Probability, Statistics and Data Analysis

Allen Caldwell, Daniel Kollár (CERN) and Kevin Kröninger (U. Göttingen)

Topics:

- **Introduction to Probability and Statistics**
- **Standard distributions**
- **Estimation of parameters**
- **Confidence, probability intervals and limits**
- **Goodness-of-fit tests, hypothesis testing**
- **Bayesian and frequentist approaches**

+

- **Introduction to BAT**
- **Analyze data sets, extract limits on right-handed W, existence of double-beta-decay**

BAYESIAN ANALYSIS TOOLKIT

Structure of Block Course

Monday:

Morning – lecture on basic distributions (binomial, Poisson, Gauss) along with examples, central limit theorem

Afternoon – introduction to BAT, installation on laptop, try out examples on your laptop

Tuesday:

Morning – lecture on learning from data, Bayes Theorem, parameter estimation, calculating probability intervals and limits

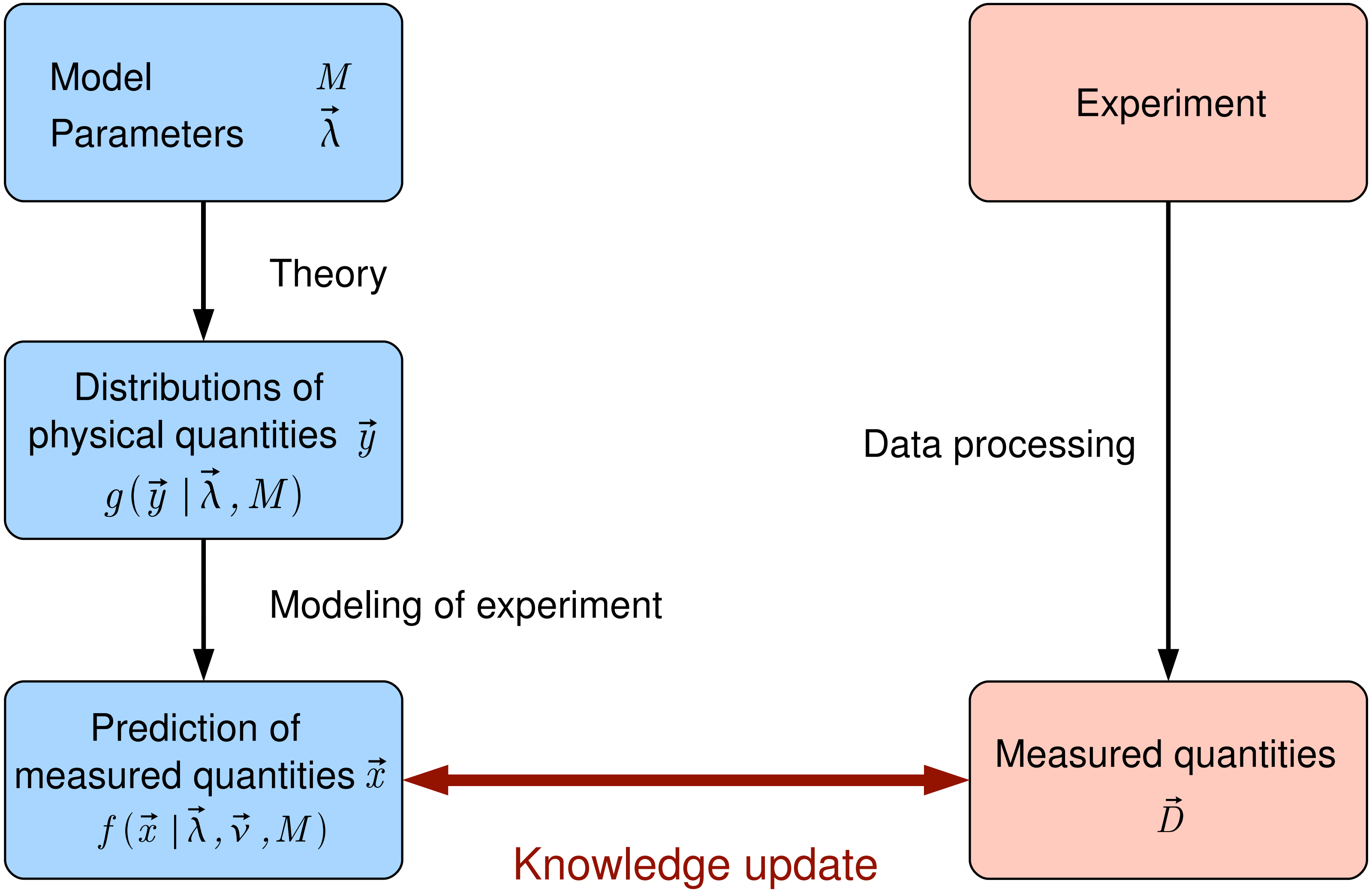
Afternoon – analyzing charged current deep inelastic scattering data on your computer to find limits on the mass of a putative right-handed W boson

Wednesday:

Morning – relationship to other approaches, hypothesis testing, goodness-of-fit tests, introduction to frequentist approach and discussion

Afternoon – analyze double beta decay spectrum to extract signal strength in the presence of background, estimate parameter values, test goodness-of-fit

Learning from Data



Types of Probability

Direct probability

• we have a model and can predict the relative frequency of possible outcomes. E.g., to evaluate the frequency of possible outcomes from flipping a coin (heads/tails), we start by making a model:

> ‘the coin is fair and does not change its properties over time; by symmetry, H, T each occur with the same frequency in the long term’

• we have frequency of outcomes from measurements, and use this to calculate future outcomes (e.g., insurance companies). Here it is assumed that the conditions under which the data were accumulated apply for the calculation of future frequencies.

We then use our models to calculate frequencies of possible outcomes.

Probability and Intuition

Imagine we flip a coin 10 times, and get the following result:

T H T H H T H T T H

We now repeat the process with a different coin and get

T T T T T T T T T T

Which outcome has higher probability ?

Probability and Intuition

Take a model where H, T are equally likely. Then,

outcome 1 $\quad prob = (1/2)^{10}$

And

outcome 2 $\quad prob = (1/2)^{10}$ $\quad$ exactly the same !

Something seem wrong with this result ? This is because we evaluate in our minds many probabilities at once (there is usually more than one probability which can be assigned to a data set). The result above is the probability for any sequence of ten flips of a fair coin. Given a fair coin, we could also calculate the chance of getting n times H:

$$\binom{10}{n}\left(\frac{1}{2}\right)^{10}$$

Probability and Intuition

And we find the following result:

n	p
0	$1 \cdot 2^{-10}$
1	$10 \cdot 2^{-10}$
2	$45 \cdot 2^{-10}$
3	$120 \cdot 2^{-10}$
4	$210 \cdot 2^{-10}$
5	$252 \cdot 2^{-10}$
6	$210 \cdot 2^{-10}$
7	$120 \cdot 2^{-10}$
8	$45 \cdot 2^{-10}$
9	$10 \cdot 2^{-10}$
10	$1 \cdot 2^{-10}$

There are many more ways to get 5 H than 0, so this is why the first result somehow looks more probable, even if each sequence has exactly the same probability in the model.

Maybe the model is wrong and one coin is not fair ? How would we test this ? This is what we do in science – test validity of models.

Types of Probability

In scientific work, what we want to know is not a direct probability, but something different:

- which model among several possible models is the best ?
- what are the best parameter values in the context of a model ?
- can we rule out certain ranges of parameters ?

Here, inductive reasoning is needed. There are different approaches employed (Bayesian reasoning, frequentist reasoning).

Types of Probability

A different (more general) kind of probability – ‘degree of belief’

e.g., actual ‘probability’ for heads is not known, and can only be determined with an ∞ number of flips: i.e., it can never be completely known. So (Bayesians) only talk about degree-of-belief for possible values. Frequentists talk about a true but unknowable probability.

Proposition Knowledge=justified belief

The purpose of experiment and analysis is to give you justification for your beliefs.

Statistical and Systematic Uncertainties

Most measurements in particle, nuclear, and astrophysics are counting experiments, where the number of counts for a particular results (e.g., heads in a coin flip) can fluctuate randomly. We call the resulting uncertainty in a result the statistical uncertainty. The model used to calculate the uncertainty depends on the physical situation. The most common distributions which result are the Binomial Distribution, the Poisson Distribution of the Gauss Distribution ➔ this lecture.

Other types of uncertainties, e.g., from miscalibrations, are called systematic uncertainties. Systematic uncertainties are 'degrees-of-belief', and require experience and hard-work to evaluate.

Binomial Distribution

Bernoulli Process: random process with exactly two possible outcomes which occur with fixed probabilities (e.g., flip coin, heads or tails, particle recorded/not recorded, …). Probabilities from symmetry argument or other information.

Definitions:

p is the probability of a 'success' (heads, detection of particle, …) $0 \le p \le 1$

N independent trials (flip of the coin, number of particles crossing detector, …)

r is the number of successes (heads, observed particles, …) $0 \le r \le N$

Then

$$f(r;N,p) = \underbrace{\frac{N!}{r!(N-r)!}}_{\text{Number of combinations - Binomial coefficient}} \overbrace{p^r q^{N-r}}^{\text{Probability of r successes in N trials}} \quad \text{where } q = 1 - p$$

Derivation: Binomial Coefficient

Ways to order N distinct objects is N!=N(N-1)(N-2)…1
N choices for first position, then (N-1) for second, then (N-2) …

Now suppose we don't have N distinct objects, but have subsets of identical objects. E.g., in flipping a coin, two subsets (tails and heads). Within a subset, the objects are indistinguishable. For the i^{th} subset, the $n_i!$ combinations are all equivalent. The number of distinct combinations is then

$$\frac{N!}{n_1!n_2!\cdots n_n!} \quad \text{where} \quad \sum_i n_i = N$$

For the binomial case, there are two subclasses (Success&failure, heads or tails, …) The combinatorial coefficient is therefore

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

Some Notation

We impose the normalization condition for probabilities and probability densities:

$$\sum_i f(x_i;\theta) = 1$$ For discrete x

$$\int_{-\infty}^{\infty} \mathrm{f}(\mathrm{x};\theta)\,\mathrm{dx} = 1$$ For continuous x

In the case that x and/or θ have multiple components:

$$\sum_j \sum_i f(x_{j,i};\vec{\theta}) = 1$$

$$\int_{-\infty}^{\infty} \mathrm{f}(\vec{\mathrm{x}};\vec{\theta})\,\mathrm{d}\vec{\mathrm{x}} = 1$$ Multidimensional integral

Notation-cont.

