GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Monte Carlo methods in experimental HEP

Kevin Kröninger – University of Göttingen

MCNet School, Maria Spring, August 5th 2013

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

Overview

Monte Carlo methods

- In general: performing random experiments and analyzing their output
- Use-cases:
  - Optimization / Integration
  - Sampling from a pdf
- Why?
  - Complicated problems
  - No closed-form solution
- In HEP:
  - Jargon: “Monte Carlo” means “generation of simulated events”
  - Differential cross-section = pdf
  - One event = one phase-space point
  - Industry of Monte Carlo generators, phase-space integrators, etc.
  - What else?

Use in (experimental) HEP

- Sampling: generate Monte Carlo events
- Statistics: test properties of an estimator, e.g. bias, pull
- Design of experiments: estimate expected sensitivity of an experiment, e.g. ILC TDR
- Correlation of measurements and uncertainties, e.g. top-mass combination
- Hypothesis testing: distribution of test statistics, p-values, type-I/II errors
- Plain math: propagation of uncertainty, e.g., ratio of two random numbers
- Model comparison: calculating the evidence of a model
- Sampling large-dimensional spaces, e.g. complex fitting in Higgs/top-sector, CKM matrix element fitting, Markov Chain Monte Carlo

All applications based on **statistics** and **Monte Carlo methods**

Overview

Analysis flow

Design phase:
- Exp. sensitivity (ensemble tests)
- Monte Carlo prod.
- Properties of estimators

↓

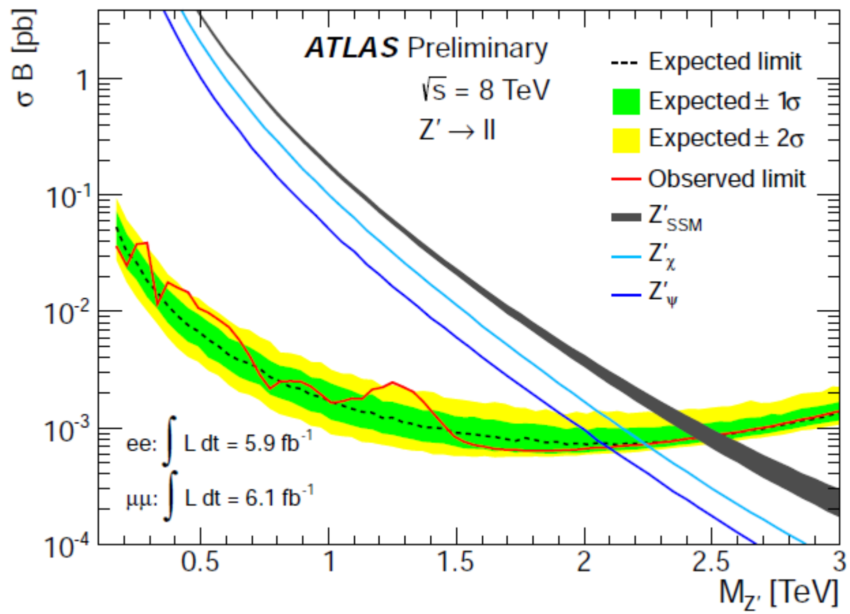
Data analysis:
- Complex fitting (sampling)
- Hypothesis testing

↓

Post-fit analysis:
- Goodness-of-fit (ensemble tests)

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

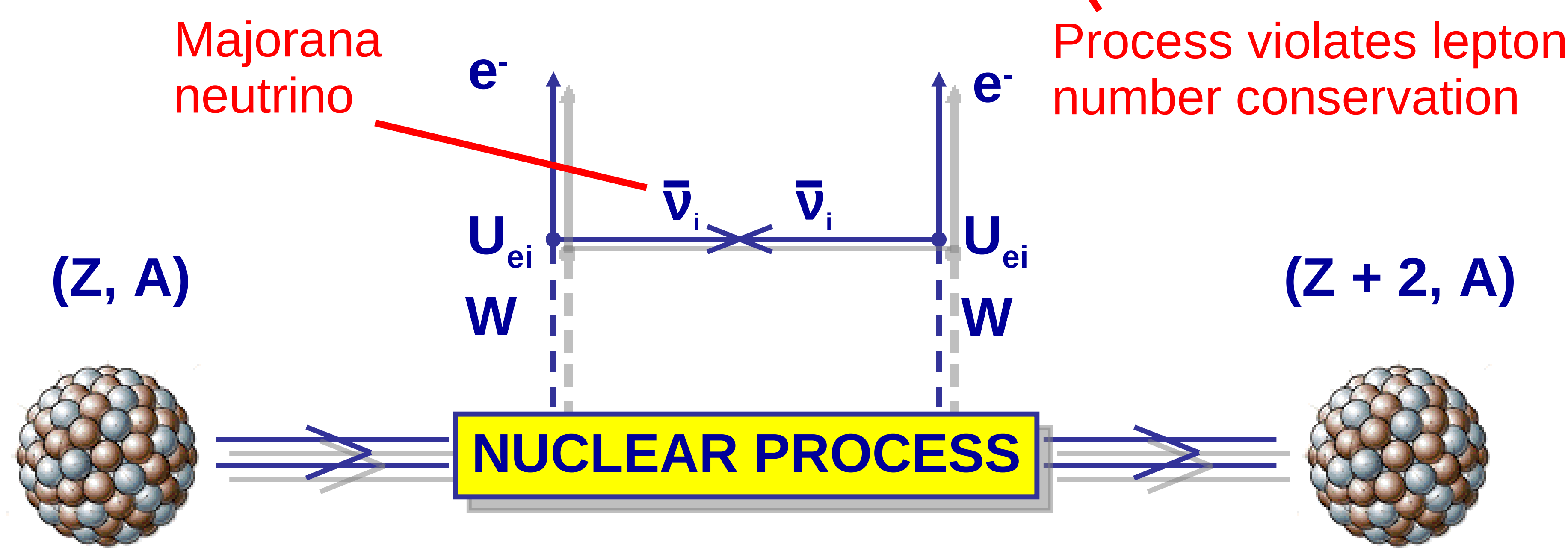
- A concrete example
- Scientific reasoning
- Probability and the Bayesian interpretation
- Parameter estimation
- Summary

Neutrinoless double beta-decay

Rare nuclear transition (2^{nd} order weak process):

