Probability and Probability Distributions

Roger Barlow
The University of Huddersfield

Terascale Statistics School, DESY, Hamburg

24^{th} February 2025

PHYSICS AT THE TERA SCALE
Helmholtz Alliance

What is Probability?

What is probability?

We all think we know, but...

From any Statistics101 Exam paper

Q1 What is meant by the probability $P(A)$ of some event A?

[1]

Please write down (or type, or whatever) your answer

4 possible answers

1. $P(A)$ is a real number between 0 and 1, obeying certain mathematical rules
2. $P(A)$ is some property of A: the larger it is, the more A happens with limits at 1 (never) and 1 (always)
3. $P(A)$ is the limit N_A/N_{Total}, as $N_{Total} \to \infty$
4. $P(A)$ expresses my belief in A, and determines the odds I will accept on a bet that A is true.

4 possible answers

1. $P(A)$ is a real number between 0 and 1, obeying certain mathematical rules
2. $P(A)$ is some property of A: the larger it is, the more A happens with limits at 1 (never) and 1 (always)
3. $P(A)$ is the limit N_A/N_{Total}, as $N_{Total} \to \infty$
4. $P(A)$ expresses my belief in A, and determines the odds I will accept on a bet that A is true.

All are possible correct answers, though only (3) and (4) are relevant

Mathematical

Kolmogorov Axioms:

A. N. Kolmogorov

For all $A \subset S$
$P_A \geq 0$
$P_S = 1$
$P_{(A \cup B)} = P_A + P_B$ if $A \cap B = \phi$ and $A, B \subset S$

From these simple axioms a complete and complicated structure can be erected. E.g. show $P_A \leq 1$....

But!!!

This says *nothing* about what P_A actually means.

Kolmogorov had frequentist probability in mind, but these axioms apply to any definition.

Classical

or Real probability

Evolved during the 18th-19th century

Developed (Pascal, Laplace and others) to serve the gambling industry.

Two sides to a coin - probability $\frac{1}{2}$ for each face

Likewise 52 cards in a pack, 6 sides to a dice...

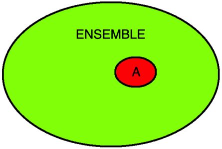
Answers questions like 'What is the probability of rolling more than 10 with 2 dice?'

Problem: can't be applied when symmetry breaks, or to continuous variables. Different answers using θ or $sin\theta$. (Bertrand's paradoxes.)

Frequentist

The usual definition taught in schools and undergrad classes

$P_A = \lim_{N\to\infty} \frac{N_A}{N}$
N is the total number of events in the ensemble (or collective)



The probability of a coin landing heads up is $\frac{1}{2}$ because if you toss a coin 1000 times, one side will come down ~ 500 times.

The lifetime of a muon is $2.2\mu s$ because if you take 1000 muons and wait $2.2\mu s$, then ~ 368 will remain.

Important

P_A is not just a property of A, but a joint property of A and the ensemble.

Problems (?) for Frequentist Probability

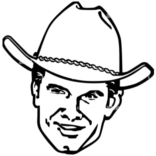
There may be several ensembles

Suppose 50% of US voters support Trump.

This varies geographically: In Texas the support is 60%.

It varies demographically: only 40% of college graduates support him

All numbers are fictitious!!



Hank is an American graduate living in Texas.

What is the probability that Hank supports Trump?

All 3 numbers are good answers

There may be no Ensemble

What is the probability that there is a SUSY particle with mass below 2 TeV?

There either is or isn't. It is either 0 or 1

CMS says $M_H = 125.35 \pm 0.15$ GeV.
What is the probability that $125.20 < M_H < 125.50$ GeV?

Either it is or it isn't. $P = 1$ or $P = 0$.

What is the probability that it will rain tomorrow?

There is only one tomorrow. It will either rain or not. P_{rain} is either 0 or 1 and we won't know which until tomorrow gets here

From Probability to Confidence

Tomorrow, here in Hamburg, it either will rain, or it won't rain.
P_{rain} is either 0 or 1.
There is only one tomorrow

Suppose the weather forecast says it will rain (actually it does).
Suppose I have checked the track-record of this forecast (actually I havn't), and they are right 90% of the time
The statement 'It will rain tomorrow' has a 90% probability of being true.
I say that it will rain tomorrow, with 90% confidence.

