Data Processing (Datenverarbeitung)

37. Herbstschule für Hochenergiephysik 2005

Maria Laach

Stefan Kluth
Max-Planck-Institut für Physik
Föhringer Ring 6
D-80805 München
skluth@mppmu.mpg.de

Overview

- Statistics I
  - foundations, error calculus
- Statistics II
  - Monte Carlo, parameter estimation, unfolding
- OOAD I
  - complex systems, object model
- OOAD II
  - dependency management, class design

1 Basics

- Sample space S is a set
- Subset A of S has probability P(A)
  - for all A in S: P(A) $\geq$ 0
  - for A, B disjoint: P(A $\cup$ B)= P(A) + P(B)
  - P(S) = 1
- From this
  - P($\underline{A}$) = 1 – P(A) with $\underline{A}$ complement of A
  - P(A $\cup$ B) = P(A) + P(B) – P(A $\cap$ B)
  - P(A | B) = P(A $\cap$ B) / P(B)

1 Bayes Theorem

Consider conditional probabilities P(A | B) and P(B | A)

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

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

With P(A $\cap$ B) = P(B $\cap$ A) one gets:

P(A | B) = P(B | A) $\cdot$ P(A) / P(B)   Bayes Theorem

P(A) is called prior probability

1 Example

Disease carried by P(yes) = 0.001 of population; test for desease with binary outcomes + and – with $P(+|yes) = 0.98$, $P(+|no) = 0.03$

Probability to have desease if tested positive using Bayes theorem:

$$P(yes|+) = P(+|yes)\, P(yes) / P(+)$$

$$P(+) = P(+|yes)\, P(yes) + P(+|no)\, P(no)$$

$$P(yes|+) = 0.98 \cdot 0.001 / (\, 0.98 \cdot 0.001 + 0.03 \cdot 0.999\,) = 0.032$$

Why $P(yes|+)$ so small? Prior probability P(yes) is small and thus contribution from false positives $P(+|no)$ dominates

1 Frequentist Statistics

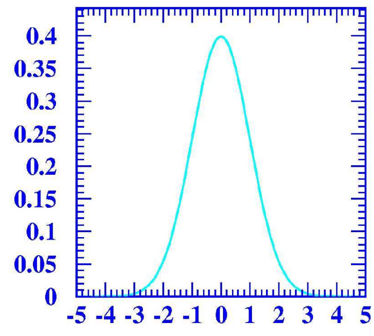
- Probability is relative frequency
  - $P(A) = \lim_{N \to \infty} n(A) / N$ with n times result A in N tries (experiments)
  - $P(A|B) = n(A \cap B) / n(B)$
- Corresponds well to quantum processes
  - predictions for ensembles, not single events
  - limited samples imply statistical uncertainties
  - correspondence estimator ↔ true value

1 Bayesian Statistics

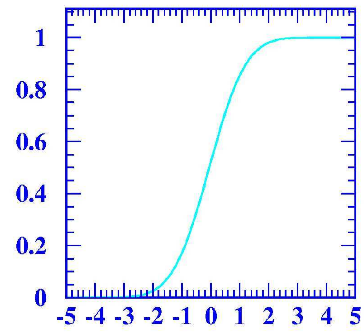
- Hypothesis testing
  - sample space consists of hypotheses
  - P(A) probability that hypothesis A correct
  - $P(A) = \lim_{N\to\infty} n(A) / N$ is also a hypothesis
- Bayes theorem relates hypothesis and data
  - $P(\text{theory} \mid \text{data}) \sim P(\text{data} \mid \text{theory})\, P(\text{theory})$
  - prior probability P(theory) subjective
  - Desease testing:
    - $P(\text{yes}) = 0.1 \Rightarrow P(\text{yes} \mid +) = 0.78$
    - $P(\text{yes}) = 0.001 \Rightarrow P(\text{yes} \mid +) = 0.03$

2 Probability Density Functions

- PDFs are functions of continuous random variables:
  - $P(x \in [x', x'+dx]) = f(x')dx$
- Normalisation:
  - $\int_S f(x)\, dx = \int_{-\infty}^{\infty} f(x)\, dx = 1$
- Cumulative distribution:
  - $F(x') = \int_{-\infty}^{x'} f(x)\, dx = P(x \leq x')$



probability density function



cumulative distribution

2 Mean, Variance, Covariance

Expectation value of random variable x (mean) defined as

$$E[x] = \int_{-\infty}^{\infty} x\, f(x)\, dx = \mu_x$$

Variance of x is $E[(x - \mu_x)^2]$:

$$\mathrm{var}(x) = E[(x - \mu_x)^2] = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x)\, dx = E[x^2] - \mu_x^{\,2}$$

$$\sigma_x = \sqrt{\mathrm{var}(x)}$$

Covariance is $E[(x - \mu_x)(y - \mu_y)]$:

$$\mathrm{cov}(x,y) = E[(x - E[x])(y - E[y])] = E[xy] - \mu_x\mu_y$$

$$= \iint xy\, f(x,y)\, dxdy - \mu_x\mu_y$$

2 Statistical Estimators

Estimators of expectation values in finite samples:

Mean: $\langle x\rangle = 1/n \sum_{i=1}^{n} x_i$ $E[\langle x\rangle] = \mu_x$

Variance: $\langle \mathrm{var}(x)\rangle = n/(n-1)\, (\langle x^2\rangle - \langle x\rangle^2)$ $E[\langle \mathrm{var}(x)\rangle] = \mathrm{var}(x)$

Covariance: $\langle \mathrm{cov}(x,y)\rangle = n/(n-1)\, (\langle xy\rangle - \langle x\rangle\langle y\rangle)$ $E[\langle \mathrm{cov}(x,y)\rangle] = \mathrm{cov}(x,y)$

Weak law of large numbers: $\langle x\rangle$ is consistent estimator of μ_x if var(x) exists

Correlation coefficient: $\rho_{xy} = \mathrm{cov}(x,y)/(\sigma_x\sigma_y)$ with $-1 < \rho_{xy} < 1$

3 Multinomial Distribution

In particle physics we “count events”, i.e. measure frequencies of outcomes and apply frequentist statistics to extract parameters

For N measured values of random variable x normalised histogram with m bins $[x_i, x_i + \Delta x]$ and n_i entries per bin is described by multinomial pdf:

$$f_M(\vec{n}, N; \vec{p}) = \frac{N!}{\prod_{i=1}^{m} (n_i!)} \prod_{i=1}^{m} p_i^{n_i}$$

Estimator for p_i: $\langle p_i \rangle = n_i / N$

Variance of p_i: $\langle var(p_i) \rangle = \langle p_i \rangle (1 - \langle p_i \rangle) / N$

Covariance of p_i: $\langle cov(p_i, p_j) \rangle = - \langle p_i \rangle \langle p_j \rangle / N$ for $i \neq j$

