Lecture 4:

Distribution Tails & Likelihood

- What is ‘Normal?’
- Robust parameter estimation
- p-values
- Combined p-values as a Statistic
- Maximum Likelihood
- Neyman-Pearson Lemma
- Fisher Information
- Wilk’s Theorem

Is That Normal?

We frequently encounter vague statements about the assumption that distributions are “sufficiently Normal,” but exactly what does that mean and how do you check that things are Normal enough?

It depends on what you're trying to do:

- For example, if you're fitting a function to a set of data, so long as the probability distributions for the data points are reasonably symmetric and tails are not very large, the derived central values for the fit parameters will generally be pretty good.

- If you want to make a precise measurement and quote Gaussian error bars, the probability distribution for the parameters should be Normal to at least ~2σ or more, as this is a tacit assumption by the reader when you quote ±1σ error bars. If this is not the case, the details should be given.

- If you want to exclude models at high confidence based on Gaussian error bars, the relevant distribution should obviously be Normal to at least that confidence level.

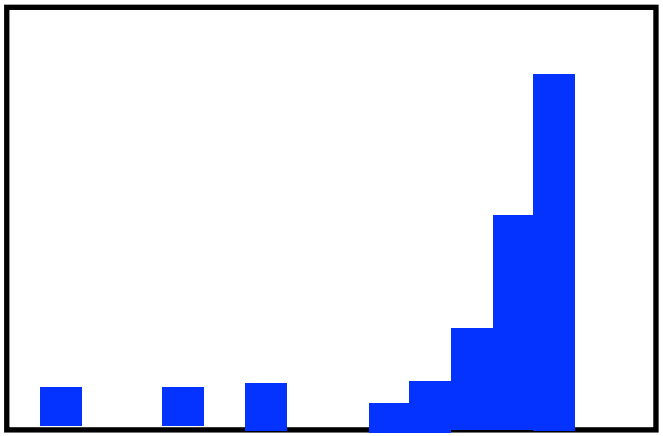
Note: the requirement on the precise Gaussian nature of individual data points may be less restrictive, since the variance of fit parameters generally arises from the accumulation of smaller deviations from the data points.

So the nature of Gaussian requirements is necessarily pragmatic, but is generally logically straight-forward.

The real issue is about:

1) Notably asymmetric distributions that can lead to systematic biases

2) Long distribution tails, whereby large deviations from the expected mean ("outliers") occur much more frequently than assumed, which can skew fits and lead to misinterpretation.



Goodness of fit parameters, such as chi-squared, can be useful indicators of issues, but these don't catch everything and won't diagnose the issue

It is always advisable to look at the distributions!

But how do you deal with very large and complex data sets, where visually inspecting every distribution is not very practical?

- If distributions are symmetric, then mean = median = peak (mode)
- The sample "skew" $\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^3}{\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2\right]^{3/2}}$ should be close to zero
- Different ways to compute the standard deviation:
  1) Perform an explicit Gaussian fit
  2) Compute the sampled RMS deviation
  3) Find the peak and then the FWHM = 2.35σ for Gaussian
  4) The central ±1σ should contain 68% of the events

Building some of these checks into your analysis is an extremely useful way to flag potential issues that warrant further investigation

Robust Parameter Estimation

The idea is to minimise the effect of distribution tails and asymmetries on the determination of derived parameters

For example, the median (or 50th percentile) is much more robust in this regard than the mean:

n odd → $x_{med} = x_{(n+1)/2}$

n even → $x_{med} = \frac{x_{n/2} + x_{n/2+1}}{2}$

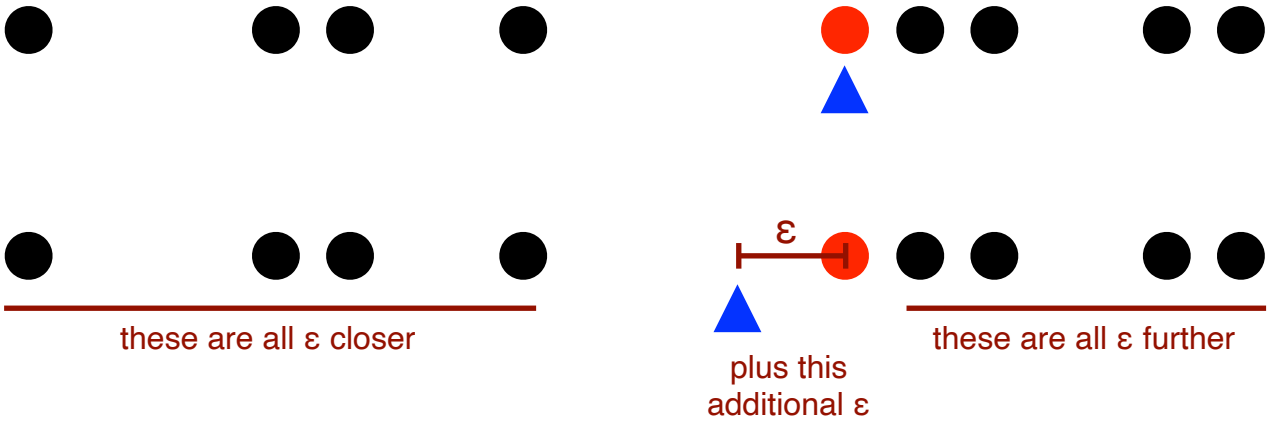
In general, distribution percentiles are robust. So, for example, one could define an equivalent distribution "width" by the 84th percentile (*i.e.* the value below which contains 84% of the distribution) minus the 16th percentile to give a region containing 68% of the distribution (roughly ±1σ for a Gaussian distribution) centred on the median.

