Statistical Data Analysis 2025/26 Lecture Week 3



London Postgraduate Lectures on Particle Physics University of London MSc/MSci course PH4515



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

Course web page via RHUL moodle (PH4515) and also

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

Some distributions

<u>Distribution/pdf</u> Example use in Particle Physics

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

Beta Prior pdf for efficiency

Gamma Sum of exponential variables

Student's *t* Resolution function with adjustable tails

Statistical Data Analysis Lecture 3-1

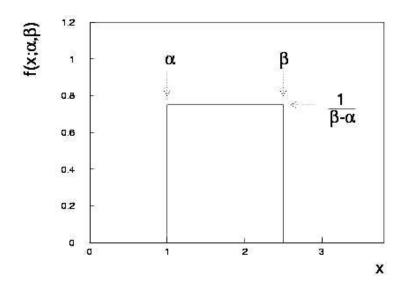
- Continuous probability density functions
 - Uniform
 - Exponential

Uniform distribution

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

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

$$V[x] = \frac{1}{12}(\beta - \alpha)^2$$



Notation: x follows a uniform distribution between α and β

write as: $x \sim U[\alpha,\beta]$

Uniform distribution (2)

Very often used with α = 0, β = 1 (e.g., Monte Carlo method).

For any r.v. x with pdf f(x), cumulative distribution F(x), the function y = F(x) is uniform in [0,1]:

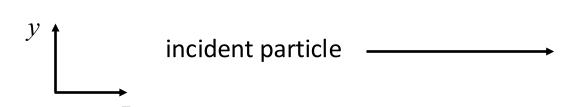
$$g(y) = f(x) \left| \frac{dx}{dy} \right| = \frac{f(x)}{|dy/dx|}$$

$$= \frac{f(x)}{|dF/dx|} = \frac{f(x)}{f(x)} = 1, \quad 0 \le y \le 1$$

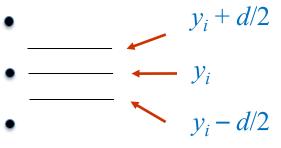
because f(x) = dF/dx = dy/dx

Uniform distribution: particle detector example

Vertical (y) position of particle's trajectory uniformly distributed over perpendicular plane of sense wires.



sense wires, spacing d



If *i*-th wire gives signal,

estimated y position is y_i ,

actual y position $\sim U[y_i - d/2, y_i + d/2]$,

$$V[y] = (y_i + d/2 - (y_i - d/2))^2 / 12 = d^2 / 12,$$

position resolution = $\sigma_y = d/\sqrt{12}$

Sense wire closest to passage of particle gives signal.

Uniform distribution: particle decay example



$$\pi^{\circ}$$
 decay isotropic:
 $cos \theta \sim U[-1,1]$
 $\phi \sim U[0,2\pi]$

In lab frame:

$$E_{\pi,i} \sim U\left[\underline{E}_{\min}, E_{\max}\right]$$

$$E_{\min} = \frac{1}{2} E_{\pi}(1-\beta)$$

$$E_{\max} = \frac{1}{2} E_{\pi}(1+\beta)$$

$$\beta = N_{\pi}/c$$

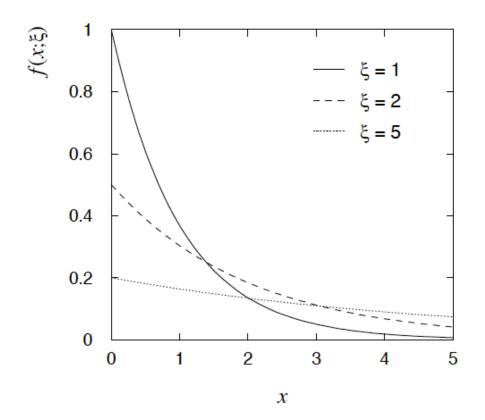
Exponential distribution

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

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

$$E[x] = \xi$$

$$V[x] = \xi^2$$



Exponential distribution (2)

Example: proper decay time *t* of an unstable particle

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

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

Question for discussion:

A cosmic ray muon is created 30 km high in the atmosphere, travels to sea level and is stopped in a block of scintillator, giving a start signal at t_0 . At a time t it decays to an electron giving a stop signal. What is distribution of the difference between stop and start times, i.e., the pdf of $t - t_0$ given $t > t_0$?

