STA 3180 Statistical Modelling: Distributions

Distributions

STA 3180 Statistical Modelling

Lecture Notes

Introduction:

Distributions are a fundamental concept in statistics. They provide a way to measure the probability of an event occurring and help us to understand the underlying structure of data. Distributions can be used to make predictions about future events, as well as to compare different sets of data. In this lecture, we will discuss the different types of distributions, how they are used in statistical modelling, and how to interpret them.

Types of Distributions:

There are many different types of distributions, each with its own characteristics and uses. The most common types of distributions are the normal, binomial, Poisson, exponential, and chi-square distributions.

Normal Distribution:

The normal distribution is one of the most commonly used distributions in statistics. It is a symmetric, bell-shaped distribution that is defined by its mean and standard deviation. The mean is the average value of the data, while the standard deviation is a measure of how spread out the data is. The normal distribution is used to model many real-world phenomena, such as height, weight, and IQ scores.

```
Start of Code
# Generate a normal distribution with mean = 0 and standard deviation = 1
import numpy as np
mean = 0
std = 1
x = np.random.normal(mean, std, 1000)
End of Code
```

Binomial Distribution:

The binomial distribution is a discrete probability distribution that is used to model the number of successes in a series of independent trials. It is defined by two parameters: the probability of success (p) and the number of trials (n). The binomial distribution is often used to model the probability of an event occurring a certain number of times in a given number of trials.

```
Start of Code
# Generate a binomial distribution with p = 0.5 and n = 10
import numpy as np
```

```
p = 0.5
n = 10
x = np.random.binomial(n, p, 1000)
End of Code
```

Poisson Distribution:

The Poisson distribution is a discrete probability distribution that is used to model the number of events occurring in a given interval of time or space. It is defined by one parameter: the average number of events per interval (λ). The Poisson distribution is often used to model the probability of an event occurring a certain number of times in a given interval.

```
Start of Code
# Generate a Poisson distribution with lambda = 5
import numpy as np
lam = 5
x = np.random.poisson(lam, 1000)
End of Code
```

Exponential Distribution:

The exponential distribution is a continuous probability distribution that is used to model the time between events occurring in a given interval. It is defined by one parameter: the average time between events (λ). The exponential distribution is often used to model the probability of an event occurring after a certain amount of time.

```
Start of Code
# Generate an exponential distribution with lambda = 5
import numpy as np
lam = 5
x = np.random.exponential(lam, 1000)
End of Code
```

Chi-Square Distribution:

The chi-square distribution is a continuous probability distribution that is used to model the sum of squares of independent random variables. It is defined by one parameter: the degrees of freedom (df). The chi-square distribution is often used to test the goodness of fit of a model to data.

```
Start of Code
# Generate a chi-square distribution with df = 5
import numpy as np
df = 5
x = np.random.chisquare(df, 1000)
End of Code
```

Interpreting Distributions:

Once a distribution has been generated, it can be interpreted in terms of its shape, central tendency, and spread. The shape of the distribution can be used to identify whether it is normal, skewed, or bimodal. The central tendency can be measured using the mean, median, and mode, while the spread can be measured using the range, interquartile range, and standard deviation.

Practice Multiple Choice Questions:

Q1. Which of the following is NOT a type of distribution?

A. Gaussian

B. Binomial

C. Logistic

D. Poisson

Answer: A. Gaussian

Explanation: The Gaussian distribution is also known as the normal distribution, which is a type of distribution.