

Hypothesis Testing & Model Comparison



Shabnaz Pashapour

Georg-August-Universität Göttingen

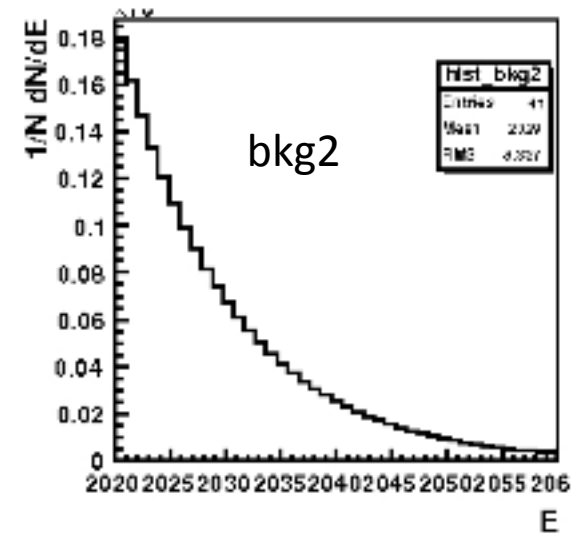
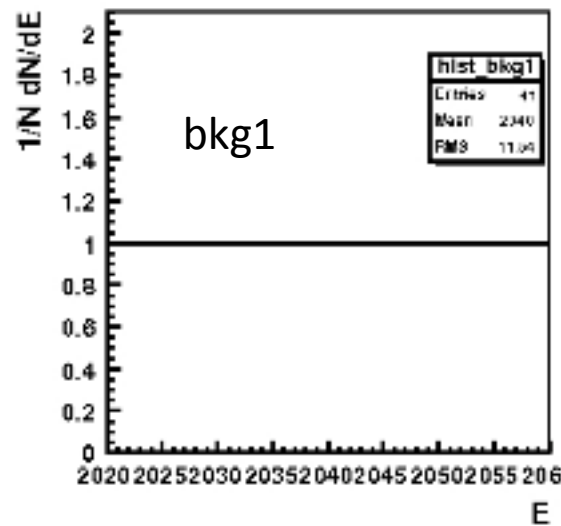
- Signal search in the presence of background
 - How to judge the validity of models
 - How to compare different models to each other
- Data is composed of different sources of background and one signal
 - The shapes are known
 - Binned likelihood fit, assuming independent Poisson fluctuations
 - Fit by scaling distributions, templates
 - The scale factors of the contributions are the parameters of the model and named N_1, N_2 , etc.

$$p(D | S, \mu, \sigma, B) = \prod_{i=1}^{N_{bins}} \frac{\lambda_i^{n_i}}{n_i!} e^{-\lambda_i}$$
$$\lambda_i = \int_{\Delta_i} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \frac{B}{\Delta_i}$$

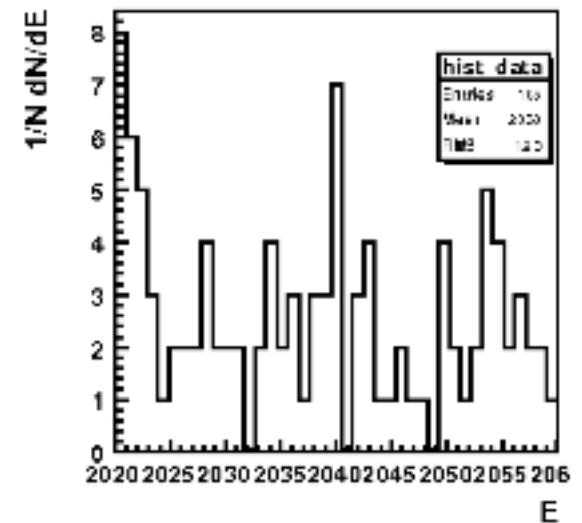
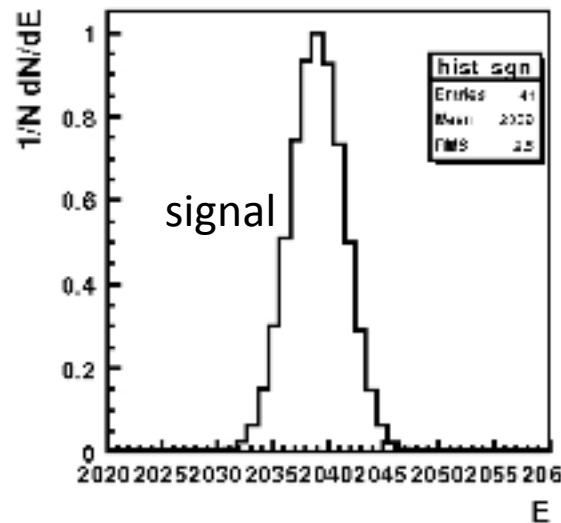
Distributions – Templates and Pseudodata



- Histograms available from a root file



- Use the templates to fit the data



“Signal search in the presence of background”

1. Getting started
2. Fitting a background-only model
3. Introduction to the goodness-of-fit tests
4. Goodness-of-fit tests for signal and background models
5. Comparing signal and background models
6. Testing more complex models

- Test the experimental result (data) given a hypothesis
 - Is the fit reasonable?
- Common hypothesis tests:
- *χ^2 -tests*
 - Sum of differences (data, prediction) squared
 - Depends on the number of degrees of freedom
 - Assume the Gaussian distribution of uncertainties
- *Kolmogorov-Smirnov test*
 - Compare the shape of the cumulative distribution
- *P-value*
 - Test how likely the observed data is given the best-fit parameters
 - Different definitions used

- What theoretical model is a better description of the data?
 - Directly compare two models. Do not judge the goodness-of-fit
- Common ways to compare models:
- *X^2 -tests*
 - Use the X^2 -probability
- *Maximum Likelihood*
 - Compare the maximum Likelihood found in the fit
- *Posterior Probability of the models*
 - Integrate over all parameters
 - Occam's razor
- *Bayes factors*
 - Use ratio of posterior probabilities of the models

- Setup the BAT environment:
source /statistics-school/BAT-0.4.2/forTutorial/setupBAT.sh
cd \$BATINSTALLDIR/tools
- Start with the tutorial *at*
http://mpp.mpg.de/bat/?page=tutorials&name=hypothesis_stack