GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Monte Carlo methods in experimental HEP

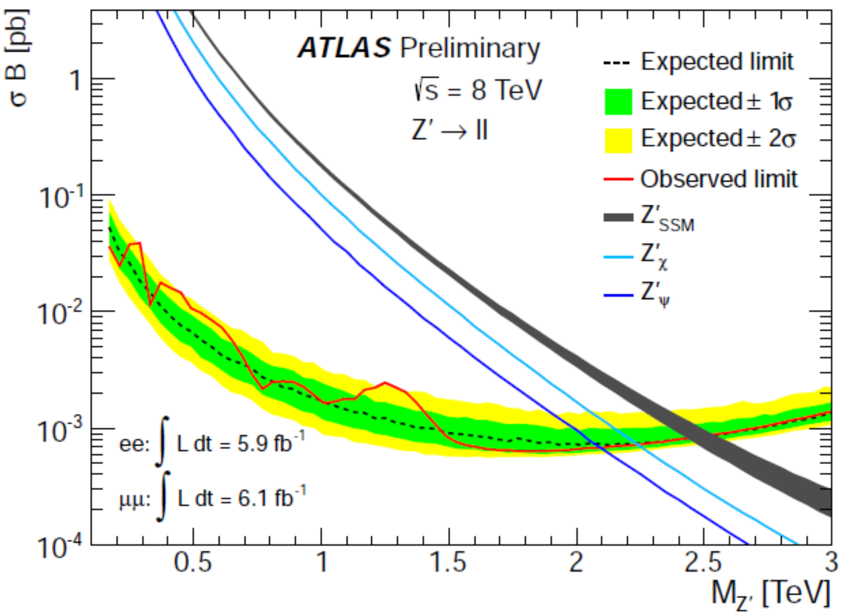
Kevin Kröninger – University of Göttingen

MCNet School, Maria Spring, August 6th 2013

Overview • Statistics • Limits, sensitivities and all that • Outlook

Outline

- Lecture 1: Crash course in (Bayesian) statistics and Markov Chain Monte Carlo
- Lecture 2: Limits, sensitivities and all that
- Summary and outlook



Outline: Statistics

- A concrete example
- Scientific reasoning
- Probability and the Bayesian interpretation
- Parameter estimation
- Markov Chain Monte Carlo
- Summary

Computational steps

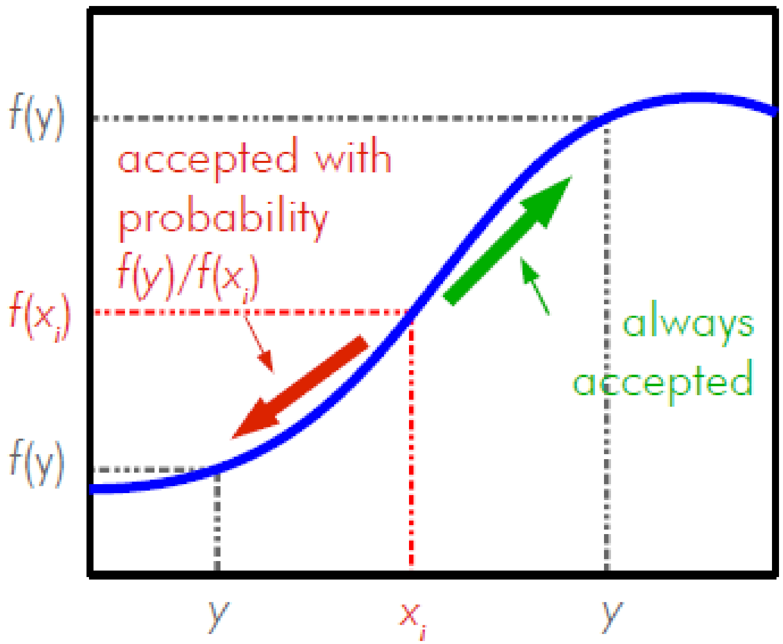
Numerical issues

- Point estimate:
  - Maximization of posterior
  - Typical tool: Minuit
  - Also: Simulated annealing
- Calculation of marginal distributions:
  - Analytical solutions usually difficult
  - Numerical integration methods, e.g. VEGAS
  - Sampling methods:
    - Hit&miss, simple Monte Carlo, ...
    - Importance sampling
    - Markov Chain Monte Carlo (MCMC)

      → Revolution in Bayesian computation

How does MCMC work?

- Output of Bayesian analyses are posterior probability densities, i.e., functions of an arbitrary number of parameters (dimensions).
- Sampling large dimensional functions is difficult.
- Idea: use random walk heading towards region of larger values (probabilities)



→ Metropolis algorithm

N. Metropolis *et al.*,
J. Chem. Phys. 21 (1953) 1087.

Start at some randomly chosen x_i
Randomly generate y around x_i
If f(y) > f(x_i) set $x_{i+1} = y$
If f(y) < f(x_i) set $x_{i+1} = y$ with prob. p=f(y)/f(x$_i$)
If y is not accepted set $x_{i+1} = x_i$
Start over

MCMC for Bayesian inference

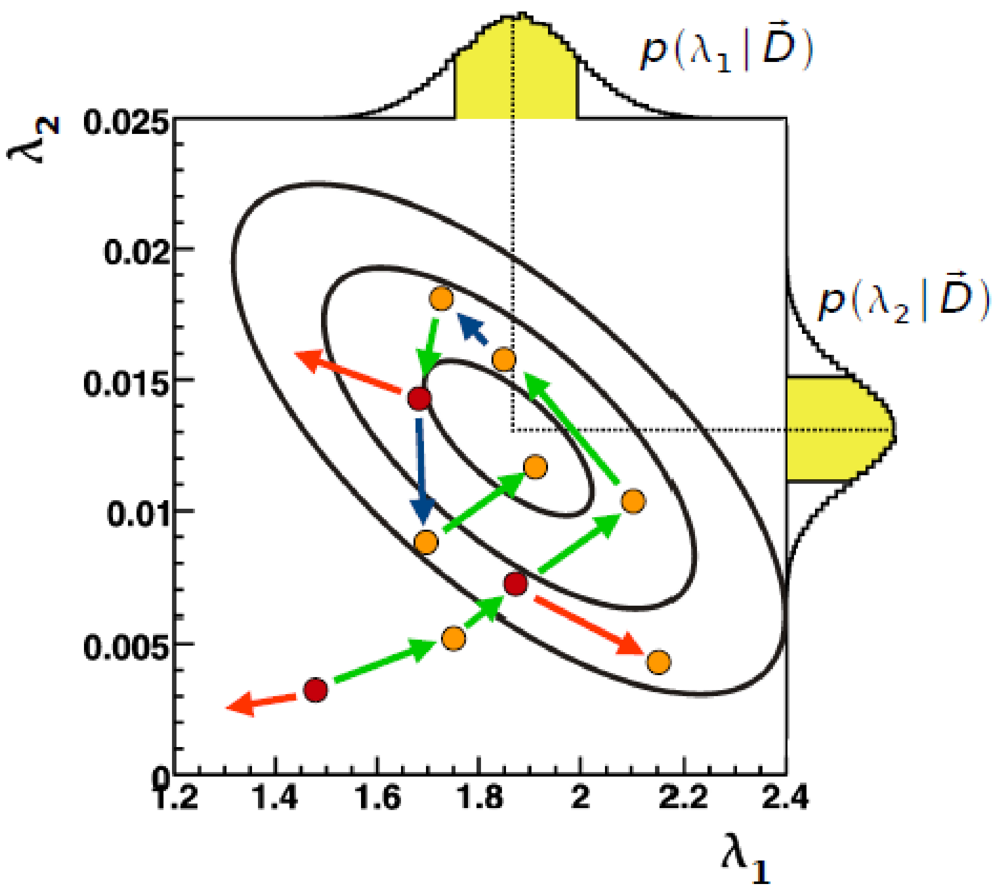
- Use MCMC to sample the posterior probability, i.e.

