



Contribution ID: 25

Type: oral presentation

HistFitter: a flexible framework for statistical data analysis

Thursday, 16 April 2015 11:15 (15 minutes)

We present a software framework for statistical data analysis, called *HistFitter*, that has been used extensively in the ATLAS Collaboration to analyze data of proton-proton collisions produced by the Large Hadron Collider at CERN. Most notably, HistFitter has become a de-facto standard in searches for supersymmetric particles since 2012, with some usage for Exotic and Higgs boson physics. HistFitter coherently combines several statistics tools in a programmable and flexible framework that is capable of bookkeeping hundreds of data models under study using thousands of generated input histograms. The key innovations of HistFitter are to weave the concepts of control, validation and signal regions into its very fabric, and to treat them with rigorous methods, while providing multiple tools to visualize and interpret the results through a simple configuration interface, as will become clear throughout this presentation.

HistFitter interfaces with the statistics tools

HistFactory and RooStats to construct parametric models and to perform statistical tests of the data, and extends these tools in four key areas:

1. Programmable framework: HistFitter puts tools from several sources together in a coherent and programmable framework, capable of performing a complete statistical analysis of pre-formatted input data samples.
2. Bookkeeping: HistFitter can perform statistical tests and scan over parameter values of hundreds of signal hypotheses in an organized way from a single user-defined configuration file.
3. Analysis strategy: HistFitter uses built-in the concepts of control, signal and validation regions which are used to constrain, extrapolate and validate data-model predictions across analysis regions. HistFitter also introduces a rigorous treatment of validation regions that is new in high-energy physics.
4. Presentation and interpretation: The HistFitter framework keeps track of data models before and after fits to the data, and includes a collection of methods to determine the statistical significance of all tested hypotheses and to produce tables and plots expressing these results with publication-quality style.

Primary authors: KOUTSMAN, Aleksej (TRIUMF (CA)); COTE, David (University of Texas at Arlington (US)); BESJES, Geert Jan (Radboud University Nijmegen (NL)); LORENZ, Jeanette Miriam (Ludwig-Maximilians-Univ. Muenchen (DE)); BAAK, Max (CERN)

Presenter: BESJES, Geert Jan (Radboud University Nijmegen (NL))

Session Classification: Track 2 Session

Track Classification: Track2: Offline software