THE UNIVERSITY OF TEXAS AT AUSTIN
McCOMBS SCHOOL OF BUSINESS

Probability review 1

Lecture 3

STA 371G

Announcements

- Homework 1 due Sunday night at 11:59 PM

Announcements

- Homework 1 due Sunday night at 11:59 PM
- First (optional) R help session will be this Friday from 11 AM to 12 PM in GSB 3.138 (note temporary room change)

Announcements

- Homework 1 due Sunday night at 11:59 PM
- First (optional) R help session will be this Friday from 11 AM to 12 PM in GSB 3.138 (note temporary room change)
- About grading

Announcements

- Homework 1 due Sunday night at 11:59 PM
- First (optional) R help session will be this Friday from 11 AM to 12 PM in GSB 3.138 (note temporary room change)
- About grading
    - QUEST: two tries on each question for full credit, then decreasing credit for each subsequent try

Announcements

- Homework 1 due Sunday night at 11:59 PM
- First (optional) R help session will be this Friday from 11 AM to 12 PM in GSB 3.138 (note temporary room change)
- About grading
  - QUEST: two tries on each question for full credit, then decreasing credit for each subsequent try
  - Learning Catalytics: answer 75% of in-class questions for full participation credit

What are the chances that two people *in this room* have the same birthday (month/day, not year)?

Probability Theory

The Concept of Probability

What do each of these have in common?

Probability Theory

The Concept of Probability

What do each of these have in common?

- Outcome of rolling a die

Probability Theory

The Concept of Probability

What do each of these have in common?

- Outcome of rolling a die
- S&P500 index at the and of January

Probability Theory

The Concept of Probability

What do each of these have in common?

- Outcome of rolling a die
- S&P500 index at the and of January
- Number of iPhone 7s to be sold over the next year

Probability Theory

The Concept of Probability

What do each of these have in common?

- Outcome of rolling a die
- S&P500 index at the and of January
- Number of iPhone 7s to be sold over the next year
- Number of unique visitors to Amazon.com over the next week

Probability Theory

The Concept of Probability

What do each of these have in common?

- Outcome of rolling a die
- S&P500 index at the and of January
- Number of iPhone 7s to be sold over the next year
- Number of unique visitors to Amazon.com over the next week
- Lifetime of your MacBook Air

Probability Theory

The Concept of Probability

What do each of these have in common?

- Outcome of rolling a die
- S&P500 index at the and of January
- Number of iPhone 7s to be sold over the next year
- Number of unique visitors to Amazon.com over the next week
- Lifetime of your MacBook Air

Probability Theory

The Concept of Probability

What do each of these have in common?

- Outcome of rolling a die
- S&P500 index at the and of January
- Number of iPhone 7s to be sold over the next year
- Number of unique visitors to Amazon.com over the next week
- Lifetime of your MacBook Air

We cannot predict any of these with certainty.

Probability Theory

The Concept of Probability

What do each of these have in common?

- Outcome of rolling a die
- S&P500 index at the and of January
- Number of iPhone 7s to be sold over the next year
- Number of unique visitors to Amazon.com over the next week
- Lifetime of your MacBook Air

We cannot predict any of these with certainty. But we can model them using probability theory and learn a lot about how these variables will behave.

Probability Theory

Definitions

Definition

A random variable expresses the outcome of a random process as a number. It is denoted by an uppercase letter.

Probability Theory

Definitions

Definition

A random variable expresses the outcome of a random process as a number. It is denoted by an uppercase letter.

- Random experiment $\longrightarrow$ Selecting a student at random from the class

Probability Theory

Definitions

Definition

A random variable expresses the outcome of a random process as a number. It is denoted by an uppercase letter.

- Random experiment $\longrightarrow$ Selecting a student at random from the class
- Random variable $\longrightarrow X$: The day of their birthday $(1, 2, \ldots, 31)$

Probability Theory
Definitions

Definition

A random variable expresses the outcome of a random process as a number. It is denoted by an uppercase letter.

- Random experiment $\longrightarrow$ Selecting a student at random from the class
- Random variable $\longrightarrow X$: The day of their birthday $(1, 2, \ldots, 31)$
- Random variable $\longrightarrow Y$: Their height, in inches

Probability Theory

Definitions

Definition

A discrete random variable is a random variable with a finite (or countably infinite) range.

Probability Theory

Definitions

Definition

A discrete random variable is a random variable with a finite (or countably infinite) range.

A continuous random variable is a random variable with an interval (either finite or infinite) of real numbers for its range.

Probability Theory

Definitions

Definition

A discrete random variable is a random variable with a finite (or countably infinite) range.

A continuous random variable is a random variable with an interval (either finite or infinite) of real numbers for its range.

