Statistical Methods for High Energy Physics
Probability

Saeid Paktinat

School of Particles and accelerators
IPM, Tehran

**Third National Workshop on
Detectors and Calculation Methods in Particle Physics**

Azar 26-28, 1387

Outline

Lecture 1 (Probability)

Probability

Random variables, probability densities, etc.

Brief catalogue of probability densities

The Monte Carlo method

Lecture 2 (Statistics)

Statistical tests

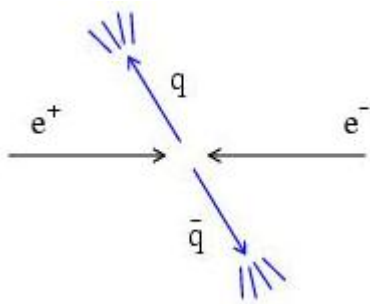Fisher discriminants, etc.

Significance and goodness-of-fit tests

Parameter estimation

Maximum likelihood and least squares

Interval estimation (setting limits)

Data analysis in particle physics

Observe events of a certain type

Measure characteristics of each event (particle momenta, number of muons, energy of jets,...)

Theories (e.g. SM) predict distributions of these properties up to free parameters, e.g., α, G_F, M_Z, α_s, m_H, ...

Some tasks of data analysis:

Estimate (measure) the parameters;

Quantify the uncertainty of the parameter estimates;

Test the extent to which the predictions of a theory are in agreement with the data ($\rightarrow$ presence of New Physics?)

Dealing with uncertainty

In particle physics there are various elements of uncertainty:

theory is not deterministic

quantum mechanics

random measurement errors

present even without quantum effects

things we could know in principle but don't

e.g. from limitations of cost, time, ...
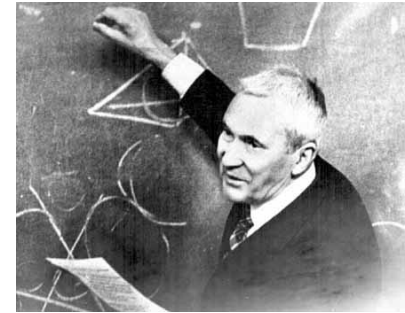
We can quantify the uncertainty using PROBABILITY

A definition of probability

Consider a set S with subsets A, B, ...

For all $A \subset S, P(A) \geq 0$

$$P(S) = 1$$

If $A \cap B = \emptyset, P(A \cup B) = P(A) + P(B)$

Kolmogorov axioms (1933)

From these axioms we can derive further properties, e.g.

$P(\overline{A}) = 1 - P(A)$

$P(A \cup \overline{A}) = 1$

$P(\emptyset) = 0$

if $A \subset B$, then $P(A) \leq P(B)$

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

Conditional probability, independence

Also define conditional probability of *A* given *B* (with *P(B) ≠ 0*):

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

E.g. rolling dice: $P(n < 3 \,|\, n \text{ even}) = \frac{P((n<3) \cap n \text{ even})}{P(\text{even})} = \frac{1/6}{3/6} = \frac{1}{3}$

Subsets *A, B* independent if: $P(A \cap B) = P(A)P(B)$

If *A, B* independent, $P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A)$

N.B. do not confuse with disjoint subsets, i.e., $A \cap B = \emptyset$

Interpretation of probability

I. **Relative frequency**

$A, B, ...$ are outcomes of a repeatable experiment

$$P(A) = \lim_{n \to \infty} \frac{\text{times outcome is } A}{n}$$

cf. quantum mechanics, particle scattering, radioactive decay...

II. **Subjective probability**

$A, B, ...$ are hypotheses (statements that are true or false)

$$P(A) = \text{degree of belief that } A \text{ is true}$$

- Both interpretations consistent with Kolmogorov axioms.
- In particle physics frequency interpretation often most useful, but subjective probability can provide more natural treatment of non-repeatable phenomena:

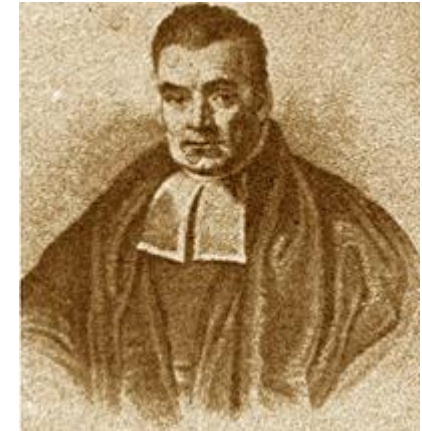systematic uncertainties, probability that Higgs boson exists,...

Bayes' theorem

From the definition of conditional probability we have,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{and} \quad P(B|A) = \frac{P(B \cap A)}{P(A)}$$

but $P(A \cap B) = P(B \cap A)$, so

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

Bayes' theorem

First published (posthumously) by the Reverend Thomas Bayes (1702−1761)

An essay towards solving a problem in the doctrine of chances, Philos. Trans. R. Soc. **53** (1763) 370; reprinted in Biometrika, **45** (1958) 293.
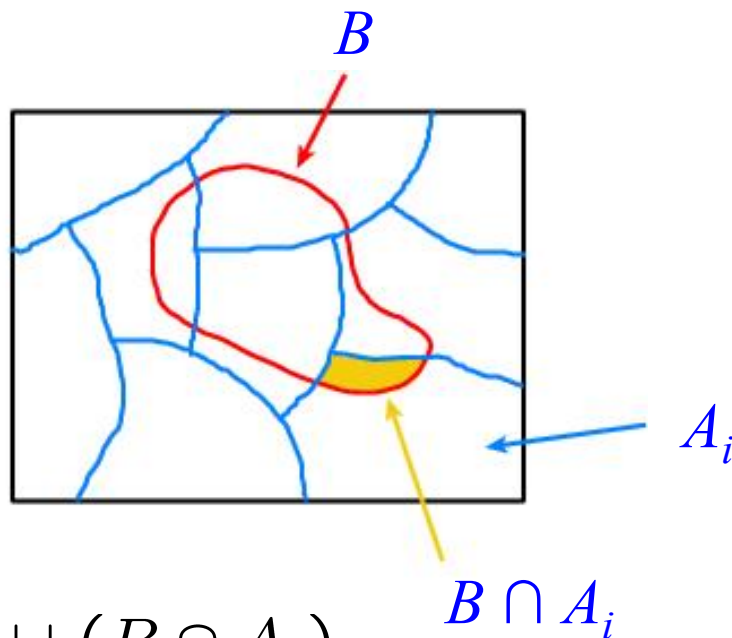
The law of total probability

Consider a subset *B* of the sample space *S*,

divided into disjoint subsets A_i such that $[_i A_i = S$,

B

S

A_i

$B \cap A_i$

$\rightarrow \quad B = B \cap S = B \cap (\cup_i A_i) = \cup_i (B \cap A_i),$

$\rightarrow \quad P(B) = P(\cup_i (B \cap A_i)) = \sum_i P(B \cap A_i)$

$\rightarrow \quad P(B) = \sum_i P(B|A_i)P(A_i)$ law of total probability

Bayes' theorem becomes

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

An example using Bayes' theorem

Suppose the probability (for anyone) to have AIDS is:

$$P(\text{AIDS}) = 0.001$$
$$P(\text{no AIDS}) = 0.999$$

← prior probabilities, i.e., before any test carried out

Consider an AIDS test: result is + or −

$$P(+|\text{AIDS}) = 0.98$$
$$P(-|\text{AIDS}) = 0.02$$

← probabilities to (in)correctly identify an infected person

$$P(+|\text{no AIDS}) = 0.03$$
$$P(-|\text{no AIDS}) = 0.97$$

← probabilities to (in)correctly identify an uninfected person

Suppose your result is +. How worried should you be?

Bayes' theorem example (cont.)

The probability to have AIDS given a + result is

$$P(\text{AIDS}|+) = \frac{P(+|\text{AIDS})P(\text{AIDS})}{P(+|\text{AIDS})P(\text{AIDS}) + P(+|\text{no AIDS})P(\text{no AIDS})}$$

$$= \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.03 \times 0.999}$$

$$= 0.032 \qquad \leftarrow \text{posterior probability}$$

i.e. you're probably OK!

Your viewpoint: my degree of belief that I have AIDS is 3.2%

Your doctor's viewpoint: 3.2% of people like this will have AIDS

Frequentist Statistics − general philosophy

In frequentist statistics, probabilities are associated only with the data, i.e., outcomes of repeatable observations (shorthand: $\vec{x}$).

Probability = limiting frequency

Probabilities such as

P (Higgs boson exists),
P ($0.117 < \alpha_\mathrm{s} < 0.121$),

etc. are either 0 or 1, but we don't know which.

The tools of frequentist statistics tell us what to expect, under the assumption of certain probabilities, about hypothetical repeated observations.

The preferred theories (models, hypotheses, ...) are those for which our observations would be considered 'usual'.

