STA-1380 – Elementary Statistics
Week #4

Hello! Welcome to the additional online **Weekly Resources** for the course of **STA-1380**. Following a traditional calendar semester, these will be some of the topics your professors will go over. If you do not see material your section is going over for the week, please look at the other resources listed for this course. In addition to these resources, there might be **Group Tutoring** for this course, please see our website for more details. These sessions will go over these materials in more detail as well as any questions about the material.

Any additional help or services can be found through the **Baylor Tutoring Website**. Visit to schedule a free 30-minute private tutoring session, drop-in times for your course, the Baylor Tutoring YouTube channel, or any additional tutoring resources.

Contacts: Sid Rich M-Th 9am-8pm (Fall and Spring class days) Office Phone: 254-710-4135

Topic of the Week:
"Continuous Probability Distributions"

Key Points:

- Probability Density Functions
- Cumulative Distribution Functions
- Uniform Distributions

Last week's material introduced the Discrete Distribution family with the PMFs, and Binomial Distribution. This week's material will cover the sister set of Distributions known as Continuous Distributions as well as introduce one of the named continuous CDFs which is Uniform Distributions.

Continuous Distributions: These models depict data sets with Continuous Random Variables as well as the ratios derived from qualitative data. In this model, probabilities are continuously spread out across the entire specified range, meaning there can be an infinite number of positions a data point can lie upon the range. Data models that can be expressed as functions, percentages, or appear without breaks on a graph are prime candidates for continuous models.

> Suppose you wish to identify the mean weight of all tree frogs. Since a frog could in theory weigh *any* value, the distribution must account for the possibility of any value and therefore must be continuous.

Another example would be a study on the correlation between gender and degree choice. Despite these two variables being categorical, the ratios gained from the data are quantitative. If "X" number of women out of the entire study have a Doctorate that turns the categorical number into a numerical ratio. There are infinitely many values that each ratio can take on, making this distribution continuous.

It is important to remember that the entire point of a distribution is to display the probabilities involved with the variables a study is targeting. Even though your variables are discrete or continuous, always check the way a question is worded to correctly identify the proper distribution needed for the data. If the question asks for a ratio or proportion
Remember that for specific distributions, there will be a certain way of identifying the distribution via symbols called Function Notation often depicted as $X \sim Graph$(Parameters)

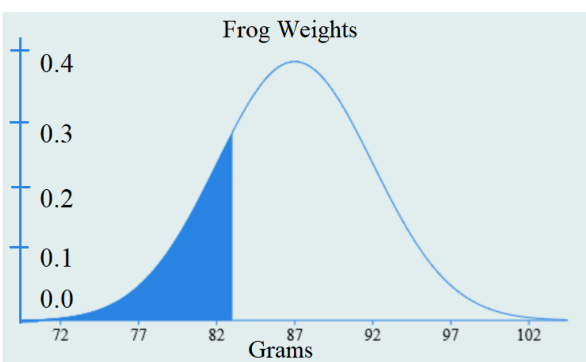
The 'X' value is a representation of all values of X that the graph can depict. The 'Graph' portion of the notation is often an abbreviation of the named distribution. In last week's guides, the Binomial distribution was shortened to 'Bin' in $X \sim Bin(n,p)$. In each distribution, what goes into the parentheses will be different. What they all share is that the values listed are the steppingstones to procuring a probability. For the case of a Binomial distribution, the probability can be found by multiplying $n*p$. Look out for what each distribution's 'steppingstones' are and memorize them.

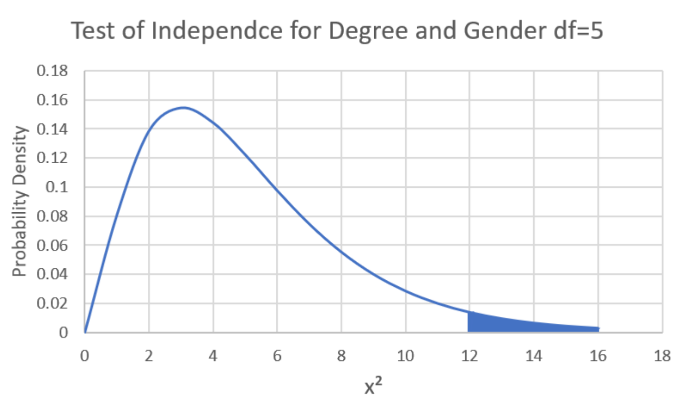
Highlight #1
"Probability Density Function"

Definition: A distribution of every possible outcome a continuous random variable can be used to express the relative likelihood of each outcome.

