RV & Probability Distribution Review

Probability Distributions

Definition: A probability distribution is a function that maps outcomes of all possible resolutions of uncertainty (elementary outcomes) to **numerical** values representing the probability of each outcome. The probability distribution is considered **discrete** when there is a countable number of possible outcomes (e.g. flipping a coin has two outcomes). The probability distribution is considered **continuous** when there is a non-countable number of possible outcomes.

We will use capital letters, like X, to represent an uncertain outcome. Many statisticians refer to this unknown result as a **random variable (RV)** representing an unknown value. Let's take an example from the craps table at a casino. Let's assume you place a $10 bet on "Craps 3", a bet that gives 15 to 1 odds. If the roll is a three, you will win $150; otherwise, you lose $10. Let X be your winnings. A mapping of winnings to outcomes of throwing two dice is as follows:

Die #1	**Die #2**	**RV Value**	**Die #1**	**Die #2**	**RV Value**	**Die #1**	**Die #2**	**RV Value**
1	1	-$10	3	1	-$10	5	1	-$10
1	2	**$150**	3	2	-$10	5	2	-$10
1	3	-$10	3	3	-$10	5	3	-$10
1	4	-$10	3	4	-$10	5	4	-$10
1	5	-$10	3	5	-$10	5	5	-$10
1	6	-$10	3	6	-$10	5	6	-$10
2	1	**$150**	4	1	-$10	6	1	-$10
2	2	-$10	4	2	-$10	6	2	-$10
2	3	-$10	4	3	-$10	6	3	-$10
2	4	-$10	4	4	-$10	6	4	-$10
2	5	-$10	4	5	-$10	6	5	-$10
2	6	-$10	4	6	-$10	6	6	-$10

What are the possible values of X?

Assuming each elementary outcome is equally likely and only one elementary outcome occurs at a time, the probability of each elementary outcome of the dice is 1/36. What is the probability of each value of X?

We have just specified a probability distribution of a random variable because it satisfies:

$$\sum_{i=1}^{k} p(v_i) = 1 \qquad and \qquad p(v_i) \geq 0 \quad and$$
$$for\ any\ two\ mutually\ exclusive\ events,$$
$$the\ probability\ that\ one\ or\ the\ other\ occurs\ is\ the\ sum\ of\ their\ probabilities$$

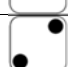
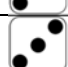
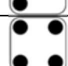
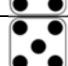
where there are k possible values of the RV, X, each value is denoted by $v_1, v_2, \ldots, v_k$, and we use a shorthand to denote the probability of a typical value, $P(X = v_i)$ or equivalently $p(v_i)$ or $f(v_i)$. Often the notation $f(v_i)$ will be referred to as the **probability mass function** (discrete distribution) or **probability density function** (continuous distribution). Others may use the notation $p(x)$, $\mathbb{P}(X = v_i),\ \ or\ \ f_X(v_i)$ for probability mass/density functions.

While the above is all that is needed to summarize a discrete probability distribution, we will sometimes find it useful to calculate cumulative probabilities. A **cumulative probability** is the probability that the random variable is less than or equal to some particular value. We denote cumulative probabilities using a capital F where $P(X \leq x)$ is denoted by $F(x)$ and refer to this as the **cumulative distribution function (CDF)**. Note that for a discrete random variable:

$$F(x) = P(X \leq x) = \sum_{v_i \leq x} f(i)$$

Let's now give this all some meaning by specifying probability mass functions and cumulative distribution functions for:

Roll of a Die (Y): 

Outcome	$\boldsymbol{y}$	$\boldsymbol{f(y) or\ P(Y = y)}$	$\boldsymbol{F(y) or\ P(Y \leq y)}$
	1		
	2		
	3		
	4		
	5		
	6		

Now that we have gone through the effort of being able to understand a probability distribution for a Random Variable, let's now create a simple summary measure for probability distributions. Mean (also known as the Expected Value) gives us the central tendency of the random variable and is calculated as follows:

Mean

$$\mu = E(X) = \sum_{i=1}^{k} (v_i \cdot f(v_i))$$

where,

What is the mean of a $10 bet on "Craps 3"?

Binomial Distribution

One of our first probability distributions modeled the flip of a coin which has two outcomes:

$$X \equiv Outcome\ of\ Coin\ Flip$$
$$X = 1\ if\ heads\ comes\ up$$
$$X = 0\ if\ tails\ comes\ up$$
$$P(X = 1) = p(1) = \frac{1}{2}$$
$$P(X = 0) = p(0) = \frac{1}{2}$$

Is the Bernoulli Distribution discrete or continuous?

While the Bernoulli distribution is mathematically important, just one flip of a coin, or one patient testing a new drug hardly gives us evidence to say anything meaningful. What we really want is to see if the coin is fair, the drug is better, or the new President will be a Democrat. Thus, we need multiple coin flips, multiple test patients for a new drug, and multiple sample voters for an election poll before we can declare something like "the next President of the United States will be a Democrat".

