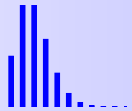
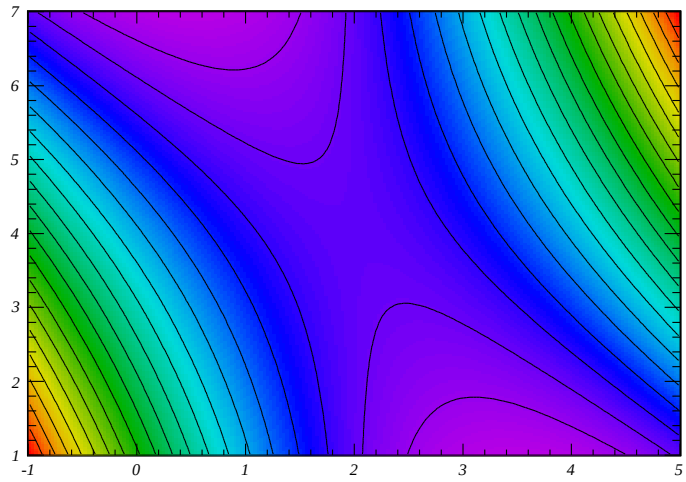


Advanced Methods in Data Analysis

Michael Schmelling – MPI for Nuclear Physics

- *Basics*
- *Monte Carlo Methods*
- *Error propagation*
- *Parameter estimates*
- *Unfolding*
- *Multivariate analysis*
- *Markov Chain Monte Carlo*



➜ *in alphabetical order. . .*

- R.J. Barlow,*Statistics*, Wiley
- S. Brand, *Data Analysis*, Springer
- G.D. Cowan,*Statistical Data Analysis*, Oxford University Press
- F. James, *Statistical Methods in Experimental Physics*, World Scientific
- H.L. Harney, *Bayesian Inference*, Springer
- D.E. Knuth, *The Art of Computer Programming*, Addison Wesley
- W.T. Press et al., *Numerical Recipes*, Cambridge University Press
- D.S. Sivia, *Data Analysis - A Bayesian Tutorial*, Oxford University Press
- plus many more . . .

➜ *statistics everywhere. . .*

(sugar served with espresso)

front side

back side

"statistics sweetens your life"

"during our lives we cover 22150 km on foot"

An introductory example . . .

➜ *the story of the cheating baker*

Once upon a time, in a holiday resort the landlord L. ran a profitable B&B, and every morning bought 30 rolls for breakfast. By law the mass of a single roll was required to be 75 g. One fine day the owner of the bakery changed, and L. suspected that the new baker B. might be cheating. So he decided to check the mass of what he bought, using a kitchen scales with a resolution of 1g.

After one month he had collected a fair amount of data. . .

```
73 79 72 62 67 60 60 67 78 68 66 75 76 73 75 64 70 69 73 59 70 73 64 72 64 69
69 71 69 71 77 69 72 71 67 72 63 66 68 76 71 76 68 71 63 65 65 66 73 73 73 67
70 65 71 69 78 67 65 69 71 71 72 73 72 69 66 66 70 60 72 62 53 65 74 65 68 69
67 75 64 76 72 76 78 67 67 67 69 79 71 67 71 68 71 65 66 65 78 76 71 70 67 65
67 64 73 67 74 79 74 71 73 67 66 76 68 74 76 65 77 67 71 67 71 77 63 66 70 62
68 74 67 67 67 77 65 68 79 72 71 77 68 70 73 67 81 70 74 71 79 62 67 63 68 76
73 81 76 73 68 72 76 61 69 73 71 80 68 70 62 76 58 68 68 64 68 78 69 65 70 70
64 75 73 72 60 86 68 68 64 60 68 71 70 75 70 67 69 67 73 65 66 71 70 70 73 66
72 71 71 64 76 75 72 72 71 72 72 71 75 68 73 70 64 76 72 75 79 70 64 70 67 70
75 70 83 69 61 70 66 69 71 72 70 76 73 62 71 60 73 74 70 68 68 70 78 71 69 71
73 73 75 65 71 67 60 70 77 71 74 64 74 73 60 77 73 70 69 66 70 78 69 75 66 71
75 75 74 69 74 70 75 77 75 66 72 68 72 61 75 65 69 68 65 73 82 67 75 67 80 71
79 72 71 68 73 70 67 75 74 69 63 63 72 70 73 63 70 70 59 78 76 66 72 79 65 71
76 72 69 69 73 70 77 73 83 66 68 67 69 73 76 65 71 70 71 65 78 71 67 70 72 75
67 79 72 64 62 79 68 70 61 65 68 71 73 60 60 68 71 74 75 69 73 70 68 ...
```

Data reduction

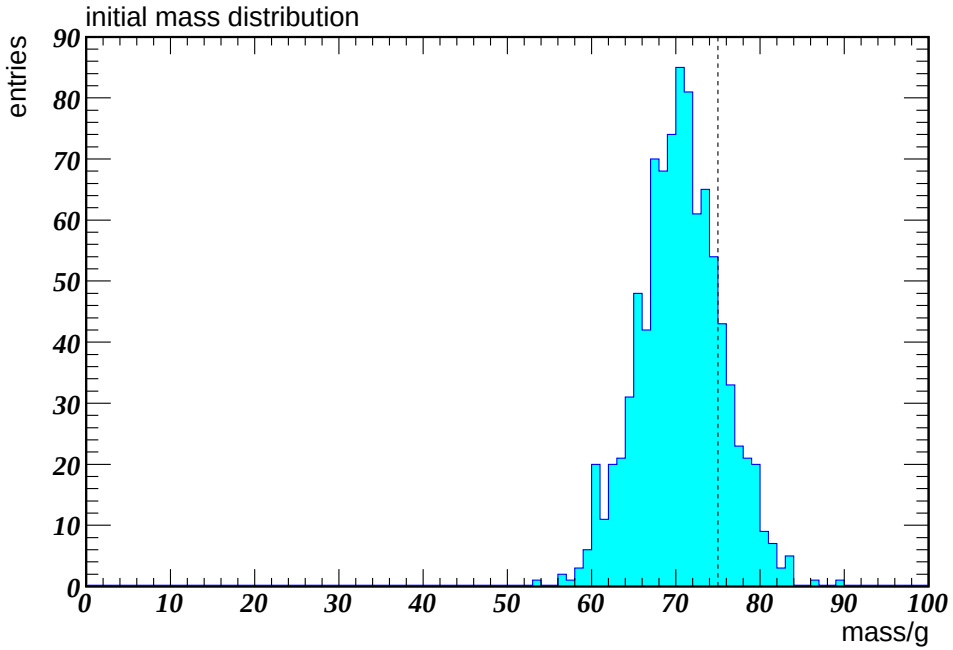
- the raw list of number is not very useful ➜ need some kind of data reduction
- assume that all measurements are equivalent
  - ➜ the sequence of the individual data does not matter (in this example)
  - ➜ all relevant information is contained in the number of counts per reading

count[50]= 0	count[60]= 20	count[70]= 85	count[80]= 9
count[51]= 0	count[61]= 11	count[71]= 81	count[81]= 7
count[52]= 0	count[62]= 20	count[72]= 61	count[82]= 3
count[53]= 0	count[63]= 21	count[73]= 65	count[83]= 5
count[54]= 0	count[64]= 31	count[74]= 54	count[84]= 0
count[55]= 0	count[65]= 48	count[75]= 43	count[85]= 0
count[56]= 2	count[66]= 42	count[76]= 33	count[86]= 1
count[57]= 1	count[67]= 70	count[77]= 23	count[87]= 0
count[58]= 3	count[68]= 68	count[78]= 21	count[88]= 0
count[59]= 6	count[69]= 74	count[79]= 20	count[89]= 1

  - ➜ much improved presentation of the collected information
  - ➜ the above numbers cover the entire data set
  - ➜ most of the measurements are lower than the legally required value...

Visualization

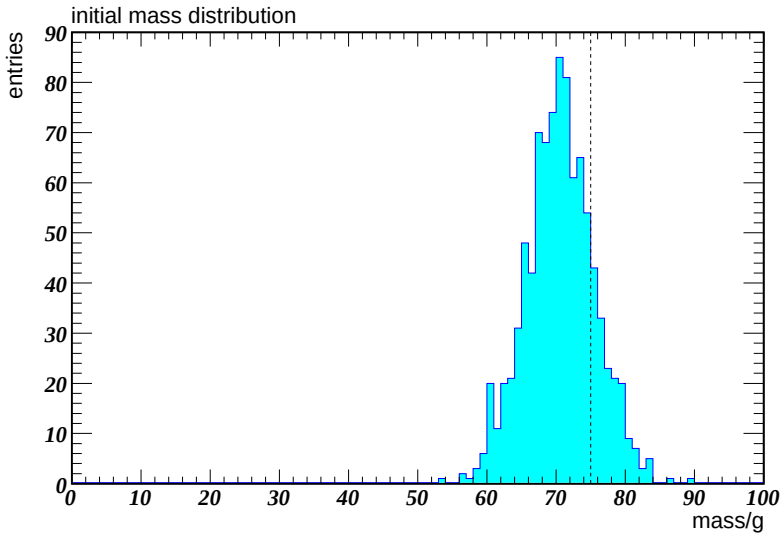
- an even better presentation of the available information: bar-chart
- example for the concept of a histogram
  - ➜ define bins for the possible values of a variable
  - ➜ plot the number of entries in each bin
  - ➜ get an immediate grasp of center and width of the distribution

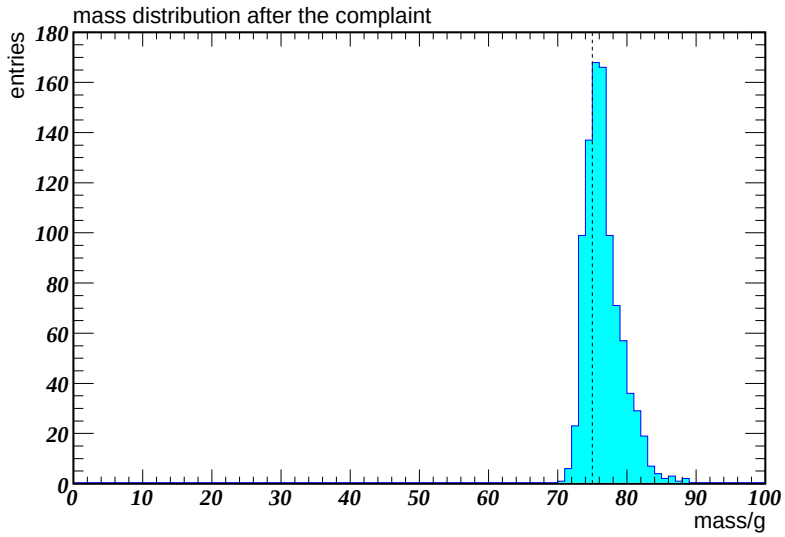


The rolls produced by baker B. definitely are too light. So L. was right in his suspcion, that B. tried to make some extra profit by cheating...

. . . and the conclusion

As a consequence of his findings, L. complained. B. apologized and claimed that the low mass of the rolls was an accident which will be corrected in the future. L., however, continues to monitor the quality delivered by the baker. One month later, B. asked whether now everything was all right. L., for his part, acknowledged that the weight of the rolls now matched his expectations, but also voiced the opinion that B. was still cheating. . .





➜*B. simply selected the heaviest rolls for L.!*

Before moving on . . .

➜ *always keep in mind:*

- the name of the game: extract meaning from a stream of numbers
- the tools: “statistical and numerical methods”
  - ➜ need know the relevant methods
  - ➜ need to understand their properties
- basic assumptions
  - ➜ measurements deviate from the respective true values
  - ➜ the deviation is a random variable
  - ➜ statistics builds on probability theory

A statistical method is neither **“right”** nor **“wrong”**.
It has **properties**, which have to be known for the interpretation of the result. Possible properties could be, that the output is the most precise estimator, or that the result is robust. The property could also be that the result is wrong, in which case use of this particular method should be discouraged...

Notations

$p(A)$	probability for A
$p(A\|B)$	conditional probability for A if B is given
$x, y, z, t, \ldots$	continuous random variable
$i, j, k, l, m, n \ldots$	discrete random variable (or index)
$\vec{x}$	vector of random variables $\{x_1, \ldots, x_n\}$
p_i, q_i	discrete probabilities
$f(x), g(x)$	probability densities functions (PDFs) of x
$F(x), G(x)$	cumulative distributions of f, g
$f(x, y)$	2-dim probability density in x und y
$f(x\|y)$	conditional PDF for x given y
$a, b, \ldots, \alpha, \beta, \ldots$	parameters
$E[x] = \langle x \rangle = \mu_x$	expectation value von x
$V[x] = \sigma_x^2$	variance von x
$\widehat{a}$	estimate for a
$\overline{x}$	arithmetic average of x
$\sum_{(i)}$	sum over all indices (i)
$\int dx$	integrate over all x

A matrix $A[m,n]$ is an array of numbers with m rows und n columns. Usually the dimensions are not given explicitly. Individual matrix elements are addressed by two indices, A_{ij}, where the first index specifies the row and the second one the column. The following is a summary of the rules for matrix manipulations:

Sum of two matrices:

$$C[m,n] = A[m,n] + B[m,n] \quad \text{or} \quad C_{ij} = A_{ij} + B_{ij}$$

Product of two matrices:

$$C[m,n] = A[m,l] \cdot B[l,n] \quad \text{or} \quad C_{ij} = \sum_{k=1}^{l} A_{ik} B_{kj}$$

Product of three matrices:

$$D[m,n] = A[m,l] \cdot B[l,k] \cdot C[k,n] \quad \text{or} \quad D_{ij} = \sum_{r=1}^{l} \sum_{s=1}^{k} A_{ir} B_{rs} C_{sn}$$

associative law of matrix multiplication:

$$A \cdot (B \cdot C) = (A \cdot B) \cdot C$$

The neutral element with respect to matrix multiplication is the unit matrix

$$\mathbf{1}[n,n] = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \qquad \text{using indices} \qquad \mathbf{1}_{ij} = \delta_{ij}$$

$$\text{giving} \qquad A[n,m] \cdot \mathbf{1}[m,m] = \mathbf{1}[n,n] \cdot A[n,m] = A[n,m]$$

Square matrices $A[n,n]$ (of rank n) have a unique inverse matrix A^{-1}:

$$A^{-1} \cdot A = A \cdot A^{-1} = \mathbf{1}$$

For the inverse of a product of square matrices on has:

$$(A_1 \cdot A_2 \cdots A_n)^{-1} = A_n^{-1} \cdots A_2^{-1} \cdot A_1^{-1}$$

Another matrix operation is transposition:

$$A[m,n]^T = B[n,m] \quad \text{or} \quad B_{ij} = A_{ji} \ .$$

For die transpose of a product of matrices one has:

$$(A_1 \cdot A_2 \cdots A_n)^T = A_n^T \cdots A_2^T \cdot A_1^T$$

For $n \times n$ matrices there exist n scalar quantities which are invariant under orthogonal transformations of the matrix. The two most important ones are determinant and trace, the product and the sum of the eigenvalues λ_i of the matrix:

$$\det(A[n,n]) = \prod_{i=1}^{n} \lambda_i \quad \text{and} \quad \mathrm{Tr}\, A[n,n] = \sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} A_{ii}$$

The trace is given by the sum of the diagonal elements. Expressed as a function of the matrix elements, the determinant of a 2×2 matrix is

$$\det(A[2,2]) = A_{11}A_{22} - A_{12}A_{21}$$

For the determinant of a product of matrices one finds:

$$\det(A_1 \cdot A_2 \cdots A_n) = \det(A_1) \cdot \det(A_2) \cdots \det(A_n)$$

The trace of a product of matrices is invariant under cyclic permutations:

$$\mathrm{Tr}(A_1 \cdot A_2 \cdots A_n) = \mathrm{Tr}(A_2 \cdots A_n \cdot A_1)$$

A special class of matrices are vectors. In the following a letter with an arrow denotes a column vector. Row vectors are obtained by transposition (T) of a column vector.

$$\vec{b} = b[n,1] \quad \text{column vector}$$

$$\vec{a}^{\,T} = a[1,n] \quad \text{row vector}$$

For two vectors $\vec{a}$ and $\vec{b}$ of dimensions n, $\vec{a}^{\,T} \cdot \vec{b}$ is a scalar and $\vec{a} \cdot \vec{b}^{\,T}$ is a matrix:

$$\vec{a} \cdot \vec{b}^{\,T} = \begin{pmatrix} a_1 b_1 & a_1 b_2 & \dots \\ a_2 b_1 & a_2 b_2 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

It follows:

$$\vec{a}^{\,T} \cdot \vec{b} = \mathrm{Tr}(\vec{a}^{\,T} \cdot \vec{b}) = \mathrm{Tr}(\vec{b} \cdot \vec{a}^{\,T})$$

Expectation values of matrices are defined by element:

$$\langle A \rangle_{ij} = \langle A_{ij} \rangle$$

The product of two sums can be written as a sum over two indices

$$\left(\sum_i x_i\right)\left(\sum_j y_j\right) = \sum_{ij} x_i y_j$$

i.e. interpreting x_i or y_i as elements of a vector $\vec{x}$ or $\vec{y}$, respectively, every element of $\vec{x}$ is multiplied with every element of $\vec{y}$ and the individual terms summed up.

Special case: $\vec{y} = \vec{x}$

$$\left(\sum_i x_i\right)\left(\sum_j x_j\right) = \left(\sum_i x_i\right)^2 = \sum_{ij} x_i x_j = \sum_{i=j} x_i^2 + \sum_{i \neq j} x_i x_j$$

Since the expectation value (formally defined later) is a linear operator sums and expectation values commute:

$$\left\langle \sum_i x_i \right\rangle = \sum_i \langle x_i \rangle$$

➜ *general problem: minimization subject to constraints*

Consider the general constrained minimization problem in 2 dimensions:

$$C(x,y) \stackrel{!}{=} \min \quad \text{with} \quad g(x,y) = 0$$

➜ *default approach:*

Use $g(x,y) = 0$ to solve for $y = G(x)$, substitute

$$\frac{\partial}{\partial x} C(x, G(x)) = 0 \quad \text{with} \quad g(x, G(x)) = 0$$

and determine $x_{\min}$ and $y_{\min} = G(x_{\min})$.

- ➜ conceptually straightforward ansatz
- ➜ minimization problem with reduced number of dimensions
- ➜ breaks the symmetry between the variables
- ➜ often impossible to do in practice

try to come up with something better...

➜ *the Lagrange multiplier approach*

Example: The Milkmaid's Problem

A milkmaid is sent to a field close to the river in order to milk a cow. Entering the field at point M, the milkmaid spots the cow at C. Normally she would go directly to the cow, – but then realizes that her bucket first needs cleaning in the river. The problem is to find the shortest path connecting M and C via the bank of the river.

❖ mathematical formulation:

cost function:

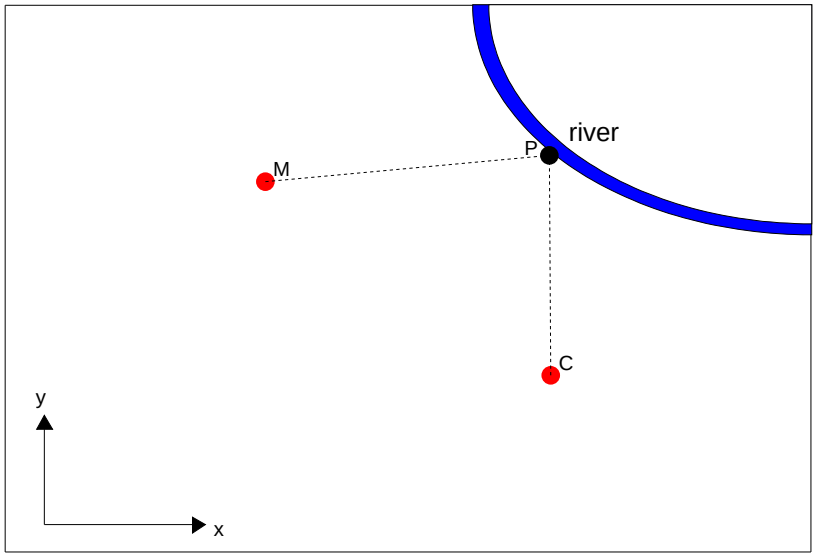
$$C(P_x, P_y) = |\vec{M} - \vec{P}| + |\vec{P} - \vec{C}|$$

description of the distance to the river:

$$g(x, y) = c$$

constraint:

$$g(P_x, P_y) = 0$$



Lagrange multipliers (iii)

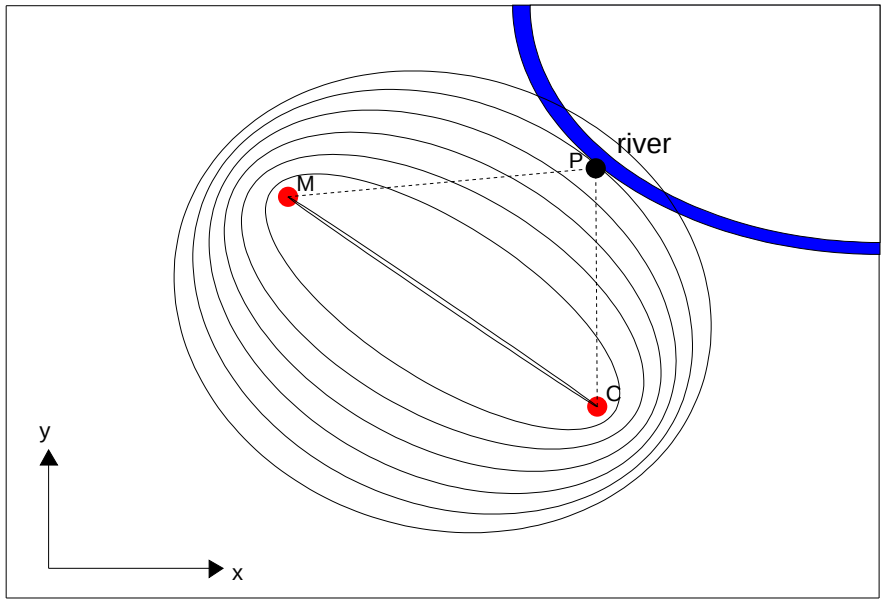
The points where the sum of the distances to two “focal” points is constant are located on an ellipse. Contours of equal cost thus are given by ellipses around C and M. The best solution is the smallest ellipse touching the river. At this point the contour lines C =const and g =const have to be parallel.

- ➜ contour lines are orthogonal to function gradients
- ➜ parallel contour lines implies parallel gradients

❖ condition for the best point P:

$$\nabla C(x,y) \propto \nabla g(x,y)$$

Exploit this to find an elegant way for solving constrained optimization problems. . .



Lagrange multipliers (iv)

➜ *insight by Lagrange*

The stationary point of a linear combination of cost function C and constraint function g is the solution of a constrained minimization. Introducing

$$F(x,y) = C(x,y) + \lambda \cdot g(x,y)$$

one finds

$$\nabla F(x,y) = 0 = \nabla C(x,u) + \lambda \cdot \nabla g(x,y) \quad \text{i.e.} \quad \nabla C(x,u) \propto \nabla g(x,y) \; .$$

➜ *discussion*

- minimization of F is usually much easier than the "default approach"
- fully symmetric in all variables
- the result is a function of λ, i.e $x(\lambda)$,$y(\lambda)$
- λ can be determined from the condition $g(x(\lambda),y(\lambda)) = 0$
- in many cases the explicit value of λ is not needed

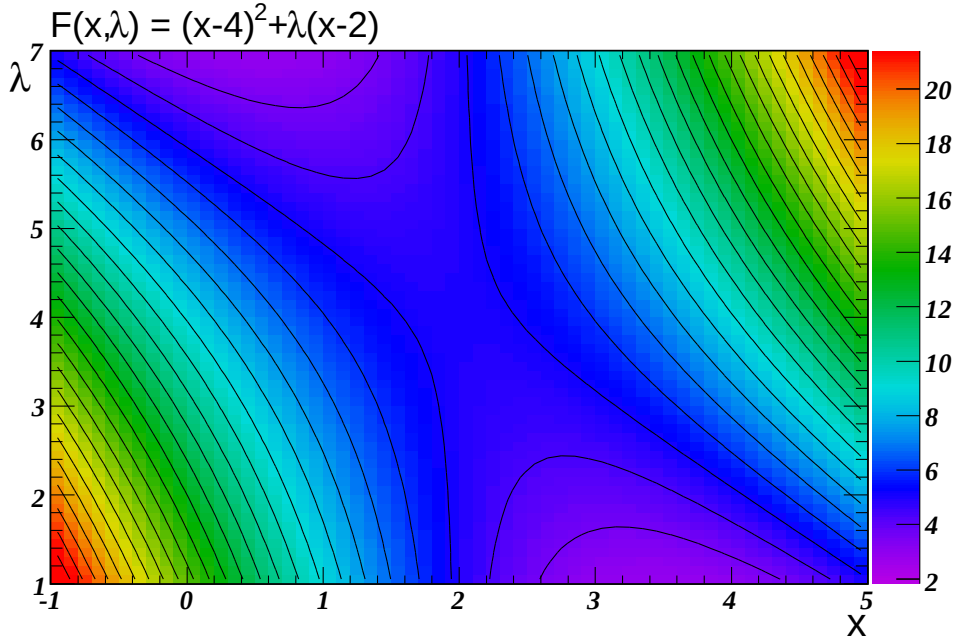
➜ *additional remarks*

Introduction of λ increases the dimension of the minimization problem and a stationary point is determined in a higher dimensional space. Since the extended cost function $F(x,y)$ is linear in λ the stationary point will be saddle point.

Example: $C(x) = (x-4)^2$ and $g(x) = x - 2$

$$F(x,\lambda) = (x-4)^2 + \lambda \cdot (x-2)$$

- local minimum in x for every λ
- no global minimum
- the saddle-point has minimum cost for constraint $g(x) = 0$
  - ➜ $x_{\min} = 2$
  - ➜ $\lambda_{\min} = 4$



Combinatorics

An important aspect of many statistical analyses is to count the number of possible results. For discrete states the solution is found by combinatorics. Some of the most important results are collected below:

➜ *words with m-characters from an alphabet with n letters:*

$$N = n^m$$

➜ *Permutations of n objects:*

$$N = n \times (n-1) \times (n-2) \times \ldots 2 \times 1 = n!$$

➜ *Possibilities to select k objects from a total of n (without putting back)*

$$\frac{n(n-1)...(n-2)(n-k+1)}{k!} = \frac{n!}{k!(n-k)!} = \left(\begin{array}{c} n \\ k \end{array} \right)$$

the "lottery-problem"

➜ *Kolmogorov's axioms on probability*

Starting from set theory, probability theory can be built on a mapping from sets E to real numbers $p(E) \in [0,1]$. Define

$$\begin{aligned} \Omega &: \text{the entire set} \\ E &: \text{partial set of } \Omega \\ p(E) &: \text{probability of } E \end{aligned}$$

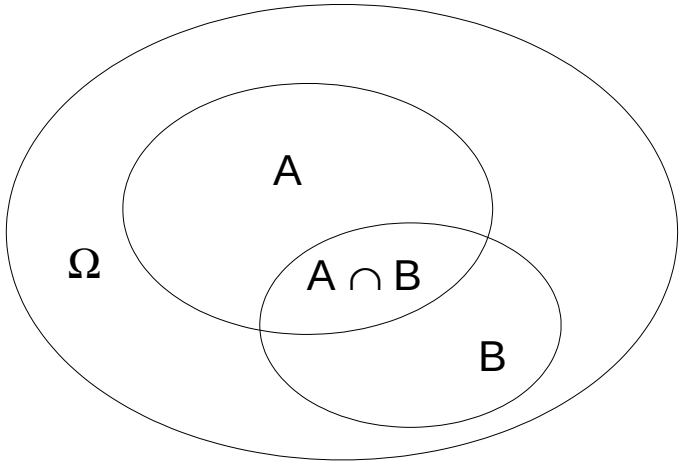
and postulate the following axioms:

1. $0 \leq p(E) \leq 1$
2. $p(\Omega) = 1$
3. $p(E_1 \cup E_2) = p(E_1) + p(E_2)$ if $E_1 \cap E_2 = 0$

Based on these axioms, calculations involving probabilities are unambiguously defined. Interpretation is left completely open . . .

