Statistical Methods in Particle Physics

2. Probability Distributions

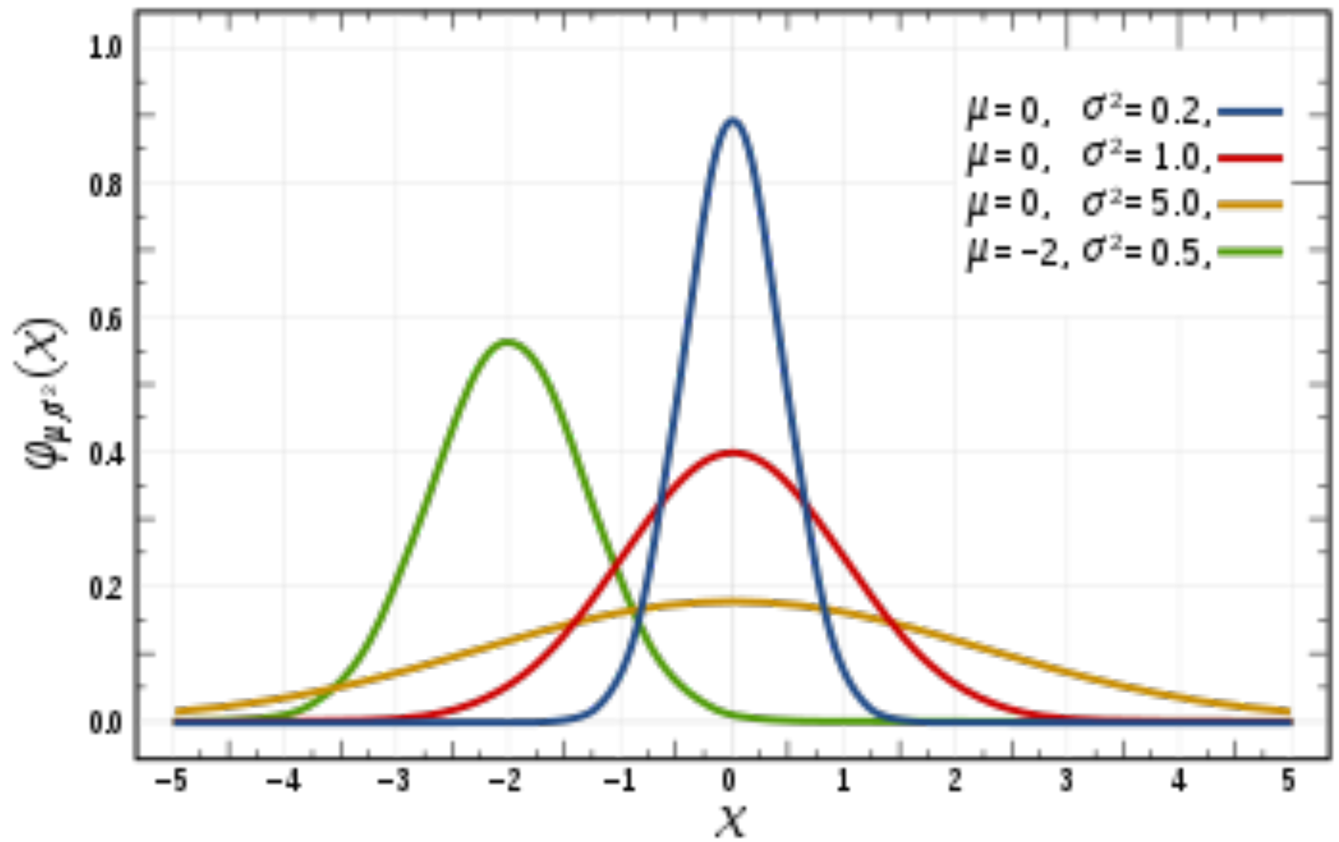
Prof. Dr. Klaus Reygers (lectures)
Dr. Sebastian Neubert (tutorials)

Heidelberg University
WS 2017/18

Gaussian

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

https://en.wikipedia.org/wiki/Normal_distribution



Mean: $E[x] = \mu$

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

μ = 0, σ = 1 ("standard normal distribution"): $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

Cumulative distribution related to error function:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} \, dz = \frac{1}{2}\left[\mathrm{erf}\left(\frac{x}{\sqrt{2}}\right) + 1\right]$$

p-value

Probability for a Gaussian distribution corresponding to [μ – Zσ, μ +Zσ]:

$$P(Z\sigma) = \frac{1}{\sqrt{2\pi}} \int_{-Z}^{+Z} e^{-\frac{x^2}{2}} \, dx = \Phi(Z) - \Phi(-Z) = \mathrm{erf}\left(\frac{Z}{\sqrt{2}}\right)$$

68.27% of area within ±1σ
95.45% of area within ±2σ
99.73% of area within ±3σ

90% of area within ±1.645σ
95% of area within ±1.960σ
99% of area within ±2.576σ

p-value:
probability that a random process produces a measurement thus far, or further, from the true mean

$$p\text{-value} = 1 - P(Z\sigma)$$

In root: `TMath::Prob`

Two-sided Gaussian p-values

Deviation	p-value (%)
1 σ	31.7
2 σ	4.56
3 σ	0.270
4 σ	0.006 33
5 σ	0.000 057 3

standard to report a “discovery” → 5 σ

Why Are Gaussians so Useful?

Central limit theorem:

When independent random variables are added, their properly normalized sum tends toward a normal distribution (a bell curve) even if the original variables themselves are not normally distributed.

More specifically:

Consider *n* random variables with finite variance σ_i^2 and arbitrary pdf:

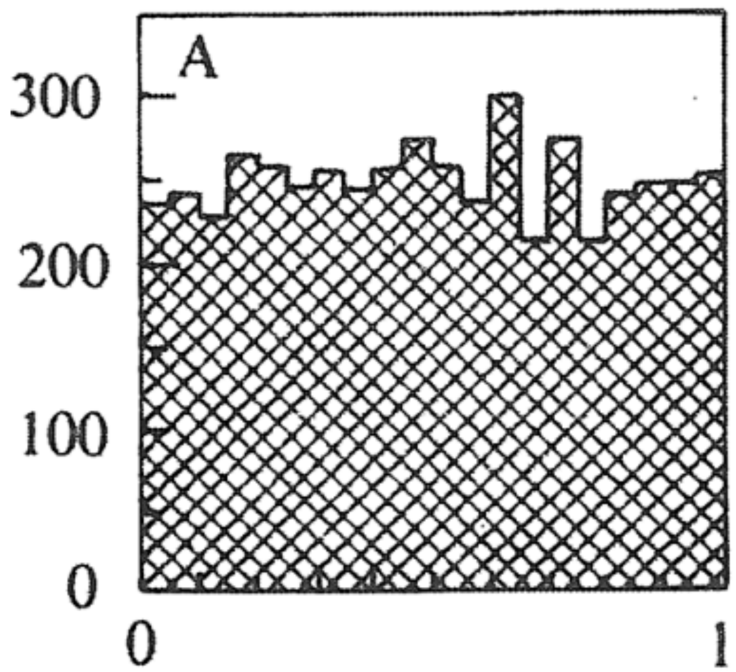
$$y = \sum_{i=1}^{n} x_i \quad \xrightarrow{n \to \infty} \quad E[y] = \sum_{i=1}^{n} \mu_i \qquad V[y] = \sum_{i=1}^{n} \sigma_i^2$$

Measurement uncertainties are often the sum of many independent contributions. The underlying pdf for a measurement can therefore be assumed to be a Gaussian.