- X : Number of stocks on NYSE whose price change today (discrete)

Probability Theory

Definitions

Definition

A discrete random variable is a random variable with a finite (or countably infinite) range.
A continuous random variable is a random variable with an interval (either finite or infinite) of real numbers for its range.

- X : Number of stocks on NYSE whose price change today (discrete)
- Y : Average price change of the stocks on NYSE (continuous)

Probability Theory

Definitions

Definition

A discrete random variable is a random variable with a finite (or countably infinite) range.

A continuous random variable is a random variable with an interval (either finite or infinite) of real numbers for its range.

- X : Number of stocks on NYSE whose price change today (discrete)
- Y : Average price change of the stocks on NYSE (continuous)
- Z : Number of people that lose money in the stock market today (discrete!)

Probability Theory

Exercise

Discrete or continuous?

Probability Theory

Exercise

Discrete or continuous?

- Number of iPhone 7s to be sold over the next year

Probability Theory

Exercise

Discrete or continuous?

- Number of iPhone 7s to be sold over the next year

Probability Theory

Exercise

Discrete or continuous?

- Number of iPhone 7s to be sold over the next year (discrete)
- Lifetime of your MacBook Air, in years

Probability Theory

Exercise

Discrete or continuous?

- Number of iPhone 7s to be sold over the next year (discrete)
- Lifetime of your MacBook Air, in years

Probability Theory

Exercise

Discrete or continuous?

- Number of iPhone 7s to be sold over the next year (discrete)
- Lifetime of your MacBook Air, in years (continuous)
- Number of unique visitors to Amazon.com over the next week

Probability Theory

Exercise

Discrete or continuous?

- Number of iPhone 7s to be sold over the next year (discrete)
- Lifetime of your MacBook Air, in years (continuous)
- Number of unique visitors to Amazon.com over the next week

Probability Theory

Exercise

Discrete or continuous?

- Number of iPhone 7s to be sold over the next year (discrete)
- Lifetime of your MacBook Air, in years (continuous)
- Number of unique visitors to Amazon.com over the next week (discrete)
- S&P 500 index at the and of January

Probability Theory

Exercise

Discrete or continuous?

- Number of iPhone 7s to be sold over the next year (discrete)
- Lifetime of your MacBook Air, in years (continuous)
- Number of unique visitors to Amazon.com over the next week (discrete)
- S&P 500 index at the and of January

Probability Theory

Exercise

Discrete or continuous?

- Number of iPhone 7s to be sold over the next year (discrete)
- Lifetime of your MacBook Air, in years (continuous)
- Number of unique visitors to Amazon.com over the next week (discrete)
- S&P 500 index at the and of January (continuous)

Probability Theory

Definitions

Definition

Probability is the measure of the likelihood that a particular outcome (or set of outcomes) will be observed. Probability is a number always between 0 and 1.

Probability Theory

Definitions

Definition

Probability is the measure of the likelihood that a particular outcome (or set of outcomes) will be observed. Probability is a number always between 0 and 1.

- X : The month number of the student's birthday, $P(X = 3) = 1/12$

Probability Theory

Definitions

Definition

Probability is the measure of the likelihood that a particular outcome (or set of outcomes) will be observed. Probability is a number always between 0 and 1.

- X : The month number of the student's birthday, $P(X = 3) = 1/12$
- Y : Lifetime of your MacBook, $P(Y > 5 \text{ years}) = 0.05$

Probability Distributions

Now that we have defined a random variable, how do we know what the probabilities are?

Probability Distributions

Now that we have defined a random variable, how do we know what the probabilities are?

For example, what is the probability that your MacBook will break down after 5 years but before 7 years? That is, $P(5 < X < 7) =?$

Probability Distributions

Now that we have defined a random variable, how do we know what the probabilities are?

For example, what is the probability that your MacBook will break down after 5 years but before 7 years? That is, $P(5 < X < 7) = ?$

Definition

The probability distribution of a random variable X is a description of the probabilities associated with the possible values of X.

Probability Distributions

Now that we have defined a random variable, how do we know what the probabilities are?

For example, what is the probability that your MacBook will break down after 5 years but before 7 years? That is, $P(5 < X < 7) = ?$

Definition

The probability distribution of a random variable X is a description of the probabilities associated with the possible values of X.
Discrete random variable $\longrightarrow$ Probability Mass Function (p.m.f.)

Probability Distributions

Now that we have defined a random variable, how do we know what the probabilities are?