$$f(\vec{\lambda}) = p(\vec{D} \mid \vec{\lambda})\, p_0(\vec{\lambda})$$

- Marginalization of posterior:

$$p(\lambda_i \mid \vec{D}) = \int p(\vec{D} \mid \vec{\lambda})\, p_0(\vec{\lambda})\, d\vec{\lambda}_{j \neq i}$$

- Fill a histogram with just one coordinate while sampling
- Error propagation: calculate any function of the parameters while sampling
- Point estimate: find mode while sampling

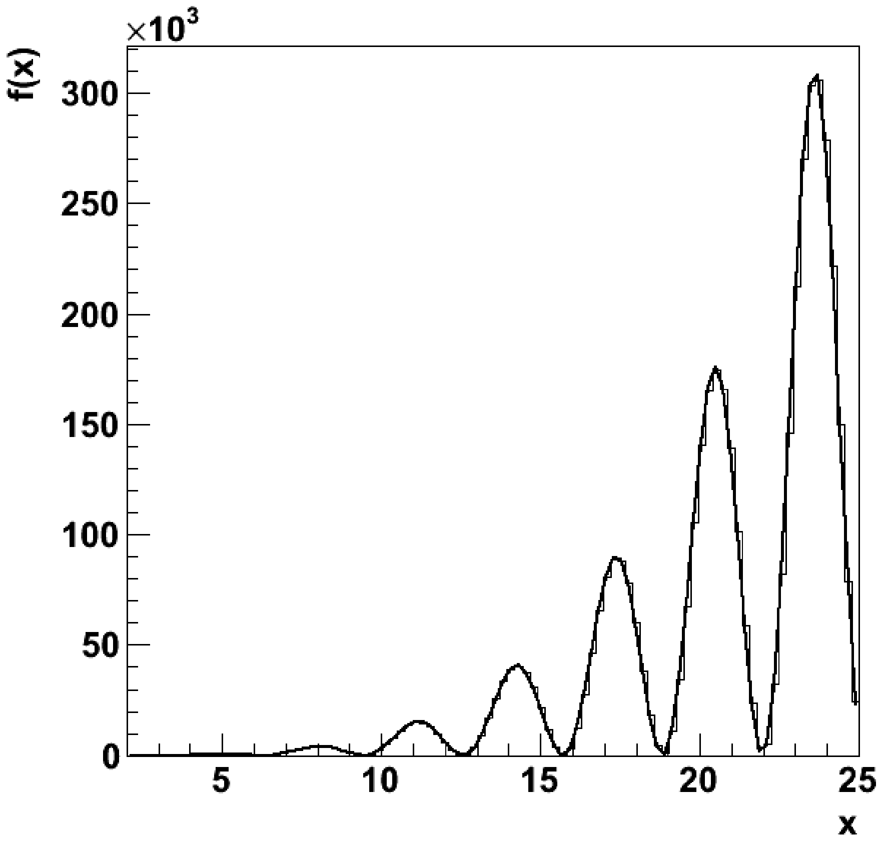


Does it work?

- Test MCMC on a function:

$$f(x)=x^4\sin^2(x)$$

- Compare MCMC distribution to analytic function
- Several minima/maxima are no problem.
- Different orders of magnitude are no problem.
- But: need to make sure that these chains converge towards the true distribution

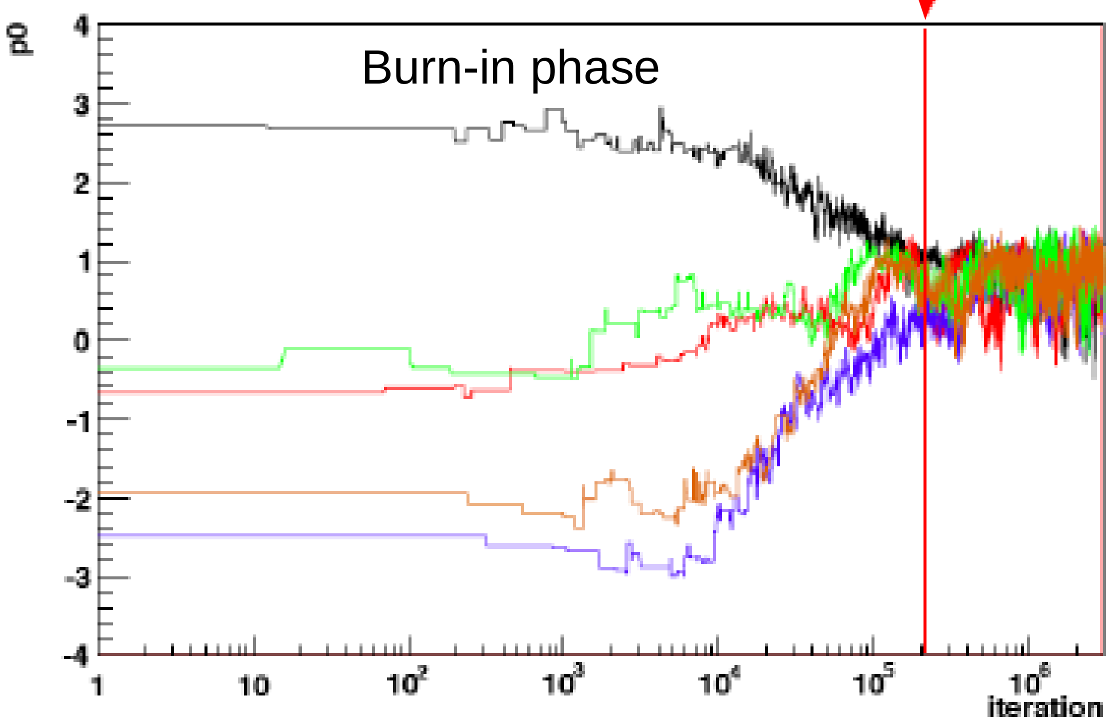


Markov Chain Monte Carlo

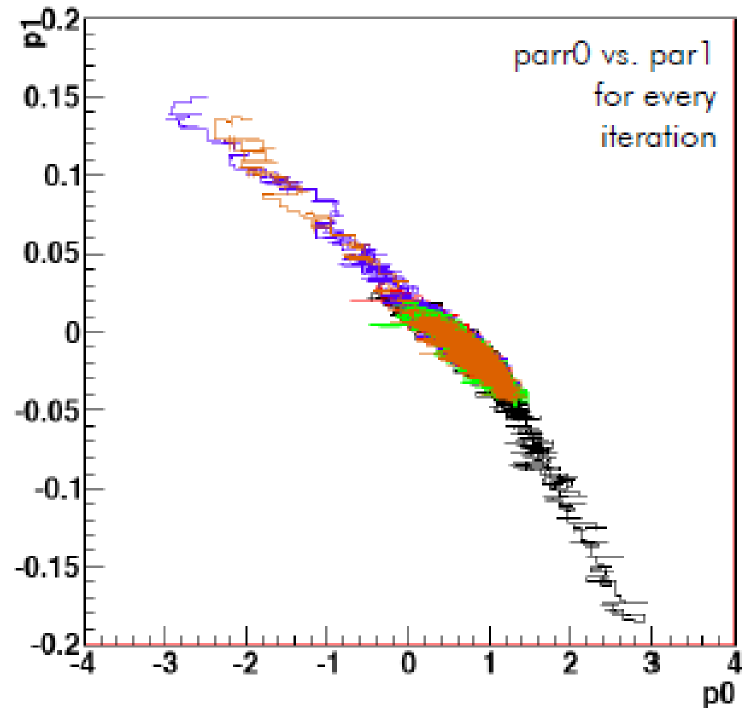
Convergence

- This is where it get’s difficult...
- Add a burn-in phase
- Use multiple chains

Convergence a la Gelman & Rubin



Parameter 0 value vs. iteration



Parameter 1 vs parameter 0

Bayesian Analysis Toolkit

- Tool for Bayesian inference written in C++
- Based on the ROOT-core functionality, interface to RooStats
- Uses MCMC for the calculation of the posterior probability
- Full control over convergence, automatic adjustment of step size
- Further algorithms: interface to CUBA, Minuit; importance sampling, simulated annealing, ...
- Pre-defined models: histogram fitter, template fitter, tool for combination of measurements, ...
- Web page: http://www.mppmu.mpg.de/bat/
- Contact: bat@mppmu.mpg.de
- Paper on BAT:

  A. Caldwell, D. Kollar, K. Kröninger, BAT - The Bayesian Analysis Toolkit
  Comp. Phys. Comm. 180 (2009) 2197-2209 [arXiv:0808.2552].