Bayesian Statistics − general philosophy

In Bayesian statistics, use subjective probability for hypotheses:

probability of the data assuming
hypothesis H (the likelihood)

prior probability, i.e.,
before seeing the data

$$P(H|\vec{x}) = \frac{P(\vec{x}|H)\pi(H)}{\int P(\vec{x}|H)\pi(H)\,dH}$$

posterior probability, i.e.,
after seeing the data

normalization involves sum
over all possible hypotheses

Bayes' theorem has an "if-then" character: If your prior
probabilities were $\pi(H)$, then it says how these probabilities
should change in the light of the data.

No unique prescription for priors (subjective!)

Random variables and probability density functions

A random variable is a numerical characteristic assigned to an element of the sample space; can be discrete or continuous.

Suppose outcome of experiment is continuous value x

$$P(x \text{ found in } [x, x+dx]) = f(x)\, dx$$

→ $f(x)$ = probability density function (pdf)

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \qquad x \text{ must be somewhere}$$

Or for discrete outcome x_i with e.g. $i = 1, 2, ...$ we have
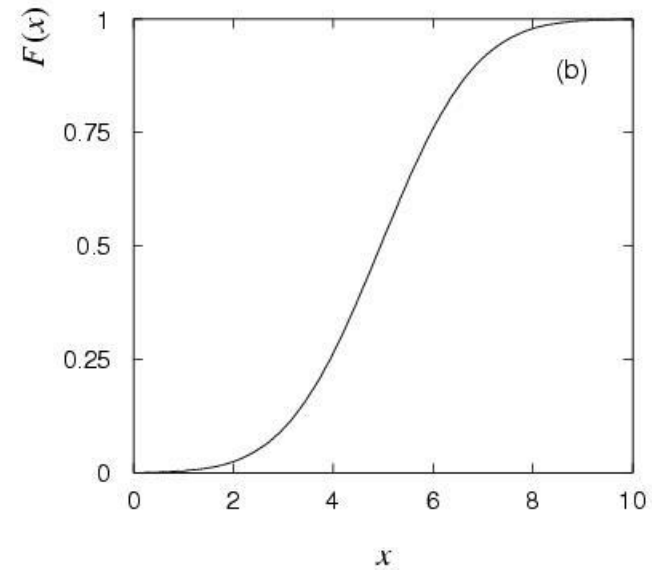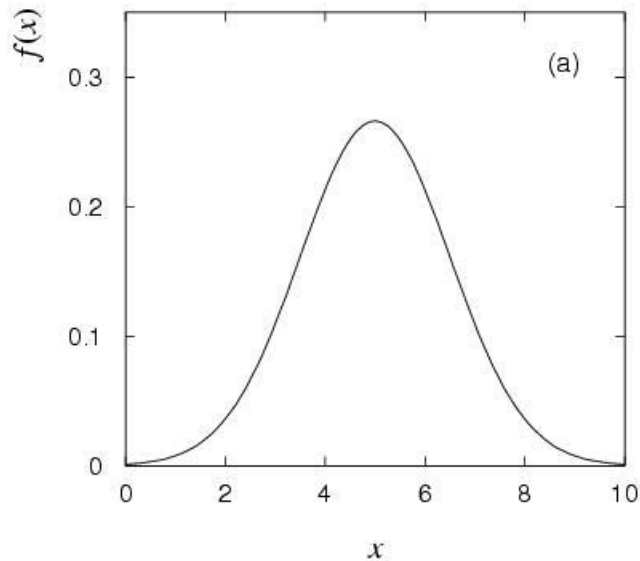
$$P(x_i) = p_i \qquad \text{probability mass function}$$

$$\sum_i P(x_i) = 1 \qquad x \text{ must take on one of its possible values}$$

Cumulative distribution function

Probability to have outcome less than or equal to x is

$$\int_{-\infty}^{x} f(x')\,dx' \equiv F(x)$$ cumulative distribution function

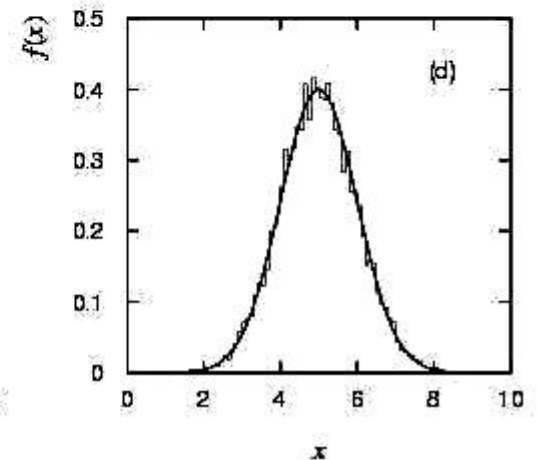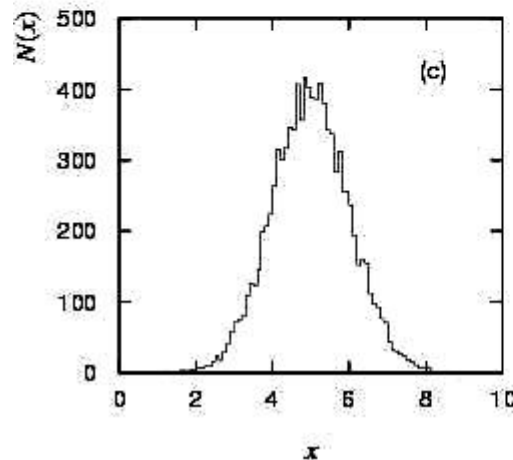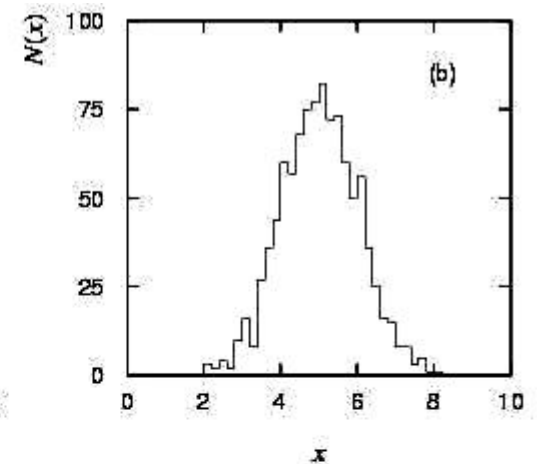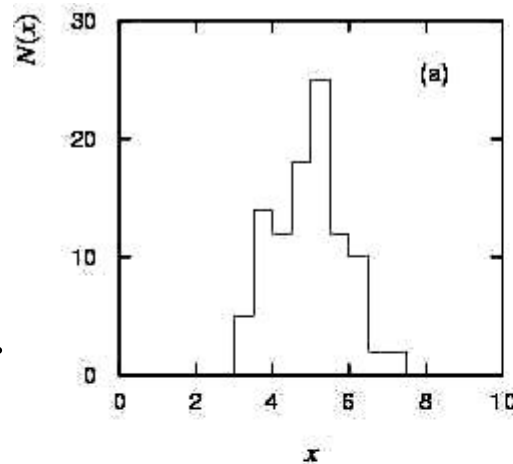Alternatively define pdf with $f(x) = \dfrac{\partial F(x)}{\partial x}$

Histograms

pdf = histogram with

infinite data sample,
zero bin width,
normalized to unit area.

$$f(x) = \frac{N(x)}{n\Delta x}$$

$n =$ number of entries

$\Delta x =$ bin width

S.Paktinat

Other types of probability densities

Outcome of experiment characterized by several values,
e.g. an n-component vector, $(x_1, ... x_n)$

$\rightarrow$ joint pdf $\quad f(x_1, \ldots, x_n)$

Sometimes we want only pdf of some (or one) of the components

$\rightarrow$ marginal pdf $\quad f_1(x_1) = \int \cdots \int f(x_1, \ldots, x_n)\, dx_2 \ldots dx_n$

x_1, x_2 independent if $\quad f(x_1, x_2) = f_1(x_1) f_2(x_2)$

Sometimes we want to consider some components as constant

$\rightarrow$ conditional pdf $\quad g(x_1|x_2) = \dfrac{f(x_1, x_2)}{f_2(x_2)}$

Expectation values

Consider continuous r.v. x with pdf $f(x)$.

Define expectation (mean) value as $\quad E[x] = \int x\, f(x)\, dx$

Notation (often): $\quad E[x] = \mu \quad$ ~ "centre of gravity" of pdf.
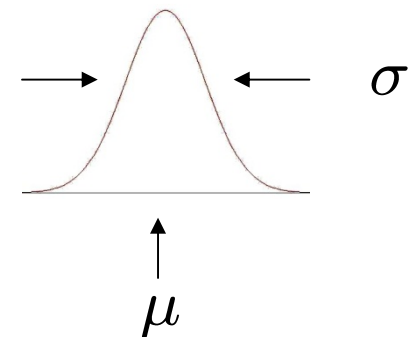
For a function $y(x)$ with pdf $g(y)$,

$$E[y] = \int y\, g(y)\, dy = \int y(x) f(x)\, dx \qquad \text{(equivalent)}$$

Variance: $\quad V[x] = E[x^2] - \mu^2 = E[(x-\mu)^2]$

Notation: $\quad V[x] = \sigma^2$

Standard deviation: $\quad \sigma = \sqrt{\sigma^2}$

σ ~ width of pdf, same units as x.

