Statistics of Measurements 2022

- Kolmogorov axioms

$P_i \geq 0 \quad P_{i \text{ or } j} = P_i + P_j \quad \sum_i P_i = 1$

↳ combined to imply all other properties of probabilities

- expectation value

↳ mean value of a distribution $E(x) = \langle x \rangle = \sum x_i P_i$

- As the number of trials $N \to \infty$ we expect to observe outcome x_i $n_i = N P_i$ times

- variance

↳ average square distance from the mean value

$$V(x) = \langle (x - \langle x \rangle)^2 \rangle = \sum_i x_i^2 P_i - \left(\sum_i x_i P_i\right)^2$$

- standard deviation

↳ indicates spread / width of distribution around mean

↳ same dimensions as x $\sigma(x) = \sqrt{V(x)}$

- Binomial distribution

↳ applies when there are 2 possible outcomes

↳ preform N independant trials and count the number n of outcomes with probability p

$$B(n; p, N) = {}^N C_n \, p^n (1-p)^{N-n} \qquad {}^N C_n = \frac{N!}{n! \, (N-n)!}$$

$E(n) = \langle n \rangle = Np \qquad V(n) = Np(1-p)$

↳ symmetric about $n = \frac{N}{2}$

- Poisson distribution
  - probability of observing n occurences if something occours many times with given average μ
  - Binomial distribution in the limit $N \to \infty$

$$P(n;\mu) = \frac{\mu^n e^{-\mu}}{n!} \qquad E(n) = \langle n \rangle = \mu \qquad V(n) = \mu$$

  - for a process with a constant rate λ (n per unit time)

$$\mu = \lambda t \qquad P(n;\mu=\lambda t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

  - rate could also be for example per area

- continuous random variables
  - x has a continuous sample space so probability of exactly $x \approx 0$ so consider interval x to $x+dx$

$$P(x+dx) = \rho(x)\,dx \implies P(x_1 < x < x_2) = \int_{x_1}^{x_2} \rho(x)\,dx$$

$$E(x) = \int_{-\infty}^{\infty} x\,\rho(x)\,dx \qquad V(x) = \int_{-\infty}^{\infty} x^2 \rho(x)\,dx - \left(\int_{-\infty}^{\infty} x\,\rho(x)\,dx\right)^2$$

- probability density function PDF
  - probability has to be dimensionless so $[\rho(x)] = \frac{1}{[x]}$
  - normalised by $\int_{-\infty}^{\infty} \rho(x)\,dx = 1$

- change of variables
  - discrete variables $P(x_i) = P_i$ and $P(y_i) = P_i$

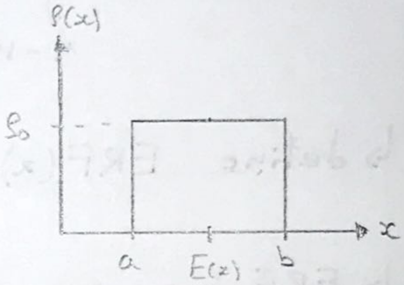
$P(y_i) = P(y(x_i))$ since $y = y(x)$

  - continuous variables $P(x \text{ to } x+dx) = P(y \text{ to } y+dy)$

express dy in terms of dx $\quad dy = \frac{dy}{dx}\,dx$

$$\rho_x(x)\,dx = \rho_y(y)\,\frac{dy}{dx}\,dx \implies \rho_y(y) = \underbrace{\left|\frac{dx}{dy}\right|}_{\text{inverted!}}\rho_x(x) \quad x = x(y)$$

- uniform distribution
  - constant PDF so $\rho(x) = \rho_0$



$$U(x; a, b) = \begin{cases} \rho_0 & \text{for } a \le x \le b \\ 0 & \text{otherwise} \end{cases} \qquad \rho_0 = \frac{1}{b-a}$$

$$E(x) = \frac{b+a}{2} \qquad V(x) = \frac{(b-a)^2}{12} \quad \text{so} \quad \sigma(x) = \frac{b-a}{\sqrt{12}}$$

- exponential distribution
  - used to determine time t between sucessive occurences with an average rate λ so $n = \lambda t$
  - derived from Poisson distribution

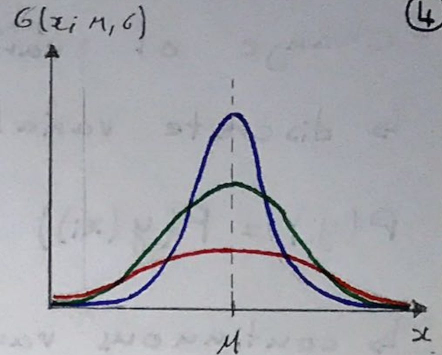
probability that $n=0$ for time $0 \to t$ $\quad P(0; \mu = \lambda t) = e^{-\lambda t}$

probability that $n=1$ for time $t \to t+dt$ $\quad P(1; \mu = \lambda t) = \lambda dt\, e^{-\lambda t}$

$$E(t; \lambda) = \lambda e^{-\lambda t} \qquad E(t; a) = \frac{1}{a} e^{-\frac{t}{a}} \qquad a = \frac{1}{\lambda}$$ averge time between occurences

$$E(x) = \frac{1}{\lambda} = a \qquad V(x) = \frac{1}{\lambda^2} = a^2 \qquad \sigma(x) = \frac{1}{\lambda} = a$$

