Probability, 2

Frank Edwards

Three kinds of probability

Joint probability

The joint probability of two events (A and B) occurring is expressed as

$$P(A \text{ and } B)$$

Marginal probability

The marginal probability of an event B is

$$P(B)$$

Conditional probability

The conditional probability of event A occurring given that event B occurred is the ratio of the joint probability of A and B divided by the marginal probability of B

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

Working with some real data

Voter files

```r
data("FLVoters")
voters<-na.omit(FLVoters)
head(voters)
```

```
##      surname county VTD age gender  race
## 1     PIEDRA    115  66  58      f white
## 2      LYNCH    115  13  51      m white
## 4    LATHROP    115  80  54      m white
## 5     HUMMEL    115   8  77      f white
## 6 CHRISTISON    115  55  49      m white
## 7      HOMAN    115  84  77      f white
```

Marginal probability

What is the probability that a randomly sampled voter in the population is Black: $P(Black) = ?$

```
voters %>%
  count(race, name = "voters") %>%
  mutate(p = voters/sum(voters))
```

```
##       race voters           p
## 1    asian    175 0.019203336
## 2    black   1194 0.131021617
## 3 hispanic   1192 0.130802151
## 4   native     29 0.003182267
## 5    other    310 0.034017338
## 6    white   6213 0.681773291
```

Is a woman: $P(Woman) = ?$

```
voters %>%
  count(gender) %>%
  mutate(n = n/sum(n))
```

```
##   gender         n
## 1      f 0.5358279
## 2      m 0.4641721
```

Joint probability

What is the probability that a voter is a Black woman:

$P(\text{Black and woman}) = ?$

```
voters %>%
  count(gender, race) %>%
  mutate(n = n/sum(n)) %>%
  pivot_wider(names_from = gender, values_from = n)
```

```
## # A tibble: 6 x 3
##   race           f       m
##   <chr>      <dbl>   <dbl>
## 1 asian    0.00911 0.0101
## 2 black    0.0744  0.0566
## 3 hispanic 0.0731  0.0577
## 4 native   0.00187 0.00132
## 5 other    0.0173  0.0167
## 6 white    0.360   0.322
```

What is the probability that a voter is a woman?

```
## # A tibble: 6 x 3
##   race           f       m
##   <chr>      <dbl>   <dbl>
## 1 asian    0.00911 0.0101
## 2 black    0.0744  0.0566
## 3 hispanic 0.0731  0.0577
## 4 native   0.00187 0.00132
## 5 other    0.0173  0.0167
## 6 white    0.360   0.322
```

Use the law of total probability:

$$P(A) = P(A \text{ and } B) + P(A \text{ and not } B)$$

put differently, for all categories of B i:

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

Conditional probability

If a voter is a man, what is the probability that he is Asian: $P(\text{Asian}|\text{man}) = ?$

```
voters %>%
  filter(gender=="m") %>%
  count(race) %>%
  mutate(n=n/sum(n))
```

```
##       race           n
## 1    asian 0.021749409
## 2    black 0.121985816
## 3 hispanic 0.124349882
## 4   native 0.002836879
## 5    other 0.035933806
## 6    white 0.693144208
```

Conditional probability

Alternatively, we can use the definition of conditional probability as the ratio of the joint probability to the marginal probability:

$$P(\text{Asian}|\text{man}) = \frac{P(\text{Asian and man})}{P(\text{man})}$$

```
voters %>%
  count(gender, race) %>%
  mutate(n = n/sum(n))%>%
  pivot_wider(names_from = gender, values_from = n)
```

```
## # A tibble: 6 x 3
##   race           f       m
##   <chr>      <dbl>   <dbl>
## 1 asian    0.00911 0.0101
## 2 black    0.0744  0.0566
## 3 hispanic 0.0731  0.0577
## 4 native   0.00187 0.00132
## 5 other    0.0173  0.0167
## 6 white    0.360   0.322
```

Conditioning on more than one variable

What is the probability that a male voter over age 60 is white?

$$P(\text{white} | \text{male and over 60})$$

```r
voters %>%
  mutate(over60=age>60) %>%
  count(over60, gender, race) %>%
  mutate(n=n/sum(n)) %>%
  pivot_wider(names_from = gender, values_from = n)
```

```
## # A tibble: 12 x 4
##    over60 race            f        m
##    <lgl>  <chr>       <dbl>    <dbl>
##  1 FALSE  asian    0.00691  0.00823
##  2 FALSE  black    0.0555   0.0435
##  3 FALSE  hispanic 0.0549   0.0436
##  4 FALSE  native   0.00121  0.000768
##  5 FALSE  other    0.0124   0.0129
##  6 FALSE  white    0.212    0.198
##  7 TRUE   asian    0.00219  0.00187
##  8 TRUE   black    0.0189   0.0132
##  9 TRUE   hispanic 0.0182   0.0142
## 10 TRUE   native   0.000658 0.000549
## 11 TRUE   other    0.00494  0.00373
## 12 TRUE   white    0.148    0.124
```

Conditioning on more than one variable

In general:

$$P(A \text{ and } B | C) = \frac{P(A \text{ and } B \text{ and } C)}{P(C)}$$

and

$$P(A | B \text{ and } C) = \frac{P(A \text{ and } B \text{ and } C)}{P(B \text{ and } C)}$$

Independence

Two events are independent if knowledge of one event gives us no information about the other event.

