

## Probability, 2

---

Frank Edwards

- Random variables assign values to events

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

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

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

- Random variables assign values to events
- Each value is mutually exclusive
- The set of all values is exhaustive (the sample space  $\Omega$ )
- 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  $[\text{pass}, \text{fail}]$  or  $[\text{plaid}, \text{stripes}]$ .



## 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  $X = 1$

$$P(X = 1) = p$$

$$X \sim \text{Bernoulli}(p)$$

$$X \sim \text{Bernoulli}(p)$$

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

$$X \sim \text{Bernoulli}(p)$$

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

```
sample_of_flips <- rbinom(5, 1, 0.5)
pander(table(sample_of_flips))
```

0	1
2	3



## Let's flip some coins

```
sample_of_flips <- rbinom(5, 1, 0.5)
pander(table(sample_of_flips))
```

0	1
2	3

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  $x$ .

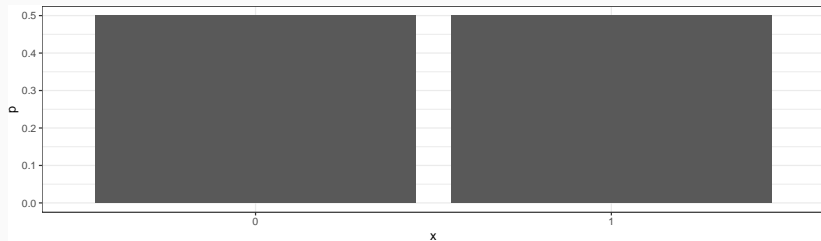
$$PMF(x) = P(X = x)$$

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

## The probability mass function for our coin flip

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

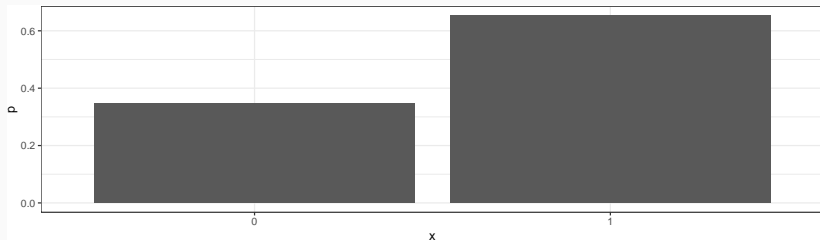
$$PMF(X = 0) = 1 - p = 0.5$$



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

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

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



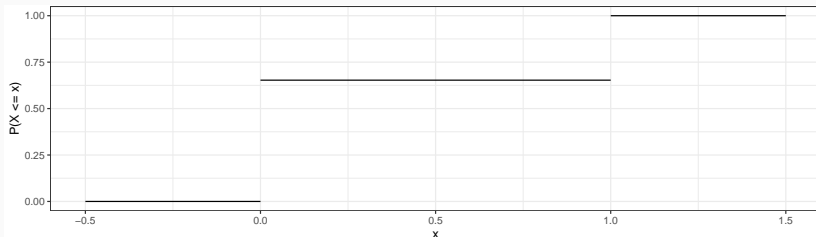
## Describing a probability distribution: cumulative probability