- Gaussian (Normal) distribution



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

$E(x) = \mu \qquad V(x) = \sigma^2 \qquad dz = \frac{dx}{\sigma}$

↳ define standard score $z = \frac{x-\mu}{\sigma} \qquad G(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

- error function (ERF)

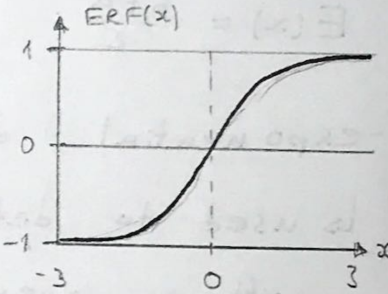
↳ to find probability of x being within $\pm n\sigma$ of mean

calculate $\displaystyle\int_{\mu-n\sigma}^{\mu+n\sigma} G(x;\mu,\sigma)\,dx = \int_{-n}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \frac{2}{\sqrt{\pi}} \int_0^{\frac{n}{\sqrt{2}}} e^{-y^2} dy$

↳ define $\mathrm{ERF}(x) = \frac{2}{\sqrt{\pi}} \displaystyle\int_0^x e^{-y^2} dy$ probability that X is in range $-x \to x$

↳ ERF has an std $\sigma_{ERF} = \frac{1}{\sqrt{2}}$ so account for factor of $\frac{1}{\sqrt{2}}$

e.g.) for $\pm n\sigma$ use $\mathrm{ERF}\left(\frac{n}{\sqrt{2}}\right)$



- Gaussian confidence intervals

↳ fraction of values which are contained within a given multiple of $n\sigma$

$\pm n$	fraction
1	0.683
2	0.954
3	0.997

fraction	$\pm n$
0.90	1.64
0.95	1.96
0.99	2.58

- sums of random variables

↳ consider N random values x_i with PDF $\rho_i(x_i)$

$X = \sum_i^N x_i \qquad E(X) = \sum_i^N E_i \qquad V(X) = \sum_i^N V_i$ (only if x_i independent)

- other properties
  - skewness $S(x) \sim x^3$ shows assynetry around mean
  - kurtosis $K(x) \sim x^4$ shows extra terms in width of peak which go beyond 2nd order

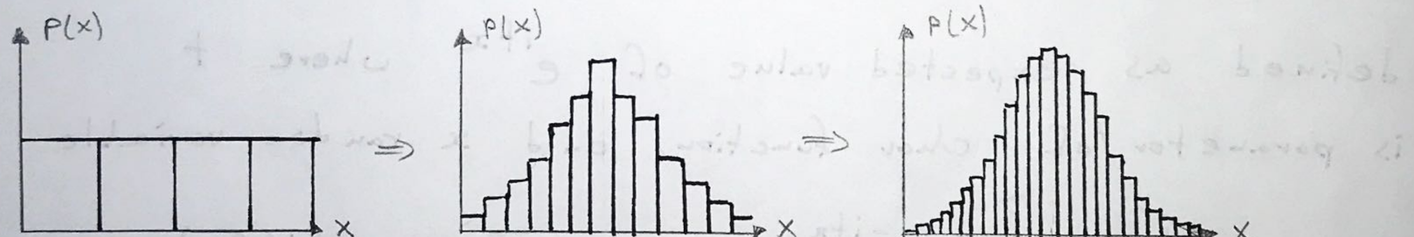
$S(x) = \sum_i^N S_i \quad K(x) = \sum_i^N K_i$

- dependance on no of trials N
  - expectation value $\bar{x} = \frac{X}{N}$

$E(\bar{x}) = E\left(\frac{X}{N}\right) = \frac{\sum E_i}{N} \implies E(\bar{x}) \sim 1 \qquad S(\bar{x}) \sim \frac{1}{N^2}$

$V(\bar{x}) = V\left(\frac{X}{N}\right) = \frac{\sum V_i}{N^2} \implies V(\bar{x}) \sim \frac{1}{N} \qquad K(\bar{x}) \sim \frac{1}{N^3}$

- central limit theorem (CLT)
  - in the limit $N \to \infty$ all random distributions tend towards a Gaussian $\lim_{N \to \infty} \rho(x) = G(x; \mu, \sigma)$
  - as N increases the width of the resultant distribution decreases uniform → gaussian → delta



- proof of CLT
  - consider N indipendant random variables x_i $\{x_1, x_2, \dots x_N\}$ with mean μ and variance σ^2
  - consider the sum $X = \sum_i^N x_i$ with mean $E(X) = N\mu$ and variance $V(X) = N\sigma^2$
  - define a z score $Z = \frac{X - N\mu}{\sqrt{N\sigma^2}} = \sum_i^N \frac{x_i - \mu}{\sqrt{N\sigma^2}} = \sum_i^N \frac{1}{\sqrt{N}} Y_i$

↳ where Y_i is a new random variable $Y_i = \frac{x_i - \mu}{\sigma}$
with $E(Y_i) = 0$ and $V(Y_i) = 1$

↳ find characteristic function of Z

$$\phi_Z(t) = \prod_i^N \phi_{Y_i}\left(\frac{t}{\sqrt{N}}\right) = \phi_{Y_1}\left(\frac{t}{\sqrt{N}}\right)\phi_{Y_2}\left(\frac{t}{\sqrt{N}}\right)\cdots\phi_{Y_N}\left(\frac{t}{\sqrt{N}}\right) = \left(\phi_Y\left(\frac{t}{\sqrt{N}}\right)\right)$$