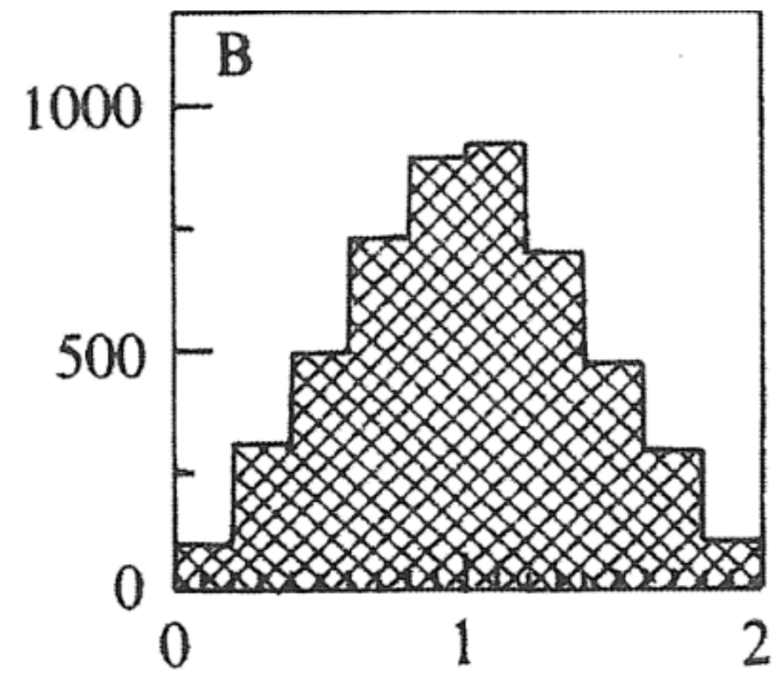
Many convenient features in addition, e.g., sum or difference of two Gaussian random variables is again a Gaussian.

The CLT at Work

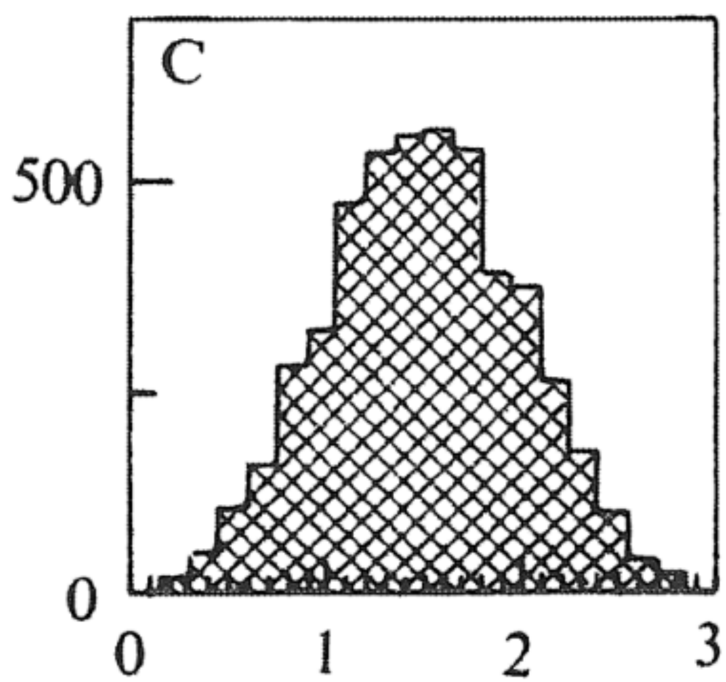
A: x taken from a uniform PD in [0,1], with $\mu=0.5$ and $\sigma^2=1/12$, N=5000



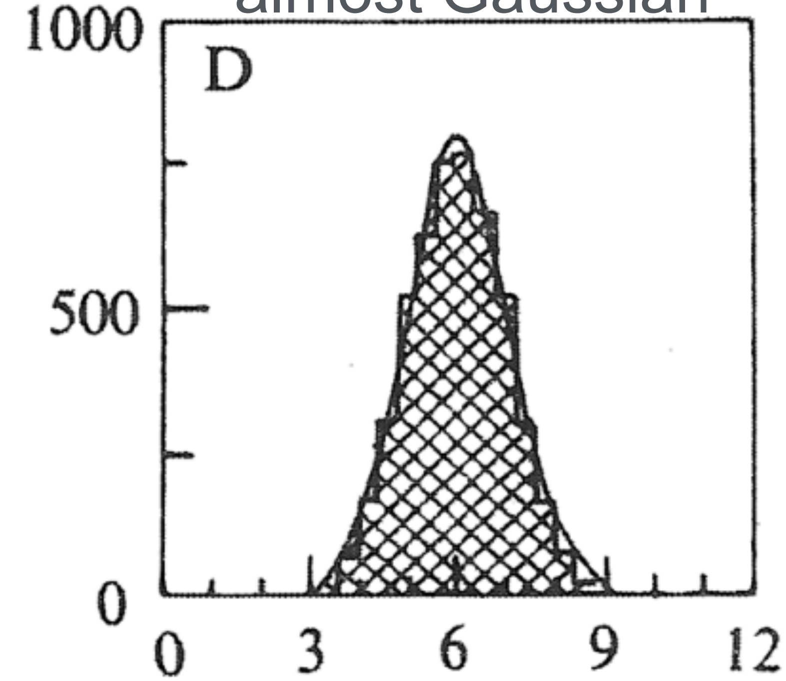
B: $X = x_1+x_2$ from A, N=5000, flat shoulders



C: $X = x_1+x_2+x_3$ from A, curved shoulders



D: $X=x_1+x_2+\ldots+x_{12}$ from A, almost Gaussian



Multivariate Gaussian

transposed (row) vectors — column vectors

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

$$\vec{x} = (x_1, \ldots, x_n), \qquad \vec{\mu} = (\mu_1, \ldots, \mu_n)$$

$$E[x_i] = \mu_i \qquad V_{i,j} = \mathrm{cov}[x_i, x_j] = \langle (x_i - \mu_i)(x_j - \mu_j) \rangle$$

For $n = 2$:

$$V = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \quad \rightsquigarrow \quad V^{-1} = \frac{1}{(1-\rho^2)} \begin{pmatrix} 1/\sigma_x^2 & -\rho/(\sigma_x\sigma_y) \\ -\rho/(\sigma_x\sigma_y) & 1/\sigma_y^2 \end{pmatrix}$$

ρ = correlation coefficient

2d Gaussian Distribution and Error Ellipse

We obtain the 2d Gaussian distribution:

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

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

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

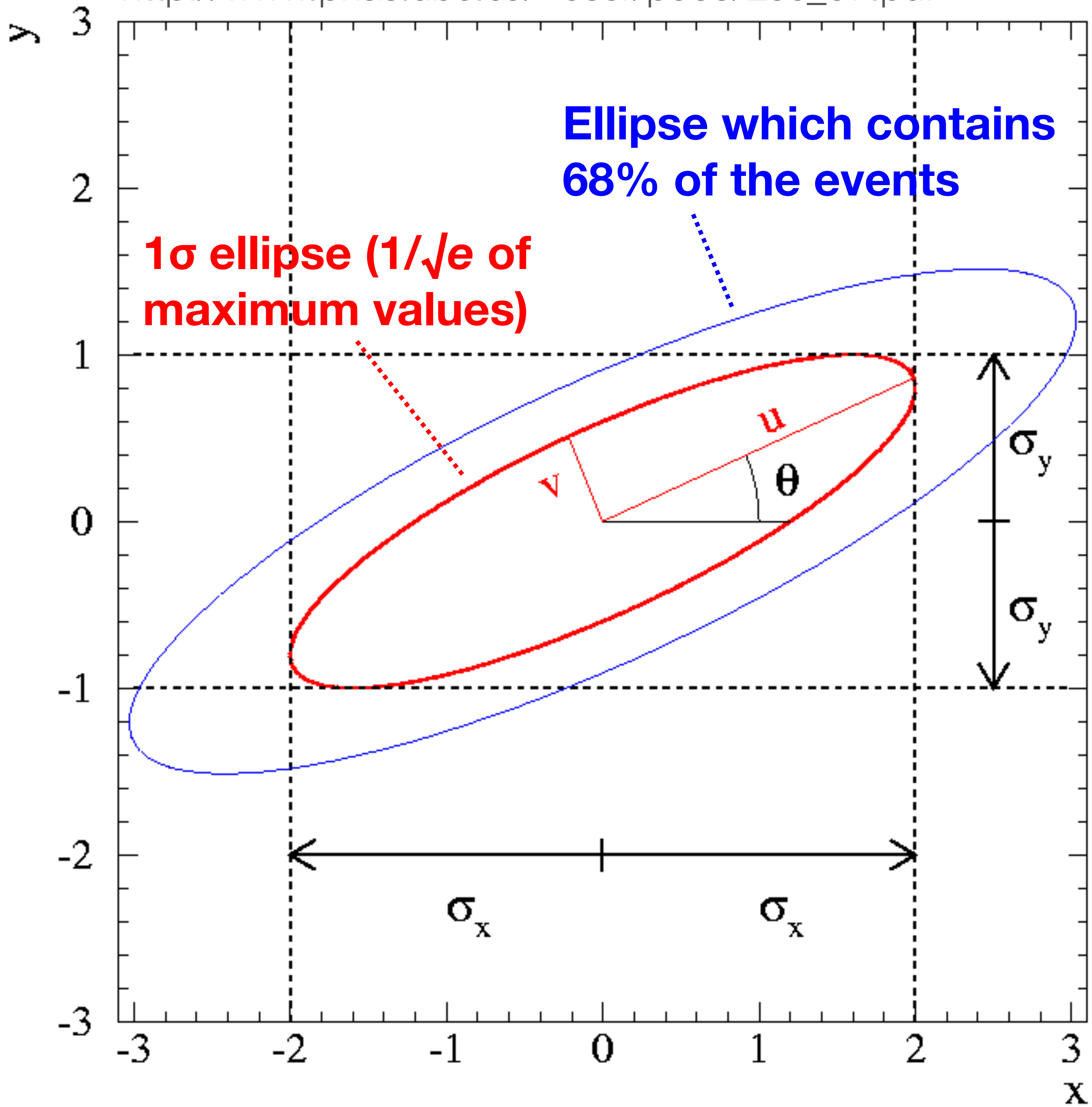
Lines of constant probability correspond to constant argument of exp
→ this defines an ellipse

1σ ellipse: $f(x_1, x_2)$ has dropped to $1/\sqrt{e}$ of its maximum value
(argument of exp is –1/2):

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

2d Gaussian: Error Ellipse

http://www.phas.ubc.ca/~oser/p509/Lec_07.pdf



$$f_y(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y$$

$$= \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right)$$

$$f_x(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left(-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right)$$

	P_{1D}	P_{2D}
1σ	0.6827	0.3934
2σ	0.9545	0.8647
3σ	0.9973	0.9889
1.515σ		0.6827
2.486σ		0.9545
3.439σ		0.9973

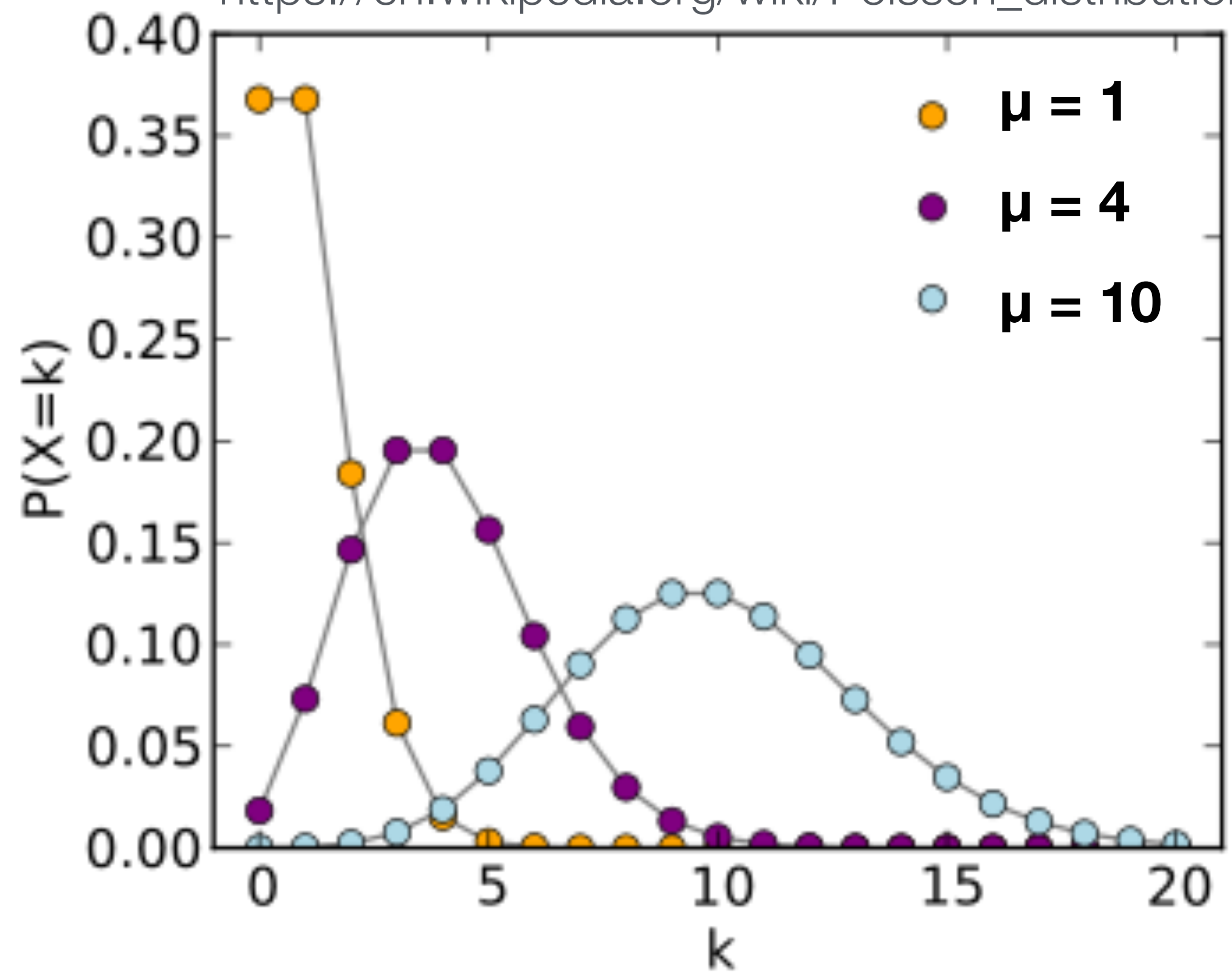
Probability for an event to be within 1σ ellipse: 39.34%

Poisson Distribution

$$p(k;\mu) = \frac{\mu^k}{k!} e^{-\mu}$$

$$E[k] = \mu, \quad V[k] = \mu$$

https://en.wikipedia.org/wiki/Poisson_distribution



Properties:

- n_1, n_2 follow Poisson distr.
  → n_1+n_2 follows Poisson distr., too
- Can be approximated by a Gaussian for large v

Examples:

- Clicks of a Geiger counter in a given time interval
- Number of Prussian cavalrymen killed by horse-kicks

Number of deaths in 1 corps in 1 year	Actual number of such cases	Poisson prediction
0	109	108.7
1	65	66.3
2	22	20.2
3	3	4.1
4	1	0.6

Binomial Distribution

N independent experiments

- Outcome of each is 'success' or 'failure'
- Probability for success is p

$$f(k; N, p) = \binom{N}{k} p^k (1-p)^{N-k} \qquad E[k] = Np \qquad V[k] = Np(1-p)$$

$$\binom{N}{k} = \frac{N!}{k!(N-k)!}$$

binomial coefficient: number of different ways (permutations) to have k successes in N tries

Use binomial distribution to model processes with two outcomes

- Example: Detection efficiency (either we detect particle or not)

