Weekly Overview Slides of Statistical Machine Learning CSE 575, Fall 2023

Moses A. Boudourides[1]

SPA and SCAI
Arizona State University

[1] Moses.Boudourides@asu.edu

Week 4

Exercises on Inferring Probability Models from Data

February 2, 2023

Agenda

- **Four New Papers Suggested for Project Proposals**
- **Exercises on Inferring Probability Models from Data: MLE and Bayesian Estimation**

Exercise 1

Exercise 1

Let X be a discrete random variable with the following probability mass function, where $0 \leq \theta \leq 1$ is a parameter:

$$P(X|\theta) = \begin{cases} \frac{2\theta}{3}, & \text{for } X = 0, \\ \frac{\theta}{3}, & \text{for } X = 1, \\ \frac{2(1-\theta)}{3}, & \text{for } X = 2, \\ \frac{1-\theta}{3}, & \text{for } X = 3. \end{cases}$$

Given the following IID data $\mathcal{D} = (3{,}0{,}2{,}1{,}3{,}2{,}1{,}0{,}2{,}1)$, what is the MLE of θ?

Solution of Exercise 1

Solution

Apparently, since the likelihood function $\mathcal{L}(\theta)$ is an order 10 polynomial:

$$\mathcal{L}(\theta) = \mathcal{P}(\mathcal{D}|\theta) = \prod_{i=0}^{3} P(X = i|\theta)$$
$$= \left(\frac{2\theta}{3}\right)^2 \left(\frac{\theta}{3}\right)^3 \left(\frac{2(1-\theta)}{3}\right)^3 \left(\frac{1-\theta}{3}\right)^2,$$

it is easier to maximize the log–likelihood function, which is:

$$\log \mathcal{L}(\theta) = \log P(\mathcal{D}|\theta) = \sum_{i=0}^{3} \log P(X = i|\theta)$$
$$= 2\left(\log \frac{2}{3} + \log \theta\right) + 3\left(\log \frac{1}{3} + \log \theta\right) + 3\left(\log \frac{2}{3} + \log(1-\theta)\right)$$
$$+ 2\left(\log \frac{1}{3} + \log(1-\theta)\right)$$
$$= 5 \log \theta + 5 \log(1-\theta) + C,$$

for a θ independent constant C. Thus, from $\frac{d}{d\theta}(\log \mathcal{L}(\theta)) = \frac{5}{\theta} - \frac{5}{1-\theta} = 0$, we find the MLE of θ to be $\hat{\theta} = \frac{1}{2}$.

Exercise 2

Exercise 2

If $X_1, X_2, \ldots, X_n$ are IID random variables with , where $\sigma > 0$ is an unknown parameter:

$$f(x|\sigma) = \frac{1}{2\sigma} \exp\left(- \frac{|x|}{\sigma} \right),$$

find the MLE of σ.

Solution

As the log–likelihood function is $\log \mathcal{L}(\sigma) = \sum_{i=1}^{n} \left(- \log 2 - \log \sigma - \frac{|X_i|}{\sigma} \right)$,

setting its derivative w.r.t. σ equal to 0,

$$\frac{d}{d\sigma}(\log \mathcal{L}(\sigma)) = \sum_{i=1}^{n} \left(- \frac{1}{\sigma} + \frac{|X_i|}{\sigma^2} \right) = -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum_{i=1}^{n} |X_i| = 0,$$

yields the MLE value

$$\hat{\sigma} = \frac{1}{n} \sum_{i=1}^{n} |X_i|.$$

Exercise 3

Exercise 3

The Pareto distribution, which is often used in economics, has a PDF with a slowly decaying tail:

$$f(x|x_0, \theta) = \theta x_0^{\theta} x^{-\theta-1}, x \geq x_0, \theta > 1.$$

If x_0 is given and $X_1, X_2, \ldots, X_n$ is an IID sample, find the MLE of θ.

