Statistical Methods in Particle Physics

2. Probability Distributions

Heidelberg University, WS 2023/24

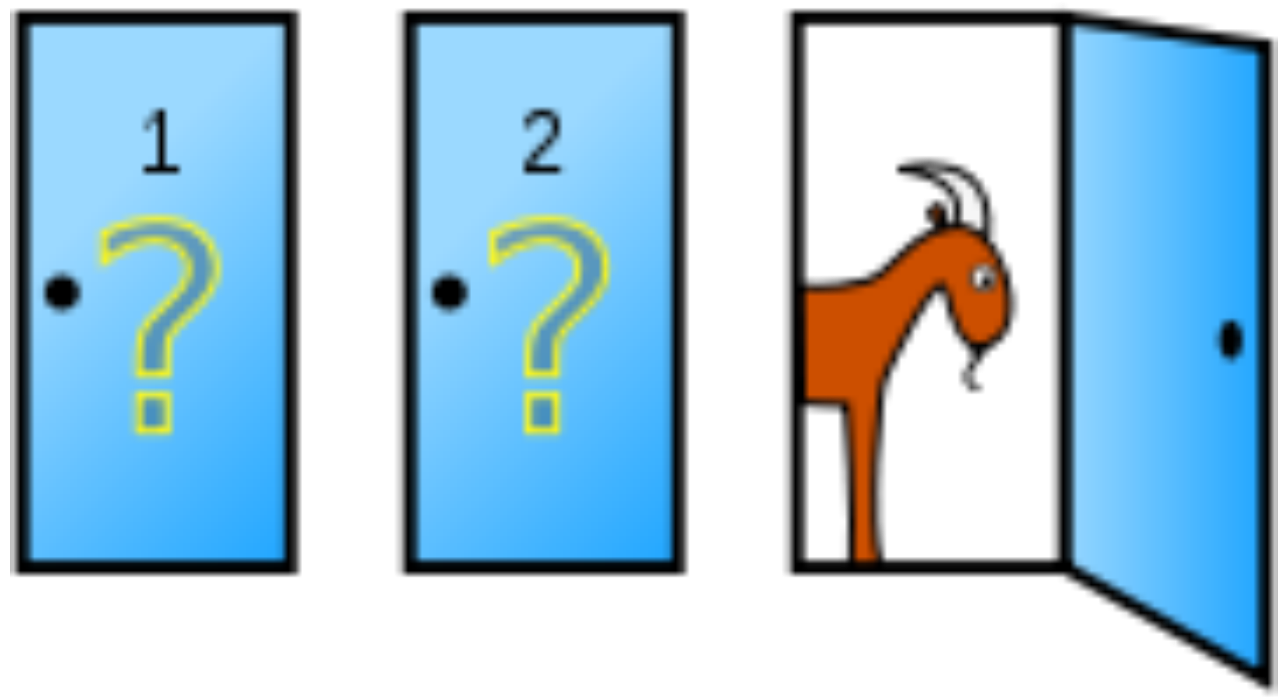
Klaus Reygers, Martin Völkl (lectures)
Ulrich Schmidt, (tutorials)

Fun with probabilities

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

Monty Hall problem ("Ziegenproblem")

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?



Standard assumptions

- The host must always open a door that was not picked by the contestant
- The host must always open a door to reveal a goat and never the car.
- The host must always offer the chance to switch between the originally chosen door and the remaining closed door.

Under these assumptions you should switch your choice!

Reminder: Frequentist and Bayesian Statistics

- Bayesian probability: degree of belief
- Start with prior $p(A)$

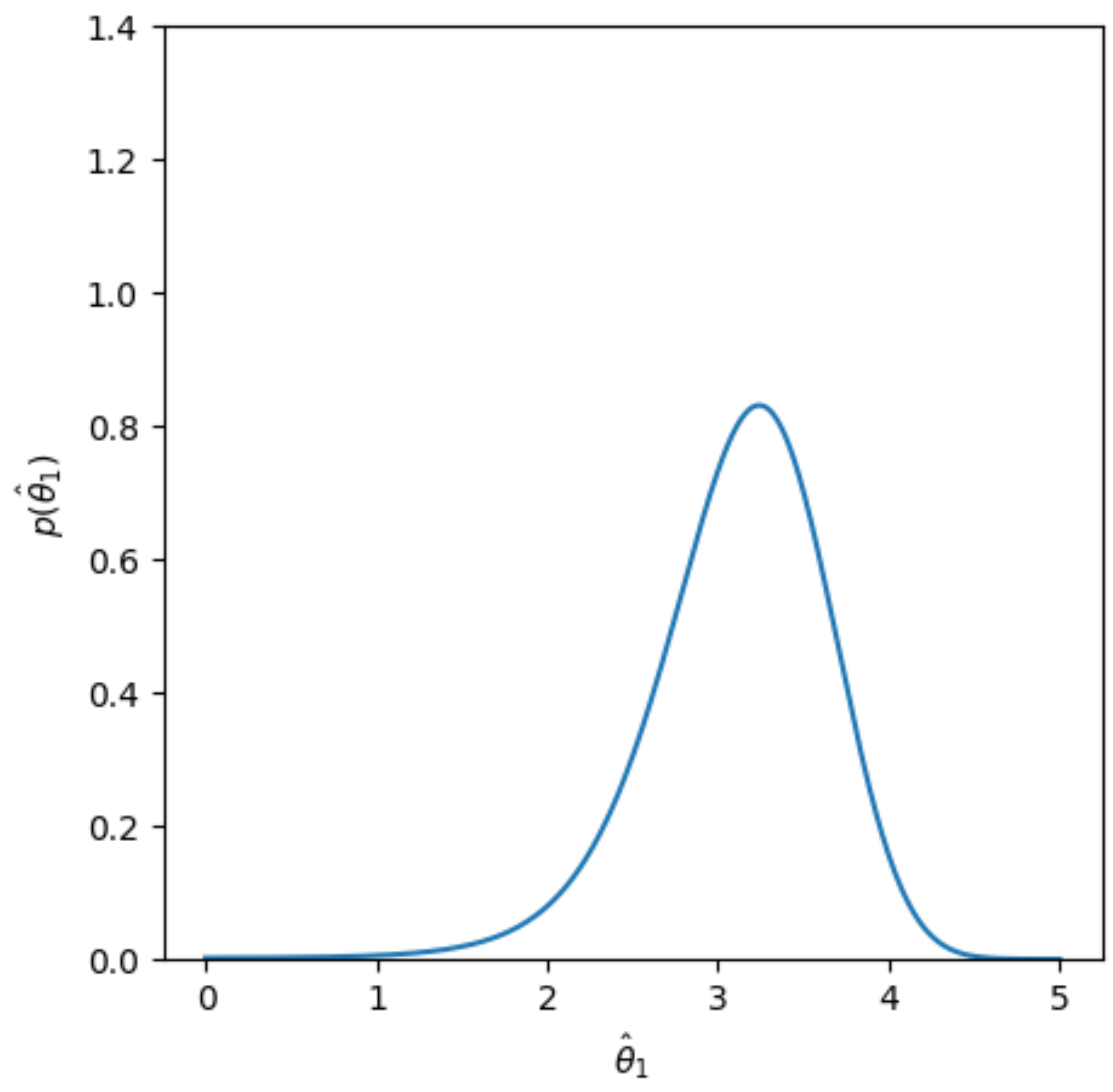
$$p(A\,|\,B) = \frac{p(B\,|\,A)\;p(A)}{p(B)}$$

- Result of statistical analysis is the posterior probability distribution (e.g. of a parameter)

- Frequentist probability: Relative frequency of outcome
- $p \equiv \lim_{N\to\infty} \frac{N_{\text{success}}}{N}$
- Outcome usually formulated in terms of what would happen if the experiment was repeated a number of times

Estimators

- Experiment with possible measured outcomes
- We “sample” the population of all possible results, giving measurement $\vec{m}$
- Probability distribution for outcomes may depend on unknown parameter(s) $p(\vec{m} \,|\, \vec{\theta})$
- Define function giving a value for parameter of interest based on measurement:
  - $\hat{\theta}_1 = \hat{\theta}_1(\vec{m})$
- In general, called a *statistic* (e.g. sample mean). Here, an *estimator* of the parameter
- For now, we will guess $\hat{\theta}_1$
- Estimate of θ_1 is measured value $\hat{\theta}_1$
- Uncertainty from standard deviation of $\hat{\theta}_1$ over several measurements
- θ_1 does not have a probability distribution, but $\hat{\theta}_1$ does!



Conjugate Priors

- Bayes: $p(\theta \,|\, m) \sim p(m \,|\, \theta)\; p(\theta)$
- Assume $p(\theta)$ is part of a class of functions with some parameters
- Depending on the likelihood, the posterior $p(\theta \,|\, m)$ can be part of the same class, but with updated parameters
- In this case, the function class is called the *conjugate prior* to the likelihood $p(m \,|\, \theta)$
- Only the parameters update, often via simple arithmetic laws
- This makes calculations easier

Sums of variables

- Reminder: Densities transform with the Jacobian:
  $\int p_a(\vec{a})\ \mathrm{d}\vec{a} = \int p_a(\vec{a}(\vec{b}))\left|J\right|\ \mathrm{d}\vec{b}$ and so $p_b = p_a\left|J_{b\to a}\right|$, with $J = \frac{\partial a_i}{\partial b_j}$
- Special case (from last time), transformation to new single variable:

$$p_\phi(\phi) = \left|\frac{\mathrm{d}\lambda}{\mathrm{d}\phi}\right|\ p_\lambda(\lambda(\phi))$$

- Now: Calculate sum of variables $z = x + y$ of bivariate distribution $p(x, y)$. Transform $\left(x, y\right) \to (z = x + y, y)$, $\left|J\right| = 1$
- Therefore $p_{z,y}(z, y) = p_{x,y}(z - y, y) \cdot 1$, now integrate out y:
- Marginalize $p_z(z) = \int p_{z,y}(z, y)\ \mathrm{d}y = \int p_{x,y}(z - y, y)\ \mathrm{d}y$; for independent variables $p_{x,y}(x, y) = p_x(x)p_y(y)$

$$\boxed{p_z(z) = \int p_x(z - y)p_y(y)\ \mathrm{d}y \equiv p_x * p_y}$$

- The *convolution* of the two distributions is the distribution of the sum of the variables

Convolutions

$$p_z(z) = \int p_x(z-y) p_y(y)\ \mathrm{d}y \equiv p_x * p_y$$

- Means are additive: $\langle z \rangle = \langle x \rangle + \langle y \rangle$
- Variances are additive: $V[Z] = V[X] + V[Y]$, $\langle (z-\mu_z)^2 \rangle = \langle (x-\mu_x)^2 \rangle + \langle (y-\mu_y)^2 \rangle$
- For families of distributions with a location and scale parameter: If convolution two distributions always yields a distribution from the same family, it is called a *stable distribution*

Linear combinations of random variables

Consider two random variables with known covariance cov(x, y):

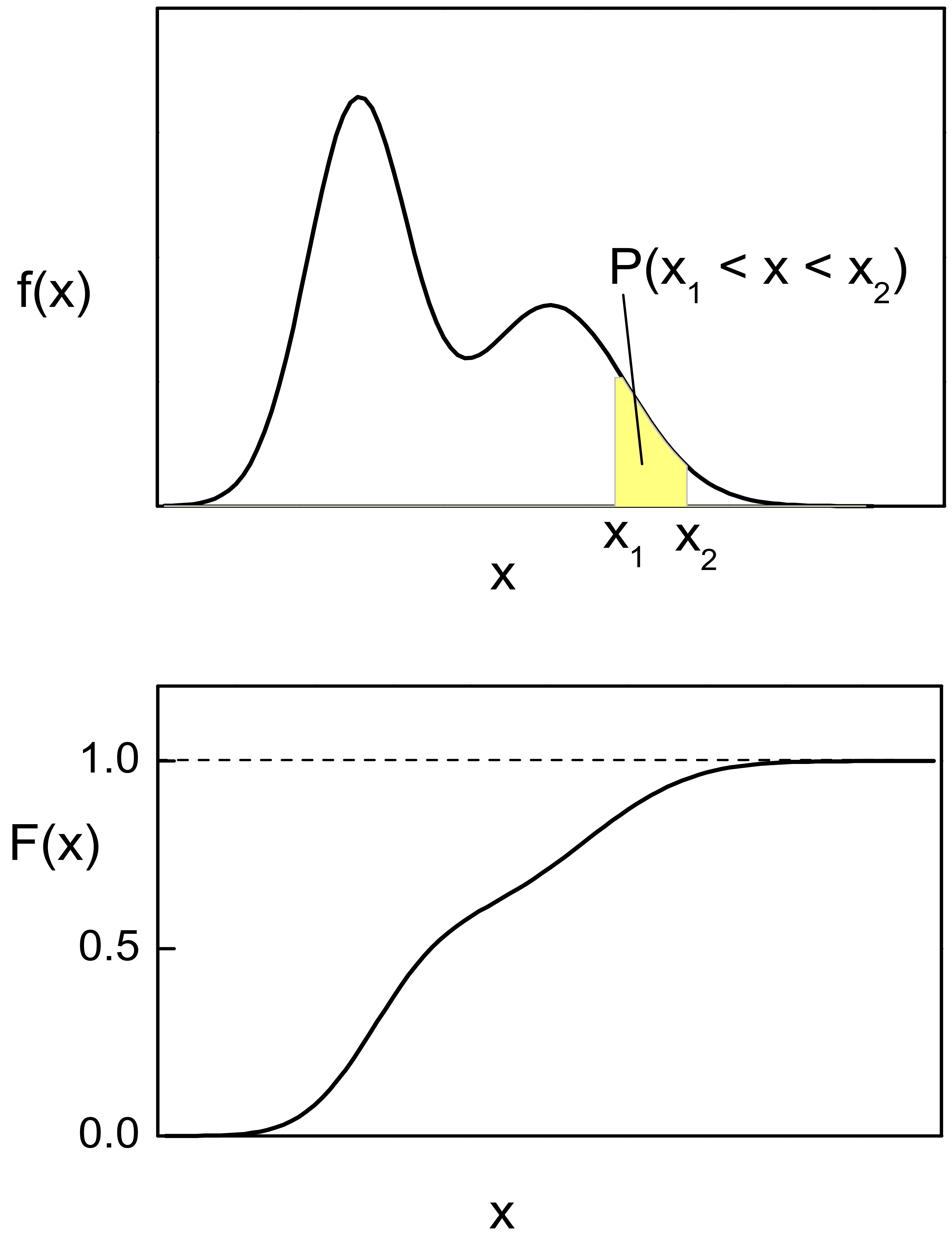
$$\langle x + y \rangle = \langle x \rangle + \langle y \rangle$$
$$\langle ax \rangle = a\langle x \rangle$$
$$V[ax] = a^2 V[x]$$
$$\mathrm{cov}(x, x) = V[x]$$
$$V[x + y] = V[x] + V[y] + 2\mathrm{cov}(x, y)$$

Example of more detailed calculation:

$$\begin{aligned} V[x+y] &= E[(x+y-\mu_x-\mu_y)^2] = E[(x-\mu_x+y-\mu_y)^2] \\ &= E[(x-\mu_x)^2 + (y-\mu_y)^2 + 2(x-\mu_x)(y-\mu_y)] \\ &= E[(x-\mu_x)^2] + E[(y-\mu_y)^2] + 2E[(x-\mu_x)(y-\mu_y)] \\ &= V[x] + V[y] + 2\mathrm{cov}(x, y) \end{aligned}$$

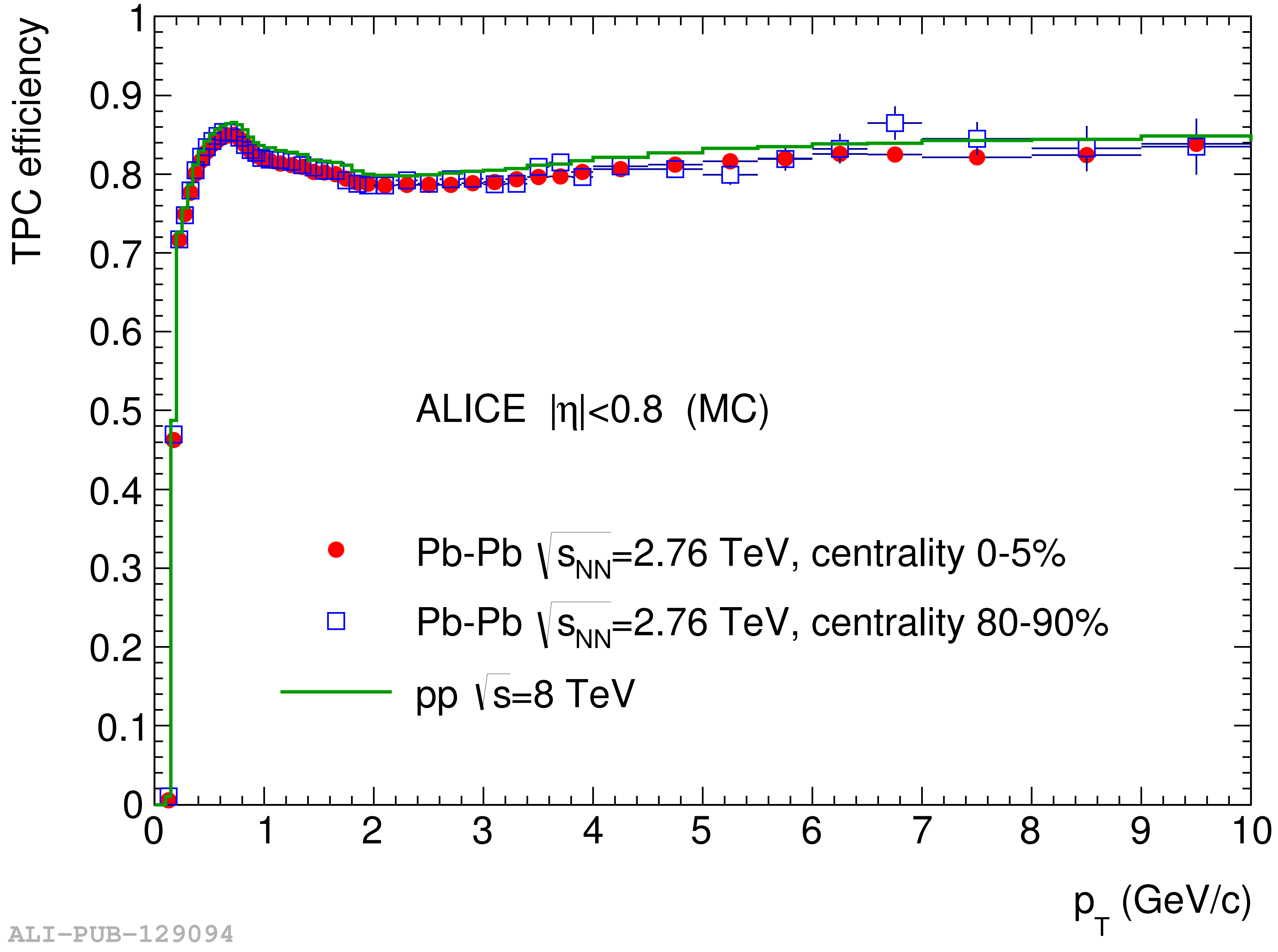
Cumulative Distribution Function (cdf)

