

SAMPLING AND REPEATED TRIALS

Consider an experiment and an event A within the sample space. We say the experiment is a success if an outcome from A occurs and failure otherwise. Let us consider the following examples:

Experiment	Sample Space	Event Description	Event A	$P(A)$
Toss a fair coin	$\{H, T\}$	Head appears	$\{H\}$	$\frac{1}{2}$
Roll a die	$\{1, 2, 3, 4, 5, 6\}$	Six appears	$\{6\}$	$\frac{1}{6}$
Roll a die	$\{1, 2, 3, 4, 5, 6\}$	A multiple of 3 appears	$\{3, 6\}$	$\frac{1}{3}$

In typical applications we would repeat an experiment several times independently and would be interested in the total number of successes achieved, a process that may be viewed as sampling from a large population. For instance, a manager in a factory making nuts and bolts, may devise an experiment to choose uniformly from a collection of manufactured bolts and call the experiment a success if the bolt is not defective. Then she would want to repeat such a selection every time and quantify the number of successes.

2.1 BERNOULLI TRIALS

We will now proceed to construct a mathematical framework for independent trials of an experiment where each trial is either a success or a failure. Let p be the probability of success at each trial. The sequence so obtained is called a sequence of Bernoulli trials with parameter p . The trials are named after James Bernoulli (1654-1705).

We will occasionally want to consider a single Bernoulli trial, so we will use the notation $\text{Bernoulli}(p)$ to indicate such a distribution. Since we are only interested in the result of the trial, we may view this as a probability on the sample space $S = \{\text{success}, \text{failure}\}$ where $P(\{\text{success}\}) = p$, but more often we will be interested in multiple independent trials. We discuss this in the next example.

EXAMPLE 2.1.1. Suppose we roll a die twice and ask how likely it is that we observe exactly one 6 between the two rolls. In the previous chapter (See Example 1.4.3) we would have viewed the sample space S as thirty-six equally likely outcomes, each of which was an ordered pair of results of the rolls. But since we are only concerned with whether the die roll is a 6 (success) or not a 6

(failure) we could also view it as two Bernoulli($\frac{1}{6}$) trials. Using notation from Example 1.4.3, note that $P(\text{success on the first roll}) = P(E) = \frac{1}{6}$ and $P(\text{success on the second roll}) = P(F) = \frac{1}{6}$. So

$$\begin{aligned} P(\{\text{success, success}\}) &= P(E \cap F) \\ &\quad (\text{using independence}) \\ &= P(E)P(F) \\ &= P(\text{success on the first roll}) \cdot P(\text{success on the second roll}) \\ &= \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}. \end{aligned}$$

We could alternately view S as having only four elements - (success,success), (success,failure), (failure,success), and (failure,failure). The four outcomes are not equally likely, but the fact that the trials are independent allows us to easily compute the probability of each. Through similar computations,

$$\begin{aligned} P(\{\text{success, failure}\}) &= 5/36, \\ P(\{\text{failure, success}\}) &= 5/36, \\ \text{and } P(\{\text{failure, failure}\}) &= 25/36. \end{aligned}$$

To complete the problem, the event of rolling exactly one 6 among the two dice requires exactly one success and exactly one failure. From the list above, this can happen in either of two orders, so the probability of observing exactly one 6 is $\frac{5}{36} + \frac{5}{36} = \frac{10}{36}$. ■

For any two real numbers a, b and any integer $n \geq 1$, it is well known that

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}. \quad (2.1.1)$$

This is the binomial expansion due to Blaise Pascal(1623-1662). It turns out when a and b are positive numbers with $a + b = 1$, the terms in the right hand side above have a probabilistic interpretation. We illustrate it in the example below.

EXAMPLE 2.1.2. After performing n independent Bernoulli(p) trials we are typically interested in the following questions:

- What is the probability of observing exactly k successes?
- What is the most likely number of successes?
- How many attempts must be made before the first success is observed?
- On average how many successes will there be?

Ans (a) - **Binomial(n,p)**: If $n = 1$, then the answer is clear, namely $P(\{\text{one success}\}) = p$ and $P(\{\text{zero successes}\}) = 1 - p$. For, $n > 1$ let $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ be an n -tuple of outcomes. So we may view the sample space S as the set of all ω where each ω_i is allowed to be either “success” or “failure”. Let A_i represent either the event {the i^{th} trial is a success} or {the i^{th} trial is a failure}. Then by independence

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = \prod_{i=1}^n P(A_i). \quad (2.1.2)$$

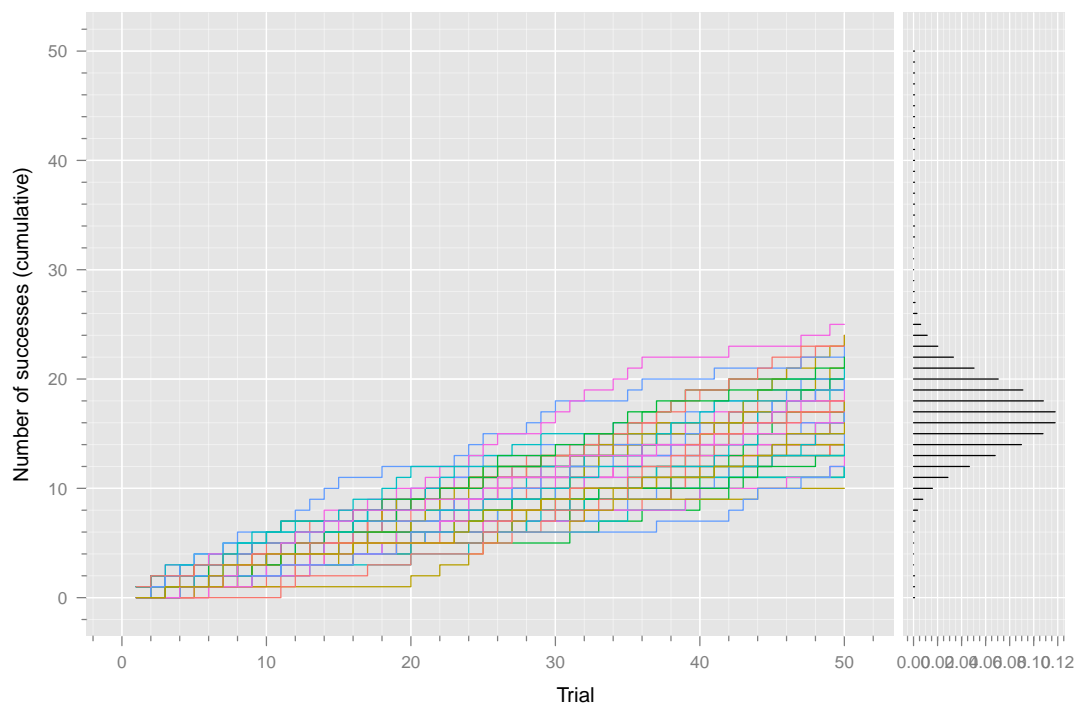


Figure 2.1: The Binomial distribution as number of successes in fifty Bernoulli ($\frac{1}{3}$) trials. The paths on the left count the cumulative successes in the fifty trials. The graph on the right show the actual probability given by the Binomial($50, \frac{1}{3}$) distribution.