Conditional probability & independent events

Rules for calculus of probabilities derived from Kolmogorov's axioms can easily be visualized using diagrams from set theory. For example:



$$p(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Consider $P(B|A)$, the probability for B if A is given

- ➜ the diagram suggests $P(B|A) \propto P(A \cap B)$
- ➜ for $A \in B$ one must have $P(B|A) = 1$
- ➜ $A \in B$ implies $P(A \cap B) = P(A)$ and thus

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$ "conditional probability"

For "independepent events", which implies $P(B|A) = P(B)$, one obtains:

$$P(A \cap B) = P(A) \cdot P(B)$$

Bayes' theorem

Consider a set of disjoint events $A_i, i = 1, \ldots, n$. It follows

$$p(A_i \cap B) = p(B|A_i)\, p(A_i) = p(A_i|B)\, p(B)$$

$$\Longrightarrow \quad p(A_i|B) = \frac{p(B|A_i)\, p(A_i)}{p(B)}$$

Bayes' theorem

The prior $p(A_i)$ for A_i is updated by the occurrence of B to become $p(A_i|B)$.

Bayes' theorem is at the heart of statistical inference based on empirical input. If the A_i are exhaustive, i.e. if one of them is realized with unit probability independently of B, then one has

$$p(B) = \sum_i p(B|A_i) p(A_i)$$

and thus

$$p(A_k|B) = \frac{p(B|A_k) p(A_k)}{\sum_i p(B|A_i) p(A_i)}$$

➜ applications . . .

Bayes' theorem - example 1

A new test for the common cold hits the market, designed to detect an infection in the early stages where an efficient cure is available. The probability to test positive in case of an infection is $p(+|I) = 0.98$, the probability for a negative result on a healthy subject is $p(-|H) = 0.97$. Series tests are performed in summer, where the a priori probability for infection is $p(I) = 0.001$.

What's the probability that a person tested positive has actually contracted a cold?

the probabilities are:

$$\begin{array}{llll} p(I) = 0.001 & p(H) = 0.999 \\ p(+|I) = 0.980 & p(-|I) = 0.020 \\ p(+|H) = 0.030 & p(-|H) = 0.970 \end{array}$$

where the rows sum up to unity. Application of Bayes' theorem then yields

$$p(I|+) = \frac{p(+|I)p(I)}{p(+|I)p(I) + p(+|H)p(H)} \approx 0.032$$

Simply administering sweets to all patients that diagnosed "infected" already will yield a "healing rate" around 97%.

Bayes' theorem - example 2

Three boxes contain each two rings made of either gold (G) or silver (S). The boxes contain (GG), (SS) and (GS). The content of a specific box is unknown. A person is allowed two draws of a single ring from any of the boxes. The first draw yields gold.

Which box for the second draw maximizes the number of gold rings?

Calculate the probability that the box of the first draw contains (GG). A priori the probabilities are $p(GG) = p(GS) = p(SS) = 1/3$. The probabilities to get (G) in the first draw become

$$p(G|GG) = 1 \quad , \quad p(G|GS) = \frac{1}{2} \quad \text{and} \quad p(G|SS) = 0 \; .$$

Bayes' theorem then yields the probability that the selected box is (GG):

$$p(GG|G) = \frac{p(G|GG)p(GG)}{p(G|GG)p(GG) + p(G|GS)p(GS) + p(G|SS)p(SS)} = \frac{2}{3}$$

The second draw should be taken from the same box.

Bayes' theorem - example 3

Two old friends A and B who have gotten out of touch accidentally meet in a pub and decide to celebrate the occasion. A suggests to flip a coin in order to determines who will pay the next round. B agrees and then pays all the drinks.

What is the probability that A is cheating each time he throws the coin?

Consider the hypotheses h and c that A is an honest guy or that he is a cheater. The probability for A to win n times in a row is

$$p(n|h) = 2^{-n} \quad \text{and} \quad p(n|c) = 1$$

With the prior probabilities $p(h)$ and $p(c) = 1 - p(h)$, Bayes' theorem allows to determine the probability that A, after having won n times, is a cheater:

$$p(c|n) = \frac{p(n|c)p(c)}{p(n|c)p(c) + p(n|h)p(h)} = \frac{p(c)}{p(c) + 2^{-n}p(h)}$$

the result depends on $p(b)$:

$$\begin{array}{rcrcrcl} p(c) & = & 0.00 & \Longrightarrow & p(c|n) & = & 0 \\ p(c) & = & 0.05 & \Longrightarrow & p(c|1) & \approx & 0.095 \\ & & & & p(c|6) & \approx & 0.771 \\ & & & & p(c|\infty) & = & 1 \end{array}$$

"bayesian" update of knowledge

➜ *definition of a probability density function (PDF)*

A function $f(x)$ can be interpreted as a PDF if

$$f(x) \geq 0 \;\; \forall \; x \quad \text{and} \quad \int\limits_{-\infty}^{+\infty} dx \, f(x) = 1 \; .$$

➜ *interpretation:*

The probability to observe an event in the infinitesimal interval $[x, x+dx]$ is:

$$p(x, x+dx) = f(x) dx \; .$$

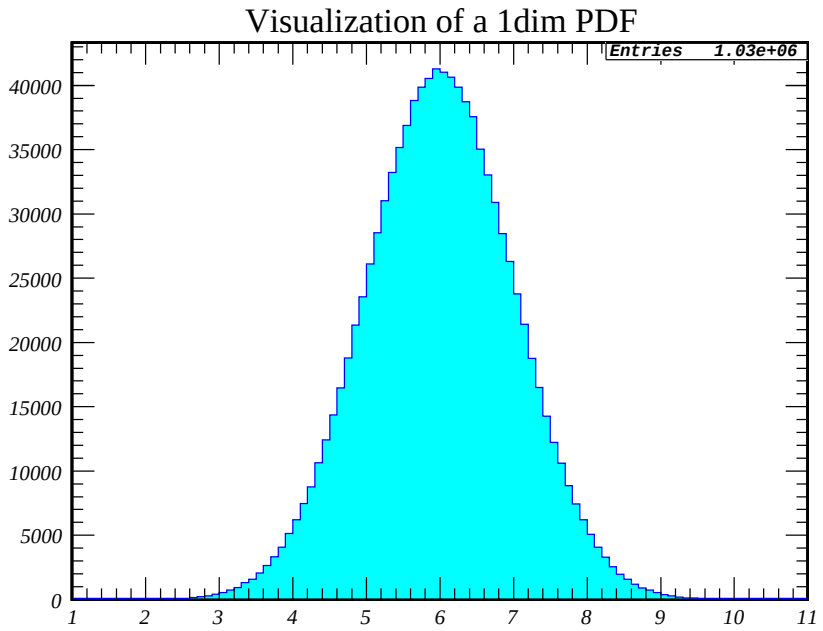
➜ *relation to discrete probabilities:*

discrete probabilities p_i, i.e. finite probabilities for discrete values, can be written as a PDF using Dirac's delta-function:

$$f(x) = \sum_{i=1}^{n} p_i \, \delta(x-i) \quad \text{where} \quad \int dx \; f(x) = \sum_i p_i = 1$$

Visualisation of 1-dim PDFs

- graphical representation of the density
- problem in practical applications
  - ➜ density function not known
  - ➜ only a random sample of size N
- obvious solution: mark the values
- better solution: histogram
  - ➜ divide the range into bins
  - ➜ count entries inside each bin
  - ➜ regarding bin limits:
    - ✔ too many bins: large fluctuations
    - ✔ too few bins: loss of information
    - ✔ use “reasonable” binning



(demo 01)

➜ *to illustrate the point. . .*

for a range $-1 \leq x \leq 1$ avoid histograms with 25 bins on the interval $[-1.1, +1.1]$. Use 20 bins between -1 and 1.

- variations:
  - ➜ density plots for small N
  - ➜ variable bin widths
  - ➜ logarithmic axes
  - ➜ . . .

Relation: histogram $\Longleftrightarrow$ PDF

➜ *given*

N : total number of entries in the histogram

h : bin width

n_k : number of entries in bin k $[x_k - h/2, x_k + h/2]$

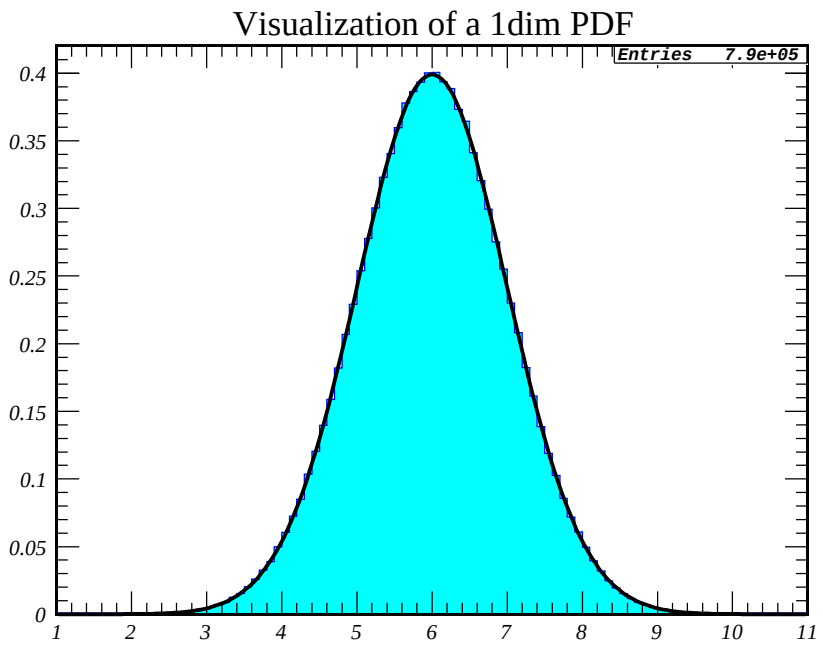
➜ *it follows*

$$n_k = N \times p(x - h/2, x + h/2)$$

$$= N \int_{x_k - h/2}^{x_k + h/2} dx\, f(x)$$

$$\approx N\, f(x_k)\, h$$

result: $$f(x_k) \approx \frac{n_k}{h \cdot N}$$



(demo)

➜ *goal:*

Summarize the properties of a PDF by (a few) numbers, so-called moments:

➜ *moments are "expectation values", defined by*

$$\int_{-\infty}^{\infty} dx\ f(x)\ w_k(x) = \langle w_k \rangle$$

i.e. as a mapping $f(x) \mapsto C$ of a PDF $f(x)$ onto a (complex) number via integral transform with a (family of) weight function(s) $w_k(x)$.

Example: cumulative distribution

$$w_X(x) = \Theta(X - x)$$

$$\langle w_X \rangle = \int_{-\infty}^{\infty} dx\ f(x)\ \Theta(X - x) = \int_{-\infty}^{X} dx\ f(x) = F(X)$$

$F(x)$ is the primitive of $f(x)$: $F(-\infty) = 0$, $F(\infty) = 1$

➜further examples . . .

A possible measure for the scatter s of x with PDF $f(x)$ around a point a is

$$s^2 = \int dx\ (x-a)^2\, f(x)$$

To use s for characterizing $f(x)$, the point a should be chosen such that s becomes minimal. Minimization of s^2 yields:

$$\frac{\partial s^2}{\partial a} = -2\int dx\ (x-a) f(x) \stackrel{!}{=} 0 \quad \text{i.e..} \quad a_{\min} = \int dx\ x\, f(x) = \langle x \rangle$$

It follows that the mean value (or "expectation value") $\langle x \rangle$ is a way to characterize the center of a PDF. For symmetric PDFs it is also the symmetry point:

$$\langle x \rangle = \int dx\ x\, f(x) = \int dx\ (x-a)\, f(x) + a\int dx\ f(x) = 0 + a \times 1 = a$$

The scatter σ around the mean value $\langle x \rangle$ is also referred to as "standard deviation" oder "rms"-scatter, its square as "variance". The following relation holds:

$$\sigma^2 = \int dx\ (x - \langle x \rangle)^2\, f(x) = \int dx\ (x^2 - 2x\langle x \rangle + \langle x \rangle^2)\, f(x) = \langle x^2 \rangle - \langle x \rangle^2$$

➜ *median*

The center of a distribution can also be taken as the median m, defined by

$$\int_{-\infty}^{m} dx\ f(x) = \int_{m}^{\infty} dx\ f(x)$$

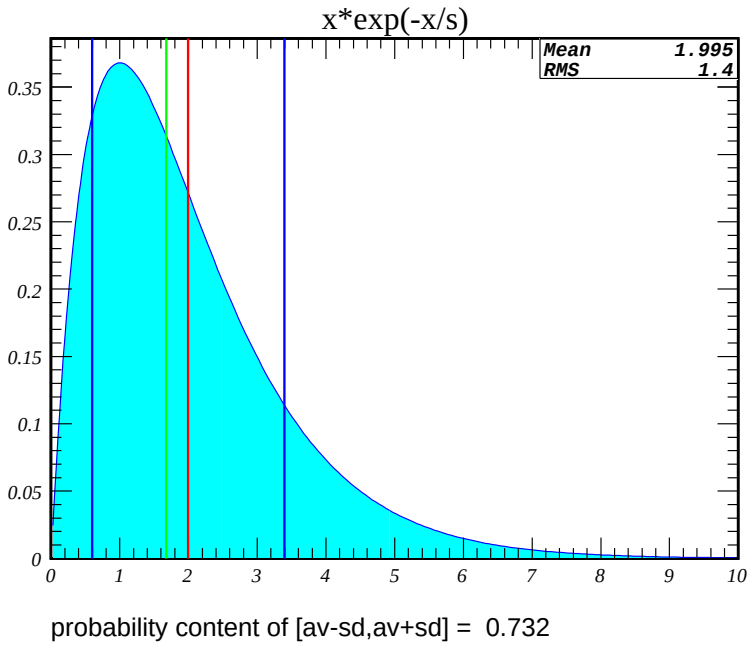
i.e. same probability on both sides. For symmetric distributions one has $\langle x \rangle = m$.

➜ *quantiles*

Quantiles are locations x_α on a PDF up to which with the probability content is $\alpha\%$. A possible measure for the width of a PDF is $x_{84} - x_{16}$.

➜ *discussion:*

- mean value and standard deviation
  - \+ linear functions of the PDF, i.e. easy to use in theoretical calculations
  - − sensitive to outliers and tails in the PDF
- median and quantiles
  - \+ insensitive against outliers and tails
  - − non-linear functions of the PDF, difficult to handle analytically



(demo 02)

➜ *for different PDFs*

- mean value $\langle x \rangle$
- standard deviation σ
- probability content of the interval $[\langle x \rangle - \sigma, \langle x \rangle + \sigma]$
- median

➜ *conclusion:*

- there are many possibilities to characterize a PDF
- other options, not discussed in detail are:
  - ➜ take as center the maximum
  - ➜ take as width the minimum interval with a given probability
- still most important: algebraic moments and derived quantities

Algebraic and central moments

➜ *algebraic moments:*

$$M_k \equiv \int dx\, f(x)\, x^k$$

- $M_0 = 1$: normalization of $f(x)$
- $M_1 = \mu$: mean value $f(x)$

➜ *central moments:*

$$Z_k \equiv \int dx\, f(x)\, (x-\mu)^k$$

- $Z_0 = 1$: normalization of $f(x)$
- $Z_1 = 0$
- $Z_2 = \sigma^2$: variance of $f(x)$

➜ *other commonly used moments:*

$$S = \frac{Z_3}{\sigma^3} \quad \text{“skewness”} \quad \text{and} \quad K = \frac{Z_4}{\sigma^4} - 3 \quad \text{“kurtosis”}$$

Normalization by σ makes S and K to quantities which depend only on the shape. For symmetric distributions one has $S = 0$. K measures how quickly the PDF drops to zero. For gaussian distributions one has $K = 0$.

➜ *probability content in the tails*

Given any PDF $f(x)$ und eine Funktion $w(x) \geq 0$, there is a relation between $\langle w \rangle$ and the probability $p(w(x) \geq C)$, to observe x in a region with $w(x) \geq C$:

$$\langle w \rangle = \int dx\, f(x)\, w(x) \geq \int\limits_{w(x) \geq C} dx\, f(x) w(x) \geq C \int\limits_{w(x) \geq C} dx\, f(x) = C\; p(w(x) \geq C)$$

$$\text{and thus} \qquad p(w(x) \geq C) \;\leq\; \frac{\langle w \rangle}{C}$$

The special choice $w(x) = (x - \mu)^2$ and $C = k^2\sigma^2$ then yields the result:

$$p_k \equiv p\left((x-\mu)^2 > k^2\sigma^2\right) \leq \frac{1}{k^2}$$

The probability content beyond $\pm k\,\sigma$ around the mean value μ is at most $1/k^2$.

- upper limit for probability in the tails of a PDF
- actual probability contents for most PDFs are much lower
  - ➜ e.g. gaussian: $\{p_1, p_2, p_3, p_4\} \approx \{0.317, 0.0555, 0.0027, 0.000063\}$

➜ *convolution of two distributions*

Given two PDFs $f_1(x_1)$ und $f_2(x_2)$, determine the PDF $g(y)$ of $y = h(x_1, x_2)$, when x_1 and x_1 are distributed according to $f_1(x_1)$ and $f_2(x_2)$, respectively.
For the cumulative distribution $G(Y)$ one has:

$$G(Y) = \int_{-\infty}^{Y} dy\ g(y) = \int dx_1\, dx_2 f_1(x_1) f_2(x_2)\ \Theta(Y - h(x_1, x_2))$$

Here the products of all probabilities $dp_1 = dx_1 f_1(x_1)$ and $dp_2 = dx_2 f_2(x_2)$ are summed which satisfy the constraint $h(x_1, x_2) < Y$. Differentiation with respect to the upper limit Y then yields the solution:

$$g(y) = \frac{d}{dY} G(Y)\Big|_{Y=y} = \int dx_1\, dx_2 f_1(x_1) f_2(x_2) \delta(y - h(x_1, x_2))$$

"general convolution integral"

For the special case $h(x_1, x_2) = x_1 + x_2$ follows the known result

$$g(y) = \int dx_1 f_1(x_1) f_2(y - x_1)$$

➜ consider moments. . .

Algebraic moments of convolutions

$$M_k(y) = \int dy\, y^k g(y) = \int dy\, y^k \int dx_1\, dx_2 f_1(x_1) f_2(x_2) \delta(y - x_1 - x_2)$$

$$= \int dx_1\, dx_2 f_1(x_1) f_2(x_2) \int dy\, y^k \delta(y - (x_1 + x_2))$$

$$= \int dx_1\, dx_2 f_1(x_1) f_2(x_2) (x_1 + x_2)^k$$

Leading order moments:

$$\langle y^0 \rangle = \int dx_1\, dx_2 f_1(x_1) f_2(x_2) = 1$$

$$\langle y^1 \rangle = \int dx_1\, dx_2 f_1(x_1) f_2(x_2) (x_1 + x_2) = \langle x_1 \rangle + \langle x_2 \rangle$$

$$\langle y^2 \rangle = \int dx_1\, dx_2 f_1(x_1) f_2(x_2) (x_1 + x_2)^2 = \langle x_1^2 \rangle + 2 \langle x_1 \rangle \langle x_2 \rangle + \langle x_2^2 \rangle$$

and thus $$\langle y^2 \rangle - \langle y \rangle^2 = \left[\langle x_1^2 \rangle - \langle x_1 \rangle^2 \right] + \left[\langle x_2^2 \rangle - \langle x_2 \rangle^2 \right]$$

➜ *convolutions are normalized, mean value and variance add up for any PDFs!*

➜ *conditions:*

- n PDFs $f_i(x_i)$ with mean values μ_i and variances $\sigma^2(x_i)$
- all algebraic moments are finite, i.e. the PDFs $f_i(x_i)$
  - ➜ drop for $|x_i| \to \infty$ faster than any power of x_i
  - ➜ or only within a finite interval one has $f_i(x_i) \neq 0$
- consider the derived variable y:

$$y = \sum_{y=1}^{n} y_i = \sum_{i=1}^{n} \frac{x_i - \mu_i}{\sigma} = h(x_1, \ldots x_n) \quad \text{with} \quad \sigma^2 = \sum_{i=1}^{n} \sigma^2(x_i)$$

  - ➜ y is a convolution of n PDFs with mean value $\mu = 0$
  - ➜ y is dimensionless
  - ➜ y is constructed such that the variances is $\sigma^2(y) = 1$

➜ *central limit theorem:*

For $n \to \infty$ the PDF of y converges towards a normal distribution $N(0,1)$:

$$g(y) = \lim_{n\to\infty} \int \prod_{i=1}^{n} dx_i \, f_i(x_i) \, \delta\left(y - h(x_1, \ldots x_n)\right) = \frac{1}{\sqrt{2\pi}} \, e^{-y^2/2}$$

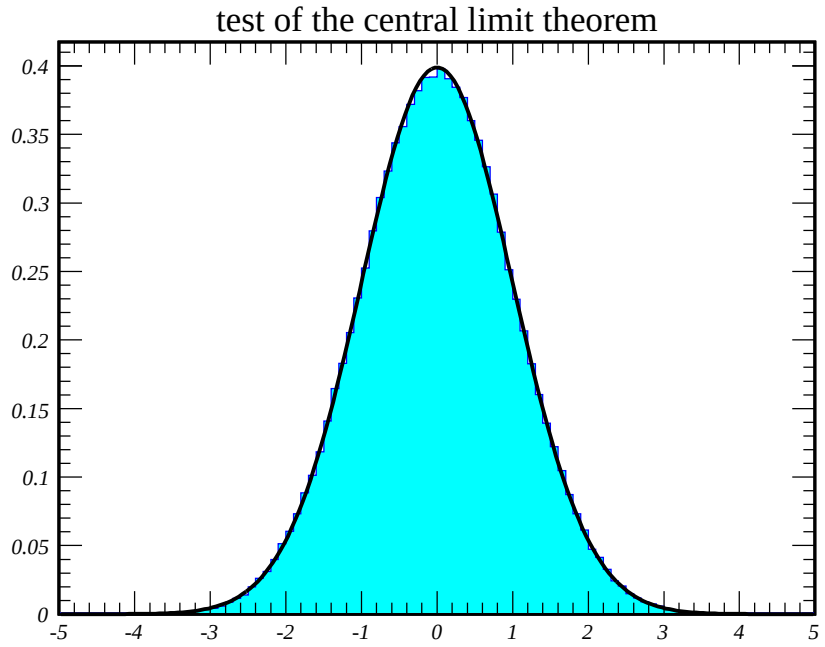
Illustration of the central limit theorem

➜ *convergence toward a normal (gaussian) distribution*

- generate n random numbers x_i according to two PDF
  - ➜ uniform distribution with $\sigma = 1/\sqrt{12}$
  - ➜ exponential distribution with $\sigma = 1$
- calculate the function $y = h(x_1, \ldots, x_n)$
  - ➜ $h = \sqrt{12/n} \sum_i x_i$ foruniform random numbers
  - ➜ $h = \sqrt{1/n} \sum_i x_i$ for exponential random numbers
- histogram y
- study convergence

A simple example how to do convolutions numerically

(demo 03)



➜ *generalization of 1-dim PDFs*

- non-negative, normalizable functions in n dimensions
- discuss the most important concepts with 2-dim PDFs

❖ 2-dim PDF:

$$f(x,y) \geq 0 \qquad \text{and} \qquad \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, f(x,y) = 1$$

❖ interpretation:

Probability for (x, y) in the (infinitesimal) rectangle $[x, x+dx] \times [y, y+dy]$

$$p(x, x+dx; y, y+dy) = f(x,y)\, dx\, dy$$

❖ independence of variables:

Two variables x and y are independent if the PDF factorizes

$$f(x,y) = f_y(x) \cdot f_x(y) = \left(\int dy\, f(x,y) \right) \cdot \left(\int dx\, f(x,y) \right)$$

➜ *look for a moment sensitive two dependencies between two variables*

normalisation	$\langle 1 \rangle$
first moments	$\langle x \rangle$, $\langle y \rangle$
second moments	$\langle x^2 \rangle$, $\langle xy \rangle$, $\langle y^2 \rangle$
third moments	$\langle x^3 \rangle$, $\langle x^2 y \rangle$, $\langle xy^2 \rangle$, $\langle y^3 \rangle$ etc.

The lowest order term sensitive to possible dependencies between x and y is $\langle xy \rangle$. For independent variables with $f(x,y) = g_1(x)\, g_2(y)$ one finds

$$\langle xy \rangle = \int dx \int dy\, x \cdot y \cdot g_1(x) \cdot g_2(y) = \left(\int dx\, x \cdot g_1(x) \right) \left(\int dy\, y \cdot g_2(y) \right)$$

and thus $\langle xy \rangle = \langle x \rangle \langle y \rangle$

❖ obvious candidat for a measure of correlation:

$$C_{xy} = \langle xy \rangle - \langle x \rangle \langle y \rangle$$ "covariance" of x and y

❖ consider the special case $y = ax + b$

$$\begin{array}{rrcccl} & \langle y \rangle & = & \langle ax + b \rangle & = & a\langle x \rangle + b \\ & \langle xy \rangle & = & \langle ax^2 + bx \rangle & = & a\langle x^2 \rangle + b\langle x \rangle \\ \text{and thus} & C_{xy} & = & \langle xy \rangle - \langle x \rangle \langle y \rangle & = & a(\langle x^2 \rangle - \langle x \rangle^2) = aC_{xx} \end{array}$$

Here the covariance is proportional to the slope between x and y, i.e. it measures linear correlation. The dimensionless correlation coefficient ρ derived from C_{xy} is a normalized measure for the correlation strength.

➜ (linear) correlation coefficient: $\rho = \dfrac{C_{xy}}{\sigma_x \sigma_y} = \dfrac{C_{xy}}{\sqrt{C_{xx} C_{yy}}}$

For $y = ax + b$ one has $C_{xy} = aC_{xx}$ and $C_{yy} = a^2 C_{xx}$ and thus:

$$y = ax + b \quad \Rightarrow \quad \rho = \text{sign}(a) = \pm 1$$

The correlation is 100%. If the linear relation only holds between x and $\langle y \rangle$, i.e. $\langle y \rangle = a\,x + b$, then one has $|\rho| < 1$.

The covariance matrix

➜ *matrix of covariances between all pairs of variables in an n-dim PDF:*

$$C_{ij} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$$

Expressed through standard deviations and correlation coefficients it is

$$C_{ij} = \rho_{ij} \cdot \sigma_i \sigma_j \quad \text{with} \quad \rho_{ii} = 1 \ .$$

➜ *note:*

- the diagonal terms C_{ii} are the variances of the individual variables
- off-diagonal terms are covariances
- the covariance matrix is symmetric and positive definite
- it can be diagonalized by rotation in the space of the variables
- C also is referred to as "error matrix"

The covariance matrix C_{ij} is the matrix of all 2nd order moments of an n-dimensional PDF $f(x_1, x_2, \ldots, x_n)$. Mean values $\langle x_i \rangle$ and C_{ij} describe the location, extension and orientation of the PDF.

