Selection: 1

| Please choose a lesson, or type 0 to return to course menu.

1: Introduction 2: Probability1 3: Probability2

4: ConditionalProbability 5: Expectations 6: Variance

7: CommonDistros 8: Asymptotics 9: T Confidence Intervals

10: Hypothesis Testing 11: P Values 12: Power

13: Multiple Testing 14: Resampling

Selection: 3

| | | 0%

| Probability. (Slides for this and other Data Science courses may be found at github

| https://github.com/DataScienceSpecialization/courses. If you care to use them, they

| must be downloaded as a zip file and viewed locally. This lesson corresponds to

| Statistical\_Inference/Probability.)

...

| |== | 3%

| In this lesson, we'll continue to discuss probability.

...

| |==== | 6%

| Recall that a random variable is a numerical outcome of an experiment. It can be

| discrete (take on only a countable number of possibilities), or continuous (take on

| any value on the real line or subset of it).

...

| |======= | 9%

| If you had a ruler of infinite precision, would measuring the height of adults around

| the world be continuous or discrete?

1: continuous

2: discrete

Selection: 1

| Excellent work!

| |========= | 11%

| Is the drawing of a hand of cards continuous or discrete?

1: discrete

2: continuous

Selection: 1

| You got it right!

| |=========== | 14%

| Continuous random variables are usually associated with measurements of time,

| distance, or some biological process since they can take on any value, often within

| some specified range. Limitations of precision in taking the measurements may imply

| that the values are discrete; we in fact consider them continuous.

...

| |============= | 17%

| A probability mass function (PMF) gives the probability that a discrete random

| variable is exactly equal to some value.

...

| |================ | 20%

| For instance, suppose we have a coin which may or may not be fair. Let x=0 represent

| a 'heads' outcome and x=1 represent a 'tails' outcome of a coin toss. If p is the

| probability of 'heads' which of the following represents the PMF of the coin toss?

| The variable x is either 0 (heads) or 1 (tails).

1: (p^x)\*(1-p)^(1-x)

2: (p^(1-x))\*(1-p)^x

Selection: 1

| Not quite, but you're learning! Try again.

| The probability p is associated with a 'heads' outcome which occurs when x=0. Which

| of the two expressions has an exponent of 1 for p when x is 0?

1: (p^x)\*(1-p)^(1-x)

2: (p^(1-x))\*(1-p)^x

Selection: 2

| Perseverance, that's the answer.

| |================== | 23%

| A probability density function is associated with a continuous random variable. To

| quote from Wikipedia, it "is a function that describes the relative likelihood for

| this random variable to take on a given value. The probability of the random variable

| falling within a particular range of values is given by ... the area under the

| density function but above the horizontal axis and between the lowest and greatest

| values of the range."

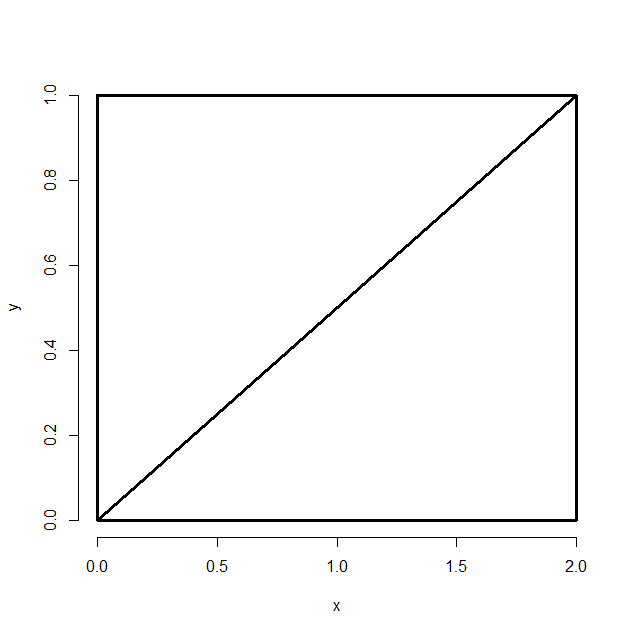
...

| |==================== | 26%

| We'll repeat two requirements of a probability density function. It must be

| nonnegative everywhere, and the area under it must equal one."

