How good is your fit?

Frederik Beaujean*, Allen Caldwell*, Daniel Kollár, Kevin Kröninger
*Max-Planck-Institut für Physik, München

DPG Frühjahrstagung 2010

Bonn, 17.3.2010

Goal: Compare data and models

1. Estimate model parameters
2. (often omitted) Check model validity

BAT (Bayesian Analysis Toolkit) → Software package to solve statistical problems using Bayesian approach

$$p(\vec{\lambda} \mid D) = \frac{p(D \mid \vec{\lambda})\, p_0(\vec{\lambda})}{\int p(D \mid \vec{\lambda})\, p_0(\vec{\lambda})\, d\vec{\lambda}}$$

- C++ based framework (flexible, modular)
- Interfaces to ROOT, Cuba, Minuit, …
- Free software: tutorials, examples, download... all at

http://mpp.mpg.de/bat/

Common tasks already implemented:

- Parameter extraction
- Uncertainty propagation
- Model comparison
- Goodness of fit
- ...

Goals for future:

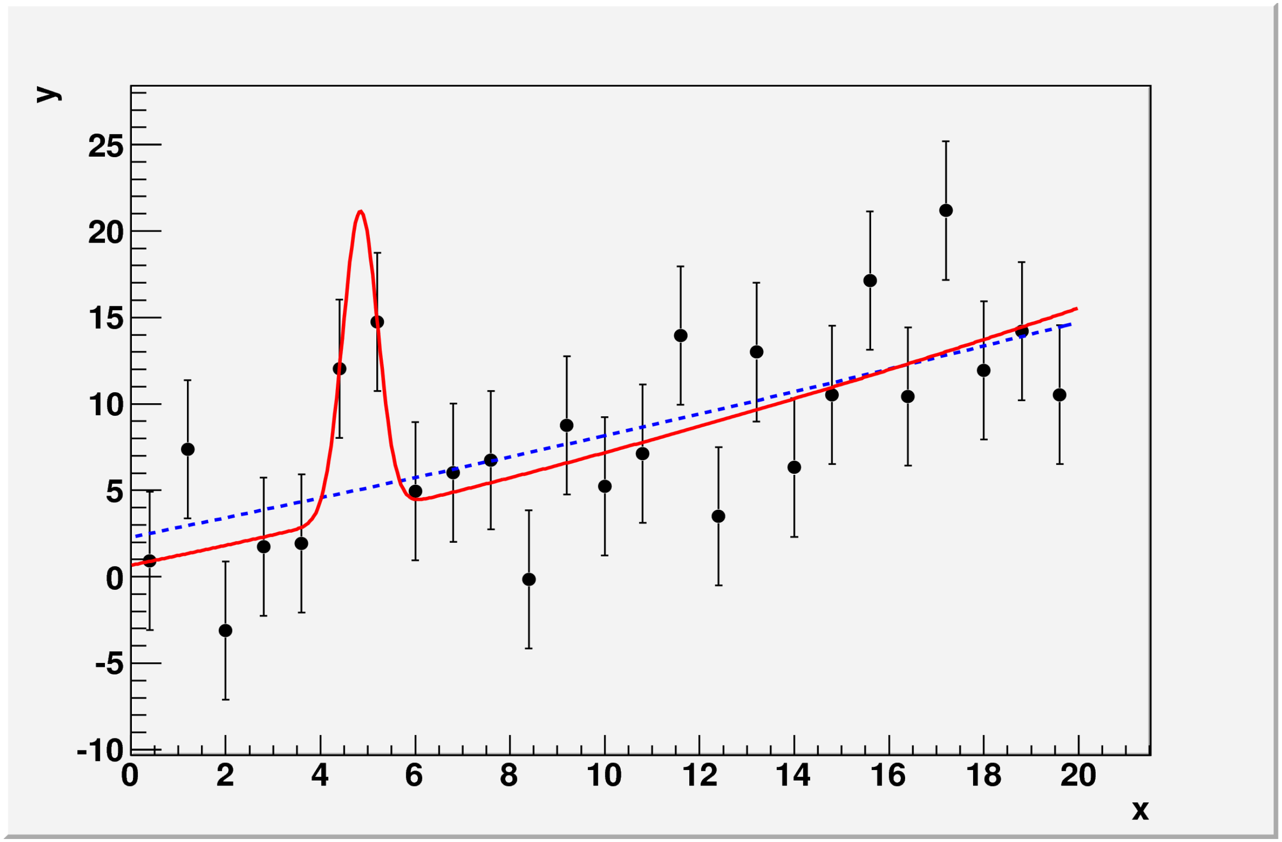
- More modular (ready for parallelization ...)
- More tools for Goodness of fit