For small p, the binomial distribution can be approximated by a Poisson distribution (more exactly, in the limit $N \to \infty$, $p \to 0$, $N \cdot p$ constant)

Negative Binomial Distribution

Keep number of successes k fixed and ask for the probability of m failures before having k successes:

$$P(m; k, p) = \binom{m+k-1}{m} p^k (1-p)^m$$

$$m = 0, 1, ..., \infty$$

$$E[m] = k\frac{1-p}{p}$$

$$V[m] = k\frac{1-p}{p^2}$$

Another representation:

$$P(m; \mu, k) = \binom{m+k-1}{m} \frac{\left(\frac{\mu}{k}\right)^m}{\left(1+\frac{\mu}{k}\right)^{m+k}}$$

$$E[m] = \mu$$

$$V[m] = \mu\left(1+\frac{\mu}{k}\right)$$

$$p = \frac{1}{1+\frac{\mu}{k}}$$ [relation btw. parameters]

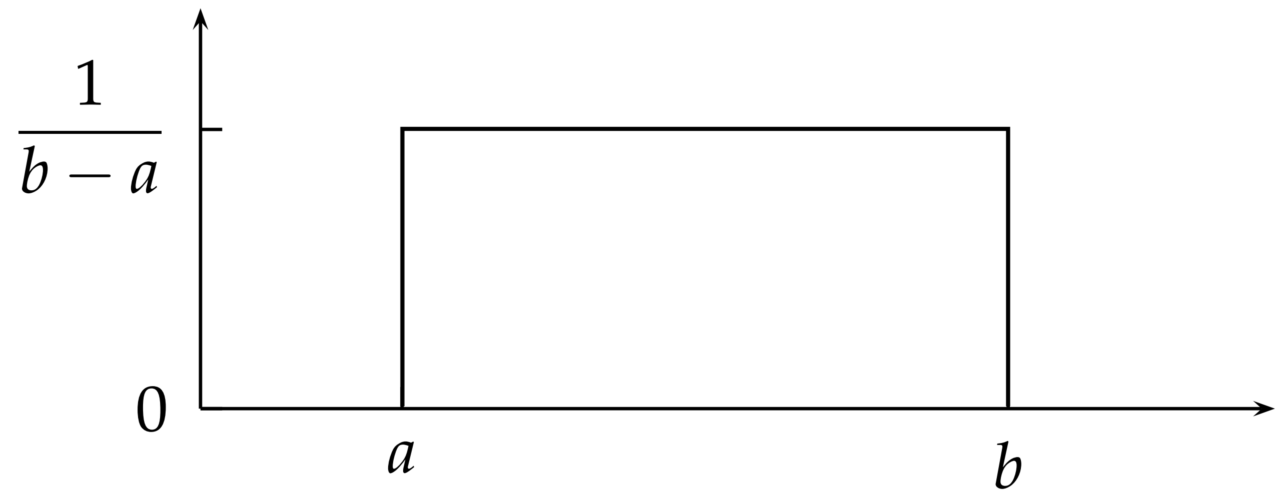
Use Gamma-fct. for non-integer values

$$x! := \Gamma(x+1)$$

Example: Distribution of the number of produced particles in e^+e^- and proton-proton collisions reasonably well described by a NBD. Why? Empirical observation, not so obvious.

Uniform Distribution

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



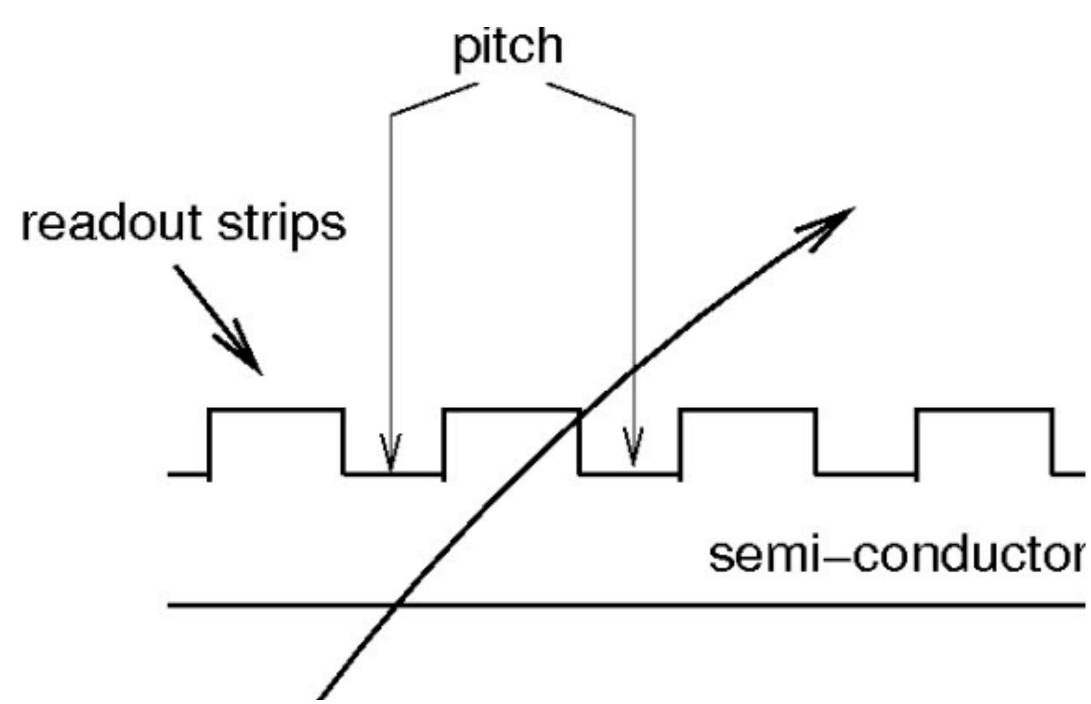
Properties:

$$E[x] = \frac{1}{2}(a+b)$$

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

Example:

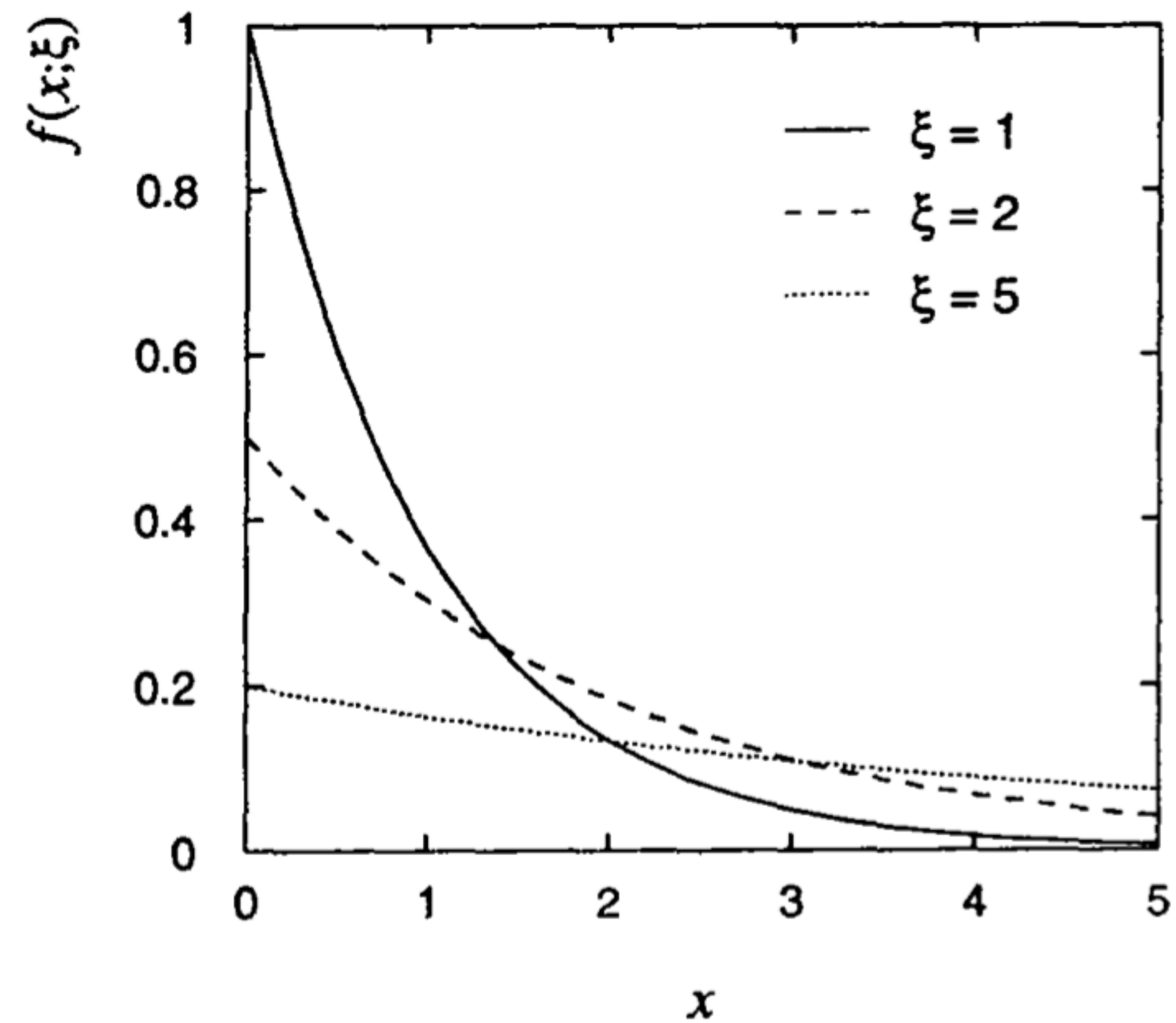
- Strip detector:
  resolution for one-strip clusters:
  pitch/$\sqrt{12}$