Cumulative distribution function: $$F(a) = \sum_{i=-\infty}^{a} f(x_i;\theta)$$

$$F(a) = \int_{-\infty}^{a} f(x;\theta)\,dx$$

$$0 \le F(a) \le 1$$

$P(a \le x \le b) = F(b) - F(a)$ Equality may not be possible for discrete case

Expectation value: $$E[x] = \sum_{i=-\infty}^{\infty} x_i f(x_i;\theta)$$

$$E[x] = \int_{-\infty}^{\infty} x\, f(x;\theta)\, dx$$

$$E[u(x)] = \int_{-\infty}^{\infty} u(x)\, f(x;\theta)\, dx$$

For u(x), v(x) any two functions of x, E[u+v]=E[u]+E[v]. For c,k any constants, E[cu+k]=cE[u]+k.

Notation-cont.

The nth moment variable is given by: $\alpha_n \equiv E[x^n] = \int_{-\infty}^{\infty} x^n f(x;\theta)\, dx$

For discrete probabilities, integrals $\Rightarrow$ sums in obvious way

$\mu \equiv \alpha_1 = E[x]$ is known as the mean

The nth central moment of x: $m_n \equiv E[(x-\alpha_1)^n] = \int_{-\infty}^{\infty} (x-\alpha_1)^n f(x;\theta)\, dx$

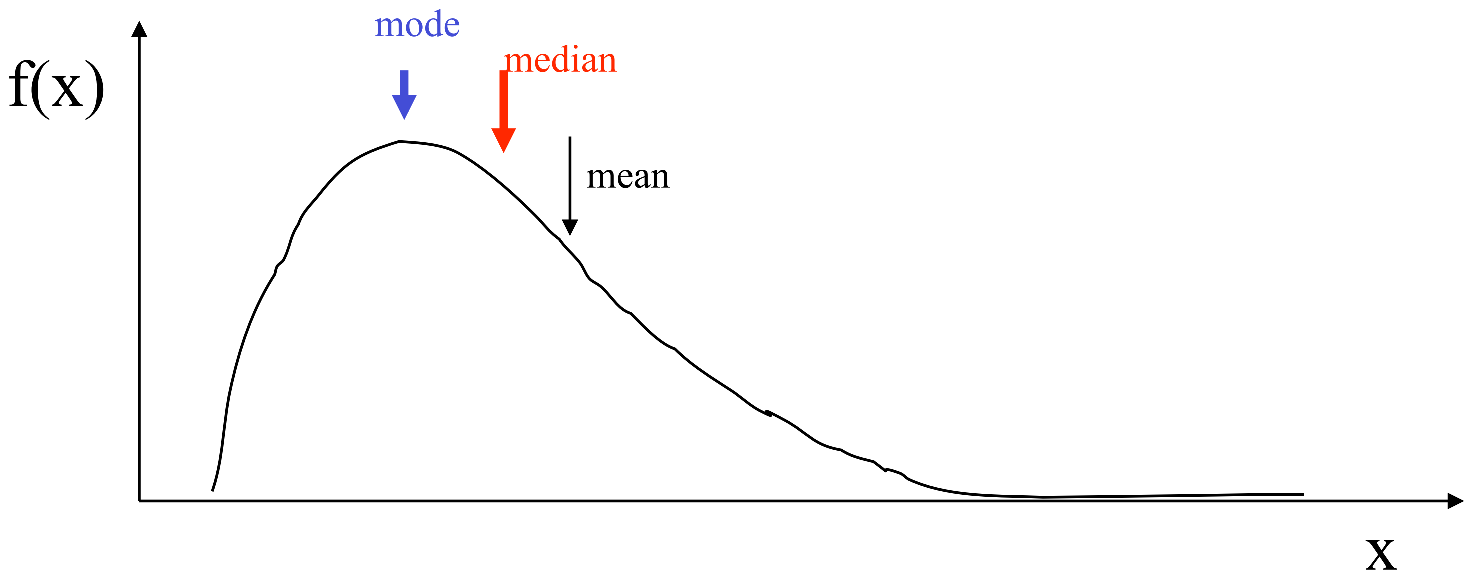
$\sigma^2 \equiv V[x] \equiv m_2 = \alpha_2 - \mu^2$ is known as the variance and σ is known as the standard deviation or 'root mean square'.

μ, σ (or σ^2) are most commonly used measures to characterize a distribution.

Notation-cont.

Other useful characteristics:

- most-probable value (mode) is value of x which maximizes $f(x;\theta)$
- median is a value of x such that $F(x_{med})=0.5$



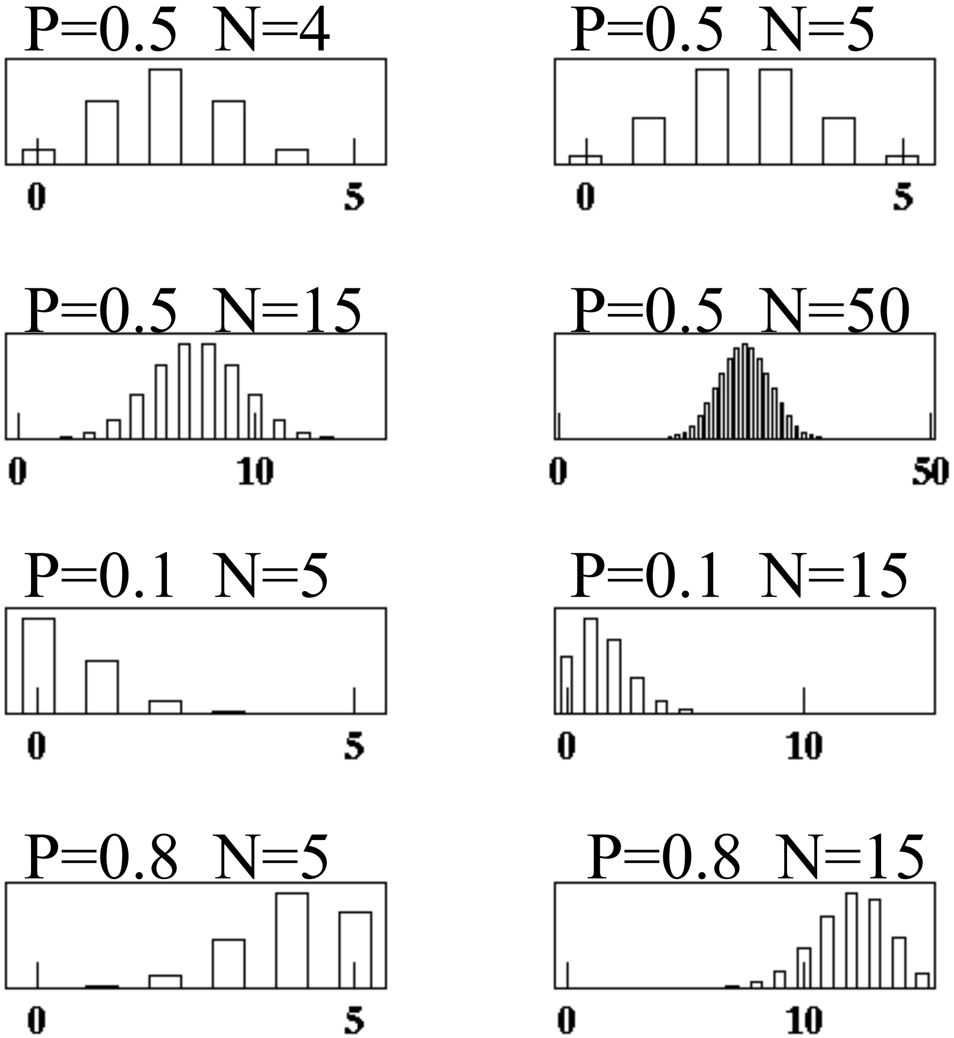
Binomial Distribution

p is the probability of a 'success' (heads, detection of particle, …) $0 \le p \le 1$
N independent trials (flip of the coin, number of particles crossing detector, …)
r is the number of successes (heads, observed particles, …) $0 \le r \le N$

Then

$$f(r;N,p) = \underbrace{\frac{N!}{r!(N-r)!}}_{\text{Number of combinations - Binomial coefficient}} \overbrace{p^r q^{N-r}}^{\text{Probability of r successes in N trials}} \quad \text{where } q = 1 - p$$

Binomial Distribution-cont.



E[r]=Np

V[r]=Np(1-p)

Notes:

- for large N, p near 0.5 distribution is approx. symmetric
- for p near 0 or 1, the variance is reduced

Example

Example:
You are designing a particle tracking system and require at least three measurements of the position of the particle along its trajectory to determine the parameters. You know that each detector element has an efficiency of 95%. How many detector elements would have to 'see' the track to have a 99% reconstruction efficiency ?

Solution:
We are happy with 3 or more hits, so our probability is

$$f(r \geq 3; N, p) = \sum_{r=3}^{N} f(r; N, p) > 0.99$$

Example-cont.

$$N=3 \quad f(3;3,0.95)=\frac{3!}{3!(3-3)!}(0.95)^3(1-0.95)^{3-3}=0.95^3=0.857$$

$$\left.\begin{aligned} N=4 \quad f(3;4,0.95)&=\frac{4!}{3!(4-3)!}(0.95)^3(1-0.95)^{4-3}=4(0.95)^3(0.05)=0.171 \\ f(4;4,0.95)&=\frac{4!}{4!(4-4)!}(0.95)^4(1-0.95)^{4-4}=(0.95)^4=0.815 \end{aligned}\right\} \quad 0.986$$

$$\left.\begin{aligned} N=5 \quad f(3;5,0.95)&=\frac{5!}{3!(5-3)!}(0.95)^3(1-0.95)^{5-3}=10(0.95)^3(0.05)^2=0.021 \\ f(4;5,0.95)&=\frac{5!}{4!(5-4)!}(0.95)^4(1-0.95)^{5-4}=5(0.95)^4(0.05)=0.204 \\ f(5;5,0.95)&=\frac{5!}{5!(5-5)!}(0.95)^5(1-0.95)^{5-5}=(0.95)^5=0.774 \end{aligned}\right\} \quad 0.999$$

With 5 detector layers, we have >99% chance of getting at least 3 hits

Law of Large Numbers

Suppose we anticipate making N measurements, what can we say about the anticipated number of successes r ? The distribution of r for many experiments will be a binomial distribution, and $f_N = r/N$ is an uncertain number.

$$< f_N > \equiv \frac{1}{N} < r > = \frac{Np}{N} = p$$

The standard deviation for f_N is

$$\sigma(f_N) \equiv \frac{1}{N}\sigma(r) = \frac{\sqrt{Np(1-p)}}{N} = \frac{\sqrt{p(1-p)}}{\sqrt{N}}$$

This 'Law of Large Numbers' states that as N becomes very large it will be very improbable to measure values of f_N very different from p.

Poisson Distribution

A Poisson distribution applies when we do not know the number of trials (it is a large number), but we know that there is a fixed probability of 'success' per trial, and the trials occur independently of each other.

High energy physics example: beams collide at a high frequency (10 MHz, say), and the chance of a 'good event' is very small. The resulting number of events in a fixed time will follow a Poisson distribution. A single trial is one crossing of the beams.

Poisson Distribution

A Poisson distribution applies in situations where the process has a constant rate, and we ask about the number of occurrences in a fixed time interval.

Nuclear physics example: radioactive decays of a large sample of an unstable isotope, where the lifetime is very long compared to the observation time; i.e., the decay rate is constant. The number of decays observed in some time period, T, follows a Poisson distribution.

Poisson Distribution-cont.

The Poisson distribution can be derived from the Binomial distribution in the limit when N $\to\infty$ and p $\to 0$, but Np fixed and finite. For accelerator example, N is the number of beam crossings observed, and p is the probability of an event occurring. The 'expected number' of events is ν=Np and the observed number is $r = n$. Then

$$f(r;N,p) \to f(n;\nu)$$

Note that ν will depend on the observation time (number of trials).

$$f(n;\frac{\nu}{N},N) = \frac{N!}{n!(N-n)!}\frac{\nu^n}{N^n}\left(1-\frac{\nu}{N}\right)^{N-n}$$

For N$\to\infty$

$$\frac{N!}{(N-n)!} = N(N-1)\cdots(N-n+1) \to N^n$$

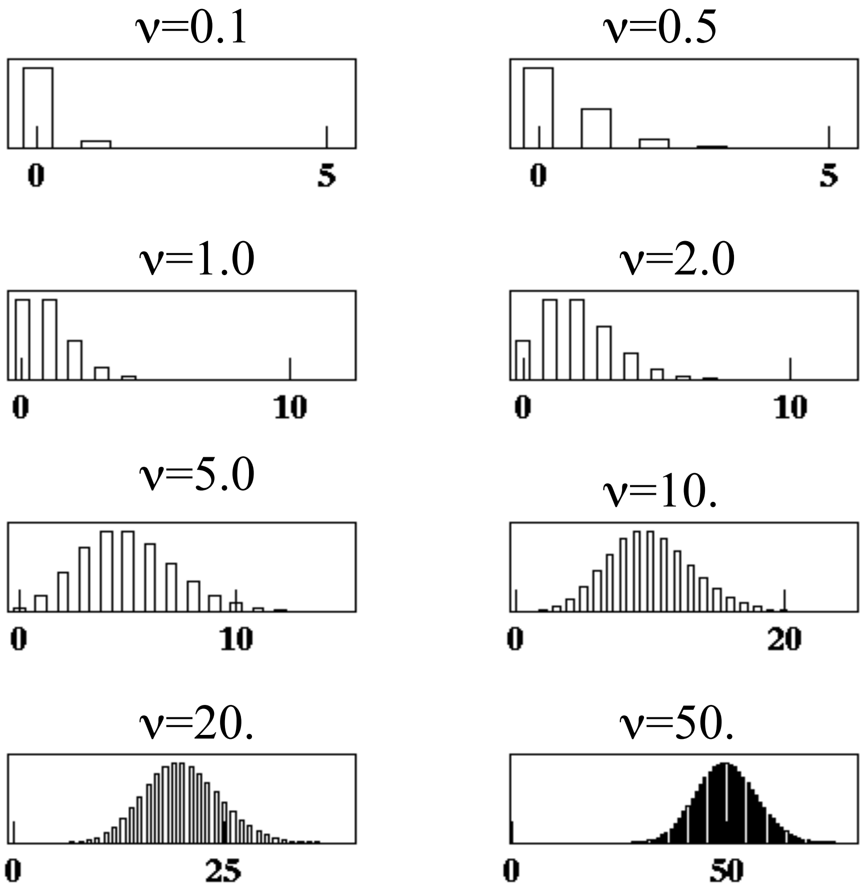
$$\left(1-\frac{\nu}{N}\right)^{N-n} \to \left(1-\frac{\nu}{N}\right)^{N} \to e^{-\nu}$$

Poisson Distribution-cont.

So,

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

$E[n]=\nu$ by definition
$\sigma^2=\nu$ variance=mean
most important property



Notes:

- As ν increases, the distribution becomes more symmetric
- Approximately Gaussian for ν>20
- Poisson formula is much easier to use that the Binomial formula.

Poisson Distribution-cont.

Proof of Normalization, mean, variance:

$$Normalization: \sum_{n=0}^{\infty} \frac{\nu^n e^{-\nu}}{n!} = e^{-\nu} \sum_{n=0}^{\infty} \frac{\nu^n}{n!} = e^{-\nu} e^{\nu} = 1$$

$$E[n] = \sum_{n=0}^{\infty} n \frac{\nu^n e^{-\nu}}{n!} = e^{-\nu} \sum_{n=1}^{\infty} \nu \frac{\nu^{n-1}}{(n-1)!} = \nu e^{-\nu} e^{\nu} = \nu$$

$$V[n] = E[n^2] - E[n]^2$$

$$E[n^2] = \sum_{n=0}^{\infty} n^2 \frac{\nu^n e^{-\nu}}{n!} = e^{-\nu} \sum_{n=1}^{\infty} \nu n \frac{\nu^{n-1}}{(n-1)!} \qquad write\ n = (n-1+1)$$

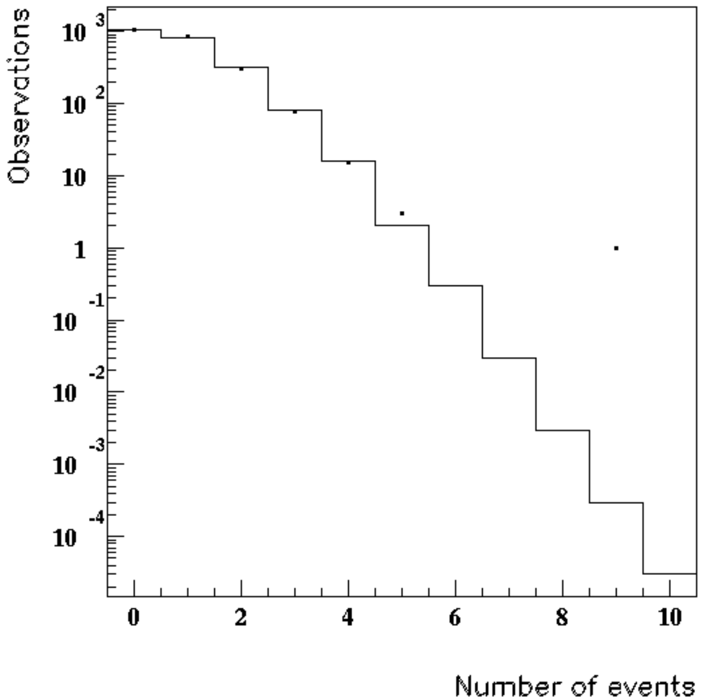
$$= \nu e^{-\nu} \left(\sum_{n=1}^{\infty} (n-1) \frac{\nu^{n-1}}{(n-1)!} + \sum_{n=1}^{\infty} \frac{\nu^{n-1}}{(n-1)!} \right) = \nu^2 + \nu$$

$$V[n] = \nu^2 + \nu - \nu^2 = \nu$$

Poisson Distribution-cont.

Example: Observation of Supernovae – IMB experiment

Number of events in 10 sec interval:	0	1	2	3	4	5	6	7	8	9
Frequency	1042	860	307	78	15	3	0	0	0	1
Poisson with mean 0.77	1064	823	318	82	16	2	0.3	0.03	0.003	0.0003



Note: a 10 sec interval contains a very large number of trials each with a very small success rate. It (the 10 sec interval) is not one trial !

Poisson Distribution-cont.

How much confidence do we have that this observation represents some new physics?

A standard quantity to calculate is Prob($r \geq n;\nu$). In this example,

Prob($r \geq 9;0.77$)

$$\sum_{n=9}^{\infty} \frac{e^{-0.77}\, 0.77^{n}}{n!} = 1.3 \cdot 10^{-7}$$

Is this the probability for observing this effect ?
More on this kind of issue later !

Poisson Distribution-cont.

We often have to deal with a superposition of two Poisson processes – the signal distribution and the background distribution, which are indistinguishable in the experiment (for example, the signal for large extra dimensions may be the observation of events where momentum balance is (apparently) strongly violated. However this can be mimicked by neutrinos, energy leakage from the detector, etc.) Usually we know the background expectations and want to know the likelihood of a signal in addition.

Poisson Distribution-cont.

Use the subscripts B for background, s for signal,

and assume n events are observed

$$P(n) = \sum_{n_s=0}^{n} f(n_s;\nu_s) f(n-n_s;\nu_B)$$

$$= e^{-(\nu_B+\nu_s)} \sum_{n_s=0}^{n} \frac{\nu_s^{n_s} \nu_B^{n-n_s}}{n_s!(n-n_s)!}$$

Binomial formula with $p = \left(\frac{\nu_s}{\nu_s+\nu_B}\right)$

$$= e^{-(\nu_B+\nu_s)} \frac{(\nu_s+\nu_B)^n}{n!} \underbrace{\sum_{n_s=0}^{n} \frac{n!}{n_s!(n-n_s)!} \left(\frac{\nu_s}{\nu_s+\nu_B}\right)^{n_s} \left(\frac{\nu_B}{\nu_s+\nu_B}\right)^{n-n_s}}_{=1 \text{ by normalization}}$$

$$= e^{-(\nu_B+\nu_s)} \frac{(\nu_s+\nu_B)^n}{n!}$$

=1 by normalization

i.e., we get another Poisson distribution for the combined expectation.

Poisson Distribution-alternate derivation

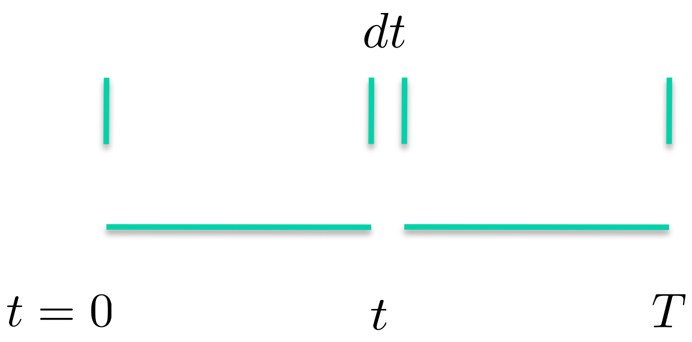
We assume a process with a constant rate, R. What is the probability that no event has occurred up to time t ?

Divide t into many very small intervals Δt, t=nΔt. Then

$$\begin{aligned} P(\text{no event}) &= (1 - R\Delta t)^n \\ &= (1 - Rt/n)^n \\ &= e^{-Rt} \text{ for } n \to \infty \end{aligned}$$

Poisson Distribution-alternate derivation

Now what is the probability to get one event in interval T ?

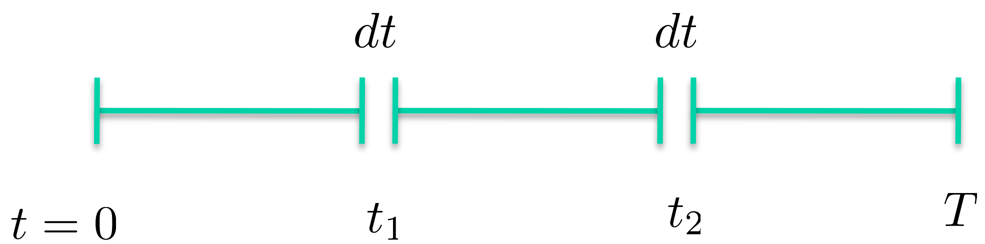


$$\int_0^T e^{-Rt} e^{-R(T-t)} (R dt) = RT e^{-RT}$$

Or, using $\nu = RT$ $\qquad P(1;\nu) = \nu e^{-\nu}$

Poisson Distribution-alternate derivation

Now what is the probability to get two events in interval T ?



$$P(2; R, T) = e^{-RT} \int_0^T R dt_1 \int_{t_1}^T R dt_2$$

$$P(2; R, T) = R^2 e^{-RT} \int_0^T (T - t_1) dt_1 = \frac{R^2 T^2 e^{-RT}}{2}$$

or $\quad P(2; \nu) = \dfrac{\nu^2 e^{-\nu}}{2} \quad$ and $\quad P(n; \nu) = \dfrac{\nu^n e^{-\nu}}{n!}$

Gaussian Distribution

The Gaussian distribution is the most widely known distribution, and the most widely used.

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

The mean is μ and the variance is σ^2.

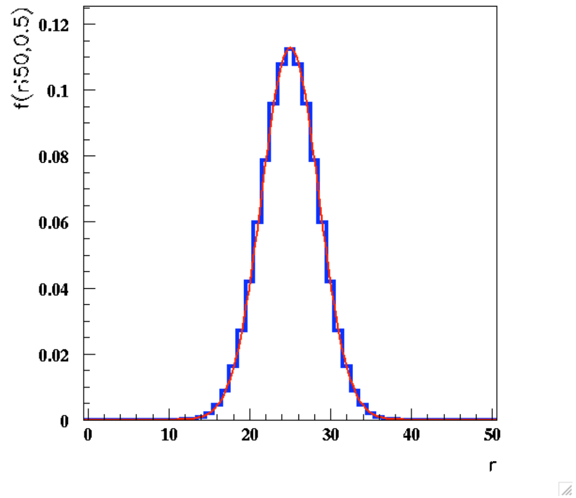
All Gaussians are similar in shape and symmetric, as opposed to the Binomial or Poisson distribution, and easily characterized. E.g.,
68.3% of the probability lies within 1 standard deviation of the mean
95.45% within 2 standard deviations
99.7% within 3 standard deviations
FWHM = 2.35σ

Gaussian Distribution

We consider two derivations of the Gauss function. First, the derivation starting from the binomial distribution. The appropriate limit in this case is N→∞ and r →∞ and p not too small and not too big. We have already seen that this leads to a symmetric distribution.

Binomial N=50, p=0.5

Gaussian μ=25,σ²=Np(1-p)



We will need Stirling's approximation

$$\ln n! \approx \ln\sqrt{2\pi n} + n\ln n - n$$

or $$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

We now substitute in the Binomial formula

Gaussian - derivation

$$f(r;N,p) = \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r} \approx \frac{\sqrt{2\pi N}(N/e)^N}{\sqrt{2\pi r}(r/e)^r \sqrt{2\pi(N-r)}((N-r)/e)^{N-r}} p^r (1-p)^{N-r}$$

$$= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{N}{r(N-r)}} \frac{N^N}{r^r (N-r)^{N-r}} p^r (1-p)^{N-r}$$

$$= \frac{1}{\sqrt{2\pi \mathrm{N}}} \frac{N^{N+1}}{r^{r+1/2}(N-r)^{N-r+1/2}} p^r (1-p)^{N-r}$$

or

$$f(r;N,p) \approx \frac{1}{\sqrt{2\pi N}} \left(\frac{r}{N}\right)^{-r-1/2} \left(\frac{N-r}{N}\right)^{-N+r-1/2} p^r (1-p)^{N-r}$$

Doesn’t look much like the Gaussian …

Gaussian - derivation

Change variables r=Np+ξ. ξ measures the distance from the mean of the binomial, Np, and the measured quantity, r. The variance of a binomial is Np(1-p), so the typical deviation of r from Np is given by

$$\sigma = \sqrt{Np(1-p)}$$

Terms of the form ξ/r will therefore be of order 1/√N and will be small.
Furthermore, $\ln(1+\xi/N) \approx \xi/N - 1/2(\xi/N)^2$

First the rewrite in terms of ξ

$$\left(\frac{r}{N}\right)^{-r-1/2} = \left(p+\xi/N\right)^{-r-1/2} = p^{-r-1/2}(1+\xi/N)^{-r-1/2}$$

$$\left(\frac{N-r}{N}\right)^{-r-1/2} = (1-p)^{-N+r-1/2}\left(1-\xi/N(1-p)\right)^{-N+r-1/2}$$

Gaussian - derivation

so
$$f(r;N,p) \approx \frac{1}{\sqrt{2\pi N}}\left(\frac{r}{N}\right)^{-r-1/2}\left(\frac{N-r}{N}\right)^{-N+r-1/2} p^r(1-p)^{N-r}$$
$$= \frac{1}{\sqrt{2\pi N}\,p(1-p)}\left(1+\frac{\xi}{Np}\right)^{-r-1/2}\left(1-\frac{\xi}{N(1-p)}\right)^{-N+r-1/2}$$

Rewrite in exponential form and use approximations from last page

$$f(r;N,p) \approx \frac{1}{\sqrt{2\pi Np(1-p)}}\exp\left[(-r-1/2)\ln\left(1+\frac{\xi}{Np}\right)+(-N+r-1/2)\ln\left(1-\frac{\xi}{N(1-p)}\right)\right]$$
$$= \frac{1}{\sqrt{2\pi Np(1-p)}}\exp\left[(-Np-\xi-1/2)\left(\frac{\xi}{Np}-\frac{1}{2}\left(\frac{\xi}{Np}\right)^2\right)\right.$$
$$\left.+\left[(-N(1-p)+\xi-1/2)\left(-\frac{\xi}{N(1-p)}-\frac{1}{2}\left(\frac{\xi}{N(1-p)}\right)^2\right)\right]\right]$$
$$\approx \frac{1}{\sqrt{2\pi Np(1-p)}}\exp\left[-Np\left(\frac{\xi}{Np}-\frac{1}{2}\left(\frac{\xi}{Np}\right)^2\right)-N(1-p)\left(-\frac{\xi}{N(1-p)}-\frac{1}{2}\left(\frac{\xi}{N(1-p)}\right)^2\right)\right]$$
$$= \frac{1}{\sqrt{2\pi Np(1-p)}}\exp\left[-\frac{\xi^2}{2Np(1-p)}\right] \qquad \sigma^2 = Np(1-p)$$

A different derivation

Here we follow the argument used by Gauss. Gauss wanted to solve the following problem: What is the form of the function $\varphi(x_i-\mu)$ which gives a maximum probability for μ=arithmetic mean of the observed values $\{x_i\}$.

$f(\vec{x} \mid \mu) = \varphi(x_1 - \mu)\varphi(x_2 - \mu)\cdots\varphi(x_n - \mu)$ is the probability to get $\{x_i\}$

Gauss wanted this function to peak at $\mu = \dfrac{\sum_{i=1}^{n} x_i}{n}$

$$\left.\frac{df}{d\mu}\right|_{\mu=\bar{x}} = 0 \quad \Rightarrow \quad \left.\frac{d}{d\mu}\prod_{i=1}^{n}\varphi(x_i - \mu)\right|_{\mu=\bar{x}} = 0$$

Assuming $f(\mu = \bar{x}) \neq 0$, $\displaystyle\sum_i \frac{\varphi'(x_i - \bar{x})}{\phi(x_i - \bar{x})} = 0$

Define $\psi = \dfrac{\varphi'}{\varphi}$ $\quad z_i = x_i - \bar{x}$

Then $\displaystyle\sum_i z_i = 0 \qquad \sum_i \psi(z_i) = 0$ for all possible z_i, so $\psi \propto z$

Gauss' derivation-cont.

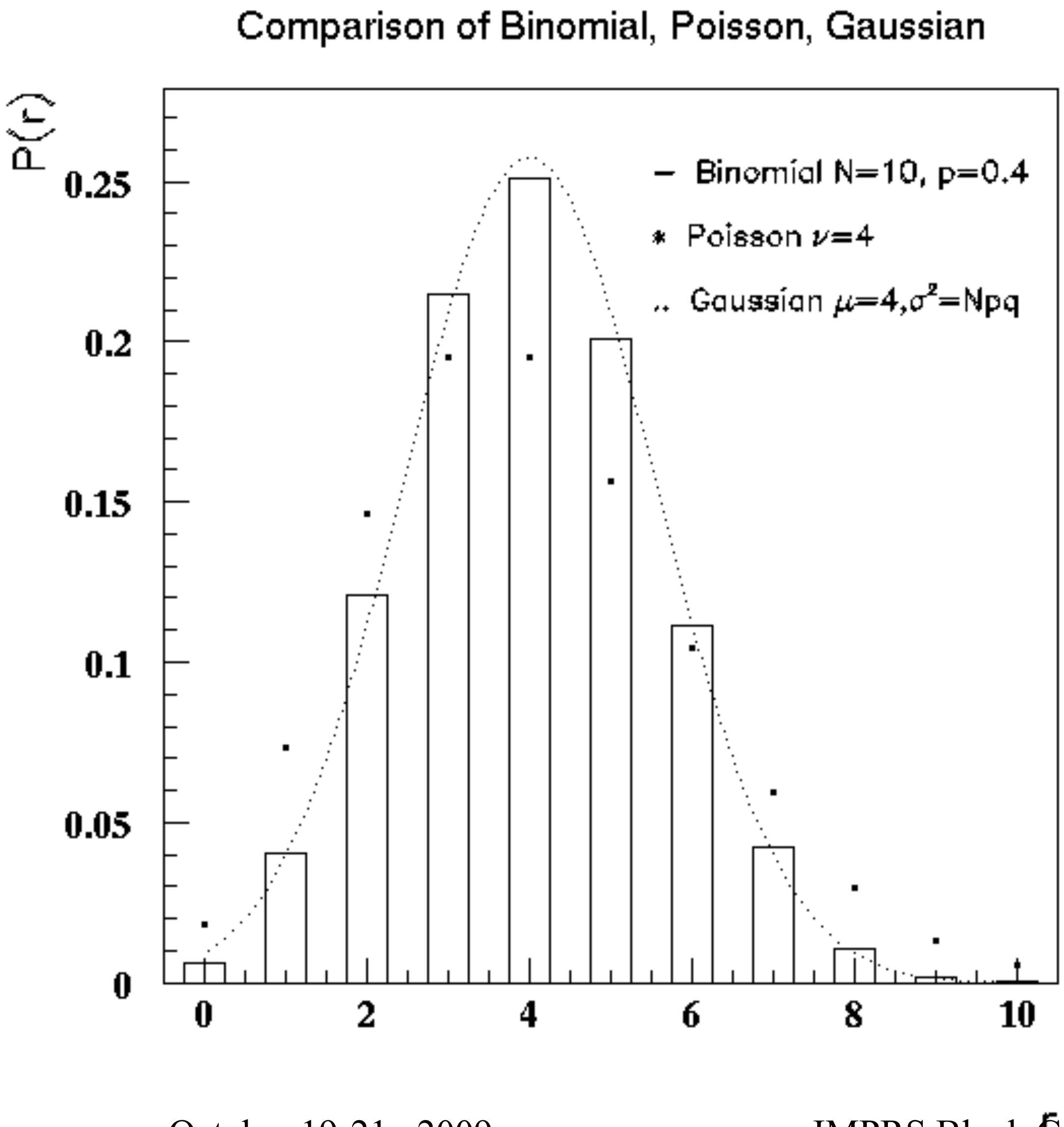
$$\psi = kz \quad \Rightarrow \quad \frac{d\varphi / dz}{\varphi} = kz, \quad \text{or} \quad \varphi(z) \propto \exp\left(kz^2 / 2 \right)$$

We get the prefactor via normalization.

Lessons:

- Binomial looks like Gaussian for large enough N,r,p
- Poisson also looks like Gaussian for large enough n
- Gauss' formula follows from general arguments (maximizing posterior probability)
- Gauss' formula is much easier to use than Binomial or Poisson, so use it when you're allowed.

Comparison Gaussian-Poisson



Four events expected

Binomial:

N	p	<r>	<(r-μ)²>	<(r- μ)³>
10	0.4	4	2.4	0.48

Poisson:

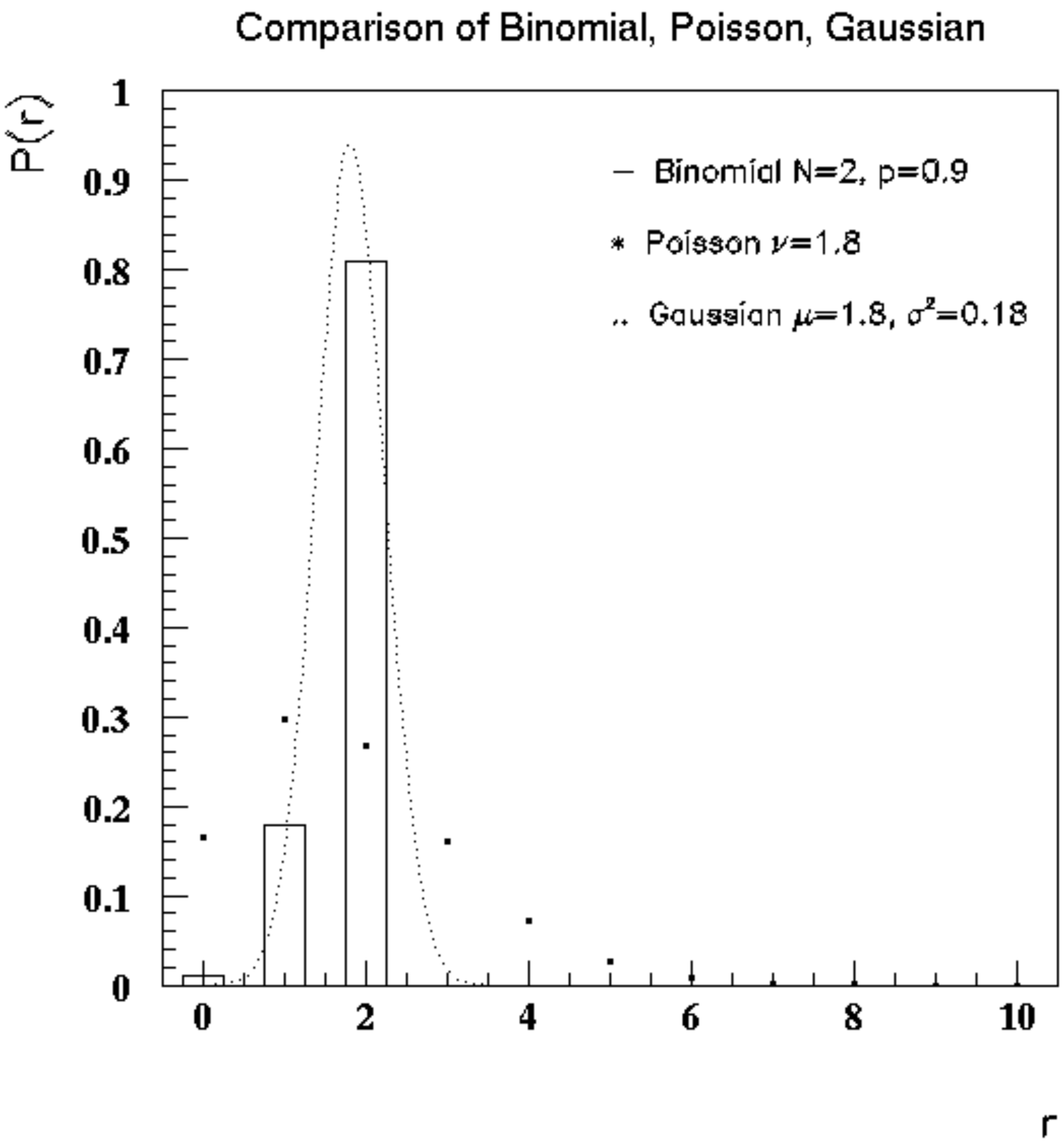
ν	<r>	<(r-μ)²>	<(r- μ)³>
4	4	4	4

Gaussian:

μ	σ²	<(r- μ)³>
4	2.4	0

- In this case, the Binomial more closely resembles a Gaussian than does the Poisson
- Note, for Binomial, can change N,p

Smaller number expected



Binomial:

N	p	<r>	$<(r-\mu)^2>$	$<(r-\mu)^3>$
2	0.9	1.8	0.18	-0.14

Poisson:

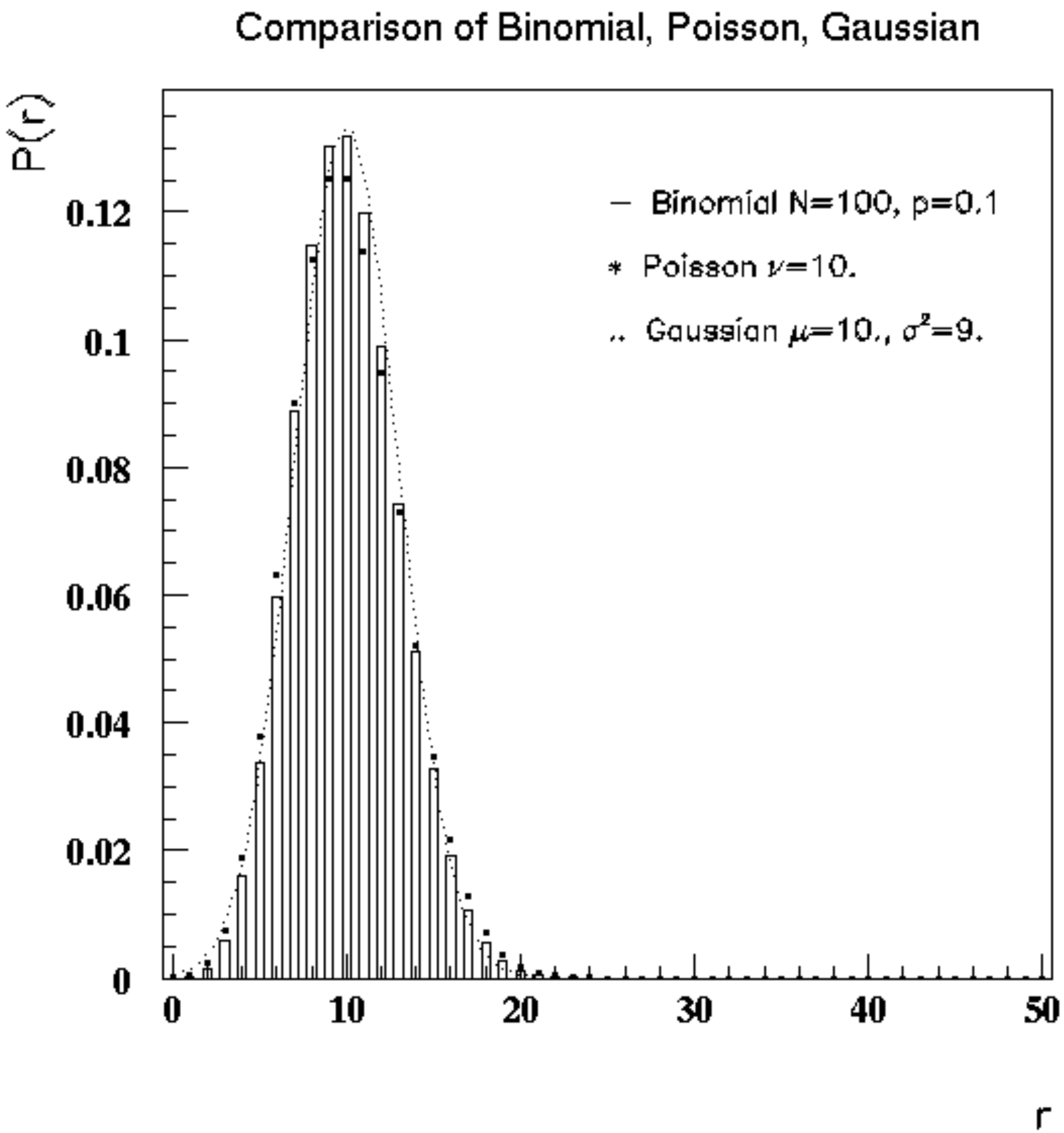
ν	<r>	$<(r-\mu)^2>$	$<(r-\mu)^3>$
1.8	1.8	1.8	1.8

Gaussian:

μ	σ^2	$<(r-\mu)^3>$
1.8	0.18	0

In general, need to use Poisson or Binomial when dealing with small statistics or p≅0,1

Larger number expected



Binomial:

N	p	$\langle r\rangle$	$\langle(r-\mu)^2\rangle$	$\langle(r-\mu)^3\rangle$
100	0.1	10	9	7.2

Poisson:

ν	$\langle r\rangle$	$\langle(r-\mu)^2\rangle$	$\langle(r-\mu)^3\rangle$
10	10	10	10

Gaussian:

μ	σ^2	$\langle(r-\mu)^3\rangle$
10	9	0

For large numbers, Gaussian excellent approximation.

Some Applications

When we don't know better, we use a Gaussian for unknown probability distributions. E.g., the degree-of-belief distribution of parameters in our systematic uncertainties. This can sometimes be justified with the Central Limit Theorem.

When reporting uncertainties on a measurement, we quote $\pm 1\sigma$ values. These are understood as Gaussian standard deviations, and therefore refer to a probability that our measurement is within the uncertainty from the true value (68.3% central probability interval).

Over-applications

From a book review of The (Mis)behavior of Markets: A Fractal View of Risk, Ruin, and Reward Benoit Mandelbrot and Richard L. Hudson. Review by Ian Kaplan:

Bachelier claimed that the change in market prices followed a Gaussian distribution. This distribution describes many natural features, like height, weight and intelligence among people. The Gaussian distribution is one of the foundations of modern statistics. If economic features followed a Gaussian distribution, a range of mathematical techniques could be applied in economics.

Unfortunately, as Mandelbrot points out in The (Mis)behavior of Markets, the foundation of this new era of economics was rotten. …There are far more market bubbles and market crashes than these models suggest.

The change in market prices does not follow a Gaussian distribution in a reliable fashion. Like income distribution, market statistics frequently follow a power law. When a graph is made of market returns (e.g., profit and loss), the curve will not fall toward zero as sharply as a Gaussian curve. The distribution of market returns has "fat tails". The "fat tails" of the return curve reflect risk, where large losses and profits can be realized.

Characteristic Function

Consider the characteristic function of a Gaussian

$$\varphi(k) = \int_{-\infty}^{\infty} dx\, e^{ikx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]}$$

$$= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} dx \exp\left(-\frac{1}{2}\left[\frac{x}{\sigma}-\left(\frac{\mu}{\sigma}+ik\sigma\right)\right]^2\right)\exp\left(ik\mu-\frac{k^2\sigma^2}{2}\right)$$

$$= e^{ik\mu} e^{-\frac{k^2\sigma^2}{2}}$$ where we have used $\int_{-\infty}^{\infty} e^{-z^2/a^2} dz = a\sqrt{\pi}$

so

$$\varphi(k) = e^{ik\mu} e^{-\frac{k^2\sigma^2}{2}}$$

Adding two Gaussians

suppose $z = x + y$

$$\varphi_z(k) = \int\int e^{ikx}\ p(x)\,dx\ e^{iky} q(y)\,dy$$

characteristic function

or

$$\varphi_z(k) = \varphi_x(k)\,\varphi_y(k)$$

The characteristic function of a sum of r.v.s is the product of the individual char. fns.

Consider the characteristic function for the sum of two Gaussians:

$$\phi_z(k) = e^{ik\mu_x - k^2\sigma_x^2/2} e^{ik\mu_y - k^2\sigma_y^2/2}$$

so

$$\phi_z(k) = e^{ik(\mu_x+\mu_y) - k^2(\sigma_x^2+\sigma_y^2)/2}$$

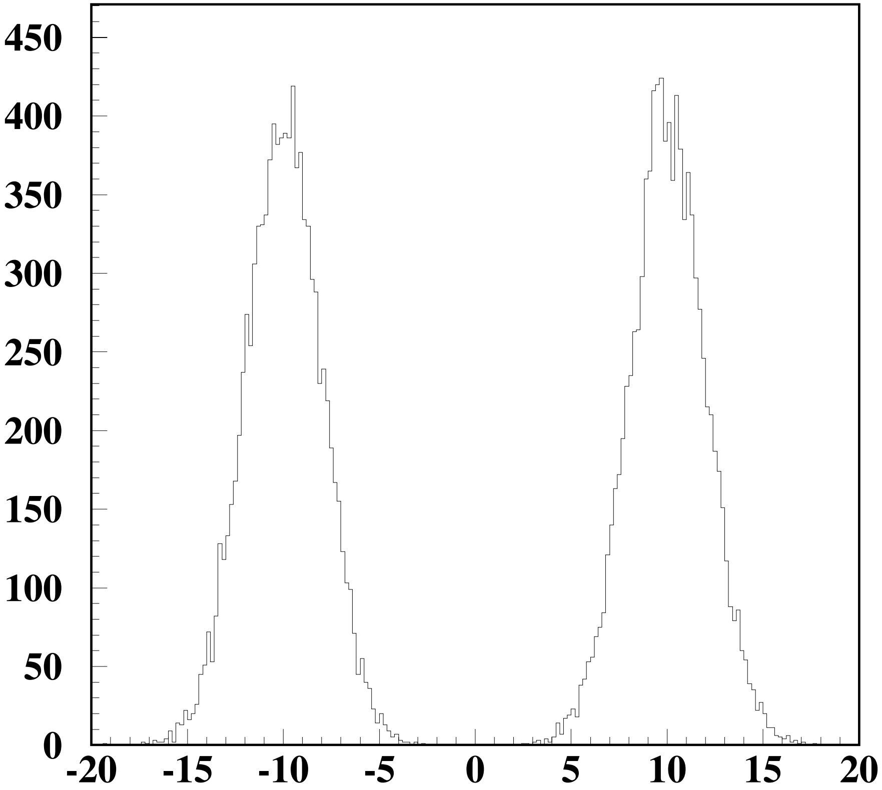
And the pdf for z is

$$f(z) = \frac{1}{\sqrt{2\pi}\sigma_z} e^{-\frac{(z-\mu_z)^2}{2\sigma_z^2}} \qquad \mu_z = \mu_x + \mu_y \quad \sigma_z^2 = \sigma_x^2 + \sigma_y^2$$

Adding two Gaussians

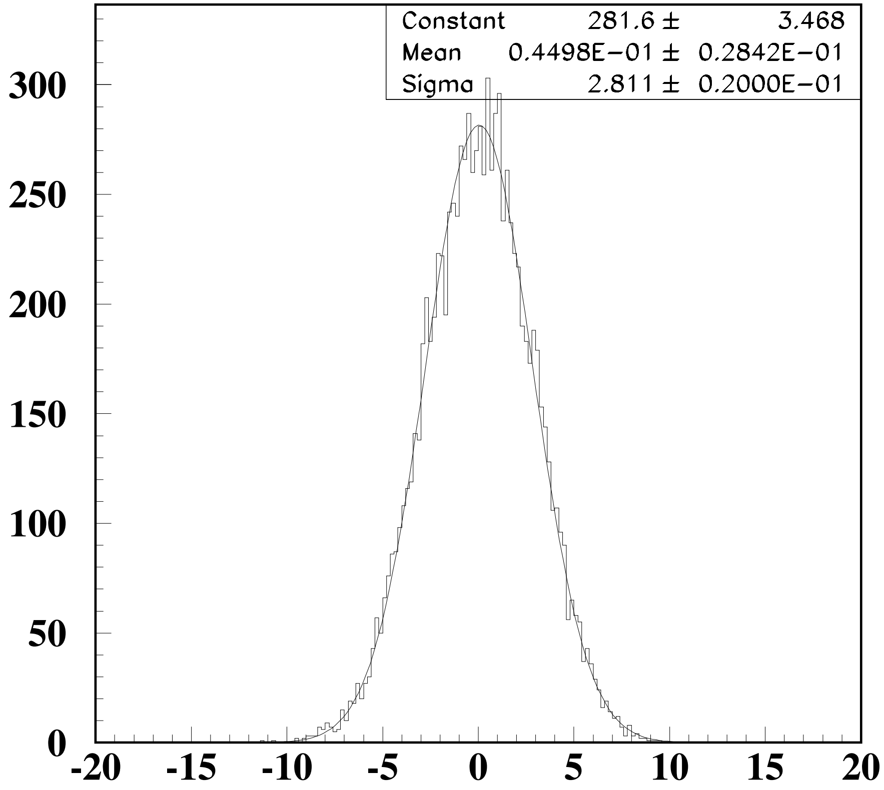
$\mu_x = -10$ $\qquad$ $\mu_y = 10$

$\sigma_x = 2.$ $\qquad$ $\sigma_y = 2.$



$\mu_z = 0$

$\sigma_z = \sqrt{2} \cdot 2$



Notation-cont.

For two random variables x,y, define *joint* p.d.f., f(x,y). The probability that x is in the range x→x+dx and **simultaneously** y is in the range y→y +dy is f(x,y)dxdy. To evaluate expectation values, etc., usually need *marginal* p.d.f. The marginal p.d.f. of x (y unobserved) is

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

The mean of x is then $\mu_x = \iint x f(x,y)\,dxdy = \int_{-\infty}^{\infty} x f_1(x)\,dx$

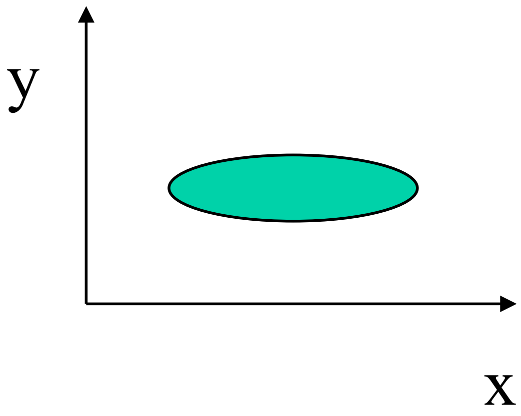
The covariance of x and y is defined as

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

And the correlation coefficient is

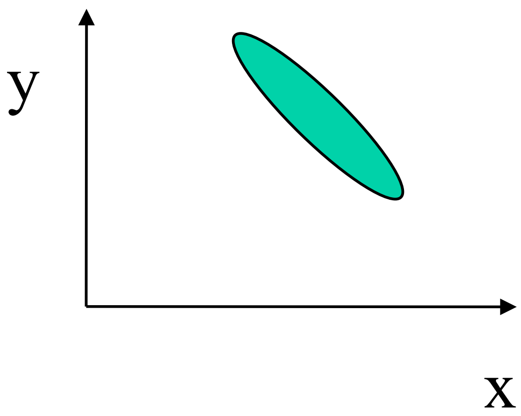
$$\rho_{xy} = \mathrm{cov}[x,y]/\sigma_x\sigma_y$$

Examples

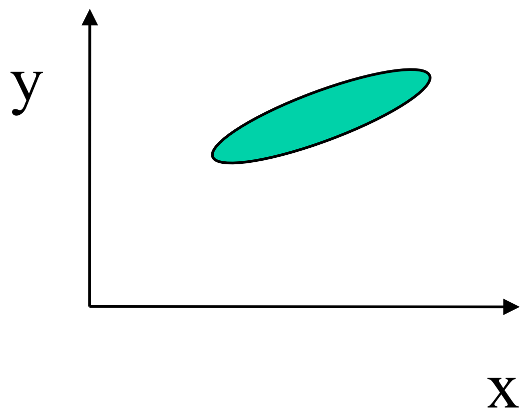


$\rho_{xy}=0$

The shading represents an equal probability density contour



$\rho_{xy}=-0.8$



$\rho_{xy}=0.2$

The correlation coefficient is limited to $-1 \leq \rho_{xy} \leq 1$

Independent Variables

Two variables are independent if and only if

$$f(x,y)=f_1(x)f_2(y)$$

Then

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

$$= \iint xy\ f(x,y)\ dxdy - \mu_x \mu_y$$

$$= \iint xy\ f_1(x) f_2(y)\ dxdy - \mu_x \mu_y$$

$$= \int x f_1(x) dx \int y f_2(y) dy - \mu_x \mu_y = 0$$

Notation-cont.

If x,y are independent, then

$$E[u(x)v(y)]=E[u(x)]E[v(y)]$$

and

$$V[x+y]=V[x]+V[y]$$

If x,y are not independent

$$V[x+y]=E[(x+y)^2]-(E[x+y])^2$$

$$=E[x^2]+E[y^2]+2E[xy] - (E[x]+E[y])^2$$

$$=V[x]+V[y]+2(E[xy]-E[x]E[y])$$

$$=V[x]+V[y]+2\mathrm{cov}[x,y]$$

Adding two Correlated Gaussians

If z=x+y and x,y are Gaussian distributed but correlated, then we again get a Gaussian for z with

$$\mu_z = \mu_x + \mu_y \quad \sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y}$$

For z=x-y, we get

$$\mu_z = \mu_x - \mu_y \quad \sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y}$$

Notation-cont.

The notation can be easily extended to any number of variables. Rather than using x,y,z, ?, we switch to a vector notation

$$\vec{x} = (x_1, x_2, \cdots, x_n)$$

The covariance between any pair of variables can be calculated. The covariance matrix is defined by $V_{ij} = \mathrm{cov}(x_i, x_j)$. It is also called the error matrix. The usual variances are on the diagonal. Similarly, a correlation matrix can be defined.

Central Limit Theorem

The characteristic function of a sum of r.v.s is the product of the individual char. fns.

