In statistical theory, a U-statistic is a class of statistics that is especially important in estimation theory. The letter “U” stands for unbiased. In elementary statistics, U-statistics arise naturally in producing minimum-variance unbiased estimators.

The theory of U-statistics allows a minimum-variance unbiased estimator to be derived from each unbiased estimator of an estimable parameter (alternatively, statistical functional) for large classes of probability distributions. An estimable parameter is a measurable function of the population’s cumulative probability distribution: For example, for every probability distribution, the population median is an estimable parameter. The theory of U-statistics applies to general classes of probability distributions.

Many statistics originally derived for particular parametric families have been recognized as U-statistics for general distributions. In non-parametric statistics, the theory of U-statistics is used to establish for statistical procedures (such as estimators and tests) and estimators relating to the asymptotic normality and to the variance (in finite samples) of such quantities. The theory has been used to study more general statistics as well as stochastic processes, such as random graphs.

Suppose that a problem involves independent and identically-distributed random variables and that estimation of a certain parameter is required. Suppose that a simple unbiased estimate can be constructed based on only a few observations: this defines the basic estimator based on a given number of observations. For example, a single observation is itself an unbiased estimate of the mean and a pair of observations can be used to derive an unbiased estimate of the variance. The U-statistic based on this estimator is defined as the average (across all combinatorial selections of the given size from the full set of observations) of the basic estimator applied to the sub-samples.

This theory was initiated in a paper by Wassily Hoeffding in 1948 (Hoeffding, W. A class of statistics with asymptotically normal distributions. Annals of Statistics, 19 (1948), 293-325). P.K. Sen provides a review of his paper, which introduced U-statistics and set out the theory relating to them, and in doing so Sen outlines the importance U-statistics have in statistical theory.

We will discuss the basic definitions, notions, inequalities (moment and maximal inequalities). The main objectives are to deal with laws of large numbers, weak convergence, functional limit theorems, approximations in limit theorems, asymptotic expansions, large deviations, and laws of the iterated logarithm. U-statistics constructed from a dependent sequence of random variables
are also discussed.

The man reference we shall use is

  Applications 273 (Translated by P. V. Malyshev and D. V. Malyshev from the 1989 Russian
  2608-3.

Other reference books are


  2.

  8253-4.

- de la Peña, V. H. and Giné, E. Decoupling. From dependence to independence. Randomly