2/25/2023

Continuous Distributions (ch 5)
Review for Exam #1

Differences between discrete distributions and continuous distributions?

Discrete distributions have values at individual outcomes. $P(X = x)$ is not zero if the outcome x is in the range of possible outcomes.

If you roll a die and the outcomes are 1, 2, 3, 4, 5, 6… only impossible outcomes have probability 0. Each of these has a non-zero probability. $P(X = 1) = \frac{1}{6}$ if the die is fair.

But for a continuous distribution, say with a range of outcomes between 1 and 6, the $P(X = 2) = 0$. Probabilities for continuous distributions can only be calculated inside a range of values, not at a single point.

When we talk about a exact distribution in discrete probability, the primary function we use is called the probability density function or the probability mass function. In the discrete case, we can use the probability density/mass function to calculate the probability at a single value. In the continuous case, the probability density function basically just draws the shape of the distribution, and so can't use it for probability calculations directly. Abbreviated pdfs/pmfs

The cumulative probability distribution (cdf) calculates probabilities over a range of values (typically for values less than (or equal to) a particular value. For continuous distributions we are only ever going to use the cdfs. The probability under a continuous probability distribution is the area under a curve. In Excel, the cumulative question will always be TRUE (unless we are drawing the curve… which I will never ask).

The other thing that is different is that $P(c < X < d) = P(c \leq X \leq d)$ when working with a continuous distribution. This is not the case with discrete distributions.

$P(X < 2)$ in the dice case, includes only X=1.
$P(X \leq 2)$ in the dice cases include both X=1, and X=2.

In the continuous case, these are the same value.

An example of a continuous distribution is the uniform distribution. Every value in the range has the same probability (density). And the curve itself looks like a rectangle. And any range of values, we can calculate the probability by calculating the area of a rectangle.

Uniform distribution: it produces a range of values between a and b. And length of the base of the distribution is $b - a$. And then the height of the rectangle (straight line representing the height of the distribution) is $\frac{1}{b-a}$. So that the area under the total curve is equal to 1.

The area under the whole curve must equal 1 because that represents all the possibly probabilities.

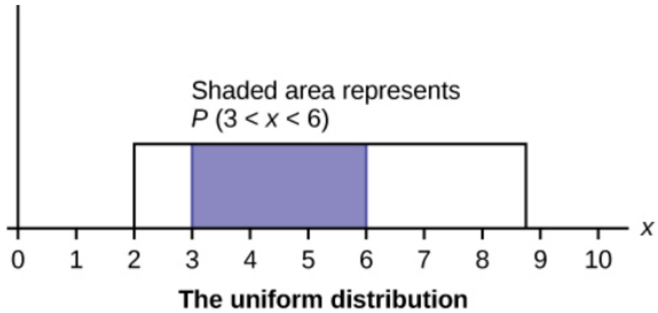

Figure 5.2 The graph shows a Uniform Distribution with the area between $x = 3$ and $x = 6$ shaded to represent the probability that the value of the random variable X is in the interval between three and six.

The height of the distribution is related to how wide the range of possible outcomes is.

$$8.8 - 2 = 6.8$$

(roughly)

The height of the distribution is $\frac{1}{6.8} \approx 0.14705\ldots = \frac{5}{34}$.

$$P(3 < X < 6) = P(3 \leq X \leq 6) = (6 - 3)\left(\frac{5}{34}\right) = 3\left(\frac{5}{34}\right) = \frac{15}{34} \approx 0.441176\ldots$$

In a typical cdf calculator/function, the probability calculated it $P(X \leq x)$. It doesn't do a range in the sense that it is between two values, it only does less than a given value (one value).

If we want to find in between two values, then the equivalent calculation is:

$$P(c < X < d) = P(X < d) - P(X < c)$$

$$P(3 < X < 6) = P(X < 6) - P(X < 3)$$

$$P(X < 6) = 4\left(\frac{5}{34}\right) = \frac{20}{34} = \frac{10}{17}$$

$$P(X < 3) = 1\left(\frac{5}{34}\right) = \frac{5}{34}$$

$$P(3 < X < 6) = \frac{20}{34} - \frac{5}{34} = \frac{15}{34}$$

In two weeks, we'll worry about the normal distribution. But for now, the uniform distribution is the only one to worry about. In Excel, there is a default uniform distribution function that produces random values between 0 and 1.

Review for Exam #1