DISTRIBUTION FITTING IN STATISTICA APPLICATION

Rodica SOBOLU, Ioana POP, Maria MICULA

University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca,
3-5 Manastur Street, Romania, rodicasobolu@yahoo.com

Keywords: distribution fitting, statistical distribution

SUMMARY

Distribution fitting is the procedure of selecting the statistical distribution which best fits to a data set generated by some random process. This allows us to develop valid models of random processes we deal with, protecting us from potential time and money loss which can arise due to invalid model selection, and enabling us to make better decisions. In some research applications one can formulate hypotheses about the specific distribution of the variable of interest. For example, variables whose values are determined by an infinite number of independent random events will be distributed following the normal distribution, whereas variables whose values are the result of an extremely rare event would follow the Poisson distribution.

For predictive purposes it is often desirable to understand the shape of the underlying distribution of the population. To determine this underlying distribution, it is common to fit the observed distribution to a theoretical distribution by comparing the frequencies observed in the data with the expected frequencies of the theoretical distribution (Normal Distribution, Bernoulli Distribution, Beta Distribution, Binomial Distribution, Cauchy Distribution, Chi-square Distribution, Exponential Distribution, Gamma Distribution, Poisson Distribution).

Once the distribution model has been chosen, it is necessary to determine how well the selected distribution fits the data. This can be done using the specific goodness of fit tests by comparing the empirical (based on sample data) and theoretical (fitted) distribution graphs. As a result, we will select the most valid model describing our data. Distribution fitting software helps us automate the data analysis and decision making process, and enables to focus on our core business goals rather than technical issues. STATISTICA application includes the Distribution Fitting tool which helps us to select the best fitting distribution and apply it to make better decisions in different fields of interest. The fit can be evaluated via the Chi-square test or the Kolmogorov-Smirnov one-sample test. STATISTICA has also Process Analysis option where one can compute maximum-likelihood parameters’ estimates for the Beta, Exponential, Gamma, Log-Normal distributions. The distribution graphs can be also very helpful in determining the best fitting model, providing an empirical way to analyze the data. It can be outlined selecting Graph in Options tab of Fitting Continuous Distributions dialog.

REFERENCES

1. www.statsoft.com