Covariance and correlation

Define covariance cov[*x*,*y*] (also use matrix notation V_{xy}) as

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

Correlation coefficient (dimensionless) defined as

$$\rho_{xy} = \frac{\text{cov}[x, y]}{\sigma_x \sigma_y}$$

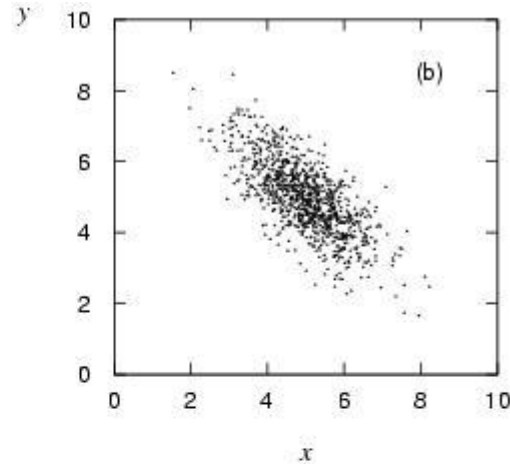If *x, y*, independent, i.e., $f(x, y) = f_x(x) f_y(y)$, then

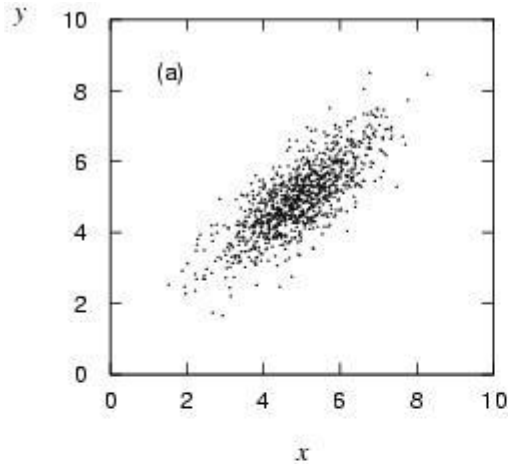$$E[xy] = \int \int xy \, f(x, y) \, dx dy = \mu_x \mu_y$$

$$\rightarrow \quad \text{cov}[x, y] = 0 \qquad x \text{ and } y, \text{ 'uncorrelated'}$$
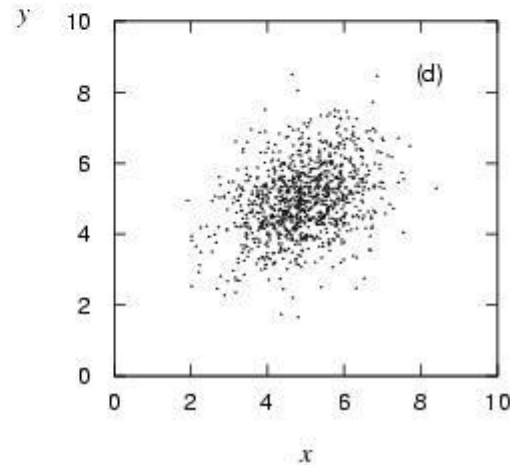
N.B. converse not always true.

Correlation (cont.)

$\rho = 0.75$

$\rho = -0.75$

$\rho = 0.95$
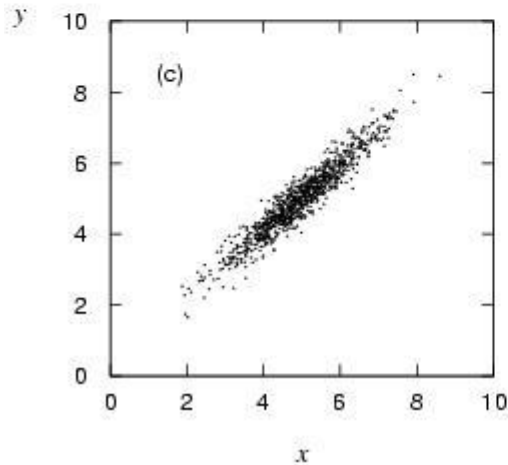
$\rho = 0.25$

Error propagation

Suppose we measure a set of values $\vec{x} = (x_1, \ldots, x_n)$

and we have the covariances $V_{ij} = \text{cov}[x_i, x_j]$

which quantify the measurement errors in the x_i.

Now consider a function $y(\vec{x})$.

What is the variance of $y(\vec{x})$?

The hard way: use joint pdf $f(\vec{x})$ to find the pdf $g(y)$,

then from $g(y)$ find $V[y] = E[y^2] - (E[y])^2$.

Often not practical, $f(\vec{x})$ may not even be fully known.

S.Paktinat

Error propagation (2)

Suppose we had $\vec{\mu} = E[\vec{x}]$

in practice only estimates given by the measured $\vec{x}$

Expand $y(\vec{x})$ to 1st order in a Taylor series about $\vec{\mu}$

$$y(\vec{x}) \approx y(\vec{\mu}) + \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i)$$

To find $V[y]$ we need $E[y^2]$ and $E[y]$.

$$E[y(\vec{x})] \approx y(\vec{\mu}) \quad \text{since} \quad E[x_i - \mu_i] = 0$$

Error propagation (3)

$$E[y^2(\vec{x})] \approx y^2(\vec{\mu}) \; + \; 2y(\vec{\mu}) \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}} E[x_i - \mu_i]$$

$$+E\left[\left(\sum_{i=1}^{n}\left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}}(x_i - \mu_i)\right)\left(\sum_{j=1}^{n}\left[\frac{\partial y}{\partial x_j}\right]_{\vec{x}=\vec{\mu}}(x_j - \mu_j)\right)\right]$$

$$= y^2(\vec{\mu}) + \sum_{i,j=1}^{n}\left[\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} V_{ij}$$

Putting the ingredients together gives the variance of $y(\vec{x})$

$$\sigma_y^2 \approx \sum_{i,j=1}^{n}\left[\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} V_{ij}$$

Error propagation (4)

If the x_i are uncorrelated, i.e., $V_{ij} = \sigma_i^2 \delta_{ij}$, then this becomes

$$\sigma_y^2 \approx \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]^2_{\vec{x}=\vec{\mu}} V_{ij}$$

Similar for a set of m functions $\vec{y}(\vec{x}) = (y_1(\vec{x}), \ldots, y_m(\vec{x}))$

$$U_{kl} = \text{cov}[y_k, y_l] \approx \sum_{i,j=1}^{n} \left[\frac{\partial y_k}{\partial x_i}\frac{\partial y_l}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} V_{ij}$$

or in matrix notation $U = A V A^T$, where

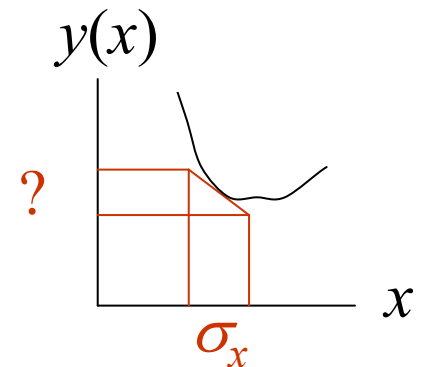$$A_{ij} = \left[\frac{\partial y_i}{\partial x_j}\right]_{\vec{x}=\vec{\mu}}$$

Error propagation (5)

The 'error propagation' formulae tell us the covariances of a set of functions $\vec{y}(\vec{x}) = (y_1(\vec{x}), \ldots, y_m(\vec{x}))$ in terms of the covariances of the original variables.

Limitations: exact only if $\vec{y}(\vec{x})$ linear. Approximation breaks down if function nonlinear over a region comparable in size to the σ_i.
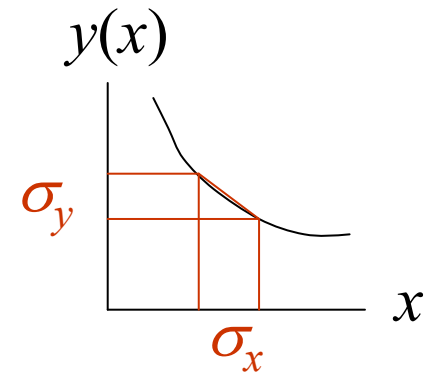
N.B. We have said nothing about the exact pdf of the x_i, e.g., it doesn't have to be Gaussian.

Error propagation − special cases

$$y = x_1 + x_2 \quad \rightarrow \quad \sigma_y^2 = \sigma_1^2 + \sigma_2^2 + 2\text{cov}[x_1, x_2]$$

$$y = x_1 x_2 \quad \rightarrow \quad \frac{\sigma_y^2}{y^2} = \frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} + 2\frac{\text{cov}[x_1, x_2]}{x_1 x_2}$$

That is, if the x_i are uncorrelated:

        add errors quadratically for the sum (or difference),

        add relative errors quadratically for product (or ratio).

But correlations can change this completely...

