Statistics for Particle Physicists
Lecture 1: Introduction, Probability

Summer Student Lectures
CERN
4 – 7 July 2023

https://indico.cern.ch/event/1254879/timetable/

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

Outline

Lecture 1: Introduction, probability,

Lecture 2: Parameter estimation

Lecture 3: Hypothesis tests

Lecture 4: Introduction to Machine Learning

Almost everything is a subset of the University of London course:

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

Some statistics books, papers, etc.

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

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

Ilya Narsky and Frank C. Porter, *Statistical Analysis Techniques in Particle Physics*, Wiley, 2014.

Luca Lista, *Statistical Methods for Data Analysis in Particle Physics*, Springer, 2017.

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

F. James., *Statistical and Computational Methods in Experimental Physics*, 2nd ed., World Scientific, 2006

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

R.L. Workman et al. (Particle Data Group), Prog. Theor. Exp. Phys. 083C01 (2022); **`pdg.lbl.gov`** sections on probability, statistics, MC.

Theory ↔ Statistics ↔ Experiment

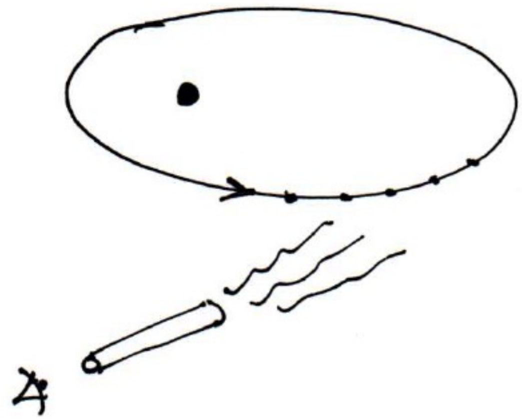
Theory (model, hypothesis):

$$F = -G\frac{m_1 m_2}{r^2}, \ldots$$

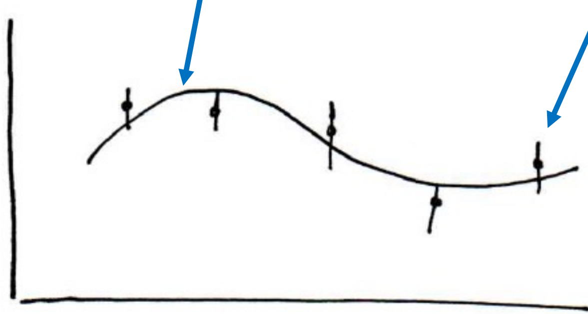
+ response of measurement apparatus

= model prediction

Experiment (observation):



data

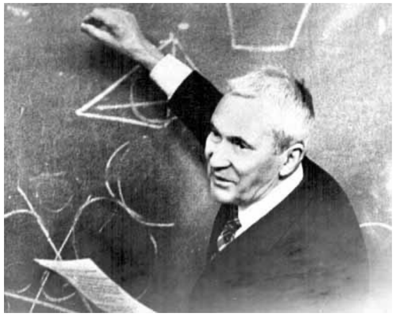


Uncertainty enters on many levels

→ quantify with probability

A definition of probability

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



$$\boxed{\begin{array}{l} \text{For all } A \subset S, P(A) \geq 0 \\ P(S) = 1 \\ \text{If } A \cap B = \emptyset, P(A \cup B) = P(A) + P(B) \end{array}}$$

Kolmogorov axioms (1933)

Also define conditional probability of A given B:

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

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

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

Interpretation of Probability

I. Relative frequency (→ "frequentist statistics")

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

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

II. Subjective probability (→ "Bayesian statistics")

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 magnetic monopoles exist,...

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

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

Bayes' theorem

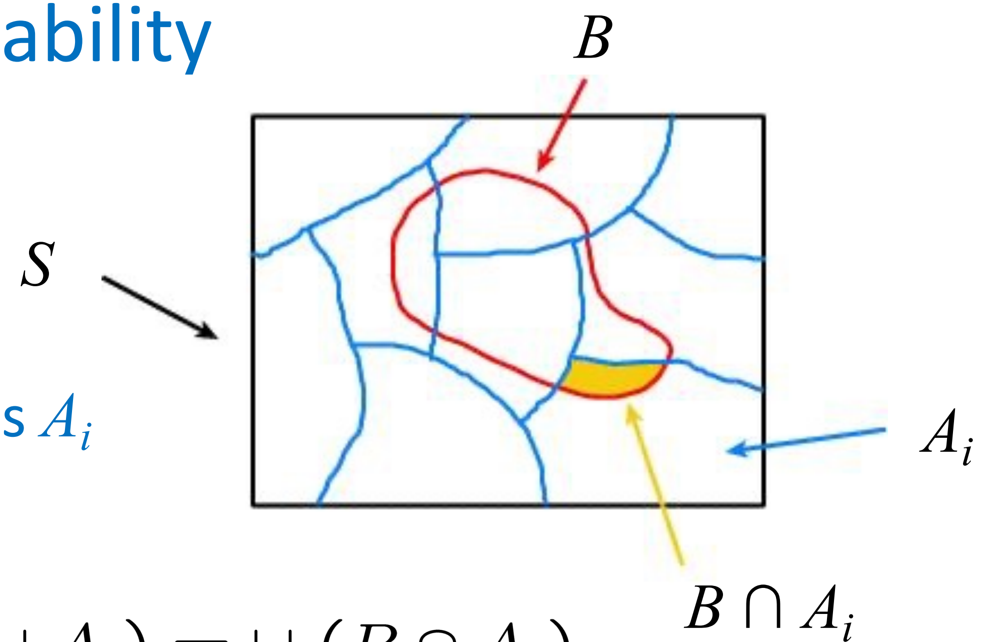
First published (posthumously) by the Reverend Thomas Bayes (1702−1761)

An essay towards solving a problem in the doctrine of chances, Philos. Trans. R. Soc. **53** (1763) 370

The law of total probability

Consider a subset B of the sample space S,

divided into disjoint subsets A_i such that $\cup_i A_i = S$,



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

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

$$\rightarrow \quad P(B) = \sum_i P(B|A_i) P(A_i)$$

Bayes' theorem becomes

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

An example using Bayes' theorem

Suppose the probability (for anyone) to have a disease D is:

$$
\begin{aligned}
P(\mathrm{D}) &= 0.001 \\
P(\text{no D}) &= 0.999
\end{aligned}
$$

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

Consider a test for the disease: result is + or –

$$
\begin{aligned}
P(+|\mathrm{D}) &= 0.98 \\
P(-|\mathrm{D}) &= 0.02
\end{aligned}
$$

← probabilities to (in)correctly identify a person with the disease

$$
\begin{aligned}
P(+|\text{no D}) &= 0.03 \\
P(-|\text{no D}) &= 0.97
\end{aligned}
$$

← probabilities to (in)correctly identify a healthy person

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

Bayes' theorem example (cont.)

The probability to have the disease given a + result is

$$p(\mathrm{D}|+) = \frac{P(+|\mathrm{D})P(\mathrm{D})}{P(+|\mathrm{D})P(\mathrm{D}) + P(+|\text{no D})P(\text{no D})}$$

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

$$= 0.032$$ $\leftarrow$ posterior probability

i.e. you're probably OK!

Your viewpoint: my degree of belief that I have the disease is 3.2%.

Your doctor's viewpoint: 3.2% of people like this have the disease.

Frequentist Statistics – general philosophy

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

Probability = limiting frequency

Probabilities such as

P (string theory is true),
P ($0.117 < \alpha_s < 0.119$),
P (Biden wins in 2024),

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.