Exponential Distribution

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

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



Example:

Decay time of an unstable particle at rest

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

Lack of memory (unique to exponential): $f(t > t_0 + t_1 | t > t_0) = f(t > t_1)$

Probability for an unstable nucleus to decay in the next minute is independent of whether the nucleus was just created or already existed for a million years

Landau Distribution

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

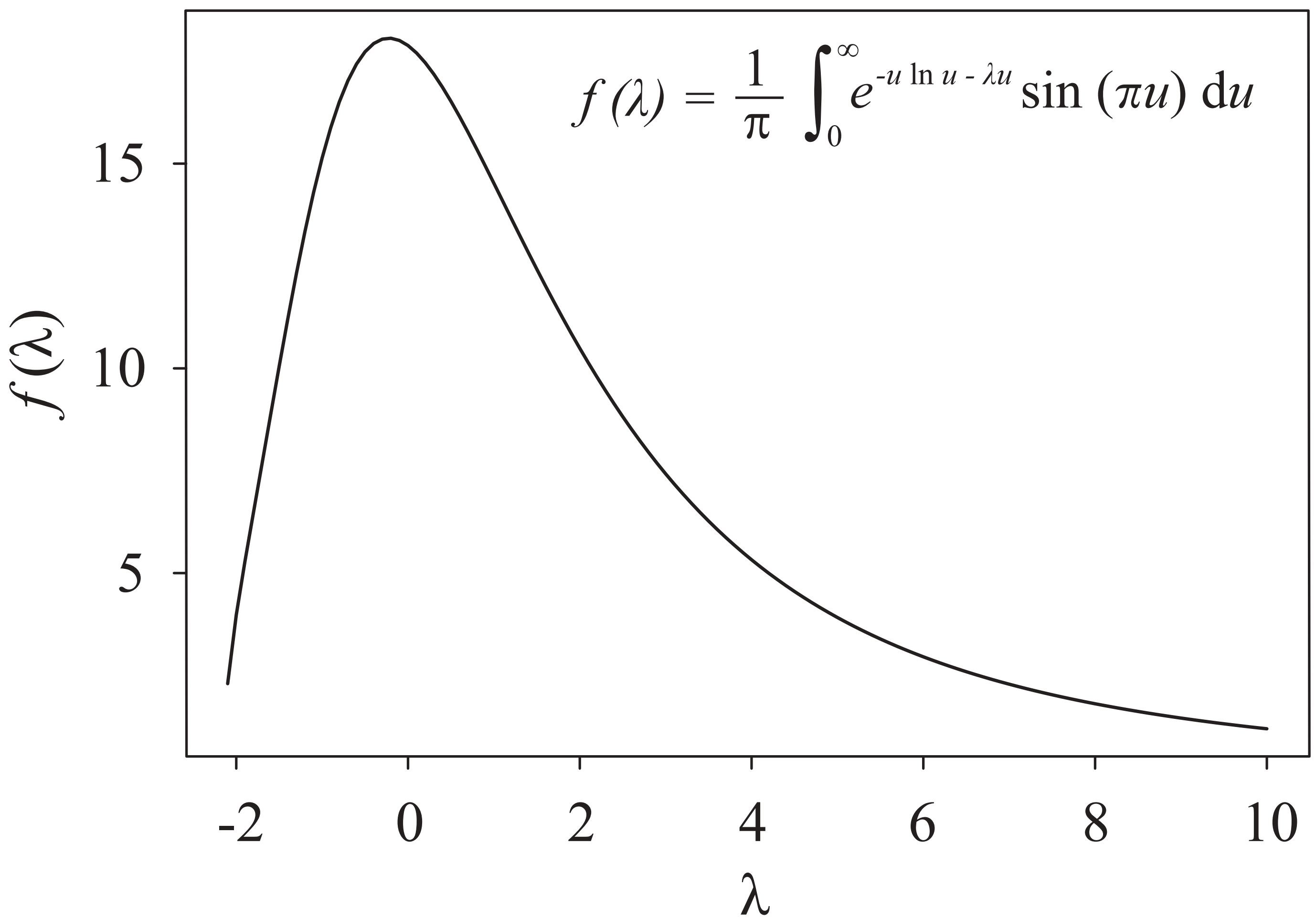
Describes energy loss of a charged particle in a thin layer of material

