

IIT Mandi

Course Number:	IC 252
Course Name:	Data Science 2 (Probability and Statistics)
Credits:	3-0-2-4 (L-T-P-C)
Prerequisites:	IC 110, IC 152.
Intended for:	First-year B.Tech
Distribution:	Institute core for B. Tech
Semester:	Even/Odd

Preamble: Uncertainty is a phenomenon that is observed in several areas – science, statistics, finance - and in daily life. Probability is the logical framework that enables one to quantify uncertainty and randomness. Though the area is quite counter-intuitive, following the rules laid out by the mathematical theory of probability helps one in avoiding mistakes while dealing with uncertainty. The probability theory part of the course provides a mathematical basis and in the statistics, part one learns how to use probability to make sense of data, which is important in a world that is largely data-driven currently.

Objective: After finishing this course the student should be able to formulate probability models and evaluate event probabilities, have a clear understanding of independence, application-level understanding of conditional probability, Bayes theorem, and random variables, do a statistical analysis of data using the WLLN, CLT; formulate and test hypothesis, and find relations between random variables by using regression. On a lighter note after finishing the course successfully the student should be able to accept and deal with uncertainty comfortably with an understanding of the common pitfalls as well as make reasonable predictions from data.

Syllabus:

1. Probability: Why probability and what is it? (give real-life situations which demand the use of probability). Counting, combinations, permutations, binomial and multinomial coefficients, Stirling's formula. Discrete probability spaces (with examples). Axiomatic definition of probability, inclusion-exclusion formula, independence, conditional probability, Bayes' rule. (Note: Cover the paradoxes and well-known problems). (6 lectures)

Lab: counting, basic probability – simulation of simple experiments, birthday paradox, conditional probability.

2. Random variables: definition, distribution function and its properties, probability mass function (binomial, Bernoulli, Poisson, geometric), probability density

function (uniform, exponential, Gaussian). Joint distributions, independence and conditioning of random variables. Function of random variables, change of variable formula. (9 lectures)

Lab: Generating random variables following a given pdf/pmf. engineering application of functions of random variables

3. Measures of central tendency, dispersion and association – expectation, median, variance, standard deviation, mean absolute deviation, covariance, correlation and entropy (definition and guidelines on how to choose a particular measure). Markov and Chebyshev inequalities. Notion of convergence in probability and distribution. Weak law of large numbers and central limit theorem (examples demonstrating the use of WLLN and CLT). Montecarlo methods (estimating value of e , π , simulation of birthday paradox). Poisson limit for rare events. (11 lectures)

Lab: Scatter plot (for independent, correlated, uncorrelated random variables), Montecarlo simulation, WLLN and CLT.

4. Statistics: Using probability to understand data (give real-life examples). Frequentist approach - point and range estimates, confidence intervals, hypothesis testing p-values, significance level, power and t-test. Bayesian inference – maximum likelihood estimation. Regression. (14 lectures)

Lab: Parameter estimation, hypothesis testing, regression.

5. Case study: Analyze a large data set (medicine/engineering/biological) using the methods covered in the course. (2 lectures)

Textbook:

1. Sheldon Ross, Introduction to Probability and Statistics for Engineers, 5/e (2014), Elsevier

Reference books:

1. Morris H. DeGroot and Mark J. Schervish, Probability and Statistics (4/e)(2012), Addison- Wesley.
2. Blitzstein and Hwang, Introduction to Probability (2015), CRC Press.
3. William Feller, An Introduction to Probability, (3/e) (2008), Volume 1, Wiley.
4. Freedman, Pisani, Purves, Statistics (4/e)(2014), W. W. Norton & Company.

Instructors:

Theory: Satyajit Thakor (satyajit@iitmandi.ac.in), Phone: 9816099683

Lead TA:

TBA (TBA), Phone: TBA

Evaluation (Theory 70%, Lab 30 % approx):

- Theory (approx. %): Quiz-1 (20%), Quiz-2 (20%), Endsem exam (50%), Assignments (10%)
- Lab (approx. %): Weekly evaluations (33%), Mid-term exam (27%), End-term exam (40%)

Lecture Hours:

Monday (08:00 – 08:50 AM), Tuesday (09:00 – 09:50 AM), Thursday (01:00 – 01:50 PM)