

Section 6: Introduction to Continuous Probability Distributions

Continuous Probability Distributions

With an understanding of discrete probability distributions, we are now ready to tackle continuous random variables. The idea of a probability distribution here does not change. We need to identify $P(x)$ for all possible x values. Of course, the difference in the continuous case is that we have an infinite number of outcomes so we cannot simply list them out as typically done in the discrete case. In mathematics we use functions to represent an infinite number of outcomes on a continuous scale and we will utilize that concept for continuous probability distributions.

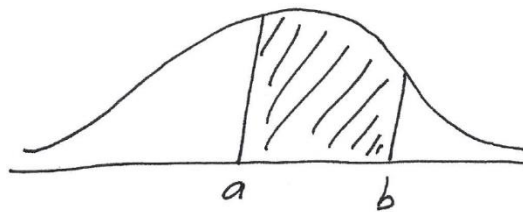
A smooth curve known as the **density function**, $f(x)$ is used to represent the probability distribution of a continuous random variable. The random variable will be denoted by x and the probability will be represented by the area under the function, $f(x)$.

Properties of a density function

- Must be a function – remember from algebra that a function is a relation where for every x (input) there is exactly one $f(x)$ (output), passes the vertical line test
- The curve must never fall below the x -axis, $f(x) \geq 0$ for all x (this should make sense because you cannot have a negative probability, $0 \leq P(x) \leq 1$ for all x)
- The total area under the curve must be 1 (this should also make sense because the total probability in a valid probability distribution is 1 and the maximum any probability can be is 1)

For continuous random variables we assign probability to intervals. (Not Points)

$P(a \leq x \leq b)$ is the area under the curve between a and b (example curve displayed below)



The probability for each point in a continuous probability distribution is theoretically zero. There are multiple ways to think about this idea. Suppose in the above graph, you calculate $P(x = a)$. You could just think about it as a rectangle. A single point has width 0, so no matter the height, the area is 0. You can also think about the ideas presented in basic probability. We know the basic idea is to take the number of outcomes that satisfy the event over the total number of outcomes. In a continuous case there is an infinite number of outcomes so the denominator goes to infinity. As the denominator gets bigger the overall quantity gets smaller and goes to 0. However you prefer to think about it, the theoretical probability of a single point in a continuous distribution is 0.

Thus for continuous variables $P(a \leq x \leq b) = P(a < x < b)$ since $P(x = a) = 0$ and $P(x = b) = 0$. The equality (in \leq) does not matter since it has a probability of 0.

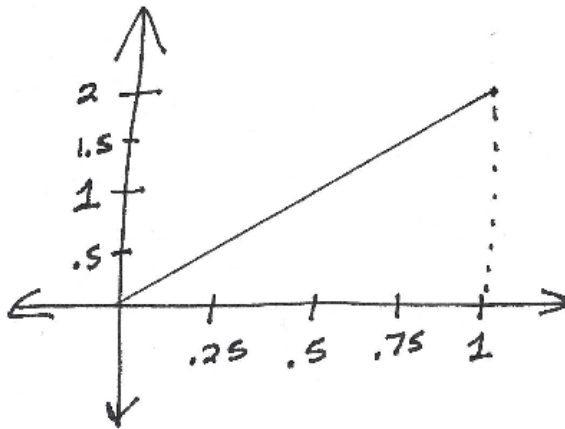
Example

Using the density function, $f(x) = 2x$ for $0 \leq x \leq 1$, do the following.

- Show the total probability (area) is 1
- Find $P(0 \leq x \leq 0.5)$.
- Find $P(0.5 \leq x \leq 1)$.
- Find $P(0.25 \leq x \leq 0.75)$.

Solutions

The first step with these kinds of problems is to draw a picture of the distribution. From the picture we can see the distribution is in the shape of a triangle. Therefore, we will use the area of a triangle formula to solve this problem. The formula is: $A = \frac{1}{2}b \cdot h$



- $A = \frac{1}{2}b \cdot h = \frac{1}{2}(1)(2) = 1$
- Here we are finding the area between 0 and 0.5 so the base is 0.5. We can find the height at 0.5 by plugging this into the function which gives us a height of 1.

$$P(0 \leq x \leq 0.5) = \frac{1}{2}b \cdot h = \frac{1}{2}(0.5)(1) = 0.25$$
- Here we are finding the area between 0.5 and 1 which is not a triangle. Therefore, we have to break the distribution up so we can find the area we want. In this case we already know that the total area is 1 and the area between 0 and 0.5 is 0.25. If we subtract these two, we will get the area we want.

$$P(0.5 \leq x \leq 1) = 1 - 0.25 = 0.75$$
- Here we are finding the area between 0.25 and 0.75 which will be done similar to the previous probability. We have to break up the area. First find the area between 0 and 0.75, then find the area between 0 and 0.25, and then subtract.

$$A = \frac{1}{2}b \cdot h = \frac{1}{2}(0.75)(1.5) = 0.5625$$

$$A = \frac{1}{2}b \cdot h = \frac{1}{2}(0.25)(0.5) = 0.0625$$

$$P(0.25 \leq x \leq 0.75) = 0.5625 - 0.0625 = 0.5$$

The Mean and Standard Deviation for Continuous Random Variables

The calculation of the mean, variance, and standard deviation for a continuous probability distribution works the same way as it does for the discrete probability distribution. The only difference is mathematical and unfortunately the required mathematics is taught in Calculus. For that reason, we will only look at specific types of continuous probability distributions.

