technische universität
dortmund

An introduction to Bayesian Reasoning and its applications

Kevin Kröninger – TU Dortmund

Terascale Statistics School 2018

DESY Hamburg

23.02.2018

Overview

- A concrete example
- Scientific reasoning
- Probability and the Bayesian interpretation
- Parameter estimation
- A concrete example (continued)
- Model comparison
- Numerical method: MCMC
- Further examples
- Summary

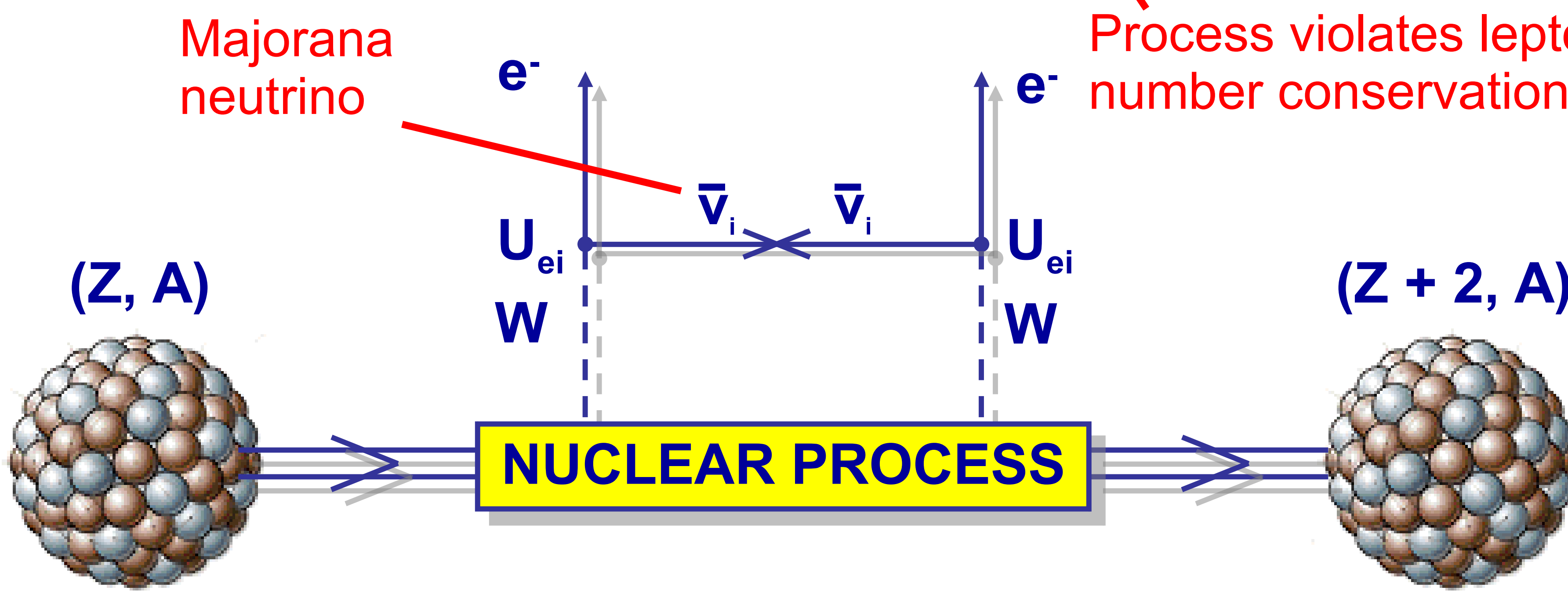
A concrete example

Neutrinoless double beta-decay

Rare nuclear transition (2nd order weak process):

$2\nu\beta\beta$:	$(Z, A) \rightarrow (Z+2, A) + 2\,e^- + 2\,\bar{\nu}$	$\Delta L = 0$	$(T_{1/2} \sim 10^{21}$ y)
$0\nu\beta\beta$:	$(Z, A) \rightarrow (Z+2, A) + 2\,e^-$	$\Delta L = 2$	$(T_{1/2} > 10^{25}$ y)

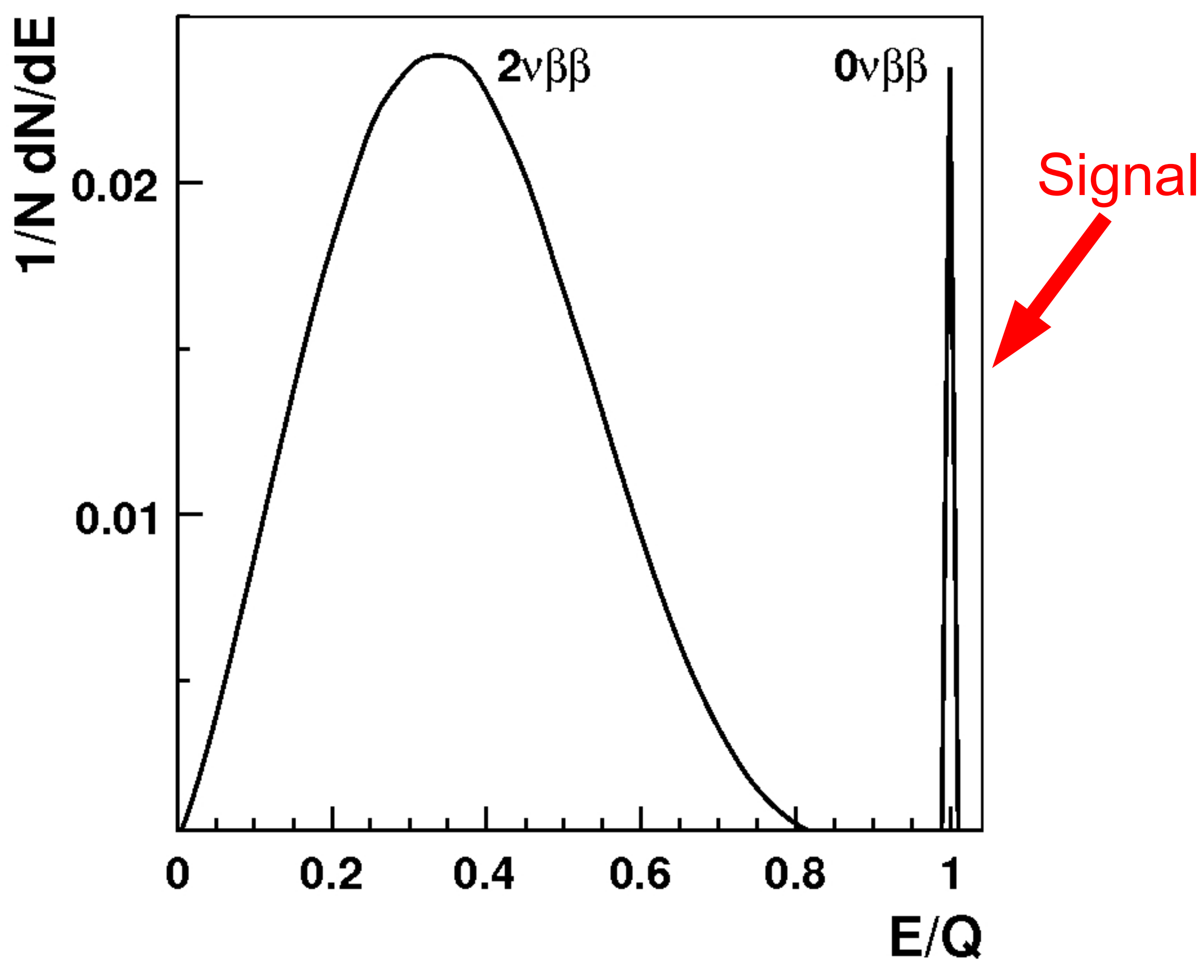
Process violates lepton number conservation



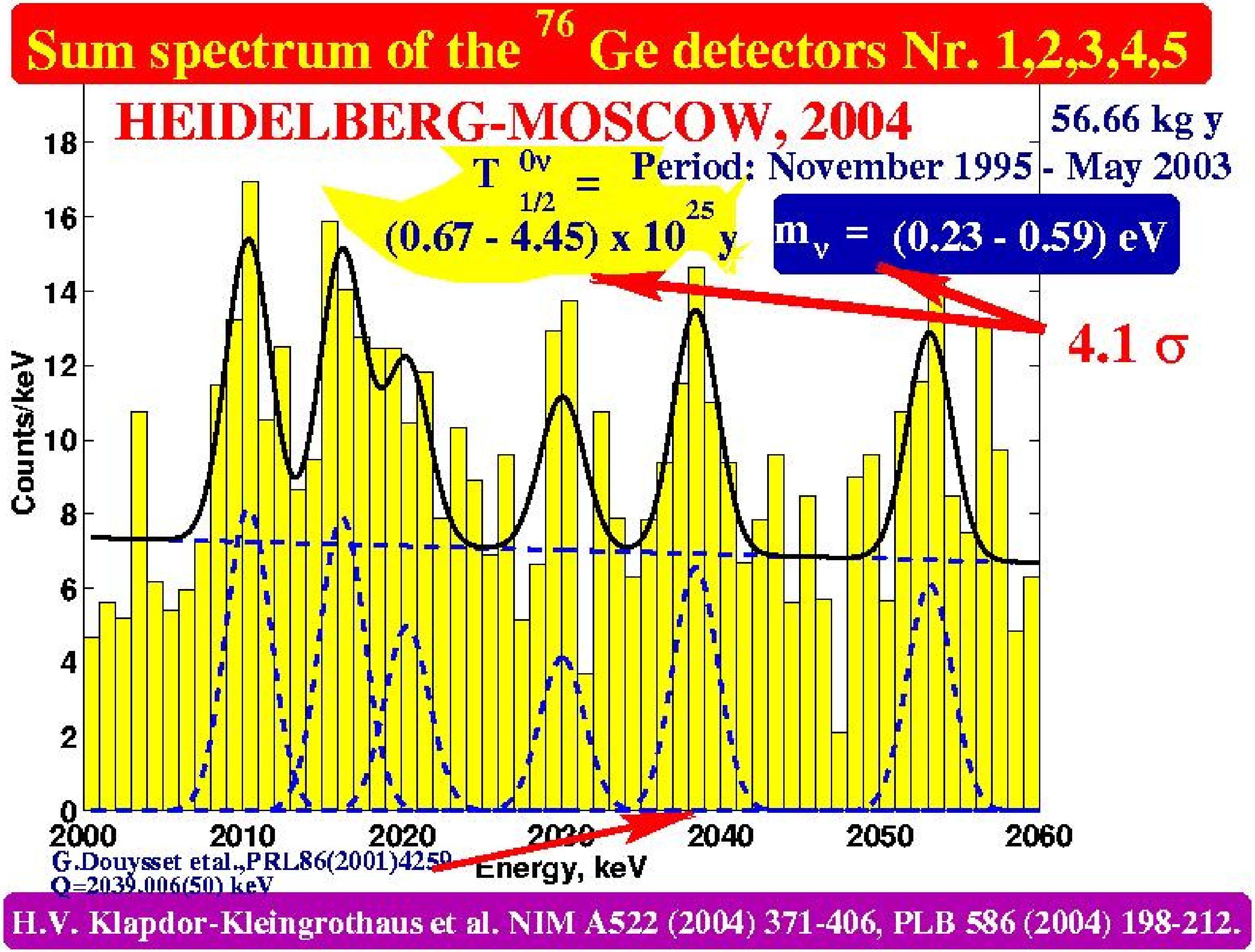
A concrete example

Neutrinoless double beta-decay

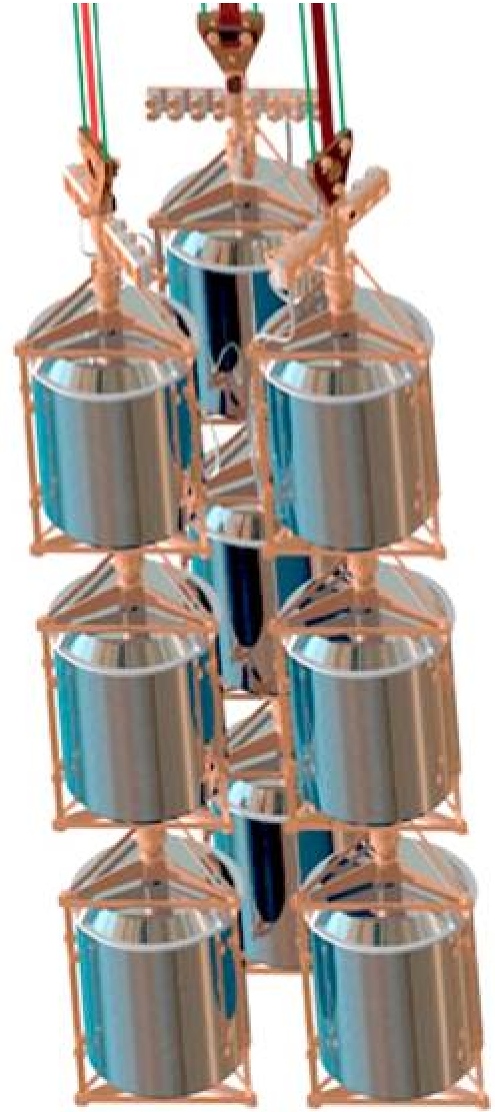
Searching for a single peak on top of a (flat) background...