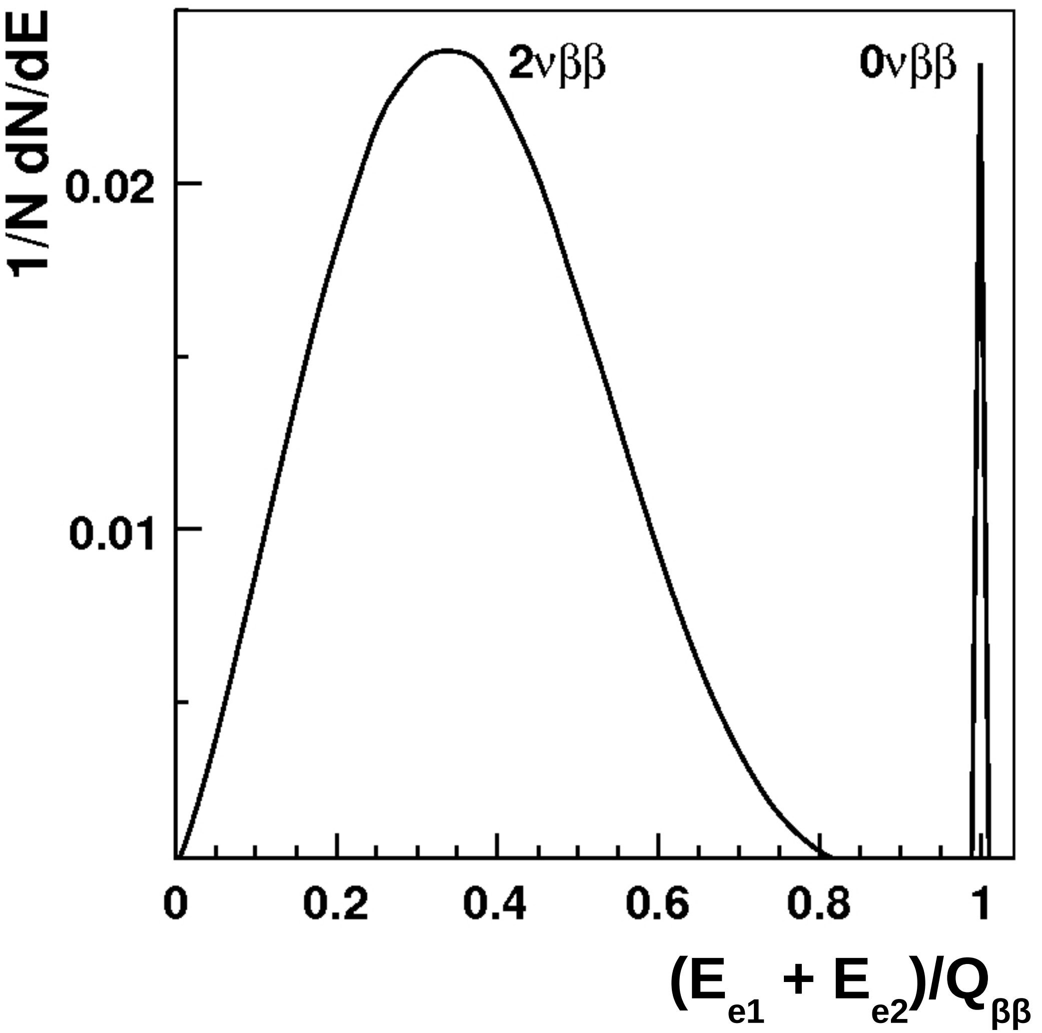
$2\nu\beta\beta$: $(Z, A) \rightarrow (Z+2, A) + 2\,e^- + 2\bar{\nu}$ $\quad \Delta L = 0 \quad (T_{1/2} \sim 10^{21}\ y)$

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



Neutrinoless double beta-decay

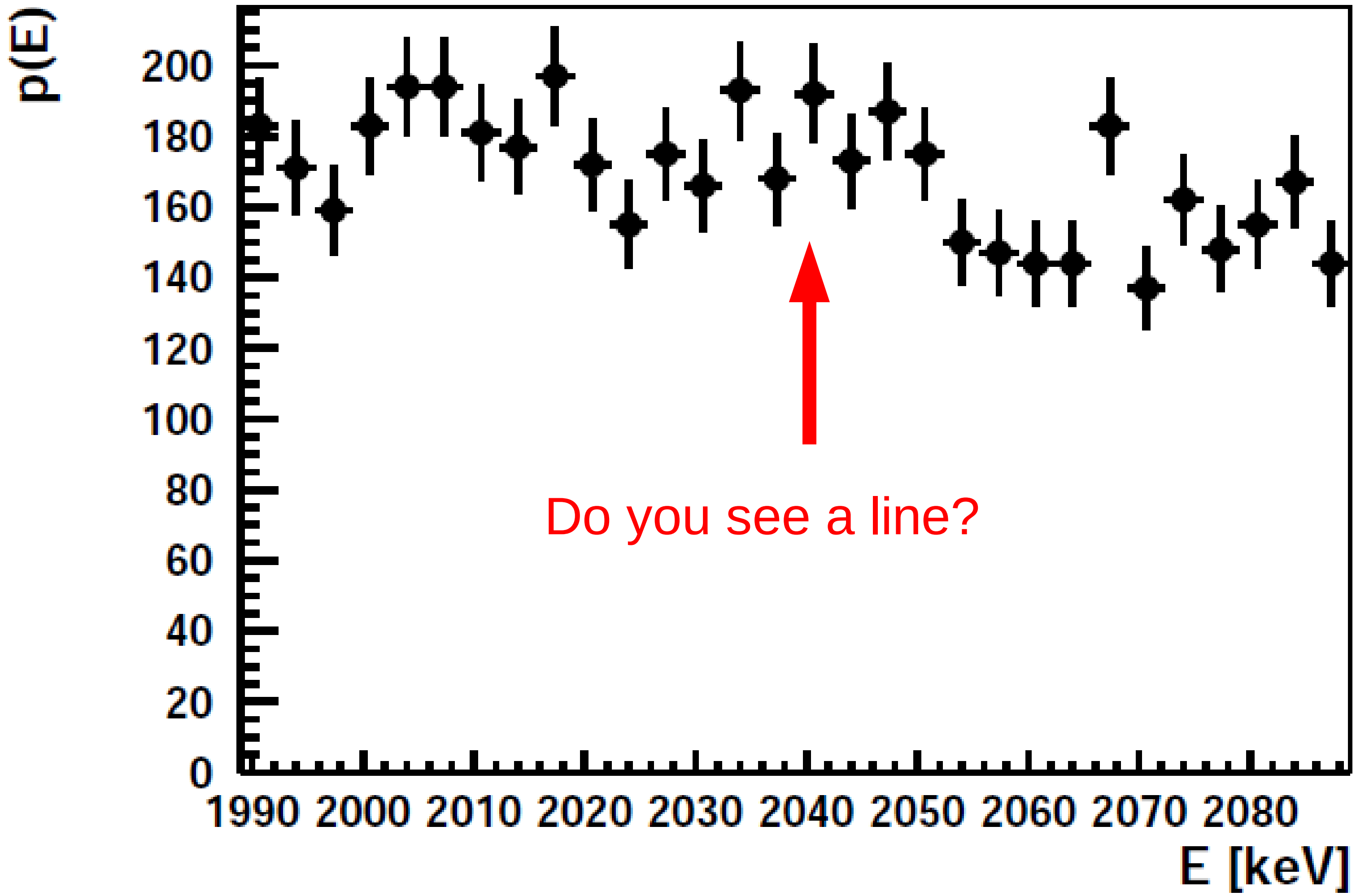
Searching for a single peak on top of an (almost flat) background...



