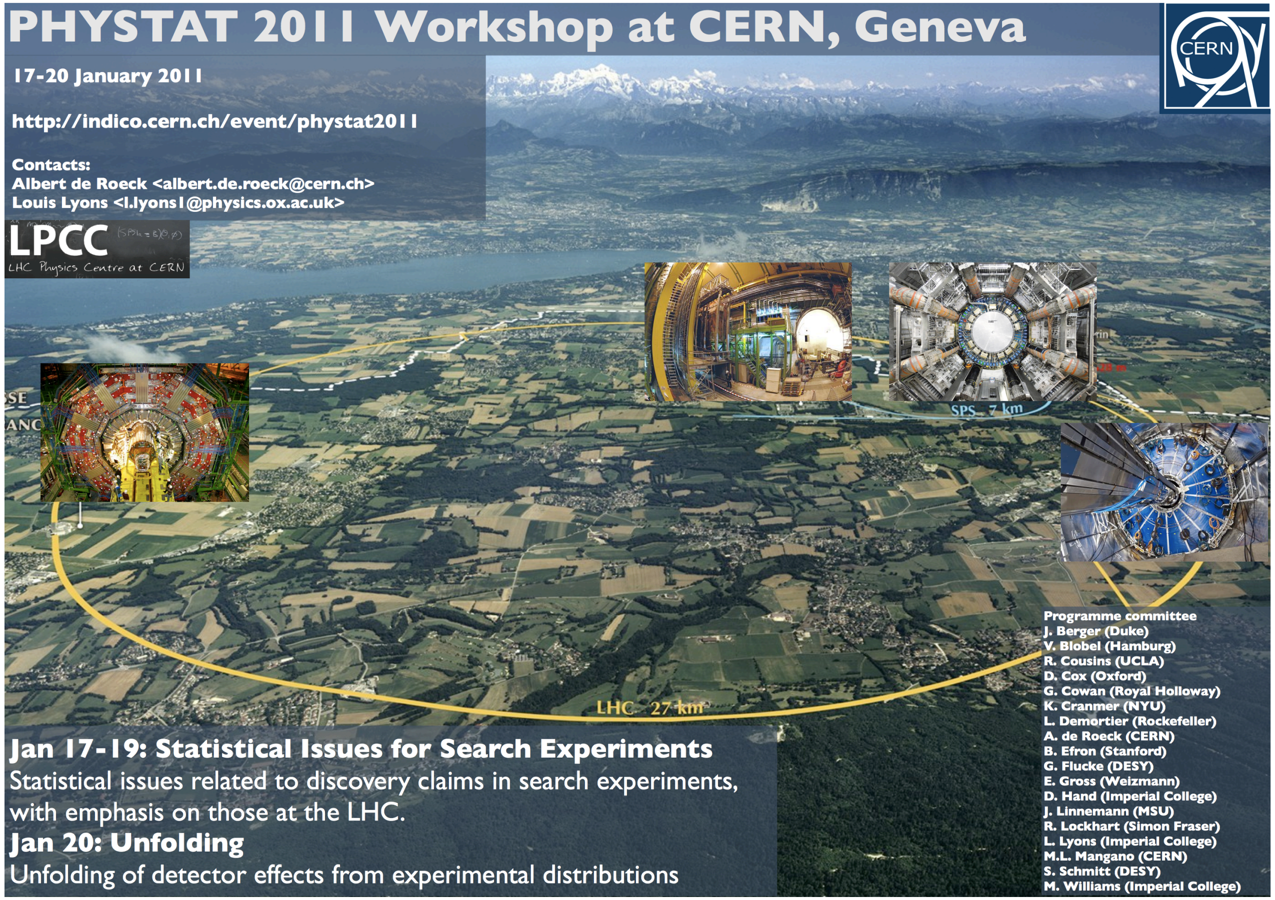


PHYSTAT 2011 Workshop Summary

CERN, 17-19 January, 2011

Glen Cowan
Physics Department
Royal Holloway, University of London

Outline

Frequentist methods

Bayesian methods

In practice: Tevatron, CMS, ATLAS, LHCb

Combining results

Look-elsewhere effect

Software tools

Applications: partons, gravity, astro

Banff Challenge 2a

Outlook

Frequentist methods

Order statistics for discovery (D. Cox)

Limits, etc. (L. Demortier, D. van Dyk)

Likelihood ratio tests (GDC, C. Roever, J. Conway)

More on p-values (F. Beaujean)

Order statistics for discovery

D. Cox

Tests at n positions. Statistical independence.

Gives a set $\{P_1, \ldots, P_n\}$. transform to $Z = -\log P$

Define the order statistics $Z_{(1)} \leq Z_{(2)} \leq \ldots \leq Z_{(n)} = \max Z_j.$

Plot of the ordered Z is helpful descriptively and is the basis for various formal tests. In particular

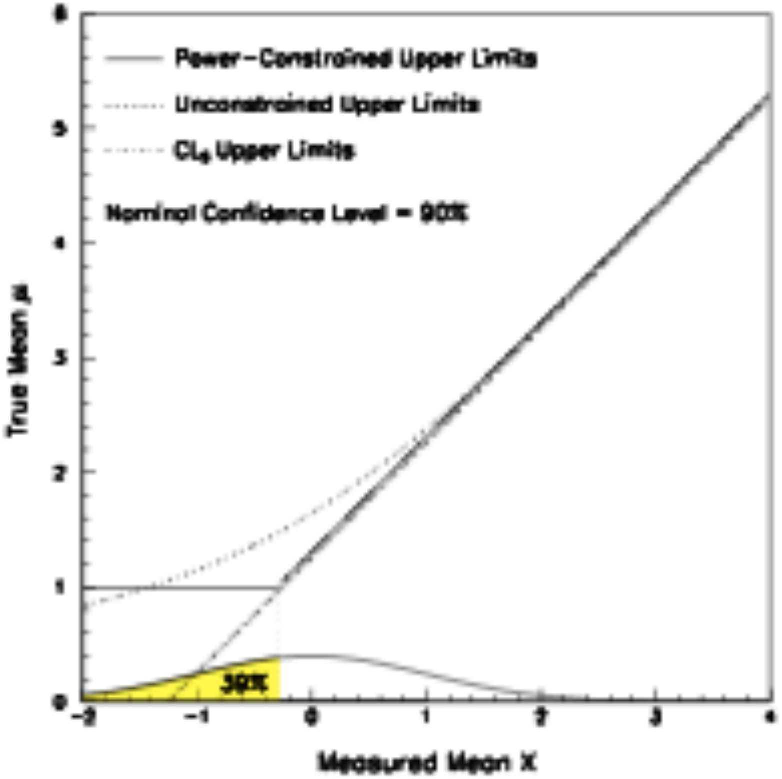
- null situation is a straight line of unit slope
- simplest alternative is one outlying point
- incorrect null distribution leads to a smooth curve
- internal correlation yields a straight line of slope different from one

Can HEP use this to look for a bump in a histogram? Need to use modified version where signal smeared over several bins.

Limits, etc.

L. Demortier

An alternative approach is to report the observed upper limit only if it is above a prespecified "sensitivity bound". If the observed limit is below the bound, only the bound itself is reported. This was proposed by V. Highland in an unpublished note in 1987. Some colleagues from ATLAS have motivated this method with a statistical power argument: you shouldn't reject a given parameter value unless you have a decent probability of detecting it when it is the true value. Hence the name "power-constrained limits" (PCL). A delicate issue here is the choice of sensitivity bound.



PCL, CLs, "unconstrained" limits for measurement of $X \sim \text{Gaussian}(\mu)$

Limits, etc. (2)

D. van Dyk

Proposed procedure:

Always report

1. Whether the source was detected.
2. A Confidence Interval for the source intensity.
   - This may be a one-sided interval taking the form of an upper limit.
3. The sensitivity, in order to quantify the strength of the experiment.

Corrections to standard UL

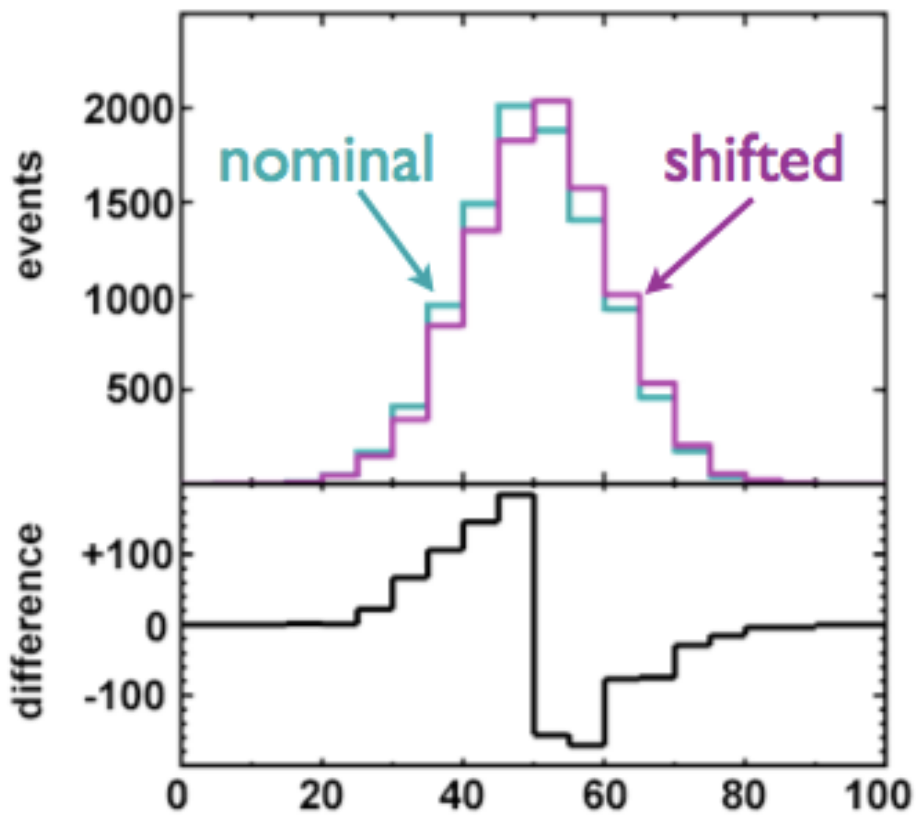
- PCL mixes a standard UL with the sensitivity.
- CL_S alters the UL for a smoothed version of PCL.

But with PCL it should be easy to communicate where the limit (bound) is the observed one, and where it is the power threshold.