BAYESIAN ANALYSIS TOOLKIT

Outline: Limits, ...

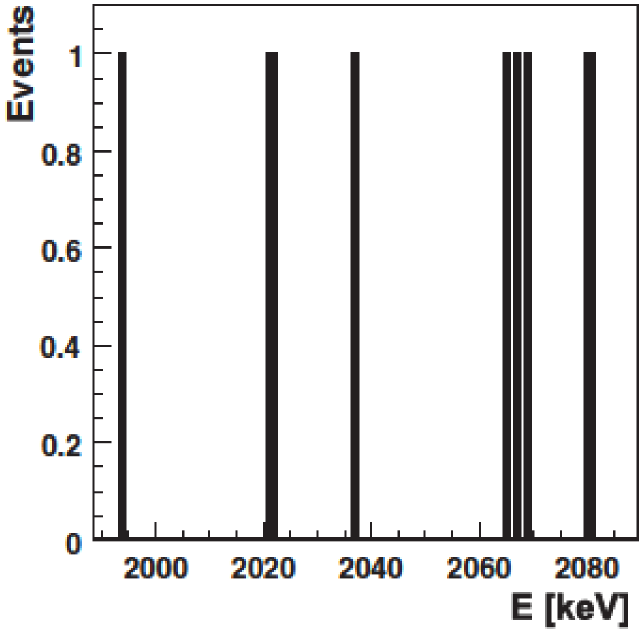
- Expected sensitivities
- Observed limits
- Goodness-of-fit and *p*-values
- Summary

Expected sensitivities

- Plan to conduct an experiment:
  - How do you get funding?
  - Explain the physics case and estimate the expected sensitivity to a process searched for
  - What does `sensitivity' mean anyway?
- Example: ILC TDR
- Three step procedure:
  - Statistical formulation: define a procedure to estimate a limit (searches) or an uncertainty (measurements)
  - Ensemble generation: generate pseudodata according to your expectation (luminosity, run-time, POT, ...). Use same settings as real data will have, but multiple statistics
  - Ensemble tests: Repeat analysis on pseudodata and quote expectation values.

Expected sensitivities

- Consider same example as yesterday:
  - Neutrinoless double β-decay: sharp Gaussian on top of a flat background
  - Expect few numbers of events
  - Now: estimate the sensitivity of the GERDA experiment: A. Caldwell, KK, Phys. Rev. D 74 (2006) 092003
- Two questions:
  - What is the probability that the observed spectrum is due to background only?
  - What is the signal contribution? Measurement or limit.
- Same idea as with search/measurement of the top quark, Higgs boson, etc.



Statistical formulation

- Hypotheses:
  - H : the spectrum is due to background only
  - $\overline{H}$: the spectrum has contributions from signal and background
  - Assume both to be equally likely
- Evidence and discovery:
  - Test hypothesis *H* using Bayes theorem:

$$p(H|data) = \frac{p(data|H)\cdot p_0(H)}{p(data|H)\cdot p_0(H) + p(data|\overline{H})\cdot p_0(\overline{H})}$$

  - If p(H|data) < cut, claim *evidence* or *discovery*
  - These are just words, but choose them and their meaning carefully, e.g., cut = 0.01: evidence, cut = 0.0001: discovery

Expected sensitivities

Statistical formulation

- Statistical models:
  - Binned likelihood of independent Poisson fluctuations
  - One or two contributions with strengths *B* and *S*

$$p(data|H) = \int dB\, p(data|H, B) \cdot p_0(B)$$
$$p(data|\overline{H}) = \int dB \int dS\, p(data|H, B, S) \cdot p_0(B) \cdot p_0(S)$$

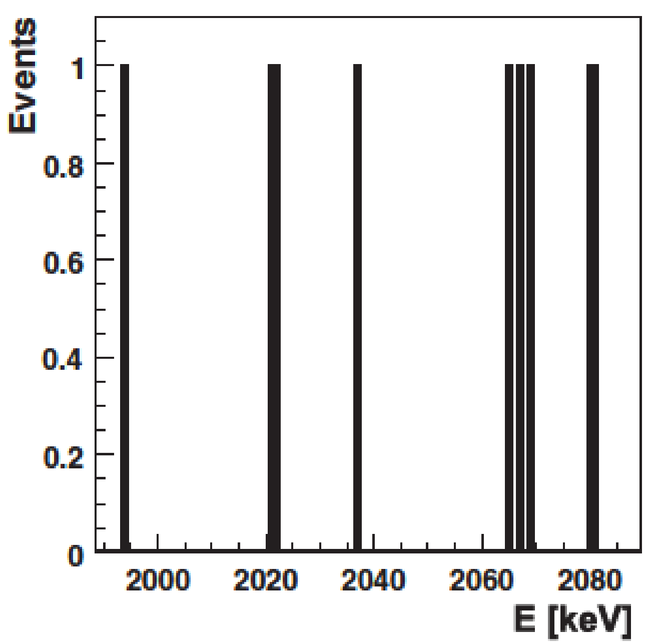
  - Assume uniform prior for S (no clue what the strength is)
  - Assume Gaussian prior for B (estimated in a sideband region)

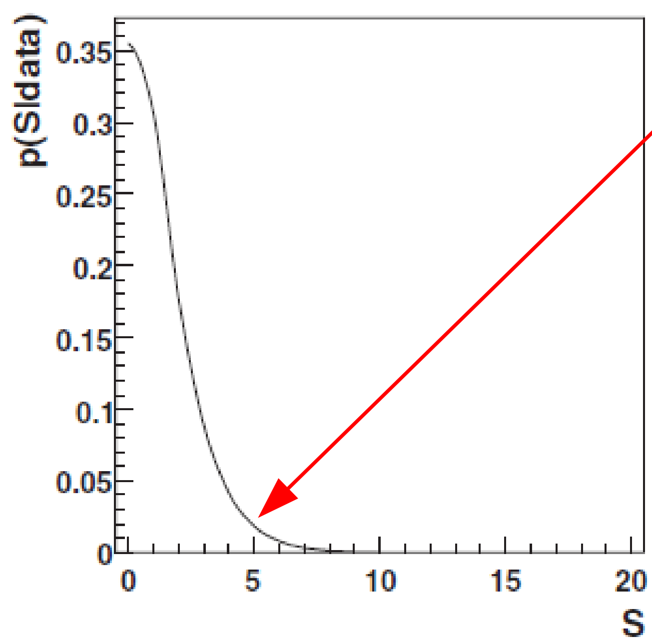
Statistical formulation

- Setting limits: calculate marginal distribution for signal

$$p(S|data,\overline{H}) = \int dB \frac{p(data|\overline{H},B,S)\cdot p_0(B)\cdot P_0(S)}{\int dB \int dS\, p(data|\overline{H},B,S)\cdot p_0(B)\cdot P_0(S)}$$

