



Contribution ID: 333

Type: oral presentation

A statistical toolkit for data analysis

Thursday, September 30, 2004 5:30 PM (20 minutes)

Statistical methods play a significant role throughout the life-cycle of HEP experiments, being an essential component of physics analysis. We present a project in progress for the development of an object-oriented software toolkit for statistical data analysis. More in particular, the Statistical Comparison component of the toolkit provides algorithms for the comparison of data distributions in a variety of use cases typical of HEP experiments, as regression testing (in various phases of the software life-cycle), validation of simulation through comparison to experimental data, comparison of expected versus reconstructed distributions, comparison of data from different sources - such as different sets of experimental data, or experimental with respect to theoretical distributions. The toolkit contains a variety of goodness-of-fit tests, from chi-squared to Kolmogorov-Smirnov, to less known, but generally much more powerful tests such as Anderson-Darling, Cramer-von Mises, Kuiper, Tiku etc.

Thanks to the component-based design and the usage of the standard AIDA interfaces, this tool can be used by other data analysis systems or integrated in experimental software frameworks. We present the architecture of the system, the statistics methods implemented and some results of its applications to the comparison of Geant4 simulations with respect to experiment.

Primary authors: PFEIFFER, A. (CERN); RIBON, A. (CERN); MASCIALINO, B. (INFN Genova); PIA, M.G. (INFN GENOVA); VIARENGO, P. (IST, Genova, Italy); DONADIO, S. (INFN Genova); GUATELLI, S. (INFN Genova, Italy)

Presenter: PIA, M.G. (INFN GENOVA)

Session Classification: Event Processing

Track Classification: Track 2 - Event processing