Ivani Patel
11809154

Assignment 8

Q1. [40 pts] Maximum Likelihood Estimation (I)

Given n = 100 (random) observations $x_1,x_2,...,x_n$ which are independently drawn from an univariate Gaussian distribution N($2\mu,7\sigma^2$) with unknown mean μ and variance $\sigma^2 > 0$.

(a) [20 pts] Derive the maximum likelihood estimation μ_{MLE} of μ as a function of n and ($x_1,x_2,...,x_n$).

Answer:

θMLE = (μMLE , σ2MLE)

dL/dμ = ∑_(i=1)^N d(– 0.5 * (xi – 2μ)2 / 7σ2) / dμ

dL/dμ = -σ -2 * ∑_(i=1)^N (xi – 2μ) d(-μ)/dμ

dL/dμ = (1 / σ^2) $\sum_{i=1}^{N}$ (x_i – 2μ)

μ_{MLE} = (2/N) $\sum_{i=1}^{N}$ x_i

(b) [20 pts] Derive the maximum likelihood estimation σ_{MLE}^2 of σ^2 as a function of n and ($x_1,x_2,...,x_n$).

Answer:

Given n independent observations x1, x2, ..., xn from a Gaussian distribution with mean μ and variance σ^2, the likelihood function can be written as:

L(μ, σ^2) = (2πσ^2)^(-n/2) * exp(-(1/2σ^2) * ∑(xi - μ)^2)

To find the maximum likelihood estimates (MLEs) of μ and σ^2, we need to find the values of μ and σ^2 that maximize the likelihood function. We can start by finding the MLE of μ as we did in the previous question:

To find the MLE of μ, we differentiate the likelihood function with respect to μ and set the derivative to zero:

d ln L / dμ = (1/σ^2) * ∑(xi - μ) = 0

Solving for μ, we get:

μ_MLE = (1/n) * ∑xi

Now, we can find the MLE of σ^2 by differentiating the likelihood function with respect to σ^2 and setting the derivative to zero:

d ln L / d(σ^2) = (-n/2σ^2) + (1/2(σ^2)^2) * ∑(xi - μ_MLE)^2 = 0

Solving for σ^2, we get:

σ^2_MLE = (1/n) * ∑(xi - μ_MLE)^2

Substituting the value of μ_MLE in the above equation, we get:

σ^2_MLE = (1/n) * ∑(xi - (1/n) * ∑xi)^2

Simplifying the above equation, we get:

σ^2_MLE = (1/n) * (∑xi^2 - (1/n) * (∑xi)^2)

Therefore, the MLE of σ^2 is (1/n) times the sum of the squared deviations of the observations from their sample mean.

Ivani Patel
11809154

Assignment 8

Q2. [30 pts] Maximum Likelihood Estimation (II)

Let $x_1,x_2,...,x_n$ denote the n independent observations which are assumed to be drawn from the same distribution $p(x \mid \theta)$ with defining parameter θ.

(a) [10 pts] Suppose $0 < \theta < 1$ and $p(x = 1 \mid \theta) = \theta$ while $p(x = 0 \mid \theta) = 1 - \theta$. Then, suppose m out of n observations ($m < n$) have value 1 while the rest has value 0.

Compute the maximum likelihood estimation θ_{MLE} of θ in terms of m and n.

Answer:

L(θ | x) = p(x | θ) = θ^m (1 - θ)^(n-m)

To find the maximum likelihood estimation of θ, we take the derivative of the log-likelihood function with respect to θ, set it to zero and solve for θ:

log L(θ | x) = m log(θ) + (n-m) log(1-θ)

d/dθ [log L(θ | x)] = m/θ - (n-m)/(1-θ) = 0

Solving for θ, we get:

m/θ = (n-m)/(1-θ)

θm - mθ = nθ - mθ

θm = nθ

θMLE = m/n

(b) [10 pts] Suppose $\theta > 0$ and assume the observations $x_1,x_2,...,x_n$ were independently drawn from Uniform$(0,1/\theta)$. Assuming all observations are positive, show that $p(x_i \mid \theta) = \mathrm{I}(\theta \leq 1/x_i) \cdot \theta$.

Here, $\mathrm{I}(\theta \leq 1/x_i) = 1$ if and only if $\theta \leq 1/x_i$ is true and if $x \sim$ Uniform(a,b) then $p(x \mid a,b) = 1/(b - a)$.