A concrete example

Neutrinoless double beta-decay

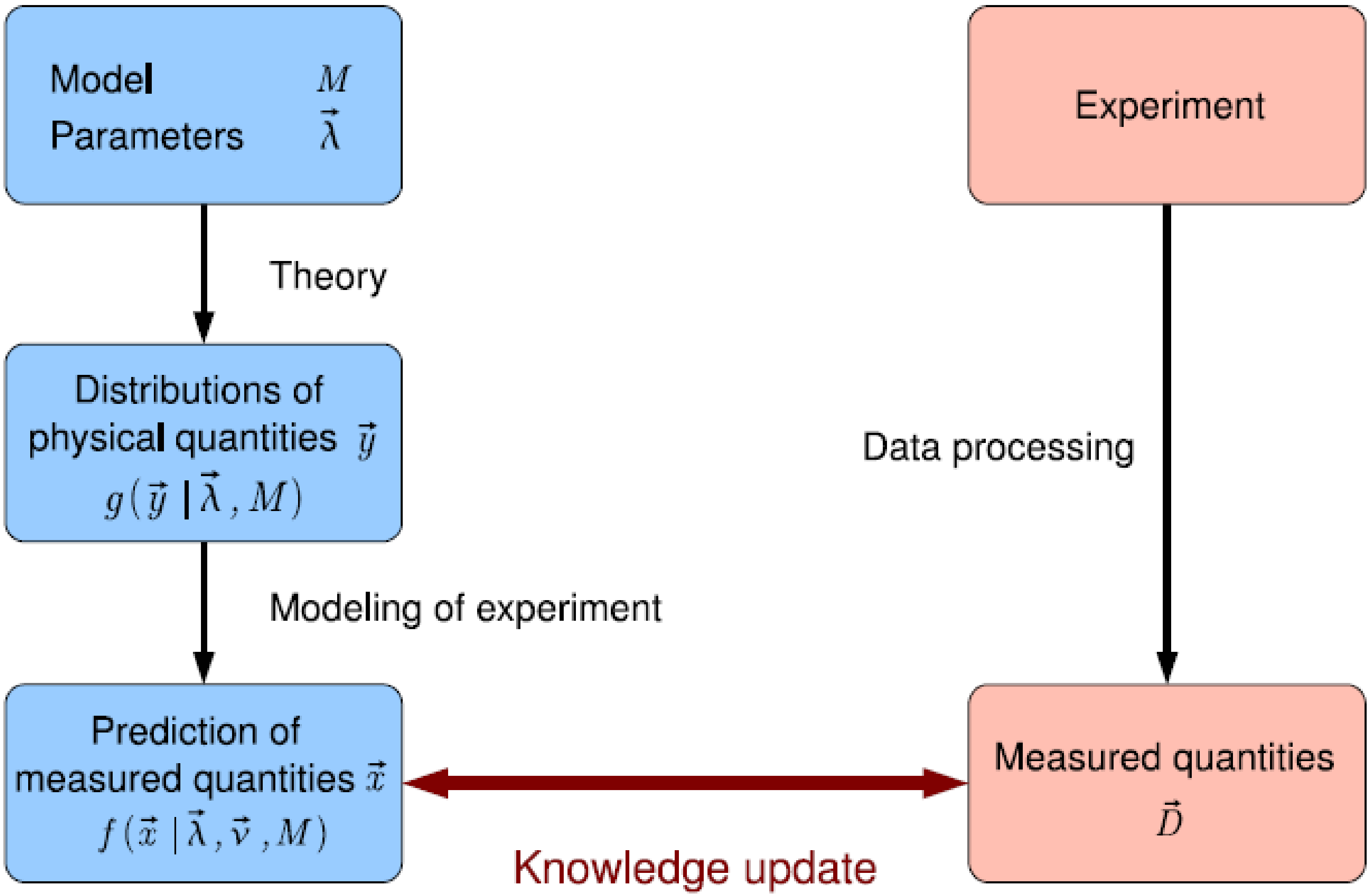
Searching for a single peak on top of an (almost flat) background...



Major subjects of data analysis

- Model comparison: Which model describes the data best?
  - SM background only?
  - Does 0υββ + background describe the data better?
- Parameter estimation: Given a model, what are the values of its free parameters?
  - What is the rate of 0υββ?
  - 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?
  - If there is a signal, does the 0υββ + background model describe the data?

Overview



Model M
Parameters $\vec{\lambda}$
Theory
Distributions of physical quantities $\vec{y}$
$g(\vec{y} | \vec{\lambda}, M)$
Modeling of experiment
Prediction of measured quantities $\vec{x}$
$f(\vec{x} | \vec{\lambda}, \vec{\nu}, M)$
Experiment
Data processing
Measured quantities
$\vec{D}$
Knowledge update

Overview

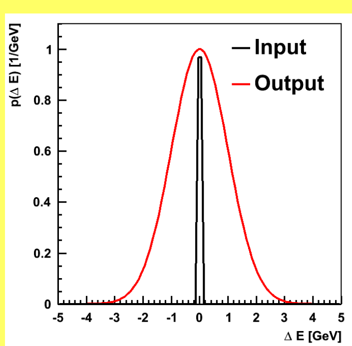
M: 0υββ

λ: signal strength *s*, background level *b*

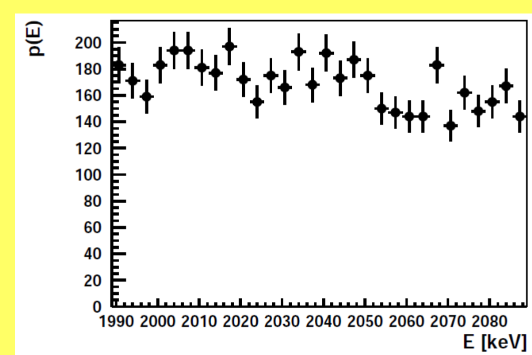
Theory

$$y(E_{e_1}+E_{e_2}|s,Q_{\beta\beta}) = b+s\cdot\delta(E_{e_1}+E_{e_2}-Q_{\beta\beta})$$

ing of experiment



Data processing



Knowledge update

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:* $0\nu\beta\beta$ + background model predicts a certain energy spectrum

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!