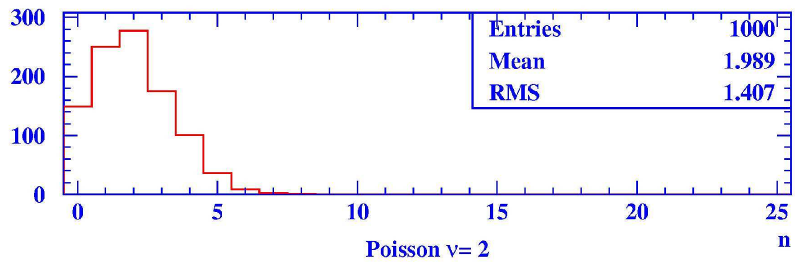
3 Poisson Distribution

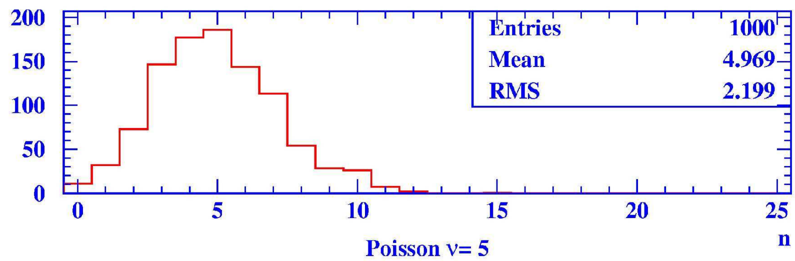
Frequency n of observations of specific event type given by Poisson distribution:

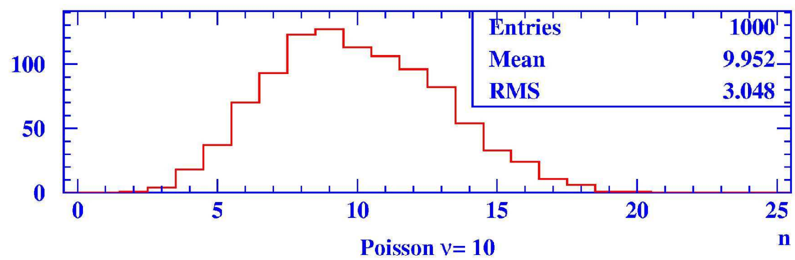
$f_P(n; \nu) = \nu^n / n!\, e^{-\nu}$

Expectation value:
$E[n] = \sum_{n=0}^{\infty} n\, \nu^n / n!\, e^{-\nu} = \nu$

Variance:
$var(n) = \sum_{n=0}^{\infty} (n-\nu)^2\, \nu^n / n!\, e^{-\nu}$
$= \nu$







3 Gauss Distribution

Normal or Gaussian distribution with mean μ and variance σ:

$$f_G(x; \mu, \sigma) = 1/\sqrt{(2\pi\sigma^2)} \exp(- 0.5 ((x-\mu)/\sigma)^2)$$

Central limit theorem: n random variables x_i with mean μ_i and variance σ_i; $X = \lim_{n\to\infty} \sum_{i=0}^{n} x_i$

$$\lim_{n\to\infty} f(X) = f_G(X; \sum_{i=0}^{n} \mu_i, \sum_{i=0}^{n} \sigma_i^2)$$

For $\mathbf{x} = (x_1, \ldots, x_n)$ with covariance matrix $V = (\mathrm{cov}(x_i, x_j))$ and expectation values $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$:

$$f_G(\mathbf{x}; \boldsymbol{\mu}, V) = 1/\sqrt{((2\pi)^n |V|)} \exp(- 0.5 (\mathbf{x}-\boldsymbol{\mu})^T V^{-1} (\mathbf{x}-\boldsymbol{\mu}))$$

3 χ^2 Distribution

The χ^2 distribution of a continous random variable z:

$$f_\chi(z; n) = z^{n/2-1} e^{-z/2} /(2^{n/2} \Gamma(n/2))$$

Parameter n often called “number of degrees of freedom (d.o.f.)

Mean $E[z] = n$ and Variance $\mathrm{var}(z) = 2n$

Important property:

$\chi^2 = (\mathbf{x}-\boldsymbol{\mu})^T \cdot V^{-1} \cdot (\mathbf{x}-\boldsymbol{\mu})$ or $\chi^2 = \sum_i ((x_i-\mu_i)/\sigma_i)^2$ if cov. matrix V diagonal

follows $F_\chi(z; n)$ → Hypothesis tests and parameter estimation

4 Error Calculus

Measurement errors of correlated and/or derived quantities

Measurements and errors described by random variables $\mathbf{x} = (x_1, ..., x_n)$ with covariance matrix $V = (\mathrm{cov}(x_i, x_j))$ and expectation values $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)$

Derived quantities: $\mathbf{y} = \mathbf{y}(\mathbf{x}) = (y_1(\mathbf{x}), ..., y_m(\mathbf{x}))$; $m \leq n$

Taylor expansion: $\mathbf{y}(\mathbf{x}) = \mathbf{y}(\boldsymbol{\mu}) + (d/d\mathbf{x} \cdot \mathbf{y}(\boldsymbol{\mu})^T)^T \cdot (\mathbf{x}-\boldsymbol{\mu})$

Jacobi matrix $(d/d\mathbf{x} \cdot \mathbf{y}(\boldsymbol{\mu})^T)^T = J$ describes changes in **y** due to deviations of measurements **x** from expectation values **μ**.
Elements of J are $J_{ij} = dy_j/dx_i$

4 Error Calculus

Covariance matrix W of **y** given by V and J:

$$W = E[\mathrm{cov}(\mathbf{y})] = E[\mathbf{y} \cdot \mathbf{y}^T] - E[\mathbf{y}] \cdot E[\mathbf{y}]^T$$

$$E[\mathbf{y}] = E[\mathbf{y}(\boldsymbol{\mu})] + J \cdot E[(\mathbf{x}-\boldsymbol{\mu})] = \mathbf{y}(\boldsymbol{\mu})$$

$$\begin{aligned} E[\mathbf{y} \cdot \mathbf{y}^T] &= E[\ (\mathbf{y}(\boldsymbol{\mu}) + J \cdot (\mathbf{x}-\boldsymbol{\mu})) \cdot (\mathbf{y}(\boldsymbol{\mu}) + J \cdot (\mathbf{x}-\boldsymbol{\mu}))^T\ ] \\ &= \mathbf{y}(\boldsymbol{\mu}) \cdot \mathbf{y}(\boldsymbol{\mu})^T + J \cdot E[(\mathbf{x}-\boldsymbol{\mu}) \cdot (\mathbf{x}-\boldsymbol{\mu})^T] \cdot J^T \end{aligned}$$

$$W = J \cdot V \cdot J^T$$

W is orthogonal transformation of V given by J
Errors of **y** given by diagonal elements of W

