The outline of the programme for our Banff Workshop is as follows:

- Monday all day:
    - Introductory talks by Particle Physicist (Louis Lyons), Astro/Cosmo person (Tom Loredo), and Statistician (Richard Lockhart tentatively)
    - 3 minutes for each participant to say who they are.

- Tuesday am: Banff Challenge 2 (Or Monday afternoon)
- Tuesday rest of day: Talks and discussion about discovery issues (see topics below)

- Wednesday am: Talks and discussion about discovery issues (see topics below)
- Wednesday pm: Free for excursion, weather permitting

- Thursday most of day: Talks and discussion about discovery issues (see topics below)
- Thursay afternoon: Future of Banff Challenge 2

- Friday am: Summing up (or possibly Thursday evening

Below is a suggestion for topics related to discovery that we can discuss in detail on Tuesday, Wednesday and Thursday. We would welcome some talks on these topics. However the idea is not to have everyone feeling obliged to give one of these talks, but rather to have a few talks in order to simulate meaningful discussion, for which we are aiming to leave plenty of time. Please let us know of any relevant talk you would be prepared to give, and send us an Abstract that would help us put together a programme.

Looking forward to hearing from you,

Richard, Jim and Louis
(lockhart@stat.sfu.ca;linnemann@pa.msu.edu;l.lyons@physics.ox.ac.uk)

BANFF TOPICS (Explanations of some of the following appear below the list)

- Choice of statistic and selections
- Limits/exclusion
    - ‘No decision'
    - Confidence Limits
    - Exclusion in range of masses
- Sensitivity of search
    - Limits and Discovery separately, or joint (Punzi)
- Significance in on-off problem:

- for distributions (New "Banff Challenge"?)
- What to do when Wilks' Theorem not valid
- Practicalities of MC estimates of significance
- 5 sigma convention

- Look elsewhere effect (False discovery rate)
  - Where is 'elsewhere'? Mass range?
- Comparing H0 and H1
  - Bayesian methods; problem of priors
  - p-values versus likelihood ratios
  - Which likelihood ratio: L(s=expected)/L(s=0) or L(s=best)/L(s=0) ?
- Dealing with nuisance parameters
  - Parton distribution functions
- If H1 has location and rate parameters, is it sensible to use 1-D raster scan for exclusion, but 2-D for parameter determination? And does it make a difference for significance?
- Combination issues e.g. combining p-values

SOME EXPLANATIONS:

Limits/exclusions

- Exclusion: We are comparing our data with the null hypothesis H_0, or with a new physics alternative H_1. Rejection of H_0 is part of the way to discovery of new physics (but we also need to confirm that H_1 provides a better explanation of the data). Non-rejection of H_0 can be divided into 2 categories: (a) Not enough separation power between the pdf's for the H_0 and H_1, and/or not enough data. Then neither hypothesis is excluded. (b) Data inconsistent with H_1, in which case H_1 is excluded. Exclusion is not as exciting as discovery, but is nevertheless regarded as progress.

- CL_s: The standard method of excluding H_1 would be to have a small p-value (e.g. below 5%) for the data. Because this gives a non-negligible probability of excluding H_1 even when we have no or little sensitivity to discriminating between H_0 and H_1, CL_s is defined as p_1/(1-p_0), where the p-values are defined in terms of the pdf tails pointing towards each other. [For statisticians: think of H_0 and H_1 as being separate hypotheses like θ=θ_0 and θ=θ_1; each can be treated as the null hypothesis with a corresponding p-value.] Because the denominator is not greater than unity, CL_s is regarded as a conservative version of p_1. Although it provides protection against exclusion claims when we have no sensitivity to the new physics, the use of CL_s is not universal in Particle Physics.{Don't ask why the ratio of p-values is called CL_s}*

Comparing H0 and H1

- p-values versus likelihood ratios: To compare H_0 and H_1, what are the relative merits of using p_0 and p_1, or the likelihood ratio L_0/L_1?
- Which likelihood ratio: L(s=expected)/L(s=0) or L(s=best)/L(s=0)?: In situations where H_1 has a specified signal rate s, and we are using a likelihood ratio to compare hypotheses, is it better to compare L(s=0) with the likelihood for the expected rate, or with the value of s, regarded as free, which maximises the likelihood for H_1? {This might be a physics question, rather than a statistics one. The use of s=best allows us to discover a signal produced at an unexpected rate.)*

Dealing with nuisance parameters

- Parton distribution functions (PDF's): In predicting expected rates of production of old or new particles in collisions involving protons, it is necessary to know the behaviour of quarks and gluons (collectively know as partons) in the proton. These are specified by 'parton distribution functions', which are obtained by fitting a large amount of data (e.g. over 2000 experimental numbers), either parametrically involving about 20 parameters, or non-parametrically. The uncertainties in these PDF's contribute a systematic uncertainty to predicted particle production rates, and could result in discovery claims being made unjustifiably, or true discoveries being missed. There are interesting statistical questions in the way the PDF uncertainties are estimated.

- Raster scan: The new physics hypothesis H_1 will often contain a location parameter for the mass M of the new particle, and also its production rate s. In exclusion analyses, typically M is held fixed in order to see whether the data excludes s=0 at this M, and this is repeated for all relevant M; this is known as a raster scan. In contrast, for determination of the best fit values of M and s, the two parameters are varied simultaneously.