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  $x$

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

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



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

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

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

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

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



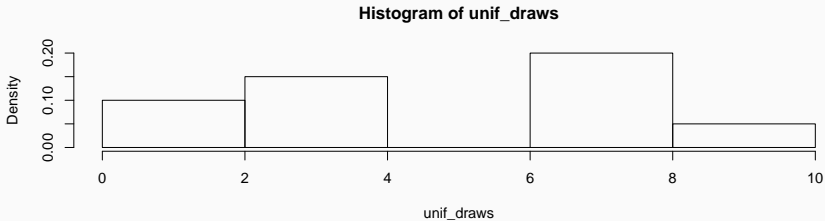
# Uniform random variables

Let's simulate it! 10 draws

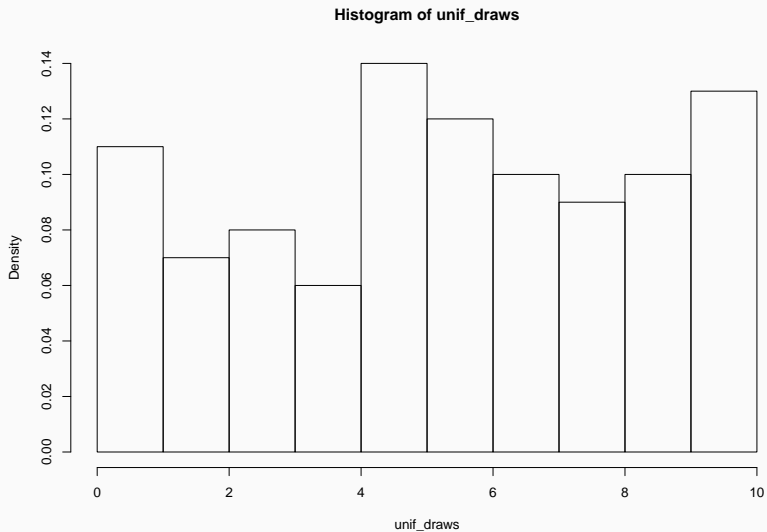
```
unif_draws <- runif(10, min = 0, max = 10)  
unif_draws
```

```
## [1] 8.92301362 1.99249196 2.20183580 7.86728455 7.27346565 6.58874232  
## [7] 3.57542959 6.68001046 3.38661812 0.03063233
```

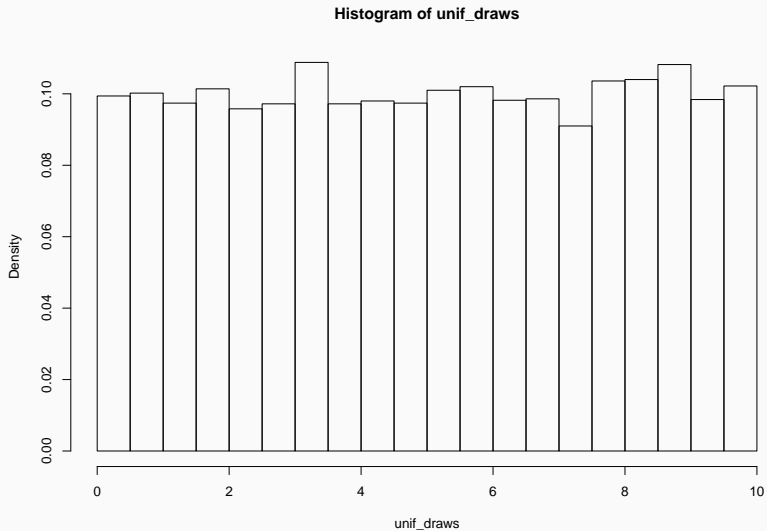
```
hist(unif_draws, freq = F)
```



## 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

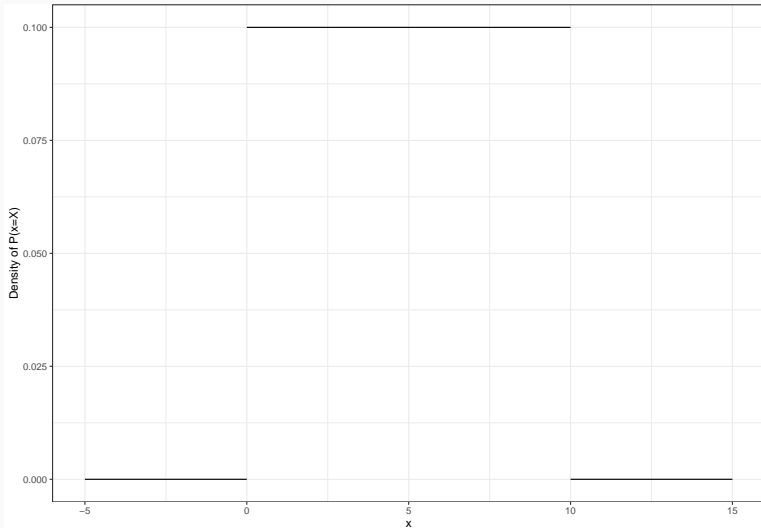
$$\frac{1}{b - a}$$

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

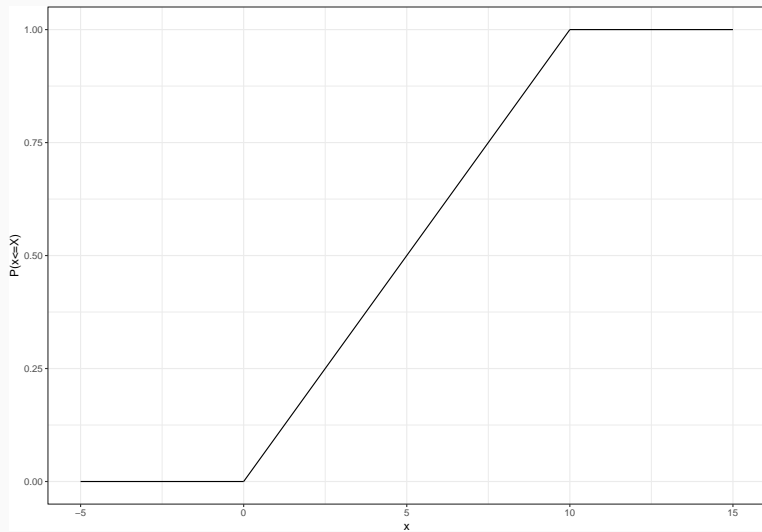
$$\text{PDF: } \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