Statistical Data Analysis Lecture 3-2

- The Gaussian (normal) distribution
 - Univariate Gaussian
 - Standardized random variables
 - Location and scale parameters
 - Central Limit Theorem
 - Multivariate Gaussian

Gaussian (normal) distribution

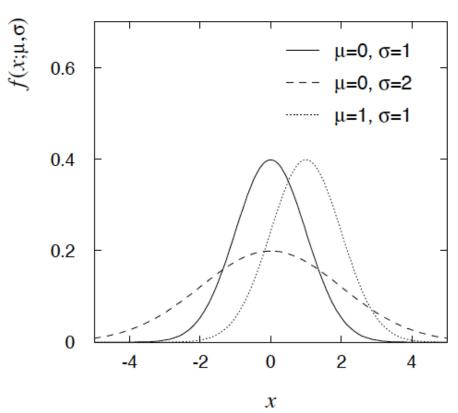
The Gaussian (normal) pdf for a continuous r.v. x is defined by:

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

$$E[x] = \mu$$

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

N.B. often $\sqrt{ }$, $\sqrt{2}$ denote mean, variance of any r.v., not only Gaussian.



Standardized random variables

If a random variable y has pdf f(y) with mean μ and std. dev. σ , then the *standardized* variable

$$x = \frac{y - \mu}{\sigma} \quad \text{has the pdf} \quad g(x) = f(y(x)) \left| \frac{dy}{dx} \right| = \sigma f(\mu + \sigma x)$$

has mean of zero and standard deviation of 1.

Often work with the *standard* Gaussian distribution (μ = 0. σ = 1) using notation:

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

Then e.g. $y = \mu + \sigma x$ follows

$$f(y) = \frac{1}{\sigma} \varphi\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2}$$

Digression: location/scale parameters

If a pdf f(x; a) depending on a parameter a can be written as

$$f(x;a) = f(x-a;0)$$

then a is called a location parameter. Adjusting a shifts the pdf to the right/left without changing its shape.

The parameter μ of the Gaussian is an example of a location parameter.

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

Digression: location/scale parameters (2)

If a pdf f(x; b) depending on a parameter b can be written as

$$f(x;b) = \frac{1}{b}f(x/b;1)$$

then b is called a scale parameter. Adjusting b changes the "units" of the random variable.

The parameter ξ of the exponential is an example of a scale parameter.

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

Or if a pdf f(x; a, b) has a location parameter a and can be written

$$f(x; a, b) = \frac{1}{b} f\left(\frac{x-a}{b}; 0, 1\right)$$

then a and b are said to be location and scale parameters.

Example: μ and σ of Gaussian.

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 f_i^2 , mean values μ_i , 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

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

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)

Versions of CLT differ in criteria for convergence and requirement (or not) of same pdf for all x_i .

See e.g. en.wikipedia.org/wiki/Central_limit_theorem

Classical CLT: all x_i independent and have same pdf with finite variance, can be proved using characteristic functions (Fourier transforms), see, e.g., SDA Chapter 10.

Physicist's CLT: for finite n, the sum $\sum_{i=1}^{n} x_i$ becomes approximately Gaussian to the extent that the fluctuation of the sum is not dominated by one (or few) terms.

Far enough in the tails the approximation generally breaks down.

