18/09/18

Scientific Method. (Data Analysis & Interpretation)

Collect data → Ask a question → Hypothesis/guess → Testable prediction

Testable prediction → Agreement → Generalise theory

Testable prediction → Disagreement → refute/alter → make new set of prediction.

Extrapolation → Validity limit.

Collect more data, pushing the boundaries of limit of validity.

Syllabus

1. Errors in an experiment.
2. Hypothesis comparison
3. Parameter Estimation.
4. Hypothesis testing.

~~If time:~~

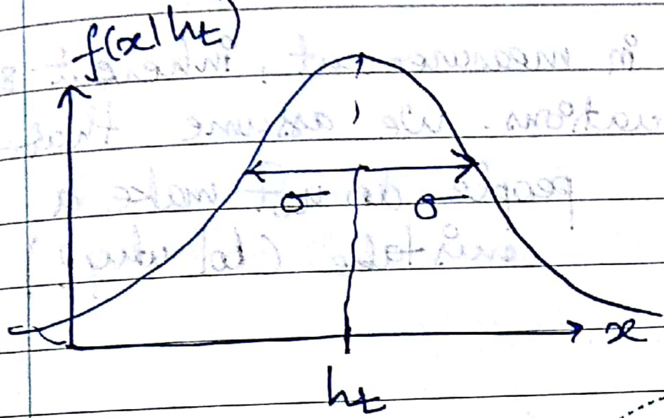
5. Confidence Limits.

- Measurements & Errors

→ ① Direct → x → h_t (height parameter)
→ ② Indirect (later)

$f(x|h_t)$

↳ We assume h_t is the mean of the fluctuation of x.
$\langle x \rangle = h_t$.



$f(x|h_t, \underbrace{\sigma, \dots}_{\text{ignore other parameters that are possible.}})$

$y(x)$ ↳ some other observable that we measure using a fixed relation with a direct measurement of x.

- Measurement: whatever we directly read off the measuring devices.

- Errors: Spread in the values of x about $\langle x \rangle = h_t$
↓
σ: standard deviation.

$x \to$ measure (actually measured)
We assume the existence of the probability distribution $f(x|h_t)$
We assume the parameter to be the mean of the pdf
$\sigma \to$ characterises the uncertainty in x.

Date:

www.navneet.com

We really want h_t, σ $\{\langle x \rangle, \sigma\}$.

Measurements → estimate h_t
Errors → estimate σ
} given our observations or the best guess for the $f(x | h_t, \sigma)$

Errors → due to limitations in measurement, inherent sources of fluctuations. We assume that the people do not make a mistake (lol why?)

Bias Vs Error
↑ Avoidable

- Errors: Unavoidable random dispersions in the data. (In a specific experimental set up)

e.g. → Result of radioactive decay experiment.

$\boxed{} \xrightarrow{(1 \text{ min})}$ 3 decays

exp. setup → x

Assuming that it obeys a Poisson distribution,

$$P[K/\lambda] = \frac{\lambda^K}{K!} e^{-\lambda}$$

$\sigma = \sqrt{\lambda}$

$\mu = \lambda$

Using only one measurement,

$\sigma = \sqrt{\lambda} = \sqrt{3}$

$\mu = \lambda = 3$

∴ Answer reported

$= 3 \pm \sqrt{3}$

$= 3 \pm 1.7$

Scanned by CamScanner

How to report measurements and errors :

$\rightarrow$ $x \pm \delta_x$

x : estimate for $\langle x \rangle$

δ_x : estimate for σ

$\boxed{x \pm \delta_x}$

(no mention of $f(x)$) (no claims on the shape of $f(x)$).

$\delta x \rightarrow$ typically 1–2 decimal/significant digits.
$x \rightarrow$ to the same significance as the errors.

(34.53 ± 0.02)

$x = 10.3^{+0.7}_{-0.3}$ } $\rightarrow$ Not using SD as a measure of the spread.

$m_e = 0.51099906 \pm 0.00000015$ MeV/c^2

$\therefore m_e = 0.51099906(15)$ MeV/c^2

0.51099906 ± 0.3 ppm $\rightarrow$ % error.

28/09/18

Errors : "spread" in the values of x. ($x_t \rightarrow$ true value)
$\rightarrow$ Estimate σ (SD) of $f(x|x_t)$

• Why assign errors and how to assign errors.

Why $\rightarrow$ to compare theory with experiments.
$\rightarrow$ to compare different experimental procedures.
and combine them and need to define the error in such a way that this comparison/combination can be done without detailed knowledge/reference to the experimental procedure.

→ overestimating the error is bad as it leads to waste resources.

→ underestimating the error is bad also as it leads to inaccuracy

Why use SD? → combining / comparison becomes easier.
(Uncorrelated experiments.
⇓
easy to combine errors $\sigma_1^2+\sigma_2^2=\sigma^2$)

- How do we estimate x_{true}?

$E[\underline{x}] = x_t$

↑

choose a quantity $\underline{x}$ s.t. expectation value of $\underline{x}$ is x_t.

Some notations

x → real number ($\mathbb{R}$)

$\underline{x}$ → random variable (RV)

$\tilde{x}$ → observation.

We use $\underline{x}$ → as an estimate for x_t.

$\underline{x}$ → before you do the experiment.

$\tilde{x}$ → experimental observation.

Similarly, for the error,

$$E[(\underline{x}-x_t)^2] = \sigma_t^2$$

We want an estimate for $\underline{\sigma_t}$.

$\underline{x}$ → RV $\quad \underline{f(x)}$; ~~Et~~ Choose f s.t. $E[f(\underline{x})] = \sigma_t^2$

↓

Also RV

$\tilde{x} \to$ observation

$\sqrt{f(\tilde{x})} \Rightarrow$ estimate of σ_t

$\therefore \; \sigma_{est}^2 = f(\tilde{x})$

E.g. We want to seperate $^{238}U_{(99\%)}$ and $^{235}U_{(0.09\%)}$

Apply magnetic field. $\leftarrow$ Difference in $\frac{Q}{M}$ ratio of nuclei.

(mass spectrometer)

Used: Centrifuge.

$\frac{dN}{dt} = N_0 e^{-t/\tau}$

If the number of decays 'K' $\to P^d(K)$

$P(\text{decay}) = \frac{dN}{N} = \frac{dt}{\tau}$ $\quad (<< 1)$

$\frac{dN}{dt} = \frac{N}{\tau}$

$\underbrace{(P \to 0)}$

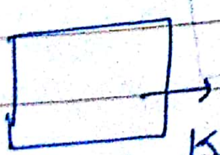
Probability of an individual particle decay.

N atoms $\Rightarrow$ $P_p^N(k) \Rightarrow B_p^N(K) \xrightarrow[p \to 0]{N \to \infty} P^d(k)$.

$\lambda = Np = N \frac{\Delta t}{\tau} = \langle K \rangle$

$K \to$ no. of decays.

$P(\underline{K} = k) = \frac{\lambda^k e^{-\lambda}}{k!}$.



$\underline{K}$. $\quad \tilde{K} \to$ actual observation.

$\lambda = k_t$ (from theory)

for one observation,

est $\lambda = \tilde{K}$.

est $\sigma_k = \sqrt{\tilde{K}}$

$E[\underline{K}] = \lambda_t$.

i.e, $f(\underline{K}) = \underline{K}$. ??

$$\tau^{est} = \frac{N\Delta t}{\lambda^{est}} \pm \Delta\tau$$

$\lambda \to$ direct measurement.
$\tau \to$ indirect measurement.

- Types of errors

→ Statistical Errors (→ errors decrease as we take more measurements)

→ Systematic Errors.

① Known distribution with one parameter. (e.g., poisson) Bernoulli

② Take a large number of measurements (IIDs)
IID Assumption.

Identical → f stays same.
independant → each trial is independant.

$$\text{est } x_t = \frac{1}{N}\sum_{i=1}^{N} \tilde{x}_i$$

(Since all of them are IIDs, mean distributes over the sum.)

$$x_{av} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad E[x_{av}] = x_t.$$

$$\text{est}(\sigma_t^2) = E[f(x_1, x_2 \ldots x_N)]$$
$$= \frac{1}{N-1}\sum_{i=1}^{N}[(x_i - x_{av})^2]$$

$$\text{The, est } \sigma_t^2 \neq \frac{1}{N-1}\sum_{i=1}^{N}(\tilde{x}_i - \tilde{x}_{av})^2.$$

What we measure is the error in individual measurements.

What we're interested in the error in the sample mean?

$$\text{est } \sigma_{\tilde{x}_{av}}^2 = \frac{1}{N} \text{ est } \sigma_{\tilde{x}}^2 = \frac{1}{N} \cdot \frac{1}{N-1} \sum_{i=1}^{N} (\tilde{x}_i - \tilde{x}_{av})^2 .$$

$$\tilde{\delta}_{\tilde{x}_{av}} = \frac{1}{\sqrt{N}} \delta_{x_i} .$$

→ We use the central limit theorem to get the factor of $\frac{1}{\sqrt{N}}$.

5/10/18

- Statistical Errors

f → is effectively a 1-parameter distribution
↳ single observation can be used to estimate mean & SD.

Multiple observations as a way to improve error estimates.

$\underbrace{\text{Multiple observations}}_{\text{IID}}$ Observations.

$f(x \mid x_t) \longrightarrow \tilde{x}_1, \tilde{x}_2, \tilde{x}_3 \ldots \tilde{x}_N$ } ← observations.

$\underbrace{f(x \mid x_t)}_{\sigma}$

$$\mu_{est} = \frac{\tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 - - - \tilde{x}_N}{N}$$

$$\sigma_{est}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\tilde{x}_i - \mu_{est})^2$$

→ Reason why this is a good estimator

"Error on the error"

$$E\left[\frac{1}{N-1} (x_i - x_{av})^2\right] = \sigma^2$$

σ ← random variate.

σ_{est} → fluctuation of each $\boxed{\tilde{x}_i}$

$\sigma_{est} \rightarrow$ not the error in μ_{est} ; σ_{est} is the error in $\tilde{\sigma}\, x_i$

$$\left(\tilde{\sigma}_{est}^{av}\right)^2 = \frac{1}{N} \tilde{\sigma}_{est}^{2} \quad \text{individual independent}$$

$$\tilde{\sigma}_{est}^{av} = \frac{1}{\sqrt{N}} \tilde{\sigma}_{est} \quad \text{individual.}$$

Report $\Rightarrow$ $\mu_{est} \pm \tilde{\sigma}_{est}^{av}$

Imp: Property of statistical errors $\Rightarrow$ Assume $\boxed{\text{IID}}$ measurements

$$\boxed{\tilde{\sigma} \sim \frac{1}{\sqrt{N}}} \qquad \tilde{\sigma} \text{ statistical}$$

- Error Estimate $\rightarrow$ $\boxed{\tilde{\sigma}_{av}^{est} \text{ on } x. \rightarrow \delta_x}$ estimated error

$\boxed{x} \rightarrow$ measurement

- Systematic Errors :

(unaccounted) ① correlations in repeated measurements.
IID assumption breaks down.

② usually these errors are poorly Known.

③ Unlike statistical errors, repeated measurements will not reduce the systematic errors.

• **Error propogation:** ↗ Direct → $f(x|x_t)$ (L, T)
↘ Indirect " $y(x)$ (g)
(derived quantities)

Error in derived things?

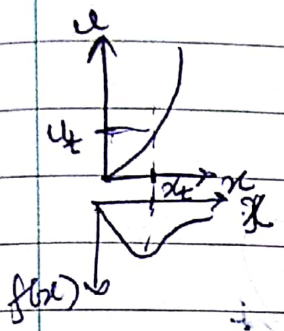
$x \rightarrow u(x)$

$\updownarrow$ close to x_t $\quad$ $\updownarrow$ close to $u(x_t) = u_t$

$u(\tilde{x}) \simeq u(x_t)$

$\tilde{x} = x_t + (\tilde{x} - x_t)$

$$u(\tilde{x}) = u(x_t) + \left.\frac{du}{dx}\right|_{x_t} (\tilde{x} - x_t) + \cancel{O((\tilde{x}-x_t)^2)}$$



$u(x) \approx u(x_t) + u'(x_t)(x - x_t) + \dots$

$y = mx + c.$

$\sigma_x^2 = E[(x - \langle x \rangle)^2]$

$\sigma_y^2 = E[(y - \langle y \rangle)^2]$

$$\sigma_y^2 = E[((mx + c) - (m\langle x \rangle + c))^2]$$

$$= m^2 E[(x - \langle x \rangle)^2] = m^2 \sigma_x^2.$$

$\sigma_y = m\sigma_x$ $\quad$ $m = u'(x_t)$

$$\boxed{\sigma_u^2 = [u'(x_t)]^2 \sigma_x^2}$$

We have assumed $|u'(x_t)| \gg |u''(x_t)(x - x_t)|$

& $|u'(x_t)| \gg |u''(x_t)\sigma_x|$

$$\therefore \boxed{|\sigma_x| \ll \left|\frac{u'(x_t)}{u''(x_t)}\right|}$$

⇒ Condition for "small" errors so as to neglect higher order terms, etc.