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

## Probability Density function for $X \sim \text{Uniform}(0, 10)$



## Cumulative Distribution Function for $X \sim \text{Uniform}(0, 10)$



## 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

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

# The binomial distribution

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.*

# 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.*

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] 1 1 1 1 0
```

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

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

Each of these trials is a sample from  $X \sim \text{Binomial}(n, p)$  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] 1 1 1 1 0
```

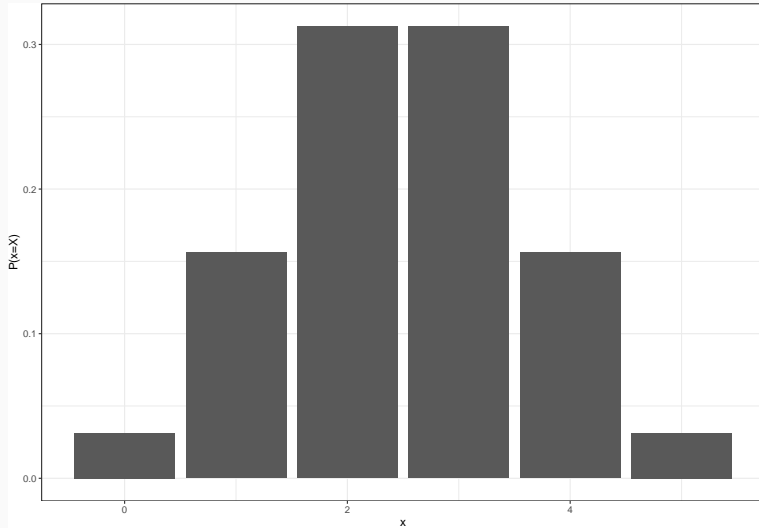
```
## [1] 1 0 1 1 0
```

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

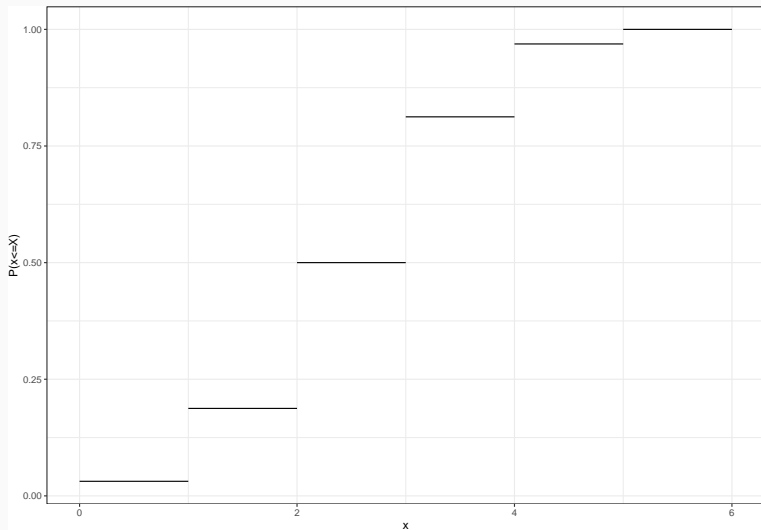
Each of these trials is a sample from  $X \sim \text{Binomial}(n, p)$  where  $n = 5$  and  $p = 0.5$

What is  $x$  for each trial?

