TRISEP 2014, Statistics and Higgs Analysis
Thomas R. Junk

Homework Assignment #1

Thursday, June 5, 2014

1) Show that the maximum-likelihood combination of several independent measurements with identical but uncorrelated Gaussian uncertainties is the average of the measurements.

2) If the uncertainties are different for each measurement, show that the maximum-likelihood combination is

$$x_{\mathrm{comb}} = \frac{\sum x_i/\sigma_i^2}{\sum 1/\sigma_i^2}$$

where the i^{th} measurement is $x_i \pm \sigma_i$. What is the total uncertainty on the combined x_{comb}?

3) For what range of μ is the median n_{med} of the Poisson probability distribution

$$P(n|\mu) = \frac{\mu^n e^{-\mu}}{n!}$$

equal to zero? For what range of μ is $n_{\mathrm{med}} = 2$? Using a Gaussian approximation, for what range of μ is $n_{\mathrm{med}} = 10000$?

4) A common exercise in HEP is to make a table of predicted values of yields itemized by process and compare the total with an observed number of events. The predictions are shown with their total uncertainties, and the total is given with its uncertainty as well. Format such a table when the uncertainties are broken down by their sources. Treat each source of uncertainty as independent from all other sources, but treat the same source of uncertainty's impact on several yield predictions as 100% correlated. The theoretical predictions are considered to have independent uncertainties, and separate Monte Carlo samples are used to predict each process's yield (where applicable). The numbers are for a fictional analysis with fictional theoretical predictions.

Diboson production – estimated using a Monte Carlo normalized to a theoretical prediction: 15.1 events is the central prediction, and it has a ±6% uncertainty from the luminosity, a ±6% uncertainty from the the theory. It has a ±3% uncertainty from jet energy scale, ±2% from ISR/FSR, and ±1% from PDF. It has a ±2% uncertainty from MC statistics. The trigger efficiency is ±1% and the lepton ID efficiency is ±1%. It has a ±5% uncertainty from the b-tag efficiency.

W+jets production – estimated using a Monte Carlo normalized to a control sample in the data: 338 events is the central prediction, and it has a ±30% uncertainty in the extrapolation from the control sample.

Z+jets production – estimated using a Monte Carlo normalized to a control sample in the data: 59 events is the central prediction, and it has a $\pm 30\%$ uncertainty in the extrapolation from the control sample, which is separate from the W+jets control sample.

$t\bar{t}$ production – estimated using a Monte Carlo normalized to a theoretical prediction: 101.2 events is the central prediction, and it has a $\pm 6\%$ uncertainty from the luminosity, a $\pm 10\%$ uncertainty from the the theory. It has a $\pm 5\%$ uncertainty from jet energy scale, $\pm 5\%$ from ISR/FSR, and $\pm 3\%$ from PDF. It has a $\pm 2\%$ uncertainty from MC statistics. The trigger efficiency is $\pm 1\%$ and the lepton ID efficiency is $\pm 1\%$. It has a $\pm 5\%$ uncertainty from the b-tag efficiency.

Single Top production – estimated using a Monte Carlo normalized to a theoretical prediction: 25.6 events is the central prediction, and it has a $\pm 6\%$ uncertainty from the luminosity, a $\pm 10\%$ uncertainty from the the theory. It has a $\pm 3\%$ uncertainty from jet energy scale, $\pm 5\%$ from ISR/FSR, and $\pm 5\%$ from PDF. It has a $\pm 2\%$ uncertainty from MC statistics. The trigger efficiency is $\pm 1\%$ and the lepton ID efficiency is $\pm 1\%$. It has a $\pm 5\%$ uncertainty from the b-tag efficiency.

QCD Multijet production – estimated from a data control sample. 50.1 events with a fractional uncertainty of 40%.

The observed number of data counts is 602. Make the yield table showing the total uncertainty on each prediction (in events, not fractional uncertainties as given above), and sum up the predictions, computing the total uncertainty on the sum.

5) Two measurements $m_1 = 10.1$ and $m_2 = 10.2$ of the same quantity have systematic uncertainties that are correlated $c_1 = 1.5$ and $c_2 = 1.6$, respectively (they come from the same source), and uncorrelated, $u_1 = 1.1$ and $u_2 = 1.0$.

a) What is the correlation coefficient ρ from the covariance matrix between the two measurements?

b) Combine the two measurements using BLUE, obtaining a central value and the total uncertainty. Find the weights w_1 and w_2.

c) Suppose now that $u_1 = 0$ and $u_2 = 0$. Combine again. What are the weights? What is the total uncertainty on the combined measurement?