Preferred theories (models, hypotheses, ...) are those that predict a high probability for data "like" the data observed.

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 general prescription for priors (subjective!)

Random variables and probability density functions

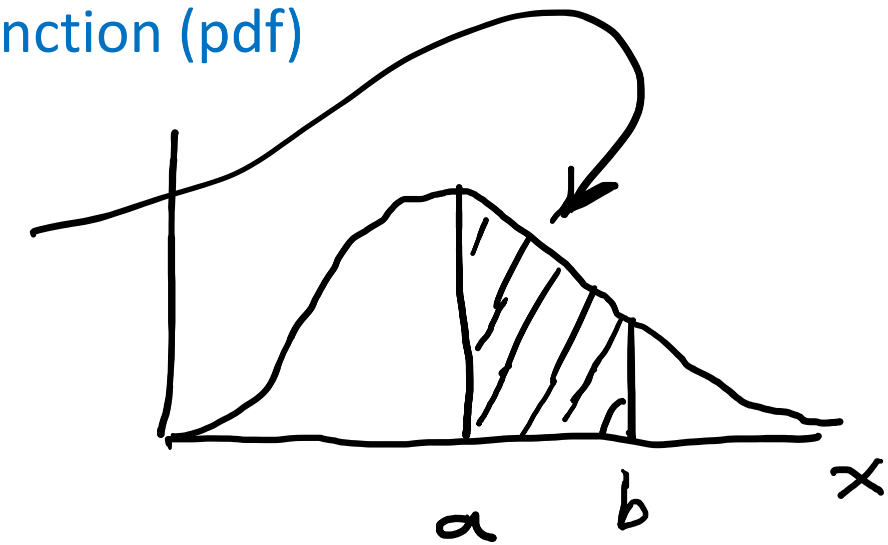
A random variable is a numerical characteristic assigned to an element of the sample space; can be discrete or continuous.

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)

$$\int_a^b f(x)\, dx = P(a \leq x \leq b)$$



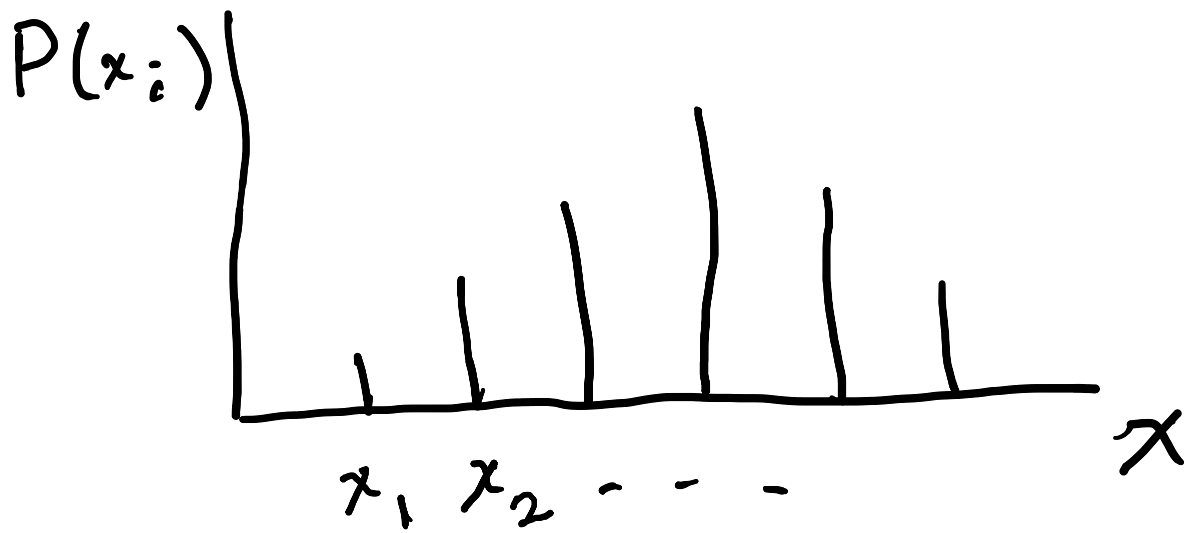
$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$ x must be somewhere

Probability mass function

For discrete outcome x_i with e.g. $i = 1, 2, \ldots$ we have

$P(x_i) = p_i$ probability (mass) function

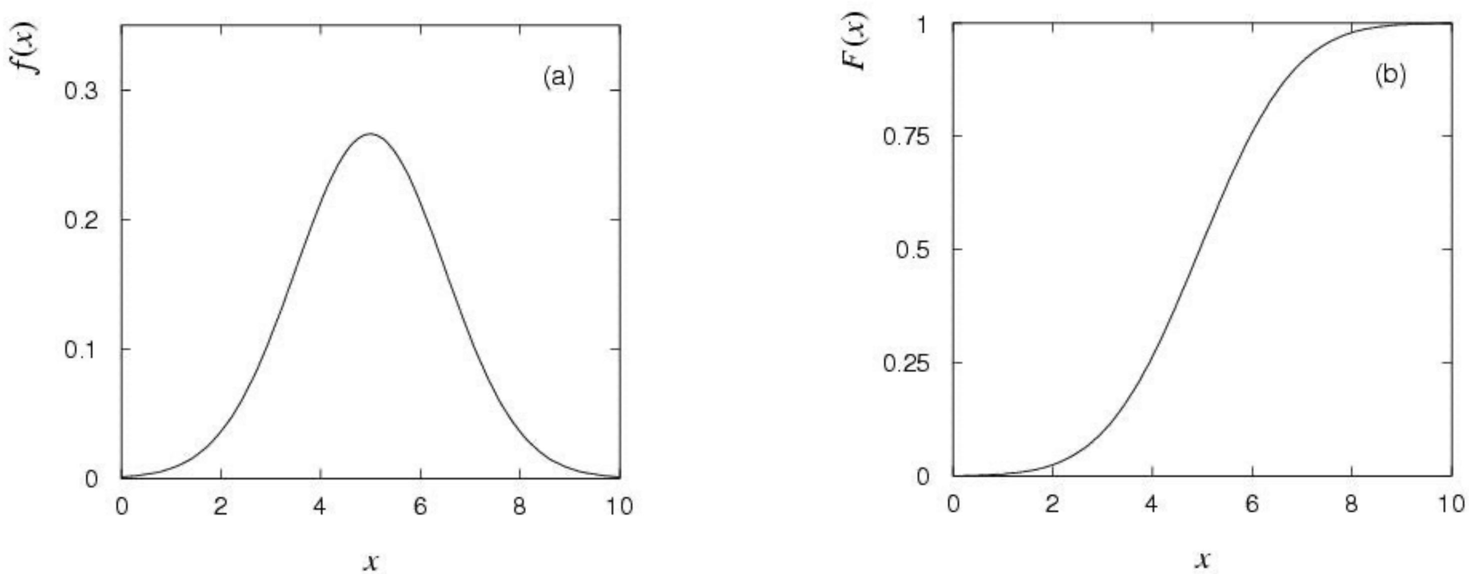
$\sum_i P(x_i) = 1$ x must take on one of its possible values



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 $f(x) = \dfrac{\partial F(x)}{\partial x}$

Some other pdfs

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

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

Marginal pdfs

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

Conditional pdfs

$$f(x|y) = \frac{f(x,y)}{f_y(y)} \qquad f(y|x) = \frac{f(x,y)}{f_x(x)}$$

Bayes' theorem:

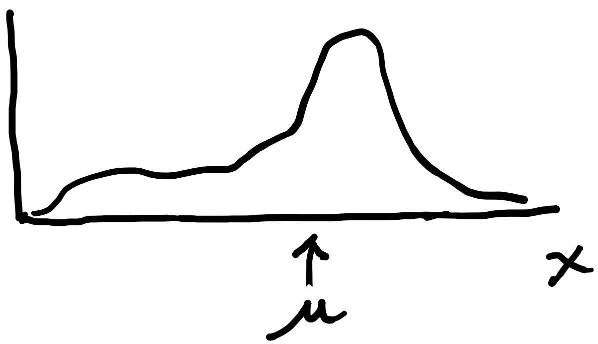
$$f(x|y) = \frac{f(y|x) f_x(x)}{f_y(y)}$$

Expectation values

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

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

Notation (often): $E[x] = \mu$ ~ "centre of gravity" of pdf.



