Bernoulli Random Variables

Recall that a *Bernoulli* random variable is a random variable whose only possible values are 0 and 1.

Bernoulli Random Variables

Recall that a *Bernoulli* random variable is a random variable whose only possible values are 0 and 1.

In general we can consider a Bernoulli random variable to be the result of an experiment with two outcomes, which for convenience we will label "success" and "failure"

Bernoulli Random Variables

Recall that a *Bernoulli* random variable is a random variable whose only possible values are 0 and 1.

In general we can consider a Bernoulli random variable to be the result of an experiment with two outcomes, which for convenience we will label "success" and "failure"

As before we define the Bernoulli random variable X by agreeing to assign the value of 1 to X if the result of the experiment is "success", and zero if the result is "failure":

$$X = \begin{cases} 1 & \text{if the outcome of the experiment is "success"} \\ 0 & \text{if the outcome of the experiment is "failure"} \end{cases}$$

Bernoulli Random Variables

To be consistent with the Kolmogorov probability axioms the probability of "success" must be a number p between zero and one (inclusive), and the probability of "failure", which is the compliment of "success", must be $1 - p$.

Bernoulli Random Variables

To be consistent with the Kolmogorov probability axioms the probability of "success" must be a number p between zero and one (inclusive), and the probability of "failure", which is the compliment of "success", must be $1 - p$.

This results in the following probability mass function $f(x)$ which we will refer to as the *Bernoulli distribution*:

$$f(x) = P(X = x) = \begin{cases} p & \text{if} \quad x = 1 \\ 1 - p & \text{if} \quad x = 0 \end{cases}$$

Bernoulli Random Variables

To be consistent with the Kolmogorov probability axioms the probability of "success" must be a number p between zero and one (inclusive), and the probability of "failure", which is the compliment of "success", must be $1 - p$.

This results in the following probability mass function $f(x)$ which we will refer to as the *Bernoulli distribution*:

$$f(x) = P(X = x) = \begin{cases} p & \text{if} \quad x = 1 \\ 1 - p & \text{if} \quad x = 0 \end{cases}$$

Most of the discrete probability distributions we will now consider are related to the Bernoulli distribution.

Discrete Distributions

Now consider a series of *independent* experiments, each of which produces a Bernoulli random variable with probability of success p (p is the same for all of the trials)

Discrete Distributions

Now consider a series of *independent* experiments, each of which produces a Bernoulli random variable with probability of success p (p is the same for all of the trials)

The following discrete probability distributions arise from this model:

If the number of trials n is fixed in advance, the number of successes X has a **binomial** distribution

Discrete Distributions

Now consider a series of *independent* experiments, each of which produces a Bernoulli random variable with probability of success p (p is the same for all of the trials)

The following discrete probability distributions arise from this model:

If the number of trials n is fixed in advance, the number of successes X has a **binomial** distribution

If trials continue indefinitely until the first success is obtained, the number of failures obtained X has a **geometric** distribution.

Discrete Distributions

Now consider a series of *independent* experiments, each of which produces a Bernoulli random variable with probability of success p (p is the same for all of the trials)

The following discrete probability distributions arise from this model:

If the number of trials n is fixed in advance, the number of successes X has a **binomial** distribution

If trials continue indefinitely until the first success is obtained, the number of failures obtained X has a **geometric** distribution.

If trials continue indefinitely until the r^{th} success is obtained, the number of failures obtained X has a **negative binomial** distribution.

Discrete Distributions

Note that the geometric distribution is a special case of the negative binomial distribution, with $r = 1$.

Discrete Distributions

Note that the geometric distribution is a special case of the negative binomial distribution, with $r = 1$.

Unfortunately, different authors define the random variable X in the negative binomial (and geometric) distribution in different ways.

Discrete Distributions

Note that the geometric distribution is a special case of the negative binomial distribution, with $r = 1$.

Unfortunately, different authors define the random variable X in the negative binomial (and geometric) distribution in different ways.

However, the characterization as a sequence of Bernoulli trials that ends at the r^{th} success is common to all definitions.