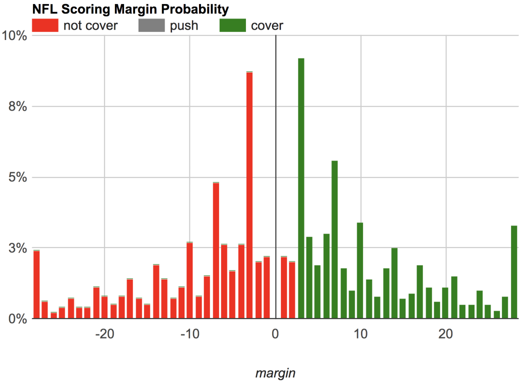
For example, what is the probability that your MacBook will break down after 5 years but before 7 years? That is, $P(5 < X < 7) = ?$

Definition

The probability distribution of a random variable X is a description of the probabilities associated with the possible values of X.
Discrete random variable $\longrightarrow$ Probability Mass Function (p.m.f.)
Continuous random variable $\longrightarrow$ Probability Density Function (p.d.f.)

Discrete Random Variables

Michael Beuoy's model for NFL games gives the following probabilities for the Super Bowl (positive margin = Patriots win):



Discrete Random Variables

Let M be a random variable representing the Patriots' margin of victory.

Patriots' margin of victory m	Probability $P(M = m)$
-28	0.024
-27	0.006
-26	0.002
-25	0.004
$\vdots$	$\vdots$
27	0.008
28	0.033

Discrete Random Variables

Definition

The expected value and variance of a random variable X are the long run mean and variance if we draw values repeatedly from X.

Discrete Random Variables

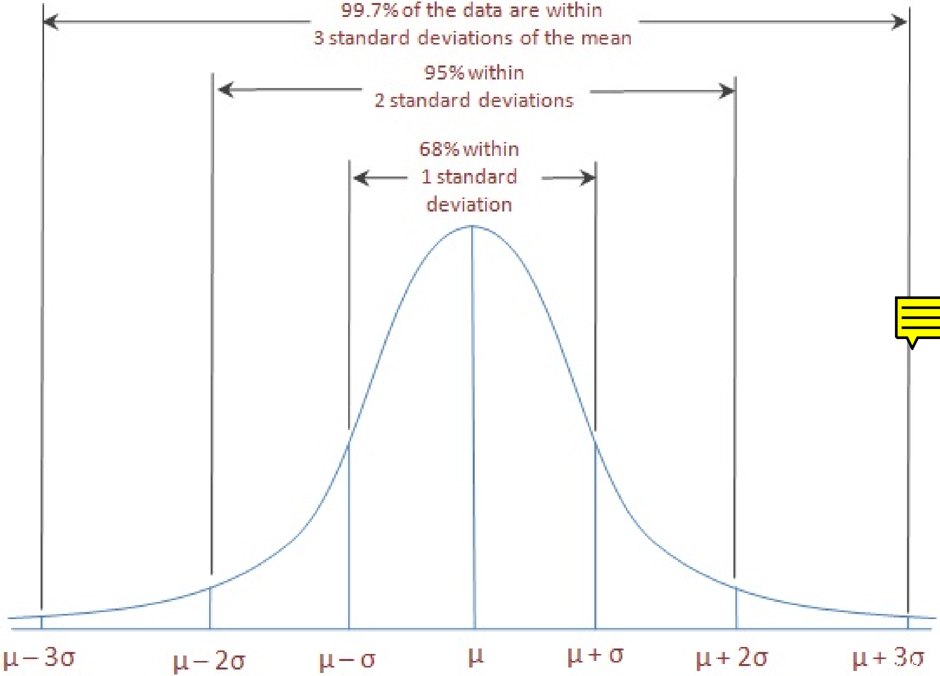
Definition

The expected value and variance of a random variable X are the long run mean and variance if we draw values repeatedly from X.

$$\begin{aligned}
E(M) &= \sum_{\text{all margins } m} m \cdot P(M = m) \\
&= (-28)(0.024) + (-27)(0.006) + \cdots \\
\mathrm{Var}(M) &= \sum_{\text{all margins } m} (m - E(M))^2 \cdot P(M = m) \\
&= (-28 - E(M))^2(0.024) + (-27 - E(M))^2(0.006) + \cdots
\end{aligned}$$

Normal Distributions (the "bell curve")

The normal distribution is a common continuous probability distribution; we will denote a normal distribution with mean μ and variance σ^2 by $N(\mu, \sigma^2)$.



