

# Statistics 100A

## Homework 4 Solutions

Ryan Rosario

### Chapter 4

39. *A ball is drawn from an urn containing 3 white and 3 black balls. After the ball is drawn, it is then replaced and another ball is drawn. This goes on indefinitely. What is the probability that of the first 4 balls drawn, exactly 2 are white?*

First, note that we have a finite number of trials,  $n = 4$  (although the game goes on forever, we are only concerned with the first 4 balls.) Each trial is a Bernoulli trial - that is, each trial has only one of two outcomes: white and not white. Define a success as the event that a white ball is drawn. Then the probability of success  $p$  is  $p = \frac{1}{2}$ . Since each ball is replaced after it is drawn, we have sampling with replacement, and thus independence.

**Since we are dealing with a finite number of independent Bernoulli trials with a constant probability of success  $p$ , we use the *binomial* distribution.**

Let  $X$  be the number of white balls (successes) that appear in  $n = 4$  trials. Then we want to find  $P(X = 2)$ .

Recall that

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Then,

$$\begin{aligned} P(X = 2) &= \binom{4}{2} \left(\frac{1}{2}\right)^2 \left(1 - \frac{1}{2}\right)^{4-2} \\ &= \binom{4}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 \\ &= \frac{6}{2^4} \\ &= \frac{3}{8} \\ &= \boxed{0.375} \end{aligned}$$

41. *A man claims to have extrasensory perception. As a test, a fair coin is flipped 10 times and the man is asked to predict the outcome in advance. He gets 7 out of 10 correct. What is the probability that he would have done at least this well if he had no ESP?*

At first this problem may seem confusing. We have  $n = 10$  trials and each flip of a coin is independent. But what is the probability of success? Take a closer look at the problem. **What is the probability that he would have done at least this well if he had no ESP?** We solve this problem on the premise that he does not have ESP. Therefore,  $p = \frac{1}{2}$ , the probability of randomly guessing the outcome. We define a success as the event that the man guesses correctly.

Since we have a finite number of independent Bernoulli trials with constant probability of success, **we use the binomial**.

Let  $X$  be the number of times the man guesses the outcome correctly (the number of successes). We want to know the probability that he does at least as well as he did in the observed study where he guessed correctly 7 times out of 10. Then,

$$\begin{aligned}
 P(X \geq 7) &= P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) \\
 &= \binom{10}{7} \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^3 + \binom{10}{8} \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^2 + \binom{10}{9} \left(\frac{1}{2}\right)^9 \left(\frac{1}{2}\right)^1 \\
 &\quad + \binom{10}{10} \left(\frac{1}{2}\right)^{10} \left(\frac{1}{2}\right)^0 \\
 &= \binom{10}{7} \left(\frac{1}{2}\right)^{10} + \binom{10}{8} \left(\frac{1}{2}\right)^{10} + \binom{10}{9} \left(\frac{1}{2}\right)^{10} + \binom{10}{10} \left(\frac{1}{2}\right)^{10} \\
 &= \left(\frac{1}{2}\right)^{10} \left[ \binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10} \right] \\
 &= \frac{176}{1024} \\
 &= \boxed{0.1719}
 \end{aligned}$$

If the man did not have ESP, then we would expect him to guess at least 7 out of the 10 outcomes correctly approximately 17.2% of the time.

For those that have taken introductory statistics, this problem can be formulated as a hypothesis test. The null hypothesis is that the man does not have ESP and the probability of guessing an outcome correctly is 0.5. The alternative hypothesis is that the man has ESP and can guess correctly more than 50% of the time. That is,

$$\begin{cases} H_0 : p = 0.5 \\ H_a : p > 0.5 \end{cases}$$

The probability we computed in this problem is called a **p-value**. **The p-value in this context is the probability that we observe a sample proportion at least as extreme as the value we observed ( $\hat{p} = 0.7$ ) assuming that the null hypothesis  $H_0$  is true.** The hypothesis test that we could have performed would have been the binomial test, not a proportions test, because the success/failure condition fails.

