begin{aligned} P(C) &= P(A^c B^c) = P(A^c)P(B^c) = (1 - P(A))(1 - P(B)) \\ &= 0.8 \cdot 0.7 = \boxed{0.56} \end{aligned}$$

- (b) Find $P(D)$.

5 pts

Solution. The event D (“exactly one of A and B occurs”) is the union of A and B minus the intersection of A and B , i.e., $A \cup B \setminus AB$. Thus,

$$\begin{aligned} P(D) &= P(A \cup B \setminus AB) = P(A \cup B) - P(A)P(B) = P(A) + P(B) - 2P(AB) \\ &= 0.2 + 0.3 - 2 \cdot 0.2 \cdot 0.3 = 0.38 = \boxed{19/50} \end{aligned}$$

- (c) Find $P(A | D)$.

5 pts

Solution. The event AD is the event “ A occurs, and exactly one of A and B occurs”, which is the same as “ A occurs, but B does not occur”, i.e., $A \setminus AB$. Thus,

$$\begin{aligned} P(A | D) &= \frac{P(AD)}{P(D)} = \frac{P(A \setminus AB)}{P(D)} = \frac{P(A) - P(A)P(B)}{P(D)} \\ &= \frac{0.2 - 0.2 \cdot 0.3}{0.38} = \frac{0.14}{0.38} = \boxed{7/19} \end{aligned}$$

- (d) Find $P(A | D^c)$.

5 pts

Solution. The event AD^c is the event “ A occurs, and either 0 or 2 of the events A and B occur”, which is the same as “ A occurs and B occurs”, i.e., AB . Thus,

$$\begin{aligned} P(A | D^c) &= \frac{P(AD^c)}{P(D^c)} = \frac{P(AB)}{1 - P(D)} \\ &= \frac{0.2 \cdot 0.3}{1 - 0.38} = \frac{0.06}{0.62} = \boxed{3/31} \end{aligned}$$

Comments: As pointed out in class, the complement rule holds for the first argument in a conditional probability, but not for the second argument. Thus, $P(A | D^c)$ is not the same as $1 - P(A | D)$, and essentially the only way to compute $P(A | D^c)$ is by using the definition of a conditional probability as above.

- (e) Are C and D independent? Justify your answer!

5 pts

Solution. We need to check whether $P(CD) \stackrel{?}{=} P(C)P(D)$. Since the events C (“none of A and B occurs”) and D (“exactly one of A and B occurs”) are mutually exclusive, we have $P(CD) = 0$. On the other hand, from above $P(C)P(D) = 0.56 \cdot 0.38 \neq 0$, so $P(CD) \neq P(C)P(D)$ and C and D are therefore **not independent**.

Comments: Independence of two events A and B is defined by the condition $(*) P(AB) = P(A)P(B)$ (or the equivalent condition $P(A | B) = P(A)$). This is the condition that needs to be checked in order to determine whether two events are independent. In particular, it is not enough to say that, since C and D are disjoint, they cannot be independent. (In fact, it is possible for two disjoint events to be independent, namely if one of them has probability 0; in this case both sides of $(*)$ are 0, so, by the definition $(*)$, the events are independent.)

2. Three Democrats, D_1 , D_2 , and D_3 , are trying to seek their party’s nomination as the Democratic candidate to run against the Republican R . According to polls the chances of being nominated are 35% for D_1 , 40% for D_2 , and 25% for D_3 . Furthermore, the chances of R of winning the election are 40% if he runs against D_1 , 35% if he runs against D_2 , and 60% if he runs against D_3 .

- (a) Express all given percentage figures as probabilities (ordinary or conditional) involving the events D_i (“Democrat i is nominated”) and R (“Republican candidate wins the election”).

5 pts

Solution. The given probabilities are:

$$P(D_1) = 0.35, \quad P(D_2) = 0.4, \quad P(D_3) = 0.25,$$

$$P(R | D_1) = 0.4, \quad P(R | D_2) = 0.35, \quad P(R | D_3) = 0.6.$$

- (b) State the general form of the average rule (involving events A and B_1, \dots, B_n) **along with any conditions that have to be satisfied for the rule to be applicable**. Then use this rule to compute the probability that R wins the election.