Let B_k denote the event that there are k successes among the n trials. Then

$$P(B_k) = \sum_{\omega \in B_k} P(\{\omega\}).$$

But if $\omega \in B_k$, then in notation (2.1.2), exactly k of the A_i represent success trials and the other $n - k$ represent the failure trials. The order in which the successes and failures appear does not matter since the probabilities are being multiplied together. So for every $\omega \in B_k$,

$$P(\{\omega\}) = p^k(1-p)^{n-k}.$$

Consequently, we have

$$P(B_k) = |B_k|p^k(1-p)^{n-k}.$$

But B_k is the event of all outcomes for which there are k successes and the number of ways in which k successes can occur in n trials is known to be $\binom{n}{k}$. Therefore, for $0 \leq k \leq n$,

$$P(B_k) = \binom{n}{k}p^k(1-p)^{n-k}. \quad (2.1.3)$$

Note that if we are only interested in questions involving the number of successes, we could ignore the set S described above and simply use $\{0, 1, 2, \dots, n\}$ as our sample space with $P(\{k\}) = \binom{n}{k}p^k(1-p)^{n-k}$. We call this a binomial distribution with parameters n and p (or a Binomial(n, p) for short). It is also worth noting that the binomial expansion (2.1.1) shows

$$\sum_{k=0}^n \binom{n}{k}p^k(1-p)^{n-k} = (p + (1-p))^n = 1,$$

which simply provides additional confirmation that we have accounted for all possible outcomes in our list of Bernoulli trials. See Figure 2.1 for a simulated example of fifty replications of Bernoulli($\frac{1}{3}$) trials.

Ans (b) - **Mode of a Binomial:** The problem is trivial if $p = 0$ or $p = 1$, so assume $0 < p < 1$. Using the same notation for B_k as in part (a), pick a particular number of successes k for which $0 \leq k < n$. We want to determine the value of k that makes $P(B_k)$ as large as possible; such a value is called the “mode”. To find this value, it is instructive to compare the probability of $(k + 1)$ successes to the probability of k successes –

$$\begin{aligned} \frac{P(B_{k+1})}{P(B_k)} &= \frac{\binom{n}{k+1}p^{k+1}(1-p)^{n-(k+1)}}{\binom{n}{k}p^k(1-p)^{n-k}} \\ &= \frac{n!}{(k+1)!(n-(k+1))!} \cdot \frac{k!(n-k)!}{n!} \cdot \frac{p^{k+1}(1-p)^{n-(k+1)}}{p^k(1-p)^{n-k}} \\ &= \frac{p}{1-p} \cdot \frac{n-k}{k+1}. \end{aligned}$$

If this ratio were to equal 1 we could conclude that $\{(k + 1) \text{ successes}\}$ was exactly as likely as $\{k \text{ successes}\}$. Similarly if the ratio were bigger than 1 we would know that $\{(k + 1) \text{ successes}\}$ was the more likely of the two and if the ratio were less than 1 we would see that $\{k \text{ successes}\}$ was the more likely case. Setting $\frac{P(B_{k+1})}{P(B_k)} \geq 1$ and solving for k yields the following sequence of equivalent inequalities:

$$\begin{aligned}
\frac{P(B_{k+1})}{P(B_k)} &\geq 1 \\
\frac{p}{1-p} \cdot \frac{n-k}{k+1} &\geq 1 \\
pn - pk &\geq k + 1 - pk - p \\
k &\leq p(n+1) - 1.
\end{aligned}$$

In other words if k starts at 0 and begins to increase, the probability of achieving exactly k successes will increase while $k < p(n+1) - 1$ and then will decrease once $k > p(n+1) - 1$. As a consequence the most likely number of successes is the integer value of k for which $k - 1 \leq p(n+1) - 1 < k$. This gives the critical value of $k = \lfloor p(n+1) \rfloor$, the greatest integer less than or equal to $p(n+1)$.

An unusual special case occurs if $p(n+1)$ is already an integer. Then the sequence of inequalities above is equality throughout, so if we let $k = \lfloor p(n+1) \rfloor = p(n+1)$ we find a ratio $P(B_k)/P(B_{k-1})$ exactly equal to 1. In this case $\{k-1$ successes $\}$ and $\{k$ successes $\}$ share the distinction of being equally likely.

Ans (c) - **Geometric(p)**: It is possible we could see the first success as early as the first trial and, in fact, the probability of this occurring is just p , the probability that the first trial is a success. The probability of the first success coming on the k^{th} trial requires that the first $k-1$ trials be failures and the k^{th} trial be a success. Let A_i be the event $\{\text{the } i^{\text{th}} \text{ trial is a success}\}$ and let C_k be the event $\{\text{the first success occurs on the } k^{\text{th}} \text{ trial}\}$. So,

$$P(C_k) = P(A_1^c \cap A_2^c \cap \dots \cap A_{k-1}^c \cap A_k).$$

As usual $P(A_i) = p$ and $P(A_i^c) = 1 - p$, so by independence

$$P(C_k) = P(A_1^c)P(A_2^c) \dots P(A_{k-1}^c)P(A_k) = (1-p)^{k-1}p$$

for $k > 0$. If we view these as probabilities of the outcomes of a sample space $\{1, 2, 3, \dots\}$, we call this a geometric distribution with parameter p (or a $\text{Geometric}(p)$ for short).

Ans (d) - **Average**: This is a natural question to ask but it requires a precise definition of what we mean by ‘‘average’’ in the context of probability. We shall do this in Chapter 4 and return to answer (d) at that point in time. ■

Bernoulli trials may also be used to determine probabilities associated with who will win a contest that requires a certain number of individual victories. Below is an example applied to a ‘‘best two out of three’’ situation.

EXAMPLE 2.1.3. Jed and Sania play a tennis match. The match is won by the first player to win two sets. Sania is a bit better than Jed and she will win any given set with probability $\frac{2}{3}$. How likely is it that Sania will win the match? (Assume the results of each set are independent).