Error propagation − special cases (2)

Consider $y = x_1 - x_2$ with

$$\mu_1 = \mu_2 = 10, \quad \sigma_1 = \sigma_2 = 1, \quad \rho = \frac{\text{cov}[x_1, x_2]}{\sigma_1 \sigma_2} = 0 .$$

$$V[y] = 1^2 + 1^2 = 2, \rightarrow \sigma_y = 1.4$$

Now suppose $\rho = 1$. Then

$$V[y] = 1^2 + 1^2 - 2 = 0, \rightarrow \sigma_y = 0$$

i.e. for 100% correlation, error in difference $\rightarrow$ 0.

Some distributions

Distribution/pdf	Example use in HEP
Binomial	Branching ratio
Multinomial	Histogram with fixed N
Poisson	Number of events found
Uniform	Monte Carlo method
Exponential	Decay time
Gaussian	Measurement error
Chi-square	Goodness-of-fit
Cauchy	Mass of resonance
Landau	Ionization energy loss

Binomial distribution

Consider *N* independent experiments (Bernoulli trials):

outcome of each is 'success' or 'failure',

probability of success on any given trial is *p*.

Define discrete r.v. *n* = number of successes ($0 \leq n \leq N$).

Probability of a specific outcome (in order), e.g. 'ssfsf' is

$$pp(1-p)p(1-p) = p^n(1-p)^{N-n}$$

But order not important; there are $\dfrac{N!}{n!(N-n)!}$

ways (permutations) to get *n* successes in *N* trials, total probability for *n* is sum of probabilities for each permutation.

Binomial distribution (2)

The binomial distribution is therefore

$$f(n; N, p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$
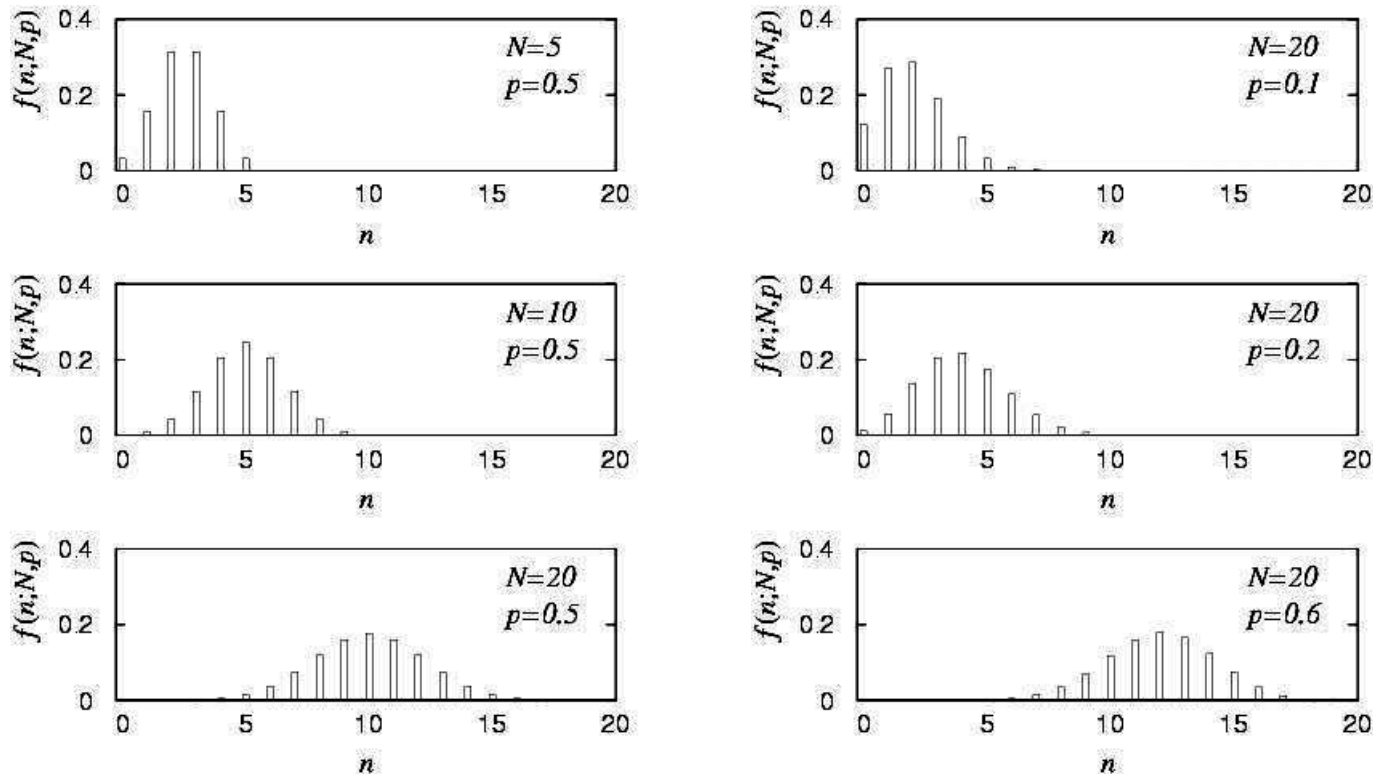
random
variable

parameters

For the expectation value and variance we find:

$$E[n] = \sum_{n=0}^{N} n f(n; N, p) = Np$$

$$V[n] = E[n^2] - (E[n])^2 = Np(1-p)$$

Binomial distribution (3)

Binomial distribution for several values of the parameters:

Example: observe N decays of $W^{\pm}$, the number n of which are $W \rightarrow \mu\nu$ is a binomial r.v., p = branching ratio.

Multinomial distribution

Like binomial but now *m* outcomes instead of two, probabilities are

$$\vec{p} = (p_1, \ldots, p_m), \quad \text{with} \quad \sum_{i=1}^{m} p_i = 1 .$$

For *N* trials we want the probability to obtain:

n_1 of outcome 1,
n_2 of outcome 2,
…
n_m of outcome *m*.

This is the multinomial distribution for $\vec{n} = (n_1, \ldots, n_m)$

$$f(\vec{n}; N, \vec{p}) = \frac{N!}{n_1! n_2! \cdots n_m!} p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}$$

Multinomial distribution (2)

Now consider outcome *i* as 'success', all others as 'failure'.

$\rightarrow$ all n_i individually binomial with parameters N, p_i

$$E[n_i] = Np_i, \quad V[n_i] = Np_i(1 - p_i) \quad \text{for all } i$$

One can also find the covariance to be

$$V_{ij} = Np_i(\delta_{ij} - p_j)$$

Example: $\vec{n} = (n_1, \dots, n_m)$ represents a histogram

with *m* bins, *N* total entries, all entries independent.

Poisson distribution
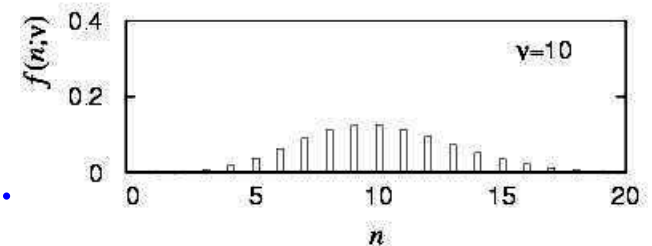
Consider binomial n in the limit
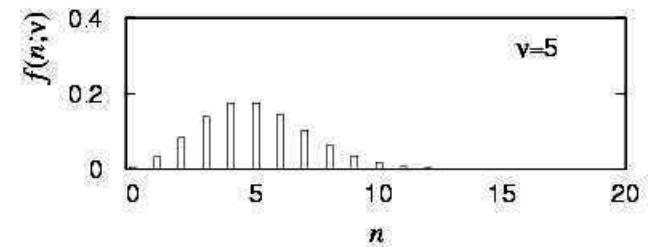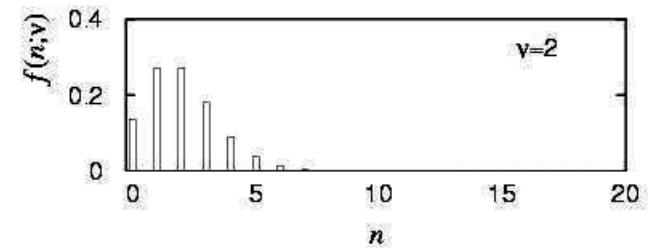
$$N \to \infty, \qquad p \to 0, \qquad E[n] = Np \to \nu \, .$$

$\to$ n follows the Poisson distribution:

$$f(n; \nu) = \frac{\nu^n}{n!} e^{-\nu} \quad (n \geq 0)$$

$$E[n] = \nu \, , \quad V[n] = \nu \, .$$

Example: number of scattering events n with cross section σ found for a fixed integrated luminosity, with $\nu = \sigma \int L \, dt$.