Key tool: Markov Chain Monte Carlo

Motivating example

Suppose:

- Measurements y_i with Gaussian uncertainty
- Standard Model (SM) background is quadratic
- New physics (NP) predicts signal peak



Goodness of Fit: standard approach

Test statistic:

- Any scalar function of data, T(D)
- Interpret: large T(D) = poor model

Example:

- Prob. of the data

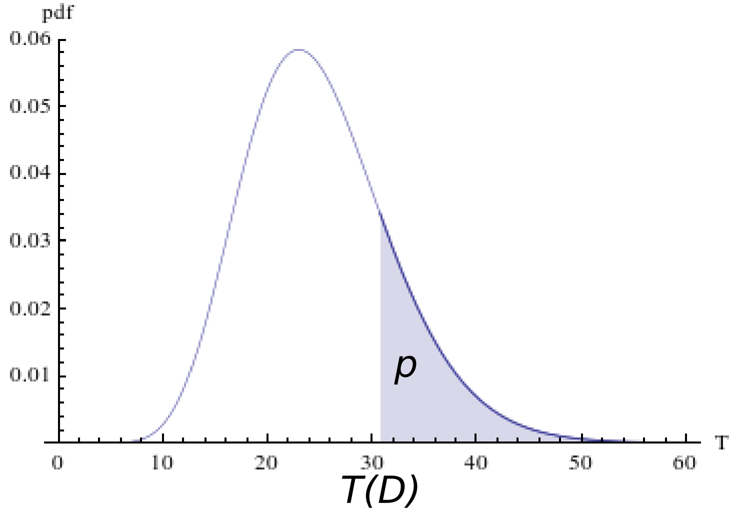
$$P(D|\vec{\lambda}) \propto \prod \exp\left\{-\frac{\left(y_i - f(x_i|\vec{\lambda})\right)^2}{2\sigma_i^2}\right\} = \exp\left\{\frac{-\chi^2}{2}\right\}$$

- Familiar choice

$$T(D) \equiv \chi^2(D)$$

p-value

Def: $p \equiv P(T > T(D))$



- Assuming the model and before data is taken: p uniform in [0,1]