A fit to parameters based on minimising the sum of RMS deviations provides an unbiased estimator for the mean:

$$\frac{d}{d\alpha}\sum_{i=1}^{n}(x_i-\alpha)^2=0 \quad =-2\sum_{i=1}^{n}(x_i-\alpha)=-2\left[\sum_{i=1}^{n}x_i-n\alpha\right]$$

$$\alpha=\frac{1}{n}\sum_{i=1}^{n}x_i$$

To instead provide an unbiased estimator for the median, minimise with respect to the sum of the absolute deviations:



So the minimum sum of absolute deviations finds the median!

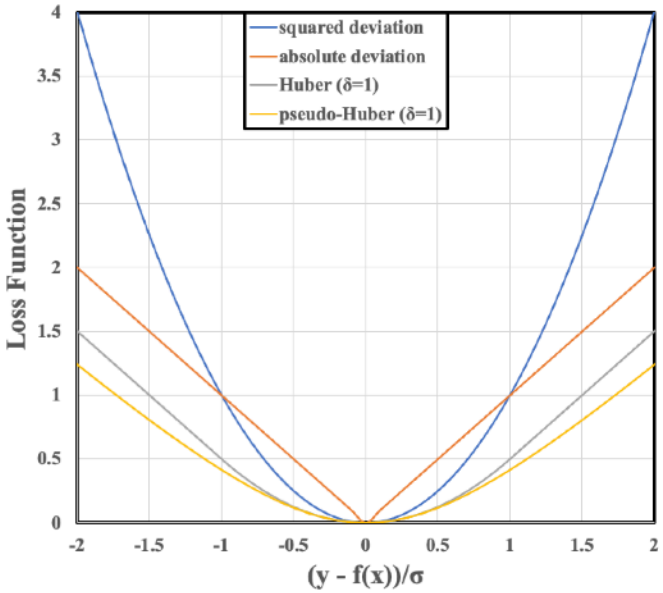
In general, the function to be minimised in order to find the best set of parameters is called the "Loss Function"

An alternative loss function suggested by Huber* provides smooth convergence in the vicinity of the minimum, while maintaining robustness from the distribution tails:

$$L_\delta = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\left(y_i - f(x_i)\right)^2 \qquad |y_i - f(x_i)| \le \delta$$

$$L_\delta = \frac{1}{n}\sum_{i=1}^{n}\delta\left(|y_i - f(x_i)| - \frac{\delta}{2}\right) \qquad |y_i - f(x_i)| > \delta$$

where δ is a tuneable parameter that would equal σ for a Gaussian distribution



A "Pseudo Huber Loss Function" provides a more convenient form that has continuous derivatives at all degrees:

$$L_\delta = \frac{1}{n}\sum_{i=1}^{n}\delta^2\left(\sqrt{1+\left(\frac{y_i - f(x_i)}{\delta}\right)^2} - 1\right)$$

*P. Huber, "Robust Estimation of a Local Parameter," Ann. Math. Statist. 35(1): 73-101 (1964)

A common quantity to compute when testing the null hypothesis:

p-value ("chance probability"):

The probability of obtaining a value of some parameter at least as extreme as that which is observed, assuming the null hypothesis is true.

"How much does this particular data set look like what is expected from the null hypothesis?"

But the p-value is **NOT** the probability of a particular hypothesis being true or false!

Example 1:

Search for Episodic X-Ray Emission

Over the course of a year, 36000 x-rays are observed to come from a particular astrophysical object. However, on one particular day, 130 events are observed. What is the statistical significance of this observed burst?

$$\langle x \rangle = \frac{36000}{365} = 98.6 \qquad \mu \simeq \langle x \rangle \qquad \sigma = \sqrt{\mu}$$

$$s \cong \frac{(130-98.6)}{\sqrt{98.6}} = 3.16\sigma$$

odds of getting at least this many events by a chance fluctuation from the average rate of emission

$$P = 8 \times 10^{-4}$$

p-value for this test, but need to look at it in the context of all other tests

Is this sufficient to claim the observation of a burst from this object?

Correct question:

What is the chance of seeing at least one burst with an excess at least as large given the number of independent tests I've done ?

Binomial !!

N Bernoulli trials where the chance of each success is **P**

$$\sum_{i=1}^{\infty} \binom{N}{i} P^i (1-P)^{N-i} = 1 - \binom{N}{0} P^0 (1-P)^{N-0}$$

$$\mathrm{P}_{\text{post-trial}} = 1 - (1-P)^N \qquad (\sim NP \quad \text{for} \quad NP << 1)$$

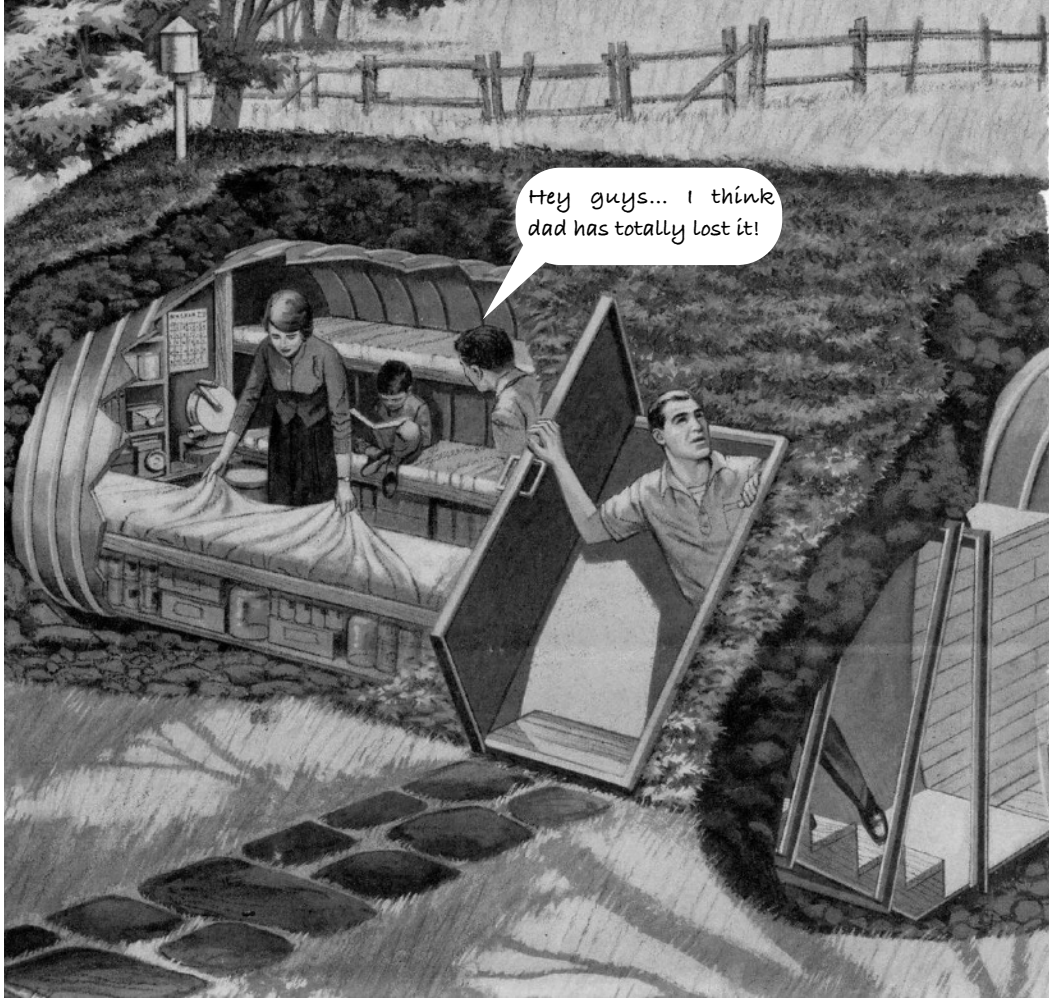
$\mathrm{P} = 8 \times 10^{-4}$, $\mathrm{N} = 365$ → $\mathrm{P}_{\text{post-trial}} = 25\%$

How many timescales were considered? How many objects examined?

Example 2:

During his year of self-isolating, Dave peered out of his bunker on six random occasions and found that it was always dark.

Assuming that the earth goes around the sun, you would expect it to be dark about half the time, averaged over the year. So the chance probability for it to be dark outside on all six occasions is:



$$P(\textit{dark all 6 times}) = (0.5)^6 = 0.0156$$

Importance of prior probabilities (more on this later)

"Gosh, That's pretty small! **Hey everyone, it looks like there's a very good chance that we're no longer going around the sun!!**"

Pragmatism!

Look carefully at context:

Very small p-values, even after careful accounting of trials, confirmed by other observations/crosschecks, which could be explained by self-consistent (plausible) alternatives…

↓

Reject H0

Combination of p-values

Two identical experiments observe evidence of the brexiton (a particle now outside of the Standard Model that inevitably then decays to a less attractive state). The first experiment assesses the odds that their observation is due to chance fluctuations as being 1%, while the second assesses their observation to have a chance probability of 10%. What is the combined chance probability that these two data sets are consistent with the null hypothesis (*i.e.* there is no brexiton)?

