POLI706: Advanced Methods of Political Analysis

Problem set 7

Exercise 1

Explain how likelihood and probability differ in terms of hypothesis and data.

Exercise 2

Let's assume that the probability density function (pdf) follows a normal distribution, $N(\theta, \theta)$. Here, we consider a restricted form of the normal distribution where both the mean and variance are equal to θ. Therefore, the form of the pdf is as follows:

$$f(x;\theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\Big(-\frac{(x-\theta)^2}{2\theta}\Big)$$

Load `australia_temp.csv` data into RStudio. This sample data shows temperatures in Wollongong at 9 a.m. and 3 p.m between 2008 and 2020. We'll use the temperature at 9 a.m. Plot the log-likelihood for a reasonable range of θ in R.

a. Set the reasonable range of θ.

b. 1. Extract the data of temperature at 9 a.m. in Wollongong at 9 a.m. and set up a function called `ll` using the `function()` construct in R. This function should take two arguments: `data` and `theta`. Note that `data` should be a vector, as it will be summed over in the function. The other argument, `theta`, should be a scalar.
   2. For the return value, use the `dnorm()` function to obtain the values of the probability density function (pdf). Normally, you would need to use the entire form of the provided pdf or calculate the log of this value yourself. However, since we are using a normal distribution with different parameters, you can use the `dnorm()` function with the mean and variance parameters. By setting `log = TRUE`, the function will automatically return the log of the pdf value, which is very convenient.
   3. The likelihood function is the product of the pdf values for all `x`, and taking the log converts this product into a sum. It is crucial to maintain the correct order of operations. Specifically, sum the log-pdf values for all data points after calculating the log of the pdf using `dnorm()`.

c. However, at this point, we have only defined the function; we have not actually calculated the likelihood function. Now, let's apply this function to our data to create the likelihood function, or more precisely, the log-likelihood function. Use the function you coded in 2b to plot the log-likelihood for a reasonable range of theta values using the `ll(data, theta)` function in R.

1. Use the `which.max()` function to find the location where the log-likelihood reaches its maximum value.
2. Plot the log-likelihoods. In the `plot()` function, the first argument should be the values for the x-axis, and the second argument should be the values for the y-axis, and both should have the same length. Therefore, place the `theta` vector on the x-axis and the log likelihood variable `ll` on the y-axis, then plot the log likelihood with respect to `theta`.
3. Explain what the plot means.