The Uniform Distribution

There are many types of continuous probability distributions. We will discuss several in this class. The uniform distribution is common and will allow us to discuss some of the mathematics of a continuous random variable. As seen previously in class, a uniform distribution is flat. Therefore, the equation is just a horizontal line which is relatively easy to deal with mathematically.

The density function for the uniform distribution is $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$.

The parameters of the uniform distribution are a and b . These are constants that define which uniform distribution we are dealing with. As you can see, the density function is just a constant function (graphically a horizontal line). Remember that x is the random variable and $f(x)$ is used to represent the probability.

Example

Suppose we are interested in a uniform distribution with parameters $a = 2$ and $b = 4$. Do the following.

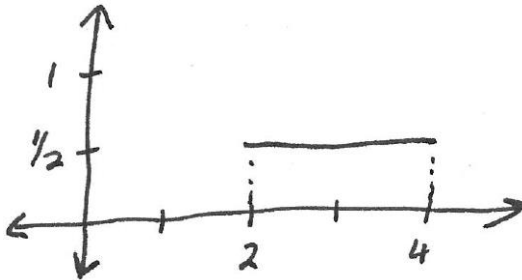
- Identify the density function.
- Draw a picture of the density function.
- Show that the total area under the density function is 1
- Find $P(2 \leq x \leq 3)$.
- Find the mean, μ .

Solutions

a. $f(x) = \frac{1}{b-a} = \frac{1}{4-2} = \frac{1}{2}$ for $a \leq x \leq b$
 $f(x) = \frac{1}{2}$ for $2 \leq x \leq 4$

(Make sure when you identify a uniform density function that you limit the domain. If you do not limit the domain then it is not a density function because the total area will not be 1.)

b.



(The density function is just a horizontal line through $\frac{1}{2}$ but the domain is limited between 2 and 4 so only that part of the line creates the density function.)

- c. The uniform distribution is in the shape of a rectangle so we can calculate the area by using $A = l \cdot w$.

$$A = l \cdot w = \left(\frac{1}{2}\right)(2) = 1$$

(You should see here how the density function of the uniform works. Basically, you limit the domain so the product of the length and width is 1.)

d. $P(2 \leq x \leq 3) = l \cdot w = \left(\frac{1}{2}\right)(1) = 0.5$

(Looking at the picture you should see that this is $\frac{1}{2}$ the rectangle.)

- e. Since we know that a uniform distribution is symmetric, the mean must be at the point of symmetry. We just showed half of the area is to the left of 3 so the other half is to the right. Therefore, the mean is just 3. You should see the graph is symmetric about 3.

$$\mu = 3$$

Descriptive Statistic Shortcut Formulas for the Uniform

Like you have seen before with the binomial distribution, when specified properties are satisfied it is often possible to come up with shortcut formulas. For the uniform, we can identify shortcuts for probability, the mean, and the variance. In application, this is a time saver. To derive shortcut formulas, we use the information we know. For example, to derive a shortcut for the mean of a uniform, you simply use the probability distribution formula for the uniform and plug it into the mean formula for continuous probability distributions. These derivations are beyond the scope of what we will cover here, but we will utilize the shortcut formulas.

- Since a and b are the parameters of the uniform, we will use c and d in the probability statement. Of course, c and d must be within the domain of the function.

$$P(c \leq x \leq d) = \frac{d-c}{b-a} \text{ where } a \leq c \leq d \leq b$$

- Population Mean = $\mu = \frac{a+b}{2}$
- Population Variance = $\sigma^2 = \frac{(b-a)^2}{12}$

Other necessary Calculations for a Uniform Distribution

$$\text{Population Standard Deviation} = \sigma = \sqrt{\sigma^2}$$

As you should be accustomed to, the standard deviation is the square root of the variance.

$$\text{Population Median} = \mu$$

The uniform distribution is symmetric so the mean and median are equal.

$$\text{Population Mode} = D.N.E.$$

The uniform distribution is flat (there is no peak) so the mode does not exist.

$$\text{Population Range} = High - Low = b - a$$

The low and high for the uniform are simply the parameters a and b , respectively.

Example

Travel time from Lexington KY to Columbus OH is a uniform distributed between 200 and 240 minutes. For this distribution, do the following.

- Identify the density function.
- Find the probability of arriving in less than 225 minutes.
- Find the mean, μ .
- Find the median.
- Find the mode.
- Find the range.
- Find the variance, σ^2 .
- Find the standard deviation, σ .

Solutions

- $f(x) = \frac{1}{b-a} = \frac{1}{240-200} = \frac{1}{40}$ for $a \leq x \leq b$
 $f(x) = \frac{1}{40}$ for $200 \leq x \leq 240$
- $P(200 \leq x < 225) = \frac{d-c}{b-a} = \frac{225-200}{240-200} = \frac{25}{40} = \frac{5}{8} = 0.625$
- $\mu = \frac{a+b}{2} = \frac{200+240}{2} = 220$
- Median = 220
- Mode = D.N.E.
- Range = $b - a = 240 - 200 = 40$
- $\sigma^2 = \frac{(b-a)^2}{12} = \frac{(240-200)^2}{12} = \frac{1600}{12} = 133.333$
- $\sigma = \sqrt{\sigma^2} = \sqrt{133.333} = 11.547$