Statistical Inference

B. Statistical Data Science 2nd Year   Indian Statistical Institute

Teacher: Soham Sarkar

Exercise Series 8

Exercise 1. In many instances, obtaining the MLE in closed-form by solving the likelihood equations $\ell'(\theta) = 0$ is not possible. In such cases, we need to use some numerical optimization technique to get a solution. One such technique, known as *Fisher's scoring*, was discussed in the class. This is an iterative method, where we find

$$\theta_{n+1} = \theta_n - \frac{\ell'(\theta_n)}{\mathbb{E}_\theta[\ell''(\theta \mid \mathbf{X})]\Big|_{\theta=\theta_n}} = \theta_n + \frac{\ell'(\theta_n)}{I(\theta_n)},$$

where $I(\theta) = -\mathbb{E}_\theta[\ell''(\theta \mid \mathbf{X})]$ is the *Fisher information* contained in the sample. Here, the derivatives are w.r.t. θ and the expectations are w.r.t. the sample $\mathbf{X}$. In this exercise, we will derive the Fisher scoring iteration steps for one particular distribution.

Let $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} f_\theta$, where $f_\theta(x) = \dfrac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}$, $\quad x \in \mathbb{R}$. Here $\theta \in \mathbb{R}$ is our parameter of interest.

(a) Write down the likelihood of θ and the likelihood equation $\ell'(\theta) = 0$. Does it admit a closed form solution?

(b) Find $\ell''(\theta)$ and $I(\theta)$.

(c) Formulate the Fisher scoring equation for finding the MLE.

(d) Write an iterative algorithm to find the MLE using your answer in part (c). You need to fix a threshold to determine convergence. Use a small value, e.g., 10^{-5}.

(e) Fix a θ and generate $n = 100$ samples from f_θ. Use your algorithm to determine the MLE for the given data.

(f) Repeat part (e) 1000 times by generating independent random samples from the same distribution. This will give you 1000 estimated values for the parameter θ. Make a histogram of these estimates. What is your assessment of the MLE as an estimator of θ.

(g) Repeat parts (e) and (f) with $n = 50, 200, 500$. What do you observe?

Exercise 2. Suppose that Y is a random variable that takes the value 0 or 1. This can be, for instance, an indicator of whether a person has a disease, or whether a person smokes, etc. Now, a natural model for such binary outcomes is the $\mathsf{Bernoulli}$ model, where we assume $Y \sim \mathsf{Bernoulli}(\theta)$ for some $\theta \in [0, 1]$. Our parameter of interest in this case is then θ.

In many situations, we also have additional information on the person, e.g., his/her vitals, medical history etc. Suppose for the i-th person, in addition to observing Y_i, we also observe an additional value x_i (for the moment, a scalar). One way to incorporate this additional information in the model is to assume $Y_i \overset{\text{indep}}{\sim} \mathsf{Bernoulli}(\theta_i)$, where $\theta_i = \theta(x_i) = \dfrac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}$ for some unknown $\alpha, \beta \in \mathbb{R}$. This particular form of θ ensures that $\theta_i \in (0, 1)$ for all values of (α, β). This model is known as the *logistic regression* model and it is captures how the probability of success depends on the covariate x.

(a) Verify that for a fixed α, $\dfrac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}$ is increasing in β. Similarly, for a fixed β, $\dfrac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}$ is increasing in α. Therefore, the sign and the magnitude of β indicates the importance of x in increasing/decreasing the probability of *success* (i.e., $Y = 1$).

(b) Write down the likelihood of α, β. Formulate the likelihood equations.

(c) Find the Fisher's scoring equations for obtaining the solutions to the likelihood equations.

Note: The iterative algorithm from the Fisher's scoring method in this context is known as *iteratively reweighted least squares* or IRLS.

Exercise 3. Let $X_1, \ldots, X_n$ be i.i.d. with mean μ and variance σ^2.

(a) Let $a_1, \ldots, a_n$ be arbitrary constants. Show that $T = \sum_{i=1}^{n} a_i X_i$ is an unbiased estimator for μ if $\sum_{i=1}^{n} a_i = 1$.

(b) Find the variance of the estimator T.

(c) Among all unbiased estimators T as in part (a), determine the one with minimum variance. Find its variance.

Exercise 4. For the following distributions, find the Fisher's information and hence the Cramér-Rao lower bound for the variance of an unbiased estimator of θ. Based on an i.i.d. sample $X_1, \ldots, X_n$ from the distributions, try to identify an estimator that attains the Cramér-Rao lower bound. Does the MLE attain the Cramér-Rao lower bound?

(a) $f_p = \mathsf{Bernoulli}(p)$, $0 \le p \le 1$.

(b) $f_\theta = \mathsf{Poisson}(\lambda)$, $\lambda > 0$.

(c) $f_p = \mathsf{Geometric}(p)$, $0 \le p \le 1$.

(d) $f_\theta = \mathsf{Uniform}(0, \theta)$, $\theta > 0$.

(e) $f_\mu = \mathsf{Normal}(\mu, 1)$, $\mu \in \mathbb{R}$.

(f) $f_\sigma = \mathsf{Normal}(0, \sigma^2)$, $\sigma > 0$.

(g) $f_{\mu,\sigma} = \mathsf{Normal}(\mu, \sigma^2)$, $\mu \in \mathbb{R}, \sigma > 0$.

Note: In this case, you need to use the multivariate version of the bound.

(h) $f_\sigma = \mathsf{Laplace}(0, \sigma)$, $\sigma > 0$.

(i) $f_\theta = \mathsf{Normal}(\theta, \theta^2)$, $\theta \in \mathbb{R}$.

(j) $f_\theta = \mathsf{Normal}(\theta, \theta)$, $\theta > 0$.

(k) $f_\theta = \mathsf{Exponential}(\text{scale} = \theta)$, $\theta > 0$.

Hint: *The pdf of* $\mathsf{Exponential}(\text{scale} = \theta)$ *is* $f_\theta(x) = \theta e^{-\theta x}, x > 0$.

Note: For each case, check whether the conditions in deriving the Cramér-Rao lower bound are satisfied. That is, whether the respective differentiation and integration operations can be interchanged. If these conditions are not satisfied, then the Cramér-Rao lower bound may be incorrect.