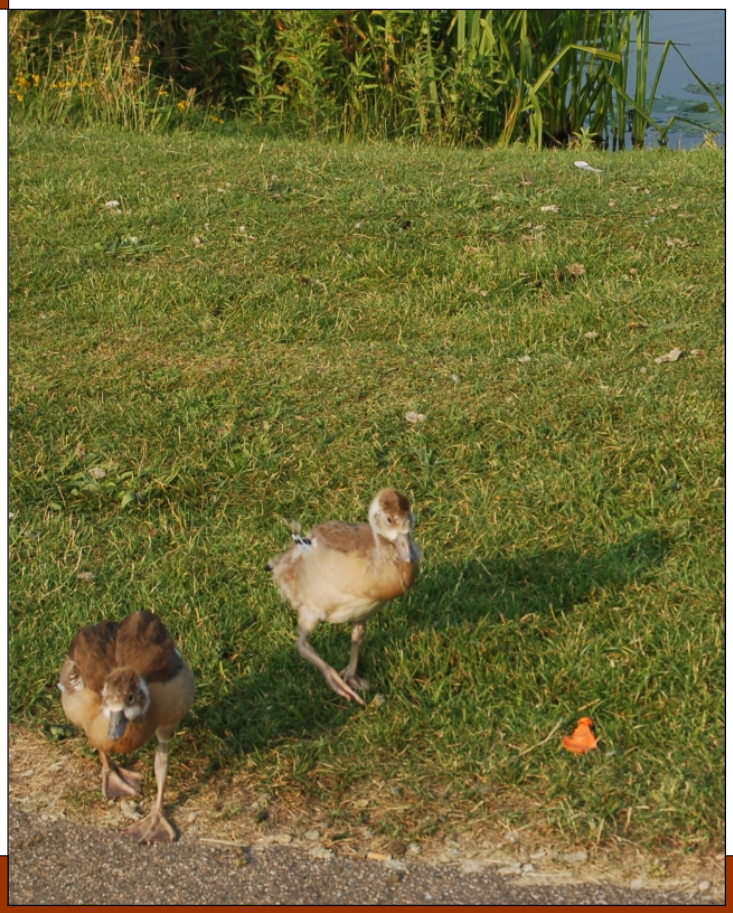

Appendix A:

Important Discrete Distributions

Forsberg, Ole J. (September 27, 2015). "Important Discrete Distributions." In *R for Starters*. Version 0.57721. Retrieved from http://rfs.kvasaheim.com/.

At their very heart, statistics are just numbers, functions of the data. However, knowing the value of a test statistic tells us little about the parameter of interest. For instance, let us suppose $\overline{x} = 5$ for a dataset. Of course our estimate of the expected value is $\mu = 5$; however, we do not know the precision of our estimate. For estimates on precision, we need to know the probability distribution of the test statistic we just measured — or at least the value of σ and have a large n.

When test statistics are first created, the statisticians attempt to create them in such a way that they have a known distribution. For instance, if we know $X \sim \mathcal{N}(\mu, \sigma^2)$, then we know $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ and $(n-1)S^2/\sigma^2 \sim \chi^2_n$.

Thus, to understand the concept of confidence intervals, the relationships between distributions, and the reason behind certain test statistic formulas, it becomes helpful to understand some of the more popular distributions. This appendix covers many of the most popular discrete distributions.

❧❧❧

I roll two fair, six-sided dice. What is the probability that the sum of the two outcomes is 11? Casinos pay 15-to-1 odds for a person rolling an 11 on a given roll. What is the casino's expected win per $10 bet?

A.1: Discrete Distributions

All discrete distributions have a sample space of countable size. Thus, working with them frequently requires knowledge of series representation. With that said, we can start with simple cases to illustrate some of the more important aspects of discrete distributions.

countable

Let us begin with a toy example: a fair, six-sided die that I roll once. As it is fair and six-sided, we can write out the probability for each of the six possible outcomes:

fair

Face; d	1	2	3	4	5	6
Probability; $\mathbb{P}[D = d]$	1/6	1/6	1/6	1/6	1/6	1/6

With this table (one way to represent the probability mass function), we know everything we need to know about the distribution. We know the expected value:

expected value

$$\begin{aligned}
\mathbb{E}[D] &:= \sum_{d} d\ \mathbb{P}[D = d] \\
&= \sum_{d=1}^{6} d\ \mathbb{P}[D = d] \\
&= 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) \\
&= 21\left(\frac{1}{6}\right) \\
\mu &= 3.5
\end{aligned}$$

We know the variance:

variance

$$\begin{aligned}
\mathbb{V}[D] &:= \sum_{d=1}^{n} \mathbb{P}[D = d]\ (d - \mathbb{E}[D])^2 \\
&= \sum_{d=1}^{6} \mathbb{P}[D = d]\ (d - 3.5)^2
\end{aligned}$$

$$= \left(\frac{1}{6}\right)(1-3.5)^2 + \left(\frac{1}{6}\right)(2-3.5)^2 + \left(\frac{1}{6}\right)(3-3.5)^2 + \left(\frac{1}{6}\right)(4-3.5)^2$$
$$+ \left(\frac{1}{6}\right)(5-3.5)^2 + \left(\frac{1}{6}\right)(6-3.5)^2$$
$$\sigma^2 = 1.3467$$

standard deviation

We know the standard deviation:

$$\begin{aligned} SD[D] &:= \sqrt{\mathbb{V}[D]} \\ &= \sqrt{1.3467} \\ \sigma &= 1.1605 \end{aligned}$$

union

We know the probability of a 1 *or* a 2 coming up:

$$\begin{aligned} \mathbb{P}\left[D \in \{1,2\}\right] &= \mathbb{P}[D=1] + \mathbb{P}[D=2] \\ &= \frac{1}{6} + \frac{1}{6} \\ &= \frac{1}{3} \end{aligned}$$

intersection

We know the probability of an outcome that is both odd *and* greater than 3:

$$\begin{aligned} \mathbb{P}\left[D \in \{1,3,5\} \cap \{4,5,6\}\right] &= \mathbb{P}\left[D \in \{5\}\right] \\ &= \mathbb{P}[D=5] \\ &= \frac{1}{6} \end{aligned}$$

negation

We know the probability of a 1 or a 2 *not* coming up:

$$\begin{aligned} \mathbb{P}\left[D \notin \{1,2\}\right] &= 1 - \mathbb{P}\left[D \in \{1,2\}\right] \\ &= 1 - \frac{1}{3} \\ &= \frac{2}{3} \end{aligned}$$

pmf

Et cetera. The probability mass function *fully defines the distribution*.

Example A.1: Let us be given that a random variable has the following probability mass function:

x	1	2	3	4
$\mathbb{P}[X=x]$	0.25	0.15	0.05	???

With a sample space of $\mathcal{S} = \{1,2,3,4\}$. Calculate the probability of $X = 4$. Also, calculate the mean, median, mode, variance, standard deviation, skew, and kurtosis of X.

Solution: We know that the probability of observing a result in the sample space is exactly 1. Thus, we know $\mathbb{P}[X=4] = 0.55$. With that, we can calculate the mean

unity

mean

$$\begin{aligned}\mathbb{E}[X] &:= \sum_{x=1}^{4} \mathbb{P}[X=x]\ x \\ &= 1(0.25) + 2(0.15) + 3(0.05) + 4(0.55) \\ &= 2.9,\end{aligned}$$

the median, $\widetilde{x}$,

median

$$\begin{aligned}\widetilde{x} &= \left\{\widetilde{x} \text{ s.t. } \mathbb{P}[X \leq \widetilde{x}] \geq 0.500 \cap \mathbb{P}[X \geq \widetilde{x}] \geq 0.500\right\} \\ &= 4,\end{aligned}$$

the mode

mode

$$\begin{aligned}\text{Mode}(X) &= \text{argmax}_x\ \mathbb{P}[X=x] \\ &= 4,\end{aligned}$$

the variance, σ^2,

variance

$$\begin{aligned}\mathbb{V}[D] &:= \sum_{x=1}^{4} \mathbb{P}[X=x]\ (x - \mathbb{E}[X])^2 \\ &= 0.25(1-2.9)^2 + 0.15(2-2.9)^2 + 0.05(3-2.9)^2 + 0.55(4-2.9)^2 \\ &= 1.69,\end{aligned}$$

the standard deviation, σ,

standard deviation

$$\begin{aligned}SD(X) &:= \sqrt{\mathbb{V}[X]} \\ &= \sqrt{1.69} = 1.3,\end{aligned}$$

skew

the skew, γ_1,

$$\begin{aligned}\gamma_1(X) &:= \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \\ &= \frac{\mathbb{E}\left[X^3\right] - 3\mu\mathbb{E}\left[X^2\right] + 2\mu^3}{\sigma^3} \\ &= \frac{38 - 3(2.9)10.1 + 2(2.9)^3}{(1.3)^3} \\ &= -21.001,\end{aligned}$$

excess kurtosis

The excess kurtosis, γ_2,

$$\begin{aligned}\gamma_2(X) &:= \frac{\mathbb{E}\left[(X-\mu)^4\right]}{\sigma^4} - 3 \\ &= \frac{4.1617}{2.8561} - 3 \\ &= -1.542873\end{aligned}$$

Since the excess kurtosis is less than zero, we know the shape of the distribution is flatter than that of the Normal distribution.

◇

support set

The probability mass function (pmf) is not always written in tabular form. Frequently, it is written as a function over a given sample space. This next example show that one can, at times, convert a functional pmf to a tabular pmf. Note, however, that it is not always possible (or desirable) to do this. Often, the sample space is infinite (but countable). In these cases, using the calculus of series is necessary to derive means, variances, etc.