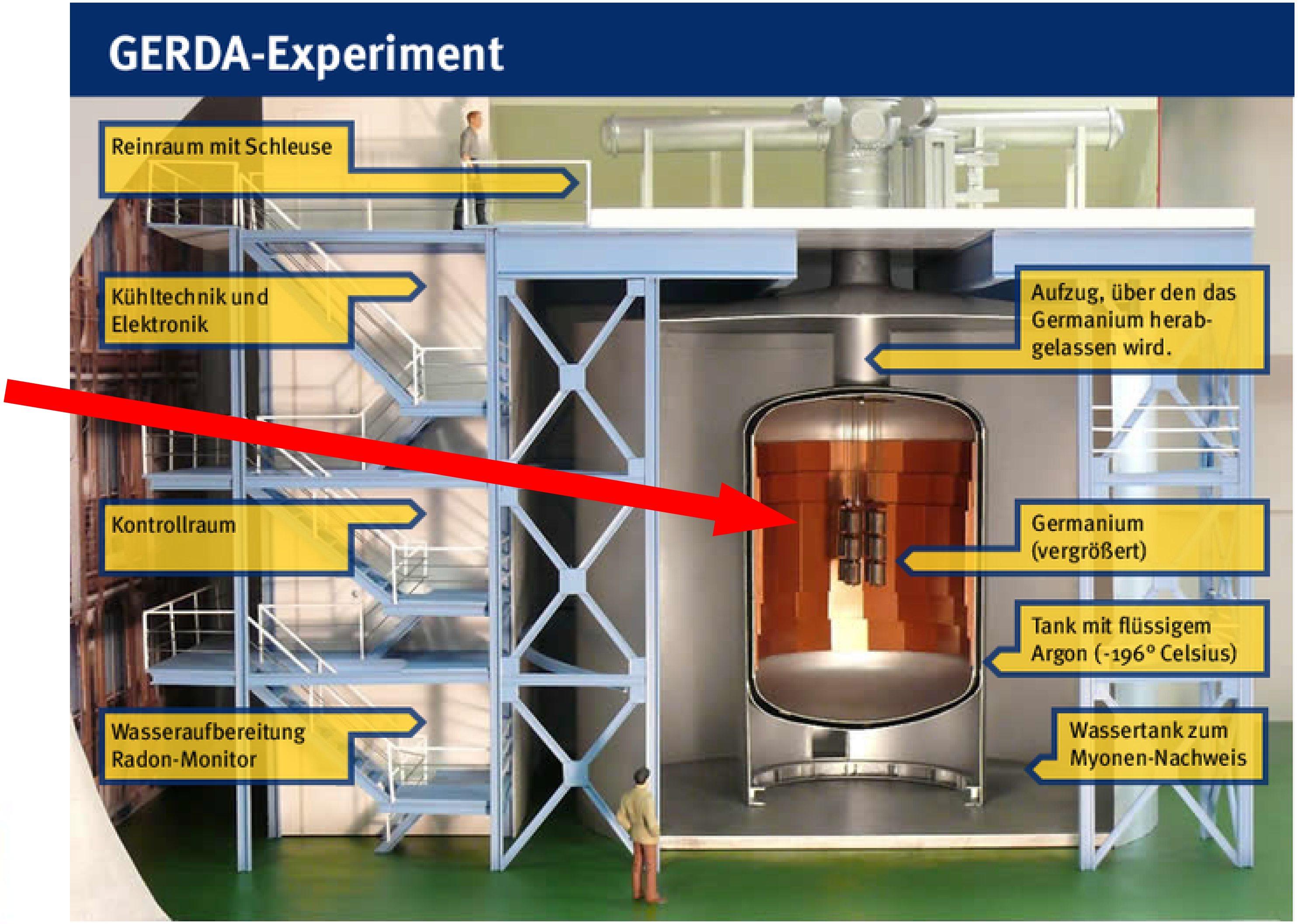


Example: Hd-Moscow

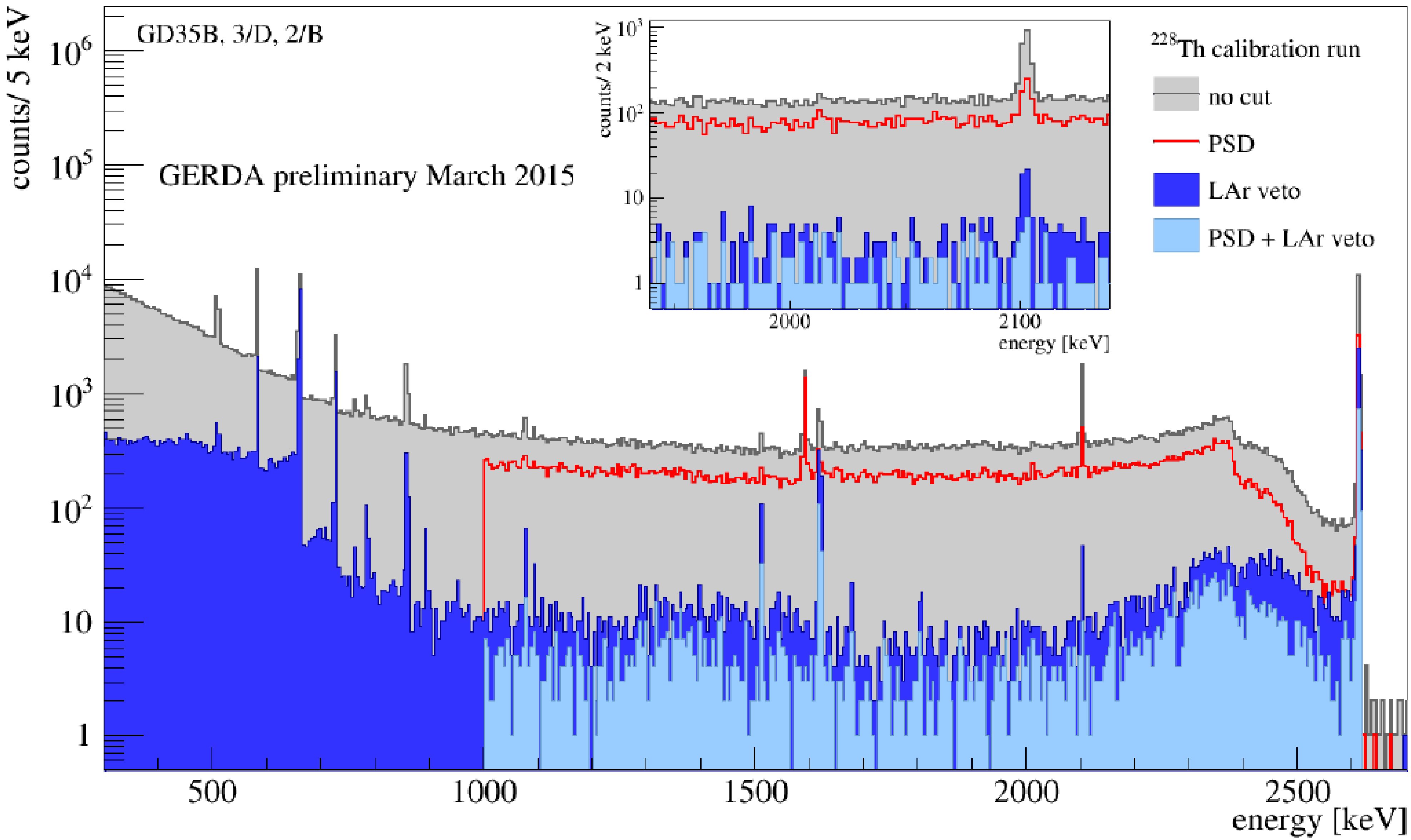


Example: GERDA



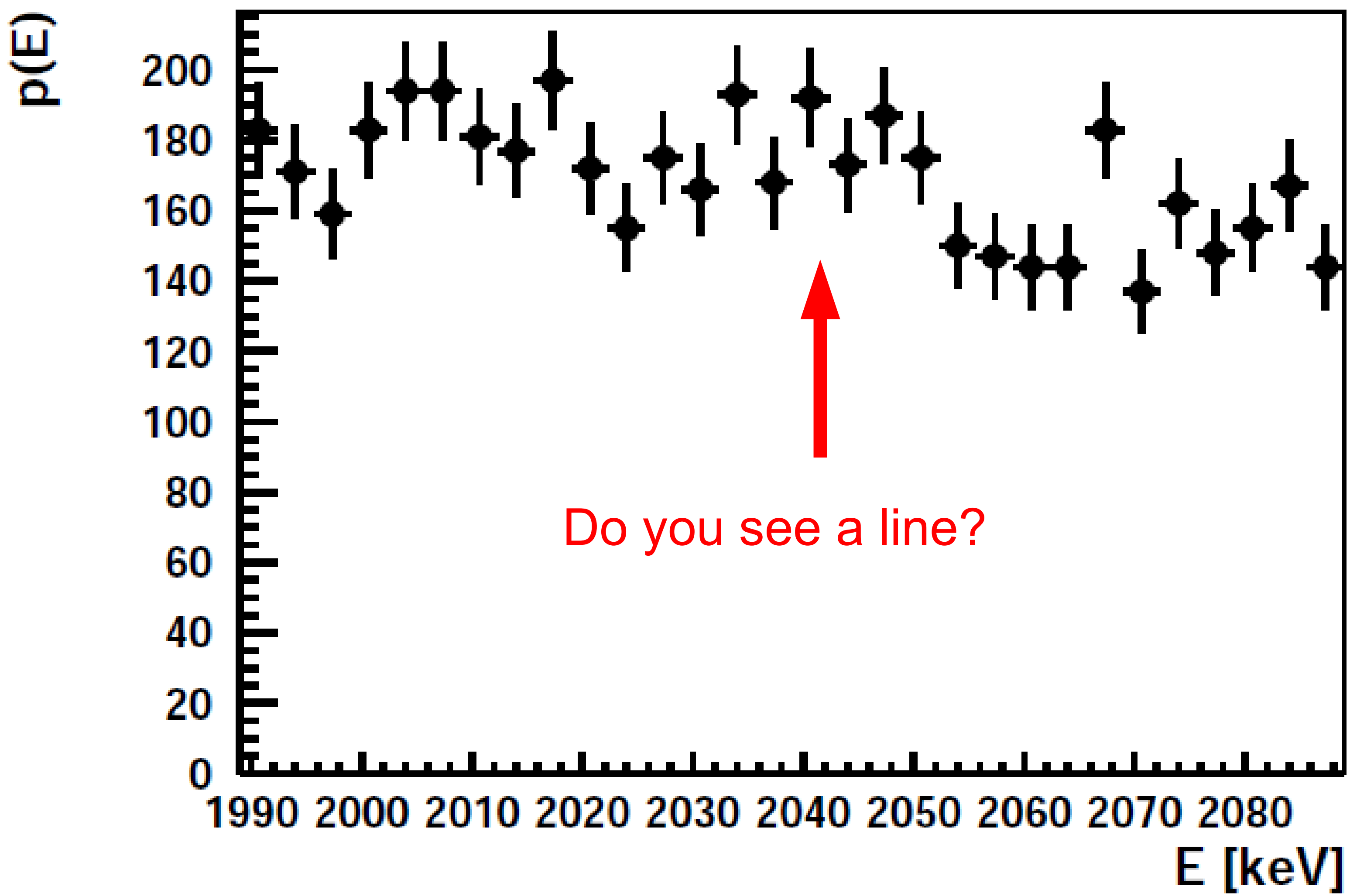


Example: GERDA



The principle

Neutrinoless double beta-decay



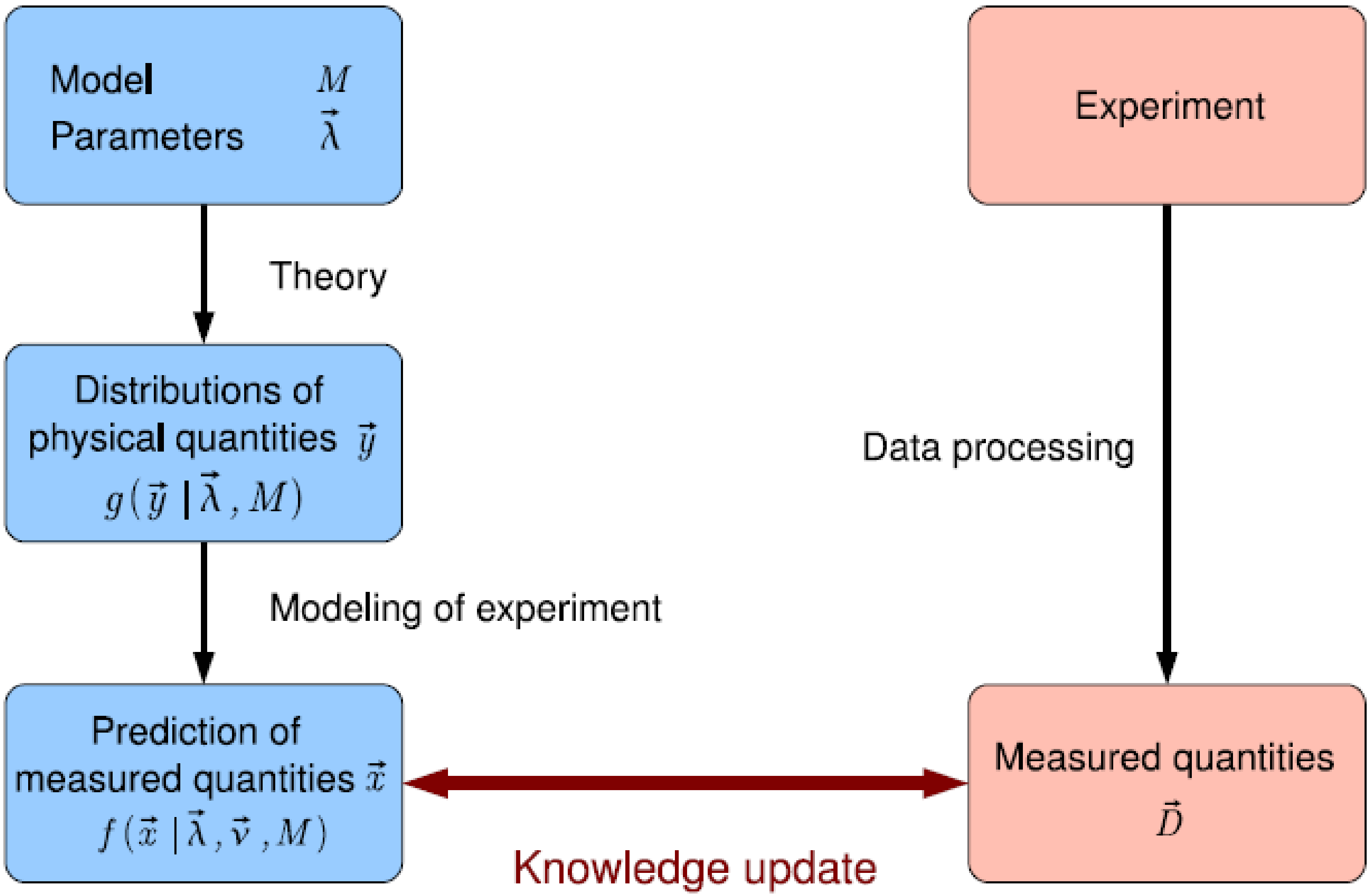
Typical questions

Major subjects of data analysis

- Model comparison: Which model describes the data best?
  - SM background only?
  - Does 0nubb + background describe the data better?
- Parameter estimation: Given a model, what are the values of its free parameters?
  - What is the rate of 0nubb?
  - What is the actual background level?
- Goodness-of-fit: Given a model, is it consistent with the data?
  - Does the background-only hypothesis describe the data reasonably well?

Scientific reasoning

Ingredients



Scientific reasoning

Deductive reasoning

- Used when making predictions from a model
- Application in data analysis:
  - Premise P (model with parameters) → Conclusion Q (observables)
  - Premise Q (observables) → Conclusion R (set of observations)
  - Thus: Premise P (model) → Conclusion R (set of observations)
- *Given a model, the outcome is specified*
- No need to argue, it's math!
- *Example:* 0nubb + background model predicts a certain energy spectrum

Scientific reasoning

Inductive reasoning

- Used when choosing a model
- Application in data analysis:
  - Premise P (model with parameters) → Conclusion R (set of observations)
  - Observe R, what does it say about P? Not much since it could have been $P_1 \rightarrow R$, $P_2 \rightarrow R$, $P_3 \rightarrow R$, ...
- *Validity* of model P?
  - If we know *all* models, and only P results in R, then we know that P is true.
  - Otherwise, can *not verify* the model.
  - Can try to *falsify* the model: if we observe something that contradicts the model, it can not be true
- Can we know which model is true? No!

Scientific reasoning

Truth and knowledge

- Plato: knowledge is *justified true belief*.
- Proposition P is known to be true if and only if
  - P is true.
  - P is believed to be true.
  - It is justified that P is believed to be true.
- But: we can not know the truth, so:

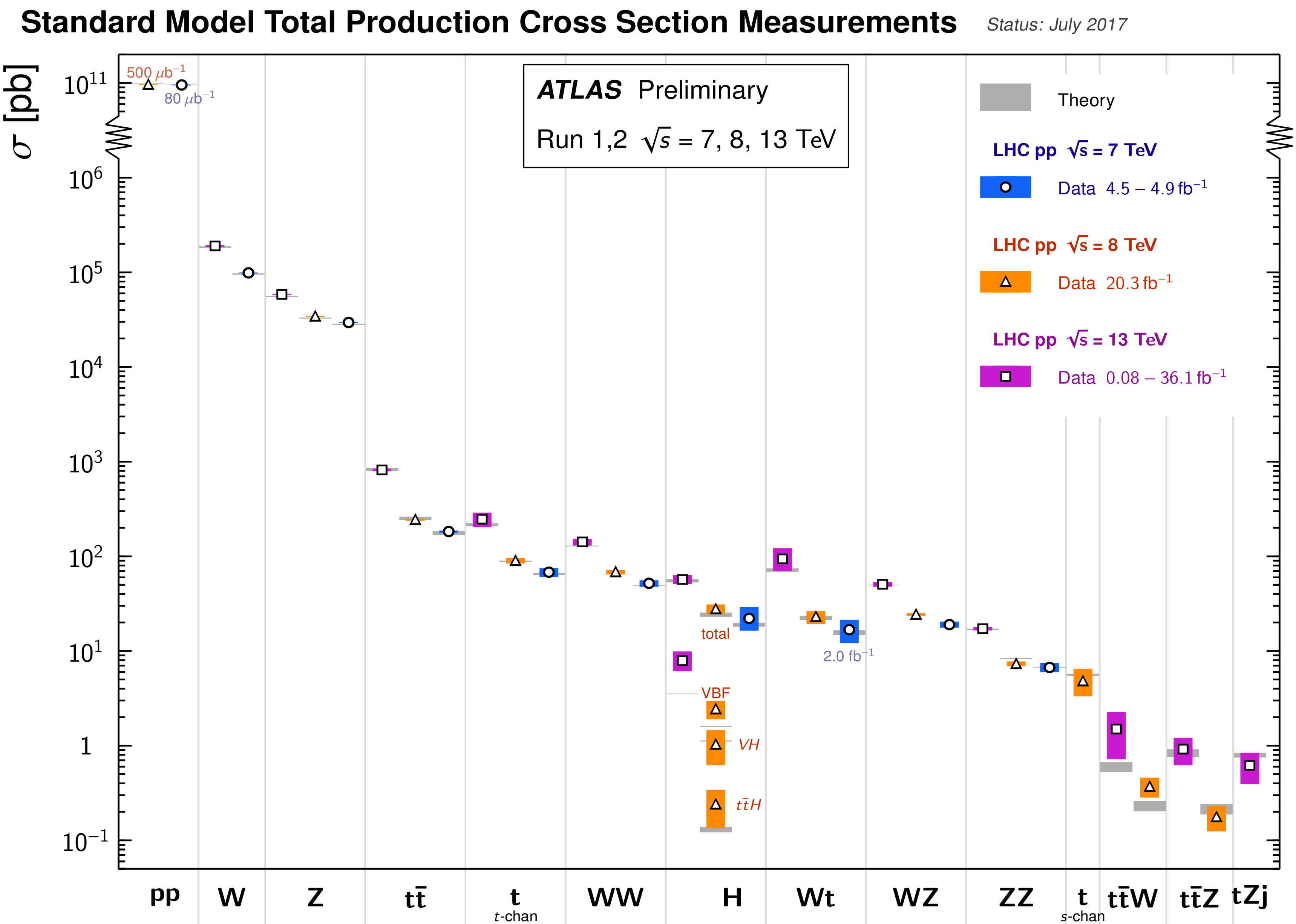
Knowledge is justified belief

- This discussion is known as the Gettier problem

- Justification comes from experimental observations:
  - Derive predictions from model and test them
  - The more tests are passed, the greater the belief in the model

Examples: SM of particle physics, general relativity, ...

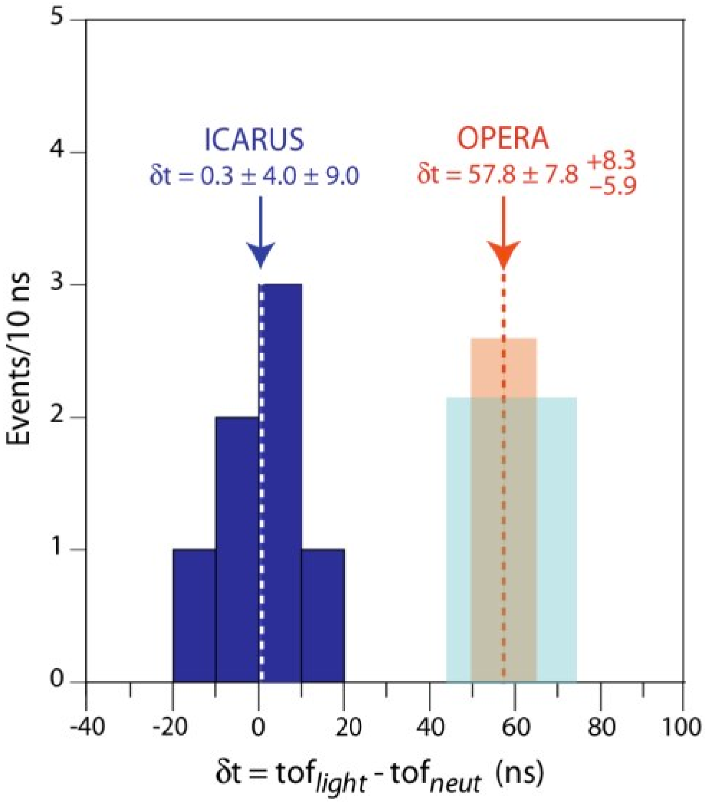
SM vs. “tensions”



SM vs. “tensions”

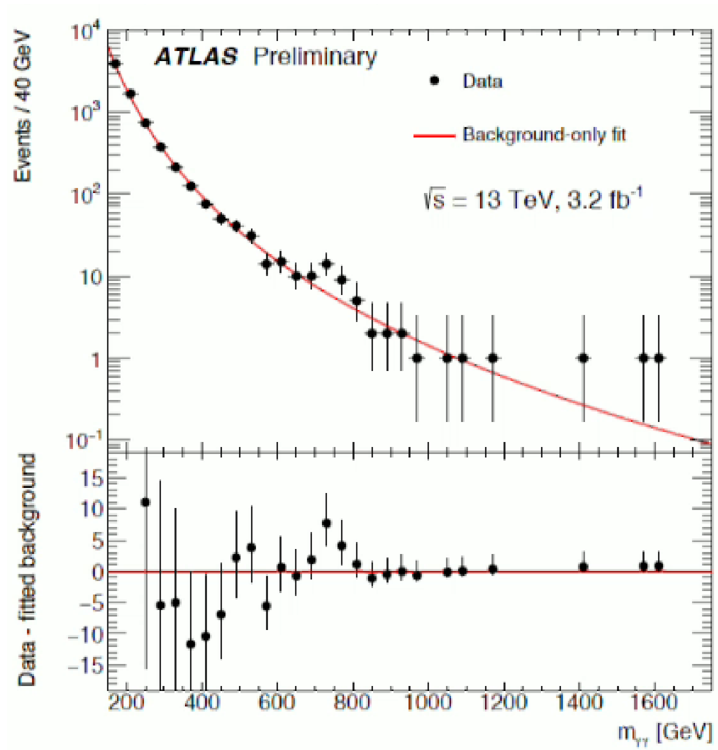
Is the SM wrong?

2011



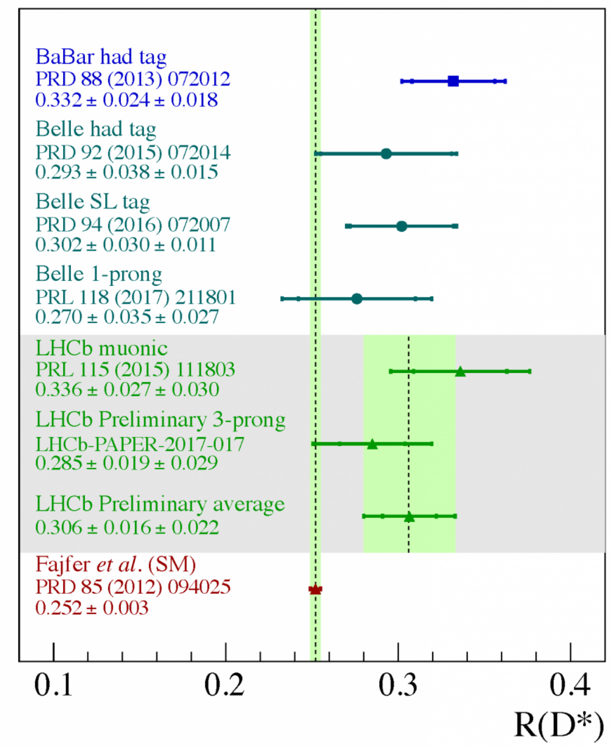
“No, can't be...”

2015



“Could be, but wasn't confirmed”

> 2016/17



“Mh, really?”

Scientific reasoning

Application in science

- How do we gain knowledge?
  - Set up models and specify their parameters (check arXiv.org!)
  - Derive (deductively) predictions from the models
  - Can not know all models, so can not verify a model
  - Good model: falsifiable, make predictions which can be proven wrong (Z' vs. SUSY vs. string theory)
  - Use data to gain knowledge about the models and parameters

Examples:

- Special relativity predicts time dilation. Atmospheric muons can thus be observed on the earth's surface.
- Neutrino postulation: Pauli was hesitant to publish his neutrino idea because he thought it would be difficult to discover.