## Probability Distribution Function for $X \sim \text{Binomial}(5, 0.5)$



## Cumulative Distribution Function for $X \sim \text{Binomial}(5, 0.5)$

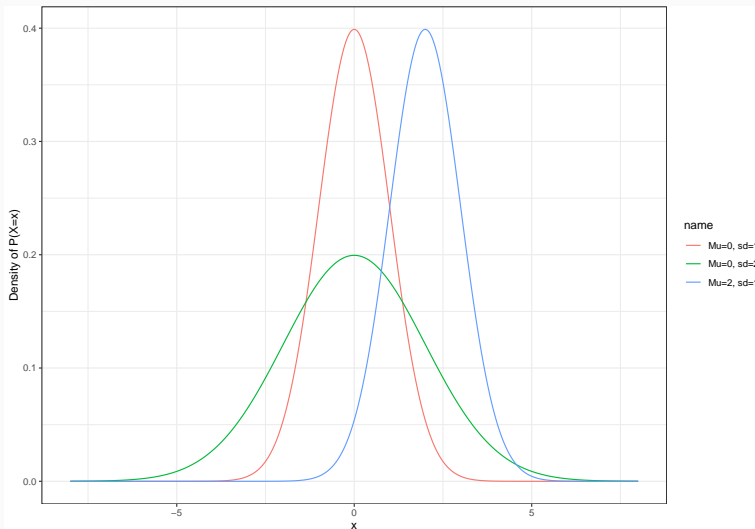


# The Normal Distribution

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

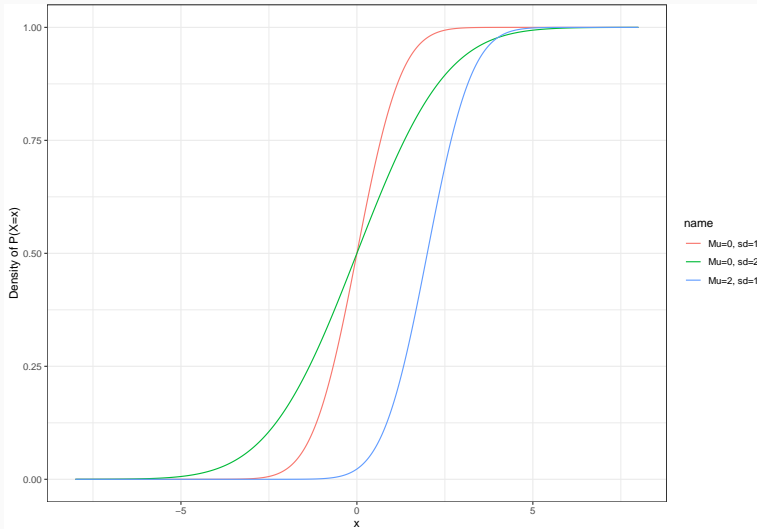
- $\mu$  determines the location of the distribution
- $\sigma$  determines the spread of the distribution

# The Normal PDF





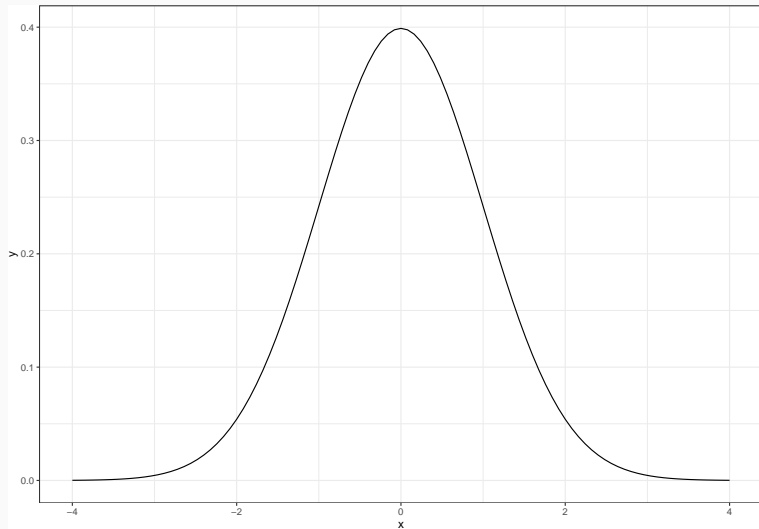
# The Normal CDF



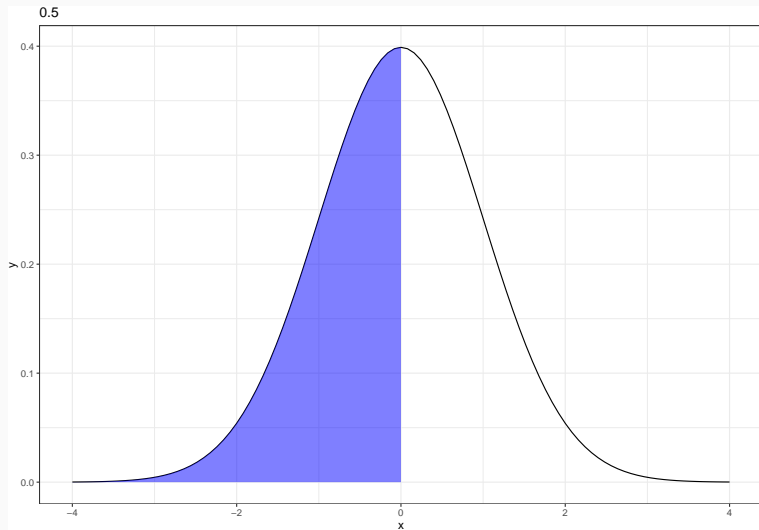
## Special features of Normal distributions:

