Exercise week 4

Gherardo Varando, gherardo.varando@math.ku.dk

12/12/2018

A trigonometric density

Load the data in the file `angles.txt`, the data have been generated from a density $f(x|k) \propto sin(x)^k$ in the interval $[0, \pi]$. In week 2 we wrote the density, CDF and quantile functions for this statistical model. In week 3 we found the MLE estimator of the parameter k.

Ex 1

Ex 1.1 Build a 99% confidence interval for k based on the data in `angles.txt`, how you can estimate the standard error? (try both parametric and non-parametric bootstrap)

Ex 1.2 Test if $k > 10$ at a confidence level $\alpha = 0.05$ for the data in the `angles.txt` file (you can use Wald test $H_0 : k \leq 10$)

A case study of neuronal data

We continue the case study of the ISI data in the `neuronspikes.txt` file. In week 3 we estimated parameters of exponential, gamma and inverse Gaussian distributions with MLE.

Ex 2

Ex 2.1 The exponential distribution is a special case of the gamma distribution when the shape parameter is equal to 1. Check this fact graphically in R.

Ex 2.2 Since the exponential model is nested in the gamma model we can perform the likelihood ratio test to select between the two models

Ex 3 We consider in total four candidate models for the ISI data: exponential, gamma, inverse Gaussian and log-normal.

Ex 3.1 Find the MLE for each model

Ex 3.2 Perform model selection using AIC and BIC

Brain cell dataset

We continue here the study of the brain cell dataset from the Allen Institute. As usual load the data from the file `cell_types.csv`.

Here we continue studying the *ramp spike time* observations.

Ex 4

Ex 4.1 Estimate the standard error of the MLE estimator of μ for the log-normal distribution applied to the *ramp spike time* data.

Ex 4.2 Obtain a 95% confidence interval for the parameter μ (try different methods)

Ex 4.3 Obtain again a 95% confidence interval for the parameter μ using only the *human* cells.

Ex 5

Ex 5.1 Transform the ramp spike time using the logarithm as we did in Week 3 and then perform a two sample t-test between the human and mouse cells. You can also try to use the argument `var.equal = TRUE` or `var.equal = FALSE`, check the documentation of `t.test`.

Ex 5.2 Perform directly a Wald test to check if $\mu_h = \mu_m$ where μ_h and μ_m are the mean-log parameters of the log-normal distributions for the human and mouse cells.

Different tests

Given two sets of observations

$$X_1, \ldots, X_m$$

$$Y_1, \ldots, Y_n$$

We want to test if the two data sets have equal mean. We moreover assume that both data sets are normally distributed and have equal variance.

$$X_i \sim N(\mu_1, \sigma^2) \quad Y_i \sim N(\mu_2, \sigma^2)$$

This is the classical scenario where we can apply the two-sample t-test.

Exercise 6

Ex 6.1 Simulate two groups of i.i.d. data following two normal distributions. For example `x = rnorm(20, 2, 4)` and `y = rnorm(40, 2.5, 4)`

Ex 6.2 Compute the p-value of the two-sample t-test with equal variance

Ex 6.3 Compute the Wald test for $H_0 : \mu_1 = \mu_2$ you can use the statistic $\delta = \overline{X} - \overline{Y}$. δ is obviously Gaussian distributed (try to prove it) and its standard error can be obtained analytically (and thus an estimator $\hat{se}(\delta)$), otherwise use bootstrap.

Ex 6.4 Perform the likelihood ratio test for $H_0 : \mu_1 = \mu_2$

Ex 6.5 Compare the results obtained in the different tests, in particular report the three p-values.