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



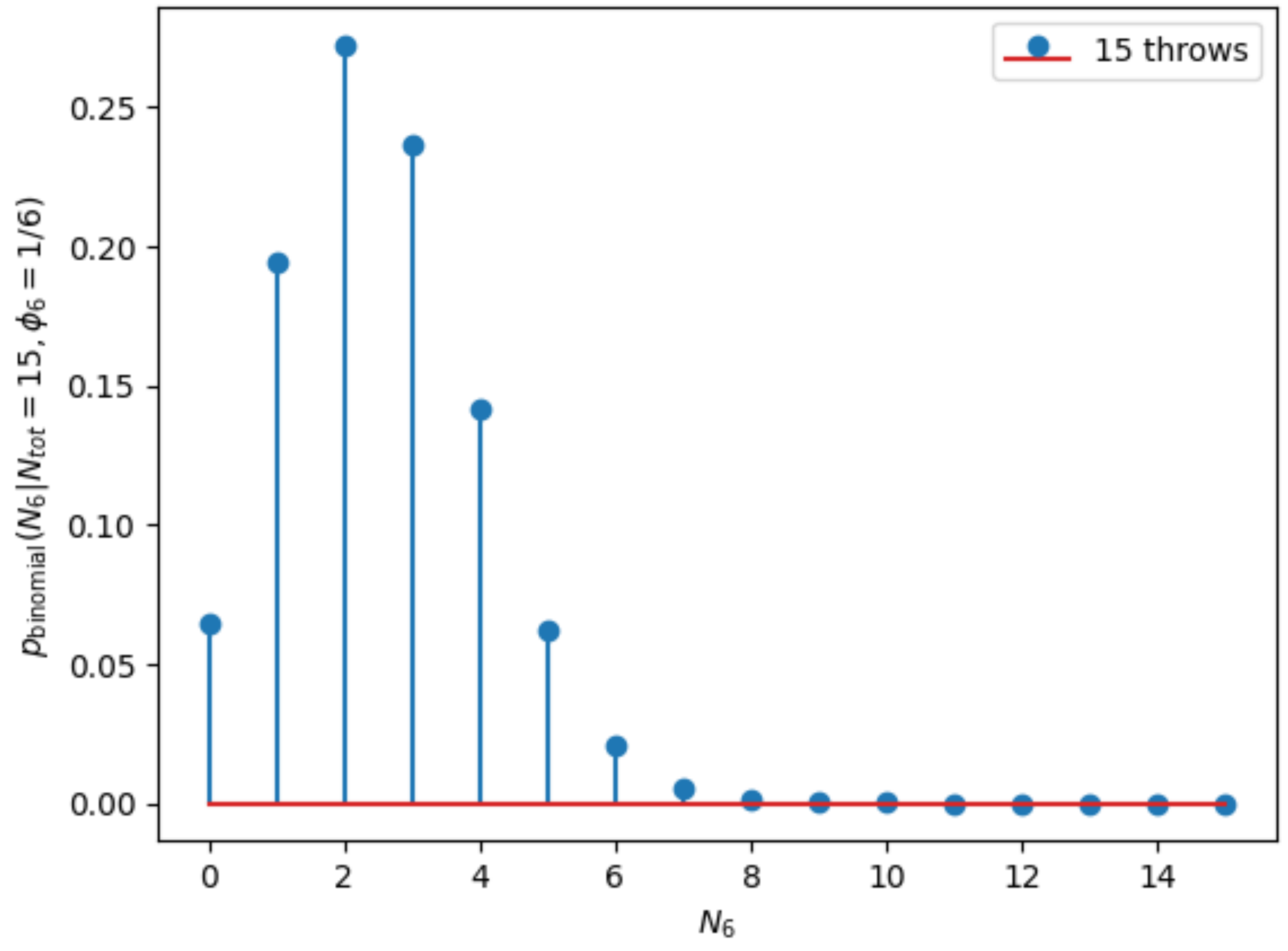
Bernoulli distribution

- Two possible outcomes, e.g. true/false, parameter is probability ϕ
- $p(\text{true}\,|\,\phi) = \phi$
- $p(\text{false}\,|\,\phi) = 1 - \phi$
- Examples:
  - throwing a coin
  - particle decaying in a particular decay channel
  - Detector successfully measuring a particle



Performance of the ALICE Experiment at the CERN LHC, ALICE Collaboration

Binomial distribution



N independent experiments

- Outcome of each is 'success' or 'failure'
- Probability for success is ϕ
- Number of ways to arrange *k* successes and (*n*-*k*) failures - binomial coefficient

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

$$p_b(k \,|\, N, \phi) = \binom{N}{k} \phi^k (1-\phi)^{N-k} \;,\; E[k] = N\phi \;,\; V[k] = N\phi(1-\phi)$$

Examples:

- Example: Detection efficiency
- Polls
- Coin throws
- Number of particles (out of a total) decaying in some channel

- $p(n_{decays} \,|\, N_{particles}, \phi_{B.R.})$ gives us the probability distribution for finding that *n* out of *N* particles decay in this particular channel
- But usually we want to know the opposite: we measure a number of decays and want to know the branching ratio
  - Or we simulate that some number of particles out of the total are measured in the detector and want to estimate the detector efficiency

Binomial parameter inference (Frequentist)

- In a test, $k = 70$ out of $N = 100$ particles were correctly reconstructed. What is the reconstruction efficiency ϕ_e?

$$p_b(k \,|\, N, \phi) = \binom{N}{k} \phi^k (1-\phi)^{N-k}$$

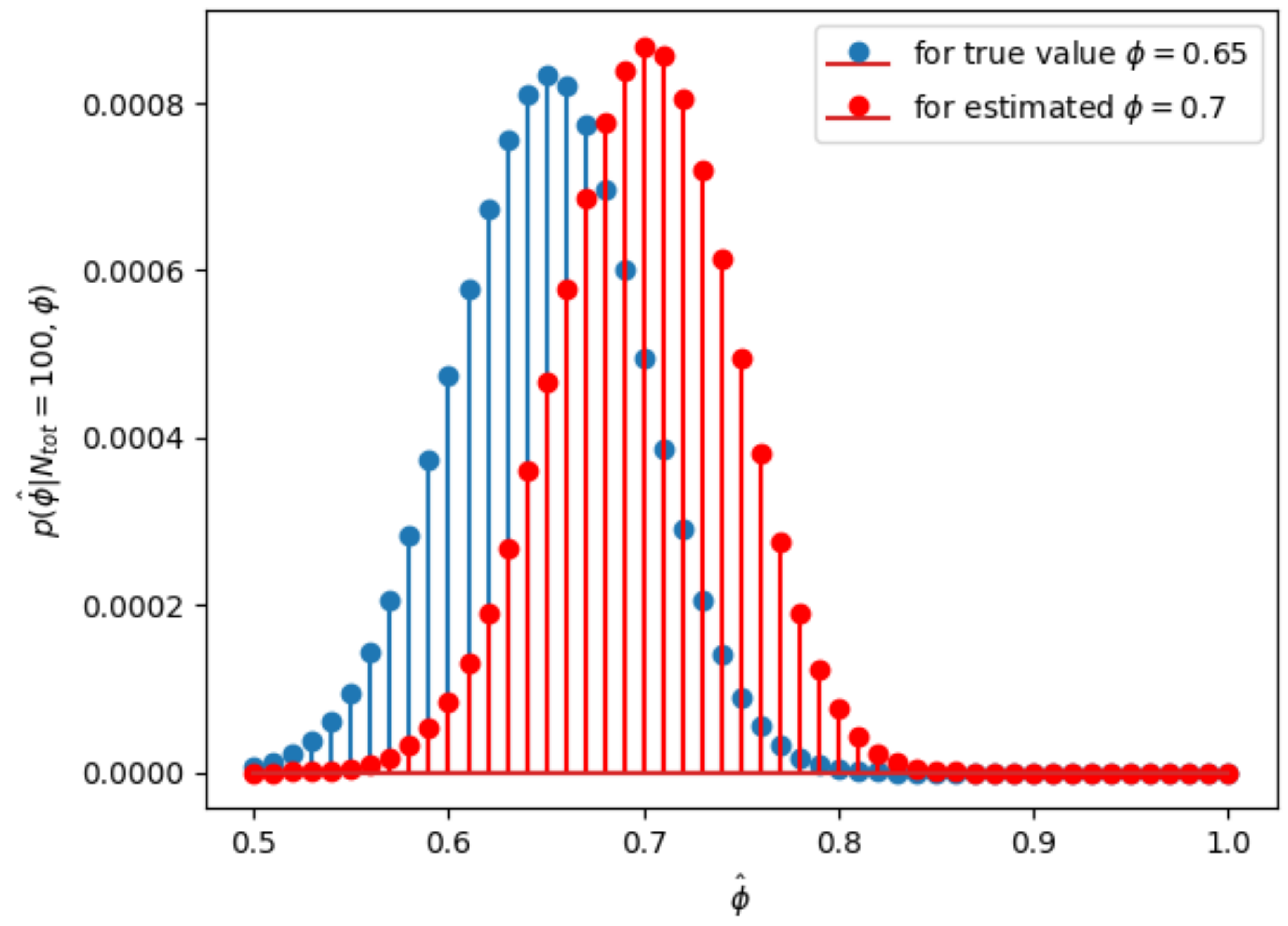
- We know: $E[k] = N\phi$
- Since the outcomes are distributed around the true value, we can guess an estimator:

$$\hat{\phi}_e = k/N$$

- The variance of k is $V[k] = N\phi(1-\phi)$, which we can approximate with our estimator $V[k] \approx N\hat{\phi}_e(1-\hat{\phi}_e)$ and so

$$V[\hat{\phi}_e] \approx \frac{\hat{\phi}_e(1-\hat{\phi}_e)}{N}, \quad \sigma_\phi \approx \sqrt{\frac{\hat{\phi}_e(1-\hat{\phi}_e)}{N}}$$

- So the result would be: $\phi_e = 0.700 \pm 0.046$

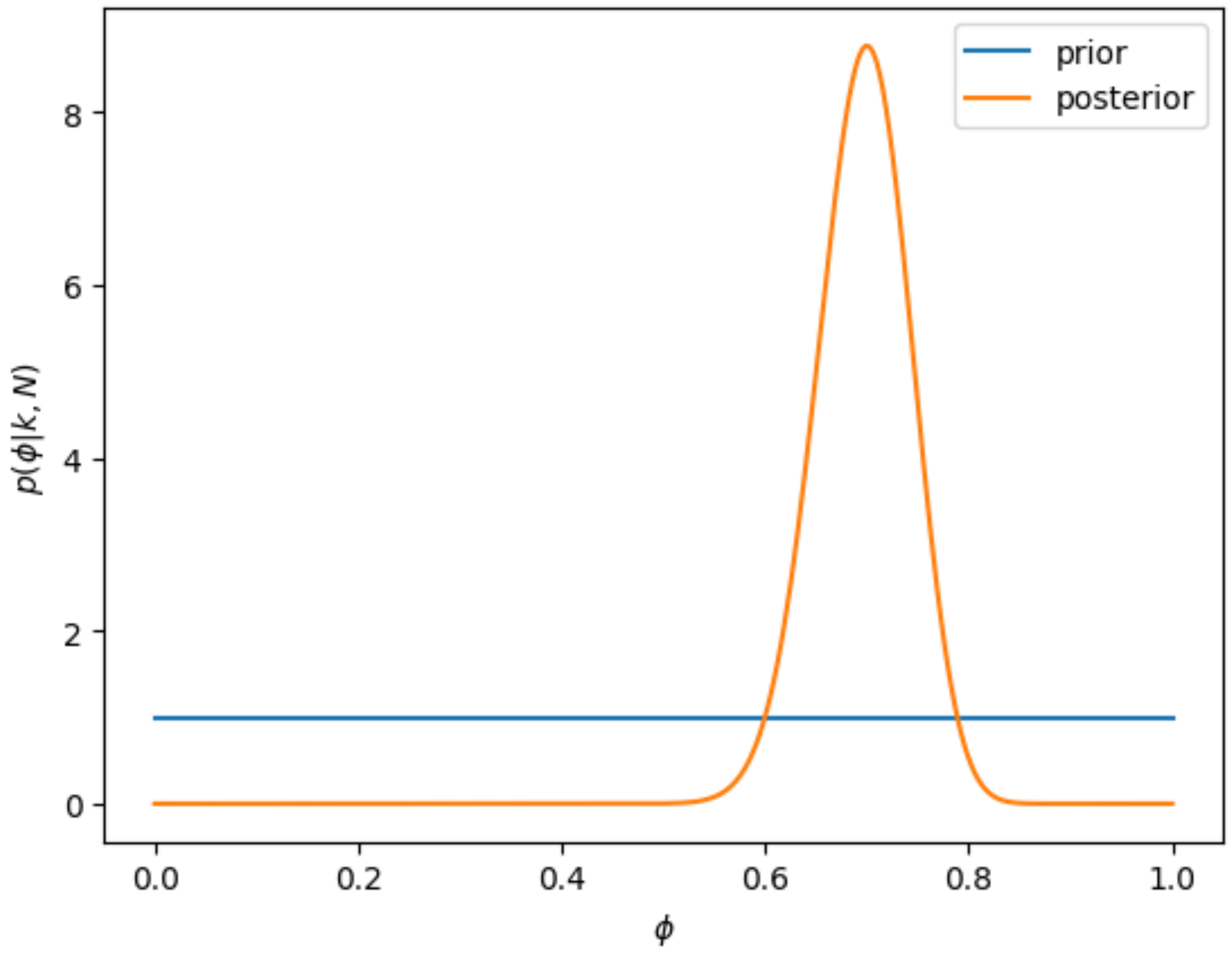


Binomial parameter inference (Bayesian)

- Assume prior $p(\phi) = 1$ (for $0 < \phi < 1$)
- Posterior is then $p(\phi \,|\, k, N) \sim \binom{N}{k} \phi^k (1-\phi)^{N-k} \cdot 1$
- Mean and standard deviation of posterior give

$$\phi = 0.696 \pm 0.045$$

- In general: For large statistics frequentist and Bayesian methods often arrive at similar results!



Reminder: the *likelihood* is the probability distribution $p(k \,|\, N, \phi)$, but considered as a function of ϕ

Small number tests

Your test 100 products from your factory and find problems with 0 of them.

- The estimator from above would suggest that the probability of producing a faulty product would be 0 ± 0
- In this case, the approximation of the variance is not very good

- The estimation only works well for sufficiently large numbers!

The Poisson distribution

- Typical case: N is large, but ϕ is very small
- Example: Radioactive material, $\mathcal{O}(10^{23})$ particles; within a time interval, each decays with a very small (independent) probability
  - Total number of expected decays, $N\phi$ is is not small
- Then Binomial distribution can be approximated by Poisson distribution with single parameter $\mu = N\phi$
- Advantage: Do not have to define N as precisely
- Example: Count gold atoms in bucket of ocean water
  - Each atom has some small probability of being gold
  - But what N do we sample from? The nearby water? All oceans in the world?

Poisson distribution

Large number of independent trials with small probability of success, total successes k

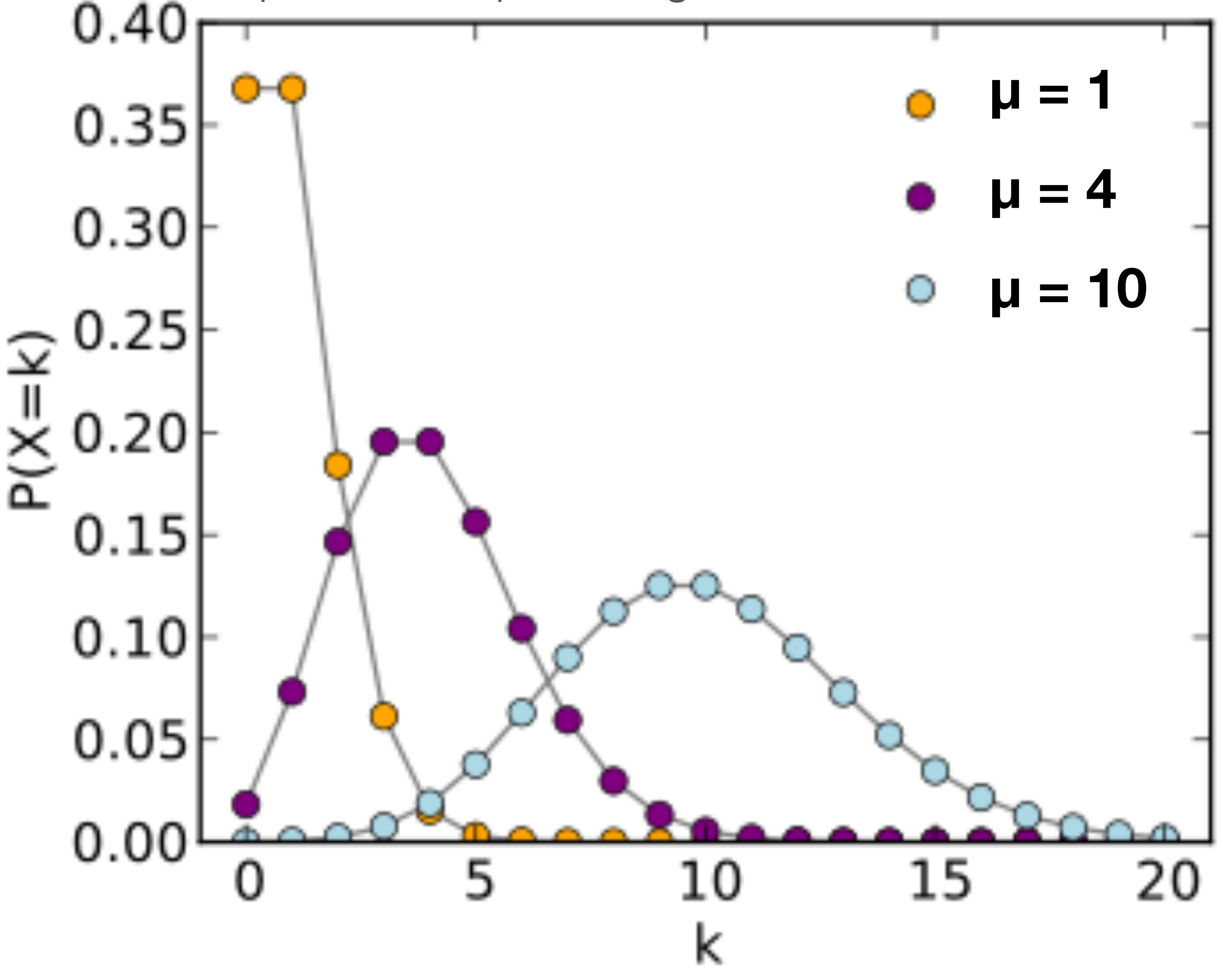
$$p(k;\mu) = \frac{\mu^k}{k!} e^{-\mu}$$

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

Properties:

- n_1, n_2 follow Poisson distr.
  → n_1+n_2 follows Poisson distr., too
- Reasonable estimator: $\hat{\mu} = k$ with variance $\sigma_k = \sqrt{\mu} \approx \sqrt{\hat{\mu}}$

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



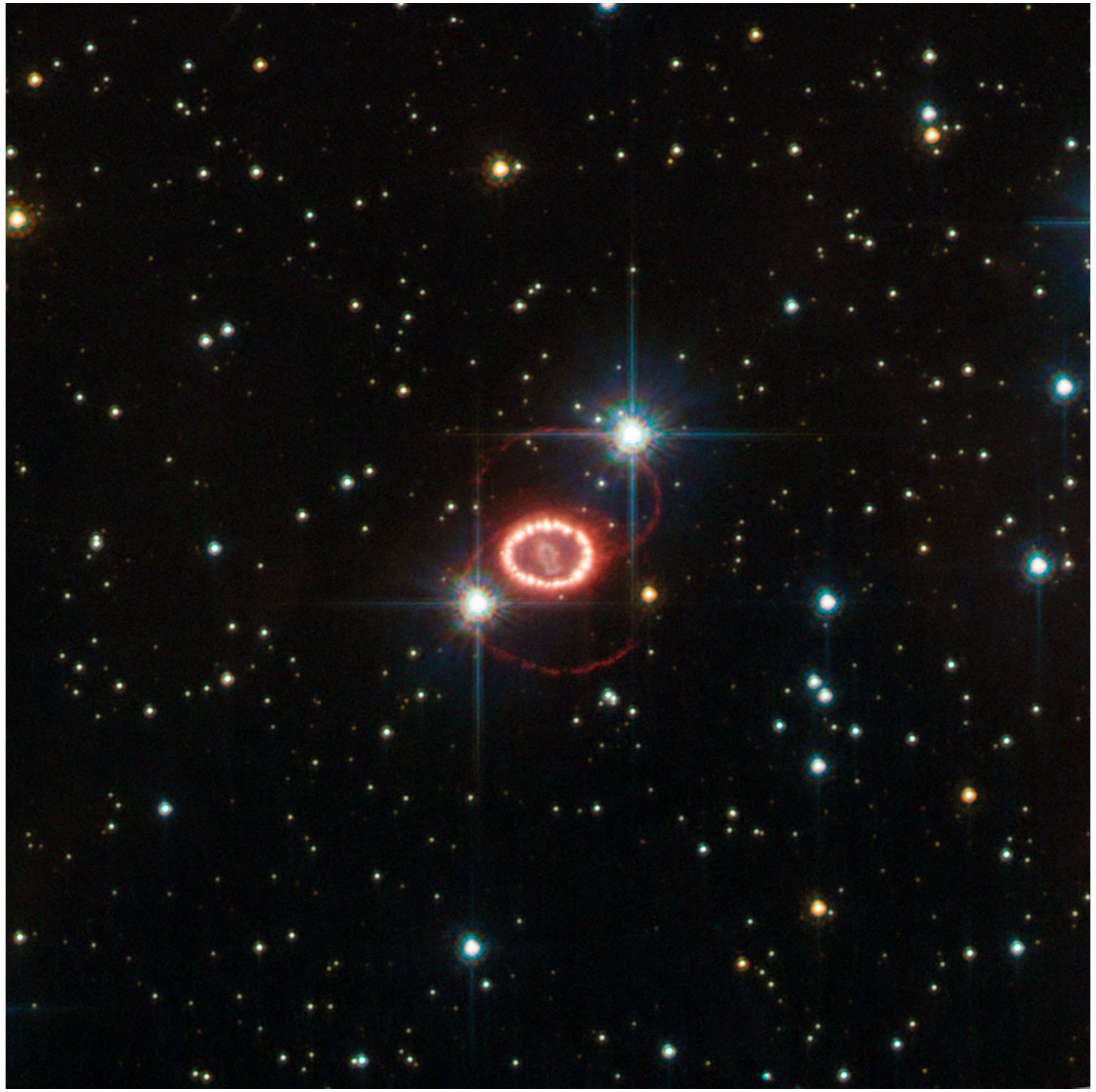
Examples:

- Clicks of a Geiger counter in a given time interval
- Number of Prussian cavalrymen killed by horse-kicks
- Goals in football(?)

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

Example for Poisson inference - SN 1987A

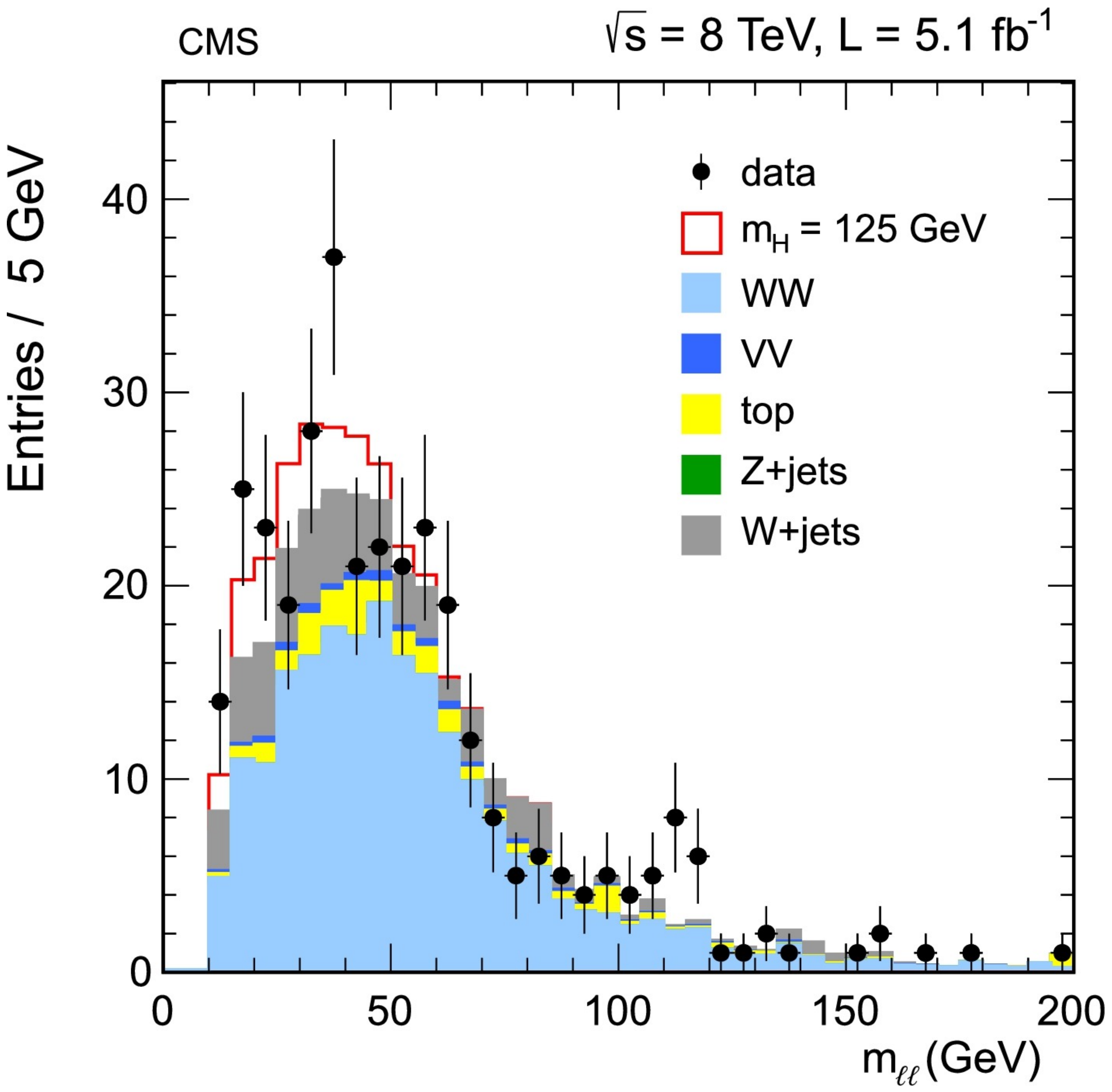
- Kamiokande II measured 12 neutrinos
- Expected number thus $12 \pm \sqrt{12}$
- Sufficiently large number for approximation?



ESA/Hubble & NASA

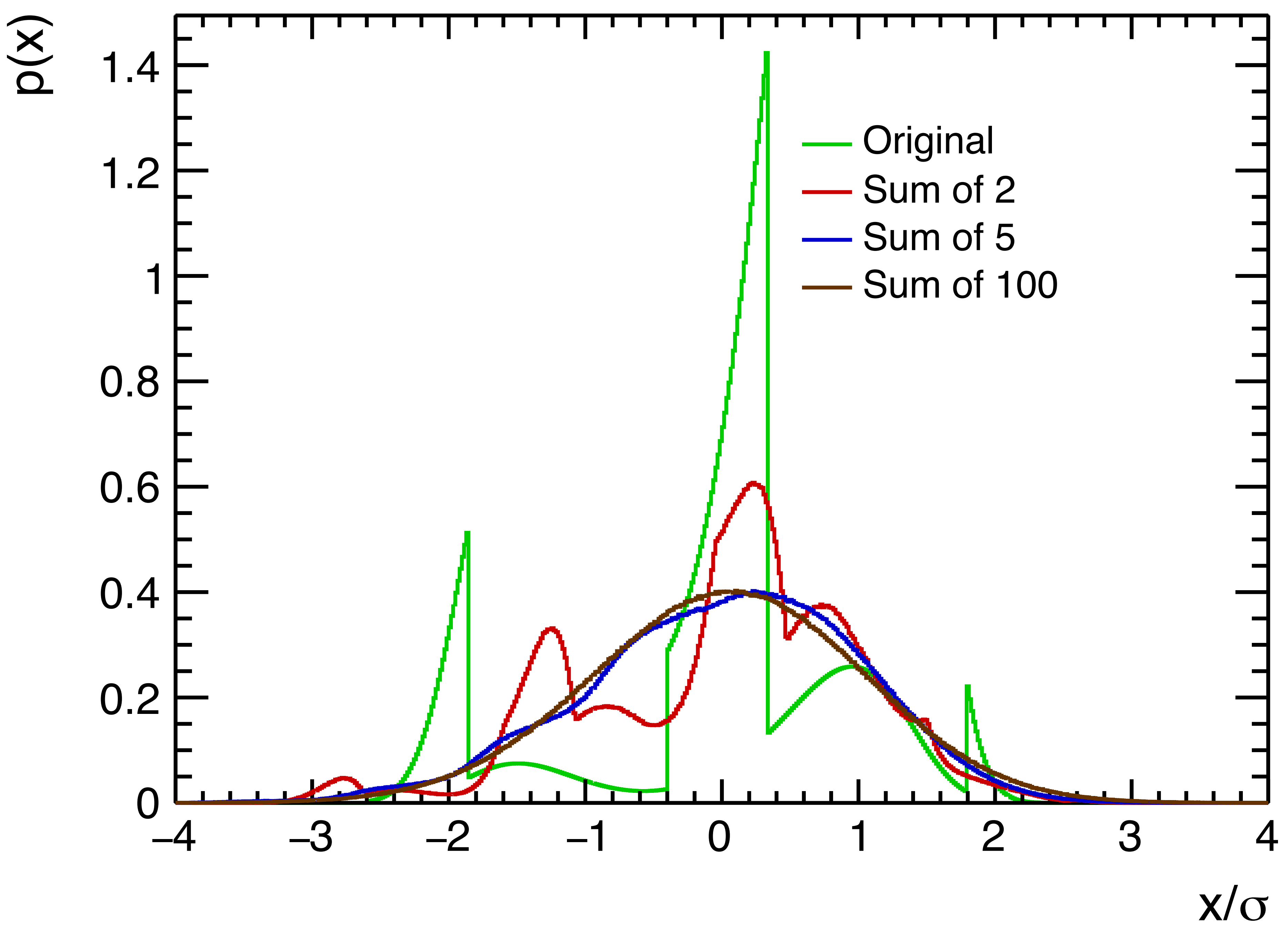
Poisson distribution - Histogram entries

- For histogram entry: each particle (or pair) has a very small chance of landing in a particular bin
- Different events don't interfere - independence
- Often error bars as $\sqrt{N}$ of the entries



Convoluting many distributions

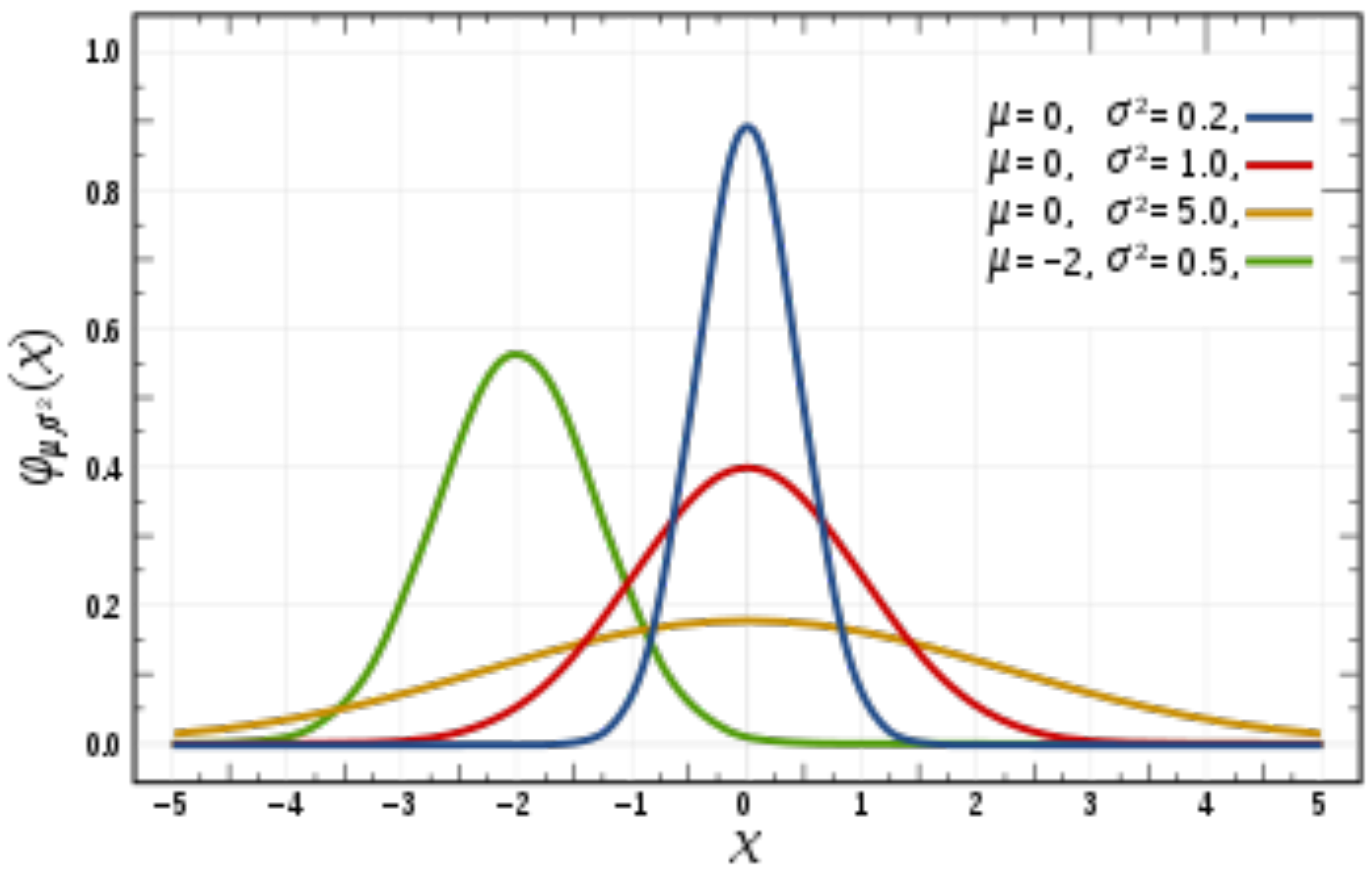
When summing up variables from a complicated distribution, the sum starts resembling a normal or Gaussian distribution



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:

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

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:

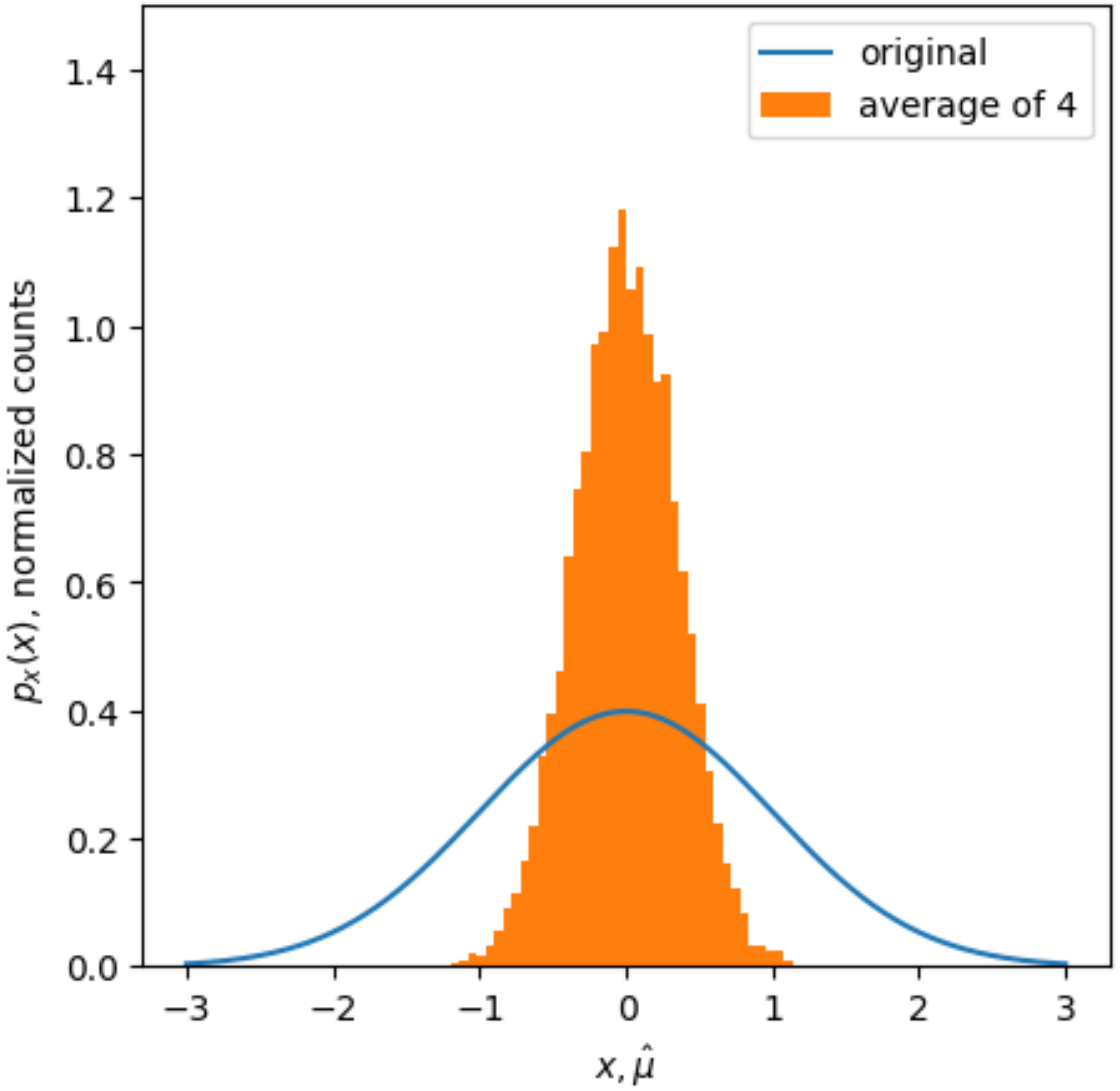
$$y = \sum_{i=1}^{n} x_i \quad \xrightarrow{n \to \infty} \quad y \text{ follows Gaussian with } \quad E[y] = \sum_{i=1}^{n} \mu_i, \; V[y] = \sum_{i=1}^{n} \sigma_i^2$$

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

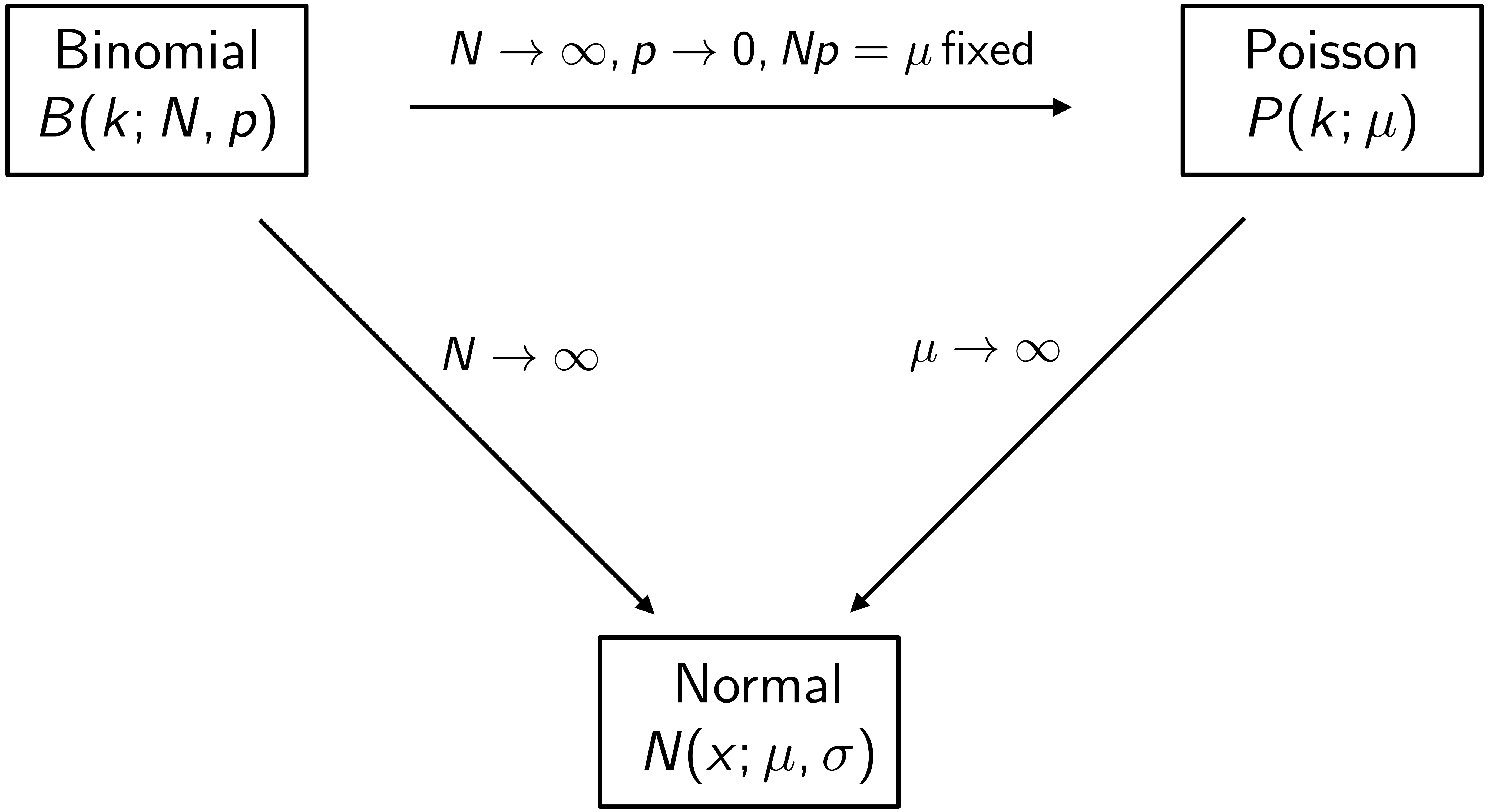
The Gaussian distribution is a stable distribution → sum or difference of two Gaussian random variables is again a Gaussian.

Averaging measurements

- Gaussian distribution $p(x) = g(x, \mu, \sigma)$
- Reasonable estimator for μ is $\hat{\mu} = x$, with standard deviation σ
- For several measurements $x_1, x_2, \ldots$, use mean $\hat{\mu} = \frac{1}{N} \sum x_i$
- We know that this is the convolution of many g scaled by $1/N$
- The variance of the sum is $V[\sum X_i] = N\sigma^2$, meaning $V[\hat{\mu}] = \sigma^2/N$
- Thus the uncertainty of the estimate is $\sigma/\sqrt{N}$
- This $1/\sqrt{N}$ of scaling appears frequently



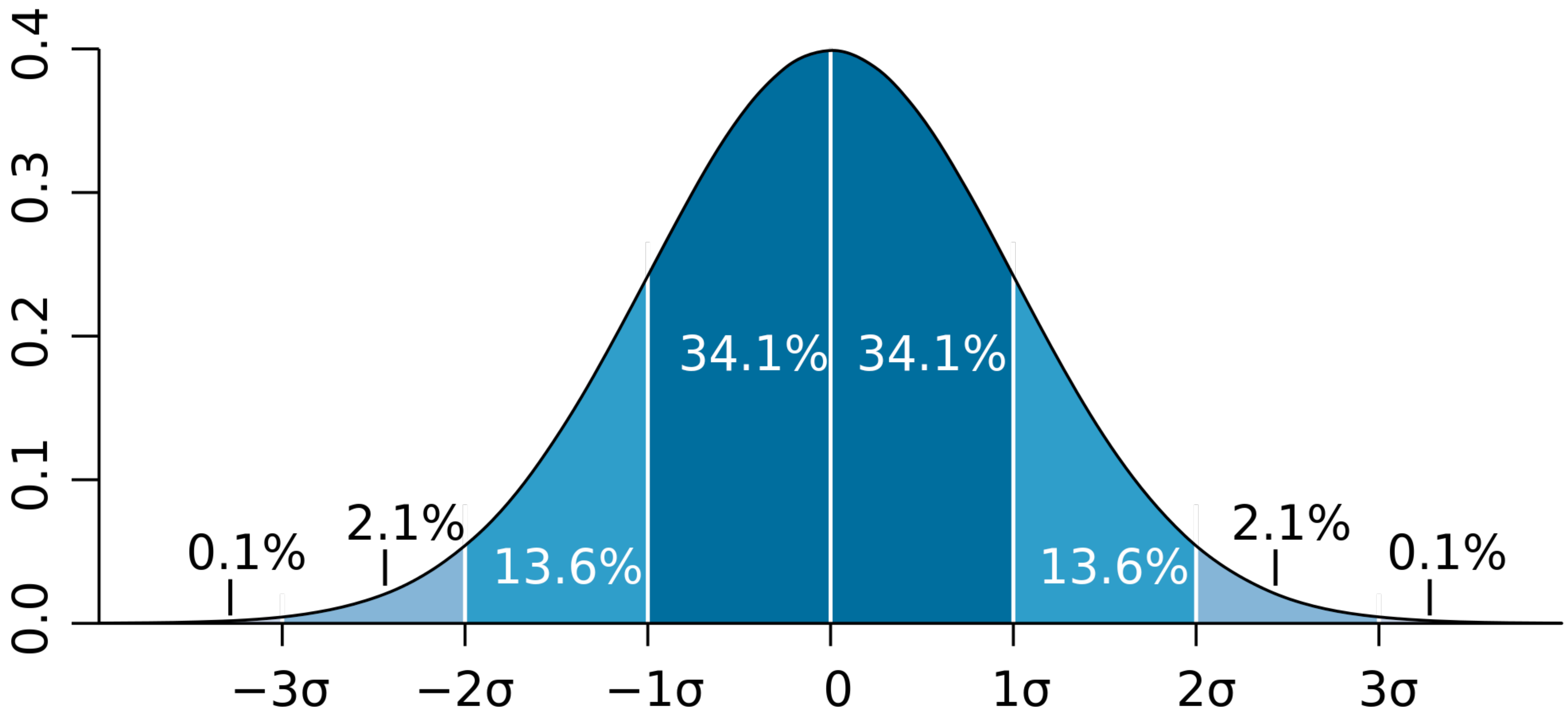
Binomial, Poisson and Normal Distribution



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

Significance of some result is often quantified as the deviation to some value relative to the uncertainty.

Multivariate normal distribution

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

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

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

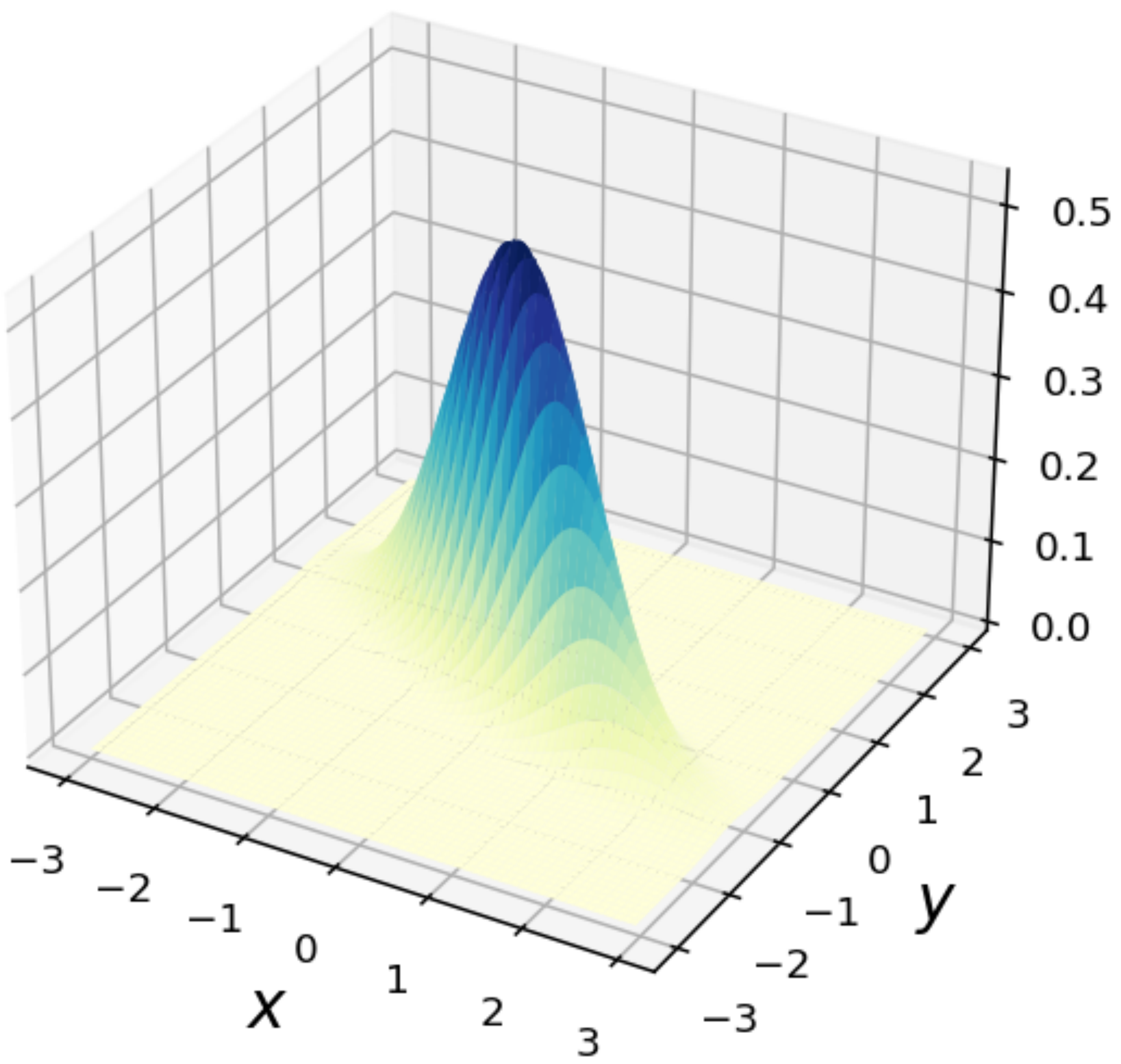
For $n = 2$:

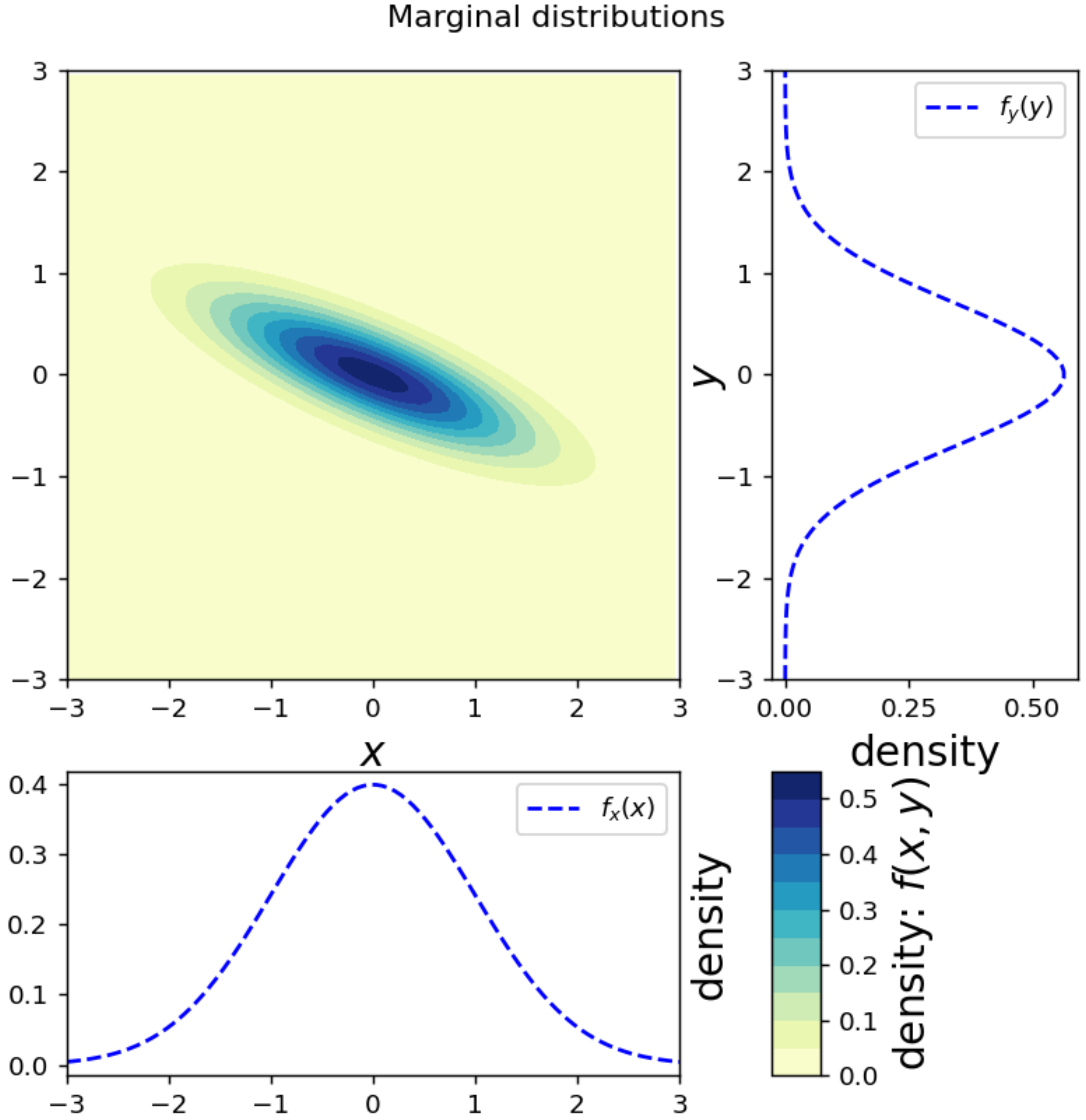
$$V = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \quad \rightsquigarrow \quad V^{-1} = \frac{1}{(1-\rho^2)} \begin{pmatrix} 1/\sigma_x^2 & -\rho/(\sigma_x\sigma_y) \\ -\rho/(\sigma_x\sigma_y) & 1/\sigma_y^2 \end{pmatrix}$$

ρ = correlation coefficient

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 distribution and error ellipse

2d Gaussian distribution:

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

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

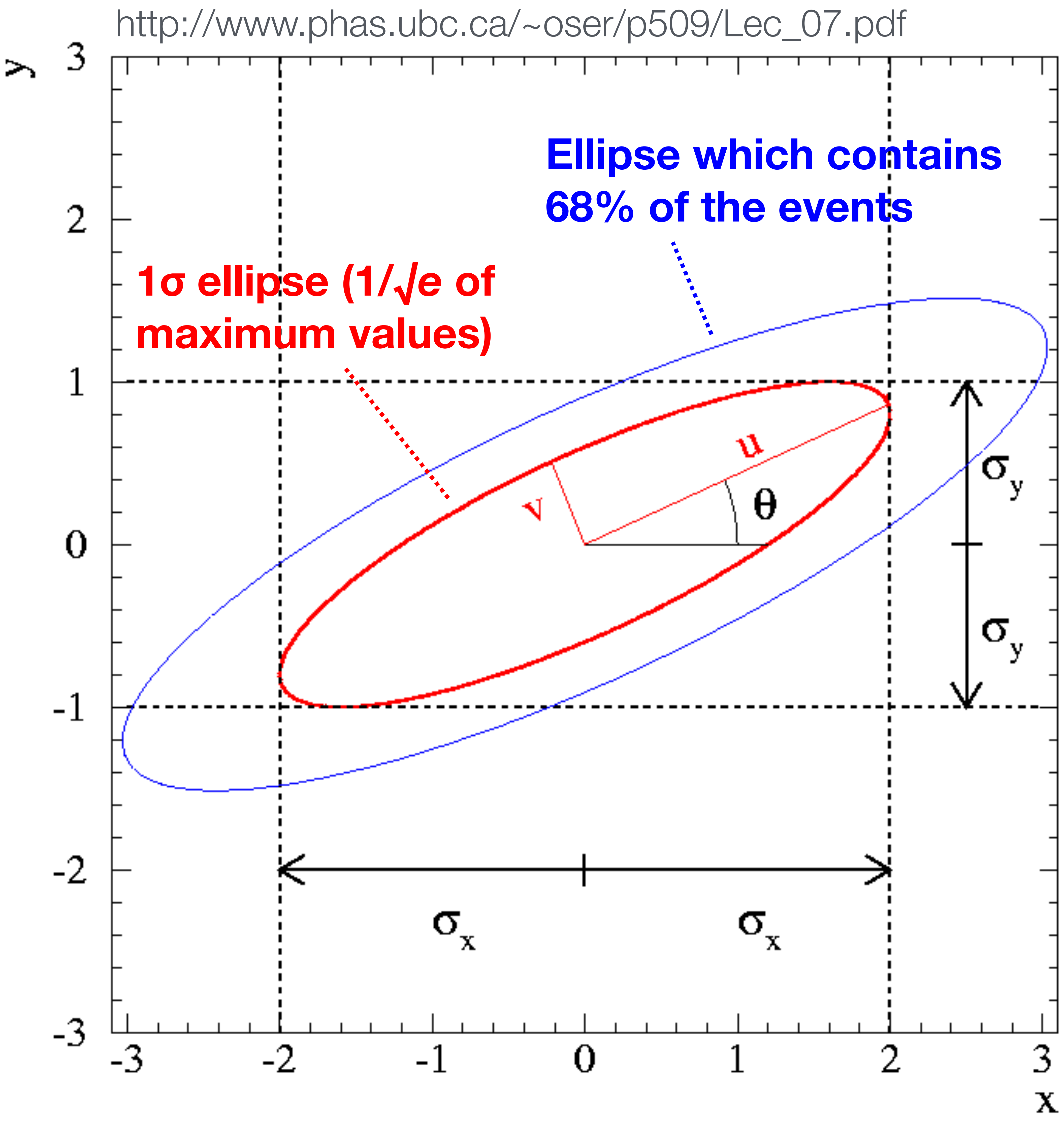
where $\rho = \mathrm{cov}(x_1, x_2)/(\sigma_1\sigma_2)$ is the correlation coefficient.

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 $-1/2$):

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

2d Gaussian: Error Ellipse



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

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

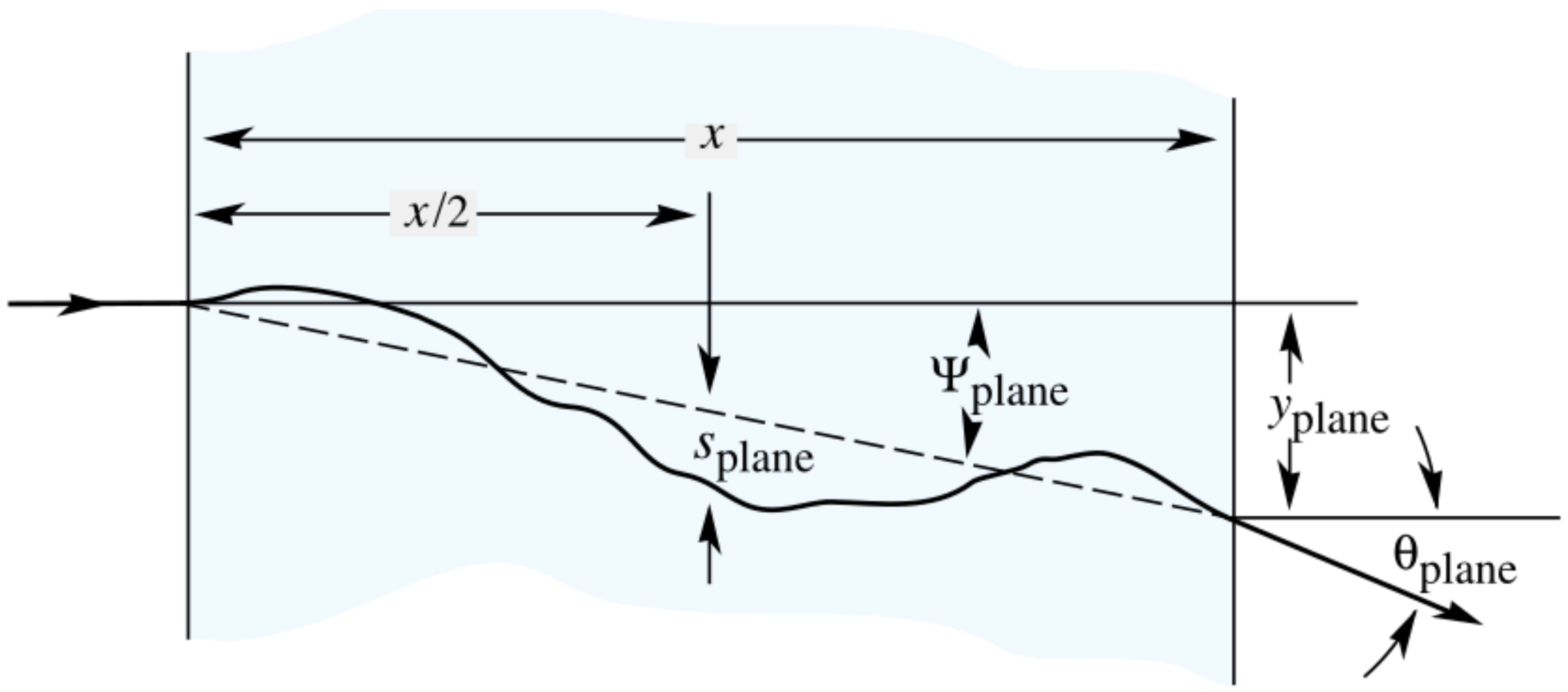
$$f_y(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left(-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right)$$

	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

Integral of probability in 1σ ellipse: 39.34%

Application of the central limit theorem: Multiple Scattering

- Particle traverses some medium
- Assume: Many independent interactions with small scattering angles
- Convolute them all for final result
  - Final distribution of directions must be a 2d Gaussian
- Derived purely from statistical principles
- All the remaining physics is then in the width of the Gaussian!



from PDG book

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:

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

$$E[m] = \mu$$

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

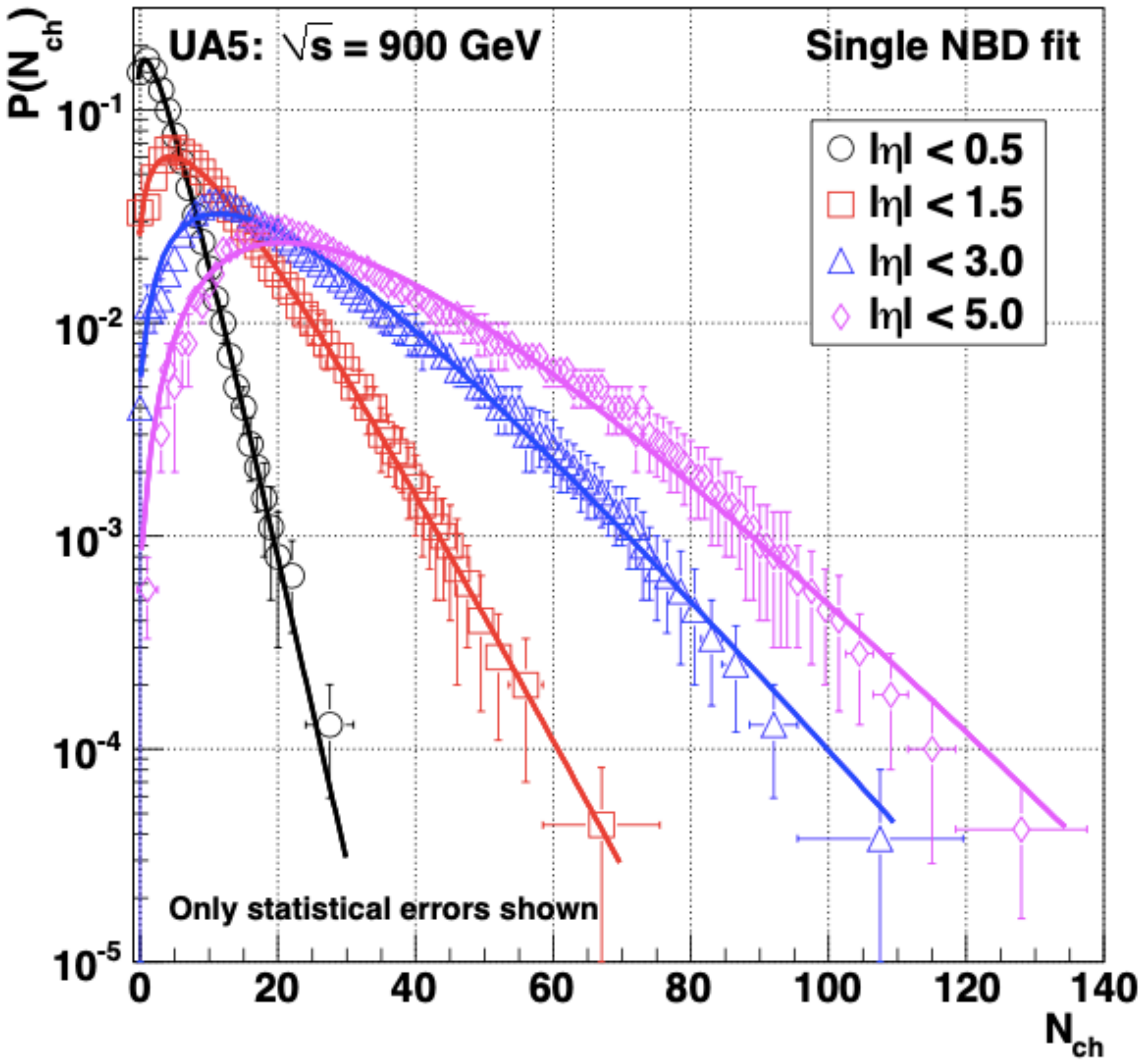
Use Gamma-fct. for non-integer values

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

$$p = \frac{1}{1+\frac{\mu}{k}}$$

[relation btw. parameters]

Example: Charged Particle Multiplicity Distribution in pp collisions



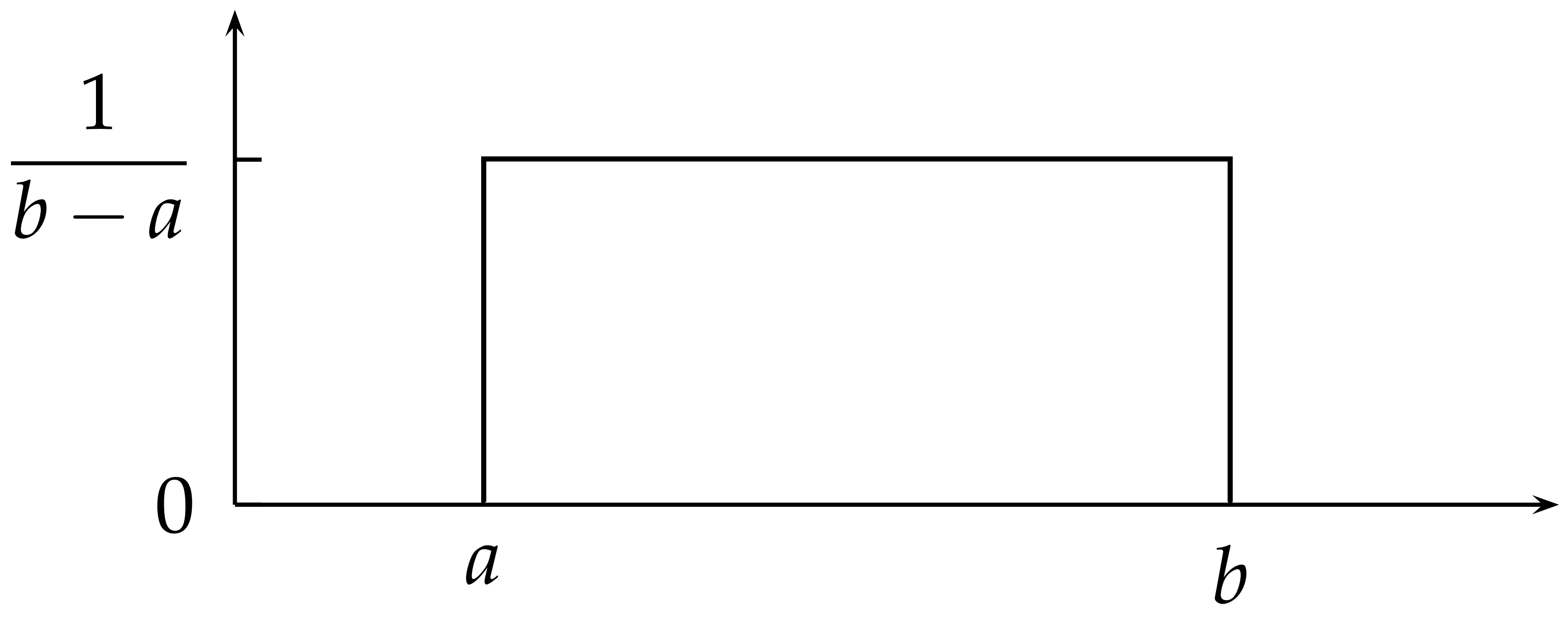
At LHC energies: Superposition of two NBD used to fit multiplicity distributions