Example A.2: Let us assume that a random variable has the following probability mass function over the sample space $\mathcal{S} = \{1, 2, 5, 10\}$:

$$f(x) = \frac{5}{9x}$$

Calculate the mean, median, mode, variance, standard deviation, skew, and excess kurtosis of X.

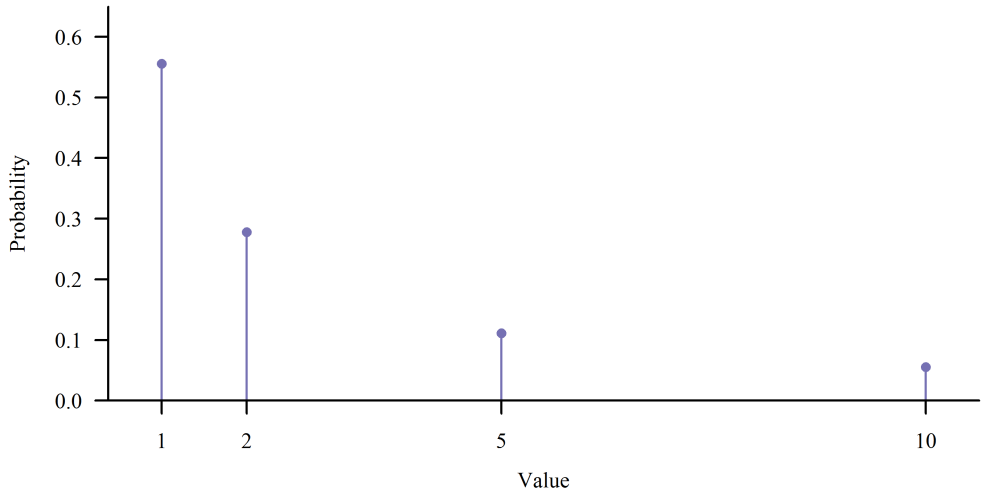

Figure A.1: *A plot of the probability mass function for the discrete distribution described in the example.*

Solution: Note that the pmf is provided in functional form, as opposed to tabular form. We can, if we so desire, write it in tabular form:

x	1	2	5	10
$\mathbb{P}[X = x]$	$\frac{5}{9}$	$\frac{5}{18}$	$\frac{1}{9}$	$\frac{1}{18}$

This form may make it easier to perform the necessary calculations. The answers are:

$$\begin{aligned} \mathbb{E}[X] &= 2.222 \\ \widetilde{X} &= 1 \\ \text{Mode}(X) &= 1 \\ \mathbb{V}[X] &= 5.062 \\ SD(X) &= 2.250 \\ \gamma_1(X) &= 2.415 \\ \gamma_2(X) &= 5.242 \end{aligned}$$

Furthermore, Figure A.1 is a plot of the probability mass function. Note its right (positive) skew. ◇

❧ ❧ ❧

These formulas describe important aspects of distributions. The mean provides a measure of the center of the distribution. The variance provides a measure of how well this mean value summarizes the entire dataset. The skew and kurtosis provide measures of how far the distribution is from Normal.

These distances from Normality will become important in the future as they give some hint to the sample size needed before the asymptotic results of the Central Limit Theorem hold (see Appendix C).

CLT

The remainder of this appendix is dedicated to some of the named distributions. These distributions are named because statisticians come across them time and again when modeling reality.

A.2: Bernoulli

One can easily argue that the Bernoulli distribution is a basis of all discrete distributions. A Bernoulli trial is a dichotomous outcome from a single experiment. For example, the number of heads resulting from flipping a coin once is a Bernoulli random variable.

dichotomous

- Symbol:

$$X \sim Bern(\pi)$$

- `R` stem:

`binom`

- Probability mass function:

$$f(x;\pi) = \pi^x \, (1-\pi)^{1-x}$$

- Cumulative distribution function:

$$F(x;\pi) = \begin{cases} 0 & x < 0 \\ 1-\pi & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$

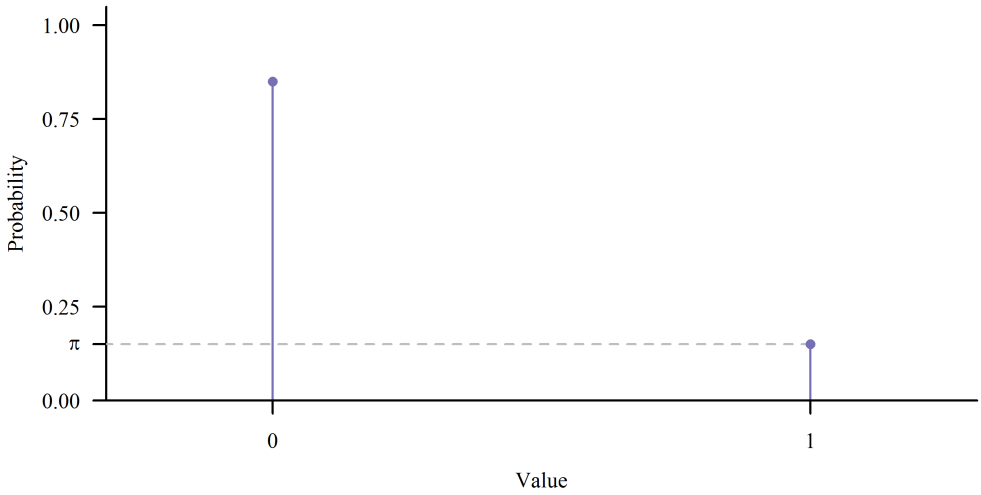

Figure A.2: *The plot of the Bernoulli distribution* $Bern(\pi = 0.15)$.

A.2.1 Parameters

$\pi \in (0,1)$ success probability

Historical note: This distribution is named after the Swiss mathematician Jakob Bernoulli, who also proved one of the most important theorems in probability: The Central Limit Theorem. In this 17th century book *Ars Conjectandi*, published posthumously, Bernoulli laid the foundations of probability theory.

A.2.2 Statistics

Mean:	π
Median:	$\begin{cases} 0 & \pi < \frac{1}{2} \\ [0,1] & \pi = \frac{1}{2} \\ 1 & \pi > \frac{1}{2} \end{cases}$
Mode:	$\begin{cases} 0 & \pi < \frac{1}{2} \\ \{0,1\} & \pi = \frac{1}{2} \\ 1 & \pi > \frac{1}{2} \end{cases}$
Variance:	$\pi(1-\pi)$
Inter-Quartile Range:	$\begin{cases} 1 & \frac{1}{4} \leq \pi < \frac{3}{4} \\ 0 & \text{otherwise} \end{cases}$
Sample space:	$\{0,1\}$
Skew:	$\frac{1-2\pi}{\sqrt{\pi(1-\pi)}}$
Excess Kurtosis:	$\frac{1-6\pi(1-\pi)}{\pi(1-\pi)}$

A.2.3 Related Distributions

Let us be given the following:

$$X \sim Bern(\pi_1) \qquad Y \sim Bern(\pi_2) \qquad Z_i \overset{\text{iid}}{\sim} Bern(\pi), \forall i \in \{1, 2, \ldots, n\}$$

Then,

- $XY \sim Bern(\pi_1 \pi_2)$.
- $\sum_{i=1}^{n} Z_i \sim Bin(n, \pi)$.

For practice, let us prove the first relationship.

Theorem A.1. *Let $X \sim Bern(\pi_x)$ and $Y \sim Bern(\pi_y)$, with X and Y independent of each other. If we define $W := XY$, then $W \sim Bern(\pi_x \pi_y)$.*

Proof. One way of proving a random variable has a specific distribution is to show that the probabilities match. Note that W has two possible outcomes, 0 and 1. Thus, W is a Bernoulli variable. All that remains is to show that the probability of $W = 1$ is $\pi_x \pi_y$.

For W to equal 1, both X and Y must be 1. The probability of this happening is $\mathbb{P}[X = 1 \cap Y = 1] = \mathbb{P}[X = 1]\mathbb{P}[Y = 1]$, because X and Y are independent. Thus, $\mathbb{P}[W = 1] = \pi_x \pi_y$. **independent**

Writing this out, we have

$$\mathbb{P}[W = w] = \begin{cases} \pi_x \pi_y & w = 1 \\ 1 - \pi_x \pi_y & w = 0 \end{cases}$$

This is the probability mass function for a Bernoulli random variable with success probability $\pi_x \pi_y$. □

A.3: Binomial

The Bernoulli distribution is the basis of all discrete distributions. The Binomial is the sum of n independent Bernoulli trials with a constant success probability.