This can almost be viewed as three Bernoulli($\frac{2}{3}$) trials where we view a success as a set won by Sania. One problem with that perspective is that an outcome such as (win, win, loss) never occurs since two wins put an end to the match and the third set will never be played. Nevertheless, the same tools used to solve the earlier problem can be used for this one as well. Sania wins the match if she wins the first two sets (which happens with probability $\frac{4}{9}$). She also wins the match with either a (win, loss, win) or a (loss, win, win) sequence of sets, each of which has probability $\frac{4}{27}$ of occurring.

So the total probability of Sania winning the series is $\frac{4}{9} + \frac{4}{27} + \frac{4}{27} = \frac{20}{27}$.

Alternatively, it is possible to view this somewhat artificially as a genuine sequence of three Bernoulli($\frac{2}{3}$) trials where we pretend the players will play a third set even if the match is over by then. In effect the (win, win) scenario above is replaced by two different outcomes - (win, win, win) and (win, win, loss). Sania wins the match if she either wins all three sets (which has probability $\frac{8}{27}$) or if she wins exactly two of the three (which has probability $3 \cdot \frac{4}{27}$).

This perspective still leads us to the correct answer as $\frac{8}{27} + 3 \cdot \frac{4}{27} = \frac{20}{27}$. ■

2.1.1 Using R to compute probabilities

R can be used to compute probabilities of both the Binomial and Geometric distribution quite easily. We can compute them directly from the respective formulas. For example, with $n = 10$ and $p = 0.25$, all Binomial probabilities are given by

```
> k <- 0:5
> choose(5, k) * 0.25^k * 0.75^(5-k)
[1] 0.2373046875 0.3955078125 0.2636718750 0.0878906250 0.0146484375
[6] 0.0009765625
```

Similarly, the Geometric probabilities with $p = 0.25$ for $k = 0, 1, 2, \dots, 10$ are given by

```
> k <- 0:10
> 0.25 * 0.75^k
[1] 0.25000000 0.18750000 0.14062500 0.10546875 0.07910156 0.05932617
[7] 0.04449463 0.03337097 0.02502823 0.01877117 0.01407838
```

Actually, as both Binomial and Geometric are standard distributions, R has built-in functions to compute these probabilities as follows:

```
> dbinom(0:5, size = 5, prob = 0.25)
[1] 0.2373046875 0.3955078125 0.2636718750 0.0878906250 0.0146484375
[6] 0.0009765625
> dgeom(0:10, prob = 0.25)
[1] 0.25000000 0.18750000 0.14062500 0.10546875 0.07910156 0.05932617
[7] 0.04449463 0.03337097 0.02502823 0.01877117 0.01407838
```

EXERCISES

Ex. 2.1.1. Three dice are rolled. How likely is it that exactly one of the dice shows a 6?

Ex. 2.1.2. A fair die is rolled repeatedly.

- (a) What is the probability that the first 6 appears on the fifth roll?
- (b) What is the probability that no 6's appear in the first four rolls?
- (c) What is the probability that the second 6 appears on the fifth roll?

Ex. 2.1.3. Suppose that airplane engines operate independently in flight and fail with probability p ($0 \leq p \leq 1$). A plane makes a safe flight if at least half of its engines are running. Kingfisher Air lines has a four-engine plane and Paramount Airlines has a two-engine plane for a flight from Bangalore to Delhi. Which airline has the higher probability for a successful flight?

Ex. 2.1.4. Two intramural volleyball teams have eight players each. There is a 10% chance that any given player will not show up to a game, independently of any another. The game can be played if each team has at least six members show up. How likely is it the game can be played?

Ex. 2.1.5. Mark is a 70% free throw shooter. Assume each attempted free throw is independent of every other attempt. If he attempts ten free throws, answer the following questions.

- How likely is it that Mark will make exactly seven of ten attempted free throws?
- What is the most likely number of free throws Mark will make?
- How do your answers to (a) and (b) change if Mark only attempts 9 free throws instead of 10?

Ex. 2.1.6. Continuing the previous exercise, Kalyani isn't as good a free throw shooter as Mark, but she can still make a shot 40% of the time. Mark and Kalyani play a game where the first one to sink a free throw is the winner. Since Kalyani isn't as skilled a player, she goes first to make it more fair.

- How likely is it that Kalyani will win the game on her first shot?
- How likely is it that Mark will win this game on his first shot? (Remember, for Mark even to get a chance to shoot, Kalyani must miss her first shot).
- How likely is it that Kalyani will win the game on her second shot?
- How likely is it that Kalyani will win the game?

Ex. 2.1.7. Recall from the text above that the R code

```
> dbinom(0:5, size = 5, prob = 0.25)
[1] 0.2373046875 0.3955078125 0.2636718750 0.0878906250 0.0146484375
[6] 0.0009765625
```

produces a vector of six outputs corresponding to the probabilities that a Binomial(5, 0.25) distribution takes on the six values 0-5. Specifically, the output indicates that the probability of the value 0 is approximately 0.2373046875, the probability of the value 1 is approximately 0.3955078125 and so on. In Example 2.1.2 we derived a formula for the most likely outcome of such a distribution. In the case of a Binomial(5, 0.25) that formula gives the result $\lfloor (5+1)(0.25) \rfloor = 1$. We could have verified this via the R output above as well, since the second number on the list is the largest of the probabilities.

- Use the formula from example 2.1.2 to find the most likely outcome of a Binomial(7, 0.34) distribution.
- Type an appropriate command into R to produce a vector of values corresponding to the probabilities that a Binomial(7, 0.34) distribution takes on the possible values in its range. Use this list to verify your answer to part (a).
- Use the formula from Example 2.1.2 to find the most likely outcome of a Binomial(8, 0.34) distribution.
- Type an appropriate command into R to produce a vector of values corresponding to the probabilities that a Binomial(8, 0.34) distribution takes on the possible values in its range. Use this list to verify your answer to part (c).

Ex. 2.1.8. It is estimated that 0.8% of a large shipment of eggs to a certain supermarket are cracked. The eggs are packaged in cartons, each with a dozen eggs, with the cracked eggs being randomly distributed. A restaurant owner buys 10 cartons from the supermarket. Call a carton “defective” if it contains at least one cracked egg.

- (a) If she notes the number of defective cartons, what are the possible outcomes for this experiment?
- (b) If she notes the total number of cracked eggs, what are the possible outcomes for this experiment?
- (c) How likely is it that she will find exactly one cracked egg among all of her cartons?
- (d) How likely is it that she will find exactly one defective carton?
- (e) Explain why your answer to (d) is close to, but slightly larger than, than your answer to (c).
- (f) What is the most likely number of cracked eggs she will find among her cartons?
- (g) What is the most likely number of defective cartons she will find?
- (h) How do you reconcile your answers to parts (g) and (h)?