Improving the model: template morphing

J. Conway

Often shift in nuisance parameters causes complicated change in a distribution – parametrize with e.g. template morphing:



$$\mu_i = \mu_i^0 + f\mathcal{M}(\mu_i^-, \mu_i^0, \mu_i^+)$$

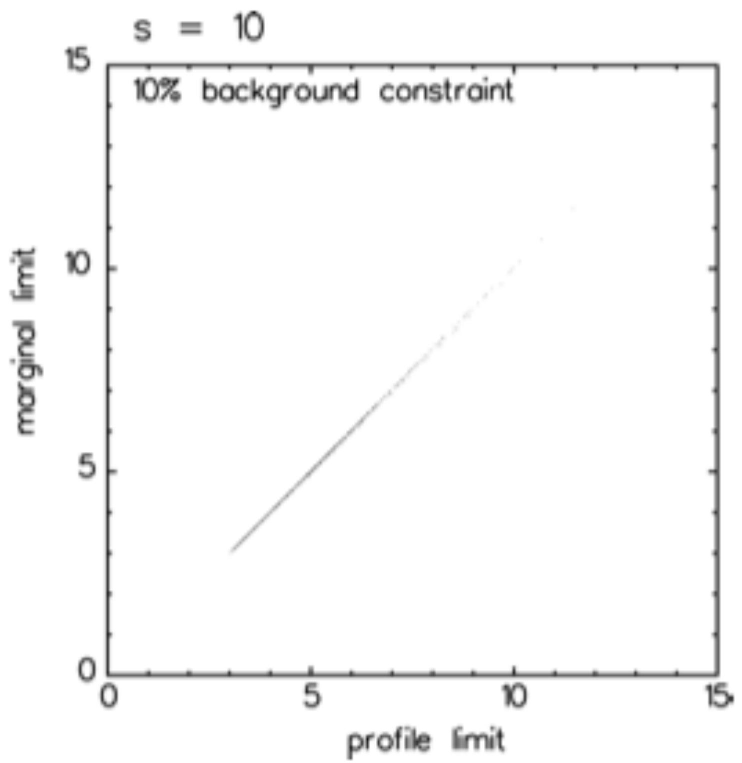
Need to do this e.g. to properly include systematics due to jet energy scale, which affects different distributions in different ways.

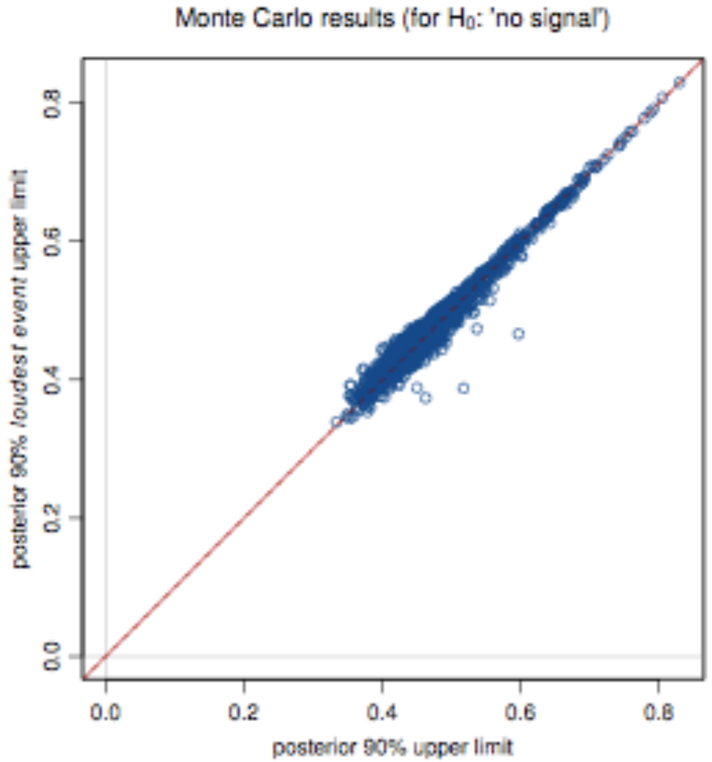
Marginalize vs. maximize

J. Conway,
C. Roever

The point was raised as to whether it is better in some sense to construct a ratio of marginalized or profile likelihoods.

Conway, Roever see little difference:





Comment on profile likelihood

Suppose originally we measure x, likelihood is $L(x|\theta)$.

To cover a systematic, we enlarge model to include a nuisance parameter ν, new model is $L(x|\theta, \nu)$.

To use profile likelihood, data must constrain the nuisance parameters, otherwise suffer loss of accuracy in parameters of interest.

Can e.g. use a separate measurement to constrain ν, e.g., with likelihood $L(y|\nu)$. This becomes part of the full likelihood, i.e.,

$$L(x, y|\theta, \nu) = L(x|\theta, \nu)L(y|\nu)$$

Comment on marginal likelihood

When using a prior to reflect knowledge of ν, often one treats this as coming from the measurement y, i.e.,

$$\pi(\nu) \propto L(y|\nu)\pi_0(\nu)$$

original prior,

Then the marginal likelihood is

$$L_{\mathrm{m}}(\theta) = \int L(x|\theta,\nu)\pi(\nu)\,d\nu$$

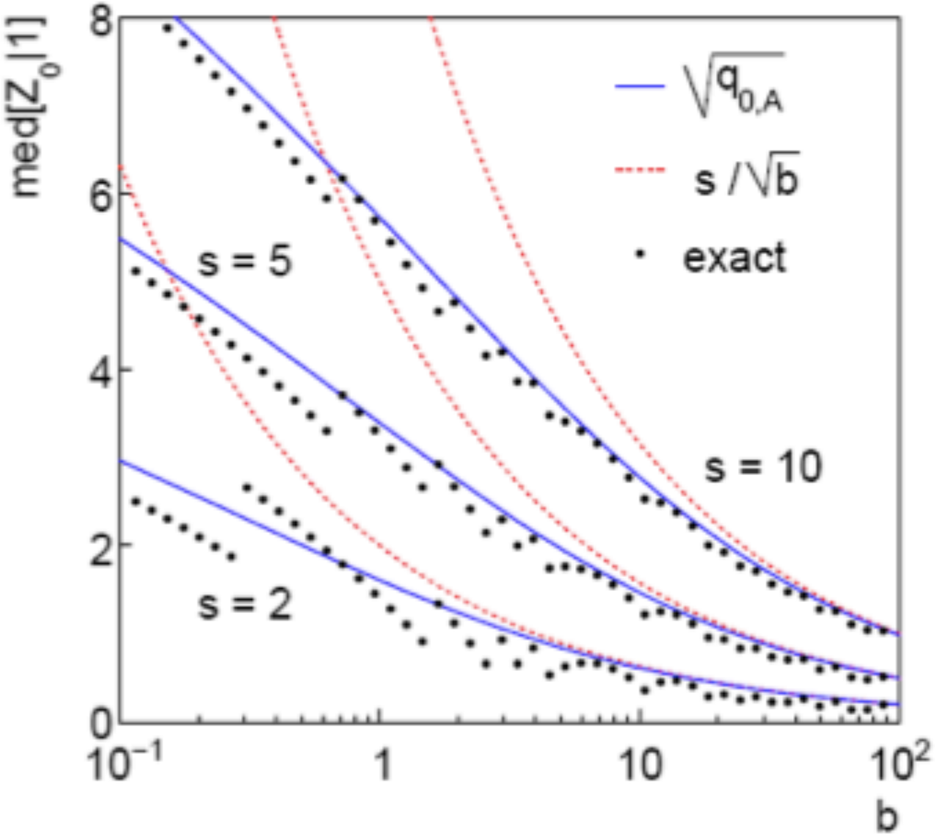
So here L in the integrand does not include the information from the measurement y; this is included in the prior.

One more comment on Asimov...

GDC

n ~ Poisson(μs+b), median significance, assuming μ = 1, with which one would reject μ = 0.

$$\mathrm{median}[Z_0|s+b] \approx \sqrt{2\left((s+b)\ln(1+s/b)-s\right)}$$



"Exact" values from MC, jumps due to discrete data.

Asimov $\sqrt{q_{0,A}}$ good approx. for broad range of s, b.

$s/\sqrt{b}$ only good for $s \ll b$.

Bayesian methods

Reference priors (L. Demortier, J. Bernardo, M. Pierini)

Bayes factors (J. Berger)

Application to sparse spectra (A. Caldwell)

Bayes factors

J. Berger

Bayes factor of H_0 to H_1: *ratio of likelihood under* H_0 *to average likelihood under* H_1 (or "odds" of H_0 to H_1)

E.g. apply to Poisson counting problem: $N \sim$ Poisson(s+b)

$$B_{01}(N) = \frac{\text{Poisson}(N \mid 0 + b)}{\int_0^\infty \text{Poisson}(N \mid s + b)\pi(s)\ ds} = \frac{b^N\ e^{-b}}{\int_0^\infty (s+b)^N\ e^{-(s+b)}\pi(s)\ ds}$$

(1) Choose $\pi(s)$ subjectively (Not easy for HEP applications)

(2) Choose $\pi(s)$ to be the 'intrinsic prior' $\pi^I(s) = b(s+b)^{-2}$.

$$B_{01} = \frac{b^N\ e^{-b}}{\int_0^\infty (s+b)^N\ e^{-(s+b)} b(s+b)^{-2}\ ds} = \frac{b^{(N-1)}\ e^{-b}}{\Gamma(N-1, b)}$$

Case 1: $p = 0.00025$ if $N = 7$, $b = 1.2$ $\longrightarrow$ $B_{01} = 0.0075$

Case 2: $p = 0.025$ if $N = 6$, $b = 2.2$. $\longrightarrow$ $B_{01} = 0.26$

Bayes factors (2)

J. Berger

(3) Lower bound on Bayes factor: make $\pi(s)$ a delta function at $\hat{s}$.

$$B_{01}(N) = \frac{\text{Poisson}(N \mid 0+b)}{\int_0^\infty \text{Poisson}(N \mid s+b)\pi(s)\, ds} \geq \frac{\text{Poisson}(N \mid 0+b)}{\text{Poisson}(N \mid \hat{s}+b)}$$

$$= \min\{1, \left(\frac{b}{N}\right)^N e^{N-b}\}.$$

Case 1: $B_{01} \geq 0.0014$ (recall $p = 0.00025$)

Case 2: $B_{01} \geq 0.11$ (recall $p = 0.025$)

So even lowest Bayes factors substantially bigger than the p-values.

(4?) Can we not simply plot the Bayes factor vs. s (i.e. report B_{0s}, not B_{01})?

Bayes factors (3)

"Bayes factor more intuitive than a *p*-value." ("5σ" ↔ B_{01} = ?)

"Automatic" inclusion of look-elsewhere effect (through prior).

But, marginal likelihoods can be difficult to compute:

$$m = \int L(\vec{x}|\vec{\theta})\pi(\vec{\theta})\,d\vec{\theta}$$

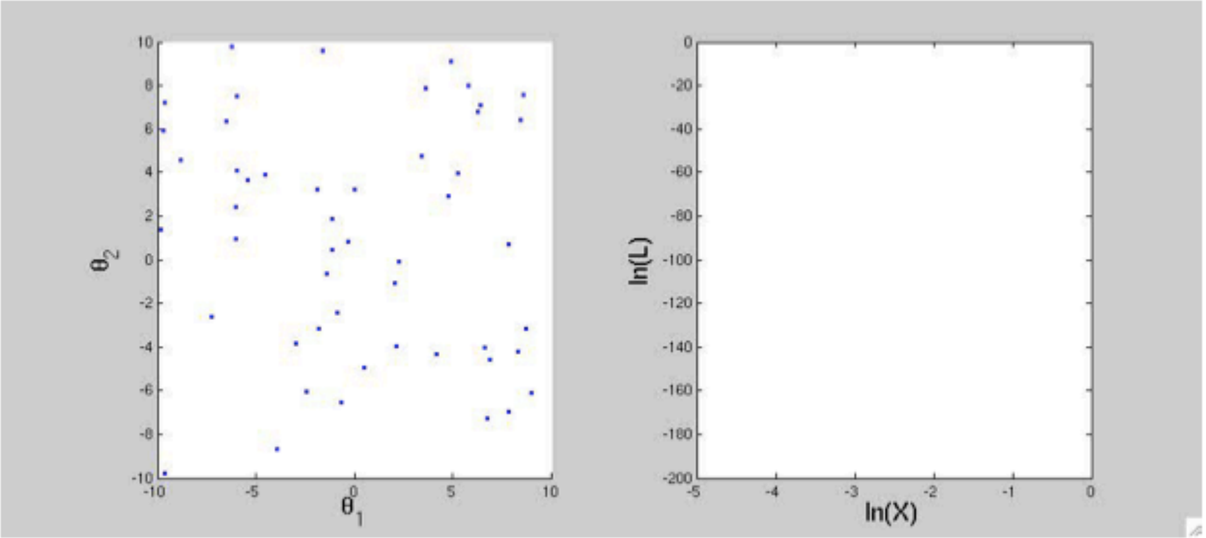
Can we use e.g.