➜ *manipulations of sums. . .*

Consider a transformation $y_k = \sum_i A_{ki} x_i$. Given the covariance matrix $C_{ij}(x)$, the covariance matrix $C_{kl}(y)$ of the transformed quantities shall be determined:

$$
\begin{aligned}
C_{kl}(y) &= \langle y_k y_l \rangle - \langle y_k \rangle \langle y_l \rangle \\
&= \left\langle \sum_i (A_{ki} x_i) \sum_j (A_{lj} x_j) \right\rangle - \left\langle \sum_i A_{ki} x_i \right\rangle \left\langle \sum_j A_{lj} x_j \right\rangle \\
&= \sum_{ij} A_{ki} A_{lj} (\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle) = \sum_{ij} A_{ki} A_{lj} C_{ij}(x)
\end{aligned}
$$

Matrix notation yields the compact expressions

$$
\vec{y} = A \cdot \vec{x} \quad \text{and} \quad C(y) = A \cdot C(x) \cdot A^T \ .
$$

➜ if $C(x)$ is positive definite, so is $C(y)$

➜ A need not be a square matrix - the number of rows is arbitrary

➜ *functional form:*

$$f(\vec{x};\vec{\mu},C) = \frac{1}{\sqrt{(2\pi)^n \det C}} \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^T C^{-1}(\vec{x}-\vec{\mu})\right]$$

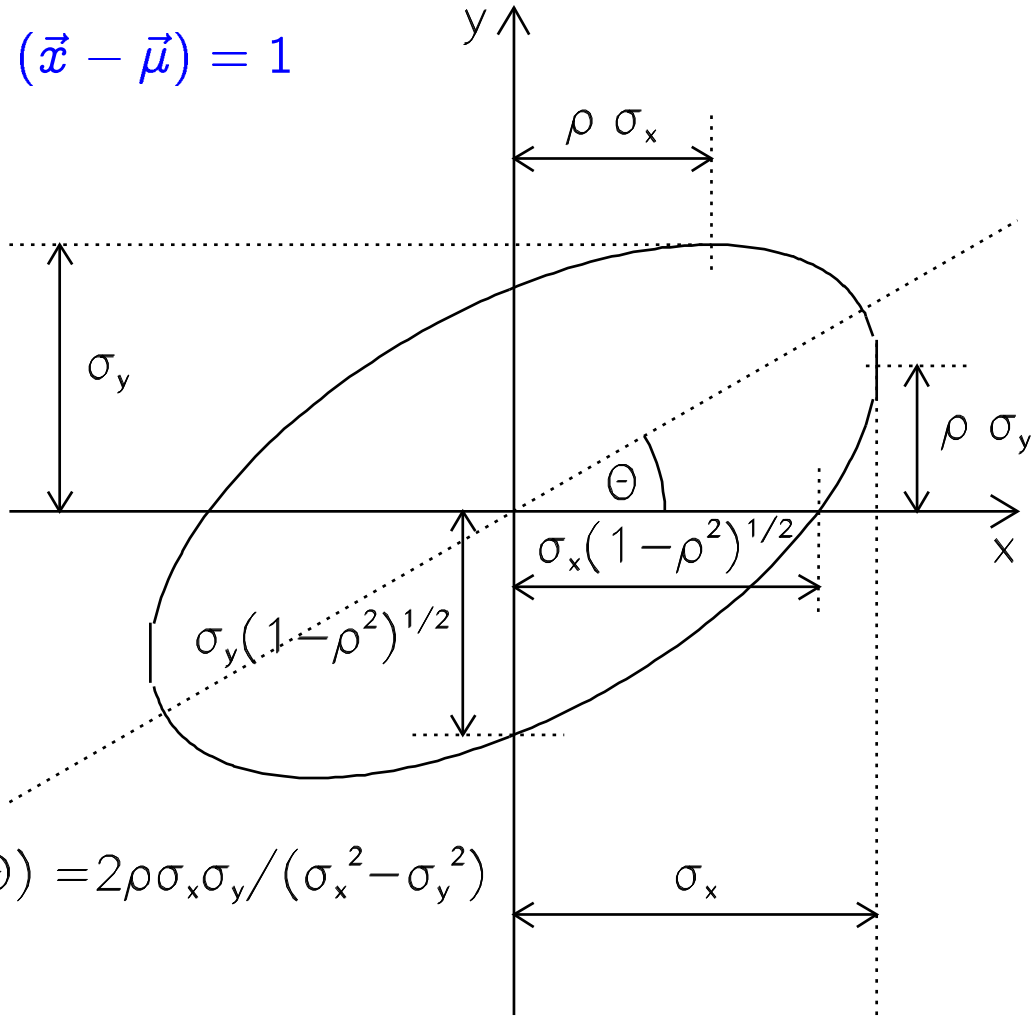
- exponential of a general n-dimensional parabola
- pre-factor guarantees proper normalization
- vector of expectation values $\vec{\mu}$
- covariance matrix C
- orientation and extension of the PDF described by C
- (hyper)plannes of constant probability density are ellipsoids
- complete description in n dimensionens:
  - ➜ n expectation values
  - ➜ n variances (diagonal elements of C)
  - ➜ $n(n-1)/2$ covariances
  - ➜ in total $n(n+3)/2$ Parameter

➜ 2-dim case

Covariance-ellipse of a 2-dim gaussian

➜ *line of constant probability density:*

$$(\vec{x}-\vec{\mu})^T\, C^{-1}\, (\vec{x}-\vec{\mu}) = 1$$



$$\tan(2\Theta) = 2\rho\sigma_x\sigma_y/(\sigma_x^2-\sigma_y^2)$$

$$\text{Hauptachsen: } 2\sigma^2=\sigma_x^2+\sigma_y^2\pm[(\sigma_x^2-\sigma_y^2)^2+4\rho^2\sigma_x^2\sigma_y^2]^{1/2}$$

➜ *basic idea:*

Study inherent statistical processes by direct simulation or map deterministic problems to statistical ones, which then are solved by simulation. The latter exploits that expectation values are defined via integrals.

➜ *needed:*

Random numbers which are distributed according to well defined PDFs.

- start with random numbers with uniform distribution in the intervall $[0, 1]$.
- the derive other distributions from those

➜ *technical realization: "pseudo random numbers"*

- generation via numerical algorithms
- not random, but hopefully indistinguishable from true random numbers
- reproducible sequence – important for debugging

Pseudo random number generators (i)

➜ *D.E. Knuths's 10-decimal-digits X "Super-random" number generator*

1. $Y = X/10^9$ iterate the next steps Y times
2. $Z = X/10^8$ mod 10 jump to step $Z + 3$
3. `if` $(X < 5 \cdot 10^9)\ \{X += 5 \cdot 10^9\}$
4. $X =$ midsquare(X)
5. $X = (X \cdot 1001001001)$ mod 10^{10}
6. `if`$(X < 100000000)\ \{X += 9814055677\}$ `else` $\{X = 10^{10} - X\}$
7. swap upper and lower 5-digit blocks
8. $X = (X \cdot 1001001001)$ mod 10^{10}
9. reduce every digit > 0 by 1
10. `if` $(X < 10^5)\ \{X = X^2 + 99999\}$ `else` $\{X -= 99999\}$
11. `while` $(X < 10^9)\ \{X *= 10\}$
12. replace X by the central 10 digits of $X(X-1)$

➜ extremely complex - sequence of steps is randomized internally

➜ properties of the generator are not discernible

➜ the generator is useless: 6065038420 is a fixed point of the algorithm

lesson learned: use only generators with known properties!

Pseudo random number generators (ii)

➜ *linear congruential generators*

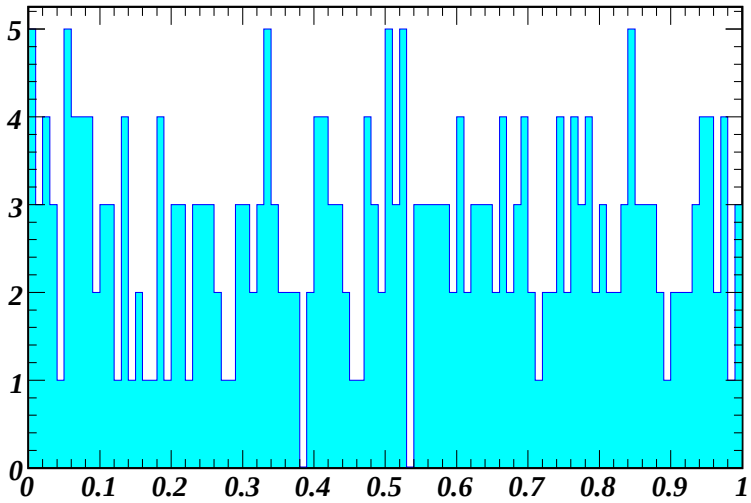
$$x_{n+1} = (a \cdot x_n + b) \text{ mod } m$$

- ➜ multiplication with a scrambles the digits of x_n
- ➜ the constant term b prevents trivial fixed points
- ➜ mod m takes care the x stays in the range $[0, m-1]$ (x/m in $[0, 1]$)
- ➜ properties/quality is determined by the parameters a, b and m

❖ study the properties of the generator for $a = 1601$, $b = 3456$ und $m = 10000$

```
seed=1601 - x=0.1601
seed=6657 - x=0.6657
seed=1313 - x=0.1313
seed=5569 - x=0.5569
seed=9425 - x=0.9425
seed=2881 - x=0.2881
seed=5937 - x=0.5937
seed=8593 - x=0.8593
seed=0849 - x=0.0849
seed=2705 - x=0.2705
etc.
```

➜ distribution of function values



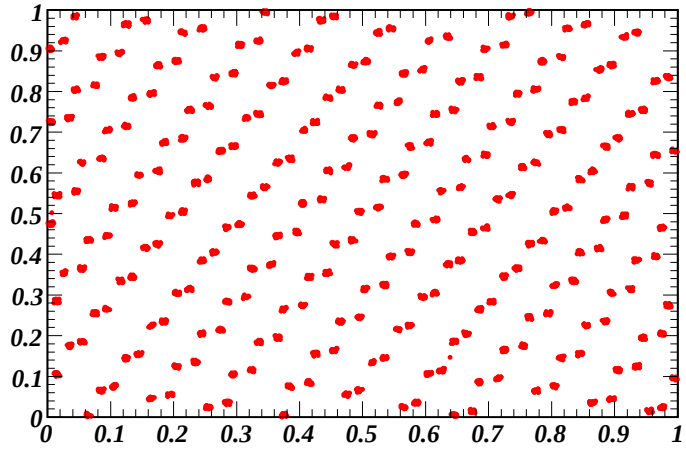
(demo 04)

➜ *result:*

- linear congruential random numbers are located on (hyper)planes
- the number of hyperplanes is a function of the plot-dimension d
- wanted: number of hyperplanes as large and period as long as possible
- both characteristics grow with the number of bits per integer
- the choice of parameters a, b, m is important, too

❖ maximum number of hyperplanes with t-bit integers: $p = (d!\ 2^t)^{1/d}$

bits	$d = 3$	$d = 4$	$d = 6$	$d = 10$
t=16	73	35	19	13
t=32	2953	566	120	41
t=36	7442	1133	191	54
t=48	119086	9065	766	126
t=60	1905376	72520	3064	290



➜ need to increase the number of bits being used . . .

➜ *example for an state-of-the-art generator:* RANLUX

- based on the Marsaglia-Zaman algorithm
  - ➜ mathematically equivalent to a linear-congruential generator
  - ➜ completely different implementation
  - ➜ for details consult Martin Lüscher, hep-lat/9309020

implementation:

$$z_n = (a \cdot z_{n-1}) \bmod m$$

with

$$m = 2^{576} - 2^{240} + 1 \quad \text{(prime)} \quad \text{and} \quad a = 2^{576} - 2^{552} - 2^{240} + 2^{216} + 1$$

❖ Discussion:

Effectively RANLUX uses 576-bit integer variables. The period is $\approx 5.2 \cdot 10^{171}$, and the number of hyperplanes in $d = 100$ dimensiones ist $h \approx 2000$. However, since the multiplier a has only very few bits set, subsequent 576-bit states are still correlated. As in deterministic chaos, the correlation decays exponentially with the distance of two numbers, and RANLUX/luxury-level 4, discards 8760 bits, before the next 576 bits are accepted.

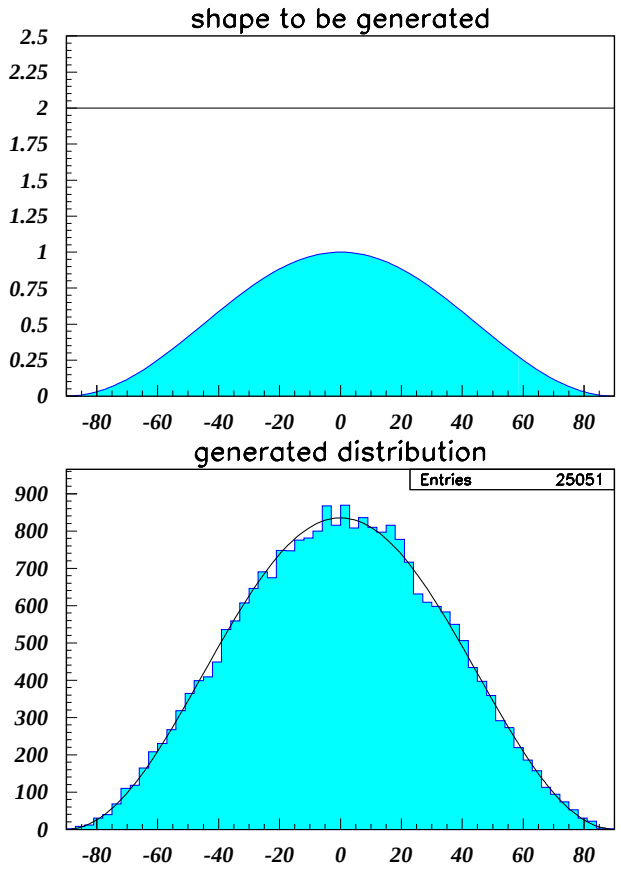
Non-uniform random numbers (i)

➜ *generate non-uniform random numbers from uniform ones*

❖ Hit-or-Miss method

➜ algorithm to generate $\rho(x) < M$ in the intervall $[a, b]$:

(1) generate x_1 uniform in $[0, 1]$
(2) scale $x = a + (b - a)x_1$
(3) generate y_1 uniform in $[0, 1]$
(4) scale $y = My_1$
(5) accept x if $y < \rho(x)$
(6) goto (1)



properties

➜ simple concept
➜ normalization of ρ not necessary
➜ small efficiency if $M \gg \langle\rho\rangle$
➜ recycling of y_1 as next x_1 is possible

example: $\rho(x) \sim \cos^2\left(x\frac{\pi}{180}\right)$

for $-90 \leq x \leq 90$; generation with $M = 2$

➜ *The transformation method*

If x is uniformly distributed in $[0,1]$, then $y = h(x)$ is a random variable with a different distribution $g(y)$. With proper choice of $h(x)$ it should be possible to realize any distribution $g(y)$. From the section about convolutions we know:

$$g(y) = \int dx\, f(x)\delta(y - h(x)) = \int_0^1 dx\, \delta(y - h(x)) = \left.\frac{1}{h'(x)}\right|_{x=h^{-1}(y)}$$

On the other hand we have

$$h(h^{-1}(y)) = y \quad \text{and, differentiating w.r.t. } y \quad h'(h^{-1}(y)) \cdot (h^{-1}(y))' = 1$$

giving

$$g(y) = \frac{1}{h'(h^{-1}(y))} = (h^{-1}(y))' \quad \text{and finally} \quad h^{-1}(Y) = \int_{y_{\mathrm{min}}}^{Y} dy\, g(y)$$

i.e. the transformation h is the inverse of the integral of the target distribution g. Note also that the transformation method outlined above can be generalized to the case that h is a function of several variables.

Non-uniform random numbers (iii)

➜ *some transformation laws for uniformly-distributed inputs* $x_1, x_2, \ldots$

$y = h(x_1, \ldots, x_n)$	➜	$g(y)$
$\sqrt{x_1}$	➜	$2y$
$-a \ln(x_1)$	➜	$\frac{1}{a} e^{-y/a}$
$\sqrt{-a \ln(x_1)}$	➜	$\frac{2}{a} y e^{-y^2/a}$
$-a \ln(x_1 x_2)$	➜	$\frac{1}{a^2} y e^{-y/a}$
$-\ln(x_1 x_2 \ldots x_n)$	➜	$y^{n-1} e^{-y}$
$\sqrt{-2 \ln(x_1)} \begin{cases} \cos(2\pi x_2) \\ \sin(2\pi x_2) \end{cases}$	➜	$\frac{1}{\sqrt{2\pi}} e^{-y^2/2}$

Verification of these relations:

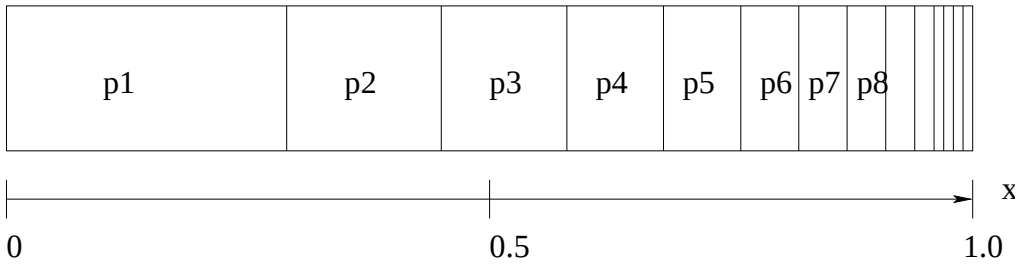
$$g(y) = \int dx_1 \cdot dx_2 \cdots x_n \, \delta(y - h(x_1, x_2, \ldots, x_n))$$

➜ *Generation of Discrete Probability Distributions*

Starting from the fact that the sum of discrete probabilities p_i is normalized,

$$\sum_i p_i = 1 \ ,$$

the individual probabilities can be arranged along $0 \leq x \leq 1$. Drawing then a uniform random number from $[0, 1]$, the interval containing the generated value x determines the discrete state to be returned, i.e. n if the interval taken by p_n is hit.



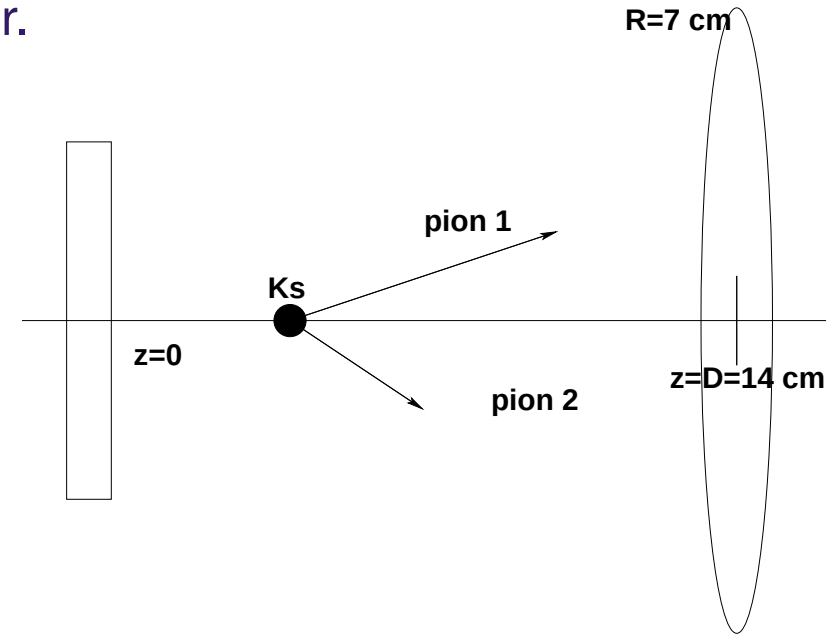
Iterative algorithm to find the hit interval:

start with $S_0 = 0$ and iterate $S_n = S_{n-1} + p_n$ until $S_n > x$.

Note that this algorithm is most efficient if the p_n are ordered in decreasing size.

Monte Carlo integration - example (i)

A thin beam of monoenergetic K_S-mesons (mass $M = 0.498$ GeV) with energy $E_K = M^2/(2m) \approx 0.886$ GeV, with $m = 0.13957$ GeV the pion mass enters an experiment. The K_S mesons decay with an average proper lifetime $\langle\tau\rangle = 8.934 \cdot 10^{-10}$ s, into $\pi^+\pi^-$ pairs. In the rest system of the K_s the decays are isotropic. Located at a distance of 14 cm behind the entry point is a silicon detector to register the decay pions. The detector is a circular disk with radius $R = 7$ cm, centered on the K_s-beam. Determine the probability that both decay pions hit the detector.



➜ *analytical calculation*

Due to the azimuthal symmetry of the problem, the monoenergetic K_s- energy spectrum and the special choice of the energy the problem can almost be solved anayltically. Only three PDFs are required:

- $f(\phi) = 1/2\pi$: azimuth of the π^+ direction in the K_s-rest system
- $g(C) = 1$: cosine of the π^+ polar angle in the K_s-rest system
- $\rho(z) = (1/\langle z \rangle) \exp(z/\langle z \rangle)$: flight distance in the lab with $\langle z \rangle \approx 4.07$ cm

With the integration limits that result from the fact that as function of C only decays in a certain z-range are registered on the detector

$$z_{\min} = D - R\,\gamma\sqrt{\frac{1-C}{1+C}} \;\leq\; z \;\leq\; D$$

one finally obtains:

$$A = \int_0^{2\pi} d\phi \int_0^1 dC \int_{z_{\min}}^{D} dz\; f(\phi) g(C) \rho(z) = \int_0^1 dC \int_{z_{\min}}^{z_{\max}} dz\; \rho(z)$$

$$= e^{-D/\langle z \rangle} \left[-1 + \int_0^1 dC \exp\left(\frac{R\gamma}{\langle z \rangle} \sqrt{\frac{1-C}{1+C}} \right) \right] \approx 0.207016$$

➜ *direct simulation*

- generate decay positions according to $\rho(z)$
- generate isotropic decays in the K_s- rest frame
- perform Lorentz transform to the lab-system
- analyze the events, i.e. check whether the pions hit the detector
  - ➜ count as "success" cases where both pions hit the detector
  - ➜ count as "failure" if at least one particle misses

❖ result of a simulation:

For N trials, the number of accepted events follows binomial statistics with approximately $p = n/N$. The number of successes n will fluctuate with a standard deviation $\sigma(n) = \sqrt{Np(1-p)}$ around its expectation value, and one finds:

$$A = \frac{n}{N} \pm \frac{\sigma(n)}{N} = \frac{n}{N} \pm \sqrt{\frac{p(1-p)}{N}} \; .$$

Given $N = 1\,000\,000$ generated events and $n = 207\,311$ successes, one has:

$A = 0.207\,311(405)$ versus the analytical result $A = 0.207\,016\ldots$

➜ *basic idea*

Exploit the definition of an expectation value as a weighted integral over a PDF and map an integration problem on to the determination of a statistical average. The idea can be applied to integrals over an arbitrary number of dimensions.

$$\int d^n x \, \rho(\vec{x}) \, f(\vec{x}) = \langle f \rangle$$

➜ *determination of the expectation value*

If vectors $\vec{x}_i$, $i = 1, \ldots, N$ are distributed according to $\rho(\vec{x})$, then the arithmetic mean provides an estimate for the required expectation value:

$$\langle f \rangle = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} f(\vec{x}_i)$$

➜ note:

The factor $1/N$ is independent of the number of dimensions in $\vec{x}$, i.e. Monte Carlo integration will converge with the same rate in any number of dimensions!

➜ *properties and limiting behavior of the arithmetic average*

$$\lim_{N\to\infty} \frac{1}{N} \sum_i f(\vec{x}_i)$$

Split the space of $\vec{x}$ into h n-dimensional (infinitesimal) volumes Δ_h^n around locations $\vec{x}_h$, such that $\rho(\vec{x}) \approx \rho(\vec{x}_h)$ and $f(\vec{x}) \approx f(\vec{x}_h)$. When generating N vectors $\vec{x}_i, i = 1, \ldots, N$ according to $\rho(\vec{x})$, the number of events in Δ_h^n will be n_h. The sum over i then can be rewritten as a sum over h:

$$\sum_{i=1}^{N} f(\vec{x}_i) = \sum_h n_h \, f(\vec{x}_h)$$

In the limit $N \to \infty$ one then has $n_h = N \, p_h = N \, \rho(\vec{x}_h) \, \Delta_h^n$, and thus

$$\lim_{N\to\infty} \frac{1}{N} \sum_i f(\vec{x}_i) = \lim_{N\to\infty} \frac{1}{N} \sum_h N \, \Delta_h^n \, \rho(\vec{x}_h) \, f(\vec{x}_h) = \sum_h \Delta_h^n \, \rho(\vec{x}_h) \, f(\vec{x}_h)$$

$$= \int d^n x \, \rho(\vec{x}) \, f(\vec{x})$$

➜ *properties of Monte Carlo estimates at finite statistics N*

The arithmetic mean $\bar{f}$ provides an estimate for $\langle f \rangle$. Since all points $\vec{x}_i$ are equivalent one has $\langle f^k(x_i) \rangle = \langle f^k \rangle$ and with $f(x_i) \equiv f_i$ one finds:

$$\langle \bar{f} \rangle = \left\langle \frac{1}{N} \sum_{i=1}^{N} f_i \right\rangle = \frac{1}{N} \sum_{i=1}^{N} \langle f_i \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle f \rangle = \langle f \rangle$$

The estimate $\bar{f}$ is unbiased. The variance of $\bar{f}$ is

$$\sigma^2(\bar{f}) = \left\langle \left(\frac{1}{N} \sum_{i=1}^{N} f_i \right) \left(\frac{1}{N} \sum_{j=1}^{N} f_j \right) \right\rangle - \left\langle \frac{1}{N} \sum_{i=1}^{N} f_i \right\rangle^2$$

$$= \frac{1}{N^2} \sum_{i=j}^{N} \langle f_i f_j \rangle + \frac{1}{N^2} \sum_{i \neq j}^{N} \langle f_i f_j \rangle - \langle f \rangle^2$$

$$= \frac{N}{N^2} \langle f^2 \rangle + \frac{N(N-1)}{N^2} \langle f \rangle^2 - \langle f \rangle^2 = \frac{1}{N} (\langle f^2 \rangle - \langle f \rangle^2) = \frac{1}{N} \sigma^2(f)$$

➜ *summary*

- Monte Carlo integration converges with an uncertainty $\sigma(f)/\sqrt{N}$
- independently of the dimension of the integration problem
- $\sigma^2(f) = \langle f^2 \rangle - \langle f \rangle^2$ a priori usually is not known
- can be estimated with uncertainty $\mathcal{O}(1/\sqrt{N})$ during the calculation
- for random vectors $\vec{x}_i$ distributed according $\rho(\vec{x})$ one has

$$\int d^n x\, \rho(\vec{x})\, f(\vec{x}) =$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(\vec{x}_i) \pm \frac{1}{\sqrt{N}} \sqrt{\left(\frac{1}{N} \sum_{i=1}^{N} f^2(\vec{x}_i) \right) - \left(\frac{1}{N} \sum_{i=1}^{N} f(\vec{x}_i) \right)^2}$$

➜ *important:*

All $f(\vec{x}_i)$ are random variables. According to the central limit theorem the Monte Carlo estimate thus will be gaussian distributed around the true value, which is contained in ca. 68% of the cases in the $\pm 1\sigma$ interval.