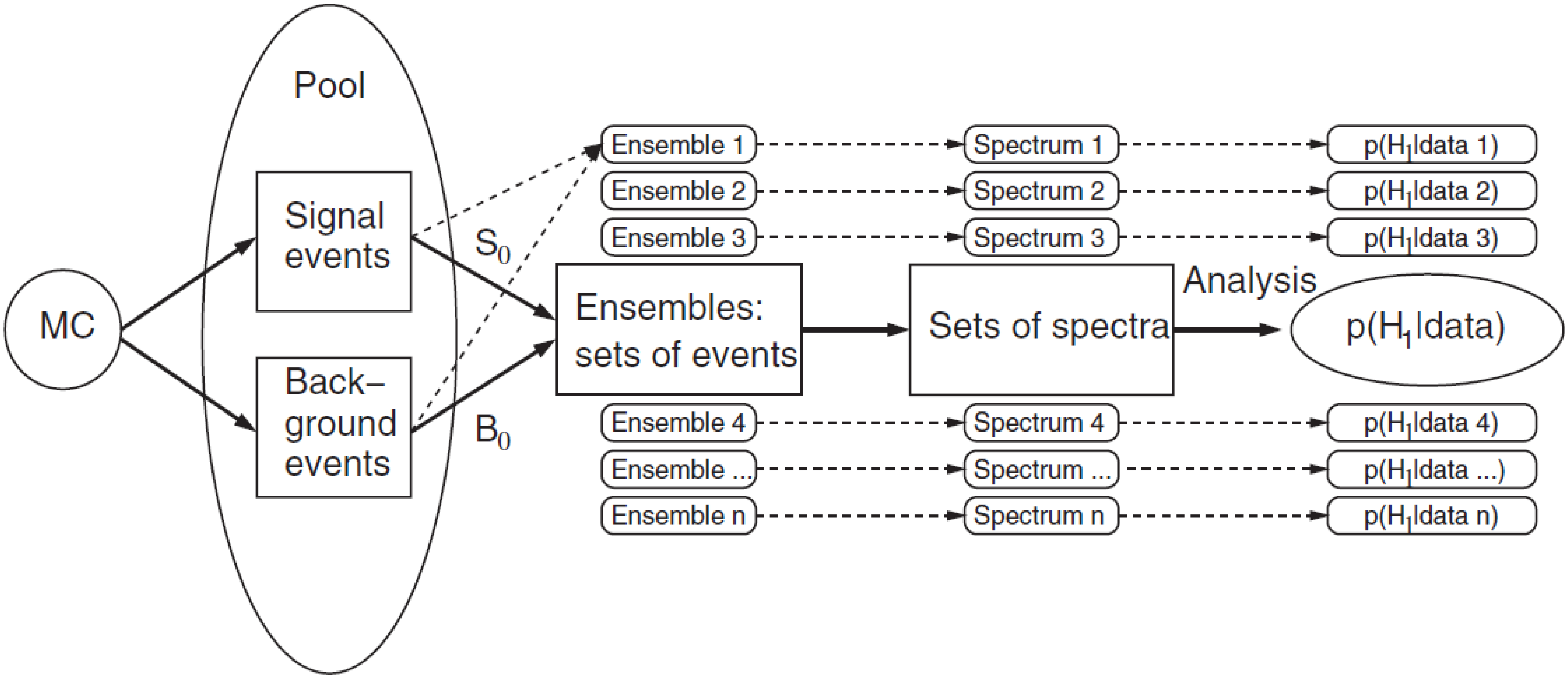
- Integrate until a certain probability is reached $\int_0^{S_{90}} dS\, p(S|data,\overline{H}) = 0.90$





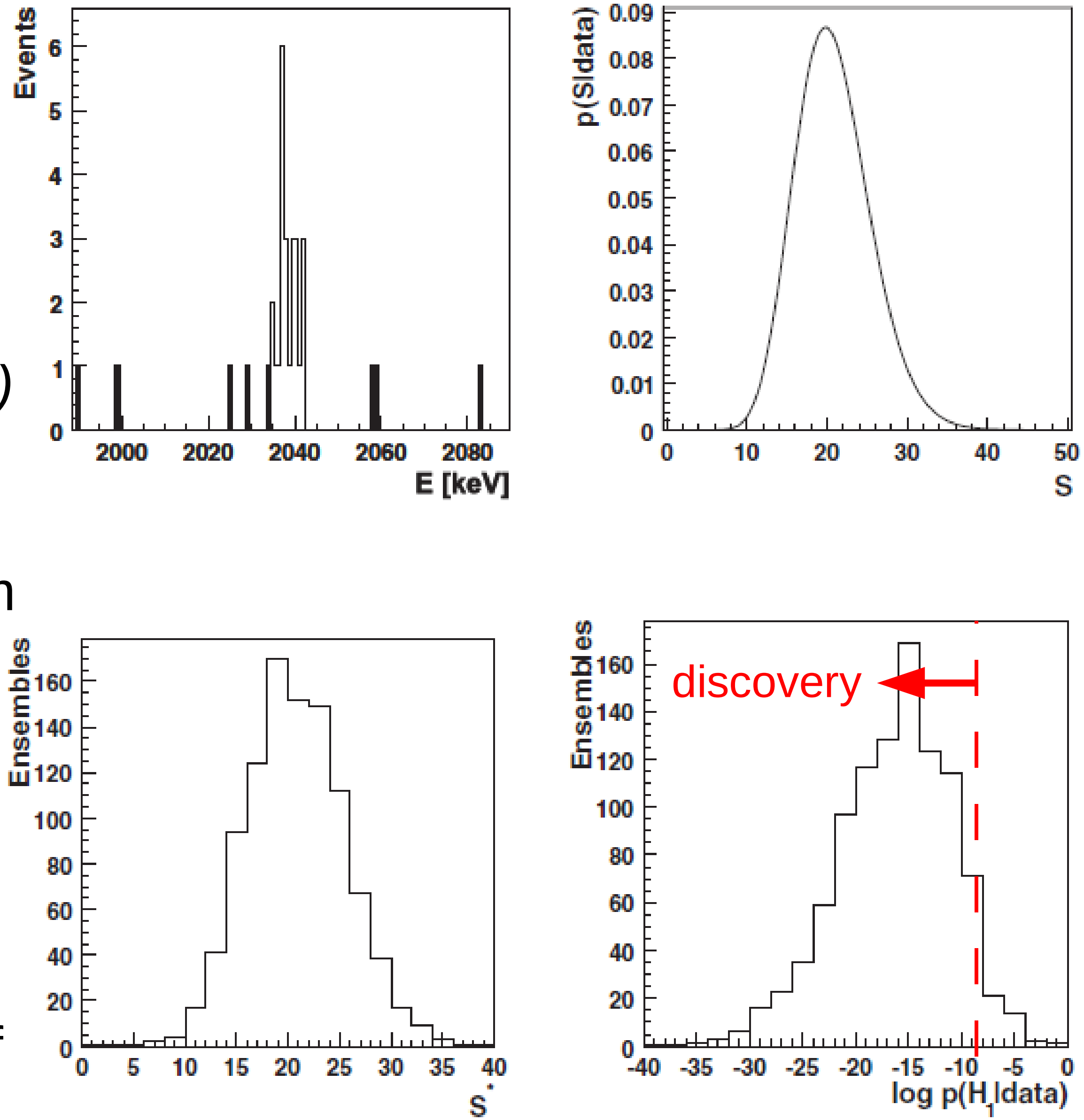
Ensemble generation

- Generate sets of spectra from Monte Carlo events



Ensemble tests

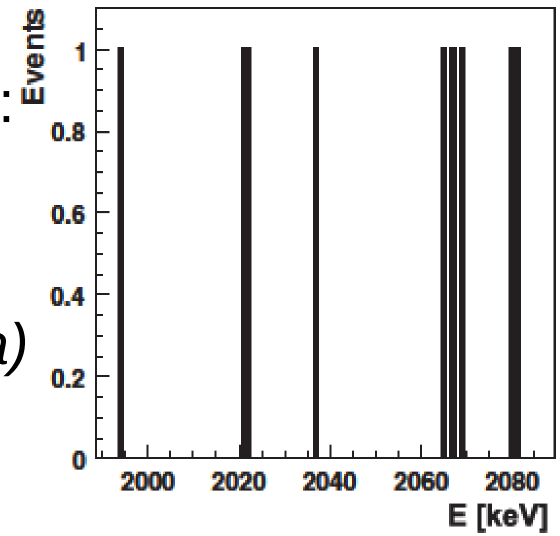
- Testing the background-only hypothesis:
  - Composition of samples: signal and background
  - Calculate probability *p(H|data)* for each spectrum
  - Calculate mode of signal contribution for each spectrum
- Example:
  - Input S = 20.4, B = 10
  - Average mode <S*>=20.3
  - More than 97% of ensembles have *p(H|data)*<0.0001, i.e. no claim of discovery in 3% of all cases