- Symbol:

$$X \sim Bin(n, \pi)$$

- `R` stem:

```
binom
```

- Probability mass function:

$$f(x; n, \pi) = \binom{n}{x} \pi^x \, (1 - \pi)^{n-x}$$

- Cumulative distribution function:

$$F(x; n, \pi) = \sum_{i=0}^{x} f(i; n, \pi)$$

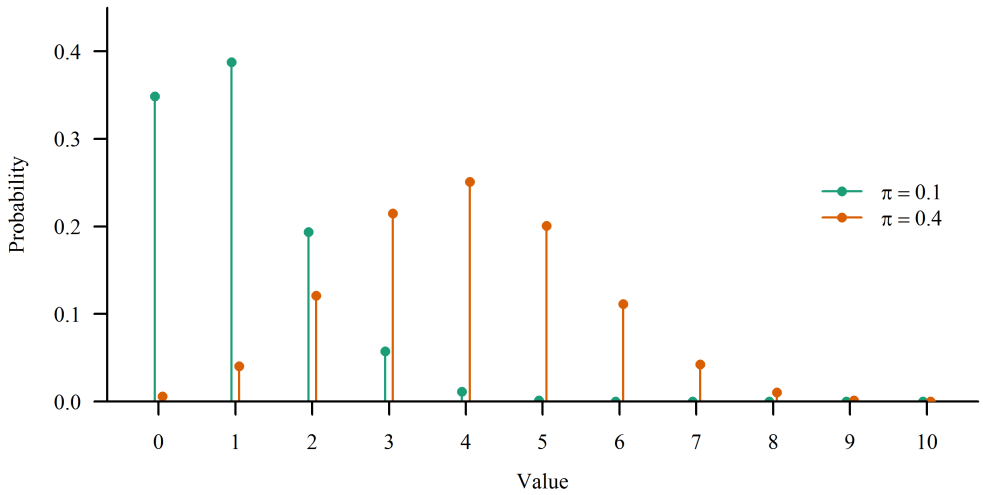

Figure A.3: *The plot of two distributions from the Binomial family, $Bin(n = 10, \pi)$.*

A.3.1 Parameters

$n \in \mathbb{N}$	number of Bernoulli trials
$\pi \in (0,1)$	success probability in each trial

Historical note: This distribution is called the Bernoulli distribution in the francophone world and the Binomial distribution elsewhere. It is called the Bernoulli distribution because Jakob Bernoulli explored it in the foundational work *Ars Conjectandi*. It is called the Binomial distribution because it makes use of the binomial coefficient, $\binom{n}{x}$.

A.3.2 Statistics

Mean:	$n\pi$
Median:	Either $\lfloor n\pi \rfloor$ or $\lceil n\pi \rceil$ (depending)
Mode:	Either $\lfloor (n+1)\pi \rfloor$ or $\lceil (n+1)\pi - 1 \rceil$ (depending)
Variance:	$n\pi(1-\pi)$
Inter-Quartile Range:	—
Sample Space:	$\{0, 1, 2, \ldots, n\}$
Skew:	$\frac{1-2\pi}{\sqrt{n\pi(1-\pi)}}$
Excess Kurtosis:	$\frac{1-6\pi(1-\pi)}{n\pi(1-\pi)}$

A.3.3 Related Distributions

Let us be given the following:

$$X \sim Bin(n_X, \pi) \qquad Y \sim Bin(n_Y, \pi)$$

Then,

- $X + Y \sim Bin(n_X + n_Y, \pi)$
- If the number of trials, n, is large enough, then $X \dot{\sim} \mathcal{N}\Big(n\pi, n\pi(1-\pi)\Big)$. This is a direct result of the Central Limit Theorem (Appendix C). The approximation is always better if $\pi \approx 0.500$. A rule of thumb for "large enough" in this case is that both $n\pi \geq 5$ and $n(1-\pi) \geq 5$.

A.3.4 Discussion

iid

requirements

The Binomial distribution is defined as the sum of n independent and identically distributed Bernoulli distributions, each with a success probability of π. This fact makes this distribution more useful than first appears. There are five requirements for an experiment to be a Binomial experiment. Any deviation from these requirements results in a different distribution:

1. The number of trials, n, is known
2. Each trial has two possible outcomes
3. The success probability, π, is constant
4. The n trials are independent
5. The random variable is the number of successes in those n trials

Example A.3: A recent 20-year average (1991–2010) suggests to us that the annual probability of at least one hurricane making landfall in Louisiana is $\pi = 0.55$. If this is true, what is the probability of a hurricane hitting Louisiana in exactly three of the next four years?

Solution: Let us define X as the event of *at least one* hurricane hitting Louisiana in a given year. This means we would like to calculate $\mathbb{P}[X = 3]$.

The next question concerns the distribution of X. Why is X a Binomial random variable? It matches the five requirements. The number of trials is known ($n = 4$); each trial has two possible outcomes (hurricane hits or not); the success probability is constant ($\pi = 0.55$); the trials are independent (the

number of hurricanes hitting this year does not depend on the number last year); and the random variable we are measuring is the number of successes (number of years in which at least one hurricane hits Louisiana).

Thus, we have that $X \sim Bin(4, 0.55)$. At this point, this problem is a straight-forward probability calculation with $n = 4$, $x = 3$, and $\pi = 0.55$:

$$\begin{aligned}\mathbb{P}[X = 3] &= \binom{4}{3} 0.55^3 \, 0.45^1 \\ &= 4(0.166375)(0.45) \\ &= 0.2995\end{aligned}$$

Thus, there is approximately a 30% chance that a hurricane will make landfall in Louisiana in exactly three of the next four years. Additionally, there is approximately a 9.2% chance that a hurricane will make landfall in Louisiana in *each* of the next four years. I leave it as an exercise for you to calculate this number. ◇

Using `R` to make this calculation is straight-forward:

```
dbinom(3, size=4, prob=0.55)
```

The `d` in `dbinom` indicates you wish to calculate the probability that the random variable **equals** a single value. The `binom` indicates you wish to calculate that probability using the Binomial pmf. This function requires three pieces of information, x, n, and π. `R` calls them `x`, `size`, and `prob`, respectively.

A.3.5 Comparing Two Binomials*

Let us suppose we have two Binomially distributed random variables

$$X \sim Bin(n_x, \pi_x) \qquad Y \sim Bin(n_y, \pi_y)$$

and we wish to determine if $\pi_x = \pi_y$. It is natural to look at the number of successes (also known as the 'realization') of X and of Y and compare them. **data**

When n is large and $\pi \approx 0.500$, this will be very straight-forward, as one can use the Normal approximation described above (*v.s.* Related Distributions) and the usual t-test (Section 6). However, when either $n\pi < 5$ or $n(1-\pi) < 5$, this approximation will not work well, and you will have to use other methods.

test stat

It would be nice if we had a test statistic, perhaps the difference between the two realizations, and an associated distribution; however, the difference of two Binomial random variables is *not* another Binomial random variable. In fact, we currently have no method for determining this distribution exactly, without making additional assumptions.

We can, however, use Monte Carlo methods to estimate (to an arbitrary degree of precision) the p-value corresponding to the test that the two distributions have the same success probability, π.

Example A.4: Continuing the previous hurricane example, let us note that Mississippi and Alabama both have approximately the same length of coastline. Thus, we may expect the probability of a hurricane making landfall in Mississippi being the same as for Alabama. However, in the 20-year period of 1991–2010, four hurricanes made landfall in Alabama, but only 2 in Mississippi. Are these numbers different enough that we can conclude the probability of a hurricane making landfall in Alabama is different than in Mississippi?

sample

population

Solution: We explicitly make the assumption that the probability of a hurricane landfall in a given year is independent of a hurricane landfall in a previous year. Thus, our null hypothesis is

null hypothesis

$$H_0 : \pi_{MS} = \pi_{AL}$$

For our sample, $p_{MS} = 0.1$ and $p_{AL} = 0.2$. Thus, the pooled proportion is $p = 0.15$. The calculation of the pooled proportion is made easier as the sample sizes are identical for the two samples ($n = 20$ years for each). The general formula for the pooled proportion is

$$p_p = \frac{(n_1\ p_1) + (n_2\ p_2)}{n_1 + n_2}$$

weighted average

This is just a weighted average.

Unfortunately, we do not have a test for comparing the means of two Binomial distributions, even if the success probabilites are the same. Furthermore, we cannot use the Normal approximation, as $p \approx 0.500$ in either case and n is small. Thus, we will have to resort to using Monte Carlo methods to approximate the p-value corresponding to our null hypothesis and our data.

Monte Carlo

To do this, we need to understand the 'Mississippi distribution' and the 'Alabama distribution.' Again, if we define X as the event of a hurricane

making landfall in a specific year, then the Mississippi distribution (*under the null hypothesis*) is

$$X_{MS} \sim Bin(20, 0.150)$$

The 20 comes from the number of years we are observing. The $\pi = 0.150$ comes from the null hypothesis that the probability of a hurricane making landfall in Mississippi equals that of Alabama, and both are equal to some common (pooled) landfall probability, 0.150.

Under the null hypothesis, the Alabama distribution is the same as the Mississippi distribution as the number of data years is the same (and the common landfall probability is the same):

$$X_{AL} \sim Bin(20, 0.150)$$

Were the sample sizes (years) different, then the Alabama distribution would reflect that.

sample size

Now that we know the distribution of the two hurricane landfalls, we need to create a test statistic. As we are not calculating the distribution for the test statistic, we can select an intuitive one. Let TS be the difference in landfalls between Mississippi and Alabama. In symbols,

test statistic

$$TS := X_{AL} - X_{MS}$$

This is not the only test statistic we could have used. We could have used the difference in their squares, the square of their difference, the tangent of their difference, etc. The key is that we need to create a test statistic that reflects what we want to test: Is the difference in their probabilities different from zero? Any test statistic that reflects this question is acceptable. The one I chose is merely simpler to interpret.

key

Now that we understand our two landfall distributions and have created a test statistic, we can perform Monte Carlo to estimate the distribution of the test statistic. The steps are similar to all other Monte Carlo experiments we have performed in the past:

steps

1. Initialize variables
2. Perform loop
    a) Draw from the Mississippi distribution
    b) Draw from the Alabama distribution
    c) Calculate the test statistic (and save it)

3. Calculate the empirical p-value of your observed test statistic
4. Calculate the $100(1-\alpha)\%$ empirical confidence interval for the test statistic

The only new step is to calculate the confidence interval from this empirical distribution (Step 2c). However, if one understands the meaning of a confidence interval, then this step should be quite easy to understand.

