Sampling in Combinatorial and Geometric Set Systems

By Nabil H. Mustafa

An understanding of basic sampling techniques and intrinsic geometric attributes of data are indispensable for students and researchers in machine learning and theoretical computer science. This book gives a survey of geometric and combinatorial techniques for studying questions surrounding problems in random sampling that deal with how to ensure that you have a large enough sampling of data to give meaningful information about a given collection of sets.

Technically, this is the study of epsilon-nets. One starts with a set-system, which is a large set of objects and a fixed collection of distinguished subsets. One would like to find the minimum proportion, epsilon, of objects from the set necessary for a random sampling of that proportional size to intersect each of the distinguished sets at least once. A commonly studied geometric version of this problem is to consider a collection of points $P$ in the plane, and the set-system defined by intersections of $P$ with half spaces defined by lines in the plane. During the last ten years there has been a lot of activity in this subject, leading to the resolution of several long-outstanding open problems.

In this book, Mustafa develops a unified theory of random sampling that leads to new and elegant proofs of classical results in the geometric setting, simpler and more efficient algorithms, and generalizations to broader combinatorial problems. An important feature of the theory is that it provides bounds on the complexity of the set-systems and on the probability that a random variable deviates from expected behavior.

The book, written in a self-contained way with ample references to related work, will be useful to graduate students and researchers interested in combinatorics, computational geometry, statistics, and machine learning. Prerequisites are a basic familiarity with discrete mathematics, probability, and combinatorics.