Notation: $f(x)=$ *Probability:* $P(a < x < b)$

Example: Graphs





First and formost the critical point that separeates a PDF from a PMF is the value of what *P(X = x)* is equal to. In a PMF, the exact value of *x* will have a value. In a PDF, the value of any given probability is infintesimal or 0. Think about it this way: A frog could weigh 83 grams or it could weigh 83.000000001 grams. Since these are two separate outcomes for the x variable, this causes the probabilities previously seen in a PMF to spread out across the entire range of the values. The term, ‘Density’ is often used to describe how much of the probability is given to a certain range. Notice how with the frog example that there is a bell-shaped curve where the probability is “denser” around the center and “sparse” around the tails. The purpose of a PDF is to give a visual representation of where the data tends to congragate.

PDFs are not usually found in table form because of the ‘0’ value property it has at any given point. Formulas do exist but are much more complicated and often require notations often associated with upper levels of Calculus. For this course you will not need to worry about any of the formulas and instead should focus on the notation styles of each distribution further down the line.

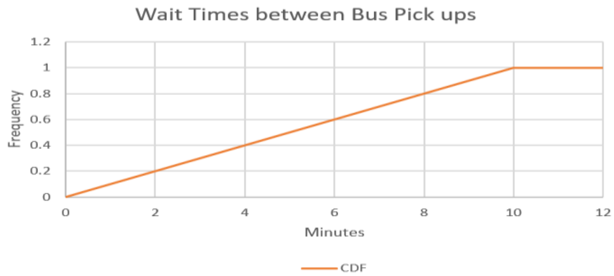
Highlight #2

“Continuous Cumulative Distribution Functions”

Definition: An accumulation of probabilities that gives the probability of having at least a set ‘X’ value identified by a certain and predetermined range.

Notations: $F(x) = P(X \leq x) = \int f(y)dy$

Example: Graphs, Functions, and Tables



*A Uniform Distribution CDF

$$F(X) = \begin{cases} 0 & X < 3 \\ \frac{3}{2}\left(\frac{x}{10}\right)^2 - \left(\frac{x}{10}\right)^3 & 3 \leq X \leq 7 \\ 1 & X > 7 \end{cases}$$

* A Function depicting a CDF

As mentioned in the previous weekly guide, the Cumulative Distribution Function or CDF is used for both discrete and continuous variables. However, the CDFs for continuous variables were saved for this week’s lesson. An important distinction between a continuous and discrete CDF is that there is no difference between *P(X < x)* and *P(X ≤ x)*. Since P(X = x) has no value for a PDF, the CDF does not consider whether the value is included and instead goes to the infinitely closest option.

A CDF's value is not derived from the area under a curve and is instead the exact value that corresponds with its point. For example, on the Bus time graph, 50% of the probability happens at 5 minutes, meaning that half of the buses can be expected to take at most 5 minutes between rotations.

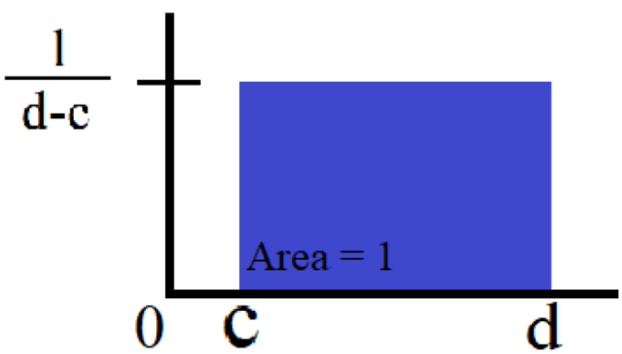
The most common type of CDF is that of the 'Table'. A very common CDF table will be shown with the Normal Distributions Highlight. These give accurate depictions of what the cumulative probability is down to 100th in specifications. Tables like these allow for detailed and accurate reads rather than having to 'eyeball' a graph or calculate the results of a function. Because of its simplicity, tables will be used throughout all STA-1380 course work. A very common CDF table will be shown with the Normal Distributions Highlight.