48. *It is known that diskettes produced by a certain company will be defective with probability .01, independent of each other. The company sells the diskettes in packages of size 10 and offers a money-back guarantee that at most 1 of the 10 diskettes in the package will be defective. If someone buys 3 packages, what is the probability that he or she will return exactly 1 of them?*

Consider this problem hierarchically. There are two levels to this problem that must be solved: within package, and between packages.

### Level 1: Within Package

First, let's find the probability that we return an arbitrary package. We return the package if more than 1 diskette is defective. Each package contains 10 independent diskettes, that is,  $n = 10$ . The probability that a diskette is defective is constant,  $p = 0.01$ . Since  $p$  is small, we could use the **Poisson**, but we use the **Binomial** since the problem does not ask for an *approximate* probability.

If we were to use the Poisson, we would fix the mean,  $\lambda = np = 10 \cdot 0.01 = 0.1$ .

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) \\ &= 1 - P(X = 1) - P(X = 0) \\ &= 1 - \frac{e^{-0.1}(0.1)^1}{1!} - \frac{e^{-0.1}(0.1)^0}{0!} \\ &= 1 - 1.1e^{-0.1} \\ &= 0.0047 \end{aligned}$$

Using the binomial is the preferred method.

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) \\ &= 1 - P(X = 1) - P(X = 0) \\ &= 1 - \binom{10}{1} (0.01)^1 (0.99)^9 - \binom{10}{0} (0.01)^0 (0.99)^{10} \\ &= \boxed{0.0043} \end{aligned}$$

The probability of returning an arbitrary package of 10 diskettes is 0.0043.

### Level 2: Between Packages

We have  $n = 3$  packages, which should be independent. We know that the probability of returning an arbitrary 10-pack of diskettes is 0.0043 and this is the probability of success,  $p = 0.0043$  and it is constant. Let  $X$  be the number of packages out of 3 that we return. We want to find  $P(X = 1)$ . By the **binomial** model,

$$\begin{aligned} P(X = 1) &= \binom{3}{1} (0.0043)^1 (0.9957)^2 \\ &= \boxed{0.0128} \end{aligned}$$

52. The monthly worldwide average number of airplane crashes of commercial airlines is 3.5. What is the probability that there will be

First we need to decide which probability model to use. All we know is the mean, which does not tell us much. However, the *context* of the problem implies which model to use. An airline crash is very rare, thus it has a small probability of occurring. We use the **Poisson** distribution for this situation.

For the Poisson distribution, recall that

$$P(X = i) = \frac{e^{-\lambda} \lambda^i}{i!}$$

Since the mean is 3.5, and the mean is  $\lambda$ , we know that  $\lambda = 3.5$ .

- (a) at least 2 such accidents in the next month?

$$\begin{aligned} P(X \geq 2) &= 1 - P(X = 1) - P(X = 0) \\ &= 1 - \frac{e^{-3.5} (3.5)^1}{1!} - \frac{e^{-3.5} (3.5)^0}{0!} \\ &= 1 - 3.5e^{-3.5} - e^{-3.5} \\ &= 1 - 4.5e^{-3.5} \\ &= \boxed{0.8641} \end{aligned}$$

- (b) at most 1 accident in the next month?

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= \frac{e^{-3.5} (3.5)^0}{0!} + \frac{e^{-3.5} (3.5)^1}{1!} \\ &= 4.5e^{-3.5} \\ &= \boxed{0.1359} \end{aligned}$$

53. Approximately 80,000 marriages took place in the state of New York last year. Estimate the probability that for at least one of these couples

(a) both partners were born on April 30.

The probability that both partners share the same birthday can be found by independence:

$$P(\text{Same Birthday, April 30}) = \frac{1}{365} \cdot \frac{1}{365} = 7.51 \times 10^{-6}$$

which is a *tiny* probability, so we use the **Poisson** approximation to the binomial.