Can we quantify the knowledge about a model? Yes, use probabilities

Axioms and interpretation

- Kolmogorov axioms: start from set S
  1. For each subset A, assign probability P(A) between 0 and 1
  2. Probability P(S) = 1
  3. For disjunct subsets A and B:

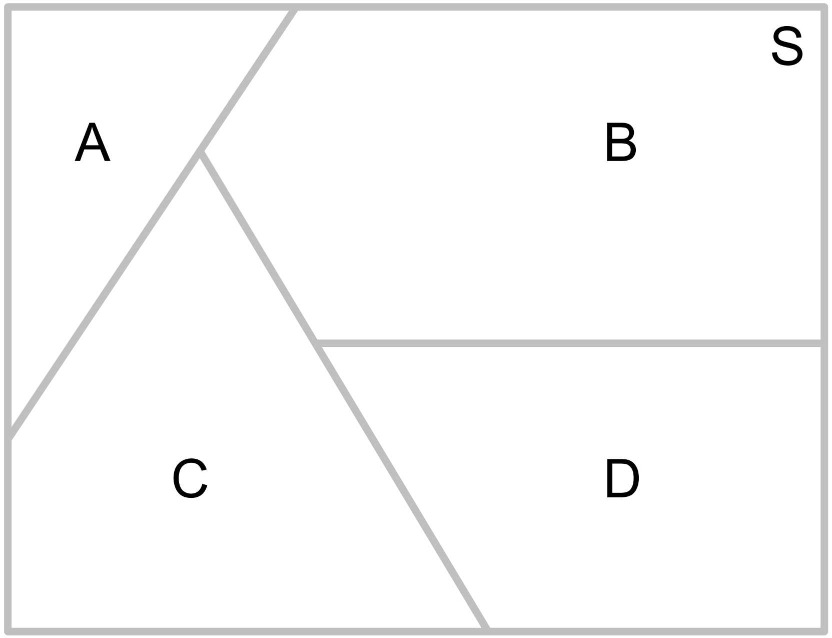
$$P(A \text{ or } B) = P(A) + P(B)$$

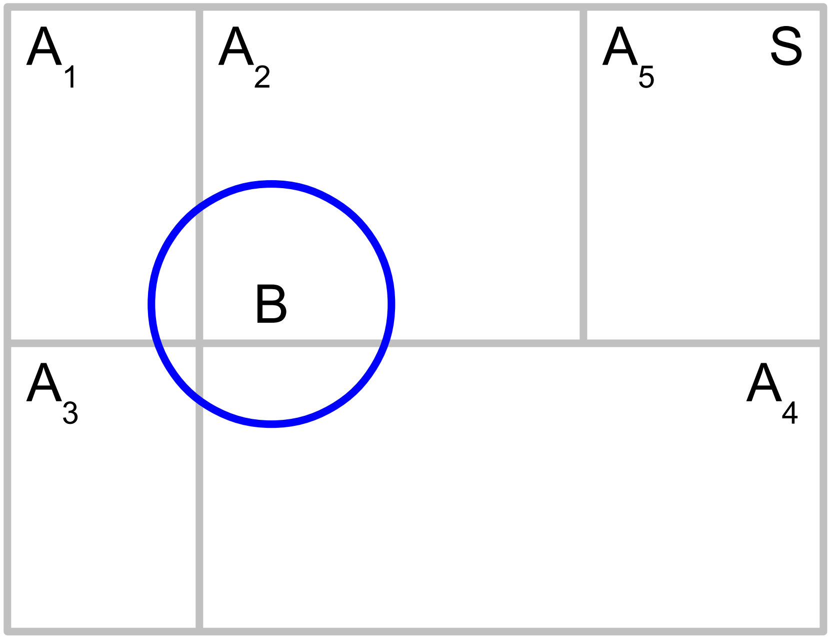
Nice mathematical formulation, but *meaningless*!

- Law of total probability:

$$P(B) = \sum P(B|A_i) \cdot P(A_i)$$

Probability





$P(A) \geq 0, \ldots$

$A \cap B = \emptyset, \ldots$

$S = A \cup B \cup C \cup D$

$P(S) = P(A) + P(B) + P(C) + P(D) = 1$

$P(A) \geq 0, \ldots$

$$P(B) = \sum P(B|A_i) \cdot P(A_i)$$

$$= P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2)$$

$$+ P(B|A_3) \cdot P(A_3) + P(B|A_4) \cdot P(A_4)$$

$$+ \underbrace{P(B|A_5)}_{= 0} \cdot P(A_5)$$

Probability

Axioms and interpretation

- Kolmogorov axioms: start from set S
  1. For each subset A, assign probability P(A) between 0 and 1
  2. Probability P(S) = 1
  3. For disjunct subsets A and B: P(A or B) = P(A) + P(B)
- Nice mathematical formulation, but *meaningless*!

Bayesian interpretation

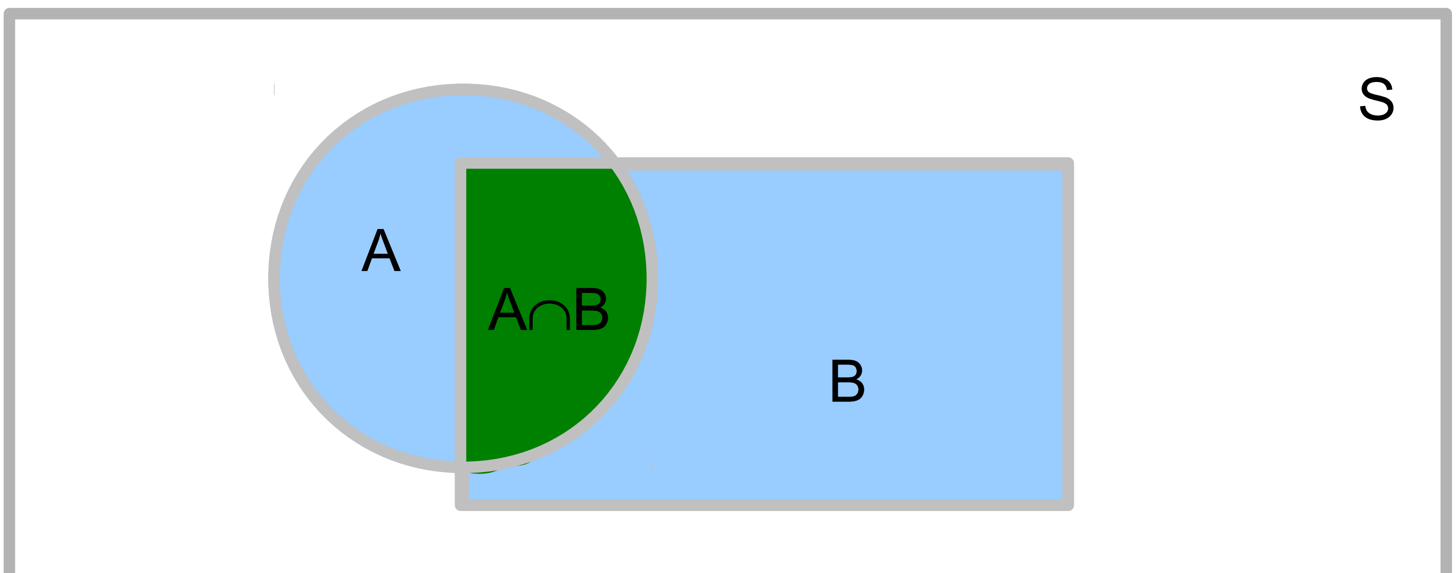
- Subsets correspond to hypotheses, i.e. a model with a particular value of the parameter. Example: SM and the electron mass
- Probability is understood as *degree-of-belief* (or *state-of-knowledge*) for this hypothesis to be true
- Interpretation fully consistent with Kolmogorov axioms

Bayes' Theorem

$$P(A|B) \cdot P(B) = P(A \wedge B) = P(B|A) \cdot P(A)$$

$$\Leftrightarrow \quad P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$\Leftrightarrow \quad P(A|B) = \frac{P(B|A) \cdot P(A)}{\sum P(B|A_i) \cdot P(A_i)}$$



Bayes' Theorem

$$P(A|B) \cdot P(B) = P(A \wedge B) = P(B|A) \cdot P(A)$$

$$\Leftrightarrow \quad P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$\Leftrightarrow \quad P(A|B) = \frac{P(B|A) \cdot P(A)}{\sum P(B|A_i) \cdot P(A_i)}$$

- Here: $P(\text{theory}|\text{data}) \propto P(\text{data}|\text{theory}) \cdot P(\text{theory})$

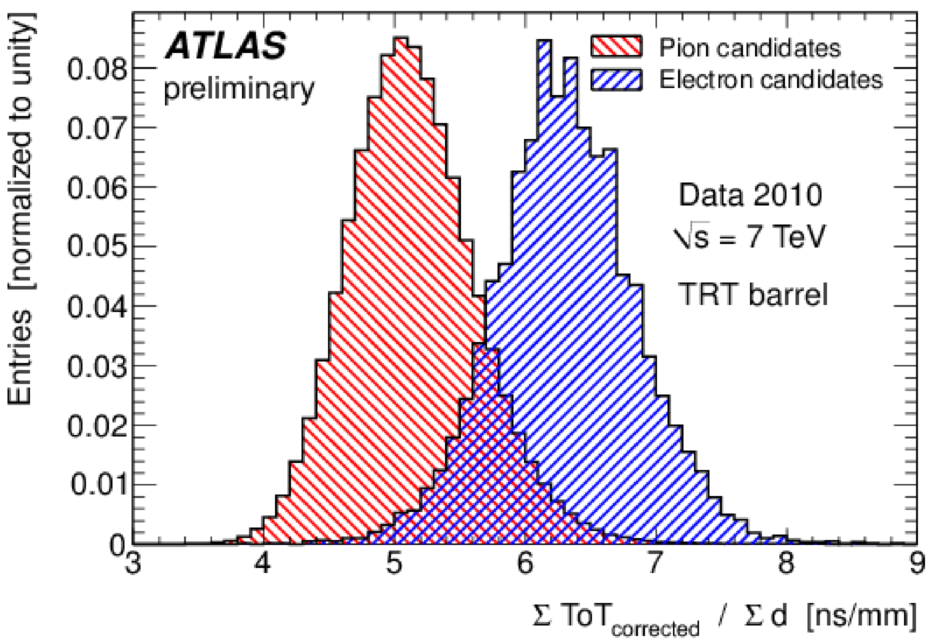
- P(theory | data): *posterior probability* (induction)
- P(data | theory): *probability of the data*, *likelihood* (deduction)
- P(theory): *prior probability*
- In words: “My degree-of-belief of a model is x%”, or

  “The parameter values lie in an interval [a,b] with x% probability”

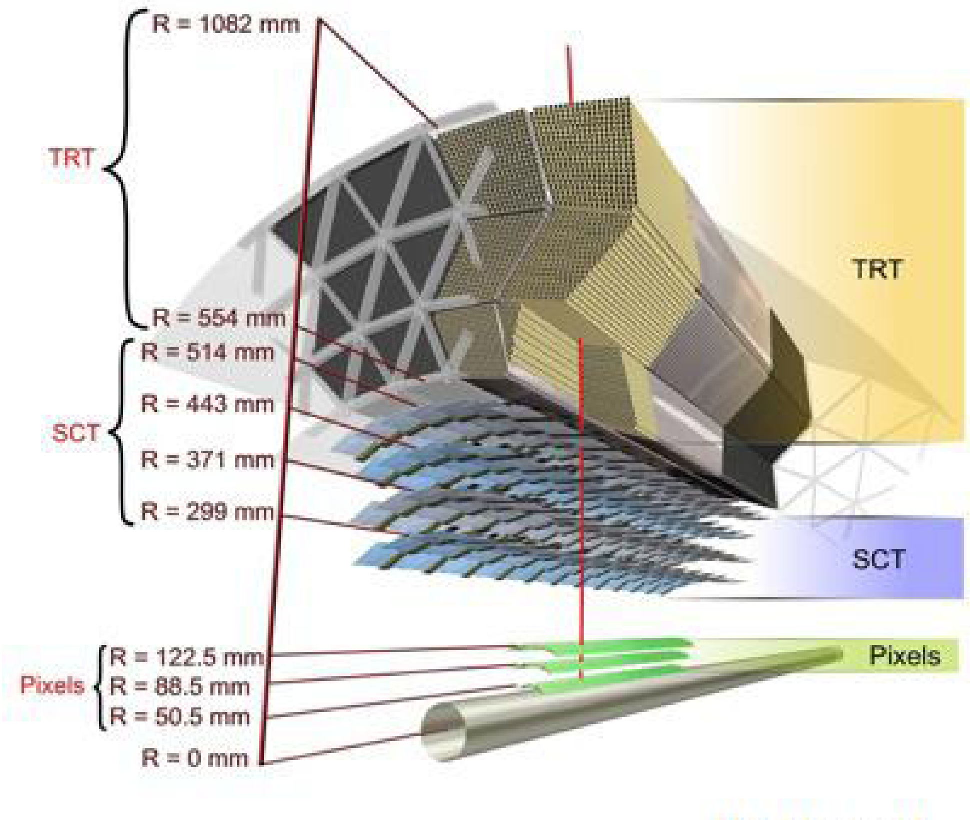
A simple example

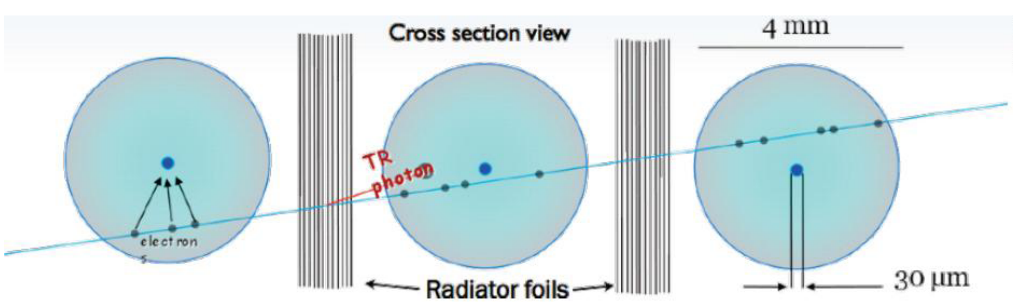
Particles in a TRT

- Particle identification based on transition radiation
- Distinguish between electrons and charged pions
- What does a "signal" mean?



[ATLAS-CONF-2011-128]





A simple example

Particles in a TRT

- Assume test beam measurement
- Test beam composition:
  - p(electron) = 90%
  - p(pion) = 10%
- Detection efficiencies:
  - p(signal | electron) = 95%
  - p(signal | pion) = 6%

Your turn! What is p(electron | signal)?

Use Bayes' Theorem:

$$P(A_j|B) = \frac{P(B|A_j)\cdot P(A_j)}{\sum P(B|A_i)\cdot P(A_i)}$$

What are A_1 and A_2 and B?

A simple example

Particles in a TRT

- Assume test beam measurement
- Test beam composition:

  - p(electron) = 90%
  - p(pion) = 10%

- Detection efficiencies:

  - p(signal | electron) = 95%
  - p(signal | pion) = 6%

$$p(e|\text{Signal}) = \frac{p(\text{Signal}|e) \cdot p(e)}{p(\text{Signal}|e) \cdot p(e) + p(\text{Signal}|\pi) \cdot p(\pi)}$$

$$p(\pi|\text{No signal}) = \frac{p(\text{No signal}|\pi) \cdot p(\pi)}{p(\text{No signal}|e) \cdot p(e) + p(\text{No signal}|\pi) \cdot p(\pi)}$$

A simple example

Particles in a TRT

- Probabilities:

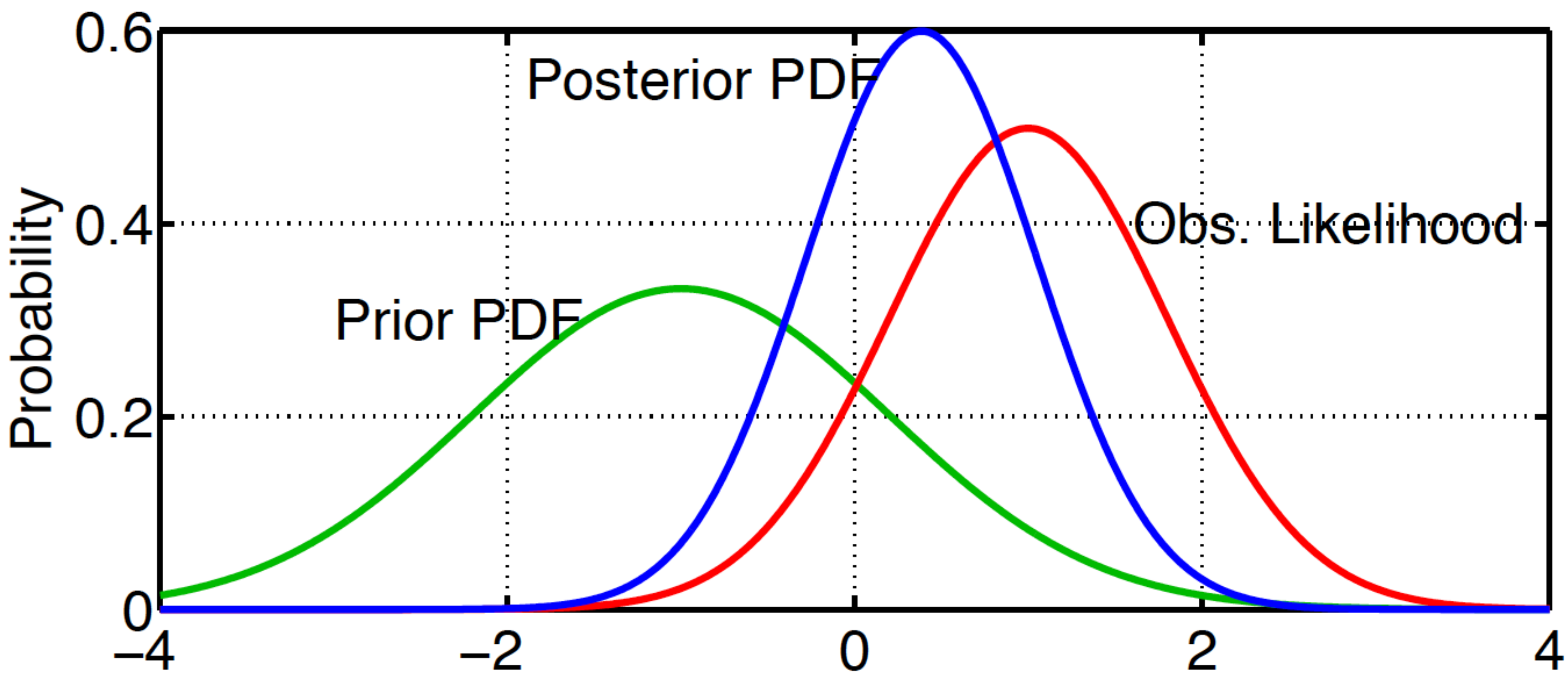
$$p(e|\text{Signal}) = \frac{0.95 \cdot 0.9}{0.95 \cdot 0.9 + 0.06 \cdot 0.10} = 0.993$$

$$p(\pi|\text{Signal}) = \frac{0.06 \cdot 0.10}{0.95 \cdot 0.9 + 0.06 \cdot 0.10} = 0.007$$

$$p(e|\text{No signal}) = \frac{0.05 \cdot 0.90}{0.05 \cdot 0.90 + 0.94 \cdot 0.10} = 0.312$$

$$p(\pi|\text{No Signal}) = \frac{0.94 \cdot 0.10}{0.05 \cdot 0.90 + 0.94 \cdot 0.10} = 0.676$$

A bit of bashing



"A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule"

Stephen Senn, Statistician & Bayesian Skeptic (mostly)

Priors

Where does prior knowledge come from?

- Prior can come from
  - personal degree-of-belief (gut feeling),
  - theoretical considerations (how badly do you want SUSY to be true?),
  - auxiliary measurements, …
  - … good arguments … (in the best case)
- Elegant update of knowledge: posterior of one experiment can be prior of another experiment. Natural way to combine measurements.

P(Model | Data 1) ~ P(Data 1 | Model) x P(Model)

and P(Model | Data 2) ~ P(Data 2 | Model) x P'(Model)

with P' = P(Model | Data 1) = P(Data 1 | Model) x P(Model)

→ P(Model | Data 2) ~ P(Data 2 | Model) x P(Data 1 | Model) x P(Model)

= P(Model | Data 1 + Data 2)

Priors

Criticism

- Priors are subjective
  - Yes, but it is made explicit
  - Objective Bayesian movement, try to find objective priors
  - reference priors minimize the "information"
- Prior depends on parametrization (lifetime τ vs. decay constant $\lambda=1/\tau$)
  - Jeffreys prior invariant under reparameterization

Remarks

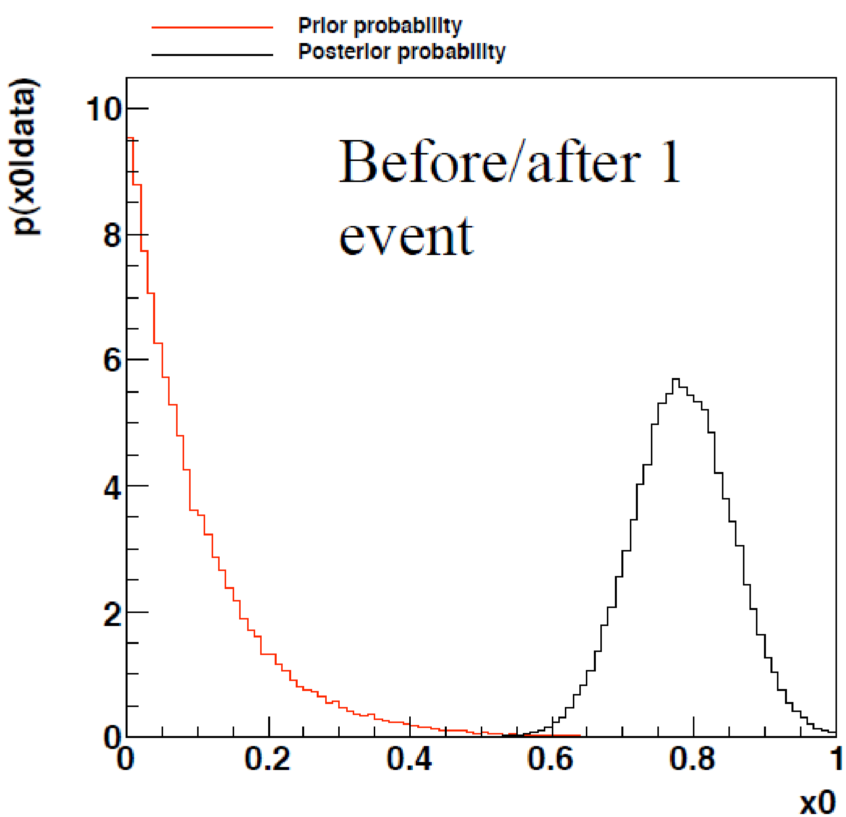
- Choice of (initial) prior should not play a strong role.
- Difficult to formulate a single prior for a collaboration of ~3.000 people
- Practical solution: subjunctive priors. Requote your result under different prior assumptions ("the optimist", "the pessimist", "the ignorant", ...)
- Write down your prior!