Ex. 2.1.9. Steve and Siva enter a bar with \$30 each. A round of drinks cost \$10. For each round, they roll a die. If the roll is even, Steve pays for the round and if the roll is odd, Siva pays for it. This continues until one of them runs out of money.

- (a) What is the Probability that Siva runs out of money?
- (b) What is the Probability that Siva runs out of money if Steve has cheated by bringing a die that comes up even only 40% of the time?

Ex. 2.1.10. Let $0 < p < 1$. Show that the mode of a Geometric(p) distribution is 1.

Ex. 2.1.11. Scott is playing a game where he rolls a standard die until it shows a 6. The number of rolls needed therefore has a Geometric($\frac{1}{6}$) distribution. Use the appropriate R commands to do the following:

- (a) Produce a vector of values for $j = 1, \dots, 6$ corresponding to the probabilities that it will take Scott j rolls before he observes a 6.
- (b) Scott figures that since each roll has a $\frac{1}{6}$ probability of producing a 6, he's bound to get that result at some point after six rolls. Use the results from part (a) to determine the probability that Scott's expectations are met and a 6 will show up in one his first six rolls.

Ex. 2.1.12. Suppose a fair coin is tossed n times. Compute the following:

- (a) $P(\{4 \text{ heads occur}\}|\{3 \text{ or } 4 \text{ heads occur}\})$;
- (b) $P(\{k - 1 \text{ heads occur}\}|\{k - 1 \text{ or } k \text{ heads occur}\})$; and
- (c) $P(\{k \text{ heads occur}\}|\{k - 1 \text{ or } k \text{ heads occur}\})$.

Ex. 2.1.13. At a basketball tournament, each round is on a "best of seven games" basis. That is, Team I and Team 2 play until one of the teams has won four games. Suppose each game is won by Team I with probability p , independently of all previous games. Are the events $A = \{\text{Team I wins the round}\}$ and $B = \{\text{the round lasts exactly four games}\}$ independent?

Ex. 2.1.14. Two coins are sitting on a table. One is fair and the other is weighted so that it always comes up heads.

- (a) If one coin is selected at random (each equally likely) and flipped, what is the probability the result is heads?

- (b) One coin is selected at random (each equally likely) and flipped five times. Each flip shows heads. Given this information about the coin flip results, what is the conditional probability that the selected coin was the fair one?

Ex. 2.1.15. For $0 < p < 1$ we defined the geometric distribution as a probability on the set $\{1, 2, 3, \dots\}$ for which $P(\{k\}) = p(1-p)^{k-1}$. Show that these outcomes account for all possibilities by demonstrating that $\sum_{k=1}^{\infty} P(\{k\}) = 1$.

Ex. 2.1.16. The geometric distribution described the waiting time to observe a single success. A “negative binomial” distribution with parameters n and p ($\text{NegBinomial}(n, p)$) is defined the number of Bernoulli(p) trials needed before observing n successes. The following problem builds toward calculating some associated probabilities.

- (a) If a fair die is rolled repeatedly and a number is recorded equal to the number of rolls until the second 6 is observed, what is the sample space of possible outcomes for this experiment?
- (b) For k in the sample space you identified in part (a), what is $P(\{k\})$?
- (c) If a fair die is rolled repeatedly and a number is recorded equal to the number of rolls until the n^{th} 6 is observed, what is the sample space of possible outcomes for this experiment?
- (d) For k in the sample space you identified in part (c), what is $P(\{k\})$?
- (e) If a sequence of Bernoulli(p) trials (with $0 < p < 1$) is performed and a number is recorded equal to the number of trials until the n^{th} success is observed, what is the sample space of possible outcomes for this experiment?
- (f) For k in the sample space you identified in part (e), what is $P(\{k\})$?
- (g) Show that you have accounted for all possibilities in part (f) by showing

$$\sum_{k \in S} P(\{k\}) = 1.$$

2.2 POISSON APPROXIMATION

Calculating binomial probabilities can be challenging when n is large. Let us consider the following example:

EXAMPLE 2.2.1. A small college has 1460 students. Assume that birthrates are constant throughout the year and that each year has 365 days. What is the probability that five or more students were born on Independence day?

The probability that any given student was born on Independence day is $\frac{1}{365}$. So the exact probability is

$$1 - \sum_{k=0}^4 \binom{1460}{k} \left(\frac{1}{365}\right)^k \left(\frac{364}{365}\right)^{1460-k}.$$

Repeatedly dealing with large powers of fractions or large combinatorial computations is not so easy, so it would be convenient to find a faster way to estimate such a probability. ■

The example above can be thought of as a series of Bernoulli trials where a success means finding a student whose birthday is Independence day. In this case p is small ($\frac{1}{365}$) and n is large (1460). To approximate we will consider a limiting procedure where $p \rightarrow 0$ and $n \rightarrow \infty$, but with limits carried out in such a way that np is held constant. The computation below is called a Poisson approximation.

THEOREM 2.2.2. Let $\lambda > 0$, $k \geq 1$, $n \geq \lambda$ and $p = \frac{\lambda}{n}$. Defining A_k as

$$A_k = \{k \text{ successes in } n \text{ Bernoulli}(p) \text{ Trials}\},$$

it then follows that

$$\lim_{n \rightarrow \infty} P(A_k) = \frac{e^{-\lambda} \lambda^k}{k!}. \quad (2.2.1)$$

Proof -

$$\begin{aligned} P(A_k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1) \dots (n-k+1) \lambda^k}{k! n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k n(n-1) \dots (n-k+1)}{k! n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!} 1\left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!} \prod_{r=1}^{k-1} \left(1 - \frac{r}{n}\right) \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}. \end{aligned}$$

Standard limit results imply that

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 - \frac{r}{n}\right) &= 1 \quad \text{for all } r \geq 1; \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} &= 1 \quad \text{for all } \lambda \geq 0, k \geq 1; \text{ and} \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n &= e^{-\lambda} \quad \text{for all } \lambda \geq 0. \end{aligned}$$

As $P(A_k)$ is a finite product of such expressions, the result is now immediate using the properties of limits. ■

Returning to Example 2.2.1 and using the above approximation, we would take $\lambda = pn = \frac{1460}{365} = 4$. So if E is the event {five or more Independence day birthdays},

$$\begin{aligned} P(E) &= 1 - \sum_{k=0}^4 \binom{1460}{k} \left(\frac{1}{365}\right)^k \left(\frac{364}{365}\right)^{1460-k} \\ &\approx 1 - \left[e^{-4} + 4e^{-4} + \frac{4^2}{2}e^{-4} + \frac{4^3}{6}e^{-4} + \frac{4^4}{24}e^{-4} \right]. \end{aligned}$$

Calculation demonstrates this is a good approximation. To seven digits of accuracy, the correct value is 0.37116294 while the Poisson approximation gives an answer of 0.37116306. These can be obtained using R as follows:

```
> 1 - sum(dbinom(0:4, size = 1460, prob = 1/365))
[1] 0.3711629
> lambda <- 1460 / 365
> 1 - sum(exp(-lambda) * lambda^(0:4) / factorial(0:4))
[1] 0.3711631
```

It also turns out that the right hand side of (2.2.1) defines a probability on the sample space of non-negative integers. The distribution is named after Siméon Poisson (1781-1840).