$P(A|B) = P(A)$ and $P(B|A) = P(B)$

$$A \perp B$$

if and only if

$$P(A \text{ and } B) = P(A)P(B)$$

Recall that a Bayesian perspective treats probability as a subjective opinion about how likely an event is. How should we change our beliefs after we make observations about the world?

Bayes' rule

Recall that a Bayesian perspective treats probability as a subjective opinion about how likely an event is. How should we change our beliefs after we make observations about the world?

Bayes' rule formalizes how we should update our beliefs based on evidence:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes' rule

Recall that a Bayesian perspective treats probability as a subjective opinion about how likely an event is. How should we change our beliefs after we make observations about the world?

Bayes' rule formalizes how we should update our beliefs based on evidence:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Prior beliefs and evidence

If we have a *prior* belief that event A has $P(A)$ chance of occurring, then we observe some data, represented as event B, we update our beliefs and obtain a *posterior probability* $P(A|B)$.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Example: Detecting breast cancer

How good is a mammogram at detecting breast cancer?

What we know: One percent of women have breast cancer. 80 percent of people who have cancer and take a mammogram test positive. 9.6 percent of people who take a mammogram get a positive result when they do not have breast cancer.

Example: Detecting breast cancer

How good is a mammogram at detecting breast cancer?

What we know: One percent of women have breast cancer. 80 percent of people who have cancer and take a mammogram test positive. 9.6 percent of people who take a mammogram get a positive result when they do not have breast cancer.

If you take a mammogram and get a positive result, what is the probability that you have breast cancer?

Rewriting as probabilities

One percent of women have breast cancer

$$P(\text{Cancer}) = 0.01$$

Rewriting as probabilities

One percent of women have breast cancer

$$P(\text{Cancer}) = 0.01$$

80 percent of people who have cancer and take a mammogram test positive

$$P(\text{Test positive}|\text{Cancer}) = 0.8$$

Rewriting as probabilities

One percent of women have breast cancer

$$P(\text{Cancer}) = 0.01$$

80 percent of people who have cancer and take a mammogram test positive

$$P(\text{Test positive}|\text{Cancer}) = 0.8$$

9.6 percent of people who take a mammogram get a positive result when they do not have breast cancer

$$P(\text{Test positive}|\text{No cancer}) = 0.096$$

Using Bayes' rule

The prior probability of having cancer is 0.01. How should we update our belief that someone has cancer based on a positive test?

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Using the law of total probability, we can rewrite the denominator as:

$$P(B) = P(B|A)P(A) + P(B|\text{ not A})P(\text{not } A)$$

Using Bayes' rule

We can apply Bayes' rule for A = Cancer, B = positive test:

$$P(\text{Cancer}|\text{Test positive}) =$$

$$\frac{P(\text{Test positive}|\text{Cancer})P(\text{Cancer})}{P(\text{Test positive})}$$

$$P(\text{Cancer}|\text{Test positive}) = \frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.096 \times 0.99}$$

```r
(0.8 *  0.01)/(0.8 *  0.01 + 0.096 * 0.99)
```

```
## [1] 0.07763975
```

The probability that someone has cancer given a prior probability of one percent and a positive test is about 0.078. What would the probability of a true positive be if the test were more sensitive? Say 0.95?

Random Variables

Random Variables

- Random variables assign values to events

Random Variables

- Random variables assign values to events
- Each value is mutually exclusive

Random Variables

- Random variables assign values to events
- Each value is mutually exclusive
- The set of all values is exhaustive (the sample space Ω)

Random Variables

- Random variables assign values to events
- Each value is mutually exclusive
- The set of all values is exhaustive (the sample space Ω)
- Discrete random variables take a finite number of values (e.g. TRUE, FALSE)

Random Variables

- Random variables assign values to events
- Each value is mutually exclusive
- The set of all values is exhaustive (the sample space Ω)
- Discrete random variables take a finite number of values (e.g. TRUE, FALSE)
- Continuous random variables are real numbers, and take on an infinite number of values

The simplest random variable: the binary Bernoulli

Any random variable with two values is called a Bernoulli random variable.

The simplest random variable: the binary Bernoulli

Any random variable with two values is called a Bernoulli random variable.

Bernoulli (binary) variables are typically represented as $[0, 1]$ or $[T, F]$. They can also be two-level character variables, like $[\text{pass, fail}]$ or $[\text{plaid, stripes}]$.

A coin flip as a Bernoulli random variable

A coin flip can be defined as a discrete random variable X

A coin flip as a Bernoulli random variable

A coin flip can be defined as a discrete random variable X

- If the coin lands on heads, $X = 1$
- If the coin lands on tails, $X = 0$

A coin flip as a Bernoulli random variable

A coin flip can be defined as a discrete random variable X

- If the coin lands on heads, $X = 1$
- If the coin lands on tails, $X = 0$

The probability of a Bernoulli variable is the probability of success, or $X = 1$

A coin flip as a Bernoulli random variable

A coin flip can be defined as a discrete random variable X

- If the coin lands on heads, $X = 1$
- If the coin lands on tails, $X = 0$

The probability of a Bernoulli variable is the probability of success, or $X = 1$

$P(X = 1) = p$

Random variable (probability distribution) notation

$$X \sim Bernoulli(p)$$

Random variable (probability distribution) notation

$$X \sim Bernoulli(p)$$

Reads: X is a Bernoulli distributed random variable with probability p

Random variable (probability distribution) notation

$$X \sim Bernoulli(p)$$

Reads: X is a Bernoulli distributed random variable with probability p

In this notation, we name the variable X, note that it is randomly distributed $\sim$, name the distribution it follows *Bernoulli*, and list the parameters for that distribution p.