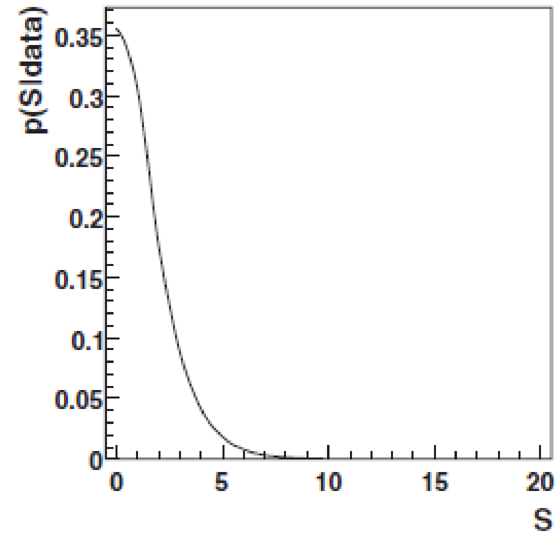


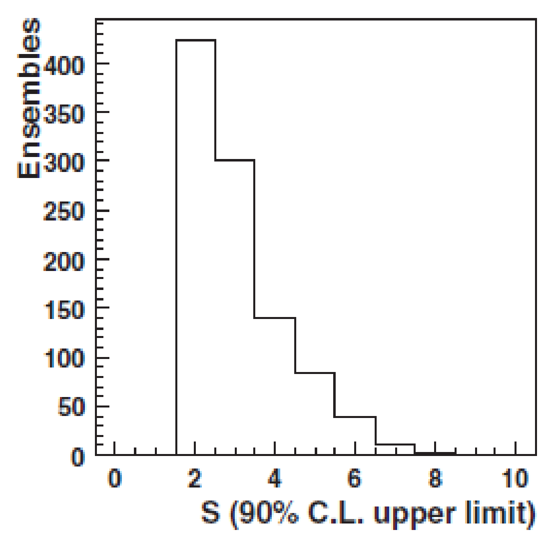
Expected sensitivities

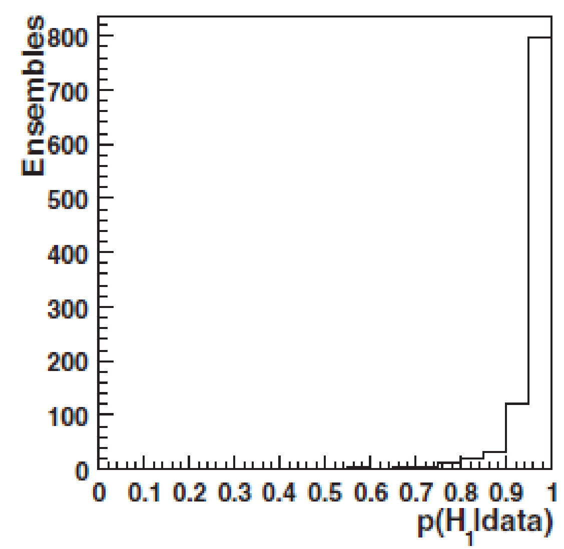
Ensemble tests

- Setting limits (on small signals):
    - Composition of samples: only background
    - Calculate probability *p(H|data)* for each spectrum
    - Calculate 90% prob. limit on signal for each spectrum
- Example:
    - Input S = 0, B = 10
    - Average limit $\langle S_{90} \rangle = 3.1$
    - No claim of discovery for any pseudodata set









Ensemble tests

- Physics interpretation:
  - Observed signal strength can be transformed into halflife of 0νββ:

$$S \approx \ln(2)\kappa\epsilon\frac{N_A}{M_A}\cdot\frac{M\cdot t}{T_{1/2}}$$

exposure

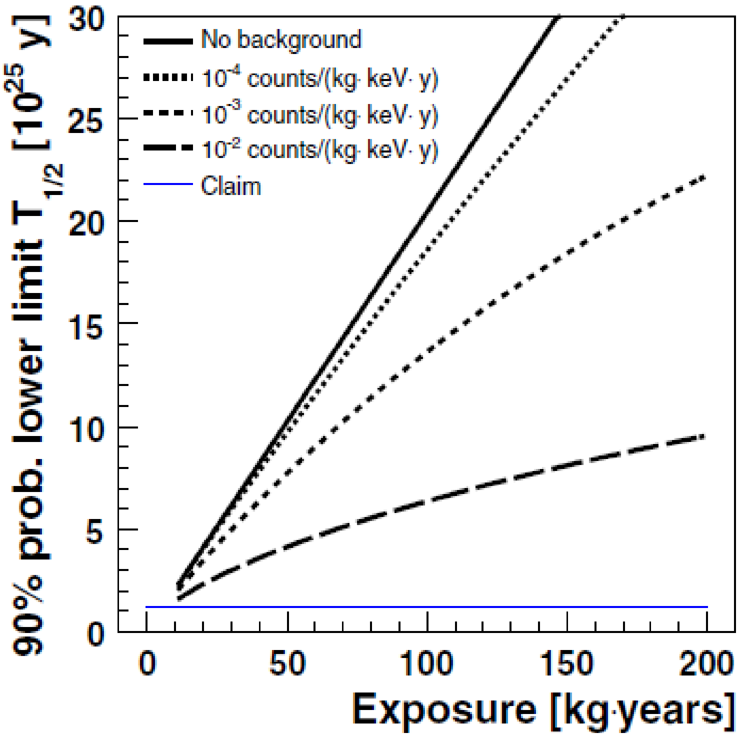
  - And into effective Majorana neutrino mass:

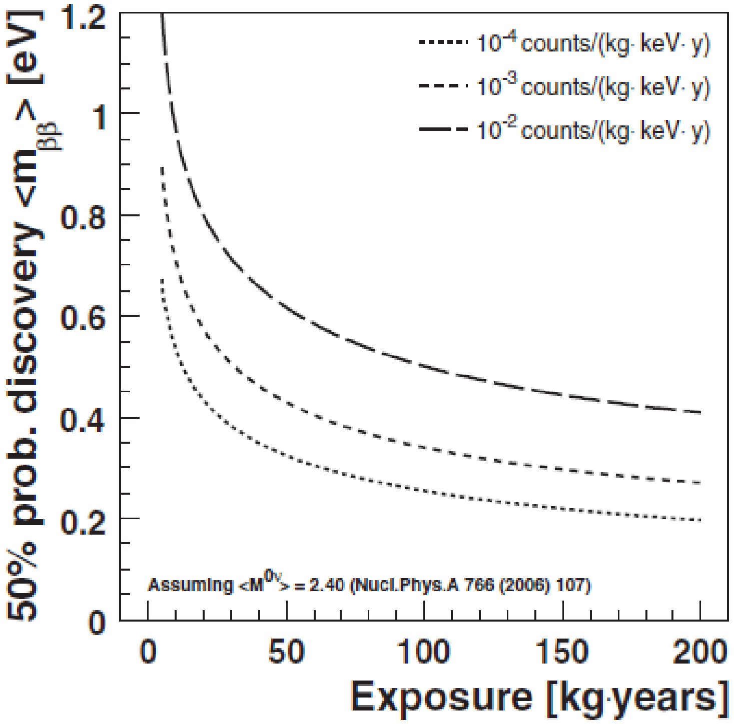
$$\langle m_{\beta\beta}\rangle = (T_{1/2}\cdot G_{0\nu})^{-1/2}\cdot\frac{1}{M_{0\nu}}$$