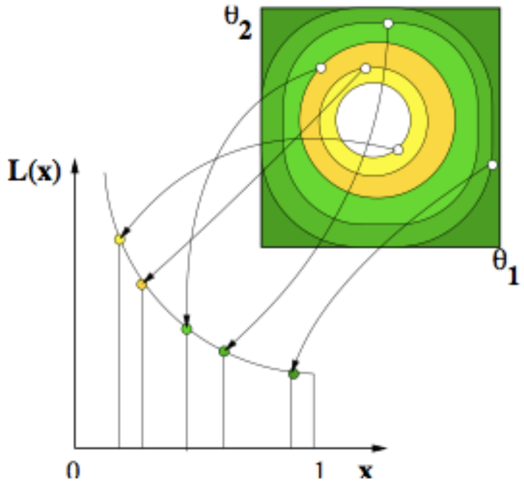
MultiNest: a multi-modal implementation of nested sampling. Also an extremely efficient sampler for multi-modal likelihoods

Feroz & Hobson (2007), RT et al (2008), Feroz et al (2008)

(K. Cranmer/R. Trotta)

The nested sampling algorithm





(animation courtesy of David Parkinson)

An algorithm originally aimed primarily at the Bayesian evidence computation (Skilling, 2006):

$$X(\lambda) = \int_{\mathcal{L}(\theta)>\lambda} P(\theta) d\theta$$

$$P(d) = \int d\theta \mathcal{L}(\theta) P(\theta) = \int_0^1 X(\lambda) d\lambda$$

Feroz et al (2008), arxiv: 0807.4512, Trotta et al (2008), arxiv: 0809.3792

Reference priors

J. Bernardo,
L. Demortier,
M. Pierini

Maximize the expected Kullback–Leibler divergence of posterior relative to prior:

$$D[\pi, p] \equiv \int p(\theta|x) \ \ln \frac{p(\theta|x)}{\pi(\theta)} \, d\theta$$

This maximizes the expected posterior information about θ when the prior density is $\pi(\theta)$.

Finding reference priors "easy" for one parameter:

Theorem 1 *Let $\boldsymbol{z}^{(k)} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k\}$ denote k conditionally independent observations from $\mathcal{M}_{\boldsymbol{z}}$. For sufficiently large k*

$$\pi_k(\theta) \propto \exp\{\mathrm{E}_{\boldsymbol{z}^{(k)}\,|\,\theta}[\log p_h(\theta \,|\, \boldsymbol{z}^{(k)})]\}$$

where $p_h(\theta \,|\, \boldsymbol{z}^{(k)}) \propto \prod_{i=1}^{k} p(\boldsymbol{z}_i \,|\, \theta)\, h(\theta)$ is the posterior which corresponds to any arbitrarily chosen strictly positive prior function $h(\theta)$ which makes the posterior proper for any $\boldsymbol{z}^{(k)}$.

Reference priors (2)

J. Bernardo,
L. Demortier,
M. Pierini

Actual recipe to find reference prior nontrivial;
see references from Bernardo's talk, website of
Berger (www.stat.duke.edu/~berger/papers) and also
Demortier, Jain, Prosper, PRD 82:33, 34002 arXiv:1002.1111:

$$
\begin{aligned}
\pi_R(\theta) &= \lim_{k\to\infty} \frac{\pi_k(\theta)}{\pi_k(\theta_0)}, \\
\text{with } \pi_k(\theta) &= \exp\left\{ \int p(x_{(k)}\,|\,\theta)\, \ln\left[\frac{p(x_{(k)}\,|\,\theta)\, h(\theta)}{\int p(x_{(k)}\,|\,\theta)\, h(\theta)\, d\theta} \right] dx_{(k)} \right\}
\end{aligned}
$$

Prior depends on order of parameters. (Is order dependence
important? Symmetrize? Sample result from different orderings?)

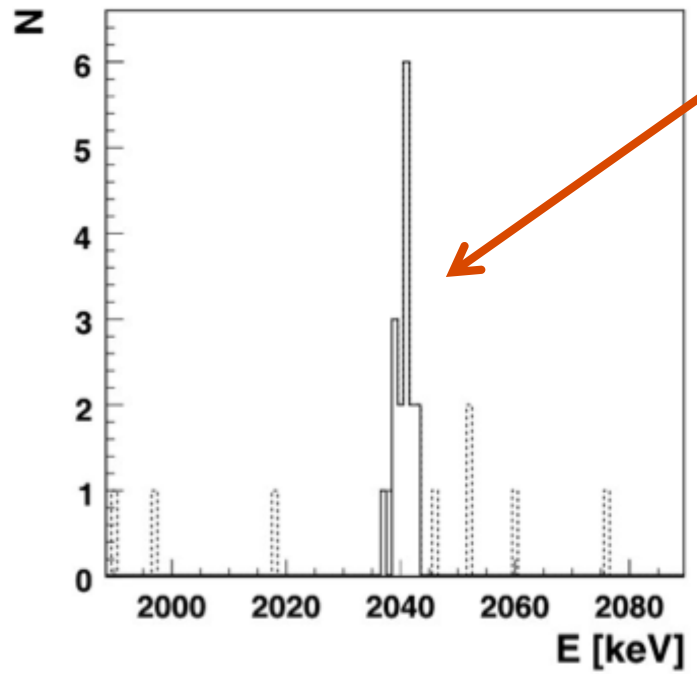
L. Demortier

There still seem to be some important puzzles regarding reference priors:

1. What is the proper probabilistic interpretation of a reference posterior?
   - Reference posterior probabilities are not subjective probabilities! So what are they then?
   - Can reference posterior inferences be reported by themselves, or should they be reported only as part of a sensitivity analysis? If the latter, how should one choose alternative priors?
2. How should we deal with the compact set normalization procedure?
   - The general definition of reference priors involves the taking of limits, and this must be done carefully in order to avoid infinities; the standard approach is to use sequences of nested compact sets that converge to the whole parameter space.
   - Unfortunately there is no unique way of choosing these compact sets, and there is no guarantee that different choices lead to the same result, or even that all choices lead to a proper posterior.
   - This ambiguity prevents us from designing a completely general numerical algorithm.
3. How should we handle implicit statistical models?
   - Can we combine ABC methods with numerical algorithms for computing reference posteriors?

Bayesian discovery with sparse data

A. Caldwell



Discovery or not ?

Meaningful elicitation of prior from community consensus, here:

$$p_0(H) = p_0(\overline{H}) = 1/2$$

Consensus priors doable in practice? (Committee?)

p(H|spectrum)<0.01, 'evidence' (better >99% belief in `new physics')

p(H|spectrum)<0.0001, 'discovery' (better >99.99% belief in `new physics') (very stringent, DoB contains our belief in the new physics)

Note: intended to be the real 'degree-of-belief'. **No fudging allowed afterwards – otherwise it implies you did not really believe your prior.**

More on p-values (model validation)

F. Beaujean

Suppose we're only given p-values – how should this influence a Bayesian's degree of belief?

- Similar prior for all models $P(M_i) \approx P(M_j)$
- Bayes Theorem: $P(M_0|p) \approx \dfrac{P(p|M_0)}{\sum_{i=0}^{K} P(p|M_i)}$

Small p

$$P(M_0|p \approx 0) \approx \frac{1}{1 + \sum_{i=1}^{K} c_i} \ll 1$$

Large p

$$P(M_0|p \approx 1) \approx 1$$

OK but... not same as $P(M_0|\text{data})$, rather a justification of p-value.

Or, justify by saying that a bizarre result prompts one to comment on (and quantify) just how bizarre it is.

J. Bernardo,
D. van Dyk

Decision theory

Statistical methods can be formulated in elegant but (for particle physicists) not fully familiar language of decision theory:

Specify loss function, minimize expected loss, etc.

How does this map onto the usual ways that physicists view discovery/limits?

What are specific benefits for HEP of this approach?

"It's a tool to be aware of." – D. van Dyk

Combining results

N. Krasnikov,
K. Cranmer

Given p-values $p_1,\ldots, p_N$ of H, what is combined p?

Rather, given the results of N (usually independent) experiments, what inferences can one draw from their combination?

Full combination is difficult but worth the effort for e.g. combined ATLAS/CMS Higgs search.

Form full likelihood function of the joint experiment;

Use to construct statistical tests (e.g. likelihood ratio)

Single common parameter of interest: $\mu = \sigma/\sigma_{\mathrm{SM}}$

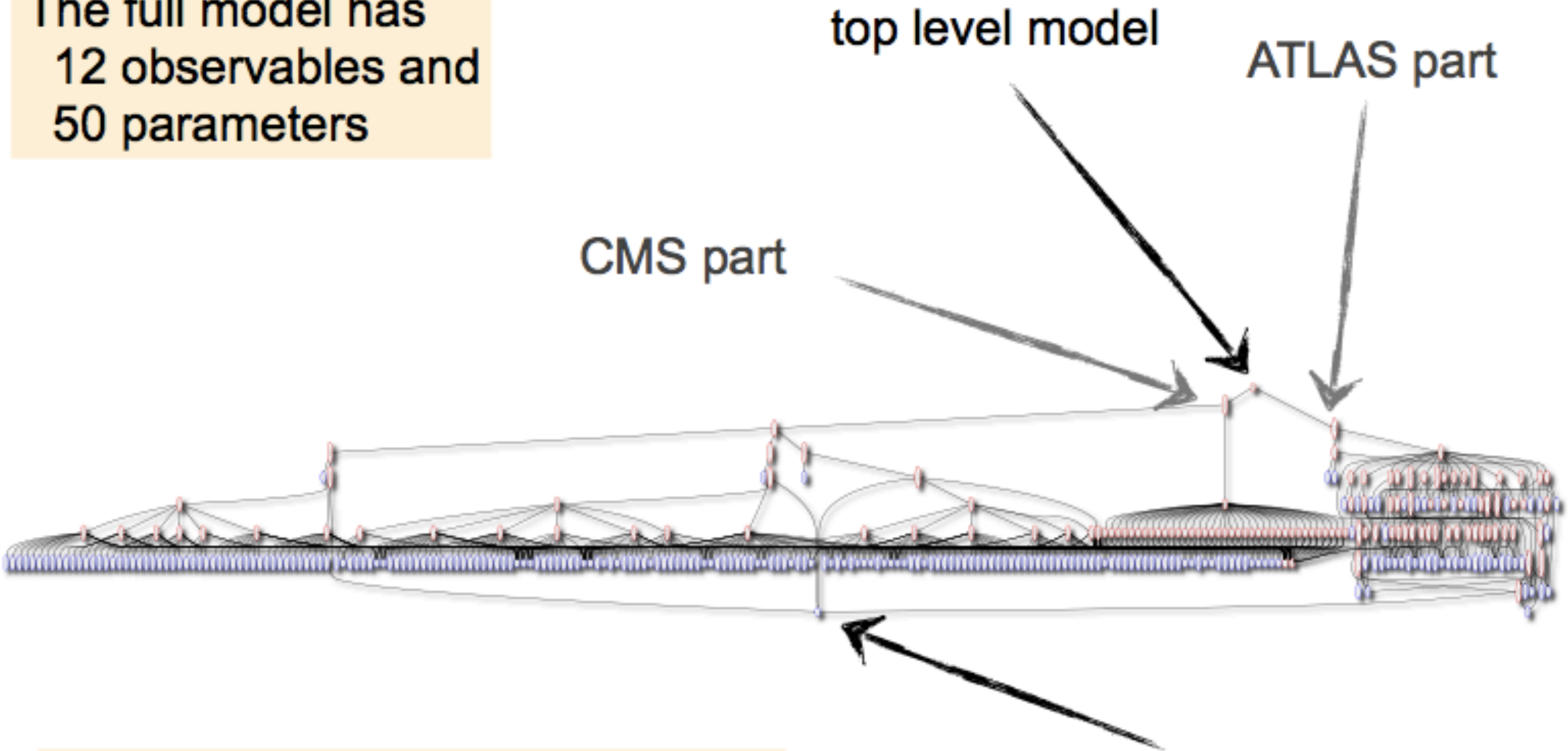
Also (in principle) common nuisance parameters (e.g., luminosity, parton uncerainties,...)