Discrete Distributions

Note that the geometric distribution is a special case of the negative binomial distribution, with $r = 1$.

Unfortunately, different authors define the random variable X in the negative binomial (and geometric) distribution in different ways.

However, the characterization as a sequence of Bernoulli trials that ends at the r^{th} success is common to all definitions.

That said, you should be prepared to encounter a different definition of X (and a different, but equivalent pmf)if you look at a different text.

Discrete Distributions

The other related distribution we will consider is the **Poisson** distribution.

Discrete Distributions

The other related distribution we will consider is the **Poisson** distribution.

The Poisson is a limiting form of the binomial distribution that you get if you let n become very large and the probability of success p very small, but always keep $np = \lambda$ the same.

Discrete Distributions

The other related distribution we will consider is the **Poisson** distribution.

The Poisson is a limiting form of the binomial distribution that you get if you let n become very large and the probability of success p very small, but always keep $np = \lambda$ the same.

Another way to say this is that we take binomial random variables with larger and larger n, but we keep the *expected number of successes* $np = \lambda$ the same for all of them.

Discrete Distributions

The other related distribution we will consider is the **Poisson** distribution.

The Poisson is a limiting form of the binomial distribution that you get if you let n become very large and the probability of success p very small, but always keep $np = \lambda$ the same.

Another way to say this is that we take binomial random variables with larger and larger n, but we keep the *expected number of successes* $np = \lambda$ the same for all of them.

The limit of the distribution of such a sequence of random variables as $n \to \infty$ is a Poisson.

The Binomial Distribution

The binomial experiment consists of:

- n independent Bernoulli trials are performed
- The random variable X is the sum of the results (i.e., the number of successes)
- The probability of success p is the same for all trials

The Binomial Distribution

The binomial experiment consists of:

- n independent Bernoulli trials are performed
- The random variable X is the sum of the results (i.e., the number of successes)
- The probability of success p is the same for all trials

Expected value: $E(X) = np$ Variance: $V(X) = np(1-p)$

The Binomial Distribution

The binomial experiment consists of:

- n independent Bernoulli trials are performed
- The random variable X is the sum of the results (i.e., the number of successes)
- The probability of success p is the same for all trials

Expected value: $E(X) = np$ Variance: $V(X) = np(1-p)$

Computation:

Value	R	Spreadsheet
$P(X = x)$	$dbinom(x, n, p)$	$= BINOMDIST(x, n, p, FALSE)$
$P(X \leq x)$	$pbinom(x, n, p)$	$= BINOMDIST(x, n, p, TRUE)$

The Binomial Distribution

The probability that the Red Sox beat the Yankees in any given game is 0.55.

In the month of June, the teams are scheduled to play each other 9 times.

Find the probability that the Red Sox win exactly 4 games.

The Binomial Distribution

The probability that the Red Sox beat the Yankees in any given game is 0.55.

In the month of June, the teams are scheduled to play each other 9 times.

Find the probability that the Red Sox win exactly 4 games.

Solution: 0.212757

The Binomial Distribution

The probability that the Red Sox beat the Yankees in any given game is 0.55.

In the month of June, the teams are scheduled to play each other 9 times.

Find the probability that the Red Sox win exactly 4 games.

Solution: 0.212757

$dbinom(4, 9, 0.55)$ or $= BINOMDIST(4, 9, 0.55, FALSE)$

The Binomial Distribution

The probability that the Red Sox beat the Yankees in any given game is 0.55.

In the month of June, the teams are scheduled to play each other 9 times.

Find the probability that the Red Sox win 4 games or fewer.

The Binomial Distribution

The probability that the Red Sox beat the Yankees in any given game is 0.55.

In the month of June, the teams are scheduled to play each other 9 times.

Find the probability that the Red Sox win 4 games or fewer.

Solution: 0.3785791

The Binomial Distribution

The probability that the Red Sox beat the Yankees in any given game is 0.55.

In the month of June, the teams are scheduled to play each other 9 times.

Find the probability that the Red Sox win 4 games or fewer.

Solution: 0.3785791

$pbinom(4, 9, 0.55)$ or $= BINOMDIST(4, 9, 0.55, TRUE)$

The Binomial Distribution

The probability that the Red Sox beat the Yankees in any given game is 0.55.