(Since we talk in terms of factors, we ignore the "2")

10

Q. $y = e^x$ $\quad \tilde{x} = 3 \quad \tilde{\delta}_x = 0.1$.

(3 ± 0.1)

① $\delta_y = ?$ & is ② ignoring subsequent terms, i.e, linear approximation justified?

$\underbrace{\sigma_y^2 = (e^3)\,\sigma_x^2}_{\text{we plug in our best estimates}} \Rightarrow \sigma_y = e^{3/2}\,0.1$.

① $\boxed{\sigma_y = \frac{e^{3/2}}{10}}$ ≈ 2.7

$u'(x_t) = u''(x_t) = e^{x_t}$

② $\underbrace{0.1 < 1}$ ∴ justified / valid

③ What is x_t? $\rightarrow x_t \neq 3$.

3 → estimate for x_t.

we dont know x_t!

$x_t \notin (3 \pm 0.1)$. (not necessary).

$\boxed{y = 27 \pm 2.7}$

(taking $e = 3$)

- Multiple variables error propogation.

(multiple direct ~~observations~~)

$x_1, x_2 \ldots x_n \longrightarrow u(x_1, x_2, x_3 \ldots x_n)$.

$$u(x_1, x_2, x_3 \ldots x_n) = u(x_1^t, x_2^t, x_3^t \ldots x_n^t) + \left.\frac{\partial u}{\partial x_1}\right|_{x_1^t, \ldots x_n^t} (x_1 - x_1^t)$$

$$+ \left.\frac{\partial u}{\partial x_2}\right|_{x_1^t, \ldots x_n^t} (x_2 - x_2^t) + \ldots\ldots$$

(various parameters, not IID).

Scanned by CamScanner

$$\sigma_u^2 = \left(\frac{\partial u}{\partial x_1}\right)\Big|_{(x_1^t \dots x_n^t)} \sigma_{x_1}^2 + \left(\frac{\partial u}{\partial x_2}\right)\Big|_{x_1^t, x_2^t \dots x_n^t} \sigma_{x_2}^2 + 2\frac{\partial u}{\partial x_1}\frac{\partial u}{\partial x_2}\sigma_{x_1 x_2}$$

(for two variables)

To check if two quantities are correlated or not, we need to check $\sigma_{x_1 x_2}$.

⟹ Generalising to n quantities :⟹

$$\sigma_u^2 = \begin{pmatrix} \frac{\partial u}{\partial x_1} & \frac{\partial u}{\partial x_2} \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} \partial u/\partial x_1 \\ \partial u/\partial x_2 \end{pmatrix}$$

$$\sigma_u^2 = \begin{pmatrix} \frac{\partial u}{\partial x_1} & \frac{\partial u}{\partial x_2} & \dots & \frac{\partial u}{\partial x_n} \end{pmatrix} \underbrace{\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots \\ \vdots & \sigma_2^2 & \\ & & \ddots \end{pmatrix}}_{\substack{\text{Covariance matrix} \\ C_n}} \begin{pmatrix} \partial u/\partial x_1 \\ \partial u/\partial x_2 \\ \vdots \\ \partial u/\partial x_n \end{pmatrix}$$

⟹ Generalising for m indirect quantities:

$$\begin{pmatrix} \sigma_{u_1}^2 \\ \sigma_{u_2}^2 \\ \sigma_{u_3}^2 \\ \vdots \\ \sigma_{u_m}^2 \end{pmatrix} \cancel{\times} \begin{pmatrix} \sigma_{u_1}^2 & \sigma_{u_1 u_2} & \dots \\ \sigma_{u_1 u_2} & \sigma_{u_2}^2 & \\ \vdots & & \ddots \\ & & & \sigma_{u_m}^2 \end{pmatrix}$$

Not exactly the Jacobian.

Page No.	
Date	

$$\begin{pmatrix} \\ \\ \end{pmatrix}_{m \times m} = \underbrace{\begin{pmatrix} \frac{\partial u_1}{\partial x_1} \frac{\partial u_1}{\partial x_2} \\ \vdots \end{pmatrix}_{m \times n}}_{M} \begin{pmatrix} C^{(x)}_{n \times n} \end{pmatrix}_{n \times n} \underbrace{\begin{pmatrix} \\ \\ \end{pmatrix}_{n \times m}}_{M^T}$$

full covarience matrix of $u_1, u_2 \ldots u_m$

$$\begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} & \cdots & \frac{\partial u_1}{\partial x_n} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} & \cdots & \frac{\partial u_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial u_m}{\partial x_1} & \frac{\partial u_m}{\partial x_2} & \cdots & \frac{\partial u_m}{\partial x_n} \end{pmatrix}$$

- How to estimate covarience?

$$f(\overset{x_t}{x}, \overset{y_t}{y}) \longrightarrow (x_1, y_1), (x_2, y_2), (x_3, y_3) \ldots\ldots (x_n, y_n)$$

[each pair is IID]

$\sigma_x^2 \quad \sigma_y^2$

σ_{xy}

$$S^2_{xy} = \text{est } \sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (\tilde{x}_i - \tilde{x}_{av})(\tilde{y}_i - \tilde{y}_{av})$$

$\mu_{x_t} = \text{est } x_t \doteq \frac{1}{N} \sum_{i=1}^{N} \tilde{x}_i$; $\mu_{y_t} \doteq \text{est } y_t \doteq \frac{1}{N} \sum_{i=1}^{N} \tilde{y}_i$

Correlation b/w our estimates μ_t^x & μ_t^y ?

$$\sigma^{av}_{est\, x_t} = \frac{1}{\sqrt{N}} \sigma^{single}_{est\, x_t} \qquad \sigma^{av}_{est\, y_t} = \frac{1}{\sqrt{N}} \sigma^{single}_{est\, y_t}$$

$$\delta^{2\, av}_{xy} = \frac{1}{N} \delta^{2\, single}_{xy}$$

Reporting $\Rightarrow$ $\begin{pmatrix} x_{est} & 0 \\ 0 & y_{est} \end{pmatrix} \pm \begin{pmatrix} \delta_x^2 & \delta_{xy}^2 \\ \delta_{xy}^2 & \delta_y^2 \end{pmatrix}$

06/10/18 Saturday

DAI — Extra (Double) Lecture.

• Systematic Errors

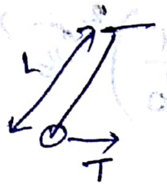
$f(x|x_t)$

$\downarrow$

$\tilde{x}_1, \tilde{x}_2, \dots \tilde{x}_n$.

Irreducible errors.

IID. $\delta^{stat} \sim \frac{1}{\sqrt{N}}$ $\qquad \delta^{sys} \not\propto \frac{1}{\sqrt{N}}$

Eg. 

$g = \frac{4\pi^2 L}{T^2}$

measure $\rightarrow$ L, T.

sources of error $\rightarrow$
in L $\Rightarrow$ • least count.
• temp. fluctuation.

in T $\Rightarrow$ • delay in reaction time.

• Take as many measurements as needed to make $\delta^{stat} \simeq \delta^{sys}$.
(no point reducing δ^{stat} further as δ^{sys} will dominate the source of errors).

$u(x, a, b)$
$\uparrow$
derived quantity

$x \rightarrow$ Known quantity. (i.e., L, T, etc.) $\rightarrow$ time
(have statistical IID errors)

$a, b \rightarrow$ unknown factors (e.g., temp).

$a_t, b_t \rightarrow$ av. value of the unknown factors.

a, b fluctuate around a_t, b_t.

$x_1, x_2 \rightarrow$ fluctuations in x.

$u_1 = u(x_1, a, b)$
$u_2 = u(x_2, a, b)$

$\Delta u_1 = u_1 - u_t$

$u_t \rightarrow$ evaluated at (x_t, a_t, b_t).

$$\Delta u_1 = \left.\frac{\partial u}{\partial x}\right|_t \Delta x + \left.\frac{\partial u}{\partial a}\right|_t \Delta a + \left.\frac{\partial u}{\partial b}\right|_t \Delta b$$

$$\Delta u_2 = \frac{\partial u}{\partial x}\Big|_{x_t, a_t, b_t} \Delta x_2 + \frac{\partial u}{\partial a}\Big|_t \Delta a + \frac{\partial u}{\partial b}\Big|_t \Delta b.$$

Δa, Δb in both cases are assumed to be the same. (also assumed x, a, b are independant of each other). (100% correlation).

- $$Var(u_1) = \left(\frac{\partial u}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial u}{\partial a}\right)^2 \sigma_a^2 + \left(\frac{\partial u}{\partial b}\right) \sigma_b^2.$$

- $$Var(u_2) = \left(\frac{\partial u}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \left(\frac{\partial u}{\partial a}\right)^2 \sigma_a^2 + \left(\frac{\partial u}{\partial b}\right)^2 \sigma_b^2$$

- $$covar(u_1, u_2) = E[\Delta u_1 \Delta u_2]$$

$$= \left(\frac{\partial u}{\partial x}\Big|_{x_t, a_t, b_t}\right)^2 E[\Delta x_1 \Delta x_2] + \left(\frac{\partial u}{\partial a}\right)^2 E[\Delta a \Delta a]$$

$$+ \left(\frac{\partial u}{\partial b}\right)^2 E[\Delta b \Delta b]$$

+ Cross terms.

(cross terms = 0. (independance)

$E[\Delta x_1, \Delta x_2] = 0$ (independant measurements).

$$\therefore \; covar(u_1, u_2) = \left(\frac{\partial u}{\partial a}\right)^2 \sigma_a^2 + \left(\frac{\partial u}{\partial b}\right)^2 \sigma_b^2$$

$$\underbrace{\quad}_{u_1, u_2 \text{ NOT independant.}} \neq 0$$

u_1, u_2 NOT independant.

$u_t = u(x_t, a_t, b_t)$.

- $a, b \rightarrow$ same fluctuation even after repeated measurements.

$\rightarrow$ will not go down with repeated measurements.

$\rightarrow$ will underestimate the error, (not taking σ_a, σ_b into account) when averaging many measurements.

$$\tilde{u}_1, \tilde{u}_2, \tilde{u}_3 \ldots \tilde{u}_n \mapsto \tilde{u}_{av} = \frac{1}{N} \sum_i \tilde{u}_i$$

$$\sigma^{est\,2} = \frac{1}{N-1} \sum_{i=1}^{N} (\tilde{u}_i - \tilde{u}_{av})^2$$

$$\sigma^{est\,av.\,2} = \frac{1}{N} \frac{1}{N-1} \sum_{i=1}^{N} (\tilde{u}_i - \tilde{u}_{av})^2.$$

comes from IID assumption of $\tilde{u}_i$.

$\hookrightarrow$ not true if we take a, b into account.

(the source of correlations are not always ~~obs~~ obvious).

- Assumption of same a, b (100% correlation) may not be true. In general, it could be $<100\%$.

