Introduction to Statistical Methods for High Energy Physics

2004 CERN Summer Student Lectures

Glen Cowan
Physics Department
Royal Holloway, University of London
g.cowan@rhul.ac.uk
www.pp.rhul.ac.uk/~cowan

- CERN course web page:

  www.pp.rhul.ac.uk/~cowan/stat_cern

- See also University of London course web page:

  www.pp.rhul.ac.uk/~cowan/stat_course

Lecture 1

1. **Probability**
2. **Random variables, probability densities, etc.**
3. **Brief catalogue of probability densities**
4. **The Monte Carlo method**

Lecture 2

1. **Statistical tests**
2. **Fisher discriminants, neural networks, etc.**
3. **Goodness-of-fit tests**
4. **The significance of a signal**
5. **Introduction to parameter estimation**

Lecture 3

1. **The method of maximum likelihood (ML)**
2. **Variance of ML estimators**
3. **The method of least squares (LS)**
4. **Interval estimation, setting limits**

Some statistics books, papers, etc.

G. Cowan, *Statistical Data Analysis*, Clarendon, Oxford, 1998

see also `www.pp.rhul.ac.uk/~cowan/sda`

R.J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, Wiley, 1989

see also `hepwww.ph.man.ac.uk/~roger/book.html`

L. Lyons, *Statistics for Nuclear and Particle Physics*, CUP, 1986

W. Eadie et al., *Statistical Methods in Experimental Physics*, North-Holland, 1971

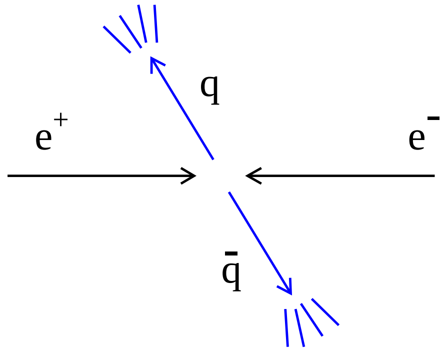
S. Brandt, *Statistical and Computational Methods in Data Analysis*, Springer, New York, 1998

with FORTRAN and C program library

S. Eidelman et al. (Particle Data Group), *Review of Particle Physics,* Physics Letters B592 (2004) 1; see also `pdg.lbl.gov`.

sections on probability, statistics, Monte Carlo

Data analysis in particle physics



Observe n events of a certain type

Measure characteristics of each event (angles, event shapes particle multiplicity, number found for a given $\int L dt, \ldots$)

Theories (e.g. SM) predict distributions of these properties up to free parameters, e.g. α, G_{F}, M_{Z}, α_{s}, $m_{\mathrm{H}}, \ldots$

Some tasks of statistical data analysis:

Estimate the parameters.

Quantify the uncertainty of the parameter estimates.

Test to what extent the predictions of a theory are in agreement with the data.

There are various elements of uncertainty :

theory is not deterministic,

random measurement errors,

things we could know in principle but don't,...

$\rightarrow$ quantify using PROBABILITY

Definition of probability

Consider a set S with subsets $A, B, \ldots$

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 one 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 $P(A) \leq P(B)$

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

Also define conditional probability of A given B (with $P(B) \neq 0$) as

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

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

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

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

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

(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

→ Data analysis in HEP: frequency interperation often most natural, but subjective probability has some attractive features, e.g. more natural treatment of phenomena that are not repeatable:

Systematic errors (same upon repetition of experiment)

The particle in this event was a positron

Nature is supersymmetric

Billionth digit of π is 7

It will rain tomorrow (uncertain future event)

It rained in Cairo on March 8, 1587 (uncertain past event)

Bayes' theorem

From the definition of conditional probability,

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

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

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

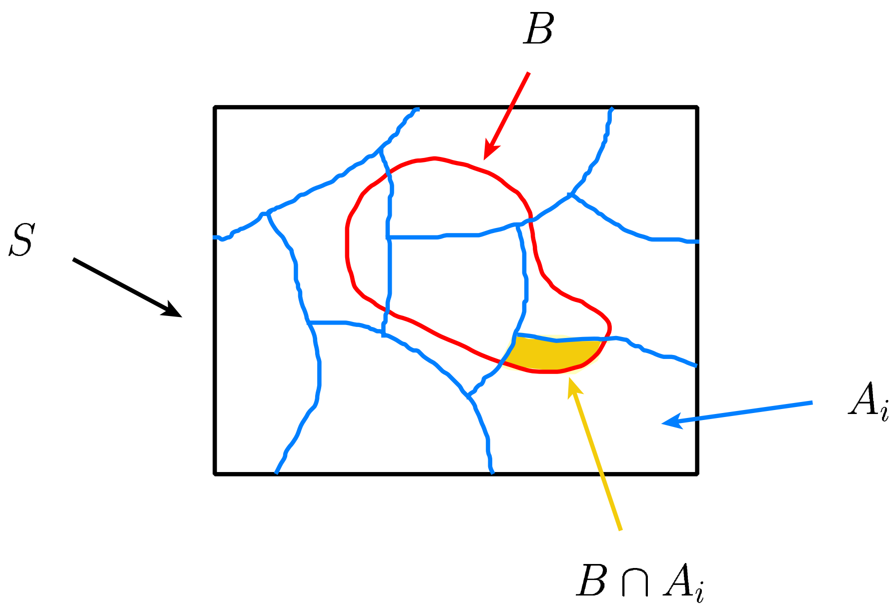
Bayes' theorem

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 $\cup_i A_i = S$,

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

$\rightarrow P(B) = P(\cup_i (B \cap A_i)) = \Sigma_i P(B \cap A_i)$ (since $B \cap A_i$ disjoint)

$\rightarrow P(B) = \Sigma_i \, P(B|A_i) \, P(A_i)$ (law of total probability)

Bayes' theorem becomes

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

An example using Bayes' theorem

Suppose the probabilities (for anyone) to have AIDS are:

$P(\text{AIDS}) = 0.001$ ← prior probabilities, i.e.

$P(\text{no AIDS}) = 0.999$ before any test carried out

Consider an AIDS test: result is + or −

$P(+|\text{AIDS}) = 0.98$ ← probabilities to (in)correctly

$P(-|\text{AIDS}) = 0.02$ identify AIDS infected person

$P(+|\text{no AIDS}) = 0.03$ ← probabilities to (in)correctly

$P(-|\text{no AIDS}) = 0.97$ identify person without AIDS

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

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

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

$$= 0.032$$ ← 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 guy will have AIDS