4 Error calculus examples

Error of product $y = x_1 x_2$:

$$W = J \cdot V \cdot J^T = (x_2 \quad x_1) \cdot \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \cdot \begin{pmatrix} x_2 \\ x_1 \end{pmatrix} = x_2^2\sigma_1^2 + x_1^2\sigma_2^2 + 2x_1x_2\rho\sigma_1\sigma_2$$

$$\left(\frac{\sigma_y}{y}\right)^2 = \left(\frac{\sigma_1}{x_1}\right)^2 + \left(\frac{\sigma_2}{x_2}\right)^2 + \frac{2\rho\sigma_1\sigma_2}{x_1x_2}$$

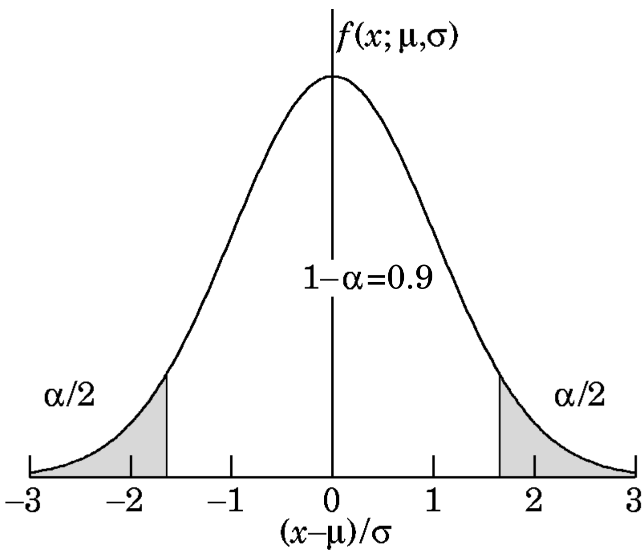
Error of sum $y = a_1x_1 + a_2x_2$:

$$W = J \cdot V \cdot J^T = (a_1 \quad a_2) \cdot \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\rho\sigma_1\sigma_2$$

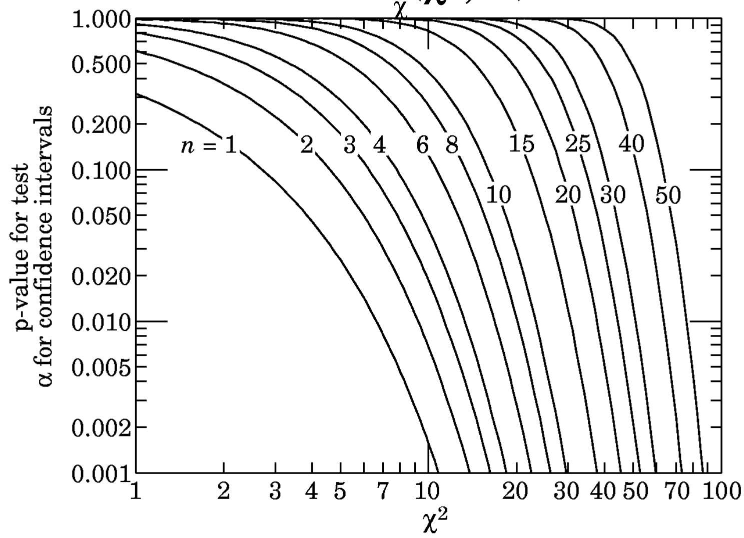
$$\sigma_y^2 = (a_1\sigma_1)^2 + (a_2\sigma_2)^2 + 2a_1a_2\rho\sigma_1\sigma_2$$

4 Error Interpretation

For normal distributed variables x errors σ_x^2 correspond to variance of Gaussian. By central limit theorem sum of many errors^2 also follows Gaussian. Coverage probability 1 - α given by $f_\chi(\chi^2, n)$ with $\chi^2 = \delta/\sigma$



$1 - f_\chi(\chi^2, n)$



5 Monte Carlo Methods

Monte Carlo: Simulation of physical processes with probability density functions (pdf) and random numbers

Example decays: $$f(t;\tau)=\frac{1}{\tau}e^{-\frac{t}{\tau}}$$

generate t_i such that distribution of t_i follows f(t; τ)

Random numbers: $n_{i+1} = \mathrm{mod}(a \cdot n_i, m)$

a = 40692, m = 2147483399

n_0 is seed, determines sequence of n_i

→ pseodo random numbers

5 Simulation of pdfs: Transformation method

Random variable r ∈ [0,1] uniform, transformed variable x(r) distributed according to f(x)

$$u(r)\,dr=f(x(r))\cdot dx$$

$$\rightarrow \int_{-\infty}^{r} u(r')\,dr'=r=\int_{-\infty}^{x(r)} f(x)\,dx$$

$$\rightarrow x(r)=F^{-1}(r)$$

draw r, evaluate $F^{-1}(r)$

⇒ x distributed as f(x)

Example: exponential distribution

$$f(t;\tau)=\frac{1}{\tau}e^{-\frac{t}{\tau}}\rightarrow r=F(t)=\int_0^t \frac{1}{\tau}e^{-\frac{t'}{\tau}}\,dt'=1-e^{-\frac{t}{\tau}}$$

$$t=-\tau\log(1-r)$$

5 Hit-or-miss Method

Problem: $f(x)$ in $[x_1, x_2]$, want x distributed as $f(x)$

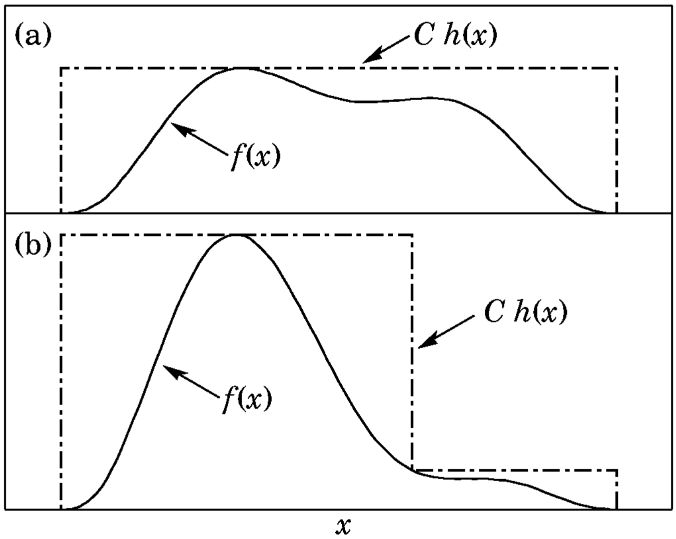
1) Generate r_1 in [0,1]

   $x = x_1 + r(x_2 - x_1)$

2) Generate r_2 in [0,1]

   $f_r = r_2 f_{max}$, accept if $f_r < f(x)$,

   or go back to 1)