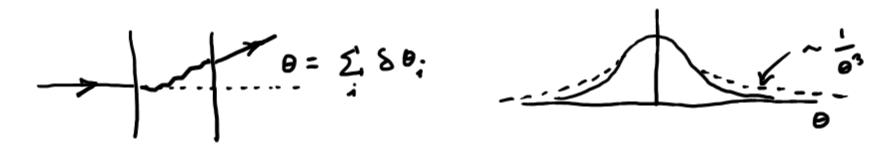
Central Limit Theorem (3)

Good example: velocity component of air molecule $v_x = \sum_i \delta v_{xi}$

If v_x , v_y , v_z ~ Gaussian, then

$$v = (v_x^2 + v_y^2 + v_z^2)^{1/2} \sim \text{Maxwell-Boltzmann}$$

OK example: total deflection of charged particle from multiple Coulomb scattering. (Rare large-angle scatters → 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 $\vec{x} = (x_1, \dots, x_n)$:

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

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

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

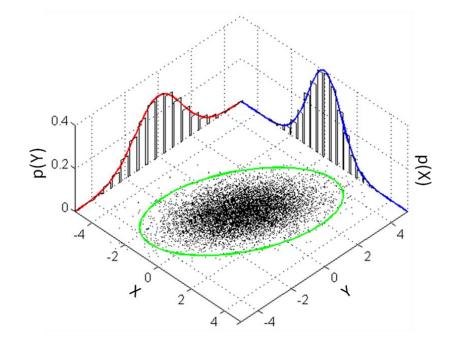
Marginal pdf of each x_i is Gaussian with mean μ_i , standard deviation $\sigma_i = \sqrt{V_{ii}}$.

Two-dimensional Gaussian distribution

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

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

where $/= cov[x_1, x_2]/(/_1/_2)$ is the correlation coefficient.



Statistical Data Analysis Lecture 3-3

- More continuous probability density functions
 - Chi-square
 - Cauchy
 - Landau
 - Beta
 - Gamma
 - Student's t

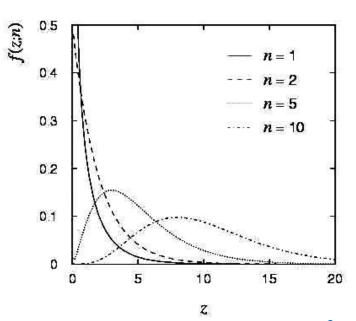
Chi-square (χ^2) distribution

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

$$f(z;n) = \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2} \stackrel{\text{(i)}}{=} \int_{0.4}^{0.5} |z|^{0.5}$$

n = 1, 2, ... = number of 'degrees of freedom' (dof)

$$E[z] = n, \quad V[z] = 2n.$$



For independent Gaussian x_i , i = 1, ..., 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 pdf with n 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

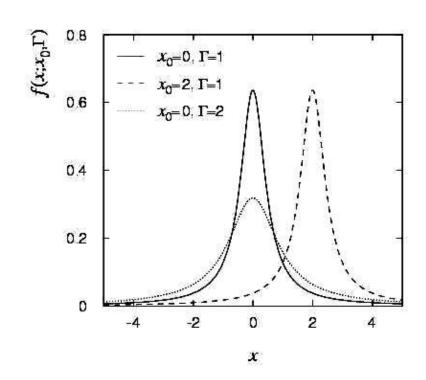
$$f(x; \Gamma, x_0) = \frac{1}{\pi} \frac{\Gamma/2}{\Gamma^2/4 + (x - x_0)^2} \quad \text{for } x_0 = \frac{1}{\pi} \frac{\Gamma/2}{\Gamma^2/4 + (x - x_0)^2}$$

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

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

 $x_0 = \text{mode (most probable value)}$

 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:

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

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

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

$$\xi = \frac{2\pi N_A e^4 z^2 \rho \sum Z}{m_B c^2 \sum A} \frac{d}{\beta^2} , \qquad \epsilon' = \frac{I^2 \exp \beta^2}{2m_B c^2 \beta^2 \gamma^2} .$$