↳ last step assumes $\phi_{Y_1} = \phi_{Y_2} = \phi_{Y_N}$ so variables Y_i are identically distributed

↳ now taylor expand $\phi_Y\left(\frac{t}{\sqrt{N}}\right) = 1 - \frac{t^2}{2N} + \dots$ around $Y=0$ so μ

↳ now take limit $N \to \infty$ and so $\frac{t}{\sqrt{N}} \to 0$

remember definition $e^x = \lim_{N\to\infty}\left(1 + \frac{x}{N}\right)^N$

so $\phi_Z(t) = \left(1 - \frac{t^2}{2N} + \dots\right)^N$ in limit $N \to \infty$ gives

$\phi_Z(t) \to e^{-\frac{1}{2}t^2}$ so a Gaussian $\boxtimes$

- characteristic function

↳ alternative to cumulative distribution function CDF

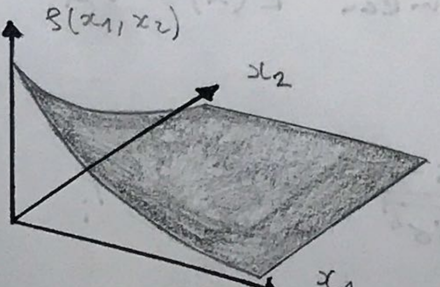
↳ defined as expected value of e^{itx} where t is parameter of char function and x random variable

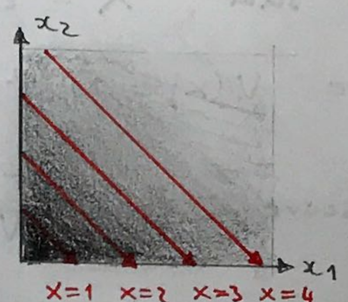
$$\phi_x(t) = \langle e^{ixt}\rangle = \int_{-\infty}^{\infty} \rho(x)\, e^{itx}\, dx \quad \text{so} \quad \phi_x(t) = \mathcal{F}(\rho(x))$$

- sum of two exponentials

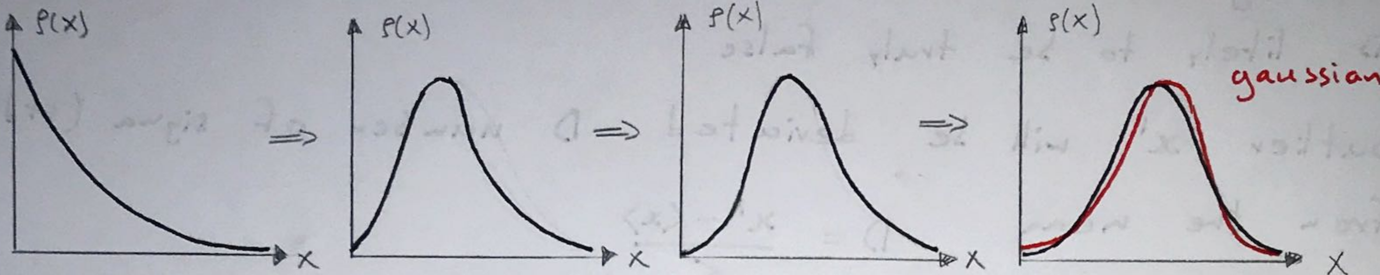
↳ one exponential along x and along y axis

↳ integrate along digonals where $x_1 + x_2 = X$





↳ the result of this summation will tend towards a gaussian as $N \to \infty$



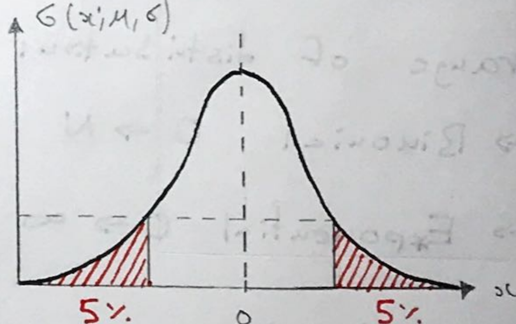
- sum of Binomial distributions
  - ↳ N trials and count n_1 outcomes then repeat N trials and count n_2 outcomes
  - ↳ combine $n = n_1 + n_2$ $\quad N \to 2N$ $\quad$ so $B(n; p, 2N)$
  - ↳ gaussian approximation at $N \to \infty$ holds for $1 \ll pN \ll N$ so mean $E(n)$ is between 1 and N
- sum of Poisson distributions
  - ↳ n_1 and n_2 outcomes both with averge μ
  - ↳ combine $n = n_1 + n_2$ $\quad \mu \to 2\mu$ $\quad$ so $P(n; 2\mu)$
  - ↳ gaussian approximation for $N_\mu \gg 1$
- range of distributions
  - ↳ Binomial $0 \to N$ $\qquad$ ↳ Poisson $0 \to \infty$
  - ↳ Exponential $0 \to \infty$ $\qquad$ ↳ Gaussian $-\infty \to \infty$
- discretisation
  - ↳ when approximating discrete distribution with continuous Gaussian discretise by integrating over integer intervals

$$\begin{tikzcd}
\text{Binomial} \arrow[rr, "N\to\infty \quad p\to 0 \quad pN\to\mu"] \arrow[dr, "1 \ll pN \ll N"'] & & \text{Poisson} \arrow[dl, "\mu \gg 1"] \\
& \text{Gaussian} &
\end{tikzcd}$$