For discrete r.v.s, replace integral by sum: $E[x] = \sum_{x_i \in S} x_i P(x_i)$

For a function $y(x)$ with pdf $g(y)$,

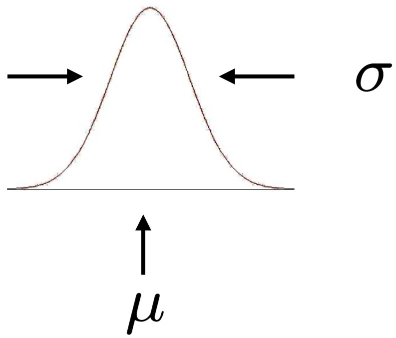
$$E[y] = \int y\, g(y)\, dy = \int y(x) f(x)\, dx$$ (equivalent)

Variance, standard deviation

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

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

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



$\sigma \sim$ width of pdf, same units as x.

Relation between σ and other measures of width, e.g., Full Width at Half Max (FWHM) depend on the pdf, e.g., FWHM = 2.35σ for Gaussian.

Covariance and correlation

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

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

Correlation coefficient (dimensionless) defined as

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

Can show $-1 \leq \rho \leq 1$.

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

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

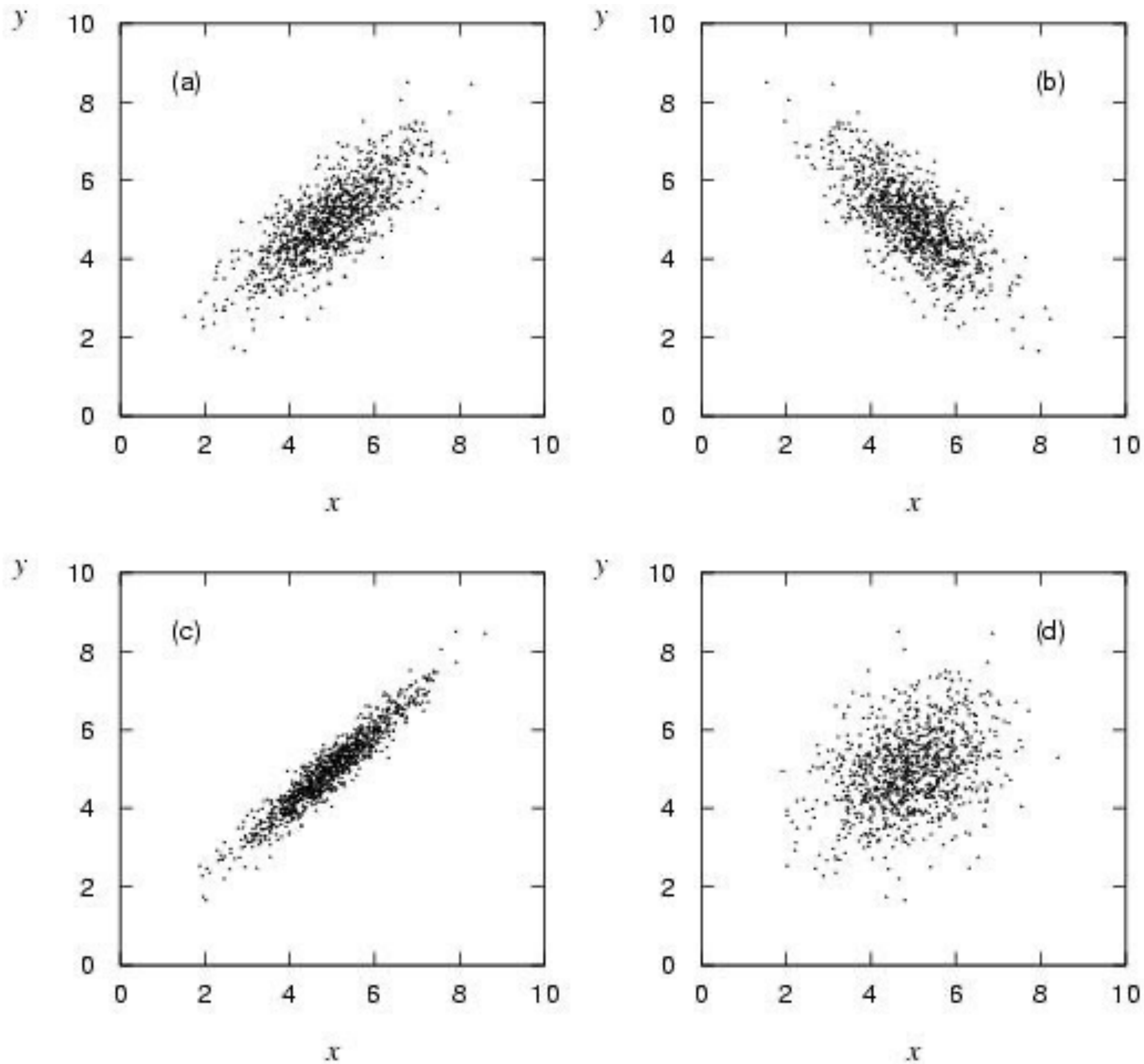
$\rightarrow$ $\mathrm{cov}[x,y] = 0$

N.B. converse not always true.

Correlation (cont.)

$\rho = 0.75$

$\rho = -0.75$



$\rho = 0.95$

$\rho = 0.25$

Covariance matrix

Suppose we have a set of n random variables, say, $x_1,\ldots, x_n$.

We can write the covariance of each pair as an n x n matrix:

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

$$V = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \ldots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \ldots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \rho_{n2}\sigma_n\sigma_2 & \ldots & \sigma_n^2 \end{pmatrix}$$

Covariance matrix is:

symmetric,

diagonal = variances,

positive semi-definite:

$$z^T V z \geq 0 \text{ for all } z \in \mathbb{R}^n$$

Correlation matrix

Closely related to the covariance matrix is the n x n matrix of correlation coefficients:

$$\rho_{ij} = \frac{\mathrm{cov}[x_i, x_j]}{\sigma_i \sigma_j}$$

$$\rho = \begin{pmatrix} 1 & \rho_{12} & \ldots & \rho_{1n} \\ \rho_{21} & 1 & \ldots & \rho_{2n} \\ \vdots & & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \ldots & 1 \end{pmatrix}$$