Random variables

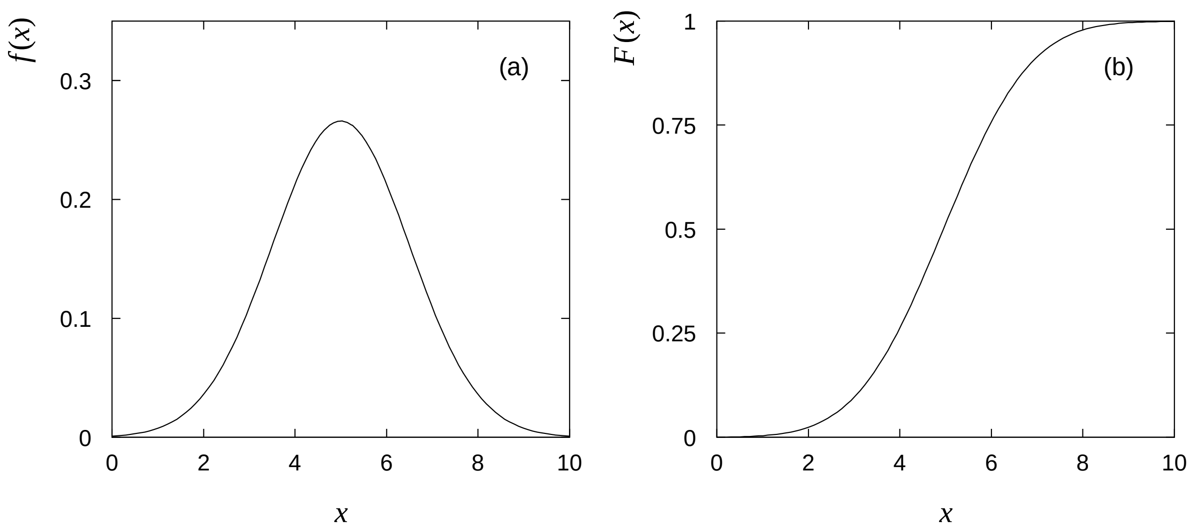
Suppose outcome of experiment is x (label for element of sample space)

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

$\rightarrow f(x) =$ probability density function (pdf)

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

$$F(x) = \int_{-\infty}^{x} f(x')\,dx' \qquad \leftarrow \text{cumulative distribution function}$$



For discrete case:

$$f_i = P(x_i)$$

$$\sum_i f_i = 1$$

$$F(x) = \sum_{x_i \le x} P(x_i)$$

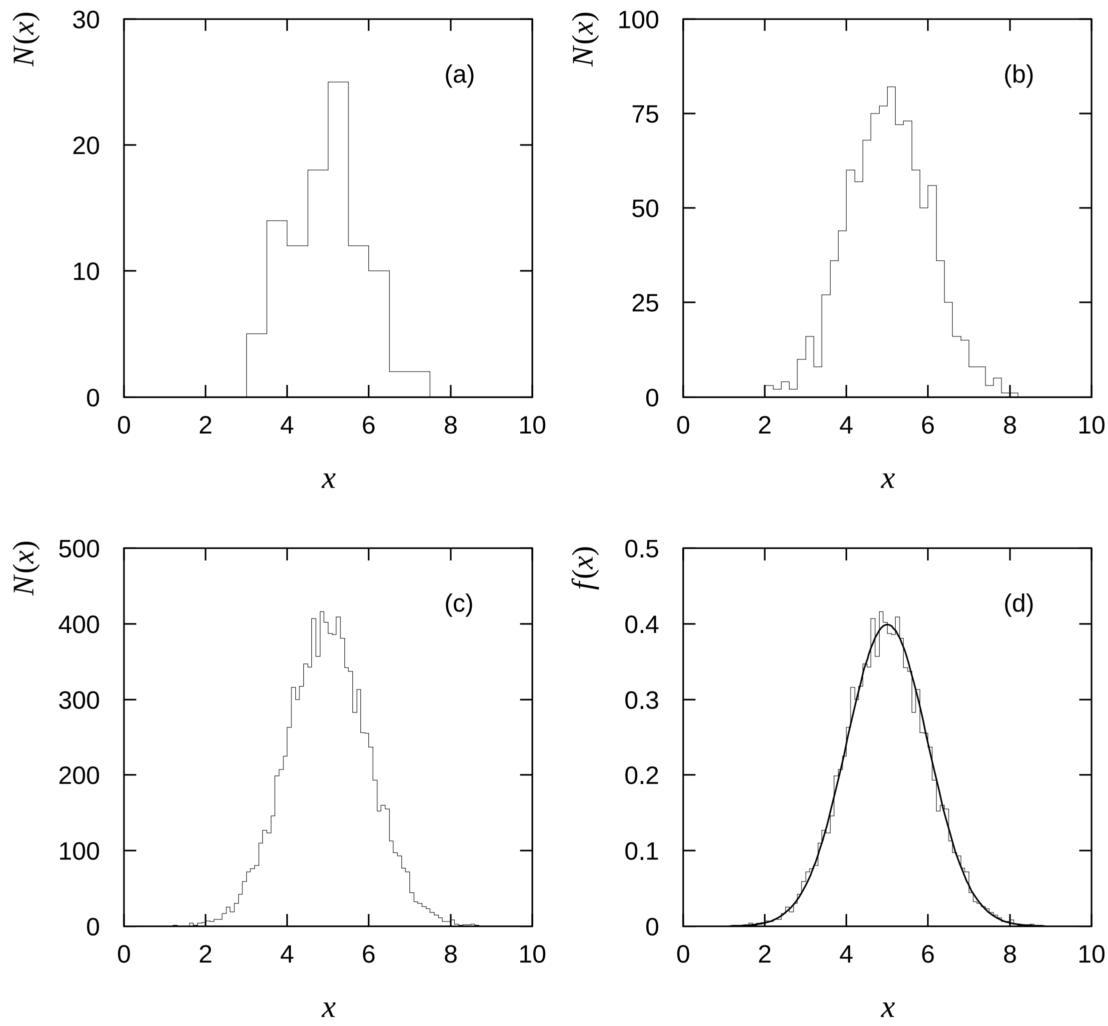
Histograms

pdf = histogram with:

infinite data sample

zero bin width

normalized to unit area



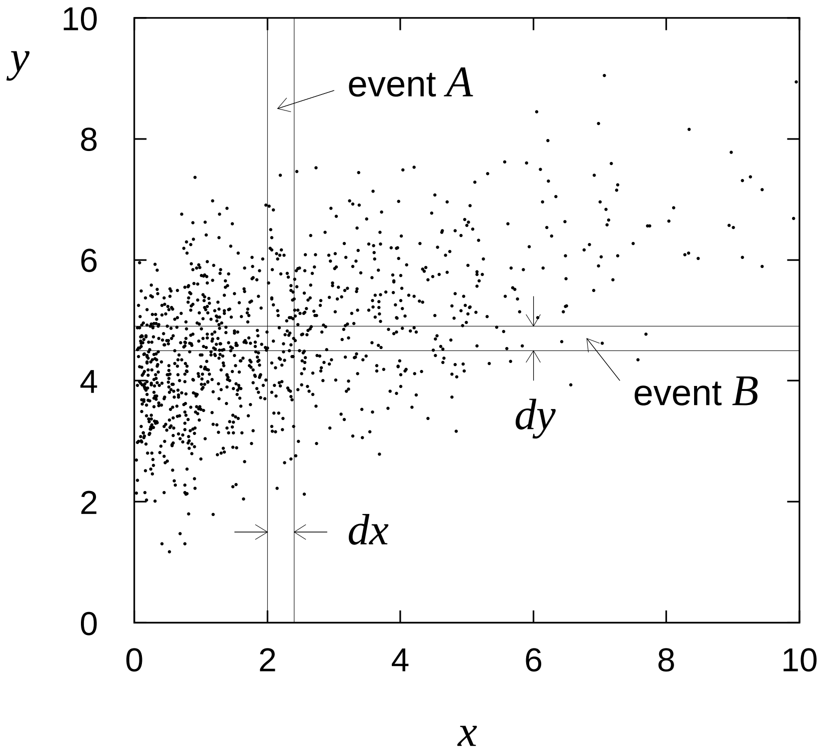
$$f(x) = \frac{N(x)}{n\Delta x}$$

n = number of entries

Δx = bin width

Multivariate case

Outcome characterized by > 1 quantity, e.g. x and y



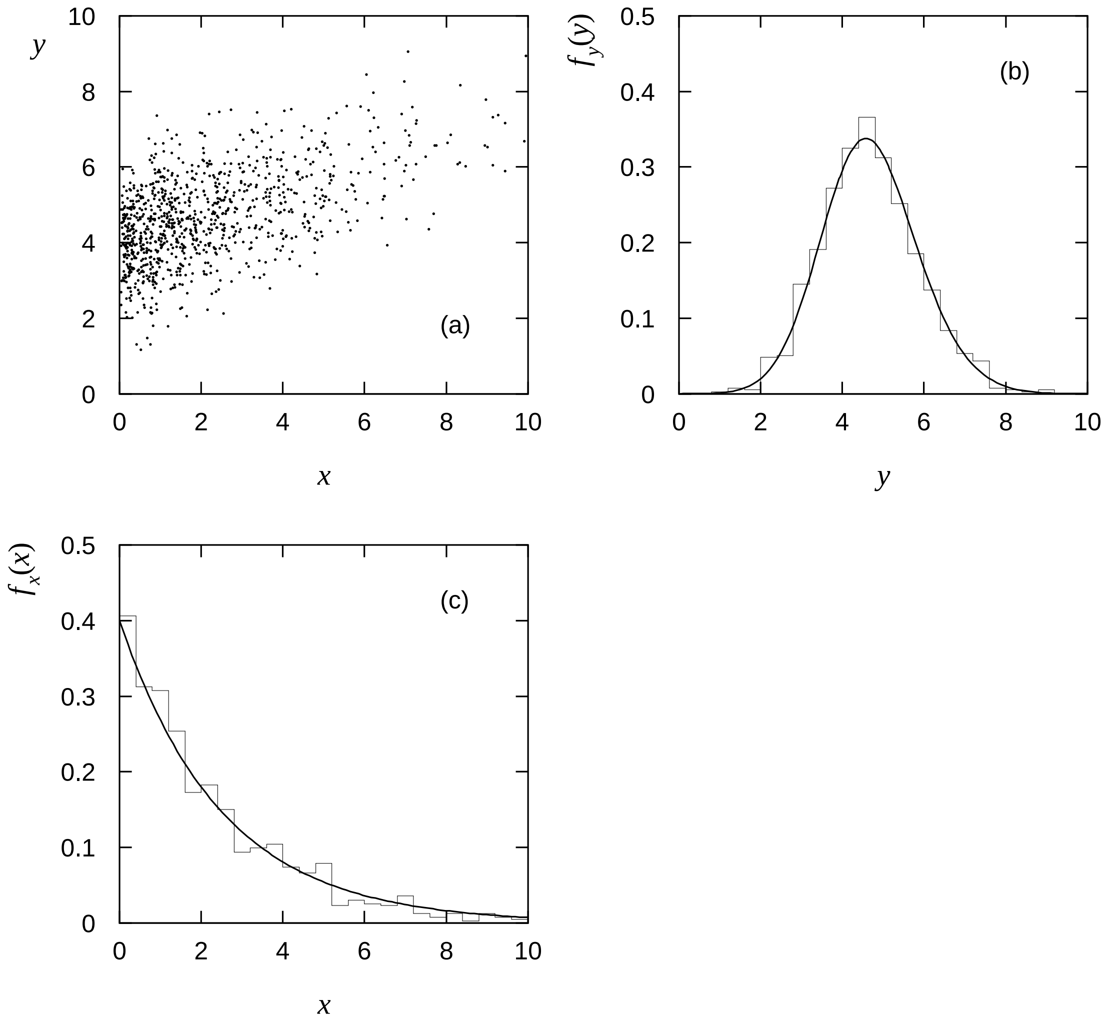
$$P(A \cap B) = f(x,y)\,dx\,dy$$

$\rightarrow f(x,y) =$ joint pdf

$$\int\int f(x,y)\,dx\,dy = 1$$

Marginal distributions

Projections of joint pdf (scatter plot) onto x, y axes:



$$f_x(x) = \int f(x,y)\,dy$$

$$f_y(y) = \int f(x,y)\,dx$$

$\rightarrow$ $f_x(x),\ f_y(y)$ = marginal pdfs

Conditional pdf

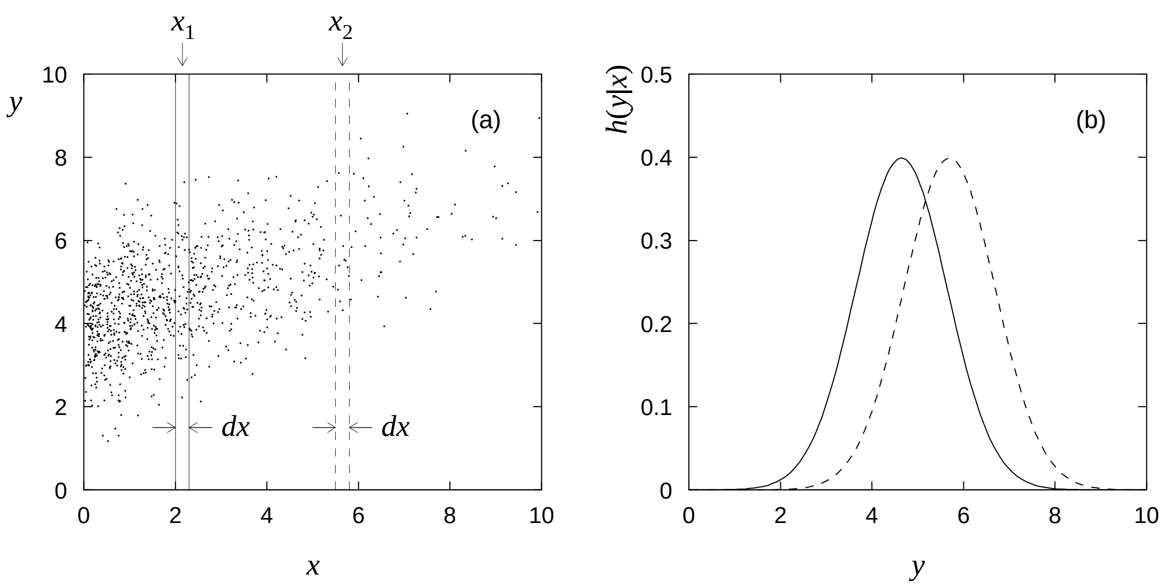
Recall conditional probability:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{f(x,y)\,dx\,dy}{f_x(x)\,dx}$$

Define $h(y|x) = \dfrac{f(x,y)}{f_x(x)}$

$g(x|y) = \dfrac{f(x,y)}{f_y(y)}$

← conditional pdfs



Bayes' theorem becomes

$$g(x|y) = \frac{h(y|x) f_x(x)}{f_y(y)}$$

Recall A, B independent if $P(A \cap B) = P(A)P(B)$

$\Rightarrow$ x, y independent if $f(x,y) = f_x(x) f_y(y)$

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

Define the expectation (mean) value as:

$$E[x] = \int x\, f(x)\, dx$$

N.B. $E[x]$ is not a function of x , rather a parameter of $f(x)$.

