Data Analysis with ROOT

Lecture 3: Bayesian statistics, propagation of errors, and unfolding

Aatos Heikkinen [1]

July 13, 2009

[1] Helsinki Institute of Physics, Helsinki, Finland. (aatos.heikkinen@cern.ch)

Outline

- Bayesian statistics
  - Learning by experience
  - Example of Bayesian reasoning
- Frequentist statistics
  - Hypothesis tests
  - Goodness of fit / p-value
- Uncertainty in physics
  - Systematic vs. statistical errors
  - Propagation of uncertainty
- An introduction to unfolding
  - Measurements are always convoluted with detector response function
  - What can be done to recover blurred signal?

Example: Who will pay the next round?

You meet an old fried at Göttingen in a pub. He proposes that the next round should be payed by whichever of the two extracts the card of lower value from a pack of cards.
This situation happens many times in the following days. What is the probability that your friend cheats if you end up paying *wins* consecutive times[2]
You assume:

- $P(cheat) = 5\%$ and $P(honest) = 95\%$. (Surely an old friend is an unlikely cheater ...)
- $P(wins|cheat) = 1$ and $P(wins|honest) = 2^{-wins}$

Bayesian solution:

$$P(cheat|wins) = \frac{P(wins|cheat)P(cheat)}{P(wins|cheat)P(cheat) + P(wins|honest)P(honest)}$$

$$P(cheat|0) = \frac{1P(cheat)}{1P(cheat) + 2^{-0}P(honest)} = \frac{0.05}{0.05 + 0.95} = 5\%$$

$$P(cheat|5) = \frac{1P(cheat)}{1P(cheat) + 2^{-5}P(honest)} = \frac{0.05}{0.05 + 0.03} = 63\%$$

[2] Adapted from G. D'Agostini, *Bayesian Reasoning in High-Energy Physics: Principles and Applications*, CERN-99-03, 1999

Example: Learning by experience

The process of updating the probability when new experimental data becomes available can be followed easily if we insert

- $P(cheat) = P(cheat|wins-1)$ and $P(honest) = P(honest|wins-1)$, where $wins-1$ indicate the propability assigned after *the previous win*
- $P(wins=1|cheat) = P(win|cheat) = 1$ and $P(wins=1|honest) = P(win|honest) = \frac{1}{2}$

Iterative aplication of the Bayes formula for $P(cheat|wins)$=

$$\frac{P(win|cheat)P(cheat|wins-1)}{P(win|cheat)P(cheat|wins-1)+P(win|honest)P(honest|wins-1)}$$

$$= \frac{P(cheat|wins-1)}{P(cheat|wins-1)+\frac{1}{2}P(honest|wins-1)}$$

P(cheat)	$P(cheat\|wins)$		
%	wins=5	10	15
1	24	91	99.7
5	63	98	99.94
50	97	99.9	99.997

When you learn from the experience, your conclusions no longer depend on the initial assumptions.

Probability[3].

Let's define a probability considering a set S, called space, and its subsets $A, B, \ldots$ as a real-valued function fulfilling the following axioms:

- For every subset A in S, $P(A) \geq 0$
- For disjoint subsets ($A \cap B = \emptyset$), $P(A \cup B) = P(A) + P(B)$
- $P(S) = 1$

From **conditional probability**
$P(A|B) = P(A \cap B)/P(B)$, using $A \cap B = B \cap A$ one obtains **Bayes' theorem**

$$P(A|B) = P(B|A)P(A)/P(B)$$

From the axioms of probability, and the definition of conditional probability we get **the law of total probability** $P(B) = \sum_i P(B|A_i)P(A_i)$.

This combined with Bayes' theorem gives

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i))}$$

[3] Adapted from W.M. Yao *et al.* Journal of Physics G 33, 1, 2006. For more details consult Particle Data Group http://pdg.lbl.gov

Bayesian probability

The subsets of the sample space S can be also interpreted as **hypothesis**, *i.e.* statements that are either true or false.

- In **the frequency interpretation**, a hypothesis is either always or never true.
- In **subjective probability** $P(A)$ is interpreted as the degree of of belief that the hypothesis A is true.

Subjective probability is used in **Bayesian** (as opposed to **frequentist**) statistics. Bayes' theorem can be written

$$P(theory|data) \propto P(data|theory)P(theory)$$

where 'theory' represents some hypothesis and 'data' is the outcome of the experiment.

Bayesian statistics provides no rule for obtaining the $P(theory)$; this is necessarily subjective and depends on theoretical considerations and previous measurements.

Bayes' theorem

Let us think of all the possible, mutually exclusive, hypotheses H_i, which could condition the event E.
The standard form of Bayes' theorem is

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_j P(E|H_j)P(H_j)}$$

where

- $P(H_i)$ is the initial (or *a priori* or briefly 'prior') propability of H_i. That is, the probability of this hypothesis using information available before the event E.
- $P(E|H_i)$ is called likelihood
- $P(H_i|E)$ is the final (or *a posteriori*) probability of H_i after the new information

Typical application of Bayesian reasoning in HEP is Kalman filter used in track finding.

Example: μ-trigger

Example[4]: The particle beam is a mixture of 90% π and 10% μ, detector has μ identification efficiency of 95%, and a probability of identifying a π as a μ of 2%. A trigger is fired, if a particle is identified as a μ.

Q: What is the probability that a trigger is falsely fired by a π?

Solution: Bayes' formula for two hypotheses μ and π conditioning the trigger T:

$$P(\mu|T) = \frac{P(T|\mu)P(\mu)}{P(T|\mu)P(\mu) + P(T|\pi)P(\pi)} = \frac{0.95 \quad 0.1}{0.95 \quad 0.1 + 0.02 \quad 0.8} = 0.84$$

$P(\pi|T) = 1 - P(\mu|T) = 1$ - $0.84 = 0.16$.

So **16%** of triggered events are π events.

[4] Adapted from G. D'Agostini, *Bayesian Reasoning in High-Energy Physics: Principles and Applications*, CERN-99-03, 1999

Example: Classifying hypotheses according to their credibility in the Bayesian framework

Consider the detector and trigger in previous slide and $P(\pi|T) = 0.16$ and $P(\mu|T) = 0.84$. **Q**: What is the signal-to-noise ratio ?