➜ *consider the 1-dim case*

- convergence is universal with $\sigma(f)/\sqrt{N}$
- only reduction of $\sigma(f)$ can improve convergence
- possible, since Monte Carlo integration integrates a product of two functions
  - ➜ weight function $f(\vec{x})$ and PDF $\rho(\vec{x})$ of the sampling point distribution
  - ➜ exploit the freedom to redistribute the factors in order to minimize $\sigma(f)$

$$I = \int_a^b dx\, f(x)$$

Evaluation of I via Monte Carlo means re-writing it as

$$I = \int_a^b dx\, \rho(x)\, f_\rho(x) \quad \text{with} \quad \rho(x) \cdot f_\rho(x) = f(x) \quad \text{and normalized PDF} \quad \rho(x)$$

Two (out of many) possibilities to distribute sampling points x over $[a, b]$:

- uniform distribution: $\rho(x) = 1/(b-a)$ and $f_\rho(x) = (b-a)f(x)$
- distribution proportional to $f(x)$: $\rho(x) = f(x)/I$ and $f_\rho(x) = I$
  - ➜ convergence at the first event - possible if $f(x) > 0$

Importance sampling (ii)

➜ *discussion*

- convergence improves if the density of sampling points matches $f(x)$
- "important" parts of $f(x)$ are visited more often ➜ "importance sampling"
- a common realization of importance sampling is direct simulation
- improvement already if $\rho(x) \sim f(x)$
- general approach:

$$\int dx\ f(x) \Rightarrow \int dx\ \rho(x)\ \frac{f(x)}{\rho(x)}$$

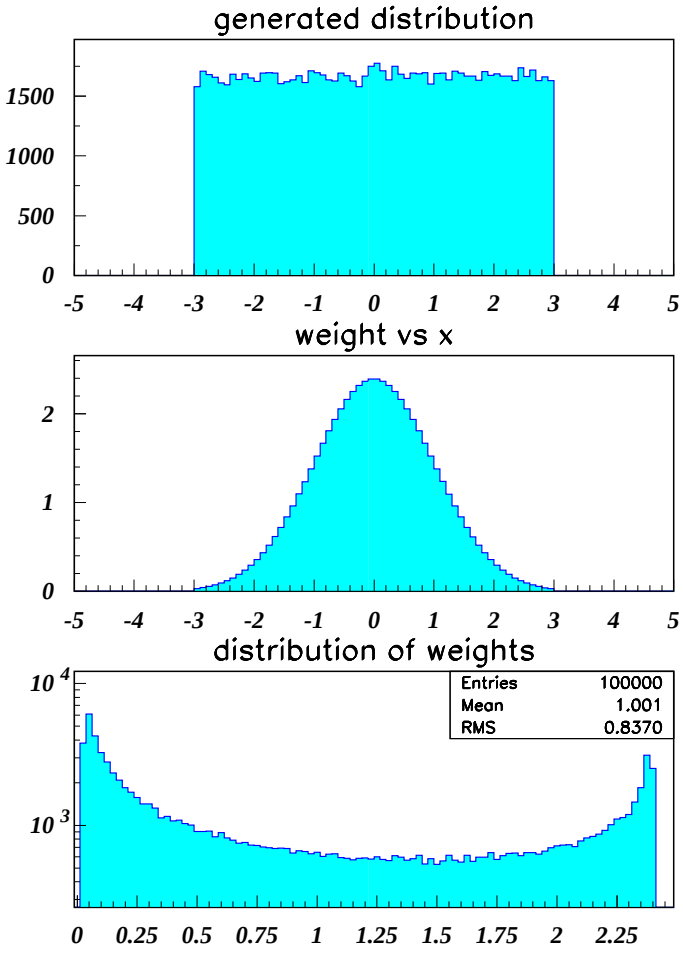
  - ➜ make sure that $\rho(x) > 0$ over the integration interval
  - ➜ $\rho(x)$ can have integrable singularities, i.e. while MC does not converge if $f(x)$ has (integrable) singularities, when the singularity is moved to $\rho(x)$ importance sampling allows MC-integration also of singular functions

➜ *numerical example: integrate a gaussian density*

$$I = \int_{-3}^{3} dx \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \Rightarrow \int_{-3}^{3} dx\ \rho(x)\, f_\rho(x)$$

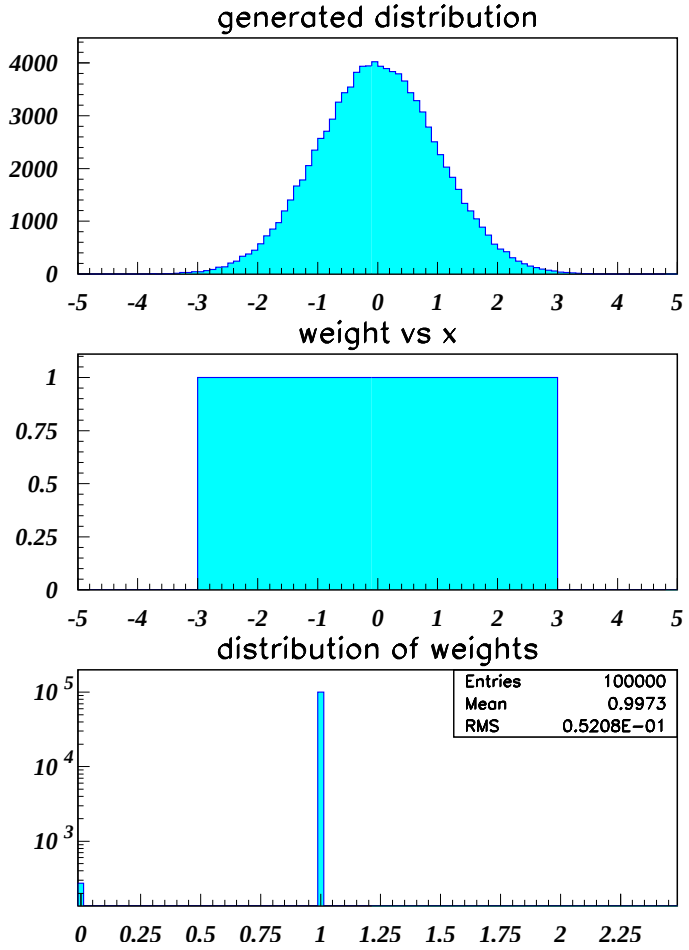
Importance sampling (iii)

❖ uniform distribution



$I = 1.0014 \pm 0.0026$

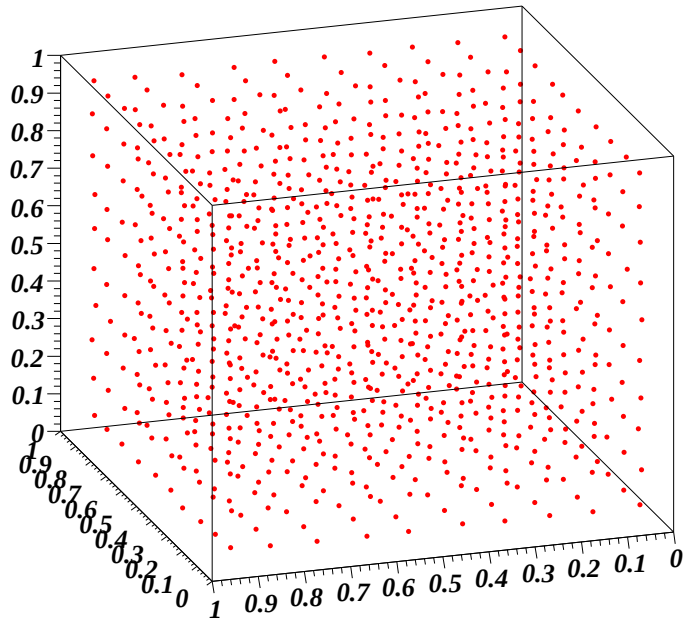
❖ importance sampling



$I = 0.99728 \pm 0.00017$

Concluding remarks

- universal convergence $\sim 1/\sqrt{N}$ in Monte Carlo integration
- grid based methods in 1-dim can have much better convergence
  - ➜ e.g. $1/N^4$ using Simpson's rule
- grid based methods degrade dramatically for higher dimensions
- cartesian grids are coarse and "irregular"
  - ➜ spacing along axes $N^{1/d}$
  - ➜ spacing along 2-d diagonals $\sqrt{2}N^{1/d}$
  - ➜ spacing along 3-d diagonals $\sqrt{3}N^{1/d}$
- much more uniform sampling through MC
- $1/\sqrt{N}$ convergence because of clustering
- $1/N$ convergence for truly uniform sampling
- attempted by "Quasi Monte Carlo" approach



➜ *note:*

Monte Carlo techniques are well understood and provide reliable error estimates.
Quasi Monte Carlo has not yet reached that level.

➜ *what are errors?*

- "errors" are uncertainties - not to be confused with "mistakes"
- meant to quantify how well one knows e.g. a constant of nature - but how?
- engineer - one preferred quantity: "tolerance"
  - ➜ maximum possible deviation
- physicist - many different conventions. . .
  - ➜ standard deviation (symmetric)
    - ✔ most common
  - ➜ 3-sigma errors (symmetric)
    - ✔ if you want to be conservative
  - ➜ frequentist confidence level intervals (asymmetric)
    - ✔ region containing the true value in a certain fraction of experiments
  - ➜ bayesian confidence level intervals (asymmetric)
    - ✔ region in which the true value is located with a certain probability
- most likely more options on the market

➜ ask the professionals. . .

➜ *Recommendation INC-1 (1980):* Expression of experimental uncertainties

Joint Committee for Guides in Metrology/WG 1 (JCGM 100:2008)

1 The uncertainty in the result of a measurement generally consists of several components which may be grouped into two categories according to the way in which their numerical value is estimated:

A those which are evaluated by statistical methods,
B those which are evaluated by other means.

There is not always a simple correspondence between the classification into categories A or B and the previously used classification into“random” and “systematic” uncertainties. The term “systematic uncertainty” can be misleading and should be avoided. Any detailed report of the uncertainty should consist of a complete list of the components, specifying for each the method used to obtain its numerical value.

2 The components in category A are characterized by the estimated variances s_i^2 (or the estimated "standard deviations" s_i) and the number of degrees of freedom ν_i. Where appropriate, the covariances should be given.

3 The components in category B should be characterized by quantities u_j^2, which may be considered as approximations to the corresponding variances, the existence of which is assumed. The quantities u_j^2 may be treated like variances and the quantities u_j like standard deviations. Where appropriate, the covariances should be treated in a similar way.

4 The combined uncertainty should be characterized by the numerical value obtained by applying the usual method for the combination of variances. The combined uncertainty and its components should be expressed in the form of "standard deviations".

5 If, for particular applications, it is necessary to multiply the combined uncertainty by a factor to obtain an overall uncertainty, the multiplying factor used must always be stated.

(end of quote)

Discussion

➜ *why define uncertainties by variances and standard deviations*

- well defined procedures how to handle them
  - ➜ when propagating uncertainties into derived variables
  - ➜ for the combination of independent measurements
- asymptotically gaussian behavior
- no (little) danger of mis-interpretation
- asymmetric errors
  - ➜ not obvious how to propagate those
  - ➜ not obvious how to combine them

❖ the following will deal with variances/standard deviations!

➜ *remark regarding error bars:*

- drawn around measured values
- reflect the scatter of the true values
  - ➜ should (but cannot) be drawn around the true values

➜ *also called "gaussian error propagation"*

the exact expression for linear transformations has been derived before:

$$\vec{y} = A \cdot \vec{x} \quad \text{and} \quad C(y) = A \cdot C(x) \cdot A^T \ .$$

➜ *analyze the 1-dim case of non-linear transformations*

$$y = q(x) = q(\vec{x}) + q'(\langle x \rangle)(x - \langle x \rangle) + \frac{q''(\langle x \rangle)}{2}(x - \langle x \rangle)^2 + \ldots$$

$$= q_0 + q_1(x - \langle x \rangle) + \frac{q_2}{2!}(x - \langle x \rangle)^2 + \ldots$$

with $q_k = q^{(k)}(\langle x \rangle)$ the k-th derivative of $q(x)$ at the expectation value of x. For the expectation value $\langle y \rangle$ one finds

$$\langle y \rangle = q_0 + \frac{q_2}{2!} Z_2 + \frac{q_3}{3!} Z_3 + \ldots$$

The expectation value of a non-linear function $q(x)$ of a random variable x is different from the function of the expectation value $q_0 = q(\langle x \rangle)$. If x is unbiased, then $q(x)$ will be biased, $\langle y \rangle = q_0 + \beta$, with $\beta = q_2 Z_2/2$.

- the bias is proportional to variance of x and second derivative of $q(x)$
- size and sign of the bias can be estimated from estimates of q_2 and Z_2
- in principle a leading order bias correction can be performed
- usually the bias is small compared to the uncertainty $\sigma(y)$ of y and ignored
- if ignored, one may still want to add it in quadrature to the error $\sigma(y)$
  - ➜ slightly lengthy but straightforward calculation yields for $\sigma^2(y)$

$$\sigma^2(y) = \left\langle y^2 \right\rangle - \left\langle y \right\rangle^2 = q_1^2 Z_2 + q_1 q_3 Z_3 + \frac{8 q_1 q_3 + 6 q_2^2}{4!} Z_4 - \frac{1}{4} q_2^2 Z_2^2 \ldots$$

  - ➜ adding the bias $(q_2 Z_2/2)$ in quadrature will cancel the last term

➜ *note:*

The series expansions tend to diverge. For numerical applications one therefore only considers the leading term. The bias term then is of the same (or higher) order as the terms which anyhow are not under control, and can in principle be ignored. The variance of y thus is estimated by $\hat{\sigma}^2(y) = q_1^2 Z_2$.

➜ *final steps*

Since $\langle x \rangle$ is not known, the derivative $q_1 = q'(\langle x \rangle)$ to calculate $q_1^2 Z_2$ is only approximately known. To leading order has

$$q'(x) = q'(\langle x \rangle) + q''(\langle x \rangle)(x - \langle x \rangle) + \ldots = q_1 + q_2(x - \langle x \rangle) + \ldots$$

and thus

$$\left\langle (q'(x))^2 \right\rangle = q_2^1 + 2\,q_1\,q_2\,\langle (x - \langle x \rangle) \rangle + q_2^2 \left\langle (x - \langle x \rangle)^2 \right\rangle + \ldots = q_1^2 + q_2^2 Z_2$$

Taking the derivative at the point of the measurement introduces a bias $O(q_2^2 Z_2^2)$ on the variance estimate which is of the same order as the higher order terms which are anyhow neglected.

➜ *lessons learned*

- non-linear transformation of unbiased variables are biased
- an estimate of the bias is $\beta = q''(x)\sigma^2(x)/2$
- leading order error propagation is done by $\hat{\sigma}^2(y) = (q'(x))^2\sigma^2(x)$
- the uncertainty due to missing higher order terms is of the same size as β^2 or the variance of $q'(x)$ caused by the fact that x is a random variable

➜ *leading order treatment in n dimensions*

- input: $x_i, i = 1, \ldots, n$ with covariance matrix $C(\vec{x})$
- output: $y_k = g_k(\vec{x}), k = 1, \ldots, m$
- wanted: covariance matrix $C(\vec{y})$ of the transformed variables
- step 1: linearization of the transformation:

$$y_k \approx g_k(\langle\vec{x}\rangle) + \sum_{i=1}^{n} \frac{\partial g_k(\langle\vec{x}\rangle)}{\partial x_i}(x_i - \langle x_i\rangle) \approx g_k(\langle\vec{x}\rangle) + \sum_{i=1}^{n} \frac{\partial g_k(\vec{x})}{\partial x_i}(x_i - \langle x_i\rangle)$$

  - ➜ first expression: leading order Taylor expansion around $\langle\vec{x}\rangle$
  - ➜ second expression: derivatives taken at $\vec{x}$
  - ➜ difference is of higher order
- step2: calculate the covariance matrix

$$C_{kl}(\vec{y}) = \langle(y_k - g_k(\langle\vec{x}\rangle))(y_l - g_k(\langle\vec{x}\rangle))\rangle$$

$$= \sum_{i,j=1}^{n} \frac{\partial g_k}{\partial x_i}\frac{\partial g_l}{\partial x_j}\langle(x_i - \langle x_i\rangle)(x_j - \langle x_j\rangle)\rangle = \sum_{i,j=1}^{n} \frac{\partial g_k}{\partial x_i}\frac{\partial g_l}{\partial x_j} C_{ij}(\vec{x})$$

- derivatives at $\vec{x}$ are taken as substitute for derivatives at $\langle\vec{x}\rangle$
  - ➜ justification to treat them as constant in the calculation of $C(\vec{y})$
- the covariance matrix $C(\vec{y})$ is determined not with respect to $\langle\vec{y}\rangle$ but with respect to the true transformed values $\vec{g}(\langle\vec{x}\rangle)$
  - ➜ $C(\vec{y})$ accounts for the bias from the non-linear transformation
  - ➜ no additional "adding in quadrature" needed

with

$$\langle y_k\rangle = g_k(\langle\vec{x}\rangle) + \beta_k \quad \text{and thus} \quad g_k(\langle\vec{x}\rangle) = \langle y_k\rangle - \beta_k$$

one finds

$$\begin{aligned}
&C_{kl}(\vec{y}) \\
&= \langle (y_k - (\langle y_k\rangle - \beta_k))(y_l - (\langle y_l\rangle - \beta_l))\rangle \\
&= \langle y_k y_l\rangle - \langle y_k\rangle(\langle y_l\rangle - \beta_l) - \langle y_l\rangle(\langle y_k\rangle - \beta_k) + \langle(\langle y_k\rangle - \beta_k)(\langle y_l\rangle - \beta_l)\rangle \\
&= \langle y_k y_l\rangle - \langle y_k\rangle\langle y_l\rangle + \beta_k\beta_l = C_{kl}^{\mathrm{true}}(\vec{y}) + \beta_k\beta_l
\end{aligned}$$

or in matrix notation:

$$C(\vec{y}) = C^{\mathrm{true}}(\vec{y}) + \vec{\beta}\vec{\beta}^T$$

➜ *matrix notation for linear error propagation*

Consider a transformation

$$\vec{y} = \vec{g}(\vec{x}) \quad \text{and its jacobian} \quad M \quad \text{with matrix elements} \quad M_{ij} = \frac{\partial g_i}{\partial x_j} \; .$$

Then error propagation transforms $C(\vec{x})$ to $C(\vec{y})$ according to

$$C(\vec{y}) = M(\vec{x}) \cdot C(\vec{x}) \cdot M^T(\vec{x})$$

where the argument to M indicates that the derivatives are taken at $\vec{x}$. If the functions $\vec{g}$ are independent and $\vec{y}$ has the same dimension as $\vec{x}$, then the transformation can be inverted. No information is lost. When chaining transformations one has:

$$\vec{y} = \vec{h}(\vec{g}(\vec{x})) \quad \text{and} \quad M_{ij} = \sum_{k=1}^{n} \frac{\partial h_i}{\partial g_k} \frac{\partial g_k}{\partial x_j} \quad \text{or} \quad M = M(\vec{g}) \cdot M(\vec{x})$$

i.e. the final covariance matrix is the same if a transformation is done at once or broken down into several steps.

➜ *linear error propagation . . .*

- is approximate in case of nonlinear functions
- becomes exact in the limit of small variances
- accounts for biases from a non-linear transformation
- must be applied to the full covariance matrix
- is always consistent
  - ➜ no information is lost for invertible transformations
  - ➜ same results when breaking down a transformation into steps

➜ *reminder: special case for a single function of independent variables*

$$\sigma^2(y) = \sum_{i=1}^{n} \left(\frac{\partial y}{\partial x_i} \right)^2 \sigma_i^2$$

alternative approach ➜

➜ *"exact" error propagation*

Linear error propagation does an (approximate) transformation of the covariance matrix. Only the first and second moments are required. If one knows or assumes the exact PDF of the input variables, one can determine the full PDF of the output.

- an assumed PDF must reproduce known mean values and covariances
  - ➜ e.g. a multivariate gaussian
- error propagation becomes transformation of variables
- the approach also works when the derivatives are not known, e.g.
  - ➜ if the transformation is very complex
  - ➜ if the transformation is only defined numerically
  - ➜ if the transformation has discontinuities
- useful cross-check for linear error propagation
  - ➜ check for biases or pathologies
  - ➜ taking mean values and covariance matrix from the transformed variables avoids some of the biases of linear error propagation
  - ➜ however, the transformation in general will not be invertible

(➜ tutorials)

➜ *common problems in data analysis. . .*

- adjust a parametric model to given data
  - ➜ models predicted by theory
  - ➜ models invented to describe the data
- data are given in different formats
  - ➜ measurements as a function of a control parameter
  - ➜ histograms
  - ➜ unbinned data
- different methods have been developed to address this issue
  - ➜ Least Squares method (Gauss,Legendre)
  - ➜ Maximum Likelihood method (Fisher)
  - ➜ bayesian approaches . . .
- generally well understood if the assumed model matches the truth
  - ➜ parameter estimates with well defined uncertainties
- unclear properties if the model is invalid
  - ➜ problem of “goodness of fit”

start with an unconventional view on Least Squares . . .

➜ *use case: straight line fit*

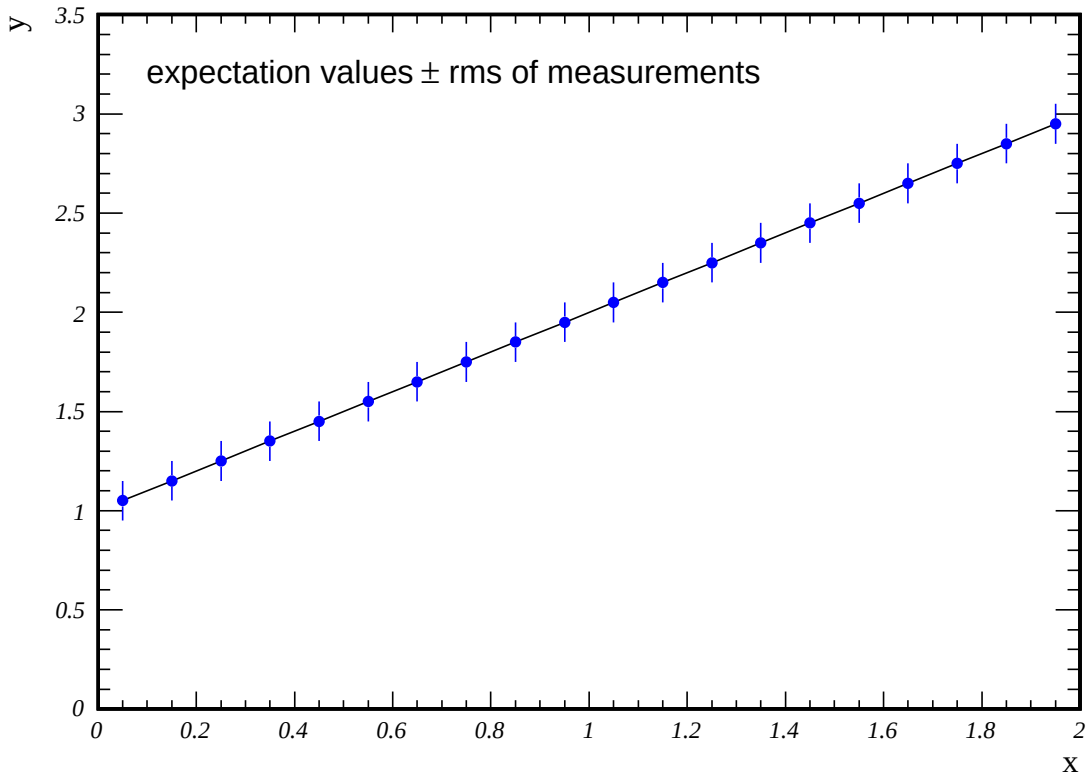
Consider uncorrelated measurements $y_i, i = 1, \ldots, n$ with known variances σ_i^2, recorded for certain values x_i of a control paramater x. The expectation value of the measurements is $\langle y_i \rangle = a_0 + a_1 x_i$, where the parameters a_0 and a_1 are not known.

➜ *wanted: a method to find estimates $\hat{a}_0$ and $\hat{a}_1$ for a_0 and a_1*

❖ discussion

- control parameters x_i are known
- the measurements y_i are unbiased
- variances σ_i^2 are known
- exact shape of PDFs describing the fluctuations of the y_i is irrelevant
  - ➜ any PDF with variance σ_i^2 would do
  - ➜ different measurements can fluctuate with different PDFs

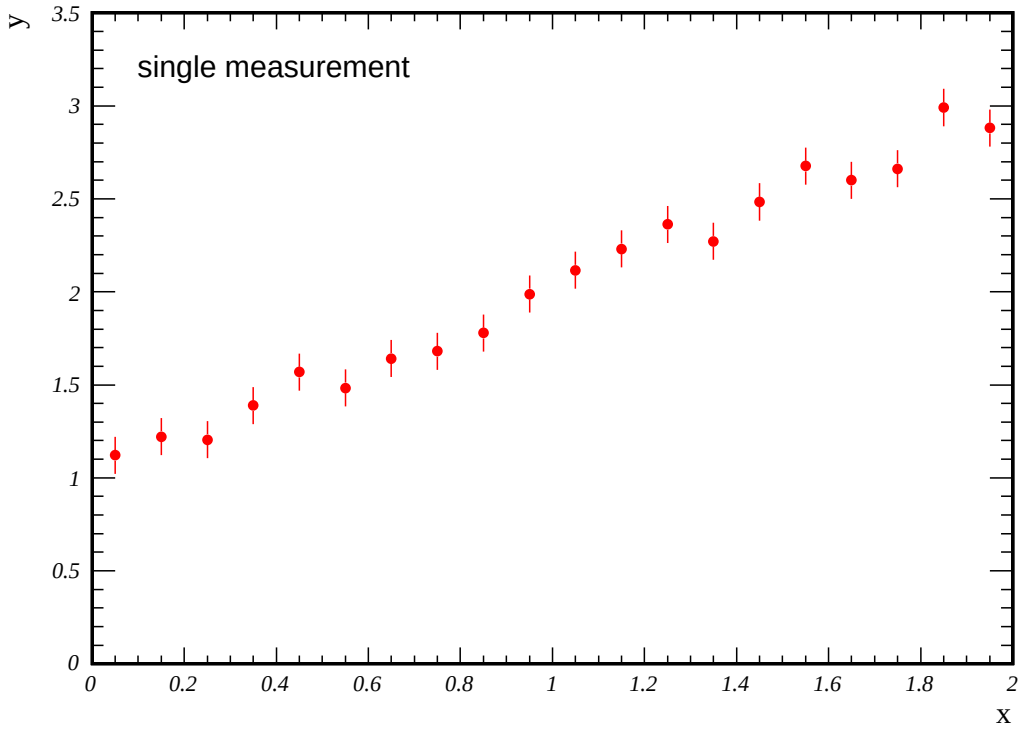
➜ *specific realization of the model discussed before*



- 20 equidistant sampling points in the range $0 < x < 2$
- model parameters $a_0 = a_1 = 1$
- same variance σ^2 for all points with $\sigma = 0.1$

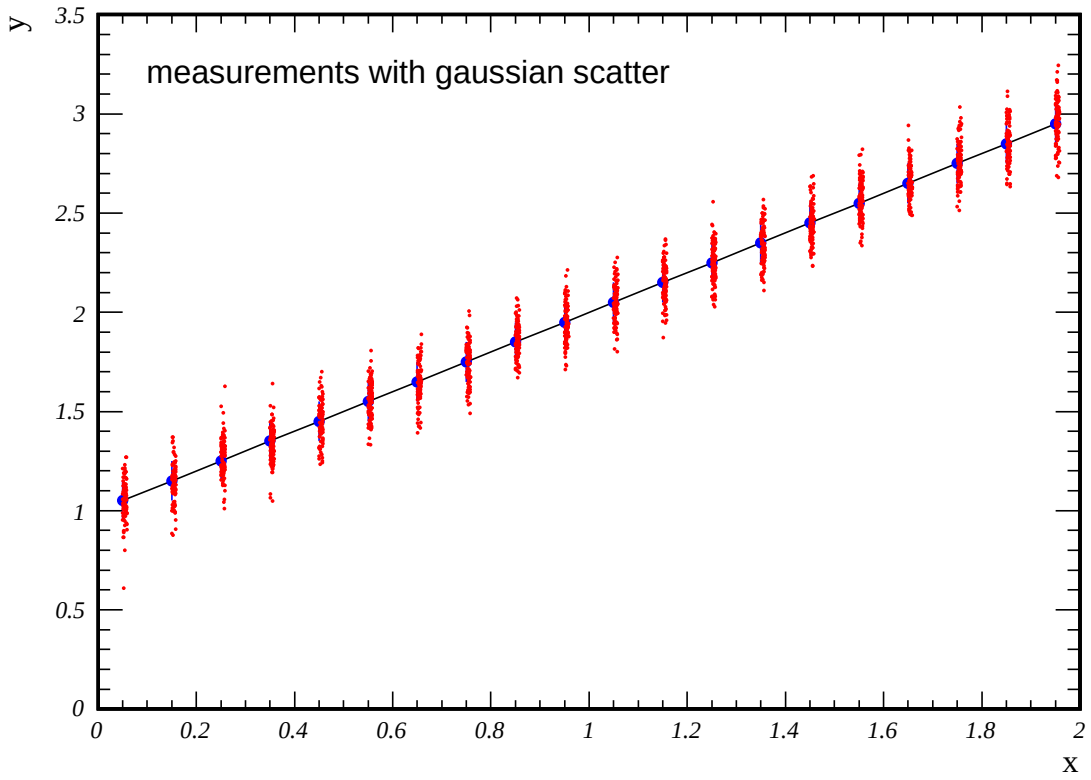
Example of a single measurement

➜ *measurements scatter around the expectation values*



- measure y at 20 equidistant sampling points in the range $0 < x < 2$
- error bars are then known rms value
- error bars indicate the distance to the expectation value
- consider different types of scattering for measurements . . .