Poisson (λ): Let $\lambda \geq 0$ and $S = \{0, 1, 2, 3, \dots\}$ with probability P given by

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

for $k \in S$. This distribution is called Poisson with parameter λ (or $\text{Poisson}(\lambda)$ for short).

As with Binomial and Geometric, R has a built-in function to evaluate Poisson probabilities as well. An alternative to the calculation above is:

```
> 1 - sum(dpois(0:4, lambda = 1460 / 365))
[1] 0.3711631
```

It is important to note that for this approximation to work well, p must be small and n must be large. For example, we may modify our question as follows:

EXAMPLE 2.2.3. A class has 48 students. Assume that birthrates are constant throughout the year and that each year has 365 days. What is the probability that five or more students were born in September? ■

The correct answer to this question is

```
> 1 - sum(dbinom(0:4, size = 48, prob = 1/12))
0.3710398
```

However, the Poisson approximation remains unchanged at 0.3711631, because $np = 48/12 = 1460/365 = 4$, and only matches the correct answer up to 3 digits rather than 6. Figure 2.2 shows a point-by-point approximation of both Binomial distributions by Poisson.

At this point we have defined many named distributions. Frequently a problem will require the use of more than one of these as evidenced in the next example.

EXAMPLE 2.2.4. A computer transmits three digital messages of 12 million bits of information each. Each bit has a probability of one one-billionth that it will be incorrectly received, independent of all other bits. What is the probability that at least two of the of the three messages will be received error free?

Since $n = 12,000,000$ is large and since $p = \frac{1}{1,000,000,000}$ is small it is appropriate to use a Poisson approximation where $\lambda = np = 0.012$. A message is error free if there isn't a single misread bit, so the probability that a given message will be received without an error is $e^{-0.012}$.

Now we can think of each message being like a Bernoulli trial with probability $e^{-0.012}$, so the number of messages correctly received is then like a $\text{Binomial}(3, e^{-0.012})$. Therefore the probability of receiving at least two error-free messages is

$$\binom{3}{3}(e^{-0.012})^3(1 - e^{-0.012})^0 + \binom{3}{2}(e^{-0.012})^2(1 - e^{-0.012})^1 \approx 0.9996.$$

There is about a 99.96% chance that at least two of the messages will be correctly received. ■

EXERCISES

Ex. 2.2.1. Do the problems below to familiarize yourself with the “sum” command in R.

- If a fair coin is tossed 100 times, what is the probability exactly 55 of the tosses show heads?
- Example 2.2.3 showed how to use R to add the probabilities of a range of outcomes for common distributions. Use this code as a guide to calculate the probability at least 55 tosses show heads.

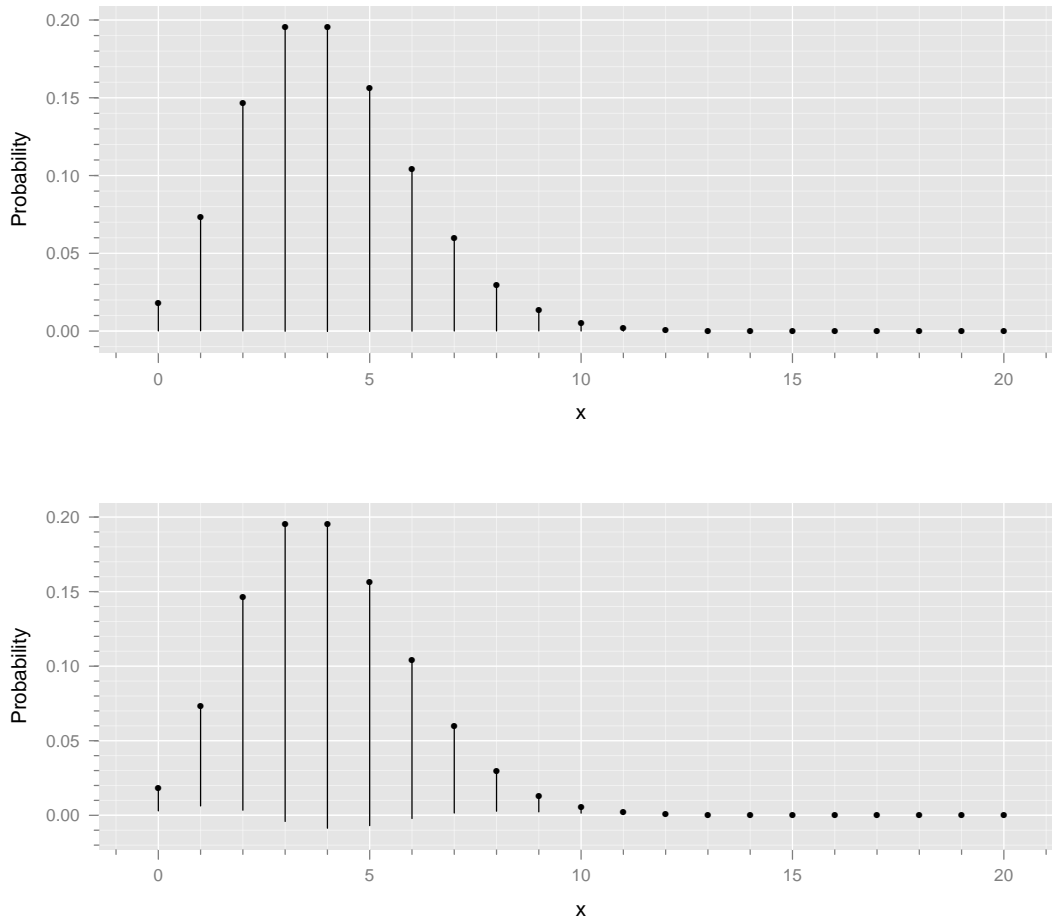


Figure 2.2: The Poisson approximation to the Binomial distribution. In both plots above, the points indicate Binomial probabilities for $k = 0, 1, 2, \dots, 20$; the top plot for $\text{Binomial}(1460, \frac{1}{365})$, and the bottom for $\text{Binomial}(48, \frac{1}{12})$. The lengths of the vertical lines, “hanging” from the points, represent the corresponding probabilities for $\text{Poisson}(4)$. For a good approximation, the bottom of the hanging lines should end up at the x -axis. As we can see, this happens in the top plot but not for the bottom plot, indicating that $\text{Poisson}(4)$ is a good approximation for the first Binomial distribution, but not as good for the second.