Recalling Bayes' formulas for two hypotheses μ and π conditioning the trigger T are:

$$P(\mu|T) = \frac{P(T|\mu)P(\mu)}{P(T|\mu)P(\mu) + P(T|\pi)P(\pi)}$$

$$P(\pi|T) = \frac{P(T|\pi)P(\pi)}{P(T|\mu)P(\mu) + P(T|\pi)P(\pi)}$$

Solution: General expression for the S/N ration if the effect E is observed:

$$S/N = \frac{P(S|E)}{P(N|E)} = \frac{P(E|S)}{P(E|N)}\frac{P(S)}{P(N)} = \frac{P(E|\mu)}{P(E|\pi)}\frac{P(\mu)}{P(\pi)} = \frac{P(\mu|T)}{P(\pi|T)} = \frac{0.84}{0.16} = \mathbf{5.3}$$

When the conditions are noisy $P(S) << P(N)$,
the experiment must be very selective $P(E|S) >> P(E|N)$.

Bayes' theorem for continuous random variables

Starting from usual Bayes formula:

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_j P(E|H_j)P(H_j)}$$

We get for continuous random variables

$$P(\nu|n) = \frac{P(n|\nu)P(\nu)}{\int_0^\infty P(n|\nu)P(\nu)d\nu}$$

For example in case of Poisson distribution

$$P(n|\nu) = \frac{\nu^n e^{-\nu}}{n!}$$

used in counting experiments, and assuming $P(\nu)$ to be constant we get

$$P(\nu|n) = \frac{\nu^n e^{-\nu}}{n!}$$

Bayesian Statistics

Typically **the goal of the signal analysis** is to:

- exclude as strongly as possible the existence of signal in its absence, or
- confirm as strongly as possible the existence of true signal

while minimizing the probabilities of

- falsely excluding a true signal, or falsely discovering a fake signal.

In **statistics** we use a data sample make inferences about a probabilistic model

- to asses the model's validity, or
- to determine the model parameters.

Frequentist statistics provides the usual tools used for reporting the outcome of an experiment, yet it does not define a probability for a hypothesis/parameter.

In Bayesian statistics probability, though, we can be speak of a **pdf for a parameter**, so it provides:

- a framework for estimating the validity of different model interpretations of the data, and
- a natural way to include additional information such as physical boundaries.

Hypothesis tests (1/2)

Consider an experiment characterized by a vector **x**, which might represent HEP events.

A H_0 statement about the distribution of **x** could represent the signal hypothesis and alternative H_1 could represent background process.

While **hypothesis test** provide a rule for accepting/rejecting hypotheses, **significance test** gives

- a probability to reject H_0 if it is true at **significance level** α.
- The quantity 1-β is called the **power** of the test to reject H_1.

In HEP terminology, the probability to accept the signal hypothesis, H_0, is the selection efficiency, i.e., $1 -$ significance level $= 1 - \alpha$.

Hypothesis tests (2/2)

The Neyman-Pearson lemma states that on can **minimize the background efficiency for a given signal efficiency** by defining the acceptance region such that for $\mathbf{x}$ in that region, the ratio of pdfs for the hypotheses

$$\lambda(\mathbf{x}) = \frac{f(\mathbf{x}|H_0)}{f(\mathbf{x}|H_1)}$$

is greater than the desired signal efficiency ϵ.

- If $\lambda > \epsilon$ we accept signal hypothesis H_0.

It is difficult to determine the ratio $\lambda(\mathbf{x})$, requiring the knowledge of joint pdfs $f(\mathbf{x}|H_0)$ and $f(\mathbf{x}|H_1)$.

There exist other multivariate classifiers that can separate different types of events:

- neural networks, and
- Fisher discriminants.
- **Recent classification methods** (see e.g. ROOT TMVA package) include:
  - support vector machines and decision trees.
- Methods such as **boosting** and **bagging** can be applied to combine weak classifiers into a stronger one with greater stability with respect to training data fluctuations.

(More details in Lecture 4.)

Goodness of fit[5]

Goodness of fit, GOF, of a statistical model describes how well it fits a set of observations, see e.g. Pearson's chi-square test.

P-value is the probability of obtaining a result as compatible with model as the observed value, when the (MC) experiment is repeated many times, assuming true H_0.

In a general procedure to prepare GOF for p-value, we need:

- a **test statistics** t, which is a function that measure **the 'distance' between the data x and the** H_0, and
- a function to map the value of the test statistic into a p-value.
  - That is, a way to calculate the probability of **exceeding** the observed value of the t, for H_0.

If the data $\mathbf{x}$ are discrete, and the test statistic $t = t(\mathbf{x})$, ($t_0 = t(\mathbf{x_0})$ for the data $\mathbf{x_0}$), the p-value would be:

$$p_{\mathbf{x}} = \sum_{\mathbf{x}: t \geq t_0} P(\mathbf{x}|H_0).$$

[5] Adapted from F. James: *Statistical Methods in Experimental Physics*, 2nd Ed. World Scientific, 2006.

Example: p-value in a counting experiment

A theory predicts the decay rate of a radioactive sample to be $\mu = 17.3$ *decays*$/h$, and we measure $N = 12$ *decays*$/h$.

Q: Is the measurement compatible with the theory?

Solution: We choose the test statistics to be the absolute difference $t = |n - \mu|$, so $t_0 = |N - \mu| = 5.3$. We use the Poisson distribution to calculate the p-value:

$$p = \sum_{\mathbf{n}:|n-\mu|\geq 5.3} \frac{e^{-\mu}\mu^n}{n!} = \sum_{n=1}^{12} \frac{e^{-17.3}17.3^n}{n!} + \sum_{n=23}^{\infty} \frac{e^{-17.3}17.3^n}{n!} = 0.23$$

Observation N is not significantly different from the theoretical prediction μ, since we have a 23% probability to measure the decay this far from the expected value.