In that case,

$$covar(u_1, u_2) = \left(\frac{\partial u}{\partial a}\right)^2 \sigma_{a_{12}} + \left(\frac{\partial u}{\partial b}\right)^2 \sigma_{b_{12}}$$

• How should we properly declare systematic errors?

$$\circ\ C_{u_1,u_2} = \begin{pmatrix} \left(\frac{\partial u_1}{\partial x}\right)^2 \sigma_{x_1}^2 & 0 \\ 0 & \left(\frac{\partial u_2}{\partial x}\right)^2 \sigma_{x_2}^2 \end{pmatrix}$$ Statistical errors

$$+ \left(\left(\frac{\partial u}{\partial a}\right)^2 \sigma_a^2 + \left(\frac{\partial u}{\partial b}\right)^2 \sigma_b^2\right) \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

(assuming 100% correlation)

$$\circ\ C_{u_1,u_2} = \begin{pmatrix} \left(\frac{\partial u_1}{\partial x}\right)^2 \sigma_{x_1}^2 & 0 \\ 0 & \left(\frac{\partial u_2}{\partial x}\right)^2 \sigma_{x_2}^2 \end{pmatrix}$$ Statistical error

$$+ \begin{pmatrix} \left(\frac{\partial u}{\partial a}\right)^2 \sigma_a^2 + \left(\frac{\partial u}{\partial b}\right)^2 \sigma_b^2 & \left(\frac{\partial u}{\partial a}\right)^2 \sigma_{a_{12}} + \left(\frac{\partial u}{\partial b}\right)^2 \sigma_{b_{12}} \\ \left(\frac{\partial u}{\partial a}\right)^2 \sigma_{a_{12}} + \left(\frac{\partial u}{\partial b}\right)^2 \sigma_{b_{12}} & \left(\frac{\partial y}{\partial a}\right)^2 \sigma_a^2 + \left(\frac{\partial u}{\partial b}\right)^2 \sigma_b^2 \end{pmatrix}$$

• Convention

measurement ± statistical errors ± systematic errors

$$u \pm \sqrt{u_x^2 \sigma_x^2} \pm \sqrt{u_a^2 \sigma_a^2 + u_b^2 \sigma_b^2}$$

$$\delta_x^{2\,stat} \Rightarrow u_x^2 \sigma_x^2 \qquad \delta_u^{2\,sys} \Rightarrow \left(\frac{\partial u}{\partial a}\right)^2 \sigma_a^2 + \left(\frac{\partial u}{\partial b}\right)^2 \sigma_b^2.$$

E.g. Initially, statistical errors dominate & we estimate systematic errors.

↳ we should keep taking measurements till we reduce the statistical errors below the systematic errors

∴ → Systematic errors are the limit of sensitivity of an experimental setup.

• Problem : ⇒ If you have large unaccounted for systematic errors → existence of errors will not be known due to fluctuations in data.

↳ Incorrect interpretation of results & comparison of results between different experiments.

• Important for experimentalists

① Check carefully for any source of systematic errors.

② Document every parameter of potential importance.

Typically $\longrightarrow$ SOURCES OF SYSTEMATIC ERRORS

① Poorly calibrated instruments.

② Influenced in an unknown way by the environment (T, P), etc.

③ Thresholds or base levels not constant in time.

④ Limited knowledge or understanding of the experimental setup.

Acceptance, sensitivity, of detector components incompletely known.

⑤ Partially known background contribution.

⑥ Use of experimental input (additional) from other ~~components~~ experiments or theory results, with large and unclear uncertainties.

$\left(\alpha = \frac{e^2}{4\pi\epsilon_0 \hbar c} \simeq \frac{1}{137}\right.$ fine structure constant $\left.\right)$ $\simeq 0.01$.

↳ Important in perturbation theory.

↳ theoretical error sources

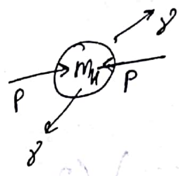
---

⑦ uncertainty in past measurements, for, e.g., when calibrating T_{calib}, P_{calib}.

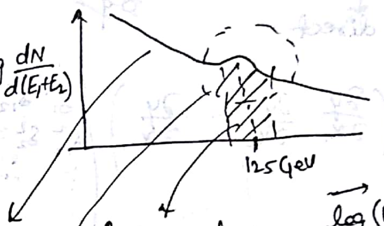
⑧ Background substraction uncertainties.

E.g. $H \to \gamma + \gamma$ 0.1%

125 GeV/c^2



$E_{\gamma_1} + E_{\gamma_2} = 125$ GeV



Actual data will be noisy

Background

Detection of the bump from removal of the background contribution.

Background substraction uncertainties $\Longleftarrow$

↳ interpolation / extrapolation.

---

⑨ Measurement of an interesting parameter correlated with the value of an uninterested parameter.

Nuisance parameter.

uncertainty in nusiance parameters $\Rightarrow$ induces a source of systematic errors.

2014 $\to$ BICEP

Cosmic microwave background.

$\to$ follows blackbody spectra.

$\to$ Wien's displacement law. ($T \simeq 2.7$K)

$\lambda_m T \Rightarrow$ = 0.29 cmK.

• Propogation of errors

$$y(\underbrace{x_1, x_2, \ldots}_{\text{direct}})$$

$\uparrow$ indirect

$$\delta_y^2 = \left(\frac{\partial y}{\partial x_1} \;\; \frac{\partial y}{\partial x_2} \;\cdots\; \frac{\partial y}{\partial x_n}\right) \begin{pmatrix} \delta_1^2 & \delta_{12}^2 & \cdots \\ \delta_{12}^2 & \delta_2^2 & \\ \vdots & & \ddots \\ & & & \delta_n^2 \end{pmatrix} \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{pmatrix}$$

E.g.

$g = \frac{4\pi^2 L}{T^2}$ $\qquad L_m \pm \delta_L$

$\qquad T_m \pm \delta_T$

$$g_m = 4\pi^2 \frac{L_m}{T_m^2}.$$

$$\delta_{g}^2 = \left(\frac{\partial g}{\partial L} \;\; \frac{\partial g}{\partial T}\right)\begin{pmatrix} \delta_1^2 & \delta_{12}^2 \\ \delta_{12}^2 & \delta_2^2 \end{pmatrix}\begin{pmatrix} \frac{\partial g}{\partial L} \\ \frac{\partial g}{\partial T} \end{pmatrix}$$

$\delta_{12} = 0$ (independent parameters)

$$= \left(\frac{\partial g}{\partial L}\right)^2 \delta_1^2 + \left(\frac{\partial g}{\partial T}\right)^2 \delta_2^2$$

$$= \left(\frac{4\pi}{T^2}\right)^2 \delta_1^2 + \left(\frac{8\pi L}{T^2}\right)^2 \delta_T^2$$

$$= g_m^2 \left[\left(\frac{\delta_L}{L}\right)^2 + \left(\frac{2\delta_T}{T}\right)^2\right]$$

$$\frac{\delta g}{g_m} = \sqrt{\left(\frac{\delta_L}{L}\right)^2 + \left(\frac{2\delta_T}{T}\right)^2}$$

In general, $y = x_1^m x_2^n$ (assume x_1, x_2 are uncorrelated)

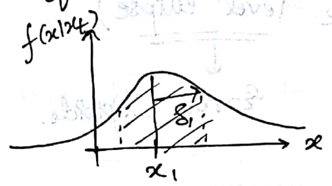
$$\frac{\delta_y}{y} = \sqrt{\left(m\frac{\delta_1}{x_1}\right)^2 + \left(n\frac{\delta_2}{x_2}\right)^2}$$

• Error Ellipsoids

$$C_{x_1 \ldots x_n} = \begin{pmatrix} \delta_1^2 & \delta_{12}^2 & \cdots \\ \delta_{12}^2 & \delta_2^2 & \cdots \\ \vdots & \vdots & \ddots \\ & & & \delta_n^2 \end{pmatrix}$$

single variable : $\boxed{x_1 \pm \delta_1}$

If x is drawn from a gaussian distributions,



$[x_1 - \delta_1, x_1 + \delta_1]$

68% confidence interval for x_t.

$[x_1 - 2\delta_1, x_1 + 2\delta_1] \leftarrow 95\%$

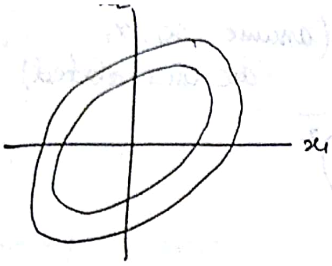
$[x_1 - 5\delta_1, x_1 + 5\delta_1] \leftarrow 99\%$

1D Gaussian $\rightarrow$ $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x^2}{2\sigma^2}\right]$

2D Gaussian $\rightarrow$ $f(x_1, x_2) = 1 \quad \exp\left[-\frac{1}{2} X^T V X\right]$

$V \rightarrow$ weight matrix $\qquad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \frac{x}{\sigma} = \eta$

$V = C^{-1}$.



$$\sqrt{(x_1\ x_2)\, C^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} = n.$$

For 2D ⇒ $n = 1$ ⇒

$n = 2$ ⇒

(larger than 1D)

For 1D ⇒ $\frac{x}{\sigma} = n = 1 \Rightarrow 68\%.$

$\frac{x}{\sigma} = n = 2 \Rightarrow 95\%.$

What is $k = \sqrt{(x_1\ x_2)\, C^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}$ so that 68% of the probability is enclosed by the corresponding contour. (68% confidence level ellipse)

↓

Error ellipsoids.

• **Averaging Measurements**

exp.1 → $x_1 \pm \delta_1$ ~~ATLAS~~ ATLAS ⇒ 125.3 GeV ± 0.1 GeV

exp.2 → $x_2 \pm \delta_2$ CMS ⇒ 125.4 GeV ± 0.2 GeV.

Uncorrelated errors.

↳ Total error should decrease on combining measurements.

↳ Higher weight to the expt with smaller error.

$x^{comb} = w_1 x_1 + w_2 x_2$

$w_1 + w_2 = 1.$

By defn, we choose the weights so that $\delta_{x\,comb}$ is minimised.

$\delta_{x\,comb} = (w_1\, \delta_{x_1})^2 + (w_2\, \delta_{x_2})^2$ ($\delta_{x_1 x_2} = 0$ uncorrelated).

$\delta_{x\,comb} = (w_1\, \delta_{x_1})^2 + ((1-w_1)\delta_{x_2})^2$ $0 \le w_1 \le 1.$

↳ minimize this by choosing the best value of w_1.

$$\boxed{\begin{aligned} &2w_1\, \delta_{x_1}^2 - 2(1-w_1)\, \delta_x^2 = 0. \\ &w_1 (2\delta_{x_1}^2 + 2\delta_{x_2}^2) = 2\delta_{x_2}^2 \\ &w_1 = \frac{\delta_{x_2}^2}{\delta_{x_1}^2 + \delta_{x_2}^2}. \end{aligned}}$$

$$\Rightarrow \delta_{x_{comb}} = \frac{1}{\left(\frac{1}{\delta_1}\right)^2 + \left(\frac{1}{\delta_2}\right)^2}$$

$$w_1 = \frac{1/\delta_{x_1}^2}{\frac{1}{\delta_{x_1}^2} + \frac{1}{\delta_{x_2}^2}} \qquad w_2 = \frac{1/\delta_{x_2}^2}{\frac{1}{\delta_{x_1}^2} + \frac{1}{\delta_{x_2}^2}}$$

$$\boxed{\begin{aligned} &x^{comb} = w_1 x_1 + w_2 x_2 + \ldots \\ &w_1 + w_2 + w_3 \ldots = 1 \end{aligned}}$$

$$\Rightarrow \delta_{x_{comb}} = (w_1\, \delta_{x_1}^2)^2 + (w_2\, \delta_{x_2}^2)^2 + \ldots (w_n \delta_{x_n})^2$$

$$w_i = \frac{1/\delta_{x_i}^2}{\frac{1}{\delta_{x_1}^2} + \frac{1}{\delta_{x_2}^2} + \ldots \frac{1}{\delta_{x_n}^2}}$$

$$\boxed{w_i = \frac{1/\delta_i^2}{\sum_{i=1}^{n} 1/\delta_i^2}}$$

$$\boxed{\delta_{comb}^2 = \frac{1}{\sum_{i=1}^{n} \frac{1}{\delta_i^2}}}$$

$\Rightarrow$ Assume all δ_{x_p}'s are the same.

$$\boxed{\delta_{comb} = \frac{\delta_x}{\sqrt{N}}}$$

• Averaging correlated measurements (Systematic)

x_1
x_2
$\vdots$
x_n

$$C = \begin{pmatrix} \delta_1^2 & \delta_{12}^2 & \cdots \\ \delta_{12}^2 & \delta_2^2 & \vdots \\ \vdots & & \ddots \\ \cdots & & \end{pmatrix}$$

$$x^{comb.} = \omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_n x_n$$

$$\omega_1 + \omega_2 + \omega_3 \cdots + \omega_n = 1$$

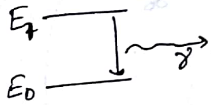
δ_x^{comb} is minimized. $\leftarrow$ Choose weights ω_i's s.t.

$$\delta_x^{comb\,2} = (\omega_1\ \omega_2 \cdots \omega_n) \begin{pmatrix} & & \\ & C & \\ & & \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \vdots \\ \omega_n \end{pmatrix}$$

$$\omega_i = \frac{\sum_{j=1}^{n} V_{ij}}{\sum_{i,j=1}^{N} V_{ij}}$$

HW. Prove this.

