



Contribution ID: 258

Type: **Poster**

A new development cycle of the Statistical Toolkit

Thursday 24 May 2012 13:30 (4h 45m)

The Statistical Toolkit is an open source system specialized in the statistical comparison of distributions. It addresses requirements common to different experimental domains, such as simulation validation (e.g. comparison of experimental and simulated distributions), regression testing in software development and detector performance monitoring.

The first development cycles concerned the provision of a wide set of non-parametric goodness-of-fit tests for the so-called two sample problem, i.e. the comparison of two distributions. The active use of the Statistical Toolkit in real-life applications, documented in the literature, has highlighted new requirements, that are addressed by a new development cycle. The new product includes extensions of the functionality of the toolkit, refinements of existing algorithms and tools and improved usability of the system.

Various sets of statistical tests have been added to the existing collection to deal with the one sample problem (i.e. the comparison of a data distribution to a function, including tests for normality), the comparison of two-dimensional distributions, categorical analysis and the estimate of randomness. Improved algorithms and software design contribute to the robustness of the results. A simple user layer dealing with primitive data types and an improved ROOT user layer facilitate the use of the toolkit both in standalone analyses and in large scale experiments. Interface to the R package extends the native functionality of the toolkit.

An overview of the new developments is presented, along with applications to concrete experimental scenarios.

Authors: Dr RIBON, Alberto (CERN); Dr PFEIFFER, Andreas (CERN); Dr PIA, Maria Grazia (Universita e INFN (IT)); Mr BATIC, Matej (Jozef Stefan Institute)

Presenter: Mr BATIC, Matej (Jozef Stefan Institute)

Session Classification: Poster Session

Track Classification: Software Engineering, Data Stores and Databases (track 5)