CS 108 - Statistics, Sections B

Fall 2019, AUA

Homework No. 08

Due time/date: Section B: 10:32 AM, 08 November, 2019

Note: Please use **R** only in the case the statement of the problem contains (R) at the beginning. Otherwise, show your calculations on the paper. Supplementary Problems will not be graded, but you are very advised to solve them and to discuss later with TA or Instructor.

Problem 1: MLE

a. MLE for the Poisson Distribution Parameter

Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim Pois(\lambda).$$

- Find the ML Estimator $\hat{\lambda}^{MLE}$;
- Prove that the obtained value is indeed the Global Maximum point of the Likelihood (or Log-Likelihood) function;
- Check if $\hat{\lambda}^{MLE}$ is Unbiased/Consistent;
- Calculate the Mean Squared Error for $\hat{\lambda}^{MLE}$.

b. MLE for the Rayleigh Distribution Parameter

Assume we have a Random sample $X_1, ..., X_n$ from the Rayleigh Distribution[1] with PDF

$$f(x|\sigma^2) = \begin{cases} \frac{x}{\sigma^2} \cdot e^{-\frac{x^2}{2\sigma^2}}, & \text{for } x \geq 0 \\ 0, & \text{for } x < 0. \end{cases}$$

It is known that if X is a r.v. with Rayleigh distribution with the above PDF, then

$$\mathbb{E}(X) = \sigma \cdot \sqrt{\frac{\pi}{2}}, \qquad \text{and} \qquad Var(X) = \sigma^2 \cdot \frac{4-\pi}{2}$$

1. Find the MLE $\widehat{\sigma^2}$ for the unknown Parameter σ^2;
2. Check if the ML Estimator is Unbiased/Consistent.

[1]See https://en.wikipedia.org/wiki/Rayleigh_distribution

3. Find the Method of Moments Estimator for the Parameter σ^2 using the first order Theoretical and Empirical Moments;
4. Find the Method of Moments Estimator for the Parameter σ^2 using the second order Theoretical and Empirical Moments;
5. Check if the MoM Estimators are Consistent;
6. (Supplementary) Check if the MoM Estimators are Unbiased;
7. (Supplementary) Prove the above formulas for $\mathbb{E}(X)$ and $Var(X)$.

c. MLE for Uniform Distribution

Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim Unif[a, b].$$

- Find the ML Estimators for a and b.
- (Supplementary) Check if the Estimators are Unbiased/Consistent.

d. MLE for a Discrete Parametric Distribution

Assume we have a Random Sample

$$X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{F}_\theta,$$

where $\mathcal{F}_\theta$ is given by its PMF:

Value of X	-1	1	2
$\mathbb{P}(X = x)$	$\frac{\theta}{4}$	$\frac{\theta}{4}$	$1 - \frac{\theta}{2}$,

with $\theta \in \Theta = (0, 2)$.

a. Find the MLE and MME for θ.
b. Assume we have the following observation from one of the $\mathcal{F}_\theta$, $\theta \in \Theta$:

$$2, 1, 1, 1, 2, -1, 2, -1$$

Estimate θ, using both MLE and MME.

e. MLE For the Categorical Distribution

We consider the generalization of the Bernoulli Distribution, assuming that our r.v. can take m different values (Bernoulli corresponds to the case $m = 2$, with the values 0 and 1). The Distribution of the Categorical r.v. X is given by its PMF

Values of X	1	2	...	m
$\mathbb{P}(X = x)$	p_1	p_2	...	p_m,

and we will write

$$X \sim Categorical(p_1, p_2, .., p_{m-1}, p_m).$$

Of course, here $p_i \geq 0$ and $p_1 + p_2 + ... + p_m = 1$. From the Statistical point of view, in the Categorical Distribution our Parameters are $p_1, p_2, ..., p_m$ (in fact, only $p_1, p_2, ..., p_{m-1}$, because $p_m = 1 - p_1 - ... - p_{m-1}$).