10 pts

Solution. The average rule states that

$$P(A) = \sum_{i=1}^n P(A | B_i)P(B_i)$$

and is valid **provided the B_i ’s form a partition of the sample space**. Applying this rule with $A = R$ and $B_i = D_i$ gives

$$P(R) = \sum_{i=1}^3 P(R | D_i)P(D_i) = 0.4 \cdot 0.35 + 0.35 \cdot 0.4 + 0.6 \cdot 0.25 = \boxed{0.43}.$$

Thus, the probability that R wins the election is 43%.

Comment: The condition that must be satisfied in order for the rule to hold is stated in boldface. For full credit this condition must be stated.

- (c) State the general form of Bayes' rule **along with any conditions that have to be satisfied for the rule to be applicable**. Then use this rule to compute the probability that, if R wins the election, his Democratic opponent was D_3 .

10 pts

Solution. Bayes' rule states that

$$P(B_i | A) = \frac{P(A | B_i)P(B_i)}{\sum_{j=1}^n P(A | B_j)P(B_j)},$$

and is valid **provided the B_i 's form a partition of the sample space**.

Applying this rule with $A = R$ and D_3 in place of B_i gives

$$P(D_3 | R) = \frac{P(R | D_3)P(D_3)}{P(R)} = \frac{0.6 \cdot 0.25}{0.43} = \boxed{\frac{15}{43}} (= 0.348837\dots).$$

3. For this problem suppose you have a fair **four-sided** die with faces labeled with 0, 1, 2, 3.

- (a) If X denotes the number showing up on a *single* roll, find $E(2^X)$ and $\text{Var}(2^X)$. (You can leave the answer as a sum/difference of rationals, such as $16/7 - 256/155$.)

10 pts

Solution. Using the formula $E(g(X)) = \sum_{x_i} g(x_i)p(x_i)$, we get

$$\begin{aligned} E(2^X) &= 2^0 \cdot \frac{1}{4} + 2^1 \cdot \frac{1}{4} + 2^2 \cdot \frac{1}{4} + 2^3 \cdot \frac{1}{4} = \boxed{\frac{15}{4}} \\ E((2^X)^2) &= 4^0 \cdot \frac{1}{4} + 4^1 \cdot \frac{1}{4} + 4^2 \cdot \frac{1}{4} + 4^3 \cdot \frac{1}{4} = \boxed{\frac{85}{4}} \\ \text{Var}(2^X) &= E((2^X)^2) - E(2^X)^2 = \frac{85}{4} - \left(\frac{15}{4}\right)^2 = \boxed{\frac{115}{16}}. \end{aligned}$$

- (b) For this and the remaining parts suppose you roll this four-sided die repeatedly, and are interested in getting 3's. How many times do you have to roll in order to be at least 98% certain to get a 3 at least once? Determine the minimal number of rolls needed with this property. (As usual, you can leave the answer in "raw" form.)

10 pts

Solution. Let n denote the required number of rolls. Interpreting the rolls as S/F trials with success meaning "get a 3" and having probability $p = 1/4$, we need to choose the smallest n so that

$$P(\text{at least one success in } n \text{ trials}) \geq 0.98,$$

Taking complements, this is equivalent to

$$\begin{aligned} P(\text{no success in } n \text{ trials}) &\leq 0.02, \\ (1 - 1/4)^n &\leq 0.02, \\ n \ln(3/4) &\leq \ln 0.02, \\ n &\geq \ln 0.02 / \ln 0.75. \end{aligned}$$

(Note that the direction of the inequality reverses in the last step, since we divide by $\ln(3/4)$, which is a negative number.) The minimal integer n satisfying this inequality is $n = \boxed{\lceil \ln 0.02 / \ln 0.75 \rceil}$.

- (c) Find the probability that the 100th 3 comes up at the 461st roll.

10 pts

Solution. Modeling the rolls as S/F trials with S meaning “get a 3” and occurring with probability $p = 1/4$, we want the probability that the 100-th success occurs at the 461st trial. By the negative binomial distribution this probability is given by

$$\binom{461-1}{100-1} \left(\frac{1}{4}\right)^{100} \left(\frac{3}{4}\right)^{361} = \binom{460}{99} \left(\frac{1}{4}\right)^{100} \left(\frac{3}{4}\right)^{361}.$$

Comments: This type of problem (asking for the probability that the k -th success occurs at the n -th trial) came up several times in class and in the HW problems (e.g., in connection with the baseball world series problem). It is important to distinguish the probability asked (“100-th success occurs at 461-st trial”) from that of getting 100 successes in 461 trials. The latter probability is given by the binomial distribution, $\binom{461}{100}(1/4)^{100}(3/4)^{361}$. Even though this looks superficially similar to the correct answer, it is based on an approach that is completely wrong and misses the main point of the problem; no credit was given for such an answer.