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.
- Can not agree: we can not know the truth, so:

Knowledge is justified belief

- 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, ...

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
  - Use data to gain (inductively) 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*
  - For each subset *A*, assign probability *P(A)* between 0 and 1
  - Probability *P(S)* = 1
  - 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

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

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 a x% probability”

Interpretation and criticism

- Prior can come from
  - personal degree-of-belief,
  - theoretical considerations,
  - auxiliary measurements, ...
- Elegant update of knowledge: posterior of one experiment can be prior of another experiment. Natural way to combine measurements.
- Subjective? Yes, but made explicit
- Objective Bayesian movement, try to find objective priors (reference priors)
- Prior depends on parametrization (lifetime vs. decay constant), Jeffreys prior
- Choice of (initial) prior should not play a strong role.
- Difficult to formulate a single prior for a collaboration of ~3.000 people

Point and interval estimation

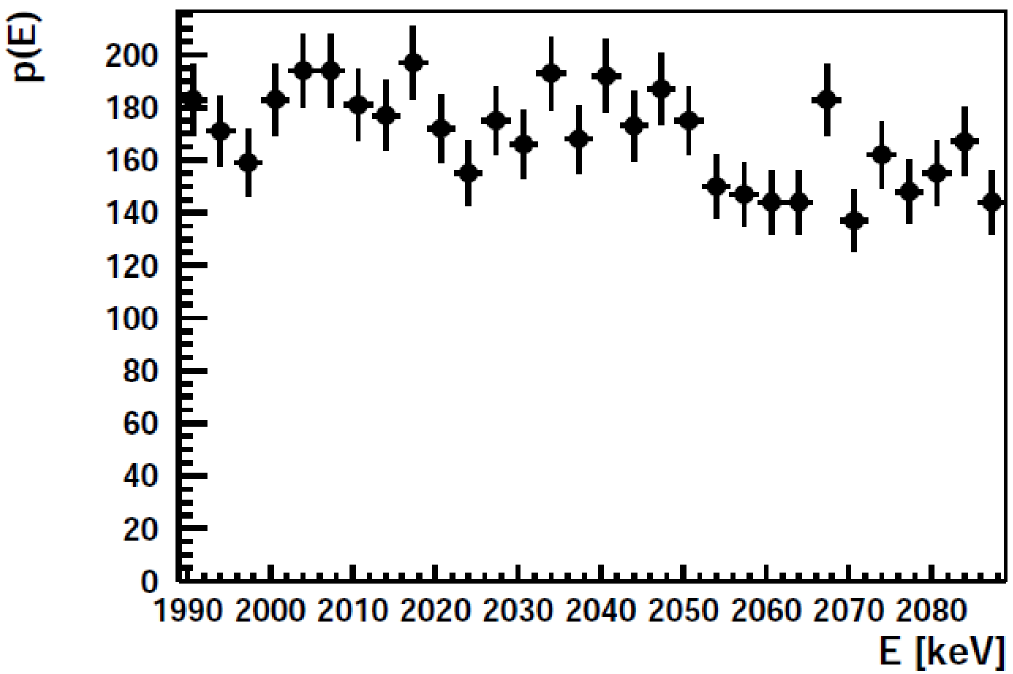
- Full solution: posterior probability
- Summarize posterior using point and interval estimates
- Common point estimators:
  - Maximum posterior probability (global mode)
  - Maximum of marginalized probability (local mode)
  - Mean value of marginalized probability
  - Median of marginalized probability

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

- Common interval estimates:
  - Smallest (set of) interval(s) covering 68% probability
  - 16% - 84% quantile
  - Standard deviation of marginalized posterior
  - Upper (lower) limits: 99%, 95%, 90% (1%, 5%, 10%) quantiles
  - Choose such that point estimator lies inside estimated interval!

