Basics 2: Distributions, Expectation Values, Moments and Hypothesis Testing

Roger Barlow
The University of Huddersfield

Terascale Statistics School, DESY, Hamburg

3^{rd} July 2023

PHYSICS AT THE TERA SCALE

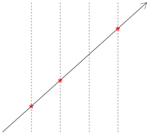
The Binomial Distribution

For N trials, each with probability p of success, the probability of r successes is

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

Proof by simple counting
Mean $\mu = Np$, standard deviation $\sigma = \sqrt{Npq}$
Example:s

- Tossing coins
- Pass/fail of components
- Hits in tracking chambers
- Particle ID



Basic, very simple, not particularly useful

The Poisson Distribution

Probability of r events occurring in some interval with a constant probability and average μ

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

Proof by taking the limit of the binomial $N \to \infty, r \to 0$ with $Nr = \mu$
Examples

- Geiger counter clicks
- Cavalrymen kicked to death by their horses
- numbers of events
- Histogram bin contents

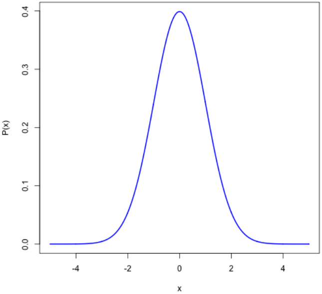
Quite common. Key fact is $\sigma = \sqrt{\mu}$

Probabilities and pdfs (Probability Density Functions)

For continuous as opposed to integer variables you need to use probability density functions: $P(x)$ rather than $P(r)$
$P(x)$ has dimensions $[x]^{-1}$. $\int P\,dx$ is a dimensionless probability

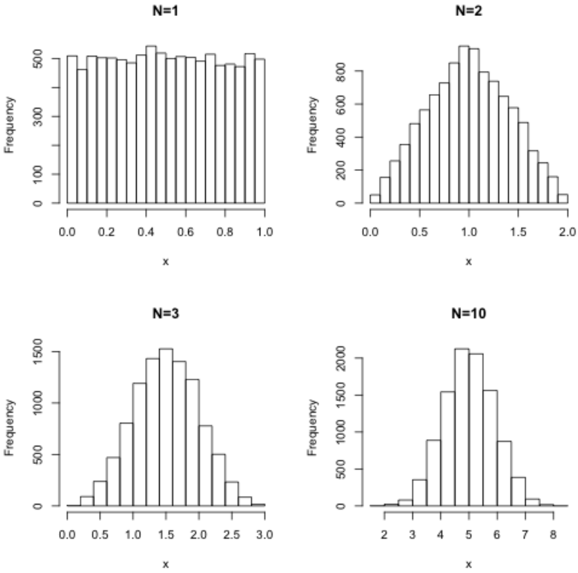
$$G(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$$

Also known as the Normal Distribution
All Gaussians related to the unit Gaussian $G(x;0,1)$ by simple shift and scale
68% probability content within $\pm\sigma$, 95% within $\pm 2\sigma$, etc



Very widespread due to the Central Limit Theorem: All distributions become Gaussian at large N

The CLT: a demonstration



Samplings from the sum of N uniform distributions

Expectation Values

Given some probability function $P(r)$ or probability distribution function $P(x)$, the Expectation value of some function $f(x)$ is the appropriate sum or integral

$$< f >= \sum_r f(r)P(r) \qquad \text{or} \qquad \int f(x)P(x)\,dx$$

Also written $E(f)$ in some texts

It's the average f you would expect after many samplings (like in quantum mechanics)

Moments

Mean $\mu =< x >$

Variance $V = \left\langle (x-\mu)^2 \right\rangle = \left\langle x^2 \right\rangle - 2 \langle x \rangle \mu + \langle x \rangle^2 = \left\langle x^2 \right\rangle - \langle x \rangle^2$
Second Central Moment
Standard deviation $\sigma = \sqrt{V}$
Note no $\sqrt{N/(N-1)}$ factor involved here.

Skew $\gamma =< (x-\mu)^3 >=< x^3 > -3 < x^2 > \mu + 2\mu^3$
Often quoted as γ/σ^3

Kurtosis. $\frac{<(x-\mu)^4>}{\sigma^4} - 3$
0 for a Gaussian
Also for several variables $cov_{xy} = \langle xy \rangle - \langle x \rangle \langle y \rangle$ and $\rho_{xy} = cov_{xy}/\sigma_x \sigma_y$

The CLT: Proof

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

Can be expanded as $1 + ik < x > + \frac{(ik)^2}{2!} < x^2 > + \frac{(ik)^3}{3!} < x^3 > \ldots$

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

This gives you a power series in ik where the coefficient K_r of each $\frac{(ik)^r}{r!}$ is made from expectation values of x with total power r

$K_1 = \langle x \rangle \,, K_2 = \left\langle x^2 \right\rangle - \langle x \rangle^2 \,, K_3 = \left\langle x^3 \right\rangle - 3 \left\langle x^2 \right\rangle \langle x \rangle + 2 \langle x \rangle^3 \ldots$

These are the semi-invariant cumulants of Thièle:

- Change in location changes only K_1
- Change in scale $x \to Ax$ gives $K_r \to A^r K_r$

CLT: if a function is convoluted with itself N times:

- Fourier transforms multiply
- Logarithms of Fourier transforms add
- $K_r \to N K_r$

Scaling this to unit standard deviation divide by $\sqrt{N K_2}$

$K_r \to N K_r / (N K_2)^{r/2} \propto N^{1-r/2}$: $K_r \to 0$ as $N \to \infty$ for $r > 2$

Log of FT is a quadratic: FT is Gaussian : Function is Gaussian. QED

Hypothesis Testing

What is it?

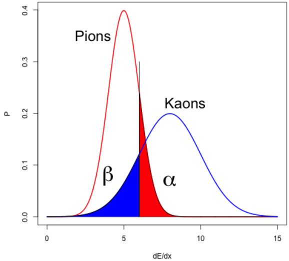
Making decisions based on statistical information

- Is this particle a pion or a kaon?
- Is this event signal or background?
- Is this patient sick or well?
- Is the accused innocent or guilty?

May be a one-off or may be one of a (large) series
Decision has to be yes or no. May be altered later if more info available
Very important part of machine learning

Basic Ideas and Notation

Suppose you want to select pions and reject kaons. The expected dE/dx measurement for pions is Gaussian with mean 5.0 and standard deviation 1.0 (in some units). For Kaons it has a mean of 8.0 and a standard deviation of 2.0.



There is a trade-off between efficiency and purity. For any cut:
α is the probability for a Type I error- wrongly rejecting a true hypothesis
β is the probability of a Type II error- wrongly accepting a false hypothesis.
Think carefully about what these probabiities mean

Where should you put the cut? You can't say. You also need to know

1. The relative numbers of pions and kaons in the data
2. The cost (or penalty) of Type I and Type II errors

The Neyman-Pearson Lemma

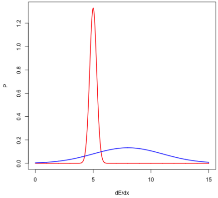
Lemma

For a given α the acceptance region which minimises β is a region where $P_0(x)/P_1(x)$ exceeds some threshold, where P_0 and P_1 are the pdfs for the desired hypothesis and the undesired alternative.

Proof.

Obvious. Given a N-P acceptance region, if some Δ at x is removed, it must be replaced by a $\Delta' = \Delta P_0(x)/P_0(x')$ for which, by hypothesis, $\Delta' P_1(x')$ is larger than $\Delta P_1(x)$. □

In a case like this you would want two cuts, to reject very low values as well as very high values. Neyman Pearson tells you how those two cuts are related: they should be at the same values of P_0/P_1. Even with complicated topologies in more than one dimension, P_0/P_1 is the only relevant quantity to cut on.



The null hypothesis H_0

To use data to support a theory, you have to show not just that the data is compatible with the theory, but that they are incompatible with the absence of the theory

- To discover the Higgs, have to show that the peak is unlikely to arise from pure background
- To show a treatment cures patients, have to show that without it they do not recover
- To establish Einstein's theory of gravity, needed data incompatible with Newton's theory

Null hypothesis H_0: there is no effect.
To build credibility for some alternative H_1 you have to try to establish H_0 and fail
Your analysis to support H_1 is, on the face of it, an analysis to support H_0
'*Every experiment is just giving the data a chance to disprove the null hypothesis*' – Ronald Fisher

Significance and power

Technical terms, better not to think about their meaning

In the language of the null hypothesis α is the probability that you will (wrongly) claim a result

α is called the *significance*. The probability under H_0 of seeing an effect this large (or larger).

Many fields publish only if significance below 5%. (1 in 20 chance that this could be a fluctuation)

Particle physics much more stringent: 0.0032% is 'evidence for' and 0.00003% is 'a discovery' (These correspond to 4 sigma and 5 sigma Gaussian deviations.)

Why so strict?

Because we've made mistakes in the past and want to avoid embarrasments in future

$1 - \beta$ is sometimes called the power of the test. Actually most of the null hypothesis procedures do not involve H_1 - except for deciding whether to use a 1-sided or 2-sided test

Good luck!

Many of you are, or will be, engaged in analyses to find new phenomena by attacking H_0 - in the form of the standard model

If you succeed, this will bring you fame and perhaps a Nobel prize

If you don't succeed, this will bring you a very solid, worthwhile and satisfactory journal publication and/or PhD thesis