E.g.



$E_\gamma = \Delta E$ $\qquad E_\gamma = \Delta E + E_0$

We are measuring ΔE $\rightarrow$ in our experiment

E_0 $\rightarrow$ Take the value from a data table.

$$E_0 = E_0 \pm \delta_0$$

G.1 $(\Delta E)_1 \pm \delta_1$ $\rightarrow E_\gamma^{(1)} = E_0 \pm (\Delta E)_1$ $\qquad \Delta_1 = \sqrt{\delta_0^2 + \delta_1^2}$

G.2 $(\Delta E)_2 \pm \delta_2$ $\rightarrow E_\gamma^{(2)} = E_0 \pm (\Delta E)_2$ $\qquad \Delta_2 = \sqrt{\delta_0^2 + \delta_2^2}$

G.3 $(\Delta E)_3 \pm \delta_3$

$\vdots$

G.N $(\Delta E)_n \pm \delta_n$ $\rightarrow E_\gamma^{(n)} = E_0 + (\Delta E)_n$ $\qquad \Delta_n = \sqrt{\delta_0^2 + \delta_n^2}$

How to combine all measurements b/w the groups?

Δ_{12}^2 = correlation / covar. b/w Groups 1 & 2.

$$E\left[\left(E_0 + (\Delta E)_1 - E_1^{true}\right)\left(E_0 + (\Delta E)_2 - E_\gamma^{true}\right)\right]$$

$$\Rightarrow E\left[\left(E_0 - E_0^{true}\right)\left(E_0 - E_0^{true}\right)\right] = \delta_0^2.$$

$\left(E_1^{true} = E_0^{true} + \Delta E^{true}\right)$

$$C = \begin{pmatrix} \Delta_1^2 & \delta_0^2 & \delta_0^2 & \cdots & \delta_0^2 \\ \delta_0^2 & \Delta_2^2 & & & \vdots \\ \delta_0^2 & & \ddots & & \\ \vdots & & & & \Delta_n^2 \\ \cdots & & & & \end{pmatrix}$$

$$C = \begin{pmatrix} \delta_0^2+\delta_1^2 & \delta_0^2 & \delta_0^2 & \cdots & \delta_0^2 \\ \delta_0^2 & \delta_0^2+\delta_2^2 & & & \vdots \\ \delta_0^2 & \cdots & \delta_0^2+\delta_3^2 & & \vdots \\ \vdots & & & & \\ \delta_0^2 & & \cdots & & \delta_0^2+\delta_n^2 \end{pmatrix}$$

$$C = \delta_0^2 \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} + \begin{pmatrix} \delta_1^2 & 0 & \cdots & 0 \\ 0 & \delta_2^2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \delta_n^2 \end{pmatrix}$$

$\Rightarrow$ $x_{comb} = \sum_{i=1}^{n} w_i x_i$

$$\delta_{comb}^2 = (w_1 \cdots w_n)\, C \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$$

$C = \underset{\downarrow C_1}{\text{diag}} + \underset{\downarrow C_2}{100\% \text{ correlated}}$

$$\delta_{comb}^2 = \sum_{i=1}^{n} w_i^2 C^1_{ii} + w_i w_j C^{(2)}_{ij}$$

$$= \sum_{i=1}^{n} w_i^2 C^1_{ii} + \left(\sum_{i,j} w_i w_j\right) \delta_0^2$$

$$= \sum_{i=1}^{n} w_i^2 \delta_i^2 + \left(\sum_{i,j} w_i w_j\right) \delta_0^2$$

$$= \sum_{i=1}^{n} w_i^2 \delta_i^2 + \left(\sum w_i\right)^2 \delta_0^2$$

$\sum w_i = 1$

$$\therefore \boxed{\delta_{comb}^2 = \sum_{i=1}^{n} w_i^2 \delta_i^2 + \delta_0^2}$$

↳ correlation only introduces a constant.

$\therefore$ choice for weights to minimise δ_{comb} remains the same.

$$\delta_{comb}^2 = \left(\frac{1}{\sum_{i=1}^{n} \frac{1}{\delta_i^2}}\right) + \delta_0^2$$

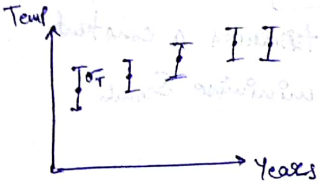
when δ_is are the same,

$$\delta_{comb}^2 = \underbrace{\left(\frac{\delta}{n}\right)^2}_{\text{stat.}} + \underbrace{\boxed{\delta_0^2}}_{\text{sys.}} \longrightarrow \text{min. limit.}$$

9/10/18

Statistical Inference

Experiments : $x_1 \pm \delta x_1$, $x_2 \pm \delta x_2$ $\longrightarrow$ Inference?



(Error-bar plot.)

} $\rightarrow$ Cannot say that the temp is rising if the errors are large

$\rightarrow$ More the uncertainties, less sure we are of the data.

- Quantitative $\boxed{\text{scores}}$ to theoritical hypothesis to compare which hypothesis best explains the dataset.
- How do we assign meaning to the quantitative scores.

$H_0 \rightarrow T\uparrow$
$H_1 \rightarrow T$ const.
$H_2 \rightarrow T\downarrow$ } Quantitative Hypothesis.

$\hookrightarrow H_0 \Rightarrow T(\text{years}) = m(\text{year}) + c.$

$H(m, c) \rightarrow$ Family of hypotheses.
$\underbrace{m, c}_{\text{parameters}}$

(has to be predictive)

Suppose we fix $m_0, c_0 \rightarrow$ gives a specific hypothesis.

$\underbrace{H(m,c) = \{H_1, H_2 \cdots H_n\}}$

Set of all theories we're testing against the data.

- Set of hypotheses : $\{H_1, H_2, H_3 \cdots H_n\}$ (each of which is quantitative)
- Assigning Scores : (ad-hoc procedure)

$\underbrace{P_1(H_1), P_2(H_2) \cdots P_n(H_n)}$

measure of our bias. (depends on what we believe)

$\hookrightarrow$ a-priori assignment of probability $\Leftrightarrow$ (bias set up quantitatively before we get the experimental data).

Given $\{P_i(H_i)\}$, we now talk about $\boxed{P(\text{data} \mid H_i)}$ (in principle, calculatable)

Let H_1, for eg, $H_1 : 1°C \left[\frac{\text{year} - 1800}{100}\right] + 30°C$.

$\boxed{P(\text{data} \mid H_P)} \Leftarrow$ Prediction from a particular hypothesis.
(well defined cal^n)

Eg. Coin toss (H, T) Hypo(p).

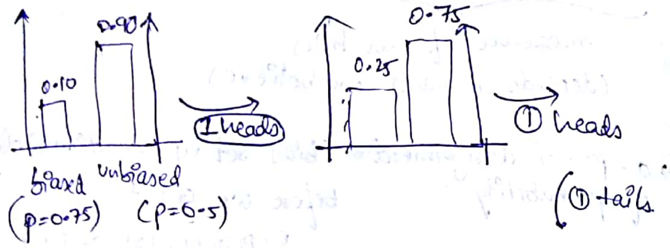
$p \rightarrow$ H
$1-p \rightarrow$ T

$H_1 : p \rightarrow 1 \quad 1-p \rightarrow 0$ (super biased)
$H_2 : p = 1/2 , 1-p = 1/2$ (unbiased)
$H_3 : p = 0.75, 1-p = 0.25$ (biased)

$P(\text{heads} \mid H_3) = 0.75$
$P(\text{tails} \mid H_3) = 0.25$

100 coins, 2 are biased $\Rightarrow P(H_2) = 0.98$
$\Rightarrow P(H_3) = 0.02$

$\{P_i(H_i)\}$ → not unique (not arbitrary)
→ depends on what the hypothesis has / what apriori knowledge.
→ conditional prob. of the data given to H_i.



Updating your priors

↳ $P(H_i \mid \text{measured data})$
→ what we want. (inference)

Bayes' theorem:
$P(A|B)P(B) = P(B|A)P(A)$

A → H_i is correct
B → certain data was obtained.

$$\underbrace{P(H_i \mid \text{data})}_{\text{a-posteriori probability}} \underbrace{P(\text{data was observed})}_{\text{net probability of observing the data}} = \underbrace{P(\text{data observed} \mid H_i)}_{\text{Calculable probability apriori}} \underbrace{P(H_i)}_{\text{apriori probability (ad-hoc)}}$$

$$P(\text{observed data}) = \sum_{i\,(\text{hypothesis})} P(H_i)\, P(\text{observed data} \mid H_i).$$

$$\therefore \boxed{P(H_i \mid \text{data}) = \frac{P(\text{data}|H_i)\,P(H_i)}{\sum_j P(H_j)P(\text{data}|H_j)}}$$
→ Independent of i. (normalisation)

∴ One possible "score" we can assign to each hypothesis is $P(H_i|\text{data})$. ← Approach 1

Problem: Dependance on priors $P(H_i)$.

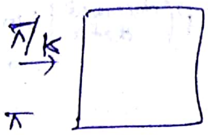
$\boxed{P(\text{data}|H_i)}$ → Reweighting our apriori probabilities.
↓
Likelihoods (associated to a hypothesis given by the data).

$\boxed{L(H_i) = P(\text{data}|H_i)}$ ← Approach 2.

↳ Candidate for being a "score"
↳ A posteriori probability with a uniform prior.

Approach 1 ⇒ Bayesian approach. (Astro Junta)
Approach 2 ⇒ ~~likelihood~~ Frequentist approach. (Particle peeps).

e.g. π / K



$P(\mu|\pi) = 0.02$

$P(\mu|K) = 0.10$.

75% π
25% K.

Data : ① μ detected
② μ not detected.

What we want to infer ?

H_0 : π entered the detector
H_1 : K entered the detector.

$\left.\begin{array}{l} P(H_0) = 0.75 \\ P(H_1) = 0.25 \end{array}\right\}$ natural choice of priors.

Data : Saw a muon (μ).

~~P(μ)~~ $P(H_0|\mu) = \dfrac{P(\mu|\pi)\ P(\pi)}{N} = \dfrac{0.02 \times 0.75}{N}$

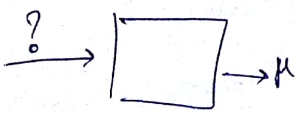
$P(H_1|\mu) = \dfrac{P(\mu|K)\ P(K)}{N} = \dfrac{0.10 \times 0.25}{N}$

Scores

$N = 0.040$.

$P(H_0|\mu) = 0.015/0.040 = 40\%$

$P(H_1|\mu) = 0.025/0.040 = 60\%$.



$H_0(\pi) = P(\mu|\pi)$
$= L(\pi) = \dfrac{0.02}{N}$

$H_1(K) = P(\mu|K)$
$= L(K) = \dfrac{0.10}{N}$

Frequentist Approach

Score $H_i \rightarrow L_{\text{ob data}}(H_i) = P(\text{obs data}|H_i)$

Continuous parameters / Hypotheses

↳ $H(\theta)$
parameter (continuous).

↳ Probability density (denoted by π) $\quad f_{prior}(\theta)d\theta$
$= \pi_{prior}(\theta)d\theta$.

$\pi_{aposteriori} \Rightarrow \pi_\theta(\theta|\text{data})$ ~~= f(data)~~

$= \dfrac{f(\text{data}|\theta) \cdot \pi_{prior}(\theta)}{f(\text{data})}$.

If expected/possible data is continuous ⇒ replace likelihood.

If $\pi(\theta)$ is continuous ⇒ replace $P(H)$ with $f(\theta)$.

