Module II revision

Statistics

Anton Afanasev

Higher School of Economics

DSBA 211
January 14, 2023

Seminar Overview

1. Quiz

2. Exam problems

3. Practice

Quiz

Find a match:

(1)	Cumulative distribution function	(A)	Waiting time
(2)	Quantile function	(B)	Separability
(3)	Pooled variance	(C)	Approximation
(4)	Exponential distribution	(D)	Antiderivative
(5)	Degrees of freedom	(E)	Weighted average
(6)	Central Limit Theorem	(F)	Goodness-of-fit
(7)	Independence	(G)	Constraints
(8)	Correlation	(H)	Inverse

Problem 1

Scrooge McDuck has a very inspiring problem! He supposes income he has is modelled by a random variable X. This random variable has c.d.f. $F(x) = x^{\theta}, \quad 0 < x < 1, \;\; \theta > 0$.
Given the sample $x_1 = 0.1, x_2 = 0.8, x_3 = 0.9$, find

1. method of moments estimate $\widehat{\theta}_{\text{MM}}$,
2. maximum likelihood estimate $\widehat{\theta}_{\text{ML}}$.

Problem 1

Here, Donald Duck came up with 3 observations

$$\begin{aligned} y_1 &= \alpha + 2\beta + \varepsilon_1, \\ y_2 &= 2\alpha + \beta + \varepsilon_2, \\ y_3 &= 3\alpha + \beta + \varepsilon_3, \end{aligned}$$

where α, β are unknown parameters, and ε_i are uncorrelated random variables with mean 0 and variance σ^2.

3. Find the least squares estimators of α, β.
4. Show that the least squares estimator of β is unbiased.

Problem 2

Walt Disney designers have completed a survey on their salaries. Wage is measured in 1000 dollars. There is a concern that teams are paid differently.

team 1	50	40	45	45	35
team 2	60	30	30	35	30

Let's assume that team 1's wage distribution is approximately normal.

1. Find 99% confidence interval for mean of team 1's wage.
2. Find 90% confidence interval for standard deviation of team 1's wage.
3. Formulate assumptions, which are necessary for construction of a confidence interval for a difference in salary between team 1 and team 2.
4. Considering assumptions from (3) true, find 90% confidence interval for a difference between means of team 1 and team 2 salaries. Can we claim that both team are paid equally?

Problem 3

Winnie-the-Pooh decided to invite his helpers in order to give all the gifts. The number of children whom the assistants should visit with gifts is randomly determined. A random variable X defines a number of kids. Despite the sawdust in his head, Winnie-the-Pooh assumes that X has a binomial distribution with $n = 10$ and $p = 0.8$.

1. Your partner asks you to estimate the probability that you will need to visit at least 8 children.
2. Piglet said that he heard a rumor that X has a binomial distribution with $n = 1000$. What is the probability that you will have to visit at least 800 children?

Problem 3

3. Rabbit came to you and said that Winnie-the-Pooh decided to look at Statistics courses on Coursera. Based on his observations last year, he realized that the number of gifts delivered during one hour is Poisson distributed and every hour, on average, gifts were given to 4 children. If we assume that this year the activity of the team will remain at the same level, then how much time will pass before the first gift from the start?
4. Christopher Robin who plan to apply to DSBA says that arrival of buses at a given bus station follows Poisson law with rate 2. The arrival of taxis at the same bus stop is also Poisson, with rate 3. What is the probability that next time he'll go to the bus stop he'll see at least two taxis arriving before a bus? Exactly two taxis?

Problem 4

A random variable X has density function

$$f_X(x) = a + bx^2,$$

over the range $(0, 2)$, where a, b are constants. The mean of X is 1.5.

1. Find a, b.
2. Find the cumulative distribution function of X.
3. Find the variance of X.

Problem 5

Find the estimator for unknown parameter θ from distribution

$$f(x) = e^{-(x-\theta)} \cdot I_{\{x>0\}}$$

using method of moments for sample $X_1, \ldots, X_n$.

Problem 6

Distribution of X is uniform $\mathcal{U}(-a, a)$. Sample of size $n = 2$ is available. Consider $\widehat{a} = c \cdot (|x_1| + |x_2|)$ as a class of estimators for the parameter a. Find c such that

1. Estimator $\widehat{a}$ is unbiased.
2. Estimator $\widehat{a}$ is the most efficient in the class. (In terms of mean square error.)

Problem 7

Consider random variables X and Y with joint density function

$$f(x,y)=\begin{cases}\dfrac{1}{2}+cx, & x+y\leq 1,\ x\geq 0,\ y\geq 0,\\ 0, & \text{otherwise.}\end{cases}$$

1. Find c.
2. Find $f_X(x)$. Evaluate $\mathsf{E}(X)$.
3. Write down an expression for $f_{Y|X}(x,y)$. Find $\mathsf{E}(Y \mid X=x)$.

Problem 8

Suppose 2000 points are selected independently at a random from the unit square $S = \{(x, y) : 0 \leq x \leq 1,\ 0 \leq y \leq 1\}$. Let W be the number of points that fall into the set $A = \left\{(x, y) :\ x^2 + y^2 < 1\right\}$.

1. How is W distributed?
2. Find the mean, variance and standard deviation of W.
3. Estimate probability that W is greater than 1600.

Problem 9

Let X and Y be two independent standard normal random variables. Find

1. $\mathsf{P}(|X+Y| > |X-Y|)$.
2. $\mathsf{P}(|X+Y| > 2|X-Y|)$.

Problem 10

Suppose that student's grade for a statistics exam, X, has continuous uniform distribution at the interval $[0, 100]$. But less then 25 points means "failed", and more than 80 points is "excellent", hence the final grade Y is calculated as follows:

$$Y = \begin{cases} 0, & X < 25 \\ X, & 25 \leq X < 80 \\ 100, & X \geq 80 \end{cases}$$

1. Find c.d.f. of Y. Sketch the plot.
2. Find p.d.f. of Y. Sketch the plot.
3. Find mean and variance of X and Y.
4. Find $\mathsf{E}(Y \mid Y > 0)$.
5. Find $\mathsf{Corr}(X, Y)$.

Problem 11

A random sample of 400 married couples was selected from a large population of married couples.

- Heights of married men are approximately normally distributed with mean 70 inches and standard deviation 3 inches.
- Heights of married women are approximately normally distributed with mean 65 inches and standard deviation 2.5 inches.
- There were 20 couples in which wife was taller than her husband, and there were 380 couples in which wife was shorter than her husband.

Problem 11

1. Find a 95% confidence interval for the proportion of married couples in the population for which the wife is taller than her husband.
2. Suppose that a married man is selected at random and a married woman is selected at random. Find the approximate probability that the woman will be taller than the man.
3. Based on your answers to (1) and (2), are the heights of wives and their husbands independent? Explain your reasoning.

Problem 12

Internal angles $\theta_1, \theta_2, \theta_3, \theta_4$ of a certain quadrilateral, located on the ground, were measured by the aerial system. It is assumed that those observations x_1, x_2, x_3, x_4 were taken with minor and independent errors, which have zero mean and identical variance σ^2.

1. Find LSE of $\theta_1, \theta_2, \theta_3, \theta_4$.
2. Find an unbiased estimate of σ^2 in the case, described in part (1).
3. Let's assume now that the considered quadrilateral is a parallelogram with $\theta_1 = \theta_3$ and $\theta_2 = \theta_4$. How values of internal angles LSE would change? Find an unbiased estimate of σ^2 in this particular case.

Look at the time!