In the month of June, the teams are scheduled to play each other 9 times.

Find the probability that the Red Sox win 4 games or more.

The Binomial Distribution

The probability that the Red Sox beat the Yankees in any given game is 0.55.

In the month of June, the teams are scheduled to play each other 9 times.

Find the probability that the Red Sox win 4 games or more.

Solution: This is one minus the probability that they win three or fewer.

0.834178

The Binomial Distribution

The probability that the Red Sox beat the Yankees in any given game is 0.55.

In the month of June, the teams are scheduled to play each other 9 times.

Find the probability that the Red Sox win 4 games or more.

Solution: This is one minus the probability that they win three or fewer.

0.834178

$1 - pbinom(3, 9, 0.55)$ or
$= 1 - BINOMDIST(3, 9, 0.55, TRUE)$

The Geometric Distribution

The geometric experiment consists of:

- Independent Bernoulli trials are performed until the first "success" is obtained
- The random variable X is the number of failures obtained
- The probability of success p is the same for all trials

The Geometric Distribution

The geometric experiment consists of:

- Independent Bernoulli trials are performed until the first "success" is obtained
- The random variable X is the number of failures obtained
- The probability of success p is the same for all trials

The probability mass function (pmf) $f(x)$ is:

$$f(x) = P(X = x) = g(x; p) = p(1-p)^x, \quad x = 0, 1, 2, 3, \ldots$$

The Geometric Distribution

The geometric experiment consists of:

- Independent Bernoulli trials are performed until the first "success" is obtained
- The random variable X is the number of failures obtained
- The probability of success p is the same for all trials

The Geometric Distribution

The geometric experiment consists of:

- Independent Bernoulli trials are performed until the first "success" is obtained
- The random variable X is the number of failures obtained
- The probability of success p is the same for all trials

The probability mass function (pmf) $f(x)$ is:

$$f(x) = P(X = x) = g(x; p) = p(1 - p)^x, \quad x = 0, 1, 2, 3, \ldots$$

The Geometric Distribution

If you sum the values of $f(x)$ over all values from zero to infinity, the sum is one.

$$\sum_{x=0}^{\infty} p(1-p)^x \;=\; p \cdot \sum_{x=0}^{\infty} (1-p)^n$$

The Geometric Distribution

If you sum the values of $f(x)$ over all values from zero to infinity, the sum is one.

$$\sum_{x=0}^{\infty} p(1-p)^x \;=\; p \cdot \sum_{x=0}^{\infty} (1-p)^n$$

The sum is now a geometric series with $r = 1 - p$. The sum of a geometric series is $1/(1-r)$, so

$$p \cdot \sum_{x=0}^{\infty} (1-p)^n \;=\; p \cdot \left(\frac{1}{1-(1-p)} \right) \;=\; p \cdot \frac{1}{p} \;=\; 1$$

The Geometric Distribution

The expected value of a geometric random variable $E(X)$ is:

$$E(X) = \sum_{x=0}^{n} x \cdot f(x)$$

The Geometric Distribution

The expected value of a geometric random variable $E(X)$ is:

$$E(X) = \sum_{x=0}^{n} x \cdot f(x)$$

$$E(X) = \sum x \cdot (1-p)^x \; = \; \frac{1-p}{p}$$

The Geometric Distribution

To find the variance $V(X)$ of a geometric random variable, first we find $E(X^2)$:

$$E(X^2) = \sum_{x=0}^{n} x^2 \cdot f(x)$$

The Geometric Distribution

To find the variance $V(X)$ of a geometric random variable, first we find $E(X^2)$:

$$E(X^2) = \sum_{x=0}^{n} x^2 \cdot f(x)$$

$$E(X^2) = \sum x^2 \cdot x(1-p)x \;=\; \frac{2-3p+p^2}{p^2}$$

The Geometric Distribution

To find the variance $V(X)$ of a geometric random variable, first we find $E(X^2)$:

$$E(X^2) = \sum_{x=0}^{n} x^2 \cdot f(x)$$

$$E(X^2) = \sum x^2 \cdot x(1-p)x \; = \; \frac{2-3p+p^2}{p^2}$$

Then

$$V(X) = E(X^2) - [E(X)]^2 \; = \; \frac{1-p}{p^2}$$

The Geometric Distribution

Now we will perform some numerical experiments.