Highlight #3
"Uniform Distribution"

Definition: A single bar graph that depicts a continuous probability across all its spanned range.

Notation: *X ~ Unif(c, d)* $f(x) = \frac{1}{d-c}$ $c \leq X \leq d$ $\mu = \frac{c+d}{2}$ $\sigma = \frac{d-c}{\sqrt{12}}$

Example: Graphs



* The PDF graph of a Uniform Distribution

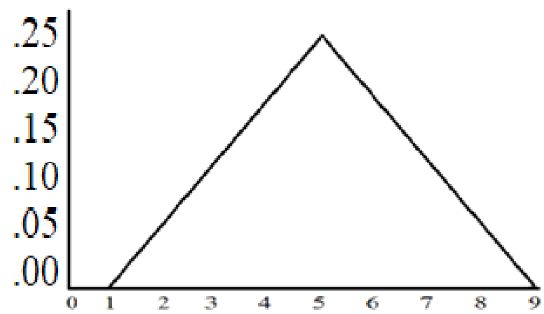
The Uniform Distribution is a simple introduction into the world of continuous distributions. Often depicted in its PDF form as seen above, the graph depicts a streamline set of equal probabilities across its entire range. This range will almost always be given in the form of (c,d) in which the value of 'c' represents the absolute minimum a value within the distribution can take and 'd' the absolute maximum.

Unlike the other functions, the Uniform distribution's style of representing data is very intuitive. In order to keep the probability equal to 1, the area of the graph, which in this case is a rectangle, must also be equal to 1. Since the formula of a rectangles area can be given by Base*Height, these values must be reciprocal. If the length of the base between c and d is 4, then the height will be $1/4^{th}$. As such, at least one of these values will always be given in order to solve a problem. The height of the graph is also equal to the probability of each 'x' value occurring.

The mean of a Uniform Distribution will always be the average sum of the c and d values, the mid-point method was given above. To figure out where the std. deviation formula comes from would require calculus. For now, stick to memorizing the formula for an easier application.

Check Your Learning

1. The time a Baylor student spends waiting to take the bus once they arrive at the bus stop is demonstrated via the given distribution:
    a. Is this distribution a PDF or CDF and how can you tell?
    b. What if a student takes 5 minutes to walk to the bus stop from their last class, what is the probability that they will wait more than 8 minutes after class?



*The Graph is partially to scale

2. An artist wishes to know what portion of their songs is longer than three minutes. They collect the durations of every song they have produced and get an average length of 2.7 minutes with a standard deviation of .2 minutes.
    a. What is the variable is this question addressing? Is it continuous or discrete?
    b. What would be the probability of a song that is exactly 2 minutes and 54 seconds long?

Things Students Struggle With

1. PMF vs PDF
    a. With acronyms sharing multiple letters, it can be difficult to remember the rules of each distribution and how to classify each one. One way to remember a PMF versus a PDF is that PDF's "DONT" have a value for P(X = x) while a PMFs "MIGHT." There is always a chance that there could be a value at P(X = x), but there is NEVER a chance that a PDF will have a non-zero value at a specific value.

2. Statistical Notation:
   a. In Statistics, many students find it difficult to remember how to write the probability of a function of ‘X’. We use CAPITAL P to denote the ‘Probability’ of a function. We use CAPITAL X to denote the function as a whole (think a graph or a PDF line). We use LOWERCASE ‘x’ to denote a SPECIFIC number or set of numbers we are targeting. We use SIGNS to denote how the probability refers to the function as a whole. If we want to find the probability of a distribution for all the potential numbers GREATER THAN 5, we will use P(X > 5). Note that the little ‘x’ has become 5 because this is a specific number.

Concluding Comments

That’s it for this week! Please reach out if you have any questions and don’t forget to visit the Tutoring Center website for further information at https://www.baylor.edu/tutoring.

Answers to CYL

1. **a.** This is a PDF because the because the total Area under the curve will equal ‘1’ and the ‘Y’ value does not reach ‘1’.
b. 87.5% P(X>3), the distribution shows that only 12.5% of the area is to the left of 3.

2. **a.** The ‘Durations’ of the songs. (Not just the songs but their times specifically)
b. 0% Chance. (Remember! PDFs “DON’T” have exact probabilities!)