**Probability distributions for continuous random variables**

Examples of continuous random variables:

* Miles per gallon (MPG) a car gets for its gas consumption
* Amount of precipitation for a city during a year
* Height for people
* GPA
* Price of diamonds
* Salary
* Distance for a field goal in football
* Lifetime before failure of an object
* How long it takes you to get to class everyday

Continuous random variables have an infinite number of values that they can take on within an interval. However, all of these examples are limited by our ability to measure them. For example, the amount of precipitation may only be measured to two decimal places. The distance of a field goal is typically rounded to the nearest yard. While these variables may be measured as discrete random variables, they often can be approximated as continuous. This is especially true if there are a lot of values the random variable can take on.

The possible values that a continuous random variable can take on can be quantified using a probability distribution. Similar to the binomial distribution, these distributions are represented by mathematical functions. However, there are some VERY IMPORTANT DIFFERENCES! These similarities and differences are explained in the next example.

Example: GPA probability distribution (gpa.R)

Let Y be a random variable representing GPAs of students on campus. Suppose the probability distribution for the random variable Y is



for 0 ≤ y ≤ 4 and f(y) = 0 for all other possible values of y. Notice the use here of Y and y again.

Below is a plot of the probability distribution

> curve(expr = 15/2048 \* x^4 \* (4 - x), xlim = c(0,

 4), col = "red", main = "Probability

 distribution for GPAs", ylab = "f(y)", xlab =

 "y")

> abline(h = 0)



Unlike probability distributions for discrete random variables, we cannot simply substitute a value for y into the mathematical function to find a probability. For example, f(3) =  P(Y = 3).

The “area” underneath the curve represents probability. For example, P(2.5 < Y < 3.5) is represented by the area underneath the plotted curve between 2.5 and 3.5 and above.

How do you find this area? Use calculus!!!

 

BUT, not all students have had a calculus course ☹. If you are interested in a calculus-based introduction to statistics, please take course that will often have the name “mathematical statistics” in them.

Fortunately, there are functions in R that can find the areas underneath the curve automatically so that you do not need to know the calculus part.

Questions:

* Suppose the interval was smaller, say 2.9 to 3.1. What will happen to the area underneath the curve?
* What happens if this interval becomes smaller and smaller and … ? What is P(Y = 3) then?

Example: Wind speed in Lincoln (wind\_speed\_distribution.R, Lincoln\_Feb\_wind.csv)

Wind speeds are often characterized by probability distributions! Continuing our past example on wind speeds, below is a probability distribution that often does a good job of accounting for Lincoln’s February wind speeds:



where are parameters. This distribution is known as a three-parameter Weibull distribution. Below is a plot of the probability distribution with the same histogram as we saw before with δ = 2.1327, γ = 1.8832, and β = 6.2926 (the y-axis values have been rescaled).



How could this probability distribution be used then?

Characteristics of a probability distribution for a continuous random variable:

* f(y) ≥ 0
* f(y) may be > 1
*  – the area underneath the curve is 1

A probability distribution for a continuous random variable is often referred to as a probability density function (PDF).

Expected values

For those of you who have had a calculus class, one can show that



and



Compare these quantities to what was shown for expected values of discrete random variables. Also, please remember that the notation Var(Y) is often used to represent .