Notes on Statistical Inference

George Baker

Statistical inference is the process of using data analysis to infer properties of an underlying distribution of probability.

Inferential statistical analysis infers properties of a population, for example by testing hypotheses and deriving estimates. It is assumed that the observed data set is sampled from a larger population.

Frequentist (Classical) Inference

Frequentist inference is a type of statistical inference that draws conclusions from sample data by emphasizing the frequency or proportion of the data.

E.g. in a poll, the percentage of voters θ who will vote for a particular option can be **inferred** by randomly sampling a large proportion of voters and calculating the **proportion** of votes for that option.

Bayesian Inference

Bayesian Inference is a type of statistical inference in which Bayes' theorem (named after Reverend Thomas Bayes) is used to update the probability for a hypothesis as more evidence or information becomes available.

In Bayesian Inference, the unknown quantity θ is assumed to be a random variable.

Random Sampling

A random sample is a subset of individuals (a sample) chosen from a larger set (a population). Each individual is chosen randomly and entirely by chance, such that each individual has the same probability of being chosen at any stage during the sampling process, and each subset of k individuals has the same probability of being chosen for the sample as any other subset of k individuals.

With Replacement

Choosing a particular member of the population more than once is allowed to happen.

This means that each member of the sample is chosen **independently** which simplifies the analysis.

When a population is very large, the probability of choosing a particular member more than once is incredibly low.

Without Replacement

In small populations and often in large ones, such sampling is typically done without replacement, i.e., one deliberately avoids choosing any member of the population more than once.

Point Estimation

Point estimation involves the use of sample data to calculate a single value (known as a point estimate since it identifies a point in some parameter space) which is to serve as a "best guess" or "best estimate" of an unknown population parameter (for example, the population mean).

More formally, it is the application of a point estimator to the data to obtain a point estimate.

Example

After creating multiple samples of a larger population, we can use a point estimator to estimate a parameter of that population.

E.g. after sampling the population of Birmingham several times (with replacement), we could represent the expectation (sample mean) of heights for each sample as $X_1, X_2, ..., X_n$.

Since we have several unknown quantities θ (random variables in Bayesian Inference), we can use $\hat{\Theta}$ to represent the combined average height of all the samples:

$$\hat{\Theta} = \bar{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

Where the random variable $\hat{\Theta}$ is a point estimator.

Definition

After collecting random samples $X_1, X_2, ..., X_n$, which all have the same type of distribution, a point estimator $\hat{\Theta}$ for an unknown parameter θ is defined as a function of the random sample:

$$\hat{\Theta} = h(X_1, X_2, ..., X_n)$$

where h can be any function e.g. mean, variance, etc.

Evaluating Point Estimators

We define three main desirable properties for point estimators:

1. Bias

The bias of an estimator $\hat{\Theta}$ tells us how far on average $\hat{\Theta}$ is from the real value of θ.

Let $\hat{\Theta} = h(X_1, X_2, ..., X_n)$ be a point estimator for θ. The **bias** of a point estimator $\hat{\Theta}$ is defined by:

$$B(\hat{\Theta}) = E[\hat{\Theta}] - \theta$$

In general, we would like to have a bias value that is close to 0, indicating that $\hat{\Theta}$ is close to θ.

2. Mean Squared Error (MSE)

The mean squared error (MSE) of an estimator $\hat{\Theta}$ tells us the average of the squares of the errors, i.e. the average squared difference between the estimated values and the actual value.

The mean squared error (MSE) of a point estimator $T\hat{h}eta$, shown as $MSE(\hat{\Theta})$, is defined by:

$$MSE(\hat{\Theta}) = E[(\hat{(}\Theta) - \theta)^2]$$

$\hat{\Theta}$ is the error that we make when we estimate θ by $\hat{\Theta}$. Thus, the MSE is a measure of the distance between $\hat{\Theta}$ and θ and a smaller MSE is generally indicative of a better estimator.

3. Consistency

We say that an estimator is consistent if as the sample size n gets larger, $\hat{\Theta}$ converges to the real value of θ.

Let $\hat{\Theta}_1, \hat{\Theta}_2, ..., \hat{\Theta}_n$ be a sequence of point estimators of θ. We say that $\hat{\Theta}_n$ is a **consistent** estimator of θ if:

$$\lim_{n\to\infty} P(|\hat{\Theta}_n - \theta| \geq \epsilon) = 0$$

for all $\epsilon > 0$.