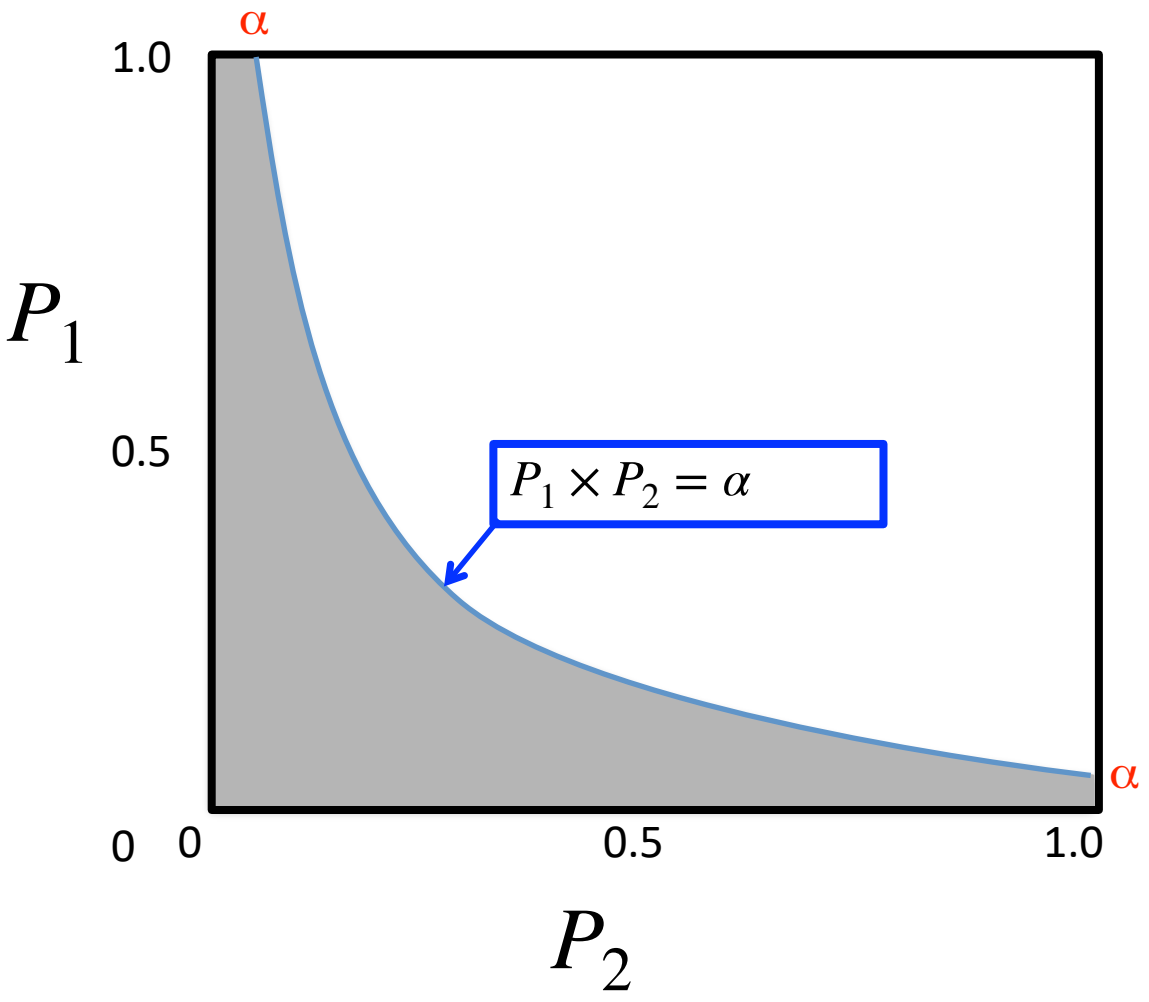
$$P_1 \times P_2 = 0.001\ ?$$

Need to look at properties of the product:

Define the statistic: $\Gamma \equiv P_1 \times P_2$

What is the chance probability for Γ

to be at least as small as some value α ?



Integrated area under the curve:

$$\alpha\,(\,1 - \ln\alpha\,)$$

$$= P(\alpha)$$

i.e. this is the chance that a background fluctuation would yield a value of Γ that is at least as small as α.

So, for the case here: $\alpha = (0.01)(0.1) = 0.001$

$$P(\leq \alpha) = 0.001(1 - \ln(0.001)) = 0.004$$

More generally…

Fisher's Method

$$F \equiv -2\ln\left(\prod_{i=1}^{n} p_i(\leq p_{obs})\right) = \sum_{i=1}^{n}\left(-2\ln p_i(\leq p_{obs})\right) \equiv \sum_{i=1}^{n} f_i$$

$$p_i(\leq p_{obs}) = e^{-\frac{f_i}{2}}$$

or $$p_i(> p_{obs}) = 1 - e^{-\frac{f_i}{2}}$$

$$p_i^{diff}(f_i) = \frac{1}{2}e^{-\frac{f_i}{2}}$$

Recall: $$P(\chi^2,2) = \frac{1}{2}e^{-\frac{\chi^2}{2}}$$

so f_i values are distributed like a χ^2 distribution with 2 DoF

and we can express:

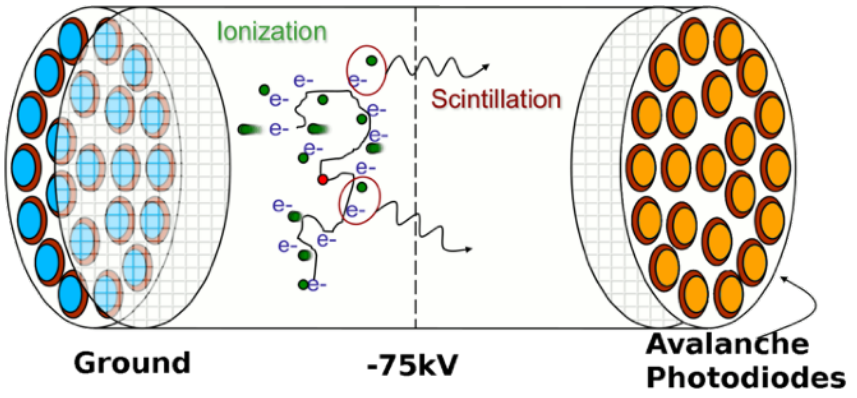
$$\chi^2 = \sum_{i=1}^{k(\equiv 2n)} g_i^2 = \sum_{i=1}^{n} \overbrace{\left(g_{2i-1}^2 + g_{2i}^2\right)}^{\chi_i^2}$$

$$P(\chi^2,2n) \equiv \sum_{i=1}^{n} P(\chi_i^2,2)$$

F is distributed like a χ^2 distribution with 2n DoF

Example:

The EXO experiment uses liquid xenon to search for evidence of neutrinoless double beta decay, which produces 2 electrons with a total energy that is well defined. The interaction produces scintillation light in the liquid xenon target, and the ionisation tracks of charged particles are also drifted to a readout plane to record the time and position of charges. Backgrounds come from radioactivity in the xenon and, to a greater extent, from the walls of the detector.



Assume that an event is observed and the chance probability for it to be background is assessed using several independent measures:

Event energy estimated from the scintillation light: $P_{scint} = 0.14$

Event energy estimated from the total charge: $P_{charge} = 0.05$

The proximity of the event to the cavity walls: $P_{charge} = 0.32$

The density of charge deposition (event topology): $P_{charge} = 0.53$

What is the overall chance probability (p-value) that this event is background?

$$-2\log(0.14 \times 0.05 \times 0.32 \times 0.53) = 13.43$$

$$P(\chi^2 > 13.43,\ DoF = 2 \times 4) = 0.10$$

Likelihood

We wish to express the likelihood for a given set of data to have resulted from a particular model of probability distributions:

conditional probability

$$L = P\left(D \,|\, H(\mathbf{q})\right)$$

likelihood — data set — assuming a particular hypothesis defined by a set of parameters $\mathbf{q}$

for independent events

$$= P(x_1|H(\mathbf{q}))P(x_2|H(\mathbf{q}))...(x_n|H(\mathbf{q})) \quad = \prod_{i=1}^{n} P(x_i|H(\mathbf{q}))$$

more practical to compute

$$\log L = \sum_{i=1}^{n} \log\left[P\left(x_i \,|\, H(\mathbf{q})\right)\right]$$

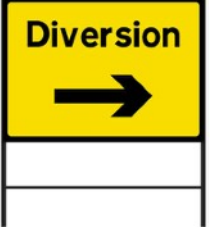
More likely data sets for $\mathrm{H}(\mathbf{q})$ will have a higher combined probability (*i.e.* likelihood)

$$\log L = \sum_{i=1}^{n} \log \left[P\left(x_i \,|\, H(\mathbf{q})\right) \right]$$

The game will then be to find the model for which the observed data set is “most likely”

Note: When used in this way, L is referred to as the “Likelihood Function” rather than a probability, because it is used to describe the *relative* probability for different models given a fixed data set… however that dependence need not be normalised to 1 over the models tested!

(the normalisation is instead defined over all possible data sets for the correct model)



Tests of Simple vs Composite Hypotheses

Simple hypothesis: All parameters of the relevant distributions are specified. (*i.e.* PDFs can be used to completely characterise the problem)

Composite hypothesis: Where this is not the case and parameters span a range of possibilities.

This is probably a university student, because they spend £20 per week on alcohol and the average student spends more than £15 per week on this.

COMPOSITE

(exact distribution not defined)

This is probably a university student, because they spend £20 per week on alcohol and the average student spends £17 per week on this with a standard deviation of ~ £13.

COMPOSITE

(distribution of alternative not defined)

This is probably a university student, because they spend £20 per week on alcohol and the average student spends £17 per week on this with a standard deviation of ~ £13, whereas this is normally what is expected for the typical UK household with an average of 1.9 adults.

SIMPLE

Statistical Power

When comparing 2 hypotheses, H0 and H1, the "statistical power" is the fraction of times that H0 is correctly rejected when H1 is true if one were to repeat the test many times with "identical" ensembles of data subject only to statistical fluctuations

Statistical Power

When comparing 2 hypotheses, H0 and H1, the "statistical power" is the fraction of times that H0 is correctly rejected when H1 is true if one were to repeat the test many times with "identical" ensembles of data subject only to statistical fluctuations

"Frequentist"

That's ridiculous… I only care whether I'VE made the right choice given THIS set of data!

Bayesian Power

When comparing 2 hypotheses, H0 and H1, the "Bayesian power" is the confidence you have in correctly rejecting H0 given the assumed probability distributions of H0 and H1 for this particular set of data

That's ridiculous… hypotheses don't have probability distributions: they are true or false!

Neyman-Pearson Lemma:

$$\Lambda \equiv \frac{L(D \mid H_0)}{L(D \mid H_1)}$$

(sometimes defined as one over this)

is

"Uniformly Most Powerful" (in a frequentist sense)
discriminate between simple hypotheses

(The exact distribution of Λ will, in general, depend on the distributions of L)

Assume that the set of possible hypotheses that describe a particular data set are distinguished only by the values of some unknown set of model parameters (*e.g.* the number of signal events, or the slope and intercept of a line, *etc.*).