The number of couples that share the same birthday, April 30, is Poisson with mean  $\lambda = np = 80,000 \cdot 7.51 \times 10^{-6} \approx 0.6$ .

Then, let  $X$  be the number of couples that share April 30 as their birthday.

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - \frac{e^{-0.6} (-0.6)^0}{0!} \\ &= 1 - e^{-0.6} \\ &= \boxed{0.4512} \end{aligned}$$

(b) both partners celebrated their birthday on the same day of the year.

This problem is different! It is a generalization of part (a). Now want to find the probability that both partners celebrated their birthday on the same day of the year, but not just April 30. The computation is similar, however, we note that we can choose from one of 365 days of the year for both partners,

$$P(\text{Same Birthday}) = 365 \cdot \frac{1}{365} \cdot \frac{1}{365} = \frac{1}{365}$$

Again, this probability is tiny so we use the **Poisson** with mean  $\lambda = np = 80,000 \cdot \frac{1}{365} = 219.2$ .

Then, let  $X$  be the number of couples that share the same birthday.

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - \frac{e^{-219.2} (219.2)^0}{0!} \\ &= 1 - e^{-219.2} \\ &\approx \boxed{1} \end{aligned}$$

59. If you buy a lottery ticket in 50 lotteries, in each of which your chance of winning a prize is  $\frac{1}{100}$ , what is the (approximate) probability that you will win a prize

We have a finite number of independent trials  $n = 50$  with a given constant probability of success (winning a prize)  $p = \frac{1}{100}$  which is small. We could use the binomial model for an exact answer, or the Poisson model for an approximate answer. Since the problem asks for an approximation, we use the **Poisson**.

- (a) *at least once?*

Using the Poisson approximation,

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - \frac{e^{-\frac{1}{2}} \left(\frac{1}{2}\right)^0}{0!} \\ &= 1 - e^{-\frac{1}{2}} \\ &\approx \boxed{0.393} \end{aligned}$$

- (b) *exactly once?*

$$\begin{aligned} P(X = 1) &= \frac{e^{-\frac{1}{2}} \left(\frac{1}{2}\right)^1}{1!} \\ &= \frac{1}{2} e^{-\frac{1}{2}} \\ &\approx \boxed{0.3033} \end{aligned}$$

- (c) *at least twice?*

$$\begin{aligned} P(X \geq 2) &= 1 - P(X = 0) - P(X = 1) \\ &= \underbrace{0.393}_{\text{Part (a)}} - \underbrace{0.3033}_{\text{Part (b)}} \\ &\approx \boxed{0.0897} \end{aligned}$$

71. Consider a roulette wheel consisting of 38 numbers - 1 through 36, 0 and double 0. If Smith always bets that the outcome will be one of the numbers 1 through 12, what is the probability that

Since Smith always bets on the numbers 1 through 12, which occupy 12 spaces on the wheel, the probability of success (winning) is  $p = \frac{12}{38}$ .

- (a) *Smith will lose his first 5 bets?*

The first five bets form a finite set of  $n = 5$  trials. Each spin of the roulette wheel is independent, and the probability of success  $p$  is constant. **We use the binomial model.** We could also solve this problem intuitively by using independence.

Let  $W$  denote the event that Smith wins his bet. Using independence, the probability that Smith loses his first 5 bets is,

$$P(W'W'W'W'W') = \left(\frac{26}{38}\right)^5 \approx \boxed{0.15}$$

Using the binomial, let  $X$  be the number of bets that Smith wins. Then,

$$\begin{aligned} P(X = 0) &= \binom{5}{0} \left(\frac{12}{38}\right)^0 \left(\frac{26}{38}\right)^5 \\ &= \left(\frac{26}{38}\right)^5 \\ &\approx \boxed{0.15} \end{aligned}$$