...



| |====================== | 29%

| Consider this figure - a rectangle with height 1 and width 2 with a diagonal line

| drawn from the lower left corner (0,0) to the upper right (2,1). The area of the

| entire rectangle is 2 and elementary geometry tells us that the diagonal divides the

| rectangle into 2 equal areas.

...

| |========================= | 31%

| Could the diagonal line represent a probability density function for a random

| variable with a range of values between 0 and 2? Assume the lower side of the

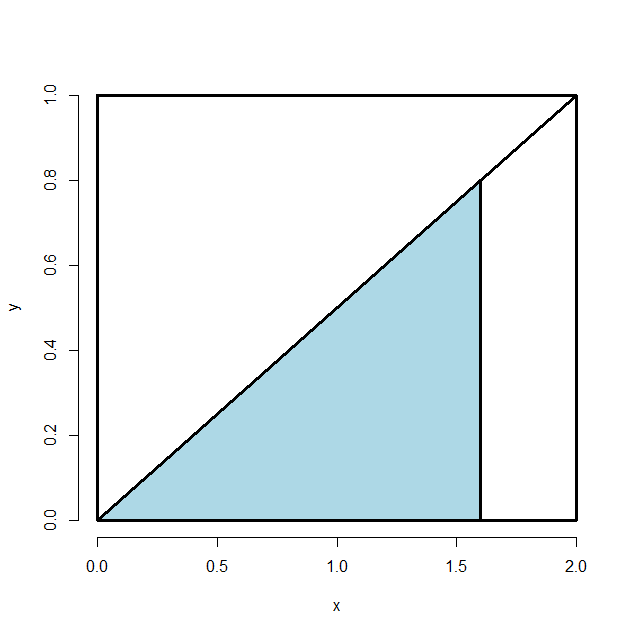
| rectangle is the x axis.

1: No

2: Yes

Selection: 2

| Keep up the great work!



| |=========================== | 34%

| Now consider the shaded portion of the triangle - a smaller triangle with a base of

| length 1.6 and height determined by the diagonal. We'll answer the question, "What

| proportion of the big triangle is shaded?"

...

| |============================= | 37%

| This proportion represents the probability that throwing a piece of cat kibble at the

| bigger triangle (below the diagonal) hits the blue portion.

...

| |=============================== | 40%

| We have to compute the area of the blue triangle. (Since the area of the big triangle

| is 1, the area of the blue triangle is the proportion of the big triangle that is

| shaded.) We know the base, but what is its height?

1: .25

2: .5

3: .8

4: I can't tell

Selection: 3

| You nailed it! Good job!

| |================================= | 43%

| What is the area of the blue triangle?

> (1.6\*.8)/2

[1] 0.64

| Excellent work!

| |==================================== | 46%

| So, what is the probability that the kibble we throw at the bigger triangle will hit

| the blue portion?

> .64

[1] 0.64

| You nailed it! Good job!

| |====================================== | 49%

| This artificial example was to meant to illustrate a simple probability density

| function (PDF). Most PDF's have underlying formulae more complicated than lines.

| We'll see more of these in future lessons.

...

| |======================================== | 51%

| The cumulative distribution function (CDF) of a random variable X, either discrete or

| continuous, is the function F(x) equal to the probability that X is less than or

| equal to x. In the example above, the area of the blue triangle represents the

| probability that the random variable was less than or equal to the value 1.6.

...

| |========================================== | 54%

| In the triangle example from above, which of the following expressions represents

| F(x), the CDF?

1: x\*2x/2

2: x^2

3: x\*x/4

4: x\*x/2

Selection: 3

| You are quite good my friend!

| |============================================= | 57%

| If you're familiar with calculus you might recognize that when you're computing areas

| under curves you're actually integrating functions.

...

| |=============================================== | 60%

| When the random variable is continuous, as in the example, the PDF is the derivative

| of the CDF. So integrating the PDF (the line represented by the diagonal) yields the

| CDF. When you evaluate the CDF at the limits of integration the result is an area.

...

| |================================================= | 63%

| To see this in the example, we've defined the function mypdf for you. This is the

| equation of the line represented by the diagonal of the rectangle. As the PDF, it is

| the derivative of F(x), the CDF. Look at mypdf now.