The R code for this algorithm is as follows

```
set.seed(30)
alpha  <- 0.05       # Typical alpha-level
trials <- 1e4        # Number of trials to run

n1 <- 20             # sample size for Mississippi
n2 <- 20             # sample size for Alabama
pp <- 0.15           # Pooled proportion

TS <- numeric()      # To hold our test statistic

for(i in 1:trials) {
  X <- rbinom(1,size=n1, prob=pp)    # MS dist
  Y <- rbinom(1,size=n2, prob=pp)    # AL dist
  TS[i] <- X-Y                       # Test stat
}

length( which(TS>2) )/trials * 2     # p-value
quantile(TS,c(alpha/2, 1-alpha/2))   # conf int
```

If you run this, you will get a p-value of 0.26 and a symmetric 95% confidence interval of $(-4, 4)$. Thus, under the null hypothesis, we will get a test statistic this extreme (or more so) 26% of the time. As such, we cannot reject the null hypothesis and must conclude, at the $\alpha = 0.05$ level, that there is no significant evidence that hurricanes make landfall in Mississippi at a rate different than in Alabama.

If we prefer to use the confidence interval to frame our conclusion, then we see that the confidence interval tells us that, if the two likelihoods are equal, then 95% of the time we will have a test statistic between -4 and $+4$. Thus, our observed test statistic of $4 - 2 = 2$ is not severe enough to cause us to conclude hurricanes make landfall in Mississippi at a rate different than in Alabama. ◇

I will leave it as an exercise to determine if Louisiana's 11 years of hurricane landfalls is statistically different from the 4 for Alabama. What

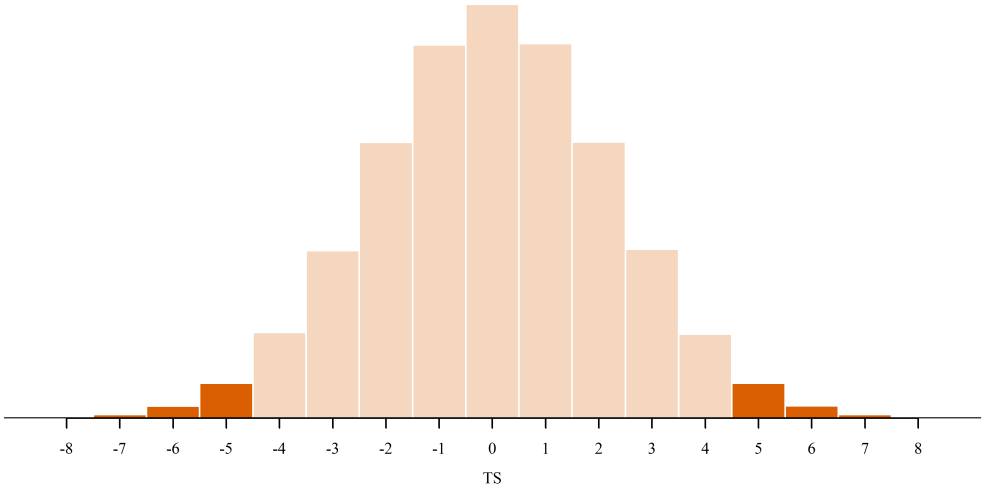

Figure A.4: *Empirical probability mass function for the difference of two Binomials corresponding to the hurricane landfall distributions for Mississippi and Alabama. The region shaded in dark orange is the rejection region corresponding to* $\alpha = 0.05$.

will you have to change in the script? What is the value of the new test statistic? The answer is that there *is* a statistically significant difference at the $\alpha = 0.05$ level ($p \approx 0.01$), with the symmetric 95% confidence interval being $(-6, 6)$.

> ***Note***: How many significant figures should we include? The number of trials in this Monte Carlo experiment is 1×10^4, thus we should only use $4/2 = 2$ decimal places. If we had used, instead, a million trials (1×10^6), we could report $6/2 = 3$ decimal places. In general, if B is the number of Monte Carlo trials performed, one can use $\frac{1}{2}(\log_{10})$ digits.
>
> This rule of thumb comes from the margin of error calculated for Binomials when using a 95% level of confidence. Recall the margin of error in this case is $E = 1.96\sqrt{\pi(1-\pi)/n}$. Using $\pi = 0.50$ (to be conservative), this becomes $1.96\sqrt{(0.5)(0.5)/n} \approx \sqrt{1/n}$.

It is actually interesting to see the empirical distribution of the test statistic (Figure A.4). Note that it is symmetric about zero. Thus, it did not matter if we decided our test statistic was $X_{AL} - X_{MS}$ or $X_{MS} - X_{AL}$; the same distribution would result. Also note that the distribution is *not* a Binomial distribution; it takes on negative values, which is outside the sample space for a Binomial.

support set

A.4: Geometric

The Geometric distribution is very similar to the Binomial distribution. Both are sums of independent and identically distributed Bernoulli random variables. However, whereas the random variable in the Binomial case is the number of successes in n trials (assumptions 2–4), the random variable in the Geometric case is the **number of failures until the first success**. Thus, the Geometric distribution fails requirements 1 and 5 for the Binomial distribution (page 526).

- Symbol:

$$X \sim \mathcal{G}\text{eom}(\pi)$$

- `R` stem:

```
geom
```

- Probability mass function:

$$f(x;\pi) = \pi(1-\pi)^x$$

- Cumulative distribution function:

$$F(x;\pi) = \mathbb{P}[X \leq x] = 1 - (1-\pi)^{x+1}$$

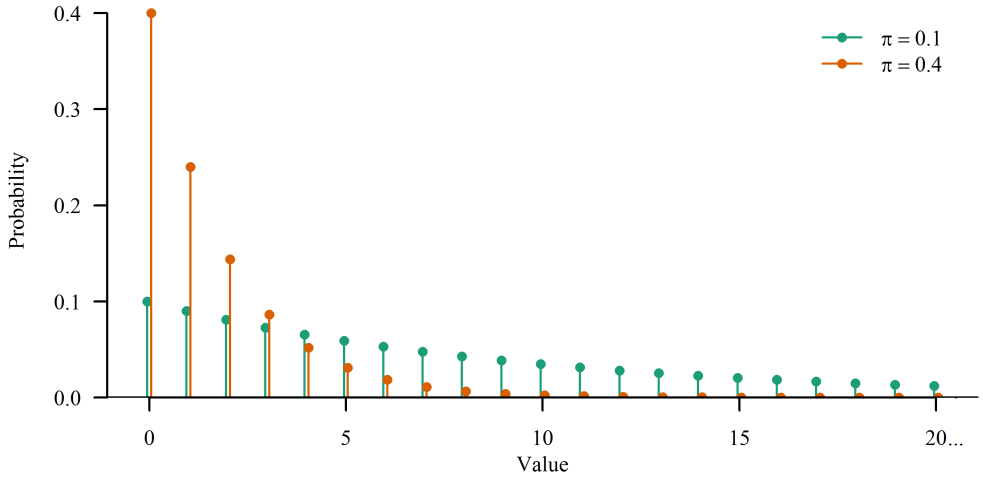

Figure A.5: *A graph of the Geometric probability mass function for various values of the success probability parameter, π.*

A.4.1 Parameter

π Success probability

Historical note: A geometric series is a series with a constant ratio between successive terms. This is the basis for the name of this distribution. Here, that constant ratio is $1 - \pi$.

A.4.2 Statistics

Mean:	$\frac{1-\pi}{\pi}$
Median:	$\lceil \frac{-1}{\log_2(1-\pi)} \rceil - 1$
Mode:	0
Variance:	$\frac{1-\pi}{\pi^2} = \frac{\mu}{\pi}$
Inter-Quartile Range:	—
Sample Space:	$\{0, 1, 2, \ldots\}$
Skew:	$\frac{2-\pi}{\sqrt{1-\pi}}$
Excess Kurtosis:	$6 + \frac{\pi^2}{1-\pi}$

Note: $\pi = \frac{\mu}{\sigma^2}$.

A.4.3 Related Distribution

- Let $X_i \overset{\text{iid}}{\sim} \mathcal{G}eom(\pi)$. If we define the random variable $Y := \sum_{i=1}^{n} X_i$, then $Y \sim NegBin(n, \pi)$.

Example A.5: Officer McGrowl patrols the mean streets of Stillwater, OK, every night looking for people who are driving under the influence of alcohol. McGrowl hypothesizes that 60% of the drivers at 2:00am are driving drunk. Assuming he is correct, what is the expected number of cars he will pull over before he catches his first drunk driver? What is the probability that he pulls over five sober drivers before his first drunk driver? If it takes five or more traffic stops until he catches his first drunk driver, can we reject his assumption that 60% of the drivers at 2:00am are driving drunk?

Solution: Let us define X as the number of sober drivers Officer McGruff pulls over before he pulls over his first drunk driver. With this, we have

$$X \sim \mathcal{G}eom(\pi = 0.60).$$

Using the appropriate formulas from above, we have

$$\begin{aligned}
\mathbb{E}[X] &= \frac{1}{0.60} - 1 = 0.6667 \\
\mathbb{P}[X = 5] &= 0.60(1 - 0.60)^5 = 0.0061
\end{aligned}$$

To calculate the p-value, we need to calculate $\mathbb{P}[X \geq 5]$. Using the cumulative distribution function, we have

$$\begin{aligned}
\mathbb{P}[X \geq 5] &= 1 - \mathbb{P}[X < 5] \\
&= 1 - \mathbb{P}[X \leq 4] \\
&= 1 - F(4; \pi = 0.60) \\
&= 1 - \left(1 - (1 - 0.60)^5\right) \\
&= (1 - 0.60)^5 \\
&= 0.01024
\end{aligned}$$

As this value is less than our usual $\alpha = 0.05$ level, we reject McGrowl's assumption about the proportion of drunk drivers at 2:00am.