Enter the **Binomial Distribution**:

A binomially distributed random variable, such as the number of heads in 50 coin flips, can be described with two parameters (as opposed to one parameter like the Bernoulli Distribution). The first parameter, usually denoted by n, represents the number of trials (e.g. coin flips). The second parameter, often denoted by θ, is the probability of success (e.g. 50% for a coin flip). The probability distribution of a binomially distributed random variable (K) represent the number of successes in n trials with the probability of success being θ is then mathematically given as:

$$p(k) = \binom{n}{k} \theta^k (1-\theta)^{n-k} \quad where \quad \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

Wow, this is one ugly formula.

Is the Binomial distribution discrete?

What are the possible outcomes?

The only nice thing about this formula is that the expected value is easily calculated; $E(K) = n\theta$. For example, suppose 300 passengers have reserved seats on a flight and the probability that each passenger shows up is 85%, how many passengers are expected to show? Notice that when $n = 300$ and $\theta = 0.85$, what is the expected value of K?

Fortunately, for this class, we will use R to do the ugly calculations of the binomial distribution. We just need to be able to interpret the output. Let's open R and use the R function `dbinom` for this purpose. As an activity, use R to fill in the below table for the following problem:

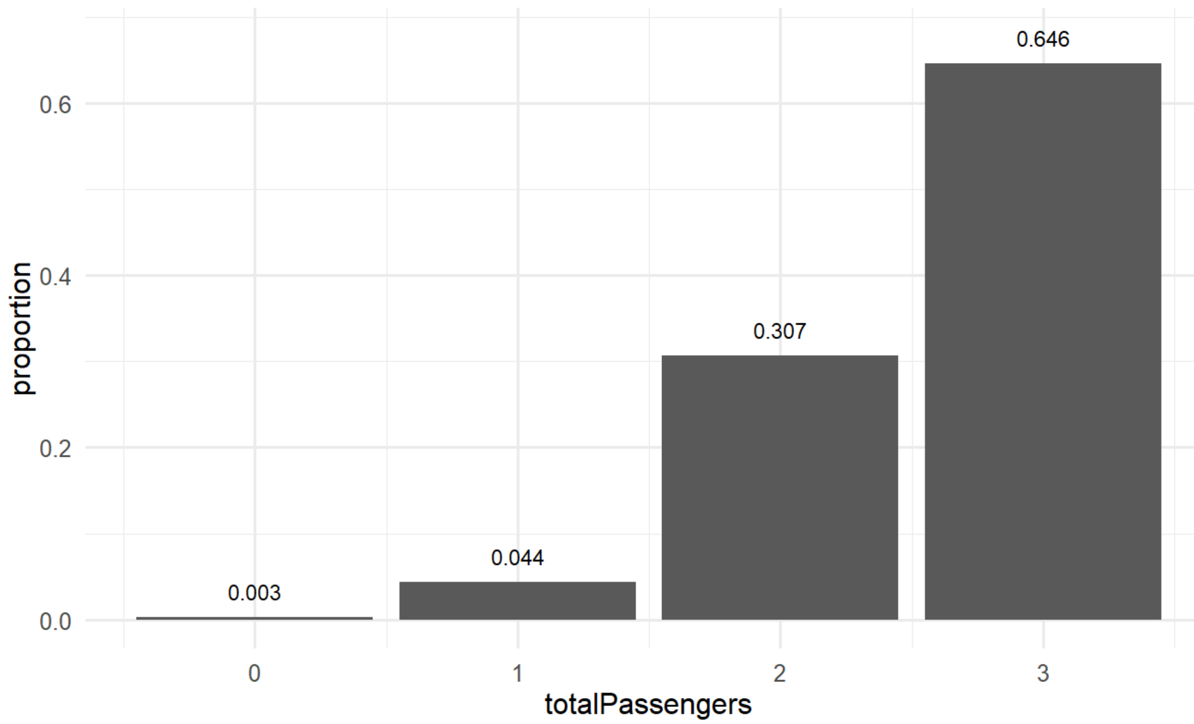
Activity for Class

Earlier, we used a generative model for this example. We will now answer the same question using the CDF of a binomial distribution in R. **Example:** The XYZ company owns and operates a 3-seater airplane to show tourists the Great Barrier Reef in Cairns, Australia. The company uses a reservation system, wherein tourists call in advance and make a reservation for aerial viewing the following day. Unfortunately, often passengers holding a reservation might not show up for their flight. Assume that the probability of each passenger not showing up for a flight is 15% and that each passenger's arrival probability is independent of the other passengers.

a) If XYZ takes three reservations, what is the probability distribution modeling the number of passengers, X, that will show up. Use the `dbinom` function in R to fill the table below.

