Exam
«Introduction to mathematical statistics»

DEADLINE: Thursday, December 19th, 23:59

Final Mark Formula: $M = \min(10, (\text{total score}) - 1)$. Scores for each problem are given below.

The solutions may be submitted either in paper form or sent to **both** of the following emails: `sashaskrip@gmail.com`, `klimenko05@mail.ru`.

In the latter case all solutions should be presented in **one** pdf-file, with pages in **the correct order and the correct orientation**. If the file is too large to be sent by e-mail, it can be put into an online storage (Yandex.Disk, Google Drive etc.) and the link should be emailed to the addresses given above. Scans/photos in other formats will be rejected.

A solution for Problem 8 should be accompanied with a file with numerical calculations, which can be a spreadsheet, computer program source code and output, etc. Your data for the problem can be found in the table presented alongside this file (see the line with your name). If you are not listed in this file, please inform us by e-mail.

Version from December 18. The fixed typos in Problem 2 and 3 are shown in red.

Problem 1 (0.5 + 0.5)**.** For a given strictly positive parameter λ consider the following function f:

$$f = \begin{cases} Cx^2, & \lambda < x < 2\lambda, \\ 0, & \text{otherwise.} \end{cases}$$

a) Determine $C = C(\lambda)$ such that f is a density function of some continuous random variable X.

b) Let X be a random variable with this density function, λ being an unknown parameter. Find $a \in \mathbb{R}$ such that $T = a/X$ is an unbiased estimator for $1/\lambda$.

Problem 2 (0.5+0.5)**.** We consider a random sample of size n taken from a discrete distribution with probability mass function $f(x;\theta) = 1/\theta$, $x = 1, 2, \ldots, \theta$ (and 0 elsewhere), where θ is an unknown positive integer.

a) Show that the largest observation of the sample (we denote it by T) is a complete sufficient statistic for θ.

b) Prove that the unique efficient estimator for θ (aka MVUE) is $\dfrac{T^{n+1} - (T-1)^{n+1}}{T^n - (T-1)^n}$.

Problem 3 (0.5 × 5)**.** Consider a random sample $(X_1, \ldots, X_n)$ where all X_i are independent identically distributed random variables with the following density function:

$$f(x, \theta) = Cx^{-(p+1)} \exp(-\theta/x) \mathbf{1}_{[0,+\infty)}(x).$$

Here $\theta > 0$ is an unknown real parameter, $p > 0$ is a known number and $\mathbf{1}_A$ is the indicator function of the set A.

a) Find the value of $C = C(p, \theta)$, prove that $U_i = 1/X_i$ have the gamma distribution, and determine its shape and scale parameters, mean and variance.

b) Find a complete sufficient statistics $T(X_1, \ldots, X_n)$ for θ.

c) Find the maximum likelihood estimator $\hat{\theta}_n$ for θ.

d) Determine $\mathbb{E}\hat{\theta}_n$ and $\mathrm{Var}(\hat{\theta}_n)$.

e) Check if $\hat{\theta}_n$ is biased? asymptotically biased? effective?

Problem 4 (0.5 + 0.5)**.** An electrical engineer made a probabilistic model for the life time of her electric circuit. She got a density function proportional to $e^{-\gamma x}$ with the support ray $[a, +\infty)$, where a and γ are two unknown strictly positive parameters.

a) Determine the estimators $\hat{a}$ and $\hat{\gamma}$ using the method of moments.

b) Check if these estimators are consistent.

Problem 5 (0.5)**.** Consider a sample $(X_1, \ldots, X_n)$ from a distribution with unknown average m and variance m^2. Find the asymptotic confidence interval for m with the given confidence level α.

Problem 6 (1)**.** Members of the Russian parliament formed a commission to re-test the scientific value of the results of famous experiments by Gregor Mendel. In his experiments with the pea plant, Mendel noticed that there are 315 round yellow seeds, 108 round green, 101 wrinkled yellow seeds and 32 wrinkled green. In accordance with Mendel's theory, the proportion in which seeds are divided into these groups is given by the ratio $9:3:3:1$. The commission has to decide whether the theory is correct or not with 95% level of confidence. Should they accept Mendel's theory or not? Please justify your answer.

Problem 7 (0.5 + 0.5 + 2)**.** We consider a random sample $(X_1, \ldots, X_n)$ where all X_i are independent identically distributed random variables with the following density function:

$$f(x) = e^{\theta - x}\mathbf{1}_{[\theta, +\infty)}(x),$$

where θ is an unknown parameter.

a) Find the estimation of maximal likelihood.

b) Find the sufficient statistics.

c) Using two previous points, construct a criterium to test the hypothesis $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$ on a level of confidence α. Check the consistency of this criterium.

Problem 8 (1 + 1 + 1)**.** Mr. Alzheimer, a farmer from Germany sowed two types of carrot seeds (A and B) into his garden, making two garden beds, but forgot to mark the lines — where which type was sowed. It is known that the logarithms of mass (in grams) of the root crops of each type have the normal distribution with the same variance σ^2 and the different means $a_A < a_B$. So, once the carrots were ready, Alzheimer collected them all and weighed carefully: the root crops taken from the first garden bed formed the following list (in grams): $X_1, \ldots, X_n$, while carrots from the second garden bed were of $Y_1, \ldots, Y_m$ grams respectively.

a) Find the optimal test that decides on which line which type of seeds was sowed.

b) Apply the test to the values from the file attached. Consider two assumptions ("A is sowed in the first garden bed" with the alternative "A is sowed in the second garden bed" and the opposite one) and find p-values for both of them.

c) Assuming that Mr. Alzheimer really made a fair random choice, so the a priori probability of each assumption being equal to 1/2, find their a posteriori probabilities.