- Chauvenet's criterion
  - ↳ deciding whether an outlier within a set of measurements is likely to be truly false
  - ↳ outlier x' will be deviated D number of sigma (std) from the mean $D = \frac{x' - \langle x \rangle}{\sigma}$
  - ↳ calculate probability that measurement will be outside of the $\pm D\sigma$ interval $P(\text{outside } D\sigma)$
  - ↳ if $N \times P(\text{outside } D\sigma)$ $< \frac{1}{2} \Rightarrow$ reject x'
    $> \frac{1}{2} \Rightarrow$ keep x'

- hypothesis testing
  - ↳ use null hypothesis H_0 with $\rho_0(x)$, $E_0(x)$ to see how consistent it is with the measured values x_m
  - ↳ if x_m lies in the rejection region reject H_0
  - ↳ even if x_m agrees with H_0 we can not confirm H_0 we can only say that we can't reject it
    (data insufficient to reject H_0 at α)

- rejection region
  - ↳ can be two-tailed or one-tailed



Rejection region (RR) $\alpha = 10\%$

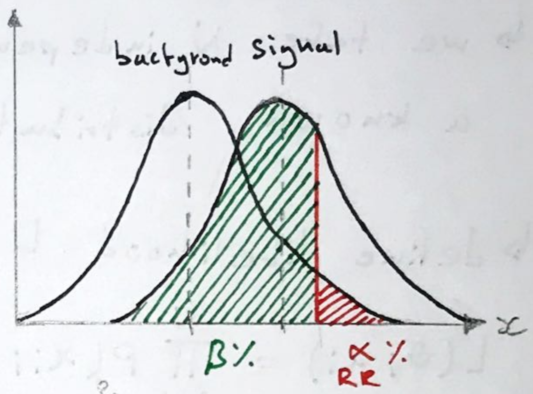
- significance level $\alpha = \int_{RR} \rho_0(x)\, dx$
  - ↳ usually use $\alpha = 0.05$ (5%) or 0.1 (10%)
  - ↳ if x_m is in RR reject H_0 at α significance level

- confidence level
  - ↳ probability outside of rejection region so $1-\alpha$

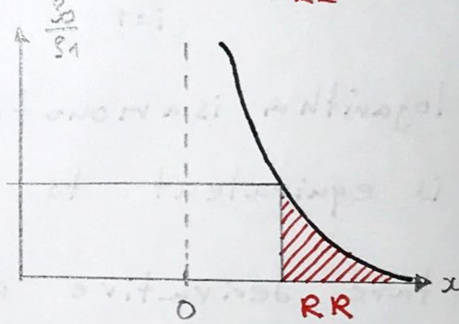
- Type 1 error
  ↳ occours if H_0 is rejected based on the hypothesis test but in reality H_0 is correct
  ↳ at α significance level there is an $\alpha\%$ chance of a type 1 error occuring

(1−β is called power of test)

- Type 2 error
  ↳ occours if H_0 is not-rejected based on the hypothesis test but in reality H_0 is false
  ↳ there is a $\beta\%$ chance of a type 2 error where β is the area under ρ_{signal} up to rejection region on $\rho_{background}$



- Neyman-Pearson PDF ratio
  ↳ ideally α, β should be as low as possible to reduce chance of type 1 and 2 errors
  ↳ if α, β are too low we could never reject a hypothesis so it is a trade off
  ↳ define ratio $\frac{\rho_{background}}{\rho_{signal}} = \frac{\rho_0}{\rho_1}$ and use it to set RR (α)



- parameter estimation
  ↳ we need to find parameters which are most consistent with the observed data
  ↳ the best estimate of a parameter is indicated by "^" $\hat{\theta}$
  ↳ this can be done using maximum likelihood or chi-squared method

- properties of estimation methods
  - consistent: as $N \to \infty$ $\quad \hat{\theta} \to \theta$
  - unbiased: averge $\quad E(\hat{\theta}) = \theta$
  - efficient: spread $\quad V(\hat{\theta}) = $ small
  - max likelihood and chi-squared are consistent and efficient but only unbiased in limit $N \to \infty$

- Maximum likelihood method
  - we take N independant measurements x_i with a known distribution $P(x_i;\theta)$ but we don't know θ
  - define likelihood L as product of probabilities of x_i

$$L(\theta; x_i) = \prod_{i=1}^{N} P(x_i;\theta) \qquad \ln(L(\theta)) = \sum_{i=1}^{N} \ln(P(x_i;\theta))$$

  - logarithm is a monotonic function so min/max in L is equivalent to min/max in $\ln(L)$
  - take derivative and set it to 0 to find the maximum of $L(\theta; x_i)$

$$\frac{d}{d\theta}\ln(\theta) = 0 \qquad \theta = \hat{\theta}$$

- Gaussian estimator

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

  - any $\ln(L(\theta))$ can be approximated as a gaussian around the peak
  - $\ln(L(\mu,\sigma)) = -N\ln(\sigma\sqrt{2\pi}) - \sum_i^N \frac{(x_i-\mu)^2}{2\sigma^2}$
  - if $N=1$ we can ignore σ so assume to be known for $N>1$ we must include σ

