Computational Statistics and Data Analysis (MVComp2)

Exercise 5

Lecturer Tristan Bereau

Semester Wi23/24
Due Nov. 23, 2023, 23:59

1 College achievement test, again (2 points)

The time required to complete a college achievement test follows an unknown distribution with mean 70 minutes and standard deviation 12 minutes. The test is terminated after 90 min. Is that enough time to allow 90% of the students to complete the test?

2 Microscope (2 points)

You look into a microscope to observe N cells at locations $\{(x_n, y_n)\}$. You would like to infer the field of view of the microscope. Assume the field of view is rectangular and that the cells' locations are independently and uniformly distributed. Use maximum likelihood to infer values of $(x_{\min}, y_{\min}, x_{\max}, y_{\max})$.

3 Fisher matrix for linear fitting (4 points)

Suppose you're fitting a linear model, $f_k(\theta) = ax_k + b$, where $\theta = (a, b)$, and you can only measure two data points. At what values of x would you choose to measure? While you may intuitively place them as far as possible, we will see that this is not necessarily the best strategy.

(a) You are given N iid data points $\boldsymbol{x} = \{x_1, \dots, x_N\}$, as well as a model for the data, f, with parameters, θ, such that $f_k(\theta)$ predicts the k^{th} data point. Assume that the model leads to Gaussian errors: deviations between data points x_k and their expected values, $f_k(\theta)$, follow a Gaussian distribution with mean 0 and standard deviation σ_k. Recall the expression for the Fisher information matrix

$$F_{ij} = -E\left[\frac{\partial^2 \log \mathcal{L}(\boldsymbol{x}|\theta)}{\partial\theta_i \partial\theta_j}\right],$$

where $\mathcal{L}(\boldsymbol{x}|\theta)$ is the likelihood function. Show that the Fisher matrix is given by

$$F = \begin{bmatrix} \frac{x_1^2}{\sigma_1^2} + \frac{x_2^2}{\sigma_2^2} & \frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2} \\ \frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2} & \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \end{bmatrix}.$$

(b) Say you place your first point at $x_1 = 1$, and assume that $\sigma_1 = \sigma_2$. Use the expression derived in (a) to show that you should place your other point at $x_2 = -1$.

4 Linear fit, continued (2 points)

Continuing on problem 3, we would like to determine where to place a *third* point so as to improve on the estimation of the parameters. This third point has identical standard deviation to the others, $\sigma_1 = \sigma_2 = \sigma_3$. Intuitively you suggest to place it at $x_3 = 0$. Write a script to compute the covariance matrix. Does this third point help you improve your confidence about the slope and/or the intercept of f?