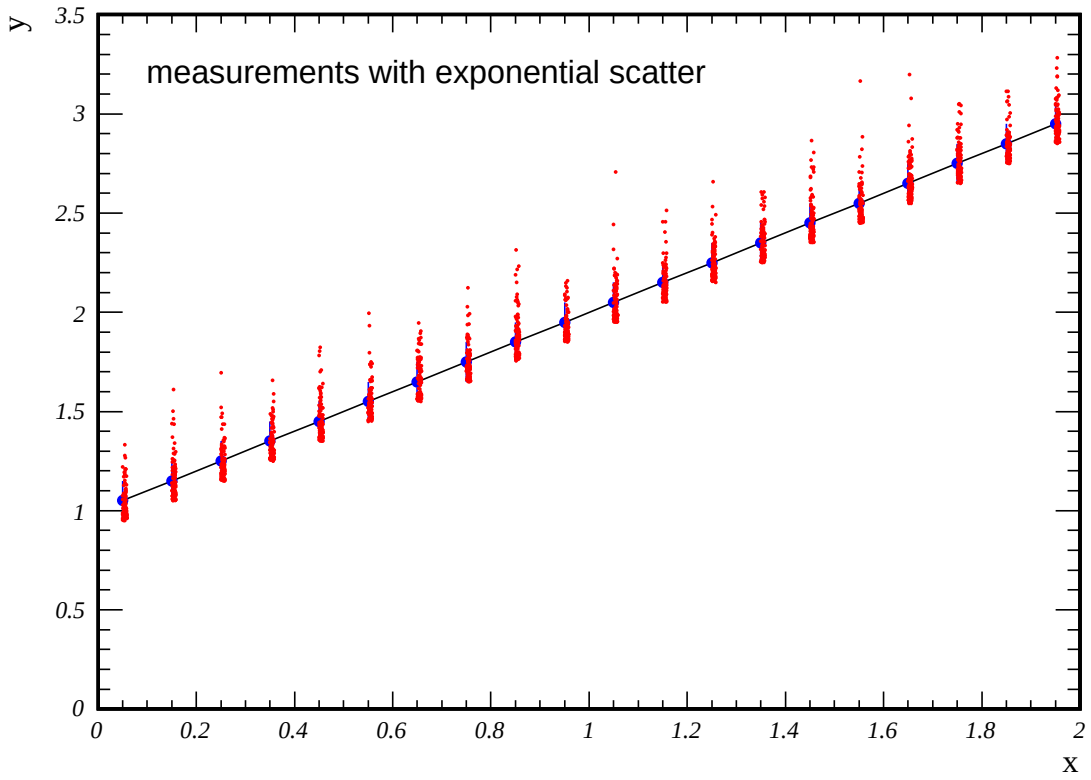
➜ *overlay of 100 measurements*



$$y = a_0 + a_1 \cdot x + r \quad \text{with} \quad \frac{dn}{dr} = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$$

Exponential fluctuations

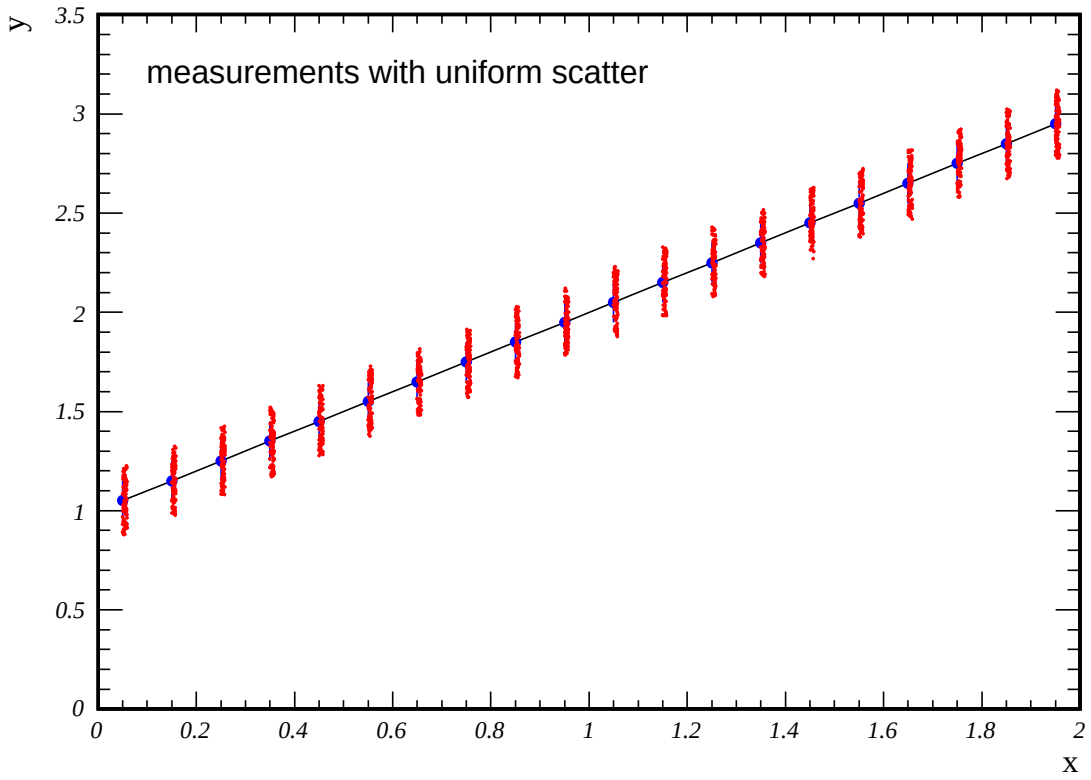
➜ *overlay of 100 measurements*



$$y = a_0 + a_1 \cdot x + (r - \sigma) \quad \text{with} \quad \frac{dn}{dr} = \frac{1}{\sigma} e^{-x/\sigma}$$

Uniform fluctuations

➜ *overlay of 100 measurements*



$$y = a_0 + a_1 \cdot x + r \quad \text{with} \quad \frac{dn}{dr} = \frac{1}{\sigma\sqrt{12}} \quad \text{for} \quad -\frac{\sigma\sqrt{12}}{2} < r < \frac{\sigma\sqrt{12}}{2}$$

➜ *the case of two measurements*

$$\langle y_1 \rangle = a_0 + a_1\, x_1 \quad \text{and} \quad y_1 = \langle y_1 \rangle + r_1$$
$$\langle y_2 \rangle = a_0 + a_1\, x_2 \quad \text{and} \quad y_2 = \langle y_2 \rangle + r_2$$

- system of linear equations relating $\langle y_i \rangle$ and x_i
- measurements y_i have random deviation r_i from $\langle y_i \rangle$
- unbiasedness of y_i implies $\langle r_i \rangle = 0$
- estimate a_0 and a_1 by assuming $r_i = 0$, i.e. make the ansatz:

$$y_1 = \hat{a}_0 + \hat{a}_1\, x_1$$
$$y_2 = \hat{a}_0 + \hat{a}_1\, x_2$$

❖ result:

$$\hat{a}_0 = y_1 - \hat{a}_1 x_1 = \frac{x_2}{x_2 - x_1}\, y_1 - \frac{x_1}{x_2 - x_1}\, y_2$$
$$\hat{a}_1 = \frac{y_2 - y_1}{x_2 - x_1} = -\frac{1}{x_2 - x_1}\, y_1 + \frac{1}{x_2 - x_1}\, y_2$$

➜ *does the estimate make sense?*

- parameter estimates are linear combinations of the measurements
- parameter estimates are random variables
- parameter estimates fluctuate with the measurements
- check the expectation values . . .

$$\langle \hat{a}_0 \rangle = \left\langle \frac{1}{x_2 - x_1}(x_2 y_1 - x_1 y_2) \right\rangle = \frac{1}{x_2 - x_1}(x_2 \langle y_1 \rangle - x_1 \langle y_2 \rangle) = a_0$$

$$\langle \hat{a}_1 \rangle = \left\langle \frac{1}{x_2 - x_1}(-y_1 + y_2) \right\rangle = \frac{1}{x_2 - x_1}(-\langle y_1 \rangle + \langle y_2 \rangle) = a_1$$

❖ conclusion:

➜ the estimates for the unknown parameters are unbiased

➜ the parameter errors can be determined by error propagation

yes, the parameter estimates make sense!

➜ *the case of $n > 2$ measurements*

Take the lessons learnt from the case $n = 2$ and try to estimate the unknown parameters by a linear combination of the measurements.

$$\hat{a}_0 = \sum_{i=1}^{n} p_i \, y_i \quad \text{and} \quad \hat{a}_1 = \sum_{i=1}^{n} q_i \, y_i$$

- this is a convenient ansatz, not derived from any “first principles”
- it is not the only possible generalization of the case $n = 2$
- nor will it give the best possible estimates for a_0 and a_1
- but it is simple and robust, requiring only minimal input
- and turns out to be surprisingly powerful . . .

➜ determine parameters p_i and q_i . . .

Optimizing the parameter estimates

➜ *exploit the freedom of the linear ansatz to. . .*

- make sure that the estimates are unbiased
- and that the estimates are as accurate as possible

❖ condition for unbiased estimates:

$$\langle \hat{a}_0 \rangle = \sum_{i=1}^{n} p_i \langle y_i \rangle = \sum_{i=1}^{n} p_i (a_0 + a_1 x_i) = a_0 \sum_{i=1}^{n} p_i + a_1 \sum_{i=1}^{n} p_i x_i \stackrel{!}{=} a_0$$

$$\langle \hat{a}_1 \rangle = \sum_{i=1}^{n} q_i \langle y_i \rangle = \sum_{i=1}^{n} q_i (a_0 + a_1 x_i) = a_0 \sum_{i=1}^{n} q_i + a_1 \sum_{i=1}^{n} q_i x_i \stackrel{!}{=} a_1$$

one obtains 4 conditions:

$$\sum_{i=1}^{n} p_i = 1 \qquad \sum_{i=1}^{n} q_i = 0 \qquad \sum_{i=1}^{n} p_i x_i = 0 \qquad \sum_{i=1}^{n} q_i x_i = 1$$

Discussion

- only 4 constraints for $2n$ parameters
- easy to satisfy both for p_i and q_i
  - ➜ start from an set of numbers e.g. for p_i
  - ➜ subtract a constant such that the “0-constraint” is satisfied
  - ➜ scale the numbers such that the “1-constraint” is satisfied
- additional criterion needed to fix the coefficients
- require minimal variance for the parameter estimates
  - ➜ constrained minimization problem

❖ variance of parameter estimates from error propagation:

$$\sigma^2(\hat{a}_0) = \sum_{i=1}^{n} \left(\frac{\partial \hat{a}_0}{\partial y_i} \right)^2 \sigma_i^2 = \sum_{i=1}^{n} p_i^2 \, \sigma_i^2 \quad \text{and} \quad \sigma^2(\hat{a}_1) = \sum_{i=1}^{n} q_i^2 \, \sigma_i^2$$

➜ constrained minimization

Fixing the coefficients p_i

➜ *minimization using Lagrange multipliers for the constraints*

$$\sum_{i=1}^{n} p_i^2\,\sigma_i^2 + 2\alpha_0\left(1-\sum_{i=1}^{n} p_i\right) + 2\beta_0\left(-\sum_{i=1}^{n} p_i\,x_i\right) \stackrel{!}{=} \min$$

requiring zero derivatives with respect to p_i then yields:

$$2p_i\,\sigma_i^2 - 2\alpha_0 - 2\beta_0\,x_i = 0 \quad \rightarrow \quad p_i = \frac{1}{\sigma_i^2}(\alpha_0 + \beta_0 x_i)$$

α_0 and β_0 follow from the constraint to have unbiased estimates:

$$\sum_{i=1}^{n} p_i = \alpha_0 \sum_{i=1}^{n} \frac{1}{\sigma_i^2} + \beta_0 \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} = \alpha_0\,S_1 + \beta_0\,S_x \;= 1$$

$$\sum_{i=1}^{n} p_i\,x_i = \alpha_0 \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} + \beta_0 \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} = \alpha_0\,S_x + \beta_0\,S_{xx} = 0$$

Fixing the coefficients q_i

➜ *minimization using Lagrange multipliers for the constraints*

$$\sum_{i=1}^{n} q_i^2\,\sigma_i^2 + 2\alpha_1\left(-\sum_{i=1}^{n} q_i\right) + 2\beta_1\left(1-\sum_{i=1}^{n} q_i\,x_i\right) \stackrel{!}{=} \min$$

requiring zero derivatives with respect to q_i then yields:

$$2\,q_i\,\sigma_i^2 - 2\alpha_1 - 2\beta_1\,x_i = 0 \quad \Rightarrow \quad q_i = \frac{1}{\sigma_i^2}(\alpha_1 + \beta_1\,x_i)$$

α_1 and β_1 follow from the constraint to have unbiased estimates:

$$\sum_{i=1}^{n} q_i = \alpha_1 \sum_{i=1}^{n} \frac{1}{\sigma_i^2} + \beta_1 \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} = \alpha_1\,S_1 + \beta_1\,S_x \;= 0$$

$$\sum_{i=1}^{n} q_i\,x_i = \alpha_1 \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} + \beta_1 \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} = \alpha_1\,S_x + \beta_1\,S_{xx} = 1$$

Solving the linear equations for the Lagrange parameters $\alpha_{\{0,1\}}$ and $\beta_{\{0,1\}}$

$$\begin{pmatrix} S_1 & S_x \\ S_x & S_{xx} \end{pmatrix} \cdot \begin{pmatrix} \alpha_0 & \alpha_1 \\ \beta_0 & \beta_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and substituting the results into p_i, q_i, with $D = S_1 S_{xx} - S_x^2$, yields

$$p_i = \frac{1}{\sigma_i^2}(\alpha_0 + \beta_0 x_i) = \frac{1}{D}\left(S_{xx}\frac{1}{\sigma_i^2} - S_x\frac{x_i}{\sigma_i^2}\right)$$

$$q_i = \frac{1}{\sigma_i^2}(\alpha_1 + \beta_1 x_i) = \frac{1}{D}\left(-S_x\frac{1}{\sigma_i^2} + S_1\frac{x_i}{\sigma_i^2}\right)$$

and thus

$$\hat{a}_0 = \frac{1}{D}(S_{xx}S_y - S_x S_{xy}) \quad \text{and} \quad \hat{a}_1 = \frac{1}{D}(S_1 S_{xy} - S_x S_y)$$

where

$$S_{\{1,x,xx,y,xy\}} = \sum_{i=1}^{n} \frac{\{1, x_i, x_i\, x_i, y_i, x_i\, y_i\}}{\sigma_i^2} \; .$$

➜ *linear error propagation*

$$C_{kl}(\hat{a}) = \sum_{i=1}^{n} \frac{\partial \hat{a}_k}{\partial y_i} \frac{\partial \hat{a}_l}{\partial y_i} \sigma_i^2$$

yields

$$C_{00}(\hat{a}) = \sum_{i=1}^{n} p_i^2 \sigma_i^2 = \frac{S_1}{D^2}(S_{xx} S_1 - S_x^2) = \frac{S_1}{D}$$

$$C_{11}(\hat{a}) = \sum_{i=1}^{n} q_i^2 \sigma_i^2 = \frac{S_{xx}}{D^2}(S_{xx} S_1 - S_x^2) = \frac{S_{xx}}{D}$$

$$C_{01}(\hat{a}) = \sum_{i=1}^{n} p_i q_i \sigma_i^2 = \frac{-S_x}{D^2}(S_{xx} S_1 - S_x^2) = \frac{-S_x}{D}$$

. . . the well known textbook formulae for straight line fits.

➜ *re-write the solution derived before. . .*

$$\hat{a}_0 = \frac{1}{D}(S_{xx}S_y - S_x S_{xy}) \quad \text{and} \quad \hat{a}_1 = \frac{1}{D}(S_1 S_{xy} - S_x S_y)$$

to make the structure more evident:

$$\begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \end{pmatrix} = \frac{1}{D}\begin{pmatrix} S_{xx} & -S_x \\ -S_x & S_1 \end{pmatrix} \cdot \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} S_1 & S_x \\ S_x & S_{xx} \end{pmatrix} \cdot \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix}$$

or

$$S_1\ \hat{a}_0 + S_x\ \hat{a}_1 - S_y\ = 0$$
$$S_x\ \hat{a}_0 + S_{xx}\ \hat{a}_1 - S_{xy} = 0$$

i.e. two equations which define the best fit parameters as the zero of a two-dimensional function. Now exploit the fact that it's always possible to interpret the zero of a function as a stationary point (e.g. minimum) of its primitive.

p.t.o. ➜

➜ *introducing $F(a_0, a_1)$ such that*

$$\left.\frac{\partial F}{\partial a_0}\right|_{\{a_0,a_1\}=\{\hat{a}_0,\hat{a}_1\}} = 0 \quad \text{and} \quad \left.\frac{\partial F}{\partial a_1}\right|_{\{a_0,a_1\}=\{\hat{a}_0,\hat{a}_1\}} = 0$$

it follows (from dimensional considerations)

$$\frac{\partial F}{\partial a_0} = S_1\, a_0 + S_x\, a_1 - S_y \quad \text{and} \quad \frac{\partial F}{\partial a_1} = S_x\, a_0 + S_{xx}\, a_1 - S_{xy}\ .$$

Integration of the first equation yields

$$F = \frac{1}{2} S_1\, a_0^2 + S_x\, a_0\, a_1 - a_0\, S_y + g(a_1)$$

where $g(a_1)$ does not depend on a_0. Taking the derivative with respect to a_1 and comparing with the known derivative determines $g'(a_1)$:

$$\frac{\partial F}{\partial a_1} = S_x\, a_0 + g'(a_1) = S_x\, a_0 + S_{xx}\, a_1 - S_{xy}$$

p.t.o. ➜

Constructing the cost function (cnt'd)

It follows

$$g'(a_1) = S_{xx}\,a_1 - S_{xy} \quad \text{and thus} \quad g(a_1) = \frac{1}{2}S_{xx}\,a_1^2 - S_{xy}\,a_1 + \frac{C}{2}$$

with an arbitrary constant C. Setting $C = \sum y_i^2/\sigma_i^2$ then yields

$$\begin{aligned} 2F &= S_1\,a_0^2 + S_{xx}\,a_1^2 + 2S_x\,a_0\,a_1 - 2S_y\,a_0 - 2S_{xy}\,a_1 + C \\ &= \sum_{i=1}^{n} \frac{1}{\sigma_i^2}(a_0^2 + a_1^2 x_i^2 + 2a_0\,a_1\,x_i - 2a_0\,y_i + 2a_1\,x_i\,y_i + y_i^2) \\ &= \sum_{i=1}^{n} \frac{(y_i - a_0 - a_1\,x_i)^2}{\sigma_i^2}\,, \end{aligned}$$

an setting $2F = \chi^2$, the cost-function becomes

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - f_i(a_0, a_1))^2}{\sigma_i^2} \quad \text{with} \quad f_i(a_0, a_1) = a_0 + a_1\,x_i\,.$$

➜ *properties of Linear Least Squares*

- formulation via the cost function . . .
  - ➜ explains the name “least squares”
  - ➜ was derived for linear models
  - ➜ easily generalizes to nonlinear problems
  - ➜ also applicale for different weight assignments
- variance of the measurements (weights) must be known/fixed
  and must be uncorrelated to the measurements
- for linear models
  - ➜ parameter estimates are linear combinations of the measurements
  - ➜ unbiased estimates of the true parameters
  - ➜ linear estimates with minimal variance
  - ➜ variance depends only on the variance of the measurement,
    independent of the shape of the PDF of the fluctuations

Numerical example

➜ *straight line fit to 20 equidistant points*

- measurements at discrete values of the control parameter x
- straight line model

$$y = a_0 + a_1\, x$$

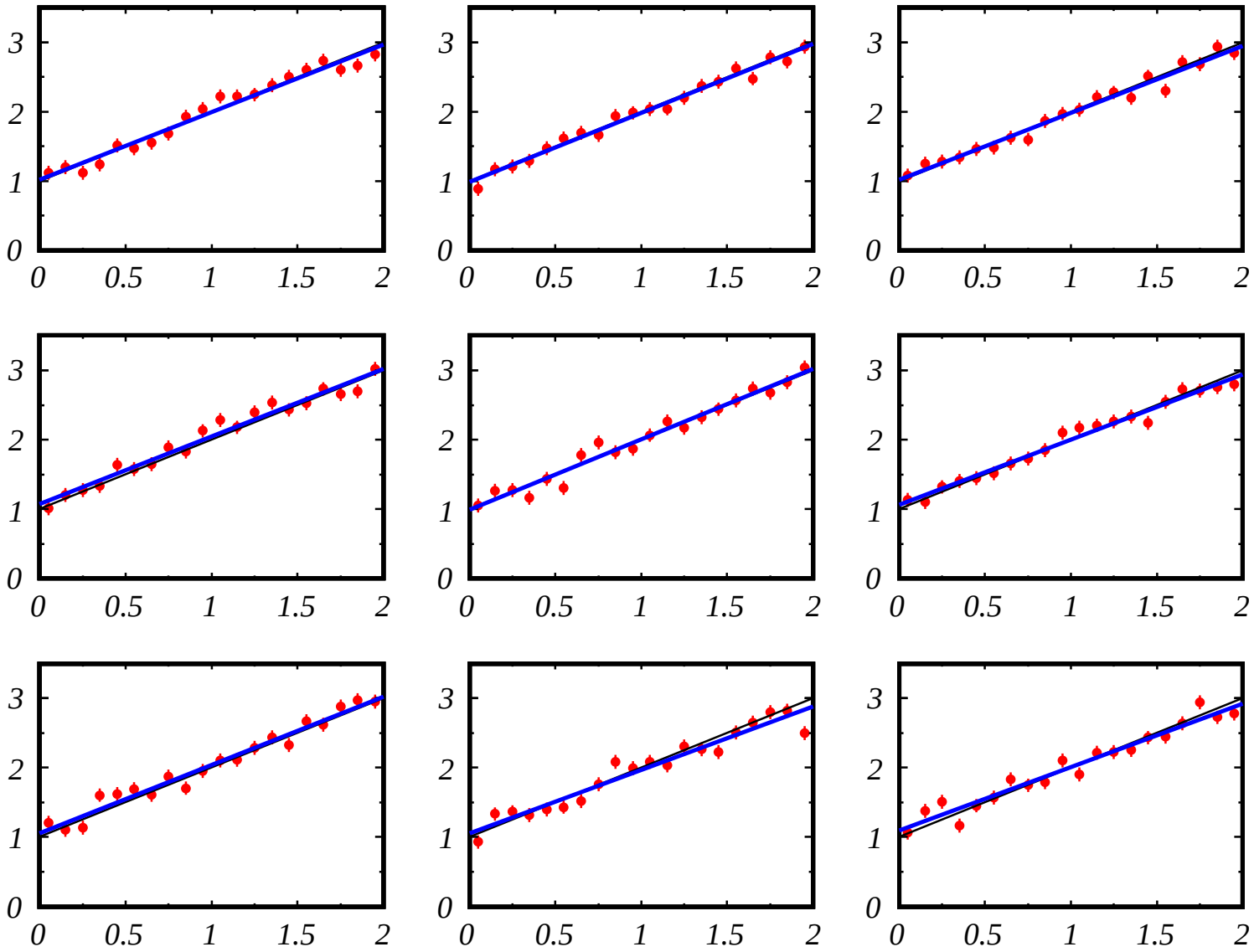
- parameters of the model: $a_0 = a_1 = 1$ and $\sigma = 0.1$
- same variance σ^2 for all points
- uncertainties of $\hat{a}_0$ and $\hat{a}_1$ are fixed:

$$\sigma(\hat{a}_0) \approx 0.038778 \qquad \sigma(\hat{a}_1) \approx 0.044763 \qquad \rho \approx -0.866296$$