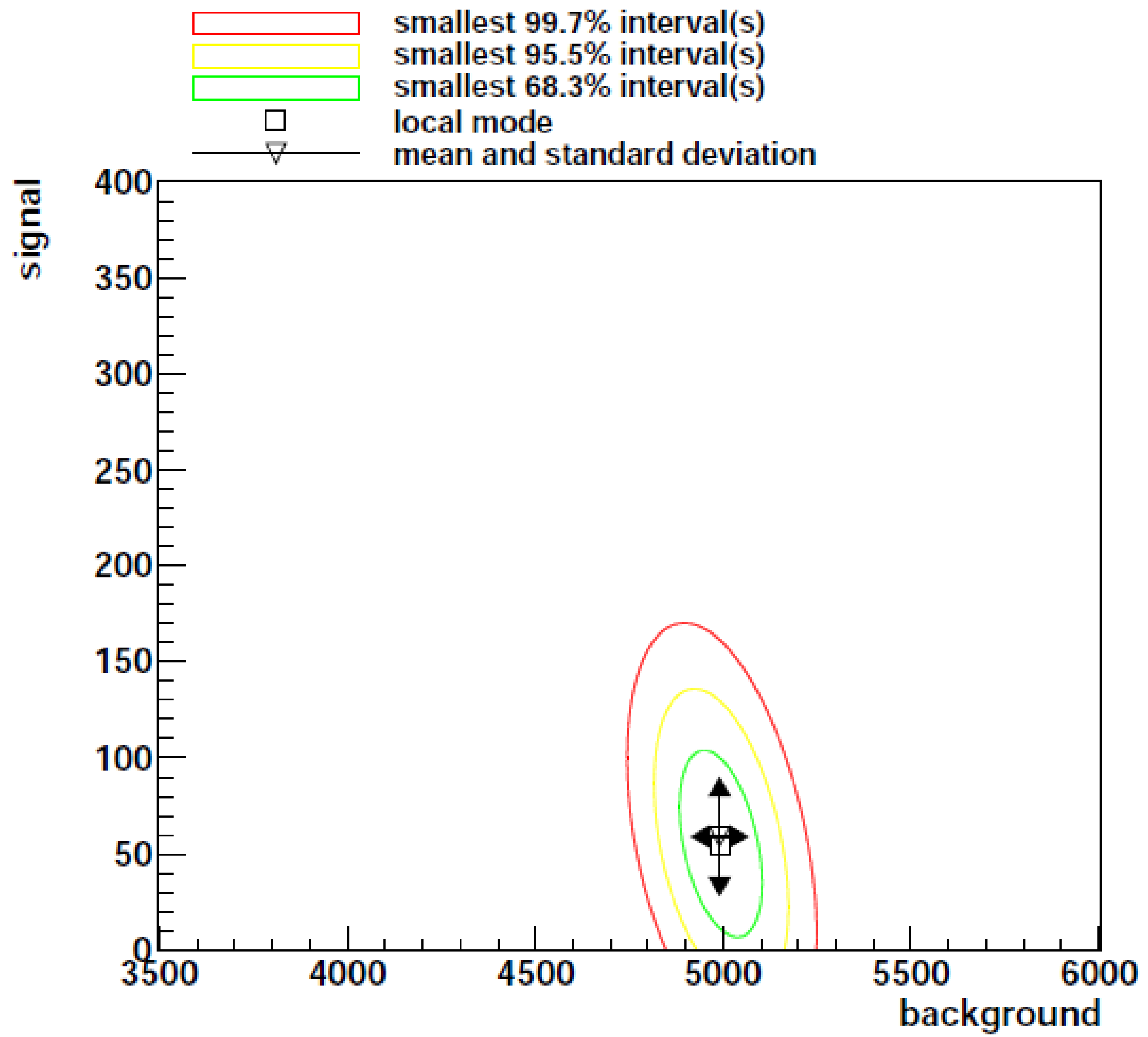
A concrete example

Concrete model

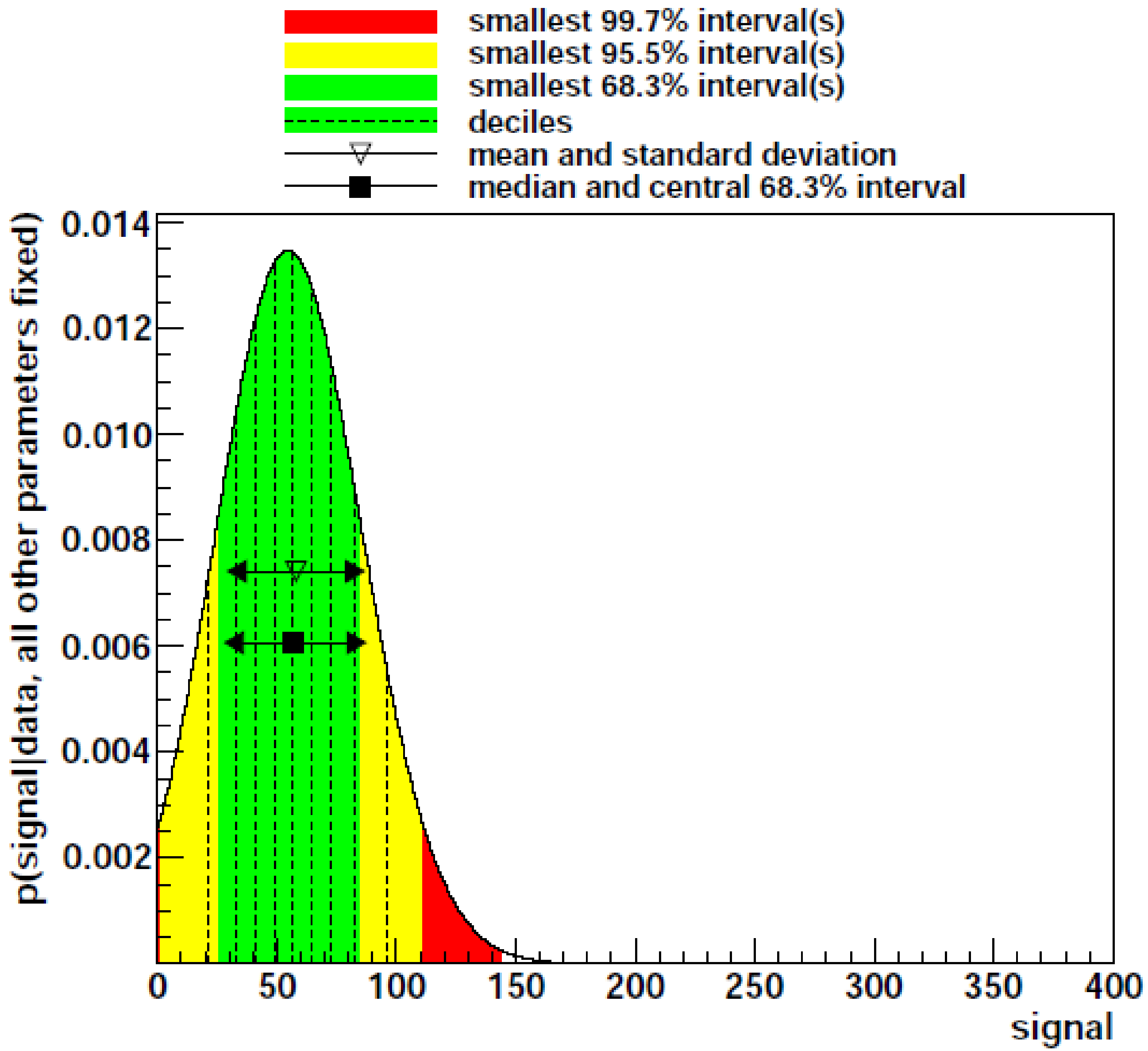


- Data:
  - Binned, number of events
- Shapes:
  - Background linearly decreasing
  - Signal: Gaussian at fixed position
- Statistical model:
  - Independent Poisson fluctuations
  - Parameter 1: background strength, Gaussian prior
  - Parameter 2: signal strength, exponentially decreasing prior
- Fit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

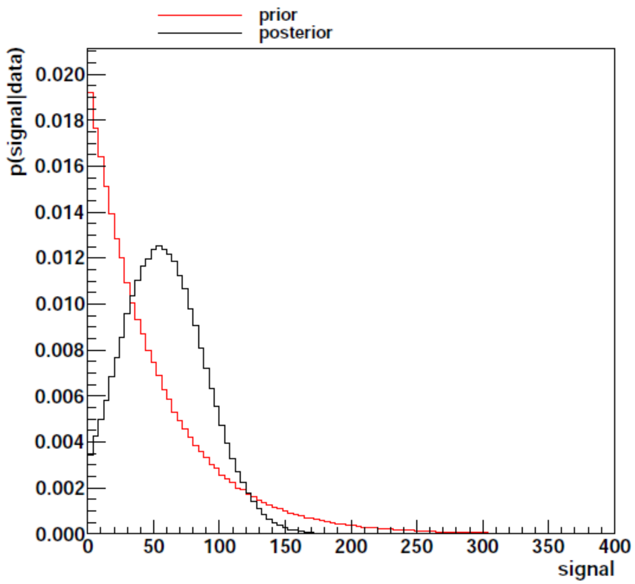
Point and interval estimates (2D)