---

posterior = $\boxed{\text{Likelihood}}\ \boxed{\text{Prior}}$ / $\boxed{\text{Normalisation}}$

make continuous if hypothesis is continuous.

make continuous if data is continuous

$L(H_i) = f(\text{data}|H_i) = f_x(x|H_i)$

if data is continuous.

THIS IS NOT A PROBABILITY DISTRIBUTION of H.

IT IS A SCORE.

→ We choose the Hypothesis with the highest score as the best explanation of the data.

→ Maximal likelihood estimation

12/10/18
Friday.

• **Statistical Inference** (continued)

$$\underbrace{P(H_i|data)}_{\text{a-posteriori}} = \frac{P(data|H_i)\,P(H_i)}{\sum_j P(data|H_j)\,P(H_j)}$$

$\{H_0, H_1, H_2 \cdots\} \Rightarrow$ ① family of hypothesis

② $H(\theta)$

③ apriori $\rightarrow P(H_i)$
↳ $f(\theta)$

④ $P(H_i) \rightarrow$ biases

⑤ $\underbrace{P(data|H_i)}_{\text{computable}}$

• $P(H_i|data) \rightarrow$ aposteriori $\rightarrow$ Score (consists of) → Calculable/data dependant, Biases.
↓
Bayesian Approach.

• $L(H_i) = P(data|H_i) \rightarrow$ score
↓
Frequentist Approach

↳ proportional to a-posteriori probability assuming equally likely priors.

e.g. $(\theta, \theta+d\theta) \rightarrow (\alpha = \theta^2, \alpha + d\alpha)$

$f(\theta)\,d\theta = g(\alpha)\,d\alpha$.
uniform prob. distribution.

$g(\alpha) = \dfrac{f(\theta)}{\left|\frac{d\alpha}{d\theta}\right|} = 2\theta f(\theta)$.

$\underbrace{g(\alpha) = 2\sqrt{\alpha}.}_{\text{not uniform.}}$

∴ "Equally likely priors" are in terms of a particular parametrisation.
(If we change the parameters of the hypothesis, the priors will also change.)

$\mathcal{L}(H(\theta)) = P(data|H(\theta)) \leftarrow$ if the data is ~~conti~~ discrete but hypothesis is continuous

$\mathcal{L}(H(\theta)) = f(data|H(\theta))$ → prob. density f^n.
↳ when both data and hypothesis are continuous.

• Multiple measurements $x_1, y_1, x_2, z_1 \cdots$

$L[H_i] = P[x_1=3, y_1=2, x_2=4, z_1=10 \mid H_i]$

$= \prod_{\text{measurements}} P[data|H_i]$ (iff measurements are independant)

$\beta \rightarrow$ denotes an independant data measurement,

$$\boxed{L[H_i] = \prod_\beta L_\beta[H_i]}$$

$\log \mathcal{L}[H_i] = \sum_\beta \log \mathcal{L}_\beta (H_i)$.
(all data) ← independant data.

$\therefore$ $\log \mathcal{L}(H_p)$ is also a good option for a score
↳ monotonic → correlates directly with $\mathcal{L}(H_p)$.

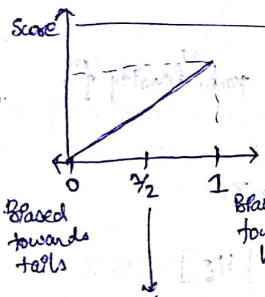
Example : Bernoulli trials. $H \to p$ $T \to 1-p$.

family : $\{H(p)\}$, $0 \le p \le 1$.

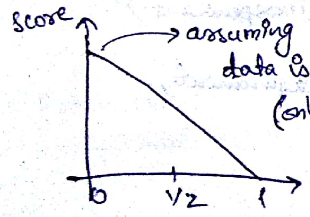
$L[H(p)] = L(p) = P(\text{heads} \mid H(p)) = p$.

we assume that the given data is the result of one trial is heads.

$L_{tails}(H(p)) = L_t(T) = 1-p$.



→ Frequentist approach.
↳ this shows that priors are important.
They reflect the common knowledge and sense we have.
(bayesian approach)

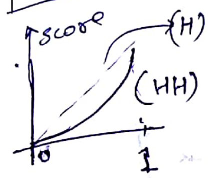


(only one measurement is highly biased and ~~[illegible]~~ insufficient)

two tosses

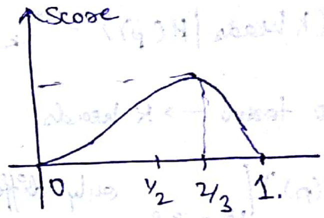
$\boxed{HH} \longrightarrow L(p) = p^2$
$\boxed{HT} \longrightarrow L(p) = p(1-p)$
$\boxed{TH} \longrightarrow L(p) = (1-p)p$.
$\boxed{TT} \longrightarrow L(p) = (1-p)^2$.

(product because independant measurements)



three tosses

$\boxed{HHT} \longrightarrow p^2(1-p)$.

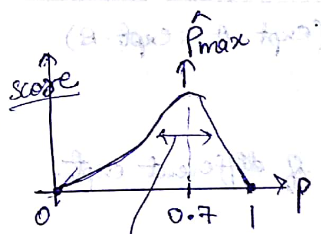


100 tosses

70 heads ; 30 tails.
↳ $p = 7/10$ claimed to have the peak.

$p^{70}(1-p)^{30}$

$\frac{\partial L}{\partial p} = 70p^{69}(1-p)^{30} - 30p^{70}(1-p)^{29} = 0$



$70p^{69}(1-p)^{30} = 30\,p^{70}(1-p)^{29}$

$70(1-p) = 30p$.

$\frac{1}{p} - 1 = \frac{3}{7}$, $\Rightarrow \frac{1}{p} = \frac{10}{7}$

$\underline{p = 0.7}$.

It would narrow down with increased number of measurements.

⇒ Maximum Likelihood Estimator (of p.)
↳ val. of p with max. score ($\hat{p}_{max}$)

Example: We could' be used binomial distribution to define our likelihoods. (for fixed number of tosses.) ↳ different definition of $H(p)$ and the scores.

$$L[H(p)] = {}^{100}C_k \, p^{70} (1-p)^{30}.$$

$$P(k \text{ heads} \mid H(p)) = {}^{100}C_k \, p^k (1-p)^{100-k}$$

100 tosses → k heads

$L(H(p))\big|_{\text{binomial}}$ only differs by the normalisation factor; irrelevant addition to the score. (equivalent scores).

$\ln L_A(H(\theta))$; $\ln L_B(H|\theta))$ (expt. A & expt. B)

$$\ln L_{AB}(\theta) = \underbrace{\ln L_A(\theta) + \ln L_B(\theta)}$$

simple combination of different expt.

→ MLE generalisation.

→ $\hat{\theta} \Rightarrow$ best "fit" hypothesis.

Scanned by CamScanner

16/10/18
Tuesday

• Maximum Likelihood Estimation:

set of Hypotheses $\{H|\Theta)\}$ → e.g. $P(H)=p$ $P(T)=1-p$.
$H(p)$

Θ → parameter.

$\mathcal{L}[\Theta] = \text{Prob}[\text{data}|\Theta]$ (with a given data from the experiment)

→ Post experimental.

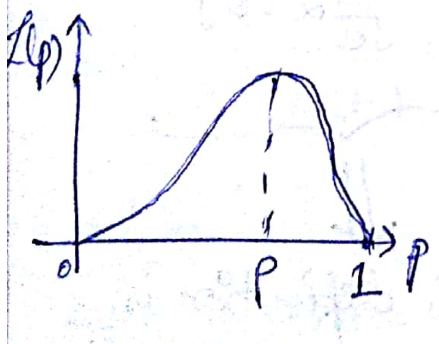
→ Score, not a probability.

HHTHT

$\mathcal{L}[\Theta] = p^3(1-p)^2$ → if order matters

$= {}^5C_3 p^3(1-p)^2$ → if order of experimental outcome does not matter.

As long as the experiments are independant, the likelihood functions differ only by a constant for ordering, and hence does not affect our inference from the scores.



For large amount of independant data,

$\mathcal{L}(p) \simeq \exp\left[\frac{-1}{2\sigma^2}(p-\hat{p})\right]^2$

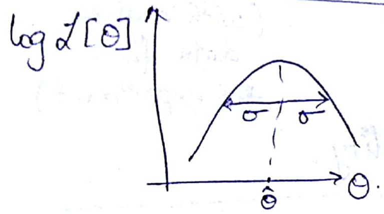
approaches a (gaussian).

$\mathcal{L}[\theta \,|\, data] \equiv Prob[data \,|\, \theta]$

$\longrightarrow \# \, e^{-\frac{1}{2}\frac{(\theta-\hat{\theta})^2}{\sigma^2}}$

for large amount of independant data.

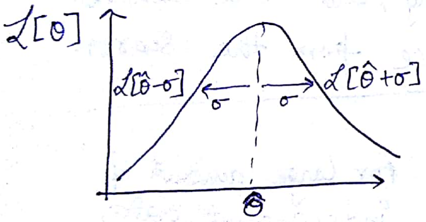
not normalised for θ!



(for large amount of independant data

$$\log \mathcal{L}[\theta] \sim \frac{1}{2}\left(\frac{\theta-\hat{\theta}}{\sigma}\right)^2 + \underset{\text{independent of } \theta.}{const.}$$

$\boxed{\hat{\theta} \Rightarrow \text{maximum likelihood estimator. (MLE)}}$



σ (called the error on $\hat{\theta}$)

$$\underbrace{\mathcal{L}[\hat{\theta}+\sigma] = \mathcal{L}[\hat{\theta}-\sigma] = \frac{1}{\sqrt{e}}\mathcal{L}[\hat{\theta}]}$$

$\Downarrow$

This definition is independant of the form of $L(\theta)$; does not have to be a gaussian.

• Mathematical, General Definitions $\Rightarrow$

$\mathcal{L}[\theta]$

① MLE $\hat{\theta}$ $\quad \boxed{\left.\frac{\partial L}{\partial \theta}\right|_{\hat{\theta}} = 0}$

② error σ (Asymmetric error.)

$$\boxed{\begin{aligned} \mathcal{L}[\hat{\theta}+\sigma_u] &= \frac{1}{\sqrt{e}}\mathcal{L}[\hat{\theta}] \\ \mathcal{L}[\hat{\theta}-\sigma_d] &= \frac{1}{\sqrt{e}}\mathcal{L}[\hat{\theta}] \end{aligned}}$$

$(\hat{\theta}-\sigma_d, \hat{\theta}+\sigma_u) \rightarrow \hat{\theta}$ lies in this interval.

($\sigma_d \neq \sigma_u \Rightarrow$ the likelihood function is asymmetric)

what do $\hat{\theta}, \sigma$ depend on?

$\mathcal{L}_{data}[\theta] = P(data \,|\, \theta)$

$\boxed{\hat{\theta}, \sigma \rightarrow \text{depend on the data (only)}}$

E.g. data = $h_1, h_2, h_3, \dots \dots h_{20}$. (heights)

$\{H[h_{av}]\}$ set of hypotheses for the av. ~~hypoth~~ height.

$$H[h_{av}] = H[\theta] = f(h \,|\, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-1}{2}\frac{(h-\theta)^2}{\sigma^2}\right]$$

$\sigma \rightarrow 10$ cms.