By construction, diagonal elements are $\rho_{ii} = 1$

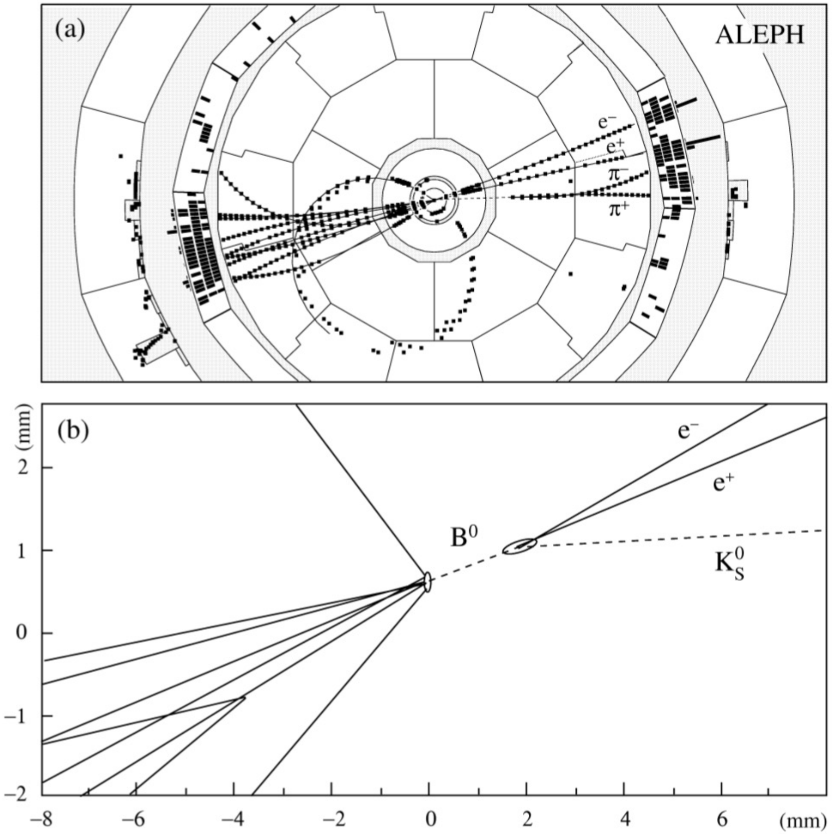
Some distributions (see extra slides)

Distribution/pdf	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

Example: decay of an unstable particle

As an example that we'll use to illustrate several statistical methods, consider measuring the proper decay time of an unstable particle such as a B meson:

R. Barate et al. / Physics Letters B 492 (2000) 259–274



Measure flight distance d and momentum p of decay products of B meson with mass m_{B}.

These are related to the proper decay time t_{p} (time in B rest frame) by

$$d = vt_{\mathrm{lab}} = \beta c \times \gamma t_{\mathrm{p}} = \frac{p_{\mathrm{B}}}{m_{\mathrm{B}}} t_{\mathrm{p}}$$

so $$t_{\mathrm{p}} = \frac{m_{\mathrm{B}} d}{p_{\mathrm{B}}}$$

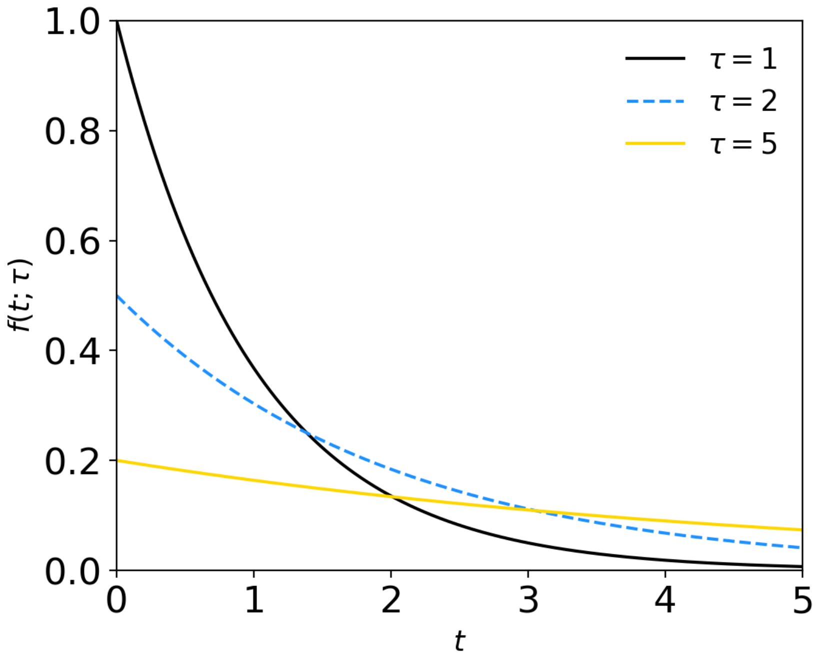
Exponential pdf for proper decay time

We can model t as following an exponential pdf:

$$f(t;\tau) = \frac{1}{\tau} e^{-t/\tau} \,, \qquad t \geq 0$$

random variable

parameter



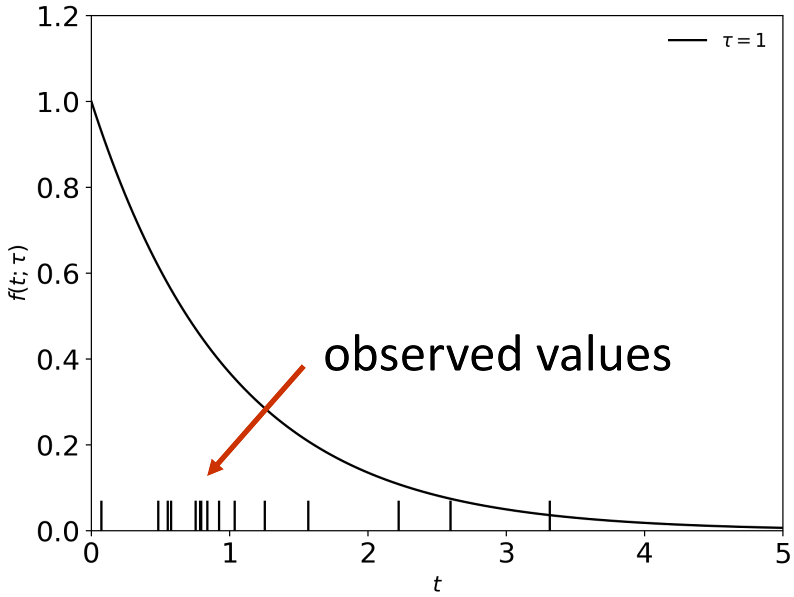
We can show (exercise) that the mean and variance of t are:

$$E[t] = \int_0^\infty t f(t;\tau)\, dt = \tau \qquad\qquad V[t] = E[t^2] - (E[t])^2 = \tau^2$$

Example: statistics with exponentially distributed data

Coming up later in the week:

Suppose the experiment is repeated n times giving data: $t_1, \ldots, t_n$.



Using the data values, estimate the mean lifetime τ.

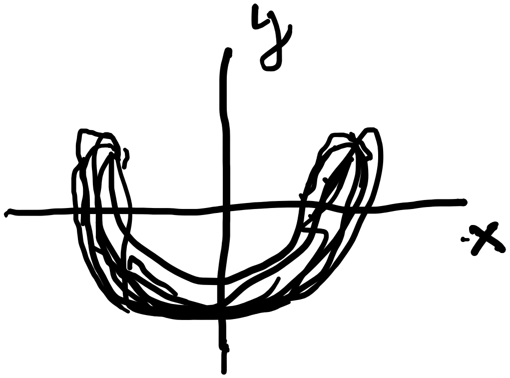
Quantify the statistical uncertainty in the estimate.

Report upper/lower limits on the the mean lifetime.

Extra slides

Correlation vs. independence

Consider a joint pdf such as:



I.e. here $f(-x,y) = f(x,y)$

Because of the symmetry, we have $E[x] = 0$ and also

$$E[xy] = \int_{-\infty}^{\infty} \int_{-\infty}^{0} xy f(x,y)\, dx\, dy + \int_{-\infty}^{\infty} \int_{0}^{\infty} xy f(x,y)\, dx\, dy = 0$$

and so $\rho = 0$, the two variables x and y are uncorrelated.