First generate a sample of $1,000,000$ observations for a geometric experiment with probability of success $p = 0.4$ at each trial:

```
x<-rgeom(1000000,0.4)
```

The Geometric Distribution

Now we will perform some numerical experiments.

First generate a sample of $1,000,000$ observations for a geometric experiment with probability of success $p = 0.4$ at each trial:

x<-rgeom(1000000,0.4)

Now plot a histogram of the results:

hist(x)

The Geometric Distribution

Now we will perform some numerical experiments.

First generate a sample of $1,000,000$ observations for a geometric experiment with probability of success $p = 0.4$ at each trial:

x<-rgeom(1000000,0.4)

Now plot a histogram of the results:

hist(x)

To get a table of the results enter

table(x)

The Geometric Distribution

Now we will perform some numerical experiments.

First generate a sample of $1,000,000$ observations for a geometric experiment with probability of success $p = 0.4$ at each trial:

x<-rgeom(1000000,0.4)

Now plot a histogram of the results:

hist(x)

To get a table of the results enter

table(x)

The results through $X = 6$ should look something like:

0	1	2	3	4	5	6
399422	240431	144595	86377	51550	31004	18720

The Geometric Distribution

0	1	2	3	4	5	6
399422	240431	144595	86377	51550	31004	18720

Now compare the frequencies to the probabilities.

First compute the probability that $X = 0$:

dgeom(0,0.4)

The Geometric Distribution

0	1	2	3	4	5	6
399422	240431	144595	86377	51550	31004	18720

Now compare the frequencies to the probabilities.

First compute the probability that $X = 0$:

dgeom(0,0.4)

The result should be something like

[1]   0.4

The Geometric Distribution

0	1	2	3	4	5	6
399422	240431	144595	86377	51550	31004	18720

Now compare the frequencies to the probabilities.

First compute the probability that $X = 0$:

dgeom(0,0.4)

The result should be something like

[1]   0.4

To get the probability that $X = 1$ enter

dgeom(1,0.4)

The Geometric Distribution

0	1	2	3	4	5	6
399422	240431	144595	86377	51550	31004	18720

Now compare the frequencies to the probabilities.

First compute the probability that $X = 0$:

dgeom(0,0.4)

The result should be something like

[1]   0.4

To get the probability that $X = 1$ enter

dgeom(1,0.4)

This time the results should look something like:

[1]   0.24

The Geometric Distribution

0	1	2	3	4	5	6
399422	240431	144595	86377	51550	31004	18720

Next compute the probability that $X = 2$:

dgeom(2,0.4)

The Geometric Distribution

0	1	2	3	4	5	6
399422	240431	144595	86377	51550	31004	18720

Next compute the probability that $X = 2$:

dgeom(2,0.4)

The result should be something like

[1]   0.144

The Geometric Distribution

0	1	2	3	4	5	6
399422	240431	144595	86377	51550	31004	18720

Next compute the probability that $X = 2$:

dgeom(2,0.4)

The result should be something like

[1]   0.144

To get the probability that $X = 5$ enter

dbinom(1,5,0.4)

The Geometric Distribution

0	1	2	3	4	5	6
399422	240431	144595	86377	51550	31004	18720

Next compute the probability that $X = 2$:

dgeom(2,0.4)

The result should be something like

[1]   0.144

To get the probability that $X = 5$ enter

dbinom(1,5,0.4)

This time the results should look something like:

[1]   0.031104

The Geometric Distribution

The expected value $E(X)$ in this case is:

$$E(X) = \frac{1-p}{p} = \frac{.6}{.4} = 1.5$$

The Geometric Distribution

The expected value $E(X)$ in this case is:

$$E(X) = \frac{1-p}{p} = \frac{.6}{.4} = 1.5$$

To compute the sample mean $\overline{x}$, enter

mean(x)

