

MODULE TITLE: 3E3 Probability and Statistics		CODE: EEU33E03
LEVEL: Junior Sophister	CREDITS: 5	PREREQUISITES: SF
LECTURER: Associate Professor Anthony Quinn		
TERM: Michaelmas	LECTURES/WEEK: 3	TUTORIALS/WEEK: 1
DURATION (WEEKS): 11	TOTAL: 33	TOTAL: 11
<p>OBJECTIVES</p> <p>The primary aim of 3E3 is to provide a secure and accessible grounding for all sophister engineering students in probability and statistics. The module equips them with consistent methods for reasoning amid the uncertainties they encounter in their professional practice. In this way, the module supports their decision-making in the uncertain contexts of real engineering practice.</p> <p>The most important context in which engineers cope with uncertainty is when they gather and process information-bearing data, and use them to plan future actions. Therefore, statistical design is the main extended content for the application of probability methods in this module. Being a module for all sophister engineers, examples are drawn liberally from their universal human experience (games, social analyses, etc.), but also from accessible contexts in engineering, notably vehicular traffic phenomena, reliability and lifetime of devices, causal inference in bio-engineering, low-count and high-count sensor outputs in imaging systems, quantization and observation noise, and telecommunications.</p> <p>The keystone of the module is a philosophical one. The profound relationship between uncertainty and information (learning) is explored from the start. An initial review of propositional logic is provided, being a familiar context for the engineer, so that the student can be confident in formulating propositions associated with an uncertain experiment, and in understanding logical relationships between propositions (sufficiency, necessity, equivalence). The probability calculus is then developed as a consistent means of quantifying and manipulating the engineer's beliefs in these uncertain propositions (i.e. the Bayesian perspective), opening a path beyond merely logical relationships. In this way, the foundation of the module is a unitary one, with such notions as logic, uncertainty, information, observation (data) and imprecision (noise) all embraced within a Bayesian notion of probability.</p> <p>The module explores engineering contexts that induce the canonical probability models, in both the discrete case (Bernoulli, geometric, binomial, multinomial, Poisson) and the continuous case (rectangular, exponential, m-Erlang, normal). Their mixtures and transformations are developed as a response to practical modelling needs. There is a special emphasis on the concept of dependence (conditioning), and its relationship to the key notions of correlation, association, causation and prediction. A main learning outcome for the student is the ability to choose the appropriate model to apply in a range of engineering contexts, such as those listed above, guided by an understanding of the assumptions that justify the deployment of each model.</p>		

A feature of the module is that it develops statistics consistently, using the same inductive inference principles described above. Often a blind spot in the formation of the engineering student, statistical inference is re-cast in this module as an application of probability modelling, with statistics derived as the necessary consequence of their chosen model, and the decisions they need to make. In this way, the student gains the skill to *design* data transformations, dispensing with the traditional cook-book of statistical recipes which has little appeal to engineers with design interests. The simple nonparametric case is emphasized, using the elegant device of the empirical distribution. This allows students to derive appropriate descriptive statistics for their data, to estimate probabilities, and to describe association and regression phenomena quantitatively. An accessible introduction to parametric estimation is also provided, via moment matching techniques, and linear regression is developed via estimation of the bivariate Gaussian parameters. Finally, some of the standard statistical hypothesis tests are motivated and applied. Together, these topics introduce the engineer to the culture, nomenclature and practice of statistics as commonly encountered in engineering practice.

The module is an invitation to the student to confront uncertainty as a fundamental phenomenon – and resource – in engineering systems, to appreciate probability as a consistent framework for the design, analysis and optimization of such systems, and to explore and exploit data in a principled way.

SYLLABUS

- **Review of Propositional Logic**
Uncertain experiments in engineering

Sample space, propositions and events

Propositional logic: equivalence, necessity, sufficiency
- **The Foundation of Probability Modelling**
The axioms of probability and the probability triple

Conditional probability; independence

Key relationships: chain rule, Bayes' rule, theorem of total probability

Conditional independence
- **Sequential Experiments**
Independent sequential experiments: geometric, binomial and multinomial probability laws
- **Univariate Random variables**
Probability functions for random variables (cdf, pdf, pmf)

Key discrete probability models (Bernoulli→geometric→binomial→Poisson)

Key continuous probability models (rectangular, exponential, m -Erlang, normal)

Functions of random variables

Expectation and the key moments of random variables

- **Multiple random variables**

Marginal and conditional distributions

Discrete-continuous case: finite mixture models

The bivariate normal distribution

Correlation and linear regression

Introduction to graphical models

- **Statistics and Data Analysis**

Random sampling: the empirical distribution and its moments (nonparametric sampling statistics)

Nonparametric probability estimation

Parameter estimation by moment-matching

Analysis of errors

Design of statistics for description, regression and hypothesis testing

ASSOCIATED LABORATORY

There is a three-hour laboratory on Matlab-assisted exploratory data analysis, emphasizing the key probabilistic and statistical descriptions.

RECOMMENDED TEXTS

The recommended text for the module is:

1. Bertsekas, D.P. and Tsitsiklis, J.N., *Introduction to Probability*, 2nd ed., Athena Scientific Press, 2008.

Secondary recommended texts are as follows:

2. Ross, S.M., *Introduction to Probability and Statistics for Engineers and Scientists*, 5th ed., Academic Press, 2014.
3. Pishro-Nik, H., *Introduction to Probability, Statistics, and Random Processes*, Kappa Research LLC, 2014. (Available online at www.probabilitycourse.com)
4. Applebaum, D., *Probability and Information*, 2nd ed., Cambridge University Press, 2008.

LEARNING OUTCOMES

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

- Quantify and manipulate beliefs in uncertain propositions related to key engineering contexts, such as traffic phenomena, device reliability, bio-imaging and noisy communication
- Distinguish between the vital notions of independence and dependence, and relate the latter to ideas of association, causation and prediction
- Apply and analyze the key parametric probability models (distributions) governing uncertainty in these contexts
- Evaluate measures of location, spread and dependence for these distributions
- Convert information-bearing experimental data into quantified beliefs, summarize these data via sampling statistics, assess dependence between data, and test competing hypotheses for the data

TEACHING STRATEGIES

There is a 3:1 ratio between lectures and tutorials. Problem-solving experience is vital, and gained primarily via the tutorial sessions, but also via regular homework sheets. The laboratory is an invitation to students to explore data in their domains using the powerful and accessible tools of Matlab, and to communicate data-driven conclusions via graphical and numerical probability and statistics summaries. Attendance at all contact sessions (i.e. a total of 47) is compulsory, as this is essential to the student's successful acquisition of the learning outcomes.

ASSESSMENT MODES

70% of the final mark is determined via the annual examination. 20% is reserved for an in-class test during the term. The remaining 10% is determined via assessment of the laboratory report, submitted a number of weeks after the lab session is completed.