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

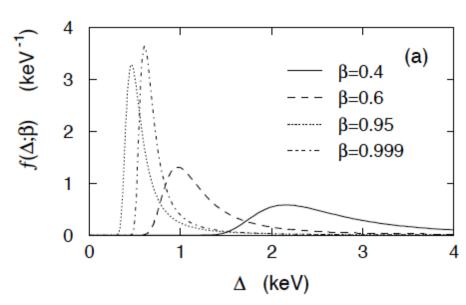
Landau distribution (2)

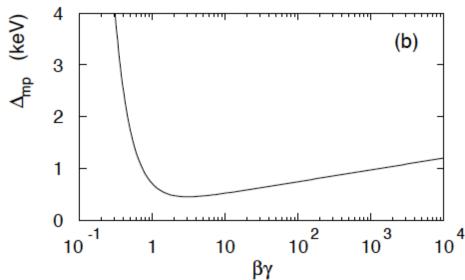
Long 'Landau tail'

 \rightarrow all moments ∞

Mode (most probable value) sensitive to β ,

 \rightarrow particle i.d.





Beta distribution

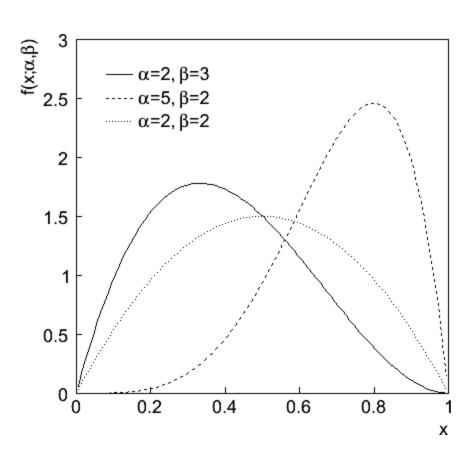
$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha - 1} (1 - x)^{\beta - 1} , \qquad 0 \le x \le 1$$

$$E[x] = \frac{\alpha}{\alpha + \beta}$$

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

Often used to represent pdf of continuous r.v. nonzero only between finite limits, e.g.,

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



Gamma distribution

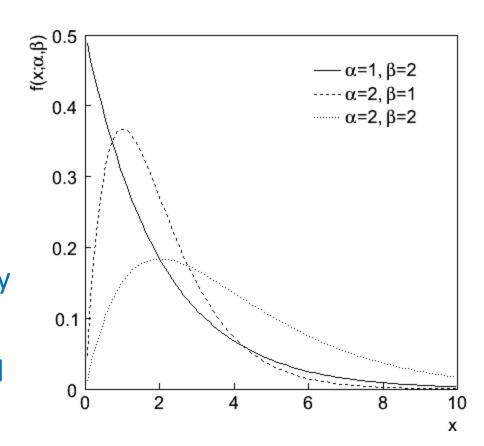
$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \qquad x \ge 0$$

$$E[x] = \alpha \beta$$

$$V[x] = \alpha \beta^2$$

Often used to represent pdf of continuous r.v. nonzero only in $[0,\infty]$.

Also e.g. sum of n exponential r.v.s or time until nth event in Poisson process \sim Gamma



Student's t distribution

$$f(x;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

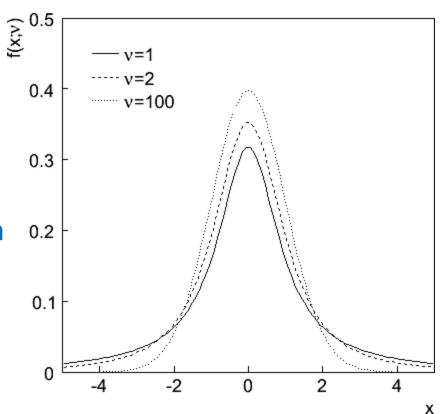
$$E[x] = 0 \quad (\nu > 1)$$

$$V[x] = \frac{\nu}{\nu - 2} \quad (\nu > 2)$$

v = number of degrees of freedom
 (not necessarily integer)

v = 1 gives Cauchy,

 $v \rightarrow \infty$ gives Gaussian.