↳ differentiate wrt μ and σ and set to 0

$$\frac{\partial}{\partial \mu}\ln(L) = \sum_i^N \frac{(x_i-\mu)}{\sigma^2} = 0 \qquad \frac{\partial}{\partial \sigma}\ln(L) = -\frac{N}{\sigma} + \sum_i^N \frac{(x_i-\mu)^2}{\sigma^2}$$

$$\Downarrow \qquad\qquad \Downarrow$$

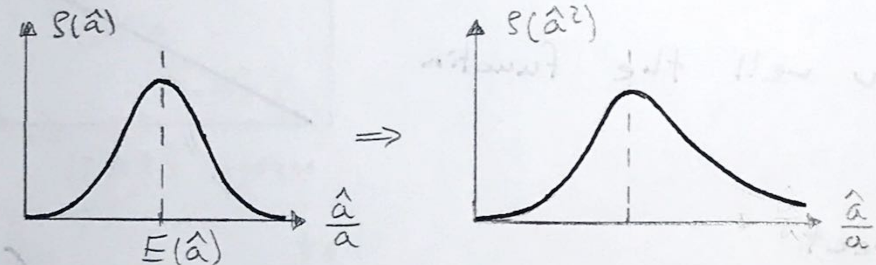
$$\hat{\mu} = \frac{1}{N}\sum_i^N x_i \qquad \hat{\sigma}^2 = \frac{1}{N}\sum_i^N (x_i-\hat{\mu})^2$$

↳ for $\hat{\sigma}^2$ is biased! since $E(\hat{\sigma}) < \sigma$ for low values of N

- Bessel correction
  ↳ to correct the bias in $\hat{\sigma}^2$ change $\frac{1}{N} \to \frac{1}{N-1}$
  ↳ the bias arises from the square since $(\hat{\sigma})^2$ doesn't vary linearly
  ↳ for large N so $N \to \infty$ the two are the same

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_i^N (x_i-\hat{\mu})^2$$

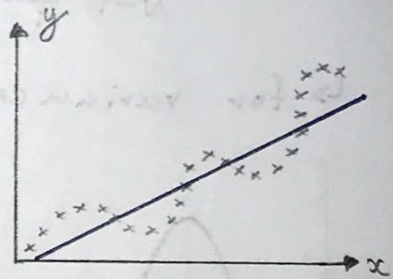
  ↳ for variance find $\hat{v}$ directly $\frac{\partial}{\partial v}\ln(L(\mu,v)) = 0$



- least square method
  ↳ measure random variables x_i and y_i which are related by $y = f(x;\theta)$ where θ is a parameter
  ↳ define residual $r_i = y_i - f(x_i;\theta)$
  ↳ form square of residual $S(\theta) = \sum_i^N r_i^2$
  ↳ vary θ to minimize $S(\theta)$

$$\frac{\partial}{\partial \theta} S(\theta) = 0 \implies \frac{\partial}{\partial \theta}\left(\sum_i^N (y_i - f(x_i;\theta))^2\right) = 0$$

- chi-squared method
  - if the measurements y_i have different uncertainties σ_i we would expect the values with lower uncertainty to be accounted for more significantly in $\hat{\Theta}$
  - same as least square method but accounts for σ_i
  - define pull $P_i = \frac{r_i}{\sigma_i} = \frac{y_i - f(x_i;\Theta)}{\sigma_i}$
  - define chi-squared $\chi^2(\Theta) = \sum_i^N P_i^2 = \sum_i^N \frac{(y_i - f(x_i;\Theta))^2}{\sigma_i^2}$
  - vary Θ to minimise $\chi^2(\Theta)$ $\quad \frac{\partial \chi^2}{\partial \Theta} = 0$
  - this only works for gaussian y_i but according to CLT we can justify it

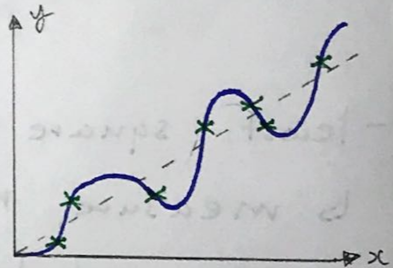


wrong $f(x,\theta)$

- goodness of fit
  - used to determine how well the function $f(x;\Theta)$ fits the data
  - in best case we expect

$$|y_i - f(x_i;\hat{\Theta})| \sim \sigma_i \implies |\hat{r}_i| \sim \sigma_i$$

$$|\hat{P}_i| \sim \frac{\sigma_i}{\sigma_i} = 1 \qquad \chi^2_{min}(\hat{\Theta}) = \sum_i^N P_i^2 = \sum 1 = N$$

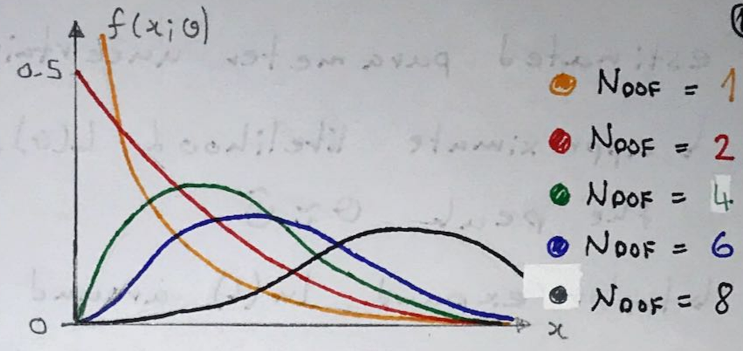


too many N_{dof}

  - so if function fits well $\chi^2_{min} \sim N$
    if $\chi^2_{min} \gg N$ $f(x;\Theta)$ is incorrect

