Fundamentals of Statistical Modeling (VT20)

Andrea Discacciati
Karolinska Institutet
Stockholm, Sweden

Lab 1

Load the dataset and the `mlci` command

{{1}}

Exercise 0 (together to get started)

We assume that $f(y_n)$ is normal. Estimate the parameters μ and σ.

{{2}}

Exercise 1

Some of the distributions we'll work with in the course/labs are:

- Normal: `https://en.wikipedia.org/wiki/Normal_distribution`
- Skew-normal: `https://en.wikipedia.org/wiki/Skew_normal_distribution`
- Gamma: `https://en.wikipedia.org/wiki/Gamma_distribution`
- Beta: `https://en.wikipedia.org/wiki/Beta_distribution`
- Exponential: `https://en.wikipedia.org/wiki/Exponential_distribution`
- Chi-squared: `https://en.wikipedia.org/wiki/Chi-squared_distribution`
- Weibull: `https://en.wikipedia.org/wiki/Weibull_distribution`
- Bernoulli: `https://en.wikipedia.org/wiki/Bernoulli_distribution`
- Negative binomial: `https://en.wikipedia.org/wiki/Negative_binomial_distribution`

Get familiar with the distributions above.

- Are they useful to model continuous or discrete variables?
- What's their support?
- By how many parameters are they parametrized?
- What does their shape look like?

Exercise 2

We assume that $f(y_n)$ is normal. Estimate the parameters μ and σ. Use the `normalden()` function.

{{3}}

Estimate again the parameters μ and σ, but this time constrain the parameter σ to be positive by replacing $\sigma = \exp(\theta)$. You'll get now an estimate of θ. Recover the MLE of σ. (Note that the MLE $\hat{\sigma}$ hasn't changed – see slide 33 – but this "trick" improves computational stability).

{{4}}

Plot the estimated density $\hat{f}(y_n)$ over the sample histogram.

{{5}}

Plot the model-based estimated CDF $\hat{F}(y_n)$ (see slide 38) over the empirical CDF (see slide 20). The function `normal()` returns the CDF of a standard normal distribution.

{{6}}

Exercise 3

Assume that $f(y_g)$ is (2-parameter) gamma. Estimate the parameters α and β using the `gammaden()` function. Fix the location parameter g (the third argument of the `gammaden()` function), equal to 0.

{{7}}

Plot the estimated density $\hat{f}(y_g)$ over the sample histogram.

{{8}}

Exercise 4

We assume that $f(y_b)$ is beta. Estimate the parameters α and β using the `betaden()` function.

{{9}}

Plot the estimated density $\hat{f}(y_b)$ over the sample histogram. (OBS: We need to be careful when plotting the estimated density function close to the boundaries of the distribution's support).

{{10}}

Exercise 5

We assume that $f(y_e)$ is exponential. Estimate the parameter λ. You'll need to code the density of an exponential distribution yourself: $f(y; \lambda) = \lambda \exp(\lambda y)$

{{11}}

Plot the estimated density $\hat{f}(y_e)$ over the sample histogram.

{{12}}

Exercise 6

We assume that $f(y_c)$ is chi-squared. Estimate the parameter k using the `chi2den()` function.

{{13}}

Plot the estimated density $\hat{f}(y_c)$ over the sample histogram.

{{14}}

Exercise 7

Install the `qplot` command (you need to be connected to the Internet)

{{15}}

Let's go back to the normal distributed variable (Exercise 1). Assume that $f(y_n)$ is normal and estimate the parameters μ and σ. Generate the transform $u = \hat{F}(y_n)$ as seen during the lecture this morning. Draw the estimated quantile plot using the `qplot` command.

{{16}}

Assume now that $f(y_n)$ is exponential and estimate the parameter λ. Generate the transform $u = \hat{F}(y_n)$. Draw the estimated quantile plot using the `qplot` command.

{{17}}

Comment the two quantile plots.

Exercise 8

Assume that y_{ber} follows a Bernoulli distribution. We want to estimate the probability of "success" ($y_{ber} = 1$). Estimate the probability η while constraining it to be bounded between 0 and 1. First, write down the likelihood by hand. Then, use the `binomialp()` function.

Are the results you obtain identical to those obtained from logistic regression?

{{18}}

Now simplify the likelihood as shown in the slides. Check, again, that the estimate of η is the same.

{{19}}