Work in Progress - Probability & Statistics in Computer Engineering Curricula

Pradip K Srimani¹ and Murali Varanasi²

Abstract – The purpose of this work in progress report is to discuss the probability and statistics body of knowledge as defined in the Computer Engineering Volume to be published by IEEECS/ACM curricula series [1]; this includes only the topics that are to be included in all the CpE programs as “core” topics.

Index Terms – Computer Engineering Curricula, probability and Statistics, Modeling, Performance Evaluation.

INTRODUCTION

In 1998, the Association for Computing Machinery (ACM) and the Computer Society of the Institute for Electrical and Electronic Engineers (IEEE-CS) convened a joint curriculum task force called Computing Curricula 2001, or CC2001 for short. A committee has been established in the beginning of 2001 to define the body of knowledge that constitutes computer engineering as well as to flesh out course outlines to suit different curricula in computer engineering. Computer Engineering embodies the science and the technology of design, construction, implementation and maintenance of the hardware and the software components of modern computing systems and computer-controlled equipments. Over the past three decades the discipline of computer engineering has emerged from the erstwhile fields of electrical engineering and computer science as a separate, although intimately related, discipline. One of the important aspects is to define the requirement of mathematics especially probability & statistics and discrete mathematics.

Mathematical techniques and formal mathematical reasoning are integral to most areas of computer engineering. The discipline depends on mathematics for many of its fundamental underpinnings. In addition, mathematics provides a language for working with ideas relevant to computer engineering, specific tools for analysis and verification, and a theoretical framework for understanding important ideas. Given the pervasive role of mathematics within computer engineering, the curriculum must include key mathematical concepts early and often. Basic mathematical concepts should appear early within a student's course work and later courses should use these concepts regularly. While different colleges and universities will need to adjust their prerequisite structure to reflect local needs and opportunities, it is important for upper-level computer engineering courses to make use of the mathematical content developed in earlier courses. A formal prerequisite structure should reflect this dependency.

Probability and Statistics is one of the essential components of the mathematics requirement as outlines in the CPE document [2]. All students should get at least some brief exposure to discrete and continuous probability, stochastic processes, sampling distributions, estimation, hypothesis testing, and correlation and regression.

CE-PRS KNOWLEDGE UNITS

The intertwined topics of probability and statistics provide important insights into a range of topics of fundamental importance to the Computer Engineer. For example all issues of reliability and dependability rely on an understanding of these topics. But additionally they play fundamental roles in testing and evaluation (of hardware, software and communications systems) where levels of performance have to be guaranteed. Further uses of the topics can be found in a wide variety of areas: searching, data structure design and implementation (hash tables), computer architecture (cache concerns), operating systems (working set models), human computer interaction (experimentation), security and in aspects of intelligent systems. We reproduce here the topics and the learning objectives from the CCCE report [2].

CE-PRS1. Discrete probability [core] [8 hrs]

Topics:
1. Randomness, finite probability space, probability measure, events
2. Conditional probability, independence, Bayes’ theorem
3. Discrete random variables
4. Binomial, Poisson, Geometric distributions
5. Mean and Variance: concepts, significance, computations, applications
6. Integer random variables

Learning outcomes:
- Calculate probabilities of events and expectations of random variables for elementary problems such as games of chance
- Differentiate between dependent and independent events
- Apply binomial theorem to independent events and Bayes’ theorem to dependent events

¹ Department of Computer Science, Clemson University, Clemson, SC 29634. ² University of North Texas, Denton, Texas.

© 2004 IEEE 0-7803-8552-7/04/$20.00
• Apply the tools of probability to solve problems such as in the Monte Carlo method, the average case analysis of algorithms and hashing

CE-PRS2. Continuous probability [core] [8 hrs]

Topics:
1. Continuous random variables, the nature of these, illustrations of use
2. Exponential and normal distribution: probability density functions, calculation of mean and variance
3. The central limit theorem and the implications for the normal distribution.

Learning outcomes:
• Recognize situations in which it is appropriate to consider the relevance of the normal distribution and/or the exponential distribution
• Calculate the mean and the variance for given distributions involving continuous random variables.

CE-PRS3. Expectation [core] [4 hrs]

Topics:
1. Moments, Transform methods, Mean time to failure
2. Conditional expectation, examples
3. Imperfect fault coverage and reliability

Learning outcomes:
• To understand the significance and be able to compute expectation of functions of more than one variable and transform.
• To be able to compute fault coverage and reliability in simple hardware and software applications

CE-PRS4. Stochastic Processes [core] [8 hrs]

Topics:
1. Introduction: Bernoulli & Poisson processes, renewal process, renewal model of program behavior
2. Discrete parameter Markov chains: transition probabilities, limiting distributions
3. Queuing: M/M1 and M/G/1, Birth and Death process
4. Finite Markov chains, program execution times.

Learning outcomes:
• To be familiar with the concepts and tools to manipulate stochastic processes
• To apply the concepts and tools of stochastic processes to analyze the performance of simple hardware and software systems

CE-PRS5. Sampling distributions [core] [4 hrs]

Topics:
1. Purpose and the nature of sampling, its uses and applications
2. Random approaches to sampling: basic method, stratified sampling and variants thereof, cluster sampling
3. Non-random approaches: purposive methods, sequential sampling
4. Data analysis; tools; graphical and numerical summaries
5. Multivariate distributions, independent random variables

Learning outcomes:
• Recognize situations in which the different approaches to sampling are relevant
• Demonstrate the ability to apply appropriate sampling methods in a range of situations.

CE-PRS6. Estimation [core] [4 hrs]

Topics:
1. Nature of estimates: point estimates, interval estimates
2. Criteria to be applied to single point estimators: unbiased estimators, consistent estimators, efficiency and sufficiency of estimators
3. Maximum likelihood principle approach, least squares approach; applicability conditions for these
4. Confidence intervals
5. Estimates for one or two samples

Learning outcomes:
• Describe the fundamental principles behind the concept of estimation and give examples that illustrate its beneficial application
• Given a distribution apply basic principles to derive estimators which exhibit desirable properties

CE-PRS7. Hypothesis tests [core] [2 hrs]

Topics:
1. Development of models and associated hypotheses, the nature of these
3. Testing hypothesis based on a single parameter, choice of test statistic; choice of samples and distributions.
4. Criteria for acceptance of hypothesis
5. t-test, chi-squared test; applicability criteria for these

Learning outcomes:
• Explain the role of hypothesis testing, describing the main steps in the process
• Given a sample situation, formulate a hypothesis and carry out appropriate tests to check its acceptability.
CE-PRS8. Correlation and regression [core] [2hrs]

Topics:
1. The nature of correlation and regression, definitions.
2. Approaches to correlation: the linear model approach, the least squares fitting approach, strengths and weaknesses of these and conditions for applicability.

Learning outcomes:
- Recognize circumstances under which it is appropriate to investigate relationships between variables.
- Given a suitable circumstance, apply correlation and regression techniques with a view to establishing relationships between variables.