Statistical Methods in Particle Physics

2. Probability Distributions

Heidelberg University, WS 2020/21

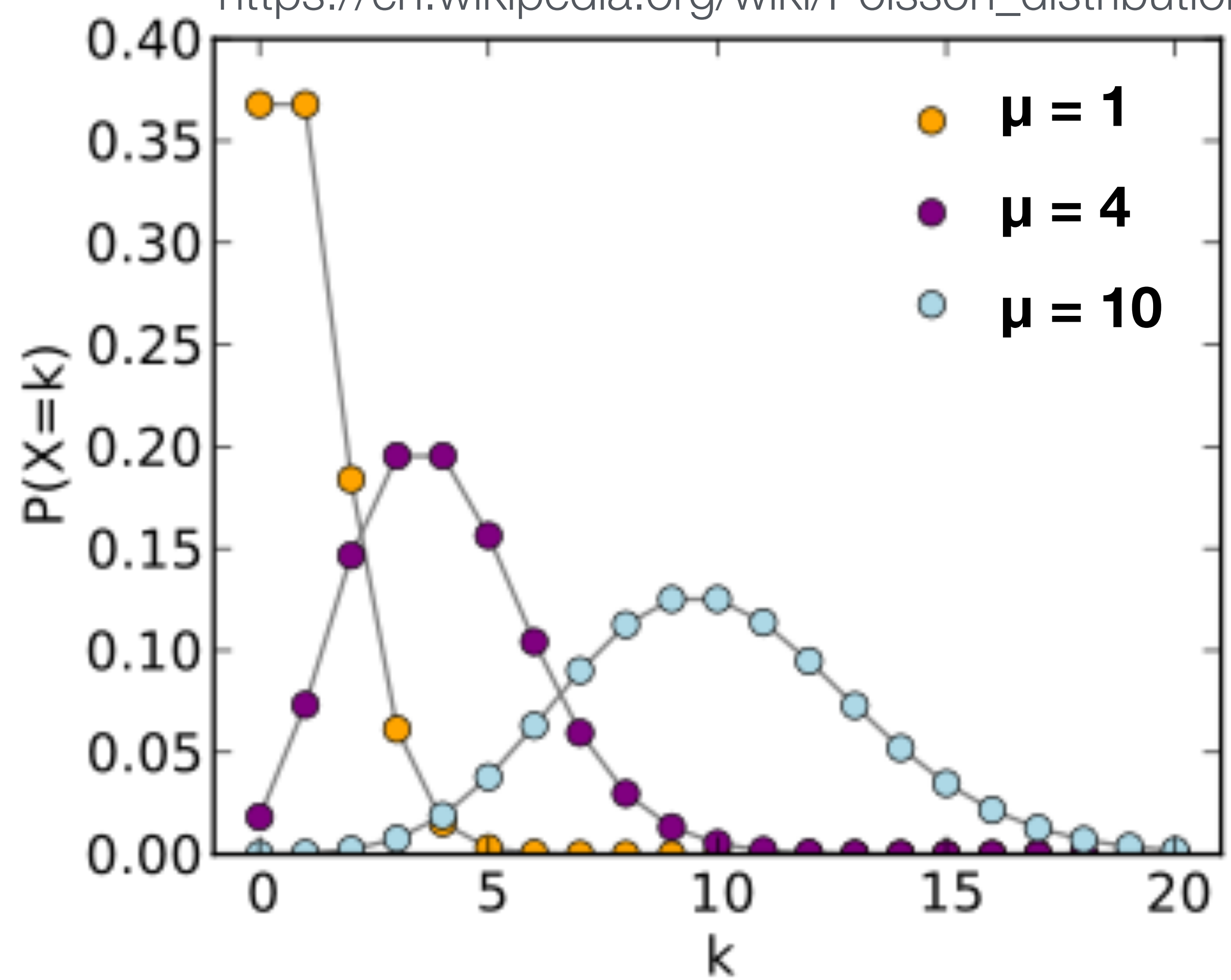
Klaus Reygers (lectures)
Rainer Stamen, Martin Völkl (tutorials)

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 μ

Counting experiment: $\sigma(n) = \sqrt{n}$

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

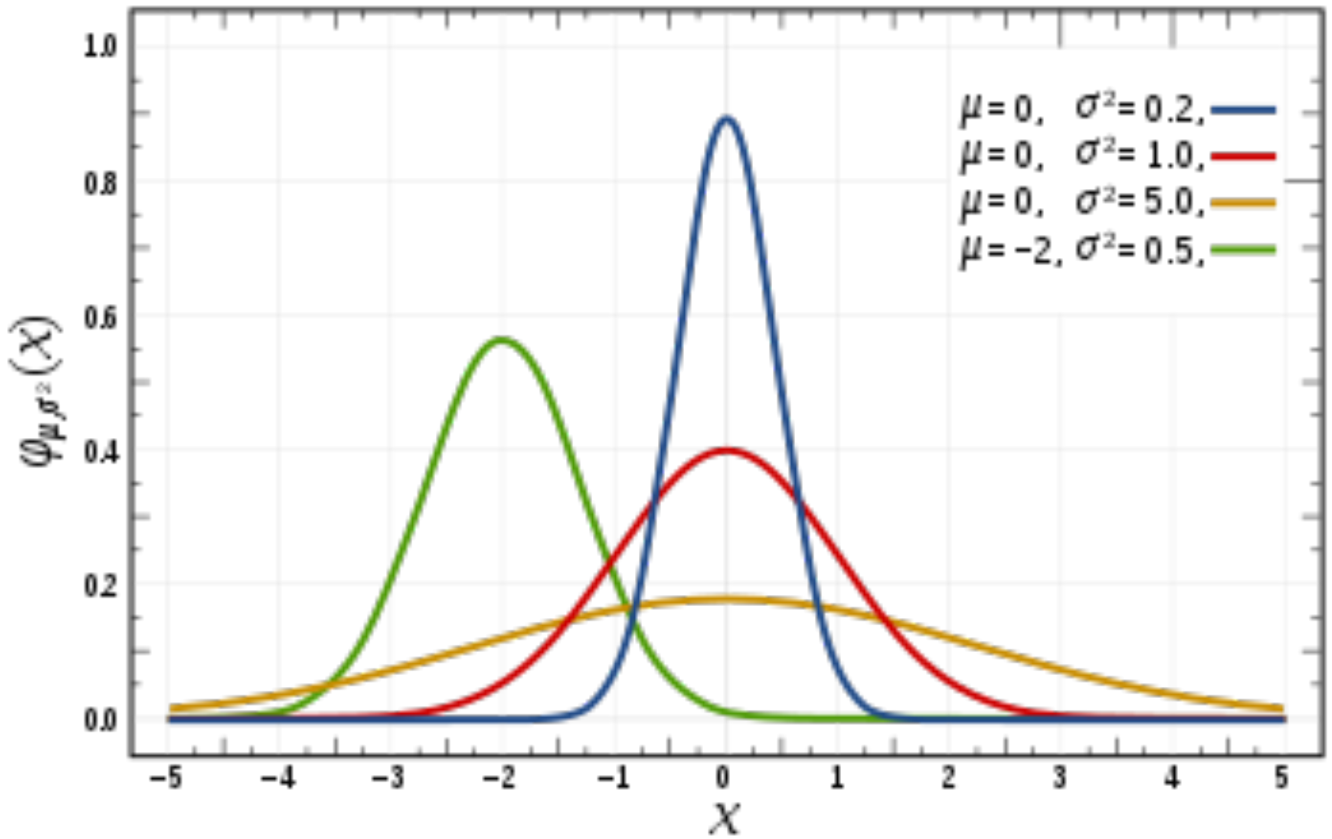
- Example: Detection efficiency (either we detect particle or not)

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)

Normal (or Gaussian) distribution

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

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



$$E[x] = \mu$$

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

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

Cumulative distribution function:

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

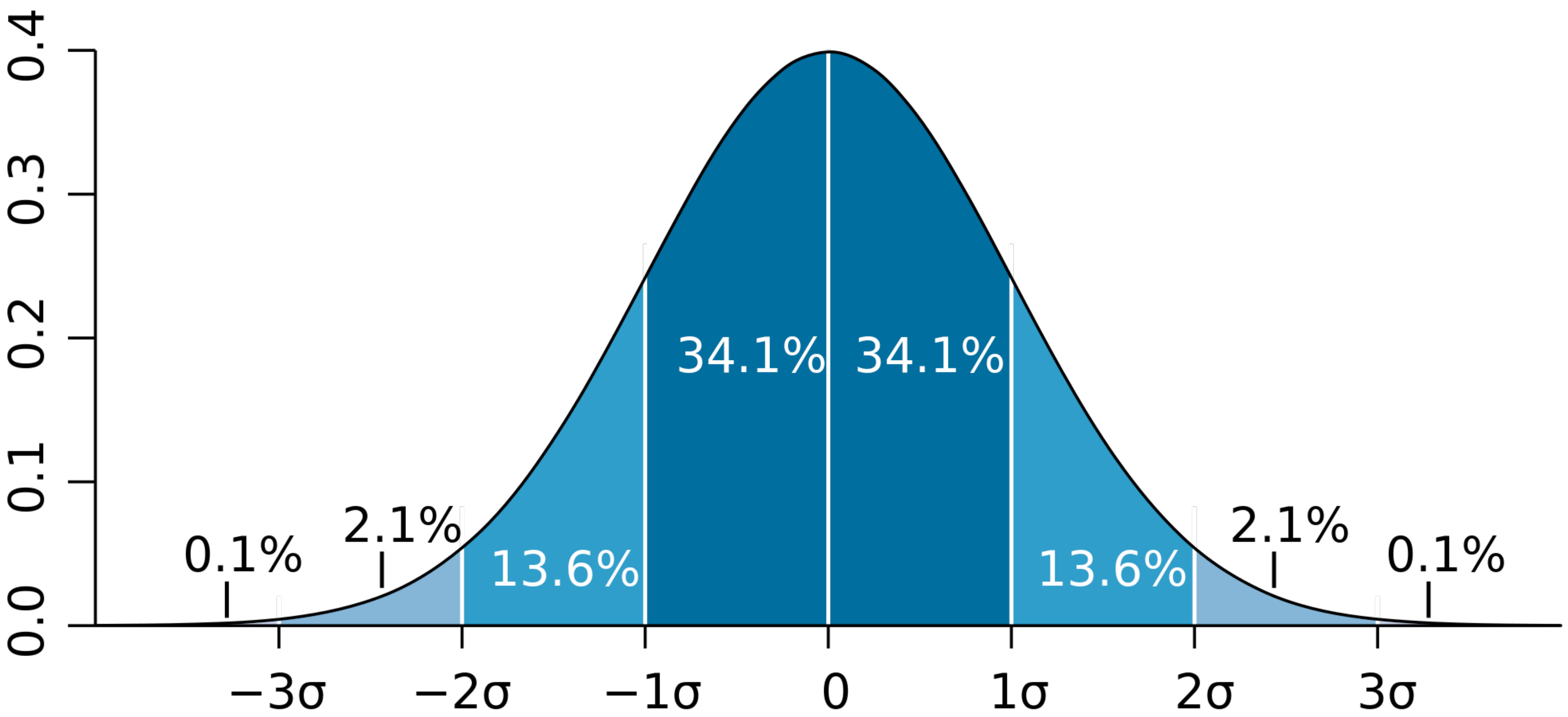
Binomial, Poisson and Normal Distribution

$$
\begin{tikzcd}[column sep=huge, row sep=huge]
\boxed{\begin{array}{c}\text{Binomial}\\ B(k;N,p)\end{array}} \arrow[rr, "N \to \infty,\ p \to 0,\ Np = \mu \text{ fixed}"] \arrow[dr, "N \to \infty"'] & & \boxed{\begin{array}{c}\text{Poisson}\\ P(k;\mu)\end{array}} \arrow[dl, "\mu \to \infty"] \\
& \boxed{\begin{array}{c}\text{Normal}\\ N(x;\mu,\sigma)\end{array}} &
\end{tikzcd}
$$

Poisson $P(k;\mu)$: $$\frac{k-\mu}{\sqrt{\mu}} \to N(0,1) \quad \text{as} \quad \mu \to \infty$$

Binomial $B(k;n,p)$: $$\frac{k-np}{\sqrt{np(1-p)}} \to N(0,1) \quad \text{as} \quad n \to \infty$$

Deviation in units of σ for a Gaussian



$$P(Z\sigma) = \frac{1}{\sqrt{2\pi}} \int_{-Z}^{+Z} e^{-\frac{x^2}{2}}\, \mathrm{d}x$$

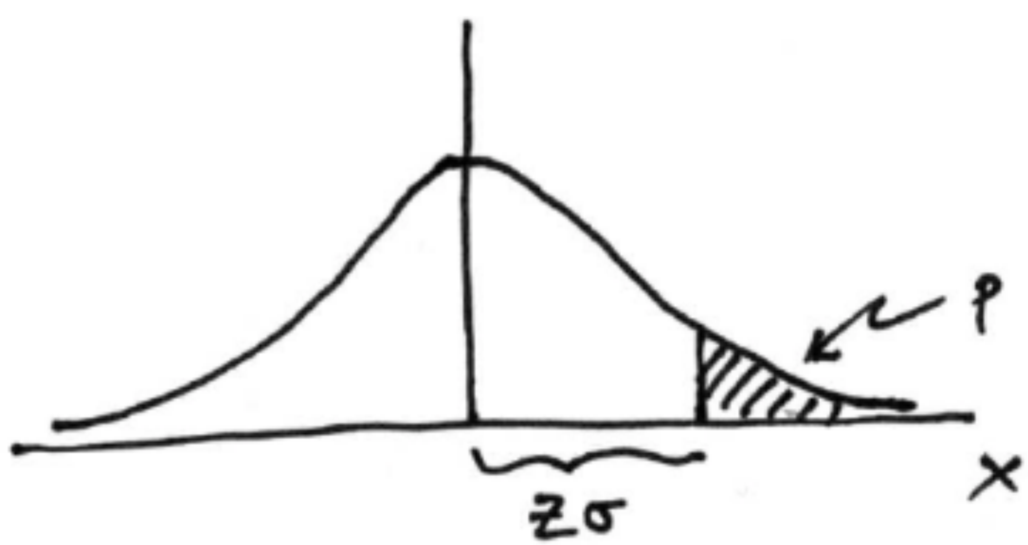
68.27% of area within ±1σ
95.45% of area within ±2σ
99.73% of area within ±3σ

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

p-value and significance

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

p-value and significance (one-tailed):



$$p = 1 - \Phi(Z), \quad Z = \Phi^{-1}(1-p)$$

One-tailed Gaussian p-values

Deviation	p-value
1 σ	0.16
2 σ	0.023
3 σ	0.0013
4 σ	3.2×10^{-5}
5 σ	2.9×10^{-7}

standard to report a "discovery" → 5 σ

https://nbviewer.jupyter.org/urls/www.physi.uni-heidelberg.de/~reygers/lectures/2020/smipp/p-values_and_n-sigma.ipynb

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 pdfs:

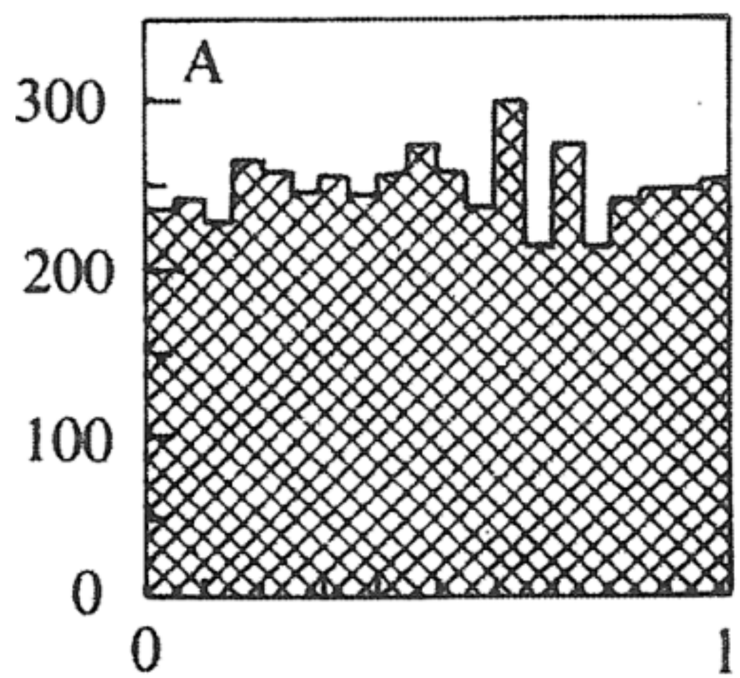
$$y = \sum_{i=1}^{n} x_i \quad \xrightarrow{n \to \infty} \quad y \text{ follows Gaussian with } E[y] = \sum_{i=1}^{n} \mu_i, \; V[y] = \sum_{i=1}^{n} \sigma_i^2$$

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

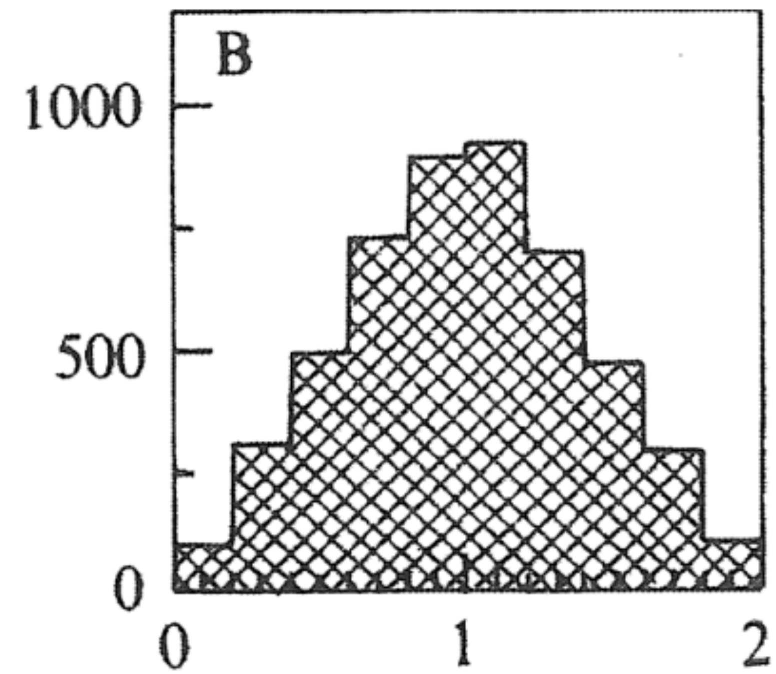
Many convenient features in addition, e.g., sum or difference of two Gaussian random variables is again a Gaussian.

The central limit theorem at work

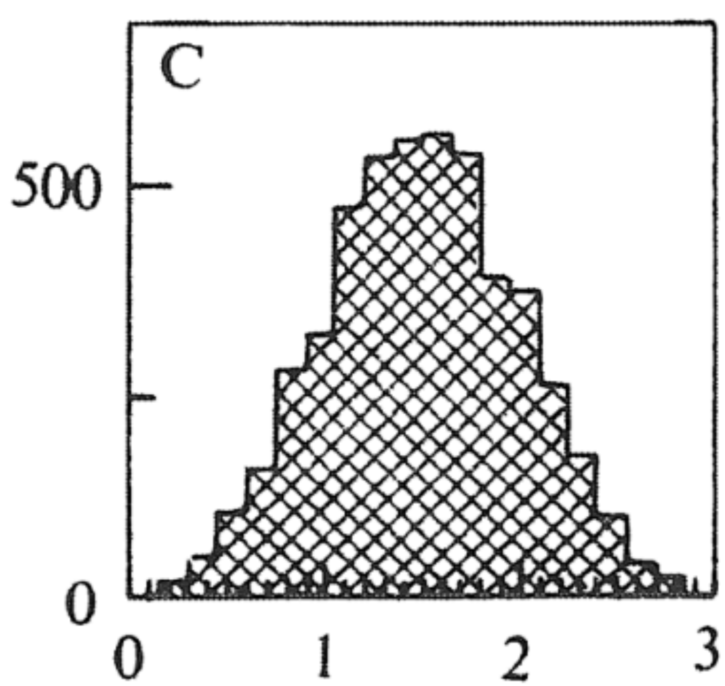
A: x taken from a uniform PD in [0,1], with μ=0.5 and σ²=1/12, N=5000



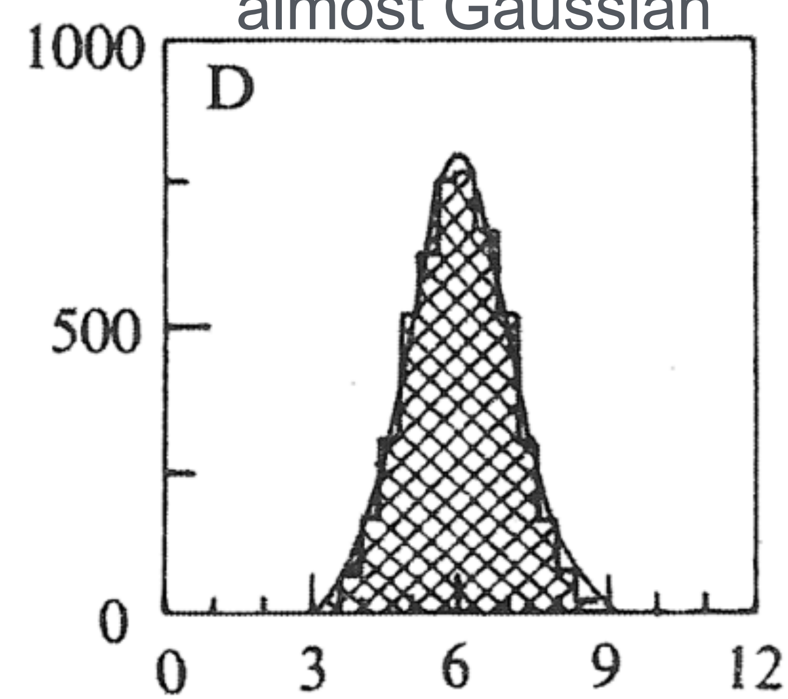
B: $X = x_1+x_2$ from A, N=5000, flat shoulders



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 normal distribution

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

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

Mean: $E[x_i] = \mu_i$ Covariance: $\mathrm{cov}[x_i, x_j] = V_{i,j}$

For $n = 2$:

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

ρ = correlation coefficient

2d Gaussian distribution and error ellipse

2d 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.

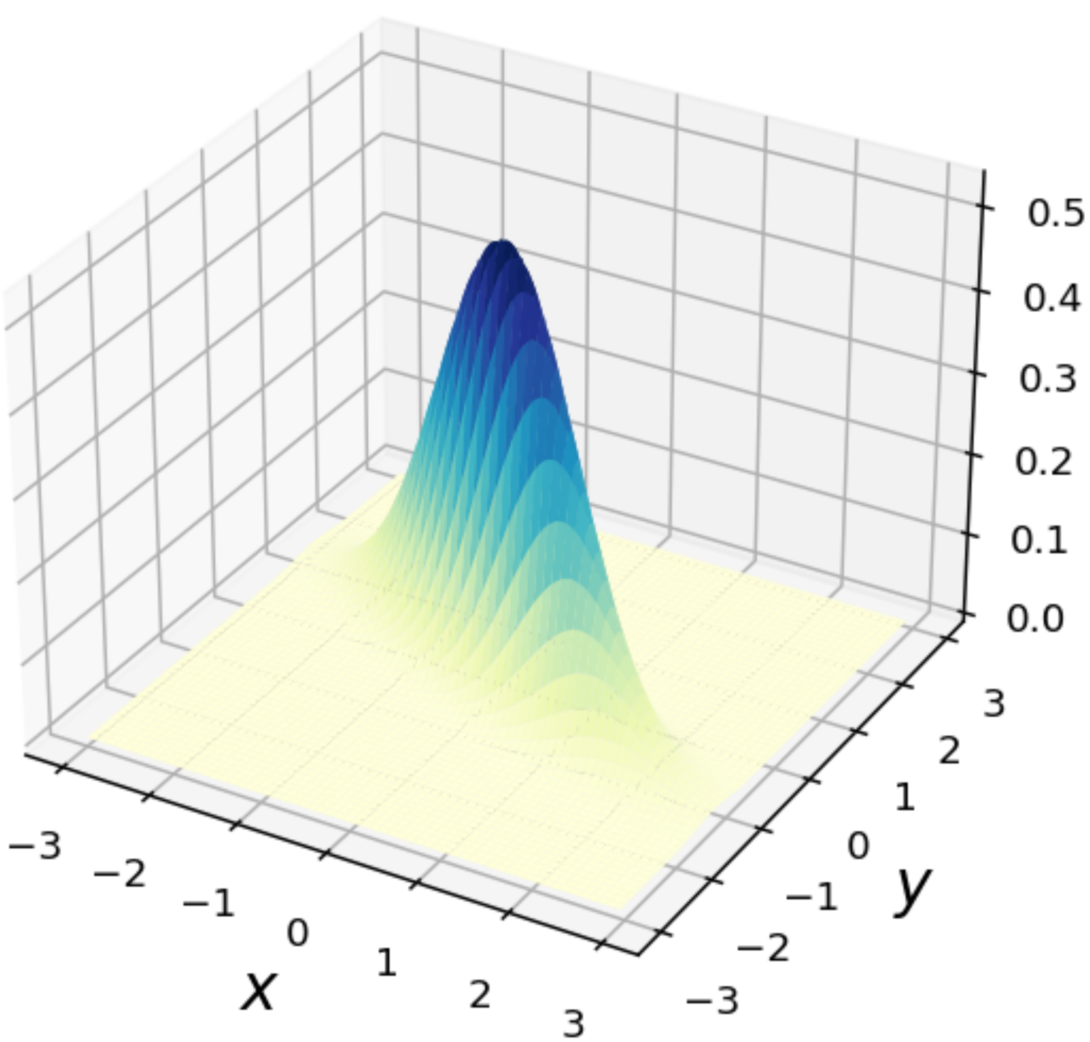
Lines of constant probability correspond to constant argument of exp
→ this defines an ellipse

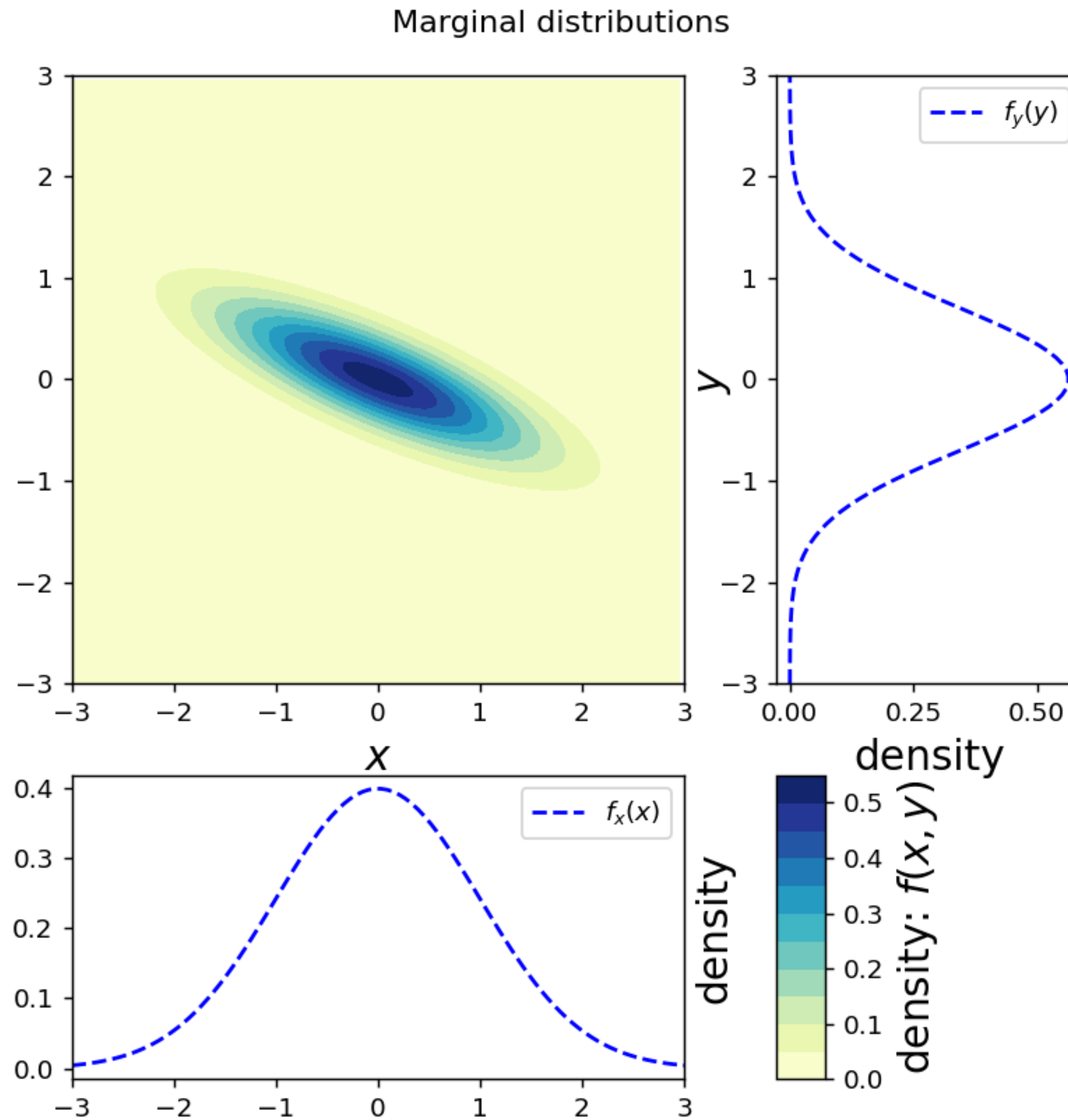
1σ ellipse: $f(x_1, x_2)$ has dropped to $1/\sqrt{e}$ of its maximum value (argument of exp is $-1/2$):

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

Visualizing the 2d Gaussian

$$\sigma_x = 1,\ \sigma_y = 0.5,\ \rho = -0.8$$

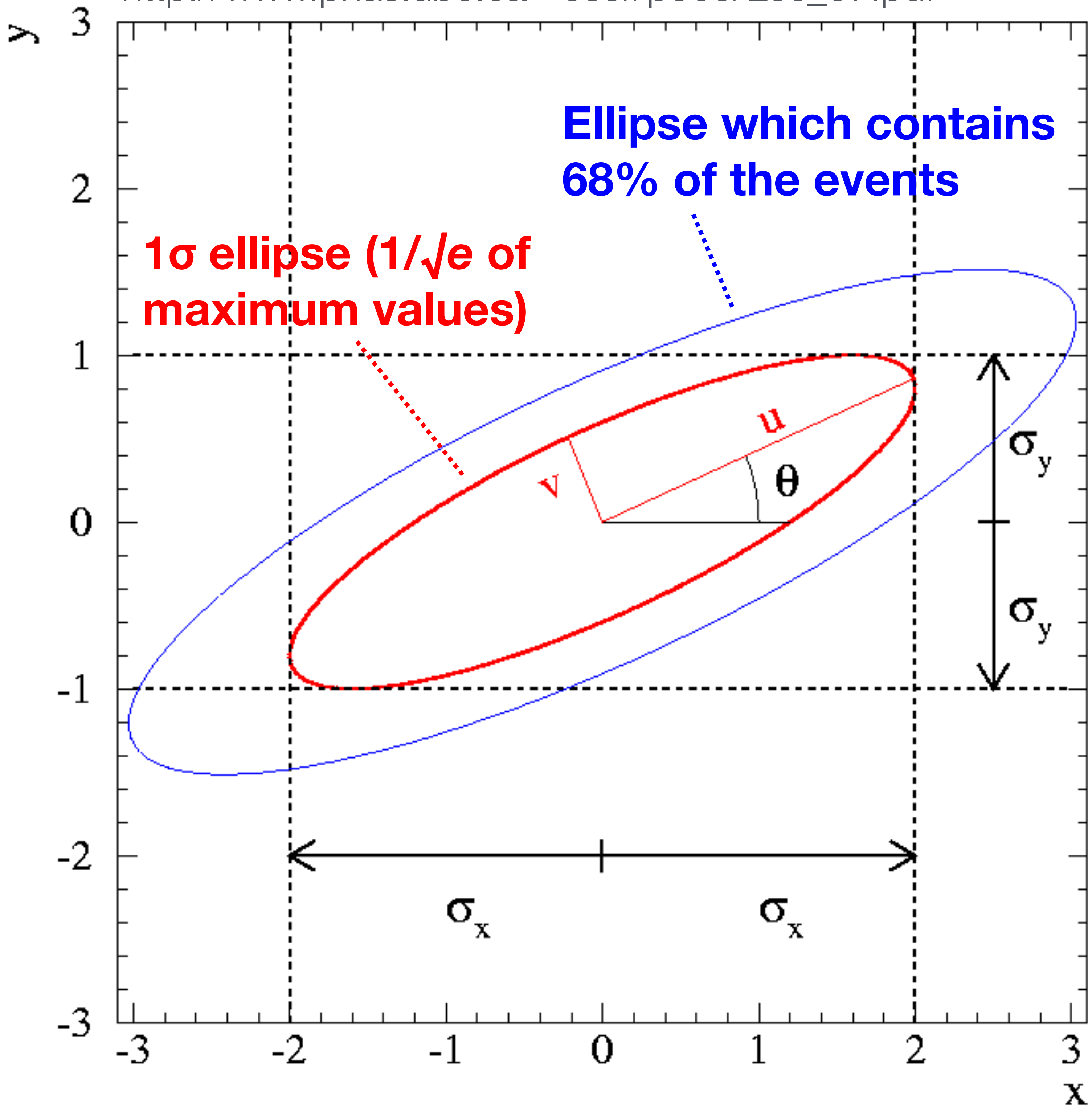




https://nbviewer.jupyter.org/urls/www.physi.uni-heidelberg.de/~reygers/lectures/2020/smipp/plot_2d_gaussian.ipynb

2d Gaussian: Error Ellipse

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



$$f_y(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y$$
$$= \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right)$$

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

	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%

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:

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

$$E[m] = \mu$$

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

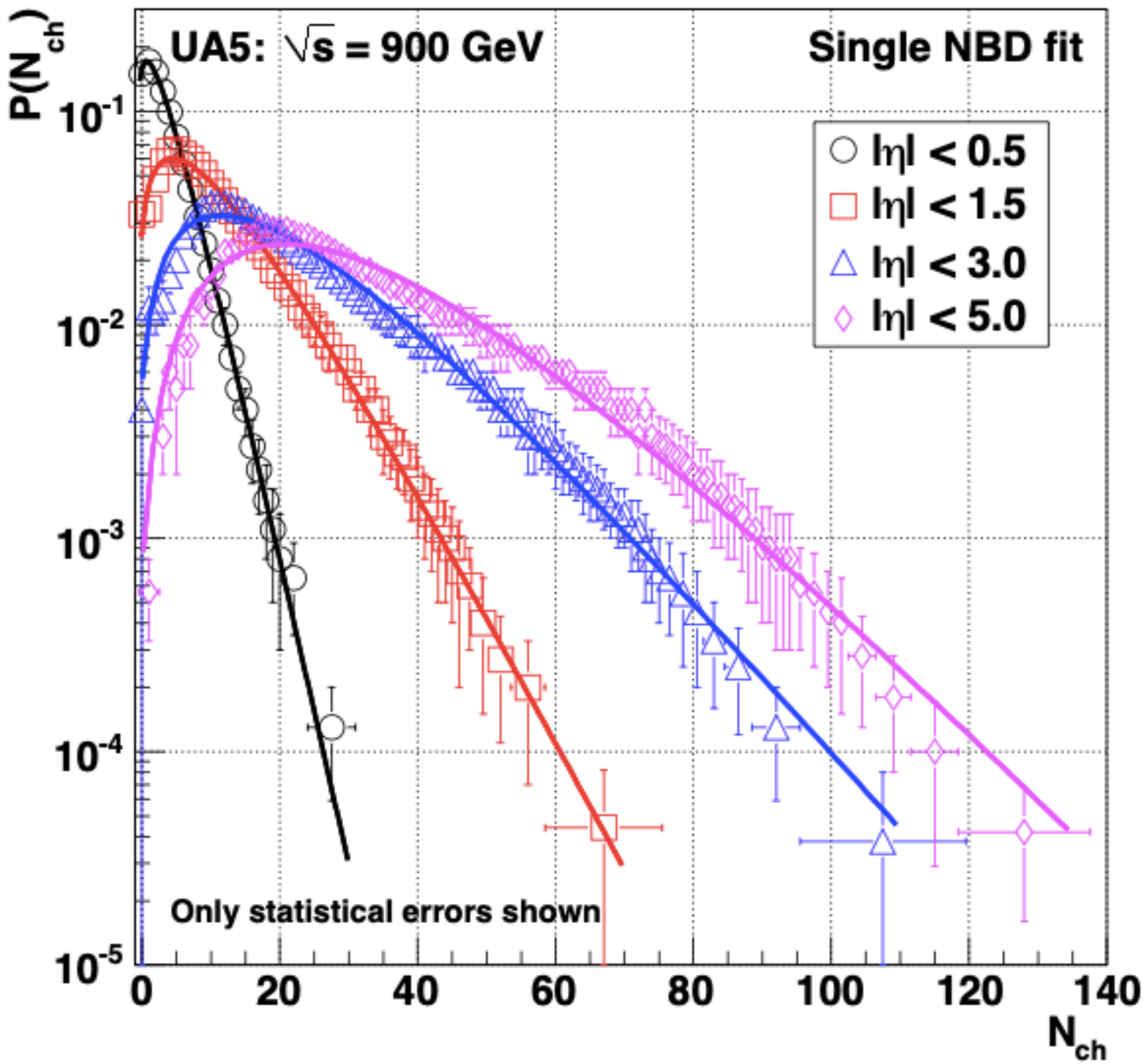
$$p = \frac{1}{1+\frac{\mu}{k}}$$ [relation btw. parameters]

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.

Example: Charged Particle Multiplicity Distribution in pp collisions

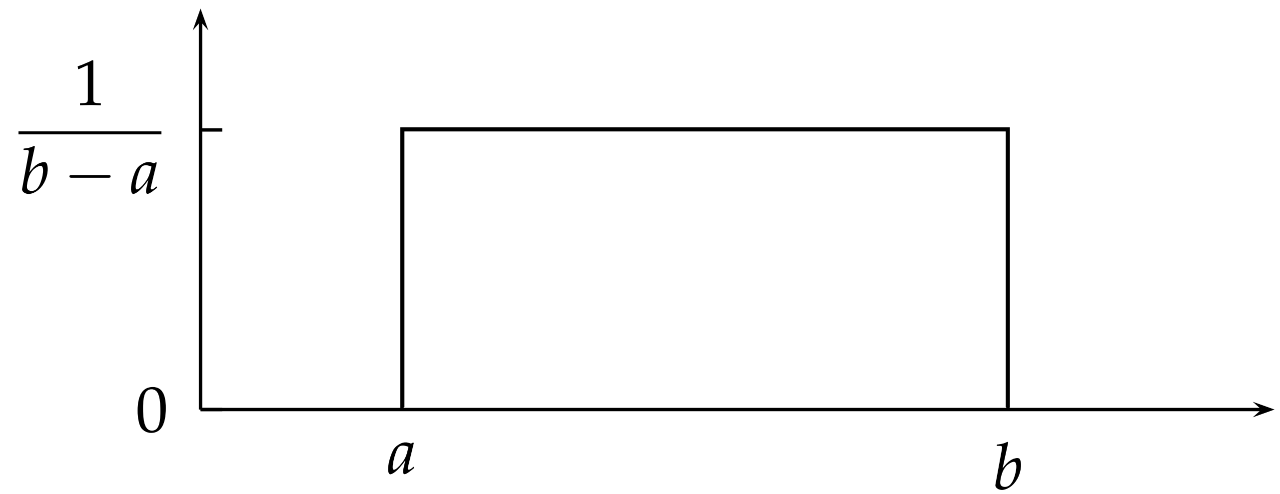


At LHC energies: Superposition of two NBD used to fit multiplicity distributions

Uniform Distribution

Properties:

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

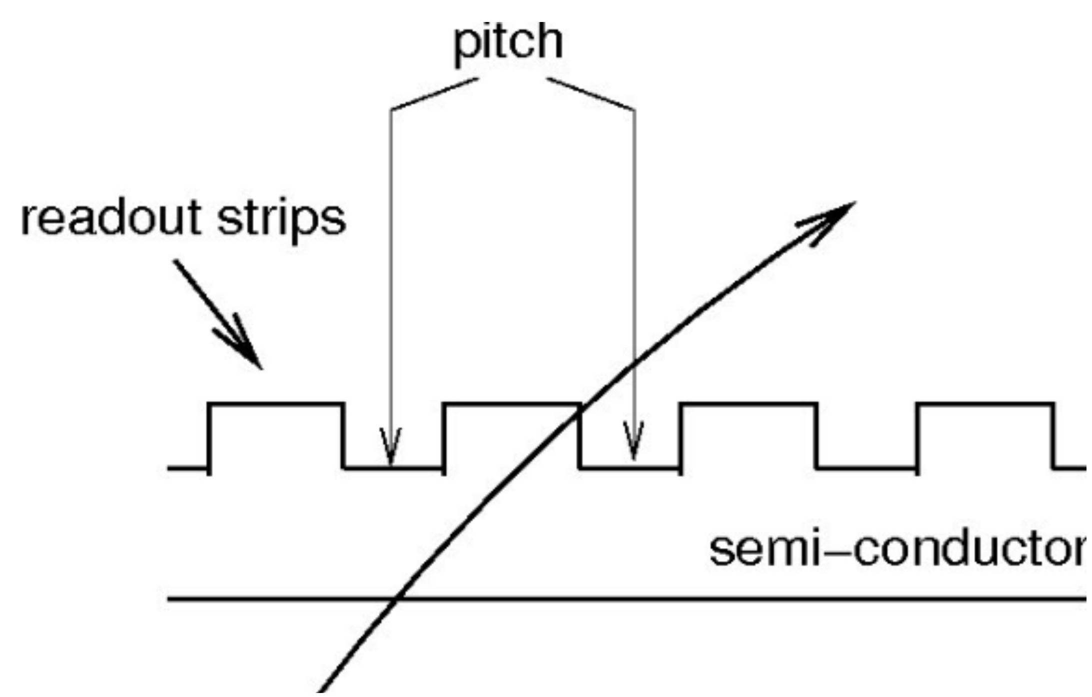


$$E[x] = \frac{1}{2}(a+b)$$

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

Example:

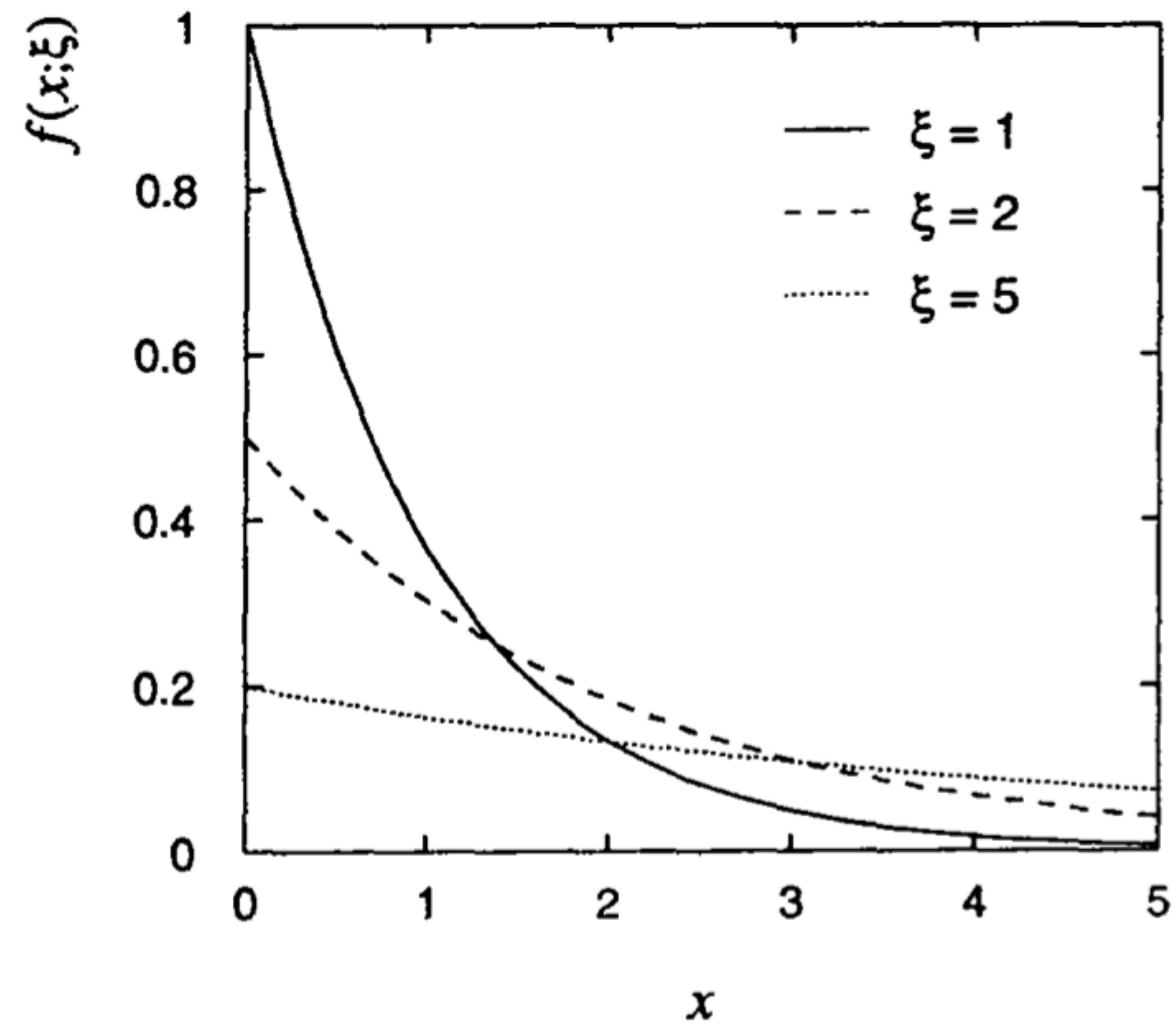
- Silicon strip detector: resolution for one-strip clusters: pitch/$\sqrt{12}$



Exponential Distribution

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

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



Example:

Decay time of an unstable particle at rest

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

Landau Distribution

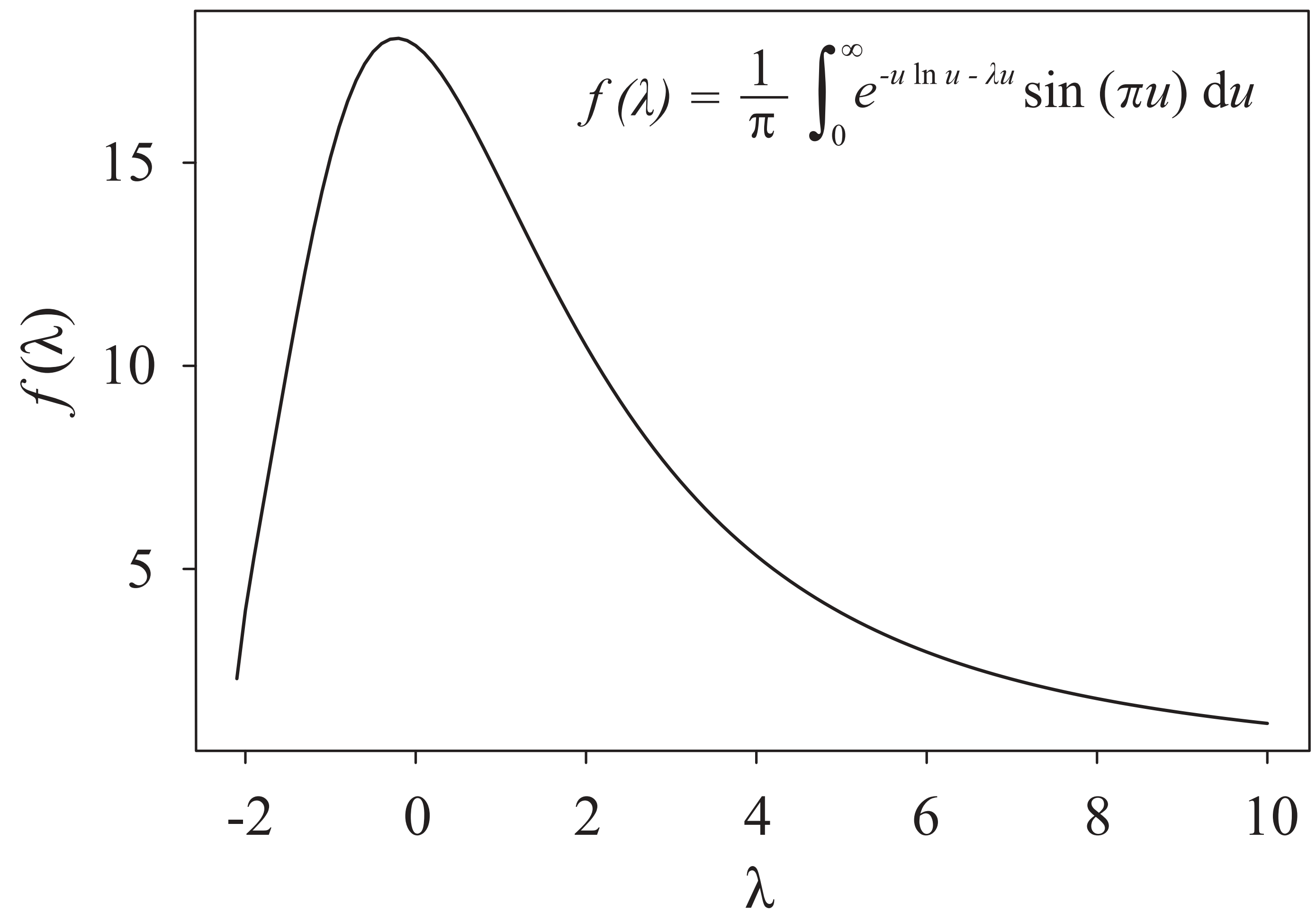
L. Landau, J. Phys. USSR 8 (1944) 201
W. Allison and J. Cobb, Ann. Rev. Nucl. Part. Sci. 30 (1980) 253.

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

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

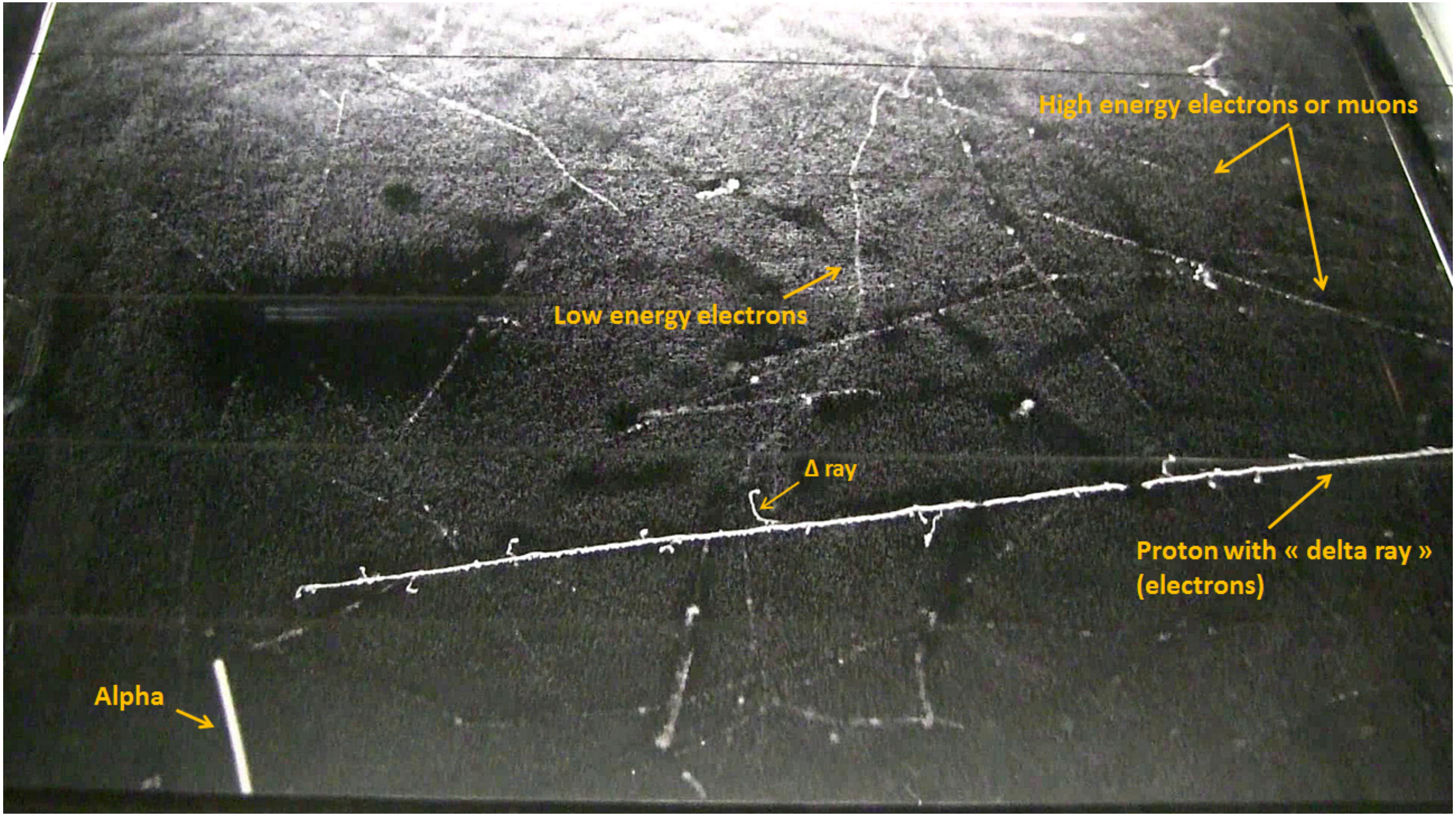
$$\lambda = \frac{\underbrace{\Delta}_{\text{actual energy loss}} - \underbrace{\Delta_0}_{\text{location parameters}}}{\underbrace{\xi}_{\text{material property}}}$$



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 t distribution arises from sampling:

$$\frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$ → follows standard normal distr. (μ=0, σ=1)

$$t := \frac{\bar{x} - \mu}{\hat{\sigma} / \sqrt{n}}$$ → follows Student's t distr. with n–1 degrees of freedom

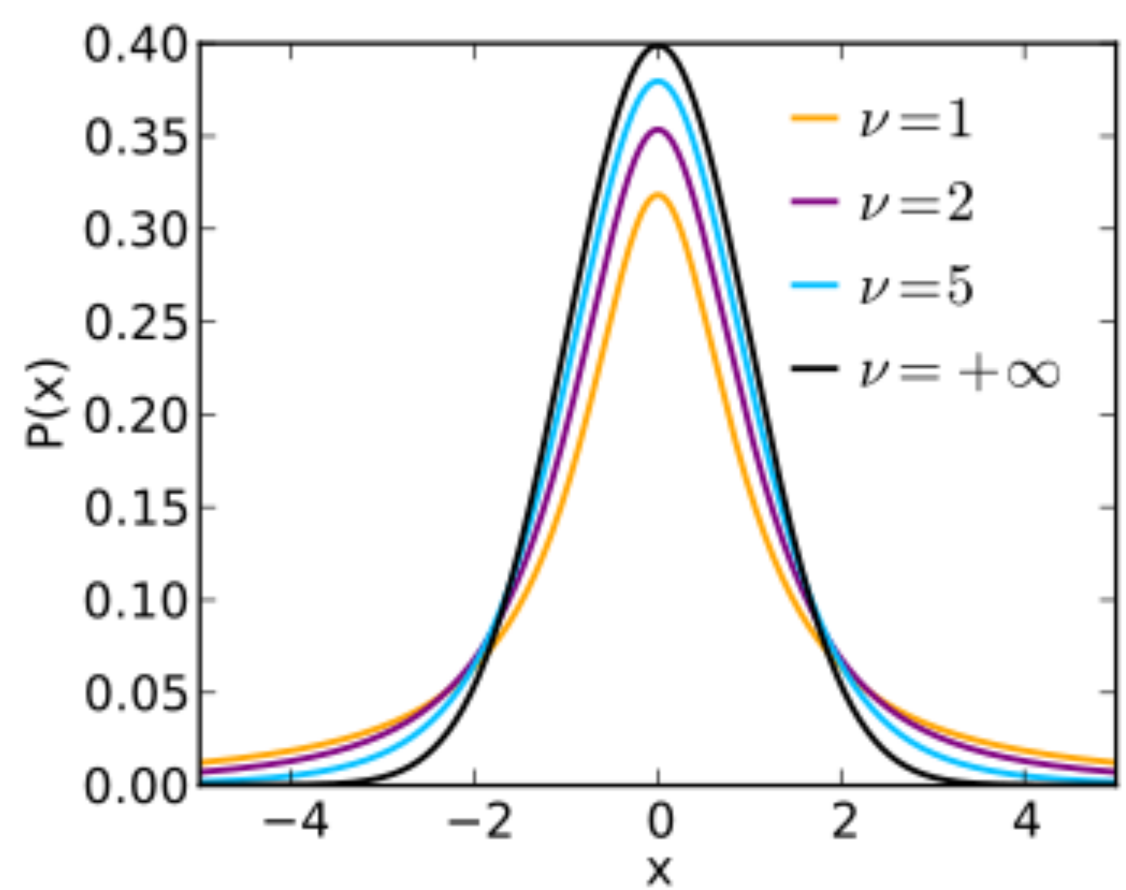
Student's t distribution:

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

With $\nu = n - 1$ for n measurements; t-distribution can be used to construct a confidence interval for the true mean

$\nu = 1: \ \text{Cauchy distr.}$

$\nu \to \infty: \ \text{Gaussian}$



χ^2 Distribution

Let x_1, …, 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:

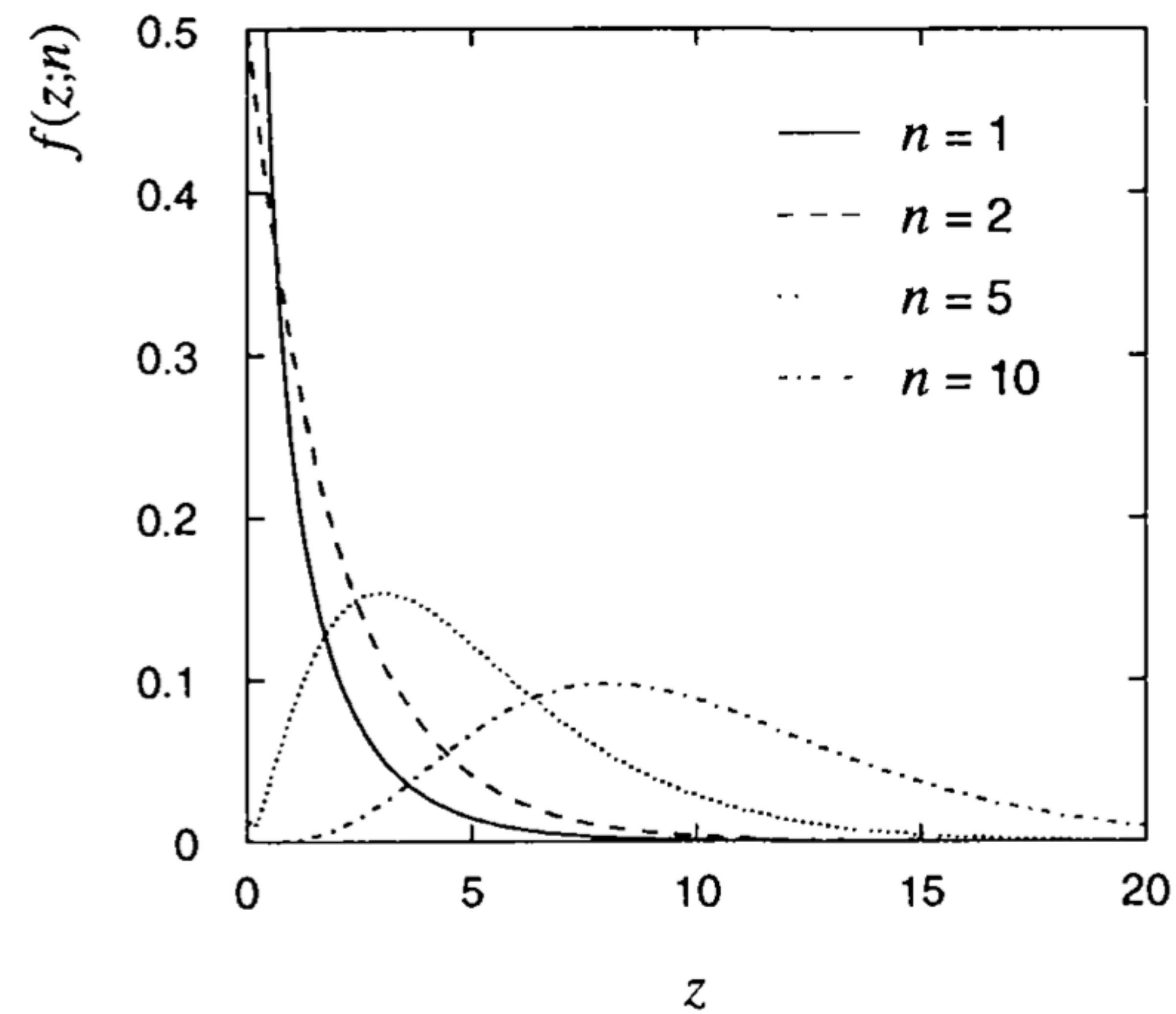
$$f(z;n) = \frac{z^{n/2-1} e^{-z/2}}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \qquad (z \geq 0)$$

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

$$\text{mode: } \max(n-2, 0)$$

Application:
Quantifies goodness of fit

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



Log-Normal Distribution

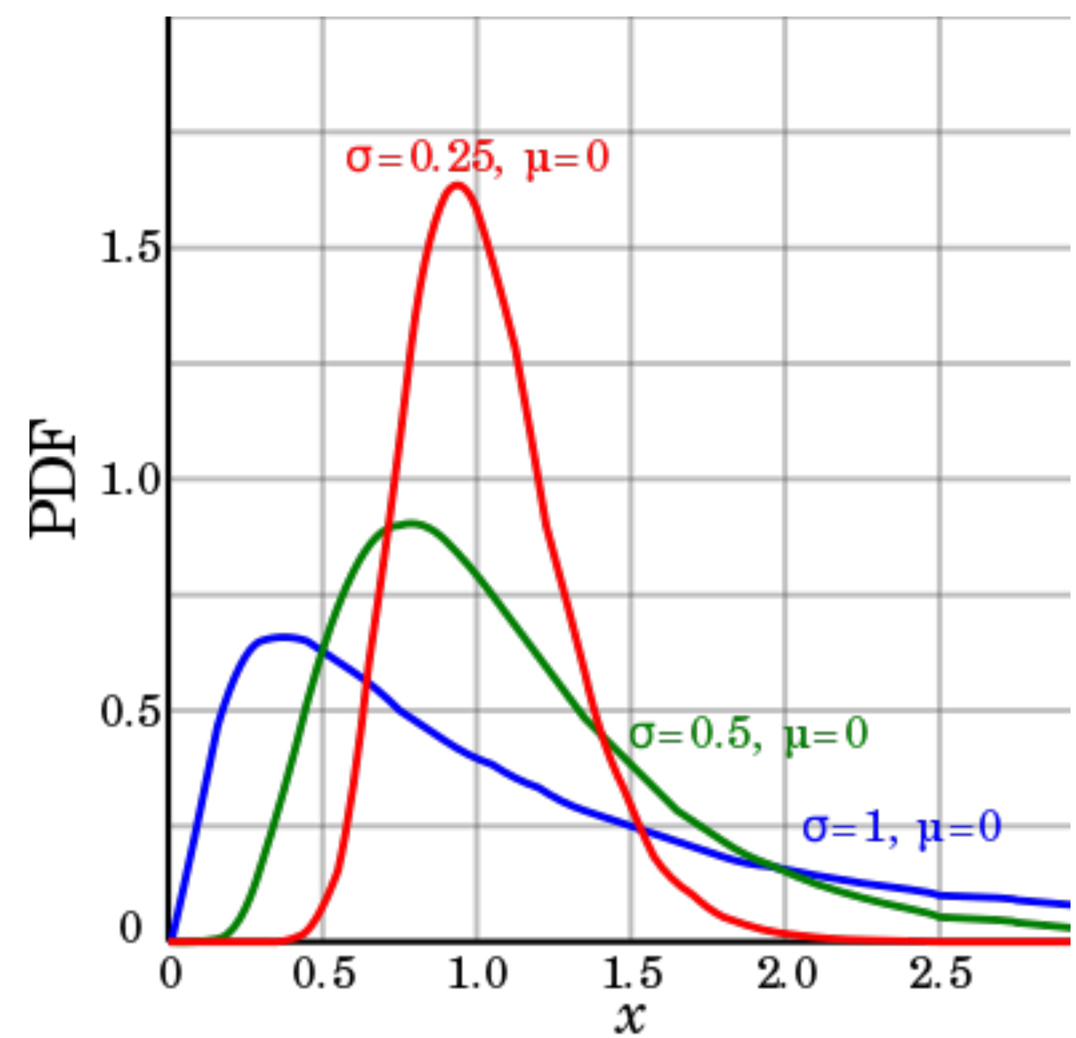
Let y be a normal (i.e. Gaussian) distributed random variable. Then $x = \exp(y)$ follows the log-normal distribution

$$f(x;\mu,\sigma) = N(y;\mu;\sigma)\left|\frac{dx}{dy}\right| = N(\ln x;\mu;\sigma)\frac{1}{x}$$

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

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

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



Multiplicative version of the central limit theorem

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

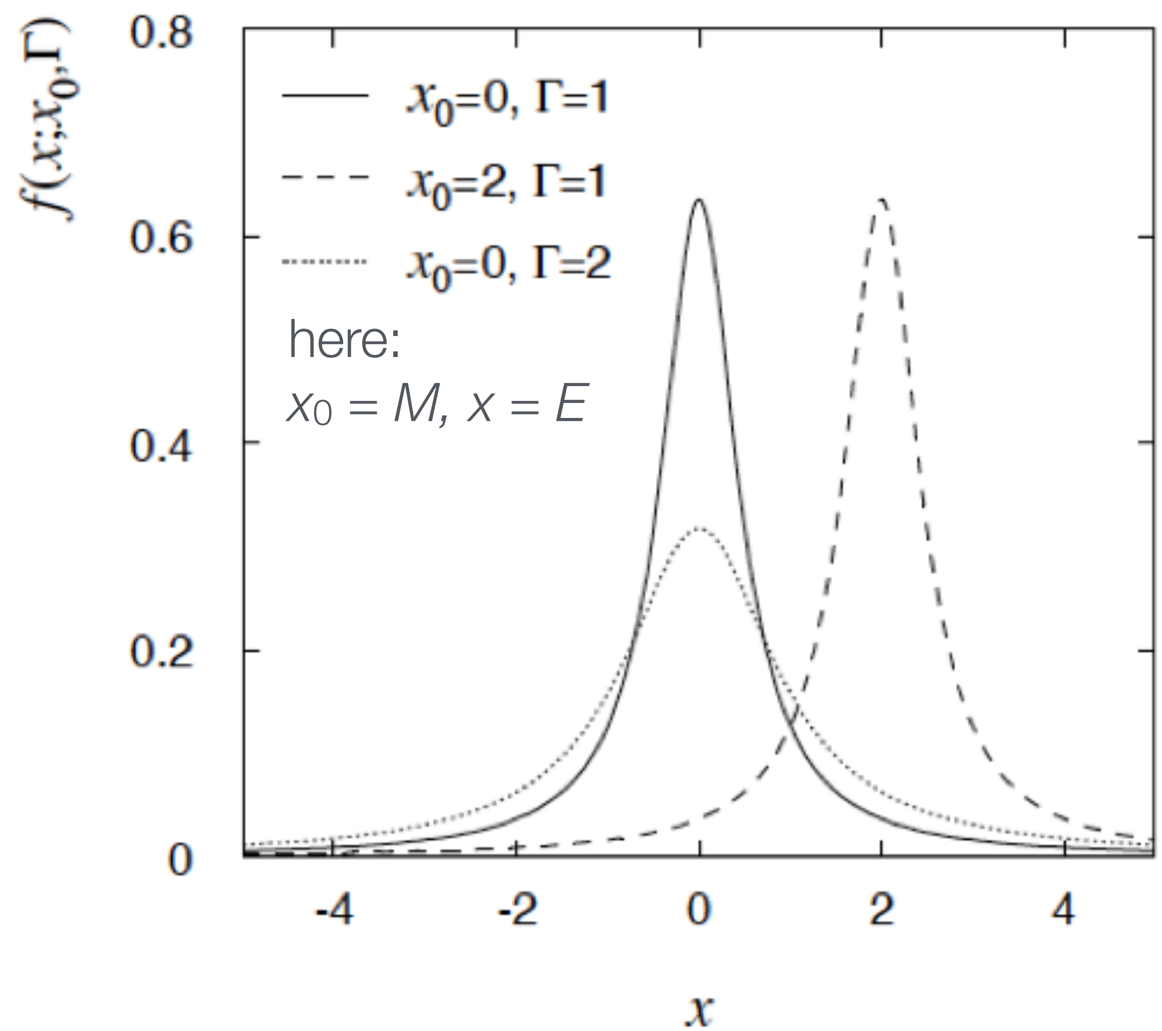
Cauchy, Breit-Wigner, or Lorentzian Distribution

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:

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



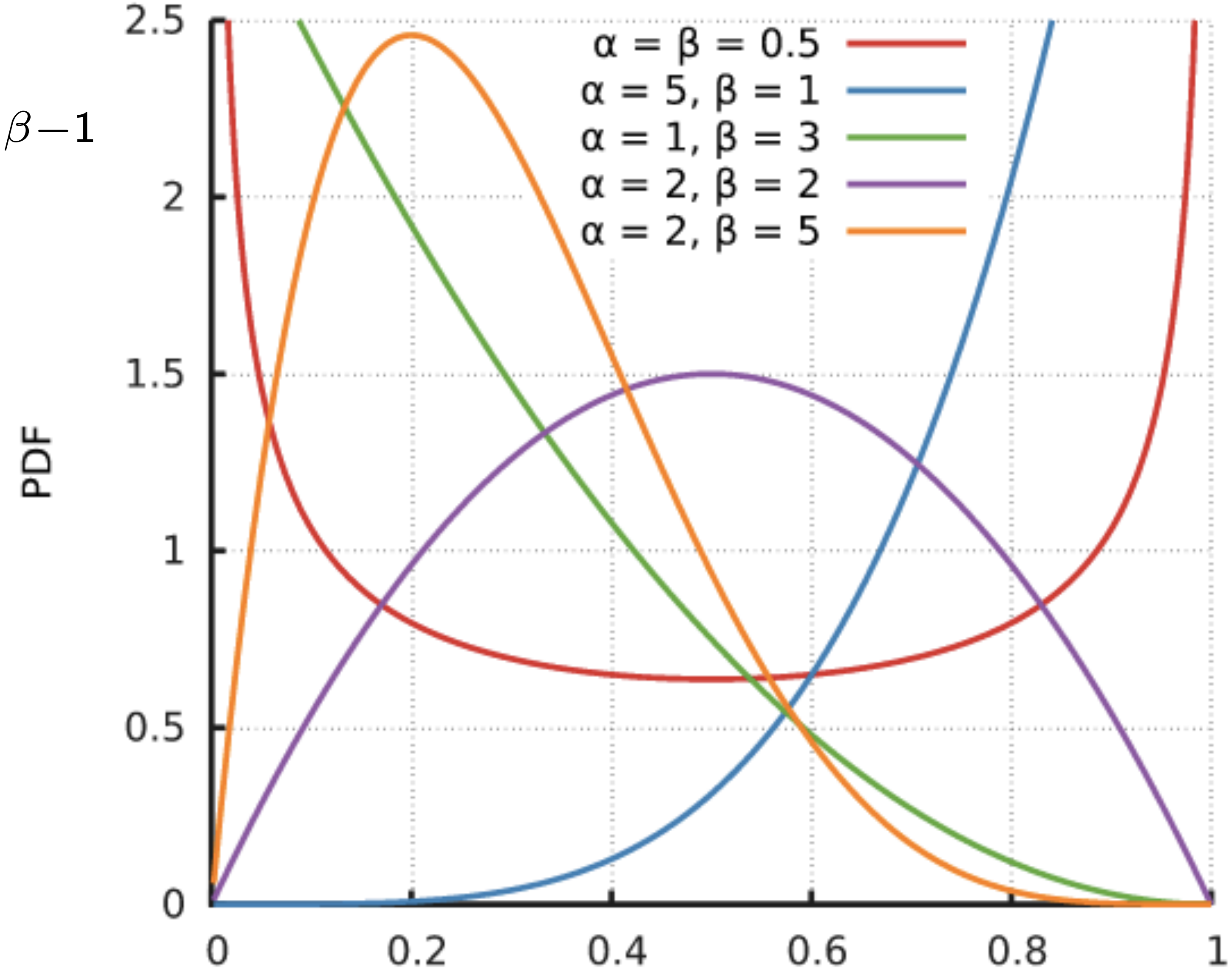
Mean and variance are undefined, mode is *M*.