Notation (often): $E[x] = \mu$

For discrete variable, $E[x] = \sum_i x_i P(x_i)$

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:

μ

$$V[x] = E\left[(x - E[x])^2\right] = E[x^2] - \mu^2$$

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

Standard deviation: $\sigma \equiv \sqrt{\sigma^2}$ (same dimension as x)

Algebraic moments: $E[x^n] = \mu'_n \quad (\mu'_1 = \mu)$.

Central moments: $E\left[(x - \mu)^n\right] \equiv \mu_n \quad (\sigma^2 = \mu_2)$

Covariance and correlation

Define the covariance $\mathrm{cov}[x, y]$ (also use matrix notation V_{xy}) as

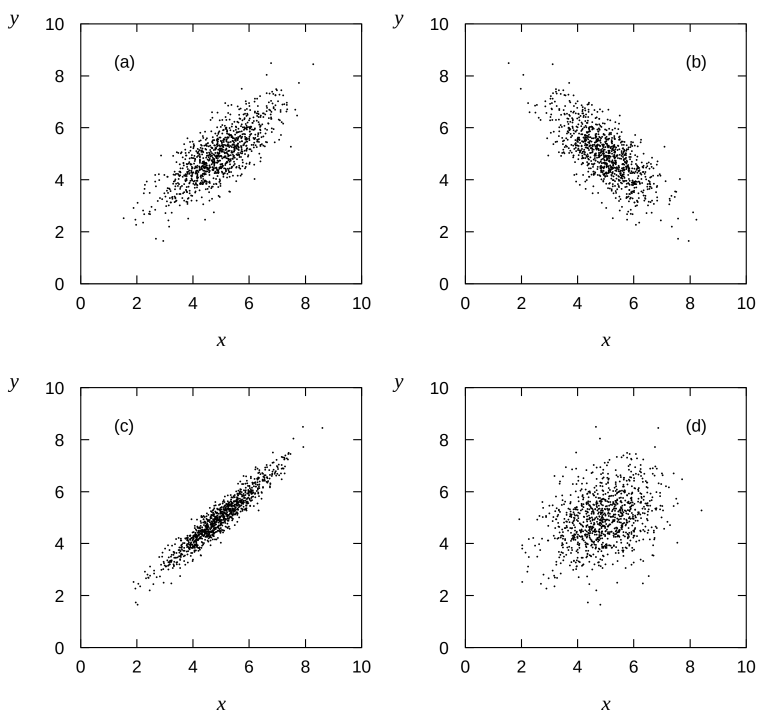
$$\mathrm{cov}[x, y] = E[(x - \mu_x)(y - \mu_y)] = E[xy] - \mu_x \mu_y$$

Correlation coefficient (dimensionless) defined as

$$\rho_{xy} = \frac{\mathrm{cov}[x, y]}{\sigma_x \sigma_y}\,, \qquad -1 \leq \rho_{xy} \leq 1$$

$\rho = 0.75$ $\rho = -0.75$

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



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

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

$\Rightarrow$ $\mathrm{cov}[x, y] = 0$ x and y ‘uncorrelated’

N.B. converse not always true.

Error propagation

Suppose $\vec{x} = (x_1, \ldots, x_n)$ follows some joint pdf $f(\vec{x})$.

$f(\vec{x})$ maybe not fully known, but suppose we have covariances

$$V_{ij} = \mathrm{cov}[x_i, x_j]$$

and the means $\vec{\mu} = E[\vec{x}]$ (in practice only estimates).

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

What is the variance $V[y] = E[y^2] - (E[y])^2$?

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

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

We need $E[y]$ and $E[y^2]$. These are:

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

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

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

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

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

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

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

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

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

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

or in matrix notation, $U = A\,V\,A^T$, where $A_{ij} = \left[\frac{\partial y_i}{\partial x_j} \right]_{\vec{x}=\vec{\mu}}$.

These are the ‘error propagation’ formulae, i.e. the covariances, which summarize the ‘errors’ in measurements of $\vec{x}$, are propagated to the new quantities $\vec{y}(\vec{x})$.

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: some special cases

$$y = x_1 + x_2$$

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

$$y = x_1 x_2$$

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

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!

Consider e.g. $y = x_1 - x_2$, with

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

Then $E[y] = \mu_1 - \mu_2 = 0$ and $V[y] = 1^2 + 1^2 = 2$,

i.e. $\sigma_y = 1.4$.

Now suppose $\rho = 1$. Then

$$V[y] = 1^2 + 1^2 - 2 = 0, \quad \text{i.e. } \sigma_y = 0.$$

i.e. for $\rho \to 1$, error in difference $\to 0$.

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 $(0 \le n \le N)$.

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.

The binomial distribution is thus

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

random variable ↗ ↖ parameters

We can show

$$\sum_{n=0}^{N} \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} = 1$$

as required.

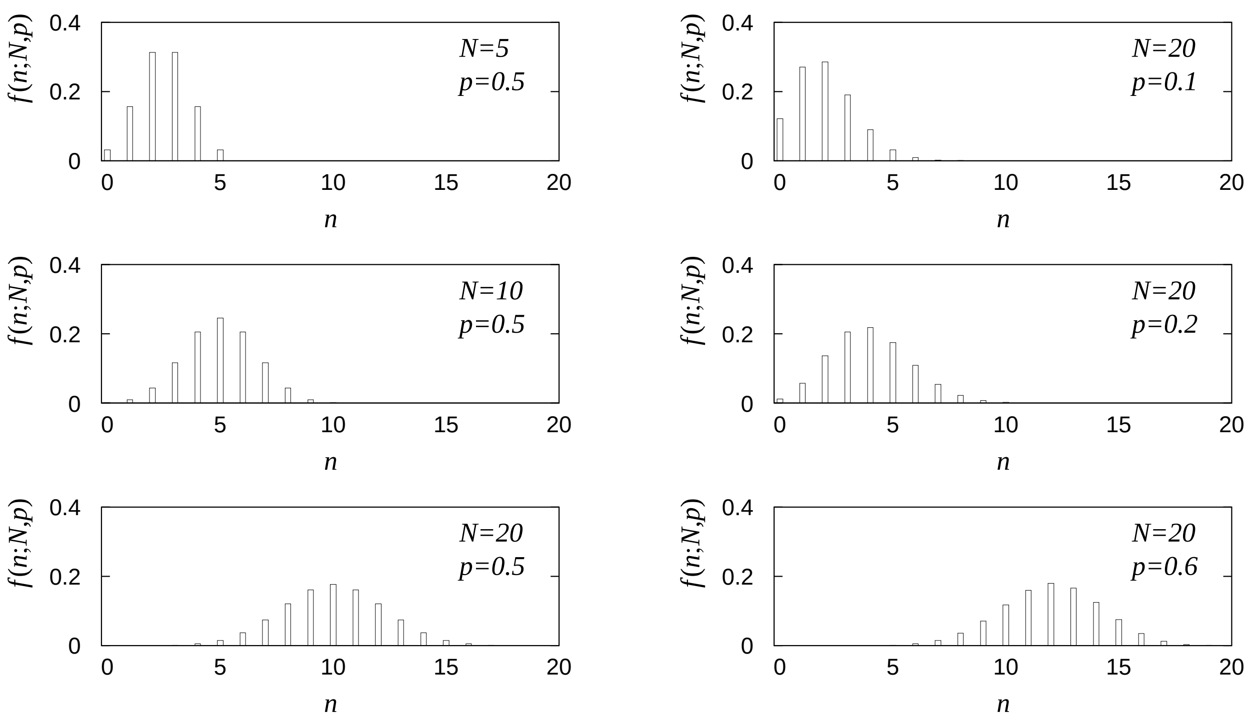
Binomial distribution (continued)

For expectation value and variance we obtain:

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

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

Recall $E[n]$, $V[n]$ are not random variables, but are constants which depend on the true (and possibly unknown) parameters N and p.