Solution

The log–likelihood function

$$\log \mathcal{L}(\theta) = \sum_{i=1}^{n} \log f(X_i|\theta) = \sum_{i=1}^{n} \left(\log \theta + \theta \log x_0 - (\theta + 1) \log X_i \right),$$

$$= n \log \theta + n\theta \log x_0 - (\theta + 1) \sum_{i=1}^{n} \log X_i.$$

Thus, setting the θ–derivative equal to 0 gives $\frac{d}{d\theta}(\log \mathcal{L}(\theta)) = \frac{n}{\theta} + n \log x_0 - \sum_{i=1}^{n} \log X_i = 0$, which results the following MLE value

$$\hat{\theta} = \frac{1}{\frac{1}{n} \sum_{i=1}^{n} X_i - \log x_0}.$$

Exercise 4

Exercise

If $X_1, \ldots, X_n$ is a random sample from a uniform distribution on the interval $(0, \theta)$, where $\theta > 0$, find the MLE of θ.

Solution

The PDF of each observation is

$$f(x|\theta) = \begin{cases} \frac{1}{\theta}, & \text{for } 0 \leq x \leq \theta, \\ 0, & \text{otherwise.} \end{cases}$$

Hence, the log–likelihood function is

$$\mathcal{L}(\theta) = \begin{cases} \frac{1}{\theta^n}, & \text{for } 0 \leq x_i \leq \theta, i = 1, \ldots, n, \\ 0, & \text{otherwise.} \end{cases}$$

It can be seen that the MLE of θ must be a value of θ for which $\theta \geq x_i$, for $i = 1, \ldots, n$, and which maximizes $\frac{1}{\theta^n}$ among all such values. Since $\frac{1}{\theta^n}$ is a decreasing function of θ, the estimate will be the smallest possible value of θ such that $\theta \geq x_i$, for $i = 1, \ldots, n$. This value is $\theta = \max(x_1, \ldots, x_n)$, and, thus, it follows that the MLE of θ is $\hat{\theta} = \max(X_1, \ldots, X_n)$. It should be remarked that in this example, the MLE $\hat{\theta}$ does not seem to be a suitable estimator of θ. We know that $\max(X_1, \ldots, X_n) < \theta$ with probability 1, and, therefore, $\hat{\theta}$ surely underestimates the value of θ.

Exercise 5

The Binomial Model

If the IID data $\mathcal{D} = (x_1, x_2, \ldots, x_n)$ all take values in $\{0, 1\}$ (such as flip of coins) and all follow the Bernoulli model, i.e., $P(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$, where the parameter $\theta \in [0, 1]$ is the probability for the value 1 (heads), find the MLE of θ.

Solution

Now, the likelihood function is $P(\mathcal{D}|\theta) = \theta^{n_1}(1-\theta)^{n_0}$, where $n_1 = \sum_{i=1}^{n} x_i$ is the number of observed 1's and $n_0 = \sum_{i=1}^{n}(1-x_i)$ is the number of observed 0's in the data. Hence, the log–likelihood function becomes

$$\log \mathcal{L}(\theta) = \sum_{i=1}^{n} \log \left(\theta^{x_i}(1-\theta)^{1-x_i}\right) = n_1 \log \theta + n_0 \log(1-\theta).$$

Clearly, $\frac{d}{d\theta}(\log \mathcal{L}(\theta)) = \frac{n_1}{\theta} - \frac{n_0}{1-\theta} = 0$ implies that (since $n_1 + n_0 = n$)

$$\hat{\theta} = \frac{n_1}{N}.$$

Exercise 6

Exercise

The Poisson PMF is defined as $\mathrm{Poi}(x|\theta) = e^{-\theta}\frac{\theta^x}{x!}$, for $x \in \{0, 1, 2, \ldots, n\}$, where $\theta > 0$ is the rate parameter. Find the MLE of θ.

Solution

The log–likelihood function is (dropping constants independent of θ):

$$\log \mathcal{L}(\theta) = \sum_{i=0}^{n} \log\left(e^{-\theta}\theta^{x_i}\right) = -n\theta + \log\theta \sum_{i=0}^{n} x_i.$$

Taking the θ derivative and equating to zero yields

$$\frac{d}{d\theta}(\log \mathcal{L}(\theta)) = -n + \frac{1}{\theta}\sum_{i=0}^{n} x_i = 0,$$

from which we get

$$\hat{\theta} = \frac{\sum_{i=0}^{n} x_i}{n}.$$

Exercise 7

Exercise

If $X_1, \ldots, X_n$ is a random sample from the distribution with PDF $f(x|\theta) = \theta e^{-\theta x}$ and if the prior is given as $P(\theta) = \mu e^{-\mu\theta}$, for some known $\mu > 0$, find the posterior.