Let’s flip some coins

```
set.seed(12345)
sample_of_flips<-rbinom(5, 1, 0.5)
table(sample_of_flips)
```

```
## sample_of_flips
## 0 1
## 1 4
```

Let's flip some coins

```r
set.seed(12345)
sample_of_flips<-rbinom(5, 1, 0.5)
table(sample_of_flips)
```

```
## sample_of_flips
## 0 1
## 1 4
```

This is the result of taking 5 draws from a Bernoulli random variable with probability 0.5.

Describing a probability distribution: probability mass

We use a probability mass function to show how likely each value is in a random variable

The probability mass function (PMF) of a variable X is defined as the probability that a variable takes on a particular value x.

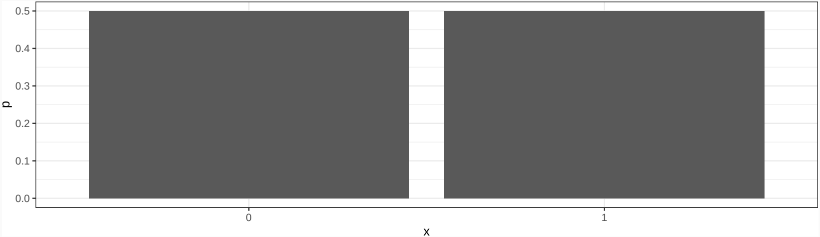
$$PMF(x) = P(X = x)$$

For a Bernoulli variable, $PMF(X = 1) = p$ and $PMF(X = 0) = 1 - p$

The probability mass function for our coin flip

$$PMF(X = 1) = p = 0.5$$

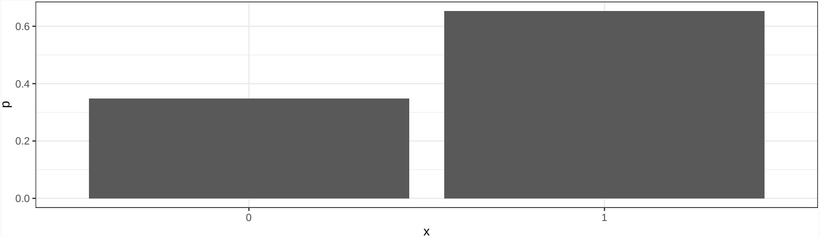
$$PMF(X = 0) = 1 - p = 0.5$$



The probability mass function for passing the bar in NJ (p=0.653)

$$PMF(X=1)=p=0.653$$

$$PMF(X=0)=1-p=0.347$$



Describing a probability distribution: cumulative probability

How likely is a variable to take a value less than or equal to a specified value?

We define the cumulative distribution function as the sum of all probabilities up to a value x

$$CDF(X) = P(X \leq x) = \sum_{k \leq x}^{x} PMF(k)$$

The CDF always ranges from 0 to 1, and never decreases as x increases.

Uniform random variables

Uniform random variables have an equal probability of taking any real value within a given interval $[a, b]$.

Uniform random variables

Uniform random variables have an equal probability of taking any real value within a given interval $[a, b]$.

What does $X \sim Uniform(0, 10)$ look like?

Uniform random variables

Uniform random variables have an equal probability of taking any real value within a given interval $[a, b]$.

What does $X \sim Uniform(0, 10)$ look like?

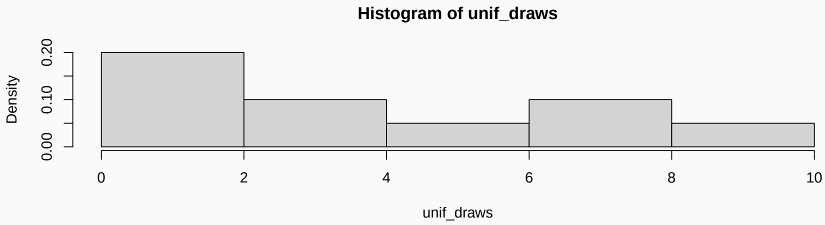
Uniform random variables

Let's simulate it! 10 draws

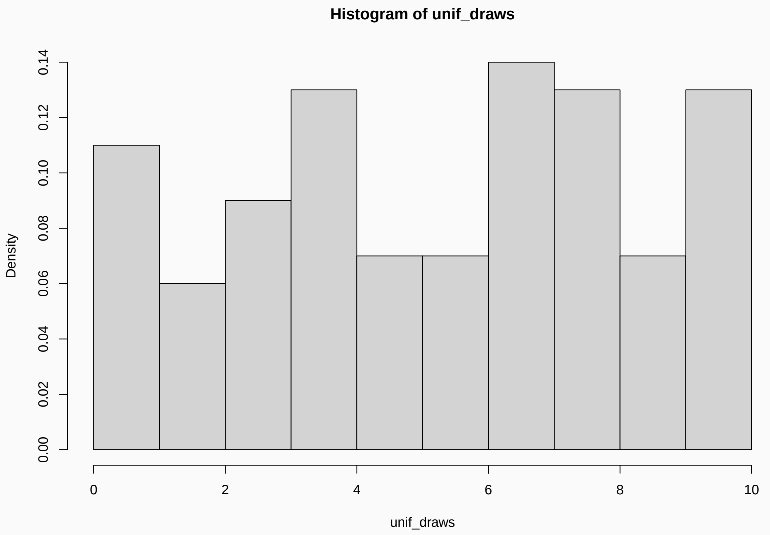
```
unif_draws<-runif(10, min=0, max=10)
unif_draws
```

```
##  [1] 1.66371785 3.25095387 5.09224336 7.27705254 9.89736938 0.34535435
##  [7] 1.52373490 7.35684952 0.01136587 3.91203335
```