- 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  $N(0, 1)$

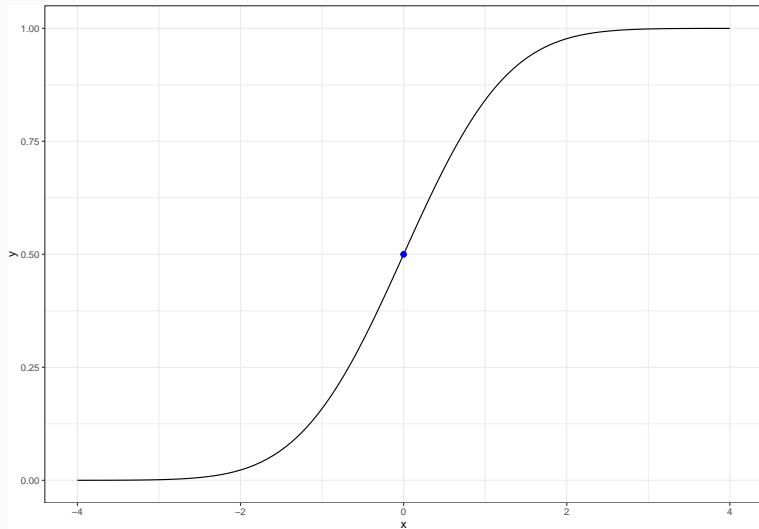
## Area under the curve: interpreting the PDF and CDF



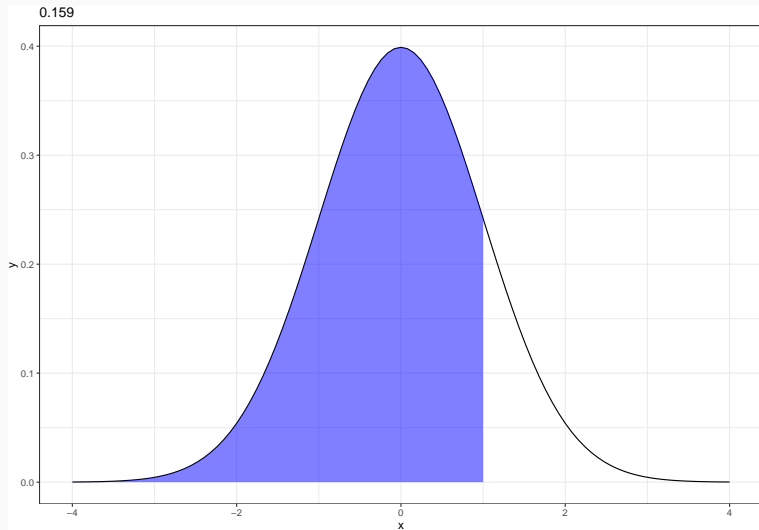
## Area under the curve: interpreting the PDF and CDF



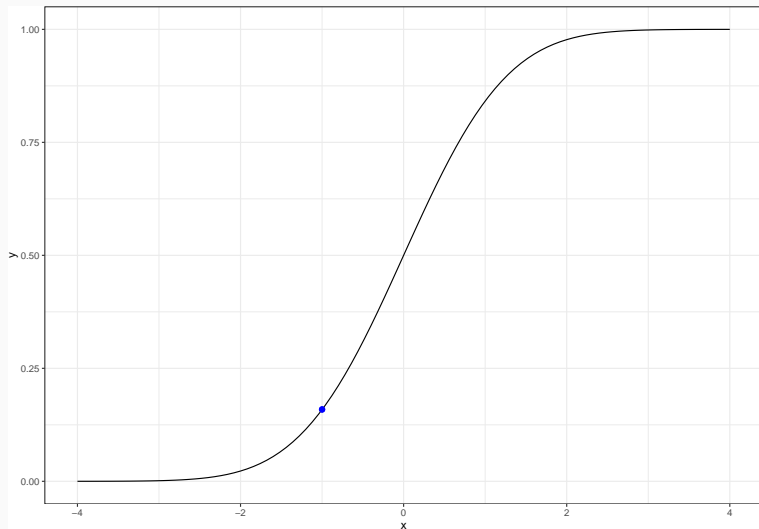
## Area under the curve: interpreting the PDF and CDF



## Area under the curve: interpreting the PDF and CDF



## Area under the curve: interpreting the PDF and CDF



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?