> mypdf

function(x){x/2}

<environment: 0x00000000063cd0c8>

| You are amazing!

| |=================================================== | 66%

| Now use the R function integrate to integrate mypdf with the parameters lower equal

| to 0 and upper equal to 1.6. See if you get the same area (probability) you got

| before.

> integrate(mypdf, 0, 1.6)

0.64 with absolute error < 7.1e-15

| That's a job well done!

| |===================================================== | 69%

| The survivor function S(x) of a random variable X is defined as the function of x

| equal to the probability that the random variable X is greater than the value x. This

| is the complement of the CDF F(x), in our example, the portion of the lower triangle

| that is not shaded.

...

| |======================================================== | 71%

| In our example, which of the following expressions represents the survival function?

1: 1-x\*2x/2

2: 1-x\*x/4

3: 1-x\*x/2

4: 1-x^2

Selection: 2

| Perseverance, that's the answer.

| |========================================================== | 74%

| The quantile v of a CDF is the point x\_v at which the CDF has the value v. More

| precisely, F(x\_v)=v. A percentile is a quantile in which v is expressed as a

| percentage.

...

| |============================================================ | 77%

| What percentile is the median?

1: I can't tell

2: 95-th

3: 25-th

4: 50-th

Selection: 4

| You're the best!

| |============================================================== | 80%

| What is the 50th percentile of the CDF F(x)=(x^2)/4 from the example above?

> (2^2)/4

[1] 1

| Not quite, but you're learning! Try again. Or, type info() for more options.

| Solve for the x such that x^2=4\*.5=2

> x^2=4\*.5=2

Error in 4 \* 0.5 = 2 :

target of assignment expands to non-language object

> (2^2)/4\*.5

[1] 0.5

| One more time. You can do it! Or, type info() for more options.

| Solve for the x such that x^2=4\*.5=2

> 2

[1] 2

| Nice try, but that's not exactly what I was hoping for. Try again. Or, type info()

| for more options.

| Solve for the x such that x^2=4\*.5=2

> x^2=4\*.5=2

Error in 4 \* 0.5 = 2 :

target of assignment expands to non-language object

> 2

[1] 2

| Give it another try. Or, type info() for more options.

| Solve for the x such that x^2=4\*.5=2

> info()

| When you are at the R prompt (>):

| -- Typing skip() allows you to skip the current question.

| -- Typing play() lets you experiment with R on your own; swirl will ignore what you

| do...

| -- UNTIL you type nxt() which will regain swirl's attention.

| -- Typing bye() causes swirl to exit. Your progress will be saved.

| -- Typing main() returns you to swirl's main menu.

| -- Typing info() displays these options again.

> nxt()

| Resuming lesson...

| What is the 50th percentile of the CDF F(x)=(x^2)/4 from the example above?

> 2

[1] 2

| You almost had it, but not quite. Try again. Or, type info() for more options.

| Solve for the x such that x^2=4\*.5=2

> 1.414214

[1] 1.414214

| Excellent job!

| |================================================================= | 83%

| What does this mean with respect to the kibble we're tossing at the triangle?

1: All of it falls on the vertical line at 1.41

2: All of it falls to the left of 1.41

3: All of it falls to the right of 1.41

4: Half of it falls to the left of 1.41

Selection: 4

| You nailed it! Good job!

| |=================================================================== | 86%

| We'll close by repeating some important points.

...

| |===================================================================== | 89%

| A probability model connects data to a population using assumptions.

...

| |======================================================================= | 91%

| Be careful to distinguish between population medians and sample medians.

...

| |========================================================================== | 94%

| A sample median is an estimator of a population median (the estimand).

...

| |============================================================================ | 97%

| Congrats! You've concluded this lesson on probability.

...

| |==============================================================================| 100%

| Would you like to receive credit for completing this course on Coursera.org?

1: Yes

2: No

Selection: 1

What is your email address? sweeyean@gmail.com

What is your assignment token? ASPiK0IHK5zyJXgM

Grade submission succeeded!

| Perseverance, that's the answer.

| You've reached the end of this lesson! Returning to the main menu...

| Please choose a course, or type 0 to exit swirl.

1: Statistical Inference

2: Take me to the swirl course repository!

Selection: