Maximum Likelihood Estimation - Linear Case Exercise

HY-673, 2023-2024, Spring

Exercise 3: MLE - Linear case (2022-2023 Assignment #1)

(a) Let $x_1, \ldots, x_n \in \mathbb{R}$ be a given dataset. Assume the parametric model that generates the samples to be $p_\theta(x) \equiv \mathcal{N}(0, \theta)$, where θ corresponds to the variance σ^2. Compute analytically the MLE solution $\hat{\theta}_{\text{MLE}}$.

<u>Solution:</u>

Letting $\mathbf{X} = x_1, x_2, \ldots, x_n,\ x_i \in \mathbb{R},\ i = 1, 2, \ldots, n$, we need to find by definition:

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} p_\theta\left(\mathbf{X}|\theta\right), \tag{1}$$

and since we are given that $p_\theta(x)$ is a univariate Gaussian, the likelihood function is:

$$\mathfrak{L} = p_\theta(\mathbf{X}|\theta) = \mathcal{N}(\mathbf{X}|\theta) = p_\theta(\mathbf{X}|\mu, \sigma^2),\ \theta = \mu, \sigma^2. \tag{2}$$

Due to the Gaussian having an exponential, it will be easier to work with the logarithmic likelihood function which will be:

$$\ln \mathfrak{L} = \ln\left(\mathcal{N}\left(\mathbf{X}|\mu, \sigma^2\right)\right) = \sum_{n=1}^{N} \ln\left(\mathcal{N}\left(x_n|\mu, \sigma^2\right)\right) = \sum_{n=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{(x_n-\mu)^2}{\sigma^2}\right)}\right) \tag{3}$$

$$= \sum_{n=1}^{N} -\ln\left(\sqrt{2\pi\sigma^2}\right) - \frac{(x_n-\mu)^2}{2\sigma^2} = -\frac{N}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2. \tag{4}$$

Since $\ln(x) : \mathbb{R}^+ \to \mathbb{R}$ is a monotonically increasing function, we can equivalently find:

$$\hat{\mu}_{\text{MLE}} = \arg\max_{\mu} \ln \mathfrak{L},\ \text{and} \tag{5}$$

$$\hat{\sigma}^2_{\text{MLE}} = \arg\max_{\sigma^2} \ln \mathfrak{L}. \tag{6}$$

It is not required for us to find $\hat{\mu}_{\text{MLE}}$, but, let's do it anyway. For that, we need the critical points of the logarithmic likelihood function with respect to μ:

$$\frac{\partial}{\partial\mu}\ln\mathfrak{L} = 0 \implies \frac{\partial}{\partial\mu}\left(-\frac{N}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right) = 0 \implies \tag{7}$$

$$\frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n-\mu) = 0 \implies \sum_{n=1}^{N}x_n - N\mu = 0 \implies \mu = \frac{1}{N}\sum_{n=1}^{N}x_n, \tag{8}$$

therefore, since it is given that $p_\theta(X) \equiv \mathcal{N}(\mu = 0, \sigma^2 = \theta)$:

$$\hat{\mu}_{\text{MLE}} = \frac{1}{N}\sum_{n=1}^{N} x_n = 0. \tag{9}$$

Likewise, to find $\hat{\sigma}^2_{\text{MLE}}$, we want:

$$\frac{\partial}{\partial \sigma^2} \ln \mathfrak{L} = 0 \implies \frac{\partial}{\partial \sigma^2}\left(-\frac{N}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu)^2\right) = 0 \implies \tag{10}$$

$$-\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{n=1}^{N}(x_n - \mu)^2 = 0 \implies -N\sigma^2 + \sum_{n=1}^{N}(x_n - \mu)^2 = 0 \implies \sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)^2, \tag{11}$$

hence, since $p_\theta(X) \equiv \mathcal{N}(\mu = 0, \sigma^2 = \theta)$, the final answer is:

$$\boxed{\hat{\theta}_{\text{MLE}} = \hat{\sigma}^2_{\text{MLE}} = \frac{1}{N}\sum_{n=1}^{N} x_n^2.} \tag{12}$$

(b) Generate samples $x_i \sim \mathcal{N}(0, \theta^*)$ with $\theta^* = 2$ and create one realization of the dataset. For $n = 10^1, 10^2, 10^3$, plot the log-likelihood as a function of θ. What do you observe?

Solution:

The requested code for the plot is in `03e_MLE_linear_case.ipynb`:

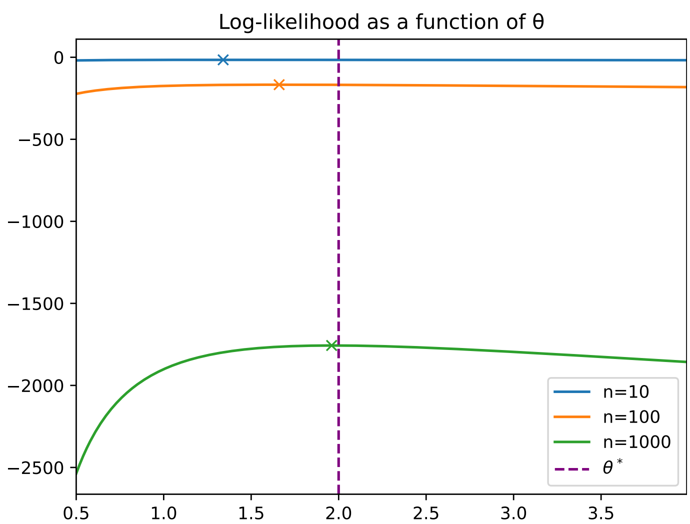

Figure 1: We observe that as n increases, it becomes easier for the log-likelihood functions reach the ground truth value $\theta^* = 2$.