Definition

A is true at X% confidence if it is a member of an ensemble of statements of which at least X% are true

What is a measurement? What does $M_H = 125.35 \pm 0.15$ mean?

$M_H \in [125.20, 125.50]$ with 68% probabiity. NO
$M_H \in [125.20, 125.50]$ with 68% confidence. YES

The small print

Note that 'at least'

1. Higher confidence statements embrace lower. If something is true at 95% confidence it is true at 90% confidence
2. It lets us handle cases with discrete data where an exact match may not be possible. A Poisson distribution with a mean of 3.2 has an 89% probability of generating 5 or less events, and 96% of generating 6 or less[1] . If you require the 90% confidence limit it is 6 events.
3. It lets us handle composite hypotheses. If an upper limit rules out a branching ratio of $Br = 0.02$ at 90% confidence, it rules out $Br \geq 0.02$ at 90% confidence.

The actual fraction of cases is called the coverage

It generally depends on some parameter of the composite hypothesis - it's a function not a simple number

Overcoverage is allowed (but inefficient)

Undercoverage is not

[1]Details later in the lecture, though they're not necessary.

Bayes' theorem

Bayes' Theorem applies (and is useful) in any probability model
Conditional Probability: $P(A|B)$: probability for A, given that B is true.
Example: $P(\clubsuit A) = \frac{1}{52}$ and $P(\clubsuit A|Black) = \frac{1}{26}$

Theorem

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

Proof.

The probabiilty that A and B are both true can be written in two ways

$$P(A|B) \times P(B) = P(A\&B) = P(B|A) \times P(A)$$

Throw away middle term and divide by $P(B)$

□

Example

$P(\clubsuit A|Black) = \frac{P(Black|\clubsuit A)}{P(Black)} P(\clubsuit A) = \frac{1}{\frac{1}{2}} \times \frac{1}{52} = \frac{1}{26}$

Example

Example: In a beam which is 90% π, 10% K, kaons have 95% probability of giving no Cherenkov signal; pions have 5% probability of giving none. What is the probability that a particle that gave no signal is a K?

$P(K|no\ signal) = \frac{P(no\ signal|K)}{P(no\ signal)} \times P(K) = \frac{0.95}{0.95 \times 0.1 + 0.05 \times 0.9} \times 0.1 = 0.68$

This uses the (often handy) breakdown:

$P(B) = P(B|A) \times P(A) + P(B|\overline{A}) \times \overline{P(A)}$

Bayesian Probability

Probability expresses your belief in A.
1 represents certainty, 0 represents total disbelief

Intermediate values can be calibrated by asking whether you would prefer to bet on A, or on a white ball being drawn from an urn containing a specific mix of white and black balls.

This avoids the limitations of frequentist probability - coins, dice, kaons, rain tomorrow, existence of SUSY can all have probabilities.

The Rev Thomas Bayes (1701-1761)

Bayesian Probability and Bayes Theorem

Re-write Bayes' theorem as

$$P(Theory|Data) = \frac{P(Data|Theory)}{P(Data)} \times P(Theory)$$

$$Posterior \propto Likelihood \times Prior$$

Works sensibly

Data predicted by theory boosts belief - moderated by probability it could happen anyway

Can be chained.

Posterior from first experiment can be prior for second experiment. And so on. (Order doesn't matter)

From Prior Probability to Prior Distribution

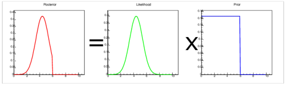
Suppose theory contains parameter a: (mass, coupling, decay rate...)

Prior probability distribution $P_0(a)$

$\int_{a_1}^{a_2} P_0(a)\,da$ is your prior belief that a lies between a_1 and a_2

$\int_{-\infty}^{\infty} P_0(a)\,da = 1$ (or: your prior belief that the theory is correct)
Generalise the number $P(data|theory)$ to the Likelihood function $L(x|a)$
Bayes' Theorem given data x the posterior is : $P_1(a) \propto L(x|a)P_0(a)$

I have measured $x = 4.5 \pm 1.0$ but I know $x < 6$.



If range of a infinite, $P_0(a)$ may be vanishingly small ('improper prior'). Not a problem. Just normalise $P_1(a)$

Shortcomings of Bayesian Probability

Subjective Probability

Your $P_0(a)$ and my $P_0(a)$ may be different. How can we compare results?

What is the right prior?

Is the wrong question.

'Principle of ignorance' - take $P(a)$ constant (uniform distribution) - is a fraud. Not constant in a^2 or $\sqrt{a}$ or $\ln a$, which are equally valid parameters.

Jefffreys' Objective Priors

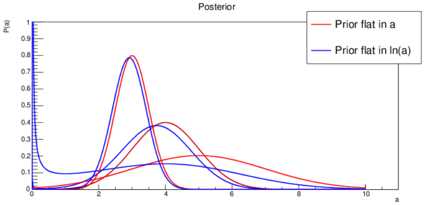
Choose a flat prior in a transformed variable a' for which the Fisher information, $-\left\langle \frac{\partial^2 L(x;a)}{\partial a^2} \right\rangle$ is flat. Not universally adopted for various reasons.

With lots of data, $P_1(a)$ decouples from $P_0(a)$. But not with little data..

Right thing to do: try several forms of prior and examine spread of results ('robustness under choice of prior')

Just an example

Measure $a = 4.0 \pm 1.0$. Likelihood is Gaussian (coming up!)



Taking a prior uniform in a gives a posterior with a mean of 4.0 and a standard deviation of 1.0 (red curve)

Taking a prior uniform in $\ln a$ (so $\propto 1/a$ for a) shifts the posterior significantly.
Also shown are 5 ± 2 (big difference) and 3 ± 0.5 (very little)

Summary

There are 4 (at least) definitions of ‘probabiity’.

The Mathematical and the Realist have severe limits. You need to use Frequentist and/or Bayesian

‘Frequentist versus Bayesian’ is the wrong language. These are not football teams!

Both have limitations.

Frequentist statistics can lead you into some quite convoluted statements

Bayesian analyses must check for robustness under choice of prior

Be ready to use them both - but always know which you are working with

Probability Distributions

Basics(1)

Data values can be: integer (discrete) or real (continuous)
(They may also be ranked or categorical but let's not go there)

Discrete values are described by *probability distributions*: P_r, pure dimensionless numbers

Real values are described by *probability density functions* or pdfs: $P(x)$
$P(x)$ has dimensions $[x]^{-1}$. $\int P(x)\,dx$ or $P(x)\,\Delta x$ are pure numbers

Parton Distribution Functions are also Probability Density Functions so no problem in calling them pdf.

You will also (sometimes) meet the Cumulative Density Function (cdf).
$C(x) = \int_{-\infty}^{x} P(x')\,dx'$

Basics(2)

Unitarity (something has got to happen)
Expressed by $\sum_r P_r = 1$ or $\int_{-\infty}^{\infty} P(x)\,dx = 1$ as appropriate

The average result, the expectation value
Expressed by $\langle r \rangle = \sum_r rP_r$ or $\langle x \rangle = \int_{-\infty}^{\infty} xP(x)\,dx$ as appropriate
Often denoted by μ

Higher Moments: $\mu_n = \langle r^n \rangle$ or $\langle x^n \rangle$
Central moments: $\mu'_n = \langle (r-\mu)^n \rangle$ or $\langle (x-\mu)^n \rangle$
Variance V is just the second central moment.. $V = \left\langle (x-\mu)^2 \right\rangle$
Notice $V = \left\langle (x-\mu)^2 \right\rangle = \left\langle x^2 \right\rangle - 2\mu \langle x \rangle + \mu^2 = \left\langle x^2 \right\rangle - \langle x \rangle^2$
V is often written as σ^2. (Physicists prefer σ, statisticians prefer V)
Skew $\gamma = \left\langle (x-\mu)^3 \right\rangle / \sigma^3$ Curtosis $K = \left\langle (x-\mu)^4 \right\rangle / \sigma^4 - 3$
Generally for any $f(x)$: $\langle f(x) \rangle = \sum_r f(r)P_r$ or $\int_{-\infty}^{\infty} f(x)P(x)\,dx$

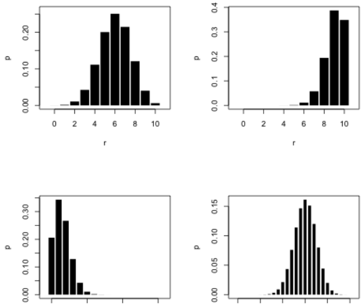
Some people use $E(f)$ rather than $\langle f \rangle$. Be prepared to meet either.

The Binomial Distribution

Binomial: Number of successes in N trials, each with probability p of success

$$P(r;p,N) = \frac{N!}{r!(N-r)!} p^r q^{N-r} \qquad (q \equiv 1-p)$$

Binomial distributions for
(1) $N = 10, p = 0.6$
(2) $N = 10, p = 0.9$
(3) $N = 15, p = 0.1$
(4) $N = 25, p = 0.6$



Mean $\mu = Np$, Variance $V = Npq$, Standard Deviation $\sigma = \sqrt{Npq}$

The Poisson Distribution

Number of events occurring at random rate λ

$$P(r;\lambda) = e^{-\lambda}\frac{\lambda^r}{r!}$$

Limit of binomial as $N \to \infty$, $p \to 0$ with $Np = \lambda = constant$:

$\frac{N^r p^r}{r!}\left(1 - \frac{\lambda}{N}\right)^N$

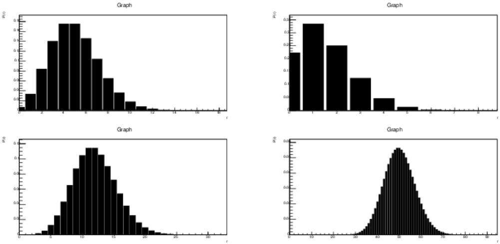
Poisson distributions for

(1) $\lambda = 5$

(2) $\lambda = 1.5$

(3) $\lambda = 12$

(4) $\lambda = 50$



Mean $\mu = \lambda$, Variance $V = \lambda$, Standard Deviation $\sigma = \sqrt{\lambda} = \sqrt{\mu}$

Meet this a lot as it applies to event counts - on their own or in histogram bins

Pop Quiz

You need to know the efficiency of your PID system for positrons

Find 1000 data events where 2 tracks have a combined mass of 3.1 GeV (J/ψ) and negative track is identified as an e^-. ('Tag-and-probe' technique)

In 900 events the e^+ is also identified. In 100 events it is not. Efficiency is 90%

What about the error?

Colleague A says $\sqrt{900} = 30$ so efficiency is $90.0 \pm 3.0\%$

Colleague B says $\sqrt{100} = 10$ so efficiency is $90.0 \pm 1.0\%$

Which is right?

Pop Quiz

You need to know the efficiency of your PID system for positrons

Find 1000 data events where 2 tracks have a combined mass of 3.1 GeV (J/ψ) and negative track is identified as an e^-. ('Tag-and-probe' technique)
In 900 events the e^+ is also identified. In 100 events it is not. Efficiency is 90%
What about the error?
Colleague A says $\sqrt{900} = 30$ so efficiency is $90.0 \pm 3.0\%$
Colleague B says $\sqrt{100} = 10$ so efficiency is $90.0 \pm 1.0\%$
Which is right?

Neither - both are wrong

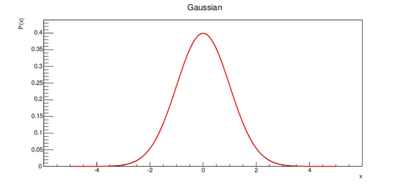
This is binomial not Poisson: $p = 0.9, N = 1000$
Error is $\sqrt{Npq} = \sqrt{1000 \times 0.9 \times 0.1}$ (or $\sqrt{1000 \times 0.1 \times 0.9}$)
$=\sqrt{90} = 9.49 \rightarrow$ Efficiency 90.0 ± 0.9 %

The Gaussian

The Formula

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

The Curve



Only 1 Gaussian curve, as μ and σ are just location and scale parameters

Properties

Mean is μ and standard deviation σ. Skew and kurtosis are 0.

The Central Limit Theorem

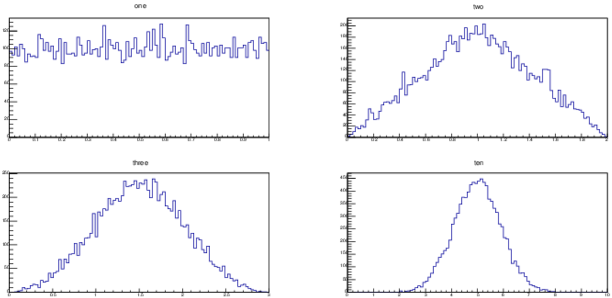
Why the Gaussian is so important

If the variable X is the sum of N independent variables $x_1, x_2 \ldots x_N$ then

1. Means add: $< X > = < x_1 > + < x_2 > + \cdots < x_N >$
2. Variances add: $V_X = V_1 + V_2 + \ldots V_N$
3. If the variables x_i are independent and identically distributed (i.i.d.) then $P(X)$ tends to a Gaussian for large N

(1) is obvious
(2) is pretty obvious, and means that standard deviations add in quadrature, and that the standard deviation of an average falls like $\frac{1}{\sqrt{N}}$
(3) applies whatever the form of the original $p(x)$

Take a uniform distribution from 0 to 1. It is flat. Add two such numbers and the distribution is triangular, between 0 and 2.



With 3 numbers, it gets curved. With 10 numbers it looks pretty Gaussian

Proof

Proof

Introduce the Characteristic Function $< e^{ikx} >= \int e^{ikx} P(x)\, dx = \tilde{P}(k)$

Proof

Introduce the Characteristic Function $< e^{ikx} >= \int e^{ikx} P(x)\, dx = \tilde{P}(k)$

Expand the exponential as a series

$$< e^{ikx} >=< 1+ikx+\frac{(ikx)^2}{2!}+\frac{(ikx)^3}{3!} \cdots >= 1+ik < x > +(ik)^2 \frac{<x^2>}{2!}+(ik^3)\frac{<x^3>}{3!} \dots$$

Proof

Introduce the Characteristic Function $< e^{ikx} >= \int e^{ikx} P(x)\, dx = \tilde{P}(k)$
Expand the exponential as a series
$< e^{ikx} >=< 1 + ikx + \frac{(ikx)^2}{2!} + \frac{(ikx)^3}{3!} \cdots >= 1 + ik < x > + (ik)^2 \frac{<x^2>}{2!} + (ik^3) \frac{<x^3>}{3!} \dots$
Take logarithm and use expansion $\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} \dots$

Proof

Introduce the Characteristic Function $< e^{ikx} >= \int e^{ikx} P(x)\, dx = \tilde{P}(k)$

Expand the exponential as a series

$< e^{ikx} >=< 1 + ikx + \frac{(ikx)^2}{2!} + \frac{(ikx)^3}{3!} \cdots >= 1 + ik < x > + (ik)^2 \frac{<x^2>}{2!} + (ik^3) \frac{<x^3>}{3!} \ldots$

Take logarithm and use expansion $\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} \ldots$

this gives power series in (ik), where coefficient $\frac{\kappa_r}{r!}$ of $(ik)^r$ is made up of expectation values of x of total power r

$\kappa_1 =< x >, \kappa_2 =< x^2 > - < x >^2, \kappa_3 =< x^3 > -3 < x^2 >< x > +2 < x >^3$

$\ldots$

Proof

Introduce the Characteristic Function $< e^{ikx} >= \int e^{ikx} P(x)\, dx = \tilde{P}(k)$
Expand the exponential as a series
$< e^{ikx} >=< 1+ikx+\frac{(ikx)^2}{2!}+\frac{(ikx)^3}{3!} \cdots >= 1+ik < x > +(ik)^2\frac{<x^2>}{2!}+(ik^3)\frac{<x^3>}{3!} \ldots$
Take logarithm and use expansion $\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} \ldots$
this gives power series in (ik), where coefficient $\frac{\kappa_r}{r!}$ of $(ik)^r$ is made up of expectation values of x of total power r
$\kappa_1 =< x >, \kappa_2 =< x^2 > - < x >^2, \kappa_3 =< x^3 > -3 < x^2 >< x > +2 < x >^3$
$\ldots$ These are called the Semi-invariant cumulants of Thièle . Under a change of scale α, $\kappa_r \to \alpha^r \kappa_r$. Under a change in location only κ_1 changes.

Proof

Introduce the Characteristic Function $< e^{ikx} >= \int e^{ikx} P(x)\, dx = \tilde{P}(k)$

Expand the exponential as a series

$< e^{ikx} >=< 1 + ikx + \frac{(ikx)^2}{2!} + \frac{(ikx)^3}{3!} \cdots >= 1 + ik < x > + (ik)^2 \frac{<x^2>}{2!} + (ik^3) \frac{<x^3>}{3!} \ldots$

Take logarithm and use expansion $\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} \ldots$

this gives power series in (ik), where coefficient $\frac{\kappa_r}{r!}$ of $(ik)^r$ is made up of expectation values of x of total power r

$\kappa_1 =< x >, \kappa_2 =< x^2 > - < x >^2, \kappa_3 =< x^3 > -3 < x^2 >< x > +2 < x >^3$

$\ldots$ These are called the Semi-invariant cumulants of Thièle . Under a change of scale α, $\kappa_r \to \alpha^r \kappa_r$. Under a change in location only κ_1 changes.

If X is the sum of i.i.d. random variables: $x_1 + x_2 + x_3 \ldots$ then $P(X)$ is the convolution of $P(x)$ with itself N times

Proof

Introduce the Characteristic Function $< e^{ikx} >= \int e^{ikx} P(x)\, dx = \tilde{P}(k)$

Expand the exponential as a series

$< e^{ikx} >=< 1+ikx+\frac{(ikx)^2}{2!}+\frac{(ikx)^3}{3!} \cdots >= 1+ik < x > +(ik)^2 \frac{<x^2>}{2!}+(ik^3)\frac{<x^3>}{3!} \dots$

Take logarithm and use expansion $\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} \dots$

this gives power series in (ik), where coefficient $\frac{\kappa_r}{r!}$ of $(ik)^r$ is made up of expectation values of x of total power r

$\kappa_1 =< x >, \kappa_2 =< x^2 > - < x >^2, \kappa_3 =< x^3 > -3 < x^2 >< x > +2 < x >^3$

$\dots$ These are called the Semi-invariant cumulants of Thièle . Under a change of scale α, $\kappa_r \to \alpha^r \kappa_r$. Under a change in location only κ_1 changes.

If X is the sum of i.i.d. random variables: $x_1 + x_2 + x_3 ...$ then $P(X)$ is the convolution of $P(x)$ with itself N times

The FT of convolution is the product of the individual FTs

Proof

Introduce the Characteristic Function $< e^{ikx} >= \int e^{ikx} P(x)\, dx = \tilde{P}(k)$
Expand the exponential as a series
$< e^{ikx} >=< 1 + ikx + \frac{(ikx)^2}{2!} + \frac{(ikx)^3}{3!} \cdots >= 1 + ik < x > + (ik)^2 \frac{<x^2>}{2!} + (ik^3) \frac{<x^3>}{3!} \ldots$
Take logarithm and use expansion $\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} \ldots$
this gives power series in (ik), where coefficient $\frac{\kappa_r}{r!}$ of $(ik)^r$ is made up of expectation values of x of total power r
$\kappa_1 =< x >, \kappa_2 =< x^2 > - < x >^2, \kappa_3 =< x^3 > -3 < x^2 >< x > +2 < x >^3$
$\ldots$ These are called the Semi-invariant cumulants of Thièle . Under a change of scale α, $\kappa_r \to \alpha^r \kappa_r$. Under a change in location only κ_1 changes.

If X is the sum of i.i.d. random variables: $x_1 + x_2 + x_3 \ldots$ then $P(X)$ is the convolution of $P(x)$ with itself N times
The FT of convolution is the product of the individual FTs
The logarithm of a product is the sum of the logarithms

Proof

Introduce the Characteristic Function $< e^{ikx} >= \int e^{ikx} P(x)\, dx = \tilde{P}(k)$
Expand the exponential as a series
$< e^{ikx} >=< 1+ikx+\frac{(ikx)^2}{2!}+\frac{(ikx)^3}{3!} \cdots >= 1+ik < x > +(ik)^2\frac{<x^2>}{2!}+(ik^3)\frac{<x^3>}{3!} \ldots$
Take logarithm and use expansion $\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} \ldots$
this gives power series in (ik), where coefficient $\frac{\kappa_r}{r!}$ of $(ik)^r$ is made up of expectation values of x of total power r
$\kappa_1 =< x >, \kappa_2 =< x^2 > - < x >^2, \kappa_3 =< x^3 > -3 < x^2 >< x > +2 < x >^3$
$\ldots$ These are called the Semi-invariant cumulants of Thièle . Under a change of scale α, $\kappa_r \to \alpha^r \kappa_r$. Under a change in location only κ_1 changes.

If X is the sum of i.i.d. random variables: $x_1 + x_2 + x_3 ...$ then $P(X)$ is the convolution of $P(x)$ with itself N times
The FT of convolution is the product of the individual FTs
The logarithm of a product is the sum of the logarithms
So $P(X)$ has cumulants $K_r = N\kappa_r$

Proof

Introduce the Characteristic Function $< e^{ikx} >= \int e^{ikx} P(x)\, dx = \tilde{P}(k)$
Expand the exponential as a series
$< e^{ikx} >=< 1 + ikx + \frac{(ikx)^2}{2!} + \frac{(ikx)^3}{3!} \cdots >= 1 + ik < x > + (ik)^2 \frac{<x^2>}{2!} + (ik^3) \frac{<x^3>}{3!} \ldots$
Take logarithm and use expansion $\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} \ldots$
this gives power series in (ik), where coefficient $\frac{\kappa_r}{r!}$ of $(ik)^r$ is made up of expectation values of x of total power r
$\kappa_1 =< x >, \kappa_2 =< x^2 > - < x >^2, \kappa_3 =< x^3 > -3 < x^2 >< x > +2 < x >^3$
$\ldots$ These are called the Semi-invariant cumulants of Thièle . Under a change of scale α, $\kappa_r \to \alpha^r \kappa_r$. Under a change in location only κ_1 changes.

If X is the sum of i.i.d. random variables: $x_1 + x_2 + x_3 ...$ then $P(X)$ is the convolution of $P(x)$ with itself N times
The FT of convolution is the product of the individual FTs
The logarithm of a product is the sum of the logarithms
So $P(X)$ has cumulants $K_r = N\kappa_r$
To make graphs commensurate, need to scale X axis by standard deviation, which grows like $\sqrt{N}$. Cumulants of scaled graph $K'_r = N^{1-r/2}\kappa_r$

Proof

Introduce the Characteristic Function $< e^{ikx} >= \int e^{ikx} P(x)\, dx = \tilde{P}(k)$
Expand the exponential as a series
$< e^{ikx} >=< 1+ikx+\frac{(ikx)^2}{2!}+\frac{(ikx)^3}{3!}\cdots >= 1+ik < x > +(ik)^2\frac{<x^2>}{2!}+(ik^3)\frac{<x^3>}{3!}\ldots$
Take logarithm and use expansion $\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3}\ldots$
this gives power series in (ik), where coefficient $\frac{\kappa_r}{r!}$ of $(ik)^r$ is made up of expectation values of x of total power r
$\kappa_1 =< x >, \kappa_2 =< x^2 > - < x >^2, \kappa_3 =< x^3 > -3 < x^2 >< x > +2 < x >^3$
$\ldots$ These are called the Semi-invariant cumulants of Thièle . Under a change of scale α, $\kappa_r \to \alpha^r \kappa_r$. Under a change in location only κ_1 changes.

If X is the sum of i.i.d. random variables: $x_1 + x_2 + x_3$... then $P(X)$ is the convolution of $P(x)$ with itself N times
The FT of convolution is the product of the individual FTs
The logarithm of a product is the sum of the logarithms
So $P(X)$ has cumulants $K_r = N\kappa_r$
To make graphs commensurate, need to scale X axis by standard deviation, which grows like $\sqrt{N}$. Cumulants of scaled graph $K'_r = N^{1-r/2}\kappa_r$
As $N \to \infty$ these vanish for $r > 2$. Leaving a quadratic.

Proof

Introduce the Characteristic Function $< e^{ikx} >= \int e^{ikx} P(x)\, dx = \tilde{P}(k)$
Expand the exponential as a series
$< e^{ikx} >=< 1 + ikx + \frac{(ikx)^2}{2!} + \frac{(ikx)^3}{3!} \cdots >= 1 + ik < x > + (ik)^2 \frac{<x^2>}{2!} + (ik^3) \frac{<x^3>}{3!} \ldots$
Take logarithm and use expansion $\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} \ldots$
this gives power series in (ik), where coefficient $\frac{\kappa_r}{r!}$ of $(ik)^r$ is made up of expectation values of x of total power r
$\kappa_1 =< x >, \kappa_2 =< x^2 > - < x >^2, \kappa_3 =< x^3 > -3 < x^2 >< x > +2 < x >^3$
$\ldots$ These are called the Semi-invariant cumulants of Thièle . Under a change of scale α, $\kappa_r \to \alpha^r \kappa_r$. Under a change in location only κ_1 changes.

If X is the sum of i.i.d. random variables: $x_1 + x_2 + x_3$... then $P(X)$ is the convolution of $P(x)$ with itself N times
The FT of convolution is the product of the individual FTs
The logarithm of a product is the sum of the logarithms
So $P(X)$ has cumulants $K_r = N\kappa_r$
To make graphs commensurate, need to scale X axis by standard deviation, which grows like $\sqrt{N}$. Cumulants of scaled graph $K'_r = N^{1-r/2}\kappa_r$
As $N \to \infty$ these vanish for $r > 2$. Leaving a quadratic.
If the log is a quadratic, the exponential is a Gaussian. So $\tilde{P}(X)$ is Gausian.

Proof

Introduce the Characteristic Function $< e^{ikx} >= \int e^{ikx} P(x)\, dx = \tilde{P}(k)$

Expand the exponential as a series

$< e^{ikx} >=< 1 + ikx + \frac{(ikx)^2}{2!} + \frac{(ikx)^3}{3!} \cdots >= 1 + ik < x > + (ik)^2 \frac{<x^2>}{2!} + (ik^3) \frac{<x^3>}{3!} \ldots$

Take logarithm and use expansion $\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} \ldots$

this gives power series in (ik), where coefficient $\frac{\kappa_r}{r!}$ of $(ik)^r$ is made up of expectation values of x of total power r

$\kappa_1 =< x >, \kappa_2 =< x^2 > - < x >^2, \kappa_3 =< x^3 > -3 < x^2 >< x > +2 < x >^3$

$\ldots$ These are called the Semi-invariant cumulants of Thièle . Under a change of scale α, $\kappa_r \to \alpha^r \kappa_r$. Under a change in location only κ_1 changes.

If X is the sum of i.i.d. random variables: $x_1 + x_2 + x_3 ...$ then $P(X)$ is the convolution of $P(x)$ with itself N times

The FT of convolution is the product of the individual FTs

The logarithm of a product is the sum of the logarithms

So $P(X)$ has cumulants $K_r = N\kappa_r$

To make graphs commensurate, need to scale X axis by standard deviation, which grows like $\sqrt{N}$. Cumulants of scaled graph $K'_r = N^{1-r/2}\kappa_r$

As $N \to \infty$ these vanish for $r > 2$. Leaving a quadratic.

If the log is a quadratic, the exponential is a Gaussian. So $\tilde{P}(X)$ is Gausian.

The FT of a Gaussian is a Gaussian. QED.

Gaussian or Normal?

Statisticians call it the 'Normal' distribution. Physicists don't. But be prepared.

Even if the distributions are not identical, the CLT tends to apply, unless one (or two) dominates.

Most 'errors' fit this, being compounded of many different sources.

From Gaussian to χ^2

Useful for Goodness-of-fit

The sum of N unit Gaussians. $\chi^2 = \sum x_i^2$

Distribution as N dimensional Gaussian, integrated over hypersphere

$P(x_1, x_2...) \propto e^{-\frac{1}{2}(x_1^2 + x_2^2...)} = e^{-\chi^2/2}$

Surface of hypershere gives factor $\chi^{N/2-1}$

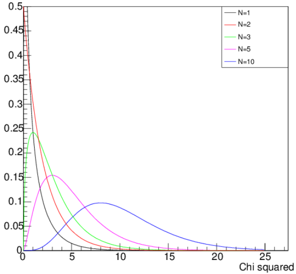
Normalised:

$p(\chi^2, N) = \frac{1}{2^{N/2}\Gamma(N/2)} \chi^{N/2-1} e^{-\chi^2/2}$

Mean N.

Rises to peak and falls again, for $N > 2$

Becomes Gaussian at very large N



Conclusions

You will use the Gaussian constantly, and the Poisson very often.
There are other distributions, but you can look up their properties: pdf, cdf, variance[2], mean[3] etc.

[2] Except for the Cauchy/Lorentz/Breit Wigner
[3] Except for the Landau