This result can be confirmed with simple MC simulation that takes samples from the Poisson distribution.

Converting the p-value into a number of standard deviations [6]

Very small p-values have little intuitive value.

Thus it is traditional to convert p-values into z-values using an error function.

- z is the number of standard deviations beyond which the integral of the tail on one side of a Gaussian distribution equals p.

$$p = \frac{1}{2}(1 - erf(\frac{z}{\sqrt{2}}))$$

The conversion is applied regardless of whether the underlying test is one-sided or two-sided.

For example a p-value 2.87e-7 corresponds to 5 sigma's.

[6] Adapted from CMS TWiki: Recommended Procedure for Searching for New Physics at CMS (11 July, 2009)

Confidence intervals[7]

For a Gaussian estimator the result of an experiment is usually expressed by

- the parameter's estimated value, plus/minus an estimate of the **standard deviation**, $\hat{\theta} \pm \sigma_{\hat{\theta}}$

If the pdf is not Gaussian, or in the presence of physical boundaries,

- one usually quotes instead an **interval**.

The quoted interval or limit should

- objectively communicate the result of the experiment,
- **communicate incorporated prior beliefs** and relevant assumptions,
- provide interval that covers the true value of the θ with specified probability,
- make possible to draw conclusions about the parameter.

These goals are satisfied in **case of large data sample** by $\hat{\theta} \pm \sigma_{\hat{\theta}}$, and in the multi-parameter case by

- the parameter estimates and covariance matrix.

For **small data sample**, or in case of constrained variables, the Bayesian or the Neyman approach can be used.

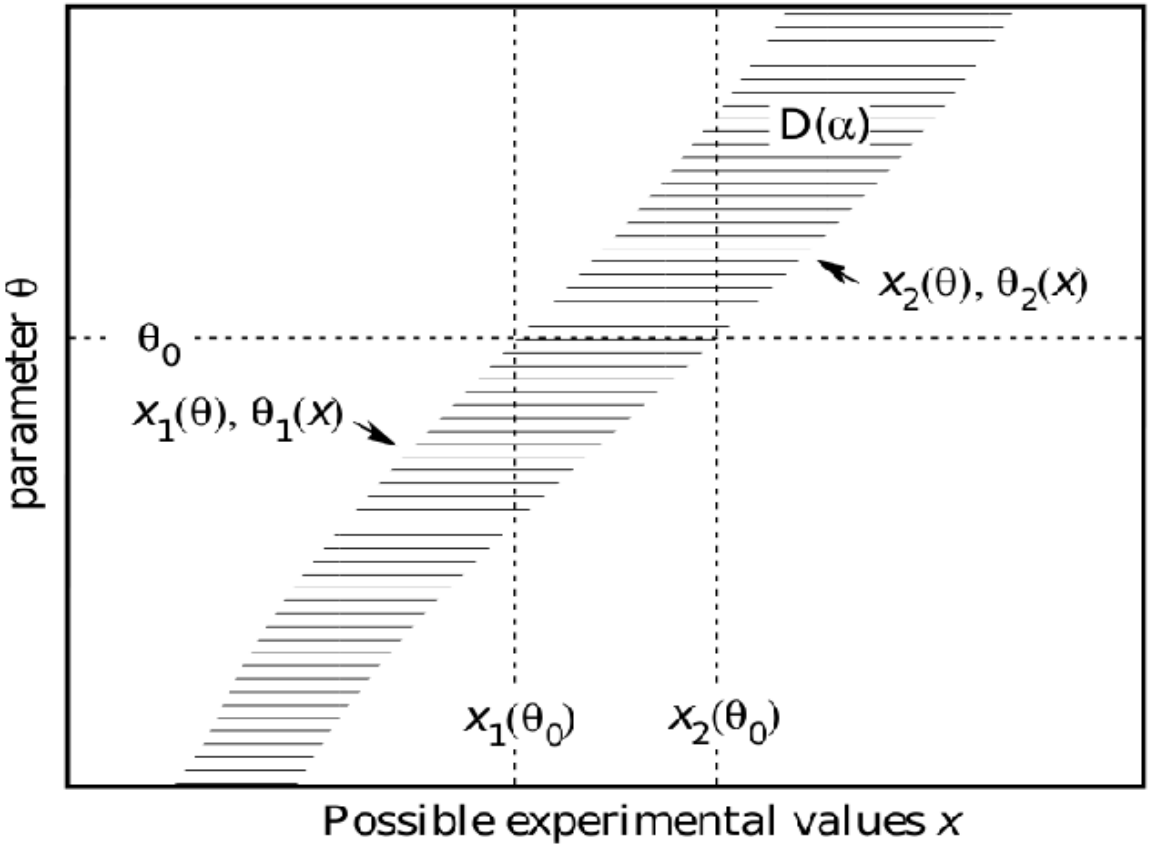
[7] Adapted from Particle Data Group.

Neyman Confidence Interval (1/2)

Using frequentist approach Neyman defines confidence interval to unknown parameter θ:

$$P(x_1 < x < x_2; \theta) = 1 - \alpha = \int_{x_1}^{x_2} f(x; \theta) dx,$$

where 1-α is pre-specified probability and x is measurement.