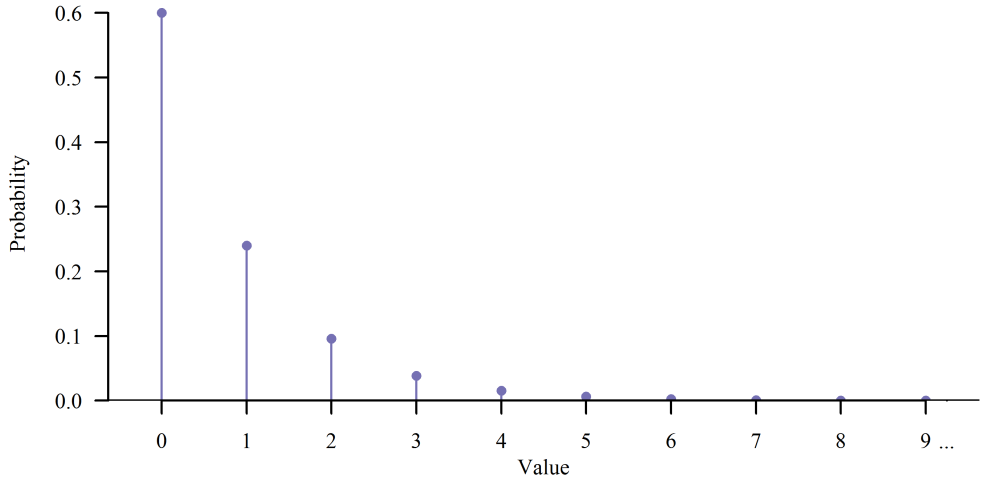

Figure A.6: *A graph of the Geometric probability mass function for the Officer McGrowl example.*

A plot of the probability mass function for this particular distribution is provided in Figure A.6, above. Note that is has strong right (positive) skew and high leptokurtosis (positive excess kurtosis). ◇

A.5: Negative Binomial (Pascal; Pólya)

The Negative Binomial distribution is an extension of the Geometric distribution. The Geometric distribution modeled the number of failures until the *first* success. The Negative Binomial distribution models the number of failures until the r^{th} success. As such, this distribution also violates requirements 1 and 5 for the Binomial distribution (page 526).

- Symbol:

$$X \sim NegBin(r, \pi)$$

- `R` stem:

```
nbinom
```

- Probability mass function:

$$f(x; r, \pi) = \binom{x + r - 1}{x} \pi^r (1 - \pi)^x$$

- Cumulative distribution function:

$$F(x; r, \pi) = \sum_{i=0}^{x} f(i; r, \pi)$$

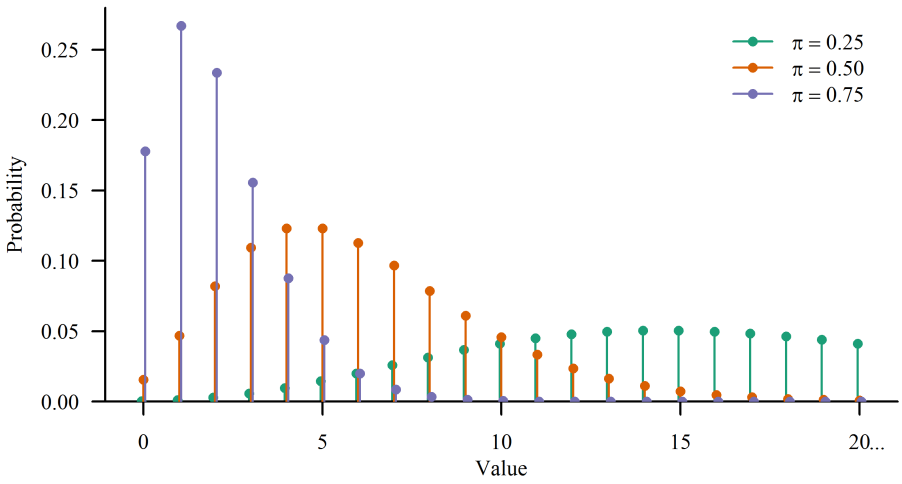

Figure A.7: *A graph of the Negative Binomial probability mass function for various values of π, with $r = 6$.*

A.5.1 Parameters

r	Number of successes until stopping
π	Success probability

Historical note: The Pascal distribution and the Pólya distribution are special cases of the Negative Binomial distribution. The former requires $r \in \mathbb{Z}$; the latter, $r \notin \mathbb{Z}$. The name of the Negative Binomial distribution comes from extending the binomial theorem to negative exponents.

A.5.2 Statistics

Mean:	$r\frac{1-\pi}{\pi}$
Median:	—
Mode:	$\lfloor\frac{(1-\pi)(r-1)}{\pi}\rfloor$
Variance:	$r\frac{1-\pi}{\pi^2}$
Inter-Quartile Range:	—
Sample Space:	$\{0,1,\ldots\}$
Skew:	$\frac{2-\pi}{\sqrt{r(1-\pi)}}$
Excess Kurtosis:	$\frac{6}{r}+\frac{\pi^2}{r(1-\pi)}$

As with the Geometric distribution, $\pi = \frac{\mu}{\sigma^2}$.

A.5.3 Related Distribution

Let us be given the following:

- If $X \sim NegBin(1, \pi)$, then $X \sim \mathcal{G}\text{eom}(\pi)$.
- Let X_i be independent and identically distributed random variables with a Geometric distribution; that is, let $X_i \overset{\text{iid}}{\sim} \mathcal{G}\text{eom}(\pi)$. This implies $\sum_i X_i \sim NegBin(r, \pi)$, where r is the number of random variables summed.

Note: This is just one way of parameterizing this distribution. Some sources have the random variable be the number of *trials* until the r^{th} success.

Example A.6: Officer McGrowl still patrols the mean streets of Stillwater, OK, looking for people who are driving under the influence of alcohol. According to official estimates, a full 10% of the drivers at 11:00pm are driving drunk.

If McGrowl begins to randomly pull over drivers, how many sober drivers can he expect to pull over before he catches his fourth drunk driver? What is the probability that he pulls over 15 sober drivers before he pulls over his fourth drunk driver? What is the probability that it takes 100 traffic stops until he catches his fourth drunk driver? If it takes 100 stops or more before he catches his fourth drunk driver, what can we conclude about the official estimates?

Solution: According to the problem, $\pi = 0.10$ and $r = 4$. Thus, we have

$$X \sim NegBin(4, 0.10).$$

With that, we have

$$\mathbb{E}[X] = r\frac{1-\pi}{\pi} = 4\frac{0.90}{0.10} = 36,$$

$$\mathbb{P}[X = 15] = \binom{x+r-1}{x}\pi^r(1-\pi)^x = \binom{15+4-1}{15}0.10^4(0.90)^{15} = 0.0168, \text{ and}$$

$$\mathbb{P}[X = 96] = \binom{x+r-1}{x}\pi^r(1-\pi)^x = \binom{96+4-1}{96}0.10^4(0.90)^{96} = 0.0006$$

These last two probabilities can be calculated using R with

```
dnbinom(15, size=4, prob=0.10), and
dnbinom(96, size=4, prob=0.10)
```

As to the question about the believability of the official estimates, we calculate $\mathbb{P}[X \geq 100]$, find that it equals 0.00572. As this value is less than our usual cut-off value of $\alpha = 0.05$, we conclude that the official estimates are in error. The proportion of drunk drivers at this time is closer to 2.5%. ◇

A.6: Hypergeometric

The Hypergeometric distribution is a generalization to the Binomial distribution. In the Binomial distribution, the probability of success did not change. This can be brought about by sampling *with* replacement or sampling from an infinite population. The Hypergeometric distribution describes the probabilities when sampling is done *without* replacement from a finite population (because π changes from Bernoulli trial to Bernoulli trial). As with the Binomial distribution, the random variable is the number of successes in n trials. Thus, the Hypergeometric distribution only fails requirement 3 for the Binomial distribution (page 526).