Student's *t* distribution (2)

If $x \sim$ Gaussian with $\mu = 0$, $\sigma^2 = 1$, and $z \sim \chi^2$ with n degrees of freedom, then $t = x / (z/n)^{1/2}$ follows Student's t with v = n.

This arises in problems where one forms the ratio of a sample mean to the sample standard deviation of Gaussian r.v.s.

The Student's t provides a bell-shaped pdf with adjustable tails, ranging from those of a Gaussian, which fall off very quickly, ($v \rightarrow \infty$, but in fact already very Gauss-like for v = two dozen), to the very long-tailed Cauchy (v = 1).

Developed in 1908 by William Gosset, who worked for the Guinness Brewery and published under the pseudonym "Student".

VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

By STUDENT.

Statistical Data Analysis Lecture 3-4

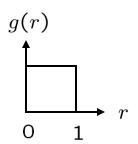
- The Monte Carlo method
 - basic ingredients
 - random number generators
 - transformation method
 - acceptance-rejection method
 - example uses

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 $r_1, r_2, ..., r_m$ independent and uniform on [0, 1].



- (2) Use this to produce another sequence $x_1, x_2, ..., x_n$ independent and 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 $\int_a^b f(x) dx$.
 - → MC calculation = integration (at least formally)

MC generated values = 'simulated data'

→ 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). → 'random number generator' = computer algorithm to generate $r_1, r_2, ..., r_n$. Example: multiplicative linear congruential generator (MLCG) $n_{i+1} = (a n_i) \mod m$, where $n_i = integer$ a = multiplierm = 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, $n_0 = 1$

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

 $n_2 = (3 \cdot 3) \mod 7 = 2$
 $n_3 = (3 \cdot 2) \mod 7 = 6$
 $n_4 = (3 \cdot 6) \mod 7 = 4$
 $n_5 = (3 \cdot 4) \mod 7 = 5$
 $n_6 = (3 \cdot 5) \mod 7 = 1 \qquad \leftarrow \text{ sequence repeats}$

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)

 $r_i = n_i/n_{\rm max}$ are in [0, 1] but are they independent and uniform?

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

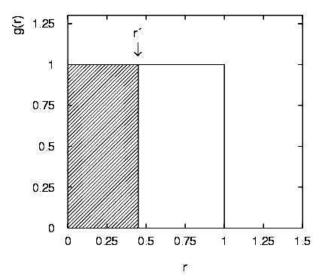
$$a = 40692$$

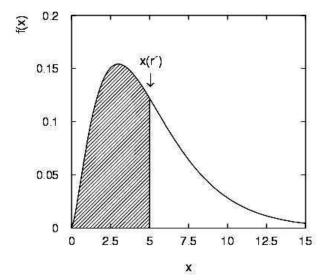
$$m = 2147483399$$

Far better generators available, e.g. **TRandom3**, based on Mersenne twister algorithm, period = $2^{19937} - 1$ (a "Mersenne prime"). 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: $P(r \le r') = P(x \le x(r'))$

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

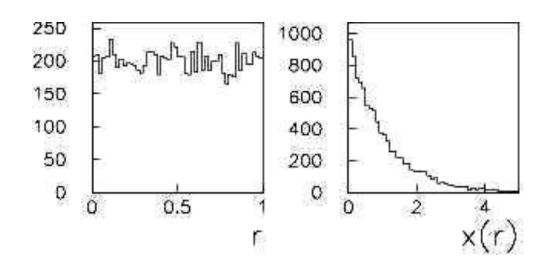
That is, set F(x) = r and solve for x(r).

