



# Introduction to the Bayesian Analysis Toolkit - BAT

---

BAT Tutorial

Rome – April 22<sup>nd</sup> 2010

Frederik Beaujean<sup>1</sup>, Allen Caldwell<sup>1</sup>, Daniel Kollár<sup>2</sup>, Kevin Kröninger<sup>3</sup>

<sup>1</sup>Max Planck Institute for Physics, Munich

<sup>2</sup>CERN

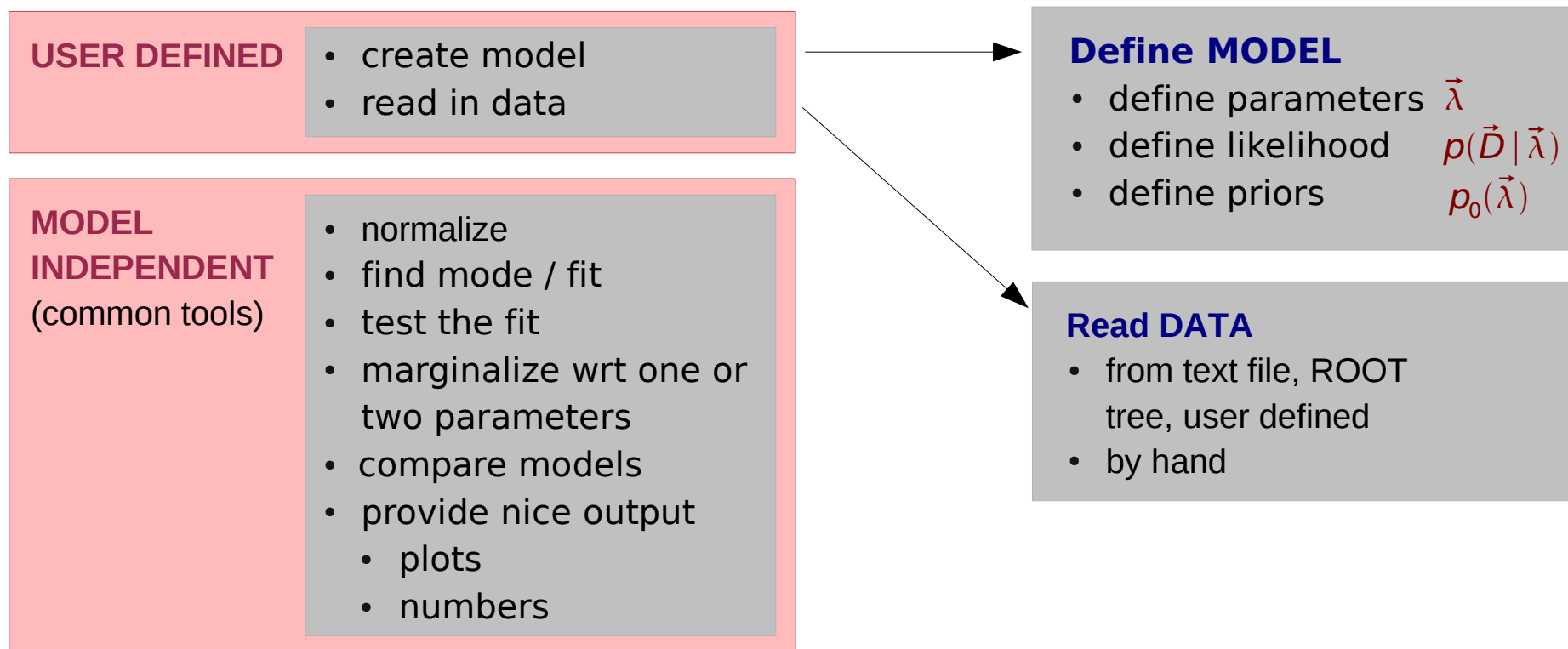
<sup>3</sup>University of Göttingen