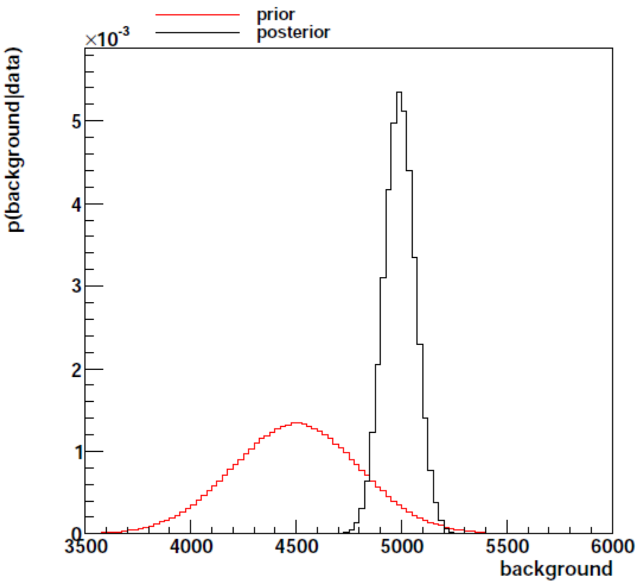


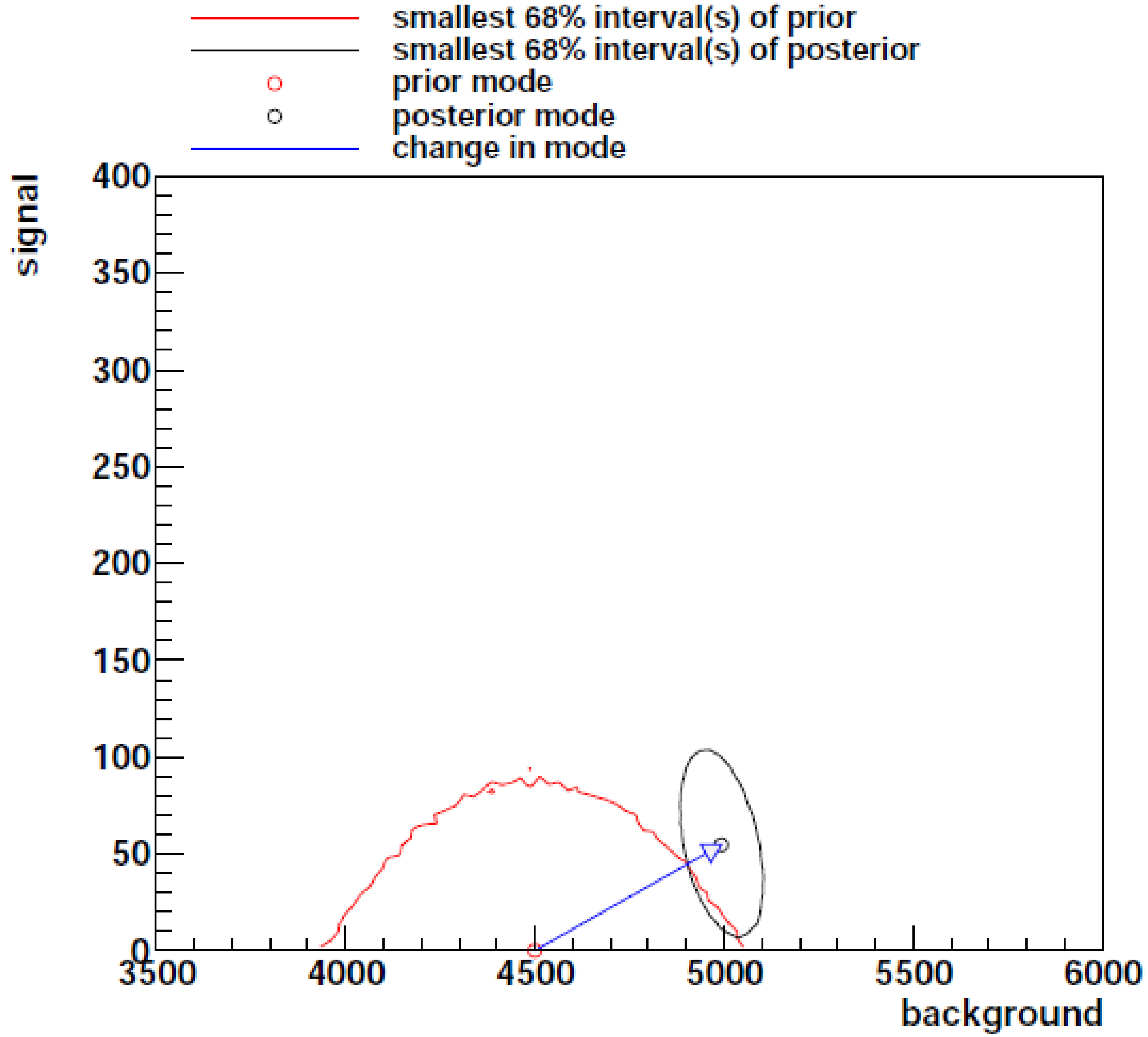
Point and interval estimates (1D)