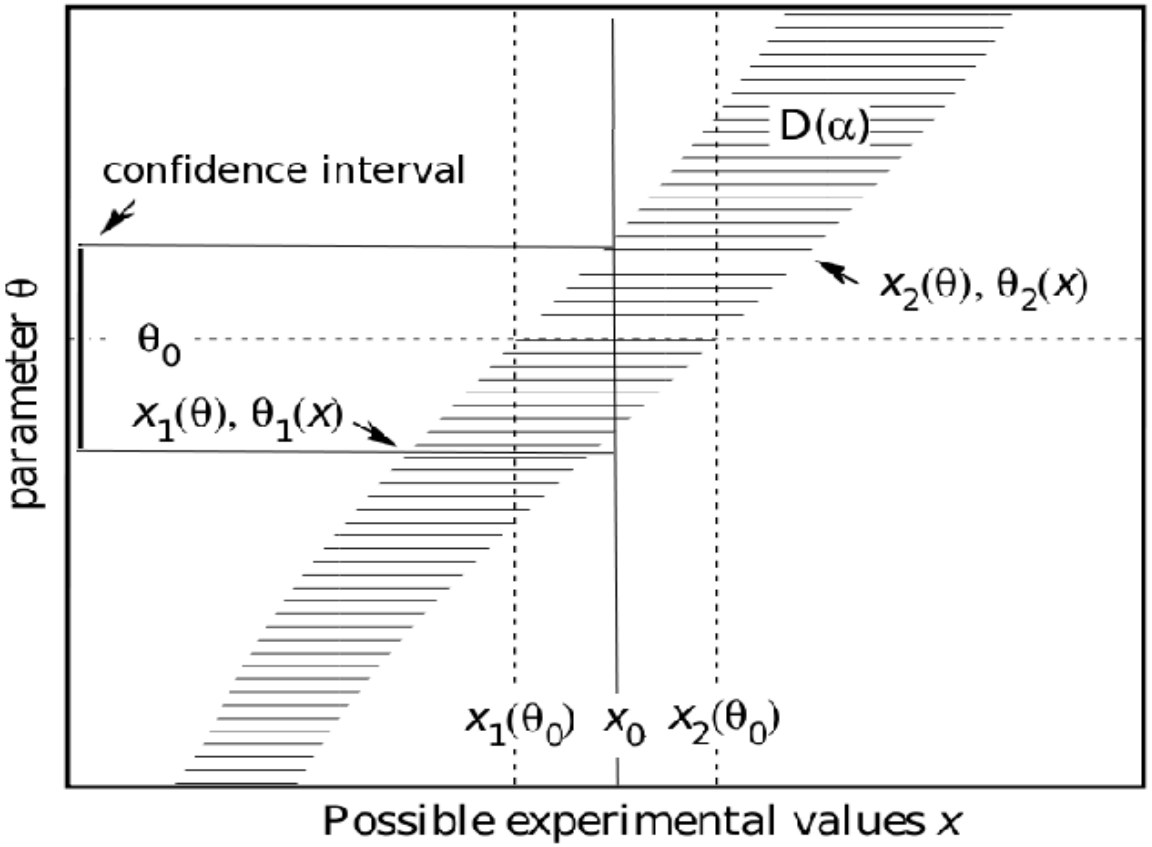


Union of $[x_1, x_2]$ segments for all values of θ is know as the **confidence belt**, D(α).

(Image courtesy of Particle Data Group.)

Neyman Confidence Interval (2/2)

Determining **confidence interval for** θ:



- Draw a vertical line through a measurement x_0.
- Inspect where the confidence belt is intercepted by the line.

Bayesian Confidence Intervals

In Bayesian statistics, all knowledge about parameter $\boldsymbol{\theta}$ is summarized by the posterior pdf $p(\boldsymbol{\theta}|\boldsymbol{x})$,

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{L(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int L(\boldsymbol{x}|\boldsymbol{\theta}')\pi(\boldsymbol{\theta}')d\boldsymbol{\theta}'}$$

which gives the degree of belief for $\boldsymbol{\theta}$ to have values in a certain region given the data $\boldsymbol{x}$.

- $\pi(\boldsymbol{\theta})$ is the prior pdf for $\boldsymbol{\theta}$, reflecting experimenter's subjective degree of belief about $\boldsymbol{\theta}$ before the measurement.
- $L(\boldsymbol{x}|\boldsymbol{\theta})$ is the likelihood function, i.e. the joint pdf for the data given a certain value of $\boldsymbol{\theta}$.
  - $L(\boldsymbol{x}|\boldsymbol{\theta})$ **should be published whenever possible**, to enable readers to calculate their own posterior pdf.
- The denominator simply normalizes the posterior pdf to unity.

Example: Non-negative constraint of a Poisson variable

Consider Poisson variable n which counts known background event with mean b, and unknown signal events with mean s constrained to be non-negative using the prior pdf

$$\pi(s) = \begin{cases} 0 & \text{if } s \leq 0 \\ 1 & \text{if } s > 0. \end{cases}$$

The likelihood function for Poisson distributed n is

$$L(n|s) = \frac{(s+b)^n}{n!} e^{-(s+b)}.$$

An upper limit s_+ at **credibility level** 1-α can be obtained by requiring

$$1 - \alpha = \int_{-\infty}^{s_+} p(s|n)ds = \frac{\int_{-\infty}^{s_+} L(n|s)\pi(s)ds}{\int_{-\infty}^{\infty} L(n|s)\pi(s)ds}$$

If $b = 0$ the equation reduces to the quantile of the χ^2 distribution

$$s_+ = \frac{1}{2} F^{-1}_{\chi^2}(1 - \alpha; n_d),$$

where $n_d = 2(n+1)$ is the number of degrees of freedom.
(*e.g.* `0.5*TMath::ChisquareQuantile(0.95,2*(10+1))` ≈ 17.0.)

Example: Bayesian Confidence Intervals in counting experiment (1/2)

Consider a counting experiment, where the number of events, $n = s + b$ is measured.