- Symbol:

$$X \sim \mathcal{H}(m, n, k)$$

- `R` stem:

`hyper`

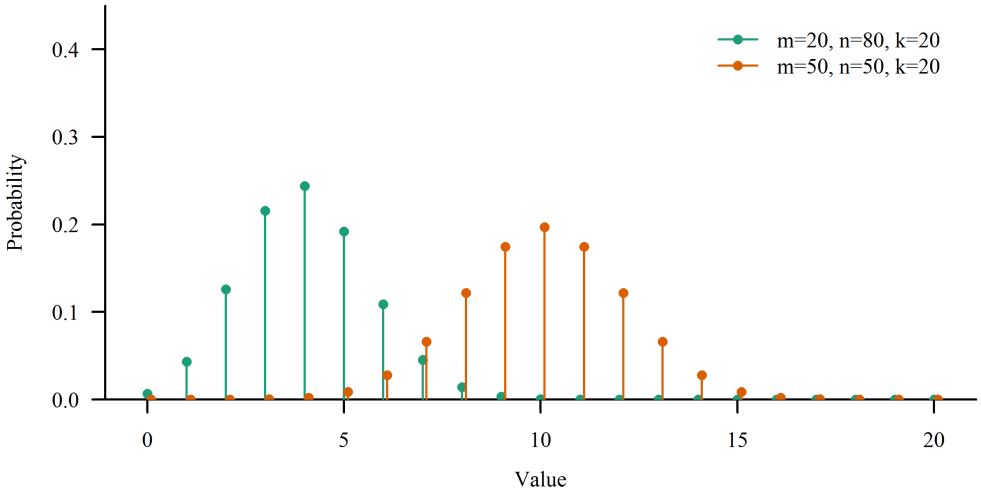

Figure A.8: *A graph of the Hypergeometric probability mass function for various values of the parameters.*

- Probability mass function:

$$f(x; m, n, k) = \frac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}}$$

- Cumulative distribution function:

$$F(x; m, n, k) = \sum_{i=0}^{x} f(i; m, n, k)$$

A.6.1 Parameters

m	Number of successes in the population
n	Number of failures in the population
k	Sample size drawn from the population

Historical note: In the 17th century, John Wallis extensively studied factorials and sequences of factorials. He named one such class of sequences hypergeometric because it grew at a rate faster than geometric. It is this use of the term *hypergeometric* that gave its name to the Hypergeometric distribution—ratios of successive terms increase.

A.6.2 Statistics

Mean:	$k\frac{m}{m+n}$
Median:	—
Mode:	$\lfloor\frac{(k+1)(m+1)}{m+n+2}\rfloor$
Variance:	$k\frac{m}{m+n}\cdot\frac{n}{m+n}\cdot\frac{m+n-k}{m+n-1}$
Inter-Quartile Range:	—
Support Set:	$\left[\max\left\{0, k-n\right\},\ \min\left\{k, m\right\}\right]$
Skew:	—
Excess Kurtosis:	—

A.6.3 Related Distribution

Let us be given the following:

- If $X \sim \mathcal{H}(m, n, 1)$, then $X \sim Bern\left(\pi = \frac{m}{m+n}\right)$

Note: As with most probability distributions, its myriad origins begat myriad parameterizations. As usual, while the letters may change, the relationships between the statistics do not.

Example A.7: According to the 2010 census, the population of Oklahoma is 3,814,820 with 168,625 Roman Catholics. A researcher calls 1000 people (without repeats) and reaches only 40 Roman Catholics. If the census is correct, what is the probability of this event?

Solution: Define the random variable X as the number of Roman Catholics telephoned, out of the 1000 Oklahomans.

First, let us solve this as a Binomial experiment. Doing so is *not* appropriate according to the information given. However, it will allow us to discover something interesting. To calculate the Binomial probability,

we need to determine n and π from the information provided. According to the problem, $n = 1000$ (the number of Oklahomans I contact) and $\pi = \frac{m}{m+n} = \frac{168{,}625}{3{,}814{,}820} \approx 0.0442$. Thus, the (approximate) distribution of X is

$$X \sim Bin(1000; 0.0442)$$

Using R,

```
dbinom(40, size=1000, prob=168625/3814820)
```

gives $\mathbb{P}[X = 40] = 0.0517819$.

Second, let us now solve this as a Hypergeometric experiment. That is, we use

$$X \sim \mathcal{H}(m = 168,625; n = 3,646,195; k = 1000)$$

Substituting these values into the probability mass function for the Hypergeometric distribution, we get $\mathbb{P}[X = 40] = 0.05178524$. In R, this is

```
dhyper(40, m=168625, n=3646195, k=1000)
```

◇

Note: These two probabilities agree to four digits. This is because a large population size ($m+n$) makes the change in success probability very slight. In other words, when the population is large, there is little reason to use the Hypergeometric over the Binomial.

Note: In this problem, we simply calculated the probability of the specific result. We did *not* calculate the p-value. Recall that the p-value is the probability of observing data this extreme *or more so*, given that the null hypothesis is true. The "or more so" part indicates that we need to calculate a cumulative probability, not a point probability, which we calculated.

p-value

A.7: Poisson

The Poisson distribution is frequently used as the basis distribution for count variables — modeling the count of an event over a period of time or a region. In this way, it is different from the Binomial in which we are counting the number of successful experiments out of the total number of experiments.

- Symbol:

$$X \sim \mathcal{P}(\lambda)$$

- `R` stem:

```
pois
```

- Probability mass function:

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

- Cumulative distribution function:

$$F(x;\lambda) = \sum_{i=0}^{x} f(i;\lambda)$$

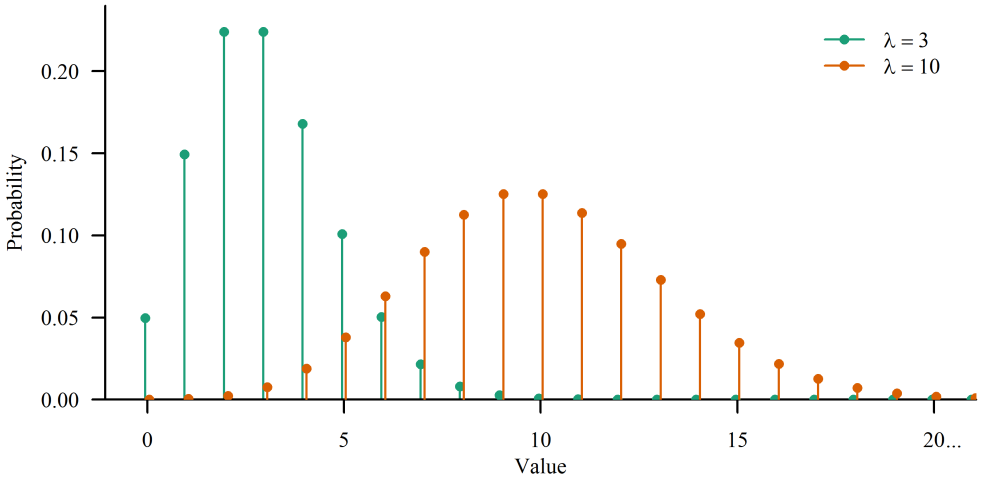

Figure A.9: *A graph of the Poisson probability mass function for various values of the rate parameter, λ.*

A.7.1 Parameter

λ rate of successes

Historical note: The distribution was first introduced by Siméon Denis Poisson (1781–1840) and published in *Recherches sur la probabilité des jugements en matière criminelle et en matière civile* (Poisson 1837). The work theorized about the number of wrongful convictions in a given country by focusing on certain random variables that count the number of discrete occurrences that take place during a time-interval of given length.

While Poisson introduced this distribution in 1837, its most famous application came in 1898 when Ladislaus Bortkiewicz modeled the number of soldiers in the Prussian army killed by accidental horse kick. In his book, *Das Gesetz der kleinen Zahlen*, Bortkiewicz examined the distribution of random variables whose success probability was small. He proved that the Poisson distribution was the limit of Binomial distributions where $n \to \infty$ and $n\pi = \lambda$ is held constant (Bortkiewicz 1898).

A.7.2 Statistics

Mean:	λ
Median:	$\lfloor \lambda + \frac{1}{3} - \frac{0.02}{\lambda} \rfloor$, approximately
Modes:	$\lceil \lambda \rceil - 1$, and $\lfloor \lambda \rfloor$
Variance:	λ
Inter-Quartile Range:	—
Sample Space:	$\{0, 1, 2, \ldots\}$
Skew:	$\frac{1}{\sqrt{\lambda}}$
Excess Kurtosis:	$\frac{1}{\lambda}$

A.7.3 Related Distributions

Let us be given the following:

- Let $X \sim Bin(n, \pi)$. Letting $n \to \infty$ with $n\pi$ remaining constant results in $X \xrightarrow{d} \mathcal{P}(\lambda = n\pi)$. This is Bortkiewicz's (1898) famous result (see Section A.7.5).
- Let $X_i \overset{\text{ind}}{\sim} \mathcal{P}(\lambda_i)$. Then $\sum X_i \sim \mathcal{P}(\sum \lambda_i)$.

inter-arrival

- If $X \sim \mathcal{P}(\lambda)$ represents the number of items arriving, then the time between arrivals, T, has an Exponential distribution, $T \sim \mathcal{E}xp(\lambda)$.
- The Negative Binomial is also used to model counts over space or time. It turns out that the Negative Binomial distribution is equivalent to the Poisson distribution where λ has a Gamma distribution (see Section B.6).

A.7.4 Discussion

Perhaps the most important thing about the Poisson distribution is that sums of independent Poisson-distributed random variables also have a Poisson distribution. That is, if $X_i \overset{\text{ind}}{\sim} \mathcal{P}(\lambda_i)$, then $\sum X_i \sim \mathcal{P}(\sum \lambda_i)$. This arises from the fact that the Poisson distribution is an infinitely divisible distribution.

The probability mass function for the Poisson distribution arises from an approximation of the Binomial distribution when the success probability π is close to 0 or 1 and n is large. One can prove this using moment generating functions or calculus (see Section A.7.5 for a proof using the latter).

Example A.8: Let us continue the previous hurricane example (Example A.4). Recall that in the past 20 years, four hurricanes made landfall in Alabama. What is the probability that Alabama will experience at least one hurricane this year? What is the probability that it will experience more than one?

Solution: Note the difference between this example and Example A.4. In Example A.4, we needed to calculate the probability of getting hit by at least one hurricane per year in three of the next four years; that is, a success is defined as Alabama is hit by a hurricane in a calendar year, and the five requirements of a Binomial experiment are satisfied. In this example, we want

the probability of multiple hits in a given year, which is not a 0/1 variable, thus requirements 2 and 5 are violated (page 526).