(c) Repeat (b) averaging over 100 realizations with new data at each realization and plot the average log-likelihood as a function of θ. For each realization, compute also $\hat{\theta}_{\text{MLE}}(n)$ and plot the respective histogram. Comment on the shape of the average log-likelihood function and the shape of the histogram of $\hat{\theta}_{\text{MLE}}(n)$.

Solution:

The requested plots are below, and the code in `03e_MLE_linear_case.ipynb`.

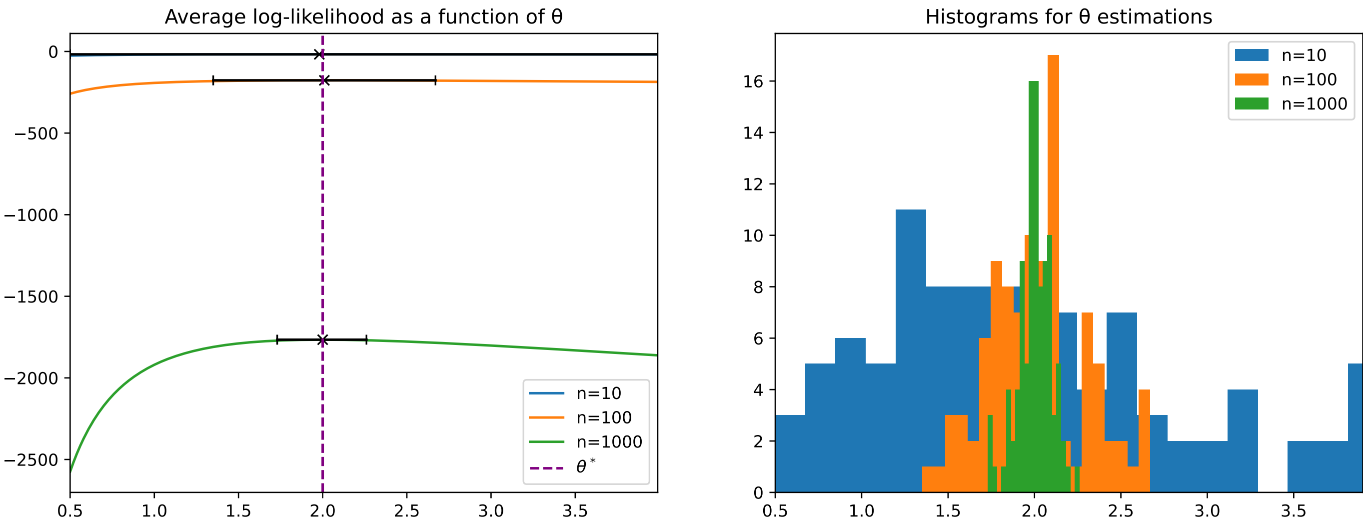

Figure 2: The shape of the average log-likelihood functions do not differ much from the shape of the log-likelihood functions that we observed in the previous question. Taking the average log-likelihood function over 100 realizations of a Gaussian dataset helps significantly approximate $\theta^* = 2$, even with a small dataset comprised of $n = 10$ samples. The histogram plot shows that by drawing more samples, our estimators become more "confident", i.e., their variance drops. The dispersion of the predictions is lower for a bigger value of n, which follows the intuition that more data points result in greater estimation confidence.

(d) Compute analytically the Fisher information:

$$\mathcal{I}(\theta) := \mathbb{E}_{p\theta}\left[\left(\frac{d}{d\theta}\log p_\theta(x)\right)^2\right]. \tag{13}$$

Solution:

Setting $p_\theta(X) \equiv \mathcal{N}(\mu = 0, \sigma^2 = \theta)$, then from eq. 3,4,10,11, the Fisher information is:

$$\mathcal{I}(\theta) = -\mathbb{E}_{p_\theta}\left[\frac{d^2}{d\theta^2}\ln p_\theta(X)\right] = -\mathbb{E}_{p_\theta}\left[\frac{d}{d\theta}\left(-\frac{1}{2\theta} + \frac{1}{2\theta^2}(X-\mu)^2\right)\right] \tag{14}$$

$$= -\mathbb{E}_{p_\theta}\left[\frac{1}{2\theta^2} - \frac{1}{2\theta^3}(X-\mu)^2\right] = \frac{1}{2\theta^3}\mathbb{E}_{p_\theta}\left[(X-\mu)^2\right] = \frac{1}{2\theta^2}. \tag{15}$$