(b) *his first win will occur on his fourth bet?*

There are three equivalent ways to formulate this problem. First, let's tackle this problem intuitively. Smith's first win occurring on the fourth bet means that he loses the first 3 bets and wins on the fourth bet. That is,

$$P(W'W'W'W) = \left(\frac{26}{38}\right) \left(\frac{26}{38}\right) \left(\frac{26}{38}\right) \left(\frac{12}{38}\right) = \left(\frac{26}{38}\right)^3 \left(\frac{12}{38}\right) \approx \boxed{0.1012}$$

We could also solve this problem using the **Geometric** distribution. Let  $X$  be the number of the trial of which the first success (win) occurs. Then,

$$P(X = i) = (1 - p)^{i-1}p$$

Thus,

$$\begin{aligned} P(X = 4) &= \left(\frac{26}{38}\right)^3 \left(\frac{12}{38}\right) \\ &\approx \boxed{0.1012} \end{aligned}$$

We could also interpret the problem as follows “the  $n$ th trial is the  $r$ th win.” We can use the **negative binomial** distribution. But since  $r = 1$ , it reduces to the **geometric**.

Let  $X$  be the number of trials required to obtain the  $r$ th success, then

$$P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n = r, r+1, \dots$$

Let  $n = 4, r = 1$ .

$$\begin{aligned}
P(X = 4) &= \binom{4-1}{1-1} \left(\frac{12}{38}\right)^1 \left(1 - \left(\frac{12}{38}\right)\right)^{4-1} \\
&= \binom{3}{0} \left(\frac{12}{38}\right) \left(\frac{26}{38}\right)^3 \\
&= \left(\frac{12}{38}\right) \left(\frac{26}{38}\right)^3 \\
&\approx \boxed{0.1012}
\end{aligned}$$

76. Solve the Banach match problem when the left-hand matchbox originally contained  $N_1$  matches and the right-hand box contained  $N_2$  matches.

First, consider the case where the right matchbox is the first to run out of matches. We started with  $N_2$  matches and ended up with 0. The left matchbox started with  $N_1$  matches and ended up with  $k$ . We note that the  $N_2 + 1$ st match is drawn from the right matchbox on the  $N_2 + 1 + N_1 - k$ th trial. Thus,  $n = N_1 + N_2 + 1 - k$ ,  $r = N_2 + 1$  and  $p = \frac{1}{2}$ . By using the **negative binomial** model, we get the following if we assume that  $E$  is the event that the right matchbox is the first one discovered empty,

$$\begin{aligned}
P(E) &= \binom{N_1 + N_2 - k}{N_2} \left(\frac{1}{2}\right)^{N_2+1} \left(\frac{1}{2}\right)^{N_1 + N_2 + 1 - k - (N_2 + 1)} \\
&= \binom{N_1 + N_2 - k}{N_2} \left(\frac{1}{2}\right)^{N_1 + N_2 - k + 1}
\end{aligned}$$

But that is only one case. Now consider the case that the left matchbook is the first one discovered empty. Denote this event  $E'$ . We started with  $N_1$  matches and ended up with 0. The right matchbox started with  $N_2$  matches and ended up with  $k$ . We note that the  $N_1 + 1$ st match is drawn from the left matchbox on the  $N_1 + 1 + N_2 - k$ th trial. Thus,  $n = N_1 + N_2 + 1 - k$ ,  $r = N_1 + 1$  and  $p = \frac{1}{2}$ . Then,

$$\begin{aligned}
P(E') &= \binom{N_1 + N_2 - k}{N_1} \left(\frac{1}{2}\right)^{N_1+1} \left(\frac{1}{2}\right)^{N_1 + N_2 + 1 - k - (N_1 + 1)} \\
&= \binom{N_1 + N_2 - k}{N_1} \left(\frac{1}{2}\right)^{N_1 + N_2 - k + 1}
\end{aligned}$$

The events  $E$  and  $E'$  are disjoint so we can add the probabilities together to yield the total probability for this variation of the Banach match problem.