New software: RooStats (RooFit/ROOT).

K. Cranmer

Combined ATLAS/CMS Higgs search

The full model has 12 observables and 50 parameters



$$\mu = \frac{\sigma BR}{\sigma_{SM} BR_{SM}}$$

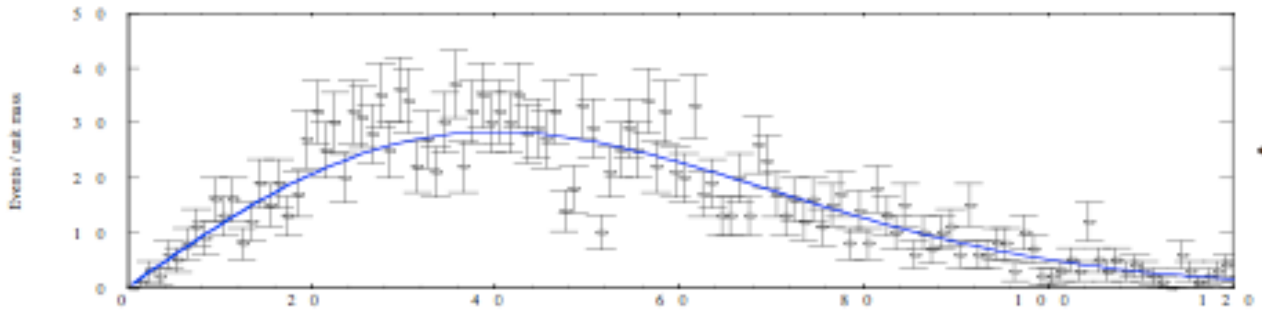
At this point, no correlated systematics across experiments

Look-elsewhere effect

O. Vitells,
G. Ranucci,
A. Caldwell

[(E. Gross and O. Vitells, Eur. Phys. J. C, 70, 1-2, (2010) , arXiv:1005.1891]

- The "look elsewhere" effect occurs when one searches for a signal in some space of parameters (mass, shape, location in the sky, etc.)
- In the language of Hypothesis testing:
  test H_0 (no signal) against $H_1(\theta)$,
  The signal parameters (θ) are not present under H_0 --
  Wilks' theorem does not apply
- The problem is to correctly estimate the p-value of a "local" excess of events, taking into account the full range.
- Monte-Carlo simulation is a straight-forward way, but can be computationally very expansive

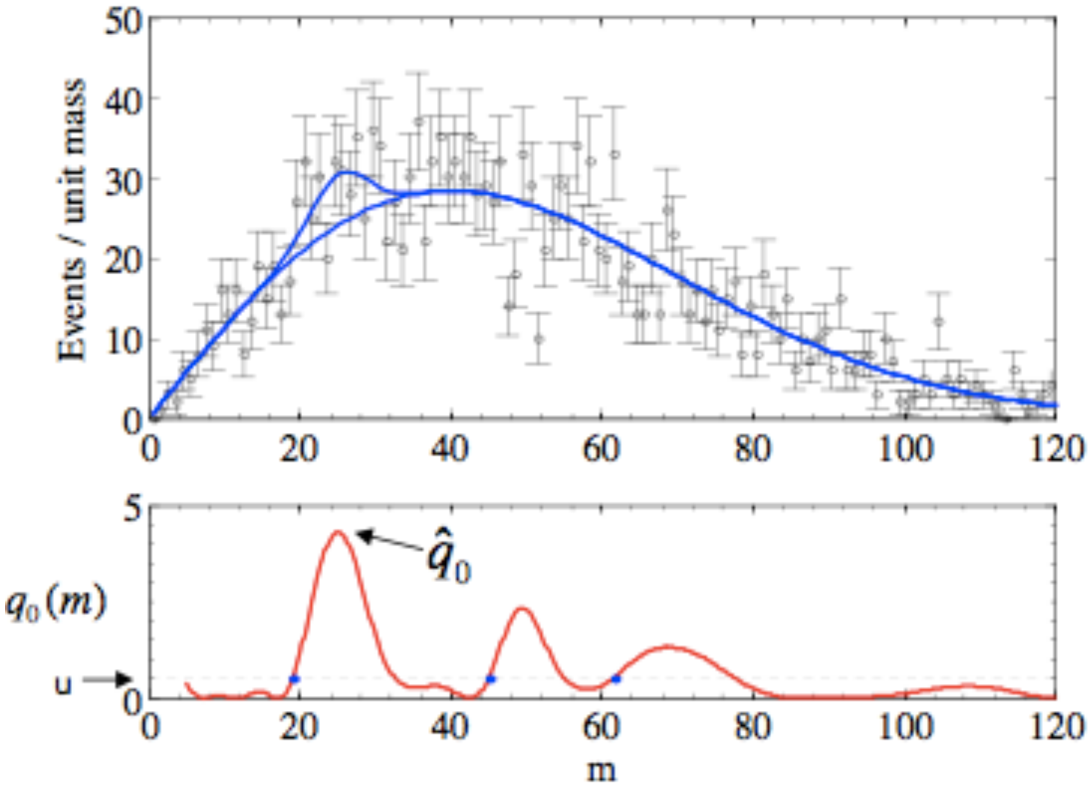


Search for a peak in some mass range

Look-elsewhere effect (2)

O. Vitells

Correction to p-value from theory of random fields; related to mean number of “upcrossings” of likelihood ratio (Davies, 1987).



$$P(q_0 > u)$$
$$\leq E[N_u] + P(q_0(0) > u)$$
$$= \mathcal{N}_1 e^{-u/2} + \frac{1}{2} P(\chi_1^2 > u)$$

estimate with MC at low reference level

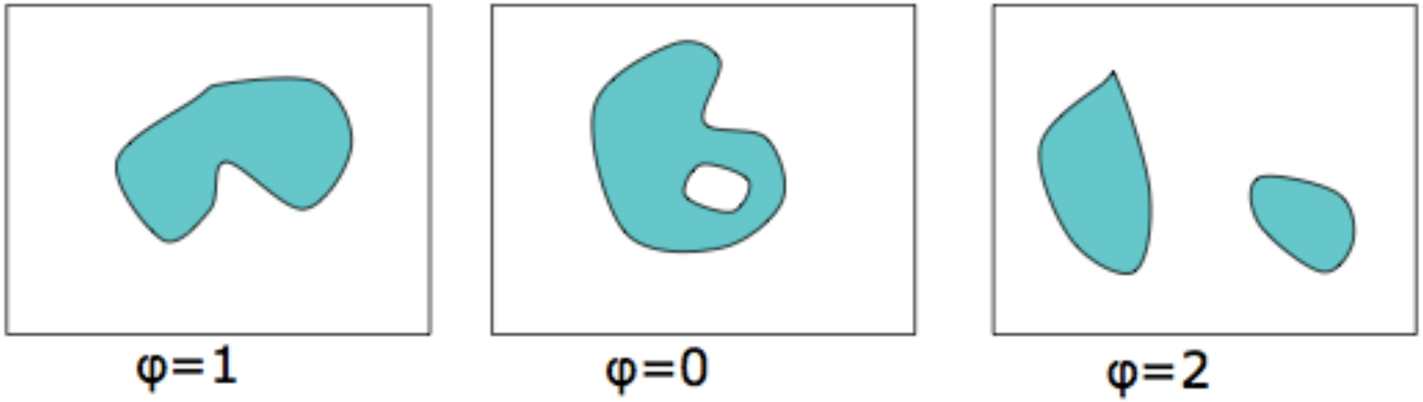
Look-elsewhere effect (3)

O. Vitells

Generalization to multiple dimensions: number of upcrossings replaced by expectation of Euler characteristic:

$$E[\varphi(A_u)] = \sum_{d=0}^{n} \mathcal{N}_d \rho_d(u)$$

- Number of disconnected components minus number of `holes'



Applications: astrophysics (coordinates on sky), search for resonance of unknown mass and width, ...

Look-elsewhere effect in time series

G. Ranucci

The correct incorporation of the **Look Elsewhere Effect** is vital while searching for a modulation hidden in time series, otherwise **"look long enough, find anything!"** (Numerical Recipes)

The **LEE** change completely the detection scenario passing from the single frequency to the multiple frequency strategy search

The sensitivity to low modulation amplitude is severely affected

Same situation as the search of a particle of unknown mass

A parallelism can be established between the frequency search and a "prototype" approach to the scan of a mass range, also **through similar analytical formalisms**

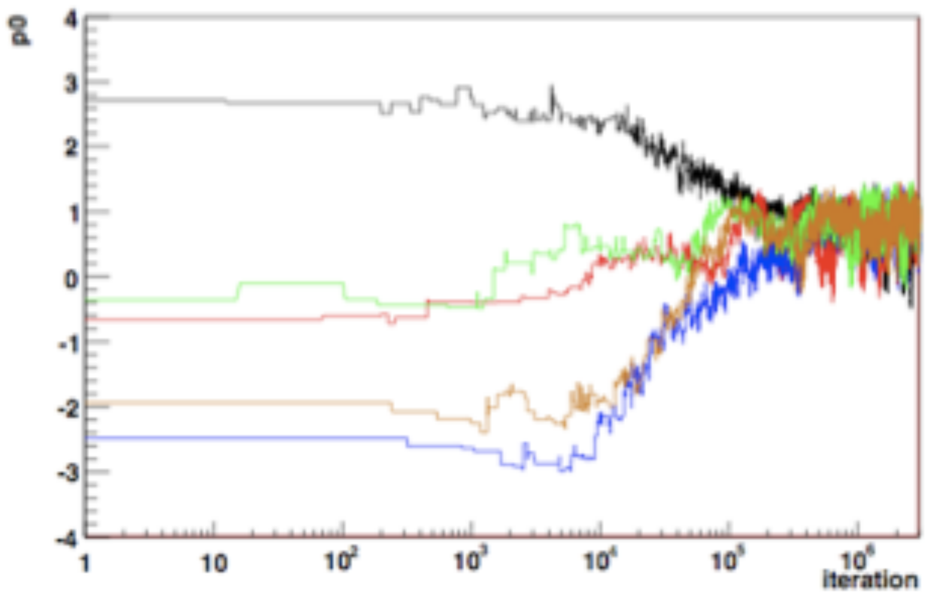
Finally what modulation in the solar neutrino data? So far, only the **annual modulation**

S. Pashapour

Bayesian Analysis Toolkit (BAT)

General framework specifically for Bayesian computation, especially MCMC for marginalizing posterior probabilities.