+: universal algorithm

−: inefficient when $f_{max} \gg f_{min}$

5 Physics and Detector Simulation

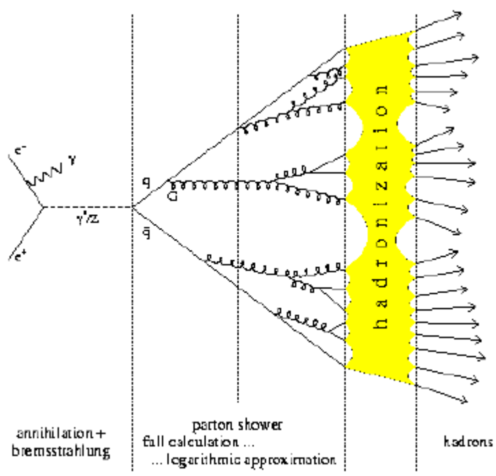
Physics simulation: Generation of 4-vectors, lifetimes final states etc. following known pdfs from theory, model or data

Example: event generators PYTHIA and HERWIG

$e^+e^- \to (Z/\gamma)^* (\gamma) \to q\bar{q}(g)$

$\to$ parton shower

$\to$ hadron/lepton decays



Detector simulation:

Response of apparatus

to 4-vectors from generator

5 Parametric Detektor Simulation

Measurement uncertainties parametrised
→ Simulation via "smearing" of input quantities

Example: simulation of tracking resolution

$$\frac{\sigma_{p_t}}{p_t} = \sqrt{0.02^2 + (0.0015 \cdot p_t[GeV])^2}$$

(OPAL Jet chamber)

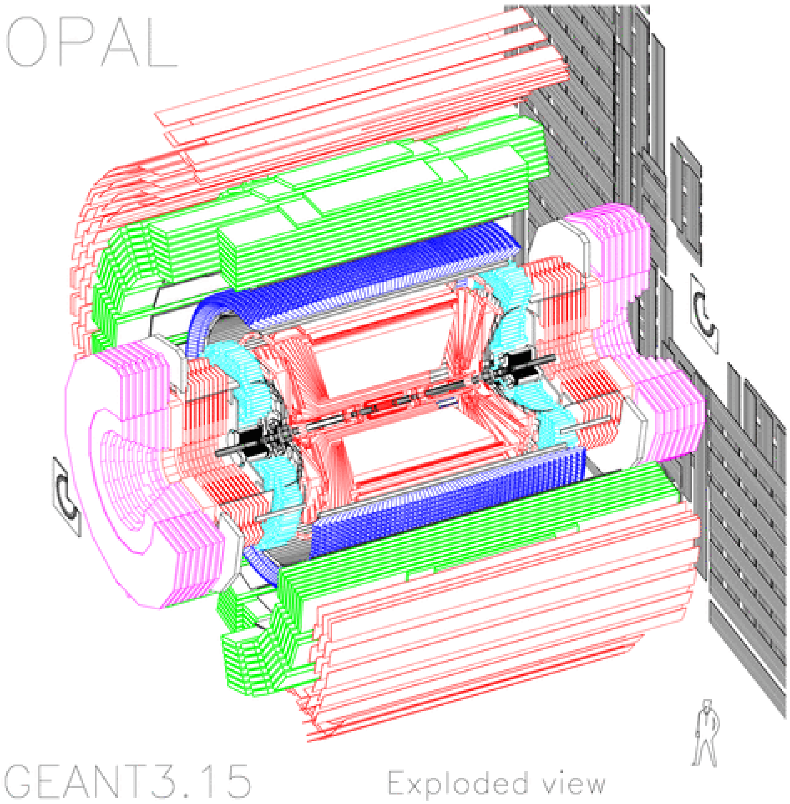
for given p_t simulate gaussian distribution:

$p_t' = p_t + \Delta p_t$

$$f(\Delta p_t) \sim e^{-\frac{1}{2}\left(\frac{\Delta p_t}{\sigma_{p_t}}\right)^2} \rightarrow \Delta p_t$$

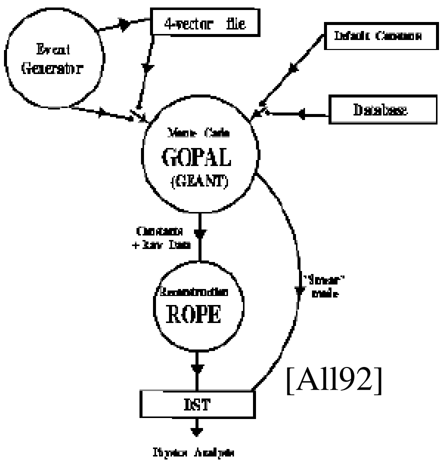
Use p_t' in following calculations

5 Detailed Detector Simulation



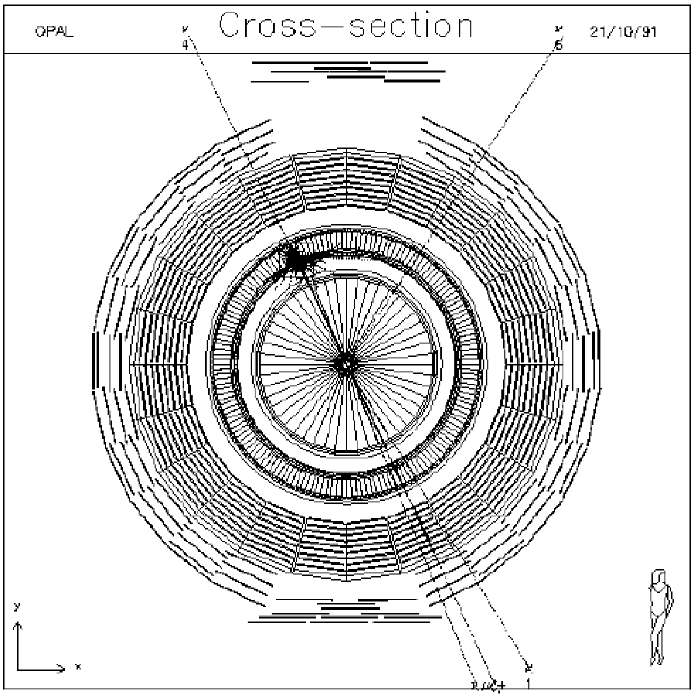
Detector geometry:
sensitive (active) elements
and passive material

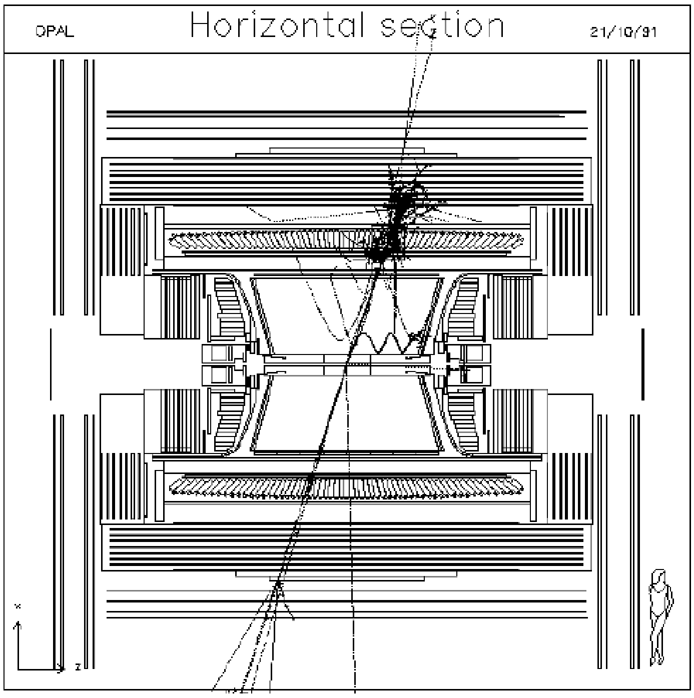
Detectorsimulation in dataflow of OPAL



[Al92]