$$\boxed{P(E) + P(E') = \binom{N_1 + N_2 - k}{N_2} \left(\frac{1}{2}\right)^{N_1 + N_2 - k + 1} + \binom{N_1 + N_2 - k}{N_1} \left(\frac{1}{2}\right)^{N_1 + N_2 - k + 1}}$$



79. Suppose that a batch of 100 items contains 6 that are defective and 94 that are non-defective. If  $X$  is the number of defective items in a randomly drawn sample of 10 items from the batch, find

This is a classic example of the **hypergeometric** model. We have some population of  $N$  items,  $m$  of which are “hot” and we draw a sample of size  $n$  from this population. We want to find the probability that the drawn sample of  $n$  contains  $X$  “hot” items. Let  $X$  be the number of “hot” items drawn.

In this context,  $N = 100$ ,  $n = 10$ ,  $m = 6$ .

$$P(X = i) = \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}}$$

(a)  $P(X = 0)$

$$P(X = 0) = \frac{\binom{6}{0} \binom{94}{10}}{\binom{100}{10}} = \boxed{0.522}$$

(b)  $P(X > 2)$

$$\begin{aligned} P(X > 2) &= 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1 - 0.522 - \frac{\binom{6}{1} \binom{94}{9}}{\binom{100}{10}} - \frac{\binom{6}{2} \binom{94}{8}}{\binom{100}{10}} \\ &= 1 - 0.522 - 0.369 - 0.0965 \\ &= \boxed{0.0125} \end{aligned}$$

## Chapter 4 Theoretical Exercises

35. An urn initially contains one red and one blue ball. At each stage a ball is randomly chosen and then replaced along with another of the same color. Let  $X$  denote the selection number of the first chosen ball that is blue. For instance, if the first selection is red and the second blue, then  $X$  is equal to 2.

First, we must understand what is going on here. The urn contains a red ball and a blue ball. We draw a ball and note its color. Then we replace the ball *and add a new ball of the same color*.

- (a) Find  $P(X > i), i \geq 1$ .

This problem is difficult. We want to find  $P(X > i)$ , that is, the probability that the first draw of a blue ball is *after* the  $i$ th trial. All we know from this is that on the first  $i$  draws we drew a red ball.

On trial  $i = 1$ , there is a red ball and a blue ball so the probability of drawing a red ball is  $\frac{1}{2}$ . On trial  $i = 2$ , there is an additional red ball added into the urn from draw 1, so the probability of drawing a red ball on the second draw after drawing a red ball on the first draw is  $\frac{2}{3}$ . Then we replace this red ball *and* add in another red ball; so for draw  $i = 3$  there are 4 balls, 3 which are red, so the probability of drawing a red ball after already drawing 2 red balls is  $\frac{3}{4}$ . Therefore, on the  $i$ th trial, there is a probability of  $\frac{i}{i+1}$  of drawing a red ball after having drawn red balls on each of the  $i - 1$  prior trials. Thus,

$$P(X > i) = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} \cdot \frac{4}{5} \cdot \dots \cdot \frac{i-1}{i} \cdot \frac{i}{i+1}$$

Note that most stuff cancels here, leaving us with the answer!

$$P(X > i) = \frac{1}{\cancel{2}} \cdot \frac{\cancel{2}}{\cancel{3}} \cdot \frac{\cancel{3}}{\cancel{4}} \cdot \frac{\cancel{4}}{\cancel{5}} \cdot \dots \cdot \frac{\cancel{i-1}}{\cancel{i}} \cdot \frac{i}{i+1} = \boxed{\frac{1}{i+1}}$$

There is a more intuitive way to solve this problem. Recall that

$$P(X > i) = 1 - P(X \leq 1) = 1 - P(X = 1) - P(X < i)$$

We can solve this problem for several values of  $i$  and notice a pattern.