But $f(y|x)$ clearly depends on x, so x and y are not independent.

Uncorrelated: the joint density of x and y is not tilted.

Independent: imposing x does not affect conditional pdf of y.

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

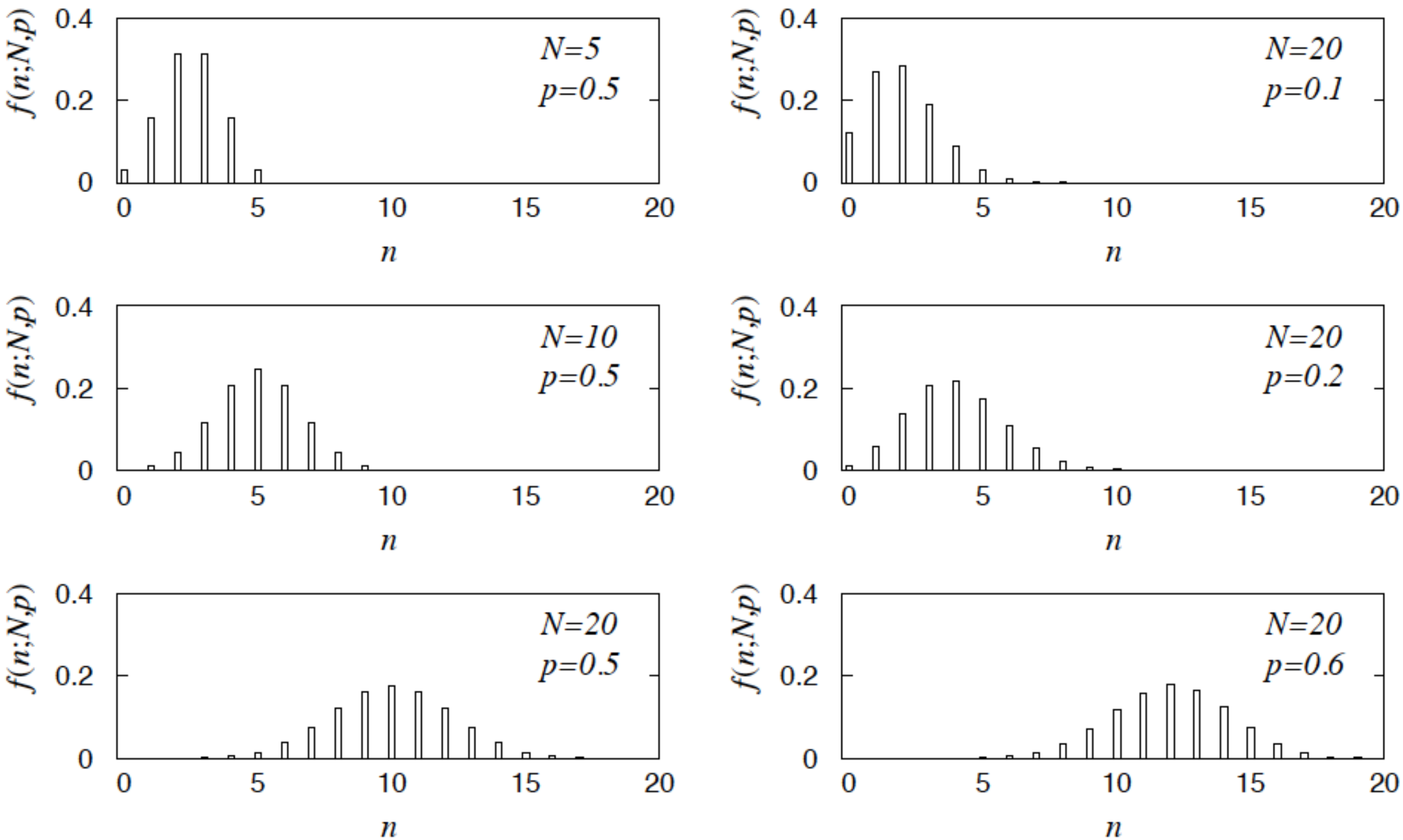
For the expectation value and variance we find:

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

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

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

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

For N trials we want the probability to obtain:

n_1 of outcome 1,
n_2 of outcome 2,
$\vdots$
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'.

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

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

Poisson distribution

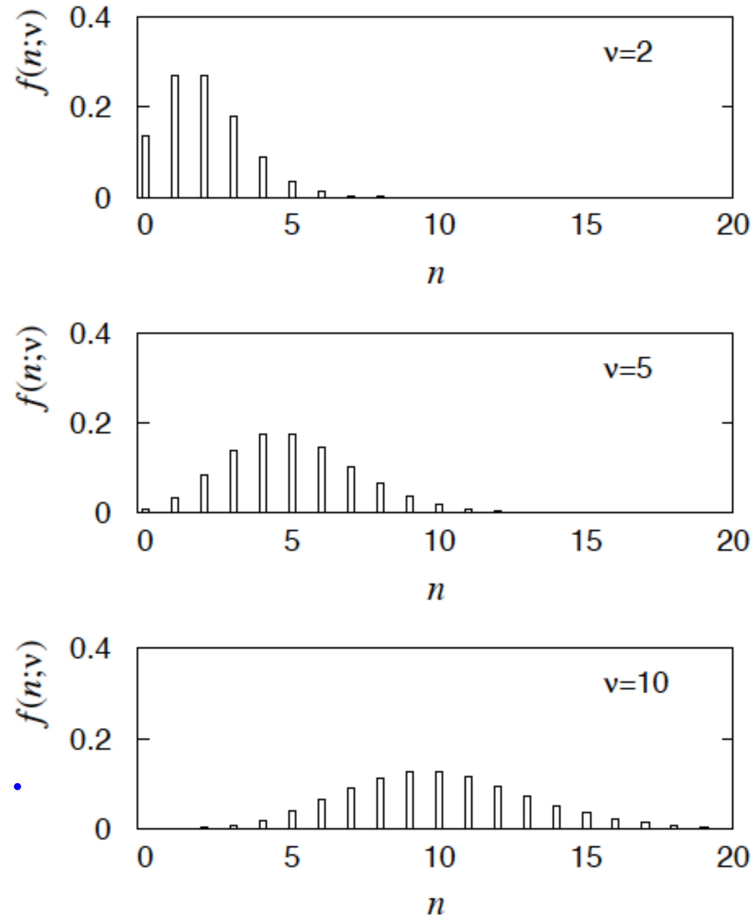
Consider binomial n in the limit

$$N \to \infty, \qquad p \to 0, \qquad E[n] = Np \to \nu \,.$$

$\to$ n follows the Poisson distribution:

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

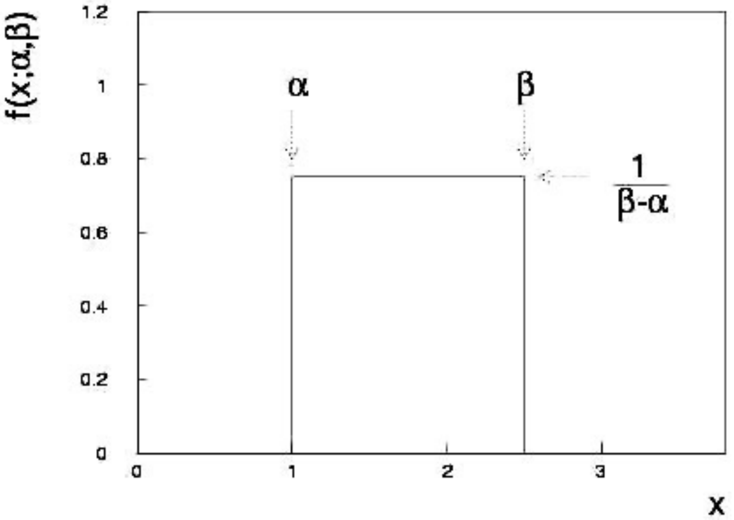
$$E[n] = \nu \,, \qquad V[n] = \nu \,.$$