# of passengers (x)	$p(x)$	
0		Hint: `dbinom(x=2,size=3,prob=1/2)` would be the formula for returning 2 success in three trials with 50% chance of success.
1		
2		
3		

After filling in the above, compare your results to our generative model results shown below:



Note: R refers to $p(x)$ as the *density function*. R refers to the $P(X \leq x)$ as a *probability distribution function* and for a binomially distributed random variable, one would use the `pbinom` function instead of `dbinom`. For example, the probability of accepting four reservations and then having enough space could be calculated as:

$P(X \leq 3) =$`pbinom(q=3, size=4, prob=0.85)` = ??

$$Is\ this\ a\ good\ business\ practice\ for\ this\ company?$$

Continuous Distribution

For a discrete probability distribution, $p(x)$, was the probability of a random variable equaling x. In the continuous world, say for modelling time, temperature, proportions, and other values that can be made infinitely accurate if the measurement tool is sophisticated enough, $p(x)$ loses its nice meaning. When x can take on infinite values, we should note two things about the probability density function:

1. The higher, $p(x)$ is, the more likely x is.
2. The $P(X = x) = 0$. This is the tricky one. The chances of getting any one particular value can be thought of as $P(X = x) = \frac{\#\ of\ outcomes\ where\ X=x}{\#\ of\ possible\ outcomes} = \frac{1}{\infty} \approx 0$.

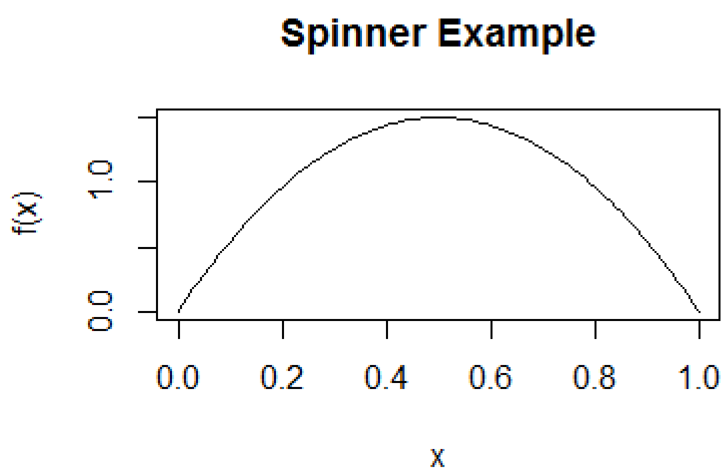
Doesn't point #2 above make the PDF for a continuous distribution pointless? The answer is sort of. We really do not care about the PDF. However, we can use the area under the probability density curve say all sorts of useful things about continuous random variables.

In a continuous world, we use calculus to calculate the area (you are not responsible for these mathematical equations of density/distribution functions, they are presented only to aid understanding and build your intuition about interpreting the area under a probability density function). From calculus, the following result for the all continuous probability distributions can be proven:

1. The **area** (not the value of $p(y)$) under **any valid probability density function** represents probability and the total area is equal to 1:

$$\int_{-\infty}^{\infty} p(y)dy = 1$$

As a graphical example of this notion of probability and area, consider a spinner with a [0,1] scale on its circumference. Suppose that the spinner is magnetized in some way such that it is biased and its probability density function is $p(x) = 6x(1-x)$. This probability density function is shown here:



The area under this curve should equal one if $\int_{x=0}^{x=1} 6x(1-x)dx = 1$. Let's check together (note: you will not be responsible for calculus in this course … only that the area under the probability density function represents probability of an outcome within that area. Calculus check:

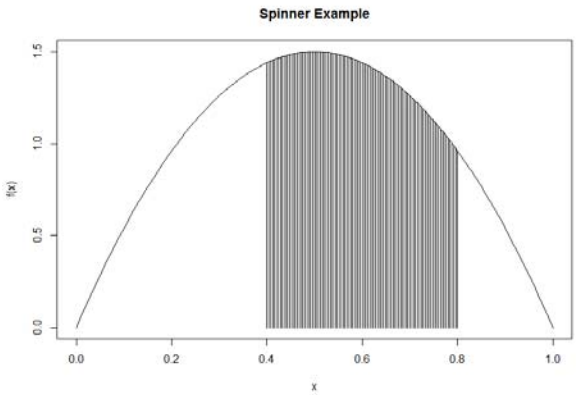
$$\int_{x=0}^{x=1} 6x(1-x)dx = 6\int_{x=0}^{x=1}(x-x^2)dx = 6\left[\frac{1}{2}x^2 - \frac{1}{3}x^3\right]_0^1 = 6\left(\left[\frac{1}{2}1^2 - \frac{1}{3}1^3\right] - \left[\frac{1}{2}0^2 - \frac{1}{3}0^3\right]\right)$$
$$= 6\left(\left(\frac{1}{2} - \frac{1}{3}\right) - (0-0)\right) = 6 \times \frac{1}{6} = 1$$

Check, it works!!!

The more useful function for making probability statements about a continuously distributed random variable is the **cumulative distribution function (CDF)** or equivalently, $F(x) = P(X \leq x)$**.** For our spinner, the CDF has the following form:

$$F(x) = P(X \leq x) = \int_0^x 6y(1-y)dy = \mathbf{3x^2 - 2x^3} \quad for\ 0 \leq x \leq 1$$

Use this function (i.e. $F(x) = 3x^2 - 2x^3$) to find: $P(0.4 \leq x \leq 0.8)$?



Key takeaway: the CDF gets us the probability we want without worrying about calculus. In fact, for any random variable, if we know its distribution, then we can use R to answer our probability questions without worrying about the mathematical details of the distribution.