E.g. convergence diagnostics à la Gelman & Rubin,



Wish list(?): computation of marginal likelihoods;
support for important types of priors (reference,...)

RooStats

G. Schott

a collaborative project with contributors from ATLAS, CMS and ROOT aimed to provide & consolidate statistical tools needed by LHC

- using same tools: compare easily results across experiments
  - not only desirable but necessary for combinations

RooStats is built on top of the RooFit toolkit :

- data modelling language (for PDFs, likelihoods, ...)

RooFit Workspaces

G. Schott

RooWorkspace class of RooFit: possibility to save it to a ROOT file

- very good for electronic publication of data and likelihood function
- and greatly help for combination (that's the format agreed to share between Atlas & CMS)

```
RooWorkspace w("w","joint workspace") ;
// Import top-level pdfs and all their components, variables
w.import("channelA.root:w:pdfA",RenameAllVariablesExcept("A","mhiggs"))
w.import("channelB.root:w:pdfB",RenameVariable("mH","mhiggs")) ;
w.import("channelC.root:w:pdfC") ;
// Construct joint pdf
w.factory("SIMUL::joint(chan[A,B,C],A=pdfA,B=pdfB,C=pdfC)") ;
```

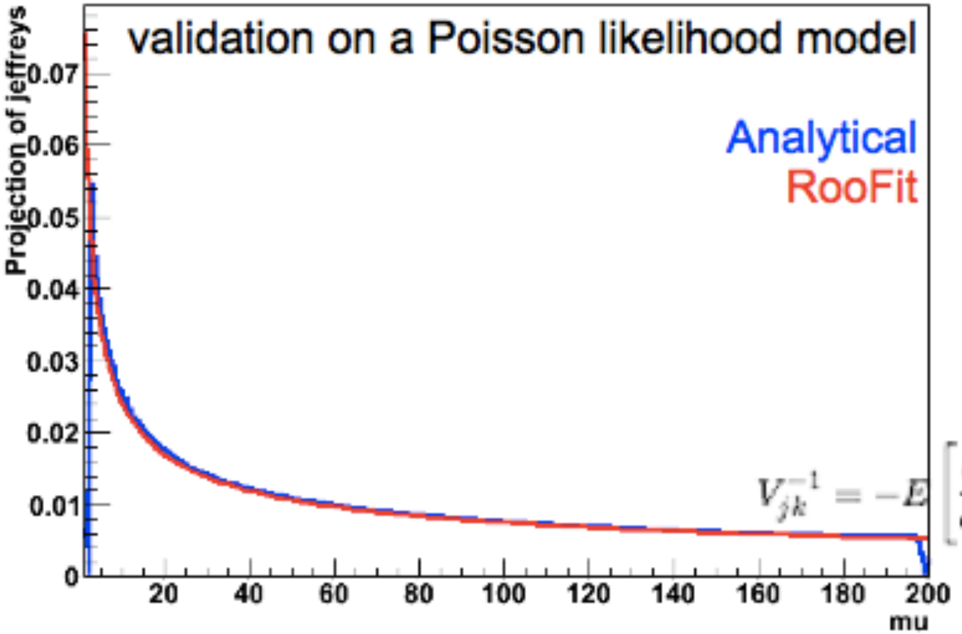
Able to construct full likelihood for combination of channels (or experiments).

RooStats

G. Schott

Example of tools: Bayesian priors from formal rules:

- New RooJeffreys class: "objective" prior based on formal rules (related to Fisher information and the Cramér-Rao bound)



- implemented for arbitrary PDF using "Asimov" dataset to help calculate the Fisher information [arXiv:1007:1727]

$$\pi(\vec{\theta}) \propto \sqrt{\det \mathcal{I}\left(\vec{\theta}\right)}.$$

$$\left(\mathcal{I}\left(\theta\right)\right)_{i,j} = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\ln f(X;\theta)\bigg|\theta\right].$$

$$V^{-1}_{jk} = -E\left[\frac{\partial^2 \ln L}{\partial\theta_j \partial\theta_k}\right] = -\frac{\partial^2 \ln L_A}{\partial\theta_j \partial\theta_k} = \sum_{i=1}^{N} \frac{\partial\nu_i}{\partial\theta_j}\frac{\partial\nu_i}{\partial\theta_k}\frac{1}{\nu_i} + \sum_{i=1}^{M} \frac{\partial u_i}{\partial\theta_j}\frac{\partial u_i}{\partial\theta_k}\frac{1}{u_i}$$

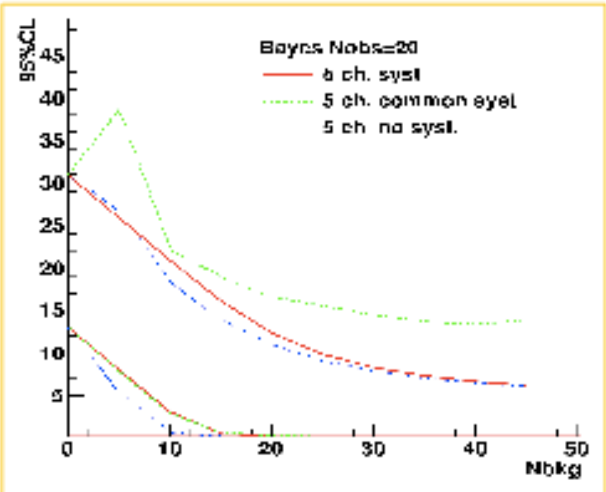
- Missing (but I heared some people are working on) a RooStats implementation of reference priors ...

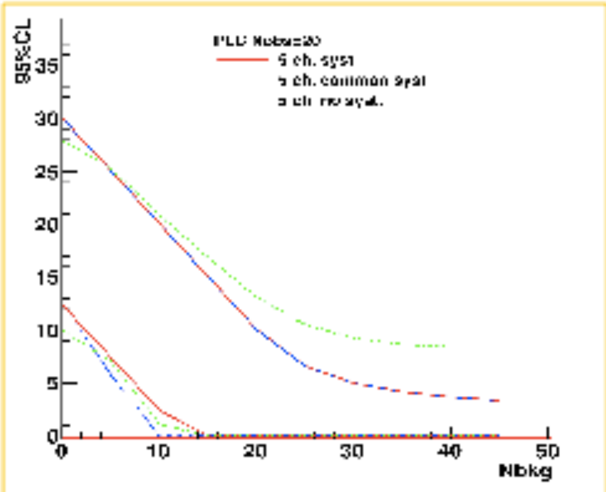
Examples of using RooStats

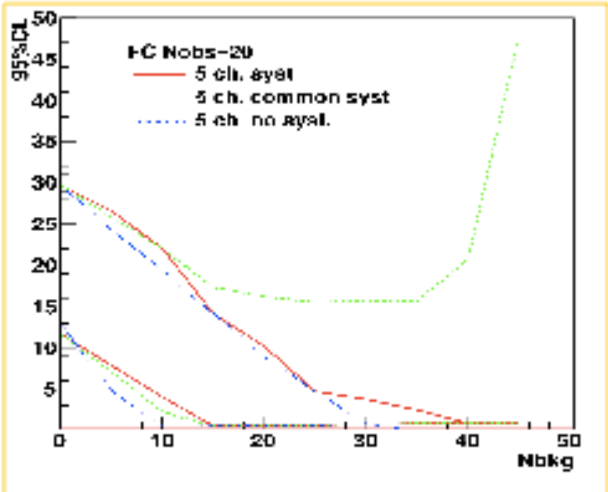
V. Zhukov,
K. Cranmer

E.g. studying effect of correlated systematics in combined fit using various methods with RooStats (V. Zhukov):

Combined model: 5 channels







And of course the previously mentioned ATLAS/CMS Higgs combination (K. Cranmer).

Lessons from the Tevatron

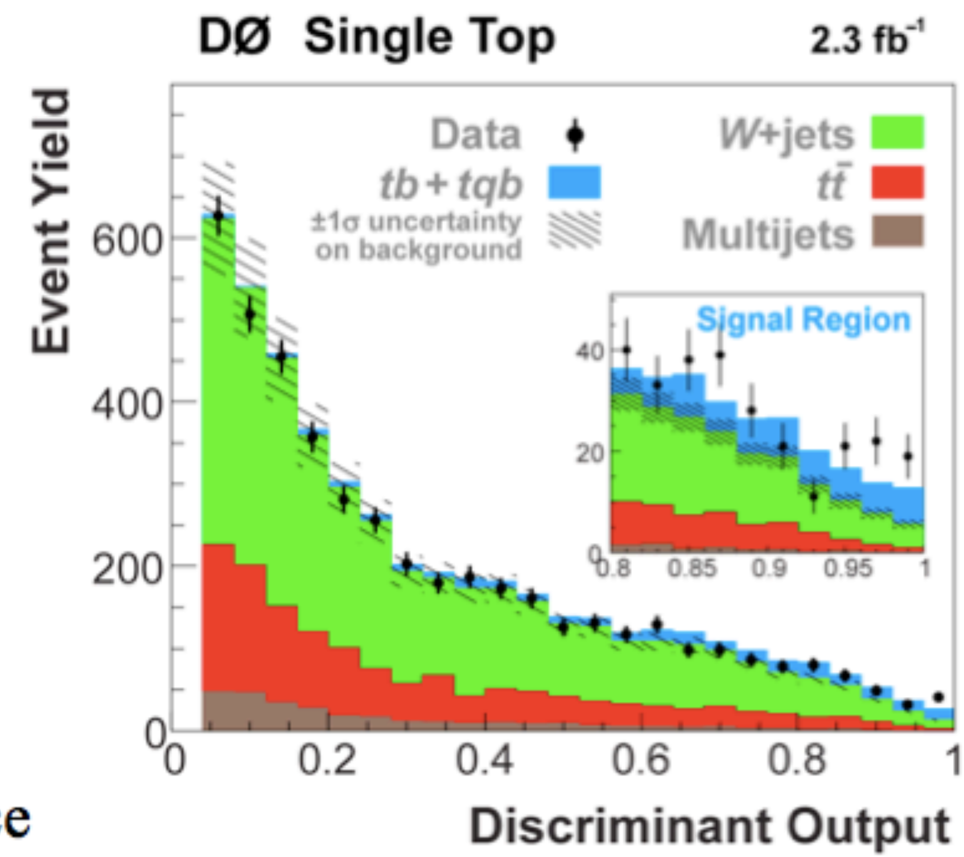
H. Prosper

Example: search for single top with multivariate analysis.
Need accurate understanding of background (mature experiments).
Would community accept MVA as readily if this were SUSY?

The data are reduced to M counts described by the likelihood

$$p(n \mid \sigma, \varepsilon, \mu) = \prod_{i=1}^{M} \text{Poisson}(n_i \mid \varepsilon_i \sigma + \mu_i)$$

where σ (the cross section) is the parameter of interest and the ε_i and μ_i are nuisance parameters.