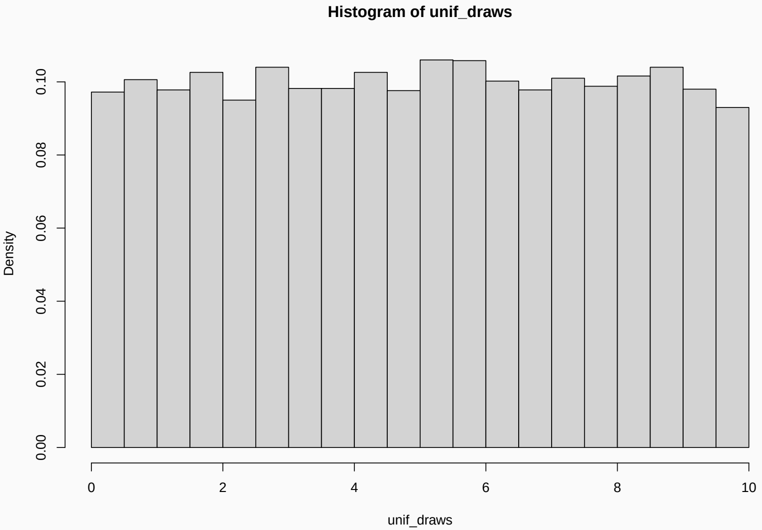
```
hist(unif_draws, freq=F)
```



Uniform random variable: 100 draws



Uniform random variable: 10000 draws



Properties of uniform random variables

For a uniform random variable on the interval $[a, b]$, the probability of drawing any value between a and b is

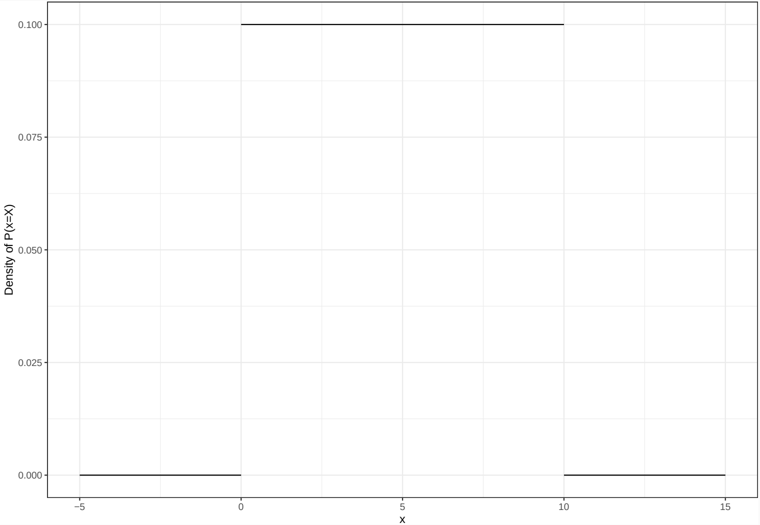
$$\frac{1}{b-a}$$

Formally, the PDF (density, not mass for continuous) and CDF are defined as

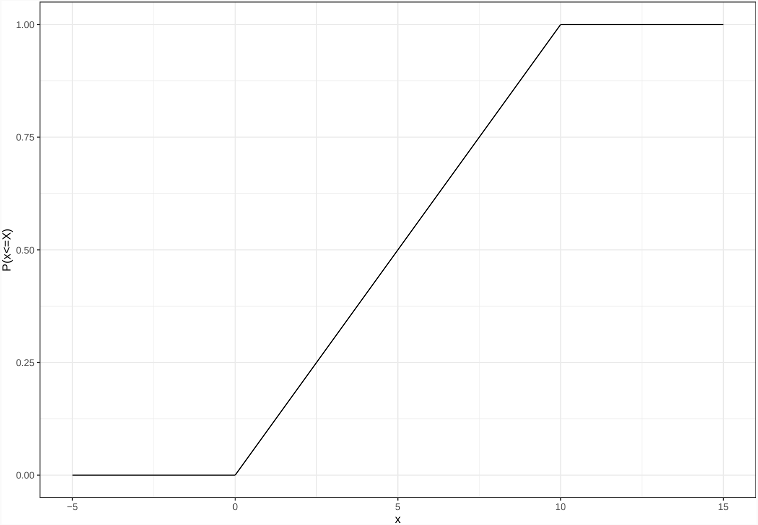
$$\text{PDF:} \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

$$\text{CDF:} \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } x \in [a, b) \\ 1 & \text{for } x \geq b \end{cases}$$

Probability Density function for $X \sim Uniform(0, 10)$



Cumulative Distribution Function for $X \sim Uniform(0, 10)$



A note on CDF for continuous variables

Recall that a CDF for a discrete variable is the sum of all probabilities for values $x \leq X$

We can't sum over each value when X is continuous. Instead, we'll take the integral

$$CDF(x) = P(x \leq X) = \int_{-\infty}^{x} PDF(x)dx$$

When we repeat Bernoulli trials many times, we get a binomial random variable.