- (d) Find the probability that the *first* 3 occurs at an *odd-numbered* roll, i.e., one of the rolls 1, 3, 5, 7, . . .

10 pts

Solution. Let X denote the roll at which the first 3 occurs. Then X has geometric distribution with $p = 1/4$, so $P(X = k) = (1 - p)^{k-1}p$ for $k = 1, 2, \dots$. The probability to compute is therefore

$$\begin{aligned} P(X = 1) + P(X = 3) + P(X = 5) + \dots \\ &= (1 - p)^0 p + (1 - p)^2 p + (1 - p)^4 p + \dots \\ &= \frac{p}{1 - (1 - p)^2} = \frac{1/4}{1 - (3/4)^2} = \boxed{\frac{4}{7}} \end{aligned}$$

by the geometric series formula.

Comments: This type of problem was discussed in class on two occasions: in connection with S/F trials and the geometric distribution, and as an illustration of the conditioning technique. The approach via the geometric distribution is the one used above. This is the most straightforward approach, but it requires summing a geometric series. The approach via conditioning is conceptually more difficult, but avoids geometric series. For this approach imagine two players alternating rolling the die; whoever rolls the first 3 wins the game. If we denote by A the event that the player who rolls first wins the game and $B (= A^c)$ the event that the player who rolls second wins the game, then the probability asked is $P(A)$. Conditioning on the result of the first roll, we have

$$\begin{aligned} P(A) &= P(A \mid 3 \text{ in 1st roll}) \cdot P(3 \text{ in 1st roll}) + P(A \mid \text{no 3 in 1st roll}) \cdot P(\text{no 3 in 1st roll}) \\ &= 1 \cdot \frac{1}{4} + P(B) \cdot \frac{3}{4} \\ &= 1 \cdot \frac{1}{4} + (1 - P(A)) \cdot \frac{3}{4} = 1 - \frac{3}{4}P(A). \end{aligned}$$

Solving this equation for $P(A)$, we get $P(A) = 4/7$, which is the same answer as the one obtained above.

4. A box contains 100 computer chips, of which exactly 5 are good and the remaining 95 are bad. *The following three questions are independent of each other.*

- (a) For this part suppose you take out chips one at a time **without replacement** and test each chip until you have found a good one. Let X be the number of chips you have to take out in order to find one that is good. Find the p.m.f. of X . (Be sure to indicate the range (i.e., the set of values of X), along with any formula for the probabilities.)

10 pts

Solution. The values of X are $1, 2, \dots, 96$. To compute the probabilities $P(X = k)$ for these values, note that $X = k$ means by definition that the k th chip taken out is the first good one, or, equivalently, that the first $k - 1$ chips are bad, but the k th chip is good. The total number of ways to draw k chips out of 100 **without replacement and taking order into account** is $100 \cdot 99 \cdots (100 - k + 1)$. The number of such draws in which the first $k - 1$ chips are among the 95 bad chips and the k th is among the 5 good chips is $95 \cdot 94 \cdots (95 - k + 2) \cdot 5$. (When $k = 1$, the product in front of the 5 is an empty product and should be interpreted as 1, so the formula reduces to 5 in this case.) Thus,

$$p(k) = P(X = k) = \frac{95 \cdot 94 \cdots (95 - k + 2) \cdot 5}{100 \cdot 99 \cdots 100 - k + 1} \quad (k = 1, 2, \dots, 96).$$

Comments: The point of this problem was to illustrate the different techniques that need to be used depending on whether the sampling is done with replacement (as in (b)) or without replacement (as in (a) and (c)) and on whether order matters (as in (a)) or does not matter (as in (c)), an issue that was discussed in some detail in class.

For part (a), one has to use **ordered** samples in order to be able to take into account the fact that the good chip found must be the **last** of the k chips checked. An answer like $\binom{5}{1} \binom{95}{k-1} / \binom{100}{k}$ would represent the probability that among k chips exactly one, **but not necessarily the last one**, is good.

- (b) Answer the same question, now assuming that the sampling is done **with replacement**. That is, you take out chips one at a time, test each chip, **then replace it**, until you have found a good one. Let X be the number of chips you have to test in order to find one that is good. Find the p.m.f. of X , again making sure to include the range.