Example: observe N decays of $W^{\pm}$,

number n which are $W \to \mu\nu$ is a binomial r.v.,

p = branching ratio

Poisson distribution

Consider binomial n in the limit

$$N \to \infty,$$

$$p \to 0,$$

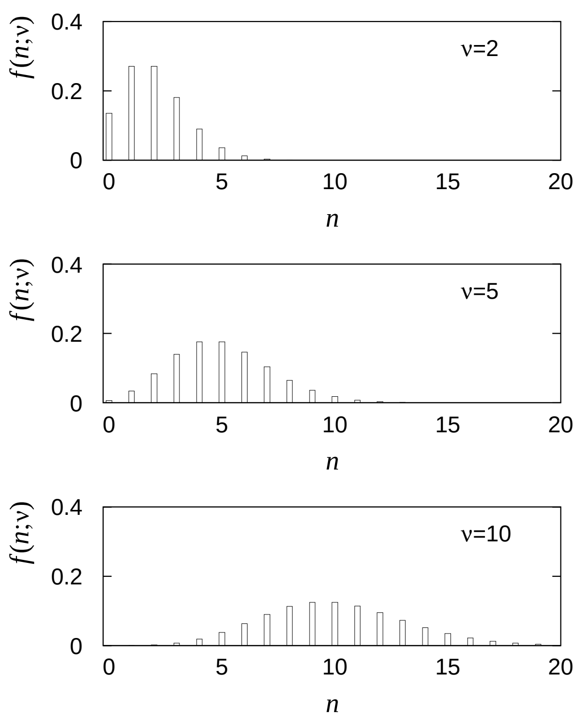
$$E[n] = Np \to \nu.$$

We can show that n then follows the Poisson distribution:

$$f(n;\nu) = \frac{\nu^n}{n!} e^{-\nu} \qquad (0 \le n < \infty)$$

$$E[n] = \nu$$

$$V[n] = \nu$$



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

Uniform distribution

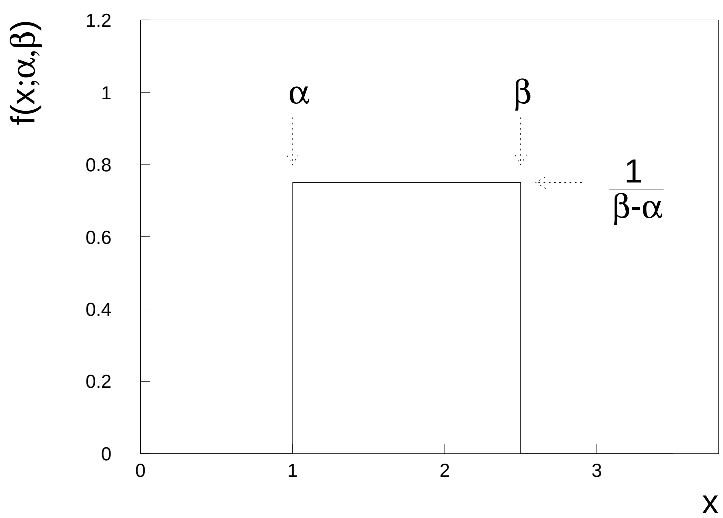
Consider a continuous r.v. x with $-\infty < x < \infty$.

The uniform distribution is defined by

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

$$E[x] = \int_\alpha^\beta \frac{x}{\beta-\alpha}\,dx = \tfrac{1}{2}(\alpha+\beta)$$

$$V[x] = \int_\alpha^\beta [x - \tfrac{1}{2}(\alpha+\beta)]^2 \frac{1}{\beta-\alpha}\,dx = \tfrac{1}{12}(\beta-\alpha)^2$$



N.B. For any r.v. x with cumulative distribution $F(x)$,

$$y = F(x) \text{ is uniform in } [0,1].$$

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

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

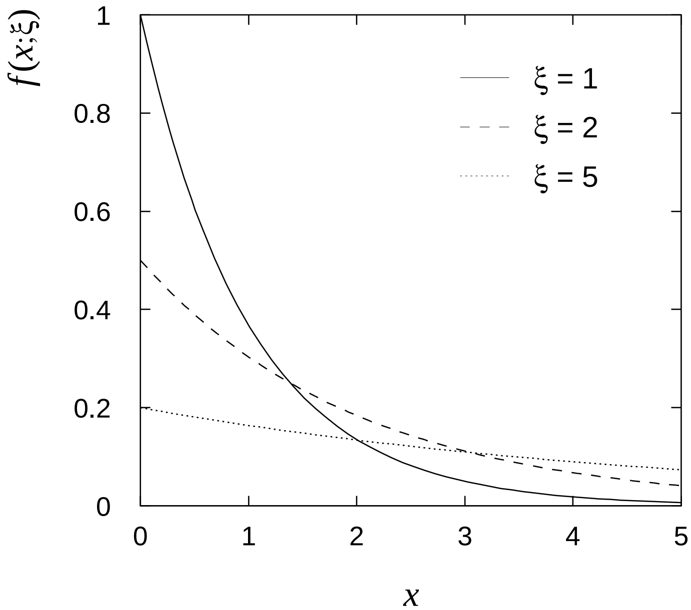
Exponential distribution

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

$$f(x;\xi) = \frac{1}{\xi} e^{-x/\xi} \qquad (x \geq 0)$$

$$E[x] = \int_0^\infty x \frac{1}{\xi} e^{-x/\xi}\, dx = \xi$$

$$V[x] = \int_0^\infty (x-\xi)^2 \frac{1}{\xi} e^{-x/\xi}\, dx = \xi^2$$



Example: proper decay time t of an unstable particle,

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

Lack of memory (unique to exponential pdf):

$$f(t - t_0 | t \geq t_0) = f(t)$$

Gaussian distribution

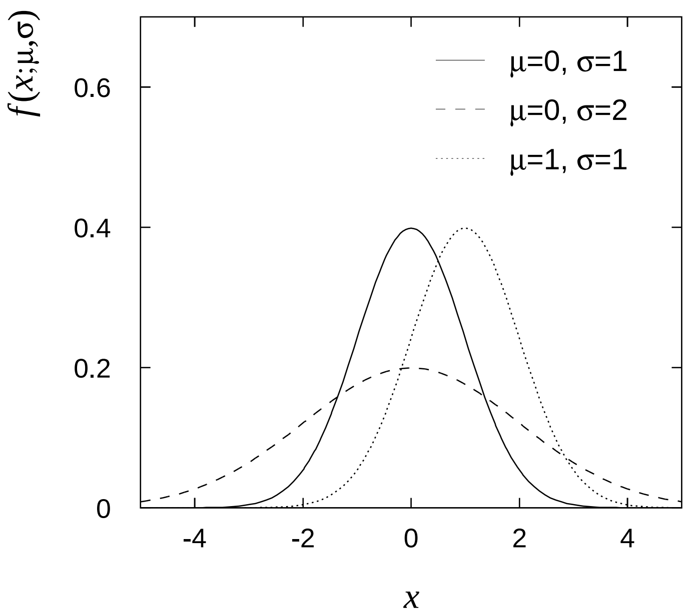
The Gaussian (or normal) pdf for the continuous r.v. x is defined by

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

$E[x] = \mu$

$V[x] = \sigma^2$

N.B. Often μ, σ^2 denote mean, variance of any r.v., not necessarily Gaussian.