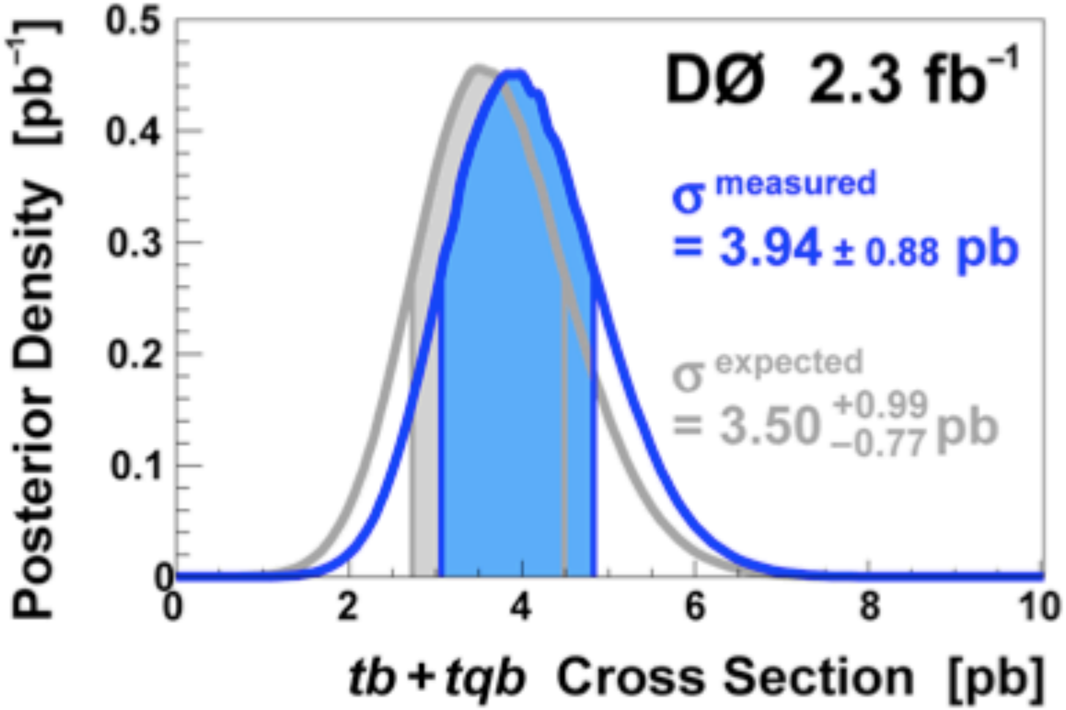


Bayesian analysis for cross section of single top

H. Prosper

D0 (and CDF) compute the posterior $p(\sigma \mid n)$ assuming:

1. a *flat* prior for $\pi(\sigma)$
2. an *evidence-based* prior for $\pi(\varepsilon, \mu)$



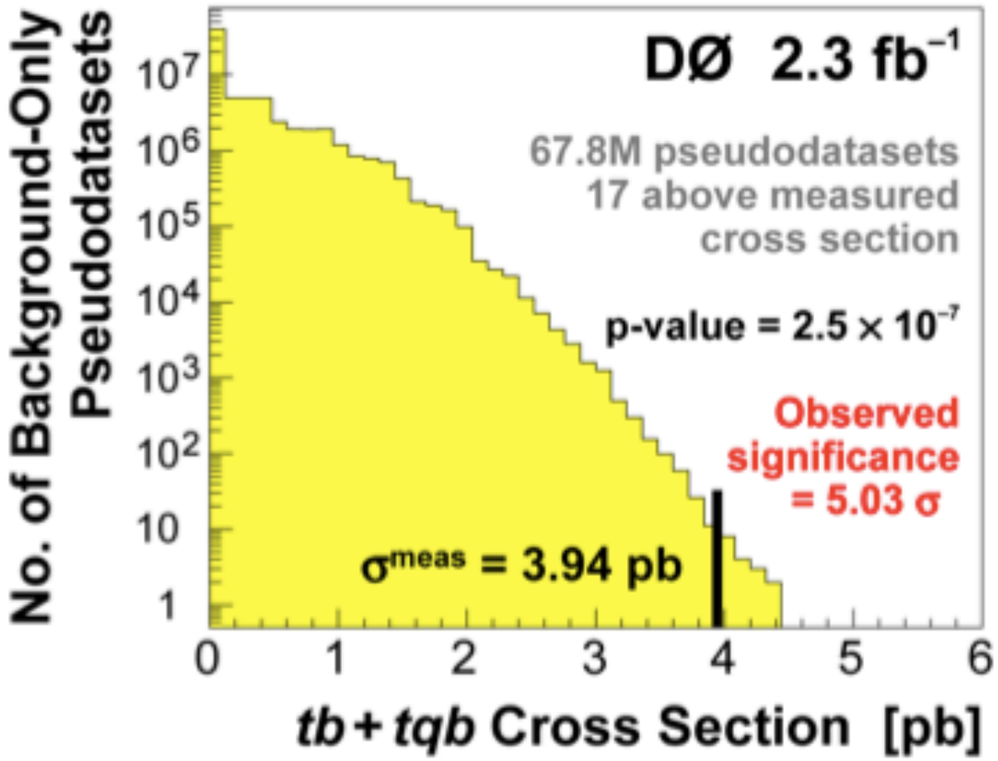
Converted to frequentist p-value for discovery of single top

H. Prosper

Estimate of "signal significance" using a p-value:

$$p_0 = \mathrm{P}[t > t_0 | H_0]$$

The statistic t is the mode of the the posterior density.



In practice: CMS, ATLAS, LHCb

A. Harel,
D. Casadei,
J. Morata

Publications from LHC have started to arrive!!!!!

The experimentalist perspective

Claiming a discovery first is the best case scenario.
But claiming a discovery is also the worst case scenario if you got it wrong.

Which of these statistical tools helps us get it right?

→A "pragmatic" approach is typical.
No standard approach. Yet.

A. Harel (CMS)

- So far, different analyses followed different routes
  - Gradually moving toward more uniformity
  - But impossible to ignore that real differences exist
    - There is no single "correct" method

D. Casadei (ATLAS)

Example: CMS W' search

A. Harel

The simplest scenario, as far as limit setting goes:

- A counting experiment (Poisson probability in each M_T bin)
- No interference between backgrounds and signal
- Systematic uncertainties factorize easily

Integrated luminosity

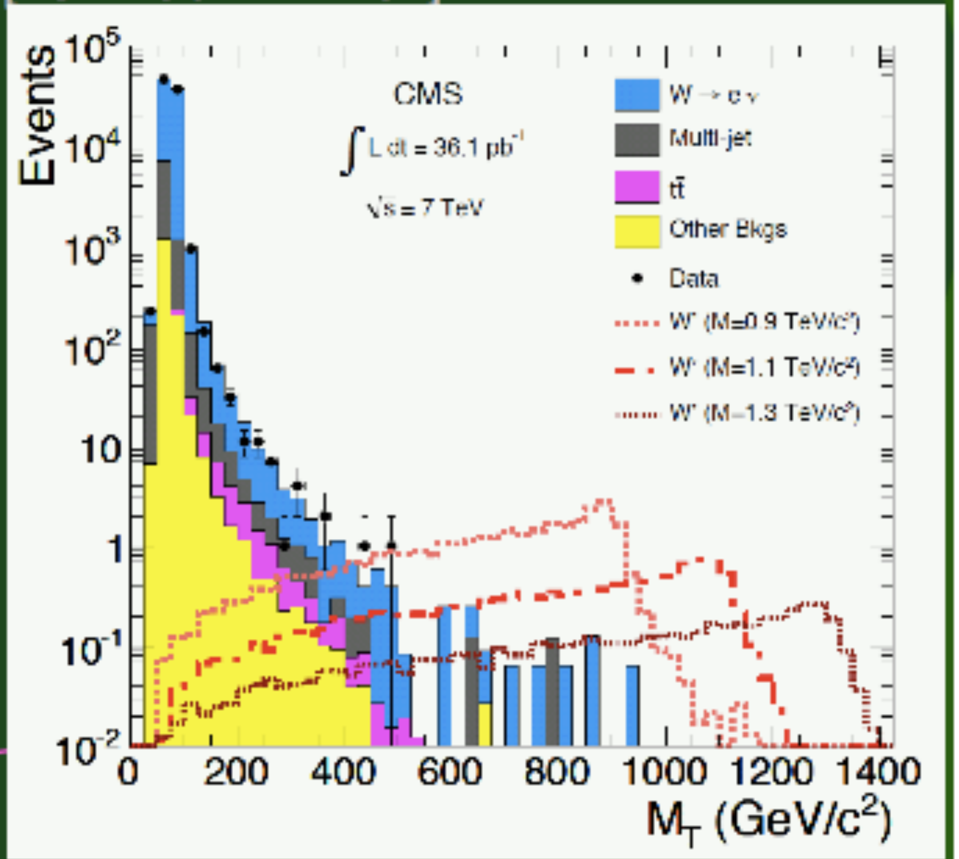
$$N_{pred} = b + \mathcal{L}\varepsilon\sigma_{\text{eff}}$$

Selection efficiency (for that bin)

Use a simple Bayesian procedure [Fermilab-TM-2104]

- nuisance parameters are integrated out
- priors:
  - $$p(\sigma_{\text{eff}}) = \begin{cases} \text{const} & \text{if } 0 < \sigma_{\text{eff}} < \sigma_{\max} \\ 0 & \text{otherwise} \end{cases}$$
  - log normal priors for the nuisance parameters $b, \mathcal{L}, \varepsilon$
    - background uncertainty (e.g. fit results) summarized in one number
    - typical approximation

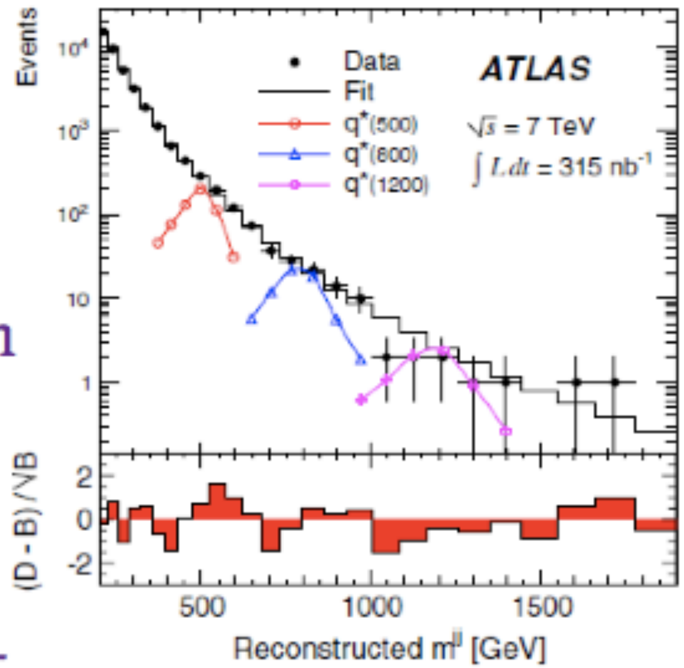
Rule out a W' with mass below 1.36TeV at 95% CL

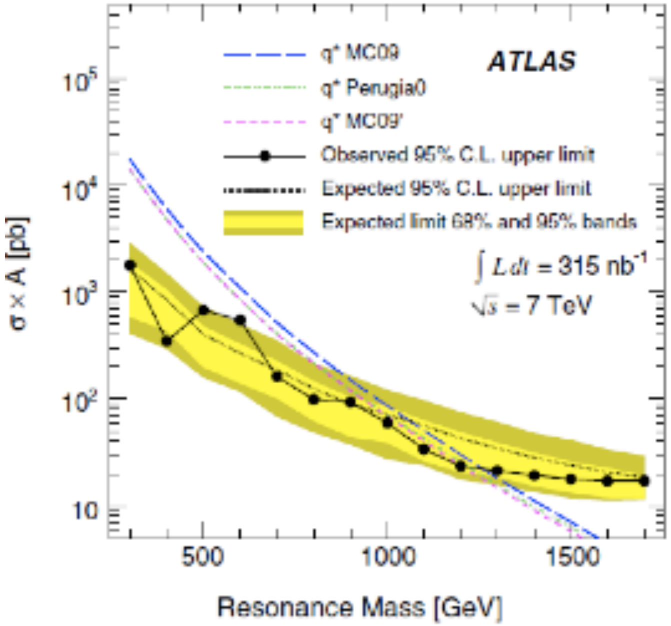


Example: ATLAS dijet resonance search

D. Casadei

- First step was to fit bkg model
  - Different statistics tested
  - No evidence for new physics
- For each hypothesized mass an upper limit has been obtained in the Bayesian approach
  - Likelihood = product of Poisson factors including both signal and background





- Coverage found by generating pseudo-experiments

[Phys. Lett. B694 (2011) 327]

Background spectrum and likelihood

$$f(x) = p_1(1-x)^{p_2} x^{p_3 + p_4 \ln x}$$

$$L_\nu(d \mid b_\nu, s) \equiv \prod_i \frac{[b_{\nu i} + s_i(\nu)]^{d_i}}{d_i!} e^{-[b_{\nu i} + s_i(\nu)]}$$

Example: LHCb search for $B_s \to \mu\mu$

J. Morata

Multivariate methods used: MLP, "$\Delta\chi^2$", BDT, etc.