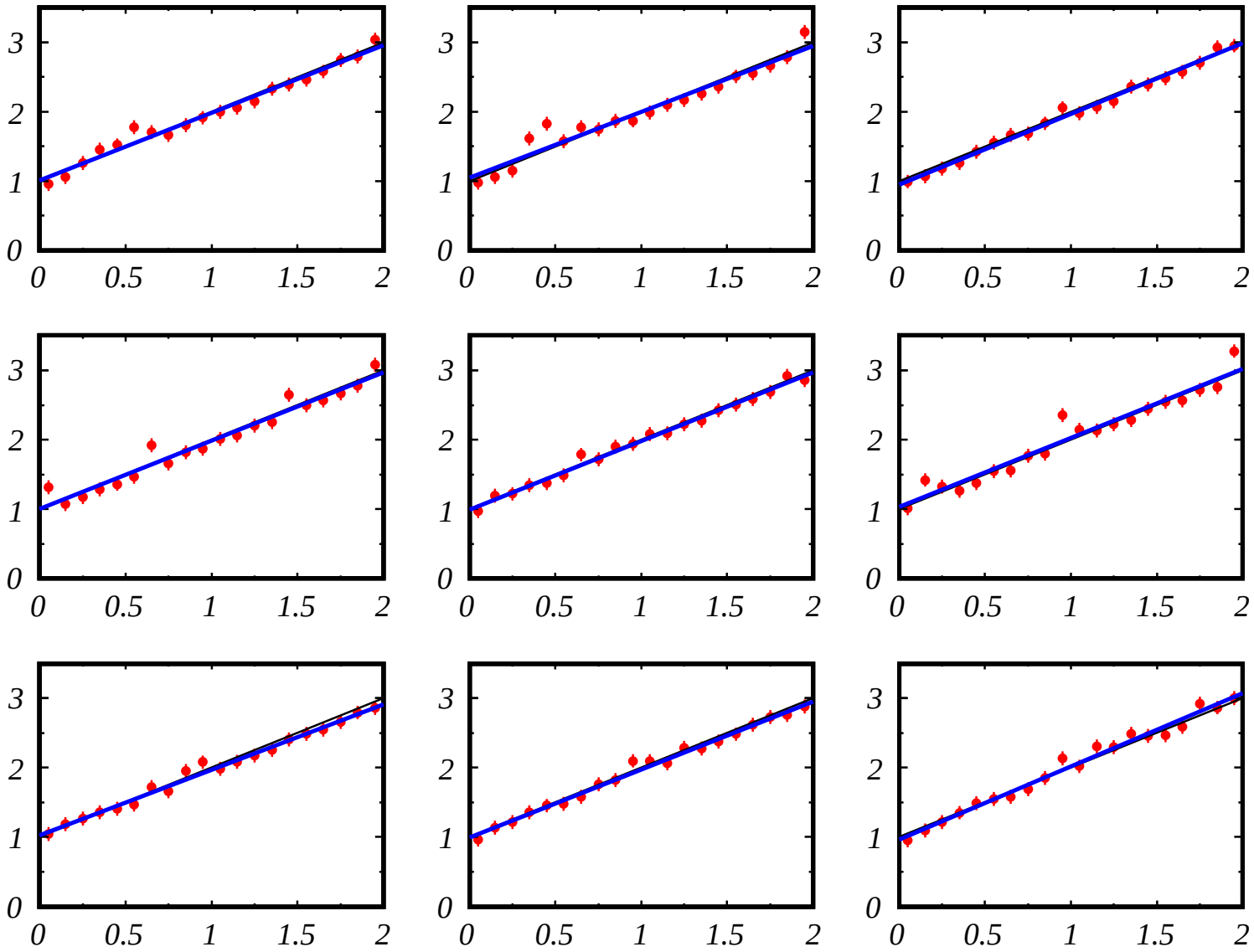
- test different types of fluctuations with the same variance
  - ➜ gaussian fluctuations
  - ➜ exponential fluctuations
  - ➜ uniform fluctuations

➜ results

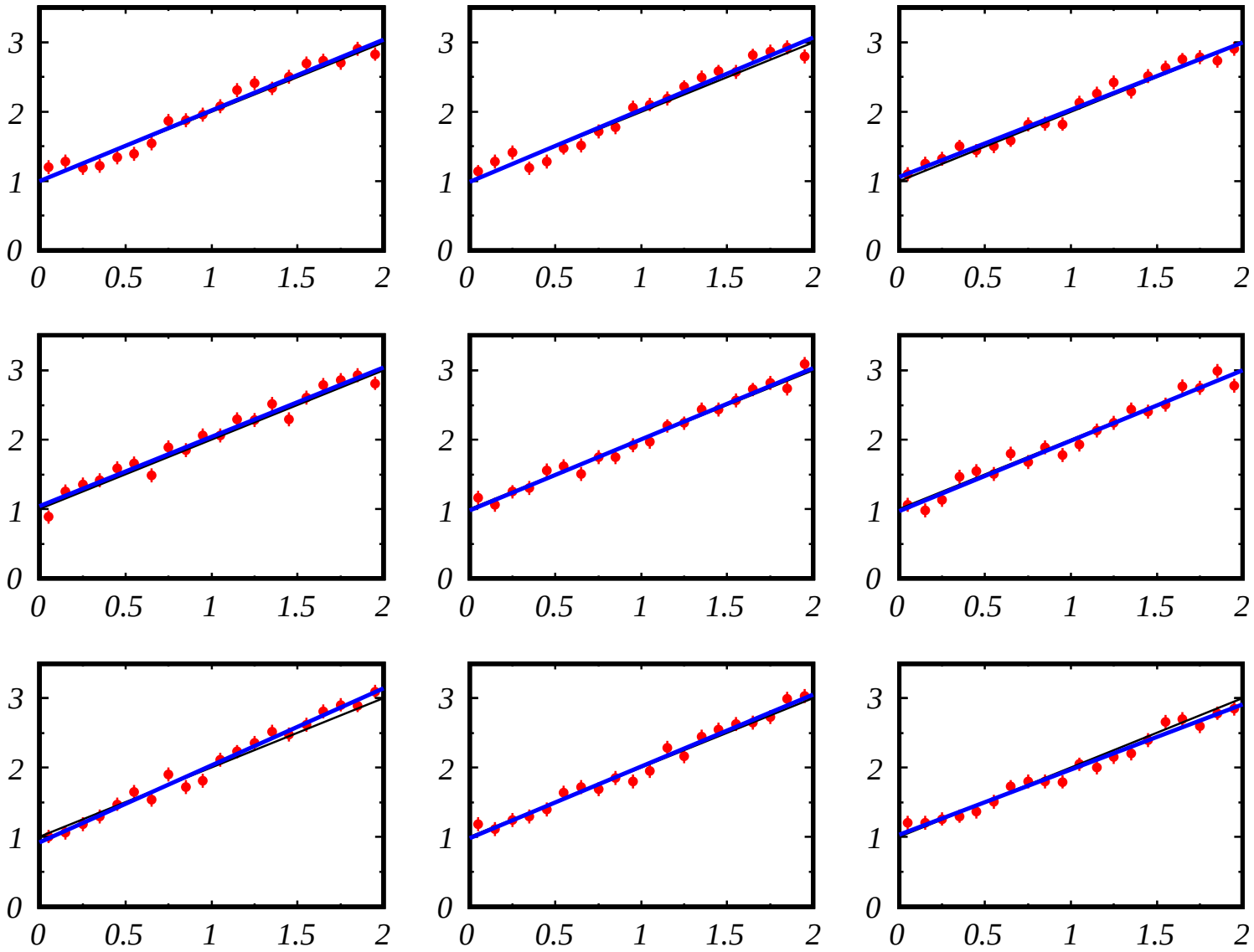
Fits to gaussian fluctuations



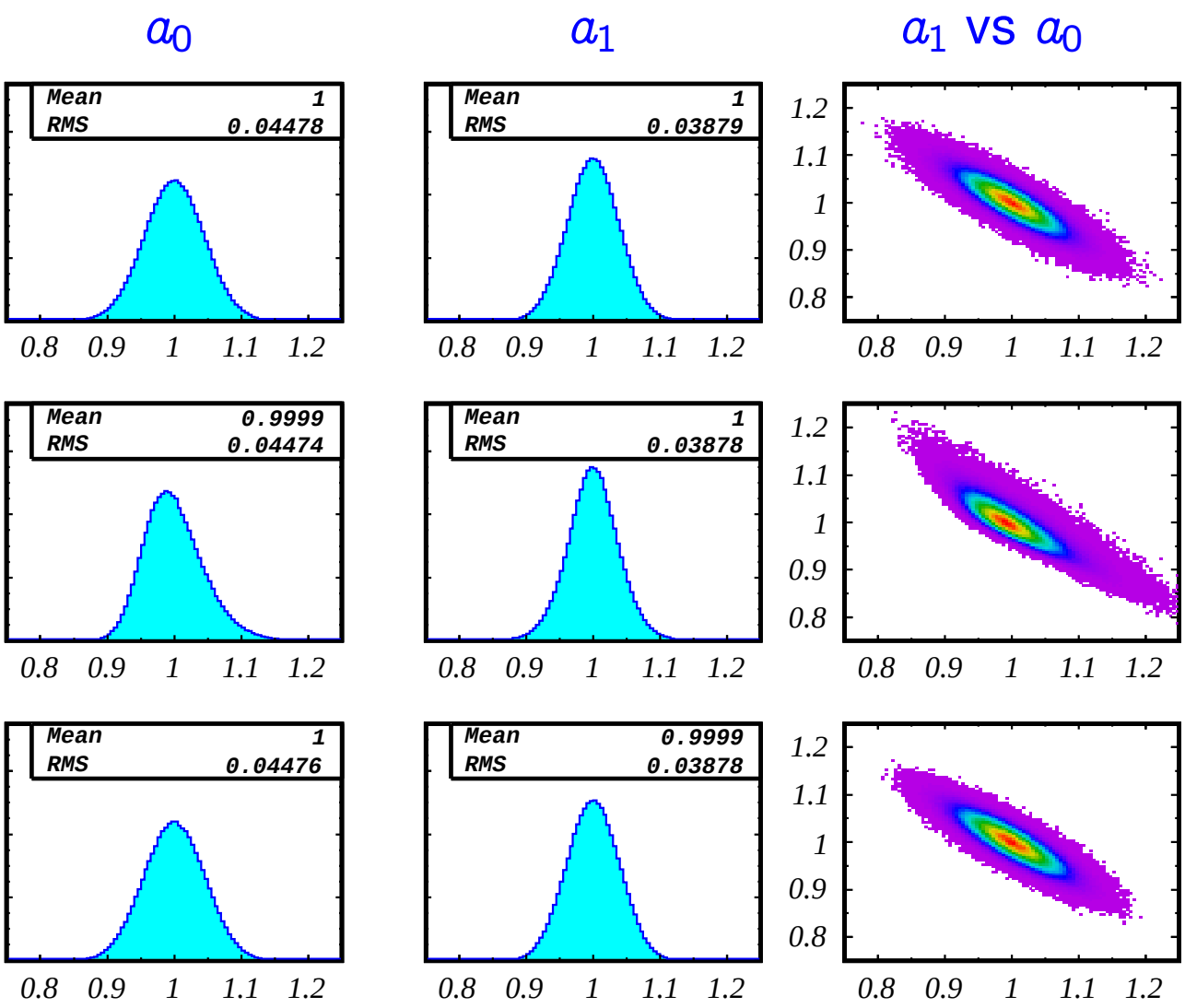
Fits to exponential fluctuations



Fits to uniform fluctuations



Distribution of parameter estimates



gaussian fluctuations

exponential fluctuations

uniform fluctuations

➜ *recap: the Least Squares Method*

Given: measurements y_i with known variances σ_i^2, a parametric model $f_i(a)$, and positive weights $w_i > 0$. Wanted: parameter estimates $\hat{a}$.

$$\text{Ansatz:} \qquad S^2(a) = \sum_i w_i (y_i - f_i(a))^2 \stackrel{!}{=} \min$$

❖ discussion

- the best fit $\hat{a}$ makes the model get "as close as possible" to the data
- the weights allow to (de)emphasize selected points
- a priori arbitrary weights are allowed
- for independent measurements the optimal weights are $1/\sigma^2$
  - ➜ $S^2 = \chi^2$ in this case
  - ➜ well measured points are emphasized
  - ➜ the best fit a has minimal variance

numerical study ➜

➜ *determination of the lifetime of an unstable particle*

- assumed lifetime distribution

$$\frac{dn}{dt} = \frac{1}{\mu} e^{-t/\mu} \quad \text{with} \quad \mu = 1\,\text{ns}$$

- run 10000 test experiments
  - ➜ histogram representation of the measurement
  - ➜ 100 bins for $0 < t < 10\,\text{ns}$
  - ➜ study for fixed number of decays $N = 10000, 1000, 100, 10$
- optimal parameter estimate:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} t_i$$

- parametric model for bin contents n_i in Least Squares fit

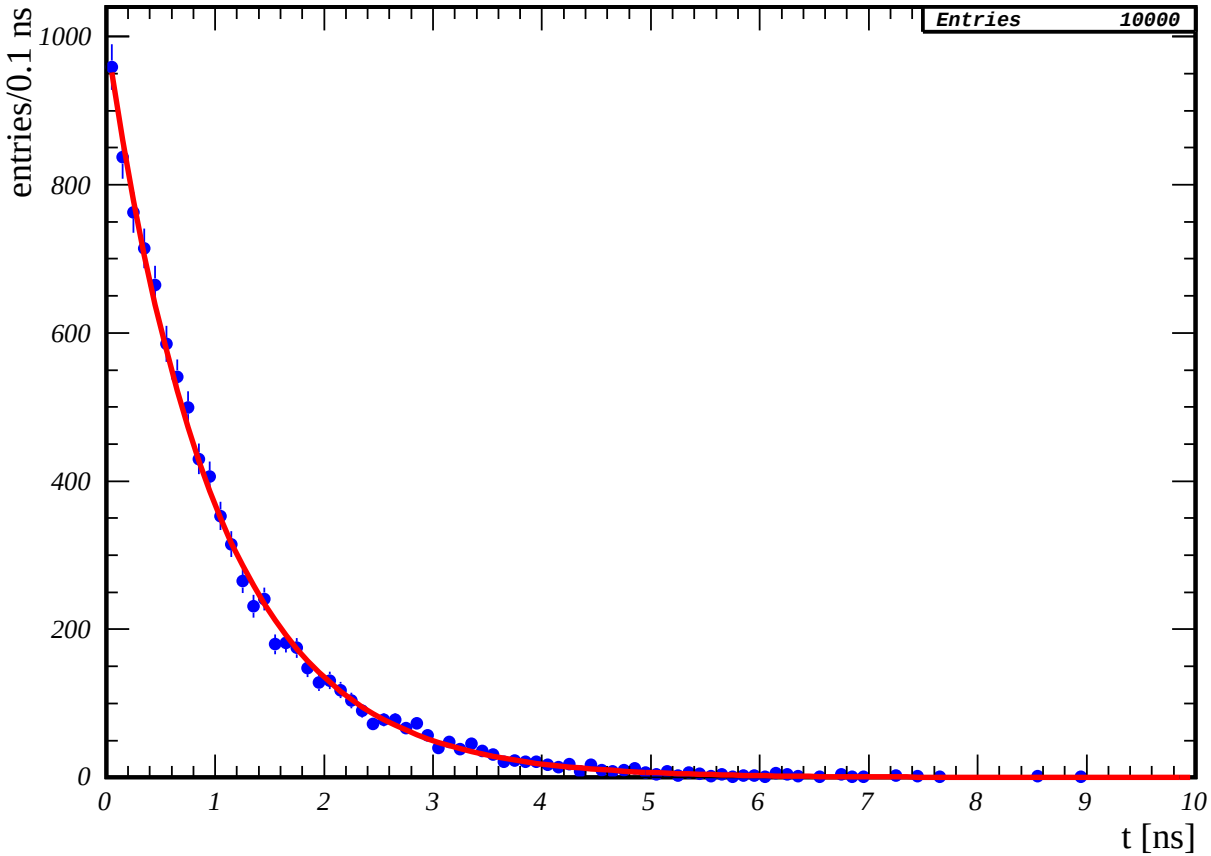
$$f_i(\mu) = N \int_{\text{bin } i} dt \; \frac{dn}{dt}$$

➜ *test different weight-assignments*

- $w_i = 1$ for all bins
  - ➜ unsophisticated but hopefully robust unweighted fit
- $w_i = 1$ for all bins with non-zero entries
  - ➜ pretend that empty bins don't have informations
- $w_i = 1/n_i$ for all bins with non-zero entries
  - ➜ use empirical variances (ROOT/PAW/HBOOK default)
- $w_i = 1/f_i$ for all bins
  - ➜ naive way to use the theoretical variances
- iterative fit with $w(0) = 1$ and $w_i(m) = 1/f_i(\hat{\mu}_{m-1})$ for all bins
  - ➜ proper way to use the theoretical variances
  - ➜ implements that variances must be fixed ("known") in minimization
- for comparison: simple arithmetic mean of all entries

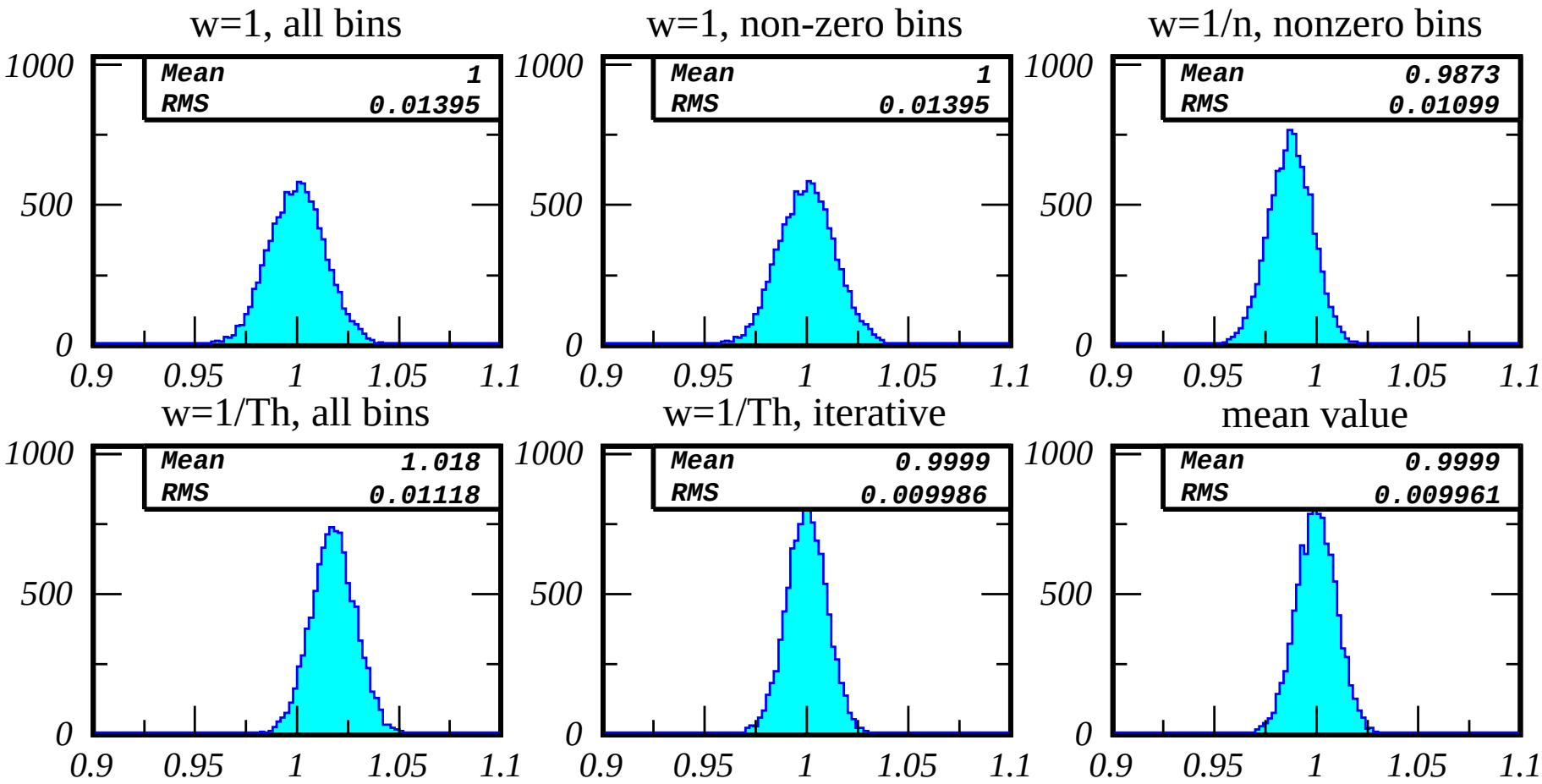
Results for $N = 10000$ events

➜ *example for a measured lifetime distribution*



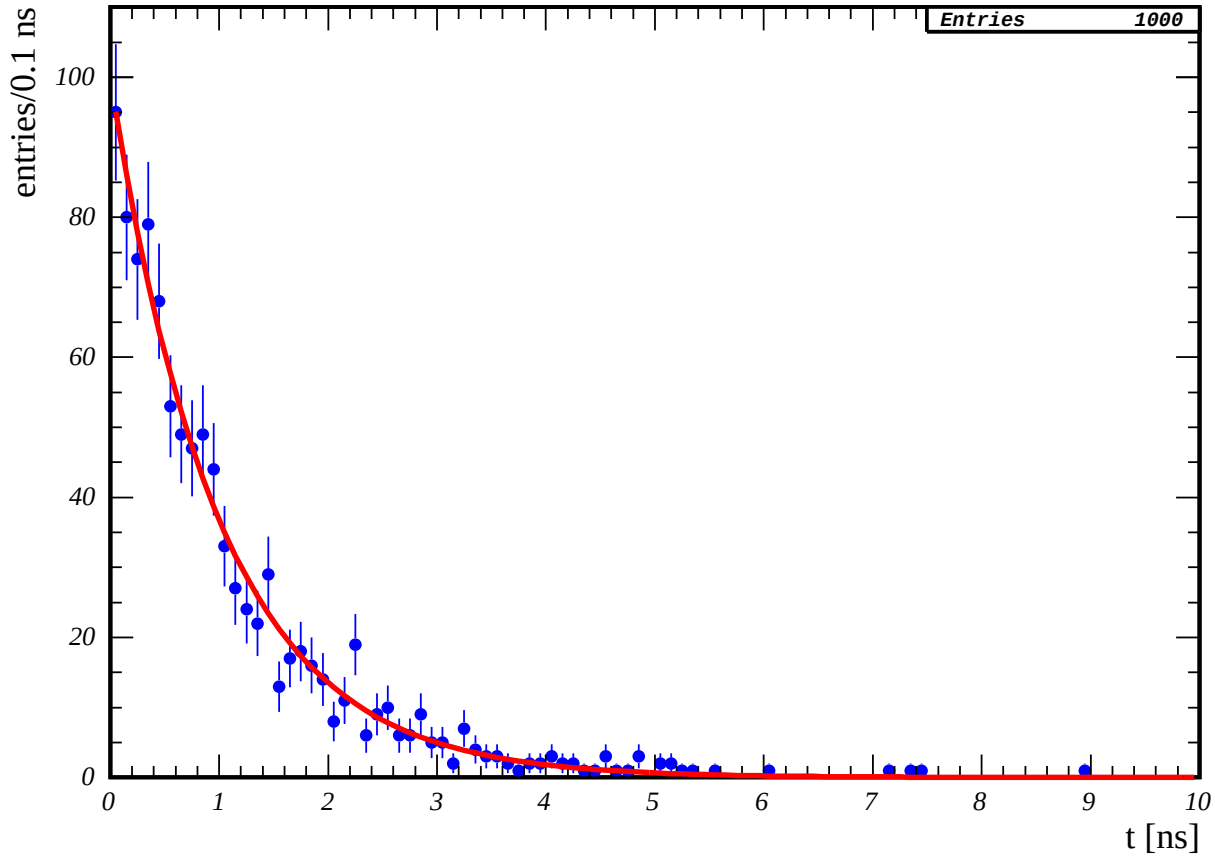
Fitted parameters for $N = 10000$ events