99.7% of the data are within
3 standard deviations of the mean
95% within
2 standard deviations
68% within
1 standard
deviation
$\mu - 3\sigma$
$\mu - 2\sigma$
$\mu - \sigma$
μ
$\mu + \sigma$
$\mu + 2\sigma$
$\mu + 3\sigma$

Continous Random Variables

Let V be a random variable indicating the SAT Verbal score of a randomly selected student in the US.

$$P(\text{SAT} > 600) =$$

Continous Random Variables

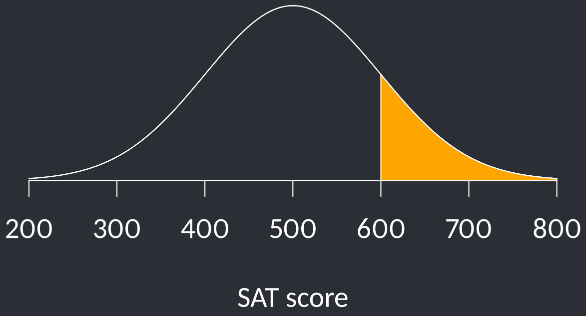
Let V be a random variable indicating the SAT Verbal score of a randomly selected student in the US.

$$P(\text{SAT} > 600) = \int_{600}^{800} P(V = v)\, dv$$

Continous Random Variables

Let V be a random variable indicating the SAT Verbal score of a randomly selected student in the US.

$$P(\text{SAT} > 600) = \int_{600}^{800} P(V = v)\, dv$$



Continuous Random Variables

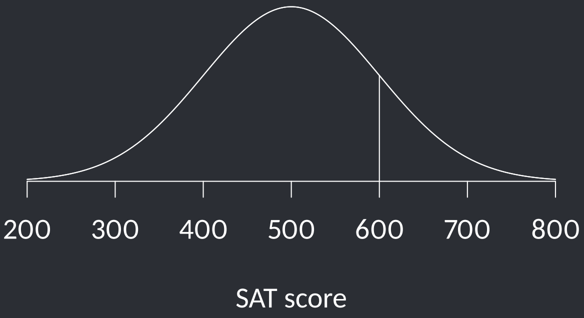
Let V be a random variable indicating the SAT Verbal score of a randomly selected student in the US.

$$P(\text{SAT} = 600) =$$

Continuous Random Variables

Let V be a random variable indicating the SAT Verbal score of a randomly selected student in the US.

$$P(\text{SAT} = 600) = 0$$



Continous Random Variables

The formulas for expected value and variance are the same for continuous random variables, but they involve integrals instead of sums:

Discrete random variable X	Continuous random variable Y with density function $f(y)$
$E(X) = \sum_x xP(X = x)$	$E(Y) = \int_y yf(y)\,dy$
$\mathrm{Var}(X) = \sum_x (x - E(X))^2 P(X = x)$	$\mathrm{Var}(Y) = \int_y (y - E(Y))^2 f(y)$

R Simulation of expected value

R Simulation

Let’s simulate rolling dice and see that the expected value really does represent the long-run average!

```
# Simulate rolling a die 1 time
sample(c(1, 2, 3, 4, 5, 6), 1, replace=T)
# Simulate rolling a die 4 times
sample(c(1, 2, 3, 4, 5, 6), 4, replace=T)
# Take the average
mean(sample(c(1, 2, 3, 4, 5, 6), 4, replace=T))
# Let's increase the number of dice
mean(sample(c(1, 2, 3, 4, 5, 6), 100, replace=T))
```

Joint probability – back to birthdays

If two events are *independent* (i.e., knowing the outcome of one tells you nothing about the outcome of the other), then the probability of two events happening at the same time is

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

Joint probability – back to birthdays

If two events are *independent* (i.e., knowing the outcome of one tells you nothing about the outcome of the other), then the probability of two events happening at the same time is

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

Let's figure out the probability of no one sharing a birthday. Let's call B_1 the first person's birthday, B_2 the second person's birthday, etc.

Joint probability – back to birthdays

If two events are *independent* (i.e., knowing the outcome of one tells you nothing about the outcome of the other), then the probability of two events happening at the same time is

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

Let's figure out the probability of no one sharing a birthday. Let's call B_1 the first person's birthday, B_2 the second person's birthday, etc.

$$\begin{aligned} P(\text{none shared}) &= P(B_2 \neq B_1) \cdot P(B_3 \neq B_1, B_2) \cdot P(B_4 \neq B_1, B_2, B_3) \cdots \\ &= \frac{364}{365} \cdot \frac{363}{365} \cdot \frac{362}{365} \cdots \end{aligned}$$

Joint probability – back to birthdays

If two events are *independent* (i.e., knowing the outcome of one tells you nothing about the outcome of the other), then the probability of two events happening at the same time is

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

Let's figure out the probability of no one sharing a birthday. Let's call B_1 the first person's birthday, B_2 the second person's birthday, etc.

$$\begin{aligned} P(\text{none shared}) &= P(B_2 \neq B_1) \cdot P(B_3 \neq B_1, B_2) \cdot P(B_4 \neq B_1, B_2, B_3) \cdots \\ &= \frac{364}{365} \cdot \frac{363}{365} \cdot \frac{362}{365} \cdots \end{aligned}$$

When $n = 70$ the probability of no shared birthdays is just 0.01%!