nuclear matrix element

Sensitivity

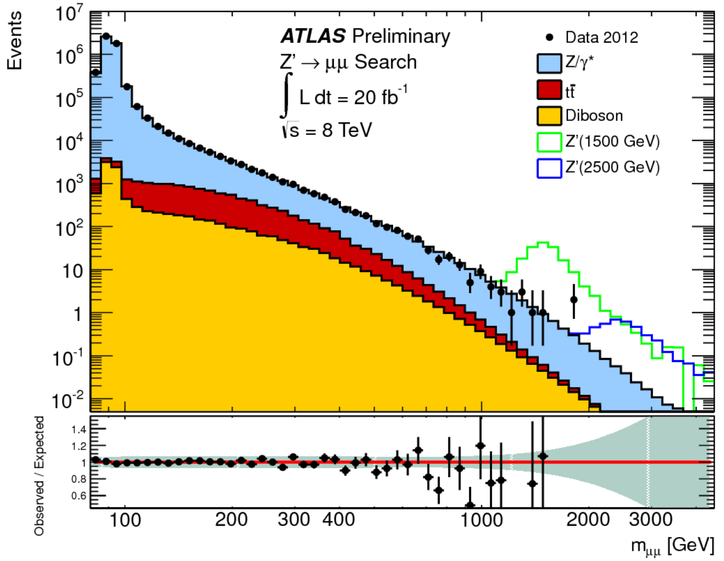
- For a given exposure, define sensitivities as
  - expected 90% prob. lower limit on halflife of 0νββ, or expected 90% prob. upper limit on neutrino mass;
  - half-life (or neutrino mass) for which 50% of all pseudodata sets claim a discovery

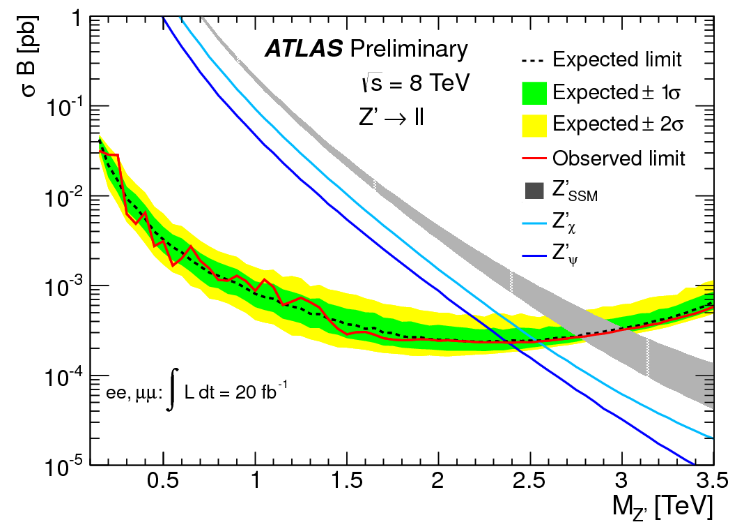




Bringing expectation and observation together

- Now add real data and set actual limits: Search for Z' boson, ATLAS-CONF-2013-017
- Repeat analysis on real data, treat pseudodata and data the same
- Note:
  - Expected upper limit calculation uses ensembles without signal contribution, but
  - fit a Z' hypothesis (template) with a certain Z' mass to the spectrum (plus background)
  - Histogram the obtained limits and quote the median value as well as the central 68% and 95% envelope
  - Exclude a Z' mass if limit is smaller than predicted cross-section (for a concrete model)





Goodness-of-fit

- Situation:
  - Measurement is done
  - Chose one model over the other
  - Fitted all parameters
- To be done: judge if the data are described by the chosen model
- How to judge the goodness-of-fit? One way: use *p*-values
- Again, a three-step procedure:
  - Define a discrepancy variable (test statistic) t such that large values suggest a discrepancy between data and model.
  - Generate ensembles and calculate the test statistic for each ensemble
  - Compare the distribution of t, $p(t)$, with the observed value of t, t_{obs} by calculating the tail-area probability to have found a larger value of t_{obs}.

p-values

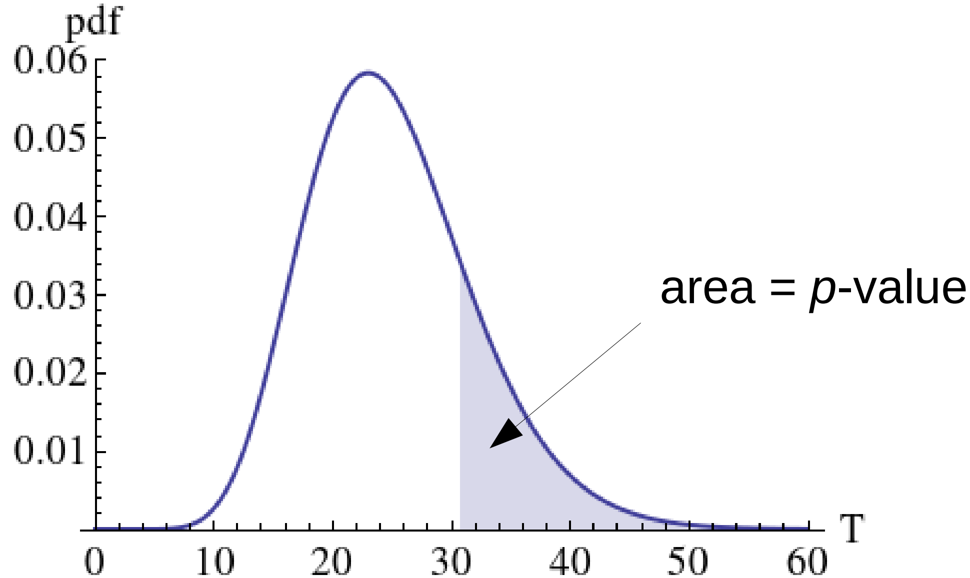
- Tail-area probability:

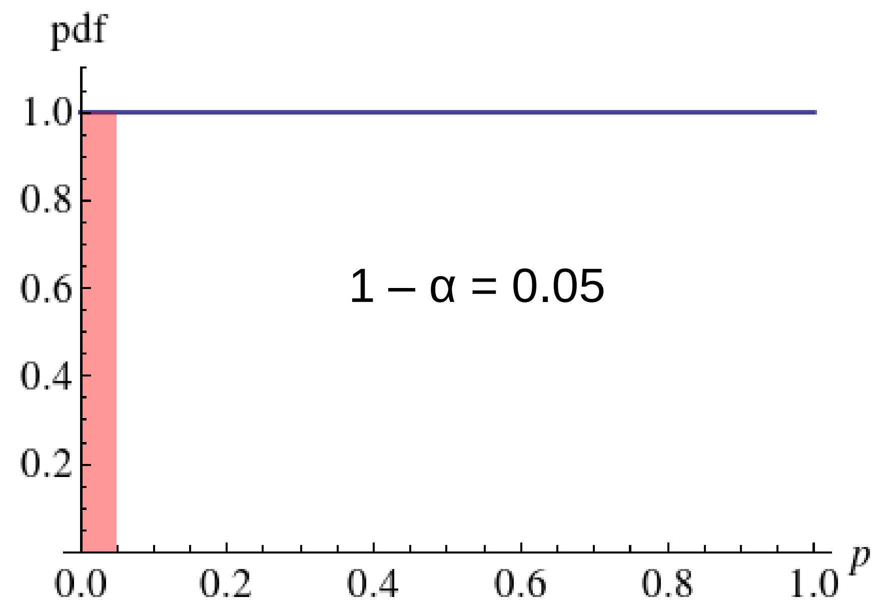
$$p(t > t_{obs} | model) = \int dt\, p(t | model)$$