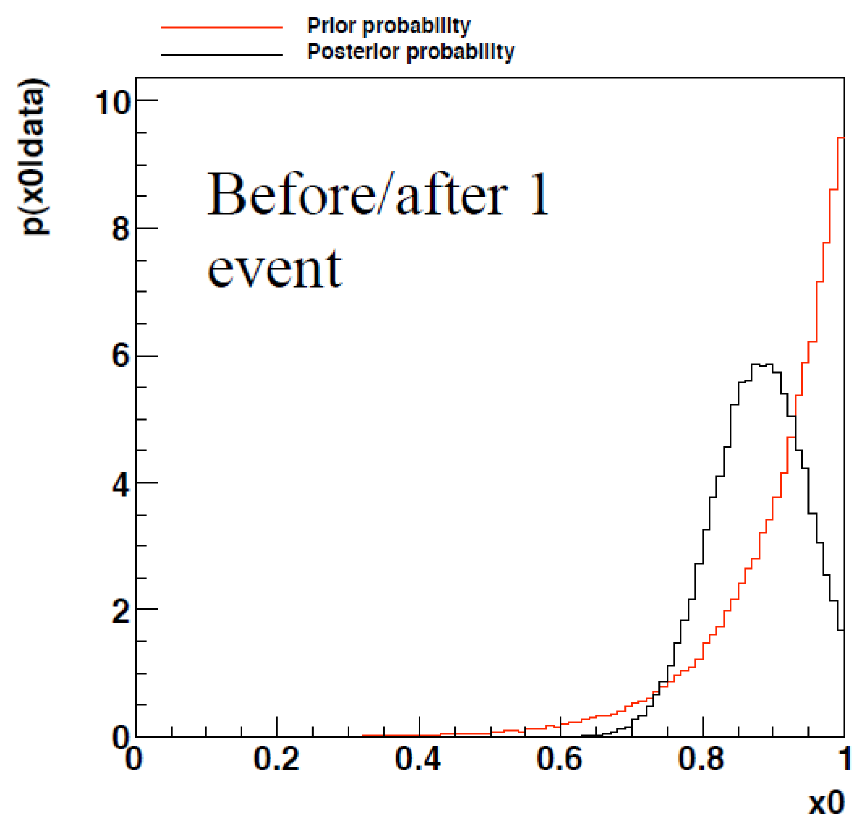
Impact of priors

Two different priors

Model: Gaussian with (unknown) mean value between 0 and 1 (truth: 0.75), and width of 0.1

Start with two different priors (optimistic / pessimistic).
→ Slightly different posteriors after one event.



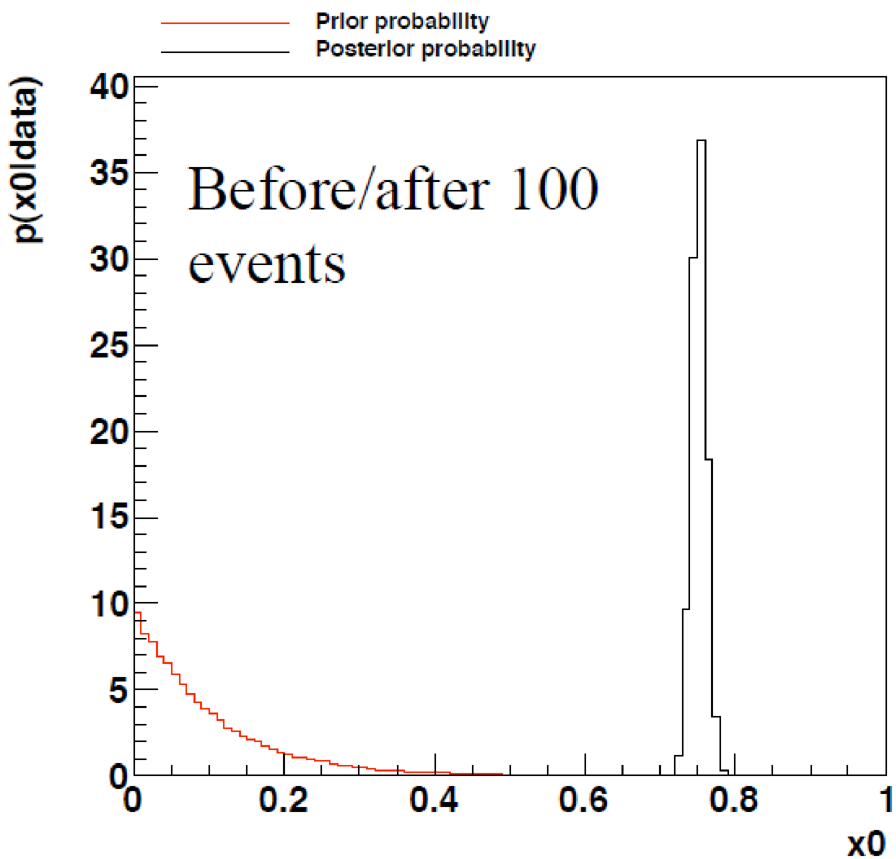


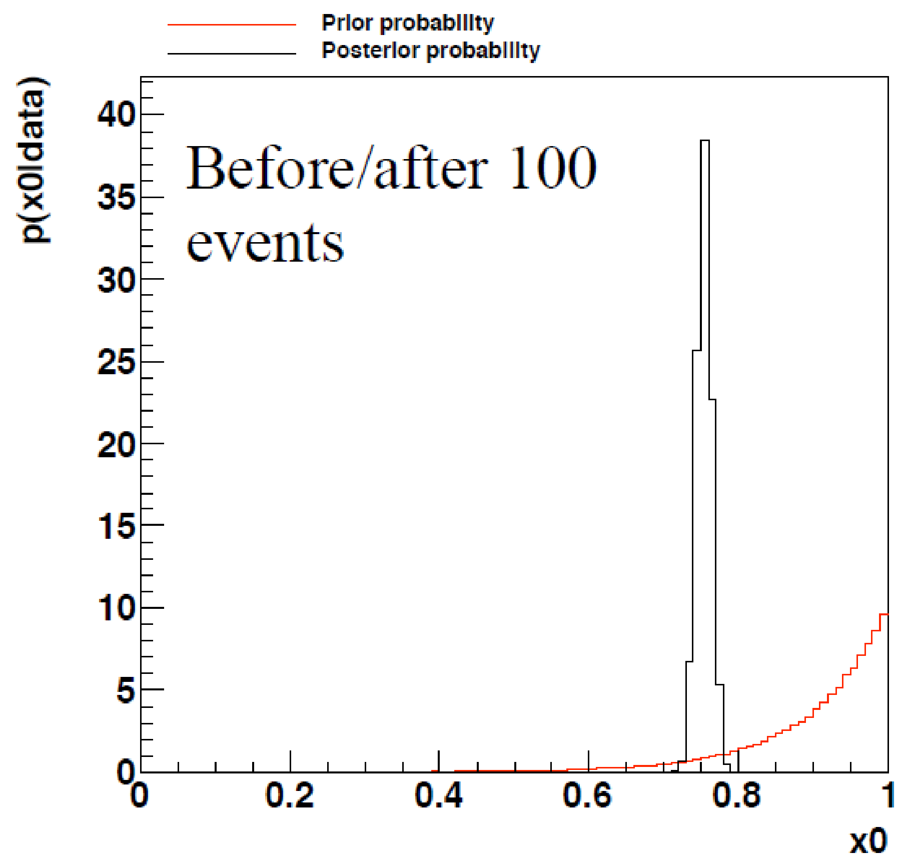
Impact of priors

Two different priors

Model: Gaussian with (unknown) mean value between 0 and 1 (truth: 0.75), and width of 0.1

→ About the same posterior after 100 events





Parameter estimation

- Full solution: posterior probability
  (nothing more than that, but difficult to write down in a paper)

- For papers/talks: summarize posterior using point and interval estimates

- Common point estimators:
  - Maximum posterior probability (global mode)
  - Maximum of marginalized probability (local mode)
  - Mean value of marginalized probability
  - Median of marginalized probability:

$$p(\lambda_i|D)=\int \prod_{i\neq j} d\lambda_j\, p(\vec{\lambda}|D)$$

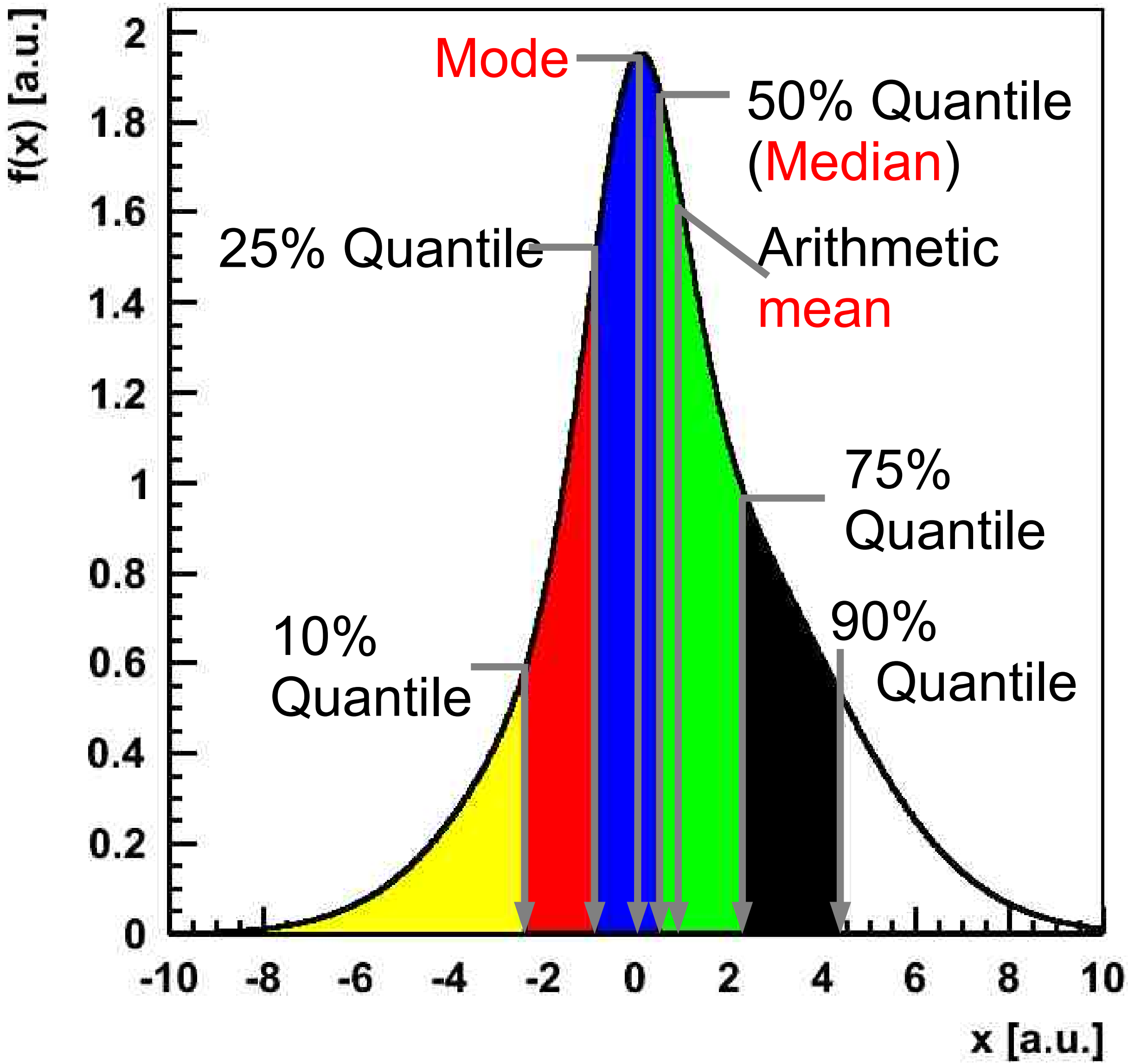
Parameter estimation

- Common interval estimates:
  - Smallest (set of) interval(s) covering 68% probability
  - (Central interval) 16% - 84% quantile
  - Standard deviation of marginalized posterior (a la Gauß)
  - Upper (lower) limits: 99%, 95%, 90% (1%, 5%, 10%) quantiles

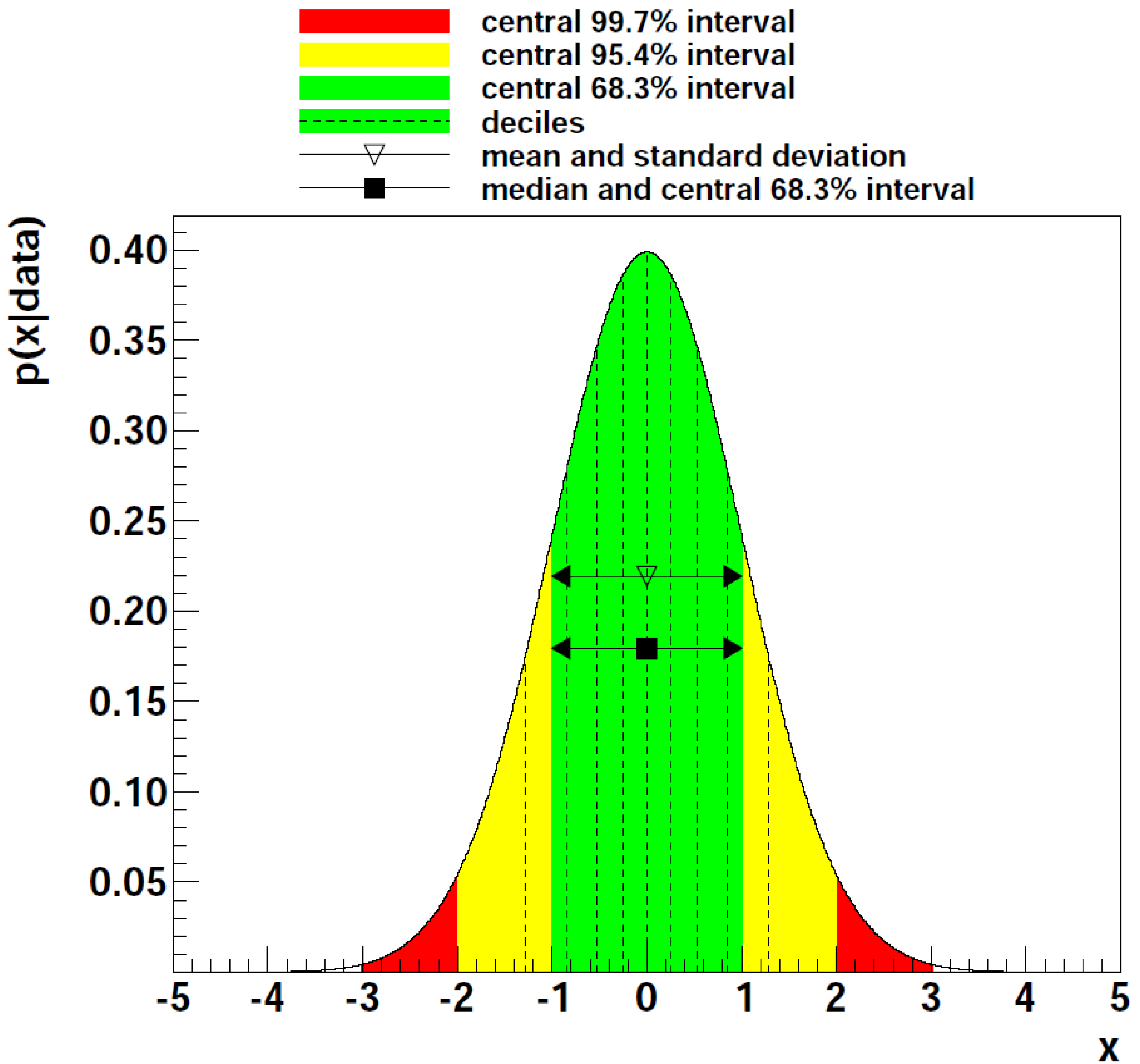
- Practical suggesstion:
  Choose such that point estimator lies inside the estimated interval!

Mean	↔	Standard deviation
Mode	↔	Smallest interval
Median	↔	Central interval

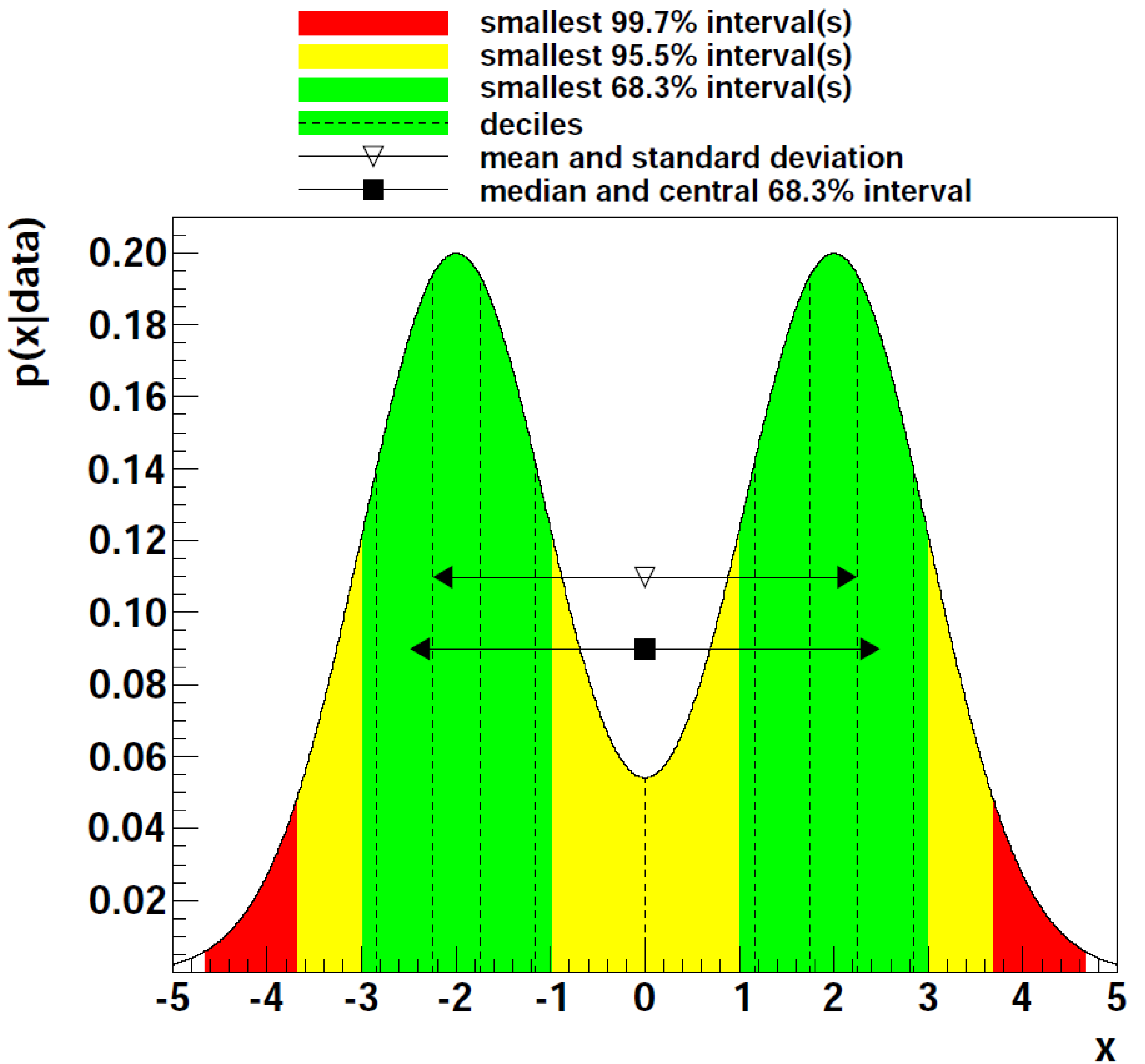
Parameter estimation



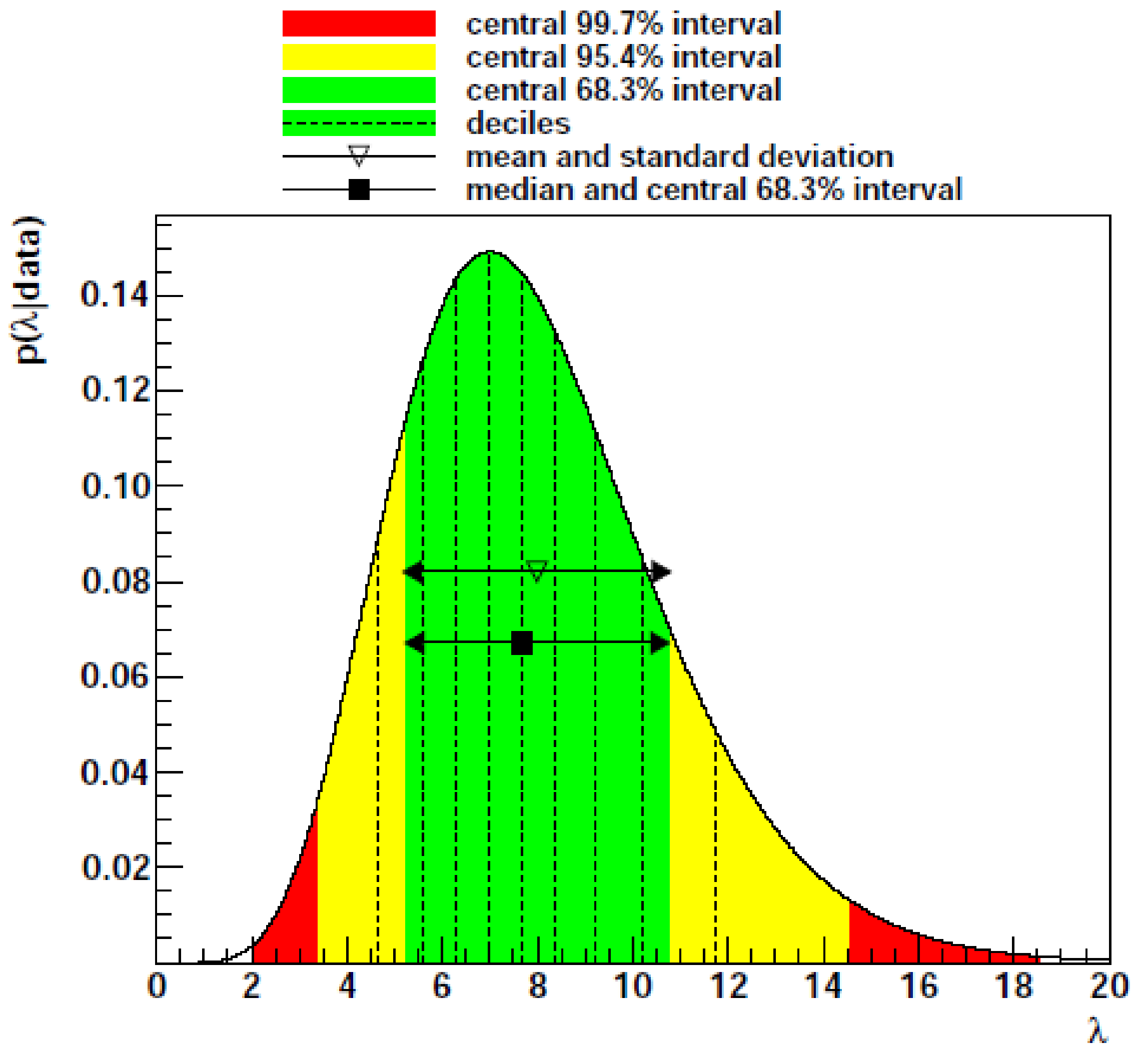
A 1-dim Gaussian



A bimodal distribution



Example: Poisson

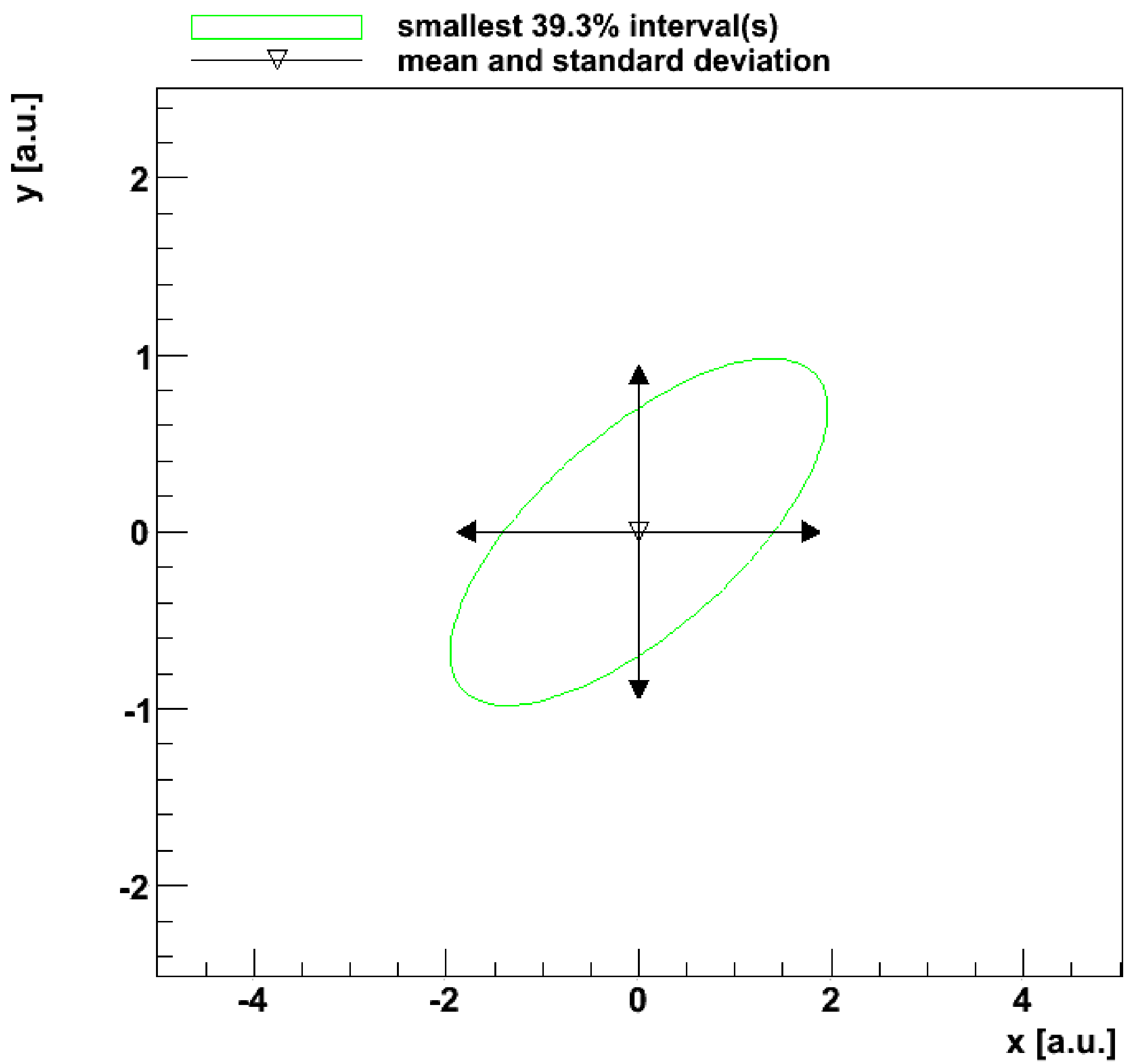


$n_{obs} = 7$

Mean: 7.986
Mode: 7.000
Median: 7.665

STD 2.805
Central 68% interval:
5.24 – 10.88
Smallest 68% interval:
4.6 - 10.2

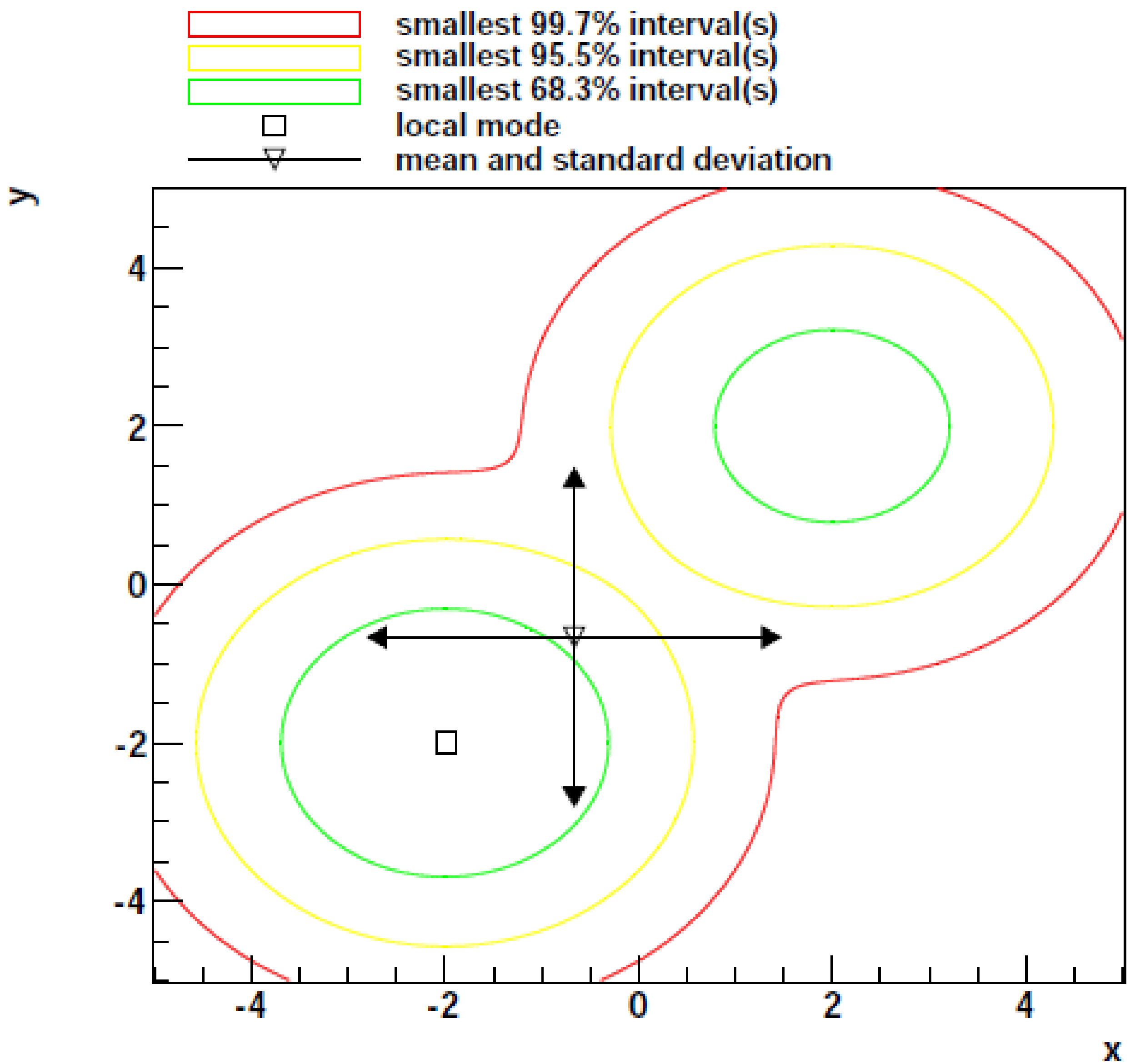
A 2-dim Gaussian



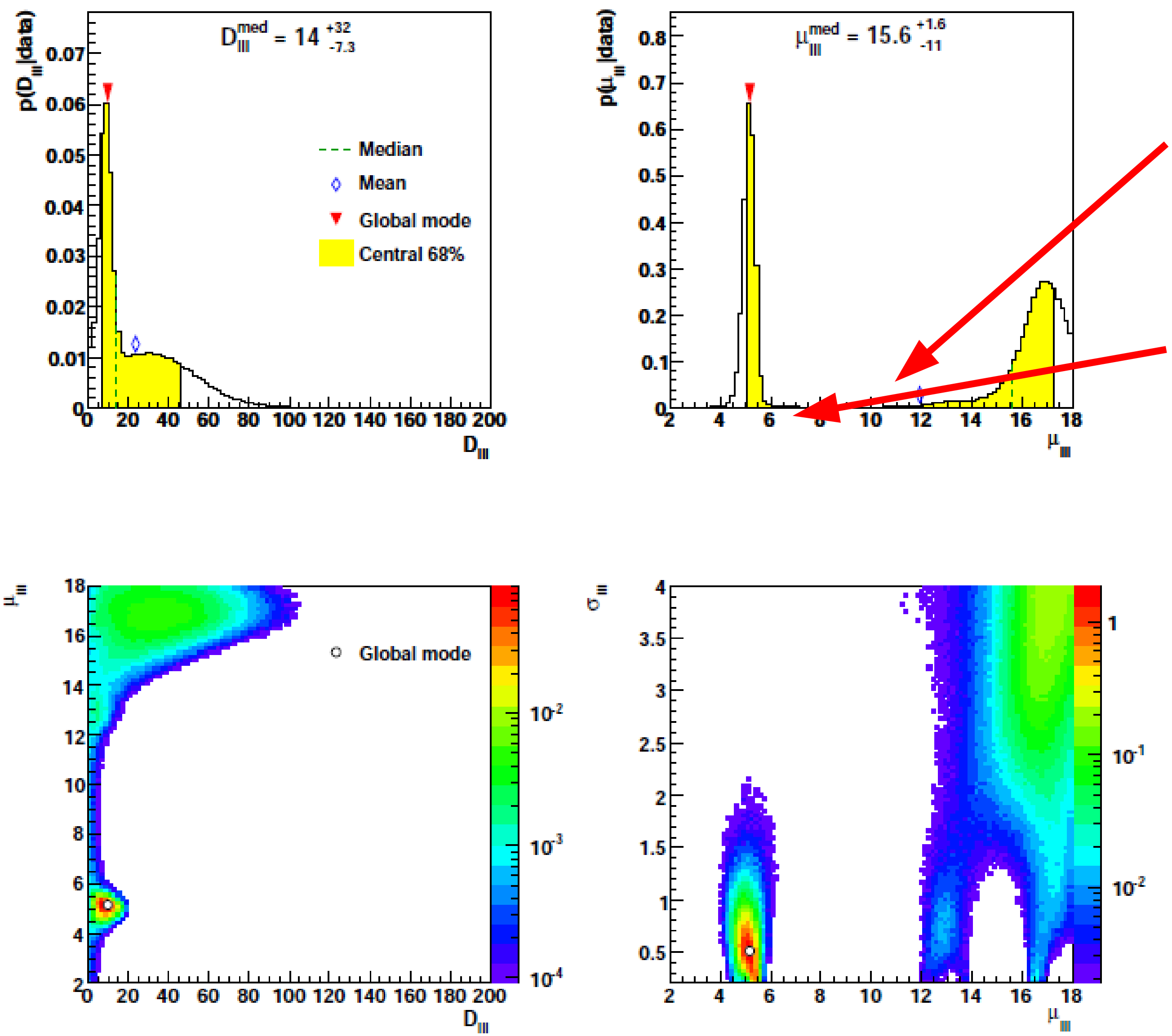
Practical hint:
1-sigma region is not the smallest region covering 68% (for 2-D problems)

(it's 39.3% instead)

A 2D bimodal distribution



Highly non-Gaussian



Median and mean in unlikely region

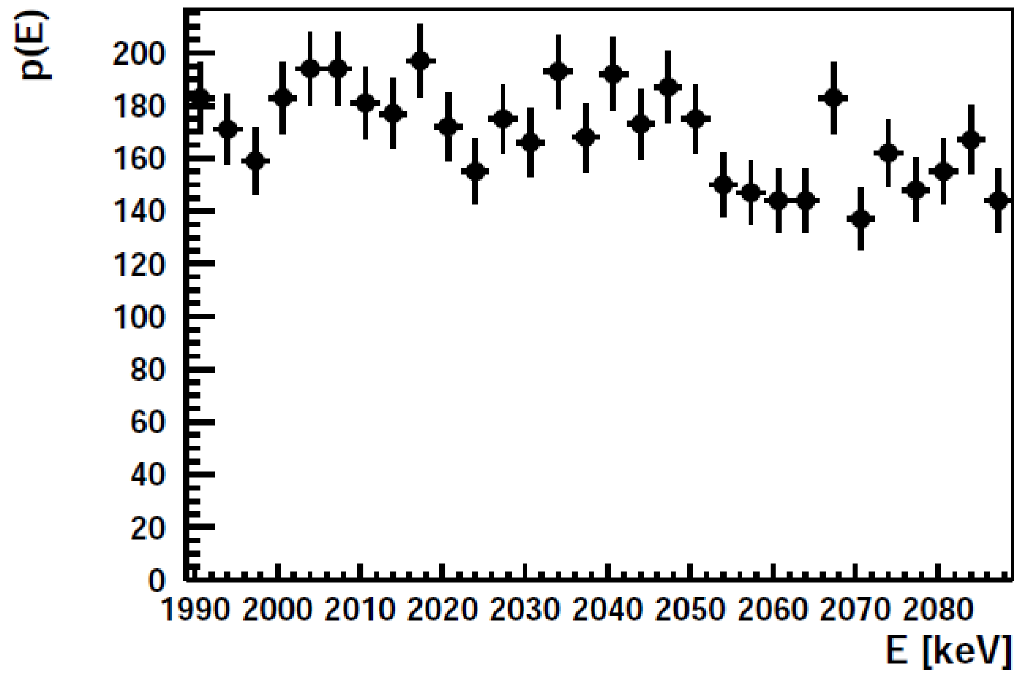
Mode only consistent with one (the most likely) "solution"