(PTO)

For the moment, we're only interested in the variation with mean. We dont consider fluctuations in σ in this example.

$$\boxed{\mathcal{L}(\theta) = \prod_{i=1}^{20} f(h_i|\theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{20} \exp\left[\frac{-1}{2\sigma^2}[(h_1-\theta)^2 + (h_2-\theta)^2 + \dots (h_{20}-\theta)^2]\right]}$$

• maximizing $L(\theta)$ = minimize $(h_1-\theta)^2 + (h_2-\theta)^2 + \dots (h_{20}-\theta)^2$

maximize $L(\theta)$ wrt θ $\equiv$ minimize $\boxed{\sum_{i=1}^{20}\left(\frac{h_i-\theta}{\sigma}\right)^2}$

$(= -2\ln\mathcal{L})$

$$\frac{\partial(-2\ln\mathcal{L})}{\partial\theta} = \sum_{i=1}^{20} \frac{2(h_i-\theta)}{\sigma^2} = 0.$$

$$\sum_{i=1}^{20} h_i = 20\hat{\theta}. \quad \therefore \boxed{\hat{\theta} = \frac{1}{20}\sum_{i=1}^{20} h_i}$$

$\hat{\theta} = \langle h \rangle \Rightarrow$ exactly what we expect the maximum likelihood for the hav to be.

$$\boxed{\chi^2 \equiv -2\ln L} = \sum_{i=1}^{N}\left(\frac{h_i-\theta}{\sigma}\right)^2 . \qquad \hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} h_i.$$

↳ We minimize χ^2

$$\mathcal{L}[\hat{\theta}+\sigma] = \frac{1}{\sqrt{e}}\mathcal{L}[\hat{\theta}]$$

$$\chi^2[\hat{\theta}+\sigma] = \chi^2[\hat{\theta}] + 1$$

$\therefore \delta$ s.t. $\chi^2[\hat{\theta}+\delta] - \chi^2[\hat{\theta}] = 1$ gives us the error in $\hat{\theta} = \sigma = \delta_{special}$.

$$\chi^2[\hat{\theta}+\delta] = \frac{\sum_{i=1}^{N} (h_i - \langle h \rangle - \delta)^2}{\sigma^2}$$

$$\chi^2[\hat{\theta}] + 1 = \frac{\sum_{i=1}^{N} (h_i - \langle h \rangle)^2}{\sigma^2} + 1.$$

$$\frac{N\delta^2 - 2\delta\underbrace{(h_i - \langle h \rangle)}_{\to 0}}{\sigma^2} = 1.$$

$$\frac{N\delta^2}{\sigma^2} = 1. \qquad \therefore \boxed{\delta = \frac{\sigma}{\sqrt{N}}}.$$

$$\left\{\hat{\theta} = \langle h \rangle, \ \delta = \frac{\sigma}{\sqrt{N}}\right\}$$

$H(\theta_1 \dots \theta_m)$

$$\mathcal{L}_{data}(\theta_1, \theta_2 \dots \theta_m) \equiv f(data|\theta_1 \dots \theta_m)$$

m equations (simultaneously) $\boxed{\left.\frac{\partial\mathcal{L}}{\partial\theta_i}\right|_{\theta_i=\hat{\theta}_i} = 0.}$

$H(\hat{\theta}_1, \hat{\theta}_2, \cdots \hat{\theta}_m) \Rightarrow$ best (maximum likelihood estimator).

- For errors, we have error ellipsoids.

$\theta \to (\theta - \sigma, \theta + \sigma)$

~~best~~ $\exp\left[\frac{-1}{2} \sum_i \frac{(\theta_i - \hat{\theta}_i)^2}{\sigma_i^2}\right]$

$\bar{\theta} = (\theta_i - \hat{\theta}_i)$

$\mathcal{L}[\hat{\theta}] \longrightarrow -\exp\left[\frac{-1}{2} \bar{\theta}^T V \bar{\theta}\right]$

$$V = \begin{pmatrix} \frac{1}{\sigma_1^2} & & & \\ & \frac{1}{\sigma_2^2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_m^2} \end{pmatrix}$$ (for large data)

We define errors such that

$$\boxed{\mathcal{L}[\vec{\hat{\theta}} + \vec{\delta}] = e^{-m/2} \mathcal{L}[\hat{\theta}]}$$ $\theta_i \in (\hat{\theta}_i - \sigma_i, \hat{\theta}_i + \sigma_i)$

↳ gives a surface of $\vec{\delta}$ values.

- ↳ best fit parameters and the uncertainty in those best fit parameters.

- Simplified error estimate

$\mathcal{L}[\theta] = e^{-\frac{1}{2}\left(\frac{\theta - \hat{\theta}}{\sigma}\right)^2}$

$\chi^2 = -2 \ln \mathcal{L} = \left(\frac{\theta - \hat{\theta}}{\sigma}\right)^2$ $\qquad \frac{\partial^2 \chi^2}{\partial \theta^2} = \frac{2}{\sigma^2}$

$$\boxed{\sigma^{-2} \simeq 2\left(\frac{\partial^2 \chi^2}{\partial \theta^2}\right)^{-1} = \left(\frac{\partial^2 \ln L}{\partial \theta^2}\right)^{-1}}$$ $\Rightarrow$ For 1 variable.

For multiple variables,

$$V_{ij} = \boxed{F_{ij} = \frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j}}\Bigg|_{\theta_k = \hat{\theta}_k}$$

⇓

Fisher information matrix.

$\vec{\delta}^T V_{ij} \vec{\delta} = m.$

$$\mathcal{L}[\theta] = \exp\left[-\frac{1}{2} \underbrace{(\vec{\theta - \hat{\theta}})^T V (\vec{\theta - \hat{\theta}})}_{m}\right]$$

Example: $h_1, h_2, \cdots h_{20}$.

$H(\theta, \sigma)$

$\theta \to$ parameter for heights

$\sigma^2 \to$ variance

$H(\theta, \sigma) \Rightarrow f(h|\theta, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-1}{2} \frac{(h - \theta)^2}{\sigma^2}\right]$

Claim: $\hat{\theta}_{guess} = h_{av} = \frac{\sum_{i=1}^{N} h_i}{20}$; $\hat{\sigma} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (h_i - \bar{h})^2}$

Using MLE, $L(\theta,\sigma) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{20} \exp\left[-\frac{1}{2}\left(\frac{h_i-\theta}{\sigma}\right)^2\right]$

① $\frac{\partial L}{\partial \theta} = 0$; $\frac{\partial L}{\partial \sigma} = 0$. $(\hat{\theta}, \hat{\sigma})$.

② Error ellipsoids. (Analytically, we just try to find the equations).
HW.

$$-2\ln L(\theta,\sigma) = \sum_{i=1}^{20}\left(\frac{h_i-\theta}{\sigma}\right)^2 + -2\ln\left[\frac{1}{\sqrt{2\pi\sigma^2}}\right]^{20}$$

$$\frac{\partial \chi^2}{\partial \theta} = 0 \Rightarrow \boxed{\hat{\theta} = \frac{1}{20}\sum_{i=1}^{20} h_i = \bar{h}}$$

$$\frac{\partial \chi^2}{\partial \sigma} = 0 \Rightarrow -2\sum_{i=1}^{20}\frac{(h_i-\theta)^2}{\sigma^3} + 40\frac{1}{\sigma}$$

$$\Rightarrow \sigma^2 = \frac{1}{20}\sum_{i=1}^{20}(h_i-\theta)^2.$$

$$\boxed{\hat{\sigma}^2_{MLE} = \frac{1}{20}\sum_{i=1}^{20}(h_i-\bar{h})^2}$$

Biased estimator of the actual σ (underlying distribution)

23/10/18
Tuesday

1 tutorial.
2 Assignments.
End Sem exams.

Parameter Estimation
↳ using MLE.

$H = \{H_1, \ldots H_m\}$ or $H(\theta) \Rightarrow$ Set of hypotheses.

$d = \{d_1, d_2 \ldots d_n\} \Rightarrow$ Data obtained.

$$\mathcal{L}_d[\theta] = \mathcal{L}_d[H(\theta)] = \text{Prob}[\text{data}\,|\,H(\theta)].$$

Likelihood maximization. $\left.\frac{\partial \mathcal{L}}{\partial \theta}\right|_{\hat{\theta}} = 0.$

$\hat{\theta} \rightarrow$ maximum likelihood ~~estimo~~

Errors: $\mathcal{L} \rightarrow$ drops by $\frac{1}{\sqrt{e}}\mathcal{L}_{max}$. for 1 parameter.

m parameters $\Rightarrow$ drops by $\frac{1}{e^{m/2}}\mathcal{L}_{max}$.

$\mathcal{L}(\theta) = \mathcal{L}[\hat{\theta}]\, e^{-m/2}$
↳ $\theta = \hat{\theta} \pm \sigma$.

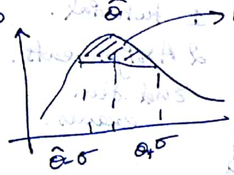
For large data, $\mathcal{L} \longrightarrow e^{-\frac{1}{2}\left(\frac{\theta-\hat{\theta}}{\sigma}\right)^2}$

$$\cancel{\left.\frac{\partial^2 \mathcal{L}}{\partial \theta^2}\right|_{\hat{\theta}} = \frac{1}{\sigma^2}}$$

$$\chi^2 = -2\ln\mathcal{L}.$$

$$\frac{\partial^2 \chi^2}{\partial \theta^2} = \frac{2}{\sigma^2}$$

$$\sigma^2 = \left(\frac{\partial^2 \ln L}{\partial \theta^2}\right)^{-1} \approx 2\left(\frac{\partial^2 \chi^2}{\partial \theta^2}\right)^{-1}.$$



→ Best agreement with data.

e.g. toss a coin 100 times ; 70h , 30 t

$p = \theta$; $p(H) = p$; $p(T) = 1-p$.

$\hat{\theta} = \frac{7}{10}$ (by observation)

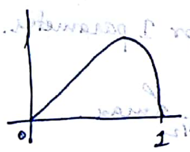
For MLE, $\mathcal{L}[p] = p^{70}(1-p)^{30}$.

$$\frac{\partial L}{\partial p} = 70\,p^{69}(1-p)^{30} - 30\,p^{70}(1-p)^{29} = 0$$

$$p^{69}(1-p)^{29}\,[\,70(1-p) - 30\,p\,] = 0$$

$$70 = 100p.$$

$$p = 0.7.$$



(not a pure gaussian, but if we look very close to the center $p = 0.7$, it is approx. gaussian).

$(\hat{p}-\sigma, \hat{p}+\sigma)$

$$\sigma^2 = -\left(\frac{\partial^2 \ln\mathcal{L}}{\partial p^2}\right)^{-1}\Bigg|_{\hat{p}}$$

$$\ln\mathcal{L} = 70\ln p + 30\ln(1-p)$$

$$\frac{\partial \ln\mathcal{L}}{\partial p} = \frac{70}{p} - \frac{30}{1-p}$$

$$\frac{\partial^2 \ln\mathcal{L}}{\partial p^2} = \frac{-70}{p^2} - \frac{30}{(1-p)^2}$$

$$\left(-\frac{\partial^2 \ln\mathcal{L}}{\partial p^2}\right)^{-1} = \frac{1}{\frac{70}{p^2} + \frac{30}{(1-p)^2}}$$

$$\sigma^2 = \frac{1}{\frac{100^2}{70} + \frac{100^2}{30}}$$

$$= \frac{70\cdot 30}{100^2\,(100)}$$

$$= 21 \times 10^{-4}$$

$$\underline{\sigma^2 = 2.1 \times 10^{-3}}.$$

$$\sigma = \sqrt{21} \times 10^{-2}$$

$$\approx 0.045$$

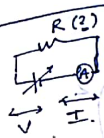
$$(0.7-0.045,\ 0.7+0.045)$$

$$\approx (0.7-0.05,\ 0.7+0.05)$$

$$\underline{(0.65,\ 0.75)}$$

width will narrow

$\sigma \propto \frac{1}{\sqrt{N}}$

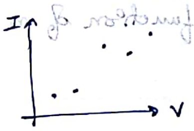


V → input (assumed to have infinite precision)

I → measurement (assumed to have finite precision $(I \pm \delta_I)$)

$V = IR$ ⟸ Hypothesis family.

$I = \frac{V}{R} = mV.$ $\quad m = \frac{1}{R}$ (slope of the line)

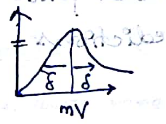


(data)

$$\mathcal{L}\left[H\left(m=\tfrac{1}{R}\right)\right] = \text{prob}\left(\text{data} \mid H\left(m=\tfrac{1}{R}\right)\right).$$

$$= f(\text{data} \mid m).$$

(assume that this distribution is gaussian).



when voltage is fixed at V_1.

~~f(data | m)~~ $\quad I_1$ → data

m → parameter

δ → independent of process (Ammeter dep.)