- Critical values: $p < 0.05, 0.01 \Rightarrow$ reject model

- Warning: p-value *not* the P. that the model is true

Example: $p_{SM} = 10\%,\ p_{NP} = 37\% \Rightarrow$ both OK

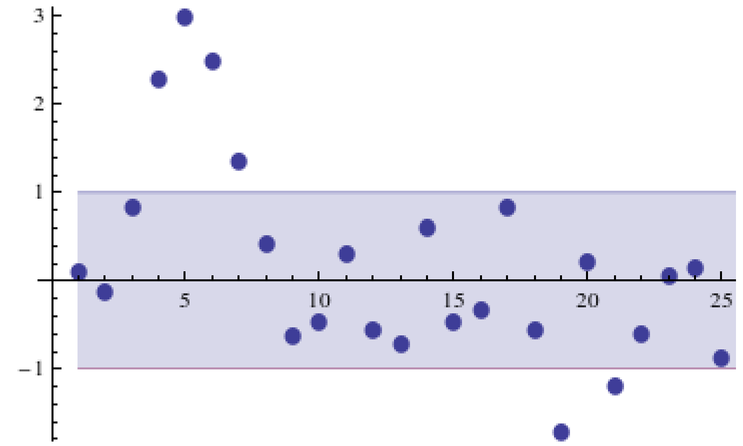
Runs

- Most statistics disrespect order of data, information wasted

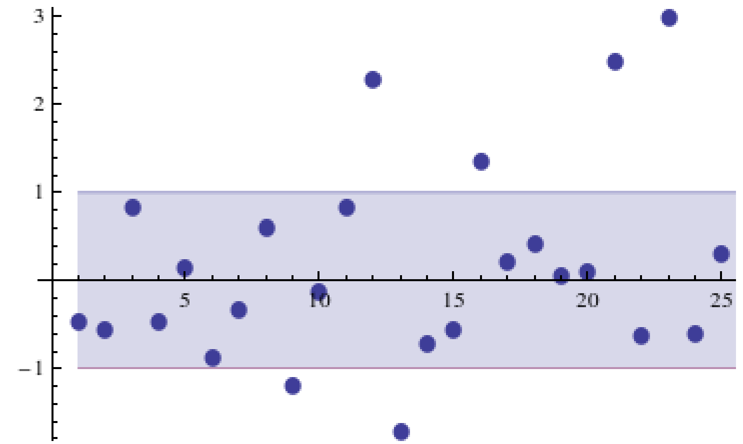
- Human brain good for simple problems

<u>Example:</u>

- N=25 datapoints
- Each Gaussian with mean = 0 and variance = 1



$$\chi^2 = 32.1 \Rightarrow p = 0.16$$

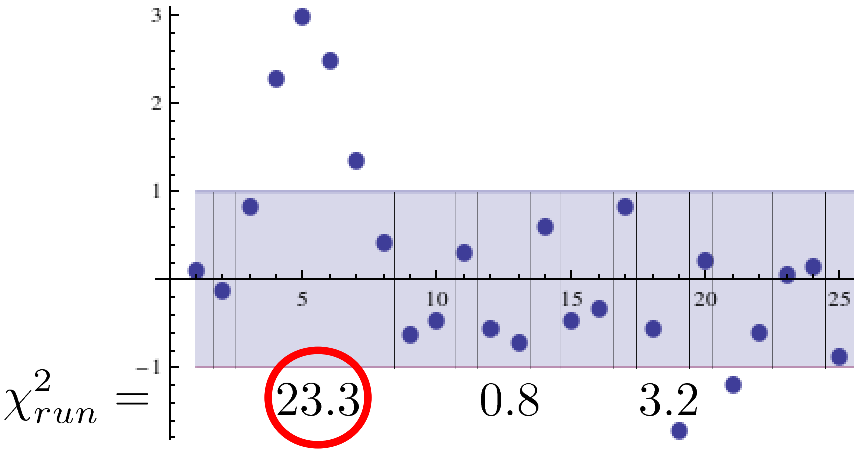


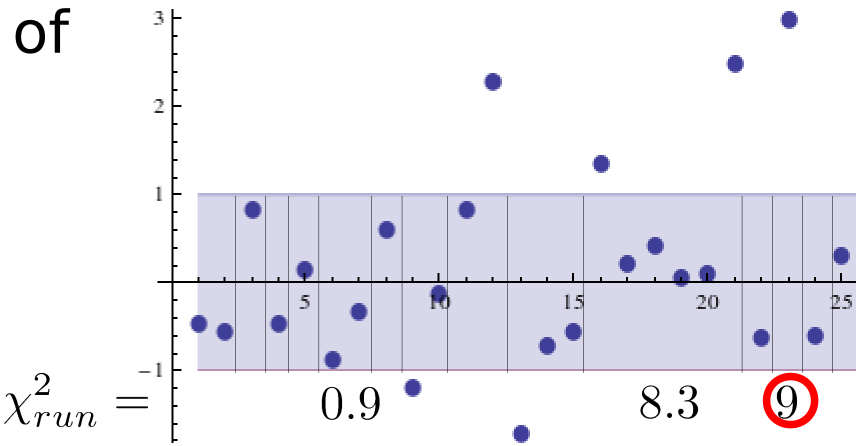
Runs statistic

Proposal:

- Split data into runs
- Each run has a weight Gaussian: take χ^2_{run}
- Test statistic: largest weight of any run $T \equiv \max\{\chi^2_{run}\}$
- p-value becomes