Bayes' formula for parameter estimation

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

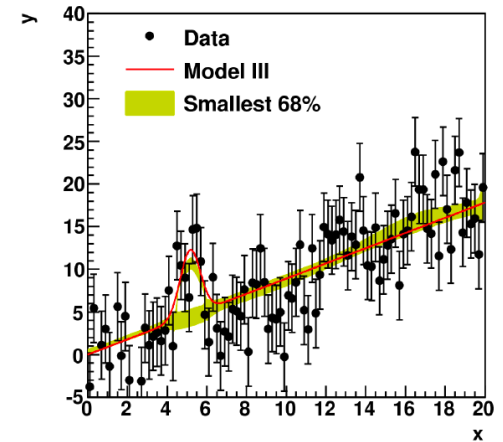
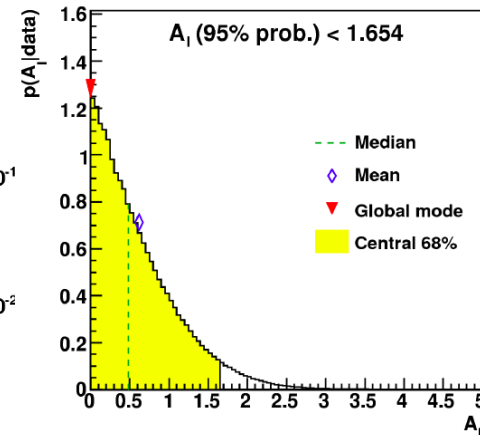
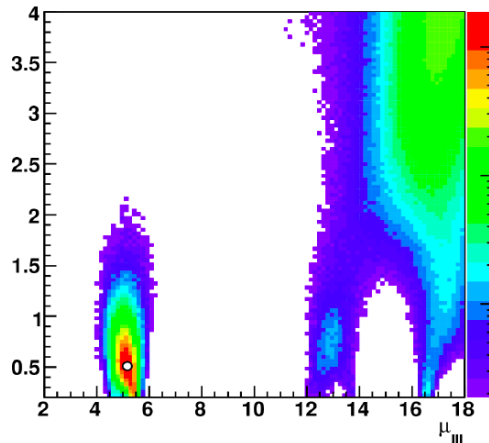
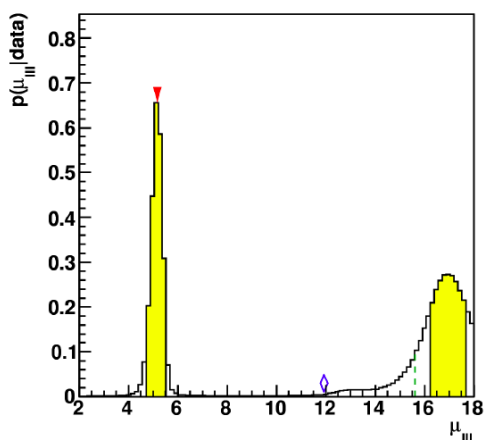
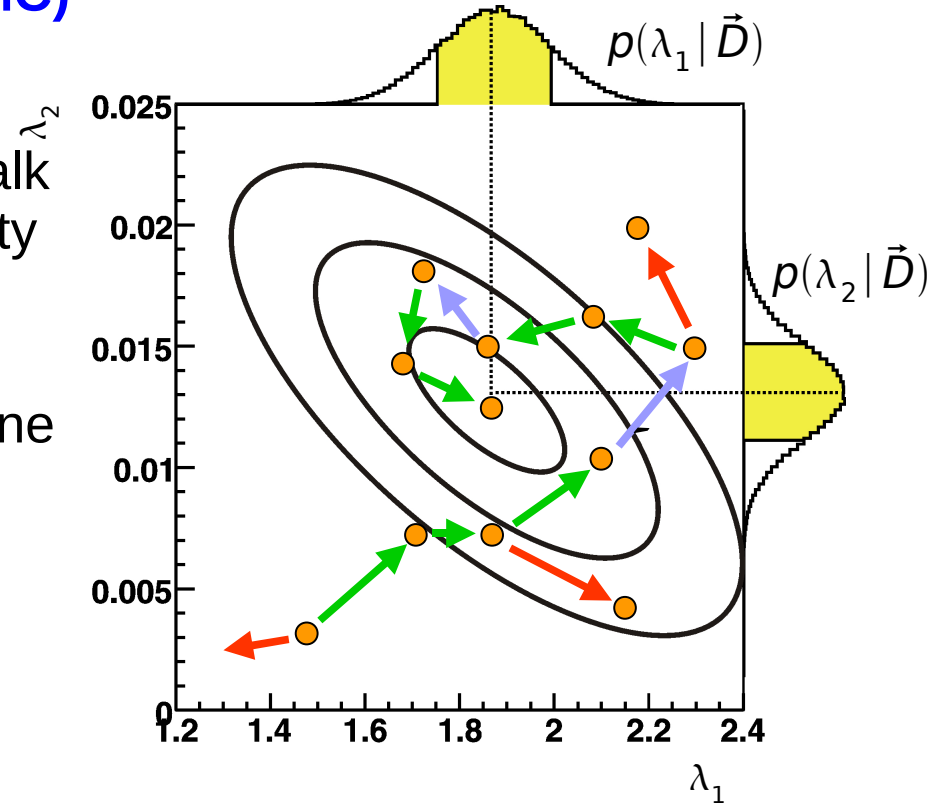
**BAT** Software package for solving of statistical problems using Bayesian approach:

- collects common tools in Bayesian analyses in a single modular framework
- C++ based, interfaced to ROOT



# Key tool: Markov Chain Monte Carlo (MCMC)

- Implemented Metropolis algorithm
  - Sample parameter space with random walk
  - Walk guided to regions of larger probability
- Calculate posterior probability for all parameters:
  - Calculate marginalized distributions wrt one or two parameters
  - Integrate out all other parameters
  - Calculate probability distribution of any function, i.e., full error propagation



## Define a model class inheriting from a base model class **BCModel**

- add parameters:
- implement likelihood:
- implement prior:

```
AddParameter("p0", -4, 4);

double LogLikelihood(vector<double> params)

double LogAPrioriProbability(vector<double> params)
```

## Main program/macro

- Create a model:
- read data:
- assign data to the model:
- do the analysis:

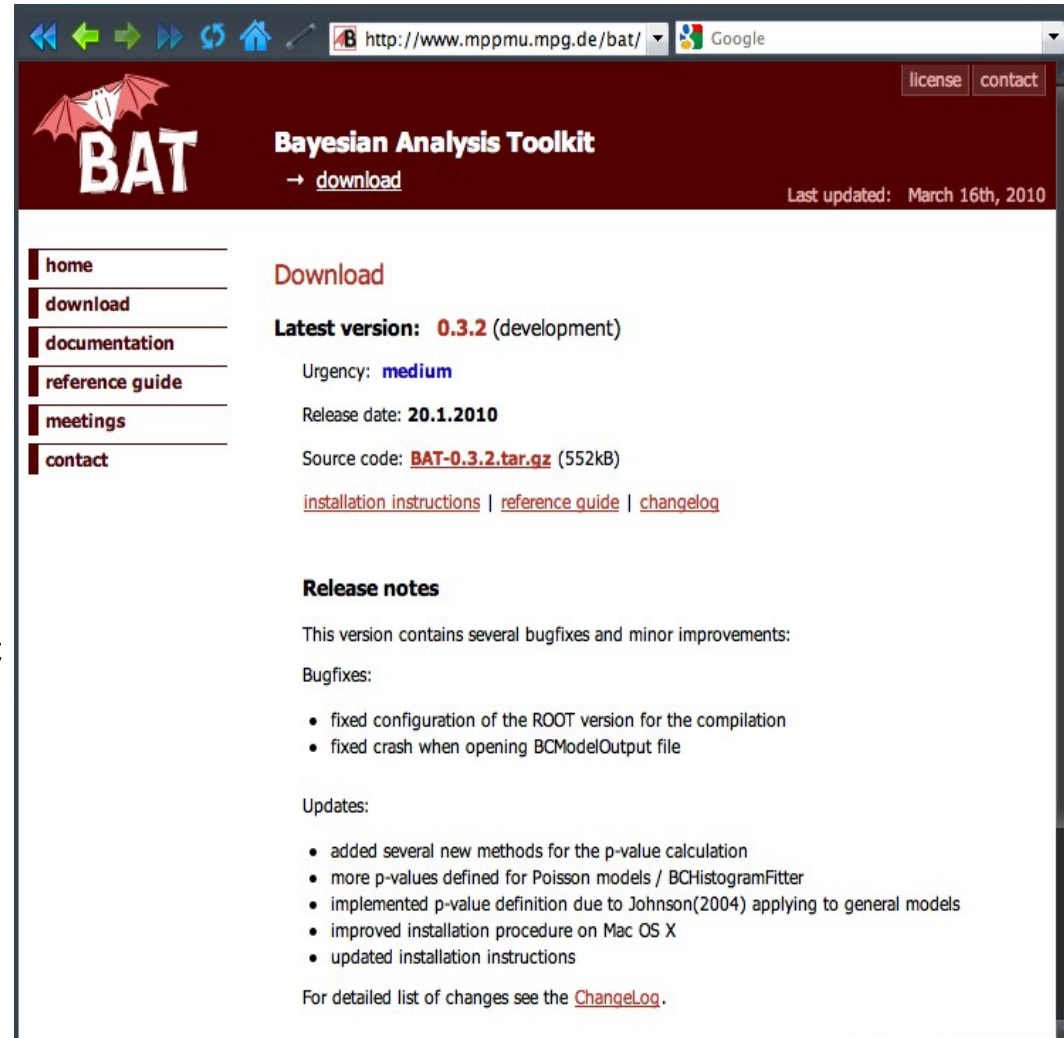
```
MyModel * mm = new MyModel("Model 0");

BCDataSet * data = new BCDataSet();
data->ReadFromFileTxt("file.txt", 4);

mm->SetDataSet(data);

mm->Normalize();
mm->MarginalizeAll();
mm->FindModeMinuit( mm->GetBestFitParameters() );
mm->PrintAllMarginalized("distributions.ps");
etc.
```