- tail with large energy loss due to occasional creation of delta rays

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

$$\lambda = \frac{\Delta - \Delta_0}{\xi}$$

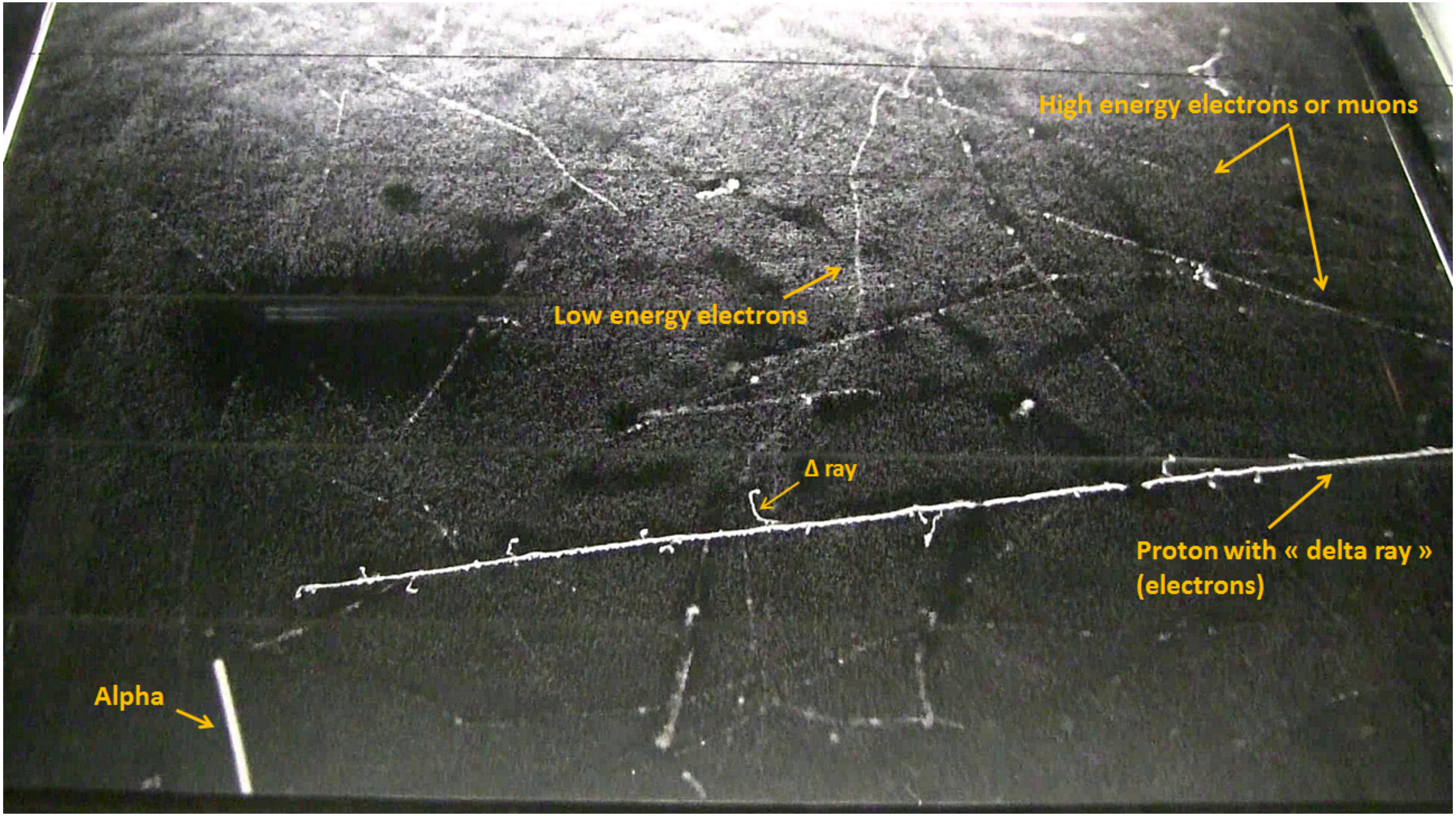
actual energy loss (Δ), location parameters (Δ_0), material property (ξ)



Unpleasant mathematical properties: mean and variance not defined

root: `TMath::Landau()`

[Delta rays]



https://en.wikipedia.org/wiki/Delta_ray

Student's t Distribution

Developed in 1908 by William Gosset for the Guinness Brewery. Published under the name "student".

Let $x_1, \ldots, x_n$ be distributed as $N(\mu, \sigma)$.

Sample mean and estimate of the variance:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

How Student's distribution arises from sampling:

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

→ follows standard normal distr. (μ=0, σ=1)

$$\frac{\bar{x}-\mu}{\hat{\sigma}/\sqrt{n}}$$

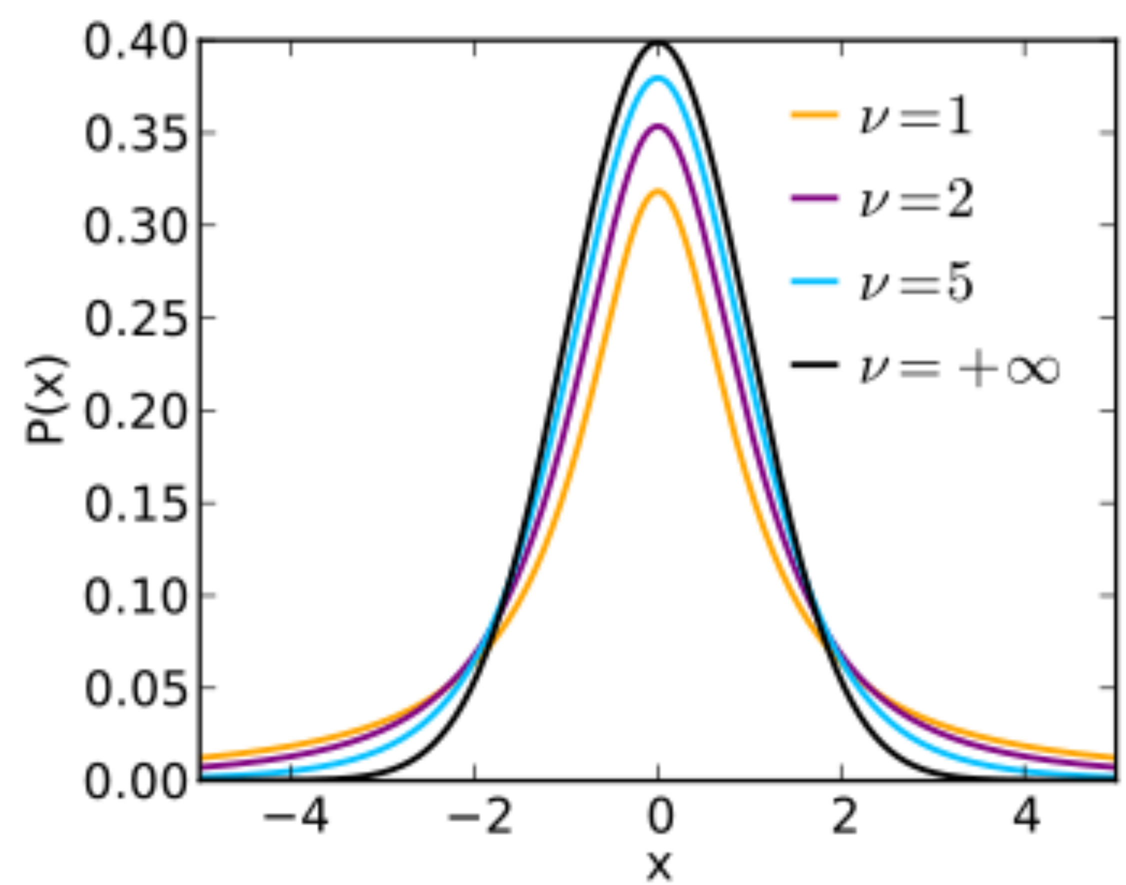
→ not Gaussian. Student's t distr. with n–1 degrees of freedom

Student's t distribution:

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

$n = 1$: Cauchy distr.

$n \to \infty$: Gaussian



χ^2 Distribution

Let $x_1, \ldots, x_n$ be n independent standard normal ($\mu = 0$, $\sigma = 1$) random variables. Then the sum of their squares

$$z = \sum_{i=1}^{n} x_i^2$$

follows a χ^2 distribution with n degrees of freedom.