Special case: $\mu = 0$, $\sigma^2 = 1$ ('standard Gaussian')

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

If y is Gaussian with μ, σ^2, then $x = \dfrac{y-\mu}{\sigma}$ follows $\varphi(x)$.

Examples: (almost) anything which is a sum of many random contributions, often the case for measurement errors.

The central limit theorem

For n independent r.v.s x_i with finite variances σ_i^2, otherwise arbitrary pdfs, in limit $n \to \infty$, $y = \sum_{i=1}^{n} x_i$ is a Gaussian r.v.

$$E[y] = \sum_{i=1}^{n} \mu_i$$

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

(As for all sums of independent r.v.s.)

For proof see e.g. GDC Ch. 10 using characteristic functions.

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

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 r.v. $\vec{x} = (x_1, \ldots, x_n)$:

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

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

$$E[x_i] = \mu_i$$

$$\text{cov}[x_i, x_j] = V_{ij}$$

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

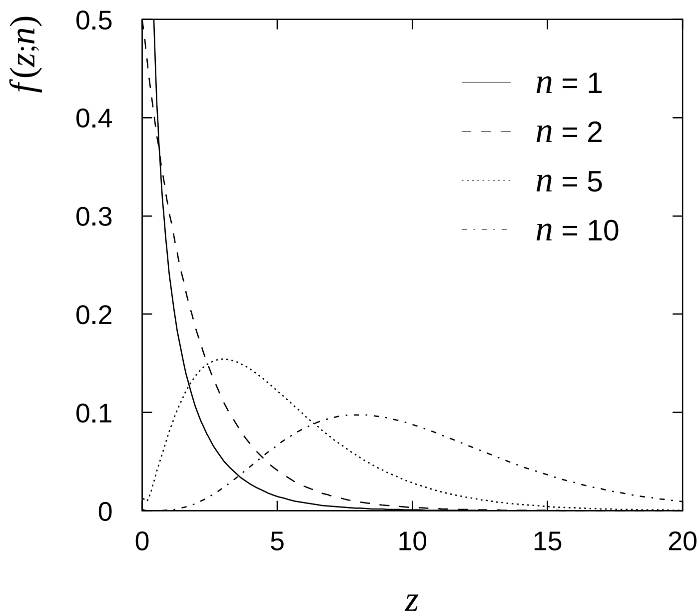
The chi-square pdf for the continuous r.v. z is defined by

$$f(z;n) = \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2} \qquad (z \geq 0)$$

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

$E[z] = n$

$V[z] = 2n$



For independent Gaussian x_i, $i = 1, \ldots, n$, means μ_i, variances σ_i^2,

$$z = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2}$$ follows χ^2 distribution with n dof.

Or for multivariate Gaussian x_i with covariance matrix V_{ij},

$$z = (\vec{x} - \vec{\mu})^T V^{-1} (\vec{x} - \vec{\mu})$$ follows χ^2 pdf.

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

Cauchy (Breit-Wigner) distribution

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

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

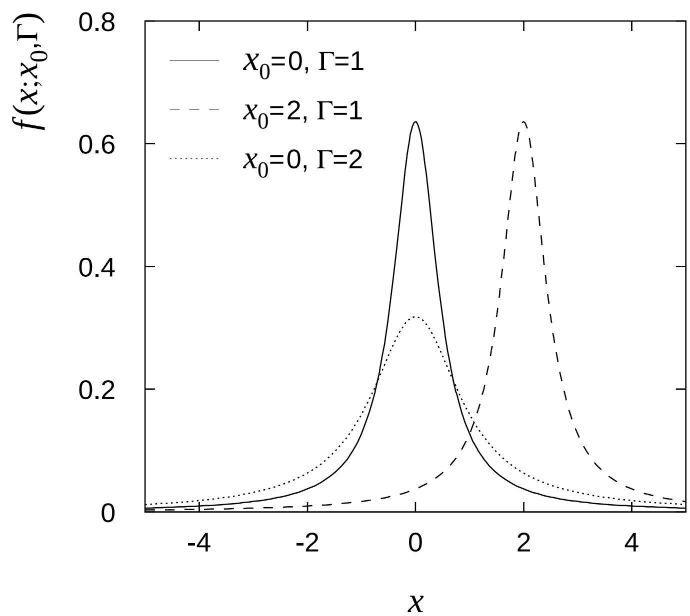
This is a special case of the Breit-Wigner pdf,

$$f(x; \Gamma, x_0) = \frac{1}{\pi} \frac{\Gamma/2}{\Gamma^2/4 + (x - x_0)^2},$$

where parameters x_0, Γ = mass, width of resonance.

$E[x]$ = not well defined

$V[x] = \infty$



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:

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

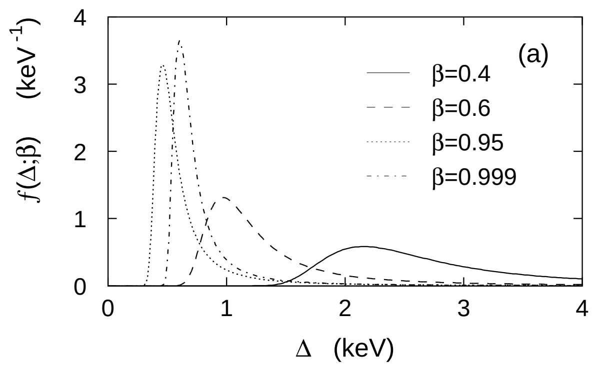
$$\phi(\lambda) = \frac{1}{\pi}\int_0^\infty \exp(-u\log u - \lambda u)\,\sin \pi u\,du,$$

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

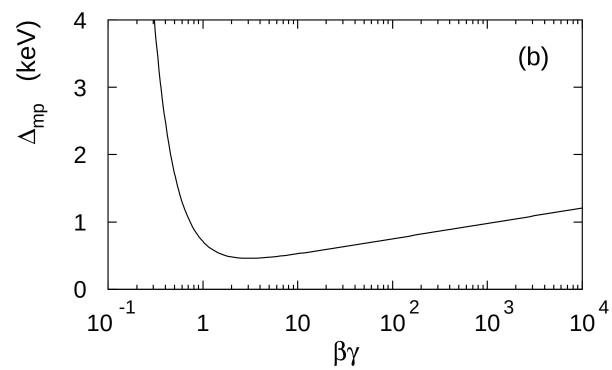
$$\xi = \frac{2\pi N_{\rm A} e^4 z^2 \rho \,\Sigma\, Z}{m_{\rm e}c^2 \,\Sigma\, A}\,\frac{d}{\beta^2}, \qquad \epsilon' = \frac{I^2 \exp(\beta^2)}{2m_{\rm e}c^2\beta^2\gamma^2}$$

(See L. Landau, *J. Phys. USSR* 8 (1944) 201;
W. Allison and J. Cobb, *Ann. Rev. Nucl. Part. Sci.* 30 (1980) 253.)

Long 'Landau tail'
⇒ all moments diverge



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:

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

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

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

$\Rightarrow$ MC calculation = integration (at least formally)

Usually trivial for 1-d: $\int_a^b f(x)\, dx$ obtainable by other methods.