- can be downloaded from:  
<http://www.mppmu.mpg.de/bat>
- BAT comes in form of shared library
- depends of the ROOT I/O functionality
- BAT contains at the moment 15 classes which provide:
  - main infrastructure
  - algorithms
  - output and logging
  - extension classes to solve specific (frequent) fitting problems
  - a set of well documented examples is included in BAT distribution
- good starting point “Introduction to BAT” document
- check the reference guide!
- BAT paper published: Computer Physics Communications **180** (2009) 2197-2209



The screenshot shows the BAT website interface. The browser address bar displays <http://www.mppmu.mpg.de/bat/>. The website header features the BAT logo and the text "Bayesian Analysis Toolkit" with a "download" link and "Last updated: March 16th, 2010". A sidebar on the left contains navigation links: home, download, documentation, reference guide, meetings, and contact. The main content area is titled "Download" and lists the "Latest version: 0.3.2 (development)" with an urgency of "medium". It provides the release date "20.1.2010" and the source code file "BAT-0.3.2.tar.gz (552kB)". Links for "installation instructions", "reference guide", and "changelog" are provided. The "Release notes" section states that this version contains several bugfixes and minor improvements, listing:
 

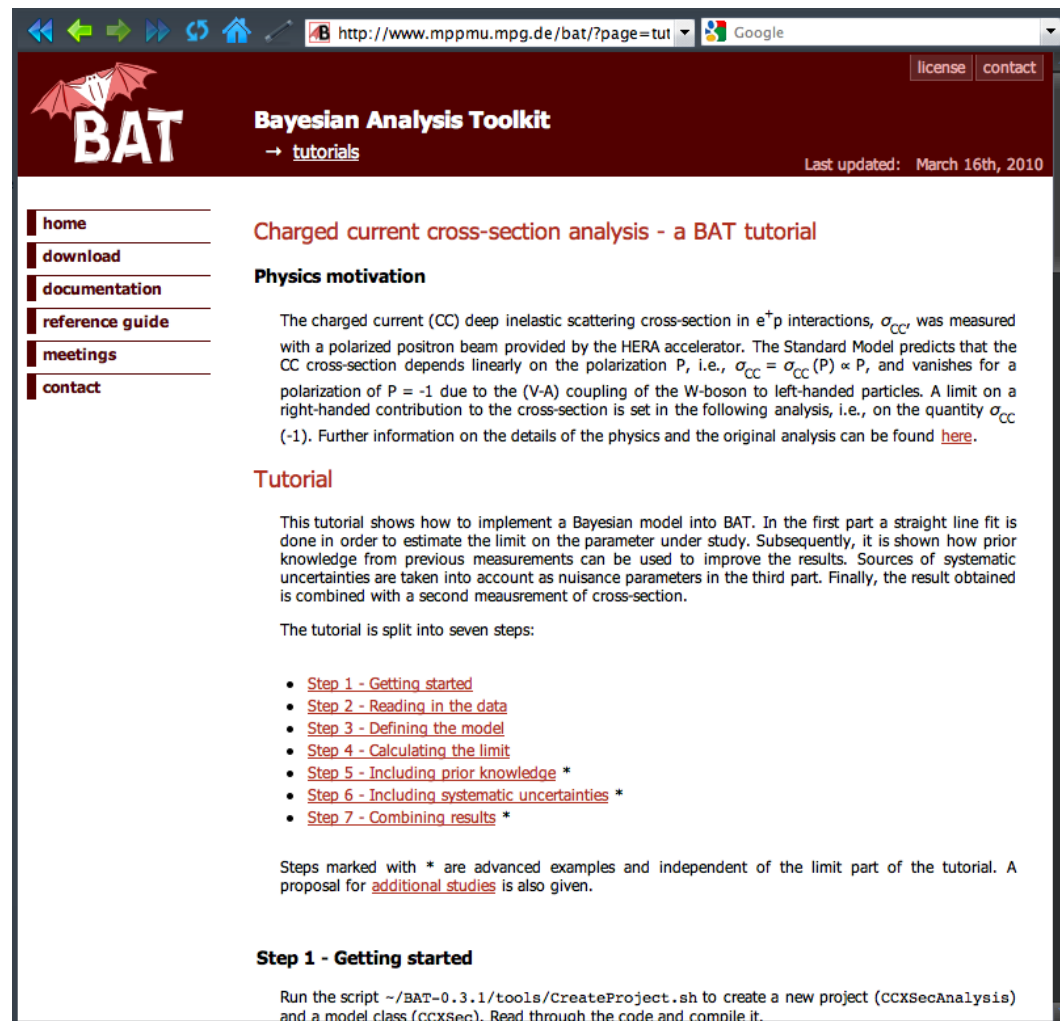
- fixed configuration of the ROOT version for the compilation
- fixed crash when opening BCModelOutput file