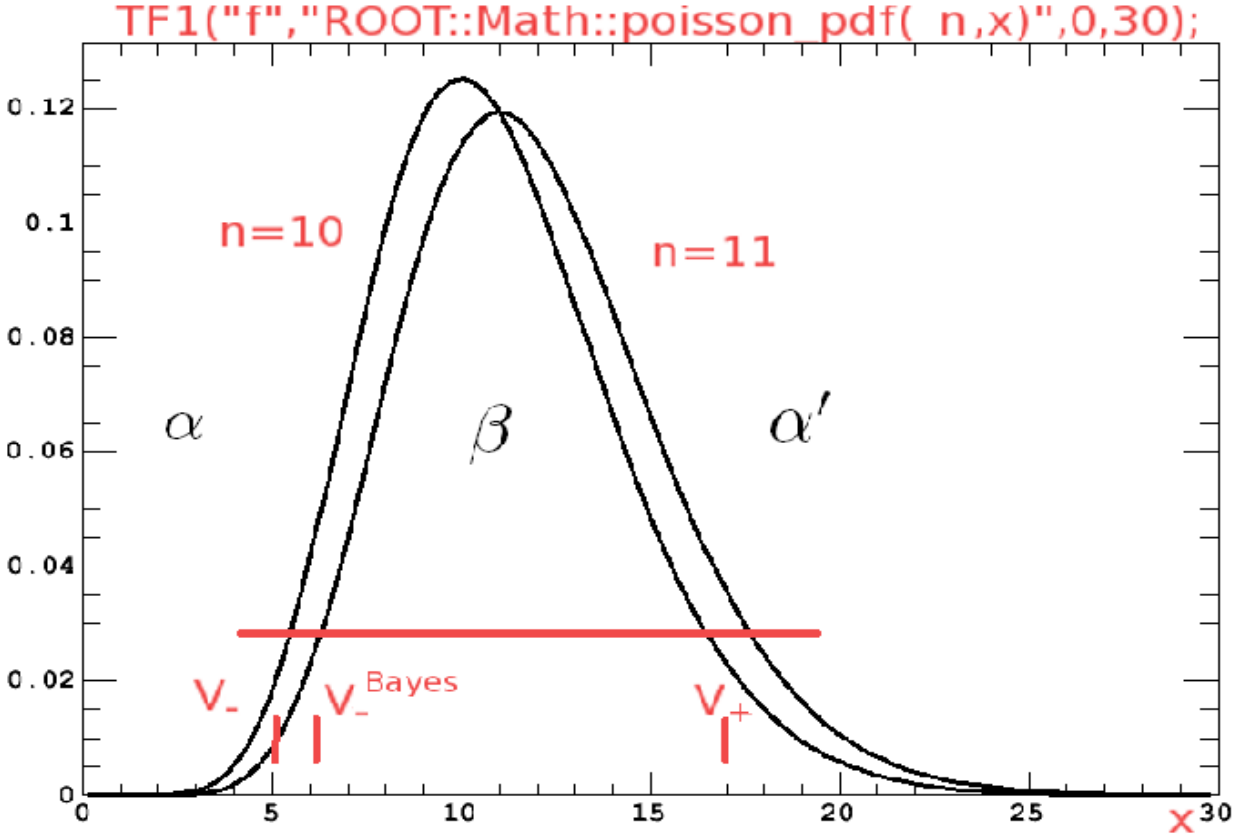
Given the posterior $P(\nu|n)$ we now choose a range $[\nu_-, \nu_+]$, for which

$$\int_{\nu_-}^{\nu^+} P(\nu|n)d\nu = \beta = \text{confidence level.}$$

Further, in our Bayesian approach we choose $P(\nu_-|n) = P(\nu_+|n)$.

- If $\alpha = \int_0^{\nu^-} Pd\nu$ and $\alpha' = \int_{\nu_+}^{\infty} Pd\nu = 1 - \beta - \alpha$,
  - symmetrical situation $\alpha = \alpha'$ is not valid in general.
- If plenty of data is available ($n > 20$) classical and Bayesian methods converge.
  - For small n, the specific method used should be reported.

Example: Bayesian Confidence Intervals in counting experiment (2/2)



In counting experiment $n = 10$ events were recorded. What are the upper and lower limits at the $\beta = 90\%$ confidence level for this measurement?

Lower limit is adjusted in Bayesian approach:

- $\nu_{-}^{Bayes}(n) = \nu_{-}(n+1)$
- $\nu_{+}^{Bayes}(n) = \nu_{+}(n)$

Classical: $5.4 < \nu < 17$

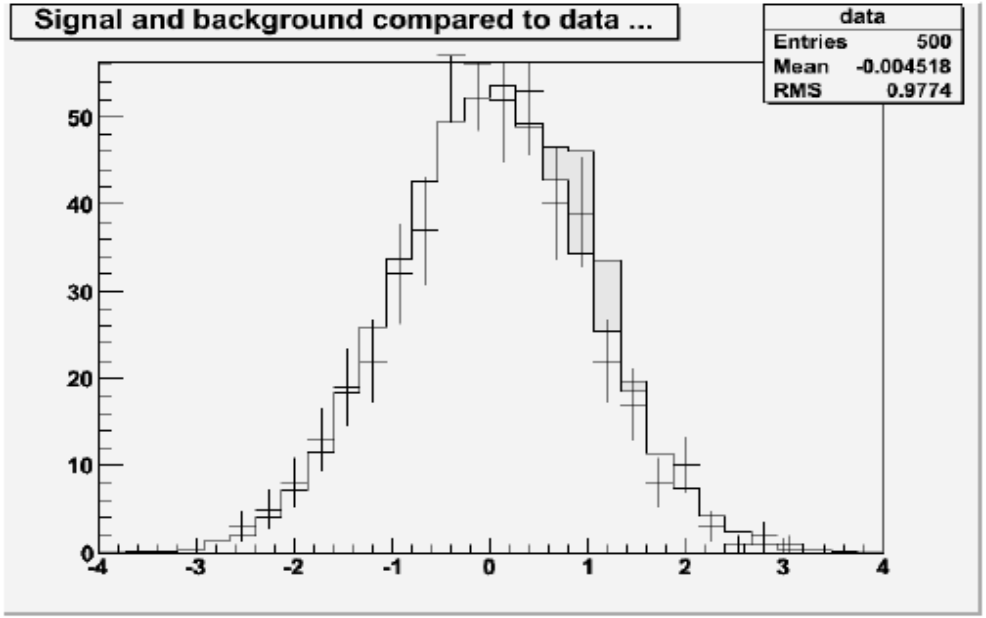
Bayes[8]: $6.2 < \nu < 17$

[8] Adapted from B. Escoubès, *Probabilités et statistiques à l'usage des physiciens*, Ellipses, 1998. For more details see CSC'09 exercises

Example: Computing confidence limits with `limit.C` (1/5)

`TLimit` computes 95% CL limits using semi-Bayesian Likelihood ratio[a].

