Cody Kimmins

Chapter 4 Write up

Chapter 4 introduces continuous variables and the probability distributions associated with them. The last few distributions in the chapter include normal, gamma, and beta. Section 4.5 discusses the normal distribution which is referred to the most widely used continuous probability distribution. This distribution has a bell-shaped curve like the empirical rule. The empirical rule is used to tell you where most of the values lie in a normal distribution. It is also known as the 68-95-99.7 rule where around 68% of values are within 1 standard deviation of the mean. Around 95% of the values are within 2 standard deviations of the mean and with 99.7% it is 3 away. This distribution encompasses two important terms that are mean and standard deviation. It is a symmetrical arrangement of a data set in which most values cluster in the mean and the rest taper off towards either extreme left or right. The normal or Gaussian probability distribution is popular because of its unique mathematical properties which facilitate its application to practically any physical problem in the real world. The expected value is equal to μ, and the variance is equal to .

Areas under the normal density function corresponding to P (a ≤ Y ≤ b) require evaluation of an integral. The normal density function is symmetric around the value μ, so areas need to be tabulated on only one side of the mean. This probability distribution can be used in datasets such as heights, blood pressure, rolling a die, and IQ scores. For a data set like the height of a population is a good example because most people have an average height. There is a number a people that are either extremely short or extremely tall compared to the average. This is a normal distribution because there are several different characteristics that determine height such as genetics or environment. Another example of this distribution is rolling a dice that is a little more familiar. The average in rolling a die is getting the result of 7 where everything less or more than that have a lower chance. The chances of rolling a 2 or 12 are extremely unlikely when compared to the average in the data set. So, with the normal probability distribution the data is symmetrically distributed with no skew hence the bell shape.

The next section 4.6 is about the gamma probability distribution is a continuous probability distribution that models right-skewed data. It is also used to predict the wait time until the k-th event occurs. This means that you can use it to model failure times, wait times, or service times. The quantity is known as the gamma function. α is sometimes called the shape parameter associated with a gamma distribution. This specifies the number of events you are modeling. The parameter β is generally called the scale parameter which represents the mean time between events. Multiplying a gamma-distributed random variable by a positive constant produces a random variable that also has a gamma distribution with the same value of α but with an altered value of β. A random variable Y is said to have a chi-square distribution with positive degrees of freedom if and only if Y is a gamma-distributed random variable with parameters α = ν/2 and β = 2. Another special case is the exponential distribution with parameter β > 0. The easiest way to compute probabilities associated with gamma-distributed random variables is to use available statistical software. The most frequent use case for the gamma distribution is to model the time between independent events that occur at a constant average rate.

Some examples of this include models like cancer rates, insurance claims, and rainfall. The age of distribution of cancer incidence often follows the gamma distribution where the shape and scale parameters predict the number of driver events and the time interval between them. There are two versions of the gamma distribution where they vary based on the number of parameters. There is one with three parameters that include shape, scale, and the threshold. When the threshold parameter is set to zero then it is a two-parameter gamma distribution.

Section 4.7 introduces the last distribution which is called the beta distribution probability. It is a continuous probability distribution often used to model the uncertainty about the probability of success of an experiment. Its density function is defined over the closed interval 0 ≤ y ≤ 1. It is often used as a model for proportions, such as the proportion of impurities in a chemical product or the proportion of time that a machine is under repair. Unlike the previous two distributions, the graphs of beta density functions assume widely differing shapes for various values of the two parameters α and β. For the distribution to work, both parameters α and β must be positive values. The beta distribution can be used to analyze probabilistic experiments that only have two possible outcomes: success with probability x or failure with probability 1 – x.The beta probability distribution is actually very close to the binomial distribution. The binomial distribution models the number of successes divided by the number of trails. Beta models the likelihood of success in Bernoulli trials and captures the uncertainty. A Bernoulli trial is an experiment with two possible outcomes such as success or failure.

Normal probability function example:

Chart, histogram

Description automatically generated

Gamma probability function example:

A picture containing diagram

Description automatically generated

Beta probability function example:

A picture containing diagram

Description automatically generated

References

<https://www.sciencedirect.com/topics/mathematics/normal-probability-distribution>

<https://www.scribbr.com/statistics/normal-distribution/>

<https://statisticsbyjim.com/probability/gamma-distribution/>

<https://www.statlect.com/probability-distributions/beta-distribution>

<https://statisticsbyjim.com/probability/beta-distribution/>