Answer:

We have the following information:

The observations x1,x2,...,xn are drawn independently from a uniform distribution U(a,b) with a = 0 and b = 1/θ.

The probability density function (pdf) of the uniform distribution U(a,b) is p(x | a,b) = 1/(b-a).

We need to show that p(xi | θ) = I(θ ≤ 1/xi) · θ, where I(θ ≤ 1/xi) = 1 if and only if θ ≤ 1/xi is true.

Let's start by finding the pdf of xi given θ. Since xi is drawn from a uniform distribution U(0,1/θ), we have:

p(xi | θ) = 1/(1/θ - 0) = θ/(1 - θxi)

Now, we need to show that p(xi | θ) = I(θ ≤ 1/xi) · θ. We have:

If θ > 1/xi, then p(xi | θ) = θ/(1 - θxi) > θ, and I(θ ≤ 1/xi) = 0. Therefore, p(xi | θ) = 0 when θ > 1/xi.

If θ ≤ 1/xi, then p(xi | θ) = θ/(1 - θxi) and I(θ ≤ 1/xi) = 1. Therefore, p(xi | θ) = θ/(1 - θxi) when θ ≤ 1/xi.

Putting these two cases together, we get:

p(xi | θ) = I(θ ≤ 1/xi) · θ

Therefore, we have shown that $p(xi \mid \theta) = I(\theta \le 1/xi) \cdot \theta$ when xi is drawn independently from Uniform$(0,1/\theta)$ and $\theta > 0$.

(c) [10 pts] Following the same setting in part (b), compute θ_{MLE} in terms of $(x_1,x_2,\ldots,x_n)$

Answer:

The likelihood function is given by:

$L(\theta \mid x) = \sum_{i=1}^n p(xi \mid \theta)$

$= \sum_{i=1}^n I(\theta \le 1/xi) \cdot \theta$

$= \theta^n \cdot \sum_{i=1}^n I(\theta \le 1/xi)$

We can see that θMLE must satisfy $\theta\text{MLE} \le 1/xi$ for all i in order for the likelihood to be non-zero. This implies that θMLE is upper bounded by the smallest value of 1/xi, i.e.,

$\theta\text{MLE} \le \min(1/xi)$

To find the maximum likelihood estimation of θ, we need to find the value of θ that maximizes the likelihood function. We know that $\theta\text{MLE} \le \min(1/xi)$, so we can set $\theta = \min(1/xi)$ to maximize the likelihood function. The likelihood function then becomes:

$L(\theta\text{MLE} \mid x) = (\min(1/xi))^n \cdot \sum_{i=1}^n I(\theta\text{MLE} \le 1/xi)$

Note that if there exists some j such that $\theta\text{MLE} < 1/xj$, then $I(\theta\text{MLE} \le 1/xj) = 0$ and the likelihood becomes 0. Therefore, θMLE must be equal to the minimum value of 1/xi in order for the likelihood to be non-zero. Thus, we have:

$\theta\text{MLE} = \min(1/xi)$

$= 1/\max(xi)$

Therefore, the maximum likelihood estimation of θ is the reciprocal of the largest observed value xi.

Ivani Patel
11809154

Assignment 8

Q3. [30 pts] Maximum a Posteriori

Let $x_1,x_2,...,x_n$ denote the $n > 2$ independent observations which are assumed to be drawn from the same distribution $p(x \mid \theta)$ where $\theta \sim \text{Beta}(a,b)$ with $a,b > 0$.

The probability density function (PDF) of the beta distribution is $p(\theta \mid a,b) = \theta^{a-1}(1-\theta)^{b-1}/\mathrm{B}(a,b)$ where $\mathrm{B}(a,b) = \mathrm{G}(a)\mathrm{G}(b)/\mathrm{G}(a+b)$ with G denote the Gamma function as mentioned in slides 46-47 of lecture 22. This density is non-zero only at $\theta \in (0,1)$.

(a) [15 pts] Assume $p(x \mid \theta)$ is the Bernoulli distribution as in part (a) of Q2. That is, $p(x = 1 \mid \theta) = \theta$ while $p(x = 0 \mid \theta) = 1 - \theta$ with $0 < \theta < 1$.

Given that m out of n observations ($1 < m < n$) has value 1, derive θ_{MAP} in terms of m, n, a and b.

