TRISEP 2014, Statistics and Higgs Analysis
Thomas R. Junk

Homework Assignment #2

Friday, June 6, 2014

1) An experiment searching for a new particle conducts a counting analysis and observes zero events. Show analytically for a Bayesian upper limit calculation that the observed upper limit on the number of signal events does not depend on the predicted background rate, assuming a uniform prior on the signal rate.

2) Show that it is also true for $\mathrm{CL_s}$ limits that the observed limit does not depend on the background prediction if zero events are observed.

3) Computer exercise: Make a plot of the median expected Bayesian upper limit assuming no signal is truly present in a counting search as a function of the predicted background rate b where $0 < b < 10$ counts. Assume no systematic uncertainties on the signal acceptance or the background prediction, and use a uniform prior on the signal rate.

4) Computer exercise – estimating the impact of the look-elsewhere effect. Let's assume we have a histogram with 20 bins numbered from 1 to 20. The background prediction is 10 events in each bin, with no systematic uncertainties. Let's use this histogram to search for a new physics signal which has an unknown mass.

a) Let's assume the new physics signal is very narrow and is expected to produce events in just one bin i, and there is no prediction for the absolute rate, and our test signals are indexed by the bin in which they produce the extra events. We run the experiment and pick the bin with the highest number of events (the lowest local p-value). Compute the global p-value as a function of the local p-value, that is, the probability of getting a smallest p-value anywhere in the histogram of p_0 or less. Show your results as a graph of the global p-value versus the smallest local p-value. Hint: The Dunn-Sidák calculation is the easiest way to do this one.

b) Repeat the exercise if signal k produces equal numbers of events in bins $k-1$, k, and $k+1$, and zero for other bins, for $2 \leq k \leq 19$, and is zero for other bins. Hint: since the s/b ratio is constant for the bins with signal, and is zero for the other bins, you may collect the bins with signal together when computing the test statistic, which then is isomorphic to just counting events in those three bins, for any choice of test statistic. Run background-only pseudoexperiments and make a distribution of the largest number of events in any three consecutive bins to compute the global p-value. To compute the local p-value, use the distribution of events in a predetermined set of three consecutive bins.

5) For the case in problem 4a, what is the median expected limit on the signal rate assuming a specific bin i that the signal may be in? For 4b, what is it? What are the median expected p-values assuming a signal yield of 5 events? For 4a, assume that the 5 events of expected signal is concentrated in one bin, and for the case of 4b, assume that it is divided evenly among the three bins.