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

Properties:

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

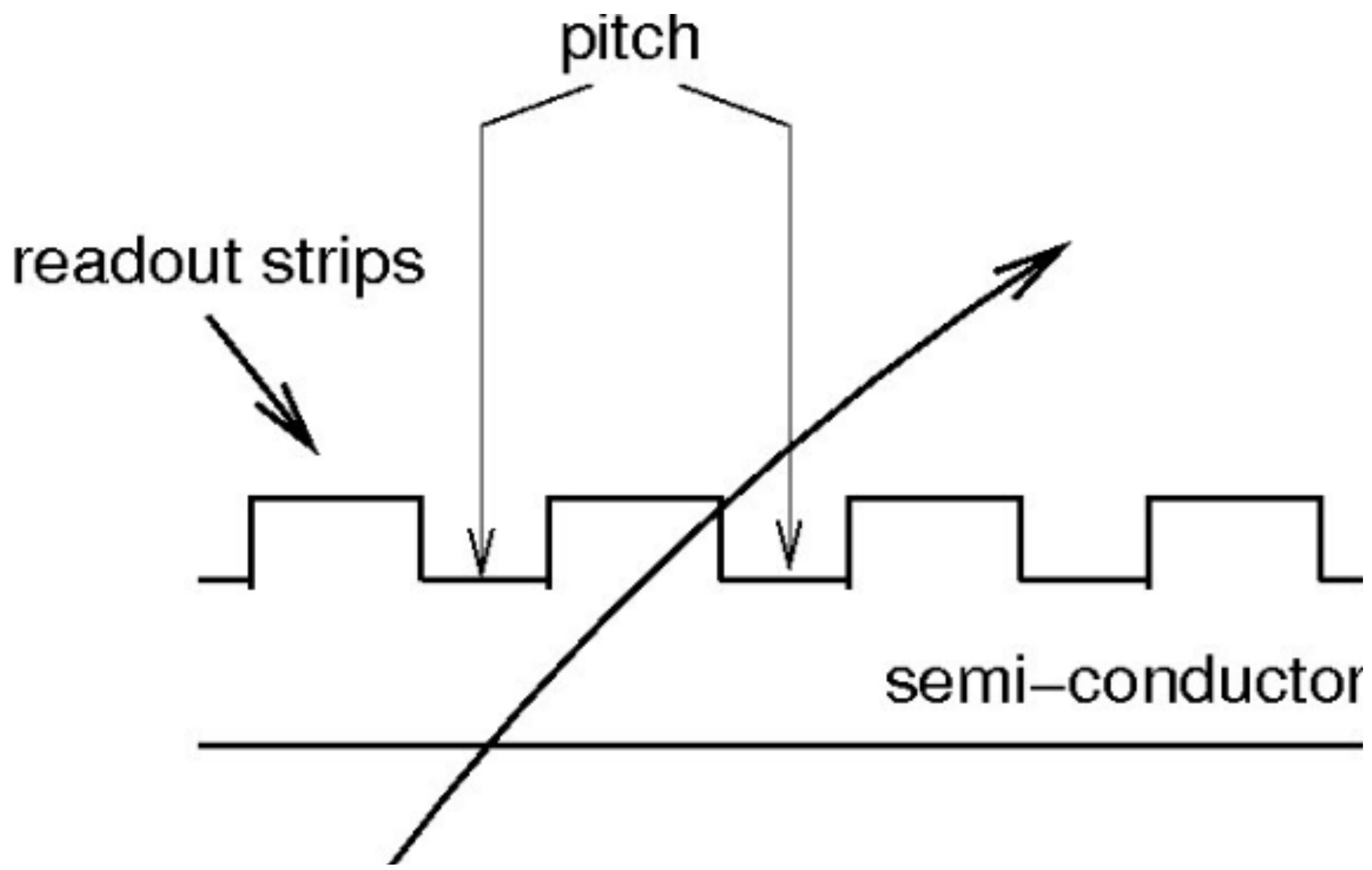


$$E[x] = \frac{1}{2}(a + b)$$

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

Example:

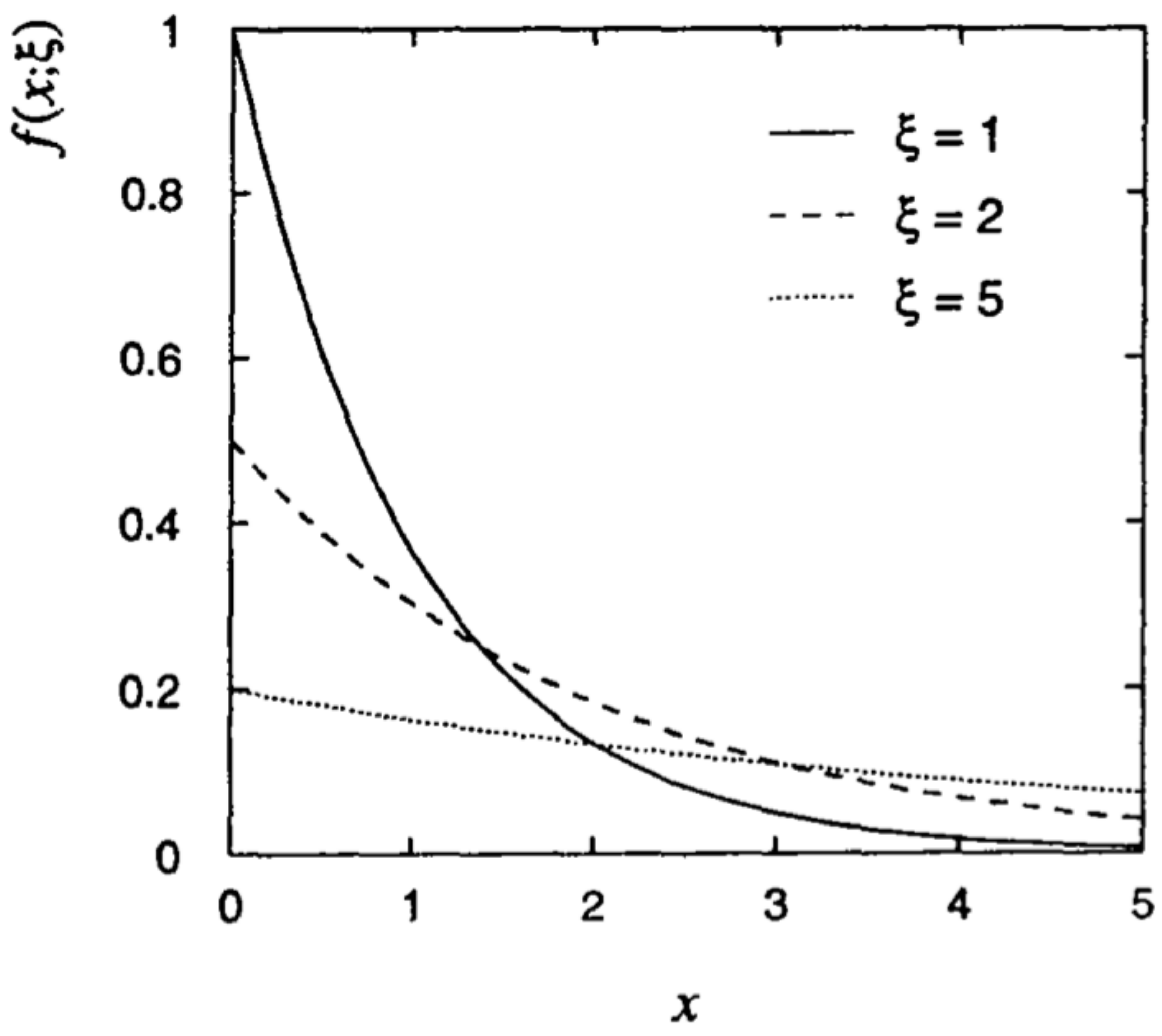
- Silicon strip detector: resolution for one-strip clusters: pitch/$\sqrt{12}$



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.

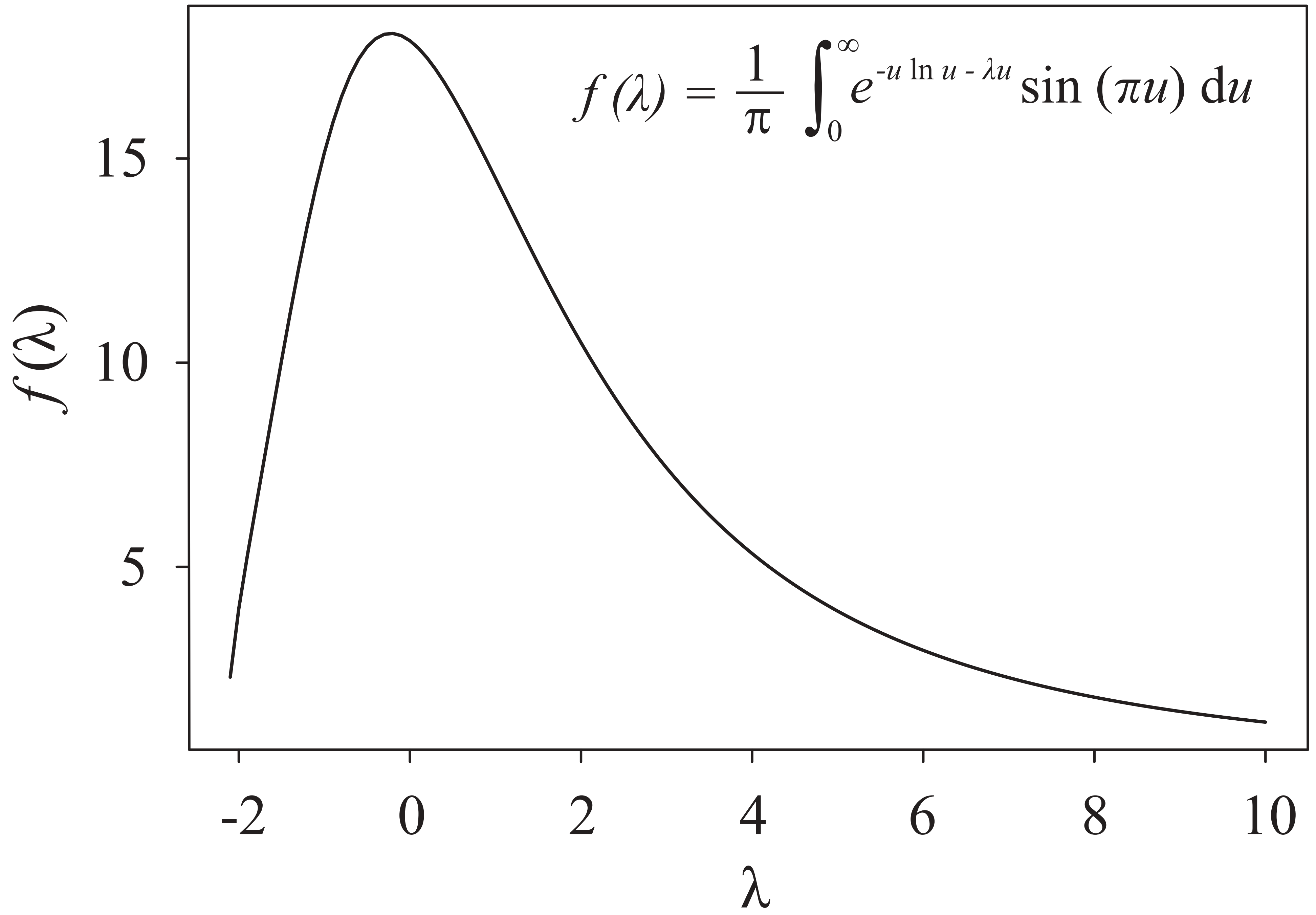
Describes energy loss of a charged particle in a thin layer of material

- Describes the sum of several Rutherford scatterings
- tail with large energy loss leads 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)



Another stable distribution.
Mean and variance not defined.

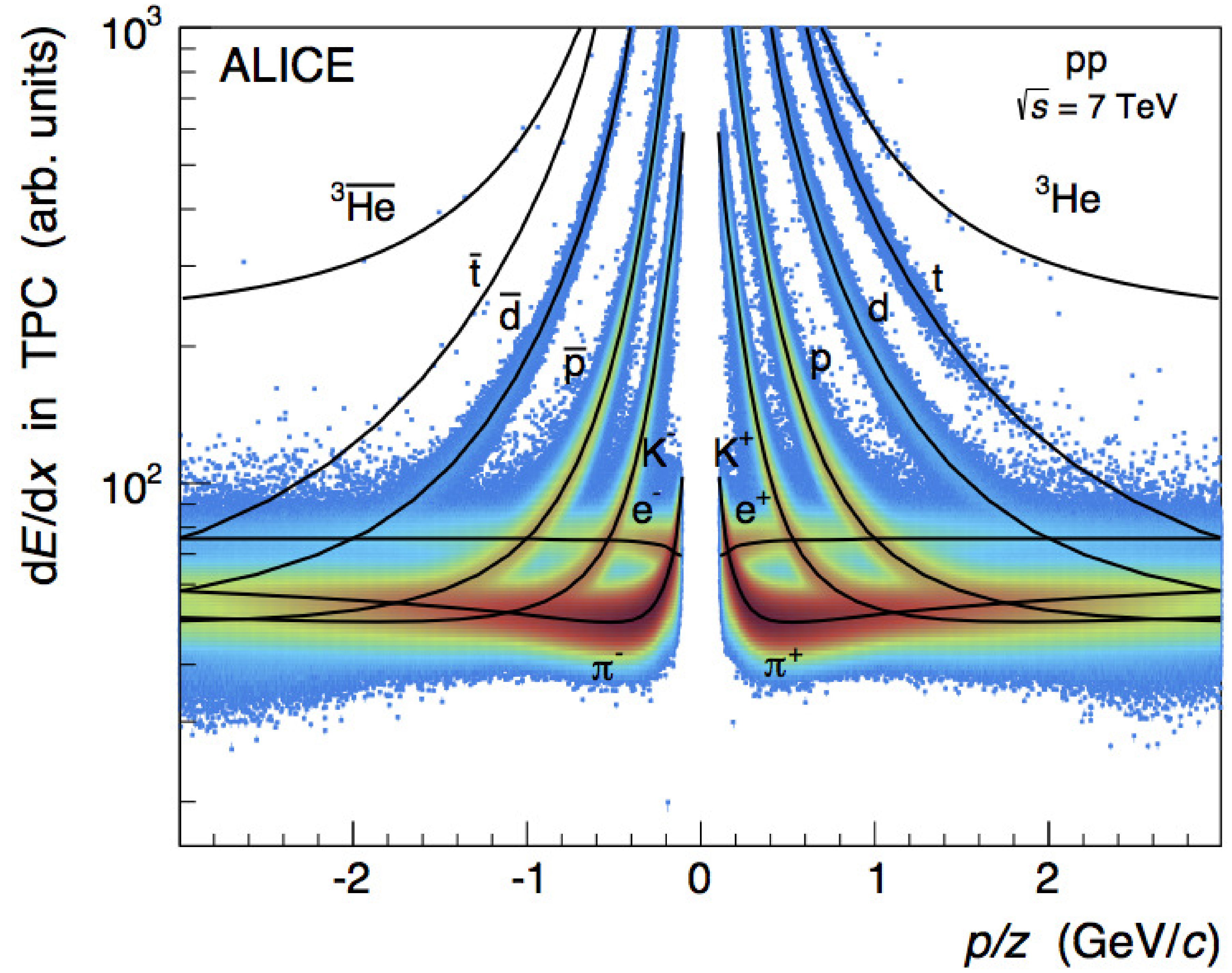
But what about the Bethe-Bloch equation?

- The Landau distribution describes fluctuations in energy loss and has no defined mean (average energy loss ≈ ∞)
- The Bethe(-Bloch) equation describes the mean energy loss of a particle

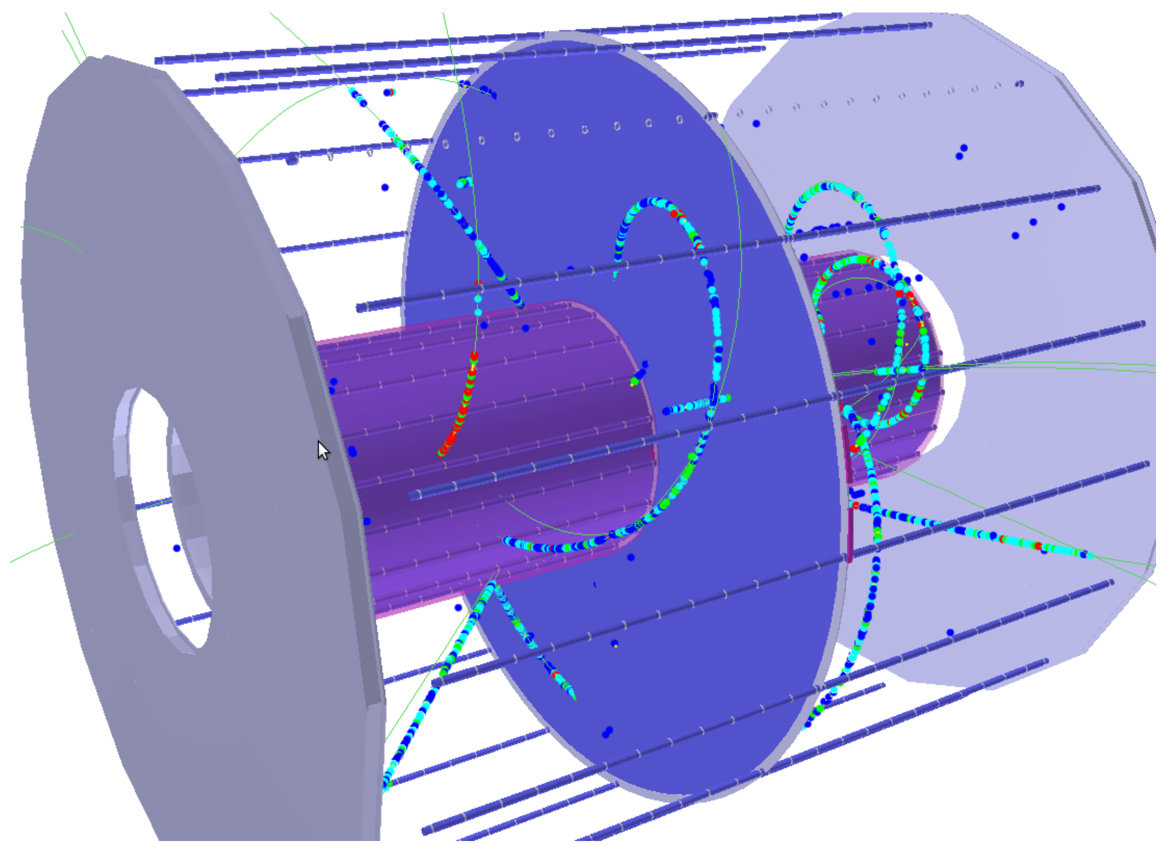
The mean of the energy loss given by the Bethe equation, […], is thus ill-defined experimentally and is not useful for describing energy loss by single particles

- PDG review Passage of Particles Through Matter

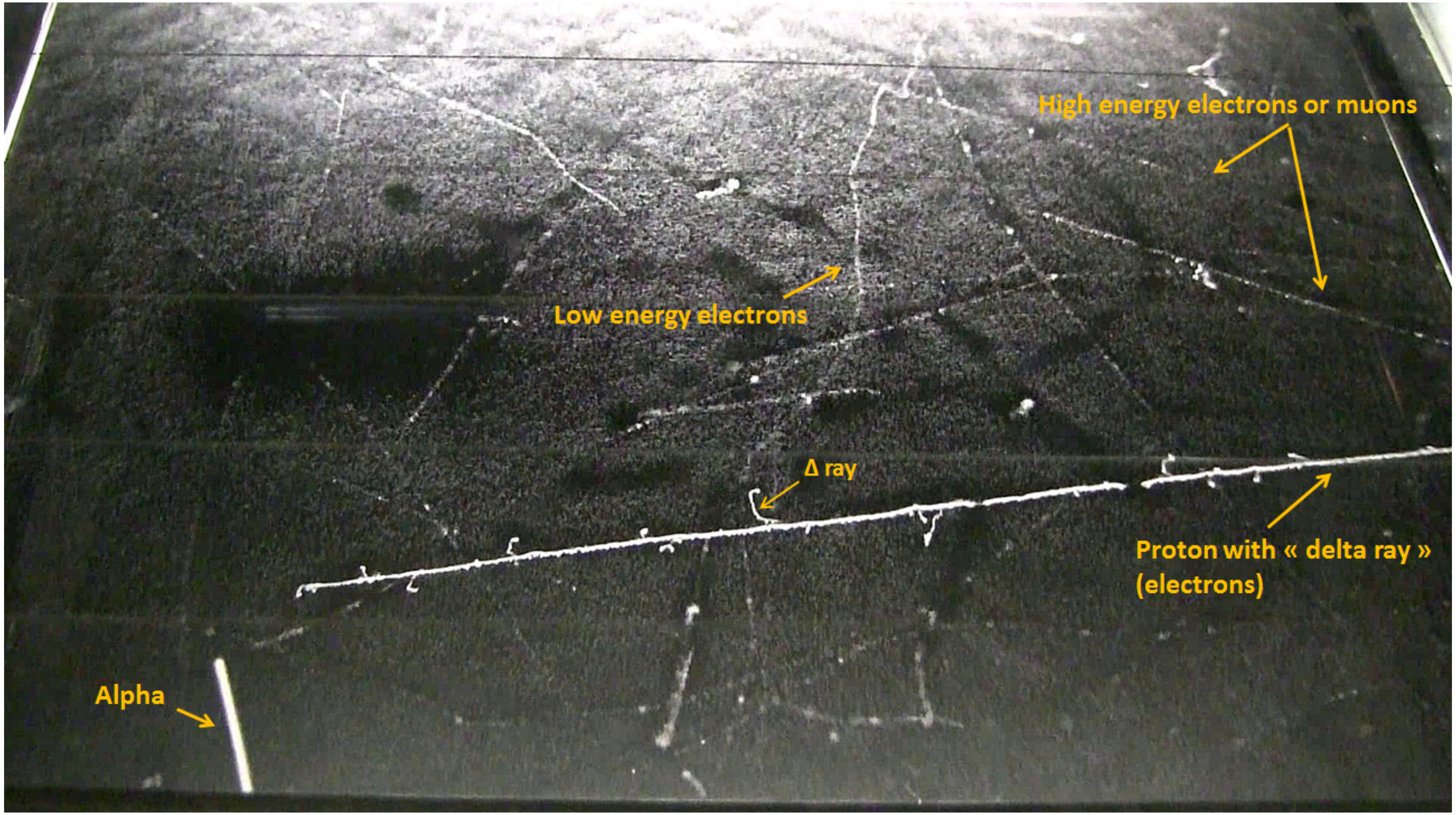
- Landau distribution assumes Rutherford goes as $1/E^2$, with divergent average - actual distribution has maximum energy transfer
- Actual distribution has mean much higher than the peak
- TPC "dE/dx" plots actually show not the mean, but the truncated mean of energy loss in reconstructed clusters -> mean of the lowest 60% of values only



ALI-PUB-108109



[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

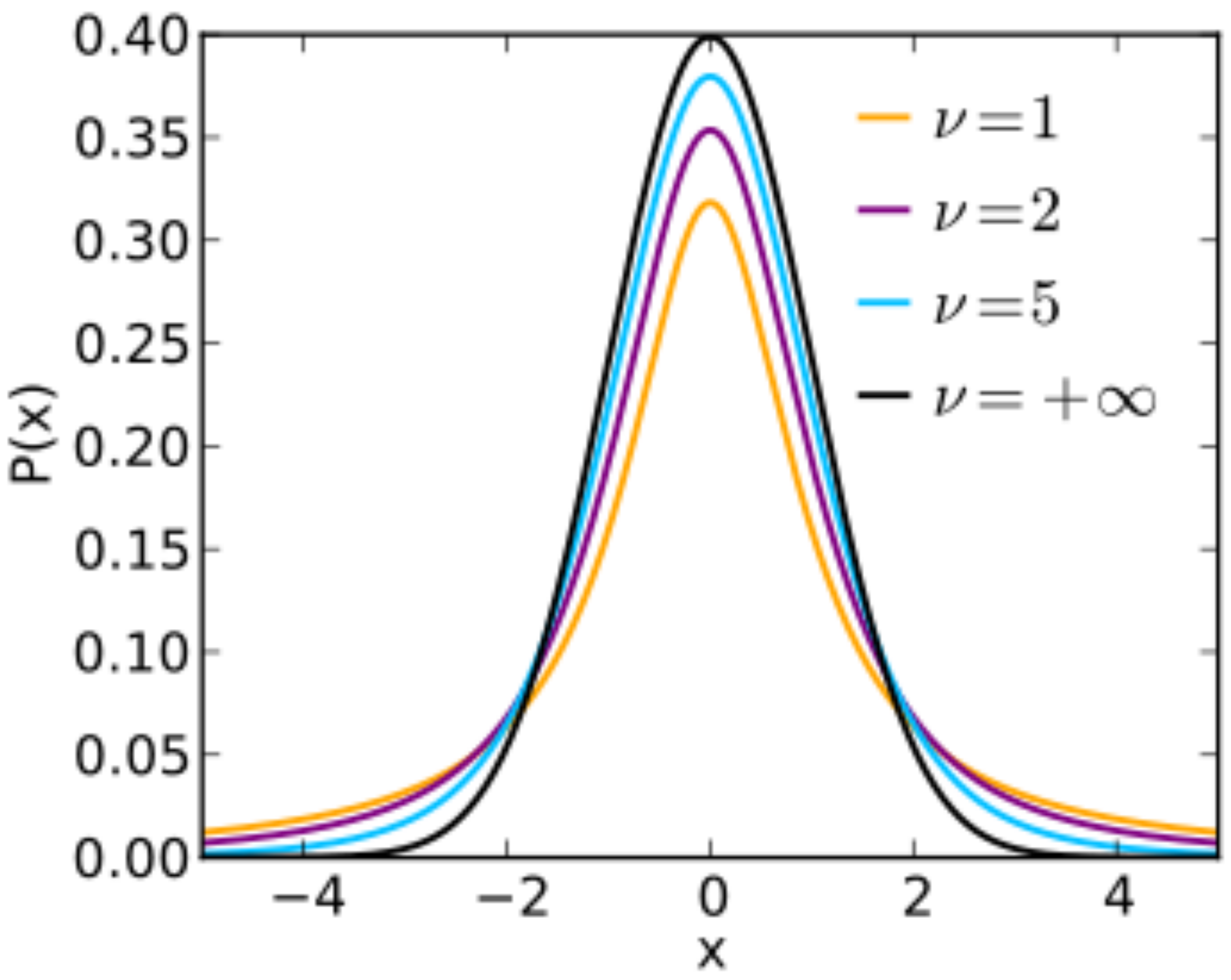
Student's t distribution:

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

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

$\nu = 1$: Cauchy distr.

$\nu \to \infty$: Gaussian



Multinomial distribution

- Binomial distribution: $p_b(k\,|\,N, \phi) = \dfrac{N!}{k!(N-k)!}\phi^k(1-\phi)^{N-k}$ for k successes, $N-k$ failures
- Can rewrite as

$$p_b(k_1, k_2\,|\,N, \phi_1, \phi_2) = \frac{N!}{k_1!k_2!}\phi_1^{k_1}\phi_2^{k_2}$$ with conditions $k_1 + k_2 = N$ and $\phi_1 + \phi_2 = 1$

- This generalizes as:

$$\boxed{p_b(k_1, k_2, \ldots\,|\,N, \phi_1, \phi_2, \ldots) = \frac{N!}{k_1!k_2!k_3!\ldots}\phi_1^{k_1}\phi_2^{k_2}\phi_3^{k_3}\ldots}$$

With conditions $\sum_i k_i = N$ and $\sum_i \phi_i = 1$

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:

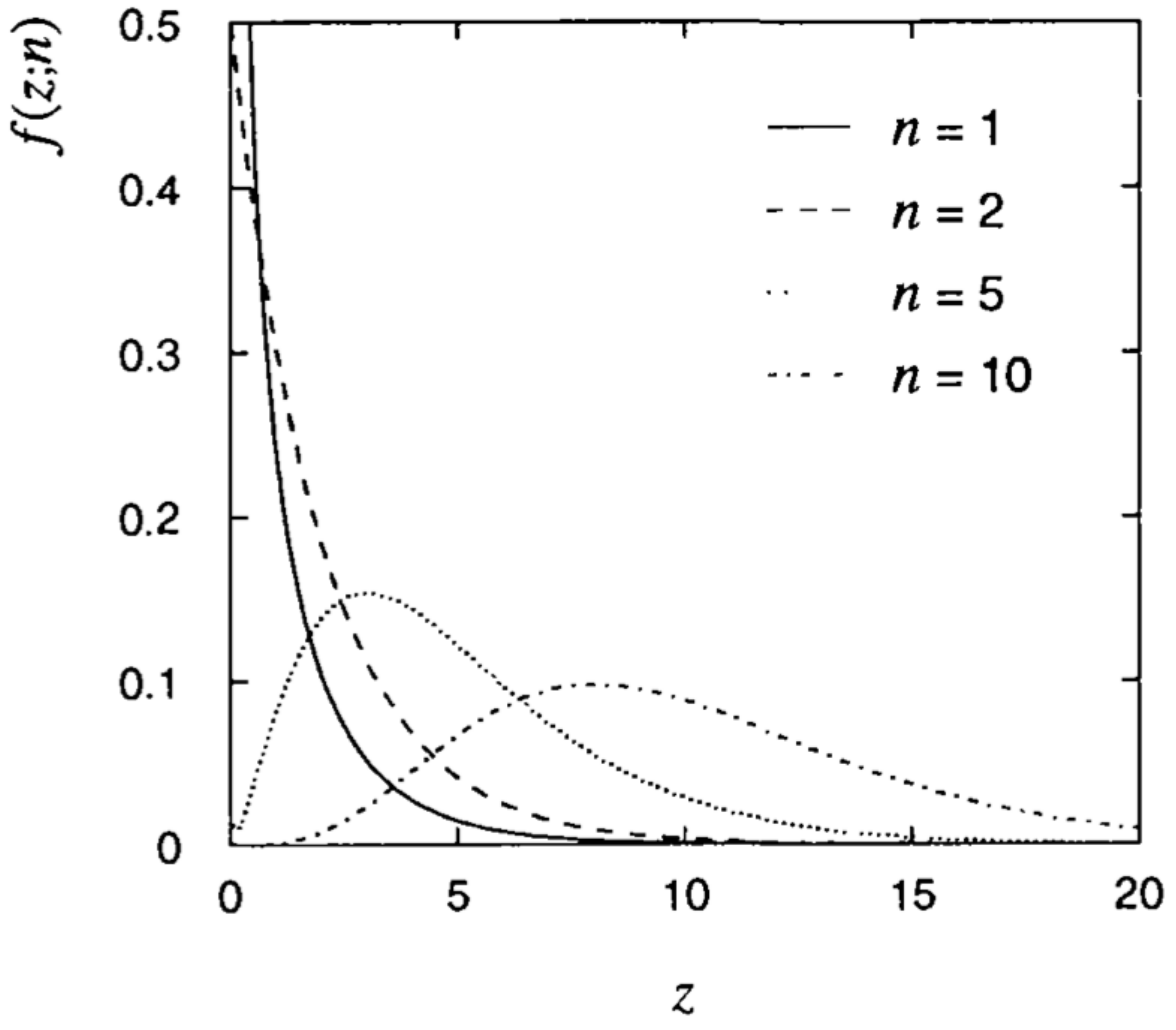
$$f(z;n) = \frac{z^{n/2-1} e^{-z/2}}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \qquad (z \geq 0)$$

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

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

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

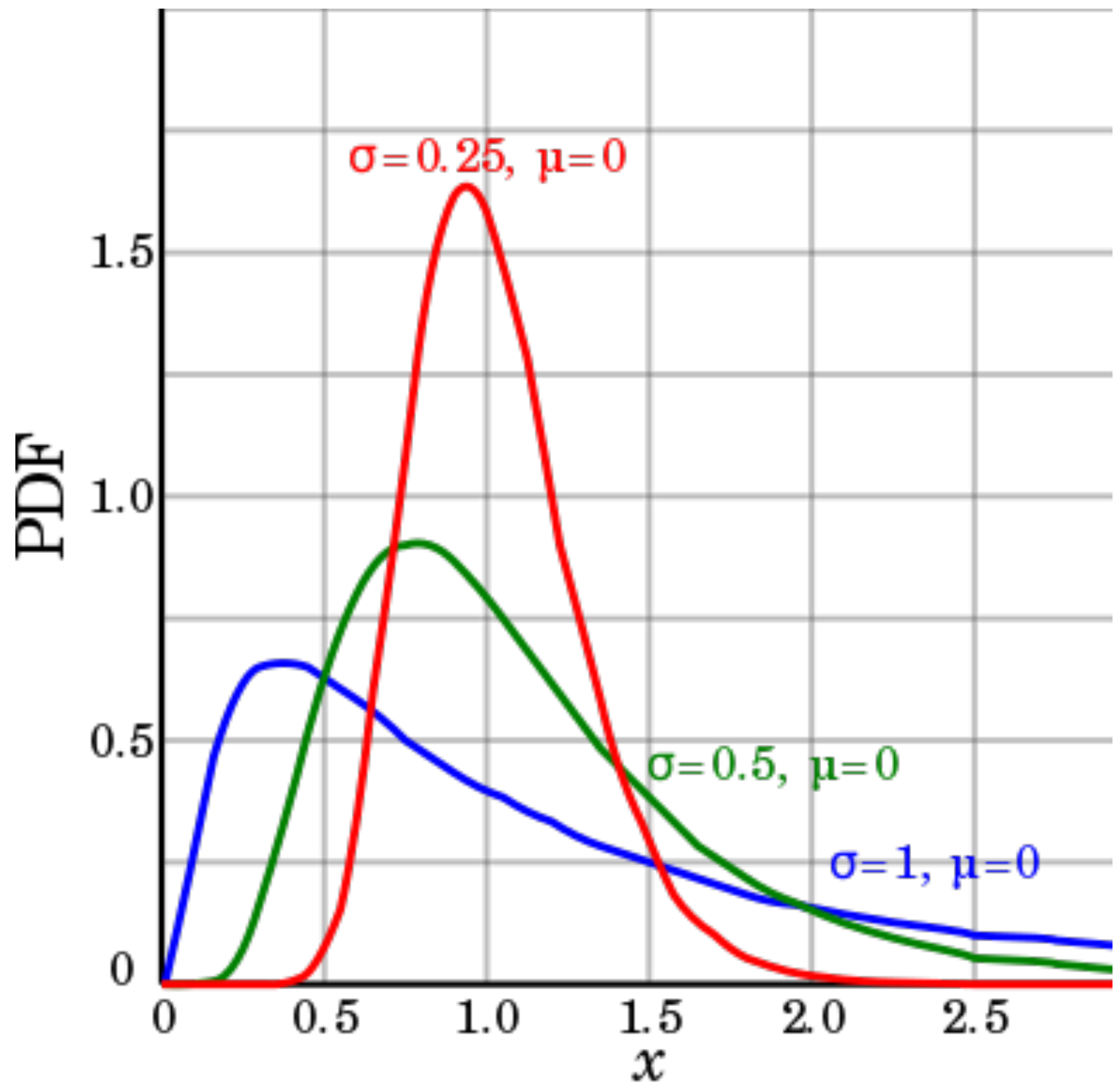
Let y be a normal (i.e. Gaussian) distributed random variable. Then $x = \exp(y)$ follows the log-normal distribution