- Signal and background histograms are wrapped in a `TLimitDataSource` as input.
- Monte Carlo experiments are run in order to compute the limits.




```
Computing limits...
CLs     : 0.0179
CLsb    : 0.0093
CLb     : 0.5181
< CLs >   : 0.0165
< CLsb > : 0.0082
< CLb >   : 0.5000
Computing limits with stat systematics...
Computing limits with systematics...
```

[a] T. Junk NIM A434, 1999, p.435-443

Example: Computing confidence limits with `limit.C` (2/5)

```
// Computation of 95 % C.L. limits
// in the precence of statistical error.
// Based on root_v5.21.06/tutorials/math/limit.C by
// Christophe.Delaere@cern.ch
// Modified for CSC'09 by A.Heikkinen 090705
#include "THStack.h"
#include "TCanvas.h"
#include "TLimit.h"
#include "TLimitDataSource.h"
#include "TConfidenceLevel.h" [...]
void CLwithErrors() {
  TCanvas *c1 = new TCanvas("c1","Dynamic Filling"
      ,200,10,700,500);
  TH1D *b = new TH1D("b","Background", 30,-4,4);
  TH1D *s = new TH1D("s","Signal"    , 30,-4,4);
  TH1D *d = new TH1D("d","Data points",30,-4,4);

[...]
```

Example: Computing confidence limits with `limit.C` (3/5)

```
[...]
TRandom2 r; // MC data generation
  Float_t bg, sig, dt;

  for (Int_t i = 0; i < 25000; i++) {
     bg  = r.Gaus(0, 1); sig = r.Gaus(1, .2);
     b->Fill(bg, 0.02); s->Fill(sig, 0.001);
  }
  for (Int_t i = 0; i < 500; i++) {
     dt = r.Gaus(0, 1); d->Fill(dt);
  }
  THStack *hs = new THStack("hs", "Signal and
     background compared to data ...");
  hs->Add(b); hs->Add(s); hs->Draw("hist");
  d->Draw("PE0, Same");

[...]
```

Example: Computing confidence limts with `limit.C` (4/5)

```
[...]
  cout << "Computing limits... " << endl;
  TLimitDataSource *dSrc=new TLimitDataSource(s,b,d);
  TConfidenceLevel *cl=TLimit::ComputeLimit(dSrc,
     50000);
  cout <<"CLs    :"    << cl->CLs()                  <<endl;
  cout <<"CLsb   :"    << cl->CLsb()                 <<endl;
  cout <<"CLb    :"    << cl->CLb()                  <<endl;

  cout <<"< CLs  > : "<< cl->GetExpectedCLs_b()  <<endl;
  cout <<"< CLsb >: "<< cl->GetExpectedCLsb_b()<<endl;
  cout <<"< CLb  > : "<< cl->GetExpectedCLb_b()  <<endl;

  // Show canonical -2lnQ histogram:
  //    for B hypothesis (full) and
  //    for S and B hypothesis (dashed).
  cl->Draw();

[...]
```

Example: Computing confidence limts with `limit.C` (5/5)

Inputs can be fluctuated according to systematics:

```
TObjString n1("B uncertainty");
TObjString n2("S uncertainty");
names->AddLast(&n1); names->AddLast(&n2);

// B error 1 (5%) and 2 (0%)
eb[0]=0.05; eb[1]=0;

// S error 1 (0 ) and 2 (1 )
es[0]=0;     es[1]=0.01;

TLimitDataSource *dSrcE = new TLimitDataSource();
dSrcE->AddChannel(s, b, d, &es, &eb, names);
```

(More details in exercises.)

Uncertainty in physics

The sources of uncertainty in measurement[9]:

- **incomplete definition** of the measurand; or its imperfect realization
- **non-representative sampling**
- inadequate knowledge of the effects of environmental conditions; or imperfect measurements of these conditions
- **personal bias** in reading instruments
- **finite instrument resolution**
- inexact values of measurement standards and reference materials
- **inexact values of constants** and other parameters obtained from external sources and used in the data-reduction algorithm
- **approximations and assumptions** incorporated in the measurement procedure
- **variations of repeated observations** of the measurand under apparently identical conditions

[9] Adapted from the The International Organization for Standardization (ISO) Guide to the Expression of Uncertainty in Measurement.

Optimal presentation of search results

Optimal presentation of search results has some desired properties[10]:

- **Uncertainties due to systematic effects should be included in a clear and consistent way.**
  - Often it is useful to quote the statistical and systematical error separately, e.g. $\sigma = 45 \pm 4 \pm 1\ mb$.
- The result should summarize completely the experiment; so that no extra information should be required for further analysis.
- Results should be easily turned into probabilistic statements.
- Analysis should be transparent, and result should be stated in such a way that it cannot be misleading. The presentation of the result should not depend on the particular application.
- **If possible full pdf-distributions and even data sets can be attached into analysis results**.
- In **unified approach to data analysis**, the transitions between exclusion, observation, discovery, and measurement are kept as small as possible.

[10] Adapted from F. James, *Workshop on Confidence Limits*, CERN-2000-005, 2000.

Example: Propagation of errors/uncertainty

Your 16 GB MicroSD has package size (HxWxD) (mm) 15 x 11 x 1. Manufacturer claims 0.1 mm deviation from the nominal dimensions.

Estimate you cards volume V = H W D and its uncertainty?

V = H W D = 15 mm 11 mm 1 mm = 165 mm^3. Assuming uncorrelated variables we have

$$(\frac{\Delta V}{V})^2 = (\frac{\Delta H}{H})^2 + (\frac{\Delta W}{W})^2 + (\frac{\Delta D}{D})^2$$

we have: $(\frac{\Delta V}{165})^2 = (\frac{0.1}{15})^2 + (\frac{0.1}{11})^2 + (\frac{\Delta 0.1}{1})^2 \approx 0.01$.

$V \pm \Delta V = 165 \pm 17\ mm^3$.

Note: Commonly the error on a quantity, here ΔV, is given as the standard deviation, σ. If the statistical probability distribution of the variable is assumed to be a normal distribution, there is a 68% probability that the true value of volume lies in the region $V \pm \Delta V$.