The binomial distribution

When we repeat Bernoulli trials many times, we get a binomial random variable.

Binomial random variables represent the count of successes in a fixed number of trials of a Bernoulli experiment.

Formally:

A binomial random variable is the sum of n independently and identically distributed (i.i.d) Bernoulli random variables.

When we repeat Bernoulli trials many times, we get a binomial random variable.

Binomial random variables represent the count of successes in a fixed number of trials of a Bernoulli experiment.

Formally:

A binomial random variable is the sum of n independently and identically distributed (i.i.d) Bernoulli random variables.

Binomial variables take on integer values between 0 and n

Back to flipping coins

Imagine we flipped a coin 5 times, and then repeated the exercise twice more

```
## [1] 0 1 0 0 0
```

```
## [1] 0 0 0 0 1
```

```
## [1] 1 1 1 1 0
```

Each of these trials is a sample from $X \sim Binomial(n, p)$ where $n = 5$ and $p = 0.5$

Back to flipping coins

Imagine we flipped a coin 5 times, and then repeated the exercise twice more

```
## [1] 0 1 0 0 0
```

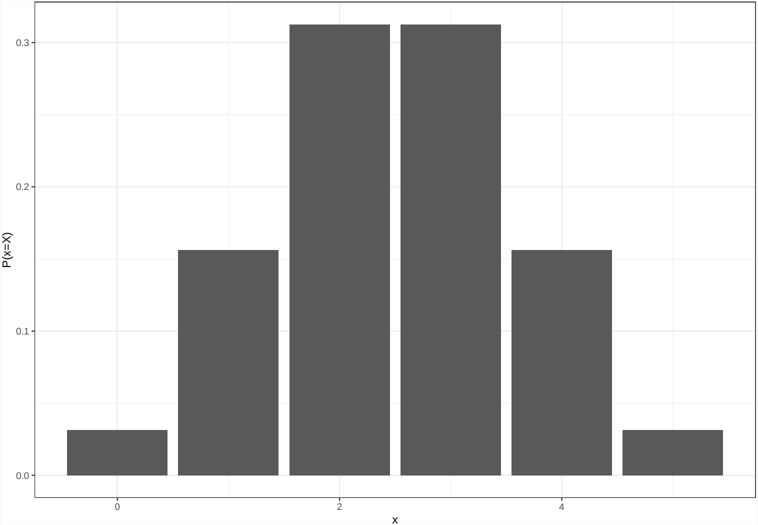
```
## [1] 0 0 0 0 1
```

```
## [1] 1 1 1 1 0
```

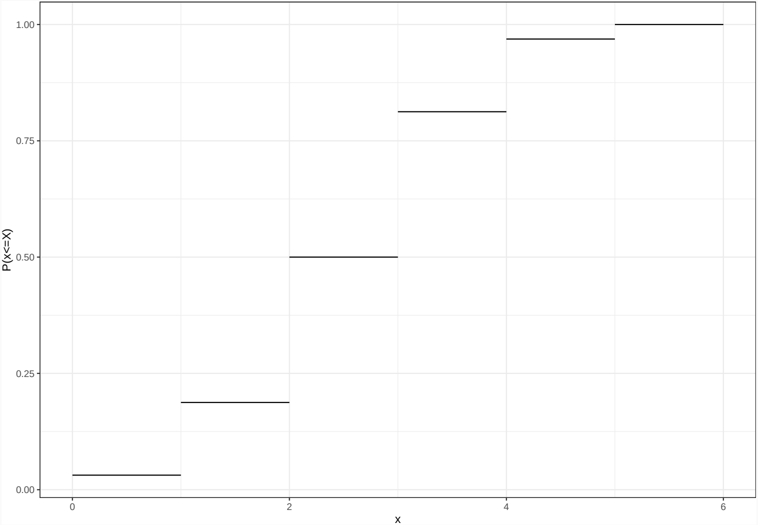
Each of these trials is a sample from $X \sim Binomial(n, p)$ where $n = 5$ and $p = 0.5$

What is x for each trial?

Probability Mass Function for $X \sim Binomial(5, 0.5)$



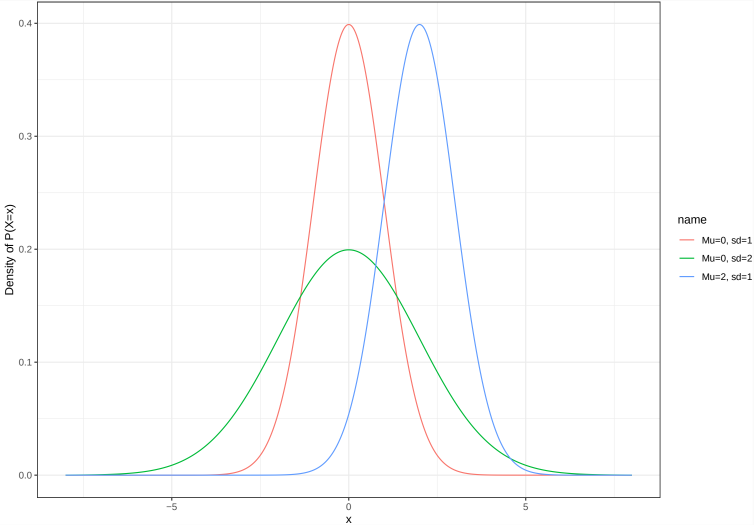
Cumulative Distribution Function for $X \sim Binomial(5, 0.5)$