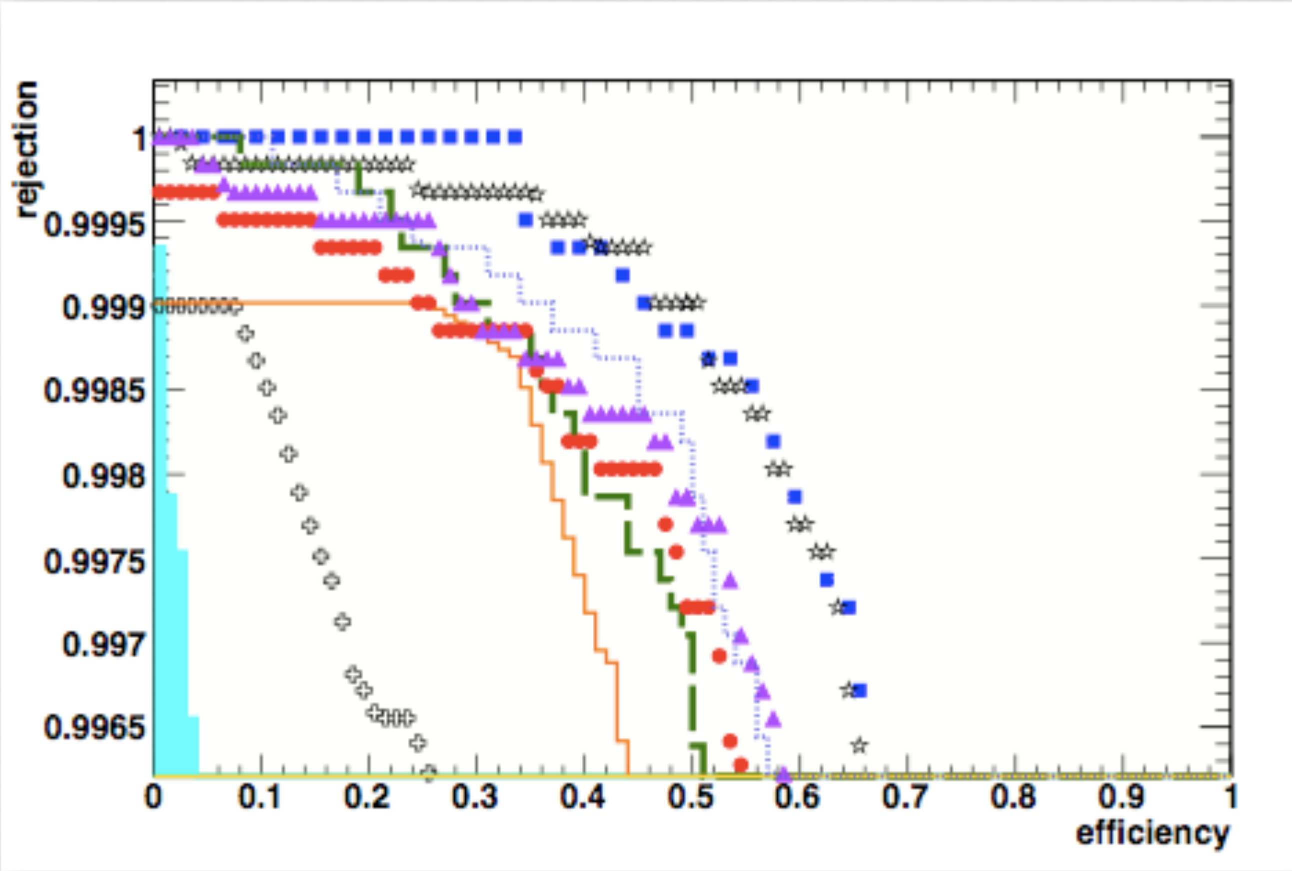


Figure 19: Performance of GL_K and set of other multivariate methods. The X axis shows the efficiency, and the Y axis the rejection. Blue squares: GL_K. Open stars: BDT. Short Dashed: PDERS-PCA. Violet triangles: Fisher Discriminant. Red cyrcles: Best performant NN.Green dashed line: Support Vector Machine.Orange solid line: RuleFit. Open crosses: less performant NN. Filled Cyan histogram: FDA-SA

Applications: parton densities

S. Forte

Fits to parton densities(e.g., MSTW, CTEQ) have found "reasonable" χ^2, but the standard procedure for errors, $\Delta\chi^2 = 1$, leads to unrealistically small variation of parton densities.

Could be consequence of:

inconsistent data;

inadequate model, e.g., parametric functions in fit (MSTW):

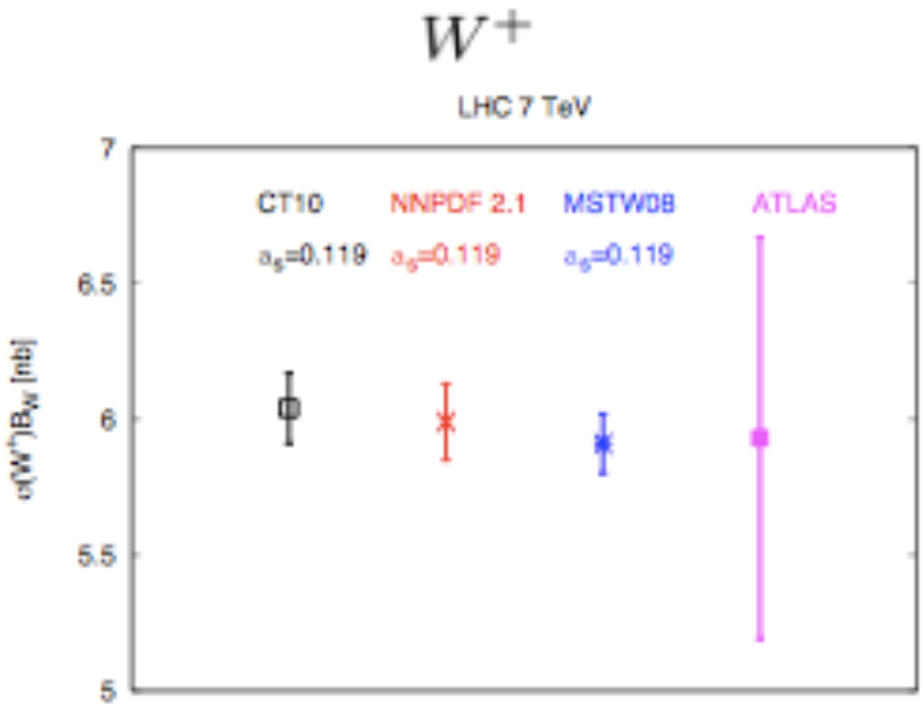
$$xq(x, Q_0^2) = A(1-x)^\eta (1+\epsilon x^{0.5}+\gamma x) x^\delta$$

NNPDF approach: parameterize using neural network; much larger number ($37\times 7 = 259$) of parameters compared to MSTW (20) or CTEQ (26); regularize using cross validation.

Partons (2)

S. Forte

NNPDF predictions quote uncertainties using “standard” Bayesian propagation of experimental uncertainties as obtained by CTEQ and MSTW when using very large $\Delta\chi^2$ (50 – 100).



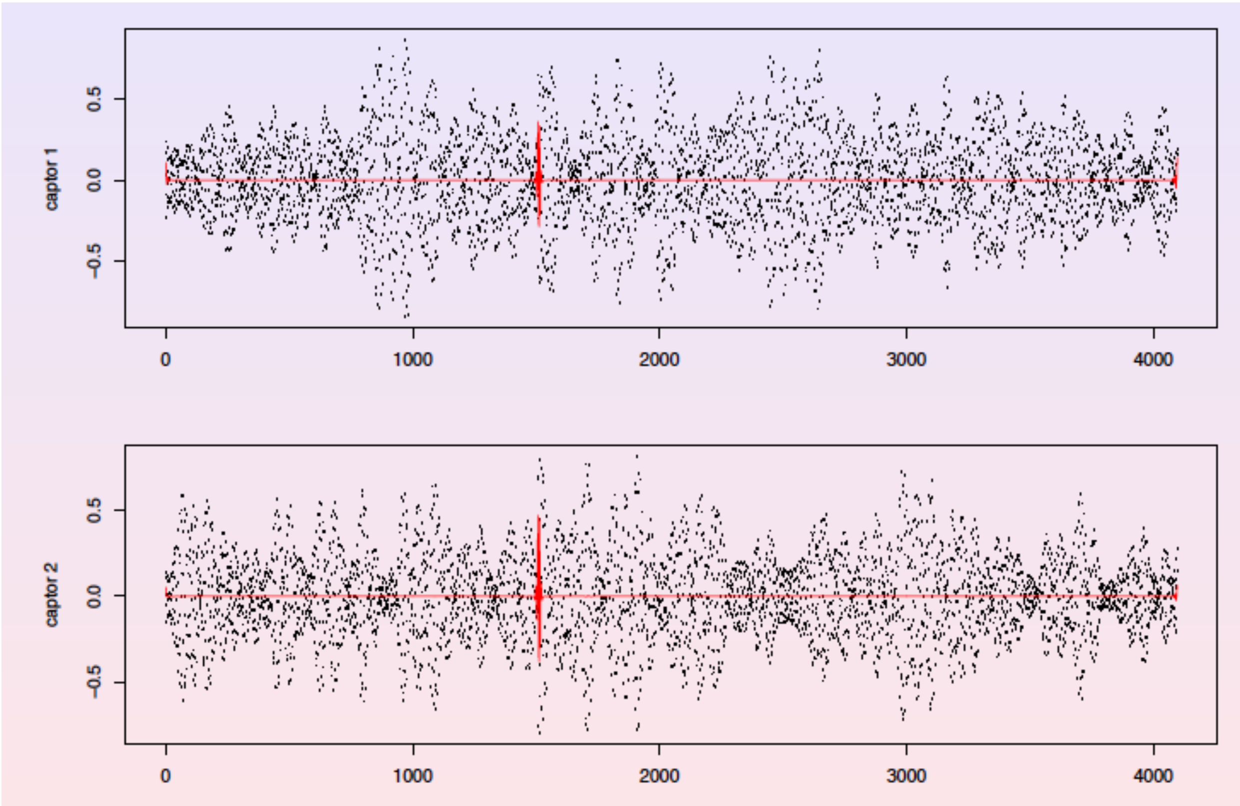
I.e. limited parametric form was important source of the “$\Delta\chi^2$=?” problem?