The Geometric Distribution

The expected value $E(X)$ in this case is:

$$E(X) = \frac{1-p}{p} = \frac{.6}{.4} = 1.5$$

To compute the sample mean $\overline{x}$, enter

mean(x) The result should be something like

[1] 1.499121

The Geometric Distribution

The variance $V(X)$ in this case is:

$$V(X) = \frac{1-p}{p^2} = \frac{.6}{.4^2} = 3.75$$

The Geometric Distribution

The variance $V(X)$ in this case is:

$$V(X) = \frac{1-p}{p^2} = \frac{.6}{.4^2} = 3.75$$

To compute the sample variance s^2, enter

var(x)

The Geometric Distribution

The variance $V(X)$ in this case is:

$$V(X) = \frac{1-p}{p^2} = \frac{.6}{.4^2} = 3.75$$

To compute the sample variance s^2, enter

var(x) The result should be something like

[1] 3.733986

The Geometric Distribution

A fair coin is tossed until the first heads comes up.

Find the probability that the first heads comes up on the fifth toss.

The Geometric Distribution

A fair coin is tossed until the first heads comes up.

Find the probability that the first heads comes up on the fifth toss.

Solution: 0.03125

The Geometric Distribution

A fair coin is tossed until the first heads comes up.

Find the probability that the first heads comes up on the fifth toss.

Solution: 0.03125

$dgeom(4, 0.5)$ or $= GEOMDIST(4, 0.5, FALSE)$

(Spreadsheet function is for GNUMERIC. EXCEL does not have this function)

The Geometric Distribution

A fair coin is tossed until the first heads comes up.

Find the probability that the first heads comes up on the fifth toss or sooner.

The Geometric Distribution

A fair coin is tossed until the first heads comes up.

Find the probability that the first heads comes up on the fifth toss or sooner.

Solution: 0.96875

The Geometric Distribution

A fair coin is tossed until the first heads comes up.

Find the probability that the first heads comes up on the fifth toss or sooner.

Solution: 0.96875

$pgeom(4, 0.5)$ or $= GEOMDIST(4, 0.5, TRUE)$

(Spreadsheet function is for GNUMERIC. EXCEL does not have this function)

The Geometric Distribution

A fair coin is tossed until the first heads comes up.

Find the probability that this takes more than 9 tosses.

The Geometric Distribution

A fair coin is tossed until the first heads comes up.

Find the probability that this takes more than 9 tosses.

Solution: 0.001953

The Geometric Distribution

A fair coin is tossed until the first heads comes up.

Find the probability that this takes more than 9 tosses.

Solution: 0.001953

$1 - pgeom(8, 0.5)$ or $= 1 - GEOMDIST(8, 0.5, TRUE)$

(Spreadsheet function is for GNUMERIC. EXCEL does not have this function)

The Geometric Distribution

A fair coin is tossed until the first heads comes up.

Find the probability that this takes 9 or more tosses.

The Geometric Distribution

A fair coin is tossed until the first heads comes up.

Find the probability that this takes 9 or more tosses.

Solution: 0.00390625

The Geometric Distribution

A fair coin is tossed until the first heads comes up.

Find the probability that this takes 9 or more tosses.

Solution: 0.00390625

$1 - pgeom(7, 0.5)$ or $= 1 - GEOMDIST(7, 0.5, TRUE)$

(Spreadsheet function is for GNUMERIC. EXCEL does not have this function)

The Geometric Distribution

A baseball player has a $.300$ batting average.

Find the probability that their first hit in a game occurs on the 4^{th} time at bat.

The Geometric Distribution

A baseball player has a $.300$ batting average.

Find the probability that their first hit in a game occurs on the 4^{th} time at bat.

Solution: 0.1029

The Geometric Distribution

A baseball player has a $.300$ batting average.

Find the probability that their first hit in a game occurs on the 4^{th} time at bat.

Solution: 0.1029

$dgeom(3, 0.5)$ or $= GEOMDIST(3, 0.5, FALSE)$

(Spreadsheet function is for GNUMERIC. EXCEL does not have this function)