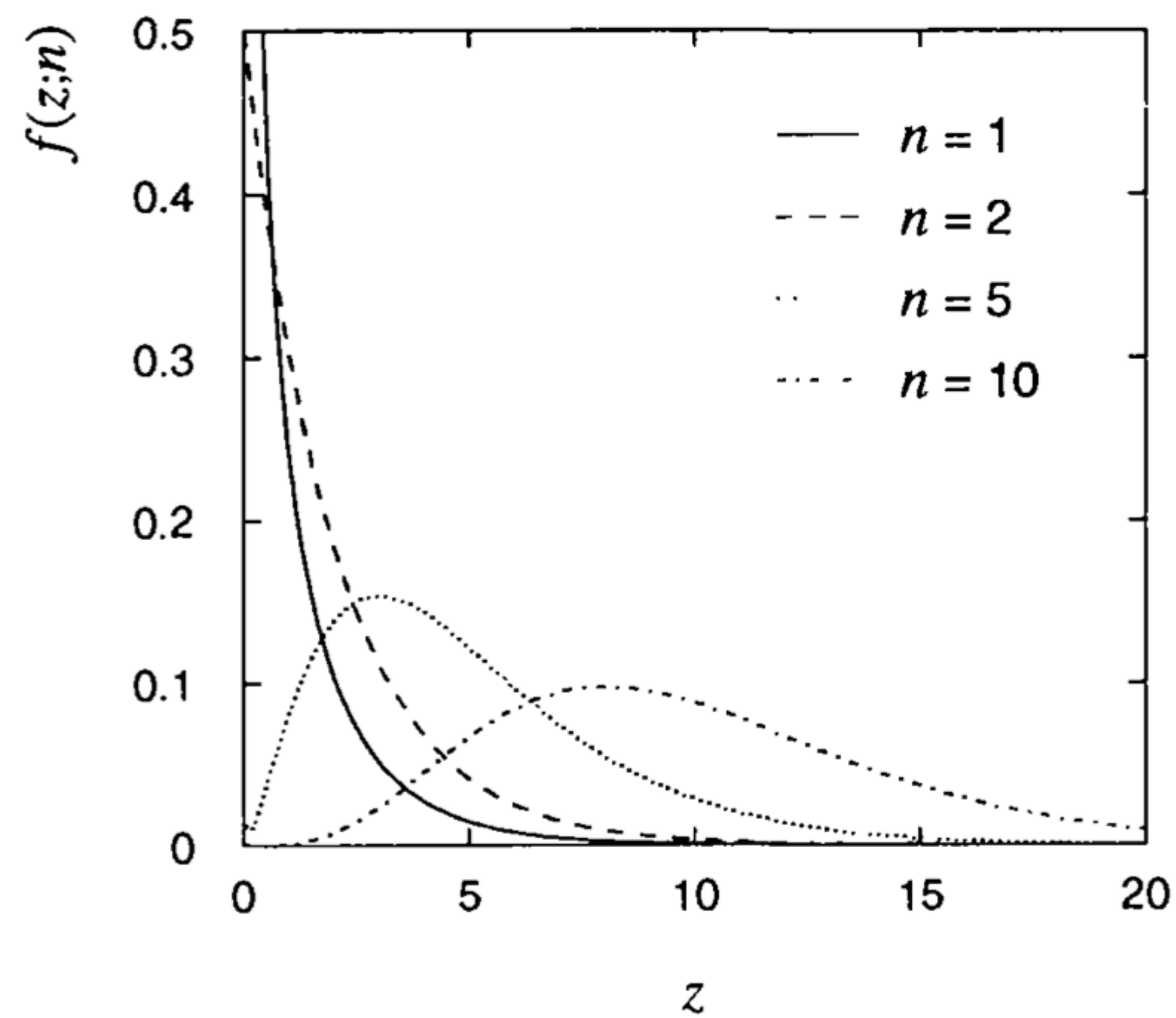
χ^2 distribution:

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

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

Application:
Quantifies goodness of fit

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{y_i - g(x_i)}{\sigma_i} \right)^2$$



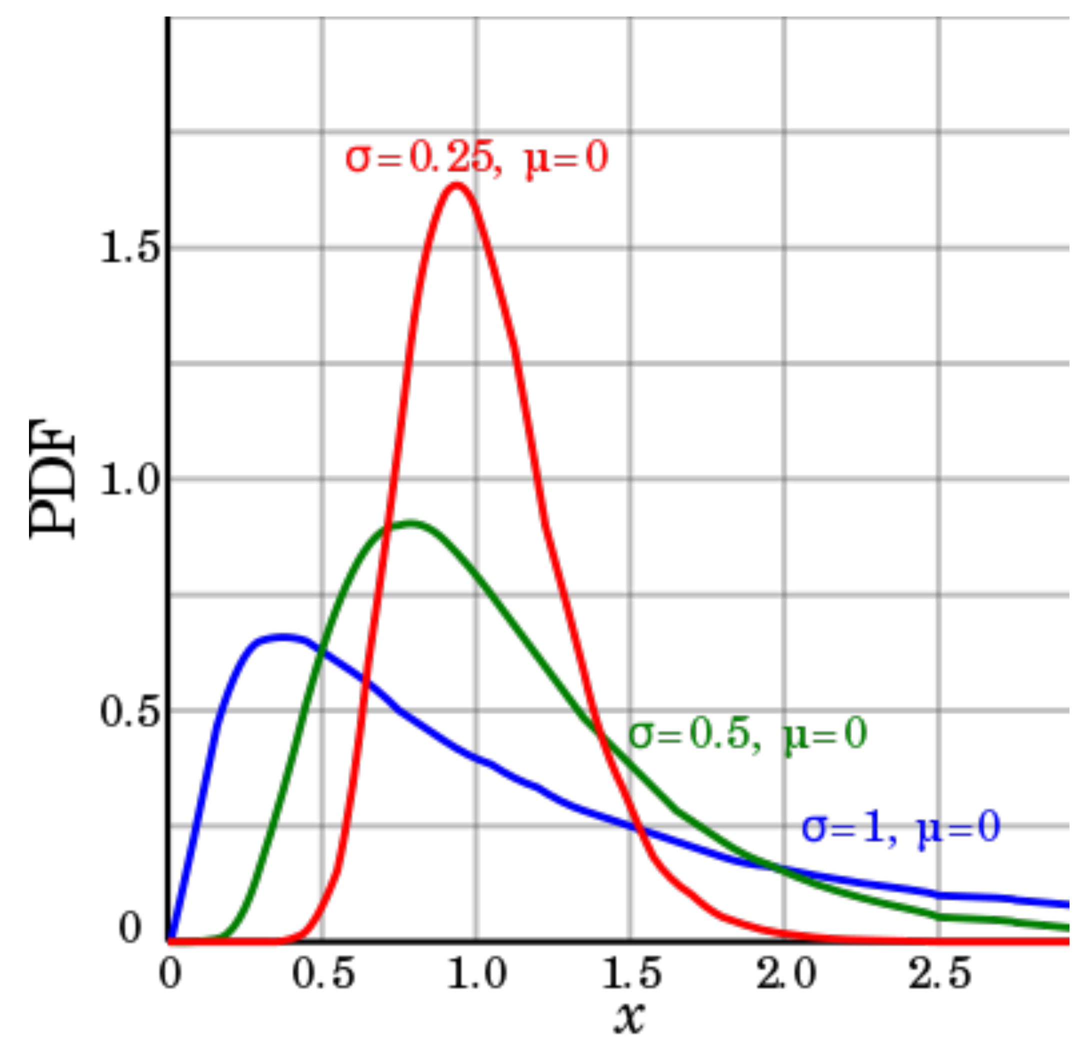
Log-Normal Distribution

Let y be a normal (i.e. Gaussian) distributed random variable. Then $x = \exp(y)$ follows the log-normal distribution