 It also lists updates:
 

- added several new methods for the p-value calculation
- more p-values defined for Poisson models / BCHistogramFitter
- implemented p-value definition due to Johnson(2004) applying to general models
- improved installation procedure on Mac OS X
- updated installation instructions

 A link to the "ChangeLog" is provided for a detailed list of changes.

- tutorials section was added to the BAT webpage
  - can be found under: Documentation → Tutorials
- four tutorials at the moment:
  - limit setting
  - hypothesis testing and template fitting
  - *efficiency fitting* ← today
  - *Poisson example* ← today
- show basic information on how to
  - set up a model
  - calculate limits
  - define prior
  - include systematic uncertainties
  - etc.
- tutorials have form of exercises with solutions
- more tutorials will come



The screenshot shows a web browser window displaying the BAT website. The URL is <http://www.mppmu.mpg.de/bat/?page=tut>. The page features the BAT logo and navigation links for 'home', 'download', 'documentation', 'reference guide', 'meetings', and 'contact'. The main content area is titled 'Charged current cross-section analysis - a BAT tutorial' and includes a 'Physics motivation' section, a 'Tutorial' section, and a list of seven steps. The first step, 'Step 1 - Getting started', is highlighted.

**home**  
**download**  
**documentation**  
**reference guide**  
**meetings**  
**contact**

**Charged current cross-section analysis - a BAT tutorial**

**Physics motivation**

The charged current (CC) deep inelastic scattering cross-section in  $e^+p$  interactions,  $\sigma_{CC}$ , was measured with a polarized positron beam provided by the HERA accelerator. The Standard Model predicts that the CC cross-section depends linearly on the polarization  $P$ , i.e.,  $\sigma_{CC} = \sigma_{CC}(P) \times P$ , and vanishes for a polarization of  $P = -1$  due to the (V-A) coupling of the W-boson to left-handed particles. A limit on a right-handed contribution to the cross-section is set in the following analysis, i.e., on the quantity  $\sigma_{CC}(-1)$ . Further information on the details of the physics and the original analysis can be found [here](#).

**Tutorial**

This tutorial shows how to implement a Bayesian model into BAT. In the first part a straight line fit is done in order to estimate the limit on the parameter under study. Subsequently, it is shown how prior knowledge from previous measurements can be used to improve the results. Sources of systematic uncertainties are taken into account as nuisance parameters in the third part. Finally, the result obtained is combined with a second measurement of cross-section.

The tutorial is split into seven steps:

- [Step 1 - Getting started](#)
- [Step 2 - Reading in the data](#)
- [Step 3 - Defining the model](#)
- [Step 4 - Calculating the limit](#)
- [Step 5 - Including prior knowledge](#) \*
- [Step 6 - Including systematic uncertainties](#) \*
- [Step 7 - Combining results](#) \*

Steps marked with \* are advanced examples and independent of the limit part of the tutorial. A proposal for [additional studies](#) is also given.

**Step 1 - Getting started**

Run the script `~/BAT-0.3.1/tools/CreateProject.sh` to create a new project (CCXSecAnalysis) and a model class (CCXSec). Read through the code and compile it.