➜ *distributions of best fit parameters from 10000 measurements*



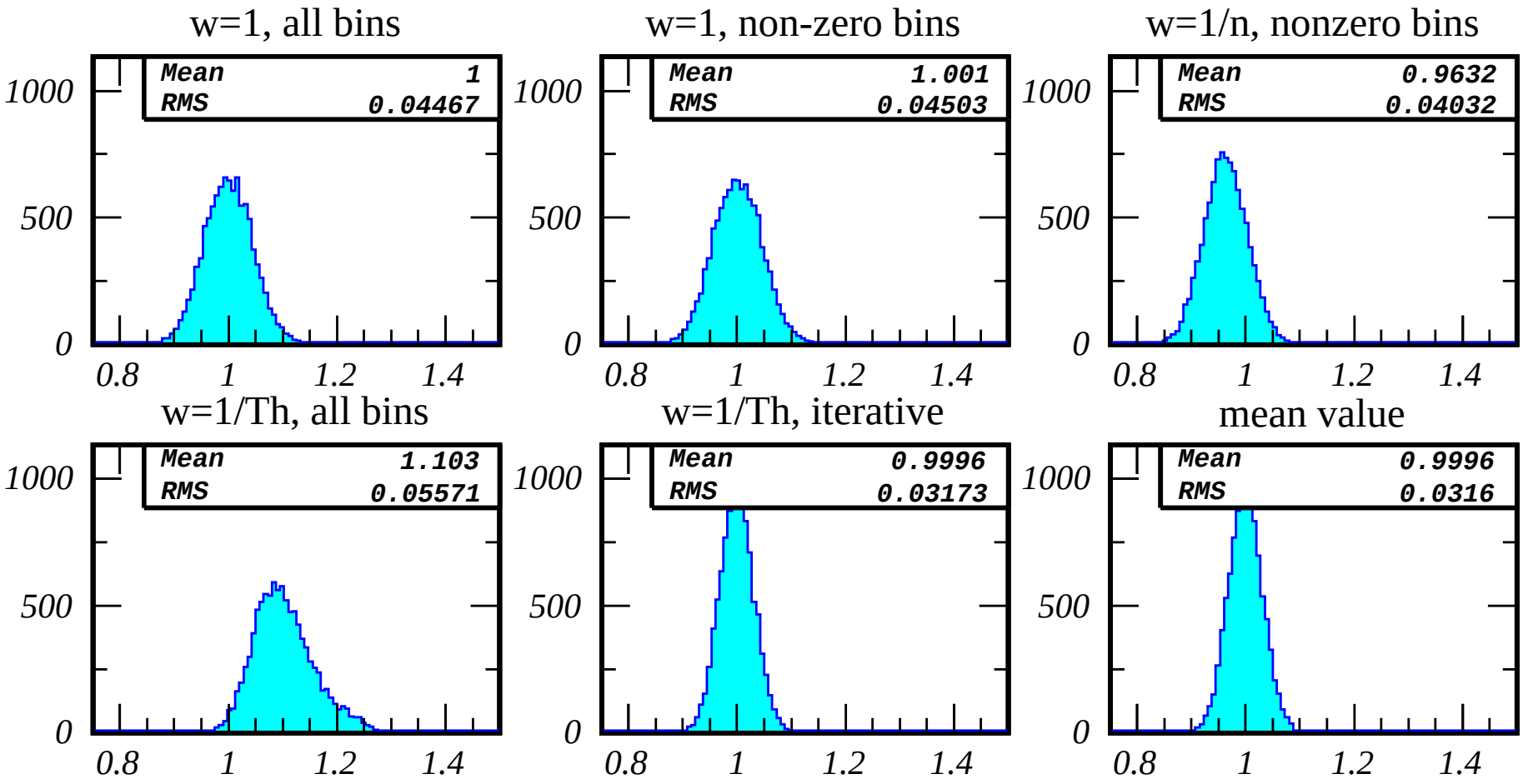
Results for $N = 1000$ events

➜ *example for a measured lifetime distribution*



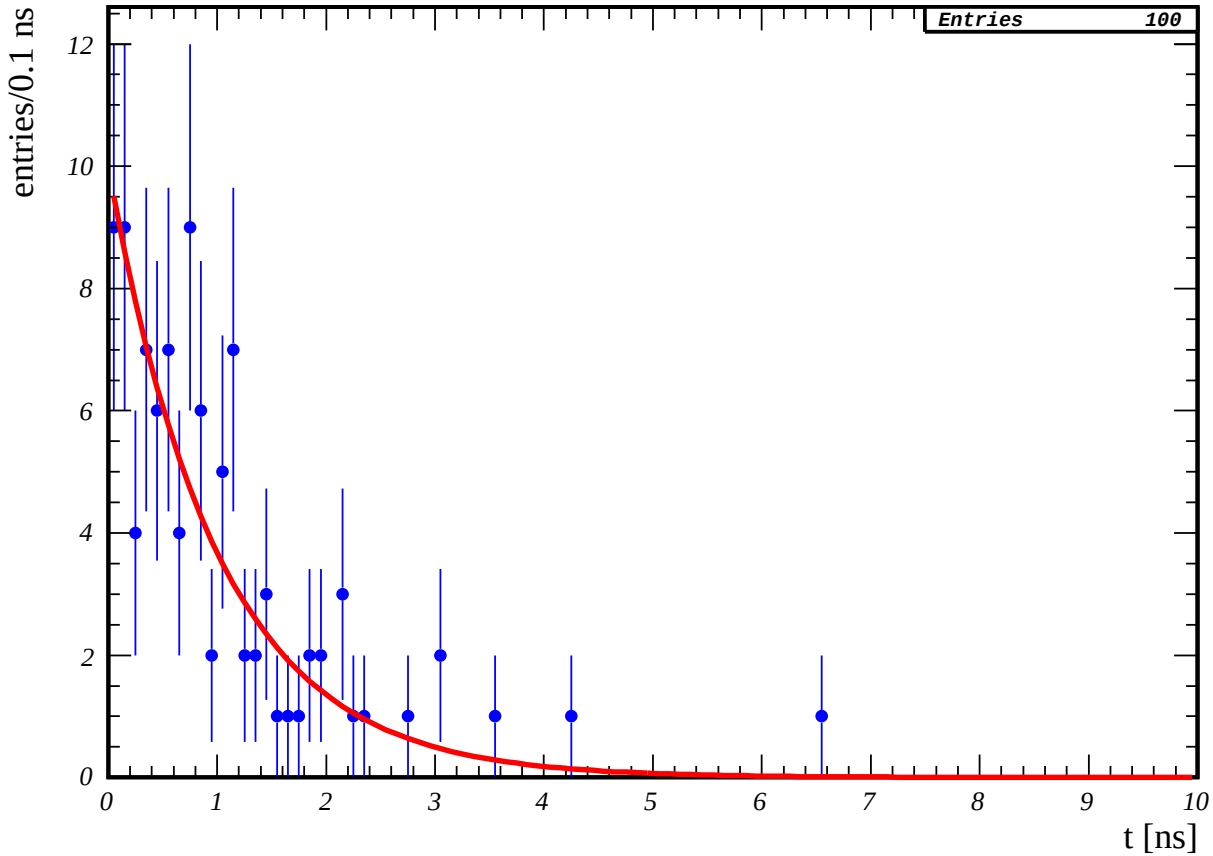
Fitted parameters for $N = 1000$ events

➜ *distributions of best fit parameters from 10000 measurements*



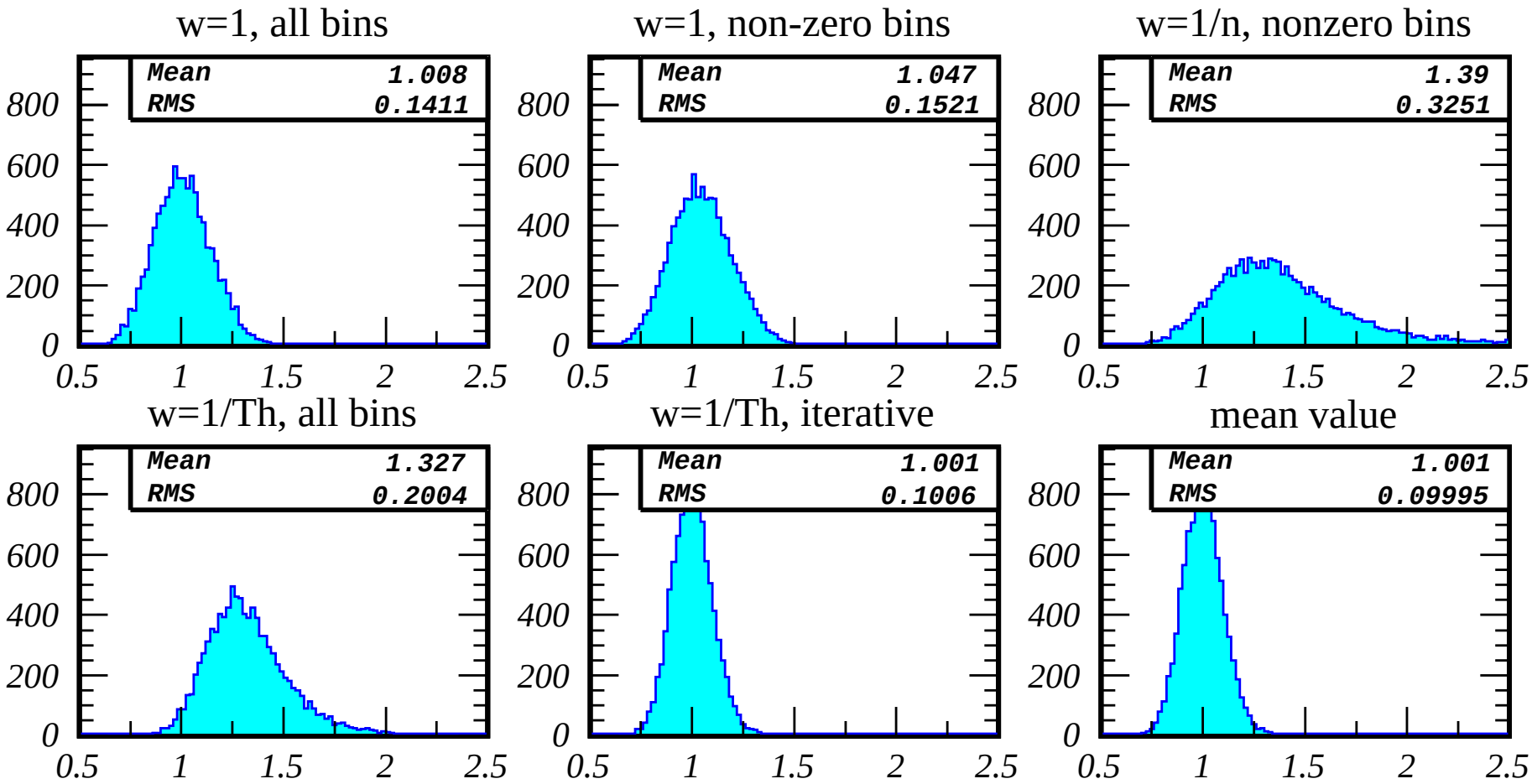
Results for $N = 100$ events

➜ *example for a measured lifetime distribution*



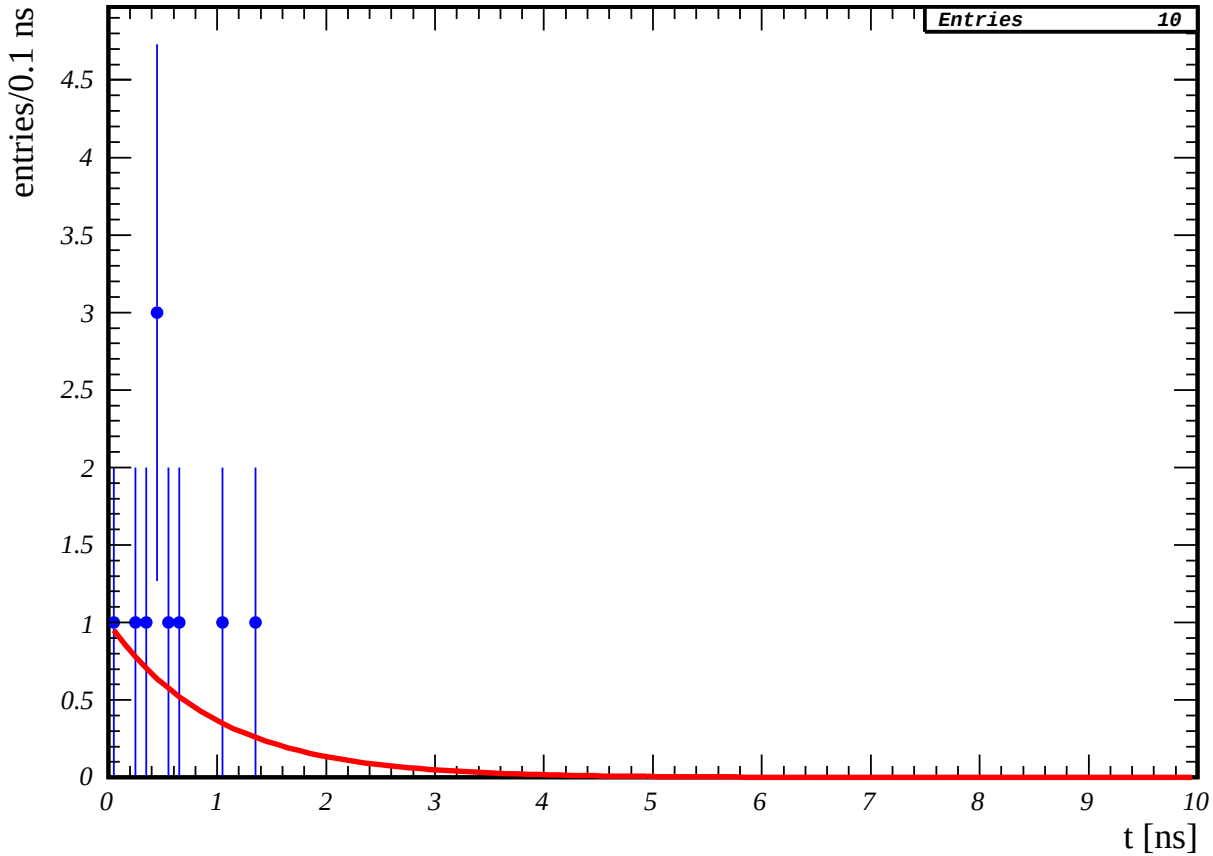
Fitted parameters for $N = 100$ events

➜ *distributions of best fit parameters from 10000 measurements*



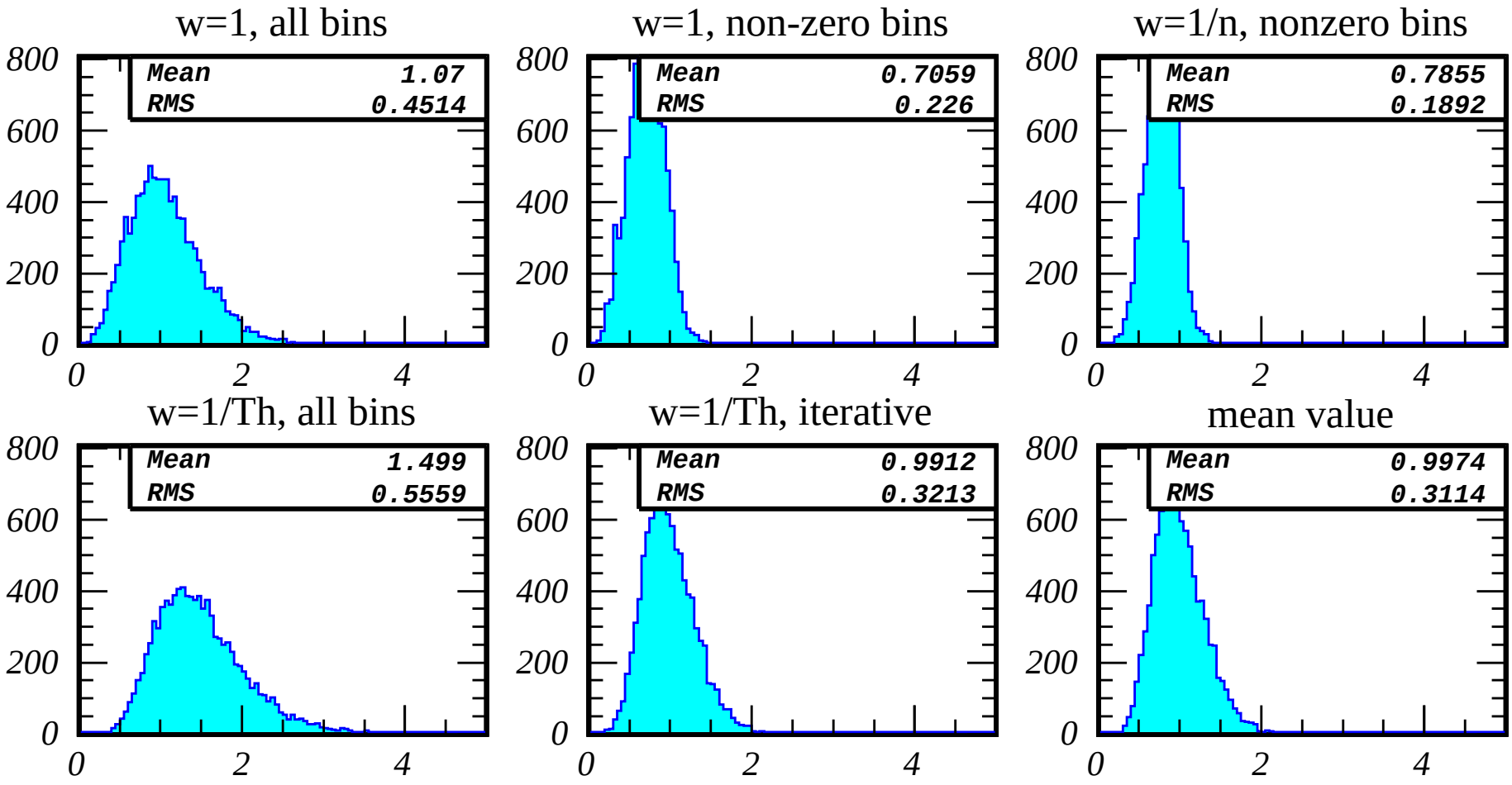
Results for $N = 10$ events

➜ *example for a measured lifetime distribution*



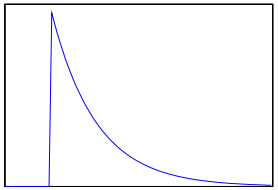
Fitted parameters for $N = 10$ events

➜ *distributions of best fit parameters from 10000 measurements*

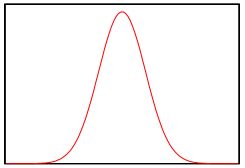


➜ *merits of different weight-assignments*

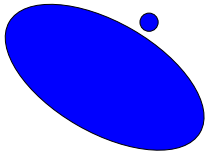
- $w_i = 1$ for all bins
  - ➜ OK, generally unbiased, but not with optimal precision
- $w_i = 1$ for non-zero bins
  - ➜ needless loss of information and bias at low statistics
- $w_i = 1/n_i$ for all bins with non-zero entries
  - ➜ biased – especially at low statistics
- $w_i = 1/f_i$ for all bins
  - ➜ violates the Least Squares ansatz
  - ➜ heavily biased – do not even think of doing this!
- iterative fit with $w(0) = 1$ and $w_i(m) = 1/f_i(\hat{\mu}_{m-1})$ for all bins
  - ➜ close to optimum (maximum likelihood fit)
  - ➜ also perfectly applicable at low statistics