Situation even worse in two dimensions

→ **Do not use summary values!**

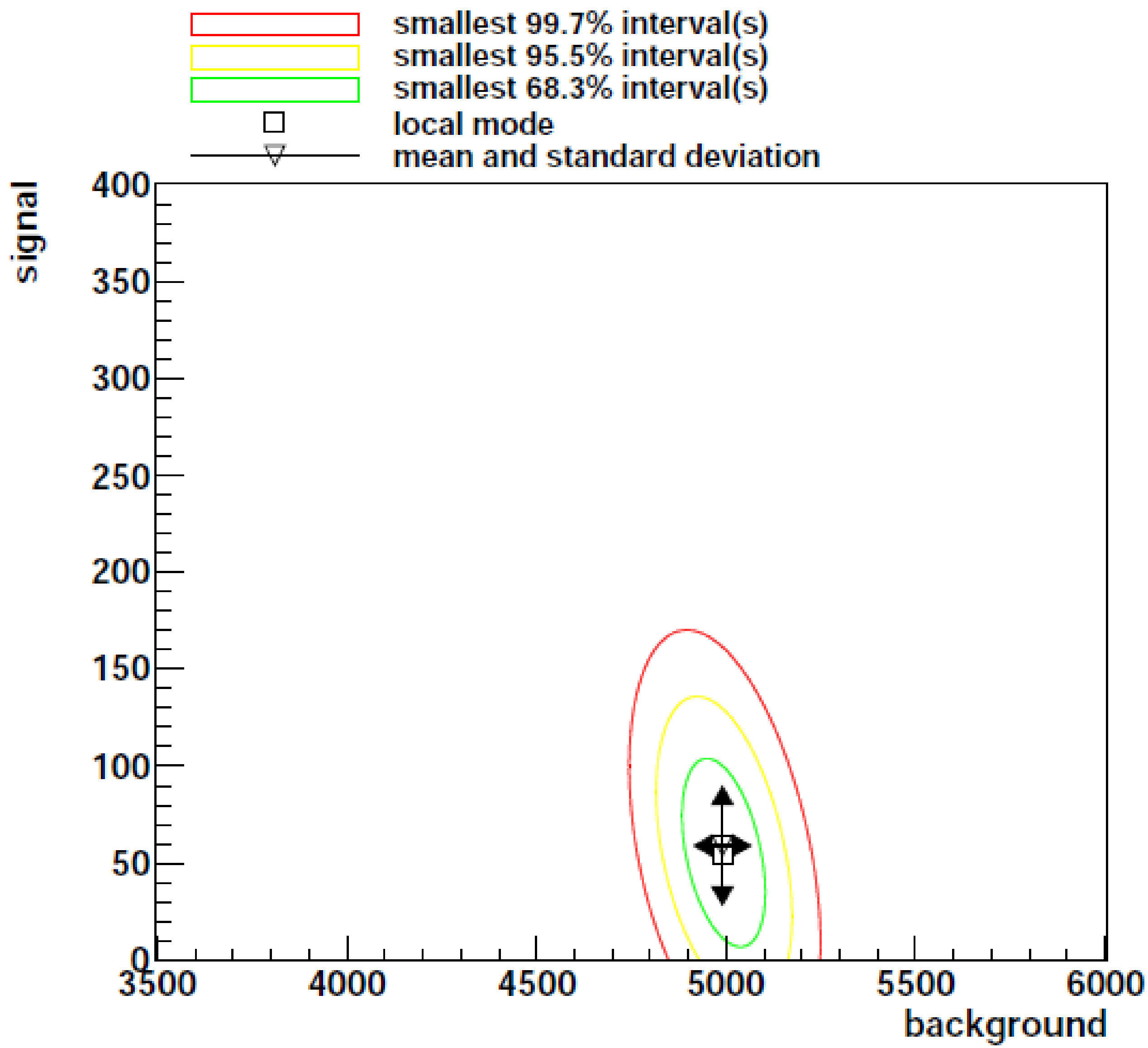
A (more complex) example

Concrete model

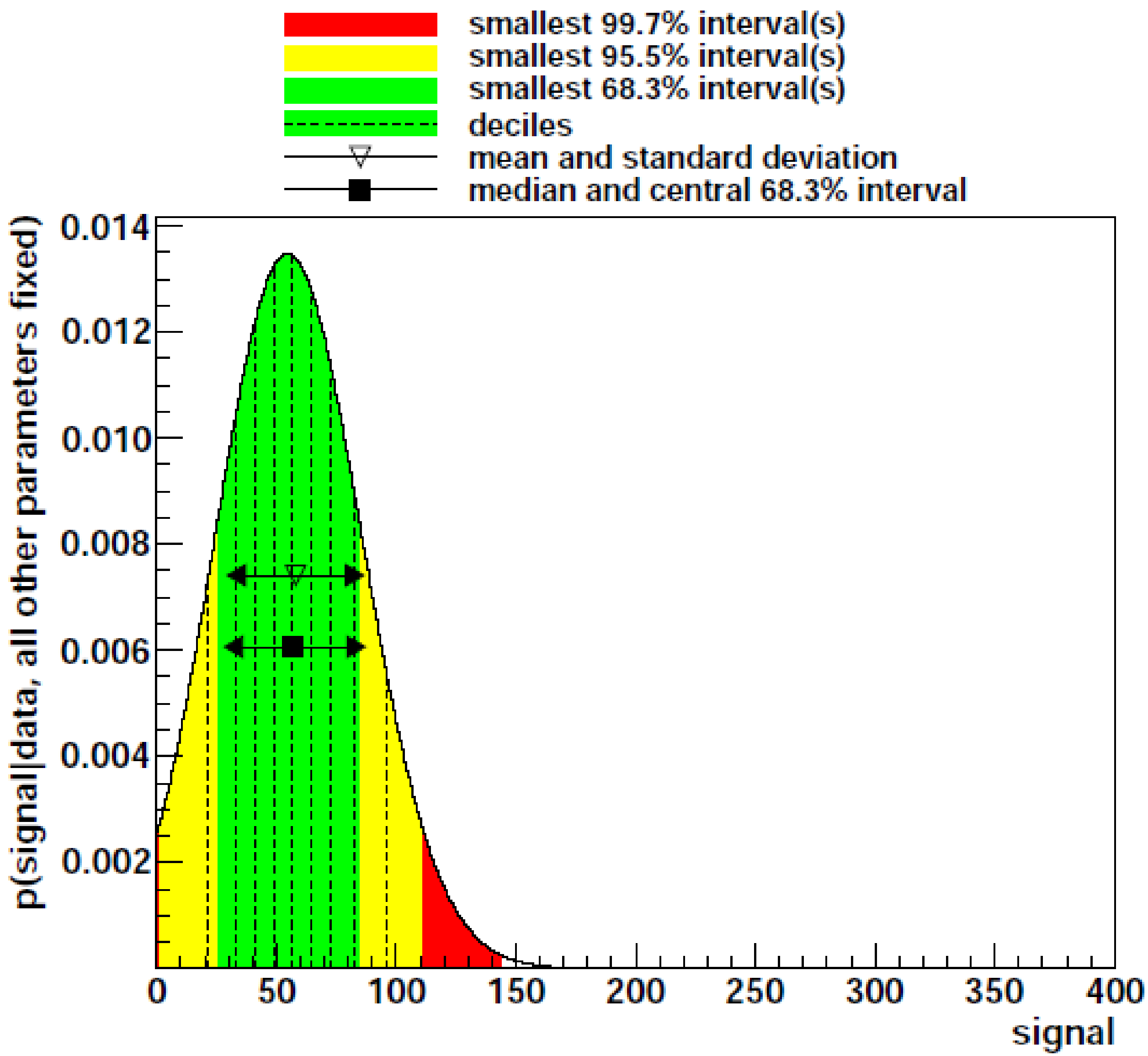


- Data:
  - Binned, number of events
- Shapes:
  - Background linearly decreasing
  - Signal: Gaussian at fixed position
- Statistical model:
  - Independent Poisson fluctuations
  - Parameter 1: background strength, Gaussian prior
  - Parameter 2: signal strength, exponentially decreasing prior
- Fit procedure
  - Template fit: scale signal and background shapes until sum of templates matches data

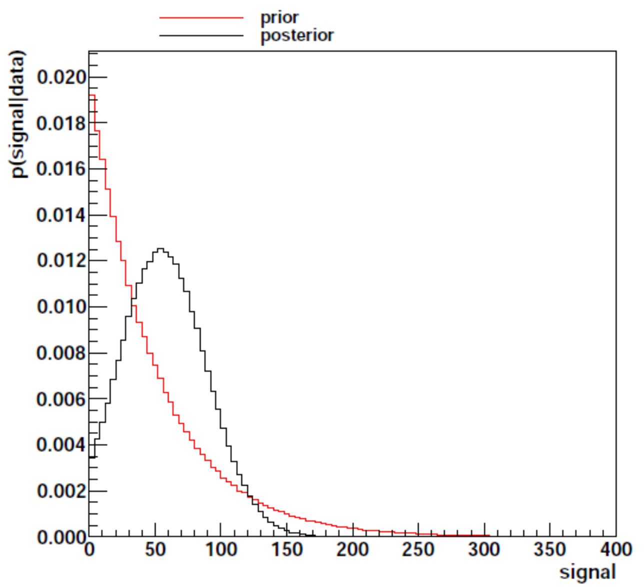
Estimation (2D)

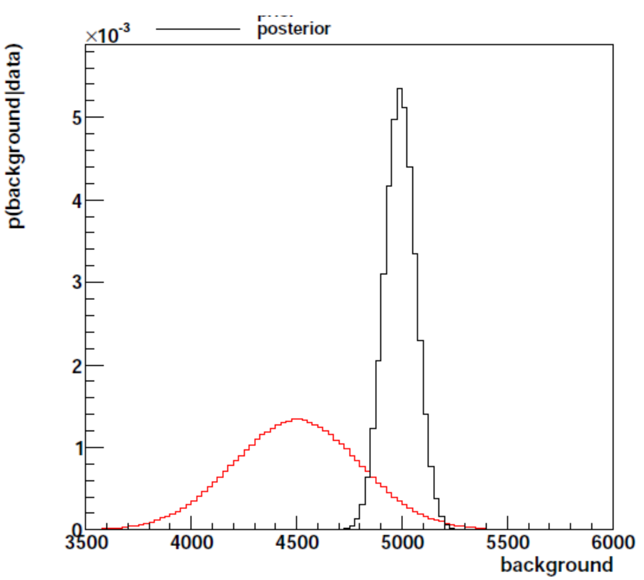


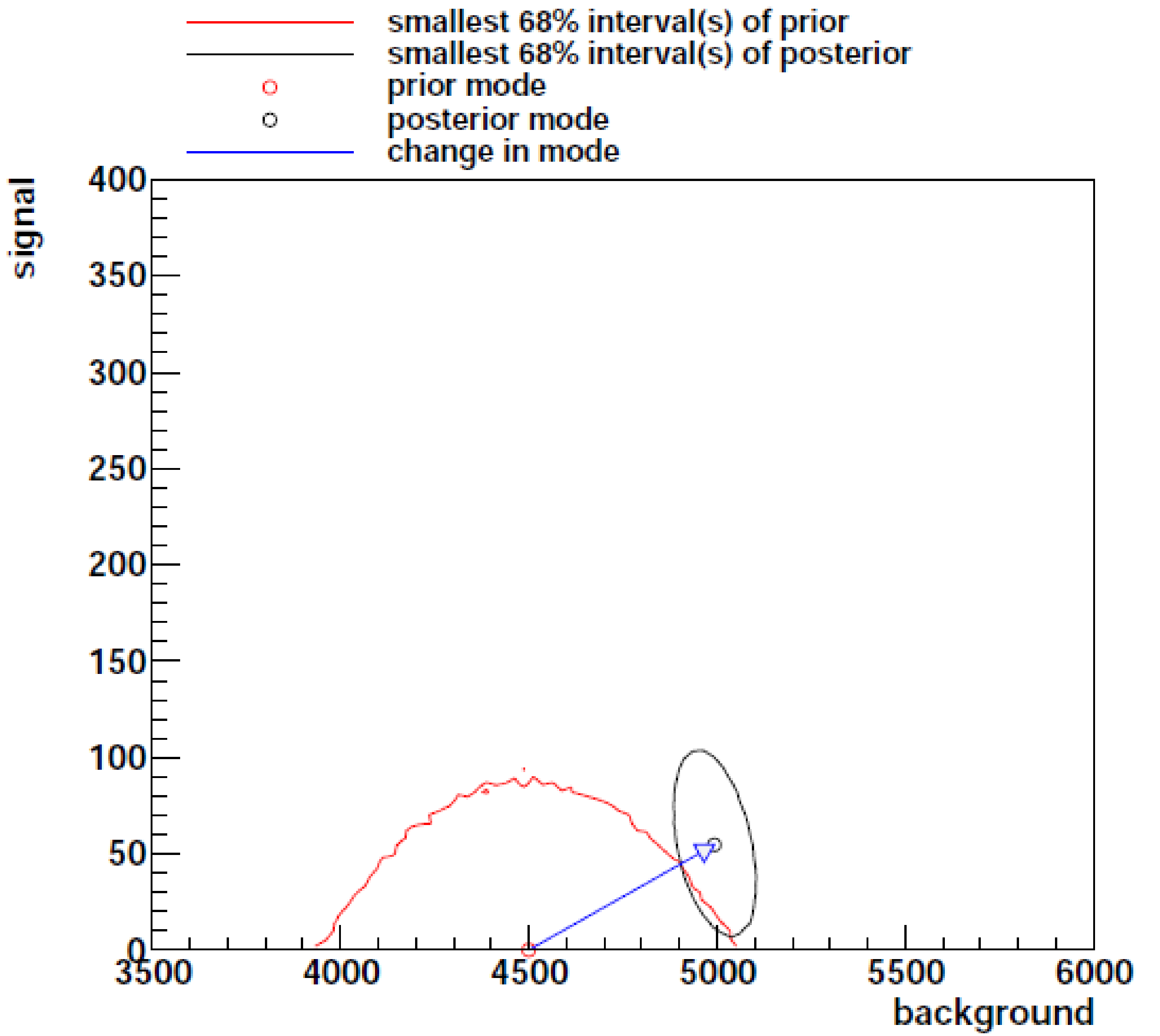
Estimation (1D)



Update of knowledge

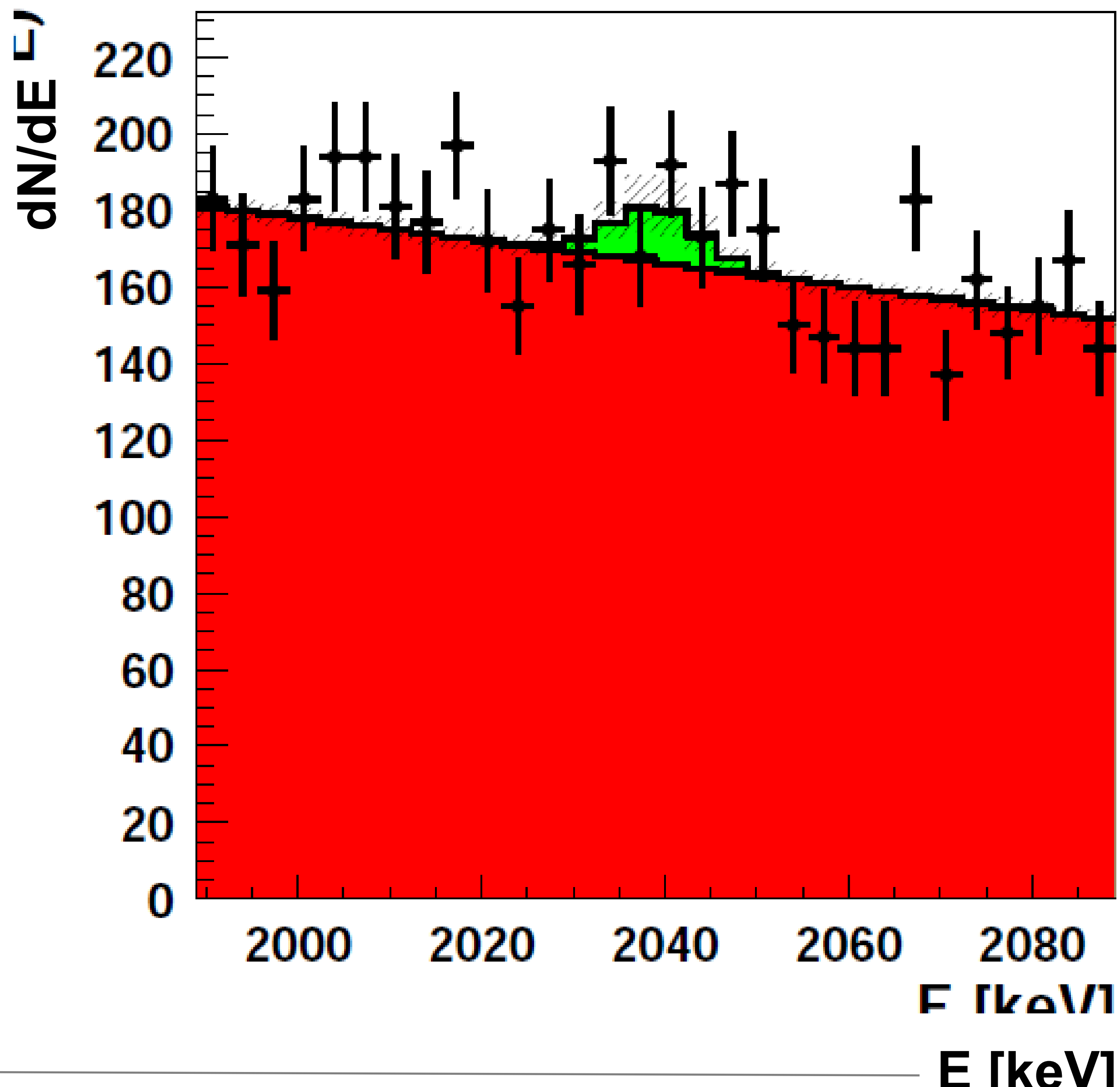






technische universität
dortmund

Fitted spectrum



Syst. uncertainties

Nuisance parameters

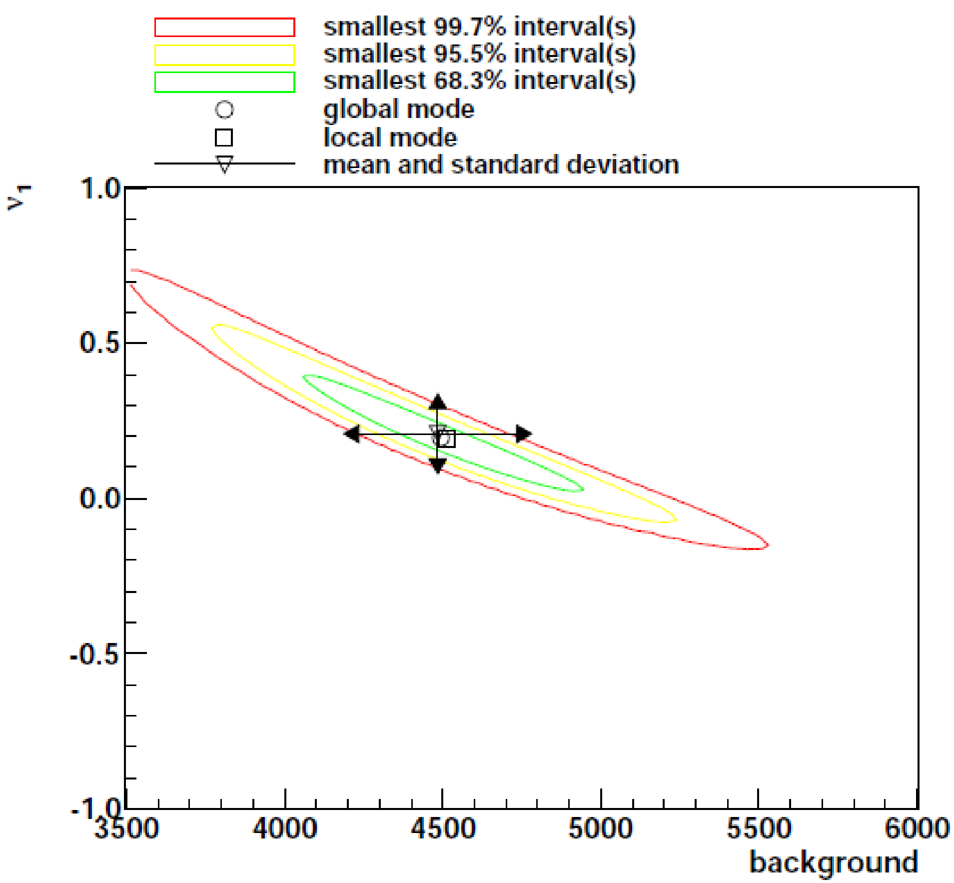
- Model = Physics model (+ par.) × Detector model (+ nuisance par.)
- Associate nuisance parameters to sources of systematic uncertainties, e.g.
  - Collider: Luminosity uncertainty (1 parameter)
  - Calorimeter: jet energy resolution (typically 3 parameters)
  - Reconstructed objects: reconstruction efficiency (*n* parameters)
  - Different physics models ?
  - ...
- Is it justified to use a nuisance parameter? Discrete vs. continuous par.
- Choose appropriate prior (typically Gaussian, sometimes flat)
- Marginalize w.r.t. all nuisance parameters
  - Remove nuisance parameter from the final answer
  - Combine systematic and statistical uncertainties

Syst. uncertainties

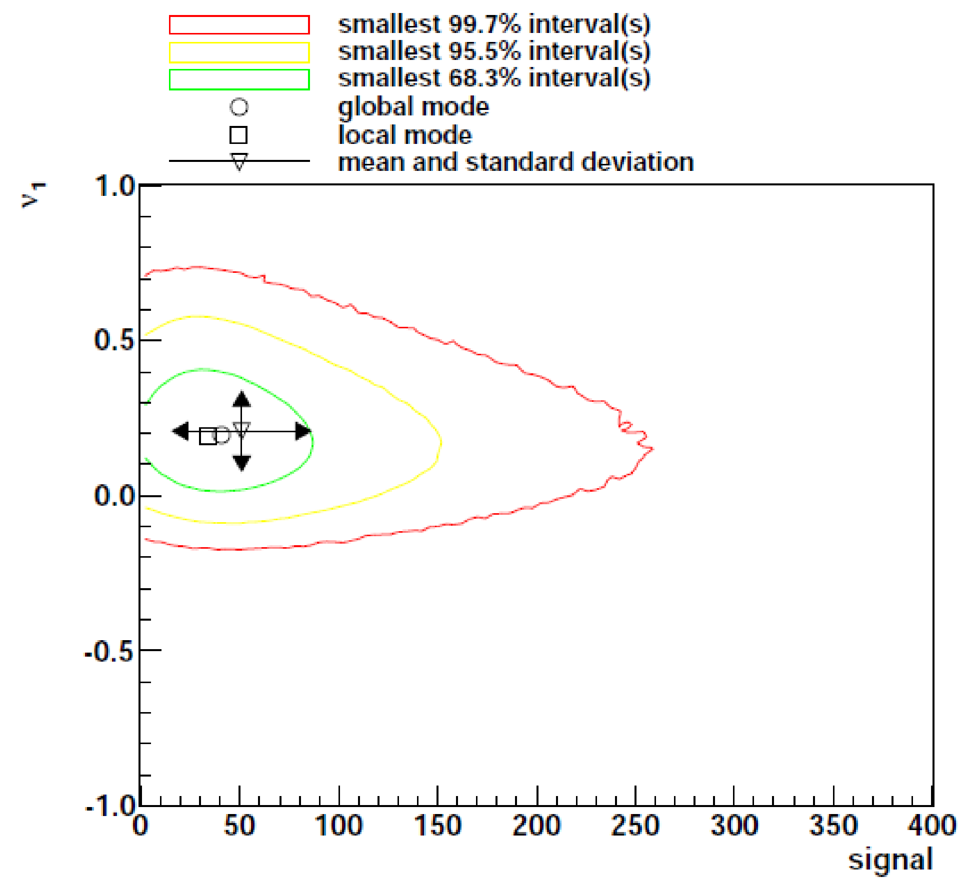
Systematic uncertainties

- Add two uncertainties:
  - Systematic 1: 10% uncertainty on signal and background yield
  - Systematic 2: 60% uncertainty on signal yield
- Priors:
  - Gaussian with mean value of 0 and width of 1 sigma

