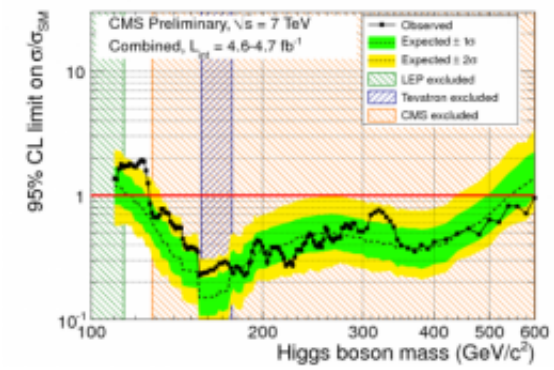
The importance of statistics in high-energy physics

January 20 2012, by Achintya Rao, CERN



The search for the Higgs boson receives regular input from the Statistics Committee, especially when it comes to defining the significances of excesses.

“If you remember, Mark Twain once said, ‘There are three kinds of lies: lies, damned lies, and statistics.’ If you use statistics in an improper way, you could get pretty much any result,” says Greg Landsberg, CMS’s new physics coordinator. “The role of the CMS Statistics Committee is to interpret what we see in a statistically correct way.”

“All of our physics results depend on statistical procedures that we perform on the data we obtain,” adds Tommaso Dorigo, the new chairperson of the CMS Statistics Committee. Terms like “statistical significance” play a key role in categorising observed signals as hints, evidence or discovery. “Upon performing a search for a new particle or process,” Dorigo continues, “you rarely observe such a huge new signal

that you do not need to quantify its size or whether it may be due to a fluctuation. You have to have a procedure with which you can give a mathematical interpretation of your data in terms of how significant your result is."

One of the tasks of the Statistics Committee is to produce guidelines on the statistical methods to be used by the many analyses within the collaboration. The team makes recommendation to the analysis teams to ensure that CMS papers do not contain statistical claims that are unsupported by the data. "It is an attempt to make our papers coherent in the way they present the results," says Dorigo.

"By now in particle physics, we have very well-defined statistical methods. And it's very important to use them uniformly across the collaboration and also to communicate with ATLAS, ALICE and LHCb with the view of combining results," adds Landsberg.

In the last year, CMS has moved into the combination regime, having combined the $B_s \to \mu\mu$ result with LHCb and the Higgs limits with half of 2011's statistics with ATLAS. Since all of the LHC experiments are designed differently and the individual results are obtained by different methods, how to go about combining the results is not always obvious. The CMS Statistics Committee has to join forces with the other experiments in order to recommend statistical methods that are acceptable to all.

Statistics are involved not only in the mathematical framework for comparing experiment with theory but also in building the right computer tools to select interesting data from collisions. This parallel branch of statistics activities is now represented at the committee: Gennadiy Kukartsev, who is co-leading the Physics Analysis Tools group at CMS, is an ex-officio member. "In this new incarnation," says Landsberg, "we hope that the Statistics Committee will not only make

recommendations on a theoretical basis but also provide more practical tools, pieces of code and examples."

Under a new chairperson, the committee has been enlarged from nine members to eleven. The reason: CMS has published over 100 physics papers so far, and the rate of papers will only increase with time. "The current size of the committee is just not sufficient to follow all the papers," points out Landsberg.

The committee also provides a helpline to answer statistics questions almost in real-time, and is looking to expand this function. As more people get involved in the analyses, the greater is the demand for help in tackling statistics problems.

The interaction between the committee and the physics groups has also changed. Plans are to establish a sustainable two-way communication between the Statistics Committee and the physics conveners through regular talks and surveys. The committee's weekly meetings are also now open to every collaborator, an important step in ensuring that the collaboration is well prepared for the statistical challenges of the coming years.

This article was originally published on the CMS website. More such articles can be found at http://cern.ch/cms.

Source: CERN