## 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  $z \sim N(0, 1)$

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

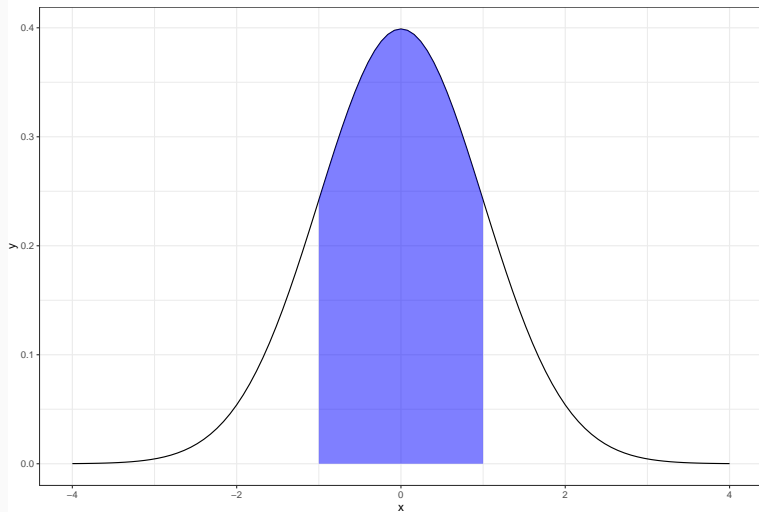
Recall that to obtain a z-score, we subtract the mean and divide by the standard deviation:

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

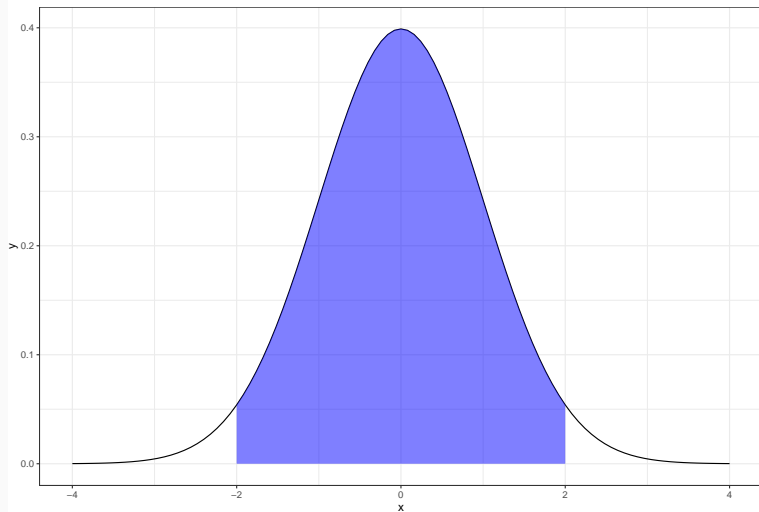
For a Normal variable, z-scores are distributed  $z \sim N(0, 1)$

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

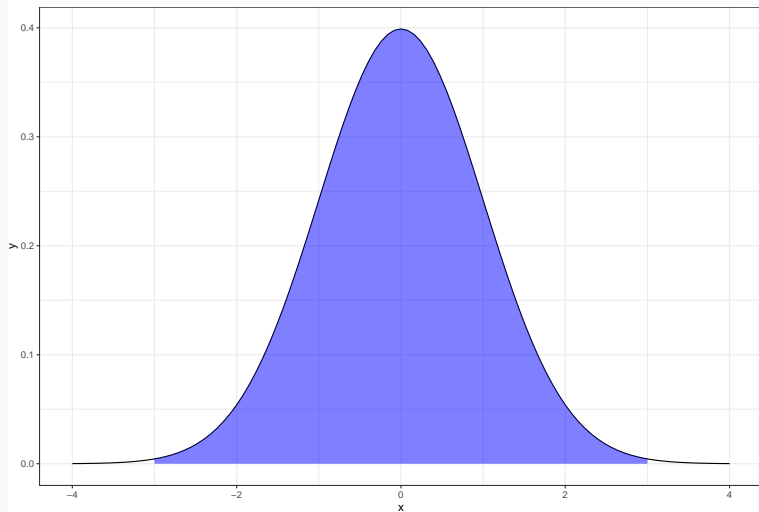
Mean  $\pm$  1 SD = 0.683



Mean  $\pm$  2 SD = 0.954



Mean  $\pm$  3 SD = 0.997



# Useful probability distribution functions

```
### Normal(0,1) probability density function
```

```
dnorm(x = 0, mean = 0, sd = 1)
```

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

```
### Normal(0,1) cumulative distribution function
```

```
pnorm(q = 0, mean = 0, sd = 1)
```