- Small *p*-values: disagreement between data and model
- Repeated calculation of *p*-value results in a uniform distribution
- Pick a confidence level α and reject the model if

$$p(t > t_{obs} | model) < 1 - \alpha$$

- For large α, this means a lot of ensembles (thus Monte Carlo events)





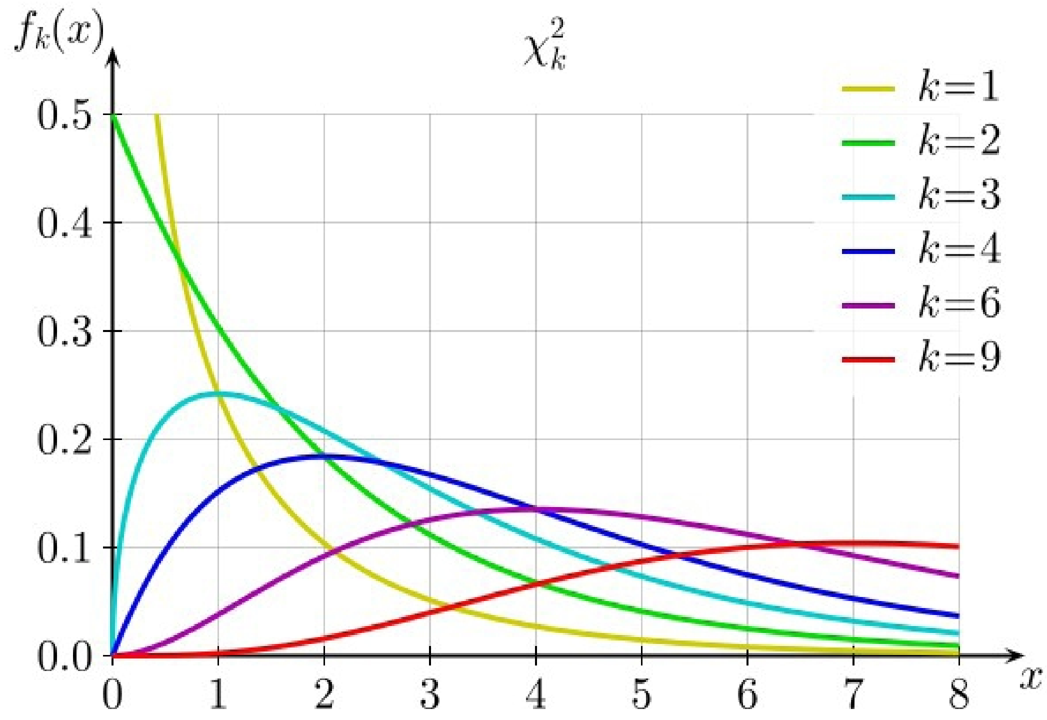
Goodness-of-fit and *p*-values

χ^2 variables

- Examples:
  - Gaussian χ^2:

$$\chi^2 = \sum_i \frac{(y_i - f(x_i))^2}{\sigma_i^2}$$

  - Assume data are Gaussian distributed around expectation
  - Ask for χ^2/dof ~ 1 because $\langle\chi^2\rangle$ = dof
  - Spread also important, thus quote χ^2-probability



- Neyman χ^2:

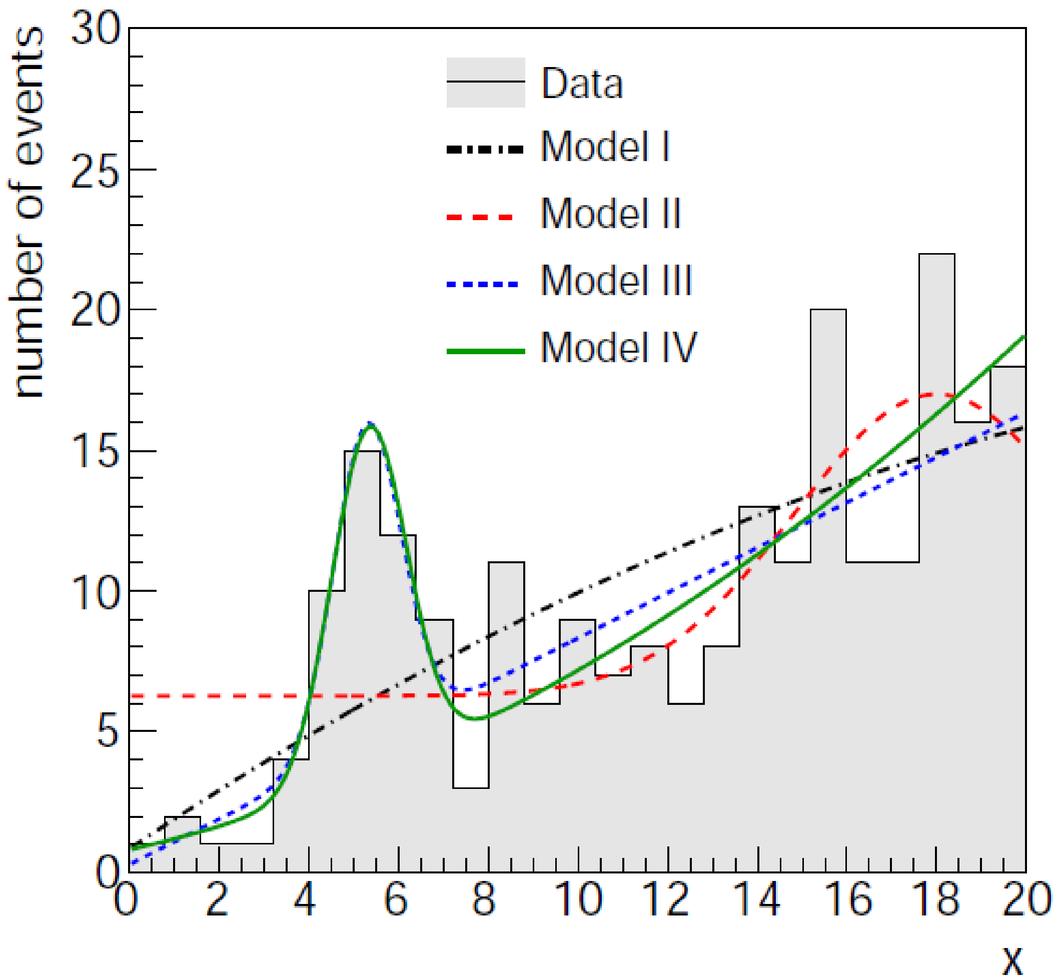
$$\chi_N^2 = \sum_i \frac{(n_i - \nu_i)^2}{n_i}$$