Propagation of errors - Formal treatment

Consider

- a set of n quantities $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)$,
- a set of m functions $\boldsymbol{\eta(\theta)} = (\eta_1(\boldsymbol{\theta}), ..., \eta_n(\boldsymbol{\theta}))$, and
- $V_{ij} = cov[\hat{\theta}_i, \hat{\theta}_j]$. (For uncorrelated quantities $V_{i \neq j} = 0$).

The purpose of **error propagation** is to determine the covariance matrix $U_{ij} = cov[\hat{\eta}_i, \hat{\eta}_j]$, where $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}(\hat{\boldsymbol{\theta}})$, for the $\boldsymbol{\eta(\theta)}$ functions.

This can be done expanding the functions $\boldsymbol{\eta(\theta)}$ about the estimates $\hat{\boldsymbol{\theta}}$ to first order in a Taylor series:

$$U_{ij} \approx \sum_{k,l} \frac{\partial \eta_i}{\partial \theta_k} \frac{\partial \eta_j}{\partial \theta_l} \bigg|_{\hat{\theta}} V_{kl}. \qquad (1)$$

Using matrix notation $U \approx AVA^T$, where $A_{ij} = \frac{\partial \eta_i}{\partial \theta_j} \Big|_{\hat{\theta}}$.

Note: In general case our previous example:
$(\frac{\Delta V}{V})^2 = (\frac{\Delta H}{H})^2 + (\frac{\Delta W}{W})^2 + (\frac{\Delta D}{D})^2 +$ correlation term due to $V_{i \neq j} \neq 0$

Example: MC estimation of errors

If Taylor series aproximation (1) fails, errors can be estimated using Monte-Carlo simulation. A real life example from neutrino physics[11], where Data Handling

deparment at CERN had to estimate value of

$$R = \frac{a}{\frac{d}{ke}(b-c) - 2(1-\frac{k^2 d}{ke})a}$$

$a = 3.84 \pm 1.33$

$b = 74 \pm 4$

$c = 9.5 \pm 3$

$d = 0.112 \pm 0.009$

$e = 0.320 \pm 0.002$

$k = 0.89$

Using (1) one finds $\hat{R} = 0.191 \pm 0.073$. What is the probability P that R is different from theoretical value $R_{th} = 0.42$? If gaussian form is assumed

$$P = \int_{\mathbf{3.14}}^{+\infty} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} = 8 \times 10^{-4}$$

`P=ROOT::Math::gaussian_cdf_c(3.14)` as introduced in `ProbFuncMathCore.h`

$(R_{th} - \hat{R})/\sigma_R = (0.42 - 0.191)/0.073 = \mathbf{3.14}$

Monte-Carlo simulation gives a quite different result. (Find it yourself in the exercises).

[11] From B. Escoubès, *Probabilités et statistiques à l'usage des physiciens*, Ellipses, 1998.

Unfolding: Hubble's Contact Lens

Immediately after Hubble went into orbit 1990, it became clear that something was wrong: images were blurred.
Hubble's primary mirror, polished so carefully, was just slightly the wrong shape.

In 1993 the Corrective Optics Space Telescope Axial Replacement (COSTAR) was installed to Hubble.

COSTAR, which was essentially a contact lens for Hubble's eye, consist of five small mirrors that corrected the defect.

This example shows how, corrective actions can be made to analyse measurements distorted by bias if the experimental setup is understood in detail.
Correspondingly, in data analysis we define a response function in order to unfold distorted measurements.

Unfolding – introduction to inverse problems [12]

The transformation of the raw measured data to the measured distribution is called unfolding:

- Linear inversion: measurement y $\longrightarrow$ x unfolded distribution.

Unfolding is an linear inverse problem with a coefficient matrix, and it is usually ill-conditioned:

- Unfolding is a complex mathematical operation and requires a good understanding of the experimental setup:
  - A correct determination of the response matrix (e.g. by Monte Carlo) is essential.

In addition to **image enhancement** and anti-blurring unfolding is of great importance in many different fields:

- **tomography in medicine** (using X-rays, positron-electron annihilation reactions, ultrasound),
- geophysics, astrophysics, and
- HEP (e.g. particle *energy spectra reconstruction* from measured pulse-height distributions.)

[12] According to V. Blobel and his lectures in CSC'84

Unfolding

The mapping of input x to output y in a system

$$\int_{\Omega} \text{system} \times \text{input}\ d\Omega = \text{output}$$

expressed with Fredholm integral

$$\int A(y,x) f(x)\ dx = g(y)$$

. With discrete x and y we get a linear equation: $\mathbf{Ax} \approx \mathbf{y}$.

- $\mathbf{x}$ = n-histogram of true variable x
- $\mathbf{y}$ = m-histogram of measured variable y
  (contains measurement errors, or is Poisson distributed)
- $\mathbf{A}$ = $m \times n$ response matrix, corresponds the A(y,x) response function
  - A_{ij} is the probability for an event originating from bin j of x, to be observed in bin i of y.

Unfolding: solving the equation

Direct solution of $\mathbf{Ax} \approx \mathbf{y}$:

- Typically not feasible since the problem is ill-conditioned.
- A small perturbation of the data can cause an arbitrary large perturbation of the solution.

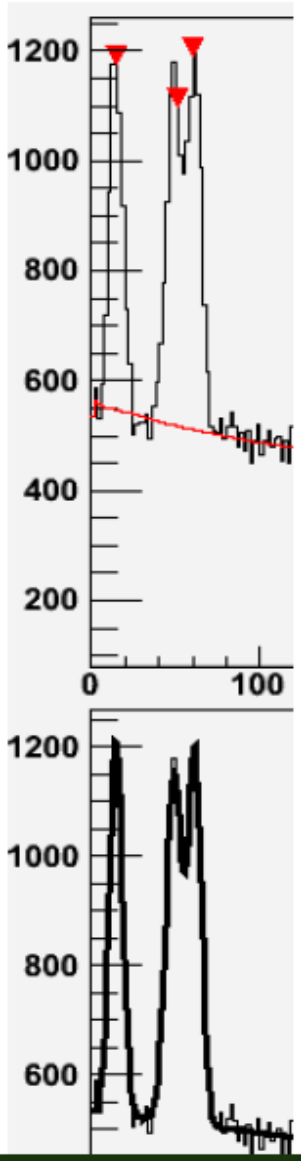
Iterative method:

- Relatively simple method, which is popular for unfolding, calibration, and alignment.
- If a single computational step is not sufficient.
- Needs a start vector $\mathbf{x}_0$
- Only method, which is applicable for large number of parameters (e.g. LHC detector alignment).