Solution

Apparently, the likelihood is:

$$P(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}$$

Hence, the Posterior $\propto$ Likelihood $\times$ Prior is:

$$\begin{aligned} P(\theta|x_1, \ldots, x_n) &\propto \theta^n e^{-\theta \sum_{i=1}^{n} x_i} \mu e^{-\mu\theta} \\ &\propto \theta^n e^{-\theta(\mu + \sum_{i=1}^{n} x_i)}. \end{aligned}$$

Exercise 8

Exercise

If $X \sim \mathrm{bionomial}(n, \theta)$, where n is known, and if the prior for θ is a $\mathrm{Beta}(\alpha, \beta)$ distribution, find the prosterior.

Solution

Now, the prior is

$$P(\theta) = \frac{1}{\mathrm{Beta}(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

and the likelihood is:

$$P(x|\theta) = \binom{n}{x} \theta^{x}(1-\theta)^{n-x}$$

Hence, the Posterior $\propto$ Likelihood $\times$ Prior is:

$$P(\theta|x) \propto \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1},$$

which is recognized to be a Beta distribution with parameters $x+\alpha$ and $n-x+\beta$.

Exercise 9

Exercise

Let $X_1, \dots, X_n$ be a random sample from the Poisson distribution $\mathrm{Poi}(x|\theta) = e^{-\theta}\frac{\theta^x}{x!}$, for $x \in \{0, 1, 2, \dots\}$, where $\theta > 0$ is the rate parameter. If the prior is given as a $\mathrm{Gamma}(\alpha, \beta)$ distribution, for known parameters $\alpha, \beta > 0$, find the posterior.

Solution

The prior is given as

$$P(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}$$

and the likelihood is:

$$P(x_1, \dots, x_n|\theta) = \prod_{i=0}^{n} e^{-\theta}\frac{\theta^{x_i}}{x_i!} = e^{-n\theta}\frac{\theta^{x_0+x_1+\ldots+x_n}}{x_0! \cdot x_1! \cdot \ldots \cdot x_n!}.$$

Therefore, the Posterior $\propto$ Likelihood $\times$ Prior becomes:

$$\begin{aligned} P(\theta|x) &\propto \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}e^{-n\theta}\frac{\theta^{x_0+x_1+\ldots+x_n}}{x_0! \cdot x_1! \cdot \ldots \cdot x_n!} \\ &\propto \theta^{\alpha+x_0+x_1+\ldots+x_n-1}e^{-(\beta+n)\theta}, \end{aligned}$$

which is recognized as the Gamma distribution with parameters $\alpha + \sum_{i=0}^{n} x_i$ and $\beta + n$.

Exercise 10

Exercise

If $X_1, \ldots, X_n \sim N(\theta, \sigma^2)$, where σ^2 is known, and if the prior $P(\theta) \sim N(\mu, \tau^2)$, for known μ and τ^2, find the prosterior.

Solution

Now, the prior is

$$P(\theta) = \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{1}{2}\frac{(\theta-\mu)^2}{\tau^2}}$$

and the likelihood is:

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x_i-\theta)^2}{\sigma^2}} = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\frac{(x_i-\theta)^2}{\sigma^2}}$$

Hence, the Posterior $\propto$ Likelihood $\times$ Prior is:

$$P(\theta|x1, \ldots, x_n) \propto e^{-\frac{1}{2}\left[\sum_{i=1}^{n}\left(\frac{(x_i-\theta)^2}{\sigma^2} + \frac{(\theta-\mu)^2}{\tau^2}\right)\right]} = e^{-\frac{1}{2}M}$$

where M is the expression in brackets. DO THE ALGEBRA, to get

$$M = a\left(\theta - \frac{b}{a}\right)^2 + \frac{b^2}{a} + c,$$

where $a = \frac{n}{\sigma^2} + \frac{1}{\tau^2}$, $b = \sum_{i=1}^{n} \frac{x_i}{\sigma^2} + \frac{\mu}{\tau^2}$, $c = \sum_{i=1}^{n} \frac{x_i^2}{\sigma^2} + \frac{\mu^2}{\tau^2}$ and then conclude that

$$P(\theta|x1, \ldots, x_n) \propto e^{-\frac{1}{2}a\left(\theta - \frac{b}{a}\right)^2},$$

which is easily recognized to be the PDF of a normal distribution with mean $\frac{b}{a}$ and variance $\frac{1}{a}$.