Proceedings of the 1983
Winter Simulation Conference
S. Roberts, J. Banks, B. Schmeiser (eds.)

457

UNIFIT: AN INTERACTIVE COMPUTER PACKAGE FOR FITTING PROBABILITY DISTRIBUTIONS TO OBSERVED DATA

Averill M. Law and Stephen G. Vincent
Simulation Modeling and Analysis Company
P.O. Box 40996
Tucson, Arizona 85717

An important problem which occurs in many different disciplines is that of determining a probability distribution which is a good representation of an observed data set. For example, in building a simulation model of a manufacturing process or of a computer system, one needs to determine appropriate probability distributions for the input random variables. A common solution to this problem is to fit standard distributions (e.g., normal or gamma) to observed system data. However, since this fitting process is rather complicated and time consuming when done by hand, it is often performed in a superficial and incorrect manner. The net effect is, of course, that the selected distributions may not be good representations of the observed data.

UNIFIT is a state-of-the-art, interactive computer package for fitting probability distributions to observed data. By combining the latest statistical techniques with graphical displays, the package allows one to perform a comprehensive analysis of a data set in significantly less time than would otherwise be possible. It employs a three-activity approach for determining an appropriate distribution. The first activity involves using heuristic techniques such as histograms or sample moments to hypothesize one or more families of distributions which might be representative of the observed data. For example, if our data are continuous and if a histogram of the data indicates that the density function of the underlying distribution is skewed to the right, then we might hypothesize that a gamma, lognormal, or Weibull distribution is an appropriate model for our observed data. However, each of these families of distributions has several parameters which must be specified in order to have a completely determined distribution. Therefore, the second activity typically involves estimating the parameters of each hypothesized family from the data, thereby specifying a number of particular distributions. In the third activity we determine which of the fitted distributions, if any, is the best representation for the data using both heuristic techniques and goodness-of-fit tests. An example of a heuristic technique provided by UNIFIT is the frequency comparison, which is a graphical display showing both the observed proportion of observations and the expected proportion of observations from a particular fitted distribution for each histogram interval. The frequency comparison is particularly useful for visually determining how well a selected probability model represents the underlying distribution for the data. In addition to heuristic techniques, UNIFIT makes available to an analyst the chi-square, the Kolmogorov-Smirnov, and the Anderson-Darling goodness-of-fit tests. These tests can be considered to be a formal approach for detecting gross discrepancies between the fitted distribution and the observed data.

UNIFIT is written in FORTRAN IV and is designed to be compatible with standard terminals on almost any moderate-sized computer. It is the culmination of more than five years of statistical research to determine the most definitive techniques for distribution fitting, and it took three years to program. Although UNIFIT is so easy to use that it requires almost no documentation, it comes with a comprehensive user's guide which carefully documents every aspect of its operation. The guide also includes what is perhaps the most complete and up-to-date treatment of distribution fitting that is available anywhere.

An analyst using UNIFIT to model observed data will realize the following benefits:

1. A considerable decrease in the time and cost required for distribution fitting--most data sets can be analyzed in less than one hour

2. An order of magnitude increase in the comprehensiveness of the analysis which can be performed

3. A reduced likelihood of making errors in your analysis--many books and software packages contain erroneous results

REFERENCE

Law AM, Vincent SG (1983), <u>UNIFIT User's Guide</u>,
    Simulation Modeling and Analysis Company,
    Tucson, Arizona.