Example of the transformation method

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

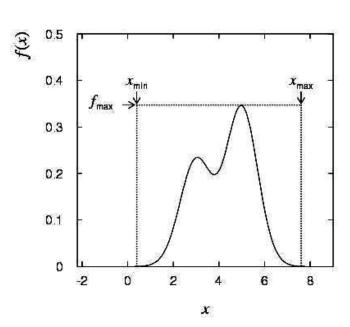
Set
$$\int_0^x \frac{1}{\xi} e^{-x'/\xi} dx' = r$$
 and solve for $x(r)$.

$$\rightarrow x(r) = -\xi \ln(1-r)$$
 $(x(r) = -\xi \ln r \text{ works too.})$



The acceptance-rejection method

Enclose the pdf in a box:



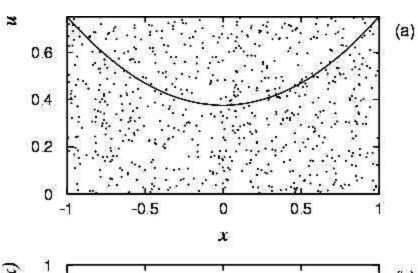
- (1) Generate a random number x, uniform in $[x_{\min}, x_{\max}]$, i.e. $x = x_{\min} + r_1(x_{\max} x_{\min})$, r_1 is uniform in [0,1].
- (2) Generate a 2nd independent random number u uniformly distributed between 0 and $f_{\rm max}$, i.e. $u=r_2f_{\rm max}$.
- (3) If u < f(x), then accept x. If not, reject x and repeat.

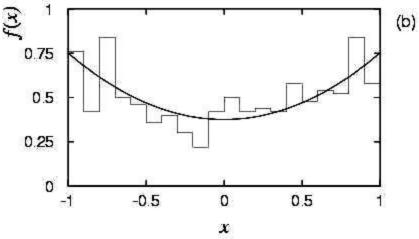
Example with acceptance-rejection method

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

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

If dot below curve, use x value in histogram.



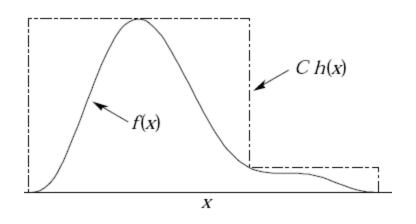


Improving efficiency of the acceptance-rejection method

The fraction of accepted points is equal to the fraction of the box's area under the curve.

For very peaked distributions, this may be very low and thus the algorithm may be slow.

Improve by enclosing the pdf f(x) in a curve C h(x) that conforms to f(x) more closely, where h(x) is a pdf from which we can generate random values and C is a constant.

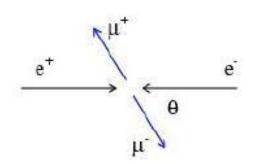


Generate points uniformly over C h(x).

If point is below f(x), accept x.

Monte Carlo event generators

Simple example: $e^+e^- \rightarrow \mu^+\mu^-$



Generate $\cos\theta$ and φ :

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

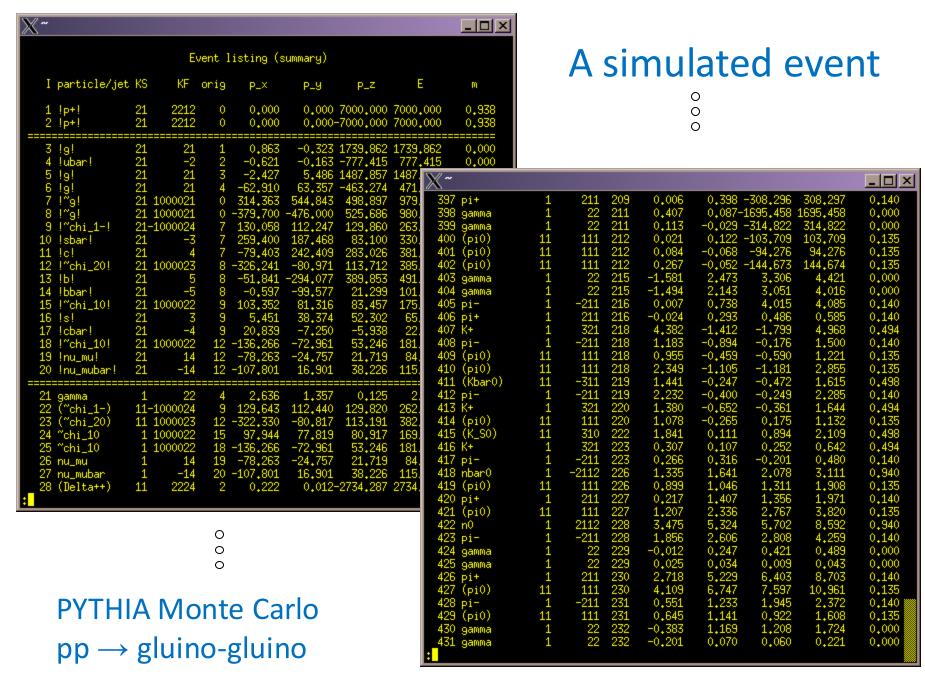
$$g(\phi) = \frac{1}{2\pi} \quad (0 \le \phi \le 2\pi)$$