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

Uniform distribution

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



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

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

Notation: x follows a uniform distribution between α and β

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

Uniform distribution (2)

Very often used with $\alpha = 0$, $\beta = 1$ (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 $[0,1]$:

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

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

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

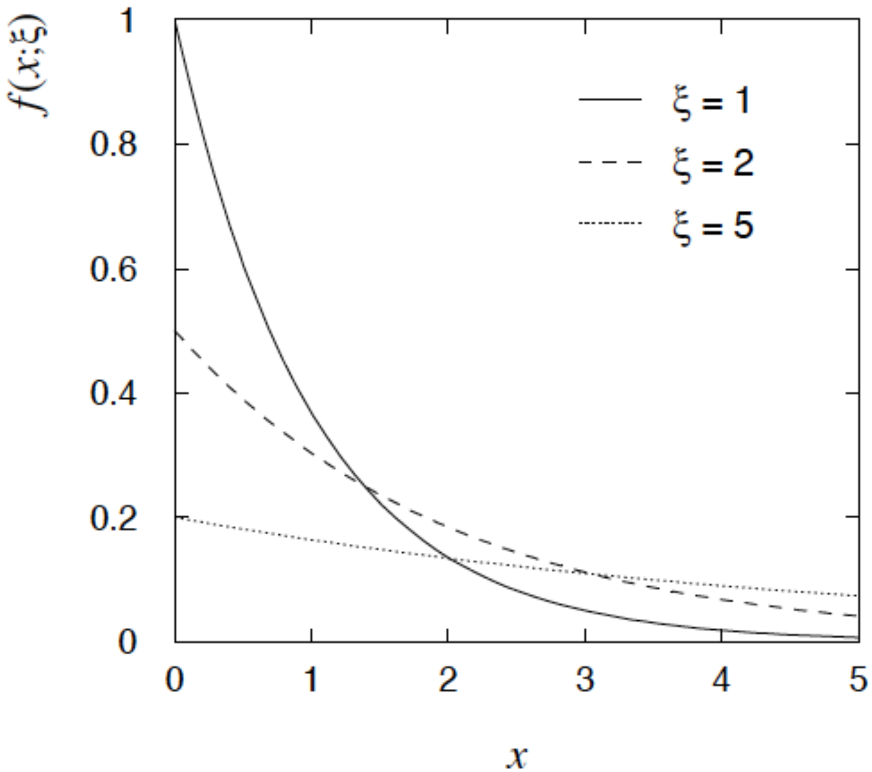
Exponential distribution

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

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

$$E[x] = \xi$$

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



Exponential distribution (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): $f(t - t_0 | t \geq t_0) = f(t)$

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 $t > t_0$?

Gaussian (normal) distribution

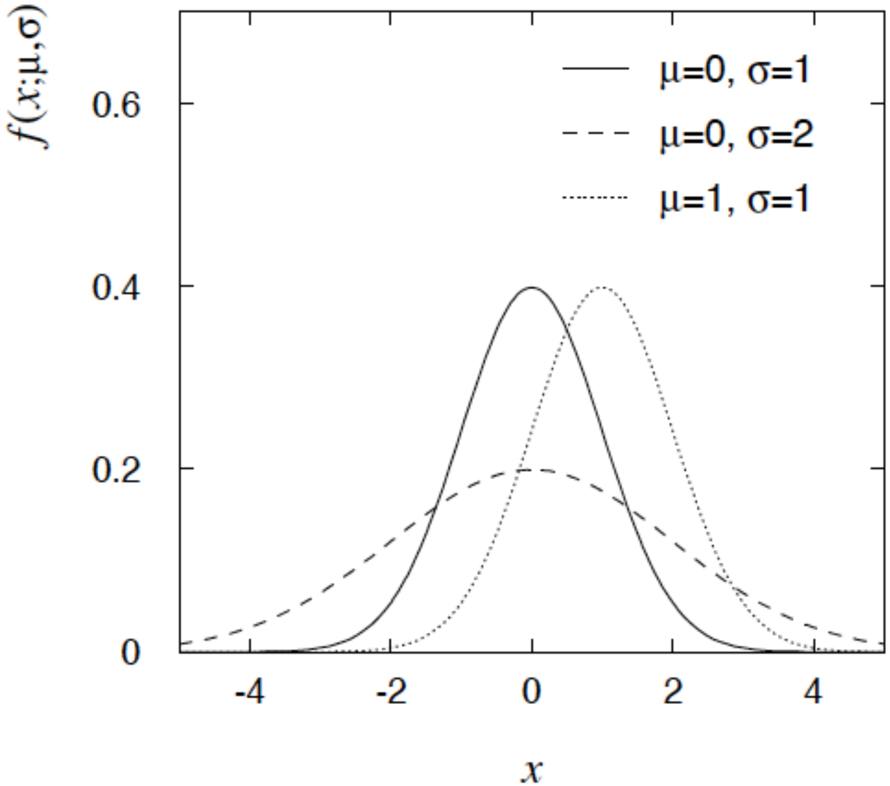
The Gaussian (normal) pdf for a continuous r.v. x is defined by:

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

$$E[x] = \mu$$

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

N.B. often μ, σ^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:

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

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

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

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,

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

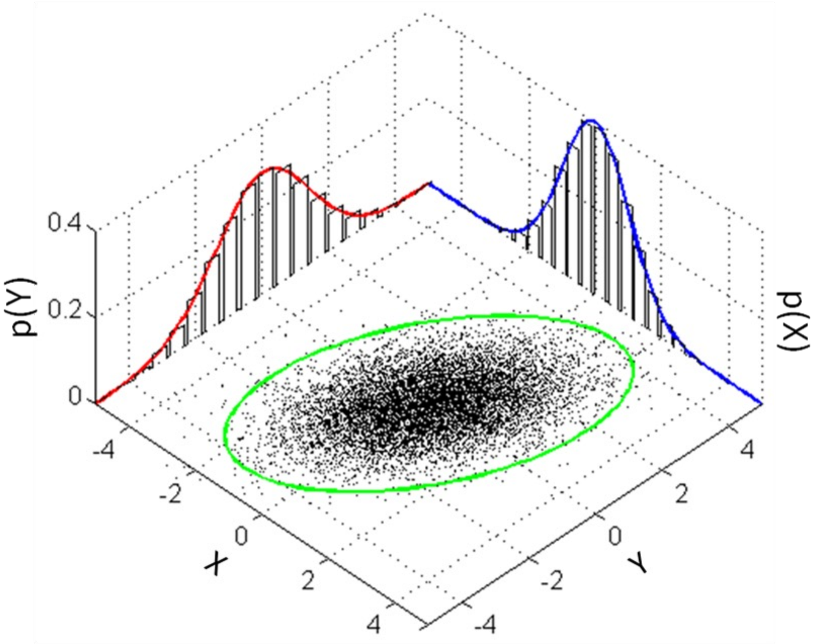
Marginal pdf of each x_i is Gaussian with mean μ_i, standard deviation $\sigma_i = \sqrt{V_{ii}}$.