- degrees of freedom (DOF)
  - having many parameters makes it easy to fit any data well
  - $N_{DOF} = N_{Data} - N_{Para}$
  - so good fit if $\chi^2_{min} \sim N_{DOF}$
    function incorrect if $\chi^2_{min} \gg N_{DOF}$

- chi-squared distribution
  - PDF of χ^2 used to find confidence intervals
  - special case of gamma distribution
  - as N_{DOF} increases it approaches a Gaussian



- relationship between L_{max} and χ^2_{min}
  - consider N measurements y_i with a Gaussian distribution and σ_i and mean $\mu_i(\theta) = f(x_i;\theta)$
  - find the log likelihood and rearrange

$$\ln(L(\theta)) = -\sum_i^N \ln(\sigma_i\sqrt{2\pi}) - \sum_i^N \frac{(y_i-\mu_i)^2}{2\sigma_i^2}$$

$$\sum_i^N \frac{(y_i-\mu_i)^2}{\sigma_i^2} = -2\sum_i^N \ln(\sigma_i\sqrt{2\pi}) - 2\ln(L)$$

$$\chi^2 = -2\left(\sum_i^N \ln(\sigma_i\sqrt{2\pi}) + \ln(L)\right) \implies \chi^2 = -2\ln(L)$$

  - since $\sum_i^N \ln(\sigma_i\sqrt{2\pi})$ is independent of θ a minimum χ^2 corresponds to a maximum $\ln(L)$ so the two methods give the same parameter estimates $\hat{\theta}$

- error propagation
  - for $y = y(x_i)$ where x_i has an uncertainty δ_i

$$\delta_y^2 = \sum_i^N \left(\frac{\partial y}{\partial x_i}\right)^2 \delta_i^2$$

  - from gaussian estimators we define uncertainty on mean

$$\hat{\delta}_\mu = \frac{\hat{\sigma}}{\sqrt{N}} \quad \text{(standard error on the mean)}$$

  - generally quote uncertainties to 1 sf if leading digit ≥ 3
    2 sf if leading digit < 3

- estimated parameter uncertainty for L_{max}

↳ approximate likelihood $L(\theta)$ as a Gaussian in θ around the peak $\theta \approx \hat{\theta}$

↳ taylor expand $\ln(L)$ around $\hat{\theta}$

$$\ln(L(\theta)) \approx \ln(L(\hat{\theta})) + \cancelto{0}{\frac{d\ln(L)}{d\theta}}\Big|_{\hat{\theta}} \frac{(\theta-\hat{\theta})}{1!} + \frac{d^2\ln(L)}{d\theta^2}\Big|_{\hat{\theta}} \frac{(\theta-\hat{\theta})^2}{2!} + \dots$$

$$\approx \ln(L(\hat{\theta})) + 0 + \frac{d^2\ln(L)}{d\theta^2}\Big|_{\hat{\theta}} \frac{(\theta-\hat{\theta})^2}{2} + \dots$$

↳ now define $\frac{d^2\ln(L)}{d\theta^2}\Big|_{\hat{\theta}} = -\frac{1}{\sigma_{\hat{\theta}}^2}$

$$\ln(L(\theta)) \approx \ln(L(\hat{\theta})) - \frac{(\theta-\hat{\theta})^2}{2\sigma_{\hat{\theta}}^2} \implies L(\theta) \approx L(\hat{\theta})\, e^{-\frac{(\theta-\hat{\theta})^2}{2\sigma_{\hat{\theta}}^2}}$$

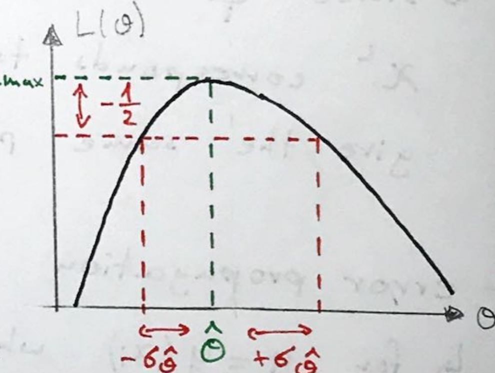
↳ so for gaussian approximation we get uncertainty $\sigma_{\hat{\theta}}$

↳ now consider $\theta = \hat{\theta} \pm \sigma_{\hat{\theta}} \implies \ln(L(\theta)) \approx \ln(L(\hat{\theta})) - \frac{(\hat{\theta} \pm \sigma_{\hat{\theta}} - \hat{\theta})^2}{2\sigma_{\hat{\theta}}^2}$

↳ so $\ln(L(\theta)) \approx \ln(L(\hat{\theta})) - \frac{1}{2}$ by introducing the uncertainty $\pm\sigma_{\hat{\theta}}$ we have decreased log likelihood by $\frac{1}{2}$

↳ so for any likelihood $L(\theta)$

for an assymetric $L(\theta)$ quote value as $\hat{\theta}\,^{+\sigma_{\hat{\theta}}}_{-\sigma_{\hat{\theta}}}$