Uniform distribution

Consider a continuous r.v. x with $-\infty < x < \infty$. Uniform pdf is:

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha \le x \le \beta \\ 0 & \text{otherwise} \end{cases}$$

$$E[x] = \frac{1}{2}(\alpha + \beta)$$

$$V[x] = \frac{1}{12}(\beta - \alpha)$$
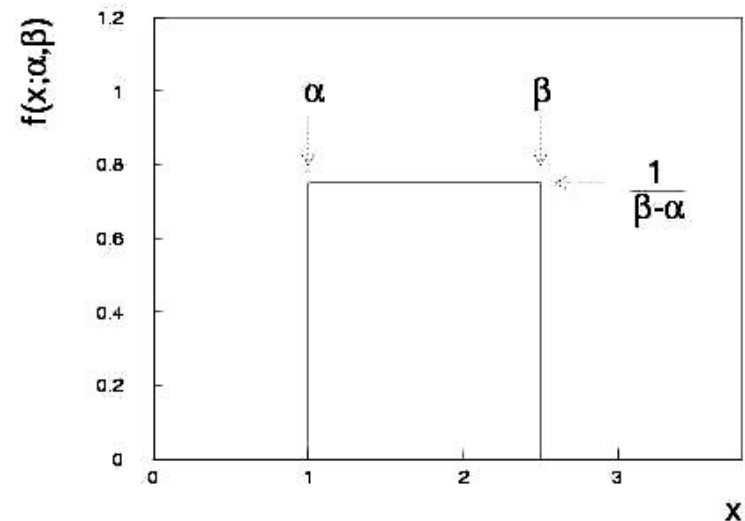
N.B. For any r.v. x with cumulative distribution $F(x)$, $y = F(x)$ is uniform in [0,1].

Example: for $\pi^0 \rightarrow \gamma\gamma$, E_γ is uniform in $[E_{min}, E_{max}]$, with

$$E_{min} = \frac{1}{2}E_\pi(1 - \beta), \qquad E_{max} = \frac{1}{2}E_\pi(1 + \beta)$$

Exponential distribution

The exponential pdf for the continuous r.v. x is defined by:

$$f(x; \xi) = \begin{cases} \frac{1}{\xi}e^{-x/\xi} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$E[x] = \xi$$

$$V[x] = \xi^2$$
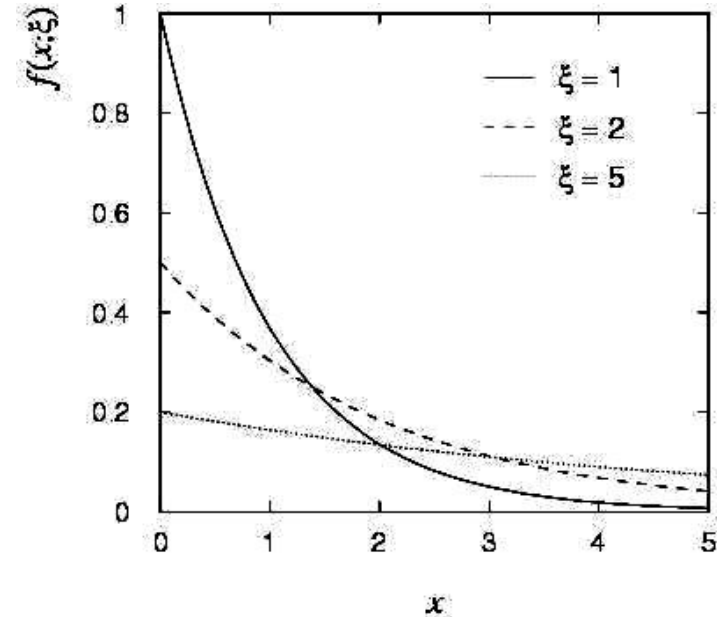
Example: proper decay time t of an unstable particle

$$f(t; \tau) = \frac{1}{\tau}e^{-t/\tau} \qquad (\tau = \text{mean lifetime})$$

Lack of memory (unique to exponential): $f(t - t_0 | t \geq t_0) = f(t)$

Gaussian distribution

The Gaussian (normal) pdf for a continuous r.v. x is defined by:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

$E[x] = \mu$

$V[x] = \sigma^2$

(N.B. often μ, σ^2 denote mean, variance of any r.v., not only Gaussian.)
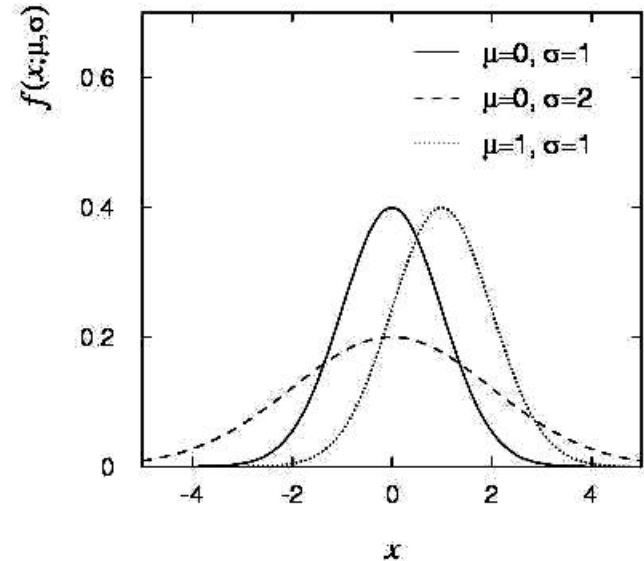
Special case: $\mu = 0$, $\sigma^2 = 1$ ('standard Gaussian'):

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} , \quad \Phi(x) = \int_{-\infty}^{x} \varphi(x') \, dx'$$

If $y \sim$ Gaussian with μ, σ^2, then $x = (y - \mu)/\sigma$ follows $\varphi(x)$.

Gaussian pdf and the Central Limit Theorem

The Gaussian pdf is so useful because almost any random variable that is a sum of a large number of small contributions follows it. This follows from the Central Limit Theorem:

For n independent r.v.s x_i with finite variances σ_i^2, otherwise arbitrary pdfs, consider the sum

$$y = \sum_{i=1}^{n} x_i$$

In the limit $n \to \infty$, y is a Gaussian r.v. with

$$E[y] = \sum_{i=1}^{n} \mu_i \qquad V[y] = \sum_{i=1}^{n} \sigma_i^2$$

Measurement errors are often the sum of many contributions, so frequently measured values can be treated as Gaussian r.v.s.

Central Limit Theorem (2)

The CLT can be proved using characteristic functions (Fourier transforms), see, e.g., Statistical Data Analysis (By. Cowan) Chapter 10.

For finite n, the theorem is approximately valid to the extent that the fluctuation of the sum is not dominated by one (or few) terms.

Beware of measurement errors with non-Gaussian tails.

Good example: velocity component v_x of air molecules.

OK example: total deflection due to multiple Coulomb scattering. (Rare large angle deflections give non-Gaussian tail.)

Bad example: energy loss of charged particle traversing thin gas layer. (Rare collisions make up large fraction of energy loss, cf. Landau pdf.)

Multivariate Gaussian distribution

Multivariate Gaussian pdf for the vector $\vec{x} = (x_1, \ldots, x_n)$:

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

$\vec{x}, \ \vec{\mu}$ are column vectors, $\vec{x}^T, \ \vec{\mu}^T$ are transpose (row) vectors,

$$E[x_i] = \mu_i, , \qquad \text{cov}[x_i, x_j] = V_{ij} .$$

For $n = 2$ this is

$$f(x_1, x_2, ; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}}$$

$$\times \exp\left\{-\frac{1}{2(1 - \rho^2)}\left[\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1 - \mu_1}{\sigma_1}\right)\left(\frac{x_2 - \mu_2}{\sigma_2}\right)\right]\right\}$$

where $\rho = \text{cov}[x_1, x_2]/(\sigma_1\sigma_2)$ is the correlation coefficient.

Chi-square (χ^2) distribution

The chi-square pdf for the continuous r.v. z ($z \geq 0$) is defined by

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

$n = 1, 2, \ldots =$ number of 'degrees of freedom' (dof)

$$E[z] = n, \quad V[z] = 2n.$$
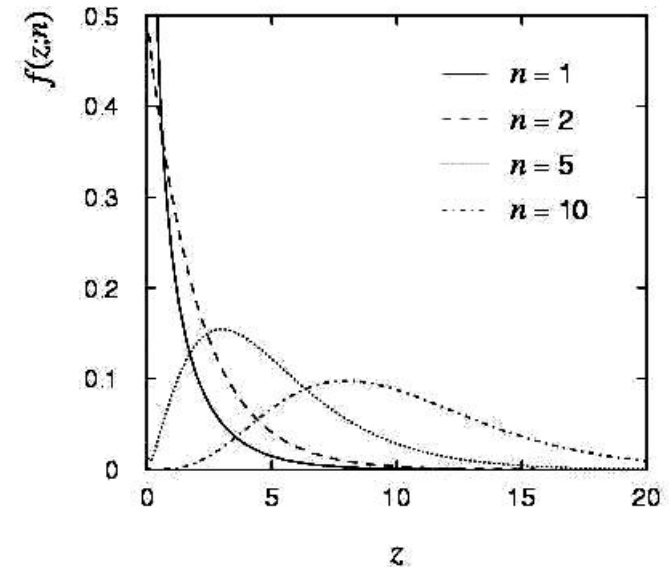
For independent Gaussian x_i, $i = 1, \ldots, n$, means μ_i, variances σ_i^2,

$$z = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2} \quad \text{follows } \chi^2 \text{ pdf with } n \text{ dof.}$$

Example: goodness-of-fit test variable especially in conjunction with method of least squares.