As our random variable is a count of successes *over a period of time*, an appropriate distribution is the Poisson. To use the Poisson distribution, we need to determine the value of the parameter λ. As λ is the rate of success over the period of time, we have $\lambda = 4 \div 20 = 0.25$ per year.

The first question asks for the probability of being hit by at least one hurricane this year. This is just

$$\begin{aligned}
\mathbb{P}[X \geq 1] &= 1 - \mathbb{P}[X < 1] \\
&= 1 - \mathbb{P}[X = 0] \\
&= 1 - \frac{e^{-\lambda}\lambda^x}{x!} \\
&= 1 - \frac{e^{-0.25}0.25^0}{0!} \\
&= 1 - 0.7788
\end{aligned}$$

Thus, there is a 22% chance of Alabama being hit by at least one hurricane this year. The probability that it will be hit by multiple hurricanes is

$$\begin{aligned}
\mathbb{P}[X \geq 2] &= \mathbb{P}[X > 1] \\
&= \mathbb{P}[X \geq 1] - \mathbb{P}[X = 1] \\
&= 0.2212 - \frac{e^{-\lambda}\lambda^x}{x!} \\
&= 0.2212 - \frac{e^{-0.25}0.25^1}{1!} \\
&= 0.2212 - 0.1947
\end{aligned}$$

There is approximately a 2.5% chance that Alabama will be hit by multiple hurricanes this (or any) year. ◇

A.7.5 Proof of the Poisson Limit

It can be shown that the Poisson distribution is actually a Binomial distribution under two requirements: The number of Bernoulli trials is infinite and the expected value is constant. To see this, here is a proof.

Theorem A.2 (Poisson as a Limit)**.** *Let $X \sim Bin(n, \pi)$. Then, if $n \to \infty$ while $\lambda := n\pi$ is constant, then $X \sim \mathcal{P}(\lambda)$.*

Proof.

$$\begin{aligned}\lim_{n\to\infty} \binom{n}{x} \pi^x (1-\pi)^{n-x} &= \lim_{n\to\infty} \frac{n!}{(n-x)!\, x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \lim_{n\to\infty} \frac{n(n-1)(n-2)\cdots(n-x+1)}{n^x} \left(\frac{\lambda^x}{x!}\right) \left(1 - \frac{\lambda}{n}\right)^{n-x}\end{aligned}$$

Now, note these three limits:

$$\begin{aligned}&\lim_{n\to\infty} \frac{n(n-1)(n-2)\cdots(n-x+1)}{n^x} = 1 \\ &\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-x} = 1^{-x} = 1 \\ &\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n} := e^{\lambda}\end{aligned}$$

Using these results and substituting them into the original limit above gives our result:

$$\lim_{n\to\infty} \binom{n}{x} \pi^x (1-\pi)^{n-x} = \frac{1}{x!}\, \lambda^x e^{-\lambda}$$

And, the result is shown. □

A.8: End of Appendix Materials

A.8.1 R Functions

In this appendix chapter, we were introduced to several R functions dealing with discrete probability distributions that may be very useful in the future.

Discrete General Distribution:

sample(n, size, replace=TRUE) Returns `size` random integers between 1 and `n`, inclusive (with replacement).

sample(c(v1,v2,...,vx), size, prob=c(p1,p2,...,px), replace=TRUE) Returns n random values from the set $\{v1, v2, , vx\}$, where the probability of each outcome is $\{p1, p2, ..., px\}$, respectively.

Binomial Distribution:

R uses the parameterization where `size` is the number of trials and `prob` is the success probability. The variable represents the number of successes in the sample.

dbinom(x, size,prob) Returns the probability for an x-value according to the specified Binomial distribution; it calculates $\mathbb{P}[X = x]$.

pbinom(x, size,prob) Returns the cumulative probability for an x-value according to the specified Binomial distribution; it calculates $\mathbb{P}[X \leq x]$.

qbinom(p, size,prob) Returns the quantile (percentile) according to the Binomial distribution specified; it calculates x_p such that $\mathbb{P}\left[X \leq x_p\right] = p$.

rbinom(n, size,prob) Returns n random numbers from the specified Binomial distribution.

Geometric Distribution: `R` uses the parameterization where `prob` is the success probability and the variable represents the number of failures until the first success.

dgeom(x, prob) Returns the probability for an x-value according to the specified Geometric distribution; it calculates $\mathbb{P}[X = x]$.

pgeom(x, prob) Returns the cumulative probability for an x-value according to the specified Geometric distribution; it calculates $\mathbb{P}[X \leq x]$.

qgeom(p, prob) Returns the quantile (percentile) according to the specified Geometric distribution; it calculates x_p such that $\mathbb{P}\left[X \leq x_p\right] = p$.

rgeom(n, prob) Returns n random numbers from the specified Geometric distribution.

Negative Binomial Distribution: `R` uses the parameterization where `size` is the number of successes sought and `prob` is the success probability. The variable represents the number of failures until the `size`$^{\text{th}}$ success.

dnbinom(x,size,prob) Returns the probability for an x-value according to the specified Negative Binomial distribution; it calculates $\mathbb{P}[X = x]$.

pnbinom(x,size,prob) Returns the cumulative probability for an x-value according to the specified Negative Binomial distribution; it calculates $\mathbb{P}[X \leq x]$.

qnbinom(p,size,prob) Returns the quantile (percentile) according to the Negative Binomial distribution specified; that is, it calculates x_p such that $\mathbb{P}\left[X \leq x_p\right] = p$.

rnbinom(n,size,prob) Returns n random numbers from the specified Negative Binomial distribution.

Hypergeometric Distribution: `R` uses the parameterization where `m` is the number of successes in the population, `n` is the number of failures in the population, and `k` is the sample size drawn from the population. The variable represents the number of successes in the sample.

dhyper(x, m,n,k) Returns the probability for an x-value according to the specified Hypergeometric distribution; it calculates $\mathbb{P}[X = x]$.

phyper(x, m,n,k) Returns the cumulative probability for an x-value according to the specified Hypergeometric distribution; it calculates $\mathbb{P}[X \leq x]$.

qhyper(p, m,n,k) Returns the quantile (percentile) according to the specified Hypergeometric distribution; it calculates x_p such that $\mathbb{P}\left[X \leq x_p\right] = p$.

rhyper(n, m,n,k) Returns n random numbers from the specified Hypergeometric distribution.

Poisson Distribution: `R` uses the parameterization where `lambda` is the rate (or expected value). The variable represents the number of successes.

dpois(x, lambda) Returns the probability for an x-value according to the specified Poisson distribution; it calculates $\mathbb{P}[X = x]$.

ppois(x, lambda) Returns the cumulative probability for an x-value according to the specified Poisson distribution; it calculates $\mathbb{P}[X \leq x]$.

qpois(p, lambda) Returns the quantile (percentile) according to the specified Poisson distribution; it calculates x_p such that $\mathbb{P}\left[X \leq x_p\right] = p$.

rpois(n, lambda) Returns n random numbers from the specified Poisson distribution.

A.8.2 Exercises and Extensions

This section offers suggestions on things you can practice from this appendix of discrete distributions.

Summary:

1. Let the random variable X have sample space $\{1, 3, 6\}$ with respective probabilities $\{0.50, 0.25, 0.25\}$. Calculate the following:
   a) $\mathbb{E}[X]$
   b) $\mathbb{V}[X]$
   c) $\mathbb{P}[X = 3]$
   d) $\mathbb{P}[X > 3]$
   e) $\mathbb{P}[X \leq 5]$

2. Let $X \sim Bin(n = 1, \pi = 0.25)$ and $Y \sim Bin(n = 1, \pi = 0.25)$.
   a) What is the distribution of $X + Y$?
   b) What is the distribution of XY?
   c) What is the distribution of X^2?
   d) What is the distribution of $X^2 Y^{0.5}$?
   e) Calculate $\mathbb{P}[XY \geq 0.5]$.
   f) Calculate $\mathbb{P}[X + Y \geq 0.5]$.