Answer:

Using Bayes' rule, we can write the posterior PDF as:

p(θ | x1, x2, ..., xn, a, b) = p(x1, x2, ..., xn | θ) * p(θ | a, b) / p(x1, x2, ..., xn | a, b)

where p(x1, x2, ..., xn | θ) is the likelihood function of the Bernoulli distribution, p(θ | a, b) is the PDF of the Beta distribution, and p(x1, x2, ..., xn | a, b) is the marginal likelihood, which is the normalizing constant that ensures the posterior PDF integrates to 1.

Using the given Bernoulli distribution, we have:

p(x1, x2, ..., xn | θ) = θ^m * (1 - θ)^(n - m)

Using the given Beta distribution, we have:

p(θ | a, b) = θ^(a-1) * (1-θ)^(b-1) / B(a,b)

where B(a,b) is the Beta function, which is defined as B(a,b) = Γ(a) * Γ(b) / Γ(a + b), where Γ(a) is the Gamma function.

Using the fact that p(x1, x2, ..., xn | a, b) = ∫ p(x1, x2, ..., xn | θ) * p(θ | a, b) dθ, we have:

p(x1, x2, ..., xn | a, b) = ∫ θ^m * (1 - θ)^(n - m) * θ^(a-1) * (1-θ)^(b-1) / B(a,b) dθ

p(x1, x2, ..., xn | a, b) = B(m+a,n-m+b) / B(a,b)

where B(m+a,n-m+b) is the Beta function evaluated at m+a and n-m+b.

Now, substituting these expressions in the expression for the posterior PDF and taking the logarithm, we get:

log p(θ | x1, x2, ..., xn, a, b) = (m + a - 1) * log(θ) + (n - m + b - 1) * log(1 - θ) - log B(m+a,n-m+b) + log B(a,b) + C

where C is a constant that does not depend on θ.

To find the MAP estimate of θ, we need to find the value of θ that maximizes the posterior PDF. Taking the derivative of the logarithm of the posterior PDF with respect to θ and setting it to zero, we get:

(d/dθ) log p(θ | x1, x2, ..., xn, a, b) = (m + a - 1) / θ - (n - m + b - 1) / (1 - θ) = 0

Simplifying this equation, we get:

Ivani Patel
11809154

Assignment 8

θ = (m + a - 1) / (n + a + b - 2)

(b) [15 pts] Now, assume instead that each observation can take on 3 values in {0,1,2} according to the following distribution: $p(x = 0 \mid \theta) = \theta$, $p(x = 1 \mid \theta) = \theta \cdot (1 - \theta)$, and $p(x = 2 \mid \theta) = (1 - \theta)^2$ with $0 < \theta < 1$.

Given that there are n_0, n_1 and n_2 observations with values 0, 1 and 2, respectively. Derive θ_{MAP} in terms of n_0, n_1, n_2, a and b assuming that $n_0, n_1, n_2 \geq 1$.

Answer:

The likelihood function can be written as:

p(x_1, x_2, ..., x_n|θ) = θ^n0 * (θ*(1-θ))^n1 * (1-θ)^2n2

where n0, n1, and n2 are the number of observations with values 0, 1, and 2, respectively.

The prior distribution of θ is given by the beta distribution:

p(θ) = θ^(a-1) * (1-θ)^(b-1) / B(a, b)

where B(a, b) is the beta function.

Therefore, the posterior distribution of θ is:

p(θ|x_1, x_2, ..., x_n) = θ^n0 * (θ*(1-θ))^n1 * (1-θ)^2n2 * θ^(a-1) * (1-θ)^(b-1) / B(a, b)

Taking the logarithm of both sides, we get:

log p(θ|x_1, x_2, ..., x_n) ∝ n0 * log(θ) + n1 * log(θ*(1-θ)) + 2n2 * log(1-θ) + (a-1) * log(θ) + (b-1) * log(1-θ) - log B(a, b)

To find the MAP estimate of θ, we need to maximize the posterior distribution, which is equivalent to maximizing its logarithm. Taking the derivative of the logarithm of the posterior with respect to θ and setting it to zero, we get:

(n0 + n1 - 1) / θ - (n1 + b) / (1 - θ) = 0

Simplifying this expression, we get:

θ = (n0 + n1 - 1 + a) / (n0 + n1 + 2n2 + a + b - 2)

θMAP = (n0 + n1 - 1 + a) / (n0 + n1 + 2n2 + a + b - 2)