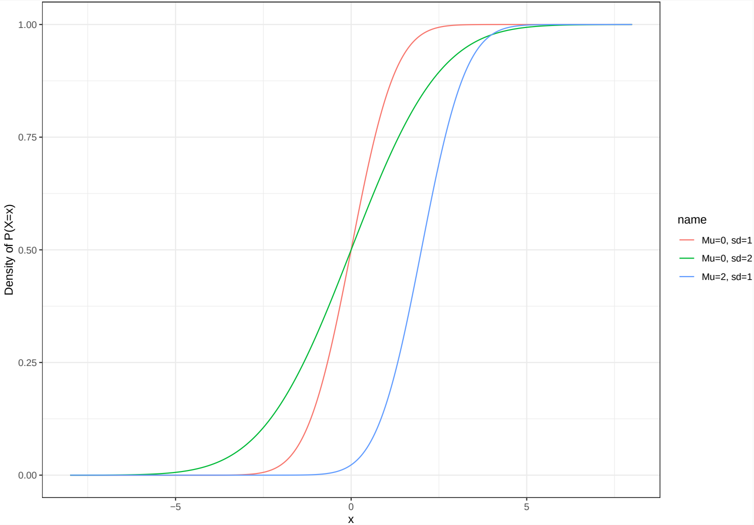
The Normal (Gaussian) distribution is continuous, and takes on values from $[-\infty, \infty]$. It has two parameters, the mean μ and standard deviation σ (or variance σ^2).

- μ determines the location of the distribution
- σ determines the spread of the distribution

The Normal PDF



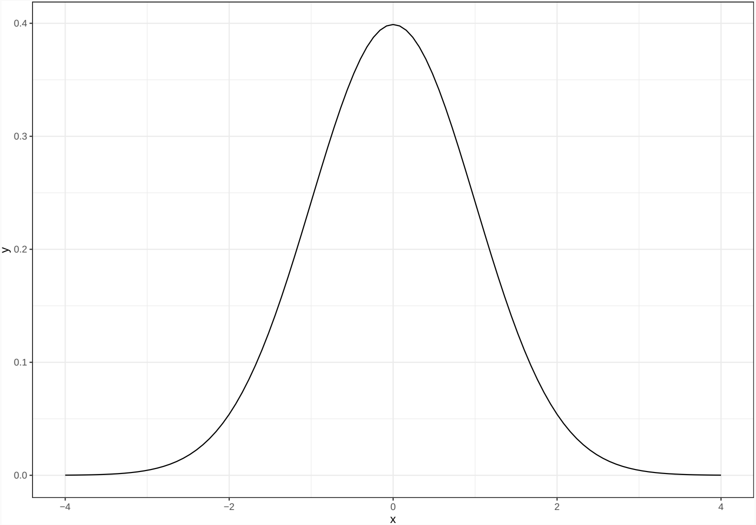
The Normal CDF



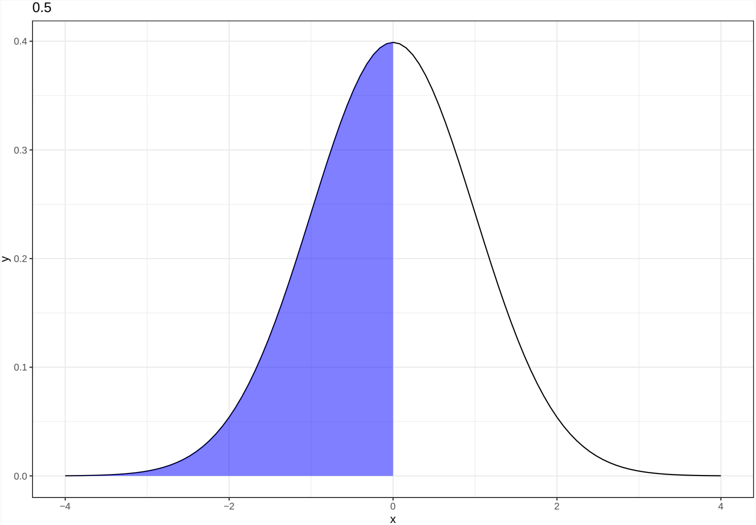
Special features of Normal distributions:

- The sum of many random variables from other distributions are often Normal
- For $X \sim N(\mu, \sigma^2)$, $Z = X + c$ is also Normal: $Z \sim (\mu + c, \sigma^2)$
- $Z = cX$ is distributed $Z \sim N(c\mu, (c\sigma)^2)$
- Z-scores of a Normal random variable are $N(0, 1)$

