# Notes on Bayesian Analysis of Linear Models

compiled by D.Gueorguiev 4/15/24

## The Parametric Inference Problem

Let be a vector of real parameters , a vector of observations, is a known design matrix with size . Then the general linear model is

(1)

where where denotes the normal distribution. Here is the precision matrix of , which has covariance matrix , and is unknown.

This is the general linear model and our objective is to provide inferences for and when observing , where is the -th observation. Inference is the procedure which extracts information about from the sample .

In Bayesian models all references are based on the posterior distribution of , which is given by Bayes theorem.

### Bayes Theorem

Suppose the prior information about is represented by a probability density function , then Bayes theorem combines this information with the information contained in the sample. The likelihood function for and is

(2)

where, as before, . The likelihood function is the inherent sample information about the parameters and and is the conditional density function of the sample random variables given the sample .

Bayes theorem gives the conditional density of and , given the sample .

(3)

The posterior density of the pair is and represents one’s knowledge of and after observing the sample . Recall that the information about of and , before is observed is contained in the prior density . Note that the posterior density is denoted with the same symbol as the prior but with the conditional notation for as - this is just a notational convention which is adopted in the current text.

One can rewrite the posterior density as

, (4)

where K is the normalizing constant and is given by

which is the marginal probability density of .

Next, we will find the posterior density of and .

### Prior Information

The prior information about the parameters and is given in two ways. The first is when is a normal-gamma prior density, namely,

, (5)

where

, (6)

and is a element vector and is a known positive definite matrix. Thus is the conditional density of given and it is normal with mean vector and precision matrix .

The marginal prior density of is Gamma with parameters and

, (7)

We can construct the marginal prior density of as

, (8)

which is a Student t density

## Appendix

### Chi-squared distribution

The chi-squared () distribution with degrees of freedom is the distribution of the squares of independent standard normal random variables.

If are independent standard normal variables, then the sum of their squares

is distributed according to the chi-squared distribution with k degrees of freedom. This fact is usually denoted with

or

The chi-squared distribution has one parameter: a positive integer that specifies the number of degrees of freedom.

The Probability Density Function of chi-squared distribution is given with

, ; for it is zero

The Cumulative Distribution Function of chi-squared distribution is given with

where is the lower incomplete gamma function. is the regularized gamma function.

Additional Notes on chi-squared distribution

The reason why chi-squared distribution is used in hypothesis-testing is its relation to normal distribution.

Many hypothesis tests use the -statistic in a -test. As the sample size increases the sampling distribution of the test statistic approaches normal distribution by the central limit theorem. The square of standard normal distribution is distributed according to the chi-square distribution.

An additional use case of the chi-squared distribution is to model the large sample distribution of generalized likelihood ratio tests (LRT).

### Erlang distribution

### Gamma distribution

The exponential distribution, the chi-squared distribution and the Erlang distribution are special cases of Gamma distribution

### Location tests and Student t-test

**Definition** A *location test* is a statistical hypothesis test that compares the location parameter of a statistical population to a given constant, or that compares the location parameters of two statistical populations to each other. Most commonly, the location parameter (or parameters) of interest are expected values but location tests based on medians or other measures of location are also used.

**Definition** *One-sample location test* compares the location parameter of one sample to a given constant. In a one-sided test, it is stated before the analysis is carried out that it is only of interest if the location parameter is either larger than or smaller than the given constant, whereas in a two-sided test, a difference in either direction is of interest.

**Definition** *Two-sample location test* compares the location parameters of two samples to each other. A common situation is where the two populations correspond to research subjects who have been treated with two different treatments (one of them possibly being a control or placebo). In this case, the goal is to assess whther one of the treatments typically yields a better response than the other. In a one-sided test, it is stated before the analysis is carried out that it is only of interest if a particular treatment yields the better responses, whereas in a two-sided test, it is of interest whether either of the treatments is superior to the other.

**Definition** *Student -test* is a location test, used to determine whether the difference between two groups of responses is statistically significant. It is most commonly applied when the test statistic follows the normal distribution if the value of the scaling term in the test statistic is estimated from data. The -test most common application is to test if the means of two populations are statistically different.

**Definition** one-sample -test

One-sample t-test is a location test of whether the mean of a population has a value specified in the null hypothesis. In testing the null hypothesis that the population mean is equal to a specified value one uses the statistic

where is the sample mean, is the sample standard deviation

### Student -distribution

Student’s -distribution (or simply ) is a continuous distribution which generalizes the standard normal distribution. Like the latter it is symmetric around zero and bell-shaped. However, has heavier tails and the amount of probability mass in the tails is controlled by the parameter .

The probability density function of is given with

For becomes the Cauchy’s distribution with and :

because . Notice that the Cauchy’s distribution has fat tails. When increases toward infinity converges to the normal distribution which has thin tails.

### Multivariate t-distribution

Let is a -dimensional vector distributed according to . Let u

## References

Bayesian Analysis of Linear Models, Lyle D. Bromeling, CRC Press, 2019 (Copyright 1985)