MC more powerful for multidimensional integrals.

MC x values = 'simulated data'

$\rightarrow$ use for testing e.g. statistical procedures.

Random number generators

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

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

$\Rightarrow$ 'random number generator'

= computer algorithm to generate $r_1, r_2, \ldots, r_n$.

Example: the multiplicative linear congruential generator (MLCG)

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

n_i = integer

a = multiplier

m = modulus

n_0 = seed

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

The n_i follow periodic sequence in $[1, m-1]$.

Example (cf. Brandt): $a = 3$, $m = 7$, $n_0 = 1$:

$$
\begin{aligned}
n_1 &= (3 \cdot 1) \bmod 7 = 3 \\
n_2 &= (3 \cdot 3) \bmod 7 = 2 \\
n_3 &= (3 \cdot 2) \bmod 7 = 6 \\
n_4 &= (3 \cdot 6) \bmod 7 = 4 \\
n_5 &= (3 \cdot 4) \bmod 7 = 5 \\
n_6 &= (3 \cdot 5) \bmod 7 = 1 \leftarrow \text{sequence repeats!}
\end{aligned}
$$

Choose a, m, to obtain long period (maximum = $m - 1$).

$r_i = \frac{n_i}{m}$ are in $[0, 1]$ (0 and 1 excluded), but are they 'random'???

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

Uniform distribution in $[0, 1]$

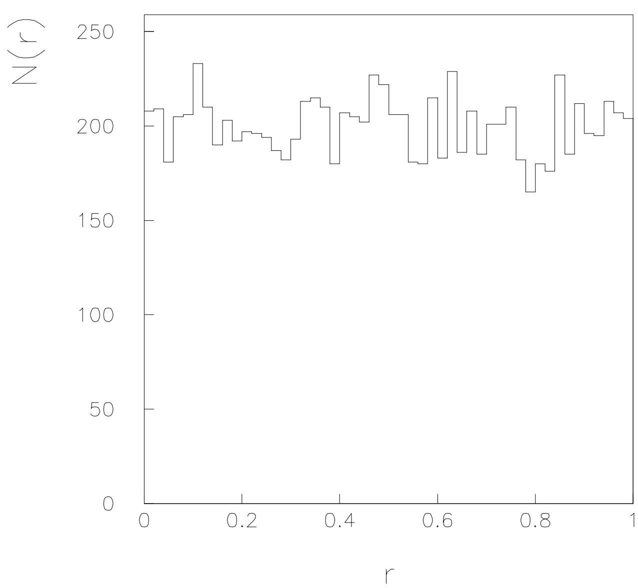
All pairs independent (no correlations)

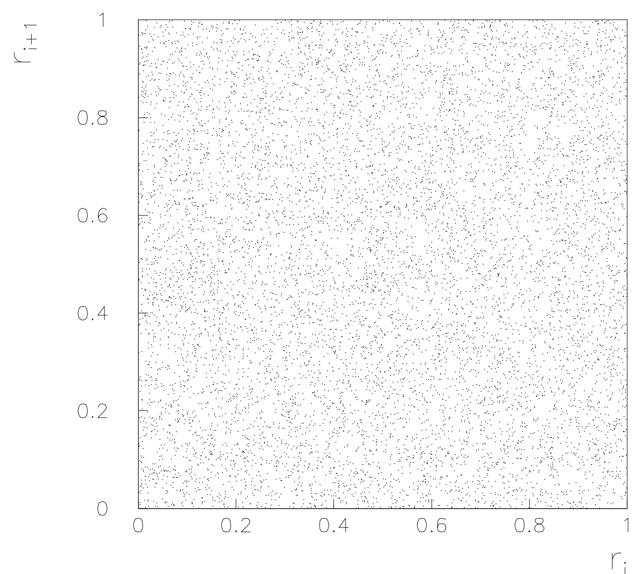
e.g. L'Ecuyer, Commun. ACM 31 (1988) 742 suggests

$a = 40692$

$m = 2147483399$

Test with 10000 generated values:





Far better algorithms available e.g. **`RANMAR`**, period $\approx 2 \times 10^{43}$.

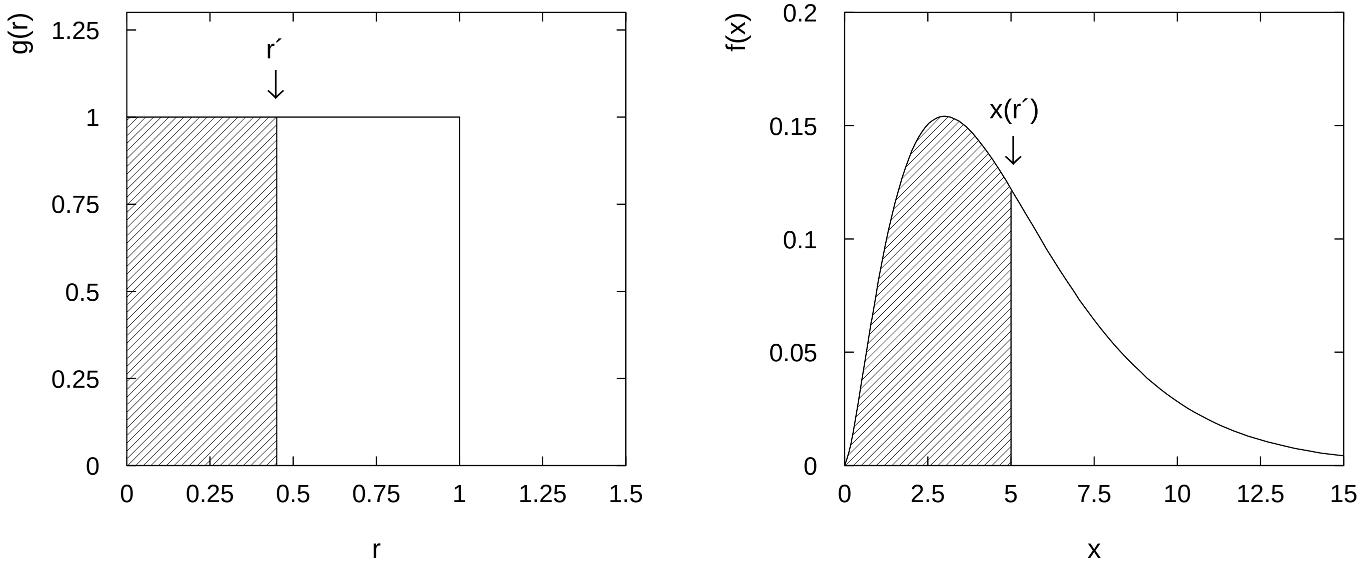
For more info see e.g.

F. James, Comput. Phys. Commun. 60 (1990) 111;

Brandt, chapter 4.

The transformation method

Given $r_1, r_2, \ldots, r_n$ uniform in $[0, 1]$, find $x_1, x_2, \ldots, x_n$ which follow $f(x)$ by finding a suitable transformation $x(r)$.