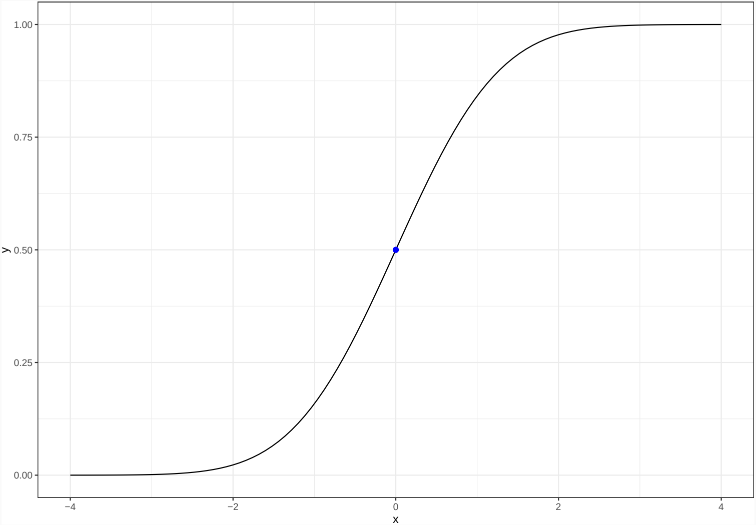
Area under the curve: interpreting the PDF and CDF



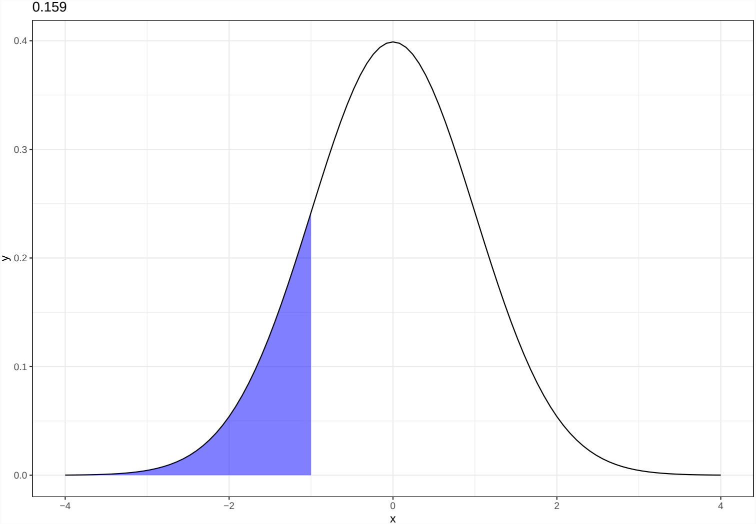
Area under the curve: interpreting the PDF and CDF



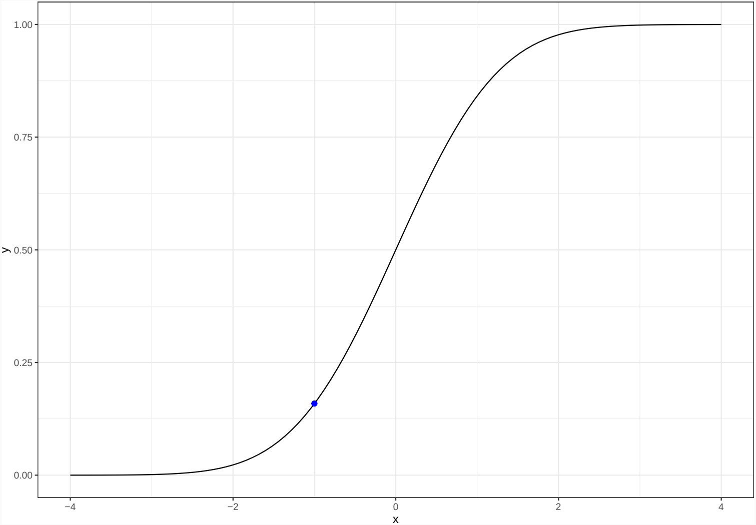
Area under the curve: interpreting the PDF and CDF



Area under the curve: interpreting the PDF and CDF



Area under the curve: interpreting the PDF and CDF



Z-scores and area under the curve

Recall that to obtain a z-score, we subtract the mean and divide by the standard deviation:

$$\text{z-score} = \frac{X - \mu}{\sigma}$$

For a Normal variable, z-scores are distributed $z \sim N(0, 1)$

Recall that to obtain a z-score, we subtract the mean and divide by the standard deviation:

$$\text{z-score} = \frac{X - \mu}{\sigma}$$

For a Normal variable, z-scores are distributed $z \sim N(0, 1)$

What does a z-score of 0 indicate?

Recall that to obtain a z-score, we subtract the mean and divide by the standard deviation:

$$\text{z-score} = \frac{X - \mu}{\sigma}$$

For a Normal variable, z-scores are distributed $z \sim N(0, 1)$

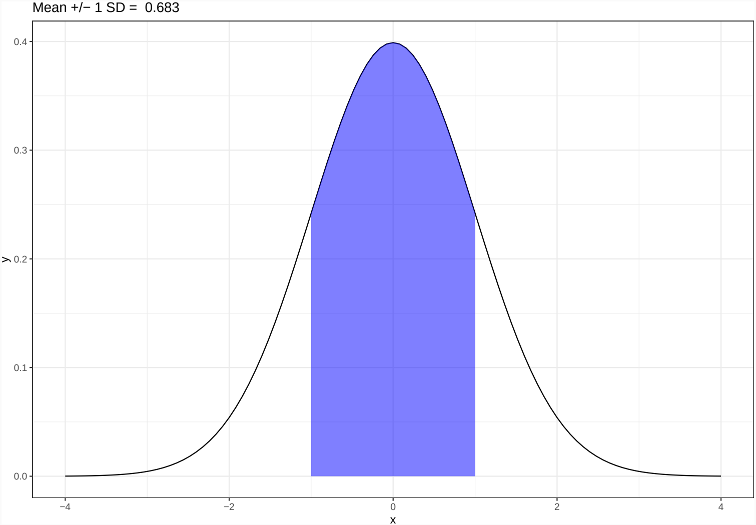
What does a z-score of 0 indicate? -1?