Cauchy (Breit-Wigner) distribution

The Breit-Wigner pdf for the continuous r.v. x is defined by
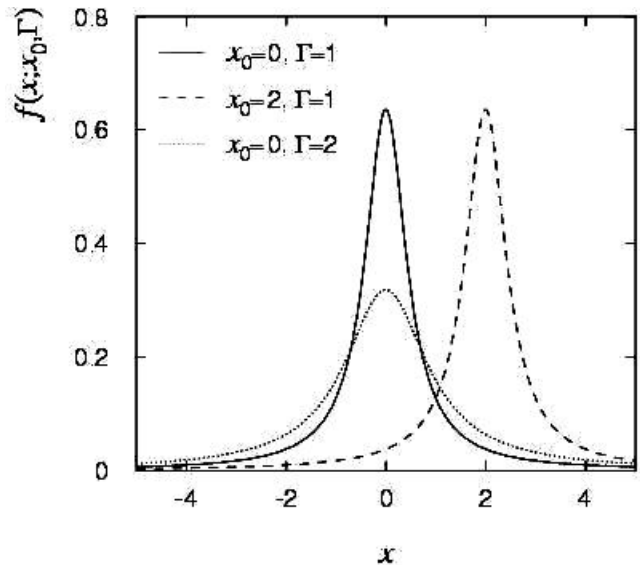
$$f(x; \Gamma, x_0) = \frac{1}{\pi} \frac{\Gamma/2}{\Gamma^2/4 + (x - x_0)^2}$$

($\Gamma = 2$, $x_0 = 0$ is the Cauchy pdf.)

$E[x]$ not well defined, $V[x] \to \infty$.

x_0 = mode (most probable value)

Γ = full width at half maximum

Example: mass of resonance particle, e.g. ρ, K^*, ϕ^0, ...

Γ = decay rate (inverse of mean lifetime)

Landau distribution

For a charged particle with $\beta = v/c$ traversing a layer of matter of thickness d, the energy loss Δ follows the Landau pdf:

$$f(\Delta; \beta) = \frac{1}{\xi}\phi(\lambda) \ ,$$

$$\phi(\lambda) = \frac{1}{\pi} \int_0^\infty \exp(-u \ln u - \lambda u) \sin \pi u \, du \ ,$$

$$\lambda = \frac{1}{\xi}\left[\Delta - \xi\left(\ln \frac{\xi}{\epsilon'} + 1 - \gamma_{\mathsf{E}}\right)\right] \ ,$$

$$\xi = \frac{2\pi N_{\mathsf{A}} e^4 z^2 \rho \sum Z}{m_{\mathsf{e}} c^2 \sum A}\frac{d}{\beta^2} \ , \qquad \epsilon' = \frac{I^2 \exp \beta^2}{2 m_{\mathsf{e}} c^2 \beta^2 \gamma^2} \ .$$
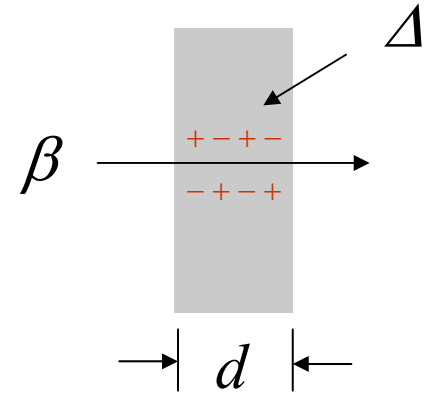
L. Landau, J. Phys. USSR **8** (1944) 201; see also
W. Allison and J. Cobb, Ann. Rev. Nucl. Part. Sci. **30** (1980) 253.

Landau distribution (2)
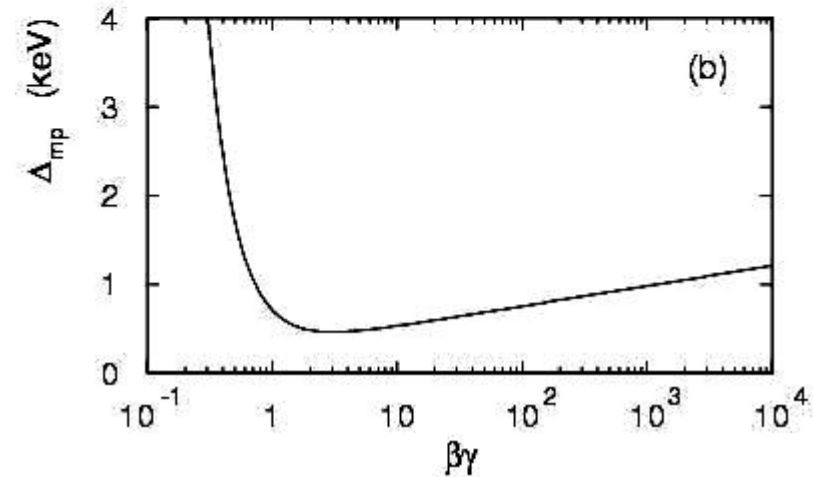
Long 'Landau tail'
→ all moments ∞

Mode (most probable value) sensitive to β,
→ particle i.d.
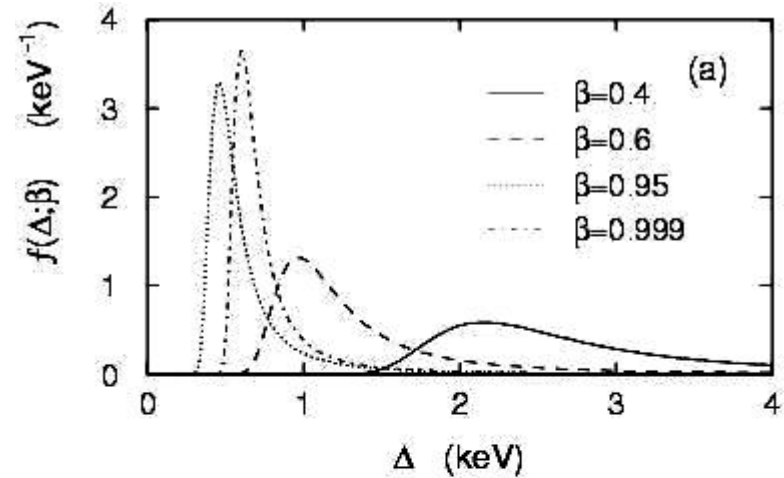
The Monte Carlo method

What it is: a numerical technique for calculating probabilities and related quantities using sequences of random numbers.

The usual steps:

$g(r)$

(1) Generate sequence r_1, r_2, ..., r_m uniform in [0, 1].

(2) Use this to produce another sequence x_1, x_2, ..., x_n distributed according to some pdf $f(x)$ in which we're interested (x can be a vector).

(3) Use the x values to estimate some property of $f(x)$, e.g., fraction of x values with $a < x < b$ gives $\int_a^b f(x)\, dx$.

$\rightarrow$ MC calculation = integration (at least formally)
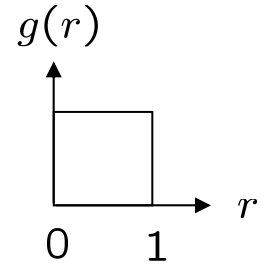
MC generated values = 'simulated data'
$\rightarrow$ use for testing statistical procedures

Random number generators

Goal: generate uniformly distributed values in [0, 1].

Toss coin for e.g. 32 bit number... (too tiring).

$\rightarrow$ 'random number generator'

= computer algorithm to generate r_1, r_2, ..., r_n.

Example: multiplicative linear congruential generator (MLCG)

$n_{i+1} = (a\, n_i) \bmod m$, where

n_i = integer

a = multiplier

m = modulus

n_0 = seed (initial value)

N.B. mod = modulus (remainder), e.g. 27 mod 5 = 2.

This rule produces a sequence of numbers n_0, n_1, ...

Random number generators (2)

The sequence is (unfortunately) periodic!

Example (see Brandt Ch 4): $a = 3$, $m = 7$, $n_0 = 1$

$$n_1 = (3 \cdot 1) \bmod 7 = 3$$

$$n_2 = (3 \cdot 3) \bmod 7 = 2$$

$$n_3 = (3 \cdot 2) \bmod 7 = 6$$

$$n_4 = (3 \cdot 6) \bmod 7 = 4$$

$$n_5 = (3 \cdot 4) \bmod 7 = 5$$

$$n_6 = (3 \cdot 5) \bmod 7 = 1 \qquad \leftarrow \text{ sequence repeats}$$

Choose *a, m* to obtain long period (maximum = $m - 1$); *m* usually close to the largest integer that can represented in the computer.

Only use a subset of a single period of the sequence.

Random number generators (3)

$r_i = n_i/m$ are in [0, 1] but are they 'random'?