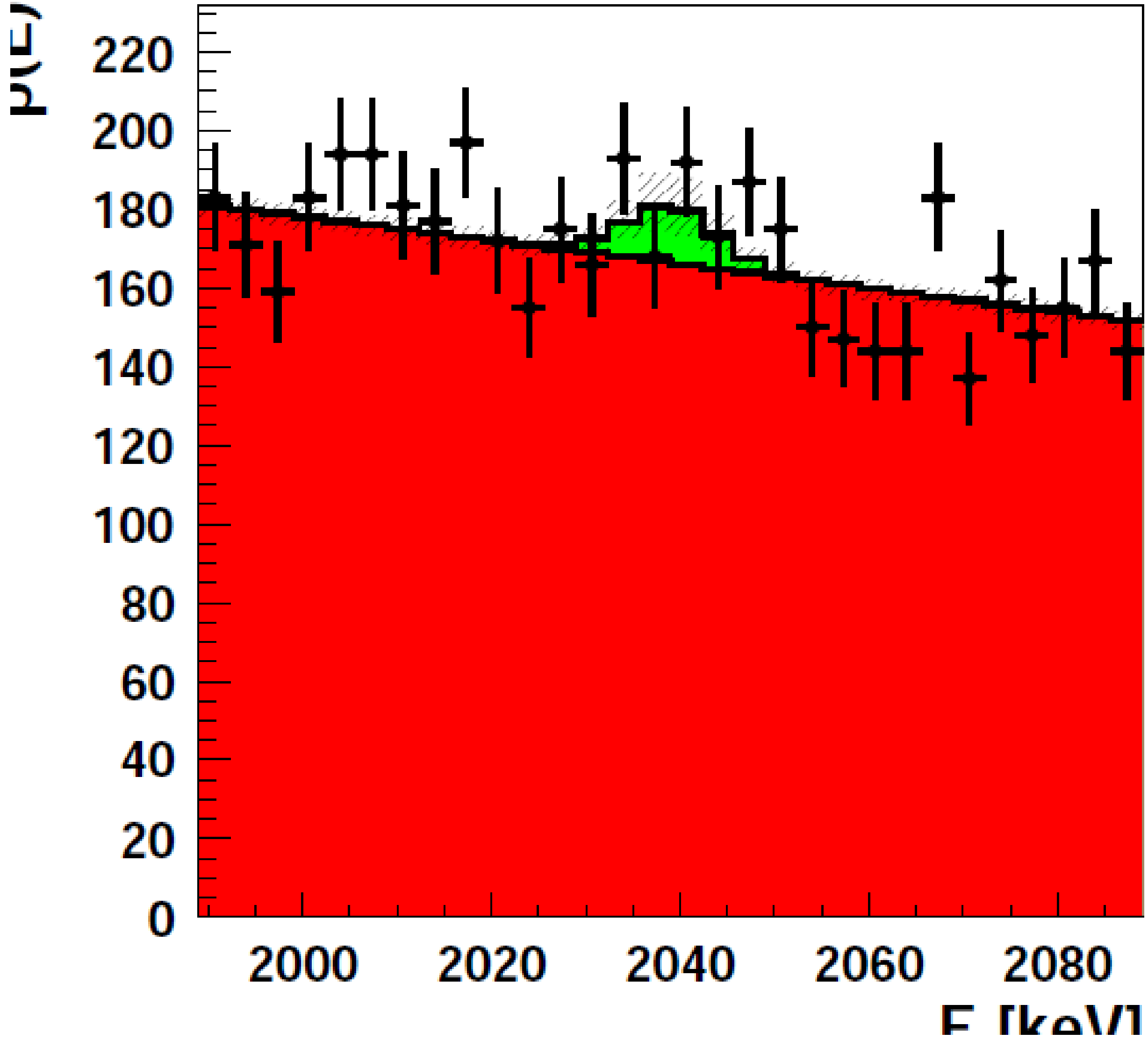
Update of knowledge







Fitted spectrum



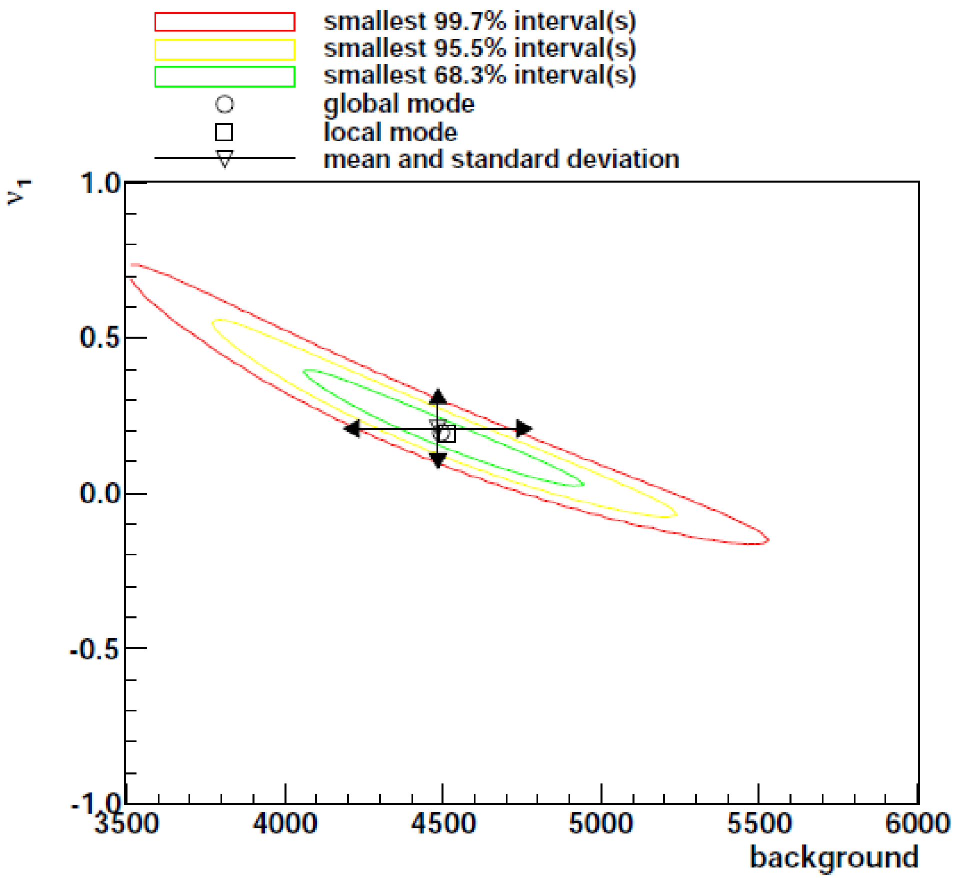
Nuisance parameters

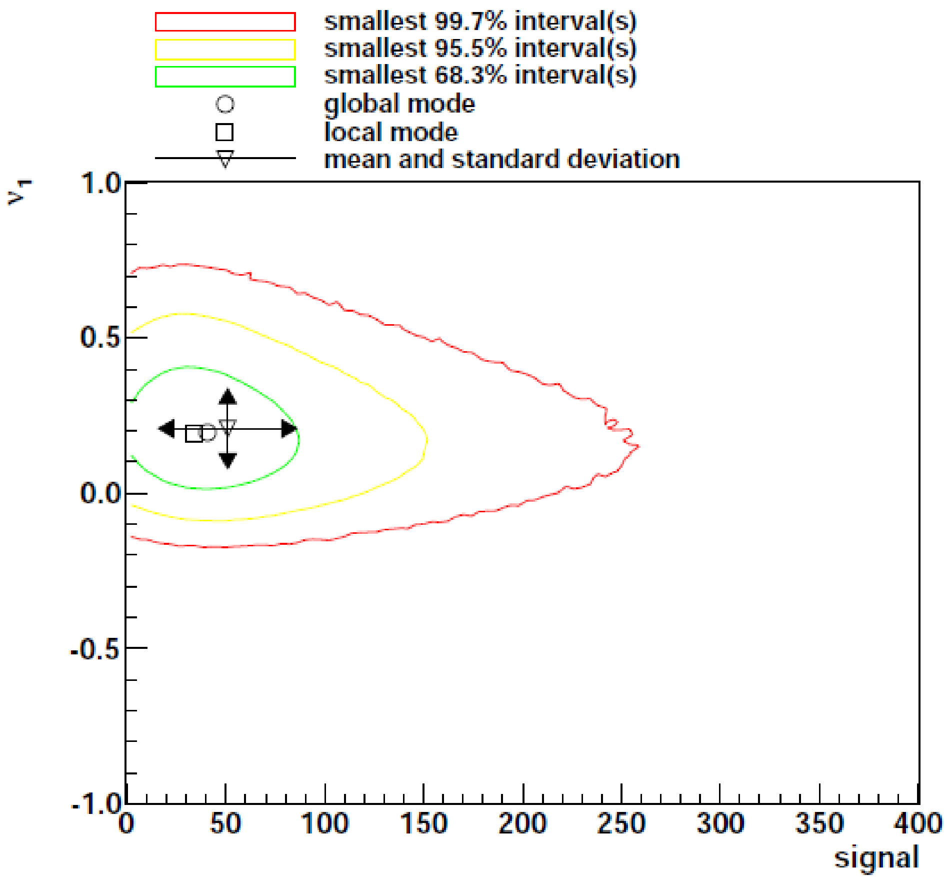
- Model = Physics model (+ par.) ⊗ Detector model (+ nuisance par.)
- Associate nuisance parameters to sources of systematic uncertainties, e.g.
  - Collider: Luminosity uncertainty (1 parameter)
  - Calorimeter: jet energy resolution (typically 3 parameters)
  - Reconstructed objects: reconstruction 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