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

$$E[x] = \exp\left(\mu+\frac{\sigma^2}{2}\right)$$

$$V[x] = [\exp(\sigma^2)-1]\exp(2\mu+\sigma^2)$$



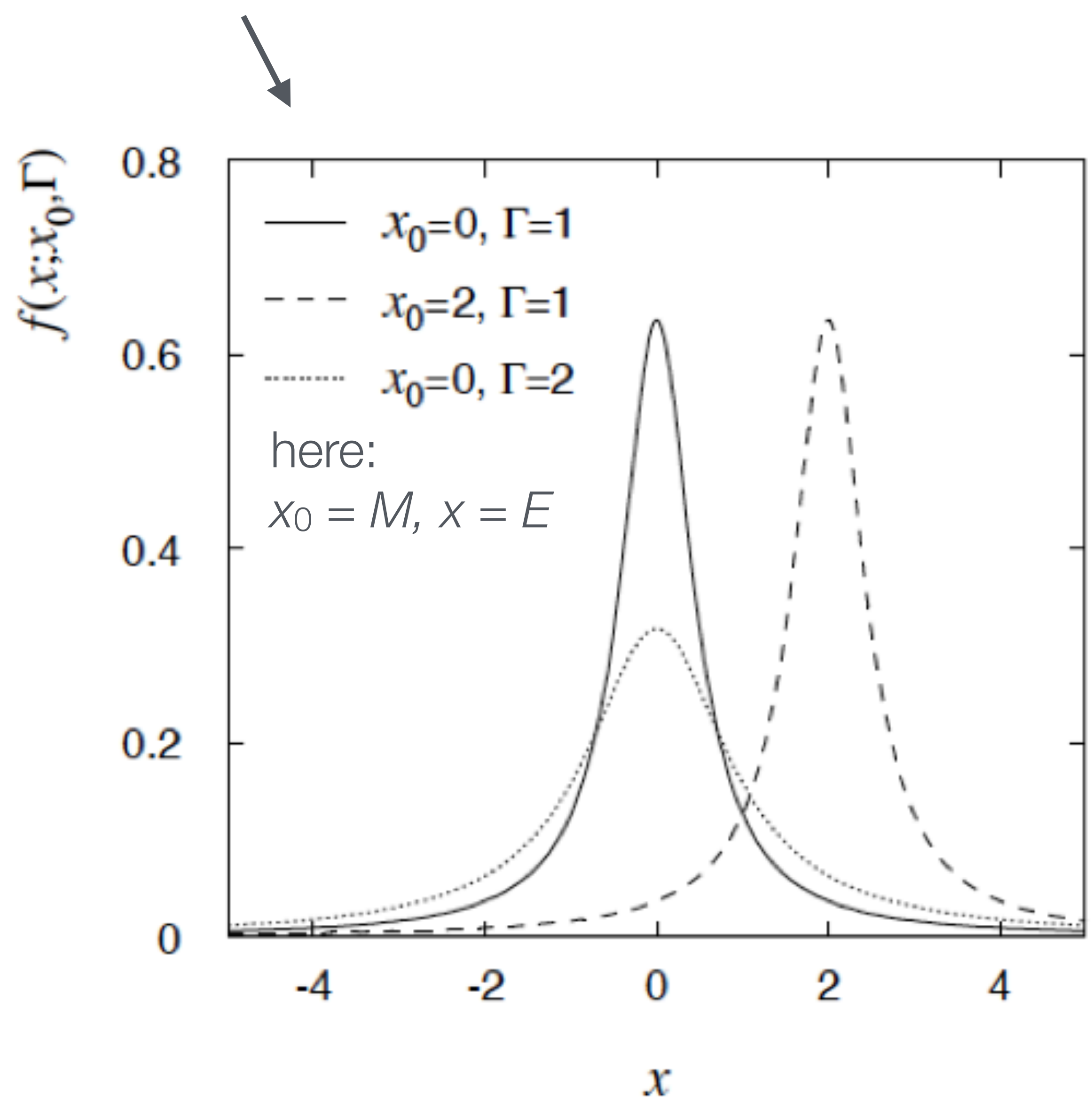
Multiplicative version of the central limit theorem

- Relevant when observable is product of fluctuating variables
- Occurs frequently, e.g., city sizes

Cauchy, Breit-Wigner, or Lorentzian Distribution

Particle physics: cross section for production of resonance with mass M and width Γ (full width at half maximum):

$$f(E; M, \Gamma) = \frac{1}{2\pi} \frac{\Gamma}{(E - M)^2 + (\Gamma/2)^2}$$



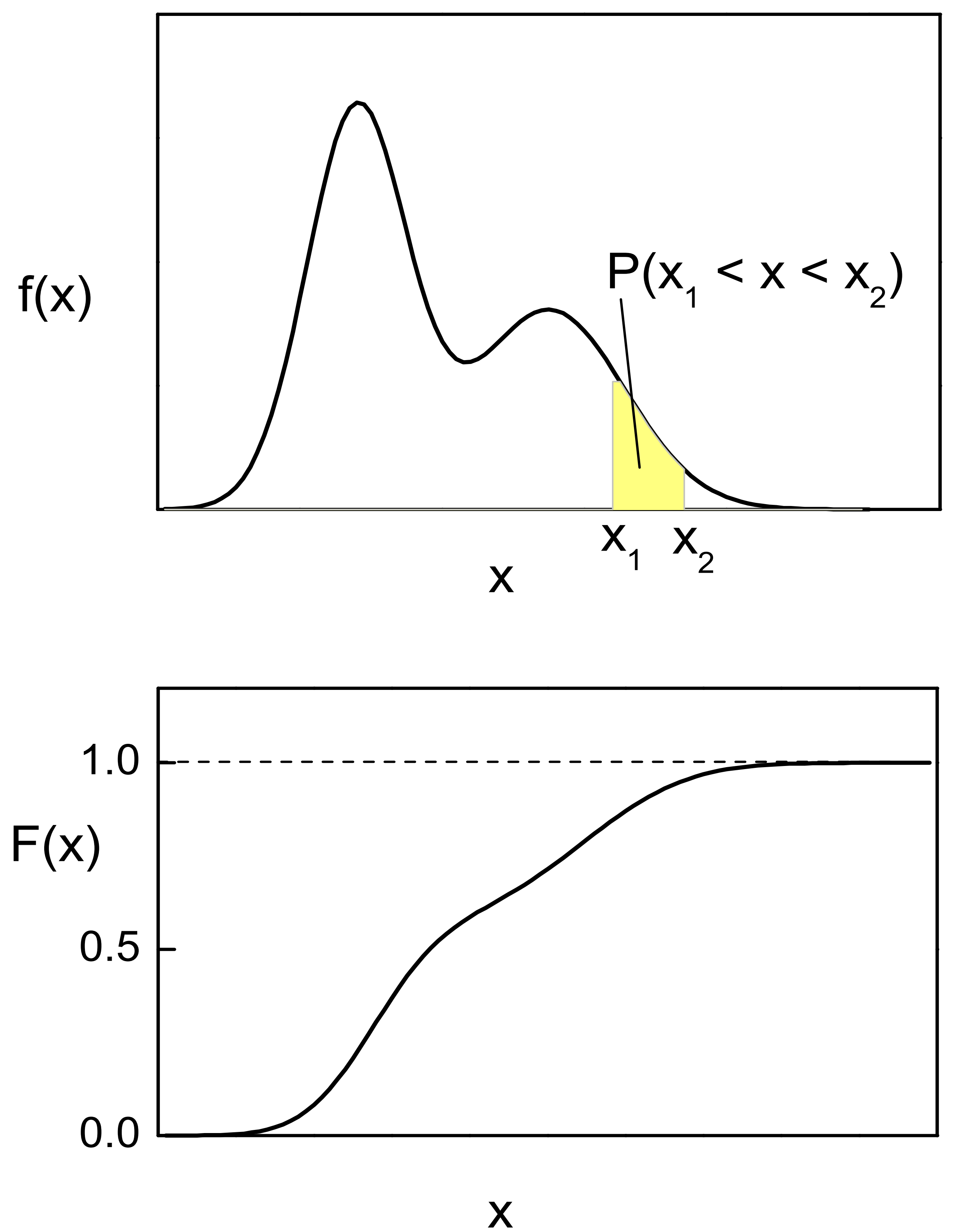
Dimensionless form:

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2} \qquad x = \frac{E - M}{\Gamma/2}$$

Mean and variance are undefined, mode is M.

Cumulative Distribution Function

$$F(X) := \int_{-\infty}^{X} f(x')\,\mathrm{d}x'$$



Convolution of Probability Distributions

$f(x)$: probability distribution of random variable x
$g(y)$: probability distribution of random variable y

PDF for sum

$$z = x + y$$

is given by:

$$h(z) = (f * g)(z) = \int_{-\infty}^{\infty} f(z-t)g(t)dt = \int_{-\infty}^{\infty} f(t)g(z-t)dt$$

Example: Two Gaussians $N(x; \mu_x, \sigma_x)$, $N(y; \mu_y, \sigma_y)$

→ Sum $z = x + y$ follows a Gaussian with $\mu = \mu_x + \mu_y, \quad \sigma = \sqrt{\sigma_x^2 + \sigma_y^2}$

Note: Product $x \cdot y$ and ratio of x/y of two Gaussian distributed random variables is not a Gaussian