Determining the best set of model parameters by comparing to find the **Maximum Likelihood** is therefore the UMP method of parameter estimation!

Simple example: You wait at a bus stop and no bus arrives for the first 10 minutes, but then 2 buses arrive in the next 10 minute interval. What is the best estimate of the mean number of buses per 10 minutes?

assume Poisson process $P(n\,|\,\mu) = \dfrac{\mu^n e^{-\mu}}{n!}$

$$L = P(0\,|\,\mu)P(2\,|\,\mu) = (e^{-\mu})(\tfrac{1}{2}\mu^2 e^{-\mu}) = \tfrac{1}{2}\mu^2 e^{-2\mu}$$

maximise the likelihood:

$$\frac{\partial L}{\partial \mu} = \mu e^{-2\mu} - \mu^2 e^{-2\mu} = 0$$

$$\rightarrow \mu_m = 1$$

(as expected)

Consider the case where uncertainties on data points are normally distributed. Assume that the mean values and variances, μ_i and σ_i, are predicted at each data point by some model. Then we have:

$$\log L = \sum_{i=1}^{N} \log \left[\frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2} \right) \right]$$

$$= \sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi}\sigma_i} - \sum_{i=1}^{N} \frac{(x_i - \mu_i)^2}{2\sigma_i^2}$$

$$-2 \log L = -2 \sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi}\sigma_i} + \sum_{i=1}^{N} \frac{(x_i - \mu_i)^2}{\sigma_i^2}$$

looks like χ^2

constant

Thus, maximising L = maximising $\log L$ = minimising $-2\log L$ is equivalent to the **Method of Least Squares** in this limit !!

Fisher Information

The "score" for data set D given a single model parameter q is defined as:

$$s(D\,|\,q) \equiv \frac{\partial \log \mathscr{L}(D\,|\,q)}{\partial q}$$

essentially the change in the fractional value of the likelihood as parameters are varied (for multiple parameters, this is a vector)

The score is zero at the maximum likelihood. So, if q represents the true model parameter, then

$$\langle s(D\,|\,q)\rangle = \left\langle \frac{\partial \log \mathscr{L}(D\,|\,q)}{\partial q} \right\rangle = 0$$

The **Fisher Information** is defined as the variance of the score.

Given the above, this is then:

$$\mathscr{I}(q) \equiv var\left[s(D\,|\,q)\right] = \left\langle \left(\frac{\partial}{\partial q} \log \mathscr{L}(D\,|\,q) \right)^2 \right\rangle$$

for k model parameters, this is a k x k covariance matrix

If the relative likelihood changes rapidly as parameters change (*i.e.* high variance), the data is therefore carrying a large information content for those parameters.

If the log likelihood is twice differentiable, then

$$\frac{\partial^2}{\partial q^2}\log \mathscr{L}(D\,|\,q) = \frac{\frac{\partial^2}{\partial q^2}\mathscr{L}(D\,|\,q)}{\mathscr{L}(D\,|\,q)} - \left(\frac{\frac{\partial}{\partial q}\mathscr{L}(D\,|\,q)}{\mathscr{L}(D\,|\,q)}\right)^2$$

$$= \frac{\frac{\partial^2}{\partial q^2}\mathscr{L}(D\,|\,q)}{\mathscr{L}(D\,|\,q)} - \left(\frac{\partial}{\partial q}\log \mathscr{L}(D\,|\,q)\right)^2$$

and $$\left\langle \frac{\frac{\partial^2}{\partial q^2}\mathscr{L}(D\,|\,q)}{\mathscr{L}(D\,|\,q)} \right\rangle = \int \frac{\frac{\partial^2}{\partial q^2}\mathscr{L}(x\,|\,q)}{\mathscr{L}(x\,|\,q)}\,\mathscr{L}(x\,|\,q)\,dx = \frac{\partial^2}{\partial q^2}\int \mathscr{L}(x\,|\,q)\,dx = 0$$

so $$\mathscr{I}(q) = \left\langle \left(\frac{\partial}{\partial q}\log \mathscr{L}(D\,|\,q)\right)^2 \right\rangle = -\left\langle \frac{\partial^2}{\partial q^2}\log \mathscr{L}(D\,|\,q) \right\rangle$$

for k model parameters, this is a k x k covariance matrix with entries:

$$\mathscr{I}(q_{ij}) = -\left\langle \frac{\partial^2}{\partial q_i \partial q_j}\log \mathscr{L}(D\,|\,\mathbf{q}) \right\rangle$$

Note that, since the log likelihood for the date set is the sum of log likelihoods of the n independent data points, then

$$\mathscr{I}(\mathbf{q}) = \sum_{i=1}^{n} \mathscr{I}_i(\mathbf{q})$$

Example: Bernoulli Distribution

(underpins all basic probability distributions)

2 possible outcomes, depends on single parameter p, the probability of success:

pass: $P(1) = p$ fail: $P(0) = 1 - p$

$$\frac{\partial \log P(1)}{\partial p} = \frac{1}{p} \qquad \frac{\partial \log P(0)}{\partial p} = -\frac{1}{1-p}$$

$$\mathcal{I}(p) = \left\langle \left(\frac{\partial \log P}{\partial p} \right)^2 \right\rangle = -\left\langle \frac{\partial^2 \log P}{\partial p^2} \right\rangle = p \times \frac{1}{p^2} + (1-p) \times \frac{1}{(1-p)^2}$$

$$= \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}$$

Note that this is equivalent to one over the distribution variance, since

$$\langle x \rangle = 1 \times p + 0 \times (1-p) = p$$

$$\langle x^2 \rangle = 1^2 \times p + 0^2 \times (1-p) = p$$

$$var = p - p^2 = p(1-p)$$

$$\mathcal{I} = \frac{1}{\sigma_{var}^2}$$

Exercise:
Show explicitly that this is also true for binomial, Poisson and Gaussian

Can we approximate the general shape of likelihood functions?

