Applied Statistic with R

Fall 2019, ASDS, YSU

Homework No. 05

Due time/date: 9:30 PM, 13 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: Fisher Information, MoM and MVUE

a. Fisher Information and MoM Estimator for the $Beta(\alpha, 1)$ Distribution

Consider a Parametric family of Distributions with PDF ($\alpha \in (0, +\infty)$)

$$f(x|\alpha) = \begin{cases} \alpha \cdot x^{\alpha-1}, & x \in (0,1) \\ 0, & otherwise, \end{cases}$$

- Calculate the Fisher Information $\mathcal{I}(\alpha)$;
- Find the MoM Estimator for α;
- Check the Consistency of $\hat{\alpha}^{MoM}$.

Note: This Distribution is a particular case of the **Beta**-Distribution[1] $Beta(\alpha, \beta)$, when $\beta = 1$.

b. MVUE for the Exponential Distribution

Sometimes, the Exponential Distribution is defined in the following way: we'll say that X is Exponentially distributed with the parameter β, if the PDF of X has the form

$$f(x|\beta) = \begin{cases} \frac{1}{\beta} \cdot e^{-x/\beta}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

In this case we will write $X \sim \widetilde{Exp}(\beta)$. Clearly, $X \sim \widetilde{Exp}(\beta)$ if and only if $X \sim Exp\left(\frac{1}{\beta}\right)$.

Clearly, if $X \sim \widetilde{Exp}(\beta)$, then

$$\mathbb{E}(X) = \beta \qquad \text{and} \qquad Var(X) = \beta^2.$$

Now,

[1] See, e.g., https://en.wikipedia.org/wiki/Beta_distribution

- Calculate the Fisher Information for the $\widetilde{Exp}(\beta)$ Model parameter β;
- Assuming we have a Random Sample

$$X_1, X_2, ..., X_n \sim \widetilde{Exp}(\beta),$$

  find the MoM Estimator for β;
- Prove that the MoM Estimator $\hat{\beta}^{MoM}$ is MVUE for β.
- (Supplementary) Calculate the MoM Estimator for our ordinary $Exp(\lambda)$ model parameter λ, and try to see why we are not doing the above for $\hat{\lambda}^{MoM}$.

Note: The above is equivalent of estimating $\dfrac{1}{\lambda}$ in the $Exp(\lambda)$ Model.

c. MVUE for the Normal Distribution

During our lecture, we have constructed the $\widehat{\sigma^2}^{MoM}$, and also, earlier, we have proved that this Estimator is a MVUE for the Variance σ^2. Now, we want to do this for the parameter μ of this model.

So we assume that σ^2 is known and fixed.

- Calculate the Fisher Information for the $\mathcal{N}(\mu, \sigma^2)$ Model parameter μ;
- Assuming we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2),$$

  find the MoM Estimator for μ;
- Prove that the MoM Estimator $\hat{\mu}^{MoM}$ is MVUE for μ.

d. MVUE for the Binomial Distribution

In fact, the Binomial Distribution is 2D-parametric Distribution, where the Parameters are m[2], the number of repetitions of a Trial, and p, the Probability of Success in one Trial. To simplify things, let us assume that m is known and fixed, and our aim is to Estimate p.

So we consider the Model $Binom(m, p)$, with m known and fixed.

- Calculate the Fisher Information for the $Binom(m, p)$ Model parameter p;
- Assuming we have a Random Sample

$$X_1, X_2, ..., X_n \sim Binom(m, p),$$

  find the MoM Estimator for p;
- Prove that the MoM Estimator $\hat{p}^{MoM}$ is MVUE for p.

[2] Here I am using m instead of n not to confuse with the Random Sample Size n.

Problem 2: The Method of Moments

a.

- Find the MoM Estimator for θ in the Parametric Model with the PDF ($\theta \in (0, +\infty)$)

$$f(x|\theta) = \begin{cases} 9 \cdot \theta^9 \cdot x^{-10}, & x \geq \theta \\ 0, & otherwise, \end{cases}$$

- Check the UnBiasedness and Consistency of the MoM Estimator.

b. MoM for Gamma Distribution

Gamma Distribution Family[3], is one of the important ones among all Distributions. It is a two-parametric Family of Distributions, denoted by $Gamma(\alpha, \beta)$. The aim of this problem is to Construct MoM Estimators for α and β, and, using observations, check how good are our Estimates.

- Find the formulae for the First and Second order Moments for this Distribution in Textbooks or Online (please be aware that there are two parametrizations, here, in ours, α is the *shape* parameter, and β is the *rate* parameter).
- Find the MoM Estimators for α and β
- (**R**) I have generated the following Dataset from Gamma Distribution, using **R**, and some parameters α and β:

```
##   [1]  1.8314358  3.5068022  1.9021544  1.9311683  2.4506474  1.3076168
##   [7]  3.2841308  2.1453588  6.3583929  0.9125459  1.0497036  3.5053462
##  [13]  5.2435455  1.8430702  3.1283204  2.1892085  0.8484110  2.5622832
##  [19]  1.7189725  2.7285723  2.3375566  2.9232977  5.3340547  1.2544215
##  [25]  8.7748185  4.3310692  0.8200883  4.2567581  0.3139263  3.9085270
##  [31]  4.9624075  1.9201225  3.3317165  2.7332516  1.3024626 10.8343385
##  [37]  4.6825457  2.1404540  0.7562919  3.6414628  2.1112214  3.4952583
##  [43]  0.5970462  3.5380416  5.2150621  1.3614640  2.6117846  1.8159541
##  [49]  0.5169138  2.5124691  3.5013392  3.3368801  0.8572007  6.4553792
##  [55]  8.6011775  1.4376073  0.4723840  2.8208109  1.4520227  1.2789404
##  [61]  0.7552889  1.1163648  0.7062065  1.1074022  6.8539689  2.0035534
##  [67]  3.5740196  5.7972806  5.1613411  7.2145042  7.1135339  4.2377204
##  [73]  1.1159071  4.5885815  3.5192624  0.4208828  1.0386598  2.3342677
##  [79]  1.3466932  5.5186891  1.5784883  3.9706863  5.6449191  8.1378222
##  [85]  4.2470504  3.8078827  2.0102668  1.7485158  2.2274422  2.4686840
##  [91]  2.1002233  4.9041384  2.2350189  0.3037267  1.7349850  8.2355609
##  [97]  4.4904943  3.3271395  1.8799883  1.7172326
```

Estimate α and β, using MoM, plot the Density Histogram of the Dataset over the interval $[0, 10]$, using 25 bins, and also plot over that the PDF of Gamma distribution with your calculated Estimates. Generate a next possible value for our Dataset.

[3]See https://en.wikipedia.org/wiki/Gamma_distribution

Problem 3: 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[4] 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.
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.

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

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[5], 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

,

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

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,. . . , 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 $X_1, ..., X_n$, . . . , 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!

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.