10 pts

Solution. If the drawing is with replacement, then after each draw we are back to the exact same situation in which we started out, with 5 out of 100 chips good and 95 out of 100 chips bad to draw from. Thus, in this case the drawings can be modelled by independent S/F trials, with “success” meaning that a good chip has been drawn and occurring with probability $5/100 = 0.05$. The random variable X then represents the trial at which the first success occurs, and therefore has geometric distribution with $p = 0.05$. Hence, X has values $k = 1, 2, \dots$ with probabilities

$$p(k) = P(X = k) = 0.95^{k-1} 0.05 \quad (k = 1, 2, \dots).$$

- (c) For this part suppose that you take out exactly 10 chips **without replacement** and then test each of these 10 chips. Let X denote the number of good chips among the 10 you have taken out. Find the p.m.f. of X (again making sure to include the range).

10 pts

Solution. In this case, the values are $0, 1, 2, \dots, 5$. The corresponding probability $P(X = k)$ is a standard box/ball type probability, namely the probability that in a sample of 10 balls from a box with 100 balls, 5 of which are red (corresponding to the good chips) and 95 of which are blue (corresponding to the bad chips, exactly k balls are red:

$$p(k) = P(X = k) = \frac{\binom{5}{k} \binom{95}{10-k}}{\binom{100}{10}} \quad (k = 0, 1, \dots, 5).$$

(This is the hypergeometric distribution.)

Comments: The appropriate model here is that of an urn/ball type problem. A success/failure model would be inappropriate here since the proportion of good chips (and hence the probability of getting a good chip) changes every time a chip is taken out.

5. A lottery has a $1/300$ chance of winning a prize. You play the lottery 200 times.

- (a) What is the probability that you win at least twice? Give an **exact** answer. (As always, your answer can be in raw/unevaluated form, but should not involve a sum of a large number of terms.)

10 pts

Solution. Interpreting the 200 lottery games as success/failure trials with success probability $p = 1/300$, we need to compute the probability $P(\geq 2)$, where “ ≥ 2 ” stands for “at least two successes”.

Taking the complement, this becomes $P(\geq 2) = 1 - P(0) - P(1)$, and the latter can be computed via the binomial distribution with $n = 200$ and $p = 1/300$:

$$\begin{aligned} P(\geq 2) &= 1 - P(0) - P(1) \\ &= 1 - \binom{200}{0} \left(\frac{299}{300}\right)^{200} - \binom{200}{1} \left(\frac{1}{300}\right)^1 \left(\frac{299}{300}\right)^{199}. \end{aligned}$$

- (b) Answer the same question (i.e., find the probability that you win at least twice), but now using a suitable approximation.

10 pts

Solution. [This is of the same type as Problem 65(a), Chapter 4, from HW 6.] Since $p = 1/300$ is small and of the order of $1/n = 1/200$, Poisson approximation is ideal for this situation. We have $\lambda = np = 2/3$, so Poisson approximation gives for the above probability

$$\begin{aligned} P(\geq 2) &= 1 - P(0) - P(1) \\ &\approx 1 - e^{-2/3} \frac{(2/3)^0}{0!} - e^{-2/3} \frac{(2/3)^1}{1!} \\ &= 1 - \frac{5}{3} e^{-2/3} \end{aligned}$$

- (c) Suppose you know that you have won at least once. What is the probability that you have won at least twice? Use a suitable approximation to answer the question.

10 pts

Solution. The probability sought is the conditional probability $P(\geq 2 \mid \geq 1)$, where “ ≥ 2 ” and “ ≥ 1 ” denote the events “win at least twice” and “win at least once”. Using again Poisson approximation, the probability asked is

$$\begin{aligned} P(\geq 2 \mid \geq 1) &= \frac{P(\geq 2)}{P(\geq 1)} = \frac{1 - P(0) - P(1)}{1 - P(0)} \\ &\approx \frac{1 - \frac{5}{3}e^{-2/3}}{1 - e^{-2/3}} \end{aligned}$$

Comments: This part is of the same type as part (b) of Problem 65 from HW 6, which asked for the conditional probability that more than one has a disease given that at least one person is known to have the disease. Part (c) of Problem 65, however, is completely different: In this part, a **specific** person (Mr. Jones) knows that he has the disease, and he asks what the probability is that at least one other person has the disease. In the lottery context, this would be like asking for the probability that you win at least one other time if you know that you have won at a **specific** trial (say, at your first lottery game). This is not the situation we have here.

Maximal Total Score: 150 points