Ex. 2.2.2. Consider an experiment described by a $\text{Poisson}(\frac{1}{2})$ distribution and answer the following questions.

- (a) What is the probability the experiment will produce a result of 0?
- (b) What is the probability the experiment will produce a result larger than 1?

Ex. 2.2.3. Suppose we perform 500 independent trials with probability of success being 0.02.

- (a) Use R to compute the probability that there are six or fewer successes. Obtain a decimal approximation accurate to five decimal places.
- (b) Use the Poisson approximation to estimate the probability that there are six or fewer successes and compare it to your answer to (a).

Now suppose we perform 5000 independent trials with probability of success being 0.002.

- (c) Use R to compute the probability that there are six or fewer successes. Obtain a decimal approximation accurate to five decimal places.
- (d) Use the Poisson approximation to estimate the probability that there are six or fewer successes and compare it to your answer to (c).
- (e) Which approximation (b) or (d) is more accurate? Why?

Ex. 2.2.4. For a certain daily lottery, the probability is $\frac{1}{10000}$ that you will win. Suppose you play this lottery every day for three years. Use the Poisson approximation to estimate the chance that you will win more than once.

Ex. 2.2.5. A book has 200 pages. The number of mistakes on each page has a $\text{Poisson}(1)$ distribution, and is independent of the number of mistakes on all other pages.

- (a) What is the chance that there are at least 2 mistakes on the first page?
- (b) What is the chance that at least eight of the first ten pages are free of mistakes?

Ex. 2.2.6. Let $\lambda > 0$. For the problems below, assume the probability space is a $\text{Poisson}(\lambda)$ distribution.

- (a) Let k be a non-negative integer. Calculate the ratio $\frac{P(\{k+1\})}{P(\{k\})}$.
- (b) Use (a) to calculate the mode of a $\text{Poisson}(\lambda)$.

Ex. 2.2.7. A number is to be produced as follows. A fair coin is tossed. If the coin comes up heads the number will be the outcome of an experiment corresponding to a $\text{Poisson}(1)$ distribution. If the coin comes up tails the number will be the outcome of an experiment corresponding to a $\text{Poisson}(2)$ distribution. Given that the number produced was a 2, determine the conditional probability that the coin came up heads.

Ex. 2.2.8. Suppose that the number of earthquakes that occur in a year in California has a Poisson distribution with parameter λ . Suppose that the probability that any given earthquake has magnitude at least 6 on the Richter scale is p .

- (a) Given that there are exactly n earthquakes in a year, find an expression (in terms of n and p) for the conditional probability that exactly one of them is magnitude at least 6.

- (b) Find an expression (in terms of λ and p) for the probability that there will be exactly one earthquake of magnitude at least 6 in a year.
- (c) Find an expression (in terms of n , λ , and p) for the probability that there will be exactly k earthquakes of magnitude at least 6 in a year.

Ex. 2.2.9. We defined a Poisson distribution as a probability on $S = \{0, 1, 2, \dots\}$ for which

$$P(\{k\}) = \frac{e^{-\lambda} \lambda^k}{k!},$$

for $k \geq 1$. Prove that this completely accounts for all possibilities by proving that

$$\sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} = 1.$$

(Hint: Consider the power series expansion of the exponential function).

Ex. 2.2.10. Consider n vertices labeled $\{1, 2, \dots, n\}$. Corresponding to each distinct pair $\{i, j\}$ we perform an independent Bernoulli (p) experiment and insert an edge between i and j with probability p . The graph constructed this way is denoted as $G(n, p)$.

- (a) Let $1 \leq i \leq n$. We say j is a neighbour of i if there is an edge between i and j . For some $1 \leq k \leq n$ determine the probability that i has k neighbours?
- (b) Let $\lambda > 0$ and n large enough so that $0 < p = \frac{\lambda}{n} < 1$ and let $A_k = \{\text{vertex } 1 \text{ has } k \text{ neighbours}\}$ what is the

$$\lim_{n \rightarrow \infty} P(A_k)?$$

2.3 SAMPLING WITH AND WITHOUT REPLACEMENT

Imagine a small town with 5000 residents, exactly 1000 of whom are under the age of eighteen. Suppose we randomly select four of these residents and ask how many of the four are under the age of eighteen. There is some ambiguity in how to interpret this idea of selecting four residents. One possibility is “sampling with replacement” where each selection could be any of the 5000 residents and the selections are all genuinely independent. With this interpretation, the sample is simply a series of four independent Bernoulli($\frac{1}{5}$) trials, in which case the answer may be found using techniques from the previous sections. Note, however, that the assumption of independence allows for the possibility that the same individual will be chosen two or more times in separate trials. This is a situation that might seem peculiar when we think about choosing four people from a population of 5000, since we may not have four different individuals at the end of the process. To eliminate this possibility consider “sampling without replacement” where it is assumed that if an individual is chosen for inclusion in the sample, that person is no longer available to be picked in a later selection. Equivalently we can consider all possible groups of four which might be selected and view each grouping as equally likely. This change means the problem can no longer be solved by viewing the situation as a series of independent Bernoulli trials. Nevertheless, other tools that have been previously developed will serve to answer this new problem.

EXAMPLE 2.3.1. For the town described above, what is the probability that, of four residents randomly selected (without replacement), exactly two of them will be under the age of eighteen?

Since we are selecting four residents from the town of 5000, there are $\binom{5000}{4}$ ways this may be done. If each of these is equally likely, the desired probability may be calculated by determining

how many of these selections result in exactly two people under the age of eighteen. This requires selecting two of the 1000 who are in that younger age group and also selecting two of the 4000 who are older. So there are $\binom{1000}{2}\binom{4000}{2}$ ways to make such choices and therefore the probability of selecting exactly two residents under age eighteen is $\binom{1000}{2}\binom{4000}{2}/\binom{5000}{4}$.

It is instructive to compare this to the solution if it is assumed the selection is done with replacement. In that case, the answer is simply the probability that a Binomial($4, \frac{1}{5}$) produces a result of two. From the previous sections, the answer is $\binom{4}{2}(\frac{1}{5})^2(\frac{4}{5})^2$.