Pearson χ^2:

$$\chi_N^2 = \sum_i \frac{(n_i - \nu_i)^2}{\nu_i}$$

An example

F. Beaujean *et al.*, Phys. Rev. D 83 (2011) 012004



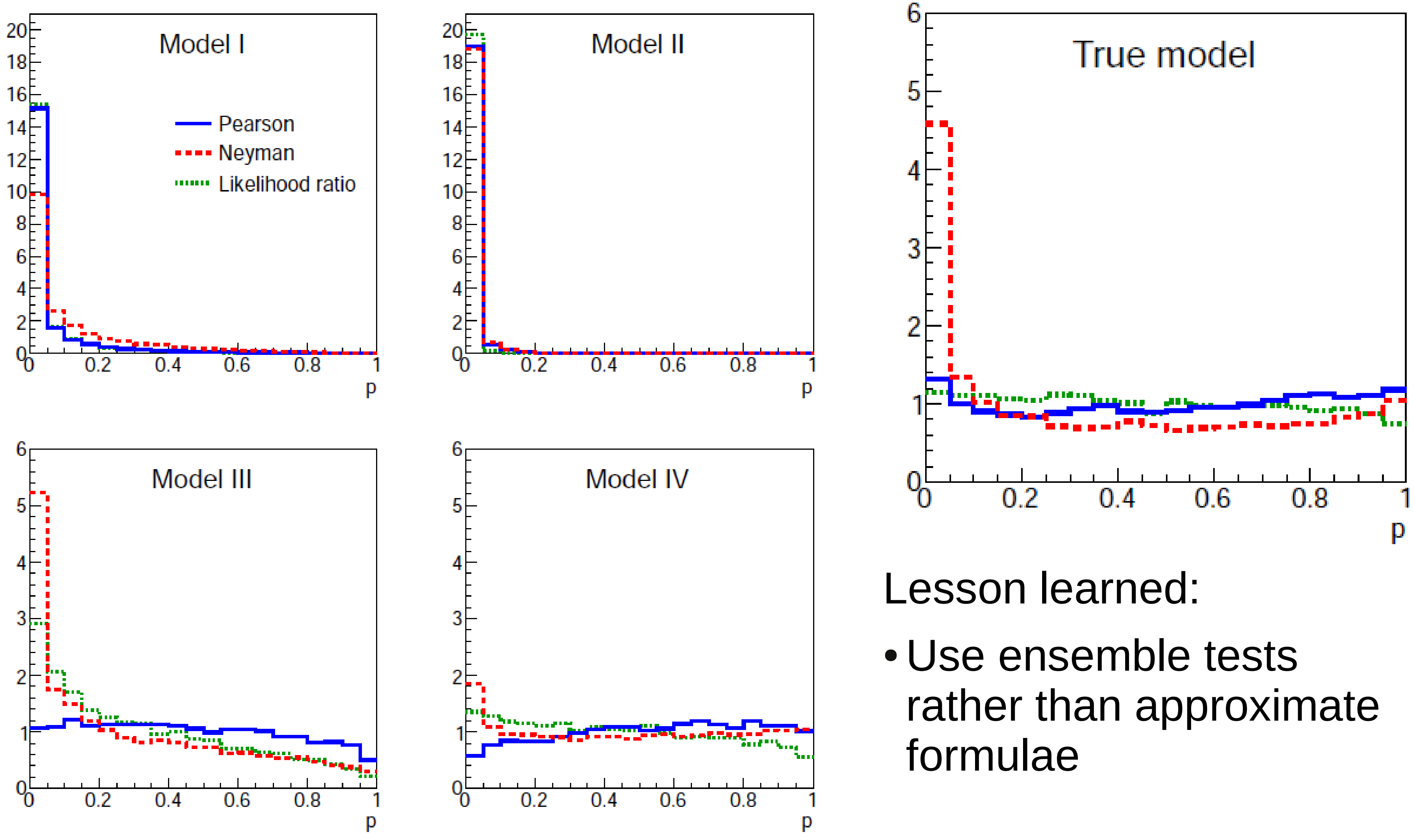
$$f(x_i) = A + B\,x_i + C\,x_i^2 + \frac{D}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} ,$$

with parameter values ($A = 0$, $B = 0.5$, $C = 0.02$, $D = 15$, $\sigma = 0.5$, $\mu = 5.0$).

An example

	Model	Par	Small range Min	Small range Max	Large range Min	Large range Max
I.	$A_{\mathrm{I}} + B_{\mathrm{I}}\, x_i + C_{\mathrm{I}}\, x_i^2$	A_{I}	0	5	-50	200
		B_{I}	0	1.2	-50	200
		C_{I}	-0.1	0.1	-50	200
II.	$A_{\mathrm{II}} + \frac{D_{\mathrm{II}}}{\sigma_{\mathrm{II}}\sqrt{2\pi}} e^{-\frac{(x_i-\mu_{\mathrm{II}})^2}{2\sigma_{\mathrm{II}}^2}}$	A_{II}	0	10	-50	200
		B_{II}	0	200	-50	200
		μ_{II}	2	18	0	50
		σ_{II}	0.2	4	0	20
III.	$A_{\mathrm{III}} + B_{\mathrm{III}}\, x_i + \frac{D_{\mathrm{III}}}{\sigma_{\mathrm{III}}\sqrt{2\pi}} e^{-\frac{(x_i-\mu_{\mathrm{III}})^2}{2\sigma_{\mathrm{III}}^2}}$	A_{III}	0	10	-50	200
		B_{III}	0	2	-50	200
		D_{III}	0	200	0	200
		μ_{III}	2	18	0	50
		σ_{III}	0.2	4	0	20
IV.	$A_{\mathrm{IV}} + B_{\mathrm{IV}}\, x_i + C_{\mathrm{IV}}\, x_i^2 + \frac{D_{\mathrm{IV}}}{\sigma_{\mathrm{IV}}\sqrt{2\pi}} e^{-\frac{(x_i-\mu_{\mathrm{IV}})^2}{2\sigma_{\mathrm{IV}}^2}}$	A_{IV}	0	10	-50	200
		B_{IV}	0	2	-50	200
		C_{IV}	0	0.5	-50	200
		D_{IV}	0	200	0	200
		μ_{IV}	2	18	0	50
		σ_{IV}	0.2	4	0	20

An example

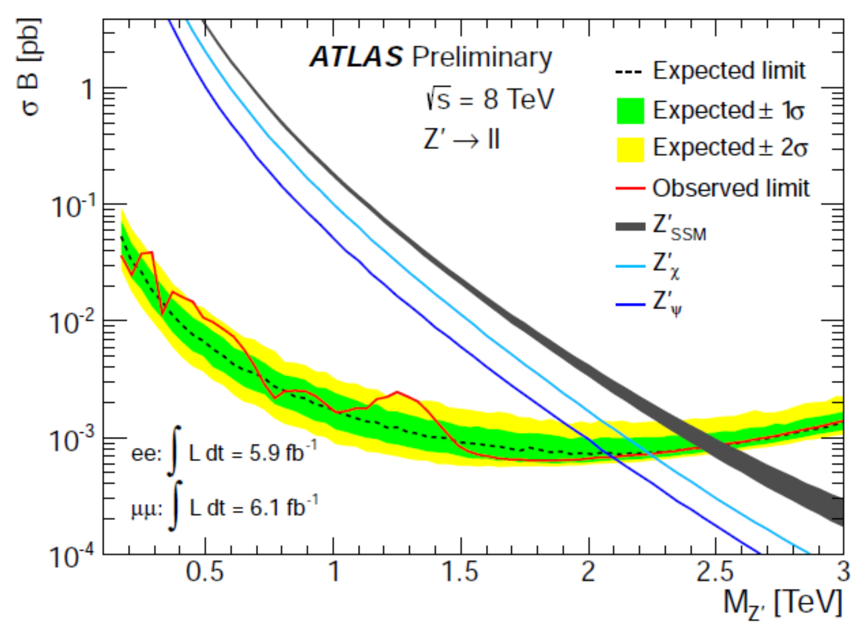


Lesson learned:

- Use ensemble tests rather than approximate formulae

Outline

- Lecture 1: Crash course in (Bayesian) statistics and Markov Chain Monte Carlo
- Lecture 2: Limits, sensitivities and all that
- Summary and outlook



Summary (Limits)

- Direct model comparison (useful for comparing different MC generators)
- Ensemble tests are a powerful tool for calculating
  - expected sensitivities (limits, observations, uncertainties, etc.),
  - properties of estimators and discrepancy variables,
  - measures of goodness-of-fit
- Often: prefer numerical solution over a bad approximation (*p*-values)

Summary

- Monte Carlo methods are crucial part in experimental HEP
- Context is often statistical inference
  - Parameter estimation, fitting large-dimensional problems
  - Ensemble tests and expected sensitivities
  - Goodness-of-fit tests and *p*-values