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

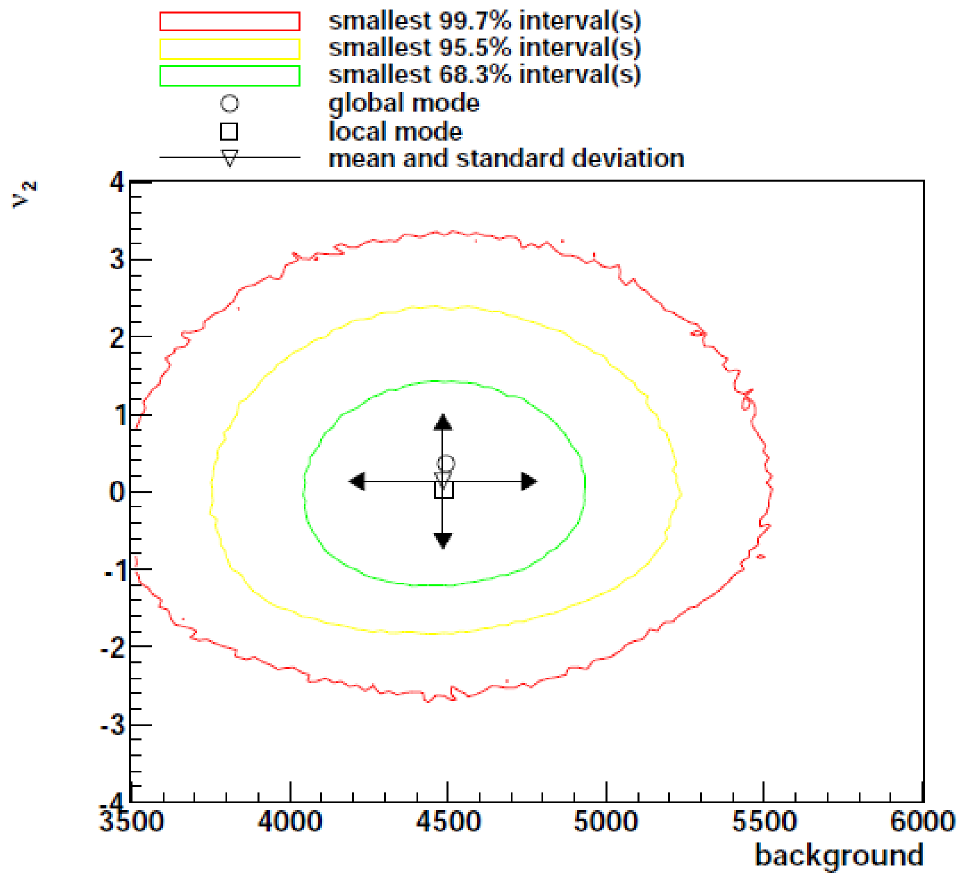


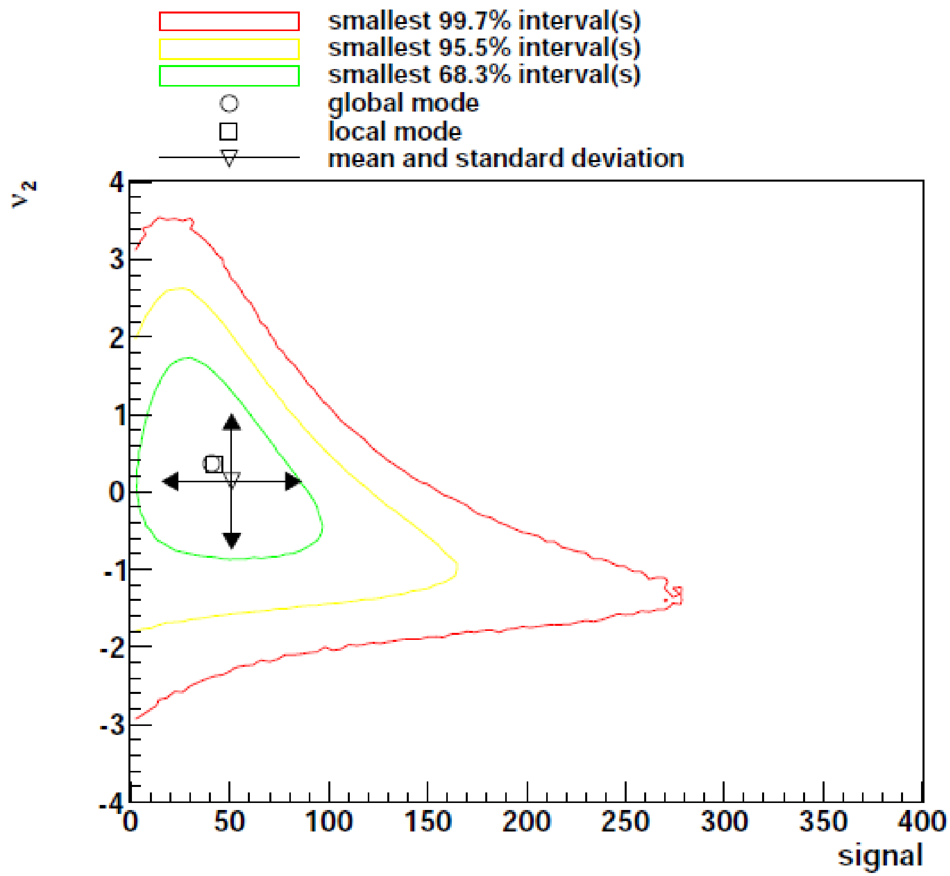


Constrained by background

Medium impact on signal

Systematic 2





Not constrained by background

Large impact on signal

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