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

```
### Random draw from a normal(0,1) distribution
```

```
rnorm(n = 1, mean = 0, sd = 1)
```

```
## [1] -0.9548283
```

```
### CDF position for a given probability (quantile)
```

```
qnorm(p = 0.75, mean = 0, sd = 1)
```

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

```
### 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  $sd$

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

And the sample variance is  $sd^2$

## Variance and standard deviation of a random variable

Recall that for a sample, the standard deviation  $sd$

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

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  $sd$

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

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

## Large sample (asymptotic) theorems

---

## The law of large numbers

As a sample of draws from a random variable increases, the sample mean converges to the population mean  $E(X)$

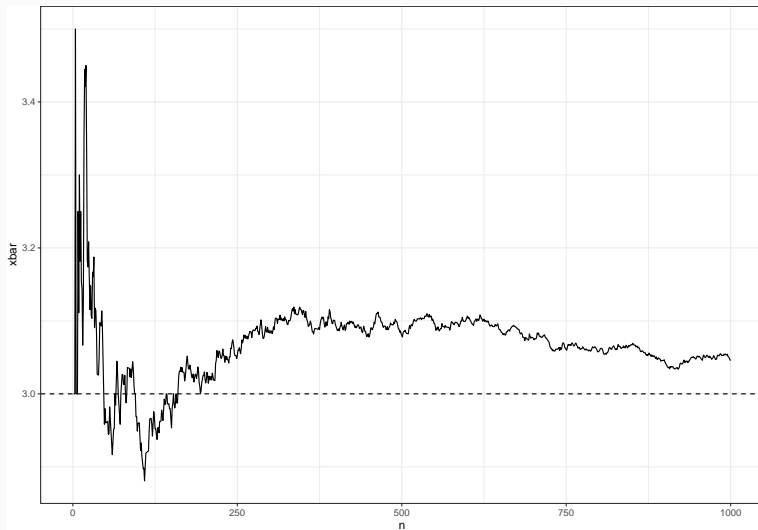
$$\bar{x}_n \rightarrow E(X)$$

# Monte Carlo simulation for the mean of a binomial variable

To test the law of large numbers, let's draw from a binomial variable with varying sample sizes.

We expect that  $\bar{x}$  will converge to  $E(X)$  as the sample size  $n$  increases