Two-dimensional Gaussian distribution

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

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

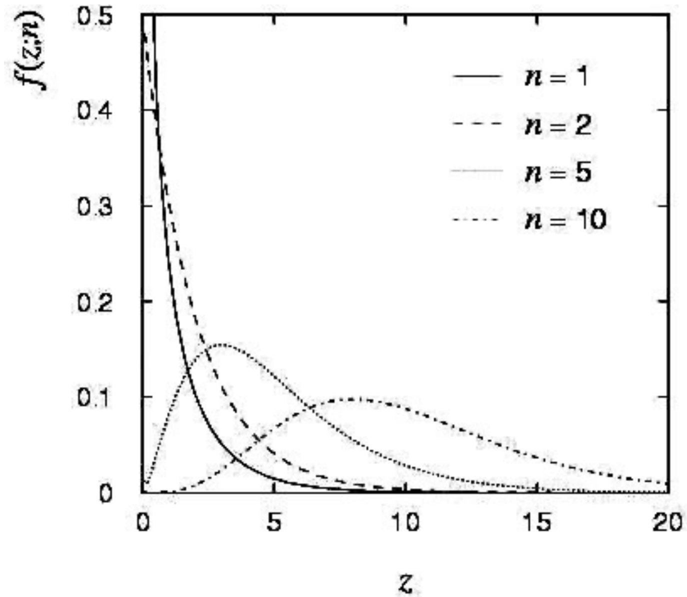
where $\rho = \mathrm{cov}[x_1, x_2]/(\sigma_1\sigma_2)$ is the correlation coefficient.



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, ... = number of 'degrees of freedom' (dof)

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

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 pdf with n dof.

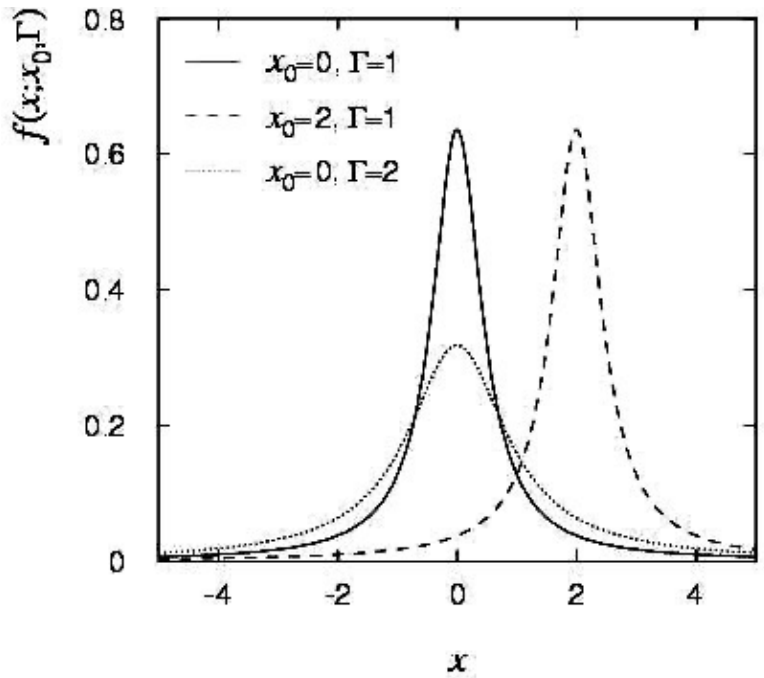
Example: goodness-of-fit test variable especially in conjunction with method of least squares.

Cauchy (Breit-Wigner) distribution

The Breit-Wigner pdf for the continuous r.v. x is defined by

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

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



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

x_0 = mode (most probable value)

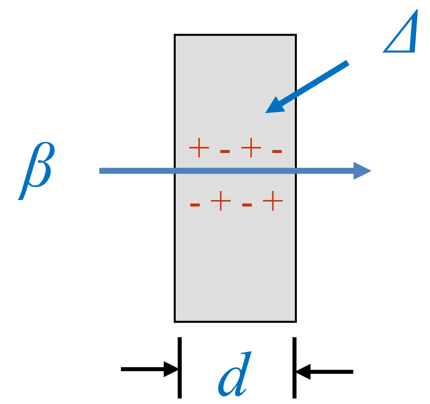
Γ = full width at half maximum

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:



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

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

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

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

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

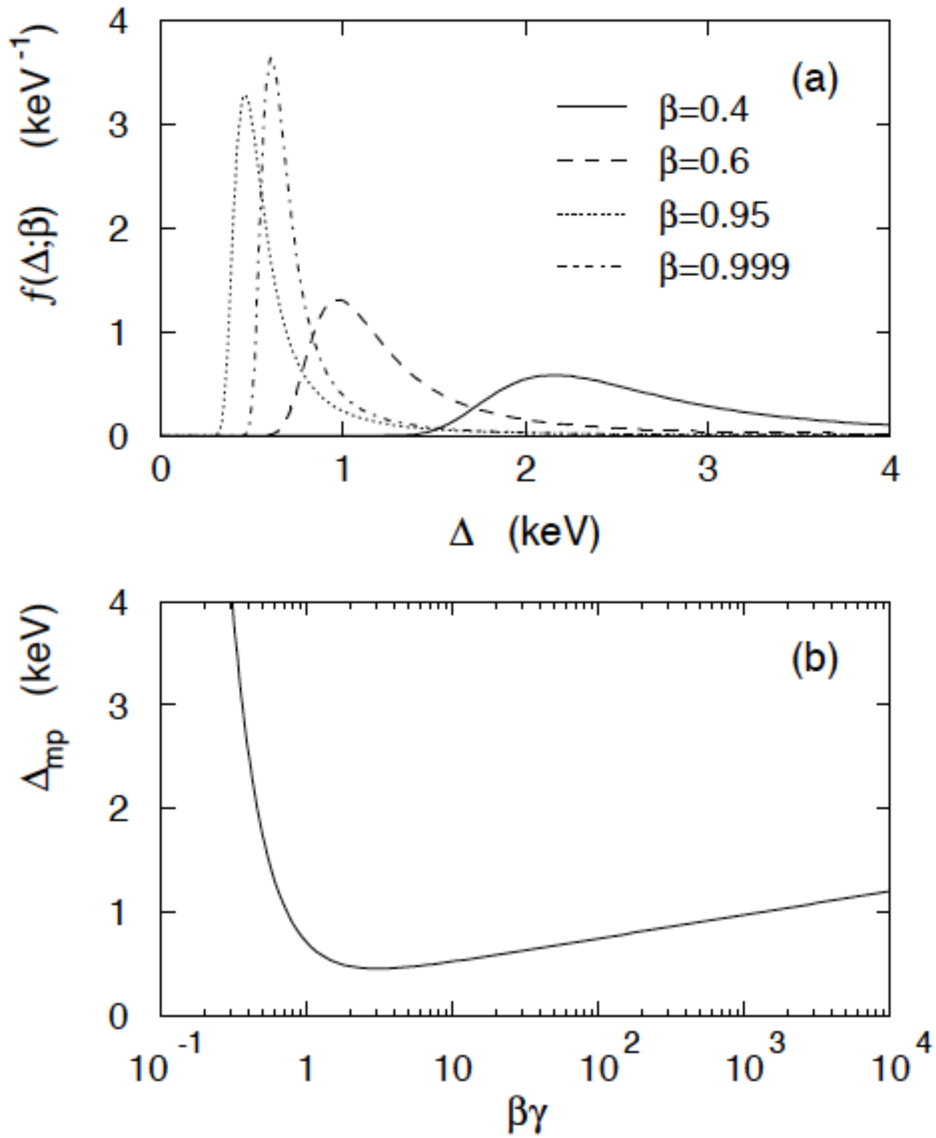
Landau distribution (2)

Long 'Landau tail'

$\rightarrow$ all moments ∞

Mode (most probable value) sensitive to β ,

$\rightarrow$ particle i.d.