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

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

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

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



Multiplicative version of the central limit theorem

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

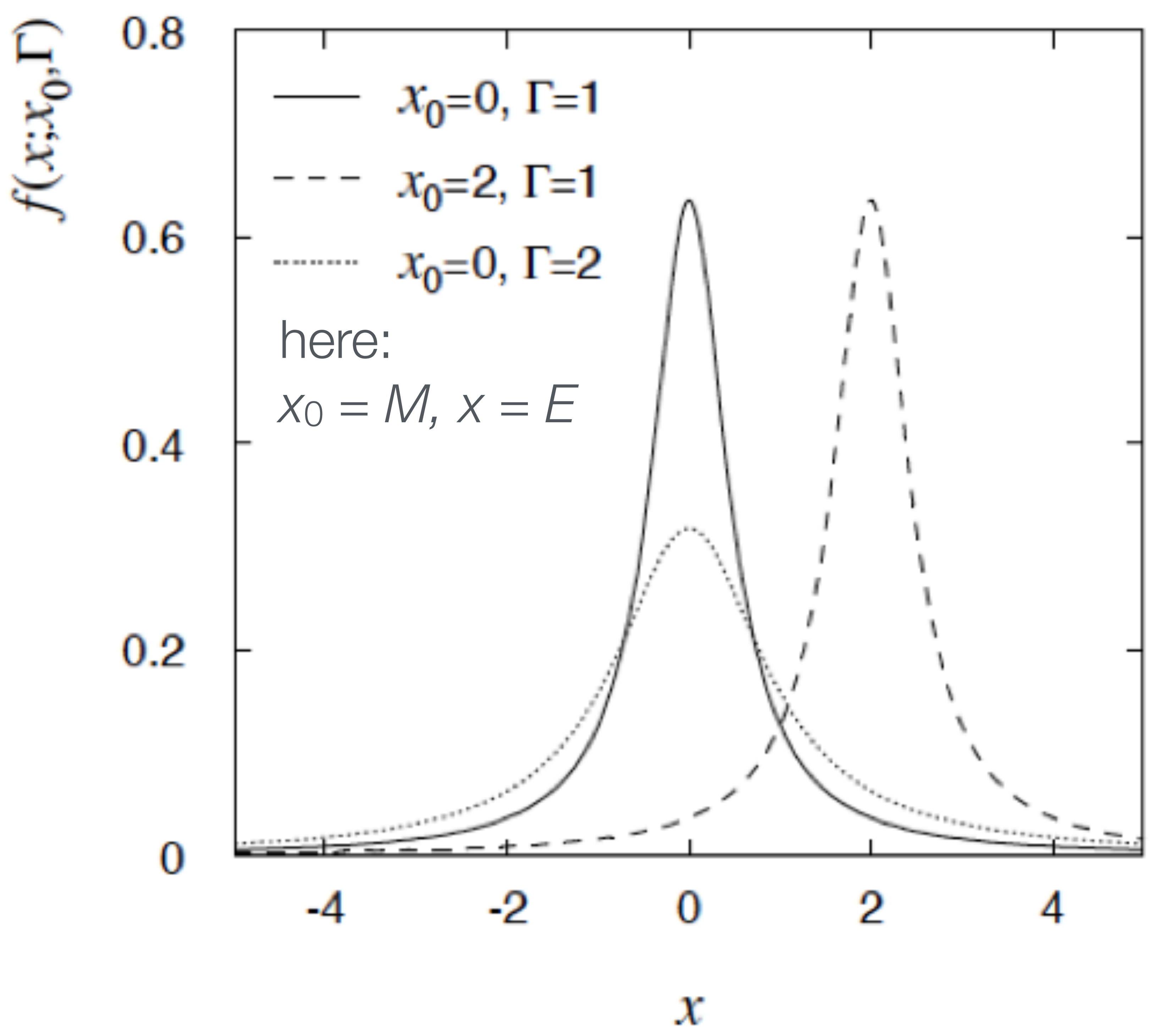
Cauchy, Breit-Wigner, or Lorentzian Distribution

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:

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



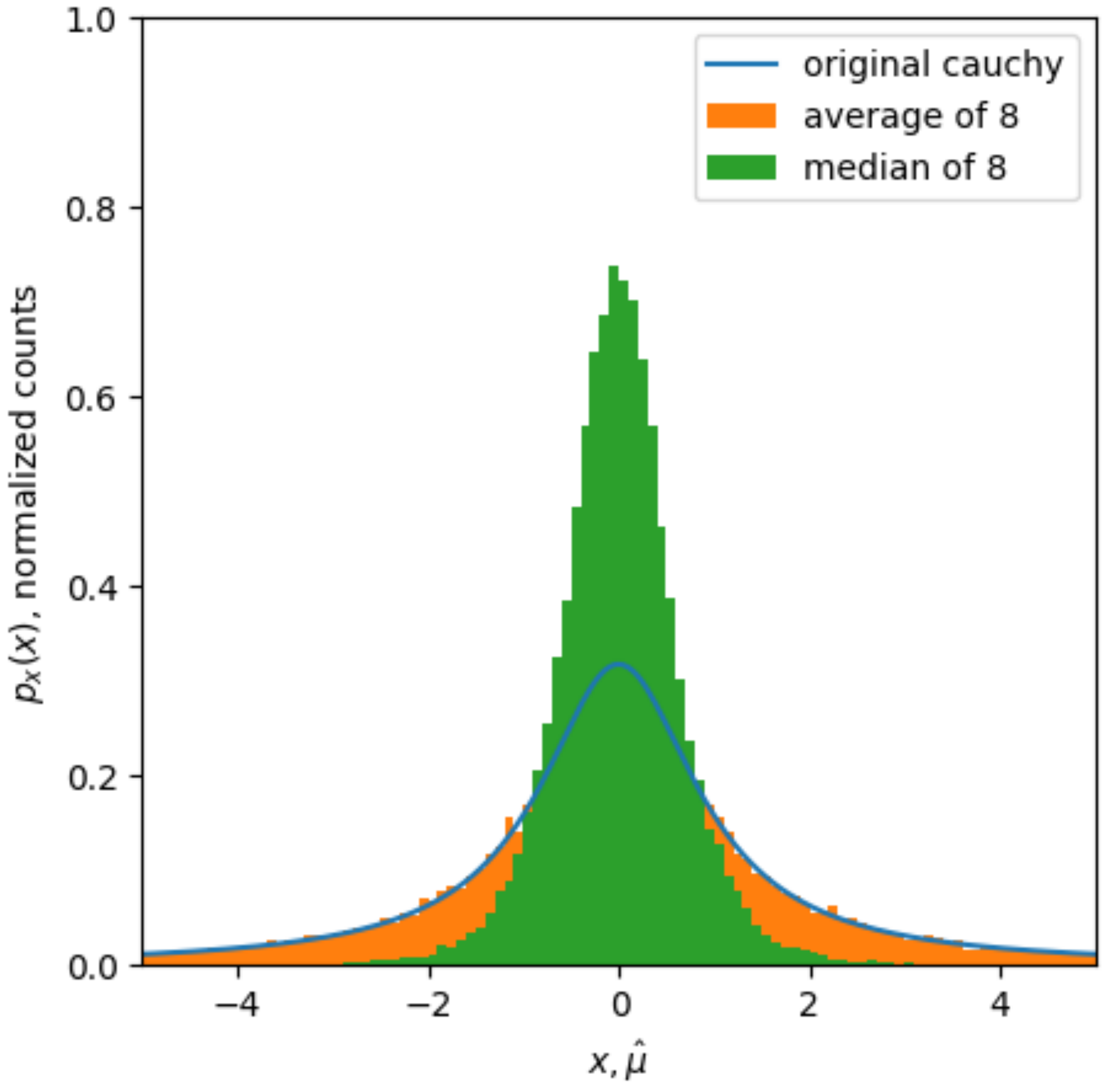
Mean and variance are undefined, mode is M.

Estimating a mass

- For Cauchy-Distribution

$$p(x\,|\,\mu,\gamma) = \frac{1}{\pi\gamma}\frac{1}{1+\left(\frac{x-\mu}{\gamma}\right)^2}$$

- Want to estimate position parameter μ (e.g. to find the mass of a decaying particle)
- Try average as estimator
- Mean and variance undefined → convolution still has infinite uncertainty
- More: Averaging does not even decrease the width γ!
- Instead using the median gives better results
  - Median often useful when distributions have wide tails

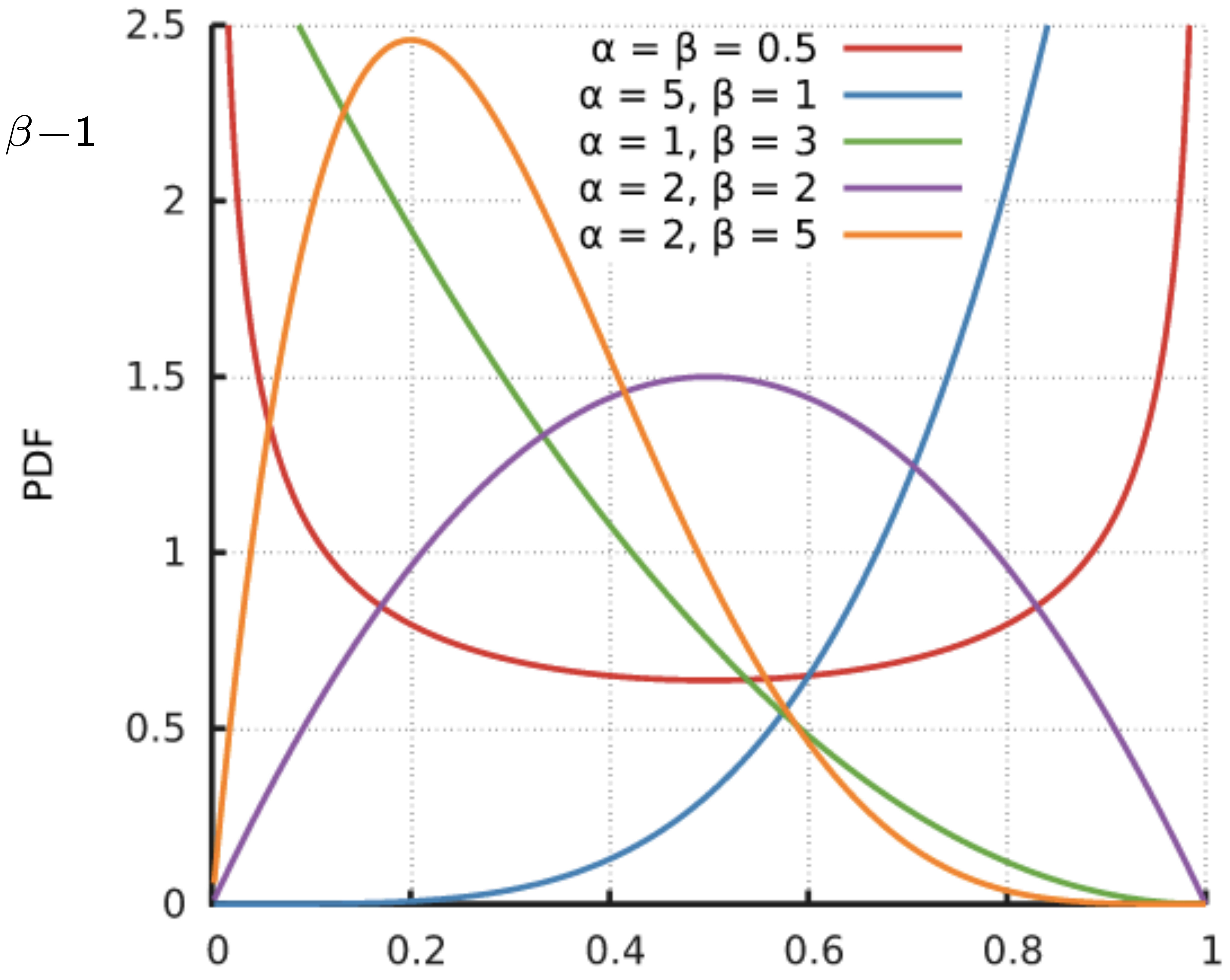


Beta Distribution

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

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

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



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.

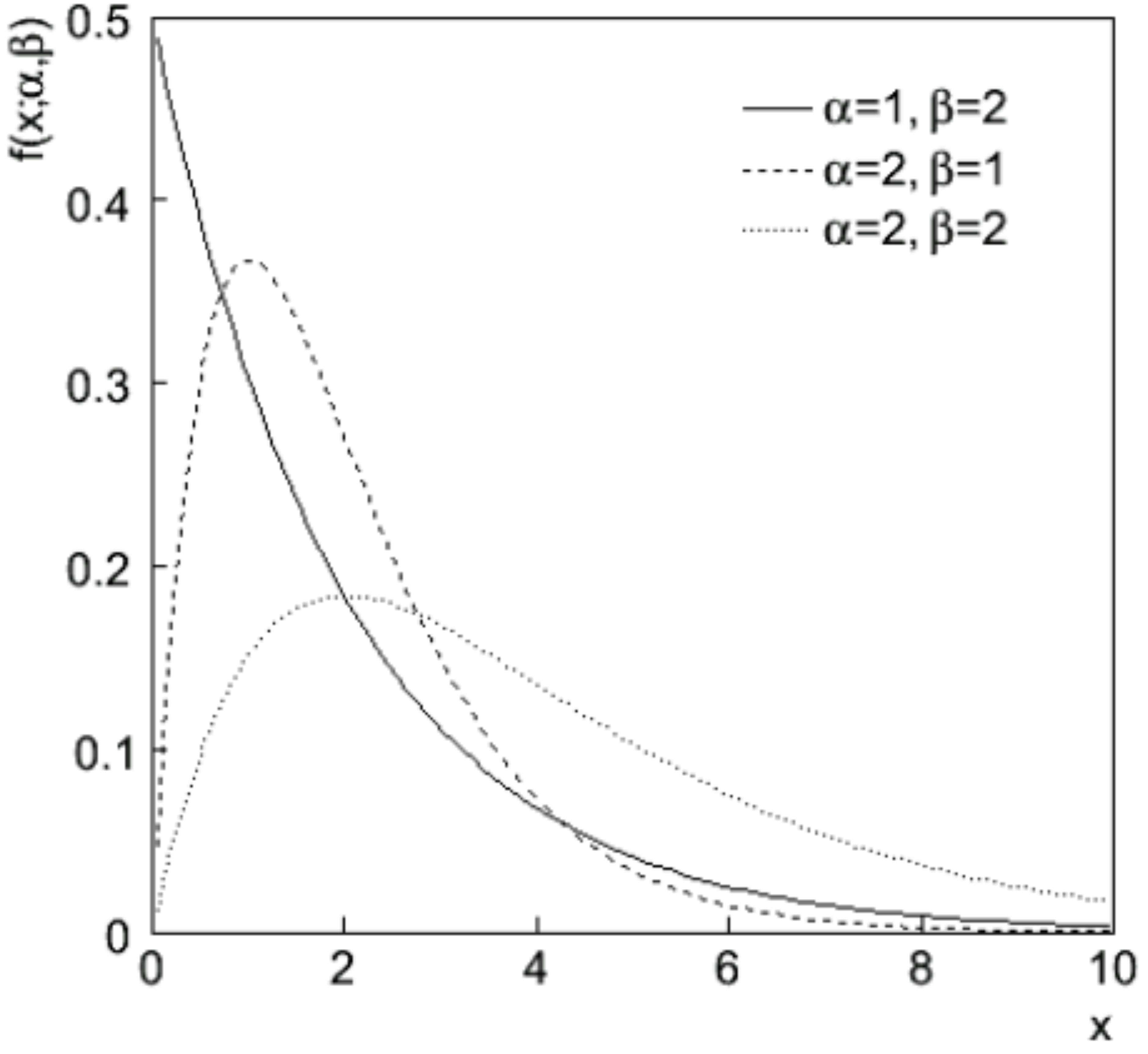
Gamma Distribution

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

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

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

Conjugate prior for Poisson likelihood and exponential likelihood



Probability of the data ⇄ likelihood

- $p(\vec{d}\,|\,\vec{\theta})$ is the probability distribution of the data for different parameters
- When considered as a function of $\vec{\theta}$ instead, it is called the *likelihood*
- Often called $\mathscr{L}$ or L with $\mathscr{L}(\vec{\theta}\,|\,\vec{d}) \equiv p(\vec{d}\,|\,\vec{\theta})$

Conclusions

- Probability distributions are the basis for mathematic modelling of measurements
- They are also important to define priors
- The likelihood is (technically) not a probability distribution but turns out to be extremely important
- In practice many distributions can be effectively modelled by Gaussians due to the central limit theorem