Exercises for Lecture 10

Statistical Computing with R, 2023-24

Exercise 1

Consider the function $f(x) = x^3 - 2x + 5$.

1. Compute $f'(x)$ and study its sign.
2. Does $f(x)$ have a global minimum and / or a global maximum? Why?
3. Does $f(x)$ have local minima and / or a local maxima? Why?
4. Apply `optimize( )` on the intervals $[-5, 0]$ and $[0, 5]$ to check whether your answers to points (2) and (3) are correct.

Exercise 2

Let X follow a Poisson distribution with probability mass function

$$f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!},\ \lambda > 0.$$

1. Write the likelihood function $L(\lambda)$ for a sample $(x_1, x_2, ..., x_n)$.
2. Compute the log-likelihood $\ell(\lambda)$.
3. Find the maximum likelihood estimator of λ analytically.
4. Write an `R` function that evaluates $\ell(\theta)$.
5. Use `optimize( )` to compute the maximum likelihood estimates of λ for the following two sample:

```r
x1 = c(9, 7, 7, 8, 10, 5, 8, 4, 3, 5, 7, 7, 9, 6)
set.seed(10)
x2 = rpois(300, lambda = 3)
```

6. Compare the estimates you obtained using `optimize( )` to the ones obtained using the formulas that you derived at (3). Do the results match?

Exercise 3

The Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ has the following density function:

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \;\; x > 0, \tag{1}$$

which can be evaluated in `R` using the function `dgamma(x, shape = alpha, scale = beta)` (see `?dgamma` for more details).

1. Write the log-likelihood $\ell(\alpha, \beta)$ for a sample $(x_1, x_2, ..., x_n)$.
2. Write an `R` function that evaluates $\ell(\alpha, \beta)$.

Let `x = iris$Petal.Width`, and suppose that this variable follows a Gamma distribution.

3. Use the `optim( )` function to compute the maximum likelihood estimates of (α, β) for `x` using the Nelder-Mead algorithm (treating the constrained problem as if it was unconstrained) and the L-BFGS-B algorithms.
4. Turn the constrained problem into an unconstrained optimization problem by setting $a = \log(\alpha)$ and $b = \log(\beta)$, and use the Nelder-Mead algorithm to solve it. Do not forget to transform the solution back to the original scale (α, β) of the problem!

Exercise 4

In Lecture 8 we compared the distribution of the variable `Sepal.Length` in the `iris` dataset between the 3 `Species` of iris plants. For the rest of the exercise, we will assume that the distribution of `Sepal.Length` in each group (X = setosa, Y = versicolor, Z = virginica) follows a normal distribution with different mean and variance:

$$X \sim N(\mu_X, \sigma_X) \;\; Y \sim N(\mu_Y, \sigma_Y) \;\; Z \sim N(\mu_Z, \sigma_Z).$$

1. Use the `split( )` function to separate the `iris` data frame by `Species` (see Lecture 8).
2. Compute 95% confidence intervals for μ_X, μ_Y and μ_Z.
3. Fix $\alpha = 0.01$ and test the null hypothesis that $\mu_X = 5$. What do you conclude?
4. Test the null hypothesis that $\mu_Y = 5$ with $\alpha = 0.05$. What do you conclude?
5. Set $\alpha = 0.01$ and test $H_0 : \mu_Y = \mu_Z$ vs $H_1 : \mu_Y \neq \mu_Z$. Do the data provide sufficient evidence to reject H_0?