This Distribution is used in a variety of situations: say, if you are interested how would be the distribution of votes between the parties A, B, C, D in the upcoming elections[2], you can model this by using the Distribution

Values of X	A	B	C	D
$\mathbb{P}(X = x)$	p_A	p_B	p_C	p_D

,

where X is the choice of a random Person (voter). Or, if we will encode $A = 1$, $B = 2$, $C = 3$, $D = 4$ (this is to ensure that X is a r.v. - the values of a r.v. need to be numerical), we will get

Values of X	1	2	3	4
$\mathbb{P}(X = x)$	p_1	p_2	p_3	p_4

,

so

$$X \sim Categorical(p_1, p_2, p_3, p_4).$$

Now, if we want to estimate p_i-s above, we will choose a Random Sample of some size n (in our example of elections, ask n persons about their preferences) from that Distribution, and Estimate the parameters p_i.

So in this problem we will assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim Categorical(p_1, p_2, ..., p_m).$$

1. What is the Probability that we will have exactly three 1-s in $X_1, X_2, ..., X_n$, i.e., what is the Probability

$$\mathbb{P}(\text{exactly three of } X_1,\ X_2,...,\ X_n \text{ are equal to } 1).$$

2. What is the Probability that among $X_1, X_2, .., X_n$, the number of 1-s will be k_1, the number of 2-s will be $k_2, \ldots$, the number of m-s will be k_m, with $k_1 + k_2 + ... + k_m = n$?

Note: You know this from your Probability Course! Do ya?

3. Find the ML Estimator for $p_1, p_2, ..., p_m$.

Note 1: To simplify things, let us denote by k_1 the number of 1-s in $X_1, X_2, ..., X_n$, by k_2 the number of 2-s in $X_1, ..., X_n$, $\ldots$, by k_m we denote the number of m-s in $X_1, ..., X_n$.

Clearly, before observing the values of X_i-s, k_i-s are Random Variables! You can use k_i-s to form the Likelihood function.

Note 2: Your likelihood function need to be a function of $m - 1$ variables $p_1, p_2, ..., p_{m-1}$.

Note 3: You need to get very intuitive result!

[2] Or cusomers preferring "Suriki Lavash"/"Sev Lavash"/"Marus Chilingaryan Lavash"/"Dietic Lavash"

4. Assume that, in our elections example, we have asked 100 persons, and 32 of them are for A, 24 are for B, 19 are for C, and the rest are for D. Find the ML Estimates for p_A, p_B, p_C and p_D.

f. (Supplementary) MLE for Multivariate Normal Distribution

Assume we have a Random Sample from the Multivariate Normal Distribution:

$$\mathbf{X}_1, ..., \mathbf{X}_n \sim \mathcal{N}(\mu, \Sigma).$$

- Find the ML Estimators for μ and Σ;
- (**R**) Test the obtained results for 2D case in **R** - take some values for the Parameters, generate a Dataset, get the Estimates, and then visualize the results.

g. (Supplementary) MLE for Gaussian Mixture Model

We consider here the Gaussian Mixture Model with 2 Gaussians. Assume our Data comes from $\mathcal{N}(\mu_1, \sigma_1^2)$ with Probability p and from $\mathcal{N}(\mu_2, \sigma_2^2)$ with Probability $1 - p$ [3]. The PDF of this Distribution will be

$$f(x|p, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = p \cdot \varphi(x|\mu_1, \sigma_1^2) + (1-p) \cdot \varphi(x|\mu_2, \sigma_2^2),$$

where φ is the PDF of the Normal Distribution (with corresponding parameters).

- Find the MLE for the Parameters $p, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ (just write the equations to find the Parameters).
- (**R**) Take some values for the Parameters, generate in **R** a Dataset from the Mixture, find the ML Estimates for Parameters, plot on the same graph the Histogram of your Dataset and the Mixture Density with Estimated Parameter values.

[3] Say, the Histogram of the Dataset is Bimodal, like our Ararat.