Federal State Autonomous Educational Institution of Higher Education "Moscow Institute of Physics and Technology (National Research University)"

APPROVED
Head of the Phystech School of Applied Mathematics and Informatics
A.M. Raygorodskiy

Work program of the course (training module)

course:	Statistics/Статистика
major:	Information Science and Computer Engineering
specialization:	Computer Science/Информатика
	Phystech School of Applied Mathematics and Informatics
	Chair of Discrete Mathematics
term:	3
qualification:	Bachelor

Semester, form of interim assessment: 6 (spring) - Grading test

Academic hours: 60 AH in total, including:
lectures: 30 AH.
seminars: 30 AH.
laboratory practical: 0 AH.

Independent work: 75 AH.

In total: 135 AH, credits in total: 3

Author of the program: A.M. Raygorodskiy, doctor of physics and mathematical sciences, head of chair

The program was discussed at the Chair of Discrete Mathematics 05.03.2020

Annotation

The statistics course is a compulsory discipline of the basic part. Within the framework of this discipline, students study the following topics: convergence of random vectors, statistics and estimates, methods for finding estimates, effective estimates, sufficient statistics and optimal estimates, Bayesian estimates, confidence estimation, linear regression model, hypothesis testing and uniformly the most powerful criteria, criteria, normality-based, goodness-of-fit criteria, independence test criteria.

1. Study objective

Purpose of the course

studying the mathematical and theoretical foundations of modern statistical analysis, as well as preparing students for further independent work in the field of analysis of statistical problems in applied mathematics, physics and economics.

Tasks of the course

- studying the mathematical foundations of mathematical statistics;
- acquisition of theoretical knowledge in the field of modern statistical analysis by students.

2. List of the planned results of the course (training module), correlated with the planned results of the mastering the educational program

Mastering the discipline is aimed at the formation of the following competencies:

Code and the name of the competence	Competency indicators
UC-1 Search and identify, critically assess and synthesize information, apply a systematic approach to problem-solving	UC-1.3 Consider various options for solving a problem, assess the advantages and disadvantages of each option
Gen.Pro.C-1 Apply fundamental knowledge acquired in the physical and mathematical fields and/or natural sciences and use it in professional settings	Gen.Pro.C-1.1 Analyze the task in hand, outline the ways to complete it
Gen.Pro.C-3 Write scientific and/or technical (technological, innovative) reports (publications, projects)	Gen.Pro.C-3.2 Employ practical methodologies for preparing scientific and technical reports (projects)
Pro.C-1 Assign, formalize, and solve tasks, develop and research mathematical models of the studied phenomena and processes, systematically analyze scientific problems, obtain new scientific outcomes	Pro.C-1.1 Locate, analyze, and summarize information on current research findings within the subject area

3. List of the planned results of the course (training module)

As a result of studying the course the student should:

know:

- basic concepts of mathematical statistics;
- basic approaches to comparing estimates of parameters of an unknown distribution;
- asymptotic and non-asymptotic properties of estimates of parameters of an unknown distribution;
- basic methods for constructing estimates with good asymptotic properties: method of moments, method of maximum likelihood, method of sample quantiles;
- the concept of effective estimates and inequality of information by Rao-Cramer;
- definition and main properties of the conditional mathematical expectation of a random variable relative to sigma-algebra or other random variable;
- definition of a general linear regression model and least squares method;
- multivariate normal distribution and its basic properties;
- basic concepts of the theory of testing statistical hypotheses;
- Neumann - Pearson lemma and monotonic likelihood ratio theorem;
- pearson chi-square test for testing simple hypotheses in the Bernoulli scheme.

be able to:

- justify the asymptotic properties of estimates using the limit theorems of probability theory;
- construct estimates with good asymptotic properties for the parameters of an unknown distribution for a given sample from it;
- find Bayesian estimates for a given prior distribution;
- calculate conditional mathematical expectations using conditional distributions;
- find optimal estimates using complete sufficient statistics;
- build exact and asymptotic confidence intervals and areas for the parameters of the unknown distribution;
- find optimal estimates and confidence regions in a Gaussian linear model;
- build uniformly the most powerful criteria in the case of a parametric family with a monotonic likelihood ratio;
- build an F-test to test linear hypotheses in a linear Gaussian model.

master:

- the main methods of mathematical statistics for constructing point and confidence estimates: the method of moments, sampling quantiles, maximum likelihood, the method of least squares, the method of central statistics.
- skills of asymptotic analysis of statistical tests;
- skills of applying the theorems of mathematical statistics in applied problems of physics and economics.