- estimated parameter uncertainty for χ^2_{min}

↳ repeat taylor expansion

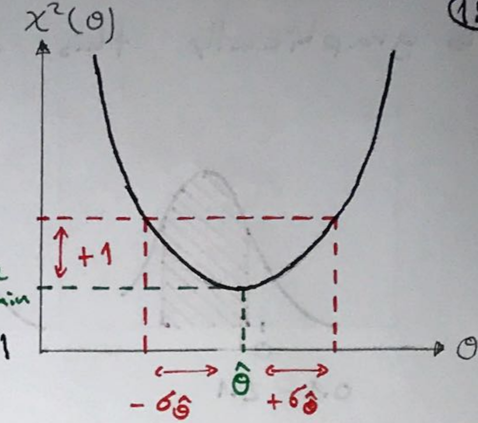
$$\chi^2(\theta) \approx \chi^2(\hat{\theta}) + \cancelto{0}{\frac{d\chi^2}{d\theta}}\Big|_{\hat{\theta}} \frac{(\theta-\hat{\theta})}{1!} + \frac{d^2\chi^2}{d\theta^2}\Big|_{\hat{\theta}} \frac{(\theta-\hat{\theta})^2}{2!} + \dots$$

$$\approx \chi^2(\hat{\theta}) + 0 + \frac{d^2\chi^2}{d\theta}\Big|_{\hat{\theta}} \frac{(\theta-\hat{\theta})^2}{2} + \dots$$

↳ we know $\ln(L) = -\frac{1}{2}\chi^2$ so $\frac{1}{2}\frac{d^2\chi^2}{d\theta^2}\Big|_{\hat{\theta}} = \frac{1}{\sigma_{\hat{\theta}}^2}$

(ignoring constant)

↳ now $\chi^2(\theta) \approx \chi^2(\hat{\theta}) + \frac{(\theta-\hat{\theta})^2}{\sigma_{\hat{\theta}}^2}$



↳ consider $\theta = \hat{\theta} \pm \sigma_{\hat{\theta}}$

$$\chi^2(\theta) \approx \chi^2(\hat{\theta}) + \frac{(\hat{\theta} \pm \sigma_{\hat{\theta}} - \hat{\theta})^2}{\sigma_{\hat{\theta}}^2} = \chi^2(\hat{\theta}) + 1$$

↳ so introducing uncertainty increased χ^2_{min} by +1

- interpretation of parameter uncertainties
  ↳ parameters aren't random variables and have a true value
  ↳ frequentist interpretation: (incorrect)
  the true parameter value is within the confidence interval $\pm n\sigma$ for α % of experiments
  e.g.) 68.3% of times within $\pm\sigma$ of $\hat{\theta}$
  ↳ this is incorrect since the parameter remains a single constant value for all experiments
  ↳ correct interpretation:
  define the uncertainty so that if we do a large number of experiments α % of the $\pm n\sigma$ confidence intervals will contain the true value so we have one θ and multiple confidence intervals

- physical constraints and frequentist estimation
  ↳ in frequentist approach physical constraints can lead to an "increased" accuracy in measured values
  ↳ e.g.) measure mass of powder in a dish $M_{dish} = 100\,g$ and the scale has an accuracy of $\pm 1g$

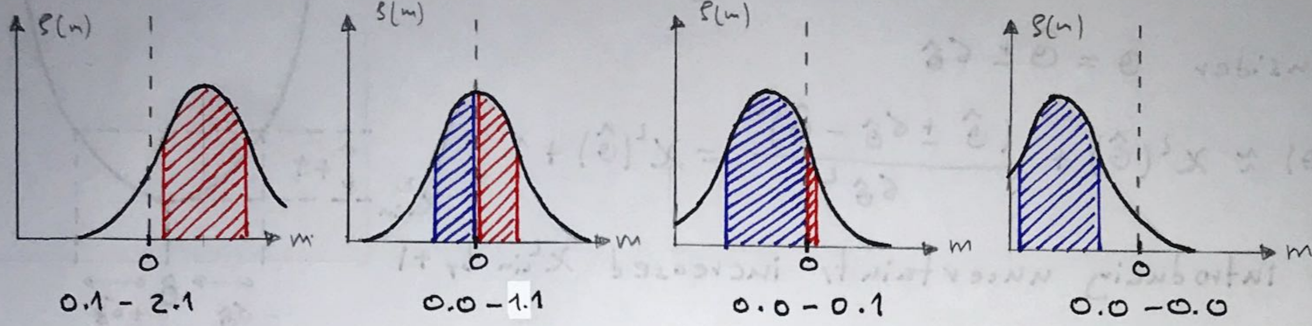
$m = 101.1 \pm 1.0g$ $\hat{\mu} = 1.1 \pm 1.0g$ $m_{powder} = 0.1 - 2.1g$

$m = 100.1 \pm 1.0g$ $\hat{\mu} = 0.1 \pm 1.0g$ $m_{powder} = 0.0 - 1.1g$

$m = 99.1 \pm 1.0g$ $\hat{\mu} = -0.9 \pm 1.0g$ $m_{powder} = 0.0 - 0.1g$

↳ are we really getting more accurate measurements?