Require: $P(r \leq r') = P(x \leq x(r'))$

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

That is,

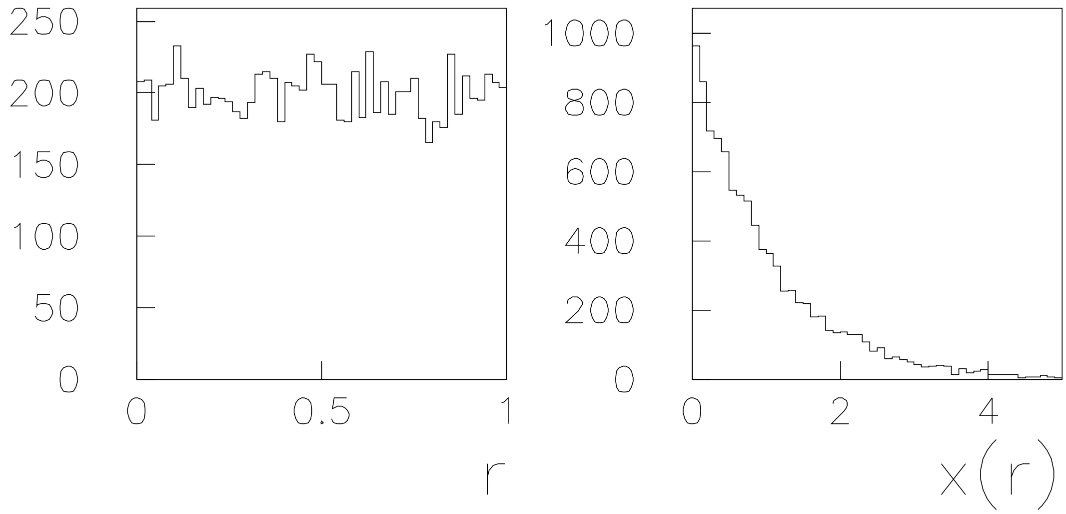
set $F(x(r)) = r$ and solve for $x(r)$.

Example of the transformation method

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

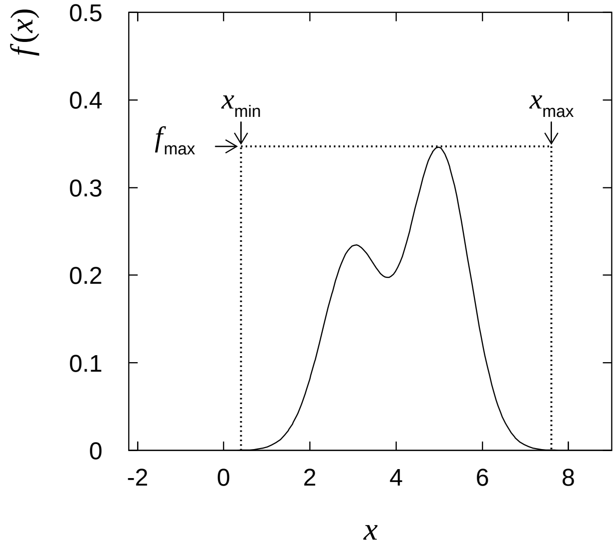
Set $\int_0^x \frac{1}{\xi} e^{-x'/\xi}\, dx' = r$ and solve for $x(r)$.

$\Rightarrow x(r) = -\xi \log(1-r) \qquad (x(r) = -\xi \log r$ works too.)



The acceptance-rejection method (von Neumann)

Enclose the pdf in a box:



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

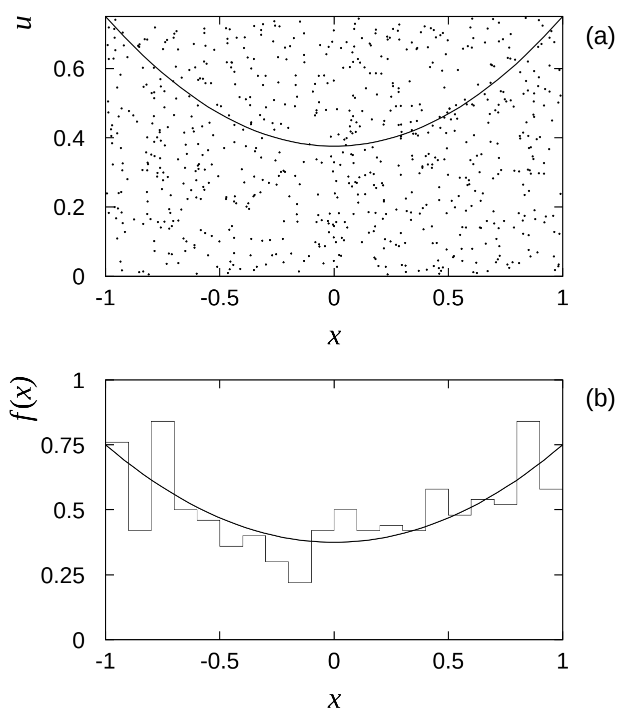
(2) Generate a second independent random number u uniformly distributed between 0 and $f_{\rm max}$, i.e. $u = r_2 f_{\rm max}$.

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

Example:

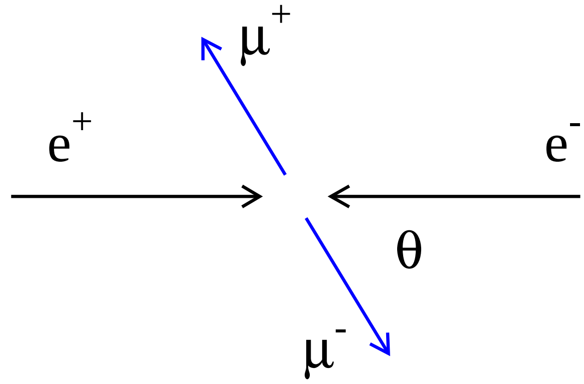
$$f(x) = \frac{3}{8}(1 + x^2)$$

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



Simple example:

$e^+e^- \to \mu^+\mu^-$



Generate θ and ϕ:

$$f(\cos\theta; A_{\rm FB}) \propto (1 + \tfrac{8}{3} A_{\rm FB} \cos\theta + \cos^2\theta)$$

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

Less simple examples:

$e^+e^- \to$ hadrons: JETSET (PYTHIA), HERWIG, ARIADNE

pp $\to$ hadrons: ISAJET, PYTHIA, HERWIG

$e^+e^- \to$ WW: KORALW, EXCALIBUR, ERATO

Output = 'events', i.e. for each event, a list of final state particles and their momentum vectors.

Monte Carlo detector simulation

Takes as input the particle list and momenta from generator.

Simulate detector response:

multiple Coulomb scattering (generate scattering angle)

particle decays (generate lifetime)

ionization energy loss (generate Δ)

EM/hadronic showers

production of signals, electronics response

$\vdots$

Output = simulated raw data

$\rightarrow$ input to reconstruction software (track finding/fitting, etc.)

Uses:

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

Estimate various 'efficiencies' $= \dfrac{\text{\# events found}}{\text{\# events generated}}$

Programming package: GEANT

Lecture 1 summary

1. **Probability**

   Definition: Kolmogorov axioms + conditional probability

   Interpretation: frequency or degree of belief

   Bayes' theorem

2. **Random variables**

   Probability density functions (pdf), e.g. $f(x)$

   Cumulative distribution functions, $F(x) = \int_{-\infty}^{x} f(x')\,dx'$

   Joint pdf, e.g. $f(x, y)$

3. **Expectation values**

   Mean, variance, covariance

   Error propagation

4. **Probability functions and densities:**

   Binomial, Poisson, uniform, exponential, Gaussian ($\rightarrow$CLT), chi-square, Cauchy, Landau

5. **The Monte Carlo method**

   Random number generators

   The transformation method

   The acceptance-rejection method

   Uses of MC in High Energy Physics