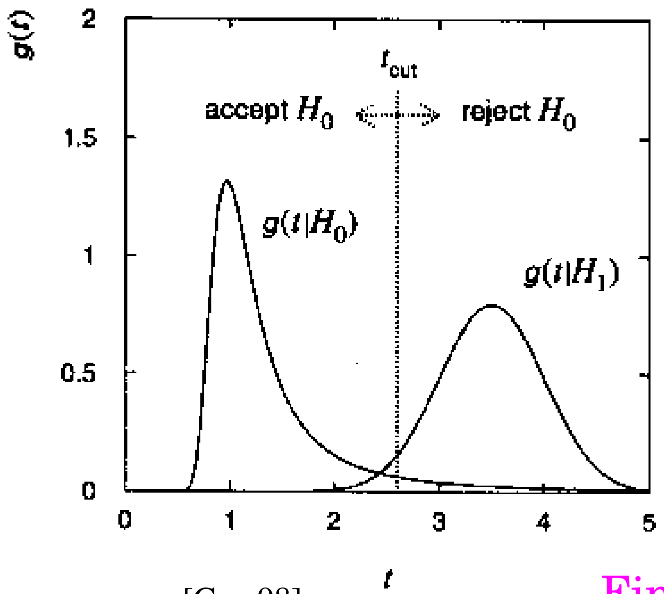
5 Example: GOPAL





[Al92]

6 Event Selection: Cuts



[Cow98]

Optimisation of t_{cut} by considering efficiency ε and purity p

$$\epsilon = \int_o^{t_{cut}} g(t\,;H_0)\,dt$$

$$p = \frac{\int_0^{t_{cut}} N_0\, g(t\,;H_0)\,dt}{\int_0^{t_{cut}} N_0\, g(t\,;H_0) + N_1\, g(t\,;H_1)\,dt}$$

Find ε and p from MC simulation of physics processes und detector response

Pragmatic approach: optimise ε · p

6 Likelihood Ratio

Construct test variables $\mathbf{x} = (x_1, \ldots, x_n)$ from reconstructed events (detector level): number of tracks, number of calorimeter objects, event shapes, total energy, vertices, etc.

Neyman-Pearson acceptance region optimises p for given ε:

$$t_L(\vec{x}) = \frac{g(\vec{x}; H_0)}{g(\vec{x}; H_1)} > c(\epsilon)$$

$g(\mathbf{x} | H_0)$ and $g(\mathbf{x} | H_1)$ from MC simulation with events of class 1 or 2 as multidimensional histograms → scales like m^n

6 Fisher Discriminant

Construct test variable t_F as linear function of $\mathbf{x}$ with expectation values $E[\mathbf{x}] = \boldsymbol{\mu}_k$ and covariance matrix V_k under hypothesis k

$$t_F(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x} \text{ with } E[t]_k = \mathbf{a} \cdot \boldsymbol{\mu}_k \text{ and } \sigma_k^2 = \mathbf{a}^T \cdot V_k \cdot \mathbf{a}$$

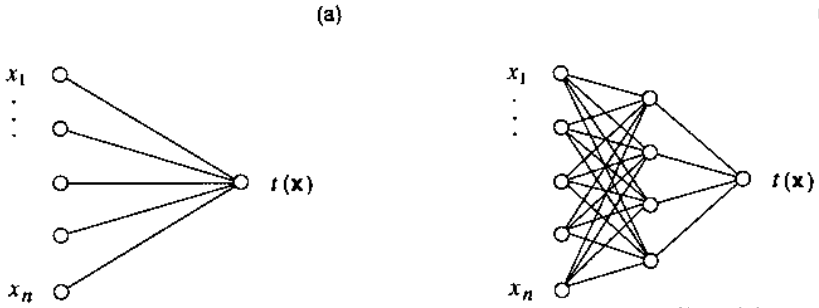
Maximise separation between two hypothesis 0 and 1:

$$J(\mathbf{a}) = (E[t]_0 - E[t]_1)^2 / (\sigma_1^2 + \sigma_2^2) = (\mathbf{a} \cdot (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1))^2 / (\mathbf{a}^T \cdot (V_0+V_1) \cdot \mathbf{a})$$

$$dJ/d\mathbf{a} = \mathbf{0} \Rightarrow (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1) - c(V_0+V_1) \cdot \mathbf{a} = \mathbf{0} \Rightarrow \mathbf{a} = 1/c\, (V_0+V_1)^{-1} \cdot (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)$$

Coefficients $\mathbf{a}$ determined from $E[\mathbf{x}] = \boldsymbol{\mu}_k$ and V_k in training samples.
$\mathbf{x}$ normal under hypotheses 0 and 1 $\Rightarrow t_F(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$ as good as $t_L(\mathbf{x})$

6 Neural Networks



Single Layer Perceptron

Two Layer Perceptron

[Cow98]

Linear combination

$$t_{SL}(\vec{x}) = s\left(a_0 + \sum_{i=1}^{n} a_i x_i\right)$$

Modeling of non-linear effects via hidden layer

$$t_{TL}(\vec{x}) = s\left(a_0 + \sum_{i=1}^{m} a_i s\left(w_{i0} + \sum_{j=1}^{n} w_{ij} x_j\right)\right)$$

$$s(z) = \frac{1}{1+e^{-z}}$$ Activation function

Training: find a_i und w_{ij} using training samples of class 1 or 2
usually from MC simulation

7 Parameter Estimation

Adjust theory or model with free parameters to data
by finding optimal values of the parameters

e.g. lifetimes: N measurements t_i, theory is:

$$f(t;\tau) = \frac{1}{\tau} e^{-\frac{t}{\tau}}$$

event shapes: distributions of event shape observables
in $e^+e^- \rightarrow$ hadrons
predictions by QCD, free parameter
is strong coupling constant $\alpha_S(M_Z)$