Choose *a, m* so that the r_i pass various tests of randomness:

uniform distribution in [0, 1],

all values independent (no correlations between pairs),

e.g. L'Ecuyer, Commun. ACM **31** (1988) 742 suggests

a = 40692
m = 2147483399
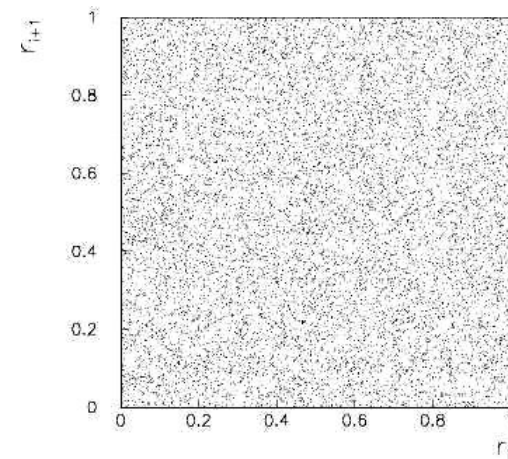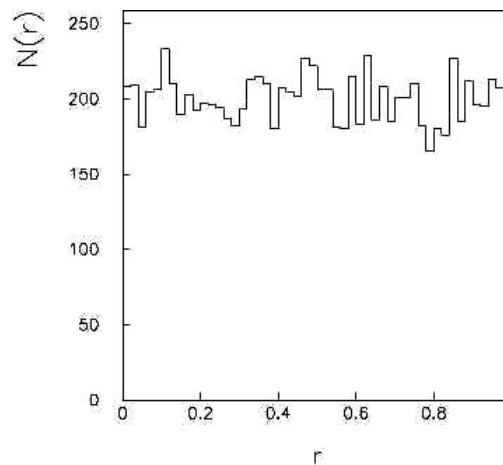
Far better algorithms available, e.g. `TRandom3`, period $2^{19937} - 1$.

See F. James, Comp. Phys. Comm. 60 (1990) 111; Brandt Ch. 4

The transformation method

Given $r_1, r_2,..., r_n$ uniform in [0, 1], find $x_1, x_2,..., x_n$
that follow $f(x)$ by finding a suitable transformation $x(r)$.
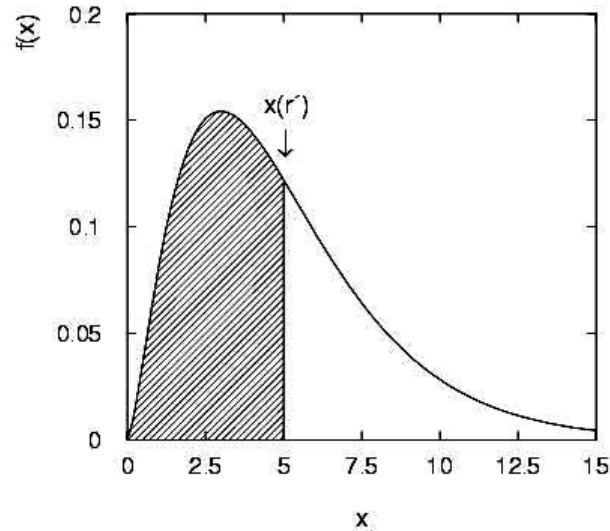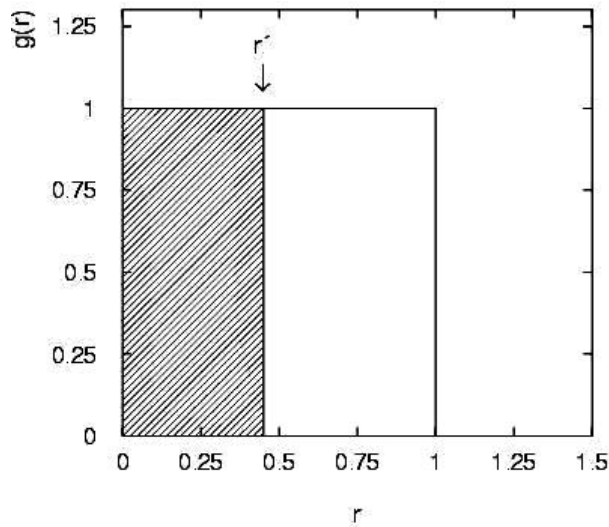
Require: $P(r \leq r') = P(x \leq x(r'))$

i.e. $\displaystyle\int_{-\infty}^{r'} g(r)\, dr = r' = \int_{-\infty}^{x(r')} f(x')\, dx' = F(x(r'))$

That is, $\boxed{\text{set } F(x) = r \text{ and solve for } x(r).}$

Example of the transformation method

Exponential pdf: $f(x; \xi) = \dfrac{1}{\xi} e^{-x/\xi}$ $(x \geq 0)$

Set $\displaystyle\int_0^x \frac{1}{\xi} e^{-x'/\xi}\, dx' = r$ and solve for $x\,(r)$.

$\rightarrow$ $x(r) = -\xi \ln(1 - r)$ ($x(r) = -\xi \ln r$ works too.)
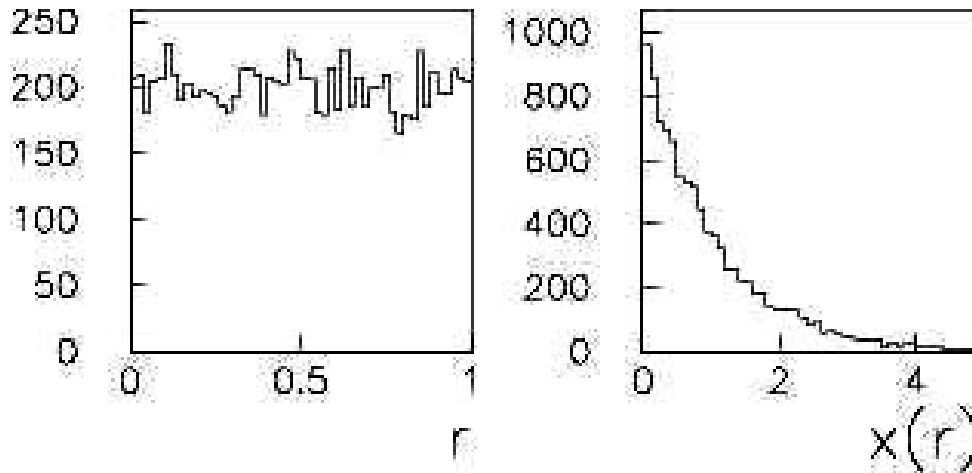
The acceptance-rejection method

Enclose the pdf in a box:
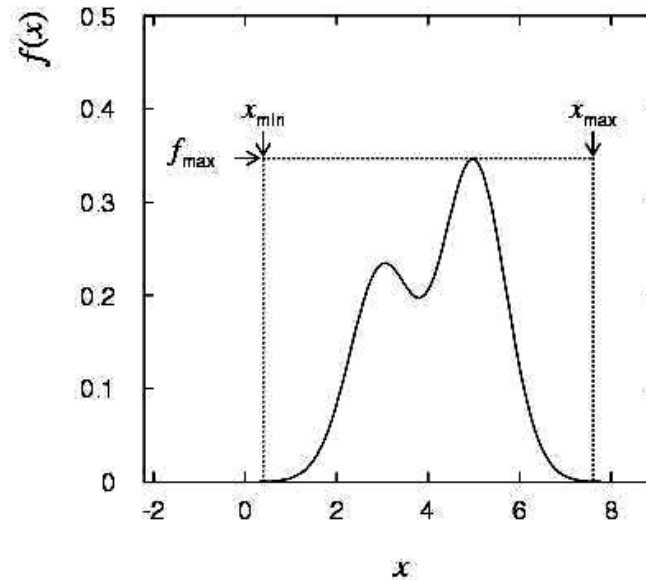
(1) Generate a random number x, uniform in $[x_{\min}, x_{\max}]$, i.e.
$x = x_{\min} + r_1(x_{\max} - x_{\min})$, r_1 is uniform in [0,1].

(2) Generate a 2nd independent random number u uniformly distributed between 0 and $f_{\max}$, i.e. $u = r_2 f_{\max}$.

(3) If $u < f(x)$, then accept x. If not, reject x and repeat.

Example with acceptance-rejection method

$$f(x) = \frac{3}{8}(1 + x^2)$$

$$(-1 \le x \le 1)$$

If dot below curve, use
x value in histogram.