Regularization method:

- Is typically more complex to implement.
- Key idea is to incorporate assumptions about the size and smoothness of the solution.

Unfolding example: spectral line analysis 1/2



`TSpectrum` class contains spectra processing functions for one-dimensional:

- background estimation,
- smoothing,
- deconvolution method to unfold/reverse the effects of convolution on measured data, and
- peak search.

```
// Use TSpectrum to find the peak candidates
TSpectrum *s = new TSpectrum(2*npeaks);
Int_t nfound = s->Search(h,2,"",0.10);
printf("Found %d candidate peaks to fit\n",nfound);
```

Unfolding example: spectral line analysis 2/2

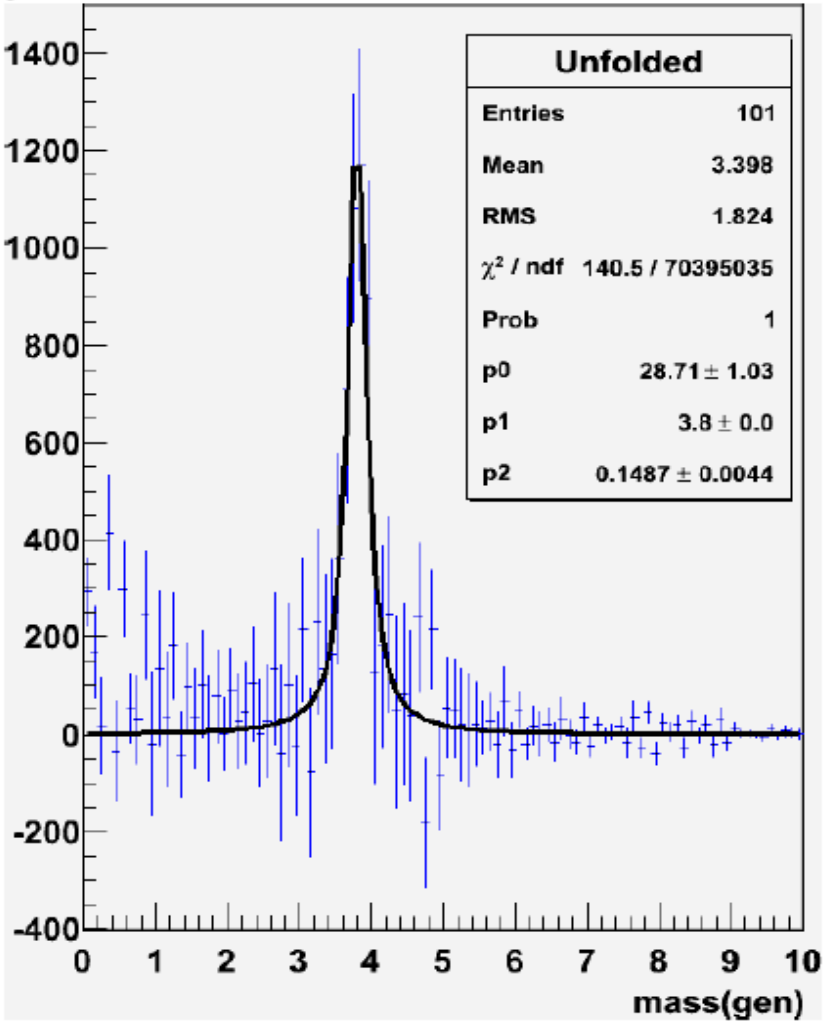
```
//Loop on all found peaks.
par[0] = fline->GetParameter(0); par[1] =fline->GetParameter(1);
Float_t *xpeaks = s->GetPositionX();
for (p=0;p<nfound;p++) {
   Float_t xp = xpeaks[p];
   Int_t bin = h->GetXaxis()->FindBin(xp);
   Float_t yp = h->GetBinContent(bin);
   if (yp-TMath::Sqrt(yp) < fline->Eval(xp)) continue;
   par[3*npeaks+2]=yp; par[3*npeaks+3]=xp; par[3*npeaks+4]=3;
   npeaks++;
}
printf("Found %d useful peaks to fit\n",npeaks);
printf("Now fitting: Be patient\n");
TF1 *fit = new TF1("fit",fpeaks,0,1000,2+3*npeaks);
TVirtualFitter::Fitter(h2,10+3*npeaks);
fit->SetParameters(par); h2->Fit("fit");
```

(See CSC'09 exercises on unfolding.)

Solving the inverse problem with TUnfold

ROOT class TUnfold solves the unfolding problem using regularisation.
In CSC'09 data analysis exercises we study a case were:

- Monte Carlo data in generate for background
- Signal is a resonance generated with a Breit-Wigner distribution smeared by a Gaussian
- Data is unfolded using TUnfold giving:
  - The background level
  - The shape of the resonance, corrected for detector effects
- Fit is done to the unfolded distribution, including the correlation matrix



Summary

- An introduction to Bayesian statistics was made.
  - Examples on subjective Bayesian reasoning were given
  - It was stressed that if plenty of data is available, Bayesian results converge with classical method.
- Concept of hypothesis test was introduced.
  - Minimizing the background efficiency was studied using Neyman-Pearson lemma.
- P-value as an example of goodness of fit was discussed.
- The sources of uncertainty in physics measurement were briefly discussed.
  - A general method to propagate errors was shown.
- Examples were given on unfolding distorted measurements.
  - The importance of defining a response function was stressed.

(In exercises these topics are studied further using ROOT data analysis software.)