Systematic 1

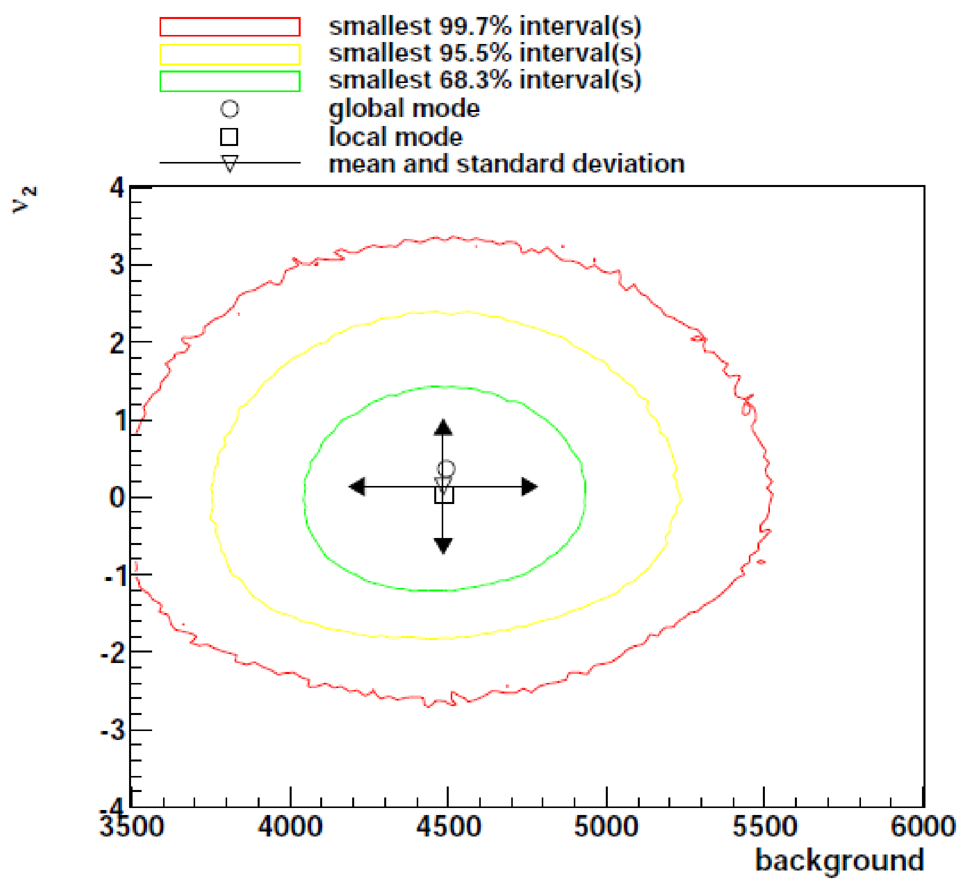


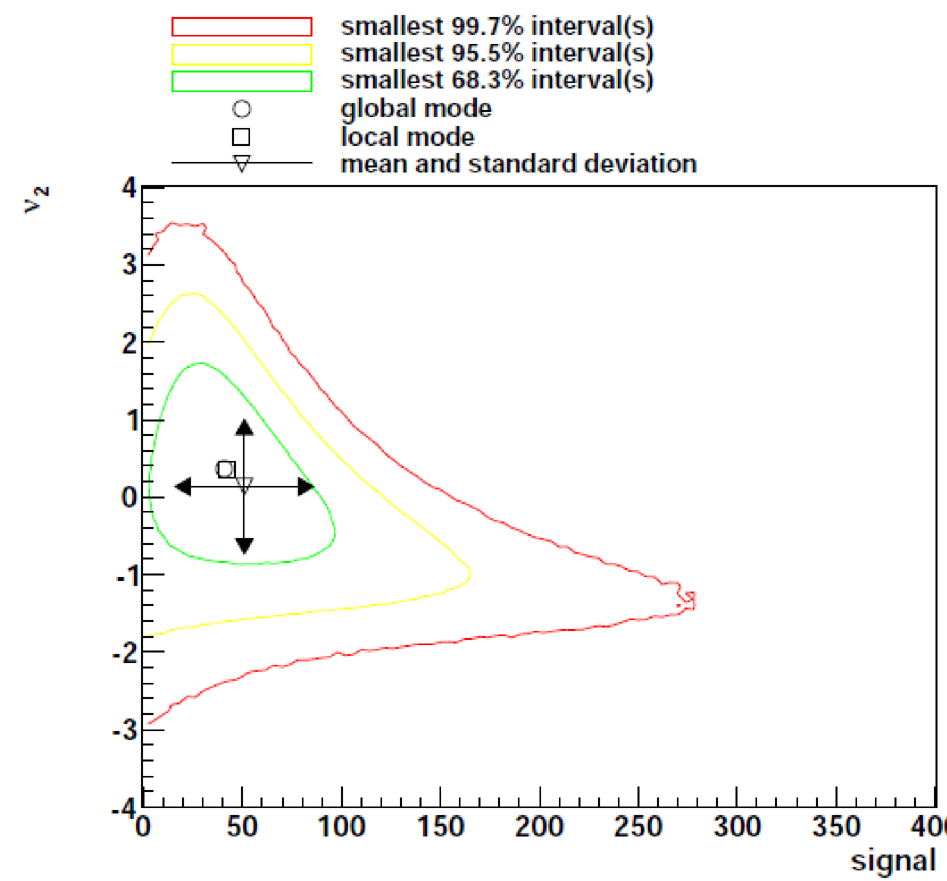
Constrained by background



Medium impact on signal

Systematic 2





Not constrained by background

Large impact on signal

Model comparison

Comparing different models

- Assume you have a full set of models which all describe the data D
- Start with the naïve Bayes ansatz:
  - Assign a probability to all models M_i: $0 \leq P(M_i) \leq 1$
  - If you have all models, then the sum of probabilities is 1: $\sum P(M_i) = 1$
  - Next, use Bayes' theorem to calculate the probability for each model:

$$P(M_i|D) = P(D|M_i) \cdot P(M_i) \,/\, P(D)\ ,$$
$$\text{where } P(D) = \sum P(D|M_i) \cdot P(M_i)$$

  - If the model has free parameters λ, then integrate them out

$$P(M_i|D) = \int p(M_i,\lambda|D)\, d\lambda$$

Model comparison

Comparing different models

- Naïve Bayes ansatz (continues):
  - The last term can be calculated as

$$p(M_i,\lambda|D) = p(D|M_i,\lambda)\cdot P(M_i,\lambda) \,/\, \sum \int p(D|M_i,\lambda)\cdot P(M_i,\lambda)\, d\lambda$$

- Note that there is no distinction between a model and a model with parameters (composite hypothesis)

- Example: A. Caldwell and K. Kröninger, "Signal discovery in sparse spectra: A Bayesian analysis", Phys. Rev. D 74 (2006) 092003

Comparing different models

- Bayes factors:
  - Assume posterior probabilites for two models M_1 and M_2:

$$p(M_i|D) = p(D|M_i) \cdot P(M_i) / (p(D|M_1) \cdot P(M_1) + p(D|M_2) \cdot P(M_2))$$

  - Calculate the posterior odds:

$$p(M_1 | D) / p(M_2 | D) = p(D|M_1) \cdot P(M_1) / p(D|M_2) \cdot P(M_2)$$

  - The ratio

$$B_{12} = p(D|M_1) / p(D|M_2)$$

  is referred to as Bayes factor.

Comparing different models

- Bayes factors (continued):
  - So we find

    "posterior odds" = "Bayes factor" times "prior odds"

  - Traditionally, for a null hypothesis H_0 and an alternative hypothesis H_1, the ratio B_{10} gives evidence against the null hypothesis. Large values of B_{10} are an indication that the hypothesis H_0 is wrong in favour of H_1
  - Bayes factors do not rely on the priors.
  - If the models do not have any free parameters, then the Bayes factors are likelihood ratios. Otherwise, the parameters have to be integrated out.

Comparing different models

- Bayes factors (continued):
  - Rough scales:

$$B_{12} = p(D|M_1) / p(D|M_2)$$

B_{10}	Evidence against H_0
< 1	none
1 – 3.2	Not worth mentioning
3.2 - 10	substantial
10 - 100	strong
> 100	decisive

  - Of course, somebody just made that up. Depends on application.

Model comparison

Occam's razor

- Principle that an explaination should not be unnecessarily complicated, i.e. chose the simpler models of the two if both describe the data reasonably well.
- Intuitively clear, although the more complex model could still be true.
- Bayesian reasoning includes Occam's razor: the prior probabilities for complex models are typically smaller than for complex ones, and thus are the posterior probabilities.
- Assume a simple model M_1 and a complex one M_2 with constant priors for their parameters

$$p(\lambda_i) = 1 / c_i \; (c_i > 1),$$

and equal prior probabilities for the models themselves, i.e.

$$p(M_1) = p(M_2) = 0.5.$$

Occam's razor

- Then we find

$$p(\vec{\lambda}, M_1|\vec{D}) = \frac{p(\vec{D}|\vec{\lambda}, M_1)\cdot p(\vec{\lambda}, M_1)}{\int \mathrm{d}\vec{\lambda}\, p(\vec{D}|\vec{\lambda}, M_1)\cdot p(\vec{\lambda}, M_1) + \int \mathrm{d}\vec{\lambda}\, p(\vec{D}|\vec{\lambda}, M_2)\cdot p(\vec{\lambda}, M_2)}$$

$$= \frac{p(\vec{D}|\vec{\lambda}, M_1)\cdot 0.5\cdot \prod_{i=1}^{N_1} 1/c_i}{\int \mathrm{d}\vec{\lambda}\, p(\vec{D}|\vec{\lambda}, M_1)\cdot 0.5\cdot \prod_{i=1}^{N_1} 1/c_i + \int \mathrm{d}\vec{\lambda}\, p(\vec{D}|\vec{\lambda}, M_2)\cdot 0.5\cdot \prod_{i=1}^{N_2} 1/c_i}$$

- So the posterior probability for M_1 will increase, the more parameters we add.

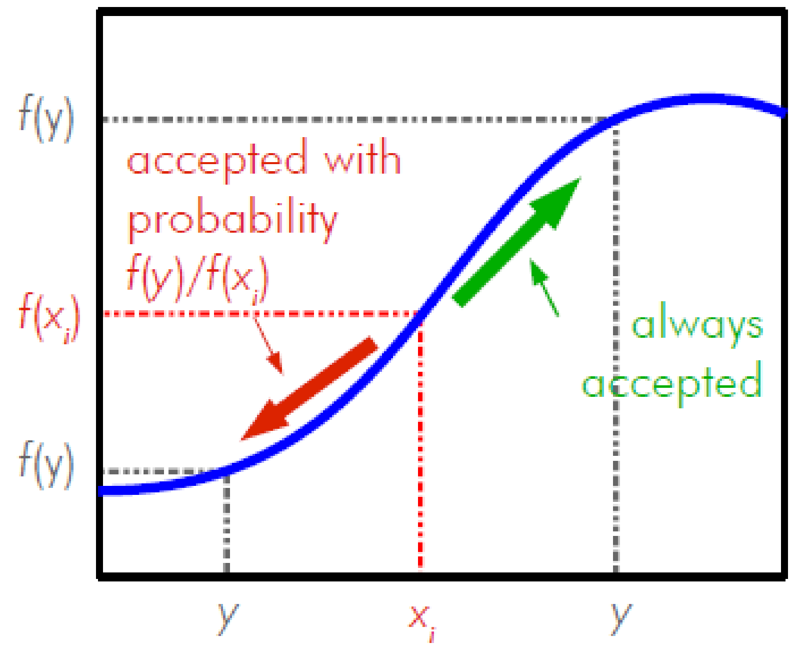
Computational steps

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 of 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



- 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

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

MCMC for Bayesian inference

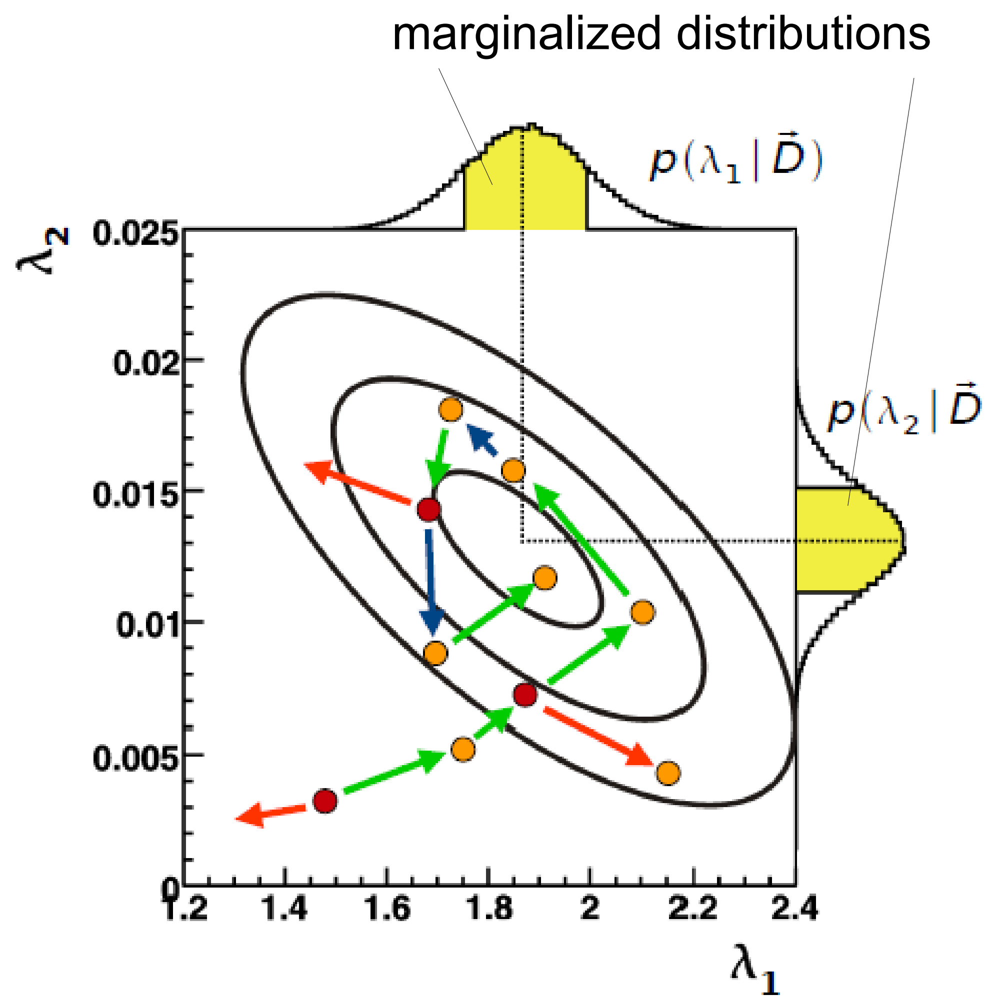
- Use MCMC to sample the posterior probability, i.e.

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

- 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

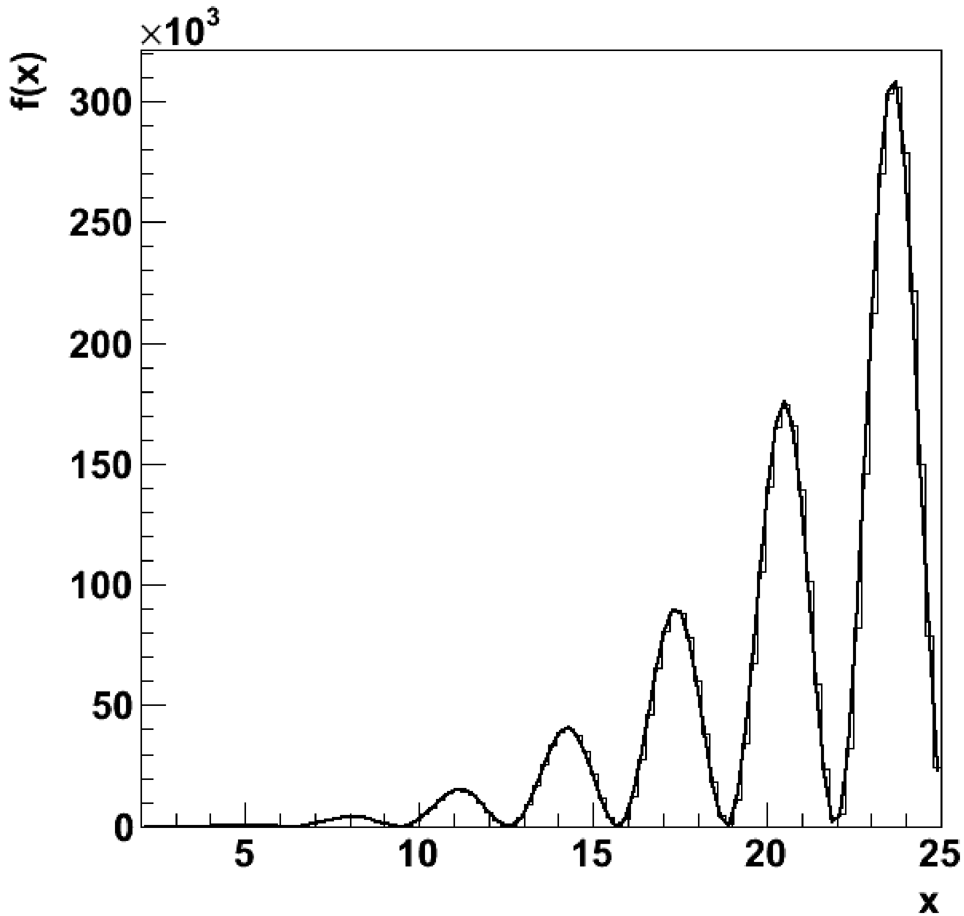


Does it work?