We use this to prove the CLT:

Suppose we make n measurements of x.The average of the measurements is

$$a = \frac{1}{n}\left(x_1 + x_2 + \cdots + x_n\right)$$

What is the distribution of a ? It's simpler to consider the distribution of

$a - \mu$, $Q(a - \mu)$, where $\mu = \langle x \rangle$

$$\Phi(k) = \int e^{ik(a-\mu)} Q(a-\mu)\, da$$

Central Limit Theorem-cont.

$$\Phi(k) = \int e^{\frac{ik}{n}[(x_1-\mu)+\cdots+(x_n-\mu)]} p(x_1)dx_1 \cdots p(x_n)dx_n$$

$$= \left(\int e^{\frac{ik}{n}(x_1-\mu)} p(x_1)dx_1\right)\cdots\left(\int e^{\frac{ik}{n}(x_n-\mu)} p(x_n)dx_n\right)$$

$$= \left[\varphi\left(\frac{k}{n}\right)\right]^n$$ where $\varphi(k)$ is the characteristic function of $x-\mu$

$$\varphi(k) = \int e^{ik(x-\mu)} p(x)\, dx$$

$$= 1 + ik\langle x-\mu\rangle - \frac{k^2}{2}\langle (x-\mu)^2\rangle + \cdots \quad = 1 - \frac{k^2\sigma^2}{2} + \cdots$$

so

$$\Phi(k) = \left[\varphi(k/n)\right]^n = \left[1 - \frac{1}{2}\frac{k^2\sigma^2}{n^2} + \cdots\right]^n \underset{n\to\infty}{\longrightarrow} e^{-\frac{k^2\sigma^2}{2n}}$$

Central Limit Theorem-cont.

To get the pdf, we use an inverse Fourier transform

$$Q(a-\mu) = \frac{1}{2\pi}\int dk\, e^{-ik(a-\mu)}\, e^{-\frac{k^2\sigma^2}{2n}} = \frac{1}{\sqrt{2\pi}}\frac{\sqrt{n}}{\sigma}\left[\frac{1}{\sqrt{2\pi}\xi}\int dk\, e^{-ik(a-\mu)}\, e^{-\frac{k^2}{2\xi^2}}\right]$$

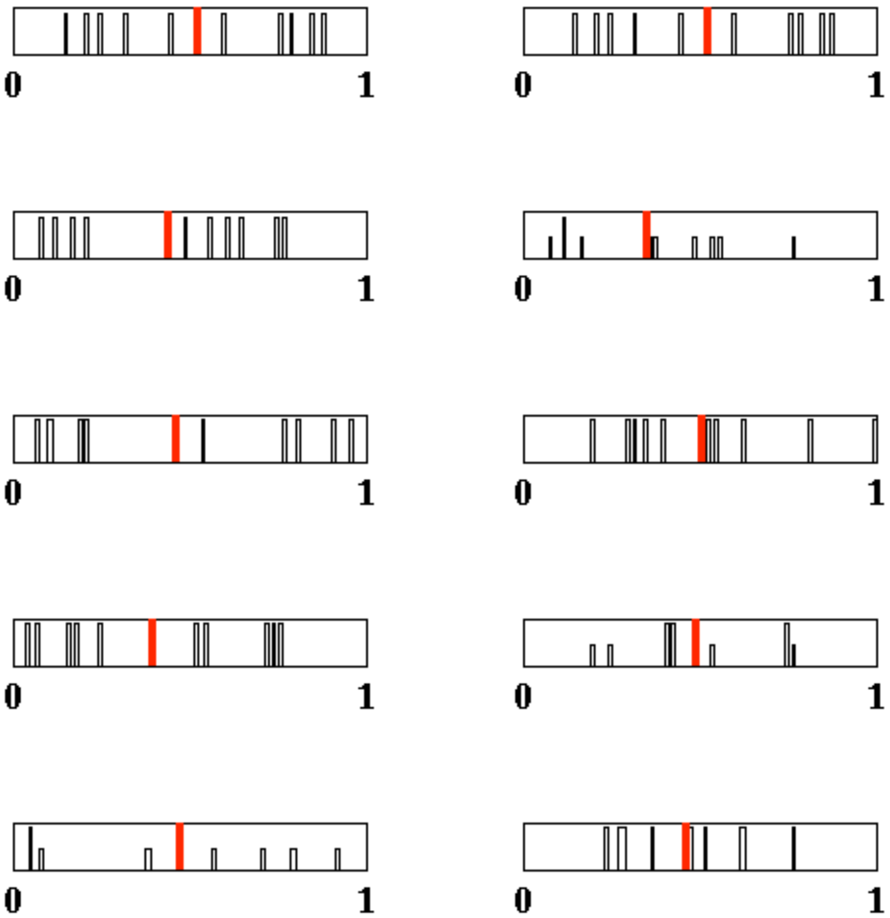
$$Q(a-\mu) = P(a) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-\frac{n}{2}\frac{(a-\mu)^2}{\sigma^2}}$$

The distribution of the **average** of a large number of measurements of a random variable x (given here by a) **follows a Gaussian distribution**. The width of the Gaussian is given by

$\xi = \frac{\sigma}{\sqrt{n}}$ where σ is the standard deviation of x

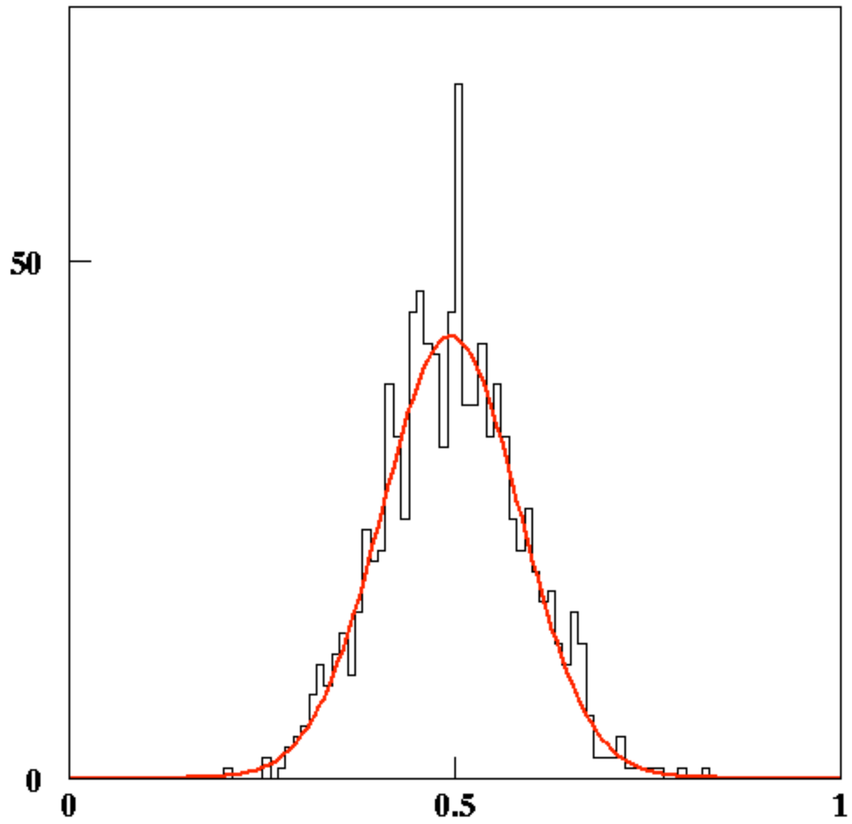
The shape of the initial distribution is unimportant !

Central Limit Theorem-Example



10 experiments where we sample 10 times randomly from a flat distribution. The data are shown as the black bars. The red bar gives the mean for the 10 samples.

Central Limit Theorem-Example



The mean value from 1000 experiments each with 10 samplings of the distribution. The red curve is a Gaussian with:

μ=0.5 and

$$\sigma = \frac{1}{\sqrt{12}}\frac{1}{\sqrt{10}}$$

Do you understand how the factors arise ?

Central Limit Theorem - conclusion

When results are presented, the uncertainties are often quoted assuming Gaussian distributions:

• For event counting, we have seen that the Binomial and Poisson reduce to the Gaussian distribution for large numbers of events (≥ 25 or so). The statistical error (1 Gaussian standard deviation) is then taken to be $\sigma=\sqrt{N}$ (from Poisson distribution).

• For other types of uncertainties (so-called systematic uncertainties or systematic errors), again a Gaussian distribution is often assumed to describe the distribution of the measured relative to the 'true'. This is usually justified with the CLT, although it is a rather indirect use. Examples of systematic uncertainties: energy calibration, alignment, time dependence, … If one dominates, then CLT not applicable.

Extras

Characteristic Function

A characteristic function is a moment generating function

$$\varphi(k) = \int dx\, e^{ikx} p(x)$$

It is simply the Fourier Transform of the p.d.f.

Expand the exponential,

$$\varphi(k) = \int dx\; p(x) \left(1 + ikx - \frac{1}{2!} k^2 x^2 - \frac{i}{3!} k^3 x^3 + \cdots \right)$$

$$= 1 + ik\langle x \rangle - \frac{k^2}{2!} \langle x^2 \rangle + \cdots + \frac{(ik)^n}{n!} \langle x^n \rangle + \cdots$$

so

$$\left. \frac{d^n \varphi(k)}{dk^n} \right|_{k=0} = i^n \langle x^n \rangle$$

Characteristic Function

Suppose x is a random variable with pdf $p_x(x)$ and

y is an independent random variable with pdf $p_y(y)$

and $z = f(x,y)$. We are interested in the probability that z lies in the interval $z \to z + dz$. Call this $p_z(z)dz$

the characteristic function of z is

$$\varphi_z(k) = \int e^{ikz} p_z(z)dz = \int\int e^{ikf(x,y)} \; p_x(x)\,dx \; p_y(y)\,dy$$

Once we have the characteristic function, we can get the pdf for z with an inverse Fourier Transform

$$p_z(z) = \frac{1}{2\pi}\int e^{-ikz}\varphi_z(k)\,dk$$