$$p_{run} \equiv P(T > T_{obs})$$

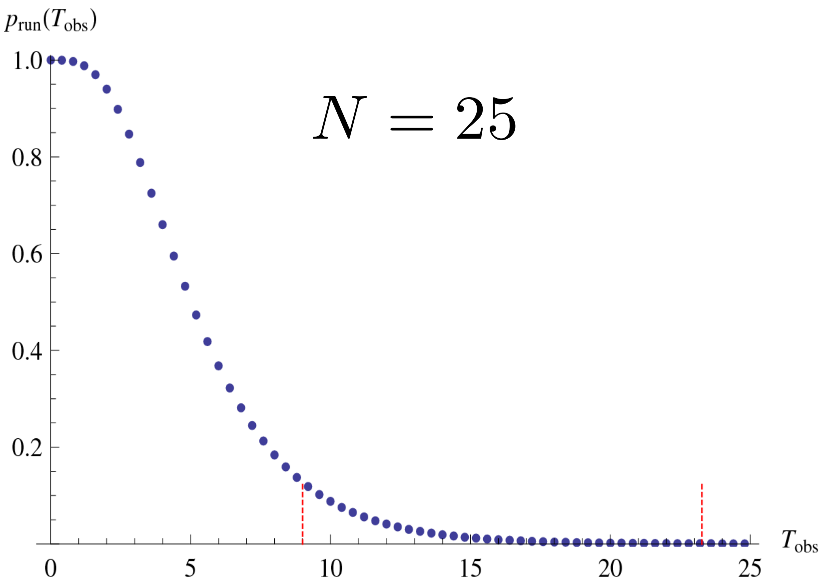




Runs distribution

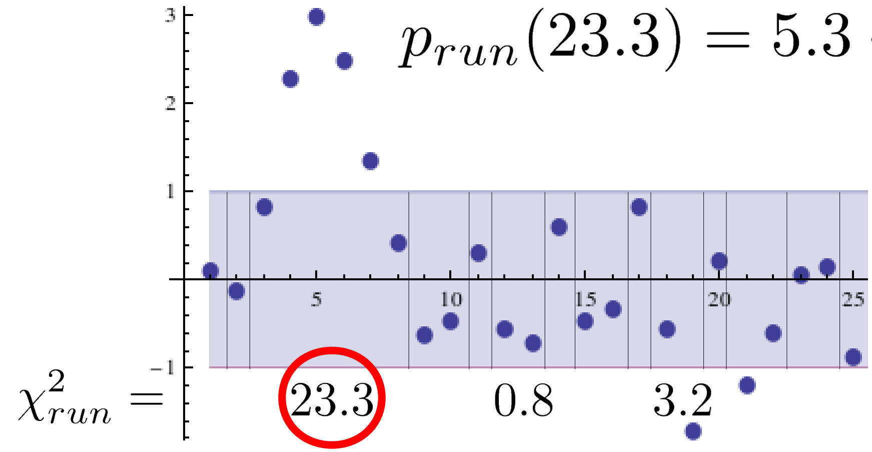
Gaussian case:

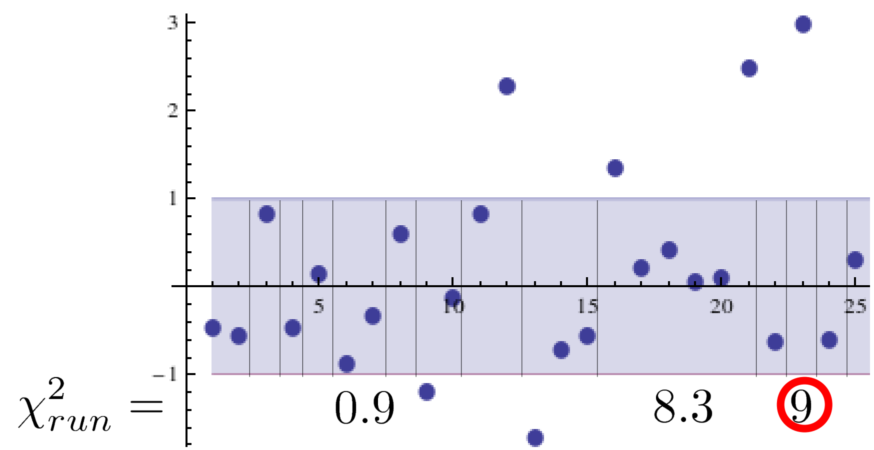
- Distribution of p_{run} exactly calculated for any N



$$N = 25$$

$$p_{run}(23.3) = 5.3 \cdot 10^{-4}$$





$$p_{run}(9) = 1.3 \cdot 10^{-1}$$

Conclusions

Model checking “=” job interview

- Ask all important questions (not just one)
- Poor answers disqualify the applicant
- Only one (bad) applicant => may have to accept anyway

FINIS

Backup

Exact runs distribution I

$$P(T \geq T_{obs}|N) = 1 - P(T < T_{obs}|N)$$

$$P(T < T_{obs}|N) = \sum_{r=1}^{N} \sum_{M=1}^{\min(r,N-r+1)} P(T < T_{obs}|M, r, N) \cdot P(M, r|N)$$

$$\begin{aligned} M &= \text{number of success runs} \\ r &= \text{number of successes} \\ N &= \text{number of datapoints} \end{aligned}$$

$$P(M, r|N) = \frac{1}{2^N - 1} \cdot R(M|r, N)$$

$$R(M|r, N) = \text{number of permutations with } M \text{ runs given } r, N$$

Exact runs distribution II

$$P(T < T_{obs}|M, r, N) = \sum_{\pi} P(T < T_{obs}|\pi)P(\pi|M, r, N)$$

$$P(\pi|M, r, N) = \frac{W_\pi}{R(M|r, N)}$$

$$\pi = (r_1, \dots r_l)$$

$$W_\pi = \binom{M}{r_1, \dots, r_l} \cdot \binom{N - r + 1}{M} = \frac{(N - r + 1)!}{(N - r - M + 1)! \cdot \prod_l r_l!}$$