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

$$e^+e^- \rightarrow \mu^+ \, \mu^-$$
, hadrons, ... pp \rightarrow 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.



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 \Delta), 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.

Programming package: **GEANT**

Extra slides

Importance Sampling

Often the goal of a Monte Carlo calculation is to determine an expectation value of a function h(x), where x is a random variable that follows a pdf f(x),

$$E_f[h(x)] = \int h(x)f(x) dx \equiv \mu . \tag{1}$$

A Monte-Carlo estimator $\hat{\mu}_{MC}$ for μ is the average of N values of h(x) where x is sampled (generated) from f(x):

$$\hat{\mu}_{\text{MC}} = \frac{1}{N} \sum_{i=1}^{N} h(x_i) ,$$
 (2)

which has a variance

$$V[\hat{\mu}_{\text{MC}}] = \frac{1}{N} V_f[h(x)] = \frac{1}{N} \left(E_f[h^2(x)] - \mu^2 \right) . \tag{3}$$

Importance Sampling (2)

By using the method of *importance sampling*, one can achieve a reduction in this variance and thus a more accurate determination of the expectation value for a given number of random values generated. The key idea is to rewrite the expectation value in Eq. (1) as

$$\mu = \int h(x)f(x) dx = \int \frac{h(x)f(x)}{g(x)} g(x) dx = E_g \left[\frac{h(x)f(x)}{g(x)} \right] , \qquad (4)$$

where g(x) is any other pdf of x with the same support as f(x) (i.e., nonzero for the same region of x). Thus the desired quantity μ is the expectation value with respect to g of h(x)f(x)/g(x). It can be estimated by generating N values of x sampled from g and computing

$$\hat{\mu}_{IS} = \frac{1}{N} \sum_{i=1}^{N} \frac{h(x_i)f(x_i)}{g(x_i)} . \tag{5}$$

Importance Sampling (3)

The variance of $\hat{\mu}_{\text{IS}}$ is given by

$$V[\hat{\mu}_{\rm IS}] = \frac{1}{N} V_g \left[\frac{h(x)f(x)}{g(x)} \right] = \frac{1}{N} \left(E_g \left[\frac{h^2(x)f^2(x)}{g^2(x)} \right] - \mu^2 \right) , \tag{6}$$

By choosing g(x) such that h(x)f(x)/g(x) is as constant as possible, one can minimize the variance of $\hat{\mu}_{IS}$. One can show (see, e.g., Refs. [1, 2]) that the variance is minimized when $g(x) \propto |h(x)|f(x)$.

An alternative importance sampling estimator can be constructed by replacing the number of generated values N in Eq. (5) by the sum $\sum_{i=1}^{N} f(x_i)/g(x_i)$. This can given an even smaller variance at the price of a small bias. It can be of further advantage in problems where the pdf f(x) is known only up to a normalization constant, which then cancels (see Refs. [1, 2]).

References

- [1] C.P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed., (Springer, New York, 2004).
- [2] J.S. Liu, Monte Carlo Strategies in Scientific Computing, (Springer, New York, 2001).