$1/\sigma d\sigma/dy = dA/dy\ \alpha_S(M_Z) + dB/dy\ \alpha_S(M_Z)^2$

7 Maximum Likelihood

Physical theory or model with pdf f(x, **a**); **x** are measurements, **a** are free parameters

$$L=\prod_{i=1}^{n} f(x_i,\vec{a})\,dx_i$$

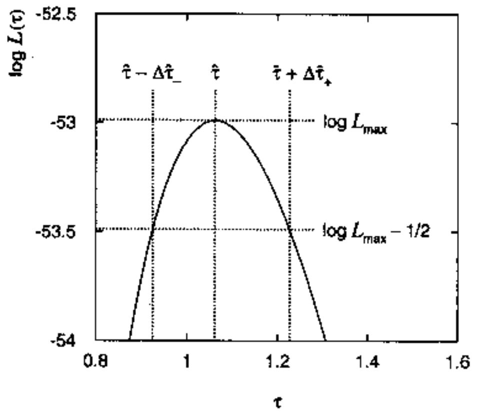
Total probability to obtain measurements x_i if x distributed according to pdf f(x, **a**)

Maximum of L → optimal values for **a**

$$\frac{d\log L(\vec{x},\vec{a})}{d\vec{a}}=\vec{0}=\sum_{i=1}^{n}\frac{d\log f(x_i,\vec{a})}{d\vec{a}}$$

$$V_{\vec{a}}=-\left(\frac{\partial^2\log L(\vec{x},\vec{a})}{\partial a_i\partial a_j}\right)^{-1}$$

[Cow98]



7 ML Example: Lifetime

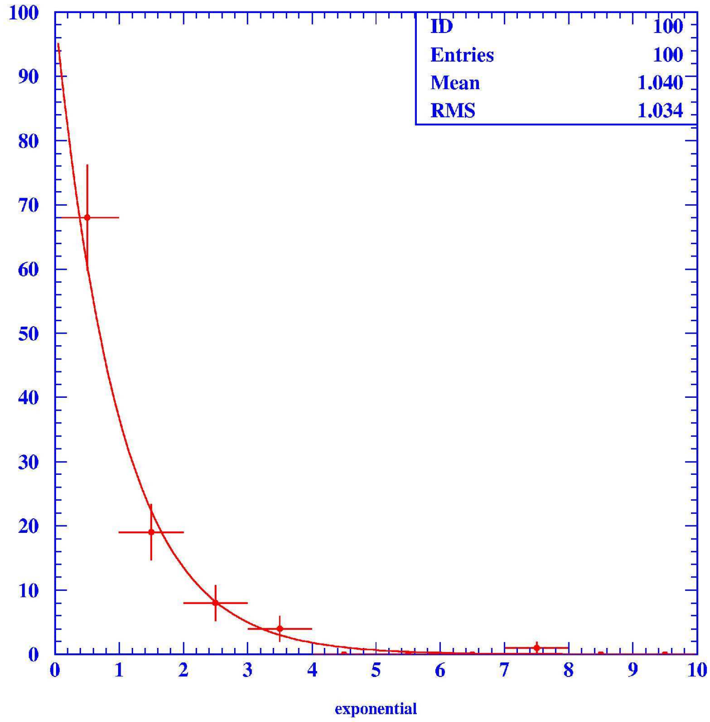
Measurements of lifetimes t_i distributed like $f(t;\tau)=1/\tau\, e^{-t/\tau}$

$$\log L=\sum_{i=1}^{n}-\log(\tau)-\frac{t_i}{\tau}$$

$$\frac{d\log L}{d\tau}=0\rightarrow\langle\tau\rangle=\frac{1}{n}\sum_{i=1}^{n}t_i$$

$$\frac{d^2\log L}{d\tau^2}=\sum_{i=1}^{n}\left(\frac{1}{\tau^2}-\frac{2t_i}{\tau^3}\right)$$

$$\simeq-\frac{n}{\langle\tau\rangle^2}\rightarrow\sigma_\tau^2=\frac{\langle\tau\rangle^2}{n}$$



7 ML with Histogram: binned ML

$$\nu_i(\vec{a}) = n_{tot} \cdot \int_{x_i}^{x_i+\Delta x} f(x,\vec{a})\,dx$$

Expectation of theory/model with pdf f(x, **a**) for frequency in interval $[x_i, x_i+\Delta x_i]$ (bin i)

$$L(n_1,\ldots,n_N;\nu_{1,}\ldots,\nu_N;\vec{a}) \sim \prod_{i=1}^{N} \left(\frac{\nu_i(\vec{a})}{n_{tot}}\right)^{n_i}$$

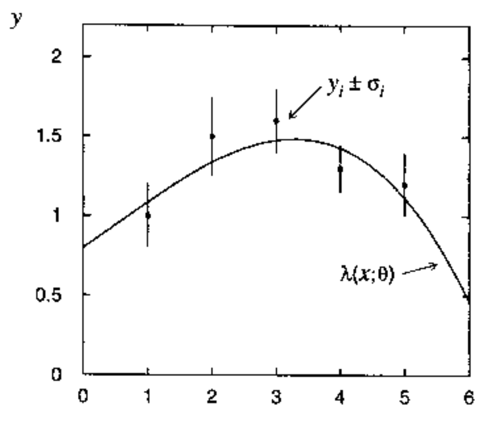
Likelihood to observe histogram with frequencies n_i given ν_i

$$\log L = \sum_{i=1}^{N} n_i \log(\nu_i(\vec{a}))$$

Maximum of log L → optimal parameters **a**

7 Least Squares

Consider N random variables **y** = $(y_1, \ldots, y_n)$ with normal pdf and theory/model **y** = **λ**(**a**)



[Cow98]

$$f(\vec{y},\vec{\sigma}_y;\vec{a}) \sim \prod_{i=1}^{N} e^{-\frac{(y_i-\lambda_i(\vec{a}))^2}{2\sigma_{y_i}^2}}$$

$$\log L(\vec{a}) = -\frac{1}{2}\sum_{i=1}^{N} \frac{(y_i-\lambda_i(\vec{a}))^2}{\sigma_{y_i}^2} = -\frac{1}{2}\chi^2(\vec{a})$$

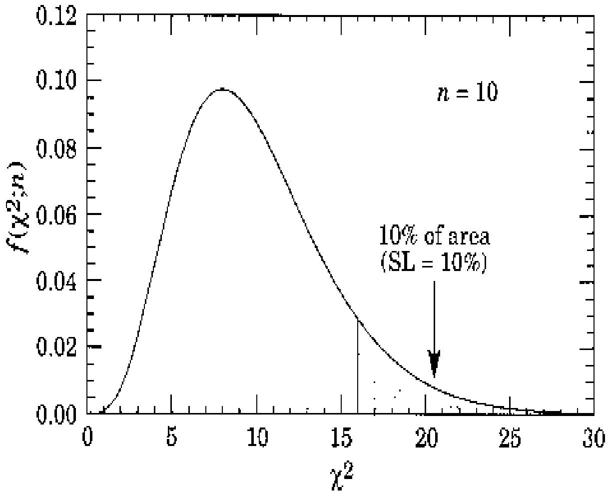
$$V_{\vec{a}} = -\frac{1}{2}\left(\frac{\partial^2\chi^2(\vec{a})}{\partial a_i \partial a_j}\right)^{-1}$$

