Normal distribution

Multivariate normal distribution

Population parameters – the parameters that define the statistical distribution fitted to the entire population, e.g. population mean and population standard deviation for a population that is normally distributed.

Given a sample from the population, we want to estimate the population parameters (like how we want to estimate the model parameters from our training data in machine learning). We want to estimate population parameters to make inferences about the population based on the sample. We use these estimates to generalize findings from the sample to the entire population. Population parameters are used to develop predictive models that can be used to forecast future outcomes or behaviors.

We estimate the population parameters from the sample parameters. The more data we have, the more confidence we have in the estimates.

P-values and confidence intervals quantify the confidence we have in population parameter estimates.

Given two different samples, we can use statistics to quantify our confidence in how different they are. P-values and confidence intervals can tell us if the differences in two samples are statistically significant. This is called hypothesis testing.

Let represent the observed values/measurements. Let there be measurements (can represent population or sample size).

In statistics, denotes the sample mean (aka estimated population mean) while denotes the population mean. In either case, mean is the average of all measurements.

Population variance is given by . It is the average of the squared differences b/w the measurements and the mean. It measures the variance of around the mean.

Population standard deviation is .

Sample variance is . Sample standard deviation is .

compensates for the fact that we are calculating the differences from the sample mean instead of population mean. Otherwise, we would consistently underestimate the population variance.

A model explores the relationship between different attributes. We use statistics to determine how useful or reliable our model is.

Hypothesis testing

MLE

MAP

Confidence interval

Distributions

Descriptive stats like mean, stdev, variance, correlation, covariance

t-tests, f-tests, ANOVA, chi^2

Expected value

Integrating over a distribution

Calculating mean of a distribution (integral of x\*P(x) I think)

Central limit theorem – this explains why we often see the normal distribution in nature.

is a random variable (a random process, where each outcome is associated with a number).

Add samples of this variable,

The distribution of this sum looks more like a normal distribution as .

<https://www.youtube.com/watch?v=zeJD6dqJ5lo&t=255s>