Computational Statistics and Data Analysis (MVComp2)

Solutions to 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?

1.1 Solution

Use Chebyshev's inequality:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

We want to ensure that at least 90% of the students can complete the test, which means we want

$$1 - \frac{1}{k^2} \geq 0.90.$$

Solving for k we get $k \geq \sqrt{10} = 3.16$.

We are concerned with the upper tail of the distribution, leading to $\mu + k\sigma \approx 108$ min. This is larger than 90 min, which may not be long enough 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 $(x_{\min}, y_{\min}, x_{\max}, y_{\max})$.

2.1 Solution

Denote the edges of the field of view by a and b on the x axis and c and d on the y axis. We consider the likelihood function, L, for a collection of N points uniformly distributed on a two-dimensional space, $\boldsymbol{X} = \{(x_1, y_1), \dots, (x_N, y_N)\}$. Independence allows us to factorize the likelihood terms

$$L_{\boldsymbol{X}}(a, b, c, d) = \prod_{i=1}^{N} \frac{1}{b-a} \mathbb{1}_{[a,b]}(x_i) \frac{1}{d-c} \mathbb{1}_{[c,d]}(y_i),$$

where $\mathbb{1}_{[a,b]}(x_i)$ is the indicator function (i.e., it is 1 in the subset $x_i \in [a, b]$ and 0 otherwise).

It is more convenient to work with the log-likelihood, l,

$$l_{\boldsymbol{X}}(a, b, c, d) = -N \log(b-a) - N \log(d-c).$$

We compute the partial derivative for each parameter

$$\begin{aligned}
\frac{\partial l}{\partial a} &= \frac{N}{b-a} \\
\frac{\partial l}{\partial b} &= -\frac{N}{b-a} \\
\frac{\partial l}{\partial c} &= \frac{N}{d-c} \\
\frac{\partial l}{\partial d} &= -\frac{N}{d-c}.
\end{aligned}$$

Taking for instance parameter a, we can see that the derivative of the log-likelihood is positive and monotonically *increases* with increasing values of a. Thus the MLE will be largest at the largest possible value. On the other hand, any sample that falls *outside* the interval $[x_{\mathrm{min}}, x_{\mathrm{max}}]$ (resp. for y axis) will lead to a likelihood of zero. Thus we obtain:

$$\begin{aligned}
a = x_{\mathrm{min}} &= \min(x_1, \dots, x_N) \\
b = x_{\mathrm{max}} &= \max(x_1, \dots, x_N) \\
c = y_{\mathrm{min}} &= \min(y_1, \dots, y_N) \\
d = y_{\mathrm{max}} &= \max(y_1, \dots, y_N)
\end{aligned}$$

3 Fisher matrix for linear fitting (4 points)

Suppose you're fitting a linear model, $f_k(\theta) = a x_k + b$, 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

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

(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 $x_2 = -1$.

3.1 Solution

(a)

The likelihood function can be written as

$$\mathcal{L}(\boldsymbol{x}|\theta) = \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x_k - f_k(\theta))^2}{2\sigma_k^2}\right)$$

which yields the log-likelihood

$$\log \mathcal{L}(\boldsymbol{x}|\theta) = -\frac{1}{2}\sum_{k=1}^{N}\left[\log(2\pi\sigma_k^2) + \frac{(x_k - f_k(\theta))^2}{2\sigma_k^2}\right].$$

Take the first derivative with respect to each parameter θ_i

$$\begin{aligned}\frac{\partial}{\partial\theta_i}\log \mathcal{L}(\boldsymbol{x}|\theta) &= -\frac{1}{2}\sum_{k=1}^{N}\left[0 + \frac{2(x_k - f_k(\theta))}{\sigma_k^2}\left(-\frac{\partial f_k(\theta)}{\partial\theta_i}\right)\right] \\ &= \sum_{k=1}^{N}\frac{(x_k - f_k(\theta))}{\sigma_k^2}\frac{\partial f_k(\theta)}{\partial\theta_i}.\end{aligned}$$

For the Fisher matrix we need the second derivatives

$$\frac{\partial^2}{\partial\theta_j\partial\theta_i}\log \mathcal{L}(\boldsymbol{x}|\theta) = \sum_{k=1}^{N}\left[\frac{(x_k - f_k(\theta))}{\sigma_k^2}\frac{\partial^2 f_k(\theta)}{\partial\theta_j\partial\theta_i} - \frac{1}{\sigma_k^2}\frac{\partial f_k(\theta)}{\partial\theta_j}\frac{\partial f_k(\theta)}{\partial\theta_i}\right].$$

Taking the expectation value, the first term will be zero, by the definition of the mean. We are left with

$$F_{ij} = \sum_{k=1}^{N}\frac{1}{\sigma_k^2}\frac{\partial f_k(\theta)}{\partial\theta_j}\frac{\partial f_k(\theta)}{\partial\theta_i}.$$

Compute the Fisher matrix for the linear model

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

(b)

Recall the expression for 2×2 matrix inversion

$$A = \begin{bmatrix} a & b \\ c & d\end{bmatrix}, \quad A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a\end{bmatrix}$$

Invert the Fisher matrix to get the covariance matrix:

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

The most information between the two data points will be where the off-diagonal components of the covariance are closest to zero. Setting them to zero, we obtain the condition

$$\frac{x_1}{x_2} = -\frac{\sigma_1^2}{\sigma_2^2}.$$

If $\sigma_1 = \sigma_2$ we obtain $x_2 = -x_1$. Thus we should set the 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. 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?

4.1 Solution

```
import numpy

def compute_covariance_matrix(xvals: list[float], sigma: float, n_par: int):
  F = numpy.zeros([n_par,n_par])
  for x in xvals:
    for i in range(n_par):
      if i==0:
        dfdpi = x
      else:
        dfdpi = 1
      for j in range(n_par):
        if j==0:
          dfdpj = x
        else:
          dfdpj = 1
        F[i,j] += dfdpi*dfdpj / (sigma**2)
  cov_mat = numpy.mat(F).I
  return numpy.mat(F).I

sigma = 0.1
n_par = 2
cov_mat_2_points = compute_covariance_matrix((-1., 1.), sigma, n_par)
cov_mat_3_points = compute_covariance_matrix((-1., 0., 1.), sigma, n_par)

print("Matrix for 2 points:")
print(cov_mat_2_points)
print("Matrix for 3 points:")
print(cov_mat_3_points)
```

```
Matrix for 2 points:
[[0.005 0.   ]
 [0.    0.005]]
Matrix for 3 points:
[[0.005      0.        ]
 [0.         0.00333333]]
```

The results show that the variance of the slope, a, remains the same, but adding the third point led to a reduction in the variance of the intercept, b.