Homework 4 Theory

October 17, 2024

1 Problem 1

Assume X is a discrete random variable that takes values in $\{1, 2, 3\}$, with probability defined by

$$\Pr(X = 1) = \theta_1, \quad \Pr(X = 2) = 2\theta_1, \quad \Pr(X = 3) = \theta_2,$$

where $\theta = [\theta_1, \theta_2]$ is an unknown parameter to be estimated. Now assume we observe a sequence $D := \{x^{(1)}, x^{(2)}, \ldots, x^{(n)}\}$ that is independent and identically distributed (i.i.d.) from the distribution. We assume the number of observations of the values: 1, 2, 3 in D are s_1, s_2, s_3, respectively.

(a) [5 points] To ensure that $\Pr(X = i)$ is a valid probability mass function, what constraint should we put on $\theta = [\theta_1, \theta_2]$? Write your answers quantitatively as expressions that include θ_1 and θ_2.

(b) [5 points] Write down the joint probability of the data sequence

$$\Pr(D \mid \theta) = \Pr(x^{(1)}, \ldots, x^{(n)} \mid \theta),$$

and the log probability $\log \Pr(D \mid \theta)$.

(c) [5 points] Calculate the maximum likelihood estimation $\hat{\theta}$ of θ based on the sequence D.

Part (a) Solution

In order for a probability mass is valid, there must be a couple conditions met.

1 The total probabilities must sum to 1.

2 Each probability must be non-negative.

Thus, we can ensure that θ meets these conditions by the following constraints:

1 $\theta_1 \geq 0$

2 $\theta_2 \geq 0$

4 $Pr(X = 1) + Pr(X = 2) + Pr(X = 3) = 3\theta_1 + \theta_2 = 1$

Part (b) Solution

The joint probability of the data sequence is defined as follows:

$$Pr(D \mid \theta) = Pr(x_1, \ldots x_n \mid \theta)$$

$$= \prod_{i=1}^{n} Pr(x_i \mid \theta)$$

Because the joint distribution is equivalent to multiplying all of the probabilities together. This is why it ends up as the product of them all.

This can be expressed as a sum by using the **log probability**:

$$log(Pr(D \mid \theta)) = log(\prod_{i=1}^{n} Pr(x_i \mid \theta))$$

$$= \sum_{i=1}^{n} log(Pr(x_i \mid \theta))$$

This makes the probability and following calculations much more manageable.

Part (c) Solution

We are given that $X \in \{1, 2, 3\}$ with probabilities:

$$P(X = 1) = \theta_1, \quad P(X = 2) = 2\theta_1, \quad P(X = 3) = \theta_2,$$

where θ_1 and θ_2 satisfy the constraint:

$$3\theta_1 + \theta_2 = 1.$$

Thus, θ_2 can be expressed as:

$$\theta_2 = 1 - 3\theta_1.$$

Given a dataset D with s_1 occurrences of $X = 1$, s_2 occurrences of $X = 2$, and s_3 occurrences of $X = 3$, we aim to find the MLE of θ based on the data.

The likelihood function is:

$$L(\theta) = \theta_1^{s_1} \cdot (2\theta_1)^{s_2} \cdot \theta_2^{s_3}.$$

Substitute $\theta_2 = 1 - 3\theta_1$ and simplify:

$$L(\theta_1) = 2^{s_2} \cdot \theta_1^{s_1+s_2} \cdot (1 - 3\theta_1)^{s_3}.$$

The log-likelihood function is:

$$\log L(\theta_1) = s_2 \log 2 + (s_1 + s_2) \log(\theta_1) + s_3 \log(1 - 3\theta_1).$$

Taking the derivative:

$$\frac{d}{d\theta_1} \log L(\theta_1) = \frac{s_1 + s_2}{\theta_1} - \frac{3s_3}{1 - 3\theta_1}.$$

Set this equal to zero and solve for θ_1:

$$\hat{\theta}_1 = \frac{s_1 + s_2}{3n}.$$

Using the relationship $\theta_2 = 1 - 3\theta_1$:

$$\hat{\theta}_2 = \frac{s_3}{n}.$$

Thus, the MLE for the vector $\theta = [\theta_1, \theta_2]$ is:

$$\hat{\theta}_1 = \frac{s_1 + s_2}{3n}, \quad \hat{\theta}_2 = \frac{s_3}{n}.$$

2 Problem 2

Let $\{x^{(1)}, \ldots, x^{(n)}\}$ be an i.i.d. sample from an exponential distribution, whose density function is defined as

$$f(x \mid \beta) = \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right), \quad \text{for } 0 \leq x < \infty.$$

Please find the maximum likelihood estimator (MLE) of the parameter β. Show your work.