Beta distribution

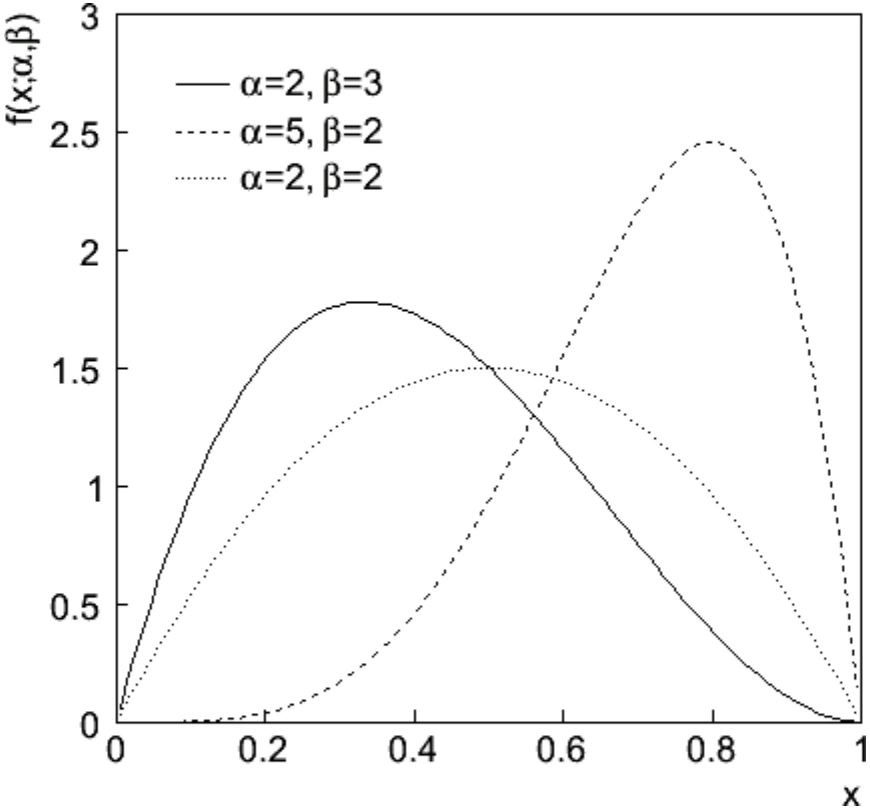
$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,x^{\alpha-1}(1-x)^{\beta-1}$$

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

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

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

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



Gamma distribution

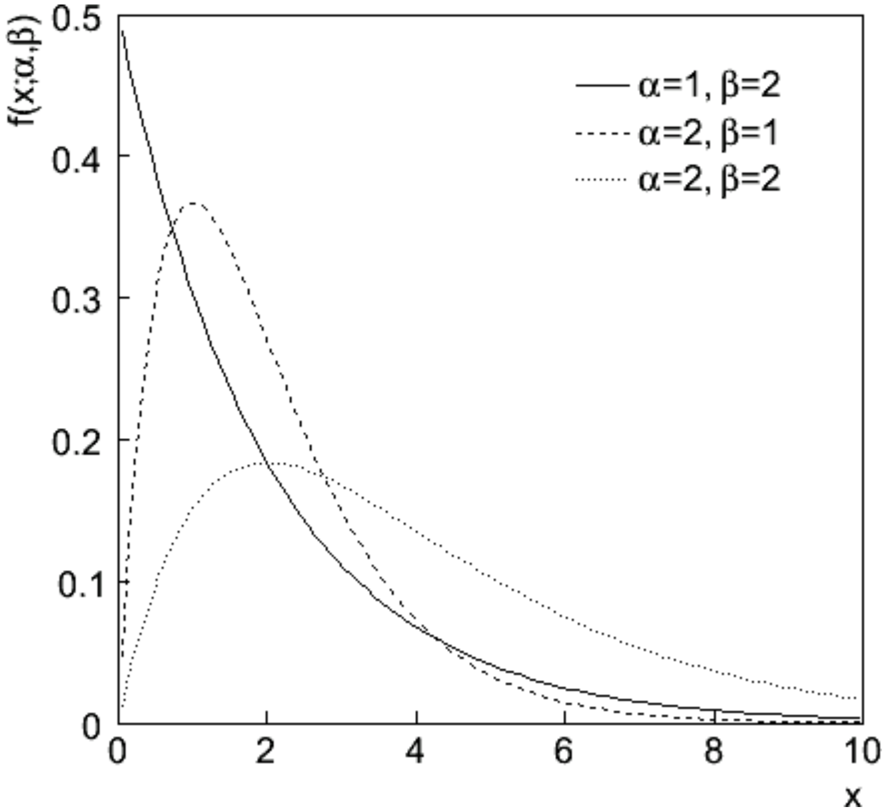
$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}\, x^{\alpha-1}\, e^{-x/\beta}$$

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

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

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

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

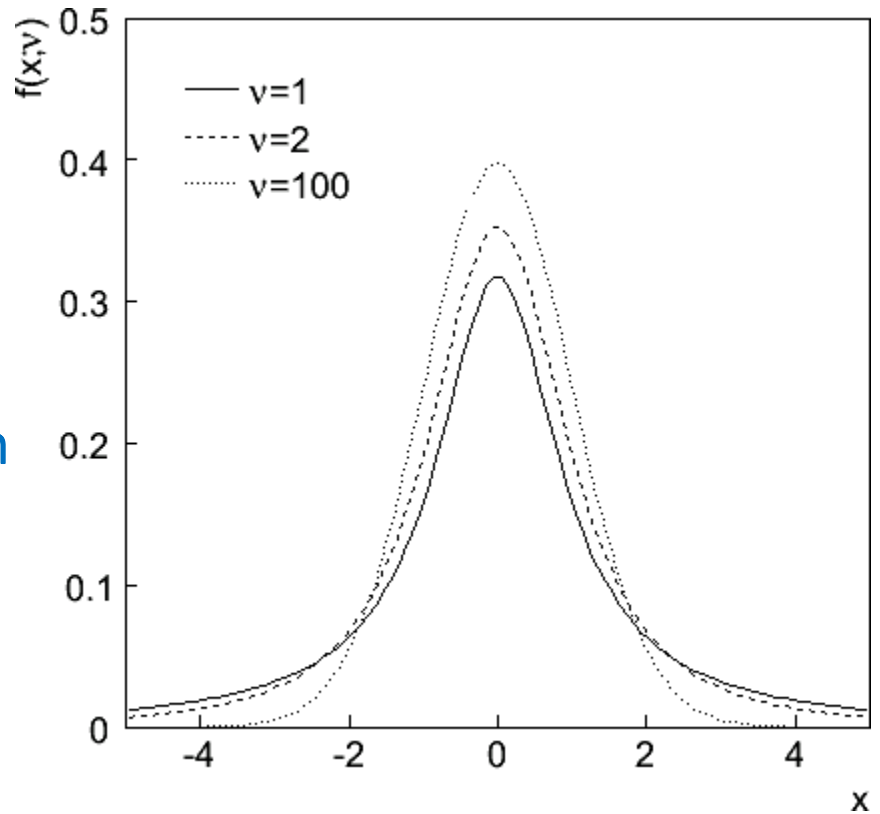
$$E[x] = 0 \quad (\nu > 1)$$

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

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

$\nu = 1$ gives Cauchy,

$\nu \to \infty$ 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 $\nu = n$.

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, ($\nu \rightarrow \infty$, but in fact already very Gauss-like for $\nu =$ two dozen), to the very long-tailed Cauchy ($\nu = 1$).

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