```
## MC simulation, 1000 reps
sims <- 1000
## Take 1000 draws from binomial(0.3, 10)
x <- rbinom(sims, p = 0.3, size = 10)
### output df
out <- data.frame(n = 1:sims, xbar = NA)
for (i in 1:sims) {
  out$xbar[i] <- sum(x[1:i])/i
}
## or use xbar<-cumsum(x)/1:sims
```





- As  $n$  increases, the distribution of the sample mean  $\bar{x}$  approaches a Normal distribution.

# The Central Limit Theorem

- As  $n$  increases, the distribution of the sample mean  $\bar{x}$  approaches a Normal distribution.
- This relationship holds for many distributions (Bernoulli, Binomial, Normal, others we'll discuss later)

# The Central Limit Theorem

- As  $n$  increases, the distribution of the sample mean  $\bar{x}$  approaches a Normal distribution.
- This relationship holds for many distributions (Bernoulli, Binomial, Normal, others we'll discuss later)
- If our samples are independent, and each observation within the sample is iid, the distribution of z-scores of sample means converges to a  $Normal(0, 1)$  distribution

# The Central Limit Theorem

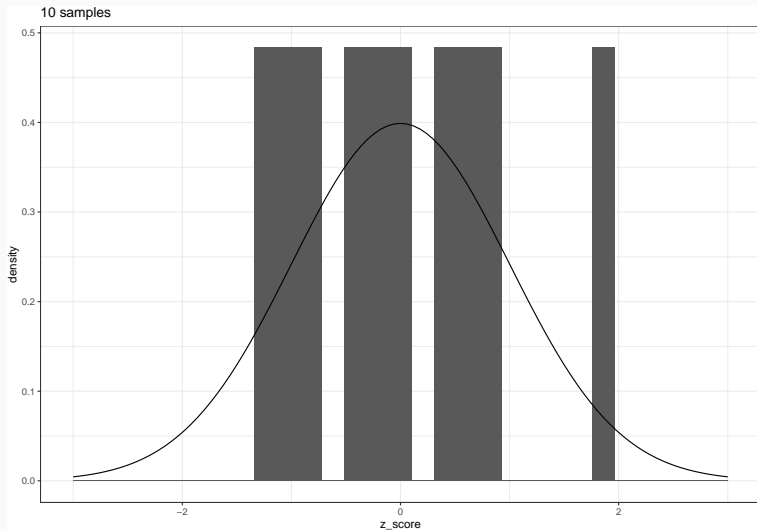
- As  $n$  increases, the distribution of the sample mean  $\bar{x}$  approaches a Normal distribution.
- This relationship holds for many distributions (Bernoulli, Binomial, Normal, others we'll discuss later)
- If our samples are independent, and each observation within the sample is iid, the distribution of z-scores of sample means converges to a *Normal*(0, 1) distribution
- The Central Limit Theorem allows us to make statements about uncertainty when we haven't observed the population mean or variance

# Monte Carlo simulations of a binomial variable $p=0.7$ , $n=10$

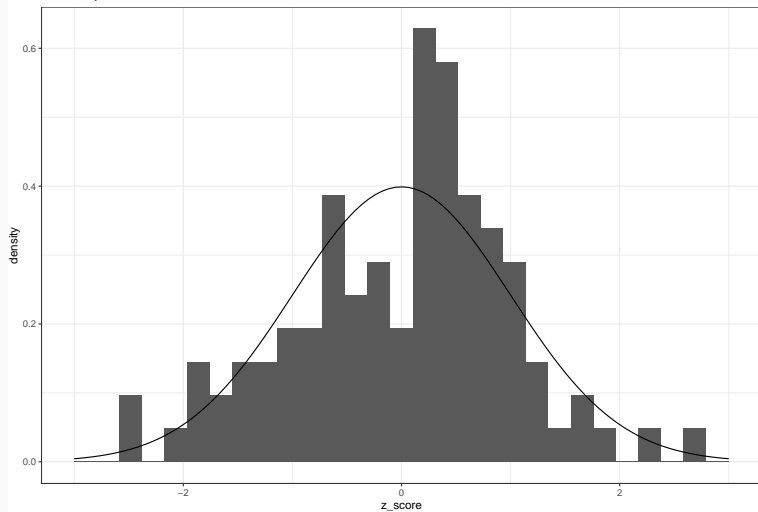
```
### Binomial random variable, 10 observations, probability of success = 0.7
###  $E(X) = np = 0.7 * 10$  pretend this is classes with size 10, probability of
### passing = 0.7 sample size = 10 classes of 10 students, size = 100 classes,
### size = 1000 classes simulate each sample size for 1000 replications
sims <- 1000
n <- 10
p <- 0.7
xbar_10 <- rep(NA, 10)
xbar_100 <- rep(NA, 100)
xbar_1000 <- rep(NA, 1000)
for (i in 1:10) {
  x_10 <- rbinom(10, p = p, size = n)
  xbar_10[i] <- (mean(x_10))
}

for (i in 1:100) {
  x_100 <- rbinom(100, p = p, size = n)
  xbar_100[i] <- (mean(x_100))
}

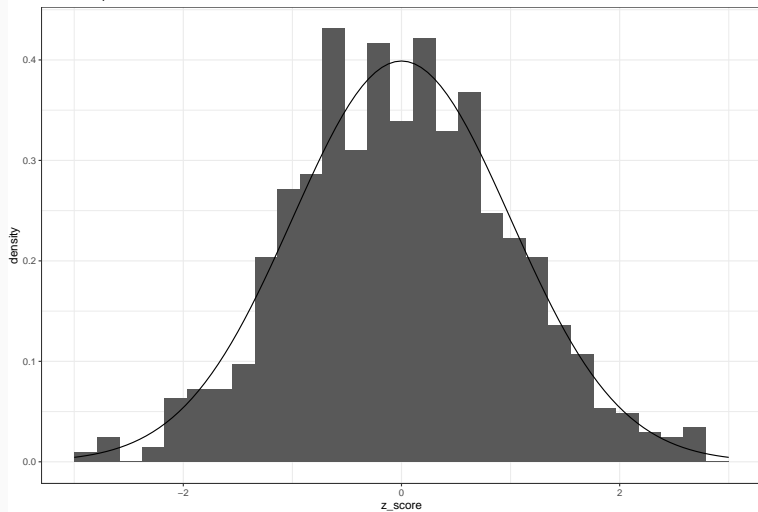
for (i in 1:1000) {
  x_1000 <- rbinom(1000, p = p, size = n)
  xbar_1000[i] <- (mean(x_1000))
}
```



100 samples



1000 samples





- Struggling with probability? Try Khan Academy probability and distributions.
- Homework (late work): Any late assignments (including this week's HW) are due without penalty by Friday. After that, I'm going to start deducting points for late work.
- Homework: Question 6.6.3. Due 12/4. Start early on this to give yourself time to ask questions
- Lab: Obama vote share example, and some hints for the homework