Solution

There are three main steps in MLE:

1. Define the likelihood function
2. Calculate the log-likelihood function
3. Maximize the log-likelihood function with respect to the target parameter

Defining the Likelihood Function

I will begin by identifying the probability density function for x:

$$f(x \mid \beta) = \frac{1}{\beta} exp(-\frac{x}{\beta})$$

Thus, the likelihood function for this problem given the observations $x_1, \ldots, x_n$, can be given by:

$$L(\beta) = \prod_{i=1}^{n} f(x_i \mid \beta)$$

$$= \prod_{i=1}^{n} \frac{1}{\beta} exp(-\frac{x_i}{\beta})$$

This is the final likelihood function that will be converted to log-likelihood in the next step.

Calculating the Log-Likelihood Function

The likelihood function derived in the last step can be converted into log-likelihood as shown below:

$$L(\beta) = \prod_{i=1}^{n} \frac{1}{\beta} exp(-\frac{x_i}{\beta})$$

$$log(L(\beta)) = log(\prod_{i=1}^{n} \frac{1}{\beta} exp(-\frac{x_i}{\beta}))$$

$$= \sum_{i=1}^{n} log(\frac{1}{\beta} exp(-\frac{x_i}{\beta}))$$

$$= \sum_{i=1}^{n} (-log(\beta) \ - \frac{x_i}{\beta})$$

$$= -n \ log(\beta) - \frac{1}{\beta} \sum_{i=1}^{n} x_i$$

This is the simplified log-likelihood function for this problem.

Maximizing the Log-Likelihood Function w.r.t. β

I will maximize this function using gradient descent since a function in the exponential distribution is convex:

Restating the log-likelihood function:

$$\log L(\beta) = -n\log(\beta) - \frac{1}{\beta}\sum_{i=1}^{n} x_i$$

Taking the derivative w.r.t. β (finding the gradient):

$$\frac{d}{d\beta}\log L(\beta) = -\frac{n}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n} x_i$$

Setting the derivative equal to 0 to find maximum:

$$-\frac{n}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n} x_i = 0$$

Eliminate fractions:

$$-n\beta + \sum_{i=1}^{n} x_i = 0$$

Solving for beta:

$$\beta = \frac{1}{n}\sum_{i=1}^{n} x_i$$

3 Problem 3

(a) Assume that you want to investigate the proportion (θ) of defective items manufactured at a production line. You take a random sample of 30 items and found 5 of them were defective. Assume the prior of θ is a uniform distribution on $[0, 1]$. Please compute the posterior of θ. It is sufficient to write down the posterior density function up to a normalization constant that does not depend on θ.

(b) Assume an observation $D := \{x^{(1)}, \ldots, x^{(n)}\}$ is i.i.d. drawn from a Gaussian distribution $\mathcal{N}(\mu, 1)$, with an unknown mean μ and a variance of 1. Assume the prior distribution of μ is $\mathcal{N}(0, 1)$. Please derive the posterior distribution $p(\mu \mid D)$ of μ given data D.

Part (a) Solution

To investigate the proportion (θ) of defective items given some prior knowledge and some observations, we can use Bayesian Inferencing. We know that to estimate the posterior distribution we can use:

$$P(\theta \mid D) = \frac{P(D \mid \theta)\ P(\theta)}{P(D)} \propto P(D \mid \theta)\ P(\theta)$$

Since our observations are of a binomial distribution, we will need that PDF, which is:

$$P(D \mid \theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

Substituting in the PDF of the binomial and the uniform distribution:

$$P(\theta \mid D) \propto \theta^5 (1-\theta)^{25}\ \frac{1}{1-0}, \quad 0 \leq \theta \leq 1$$

Since $\frac{1}{b-a}$ is a constant for the purposes of solving the posterior distribution with respect to θ, we can omit it from the calculation:

$$P(\theta \mid D) \propto \theta^5 (1-\theta)^{25}, \quad 0 \leq \theta \leq 1$$

Above is the final posterior distribution when θ falls on the interval [0,1]. θ will be assigned 0 otherwise based on the prior.

Part (b) Solution

Starting with the formula for Bayes Inferencing:

$$P(\theta \mid D) = \frac{P(D \mid \theta)\ P(\theta)}{P(D)} \propto P(D \mid \theta)\ P(\theta)$$

Deriving the likelihood function for the observations, which was drawn from a Gaussian distribution $\mathcal{N}(\mu, 1)$:

$$P(D \mid \theta) = \prod_{i=1}^{n} P(x_i \mid \mu) = (\frac{1}{\sqrt{2\pi}})^n exp(-\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2)$$

The prior for μ, given a normal distribution:

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

Combining the likelihood and the prior (ignoring all constants) to find the posterior:

$$p(\mu \mid D) \propto exp(-\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2)\ \ exp(-\frac{\mu^2}{2})$$