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 for random variable bounded at both sides.

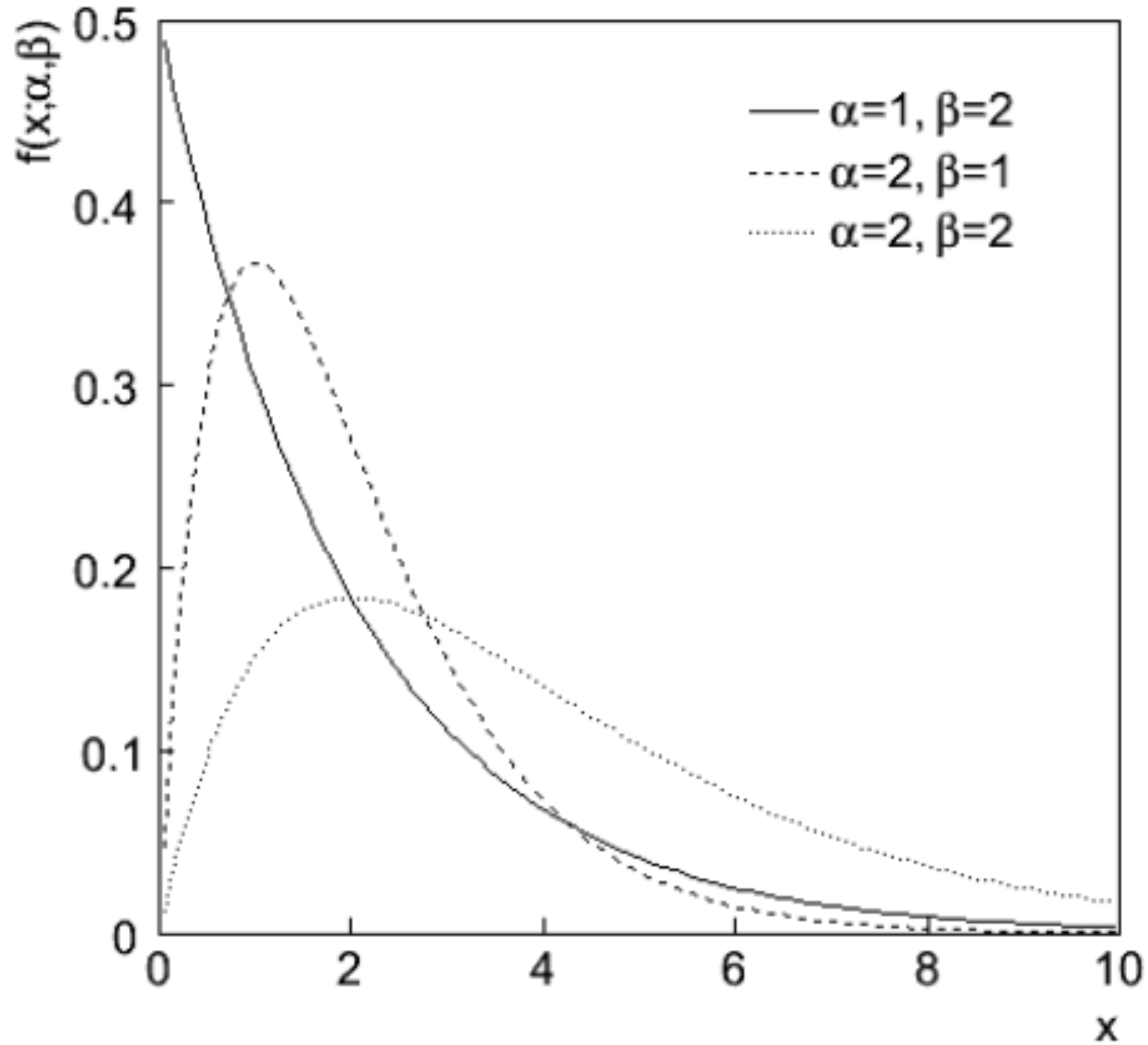
$\alpha = \beta = 1$: uniform distribution

Conjugate prior for the binomial distribution, i.e., if the likelihood function is binomial, then a beta prior gives a beta posterior. Bayesian updating then corresponds to modifying the parameters of the prior.

Gamma Distribution

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

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

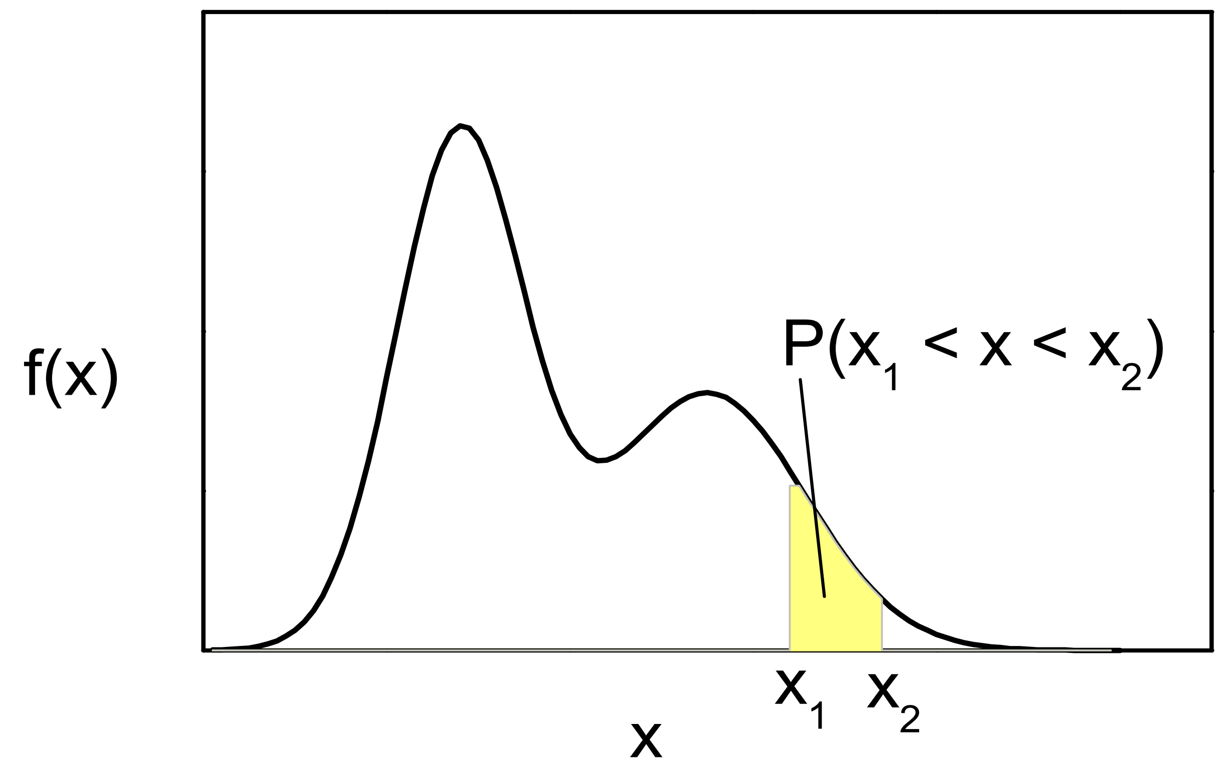


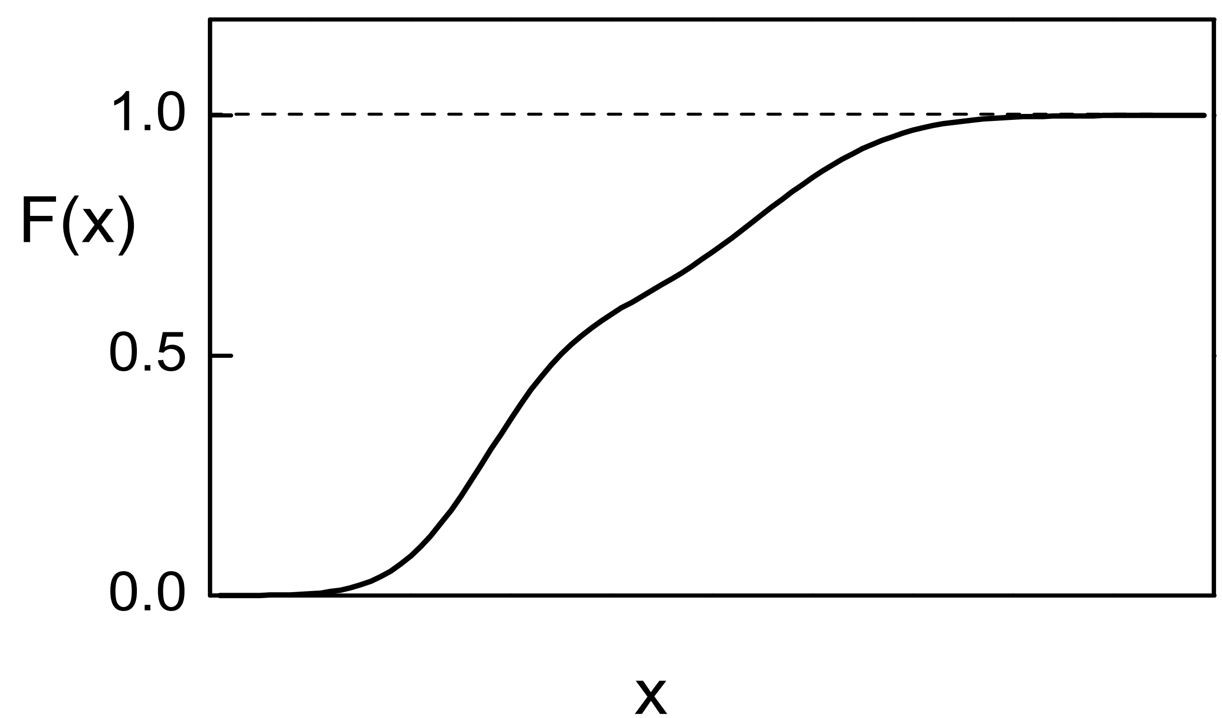
Exponential and χ^2 distributions are special cases of the gamma distribution

Conjugate prior for Poisson likelihood and exponential likelihood

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