3. Let us assume $X \sim Bin(n = 3, \pi = 0.25)$. Calculate the following:
   a) $\mathbb{E}[X]$
   b) $\mathbb{V}[X]$
   c) $\mathbb{P}[X = 0]$
   d) $\mathbb{P}[X > 0]$
   e) $\mathbb{P}[X \leq 5]$

4. Let us assume $X \sim Bin(10, 0.50)$. Calculate the following:

    a) $\mathbb{E}[X]$
    b) $\mathbb{V}[X]$
    c) $\mathbb{P}[X = 0]$
    d) $\mathbb{P}[X > 0]$
    e) $\mathbb{P}[X \leq 5]$

5. Let us assume $X \sim \mathcal{H}(m = 50, n = 50, k = 20)$. Calculate the following:

    a) $\mathbb{E}[X]$
    b) $\mathbb{V}[X]$
    c) $\mathbb{P}[X = 10]$
    d) $\mathbb{P}[X > 10]$
    e) $\mathbb{P}[X \leq 12]$

6. Let us assume $X \sim \mathcal{H}(m = 10, n = 5, k = 2)$. Calculate the following:

    a) $\mathbb{E}[X]$
    b) $\mathbb{V}[X]$
    c) $\mathbb{P}[X = 1]$
    d) $\mathbb{P}[X > 1]$
    e) $\mathbb{P}[X \leq 2]$

7. Let us assume $X \sim \mathcal{G}\text{eom}(\pi = 0.10)$. Calculate the following:

    a) $\mathbb{E}[X]$
    b) $\mathbb{V}[X]$
    c) $\mathbb{P}[X \leq 10]$
    d) $\mathbb{P}[X \leq 20]$
    e) $\mathbb{P}[5 < X \leq 10]$

8. Let us assume $X_i \overset{\text{iid}}{\sim} \mathcal{G}\text{eom}(\pi = 0.10)$. Let us define T as the sum of five such X values; i.e. $T := \sum_{i=1}^{5} X_i$. What is the exact distribution of T? Calculate the following:

   a) $\mathbb{E}[T]$
   b) $\mathbb{V}[T]$
   c) $\mathbb{P}[X \leq 10]$
   d) $\mathbb{P}[T \leq 40]$
   e) $\mathbb{P}[20 < T \leq 40]$

9. Let us assume $X \sim \mathcal{P}(\lambda = 1)$. Calculate the following:

   a) $\mathbb{E}[X]$
   b) $\mathbb{V}[X]$
   c) $\mathbb{P}[X \leq 1]$
   d) $\mathbb{P}[X \leq 2]$
   e) $\mathbb{P}[2 < X \leq 4]$

Data:

10. In Oklahoma, standard license plates consist of six alphanumeric characters. The first three are digits (0–9); the second three are letters (A–Z). All digit and letter combinations are possible (equally likely).

    a) How many such license plates can Oklahoma issue?
    b) With that information, what is the probability that a randomly selected license plate has exactly one 'A' on it?
    c) What is the probability that a randomly selected plate contains no '3' on it?

11. A research scientist desires to test for the presence of a lethal amount of dissolved hydrogen cyanide (HCN) in a sample of water from a water bottler. Unfortunately, this experiment is dangerous, and the scientist is clumsy. Each time he performs this experiment, he has a 5 percent chance of causing some damage to himself and the laboratory. Regardless, he performs the experiment 10 times to get a better estimate of the HCN level in the water sample. Assuming the test results are independent, we know the distribution of the number of experiments that cause damage.

    a) What is the probability that the first experiment causes damage?
    b) What is the probability that the second experiment causes damage, given that the first did not?
    c) What is the probability that none of the 10 experiments cause damage?
    d) What is the probability that at least one of those 10 experiments causes damage?
    e) What is the probability that all 10 of the experiments cause damage?

12. At the beginning of the NCAA Football season, an infamous commentator stated that Oklahoma State University (OSU) only has a 90% chance of winning each of its 12 games this season. Let us assume the commentator is correct, which means the number of games OSU wins this season is a Binomial random variable with $n = 12$ and $\pi = 0.90$.

    a) What is the expected number of games OSU will win this season?
    b) The last game of the regular season is the Bedlam Game against the University of Oklahoma (OU). What is the probability that OSU beats OU?
    c) Another enjoyable game will against the Longhorns of the University of Texas. What was the probability that OSU will beat Texas?
    d) Given that OSU beats Texas, what is the probability that it beats OU?
    e) What is the probability that OSU beats *both* Texas and OU?
    f) What is the probability that OSU beats *neither* Texas nor OU?

13. You professor loves M&M's. According to its website, the distribution of the colors of milk chocolate M&M's is

Color:	Brown	Yellow	Red	Blue	Orange	Green
Proportion:	0.13	0.14	0.13	0.24	0.20	0.16

a) To determine if this is the correct distribution of the colors, your professor decided to perform an experiment, buying a small bag of 20 M&M's. What is the expected number of orange M&M's in the bag? What is the probability of finding exactly 2 orange M&M's? What is the probability of finding no orange M&M's?

b) Now, let us pretend that the provided color distribution is correct. Your professor reaches in the bag and draws out a single M&M. What is the probability that it is green? Placing the M&M back in the bag (whatever its color), shaking it up, and drawing an M&M, what is the probability that this second M&M is green?

c) To continue the experiment, your professor buys another large bag of M&M's with exactly 13 brown, 14 yellow, 13 red, 24 blue, 20 orange, and 16 green M&M's. Reaching in the bag, he draws out a single M&M. What is the probability that it is blue? He eats that M&M (whatever the color) and pull out another random M&M. What is the probability that *this* M&M is blue?

d) Let us still pretend that the provided color distribution is correct and that a new bag of M&M's has exactly 13 brown, 14 yellow, 13 red, 24 blue, 20 orange, and 16 green M&M's. He draws an M&M, records its color, eats it, draws an M&M, records its color, and eats it. What is the probability that he drew a red followed by a green M&M? What is the probability that he drew 2 blue M&M's? Given that the first M&M was yellow, what is the probability that the second is also yellow? What is the probability that both M&M's are yellow?

14. During the 2010 census, the population of Stillwater, OK, was 46,048. Only 432 of this group knows the official state beverage. If one randomly calls 1000 people in Stillwater, what is the probability that none of them knows the official state beverage?

    - What data-collection scheme would produce the Binomial result?
    - What data-collection scheme would produce the Hypergeometric result?
    - Solve this problem both as a Binomial problem and as a Hypergeometric problem. Comment on the difference in probability estimates between the two data-collection schemes.

15. The toll booth on the Stillwater spur of the Cimarron Turnpike is the main entrance to Stillwater from Tulsa. The traffic through the toll booth tends to be rather steady throughout the year — except for the Saturdays that OSU has a home game. On those mornings, the average rate of cars passing through the tollbooth is six cars every 10 minutes. Answer the following for a gameday morning.

    a) What is the probability that no cars will pass through in a 10-minute period?

    b) What is the probability that more than 5 cars will pass through in a 10-minute period?

    c) What is the probability that at least 36 cars will pass through in a given hour?

16. In the Pick 3 lottery game in Oklahoma, the player selects a three-digit number (000 through 999). The lottery authority uses a random number generator to select the winning three digits. If the player's three digits match the winning digits, in order, the player wins $500.

    a) What is the probability of a player winning on a single ticket?

    b) What is the probability of a player winning at least once with three randomly selected tickets (with replacement)?

    c) What is the probability of a player winning at least once with 100 randomly selected tickets (with replacement)?

17. Answer the previous problem's questions where the player ensures that the tickets do not repeat (i.e. *without* replacement) and comment on the difference in probabilities between this problem and the previous one.

18. When opinion polls call residential landlines, only 15% of the calls are answered by a human. You are hired to contact 1048 people to determine their position on a current news topic. What is the expected number of calls you can expect to make in order to reach 1048 people? What is the standard deviation?

    People who make these types of telephone calls for a living charge by the call *and* by the person contacted. A typical rate is $1 per call plus $5 per person interviewed. What is the expected cost of the above poll? What is the standard deviation of that cost?

19. A person claims to have extrasensory perception (ESP). To test this, a researcher places six cards face down on the table in front of the claimant. Each of these cards can be one of the following four shapes: waves, plus, circle, and square. The cards are independent of each other; that is, knowing that the first is a wave does not change the probability that any of the others are waves.

    Assuming that the person does not have ESP and merely guesses randomly, write the probability mass function for the number of cards he guesses correctly. Under this same assumption, what is the expected number of cards he will guess correctly? If he guesses three cards correctly, do we have sufficient proof that he has ESP? What if he guesses all six cards correctly?

20. Officer McGrowl patrols the mean streets of Stillwater, OK, every night looking for people who are driving under the influence of alcohol. McGrowl estimates that 60% of the drivers at 2:00am are driving drunk. If McGrowl begins to randomly pull over drivers, how many sober drivers can he expect to pull over before he catches his first drunk driver? How many sober drivers can he expect to pull over before he gets his fourth drunk driver?

21. Referring to the previous problem, what is the probability that Officer McGrowl pulls over 5 sober drivers before he pulls over his first drunk

driver? What is the probability that he pulls over 15 sober drivers before he pulls over his first drunk driver? What is the probability that he pulls over 10 sober drivers out of the 10 drivers he pulls over in that night?

22. According to the US Census Bureau (2010), 15.3% of all Americans aged 25 and older had neither a high school diploma nor a GED. As a part of your research, you need to contact 500 people from this population. Unfortunately, a complete sampling frame does not exist. Thus, you decide to purchase a mailing list of the general population for $9000. To keep costs down, you only want to send to subsample of this list.

    How many people on the list will you need to contact in order to have a 50% probability of contacting 500 people lacking a diploma and GED? If you decide you need a 95% probability of achieving your goal of 500, how many will you need to mail?

Monte Carlo:

23. Let $X \sim Bin(n = 3, \pi = 0.50)$. Define $Y = X^2$.

    a) Calculate the expected value of X.

    b) Estimate the expected value of Y.

    c) Estimate the standard deviation of Y.

    d) Estimate the first quartile of X and of Y.

24. Let $X \sim \mathcal{P}(\lambda = 2)$ and $Y \sim \mathcal{P}(\lambda = 5)$. Define $W = X + Y$. Determine the exact distribution of W. Simulate from $X + Y$ to check that your answer is reasonable.

25. Is Louisiana's $x_{LA} = 11$ years of hurricane landfalls statistically different from Alabama's $x_{AL} = 4$ for those 20 year

A.8.3 References and Additional Readings

This section provides a list of statistical works. Those works cited in the chapter are here. Also here are works that complement the chapter's topics.

- Ladislaus von Bortkiewicz. (1898) *Das Gesetz der kleinen Zahlen*. Leipzig, Germany: B.G. Teubner.
- Siméon Denis Poisson. (1837) *Probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilitiés*. Paris, France: Bachelier.