↳ graphically this situation is:



- conditional probability

↳ $P(y|x)$ - probability of y given that x $\quad P(y|x) = \frac{P(y \cap x)}{P(x)}$

- Bayes theorem $\quad P(y|x) = \frac{P(x|y)\,P(y)}{P(x)}$

- Bayesian estimation of parameters

↳ use Bayes theorem $\quad x \rightarrow$ Data $\quad y \rightarrow \theta$

$$P(\theta|\text{Data}) = \frac{P(\text{Data}|\theta)\,P(\theta)}{P(\text{Data})}$$

$P(\theta|\text{Data})$ - posterior - probability of θ given collected Data

$P(\theta)$ - prior - probability of θ without experimental evidence

$P(\text{Data}|\theta)$ - likelihood - probability of obtaining Data given θ

$P(\text{Data})$ - evidence - probability of obtaining Data independent of parameter θ

↳ experiment yields data which update the PDF of θ

previous posterior $\longrightarrow$ new prior

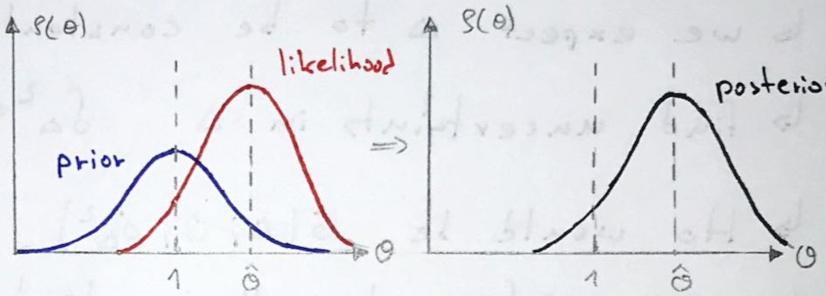
$P(\theta|\text{Data}) \qquad P(\theta)$

- Marginalisation
  - ↳ summing over all possible values of a parameter to determine the marginal contribution of another

$$P(y) = \sum_i P(y|x)\,P(x) = \int P(y|x)\,\rho(x)\,d\theta \quad \text{where } dP(x) = \rho(x)\,dx$$

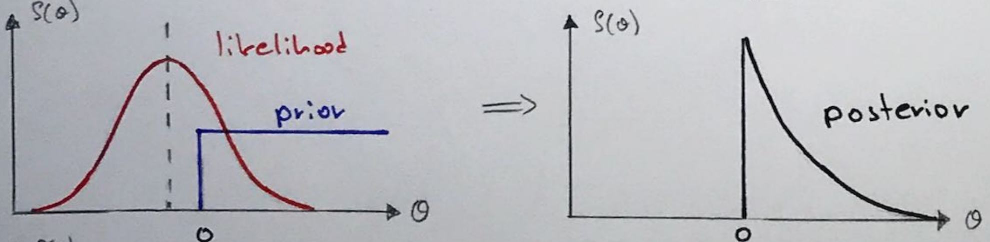
  - ↳ used to find P(Data) $\quad P(\text{Data}) = \sum_i P(\text{Data}|\theta)\,P(\theta)$

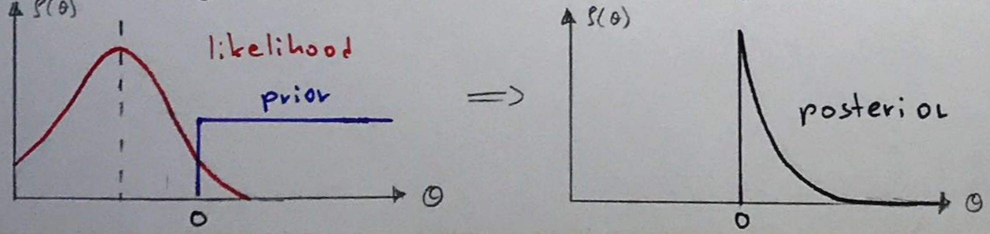
- priors
  - ↳ we have to guess $P(\theta)$ before inputting experimental data
  - ↳ gaussian prior with a gaussian likelihood will always give a gaussian posterior



  - ↳ usually assume prior to be uniform in θ since a uniform prior doesn't lead to any skewness of the posterior so $E(\hat{\theta}) = \theta$
  - ↳ if estimated parameter varies as θ^2 we need a prior uniform in θ^2
  - ↳ in the limit $N \to \infty$ the prior converges to the posterior

- Physical constraints in Bayesian approximation
  - ↳ use a prior such as a tophat which gives 0 in regions of non-physical θ
  - ↳ for example for mass $\rho(\theta) = \begin{cases} \rho_0 & \text{for } \theta \geq 0 \\ 0 & \text{for } \theta < 0 \end{cases}$





extra things to note:

- when changing multiple random variables use the Jacobian $J = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix}$

- comparing two experimental measurements of the same value
  - ↳ calculate the difference in estimated means

    $\Delta = \hat{\mu}_2 - \hat{\mu}_1$
  - ↳ we expect Δ to be constant with $E(\Delta) = 0$
  - ↳ find uncertainty in Δ $\quad \delta_\Delta^2 = \delta\hat{\mu}_2^2 + \delta\hat{\mu}_1^2 \Rightarrow \sigma_\Delta^2 = \sigma_{\hat{\mu}_2}^2 + \sigma_{\hat{\mu}_1}^2$
  - ↳ H_0 would be $G(\Delta; 0, \sigma_\Delta)$
  - ↳ now preform hypothesis test to see if H_0 is acceptable