Probability Theory and Statistics

Module Description

Module code	MAT202-S28	**Credits**	8 ECTS
Programme	Foundation programme	**Level**	Semester 4
Semester	Spring 2028	**Prerequisites**	MAT102-S27 Linear Algebra; MAT231-F27 Calculus II
Module leader	Dr. Ali Khudiyev	**Email**	ali.khudiyev@wosm.academy
Office hours	Monday–Friday, 10:00–17:00 (Baku time, UTC+4)		

1. Module Overview

Probability Theory and Statistics introduces the mathematics of uncertainty and of learning from data. The first part builds probability theory from axioms; the second part uses it to draw conclusions from data with estimation, confidence intervals, hypothesis tests and regression.

The module covers:

- Probability spaces, conditional probability, random variables and their distributions (first 5–6 weeks)
- Expectation, variance, joint distributions, tail bounds and the limit theorems
- Statistical inference: estimation, confidence intervals, hypothesis testing and linear regression
- Computational methods: simulation, Monte Carlo and the bootstrap

Examples come from computer science throughout: randomised algorithms and hashing, the analysis of average-case running time, A/B testing, and the probabilistic foundations of machine learning. Students run simulations and analyse data in short programming assignments.

2. Aims

- Give a rigorous introduction to probability theory, using the calculus and linear algebra of earlier modules.
- Develop the probabilistic reasoning needed to design and analyse randomised algorithms and systems.
- Introduce the main methods of statistical inference and the assumptions behind them.
- Prepare students for machine learning, data science, information theory and networking.

3. Learning Outcomes

On successful completion of this module, students will be able to:

LO1. Model random experiments with probability spaces and compute probabilities using counting, conditional probability, independence and Bayes' theorem.

LO2. Work with standard discrete and continuous distributions and compute expectations, variances and distributions of functions of random variables.

LO3. Handle joint distributions, covariance and conditional expectation, including the multivariate normal distribution.

LO4. Apply linearity of expectation, indicator variables and the Markov, Chebyshev and Chernoff bounds to analyse randomised algorithms.

LO5. State and apply the law of large numbers and the central limit theorem.

LO6. Construct estimators by the method of moments and maximum likelihood, assess their bias and variance, and build confidence intervals.

LO7. Carry out and interpret hypothesis tests and simple and multiple linear regression, and recognise common misuses of p-values.

LO8. Use simulation, Monte Carlo methods and the bootstrap to approximate probabilities and quantify uncertainty in code.

4. Syllabus

Unit	Topic
1	Sample spaces, events, the axioms of probability
2	Counting and equally likely outcomes; birthday problem
3	Conditional probability; law of total probability; Bayes' theorem
4	Independence; randomised algorithms: verifying matrix products, min-cut
5	Discrete random variables; Bernoulli, binomial, geometric, Poisson
6	Expectation; linearity of expectation; indicator variables
7	Variance and moments; coupon collector; balls into bins and hashing
8	Continuous random variables; uniform, exponential, normal
9	Functions of random variables; transformations of densities
10	Joint distributions; marginal and conditional distributions
11	Covariance and correlation; the multivariate normal distribution
12	Conditional expectation; law of total expectation; quicksort analysis
13	Moment generating functions
14	Tail bounds: Markov, Chebyshev, Chernoff
15	Laws of large numbers
16	The central limit theorem
17	Simulation and Monte Carlo methods
18	Introduction to Markov chains
19	Descriptive statistics and exploratory data analysis
20	Point estimation: method of moments, maximum likelihood
21	Bias, variance, mean squared error, consistency
22	Confidence intervals; the t-distribution
23	Hypothesis testing: z- and t-tests, p-values, errors of type I and II
24	Chi-squared tests; A/B testing; multiple testing
25	Linear regression: least squares, inference on coefficients
26	The bootstrap; introduction to Bayesian inference

5. Teaching and Learning Methods

The module is self-paced. All materials are provided on Moodle: lecture notes and recorded lectures for each unit, tutorial problem sheets with worked solutions, short quizzes, and the briefs for the two projects. Problem sheets are for practice and are not marked. Questions are discussed in the module channel on Discord and in office hours. Independent study includes reading, problem sheets and project work.

Activity	Hours
Lectures	52
Labs / Tutorials	38
Independent study	150
Total	**240**

6. Assessment

Knowledge is assessed with weekly unit quizzes, two projects and a final exam. Working with simulation and real data is central to the module, so the projects carry slightly more weight.

- **Project 1 – Simulation study:** simulate a randomised algorithm or hashing scheme, compare the observed behaviour with the expectation and tail bounds derived in the module, and illustrate the law of large numbers and the central limit theorem.
- **Project 2 – Data analysis:** pose a question about a public data set, estimate parameters, build confidence intervals, carry out a hypothesis test or regression, and use the bootstrap; written up as a short report.

The final exam tests whether students can build a probabilistic model for a problem, carry out the calculations, and choose and interpret an appropriate statistical method.

Component	Weight	Due	LOs
Unit quizzes (best 10 of 12)	15%	Weekly	LO1–LO7
Project 1: Simulation study	20%	Week 7	LO4, LO5, LO8
Project 2: Data analysis	25%	Week 12	LO6–LO8
Final exam (3 hours, written)	40%	End of term	LO1–LO7
Total	**100%**		

7. Reading List

Essential

- J. K. Blitzstein, J. Hwang: *Introduction to Probability*, 2nd ed., CRC Press, 2019. Free online at `https://probabilitybook.net`.
- L. Wasserman: *All of Statistics: A Concise Course in Statistical Inference*, Springer, 2004.

Recommended

- S. M. Ross: *A First Course in Probability*, 10th ed., Pearson, 2019.
- C. M. Grinstead, J. L. Snell: *Introduction to Probability*, 2nd ed., American Mathematical Society, 1997. Free online.
- M. Mitzenmacher, E. Upfal: *Probability and Computing*, 2nd ed., Cambridge University Press, 2017.
- M. H. DeGroot, M. J. Schervish: *Probability and Statistics*, 4th ed., Pearson, 2012.

8. Materials and Licence

Course materials are published under the Creative Commons Attribution-ShareAlike 4.0 licence unless stated otherwise. See wosm.academy/licence for details.

9. Policies

- **Attendance:** Attendance is not mandatory. The module is self-paced, so you can study the materials when it suits you.
- **Late submissions:** Work submitted after the deadline is penalised.
- **Academic integrity:** All submitted work must be your own. See the Code of Conduct.