Statistical Inference

B. Statistical Data Science 2nd Year   Indian Statistical Institute

Teacher: Soham Sarkar

Exercise Series 7

Exercise 1. Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} f_\theta$. For the following cases, find the method-of-moments (MoM) estimator and the maximum likelihood estimator (MLE) of the unknown parameter. Are the two estimators functions of minimal sufficient statistics? Verify whether the MoM estimator is also an MLE. If not, which one would you prefer and why?

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

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

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

(d) $f_\theta = \mathsf{Uniform}(\theta, 1)$, $\theta < 1$.

(e) $f_\theta = \mathsf{Uniform}(\theta, \theta + 1)$, $\theta \in \mathbb{R}$.

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

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

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

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

(j) $f_{\mu,\sigma} = \mathsf{Laplace}(\mu, \sigma)$, $\mu \in \mathbb{R}, \sigma > 0$.

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

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

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

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

Exercise 2. Let $(X_1, Y_1), \ldots, (X_n, Y_n) \overset{\text{i.i.d.}}{\sim} \mathsf{BVN}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, where $\mu_x, \mu_y \in \mathbb{R}$, $\sigma_x, \sigma_y > 0$ and $-1 < \rho < 1$.

(a) Assuming all the parameters to be unknown, write down the likelihood and find MLE of the parameters. Do they coincide with the MoM estimators?

(b) Now, suppose that ρ is known. Find MLE and MoM estimators of the other parameters in this case. Are they the same as in part (a)?

(c) Assume now that $\mu_x = \mu_y = 0, \sigma_x^2 = \sigma_y^2 = 1$ and ρ is unknown. Write down the likelihood of ρ. Verify that the likelihood equation is cubic in ρ. Conclude that, in this case, multiple roots of the likelihood equation are possible. What happens to the MoM estimator in this case?

(d) Check what happens in part (c) if $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2$ are known, but are not necessarily $0, 0, 1, 1$, respectively.

Exercise 3. Let $X_1, \ldots, X_n$ be i.i.d. with pdf f_θ, $\theta \in \{0, 1\}$, with

$$f_0(x) = \begin{cases} 1 & \text{if } 0 < x < 1, \\ 0 & \text{otherwise,} \end{cases} \qquad f_1(x) = \begin{cases} \dfrac{1}{2\sqrt{x}} & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Write down the likelihood of θ for this model.

(b) Find the maximum likelihood estimator of θ.

Exercise 4. Let the random variable X take three values $0, 1$ and 2 with

$$\mathbb{P}_\theta(X = 0) = 6\theta^2 - 4\theta + 1, \quad \mathbb{P}_\theta(X = 1) = \theta - 2\theta^2, \quad \mathbb{P}_\theta(X = 2) = 3\theta - 4\theta^2, \qquad 0 \leq \theta \leq \frac{1}{2}.$$

(a) Write down the likelihood of θ based on a single observation x.

(b) Verify whether the MLE of θ always exists. If so, find the MLE.

(c) Verify whether the method of moments estimator of θ always exists. If so, find the MoM estimator.

Exercise 5. Let $x_1, \ldots, x_n$ be fixed constants. The random variables $Y_1, \ldots, Y_n$ satisfy

$$Y_i = \beta x_i + \epsilon_i, \qquad i = 1, \ldots, n,$$

where $\epsilon_1, \ldots, \epsilon_n \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0, \sigma^2)$, $\beta \in \mathbb{R}, \sigma > 0$ are unknown.

(a) Write down the likelihood of β and σ.

(b) Find the MLEs of the unknown parameters.

(c) How would your results change if $\epsilon_1, \ldots, \epsilon_n \overset{\text{i.i.d.}}{\sim} \mathsf{Laplace}(0, \sigma)$?

(d) Redo parts (a) and (b) for $Y_i = \alpha + \beta x_i + \epsilon_i$, $i = 1, \ldots, n$, where $\epsilon_1, \ldots, \epsilon_n \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0, \sigma^2)$, $\alpha \in \mathbb{R}, \beta \in \mathbb{R}, \sigma > 0$ are unknown parameters.

Note: The models described above are known as *linear regression* models.

Exercise 6. Recall Exercise 4 from Series 2, where we modeled the prevalence of antigens `A` and `B` in humans. The determination of the four blood groups, `A`, `B`, `AB` and `O`, depending on the antigens according to the following table.

	Antigen `A` present	Antigen `A` absent
Antigen `B` present	Blood group `AB`	Blood group `B`
Antigen `B` absent	Blood group `A`	Blood group `O`

Let $p_{\mathtt{A}}, p_{\mathtt{B}}, p_{\mathtt{AB}}$ and $p_{\mathtt{O}}$ denote the probabilities of a randomly selected person having blood group `A`, `B`, `AB`, or `O`, respectively. We select 500 people at random and note down their blood groups.

(a) Using the $\mathsf{Multinomial}$ distribution for the data, write down the likelihood of the parameters $p_{\mathtt{A}}, p_{\mathtt{B}}, p_{\mathtt{AB}}$ and $p_{\mathtt{O}}$ in terms of $N_{\mathtt{A}}, N_{\mathtt{B}}, N_{\mathtt{AB}}$ and $N_{\mathtt{O}}$; the number of people (out of 500) having blood groups `A`, `B`, `AB` and `O`, respectively.

Hint: *This is a singular/non-full-rank model since $p_A + p_B + p_{AB} + p_O = 1$. You can make the model full rank by removing one of the parameters and expressing it in terms of the others. For instance, you may use p_A, p_B, p_{AB} and $1 - p_A - p_B - p_{AB}$. See Exercise 2(m) in Series 6.*

(b) Let $q_{\mathtt{A}}$ and $q_{\mathtt{B}}$ denote the probability of the presence of antigen `A` and antigen `B`, respectively, and the presence/absence of the two antigens do not affect each other. Write the likelihood of the updated model with this additional piece of information and find the MLE of the parameters $q_{\mathtt{A}}$ and $q_{\mathtt{B}}$.

Exercise 7. Maximum likelihood estimator depends heavily on the parameter space. In the following, we revisit some of the distributions considered in Exercise 1 but with a different parameter space. Determine the MLE in the respective cases.

(a) $f_p = \mathsf{Bernoulli}(p)$, $\dfrac{1}{4} \leq p \leq \dfrac{3}{4}$.

(b) $f_\lambda = \mathsf{Poisson}(\lambda)$, $\lambda \in \mathbb{N}$. ($\mathbb{N} = \{1, 2, 3, \ldots\}$ is the set of *natural numbers*)

(c) $f_\theta = \mathsf{Uniform}(\theta, \theta + 1)$, $\theta \in \mathbb{I}$. ($\mathbb{I} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ is the set of *integers*)

(d) $f_\mu = \mathsf{Normal}(\mu, 5)$, $\mu > 0$.