$I_1^{obs.}$ (for one data point)

$$\boxed{f(\text{data} \mid m) = \frac{1}{\sqrt{2\pi\delta^2}} \exp\left(-\frac{1}{2}\frac{\left(I_1^{obs} - I_1^{pred}(m)\right)^2}{\delta^2}\right)}$$

Ammeter

$I_1^{obs} = $ data $\quad I_1^{pred} = mV_1$ → hypothesis

$$\mathcal{L}[m] = \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi\delta^2}} \exp\left[-\frac{1}{2}\left[\frac{I_j^{obs} - I_j^{pred}(m)}{\delta}\right]^2\right]$$

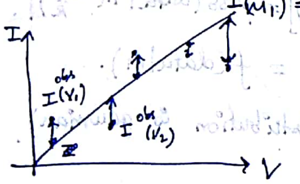
$N \to$ no. of data points.

$\delta \to$ ammeter error (independent of measurement)

$$\chi^2 = -2\log\mathcal{L}[m] = \sum_{j=1}^{N}\left(\frac{I_j^{obs} - I_j^{pred}(m)}{\delta}\right)^2$$

$$\chi^2 = \frac{1}{\delta^2}\left[\sum_{j=1}^{N}\left(I_j^{obs} - I_j^{pred}(m)\right)^2\right]$$

- Minimize χ^2 or maximize $\mathcal{L}$ as a function of m to find $\hat{m}$.



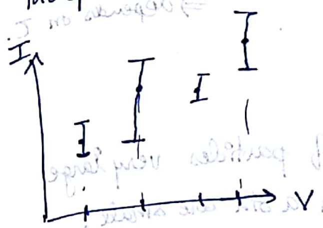
essentially, we're minimizing the square distance for the observed points from the predictions.

$\Rightarrow$ **Least square fit**

$$\frac{\partial\chi^2}{\partial m} = 0 \text{ for } \hat{m}.$$

for errors, $\boxed{\sigma^2 = \frac{1}{2}\left(\frac{\partial^2\chi^2}{\partial m^2}\right)^{-1}_{\hat{m}}.}$

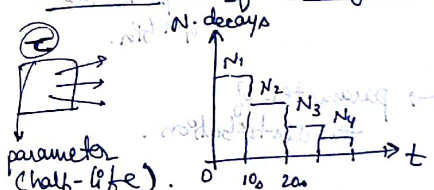
Now, we may a case where δ_j is not independent of the voltage.



$$\chi^2 = \sum_{j=1}^{N}\left(\frac{I_j^{obs} - I_j^{pred}(m)}{\delta_j}\right)^2$$

If certain points are precisely measured while others have larger errors, best fit is driven by the precise measurements.

- Likelihoods of histograms



- Instead of doing the complicated likelihood function of when the particle decays, we divide the time into bins.

- We throw away some information.

(from full data $\longrightarrow$ binned data / coarse grained data)

$\tau \longrightarrow$ parameter for the family of Hypotheses.

we predict $P(K_1 \mid 0 \le t \le 10s) \Rightarrow$ prob. of getting K_1 decays b/w 0–10s.

$P_1(K_1|\tau)$ $\qquad P(K_2 \mid 10 \le t \le 20s)$ $\Rightarrow$ depends on τ.

$\therefore$ $P_2(K_2|\tau)$

$\vdots$

- If no. of bins are large, no. of particles very large and the prob. of landing in a bin are small;

$\Rightarrow$ large amount of data ①

$\Rightarrow$ for small bernoulli probabilities of data to fall in a bin ②

$\Rightarrow$ large no. of bins ③

Then,

$$\underset{\substack{\uparrow \\ n^{th}\text{ bin}}}{P_n}(K_n) = \text{poisson } (\lambda_n)$$

$P_n(\underline{K_n}) \rightarrow$ prob. of getting K_n decays in the n^{th} bin.

$\lambda_n(\tau) \rightarrow$ parameter of the distribution.

$$\left\{ \begin{array}{l} N_{decays}^{(t_1,t_2)} = N_0\left(e^{-t_2/\tau} - e^{-t_1/\tau}\right) \\ N_{decays} = N_0\left(1-e^{-t/\tau}\right) \qquad \frac{dN}{dt} = \frac{-N}{\tau} \\ N = N_0\, e^{-t/\tau} \end{array} \right\}$$

data $\rightarrow$ N bins.

$K_1^{obs}, \ldots\ldots K_N^{obs}$.

$$\text{hyp} \rightarrow P(K_i) = \frac{[\lambda_i(\tau)]^{K_i}}{K_i!} e^{-\lambda_i(\tau)}.$$

$$\mathcal{L}[\tau] = \prod_{i=1}^{N} P(K_i^{obs} \mid \lambda_i(\tau))$$

$$= \prod_{i=1}^{N} \frac{(\lambda_i(\tau))^{K_i^{obs}}}{(K_i^{obs})!} e^{-\lambda_i(\tau)}.$$

$$\therefore \quad \log \mathcal{L} = \sum_{i=1}^{N} \left(K_i^{obs} \log(\lambda_i(\tau)) - \lambda_i(\tau) - \log(K_i^{obs}!)\right)$$

To maximize $\mathcal{L}$,

maximize $\displaystyle\sum_{i=1}^{N} \left[K_i^{obs} \log(\lambda_i(\tau)) - \lambda_i(\tau)\right]$

- Approximate predictions : $\frac{dN}{dt} = -\frac{N}{\tau} \approx \frac{\Delta N}{\Delta t}$.

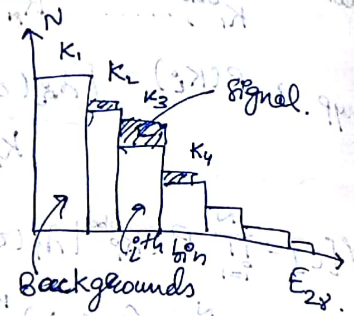
$$\therefore \quad \underbrace{N^i_{decays} \approx \frac{\Delta t}{\tau} N(t,\tau)}_{\text{for the average distribution in the } i^{th} \text{ bin.}}$$

$N(t) = N_0(1-e^{-t/\tau})$.

$\lambda_n(\tau) \rightarrow$ mean no. of decays in the n^{th} bin as a function of $\tau \Rightarrow N^i_{decays} = \frac{\Delta t}{\tau} \cdot N(t,\tau)$

Histograms with background

$H \to \gamma\gamma$ [Background due to other processes that result in 2 photons]



Prediction

$\Rightarrow b_i + s_i(m_h, \alpha \ldots)$

b_i: background distribution with Higgs boson

s_i: signal due to higgs boson (depends on m_h when $m_h \approx E_{2\gamma}$)

observations $= \{K_1, K_2, \ldots K_N\}$

$$\mathcal{L}[m_h, \alpha] = P(K_i \mid m_h, \alpha)$$

we assume this distribution to be poisson around $b_i + s_i(m_h, \alpha)$

$$\therefore \mathcal{L}[m_h, \alpha] = \prod_{i=1}^{N} \frac{[b_i + s_i(m_h, \alpha)]^{K_i} e^{-b_i - s_i(m_h,\alpha)}}{K_i!}$$

$$\log \mathcal{L} = \sum_{i=1}^{N} K_i \log[b_i + s_i(m_h, \alpha)] - b_i - s_i(m_h, \alpha) - \log K_i!$$

Maximize wrt both parameters.

30/10/18

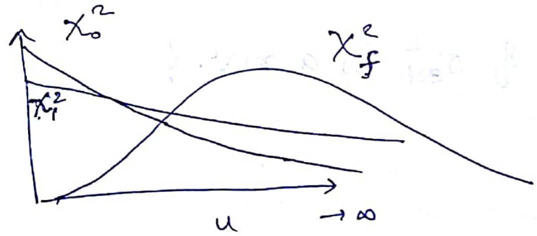
- χ^2_f distributions

↳ N independant gaussian RVs : μ, σ.

$$u = \sum_{i=1}^{N} \left(\frac{x_i - \mu_i}{\sigma_\mu}\right)^2$$

u has a χ^2_f distribution

$$f(u) = \frac{1}{2^{f/2}\,\Gamma(f/2)}\, u^{f/2 - 1} e^{-u/2}$$



$E[u] = f$; $\text{var}[u] = 2f$.

- Important for Hypothesis testing.

Sample variances:

$f(x) \rightarrow$ gaussian $\qquad \mu^{est}, \sigma_{est}^2$

$x_1 \dots x_N$ IID

$$\mu^{est} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma_{est}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_{est})^2.$$

• Distribution of μ^{est} as a r.v.?

$\Downarrow$

Gaussian (true for finite N also; if x_i are gaussian).

• Distribution of σ_{est}^2 as a r.v.?

$\Downarrow$

$\propto \boxed{\chi^2_{N-1}}$

$$\sigma_{est}^2 = \underbrace{\frac{\sigma^2}{N-1}}_{\text{Constant scaling.}} \underbrace{\left[\sum_{i=1}^{N}\frac{(x_i - \mu_{est})^2}{\sigma^2}\right]}_{\chi^2_{N-1}\text{ distributed quantity}}$$

~~quant~~

$$E[\sigma_{est}^2] = \sigma^2.$$

• What is the error on the error?

$\Rightarrow$ S.D. of σ_{est}^2 or $\sqrt{var[\sigma_{est}^2]}$.

$\therefore$ Expected error on the error $= \sqrt{var[\sigma_{est}^2]}$

$$SD = \sqrt{var[\sigma_{est}^2]} = \sqrt{\left(\frac{\sigma^2}{N-1}\right)^2 \cdot 2(N-1)}$$

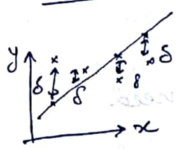
$$= \frac{\sigma^2}{\sqrt{N-1}}\sqrt{2}$$

$$\therefore \sigma_{est}^2 \sim \sigma^2 \pm \frac{\sigma^2\sqrt{2}}{\sqrt{N-1}}.$$

(not perfectly symmetric).

• Error on the error also decreases as we take more measurements.

• Distribution of the maximum log likelihood?



$\mathcal{H}(m,c) \rightarrow$ we try to fit the "best fit line" for various hypotheses.

$\mathcal{L}_{data}(m,c)$?

$y_1^{pred} = m x_1 + c$ (from hypothesis)

observed $\Rightarrow y_1^{obs}$. (observation)

Assuming Gaussian distribution about $y_1^{obs} \Rightarrow$

$$\mathcal{L}(m,c) = \prod_{i=1}^{N}\frac{1}{\sqrt{2\pi\delta^2}}\, e^{-\left(\frac{y_i^{obs} - y_i^{pred}(m,c)}{2\delta^2}\right)^2}$$

$N \rightarrow$ no. of data points.

$$\chi^2 = -2\ln \mathcal{L}_{data}(m,c) = \sum_{i=1}^{N} \frac{\left(y_i^{obs} - y_i^{pred}(m,c)\right)^2}{\delta^2}$$

$$\underbrace{-2\ln\left[\cancel{\left(\frac{1}{2\pi\delta^2}\right)^{N/2}}\right]}_{\text{independant of } (m,c)}$$

so, ignore.

But $\underbrace{\sum_{i=1}^{N}\left(\frac{y_i^{obs} - y_i^{pred}(m,c)}{\delta^2}\right)}_{\text{square distance b/w observation and hypothesis.}}$

$\therefore$ We minimize χ^2 or maximize $\mathcal{L}$.

- N = No. of data points.
  p = No. of parameters.

To have accurate hypothesis testing, we need maximum freedom.

$f = N - p$.
↳ No. of degrees of freedom.

For $N=2$; $p=2$ $\Rightarrow$ $\chi^2 = 0$ (line passing through both these points)

for $p=2$; $N>2$ $\Rightarrow$ $\chi^2 \neq 0$ (line may not pass through any of the data points).

$$\left.\frac{\partial \chi^2}{\partial m}\right|_{m,c} = 0 \quad ; \quad \left.\frac{\partial \chi^2}{\partial c}\right|_{m,c} = 0.$$

$$\chi^2 = \sum_{i=1}^{N}\left(\frac{y_i^{obs} - (mx_i + c)}{\delta}\right)^2$$

$$\frac{\partial \chi^2}{\partial m} = \sum_{i=1}^{N} 2\left(\frac{y_i^{obs} - (mx_i + c)}{\delta}\right) \cdot -\frac{x_i}{\delta}$$

$$\frac{\partial \chi^2}{\partial c} = 2\sum_{i=1}^{N}\left(\frac{y_i^{obs} - (mx_i + c)}{\delta}\right) \cdot \frac{-1}{\delta}$$

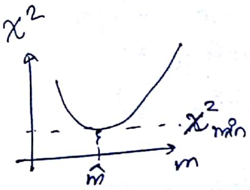
$$\therefore \sum_{i=1}^{N}\left(x_i y_i^{obs} - (mx_i + c)\, x_i\right) = 0$$

$$\sum_{i=1}^{N}\left(y_i^{obs} - mx_i + c.\right) = 0.$$

Solve for $\hat{m}, \hat{c}$. $\Rightarrow$ best describes the data.