4. Content of the course (training module), structured by topics (sections), indicating the number of allocated academic hours and types of training sessions

4.1. The sections of the course (training module) and the complexity of the types of training sessions

№	Topic (section) of the course	Types of training sessions, including independent work			
		Lectures	Seminars	Laboratory practical	Independent work
1	Probabilistic-statistical model.	7	9		15
2	The main task of mathematical statistics.	8	9		15
3	Various kinds of convergence of random vectors.	5	4		15
4	Statistics and estimates.	5	4		15
5	Empirical distribution and empirical distribution function.	5	4		15
AH in total		30	30		75
Exam preparation		0 AH.			
Total complexity		135 AH., credits in total 3			

4.2. Content of the course (training module), structured by topics (sections)

Semester: 6 (Spring)

1. Probabilistic-statistical model.

Examples of unbiased and consistent estimates (moments, variance); biased but consistent estimates; inconsistent but unbiased estimates. Estimates of functions from parameters. An example of a situation in which there is no unbiased estimate for some function of a parameter.

2. The main task of mathematical statistics.

Bayesian and minimax strategies. Minimaxity of Bayesian strategy with constant risk.

3. Various kinds of convergence of random vectors.

Theorems on the asymptotic normality of the sample mean and median in a symmetric distribution model with an unknown shift parameter.

4. Statistics and estimates.

A reminder of the Three Sigma Rule and an explanation in terms of this rule. Example with "mixed" normal distribution (median vs. sample mean).

5. Empirical distribution and empirical distribution function.

Maximum likelihood estimates (m.p.) and their properties (consistency, asymptotic normality, and efficiency). Omp for the shift parameter in the Laplace distribution as an example of an asymptotically normal omp in an irregular pattern.

5. Description of the material and technical facilities that are necessary for the implementation of the educational process of the course (training module)

Standard classroom.

6. List of the main and additional literature, that is necessary for the course (training module) mastering

Main literature

1. Математическая статистика [Текст] : [учебник для вузов] / А. А. Боровков .— [3-е изд., испр.] .— М. : Физматлит, 2007 .— 704 с.
2. Введение в математическую статистику [Текст] : [учебник для вузов] / Г. И. Ивченко, Ю. И. Медведев .— М. : ЛКИ, 2010, 2014, 2015 .— 600 с.
3. Теория вероятностей и математическая статистика [Текст] : учеб. пособие для вузов / П. П. Бочаров, А. В. Печинкин .— М. : Физматлит, 2005 .— 295 с. : ил. + pdf-версия. - Библиогр.: с. 292. - ISBN 5-9221-0633-3. — Полный текст (Доступ из сети МФТИ / Удаленный доступ).

Additional literature

1. Наглядная математическая статистика [Текст] : учеб. пособие для вузов / М. Б. Лагутин .— 2-е изд., испр. — М. : Бином. Лаб. знаний, 2009 .— 472 с.

7. List of web resources that are necessary for the course (training module) mastering

Not used

8. List of information technologies used for implementation of the educational process, including a list of software and information reference systems (if necessary)

Multimedia technologies can be used in lectures and practical exercises, including presentations.

9. Guidelines for students to master the course

1. It is recommended to successfully pass test papers, as this simplifies the final certification in the subject.
2. To prepare for the final certification in the subject, it is best to use the lecture materials.

SUPPLEMENT

Assessment funds for course (training module)

major:	Information Science and Computer Engineering
specialization:	Computer Science/Информатика
	Phystech School of Applied Mathematics and Informatics
	Chair of Discrete Mathematics
term:	3
qualification:	Bachelor

Semester, form of interim assessment: 6 (spring) - Grading test

Author: A.M. Raygorodskiy, doctor of physics and mathematical sciences, head of chair

1. Competencies formed during the process of studying the course

Code and the name of the competence	Competency indicators
UC-1 Search and identify, critically assess and synthesize information, apply a systematic approach to problem-solving	UC-1.3 Consider various options for solving a problem, assess the advantages and disadvantages of each option
Gen.Pro.C-1 Apply fundamental knowledge acquired in the physical and mathematical fields and/or natural sciences and use it in professional settings	Gen.Pro.C-1.1 Analyze the task in hand, outline the ways to complete it
Gen.Pro.C-3 Write scientific and/or technical (technological, innovative) reports (publications, projects)	Gen.Pro.C-3.2 Employ practical methodologies for preparing scientific and technical reports (projects)
Pro.C-1 Assign, formalize, and solve tasks, develop and research mathematical models of the studied phenomena and processes, systematically analyze scientific problems, obtain new scientific outcomes	Pro.C-1.1 Locate, analyze, and summarize information on current research findings within the subject area

