Exercise week 2

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

27/11/2019

1 Brain cell dataset

We continue here the study of the brain cell dataset from the Allen Institute.

Ex 4

Ex 4.1 In the `specimen__hemisphere` variable is marked if the cell belongs to the left or right hemisphere. Plot histograms for the distribution of the *ramp spike time* variable in the left and right hemisphere separately.

Ex 4.2 Plot now the kernel density estimations for the *ramp spike time* variable in the left and right hemisphere in the same plot, use as usual different colors and plot a legend. Can you observe any difference?

Ex 4.3 Repeat Ex 4.2 just for the human cells, differentiating now between male and female cells (not left and right hemisphere).

Ex 5

Ex 5.1 In exercise 6.5 of week 1 we found values for the parameters of the log-normal distribution that fit the *ramp spike time* distribution. Compare now the theoretical and empirical quantiles (Q-Q plot).

Ex 5.2 (A wrong guess) We assume that the distribution of the *ramp spike time* is a Gaussian. The best parameters of the Gaussian are then (we will see it next week) `mean = 6.41` and `sd = 4.32` (empirical mean and standard deviation). Plot the Gaussian density on top of the histogram of the ramp spike time (for all the cells) and comment why the choice of the distribution is incorrect.

Ex 5.3 Obtain the Q-Q plot for the *ramp spike time* against the Gaussian distribution (with parameters as in 5.2).

2 Empirical mean and variance

Ex 6 In this exercise we are going to test the fact that the empirical mean is an estimator of the true mean value.

Ex 6.1 Sample n values of a Binomial random variable (parameter size = 100, prob = 0.3), and compute the empirical mean, empirical variance and empirical standard deviation. Repeat it for $n = 10, 100, 1000, 10000$. Compare the empirical values and the true values, you can also make a plot.

Ex 6.2 For a given value of n fixed, repeat the sampling and estimation of Ex 6.1 a large number of times (e.g. 1000). So you will obtain 1000 empirical means (and variances and sd). Plot the histogram of the empirical means. We can also obtain the variance and standard deviation of the empirical means.

Ex 6.3 Compute the standard error of the mean using the formula we have seen in class and compare it with the value of the standard deviation of the empirical means obtained in Ex 6.2.

Ex 6.4 On top of the histogram plotted in Ex 6.2, draw the density of Gaussian distribution with parameters mean = 100×0.3 and sd equal to the sem, check the results for different values of n.

Hint: Since from Ex 6.2 we are fixing a value of n, the easier solution is to write Ex 6.2, 6.3, 6.4 as functions of n.