Consider a single parameter, q, which maximises the likelihood at $q=q_m$.
Now Taylor expand around the maximum likelihood point:

$$\ln L(q) = \ln L(q_m) + \left[\frac{\partial \ln L}{\partial q}\right]_{q=q_m} (q - q_m) + \frac{1}{2!}\left[\frac{\partial^2 \ln L}{\partial q^2}\right]_{q=q_m} (q - q_m)^2 + \ldots$$

zero by definition

must be negative, since we are at the maximum and, as $n \to \infty$, the magnitude tends to $\sim \mathcal{I}(q_m) = 1/\sigma^2_{q_m}$

$$\ln L(q) \sim \ln L(q_m) - \frac{1}{2}\frac{(q - q_m)^2}{\sigma^2_{q_m}}$$

$$q \to q_m \pm \sigma_{q_m}$$

looks like $\Delta\chi^2$!!

$$\ln L(q_m \pm \sigma_{q_m}) \sim \ln L(q_m) - \frac{1}{2} \quad \text{or} \quad -2[\ln L(q_m \pm \sigma_{q_m}) - \ln L(q_m)] \sim 1$$

Wilks' Theorem

more generally:

$$-2[\ln L(\mathbf{q_o}) - \ln L(\mathbf{q})] = -\mathbf{2}\ln\left(\frac{L(\mathbf{q_o})}{L(\mathbf{q})}\right) \equiv -\mathbf{2}\ln L_{\mathbf{R}} \sim \chi^2_d$$

where $\mathbf{q_o}$ are the set of model parameters that define the default (null) hypothesis, and the d = DoF = the difference in the number of model parameters constrained

(*i.e.* how many extra degrees of freedom one model has compared to the other)

Legal Statement:

- *For nested hypotheses (i.e. a continuous transition from one hypothesis to the next)*
- *Away from boundaries in likelihood space*
- *In the limit of large amounts of data*

However, for example, in the case of Poisson distributions, this actually works pretty well even for small numbers of events and also near μ=0. But generally need to check. Can do this, for example, by generating simulated data sets under a given hypothesis to directly look at the distribution of likelihood estimates.

Regarding the requirement of continuity for Wilks' Theorem…

Example 1: A model where you have some number of signal and some number of background, and you allow the relative fractions of these to change continuously while finding the most likely values.

Perfectly fine!!

Example 2: Using neutrino oscillation measurements to try to determine whether they have a normal or inverted mass ordering.

Violates Wilks' !!
The likelihood ratio still provides the UMP test, but the distribution of $-2\log\mathscr{L}_R$ cannot be assumed to follow χ^2

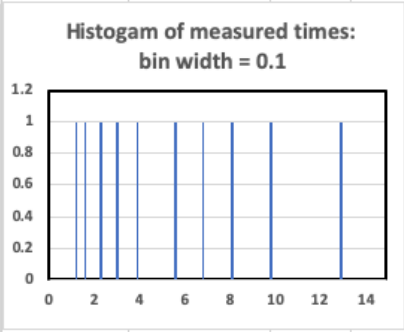
Example:

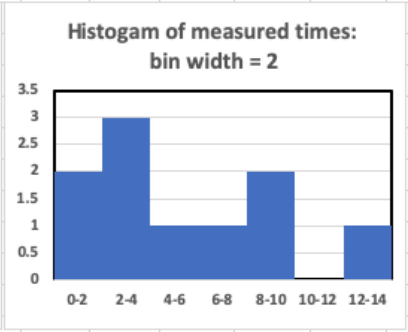
A newly commissioned underground neutrino detector sees a rate of internal radioactive contamination decreasing as a function of time. 10 events are observed over a period of 15 consecutive days. Determine the best fit mean decay time in order to determine the source of the contamination.

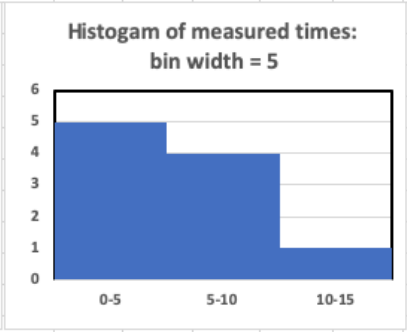
decay probability:

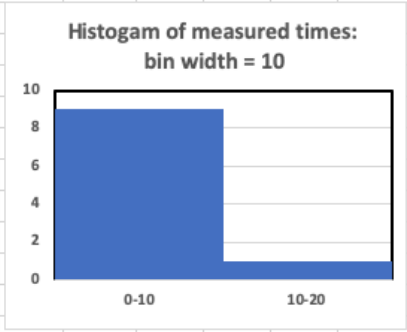
$$P(t) = \frac{1}{t_o} e^{-\frac{t}{t_o}}$$

t_o = mean decay lifetime

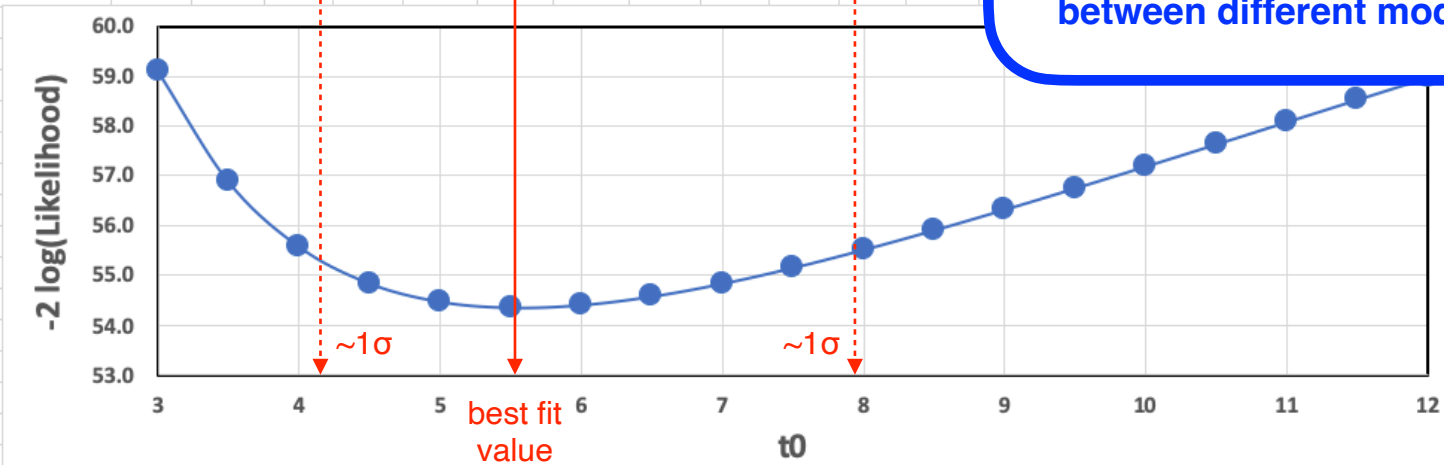








Measured Time (days)	t0:	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10	10.5	11	11.5	12
		Table of Probabilities: P(t) = (1/t0)*exp(-t/t0) for different assumed values of t0																		
5.6		0.0515	0.058	0.062	0.064	0.065	0.066	0.066	0.065	0.064	0.063	0.062	0.061	0.06	0.058	0.057	0.056	0.055	0.053	0.052
1.3		0.2161	0.197	0.181	0.166	0.154	0.144	0.134	0.126	0.119	0.112	0.106	0.101	0.096	0.092	0.088	0.084	0.081	0.078	0.075
2.4		0.1498	0.144	0.137	0.13	0.124	0.118	0.112	0.106	0.101	0.097	0.093	0.089	0.085	0.082	0.079	0.076	0.073	0.071	0.068
12.9		0.0045	0.007	0.01	0.013	0.015	0.017	0.019	0.021	0.023	0.024	0.025	0.026	0.027	0.027	0.028	0.028	0.028	0.028	0.028
6.8		0.0346	0.041	0.046	0.049	0.051	0.053	0.054	0.054	0.054	0.054	0.053	0.053	0.052	0.051	0.051	0.05	0.049	0.048	0.047
1.7		0.1891	0.176	0.163	0.152	0.142	0.133	0.126	0.118	0.112	0.106	0.101	0.096	0.092	0.088	0.084	0.081	0.078	0.075	0.072
9.8		0.0127	0.017	0.022	0.025	0.028	0.031	0.033	0.034	0.035	0.036	0.037	0.037	0.037	0.038	0.038	0.037	0.037	0.037	0.037
4		0.0879	0.091	0.092	0.091	0.09	0.088	0.086	0.083	0.081	0.078	0.076	0.073	0.071	0.069	0.067	0.065	0.063	0.061	0.06
8.1		0.0224	0.028	0.033	0.037	0.04	0.042	0.043	0.044	0.045	0.045	0.045	0.045	0.045	0.045	0.044	0.044	0.044	0.043	0.042
3.1		0.1186	0.118	0.115	0.112	0.108	0.103	0.099	0.095	0.092	0.088	0.085	0.082	0.079	0.076	0.073	0.071	0.069	0.066	0.064
Product of probabilities:		1E-13	4E-13	9E-13	1E-12	1E-12	2E-12	2E-12	1E-12	1E-12	1E-12	9E-13	7E-13							
Sum of $\log_e$(probabilities):		-29.55	-28.4	-27.8	-27.4	-27.23	-27.17	-27.2	-27.3	-27.4	-27.6	-27.8	-28							
-2 x Sum of the logs :		59.106	56.88	55.58	54.84	54.47	54.35	54.4	54.57	54.83	55.15	55.51	55.91							



No absolute goodness-of-fit, just the "relative goodness" between different models