Skoltech
Skolkovo Institute of Science and Technology

Course Syllabus

| | |
|---|---|
| **Course Title (in English)** | Principles of Applied Statistics |
| **Course Title (in Russian)** | Прикладная статистика |
| **Lead Instructor(s)** | Panov, Maxim |
| **Is this syllabus complete, or do you plan to edit it again before sending it to the Education Office?** | The syllabus is a work in progress (draft, won't appear anyware) |
| **Contact Person** | Maxim Panov |
| **Contact Person's E-mail** | m.panov@skoltech.ru |

# 1. Annotation

**Course Description**

Standard courses in mathematical statistics focus on classical statistical methods. However, in practice, modern statistical methods are often used, for example, bootstrap, nonparametric estimation, smoothing based on decomposition in orthogonal bases, methods for reducing dimensionality and sensitivity analysis, etc. Understanding the theory underlying these methods, as well as the ability to apply them in practice, is absolutely necessary for anyone working in mathematical statistics and data analysis.

**Аннотация**

В типичных курсах математической статистики основное внимание уделено классическим статистическим методам. Однако, на практике зачастую применяются современные методы статистики, например, бутстреп, непараметрическое оценивание, сглаживание на основе разложения по ортогональным базисам, методы снижения размерности и анализа чувствительности и т.д. Понимание теории, лежащей в основе этих методов, а также умение применить их на практике, абсолютно необходимо каждому, кто занимается математической статистикой и анализом данных.

# 2. Structure and Content

| | |
|---|---|
| **Course Academic Level** | Master-level |
| **Number of ECTS credits** | 3 |

| Topic | Summary of Topic | Lectures (# of hours) | Seminars (# of hours) | Labs (# of hours) |
|---|---|---|---|---|
| Main problems and methods of mathematical statistics | Parametric and nonparametric models. Main problems: point estimates, confidence sets, hypothesis thesting, prediction. Empirical distribution function. Statistical functionals. | 2 | | |
| Bootstrap | Monte-Carlo modelling, bootstrap. Variance Estimation based on bootstrap. Confidence intervals based on bootstrap. Jackknife. | 1 | 2 | |
| Parametric estimation | Method of moments. Maximum likelihood estimation. Delta-method. Parametric bootstrap. Confidence intervals. Sufficient statistic. Exponential family. EM-algorithm. | 2 | 2 | |
| Hypothesis testing | Basics of hypothesis testing. Wald criterion. p-values. Chi-squared distribution and Pearson criterion. Permutation test. Likelihood ratio test. Multiple testing. | 2 | 2 | |
| Nonparametric estimation | Bias-variance tradeoff. Histograms. Kernel density estimation. Non-parametric regression. | 2 | 2 | |
| Bayesian estimation | Bayes formula. Gaussian processes. | 2 | 2 | |
| Design of Experiments | Space-filling Design of Experiments. Optimal design of experiments for linear regression models. Design of experiments based on Gaussian processes. Bayesian optimization. | 2 | 2 | |
| Sampling and MCMC | Markov Chain Monte Carlo | 2 | 2 | |
| Variational Inference | Variational Inference. Normalizing Flows | 2 | 2 | |

# 3. Assignments

| Assignment Type | Assignment Summary |
|---|---|
| Homework | Series of problems on hypothesis testing |
| Homework | Experimental comparison of different methods for density estimation |
| Homework | Experimental comparison of different methods for sampling |
| Team Project | Final project |

# 4. Grading

| | |
|---|---|
| **Type of Assessment** | Graded |
| **Grade Structure** | |

| Activity Type | Activity weight, % |
|---|---|
| Homework Assignments | 50 |
| Midterm Exam | 20 |
| Team Project | 30 |

## Grading Scale

| | |
|---|---|
| **A:** | 86 |
| **B:** | 76 |
| **C:** | 66 |
| **D:** | 56 |
| **E:** | 46 |
| **F:** | 0 |
| **Attendance Requirements** | Optional with Exceptions |

# 5. Basic Information

**Maximum Number of Students**

| | Maximum Number of Students |
|---|---|
| **Overall:** | 50 |
| **Per Group (for seminars and labs):** | 25 |

| | |
|---|---|
| **Course Stream** | Science, Technology and Engineering (STE) |
| **Course Term (in context of Academic Year)** | Term 2 |
| **Course Delivery Frequency** | Every year |

**Students of Which Programs do You Recommend to Consider this Course as an Elective?**

| Masters Programs | PhD Programs |
|---|---|
| Data Science<br>Information Science and Technology | |

| | |
|---|---|
| Course Tags | Math<br>Statistics |

## 6. Textbooks and Internet Resources

| Required Textbooks | ISBN-13 (or ISBN-10) |
|---|---|
| Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: data mining, inference, and prediction. Springer, 2001. | 9780387848570 |
| Wasserman L. All of Nonparametric Statistics. Springer, 2006. | 9780387251455 |

## 7. Facilities

| Software |
|---|
| Python |

## 8. Learning Outcomes

| Knowledge |
|---|
| How the ideas from mathematical statistics can be applied in modern methods of data analysis and processing |

| Skill |
|---|
| Be able to formulate in mathematical terms a real-world problem, built a corresponding probabilistic model, select an appropriate statistical method |

| Experience |
|---|
| Obtain a sufficient experience during practical exercises and project activities to become a qualified user of statistical methods. |

# 9. Assessment Criteria

Input or Upload Example(s) of Assigment 1:

Input or Upload Example(s) of Assigment 2:

Input or Upload Example(s) of Assigment 3:

Input or Upload Example(s) of Assigment 4:

Input or Upload Example(s) of Assigment 5:

## 10. Additional Notes