$\hat{m}(data)$, $\hat{c}(data)$.

$\chi^2(min) = \chi^2(\hat{m}, \hat{c})$ $\Rightarrow$ Score.
$= \chi^2(data)$ $\Rightarrow$ No free parameters, fixed parameters depend on the data themselves.



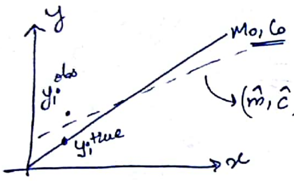
$\underbrace{\chi^2(min)}_{RV.}$

whose distribution depends on the distribution of y_{obs} values.

• Assuming y^i are generated by gaussian fluctuations about the line m_0, c_0, then, $\chi^2_{min}(\hat{m}, \hat{c})$ is χ^2_f distributed.

$$E[\chi^2_{min}] = f = N-p.$$

for a line $\Rightarrow N-2$.



If the distribution about m_0, c_0 is gaussian, y_i^{obs} will be roughly σ away from y_i^{true}.

∴ Av. value of $E\left[\sum_{i=1}^{N} \frac{(y_i^{obs} - y_i^{pred}(m,c))^2}{\delta^2}\right] = N-2$

HYPOTHESIS TESTING

data $\longrightarrow \{H(\theta)\} \longrightarrow \mathcal{L}[\theta] \longrightarrow \hat{\theta}_{MLE}$.

$\hat{\theta}_{MLE} \subseteq \{H(\theta)\}$.

H_0 ↳ Does H_0 describe the data well?
(Even if its the best amongst the assumed family of hypotheses).

⇒ Need a quantitative measure.

$H_0 \rightarrow \epsilon^{data}$

• We say that if $\epsilon^{data} < \epsilon_0$ (threshold),
– H_0 does not describe the data well,
if $\epsilon^{data} > \epsilon_0$ (threshold), we will claim that H_0 describes the data well. (good agreement).

⇒ Hypothesis "test":

• cannot prove the validity of a hypothesis; it can only indicate a problem with it.

• require two choices
↳ ϵ threshold.
↳ ϵ^{data} ⇒ calculate probabilities.

• ∴ Not a procedure of discovery; process of ruling out other hypothesis.

• Alternative hypotheses may not be clearly specified.

- Test should have meaning for an Hypothesis without needing comparison with anything else.
- MUST FIX all parameters of your test before looking at the ~~to~~ data!. (Blind Analysis)

E.g. H_0: Drug A cures headache
N : 100 participants
ϵ : No. of participants who self report curing of headache

Claim $\epsilon < 50$ → drug has no effect
$\epsilon > 50$ → drug may have some effect.

ϵ : Test statistic

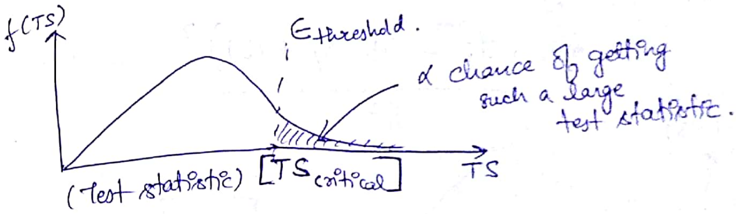
If we fix the threshold after looking at the data → BADDDD.

Sample (take a small set of data)
(Calibration Sample)
↳ use a small set of data to calibrate parameters of testing.
↳ use seperate set of data for actual hypothesis testing.

PTO : TEST PROCEDURE (Before data) ⇒

① H_0 → Null Hypothesis. (Hypothesis to be tested)
② Define a procedure to create a test statistic
$\boxed{t(x_i)} \Rightarrow \epsilon$.
(calibr. sample) data.

③ Look at the distribution of the Test statistic assuming H_0 is the true underlying hypothesis.



④ Define a range of TS, $TS > TS_{critical}$ such that $P[TS > TS_{critical}] = \alpha$.
(Let $\alpha = 0.05$ for example)

⑤ Define your test such that if $TS_{crit} < TS^{obs}$, then H_0 is ruled out at the $(1-\alpha)$ confidence level (e.g. 95% confidence).

- If $TS^{obs} < TS_{crit}$; then H_0 cannot be ruled out at the 95% confidence level
$(1-\alpha)$

α → arbitrary (95%–99% conventionally chosen)

• Define a procedure for collecting data.

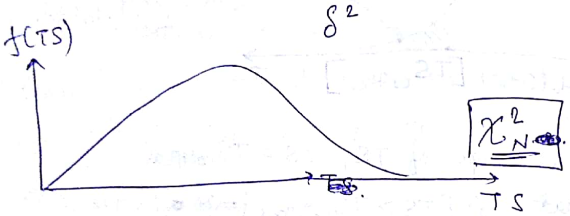
$H_0 : m_0, c_0$

$y = m_0 x + c$

defined a procedure for collecting data $\Rightarrow \{x_1, x_2, \ldots x_N\}$

data $\Rightarrow \{y_1^{obs}, y_2^{obs} \ldots y_N^{obs}\}$

$$TS \equiv \chi^2 = \frac{\sum_{i=1}^{N} \left(y_i^{obs} - (m_0 x_i + c_0)\right)^2}{\delta^2}$$



02/11/18

↳ Goodness of fit test.

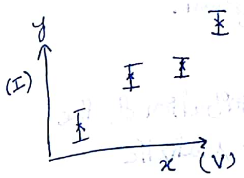
H_0 (data)

Q: Is H_0 consistent with the observed data?

$H_0 : R_0 = 670\Omega$

$V = IR$.

$I = \left(\frac{1}{R_0}\right) V$



$$f(I_{mean}(V)) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(I_{obs} - I_{pred})^2}{2\delta^2}\right]$$

① Define TS:

Test Statistic (TS)(data) $= -2 \ln \mathcal{L}$

$= \chi^2$ (data)

D → no. of data points.

$$= \sum_{a=1}^{D} \frac{(I^a_{obs} - I^a_{pred})^2}{\delta^2}$$

observations

H_0.

$$= \sum_a \frac{\left(y^{obs}(x_a) - y^{pred}(x_a)\right)^2}{\delta^2}$$

(IIDs)

(assuming a constant δ for all measurements)

Procedure for constructing TS specified before looking at the data.

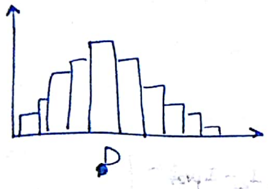
② ⇒ Perform pseudo-experiments. (for large N.)

• Pseudo-experiments : generate fake data assuming the hypothesis is true.

③ ⇒ For each pseudo-experiment ; generate the test statistic.

④ ⇒ For TS ; plot a histogram.

TS



} Distribution of the test statistic

⑤ Define a critical region for step 6. (away from the distribution of TS expected)

⑥ Comparison of observed TS from real experiments with the distribution of TS from pseudo experiments, assuming H_0 is true.

⑦ Does the observed test statistic lie inside the critical region?

↳ Yes ⇒ say H_0 is false.

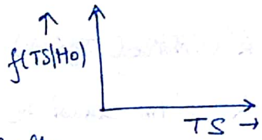
↳ No ⇒ say H_0 cannot be ruled out.

exp → data → TS

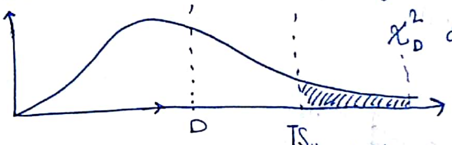
$$TS = \sum_a \left(\frac{y_a^{obs} - y_a^{pred}}{\delta a} \right)^2$$

Instead of the Histogram, we may also use $f(TS|H_0)$ and plot

↳ sometimes, it is also calculated analytically.



• How do we construct the critical region?



critical region. ($\underline{\kappa}$)

(extremely values of the TS)

$P(TS > TS_{threshold}) = \alpha = 0.01$ (1% chance for $TS > TS_{threshold}$)

$$f(TS|H_0) = \frac{1}{\Gamma(\frac{D}{2})(N)} (TS)^{\frac{D}{2}-1} e^{-\left(\frac{TS}{2}\right)}$$

Normalisation factors.

For $TS_{threshold}$ ⇒

$TS = x$ $\quad 1 - \int_0^{x_{th}} f(x|x_0)\,dx = 0.01. = \alpha.$

Integral does not exist analytically, have to calculate numerically.

Notation:

• $P(TS > TS_{th}) = 0.01 = \alpha$ (1% chance of $TS > TS_{th}$)

$\Downarrow$

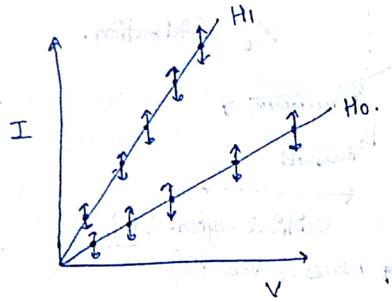
H_0 is said to be ruled out at >99% Confidence Level (CL) (if at all).

• $TS^{obs} \in K$ (critical region) $\Rightarrow$ ruled out.

$TS^{obs} \notin K \Rightarrow H_0$ cannot be ruled out.

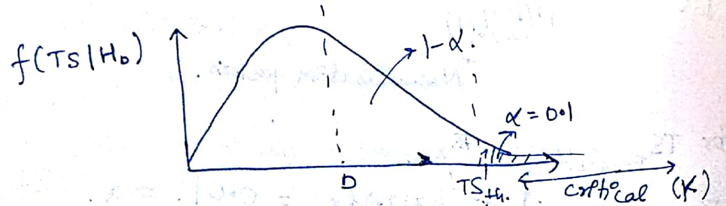
Example: H_0 : R is 670 Ω H_1 : R is 300 Ω.

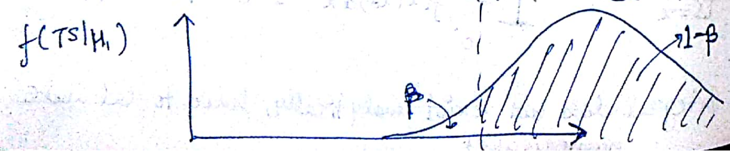
I → has gaussian fluctuations of size δ.



$$TS_{H_0} = \sum_i \left(\frac{y_i^{obs} - y_i^{pred}}{\delta}\right)^2 \leftarrow H_0$$

$f(TS|H_0)$





① H_0 is true and $TS \notin K \Rightarrow$ Conclude H_0 cannot be ruled out.
TRUE POSITIVE

② H_0 is true but $TS \in K \Rightarrow$ Conclude H_0 is ruled out.
FALSE NEGATIVE

③ H_1 is true and $TS \in K \Rightarrow$ Conclude H_0 is ruled out.

④ H_1 is true and $TS \notin K \Rightarrow$ Conclude H_0 cannot be ruled out.
FALSE POSITIVE

(for $\alpha >> 1$)

$P(①) = 1-\alpha$; $P(②) = \alpha$; $P(③) = 1-\beta$; $P(④) = \beta$.

$P(TS \notin K|H_0) = 1-\alpha$ → Right conclusion. True positive

$P(TS \in K|H_0) = \alpha$ → False negative

$P(TS \in K|H_1) = 1-\beta$ → ~~Wrong~~ Right conclusion

$P(TS \notin K|H_1) = \beta$. → Not able to make an effective statement. False positive

• Both false positives and false negatives are errors we have.

• If we make α too small, false negatives will decrease, but then false positives will increase; errors cannot be made arbitrarily small.

α → inefficiency of selection of H_0 events.

$\beta \rightarrow$ Background contaminating fake H_0 type events.

$1-\beta \rightarrow$ power of the test.

(optimisation b/w the power of the test and the inefficiency of the test).