2. Competency assessment indicators

As a result of studying the course the student should:

know:

- basic concepts of mathematical statistics;
- basic approaches to comparing estimates of parameters of an unknown distribution;
- asymptotic and non-asymptotic properties of estimates of parameters of an unknown distribution;
- basic methods for constructing estimates with good asymptotic properties: method of moments, method of maximum likelihood, method of sample quantiles;
- the concept of effective estimates and inequality of information by Rao-Cramer;
- definition and main properties of the conditional mathematical expectation of a random variable relative to sigma-algebra or other random variable;
- definition of a general linear regression model and least squares method;
- multivariate normal distribution and its basic properties;
- basic concepts of the theory of testing statistical hypotheses;
- Neumann - Pearson lemma and monotonic likelihood ratio theorem;
- pearson chi-square test for testing simple hypotheses in the Bernoulli scheme.

be able to:

- justify the asymptotic properties of estimates using the limit theorems of probability theory;
- construct estimates with good asymptotic properties for the parameters of an unknown distribution for a given sample from it;
- find Bayesian estimates for a given prior distribution;
- calculate conditional mathematical expectations using conditional distributions;
- find optimal estimates using complete sufficient statistics;
- build exact and asymptotic confidence intervals and areas for the parameters of the unknown distribution;
- find optimal estimates and confidence regions in a Gaussian linear model;
- build uniformly the most powerful criteria in the case of a parametric family with a monotonic likelihood ratio;
- build an F-test to test linear hypotheses in a linear Gaussian model.

master:

- the main methods of mathematical statistics for constructing point and confidence estimates: the method of moments, sampling quantiles, maximum likelihood, the method of least squares, the method of central statistics.
- skills of asymptotic analysis of statistical tests;
- skills of applying the theorems of mathematical statistics in applied problems of physics and economics.

3. List of typical control tasks used to evaluate knowledge and skills

Examples of homework assignments:

1.Find the optimal estimate of the parameter> 0 based on a sample from the distribution: a) N (, 1), b) R (0,), c) Pois (), d) Bin (1,) (here (0,1)) ... 6. Let X1, ..., Xn be a sample from a uniform distribution on the interval [0,],> 0. Construct the confidence interval for the confidence level using the statistics a) X, b) X (1), c) X (n) ...

2 .. Let X1, ..., Xn be a sample from a normal distribution with parameters (, 1). Find the Bayesian estimate of the parameter if the prior distribution is Bin (1, p). Will the resulting estimate be a consistent estimate of the parameter?

3. There are 2 objects with weights a and b. We weighed the first, second, and both objects together with errors, and the variance of the error in the latter case was 4 times greater. Reduce the problem to a linear regression model and find the least squares estimates for a and b.

4. X1, ..., Xn sample from exponential distribution with parameter. Construct evenly the most powerful criterion for the significance level of testing the hypothesis H0: = 0 against the alternative a) H1:> 0, b) H1: <0.

4. Evaluation criteria

1. The main task of mathematical statistics. Parametric statistics. Estimates ("statistics"). Variational series and ordinal statistics. Reminder of the distribution of ordinal statistics.
2. Unbiasedness and consistency of estimates. Examples of unbiased and consistent estimates (moments, variance); biased but consistent estimates; inconsistent but unbiased estimates. Estimates of functions from parameters. An example of a situation in which there is no unbiased estimate for some function of a parameter.
3. Empirical distribution function. Its unbiasedness and consistency in relation to the theoretical distribution function.
4. Penalty (loss) function and risk function. Bayesian and minimax strategies. Minimaxity of Bayesian strategy with constant risk.
5. Bayesian estimate in the Bernoulli scheme. Minimax estimate in the Bernoulli scheme (Hodges - Lehmann statistics). Comparison of the Hodges - Lehmann statistics with the sample mean.
6. Asymptotic normality of the estimate. Theorems on the asymptotic normality of the sample mean and median in a symmetric distribution model with an unknown shift parameter.
7. Relative asymptotic efficiency. A reminder of the Three Sigma Rule and an explanation in terms of this rule. Example with "mixed" normal distribution (median vs. sample mean).
8. Regularity conditions and the Rao - Cramer inequality (the case of an unbiased estimate). Fischer's information. Effective and super effective assessments.
9. Method of moments. Consistency of the assessment.
10. Maximum likelihood method. Maximum likelihood estimates (m.p.) and their properties (consistency, asymptotic normality, and efficiency). Omp for the shift parameter in the Laplace distribution as an example of an asymptotically normal o.m.s. in an irregular pattern.
11.O.m.p. in standard models (binomial, normal, Cauchy, etc.).
12. Sufficient statistics in the discrete case. Factorization criterion. Examples.
13. Conditional expectation of the partition. Conditional densities. Properties of the conditional expectation (b / d in the "continuous" case, but with a proof in the case of the conditional expectation with respect to the partition).
14. Sufficient statistics in general: general factorization criterion as a definition. Examples.
15. Exponential family. Pessimistic theorem (b / d).
16. The Kolmogorov - Blackwell - Rao theorem on improving an unbiased estimate.
17. Complete statistics. Completeness and sufficiency criterion for an exponential family (b / d).
18. The Glivenko - Cantelli theorem (formulation). General problem of uniform convergence in the laws of large numbers. The problem of triangles in the plane. Vapnik - Chervonenkis dimension. An example with space and open half-spaces. Two theorems of Vapnik - Chervonenkis (b / d). The Glivenko - Kantelli theorem as a special case of the Vapnik - Chervonenkis theorem. The triangle theorem as a special case of the Vapnik - Chervonenkis theorem.
19. Kolmogorov statistics and its independence from the form of a continuous distribution function (only strictly monotone case). Kolmogorov's asymptotic theorem (b / d).

