Continuous Probability

Difference between Discrete and Continuous Probability

- Discrete Probability: Deals with events that occur in a distinct or separate manner.
- Continuous Probability: Describes the probabilities of possible outcomes for events that can take on a continuous range of values.

Probability Density Function (PDF)

The **probability density function** (pdf) of a continuous random variable is a function that describes the relative likelihood for this random variable to take on a given value. We write the pdf as

$$f_X(x)$$

- f denotes a function
- X is the random variable for which the pdf is for
- x is an argument

Probability Density Function (PDF)

The pdf, unlike the probability mass function for discrete distributions, does not correspond to probabilities directly.

$$Pr(X = a) \neq f_X(a)$$

In fact, if X is a continuous random variable, then $Pr(X = a) = 0$ for any a. Instead, we find probabilities of a region.

Probability Density Function (PDF)

The probability that the random variable falls within a particular range can be found by integrating the PDF over that range.

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

*Note: Calculus is not required for this class. We will use software to find probabilities. But just so you know, continuous probabilities are found analytically using calculus

Cumulative Distribution Function (CDF)

The **cumulative distribution function** (CDF) describes the probability that a random variable X will take a value less than or equal to x. It is written as

$$P(X \leq x) = F_X(x) = \int_{-\infty}^{x} f(t)dt$$

CDFs are also. very useful for finding probabilities of ranges.

$$P(a \leq X \leq b) = F_X(b) - F_X(a)$$

Normal Distribution

The normal distribution is a continuous probability distribution that is symmetrical around its mean, usually denoted as μ, and is characterized by its standard deviation, denoted as σ.
The probability density function of the normal distribution is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Uses:

- Modeling natural phenomena like heights of individuals.
- In statistical inference.
- As a foundational block in various statistical methods.

Normal Distribution

We can write the normal distribution using

$$X \sim N(\mu, \sigma)$$

This means that X is a normal random variable with mean μ and standard deviation σ.

Normal Distribution

A **parameter** of a distribution is an unknown quantity that is derived from the population represented by the distribution. We've already seen a parameter in the Binomial and Geometric distributions, p. For the normal distribution, the parameters are μ and σ (or σ^2).

Often a goal of a statistical model is to estimate a parameter. Because a parameter is derived from a population, if we can estimate the parameter well, it can often tell us something about the population.

Regression Model and Normal Distribution

In regression analysis, we often assume that the errors (or residuals) follow a normal distribution.
For a simple linear regression model:

$$Y \sim N(\beta_0 + \beta_1 X, \sigma)$$

This essentially replaces μ with $\beta_0 + \beta_1 X$. What are the parameters of a regression model?

Visualizing the Normal Distribution PDF

```
import matplotlib.pyplot as plt
import numpy as np

# Generate a range of values for X
x = np.linspace(-4, 4, 1000)

# Parameters for the normal distribution
mu = 0  # mean
sigma = 1  # standard deviation

# Generate the corresponding pdf values
y = (1/(sigma*np.sqrt(2*np.pi))) * np.exp(-0.5*((x-mu)/sig

# Plot the distribution
plt.figure()
plt.plot(x, y)
```

Using Python to Find Probabilities

We can use Python functions to find probabilities for continuous distributions. Here, we use the normal distribution as an example.

```
from scipy.stats import norm

# Define the parameters for the normal distribution
mean = 0
std_dev = 1

# Calculate the probability that X is between -1 and 1
probability = norm.cdf(1, mean, std_dev) - norm.cdf(-1, mea
```