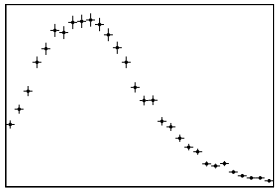
Measurement



Physics



Data



Analysis

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
...
```

➜ *starting point*

A physical distribution $b(y)$ is measured by a detector. The detector distorts $b(y)$. It is described by a response function, $g(x, y)$, which for every true y determines the distribution of the measured value x. Alternative names for $g(x, y)$ are:

- point-spread function
- green's function
- kernel

The relation between $b(y)$ and the observed distribution $a(x)$ is

$$a(x) = \int_{y_{\rm min}}^{y_{\rm max}} dy \; g(x, y) b(y)$$

a typical response function:

$$g(x, y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-y)^2}{2\sigma^2}}$$

every-day example: out-of-focus photograph . . .

➜ *PDFs of true distributions on* $0 \leq y \leq 1$*:*

- two Breit-Wigner peaks on top of background

$$b_1(y) = \frac{20.334}{100 + (10y - 2)^2} + \frac{2.0334}{1 + (10y - 4)^2} + \frac{4.0668}{4 + (20y - 15)^2}$$

- two narrow gaussian peaks

$$b_2(y) = 5.31923 \exp\left(-200(y - 0.35)^2\right) + 2.659615 \exp\left(-200(y - 0.65)^2\right)$$

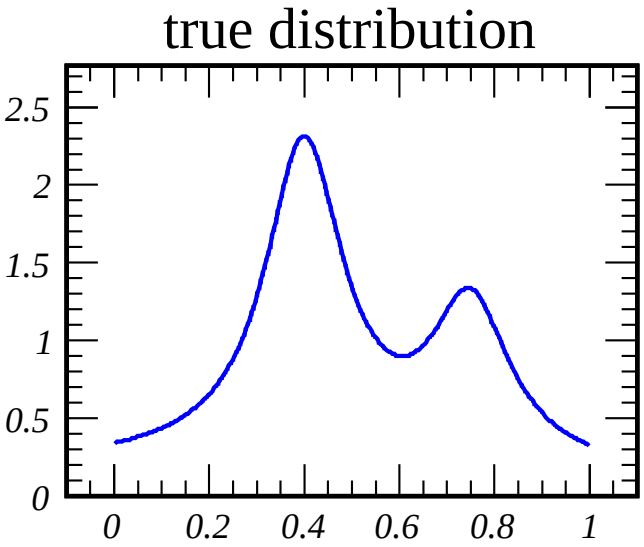
- a step-fuction

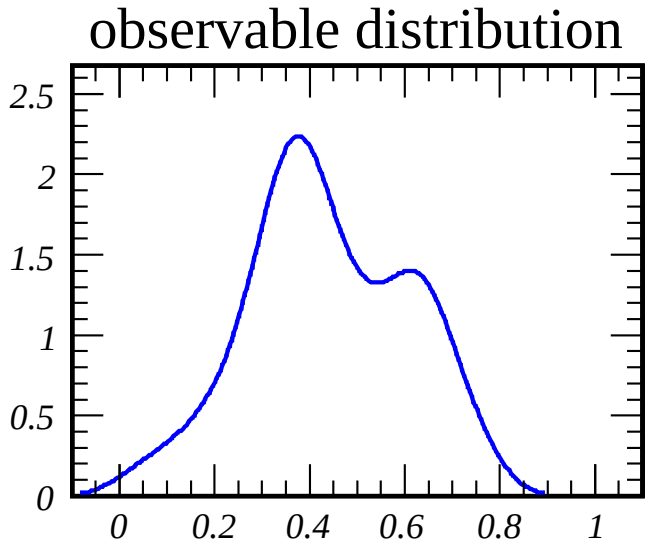
$$b_3(y) = \begin{cases} 2 & \text{for} \quad 0.25 < y < 0.75 \\ 0 & \text{else} \end{cases}$$

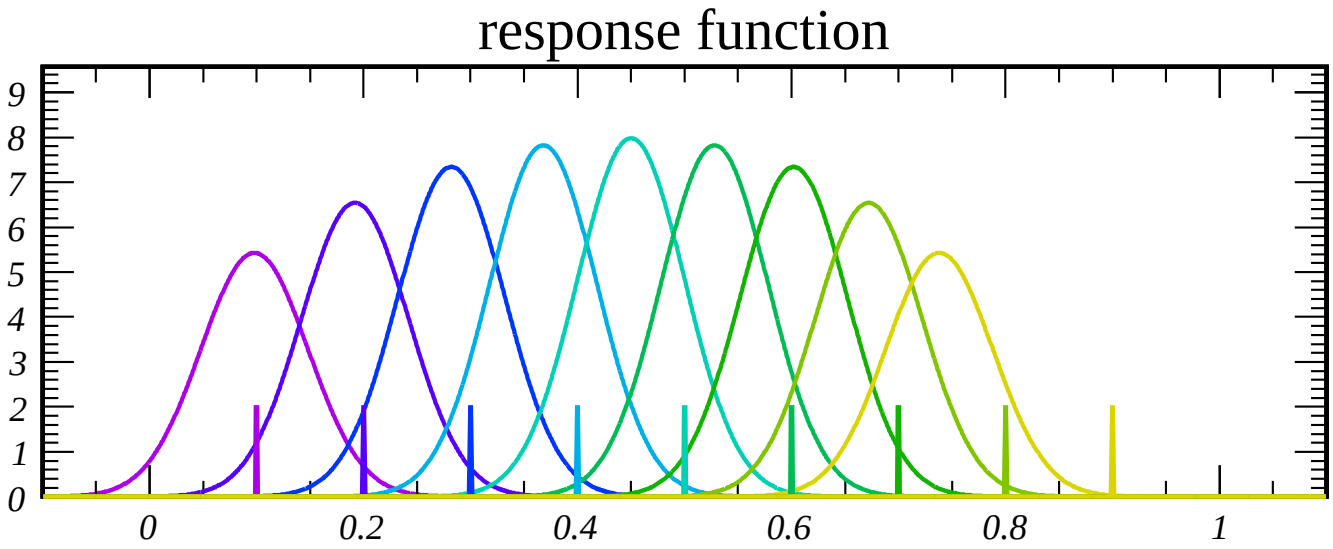
➜ *model for the response function with parameters* σ*,* α *and* β

- gaussian resolution (σ)
- quadratically rising bias of the expectation value of the measurement (β)
- symmetric-parabolic efficiency drop towards the phase-space limit (α)

$$g(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left(x - \left[y - \beta y^2\right]\right)^2\right) \cdot \left(1 - 4\alpha\left(y - \frac{1}{2}\right)^2\right)$$


true distribution
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1


observable distribution
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1


response function
9
8
7
6
5
4
3
2
1
0
0
0.2
0.4
0.6
0.8
1

(demo u0)

➜ *known*

- ☞ response function $g(x, y)$
- ☞ a finite sample of n measurements distributed according to $a(x)$

➜ *wanted:*

- ☞ an estimate for the true distribution $b(y)$

❖ Discussion:

- Events with the same true value y will be measured with different x. This does not mean that the measurements are correlated in x. Independent events are by definition uncorrelated.
- If a theoretical model with few free parameters exists, e.g. $b(y) = \alpha_0 + \alpha_1 y$, then the unfolding problem is best solved by fitting the parameters.
- In any application only a finite sample $x_i, i = 1, \ldots, n$ is given. Since a finite amount of data cannot determine a PDF for all points of a continuum, a completely model independent estimate of $b(y)$ is impossible. One only can find a description based on a finite number of parameters. Some discretization of the problem is required.

➜ *general approach:*

representation of the PDFs by suitable basis functaion $\alpha_k(x)$ and $\beta_l(y)$:

$$a(x) = \sum_{k=1}^{n_a} a_k \; \alpha_k(x) \quad \text{and} \quad b(y) = \sum_{l=1}^{n_b} b_l \; \beta_l(y)$$

When using orthogonal functions the coefficients a_k and b_l are determined by integral-transforms of the respective densities:

$$\int_{x_{\min}}^{x_{\max}} dx \; \alpha_k(x)\alpha_l^*(x) = \delta_{kl} \quad \text{and} \quad \int_{y_{\min}}^{y_{\max}} dy \; \beta_k(y)\beta_l^*(y) = \delta_{kl}$$

and thus e.g.

$$\int_{x_{\min}}^{x_{\max}} dx \; a(x)\alpha_k^*(x) = \int_{x_{\min}}^{x_{\max}} dx \; \left(\sum_{i=1}^{n_a} a_i \alpha_i(x) \right) \alpha_k^*(x)$$

$$= \sum_{i=1}^{n_a} a_i \int_{x_{\min}}^{x_{\max}} dx \; \alpha_i(x)\alpha_k^*(x) = \sum_{i=1}^{n_a} a_i \delta_{ik} = a_k$$

Discretization by non-orthogonal functions is possible, too, but less convenient.

❖ note: there is considerable freedom in the choice of the base functions . . .

- ➜ harmonic functions (➜ Fourier components)
- ➜ orthogonal polynomials
- ➜ histogram bins
- ➜ B-Splines (nicht orthogonal)

For orthogonal base functions discretization yields

$$a_k = \int_{x_{\min}}^{x_{\max}} dx\ \alpha_k^*(x) \cdot a(x) = \int_{x_{\min}}^{x_{\max}} dx\ \alpha_k^*(x) \int_{y_{\min}}^{y_{\max}} dy\ g(x,y) b(y)$$

$$= \int_{x_{\min}}^{x_{\max}} dx\ \alpha_k^*(x) \int_{y_{\min}}^{y_{\max}} dy\ g(x,y) \sum_{i=1}^{n_b} b_i \cdot \beta_i(y)$$

$$= \sum_{i=1}^{n_b} \left(\int_{x_{\min}}^{x_{\max}} dx \int_{y_{\min}}^{y_{\max}} dy\ g(x,y)\ \alpha_k^*(x)\ \beta_i(y) \right) \cdot b_i \equiv \sum_{i=1}^{n_b} G_{ki} b_i$$

➜ *unfolding* $\equiv$ *solving a linear system* $\vec{a} = G\,\vec{b}$ *with response matrix* G_{ik}

➜ *base functions used in the following*

$$\alpha_k^*(x) = \begin{cases} 1 & \text{if} \quad x_{k-1} \leq x < x_k \\ 0 & \text{else} \end{cases}$$

$$\beta_i(y) = \begin{cases} 1/(y_i - y_{i-1}) & \text{if} \quad y_{i-1} \leq y < y_i \\ 0 & \text{else} \end{cases}$$

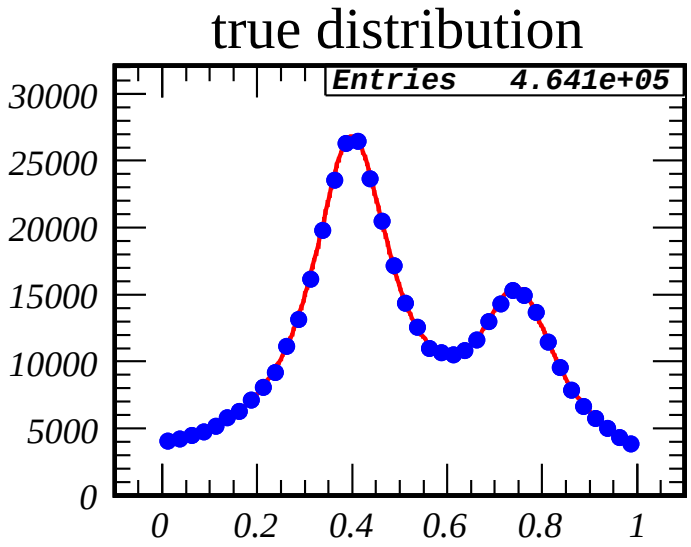
➜ response matrix:

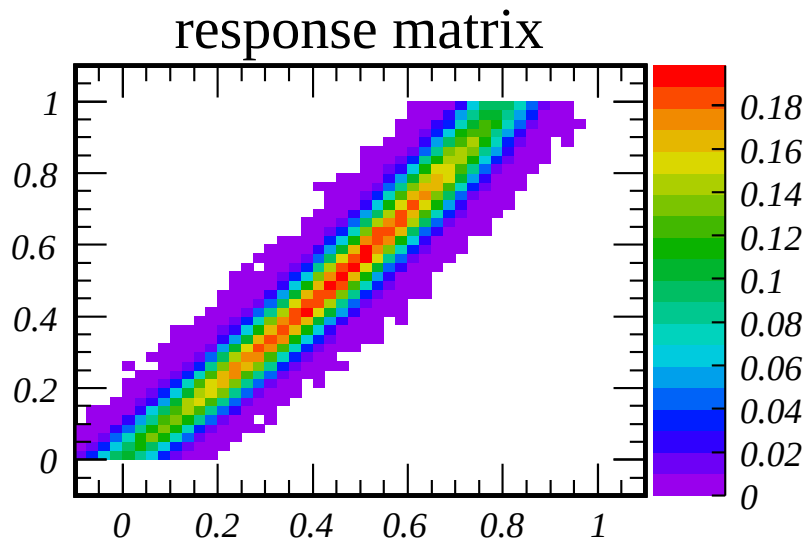
$$G_{ki} = \frac{1}{y_i - y_{i-1}} \int_{x_{k-1}}^{x_k} dx \int_{y_{i-1}}^{y_i} dy \; g(x,y)$$

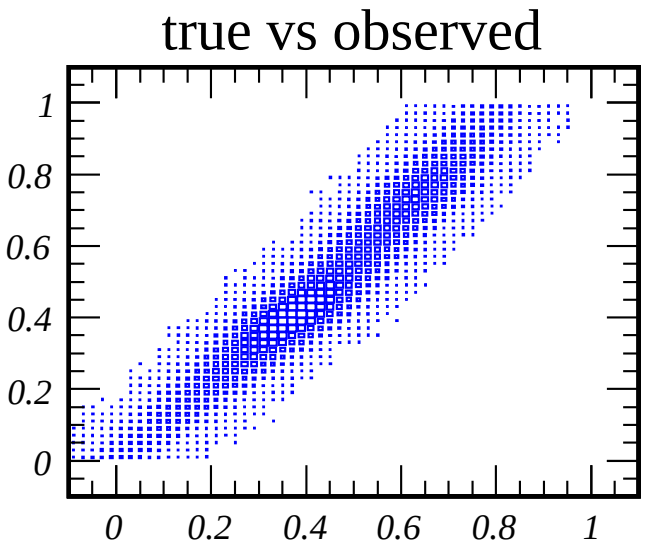
➜ discretized distributions:

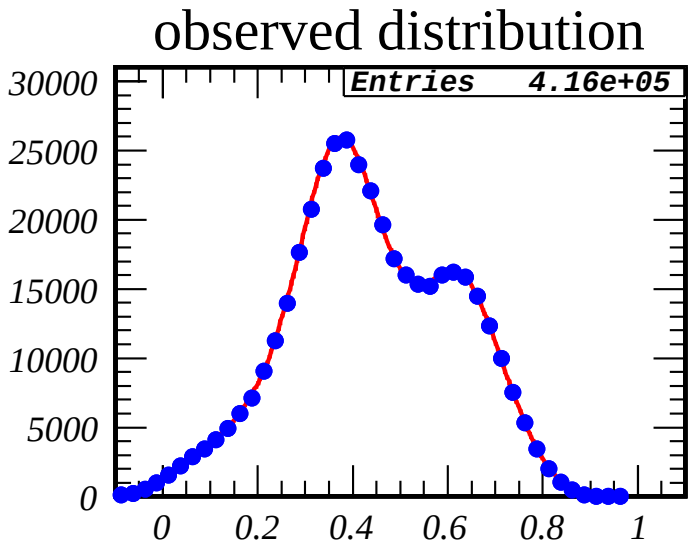
$$a_k = \int_{x_{k-1}}^{x_k} dx \; a(x) \quad \text{and} \quad b_i = \int_{y_{i-1}}^{y_i} dy \; b(y)$$

- simple intuitive interpretation for a_k and b_i
- no assumptions made about smoothness of curvature of the distributions
- note: the numbers of bins in b_i should be large enough, such that the discretization erros in the replacement $b(y) = \sum_i b_i \beta_i(y)$ is negligible.


true distribution
Entries 4.641e+05
30000
25000
20000
15000
10000
5000
0
0 0.2 0.4 0.6 0.8 1


response matrix
1
0.8
0.6
0.4
0.2
0
0 0.2 0.4 0.6 0.8 1
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0


true vs observed
1
0.8
0.6
0.4
0.2
0
0 0.2 0.4 0.6 0.8 1


observed distribution
Entries 4.16e+05
30000
25000
20000
15000
10000
5000
0
0 0.2 0.4 0.6 0.8 1

(demo u1)

➜ *three typical unfolding problems*

	$b(y)$	σ	α	β
Problem 1	$b_1(y)$	$1/20$	$1/2$	$1/10$
Problem 2	$b_2(y)$	$1/12$	0	0
Problem 3	$b_3(y)$	$1/8$	0	0

❖ realization and simplifying assumptions

- (n_a, n_b) bins for the observed and true distribution $(a(x), b(y))$
- no discretization errors
- response matrix

$$G_{ki} = \frac{1}{y_i - y_{i-1}} \int_{x_{k-1}}^{x_k} dx \int_{y_{i-1}}^{y_i} dy \; g(x,y)$$

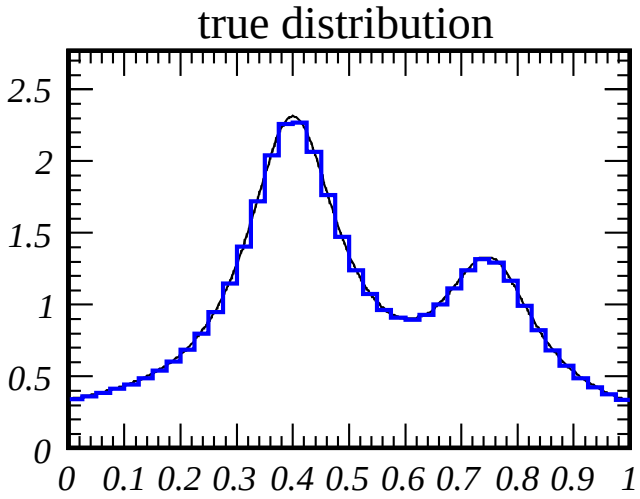
- relation between observables and true distribution

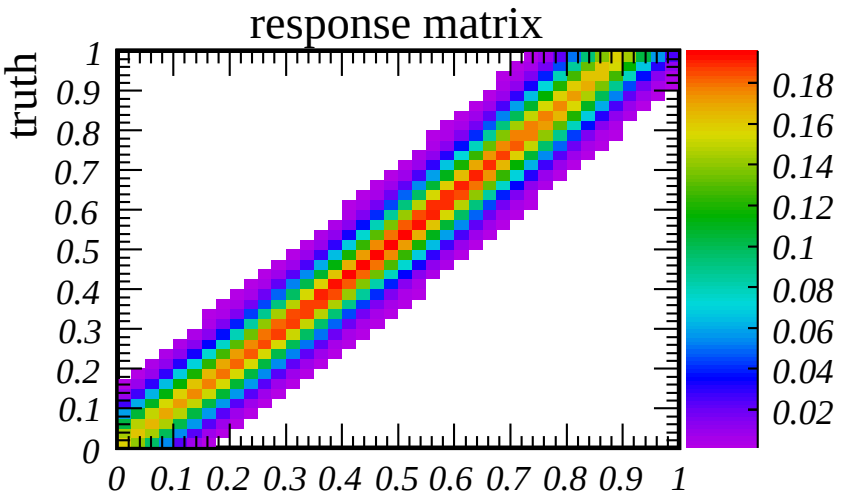
$$\langle \vec{a} \rangle = G \cdot \vec{b}$$

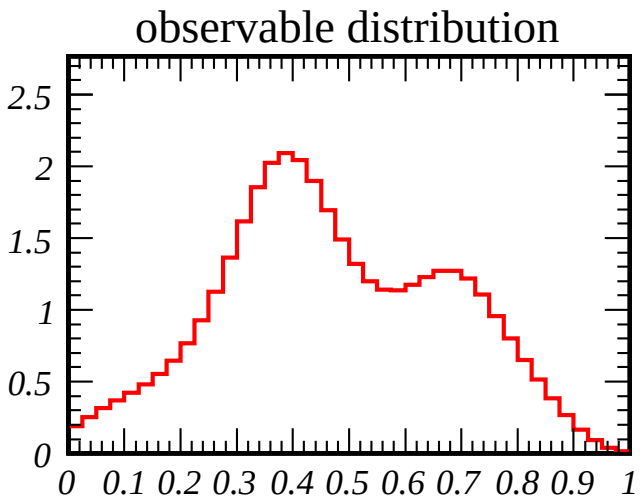
- actual measurements scatter randomly around the expectation values

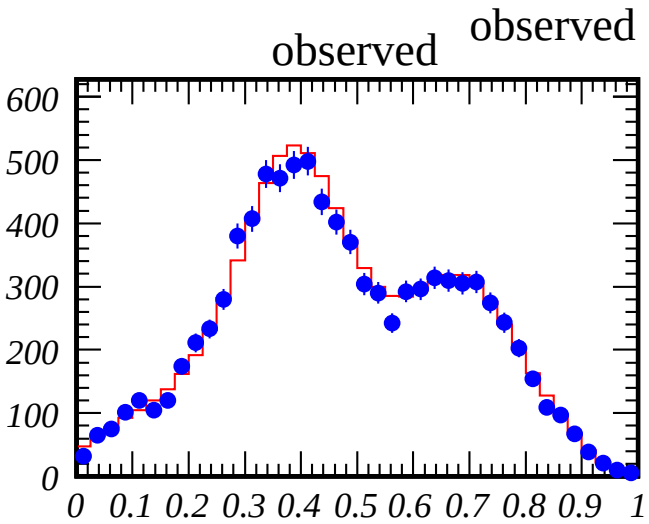
$$\vec{a} = \langle \vec{a} \rangle + \vec{r} \quad \text{with} \quad \langle \vec{r} \rangle = 0 \quad \text{and} \quad \left\langle \vec{r} \cdot \vec{r}^{\,T} \right\rangle = C$$

Illustration









(demo u2)

➜ apply different unfolding methods. . .

➜ *simplest and most common correction method*

- same binning for true and observed distribution
- bin-by-bin correction factors c_k

$$b_k = a_k \cdot c_k$$

- determination of the correction factors:
  - ➜ start with an assumption b_k
  - ➜ determine a_k by multiplication with the response matrix
  - ➜ use $c_k = b_k / a_k$

$$c_k = \frac{b_k}{\sum_{l=1}^{n_b} G_{kl} b_l}$$

- correction factors depend on the assumed b_k - possible choices:
  - ➜ (approximately/expected) true distribution (unknown)
  - ➜ uniform distribution (“objective”)
  - ➜ measured distribution (probably not so bad . . .)
- correct unfolding result guaranteed only for $b_k = b_k^{\mathtt{true}}$
  - ➜ in general at least partial correction should be achieved

➜ *become independent from a specific choice for the initial b_k*

initial setting: (for example)

$$b_k^{(0)} = a_k$$

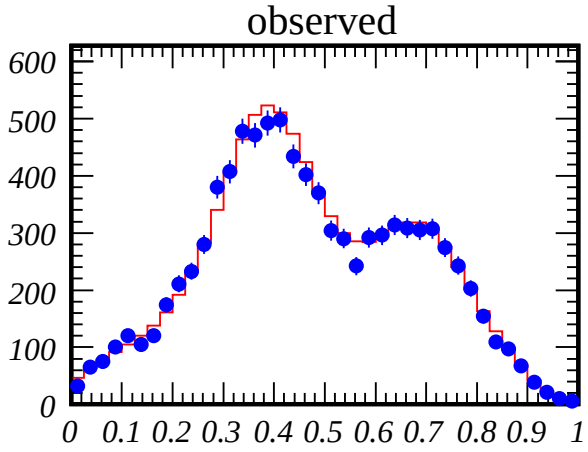
iteration:

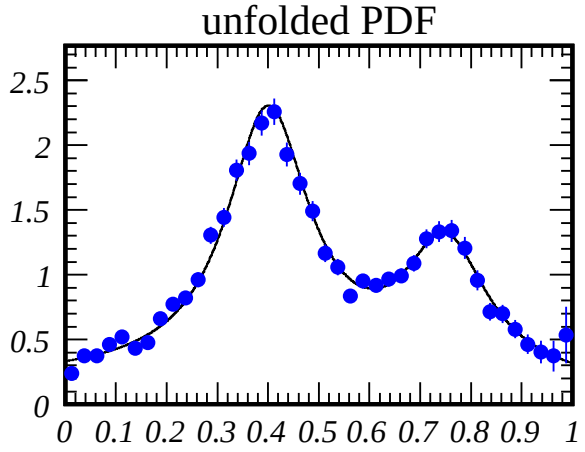
$$b_k^{(n+1)} = a_k \cdot c_k^{(n+1)} = a_k \cdot \frac{b_k^{(n)}}{\sum_{l=1}^{n_b} G_{kl}\, b_l^{(n)}}$$

error estimate for the unfolded distributions:

$$\sigma(b_k) = c_k \cdot \sigma(a_k) \quad \text{would be OK if } C(\vec{c}) = 0$$

➜ *application to sample problems*





(demo u3)

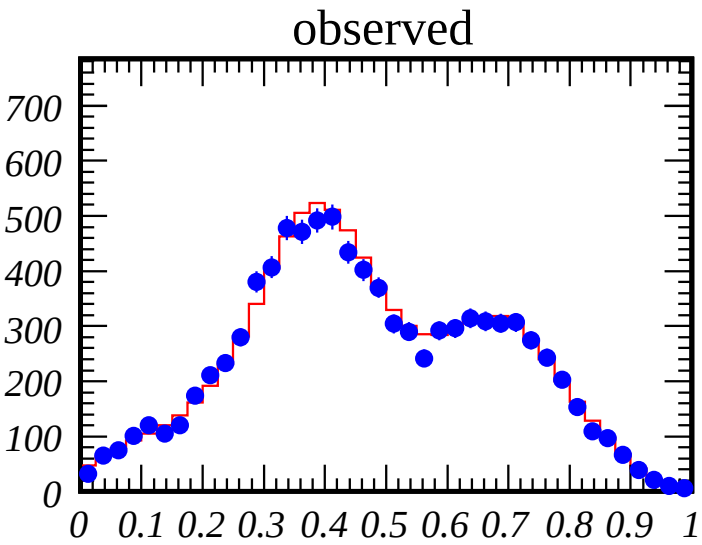
➜ *toy Monte Carlo error propagation*

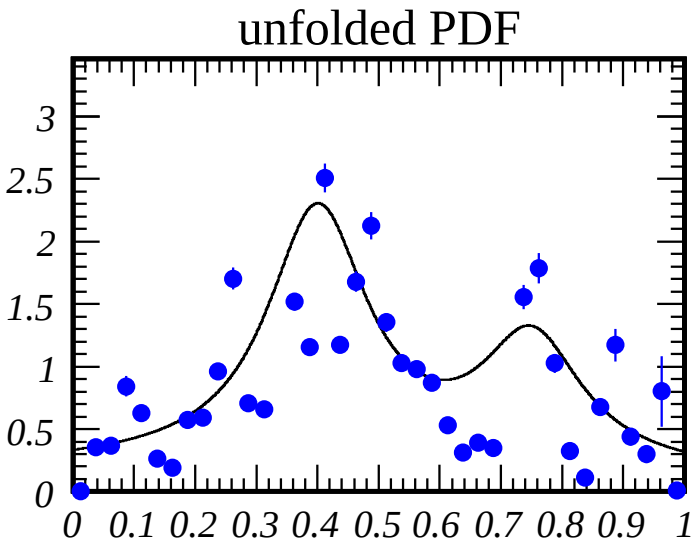
- Fluctuate the measurements $\vec{a}$ according to their errors
  - ➜ generate N pseudo-samples $\vec{a}_n$ with $n = 1, \ldots, N$
- For every pseudo-sample $\vec{a}_n$ determine an unfolded distribution $\vec{b}_n$ by iteration of the correction factors
- Take as result of the unfolding the mean value of the $\vec{b}_n$. Get the error from the empirical covariance matrix.

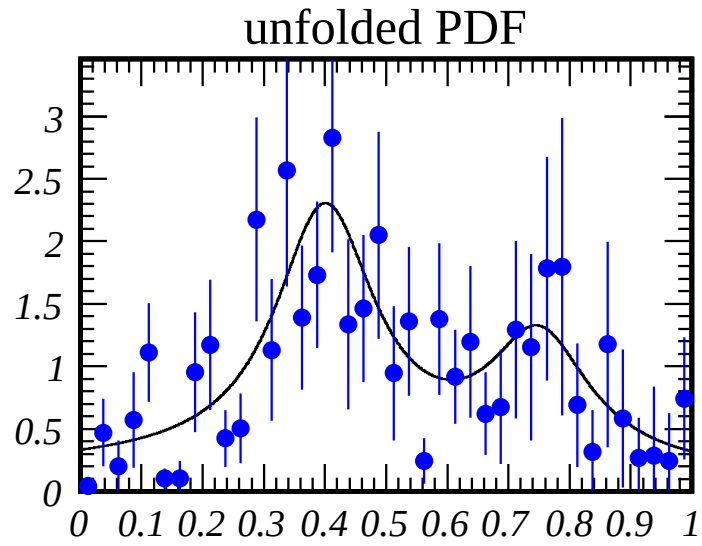
$$\vec{b} = \frac{1}{N} \sum_{n=1}^{N} \vec{b}_n \quad \text{and} \quad C(b) = \left(\frac{1}{N} \sum_{n=1}^{N} \vec{b}_n \cdot \vec{b}_n^T \right) - \vec{b} \cdot \vec{b}^T$$

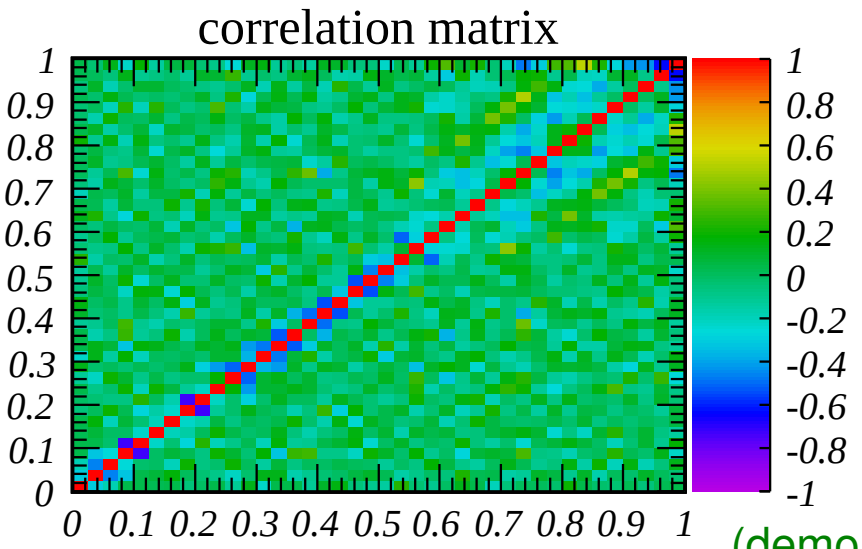
  - ➜ proper covariance matrix for the bins of the unfolded distribution
- observations from numerical studies
  - ➜ surprisingly large statistical errors in the unfolded distribution
  - ➜ strong correlations between neighboring bins
  - ➜ errors and correlations grow with the number of iterations

Numerical study









(demo u4)

➜ *use Bayes' theorem to "invert" the response matrix*

decompose the true distribution int discrete probabilities p_i and normalization B

$$b_i = B \cdot p_i \quad \text{and} \quad a_i = \sum_{k=1}^{n_b} G_{ik} b_k = B \sum_{k=1}^{n_b} G_{ik} p_k$$

interpret the elements of the column-normalized G as conditional probabilities

$$G_{ik} = p(\text{observed } i | \text{true } k)$$

use Bayes' theorem to construct an unfolding matrix H_{ik}:

$$H_{ik} = p(\text{true } k | \text{observed } i) = \frac{p(\text{observed } i | \text{true } k) \cdot p(\text{true } k)}{p(\text{observed } i)}$$

$$= \frac{p(\text{observed } i | \text{true } k) \cdot p(\text{true } k)}{\sum_j p(\text{observed } i | \text{true } j) \cdot p(\text{true } j)} = \frac{G_{ik} \cdot p_k}{\sum_{j=1}^{n_b} G_{ij} p_j}$$

- H_{ik} depends on b_k - like the correction factors
- H_{ik} corrects only resolution effects - no efficiencies

implementation ➜

➜ *strategy:*

- use the unfolding matrix to correct smearing effects
- then correct for efficiencies (column normalization)

$$q_j = \frac{1}{\epsilon_j} \sum_{j=1}^{n_a} a_i \cdot H_{ij} \quad \text{mit} \quad \epsilon_j = \sum_{k=1}^{n_a} G_{kj} \quad \text{und} \quad p_j = \frac{q_j}{\sum_{i=1}^{n_b} q_i}$$

- finally adjust the normalization from e.g. the condition

$$\sum_i B \sum_k G_{ik} \cdot p_k = \sum_i a_i$$

- iterate the above procedure to become independent of the initial p_k
- naive error propagation for q_j

$$C_{ij}(q) = \sum_{k,l=1}^{n_a} \frac{\partial q_i}{\partial a_k} \frac{\partial q_j}{\partial a_l} C_{kl}(a) = \frac{1}{\varepsilon_i \varepsilon_j} \sum_{k,l=1}^{n_a} H_{ki} H_{lj} C_{kl}(a)$$

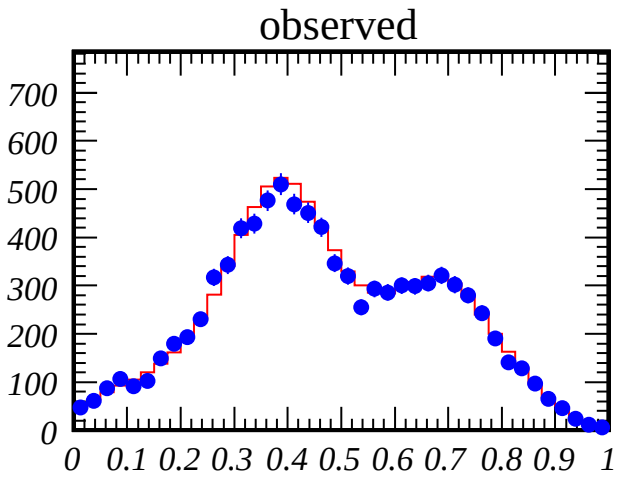
  - ➜ ignores dependence of an iterated unfolding matrix on $\vec{a}$
  - ➜ will in general have correlations

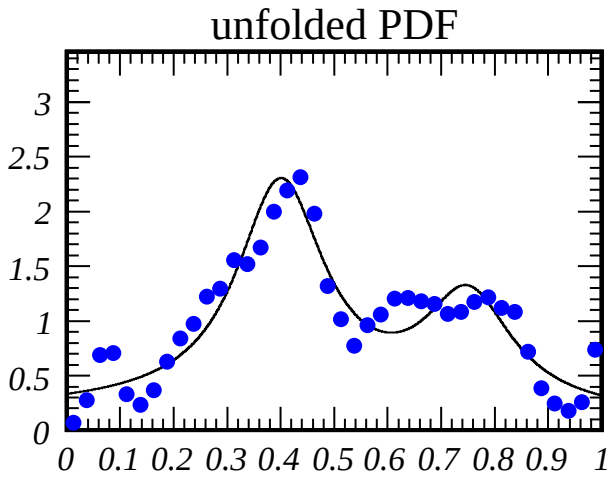
➜ *characteristics of bayesian unfolding*

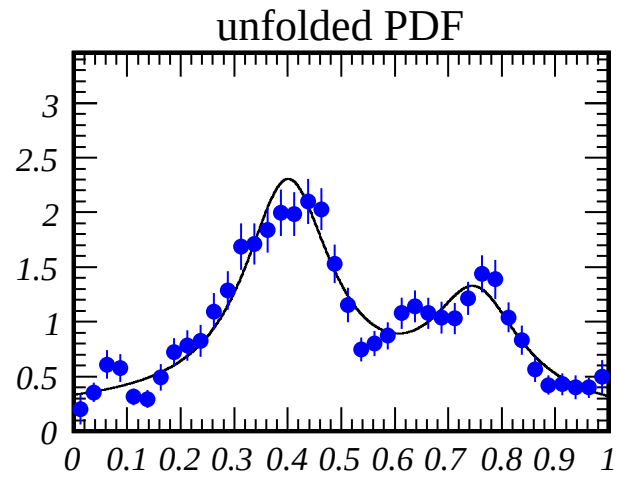
- mathematically well founded approach
- positivity of probabilities is used explicitly
- measurements in unphysical regions are considered consistently
- no matrix inversion required
- unfolding also works for non-square response matrices
- same initial-values-problem as for correction factor
- iteration allows to reduce sensitivity to initial values
- need pseudo-experiments for proper error propagation
  - ➜ determination of the covariance matrix of the unfolded distribution
  - ➜ stabilisation against statistical fluctuations in the measurements

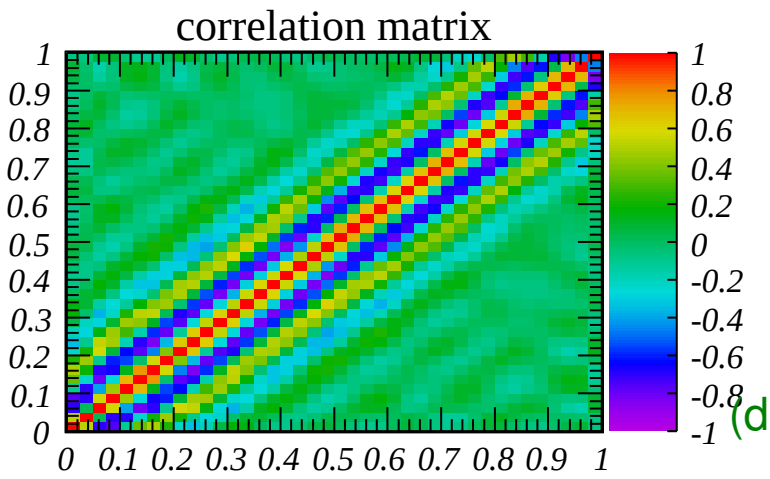
➜ application to sample problems. . .

Numerical tests









(demo u5)

- slow convergence with the number of iterations
- number of iteration steps affects errors and correlations

➜ *problem*

- too few steps: result depends on initial assumptions
- too many steps: result become unstable

➜ *towards a solution*

The number of iterations is arbitrary, but the consequences of a certain choice need to be quantified. Here the covariance matrix can give some guidance. schematically, assuming a quadratic response matrix, one has:

$$\vec{b}_{unf} = H \cdot \vec{a} = H \cdot (G \cdot G^{-1}) \cdot \vec{a} = (H \cdot G) \cdot \vec{b}_{true} = G_{res} \cdot \vec{b}_{true}$$

The unfolded distribution the true distribution smeared by a residual response matrix $G_{res} = H \cdot G$. For $H = G^{-1}$ one would obtain perfect correction. The residual resolution can be estimated from the correlation length in the covariance matrix.