20. Exact and asymptotic confidence intervals. One-sided and two-sided spacing.
21. Quantiles of distributions. Chi-square distribution reminder. Student's distribution.
22. Construction of confidence intervals for the parameters of the normal distribution: for the mean with a known variance, for the variance with a known mean and for simultaneously unknown mean and variance.
23. Confidence intervals for quantiles.
24. Kolmogorov's agreement criterion.
25. Criterion of agreement Pearson's chi-square.
26. Comparison of two simple hypotheses. Errors of the first and second kind.
27. The case of absolutely continuous distributions: likelihood ratio and the Neumann - Pearson theorem.
28. Smirnov's homogeneity criterion for two independent samples.
29. Criterion of homogeneity chi-square for several independent samples.
30. Checking the equality of means of two normal samples with known variances; with unknown variances. Checking the equality of variances in two normal samples with unknown means. Fisher's distribution.

- the mark "excellent (10)" is given to a student who has shown comprehensive, systematized, in-depth knowledge of the curriculum of the discipline and the ability to confidently apply them in practice in solving specific problems, free and correct justification of the decisions

- the mark "excellent (9)" is given to a student who has shown comprehensive, systematized, in-depth knowledge of the curriculum of the discipline and the ability to apply them in practice in solving specific problems, free and correct justification of decisions

- the mark "excellent (8)" is given to a student who has shown a comprehensive systematized, deep knowledge of the curriculum of the discipline and the ability to apply them in practice when solving specific problems, and the correct justification of the decisions made

- the mark "good (7)" is given to a student if he firmly knows the material, expresses it competently and to the point, knows how to apply the knowledge gained in practice, but makes some inaccuracies in the answer or in solving problems;

- the mark "good (6)" is given to the student if he knows the material, presents it competently and in essence, knows how to apply the knowledge gained in practice, but makes some inaccuracies in the answer or in solving problems;

- the mark "good (5)" is given to a student if he knows the material, and essentially expounds it, knows how to apply the knowledge gained in practice, but makes some inaccuracies in the answer or in solving problems;

- the mark "satisfactory (4)" is given to a student who has shown a fragmented, scattered nature of knowledge, insufficiently correct formulations of basic concepts, a violation of the logical sequence in the presentation of the program material, but at the same time he owns the main sections of the curriculum necessary for further education and can apply the obtained knowledge by model in a standard situation;

- the mark "satisfactory (3)" is given to a student who has shown a fragmentary, scattered nature of knowledge, insufficiently correct formulations of basic concepts, violation of the logical sequence in the presentation of program material, but at the same time he has fragmentary knowledge of the main sections of the curriculum necessary for further education and can apply the knowledge gained by the model in a standard situation;

- the mark "unsatisfactory (2)" is given to a student who does not know most of the main content of the curriculum of the discipline, makes gross mistakes in the formulation of the basic concepts of the discipline and does not know how to use the knowledge gained in solving typical practical problem;

- grade "unsatisfactory (1)" is given to a student who does not know the formulations of the basic concepts of the discipline.

5. Methodological materials defining the procedures for the assessment of knowledge, skills, abilities and/or experience

During the test, students can use the discipline program.