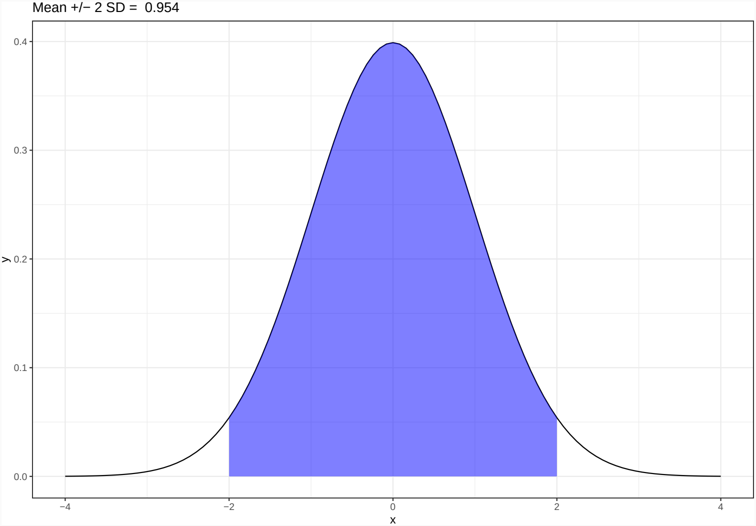
Recall that to obtain a z-score, we subtract the mean and divide by the standard deviation:

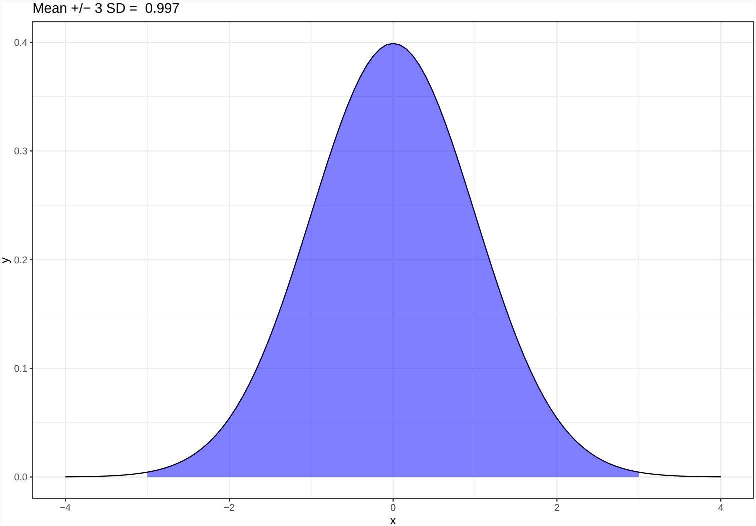
$$\text{z-score} = \frac{X - \mu}{\sigma}$$

For a Normal variable, z-scores are distributed $z \sim N(0, 1)$

What does a z-score of 0 indicate? -1? 2?


Mean +/− 1 SD = 0.683
0.4
0.3
0.2
0.1
0.0
y
−4
−2
0
2
4
x


Mean +/− 2 SD = 0.954
0.4
0.3
0.2
0.1
0.0
y
−4
−2
0
2
4
x



Mean +/− 3 SD = 0.997
0.4
0.3
0.2
0.1
0.0
y
−4
−2
0
2
4
x

Useful probability distribution functions

```r
### Normal(0,1) probability density function
dnorm(x = 0, mean = 0, sd = 1)
```

```
## [1] 0.3989423
```

```r
### Normal(0,1) cumulative distribution function
pnorm(q = 0, mean = 0, sd = 1)
```

```
## [1] 0.5
```

```r
### Random draw from a normal(0,1) distribution
rnorm(n = 1, mean = 0, sd = 1)
```

```
## [1] 1.231011
```

```r
### CDF position for a given probability (quantile)
qnorm(p = 0.75, mean = 0, sd = 1)
```

```
## [1] 0.6744898
```

```r
### You can also use dbinom(), pbinom(), rbinom(), qbinom()
```

The expectation of a random variable

The expectation of a random variable $E(X)$ is the mean of a random variable.

Be careful not to confuse $E(X)$ and $\bar{x}$.

The expectation of a random variable

The expectation of a random variable $E(X)$ is the mean of a random variable.

Be careful not to confuse $E(X)$ and $\bar{x}$.

For a discrete variable, the expectation is the sum of all values of x weighted by their probability, given by the PDF $f(x)$.

$$E(X) = \sum_x x \times f(x)$$

Because continuous variables take on an infinite number of values, we compute the expectation with an integral

$$\int x \times f(x) dx$$

Variance and standard deviation of a random variable

Recall that for a sample, the standard deviation *sd* is

$$sd = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

And the sample variance is sd^2

Variance and standard deviation of a random variable

Recall that for a sample, the standard deviation *sd* is

$$sd = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

And the sample variance is sd^2

For a random variable X, the variance is defined via the expectation instead of sample mean

$$V(X) = E[\{X - E(X)\}^2]$$

Variance and standard deviation of a random variable

Recall that for a sample, the standard deviation *sd* is

$$sd = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

And the sample variance is sd^2

For a random variable X, the variance is defined via the expectation instead of sample mean

$$V(X) = E[\{X - E(X)\}^2]$$

Note the similarities in the two equations