- Test MCMC on a function:

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

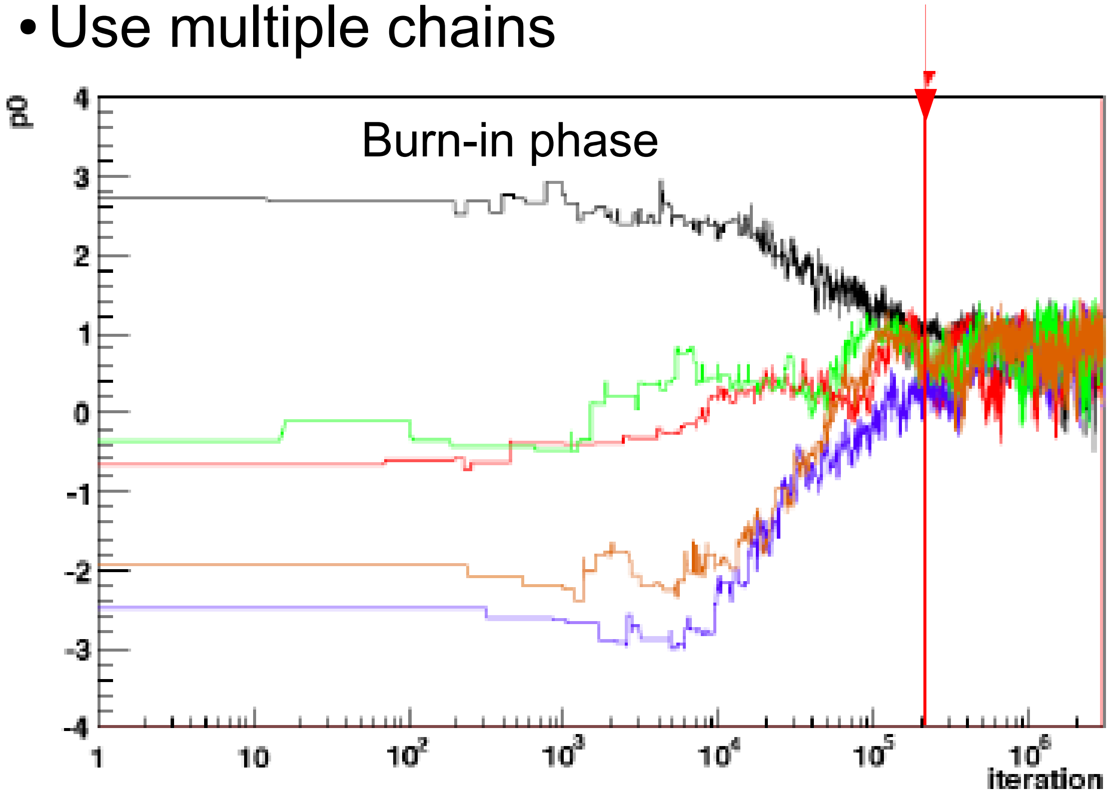
- Compare MCMC distribution to analytic function
- Several minima/maxima are no problem.
- Different orders of magnitude are no problem.
- But: need to make sure that these chains converge towards the true distribution



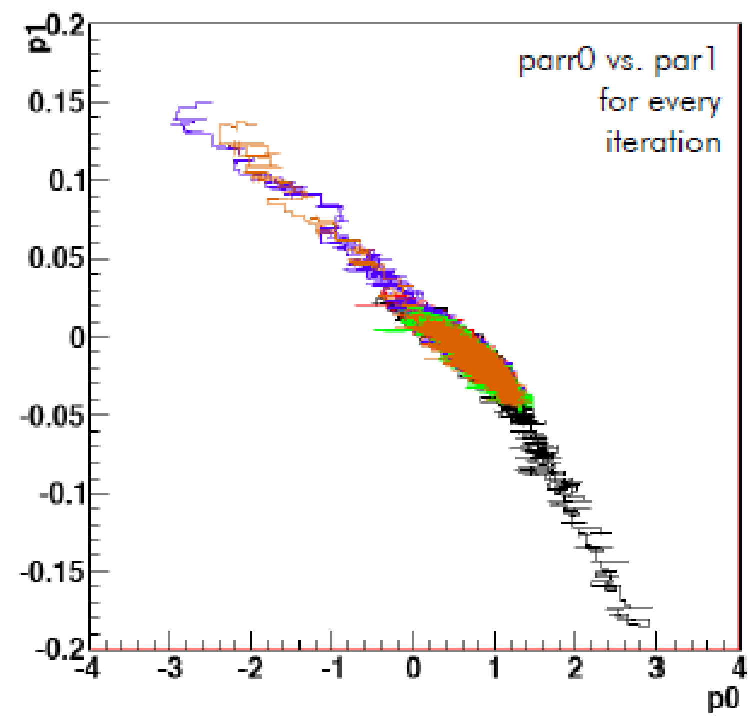
Convergence

- 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:

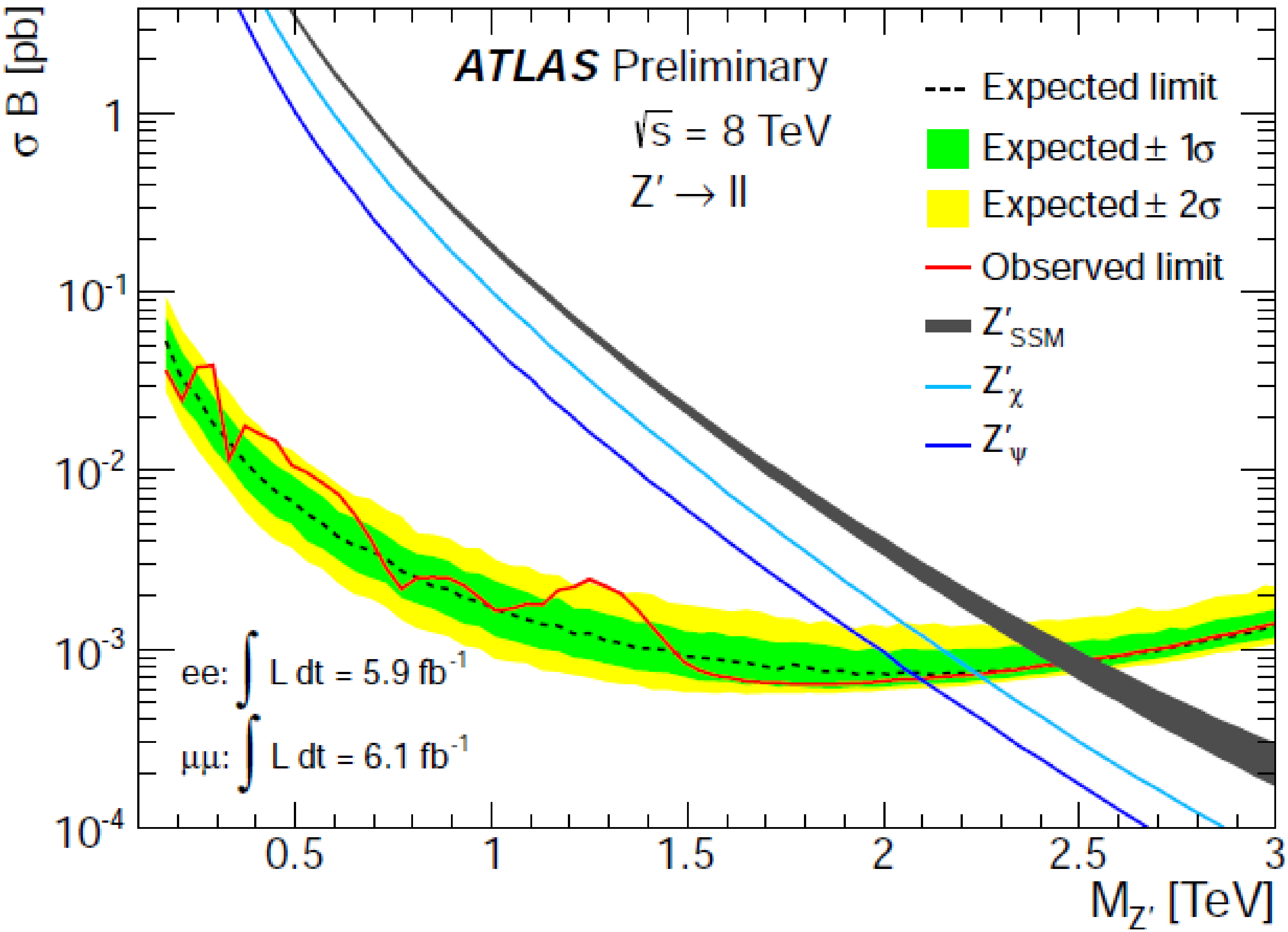
  A. Caldwell, D. Kollar, K. Kröninger, BAT - The Bayesian Analysis Toolkit
  Comp. Phys. Comm. 180 (2009) 2197-2209 [arXiv:0808.2552].

BAYESIAN ANALYSIS TOOLKIT

https://github.com/bat/bat

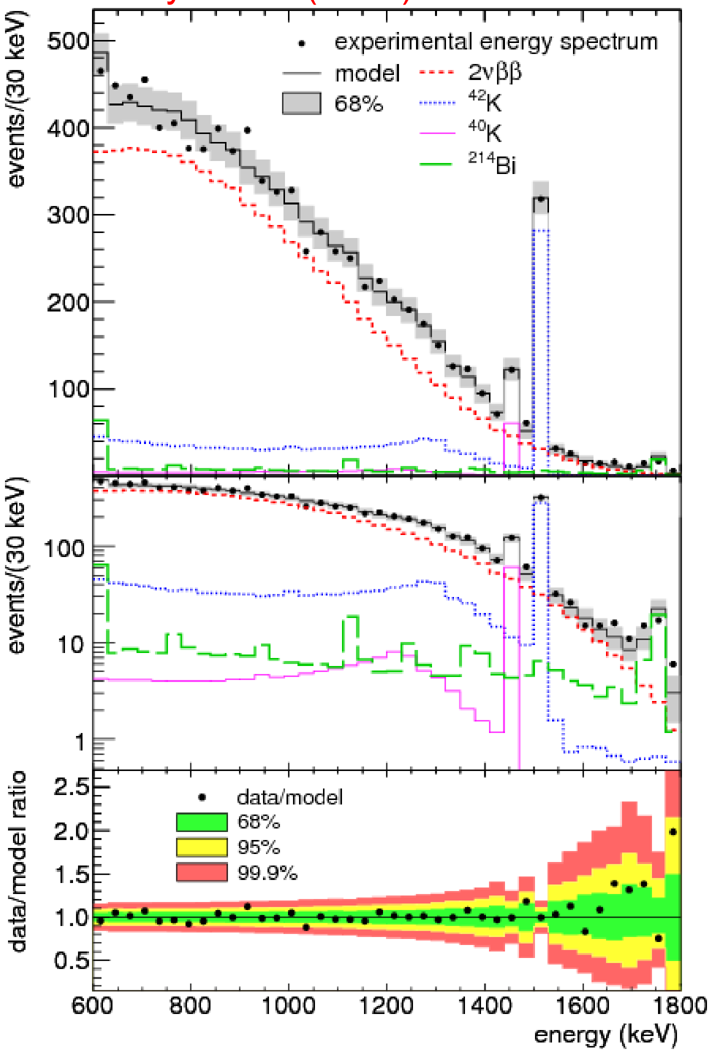
HEP examples

Dilepton resonances
ATLAS-CONF-2012-129

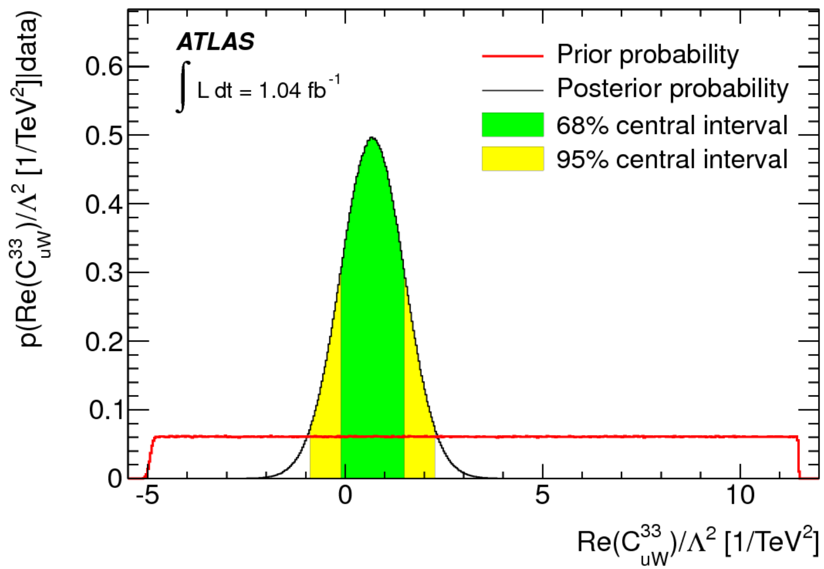


HEP examples

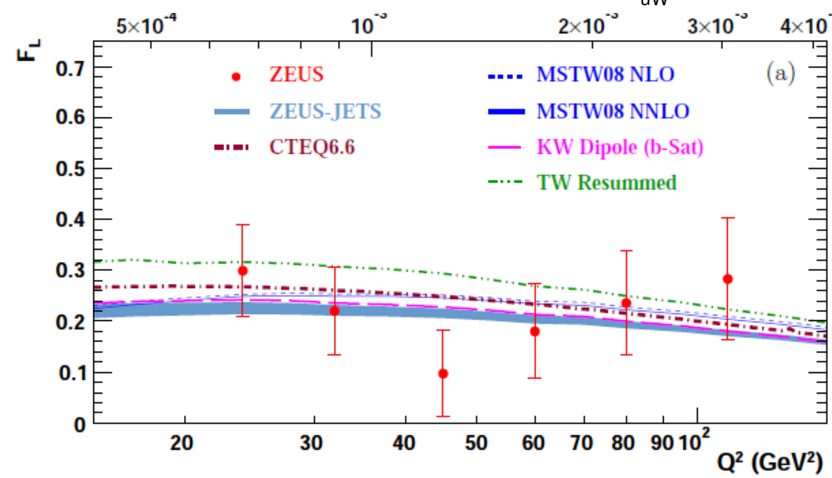
Classical template fit
J.Phys. G40 (2013) 035110



Anomalous couplings
JHEP 1206 (2012) 088



Longitudinal structure function (48-dim. fit)
Phys.Lett. B682 (2009) 8

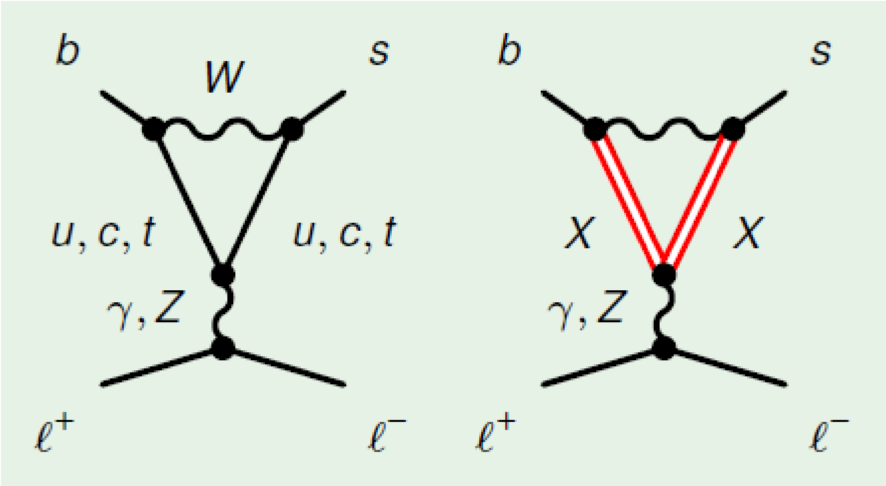


A complex example

Rare b-meson decays

Frederic Beaujean *et al.*
JHEP 08(2012)030

- Tree-level FCNC forbidden in the SM
- Effective field theory: add eff. operators to Lagrangian
- Similar to Fermi's four-point interaction
- Physics case: search for non-SM contributions
- Model parameters:
  - 3 Wilson coefficients
  - 25 nuisance parameters
- Input:
  - 59 measurements from BaBar, Belle, CDF, LHCb
  - Theory calculations, quark masses, CKM parameters, ...
- Numerically difficult ~ impossible with MCMC



A complex example

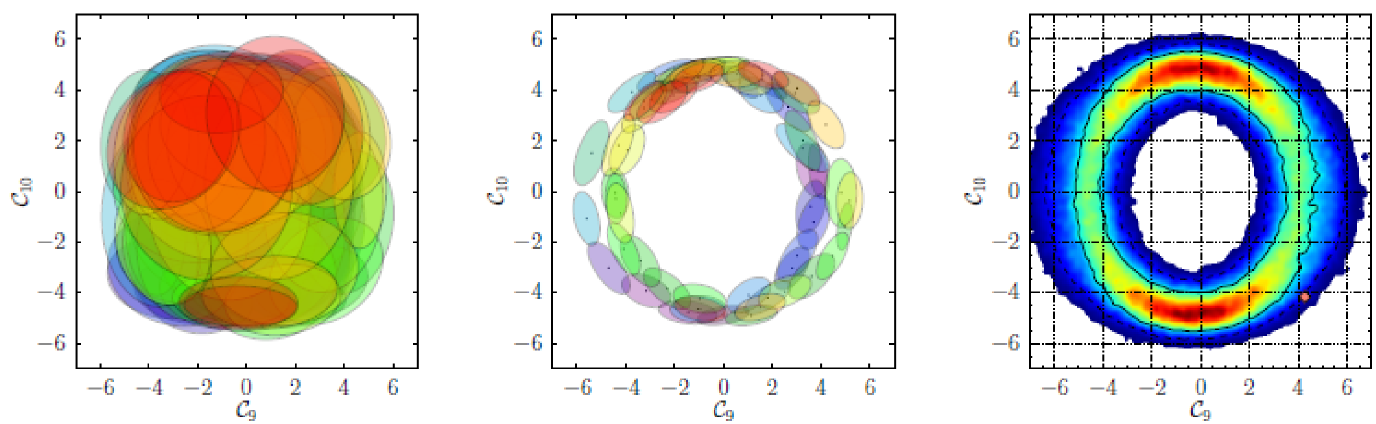
Rare b-meson decays

Frederic Beaujean *et al.*
JHEP 08(2012)030

- Use MCMC plus population MC

	Pro	Con
MCMC	local exploration, learns on the fly	trapped in local maxima
PMC	massive parallelization, yields normalization, multiple modes OK	very sensitive to initialization

- Dominant contribution: $BR(B \rightarrow Kll) \propto |C_9{}^2| + |C_{10}{}^2|$



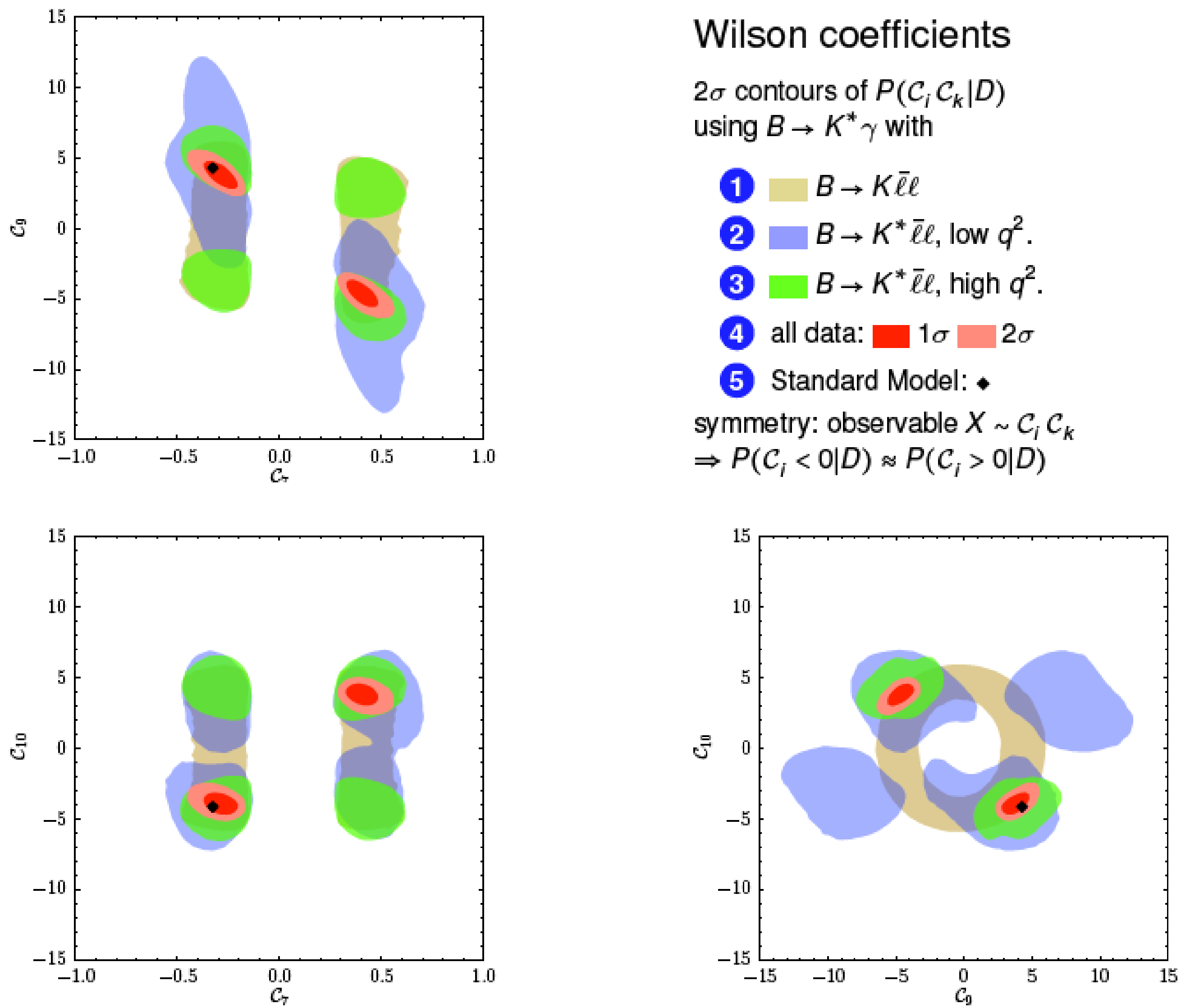
Initial proposal from chains — Final proposal, 10 PMC updates — Smoothed posterior

A complex example

Rare b-meson decays

- Posterior probability well sampled
- No hint for new physics
- Showed necessity to implement new numerical algorithms
- Example:
  - F. Beaujean *et al.*, *Initializing adaptive importance sampling with Markov chains*, arXiv:1304.7808

Frederic Beaujean *et al.*
JHEP 08(2012)030



Summary

- Knowledge is justified belief
- Bayesian probability is degree-of-belief
- Bayes' theorem allows easy update of knowledge
- Everything else is about math and numerical methods:
  - Parameter (point and interval) estimation
  - Treatment of systematic uncertainties
  - Calculation of marginalized distributions
  - Also: model comparison and goodness-of-fit (not covered)
- Numerical methods necessary for complex fit with a large number of parameters