$$P(T < T_{obs}|\pi) = \prod_l \left[P\left(T < T_{obs}|l\right)^{r_l}\right]$$

$$P\left(T < T_{obs}|l\right) = \frac{\gamma(l/2, T_{obs}/2)}{\Gamma(l)}$$ CDF of χ^2 distribution

$$\gamma(a,x) = \int_0^x t^{a-1}e^{-t}\mathrm{d}t$$ Lower incomplete gamma function

$$\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}\mathrm{d}t$$ Complete gamma function

Goodness of Fit: Bayesian approach

Model selection:

- Need explicit alternatives M_1, M_2
- Posterior odds

$$\frac{P(M_1|D)}{P(M_2|D)} = \frac{P(M_1)}{P(M_2)} \times \frac{P(D|M_1)}{P(D|M_2)}$$

Bayes factor:

- (very) sensitive to parameter range

$$P(D|M_1) = \int p(D\,|\,\vec{\lambda})\; p_0(\vec{\lambda})\; d\vec{\lambda}$$

- Occam's razor built in

Example:

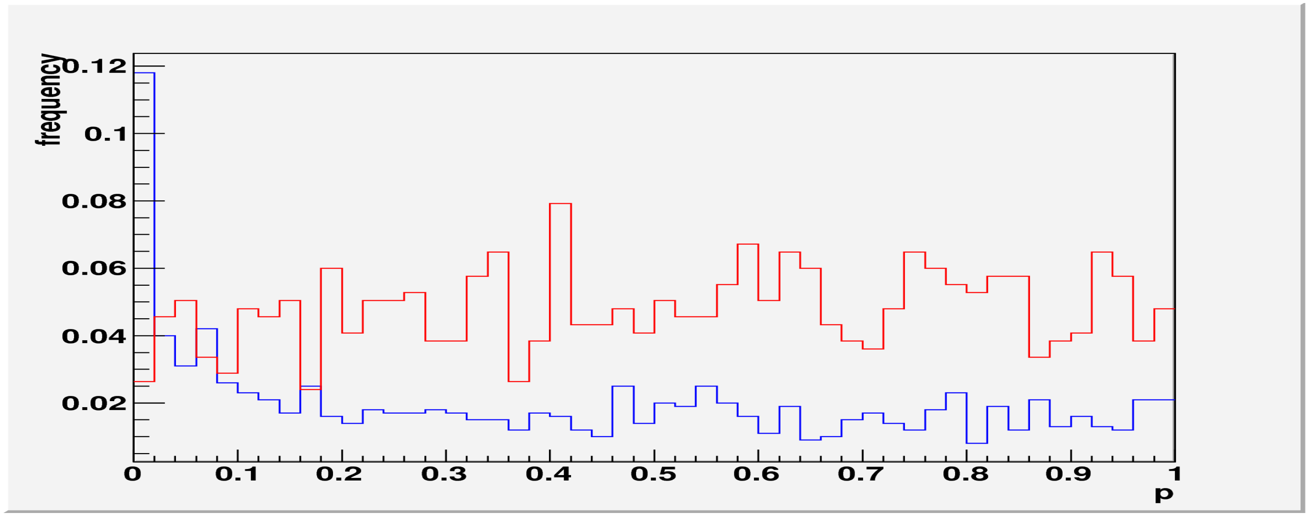
- Six (NP) vs three (SM) parameters

$$\frac{P(\mathrm{SM}|D)}{P(\mathrm{NP}|D)} = \frac{P(\mathrm{SM})}{P(\mathrm{NP})} \times 61.7$$

More on p values

Example revisited: Poisson instead of Gaussian

$$\chi^2 = \sum_i \frac{(n_i - \nu_i)^2}{\nu_i} \quad \text{vs.} \quad \chi^2 = \sum_i \frac{(n_i - \nu_i)^2}{n_i}$$



=> When checking a model, have to take uncertainties from the model, not from observation

p value is **not** the probability that the model is true

Common tools

- Integration
  - Monte Carlo (sampled mean)
  - Importance sampling
  - CUBA (Vegas,...)
- Error propagation
  - Calculate any value of the parameters during the run
- Optimization
  - Monte Carlo (hit & miss)
  - Metropolis
  - Interface to Minuit
  - Simulated Annealing
- Validation
  - Ensemble testing and *p*-value

Key tool: Markov Chain Monte Carlo