Still need to address other systematics (e.g. theory erorrs), data compatibility, etc.

Applications: gravitational waves

C. Roever,
S. Sardy

Wavelet transforms to search for blips in time series:



Regularization to suppress noisy wavelet coefficients,
→ bias-variance trade-off; mathematics similar to unfolding.

Applications: astrophysics

O. Lahav

Wide variety of problems – mostly Bayesian approach.

Parameters can be: fundamental (e.g., curvature of universe);
"geographical" (eccentricity of an orbit).

Wide use of Bayes factors for model selection.

$$B_{01} \equiv \frac{p(d|\mathcal{M}_0)}{p(d|\mathcal{M}_1)} \qquad p(d|\mathcal{M}) \equiv \int_{\Omega_{\mathcal{M}}} p(d|\theta, \mathcal{M}) p(\theta|\mathcal{M}) \mathrm{d}\theta$$

e.g. B=1 weak; B=5 strong

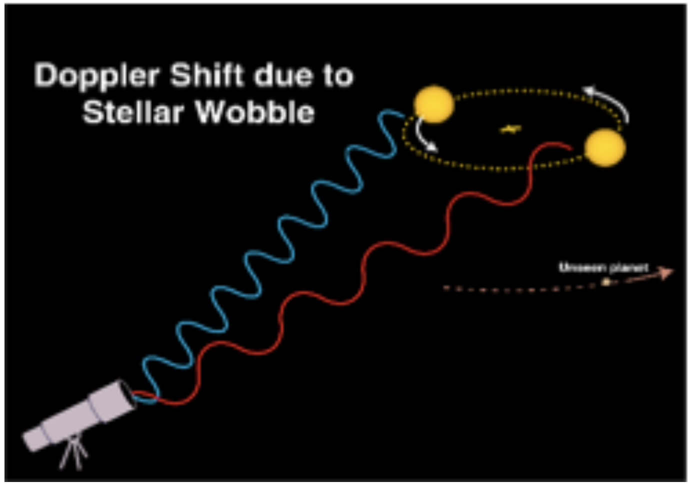
But how sensitive to assumed priors?

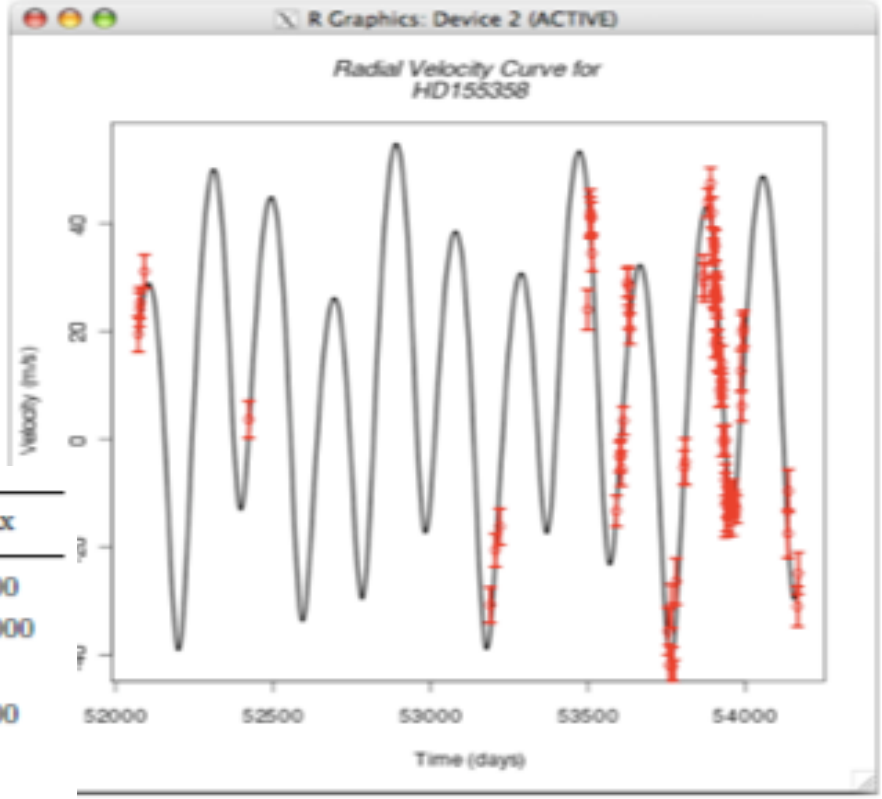
Variations: AIC, BIC, DIC,…

HEP should look at astro community's tools for computing these (e.g. MultiNest).

Bayesian exoplanet search

O. Lahav





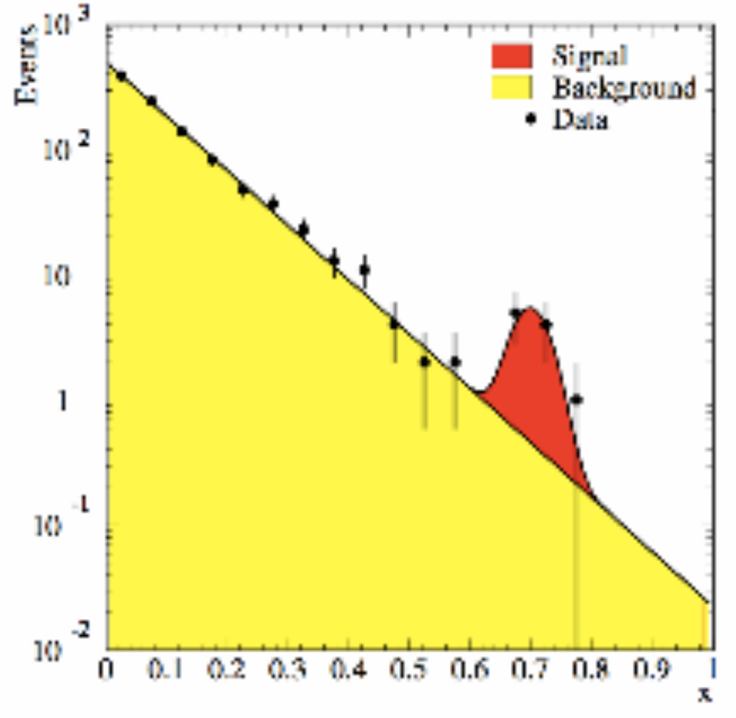
Para.	Prior	Mathematical Form	Min	Max
$V(ms^{-1})$	Uniform	$\frac{1}{V_{max}-V_{min}}$	-2000	2000
$T_1(days)$	Jeffreys	$\frac{1}{T_1 \ln\left(\frac{T_{1\,max}}{T_{1\,min}}\right)}$	0.2	15000
$K_1(ms^{-1})$	Mod. Jeffreys	$\frac{(K_1+K_{1\,0})^{-1}}{\ln\left(\frac{K_{1\,0}+K_{1\,max}}{K_{1\,0}}\right)}$	0.0	2000
e_1	Uniform	1	0	1
ϖ_1	Uniform	$\frac{1}{2\pi}$	0	2π
χ_1	Uniform	1	0	1
$T_2(days)$	Jeffreys	$\frac{1}{T_2 \ln\left(\frac{T_{2\,max}}{T_{2\,min}}\right)}$	0.2	15000
$K_2(ms^{-1})$	Mod. Jeffreys	$\frac{(K_2+K_{2\,0})^{-1}}{\ln\left(\frac{K_{2\,0}+K_{2\,max}}{K_{2\,0}}\right)}$	0.0	2000
e_2	Uniform	1	0	1
ϖ_2	Uniform	$\frac{1}{2\pi}$	0	2π
χ_2	Uniform	1	0	1
$s(ms^{-1})$	Mod. Jeffreys	$\frac{(s+s_0)^{-1}}{\ln\left(\frac{s_0+s_{max}}{s_0}\right)}$	0	2000

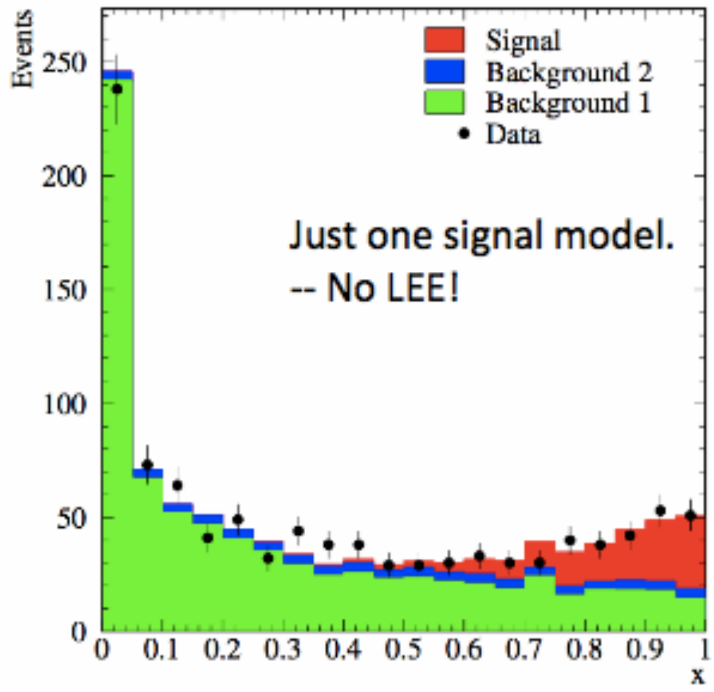
← priors

T. Junk

The Banff Challenge 2a

Two problems:





The winners:

Mark Allen
Stefan Schmitt
Wolfgang Rolke
Eilam Gross and Ofer Vitells
Stanford Challenge Team

Stefan Schmitt
Eilam Gross & Ofer Vitells

D. van Dyk

Some thoughts on 5σ (van Dyk)

Using 5σ is really not the answer:

- We don't know the actual effect of Systematics and LEE.
- "No distribution is valid to the 5σ tail!"
- Sampling distributions are only asymptotic approximations.
- Must calculate extreme-tail probabilities.

We have ***no*** *idea what the actual level is.*

5σ simply sweeps the problem under the rug.

D. van Dyk

Some more thoughts

- Focus on model diagnostics and model improvement.
- View prior distributions as a way to illuminate assumptions, not as a source of assumptions.
- Focus on ultimate scientific goals, not superficial properties of procedures.

Need to develop more ways to insert nuisance parameters into models in clever, physically motivated, controlled ways.

And figure out how to constrain these parameters with control measurements; assign meaningful priors to them.

Outlook

Great progress in methods/software/sophistication since the first "Confidence Limits" workshop at CERN in 2000.

And many areas where progress still being made:

Spurious exclusion when no sensitivity (CLs, PCL, other?)
Bayesian priors (reference and otherwise)
Look-elsewhere effect (solved?)
Software tools (RooStats, BAT)
Ways to improve models
...
...

The LHC data floodgates are open – we need to continue to improve and develop our analysis methods, taking full advantage of the lessons from the statistics community and other fields.

Thanks

Thanks to all the speakers for contributing highly interesting talks.

Thanks to the non-particle physicists and especially the professional statisticians for sharing their insights and expertise.

Thanks to the organisers, Louis, Albert, Michelangelo, for arranging such a stimulating and productive meeting.