To compare these answers we give decimal approximations of both. To six digits of accuracy

$$\frac{\binom{1000}{2}\binom{4000}{2}}{\binom{5000}{4}} \approx 0.153592 \quad \text{and} \quad \binom{4}{2}\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^2 = 0.1536,$$

so while the two answers are not equal, they are very close. This is a reflection of an important fact in statistical analysis – when samples are small relative to the size of the populations they came from, the two methods of sampling give very similar results. ■

2.3.1 The Hypergeometric Distribution

Analyzing such problems more generally, consider a population of N people. Suppose r of these N share a common characteristic and the remaining $N - r$ do not have this characteristic. We take a sample of size m (without replacement) from the population and count the number of people among the sample that have the specified characteristic. This experiment is described by probabilities known as a hypergeometric distribution. Notice that the largest possible result is $\min\{m, r\}$ since the number cannot be larger than the size of the sample nor can it be larger than the number of people in the population with the characteristic. On the other extreme, it may be that the sample is so large it is guaranteed to select some people with the characteristic simply because the number of people without has been exhausted. More precisely, for every selection over $N - r$ in the sample we are guaranteed to select at least one person who has the characteristic. So the minimum possible result is the larger of 0 or $(m - (N - r))$.

HyperGeo(N, r, m): Let r and m be non-negative integers and let N be an integer with $N > \max\{r, m\}$. Let S be the set of integers ranging from $\max\{0, m - (N - r)\}$ to $\min\{m, r\}$ inclusive with probability P given by

$$P(\{k\}) = \frac{\binom{r}{k}\binom{N-r}{m-k}}{\binom{N}{m}}$$

for $k \in S$. Such a distribution is called hypergeometric with parameters N , r , and m (or HyperGeo(N, r, m)).

Of course, R can be used to compute hypergeometric probabilities as well. Example 2.3.1 can be phrased in terms of a HyperGeo(5000, 1000, 4) distribution, with $P(\{k\})$ being the desired answer. This probability can be computed as:

```
> dhyper(2, 4000, 1000, 4)
[1] 0.1535923
```

Note however, that instead of N , the parameter used by R is $N - r$.

2.3.2 Hypergeometric Distributions as a Series of Dependent Trials

It is also possible (and useful) to view sampling without replacement as a series of dependent Bernoulli trials for which each trial reduces the possible outcomes of subsequent trials. In this case

each trial is described in terms of conditional probabilities based on the results of the preceding observations. We illustrate this by revisiting the previous example.

Example 2.3.1 Continued: We first solved this problem by considering every group of four as equally likely to be selected. Now consider the sampling procedure as a series of four separate Bernoulli trials where a success corresponds to the selection of a person under eighteen and a failure as the selection of someone older. We still want to determine the probability that a sample of size four will produce exactly two successes. One complication with this perspective is that the successes and failures could come in many different orders, so first consider the event where the series of selections follow the pattern “success-success-failure-failure”. More precisely, for $j = 1, 2, 3, 4$ let

$$A_j = \{\text{The } j^{\text{th}} \text{ selection is a person younger than eighteen}\}.$$

Clearly $P(A_1) = \frac{1000}{5000}$. Given that the first selection is someone under eighteen, there are now only 4999 people remaining to choose among, and only 999 of them are under eighteen. Therefore $P(A_2|A_1) = \frac{999}{4999}$. Continuing with that same reasoning,

$$P(A_3^c|A_1 \cap A_2) = \frac{4000}{4998}$$

and

$$P(A_4^c|A_1 \cap A_2 \cap A_3^c) = \frac{3999}{4997}.$$

From those values, Theorem 1.3.8 may be used to calculate

$$\begin{aligned} P(\text{success} - \text{success} - \text{failure} - \text{failure}) &= P(A_1 \cap A_2 \cap A_3^c \cap A_4^c) \\ &= \frac{1000}{5000} \cdot \frac{999}{4999} \cdot \frac{4000}{4998} \cdot \frac{3999}{4997}. \end{aligned}$$

Next we must account for the fact that this figure only considers the case where the two younger people were chosen as the first two selections. There are $\binom{4}{2}$ different orderings that result in two younger and two older people, and it happens that each of these has the same probability calculated above. For example,

$$\begin{aligned} P(\text{failure} - \text{success} - \text{success} - \text{failure}) &= P(A_1^c \cap A_2 \cap A_3 \cap A_4^c) \\ &= \frac{4000}{5000} \cdot \frac{1000}{4999} \cdot \frac{999}{4998} \cdot \frac{3999}{4997}. \end{aligned}$$

The individual fractions are different, but their product is the same. This will always happen for different orderings of a specific number of successes since the denominators (5000 through 4997) reflect the steady reduction of one available choice with each additional selection. Similarly the numerators (1000 and 999 together with 4000 and 3999) reflect the number of people available from each of the two different categories and their reduction as previous choices eliminate possible candidates. Therefore the total probability is the product of the number of orderings and the probability of each ordering.

$$P(\text{two under eighteen}) = \binom{4}{2} \cdot \frac{4000}{5000} \cdot \frac{1000}{4999} \cdot \frac{999}{4998} \cdot \frac{3999}{4997}.$$

We leave it to the reader to verify that this is equal to $\binom{1000}{2} \binom{4000}{2} / \binom{5000}{4}$, the answer we found when we originally solved the problem via a different method.

The following theorem generalizes this previous example.

THEOREM 2.3.2. *Let S be a sample space with a hypergeometric distribution with parameters N , r , and m . Then $P(\{k\})$ equals*

$$\binom{m}{k} \left[\frac{r}{N} \frac{r-1}{N-1} \cdots \frac{r-(k-1)}{N-(k-1)} \right] \left[\frac{N-r}{N-k} \frac{N-r-1}{N-k-1} \cdots \frac{N-r-(m-1-k)}{N-(m-1)} \right]$$

for any $k \in S$.

Proof- Following the previous example as a model, this can be proven by viewing the hypergeometric distribution as a series of dependent trials. The first k fractions are the probabilities the first k trials each result in successes conditioned on the successes of the preceding trials. The remaining $m - k$ fractions are the conditional probabilities the remaining trials result in failures. The leading factor of $\binom{m}{k}$ accounts for the number of different patterns of k successes and $m - k$ failures, each of which is equally likely. It is also possible to prove the equality directly using combinatorial identities and we leave this as Exercise 2.3.4. ■

2.3.3 Binomial Approximation to the Hypergeometric Distribution

We saw with Example 2.3.1 that sampling with and without replacement may give very similar results. The following theorem makes a precise statement to this effect.