χ^2 follows χ^2-distribution
σ_a given by $\chi^2 = \chi^2_{min} + 1$

N normal and correlated variables → N-dimensional Gaussian

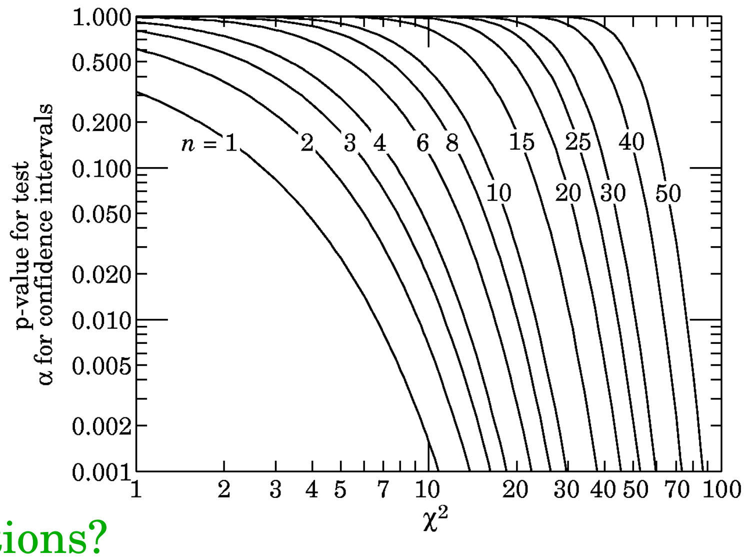
$$\chi^2(\vec{a}) = (\vec{y}-\vec{\lambda}(\vec{a}))^T \cdot V_{\vec{y}}^{-1} \cdot (\vec{y}-\vec{\lambda}(\vec{a}))$$

7 Interpretation of χ^2



[Pdg00]

$P(\chi^2) = P(\chi^2 \geq \chi^2_{obs})$ caused by statistical fluctuations if theory/model is correct



$P(\chi^2)$ small: theory/model?,
errors?, correlations?

$P(\chi^2) \simeq 1$: errors?, correlations?

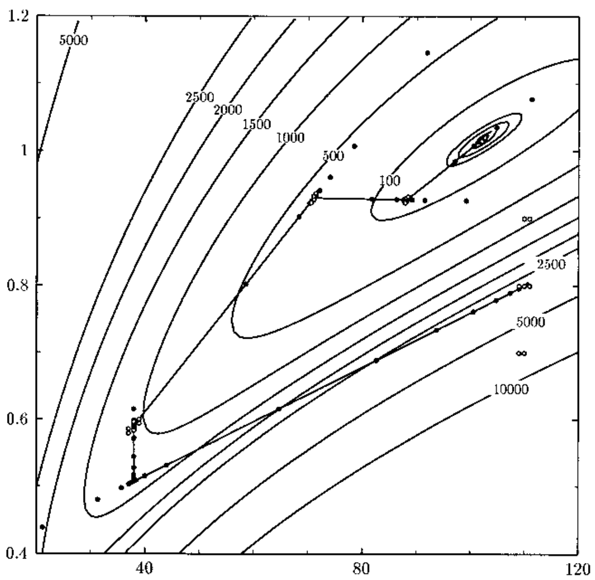
7 Numerical Optimisation

$$f(\vec{x}_i + \Delta\vec{x}) \simeq f(\vec{x}_i) + \frac{\partial f(\vec{x}_i)}{\partial \vec{x}} \cdot \Delta\vec{x} + \frac{1}{2} \Delta\vec{x}^T \cdot H_i \cdot \Delta\vec{x}$$

Newton method:

$$\frac{\partial f(\vec{x}_i + \Delta\vec{x})}{\partial \vec{x}} \simeq \frac{\partial f(\vec{x}_i)}{\partial \vec{x}} + H_i \cdot \Delta\vec{x} = \vec{0}$$

$$\Delta\vec{x} = -H_i^{-1} \cdot \frac{\partial f(\vec{x}_i)}{\partial \vec{x}}$$



[Blo98]

Algorithm:

1) Find f, df/d**x** und H at $\mathbf{x}_i$
2) Find $\Delta\mathbf{x}$ and thus $\mathbf{x}_{i+1}$
3) Find f, df/d**x** and H at $\mathbf{x}_{i+1}$
4) Test convergence

8 Correcting Data

- Distribution of random variable determined by several factors
  - acceptance: experimental coverage of phase space
  - efficiency: probability of successful measurement
  - resolution: additional random fluctuations caused by measurement
  - background: events not from target sample
- Use control samples or MC for corrections

8 Correction factors

Measurements n_i with background estimate β_i

MC samples before (generator level) and after detector simulation (detector level) give correction factors:

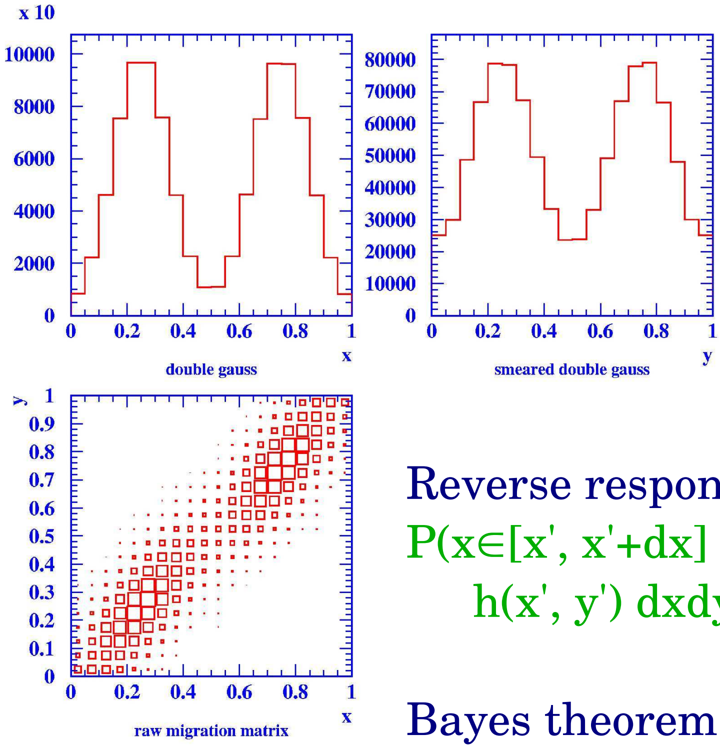
$$C_i = \mu_i^{MC} / (n_i^{MC} - \beta_i^{MC}) = \mu_i^{MC} / n_i^{MC,sig}$$

Estimators given by: $\langle\mu_i\rangle = (n_i - \beta_i)\, C_i = n_i^{sig}\, C_i$

$$E[\mu_i] = (C_i - \mu_i / n_i^{sig})\, n_i^{sig} + \mu_i$$

$\Rightarrow$ results $\langle\mu_i\rangle$ biased if μ_i / n_i^{sig} not well modelled by MC

8 Migration & Response Matrix



Migration probabilities:

$P(y \in [y', y'+dy], x \in [x', x'+dx]) =$
$\quad h(x', y')\, dxdy$

Response:

$P(y \in [y', y'+dy] \mid x \in [x', x'+dx]) =$
$\quad h(x', y')\, dxdy \,/ \int h(x', y'')dy''dx =$
$\quad r(y \mid x)\, dy$

Reverse response:

$P(x \in [x', x'+dx] \mid y \in [y', y'+dy]) =$
$\quad h(x', y')\, dxdy \,/ \int h(x'', y')dx''dy = s(x \mid y)\, dx$

Bayes theorem $\Leftrightarrow$ unfolding equations:

$g(y) = \int h(x, y)dx = \int r(x \mid y)f(x)dx = r(y \mid x) \otimes f(x)$
$f(x) = \int h(x, y)dy = \int s(y \mid x)g(y)dy = s(x \mid y) \otimes g(y)$

8 Unfolding

Obtaining f(x) from g(y) is task of unfolding. After discretisation (expected data $\boldsymbol{\nu}$, true values $\boldsymbol{\mu}$, background $\boldsymbol{\beta}$, observed data $\mathbf{n}$):

$$\boldsymbol{\nu} = R \cdot \boldsymbol{\mu} + \boldsymbol{\beta} \text{ and } \boldsymbol{\mu} = S \cdot (\boldsymbol{\nu} - \boldsymbol{\beta}) \text{ with } R^{-1} \neq S$$

Estimators for $\boldsymbol{\mu}$: $\langle \boldsymbol{\mu} \rangle = R^{-1} \cdot (\mathbf{n} - \boldsymbol{\beta})$ and $\langle \boldsymbol{\mu} \rangle = S \cdot (\mathbf{n} - \boldsymbol{\beta})$

Estimators are LS solution of $\chi^2 = (\boldsymbol{\nu} - \mathbf{n})^T \cdot V^{-1} \cdot (\boldsymbol{\nu} - \mathbf{n})$
with V covariance matrix of data $\mathbf{n}$ $\Rightarrow$ no bias, optimal variance

Find R (S) by normalising migration matrix column- (row-) wise
Obtaining R^{-1} is an inverse problem → regularised unfolding

8 Unfolding Example

Reference sample 10^6 events:

Double Gauss $\mu_1 = 0.25$, $\mu_2 = 0.75$,
$\sigma = 0.1$

Resolution given by $\sigma_x = 0.075$

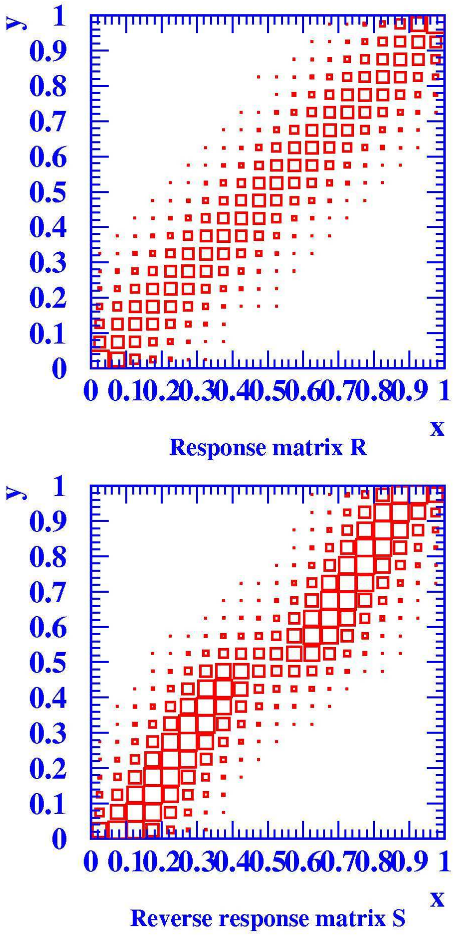
Find R and S

Test samples 10^3 events:

Generate "smeared" distribution

Unfold with R^{-1} or S

Compare with reference



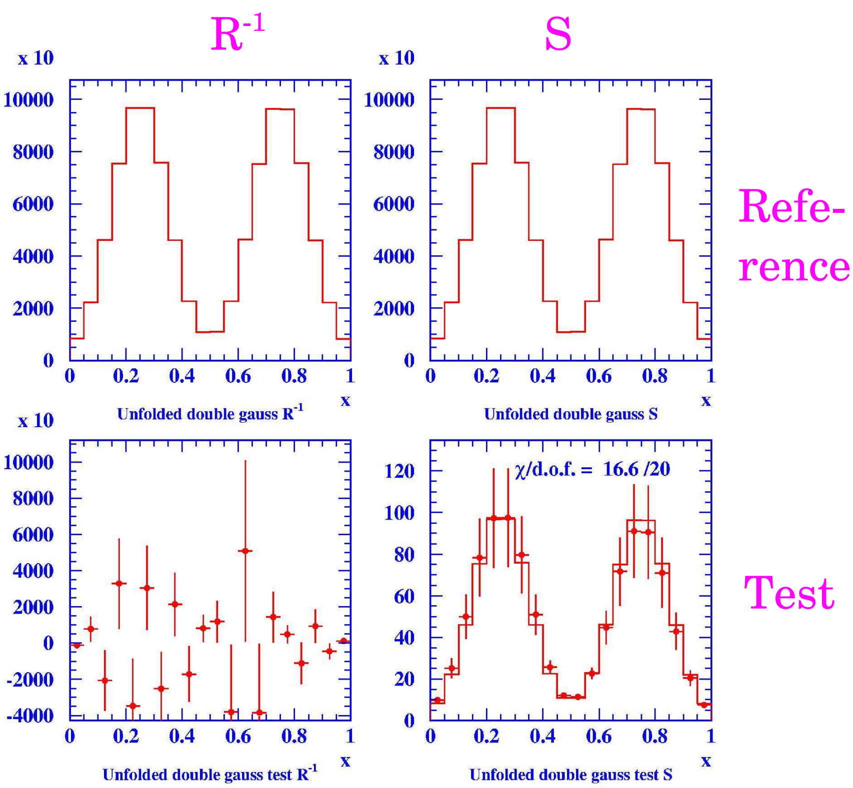
8 Unfolding Example

Identical results with R^{-1} and S in reference

Naive R^{-1} failes in test → statistical fluctuations of test sample amplified

S gives consistent results

Can avoid regularised unfolding if joint pdf h(x,y) (migration matrix) known



8 Unfolding Example

Upper row:

$$\chi^2 = (\mu - \langle\mu\rangle)^T \cdot W^{-1} \cdot (\mu - \langle\mu\rangle)$$
$$W = S \cdot V \cdot S^T$$

W often singular → χ^2 and $P(\chi^2)$ unreliable

Lower row:

$$\chi^2 = (\boldsymbol{\nu} - \mathbf{n})^T \cdot V^{-1} \cdot (\boldsymbol{\nu} - \mathbf{n})$$

V diagonal → χ^2 stable