S.Paktinat
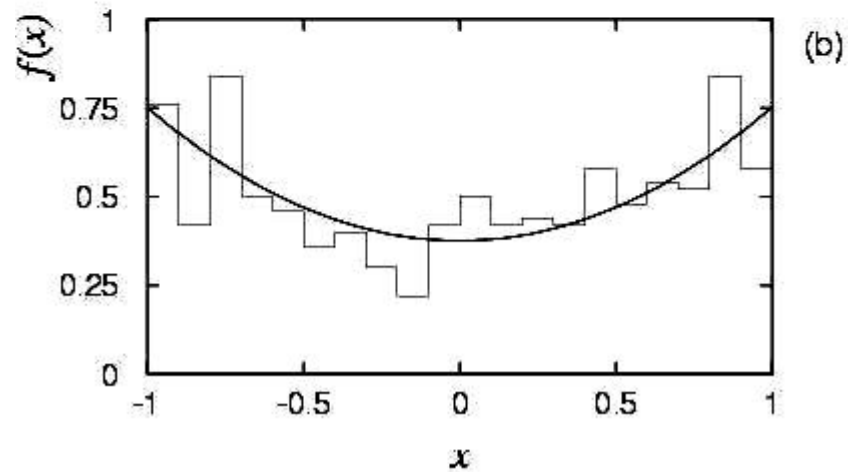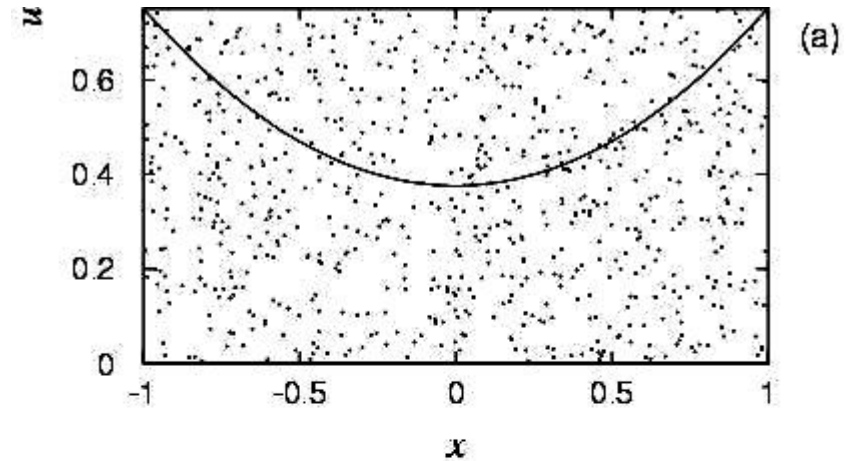
Monte Carlo event generators

Simple example: $e^+e^- \to \mu^+\mu^-$
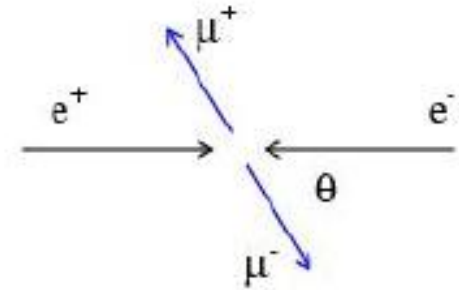
Generate $\cos\theta$ and ϕ:

$$f(\cos\theta; A_{\mathrm{FB}}) \propto (1 + \frac{8}{3}A_{\mathrm{FB}}\cos\theta + \cos^2\theta) \, ,$$

$$g(\phi) = \frac{1}{2\pi} \quad (0 \le \phi \le 2\pi)$$

Less simple: 'event generators' for a variety of reactions:

$e^+e^- \to \mu^+\mu^-$, hadrons, ...

$pp \to$ hadrons, D-Y, SUSY,...

e.g. PYTHIA, HERWIG, ISAJET...

Output = 'events', i.e., for each event we get a list of generated particles and their momentum vectors, types, etc.

A simulated event

⋮

Event listing (summary)

I	particle/jet	KS	KF	orig	p_x	p_y	p_z	E	m
1	!p+!	21	2212	0	0.000	0.000	7000.000	7000.000	0.938
2	!p+!	21	2212	0	0.000	0.000	-7000.000	7000.000	0.938
3	!g!	21	21	1	0.863	-0.323	1739.862	1739.862	0.000
4	!ubar!	21	-2	2	-0.621	-0.163	-777.415	777.415	0.000
5	!g!	21	21	3	-2.427	5.486	1487.857	1487	
6	!g!	21	21	4	-62.910	63.357	-463.274	471	
7	!~g!	21	1000021	0	314.363	544.843	498.897	979	
8	!~g!	21	1000021	0	-379.700	-476.000	525.686	980	
9	!~chi_1-!	21	-1000024	7	130.058	112.247	129.860	263	
10	!sbar!	21	-3	7	259.400	187.468	83.100	330	
11	!c!	21	4	7	-79.403	242.409	283.026	381	
12	!~chi_20!	21	1000023	8	-326.241	-80.971	113.712	385	
13	!b!	21	5	8	-51.841	-294.077	389.853	491	
14	!bbar!	21	-5	8	-0.597	-99.577	21.299	101	
15	!~chi_10!	21	1000022	9	103.352	81.316	83.457	175	
16	!s!	21	3	9	5.451	38.374	52.302	65	
17	!cbar!	21	-4	9	20.839	-7.250	-5.938	22	
18	!~chi_10!	21	1000022	12	-136.266	-72.961	53.246	181	
19	!nu_mu!	21	14	12	-78.263	-24.757	21.719	84	
20	!nu_mubar!	21	-14	12	-107.801	16.901	38.226	115	
21	gamma	1	22	4	2.636	1.357	0.125	2	
22	(~chi_1-)	11	-1000024	9	129.643	112.440	129.820	262	
23	(~chi_20)	11	1000023	12	-322.330	-80.817	113.191	382	
24	~chi_10	1	1000022	15	97.944	77.819	80.917	169	
25	~chi_10	1	1000022	18	-136.266	-72.961	53.246	181	
26	nu_mu	1	14	19	-78.263	-24.757	21.719	84	
27	nu_mubar	1	-14	20	-107.801	16.901	38.226	115	
28	(Delta++)	11	2224	2	0.222	0.012	-2734.287	2734	

					p_x	p_y	p_z	E	m
397	pi+	1	211	209	0.006	0.398	-308.296	308.297	0.140
398	gamma	1	22	211	0.407	0.087	-1695.458	1695.458	0.000
399	gamma	1	22	211	0.113	-0.029	-314.822	314.822	0.000
400	(pi0)	11	111	212	0.021	0.122	-103.709	103.709	0.135
401	(pi0)	11	111	212	0.084	-0.068	-94.276	94.276	0.135
402	(pi0)	11	111	212	0.267	-0.052	-144.673	144.674	0.135
403	gamma	1	22	215	-1.581	2.473	3.306	4.421	0.000
404	gamma	1	22	215	-1.494	2.143	3.051	4.016	0.000
405	pi-	1	-211	216	0.007	0.738	4.015	4.085	0.140
406	pi+	1	211	216	-0.024	0.293	0.486	0.585	0.140
407	K+	1	321	218	4.382	-1.412	-1.799	4.968	0.494
408	pi-	1	-211	218	1.183	-0.894	-0.176	1.500	0.140
409	(pi0)	11	111	218	0.955	-0.459	-0.590	1.221	0.135
410	(pi0)	11	111	218	2.349	-1.105	-1.181	2.855	0.135
411	(Kbar0)	11	-311	219	1.441	-0.247	-0.472	1.615	0.498
412	pi-	1	-211	219	2.232	-0.400	-0.249	2.285	0.140
413	K+	1	321	220	1.380	-0.652	-0.361	1.644	0.494
414	(pi0)	11	111	220	1.078	-0.265	0.175	1.132	0.135
415	(K_S0)	11	310	222	1.841	0.111	0.894	2.109	0.498
416	K+	1	321	223	0.307	0.107	0.252	0.642	0.494
417	pi-	1	-211	223	0.266	0.316	-0.201	0.480	0.140
418	nbar0	1	-2112	226	1.335	1.641	2.078	3.111	0.940
419	(pi0)	11	111	226	0.899	1.046	1.311	1.908	0.135
420	pi+	1	211	227	0.217	1.407	1.356	1.971	0.140
421	(pi0)	11	111	227	1.207	2.336	2.767	3.820	0.135
422	n0	1	2112	228	3.475	5.324	5.702	8.592	0.940
423	pi-	1	-211	228	1.856	2.606	2.808	4.259	0.140
424	gamma	1	22	229	-0.012	0.247	0.421	0.489	0.000
425	gamma	1	22	229	0.025	0.034	0.009	0.043	0.000
426	pi+	1	211	230	2.718	5.229	6.403	8.703	0.140
427	(pi0)	11	111	230	4.109	6.747	7.597	10.961	0.135
428	pi-	1	-211	231	0.551	1.233	1.945	2.372	0.140
429	(pi0)	11	111	231	0.645	1.141	0.922	1.608	0.135
430	gamma	1	22	232	-0.383	1.169	1.208	1.724	0.000
431	gamma	1	22	232	-0.201	0.070	0.060	0.221	0.000

⋮

PYTHIA Monte Carlo
pp → gluino-gluino

S.Paktinat

Monte Carlo detector simulation

Takes as input the particle list and momenta from generator.

Simulates detector response:
multiple Coulomb scattering (generate scattering angle),
particle decays (generate lifetime),
ionization energy loss (generate Δ),
electromagnetic, hadronic showers,
production of signals, electronics response, ...

Output = simulated raw data $\rightarrow$ input to reconstruction software:
track finding, fitting, etc.

Predict what you should see at 'detector level' given a certain hypothesis for 'generator level'. Compare with the real data.

Estimate 'efficiencies' = #events found / # events generated.

Programming package: **GEANT**