THEOREM 2.3.3. *Let N , m , and r be positive integers for which $m < r < N$ and let k be a positive integer between 0 and m . Define*

$$p = \frac{r}{N}, \quad p_1 = \frac{r-k}{N-k}, \quad \text{and} \quad p_2 = \frac{r-k}{N-m}.$$

Letting H denote the probability that a hypergeometric distribution with parameters N , r , and m takes on the value k , the following inequalities give bounds on this probability:

$$\binom{m}{k} p_1^k (1-p_2)^{m-k} < H \leq \binom{m}{k} p^k (1-p_1)^{m-k}.$$

Proof- The inequalities may be verified by comparing p , p_1 , and p_2 to the fractions from Theorem 2.3.2. Specifically note that the k fractions

$$\frac{r}{N}, \frac{r-1}{N-1}, \dots, \frac{r-(k-1)}{N-(k-1)}$$

are all less than or equal to p . Likewise the $m - k$ fractions

$$\frac{N-r}{N-k}, \frac{N-r-1}{N-k-1}, \dots, \frac{N-r-(m-1-k)}{N-(m-1)}$$

are all less than or equal to $\frac{N-r}{N-k}$ which itself equals $1 - p_1$. Combining these facts proves the right hand inequality. The left hand inequality may be similarly shown by noting that the fractions

$$\frac{r}{N}, \frac{r-1}{N-1}, \dots, \frac{r-(k-1)}{N-(k-1)}$$

are all greater than p_1 while the fractions

$$\frac{N-r}{N-k}, \frac{N-r-1}{N-k-1}, \dots, \frac{N-r-(m-1-k)}{N-(m-1)}$$

all exceed $\frac{N-r-(m-k)}{N-m}$ which equals $1 - p_2$. ■

When m is small relative to r and N , both fractions p_1 and p_2 are approximately equal to p . So this theorem justifies the earlier statement that sampling with and without replacement yield similar results when samples are small relative to the populations from which they were derived.

EXERCISES

Ex. 2.3.1. Suppose there are thirty balls in an urn, ten of which are black and the remaining twenty of which are red. Suppose three balls are selected from the urn (without replacement).

- What is the probability that the sequence of draws is red-red-black?
- What is the probability that the three draws result in exactly two red balls?

Ex. 2.3.2. This exercise explores how to use R to investigate the binomial approximation to the hypergeometric distribution.

- A jar contains forty marbles – thirty white and ten black. Ten marbles are drawn at random from the jar. Use R to calculate the probability that exactly five of the marbles drawn are black. Do two separate computations, one under the assumption that the draws are with replacement and the other under the assumption that the draws are without replacement.
- Repeat part (a) except now assume the jar contains 400 marbles – 300 white and 100 black.
- Repeat part (a) except now assume the jar contains 4000 marbles – 3000 white and 1000 black.
- Explain what you are observing with your results of parts (a), (b), and (c).

Ex. 2.3.3. Consider a room of one hundred people – forty men and sixty women.

- If ten people are selected from the room, find the probability that exactly six are women. Calculate this probability with and without replacement and compare the decimal approximations of your two results.
- If ten people are selected from the room, find the probability that exactly seven are women. Calculate this probability with and without replacement and compare the decimal approximations of your two results.
- If 100 people are selected from the room, find the probability that exactly sixty are women. Calculate this probability with and without replacement and compare the two answers.
- If 100 people are selected from the room, find the probability that exactly sixty-one are women. Calculate this probability with and without replacement and compare the two answers.

Ex. 2.3.4. Use the steps below to prove Theorem 2.3.2

- Prove that $\frac{r!(N-k)!}{N!(r-k)!}$ equals

$$\frac{r}{N} \cdot \frac{r-1}{N-1} \cdots \frac{r-(k-1)}{N-(k-1)}.$$

- Prove that $\frac{(N-r)!(N-m)!}{(N-k)!((N-r)-(m-k))!}$ equals

$$\frac{N-r}{N-k} \cdot \frac{N-r-1}{N-k-1} \cdots \frac{N-r-(m-1-k)}{N-(m-1)}.$$

- Use (a) and (b) to prove Theorem 2.3.2.

Ex. 2.3.5. A box contains W white balls and B black balls. A sample of n balls is drawn at random for some $n \leq \min(W, B)$. For $j = 1, 2, \dots, n$, let A_j denote the event that the ball drawn on the j^{th} draw is white. Let B_k denote the event that the sample of n balls contains exactly k white balls.

- Find $P(A_j|B_k)$ if the sample is drawn with replacement.
- Find $P(A_j|B_k)$ if the sample is drawn without replacement.

Ex. 2.3.6. For the problems below, assume a HyperGeo(N, r, m) distribution.

- Calculate the ratio $\frac{P(\{k+1\})}{P(\{k\})}$.
(Assume that $\max\{0, m - (N - r)\} \leq k \leq \min\{r, m\}$ to avoid zero in the denominator).
- Use (a) to calculate the mode of a HyperGeo(N, r, m).

Ex. 2.3.7. Biologists use a technique called “capture-recapture” to estimate the size of the population of a species that cannot be directly counted. The following exercise illustrates the role a hypergeometric distribution plays in such an estimate.

Suppose there is a species of unknown population size N . Suppose fifty members of the species are selected and given an identifying mark. Sometime later a sample of size twenty is taken from the population and it is found that four of the twenty were previously marked. The basic idea behind mark-recapture is that since the sample showed $\frac{4}{20} = 20\%$ marked members, that should also be a good estimate for the fraction of marked members of the species as a whole. However, for the whole species that fraction is $\frac{50}{N}$ which provides a population estimate of $N \approx 250$.

Looking more deeply at the problem, if the second sample is assumed to be done at random without replacement and with each member of the population equally likely to be selected, the resulting number of marked members should follow a HyperGeo($N, 50, 20$) distribution.

Under these assumptions use the formula for the mode calculated in the previous exercise to determine which values of N would cause a result of four marked members to be the most likely of the possible outcomes.

Ex. 2.3.8. The geometric distribution was first developed to determine the number of independent Bernoulli trials needed to observe the first success. When viewing the hypergeometric distribution as a series of dependent trials, the same question may be asked. Suppose we have a population of N people for which r have a certain characteristic and the remaining $N - r$ do not have that characteristic. Suppose an experiment consists of sampling (without replacement) repeatedly and recording the number of the sample that first corresponds to selecting someone with the specified characteristic. Answer the questions below.

- What is S , the list of possible outcomes of this experiment?
- For each $k \in S$, what is $P(\{k\})$?
- Define $p = \frac{r}{N}$ and $p_1 = \frac{r - (k-1)}{N - (k-1)}$. Using the result from (b) prove the following bounds on the probability distribution:

$$p(1 - p_1)^{k-1} \leq P(\{k\}) \leq p_1(1 - p)^{k-1}$$

(As a consequence, when k is much smaller than r and N , the values of p_1 and p are approximately equal and the probabilities from (b) are closely approximated by a geometric distribution).