$X = 1$  is the case that we draw a blue ball on the first draw. Since there are 2 balls, 1 red and 1 blue, the probability of drawing a blue ball on the first draw is simply  $\frac{1}{2}$ .

$$\begin{aligned} P(X > i) &= 1 - P(X = 1) - \underbrace{P(X < i)}_{=0} \\ &= 1 - \frac{1}{2} \\ &= \frac{1}{2} \end{aligned}$$

$X = 2$  is the case that the first blue ball drawn is chosen on the second round. This means the first draw was red (which occurs with probability  $\frac{1}{2}$ ) and the second is blue. However, on the first draw, after we put the red ball back into the urn, we add in *another* red ball. So the number of balls, and the number of red balls increases for this round. The probability of drawing a blue ball is now  $\frac{1}{3}$ , so  $P(X = 2) = \frac{1}{2} \cdot \frac{1}{3}$ .

$$\begin{aligned}
 P(X > 2) &= 1 - P(X = 2) - \underbrace{P(X < 2)}_{=P(X=1)} \\
 &= 1 - \frac{1}{2} \cdot \frac{1}{3} - \frac{1}{2} \\
 &= 1 - \frac{2}{3} \\
 &= \frac{1}{3}
 \end{aligned}$$

$X = 3$  is the case that the first blue ball drawn is chosen on the third round. This means the first two draws were red (which occurs with probability  $\frac{1}{2} \cdot \frac{2}{3}$ ) and the third draw is the blue ball, which occurs with probability  $\frac{1}{4}$ . Thus,  $P(X = 3) = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{4}$

$$\begin{aligned}
 P(X > 3) &= 1 - P(X = 3) - \underbrace{P(X < 3)}_{=P(X=1)+P(X=2)} \\
 &= 1 - \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{4} - \frac{1}{2} - \frac{1}{2} \cdot \frac{1}{3} \\
 &= 1 - \frac{3}{4} \\
 &= \frac{1}{4}
 \end{aligned}$$

and so on... If you are bored, you can prove this using mathematical induction, but we will just accept it here. We note that,

$$P(X > i) = 1 - \frac{i}{i+1} = \frac{1}{i+1}$$

(b) *Show that with probability 1, a blue ball is eventually chosen. (Show  $P(X < \infty) = 1$ .)*

*Proof.* We can use the result from part (a) to instead prove that  $P(X > \infty) = 0$ . We take the limit as  $i \rightarrow \infty$ .

$$\begin{aligned}
 P(X > \infty) &= \lim_{i \rightarrow \infty} \frac{1}{i+1} \\
 &= \lim_{i \rightarrow \infty} \frac{\frac{1}{i}}{1 + \frac{1}{i}} \\
 &= 0, \text{ so} \\
 P(X < \infty) &= 1
 \end{aligned}$$

□

(c) Find  $E(X)$ .

Recall that the definition of  $E(X)$  is

$$E(X) = \sum_{i=1}^n x_i P(X = x_i)$$

In this context,  $E(X) = \sum_{i=1}^{\infty} i P(X = i)$ , but we need to find a representation for  $P(X = i)$ . For  $X = i$ , this means that the first  $i - 1$  draws were all red, and the  $i$ th draw was blue. Then,

$$P(X = i) = \underbrace{\frac{1}{2} \cdot \frac{2}{3} \cdot \dots \cdot \frac{i-1}{i}}_{i-1 \text{ red balls drawn}} \cdot \underbrace{\frac{1}{i+1}}_{\text{blue ball drawn}} = \frac{1}{i(i+1)}$$

Then,

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i) = \sum_{i=1}^{\infty} i \frac{1}{i(i+1)} = \sum_{i=1}^{\infty} \frac{1}{i+1}$$

which diverges by the basic comparison test since  $\sum_{i=1}^{\infty} \frac{1}{i}$  diverges.

$$\boxed{E(X) = \infty}$$

**This is a variation of the Polya urn model.**