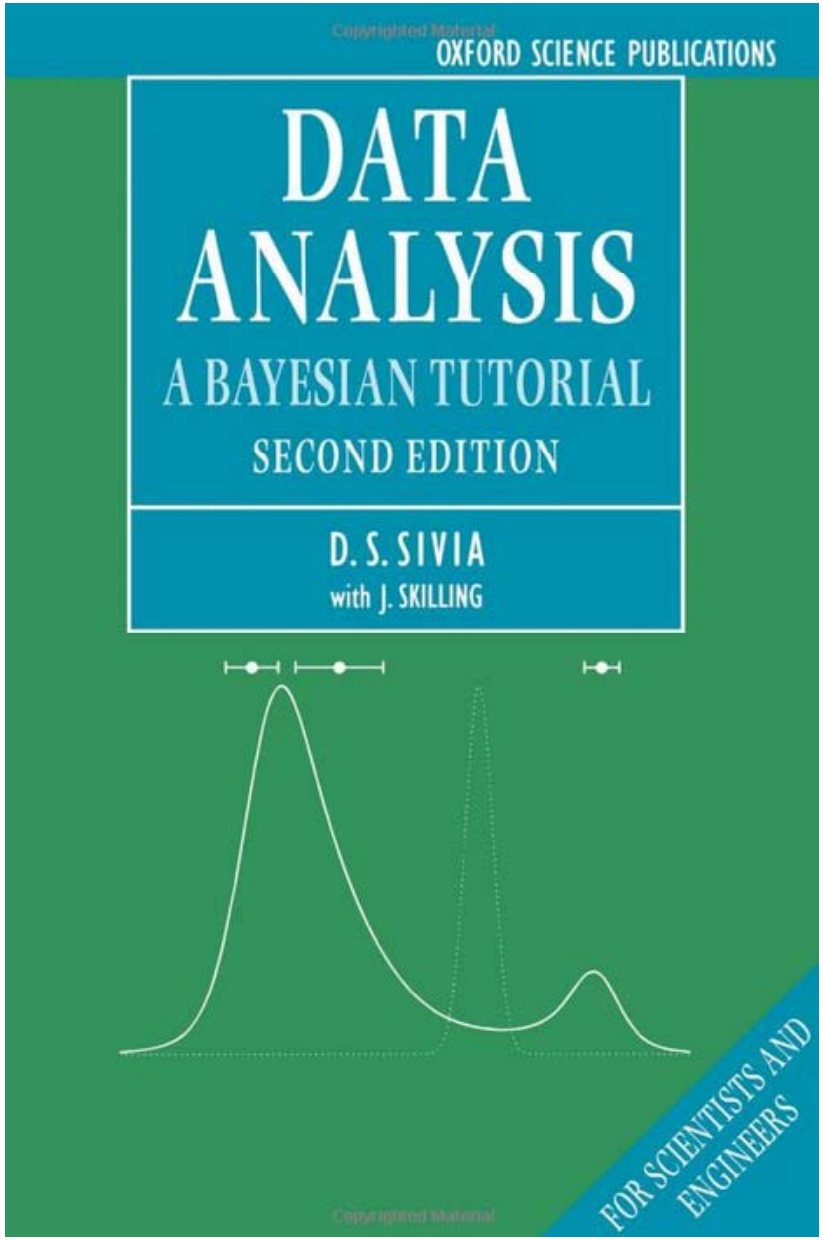


Problems of classical statistics

- Statistics seems a disconected series of receipts...
- Central limit theorem

  Statistic is a theory from which if you have two cars and I have none, each has one car

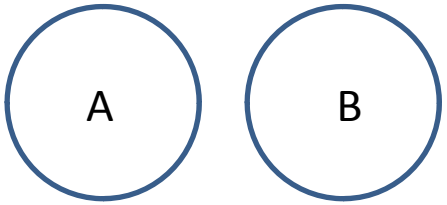
- And therefore is “firmly” based on gaussian distribution
- And therefore the “solutions” are also “gaussians”

Let’s stick to probability

Some basics of probability

A random variable A has a probability prob(X) of taking place between 0 and 1, being 1 when the event is sure.

If A and B are exclusive events then:

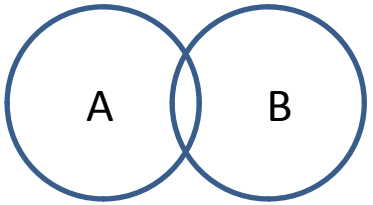


- prob(X or Y)=prob(X)+prob(Y)
- prob(X and Y)$\equiv$prob(X,Y)=prob(A)·prob(B)
- prob(X knowing Y) $\equiv$ prob(X|Y)=prob(X)

(since the information of B does not affect to our knowledge of A, since they are independent)

If A and B are not exclusive events then:

$$prob(X \mid Y) = \frac{prob(X,Y)}{prob(Y)}$$



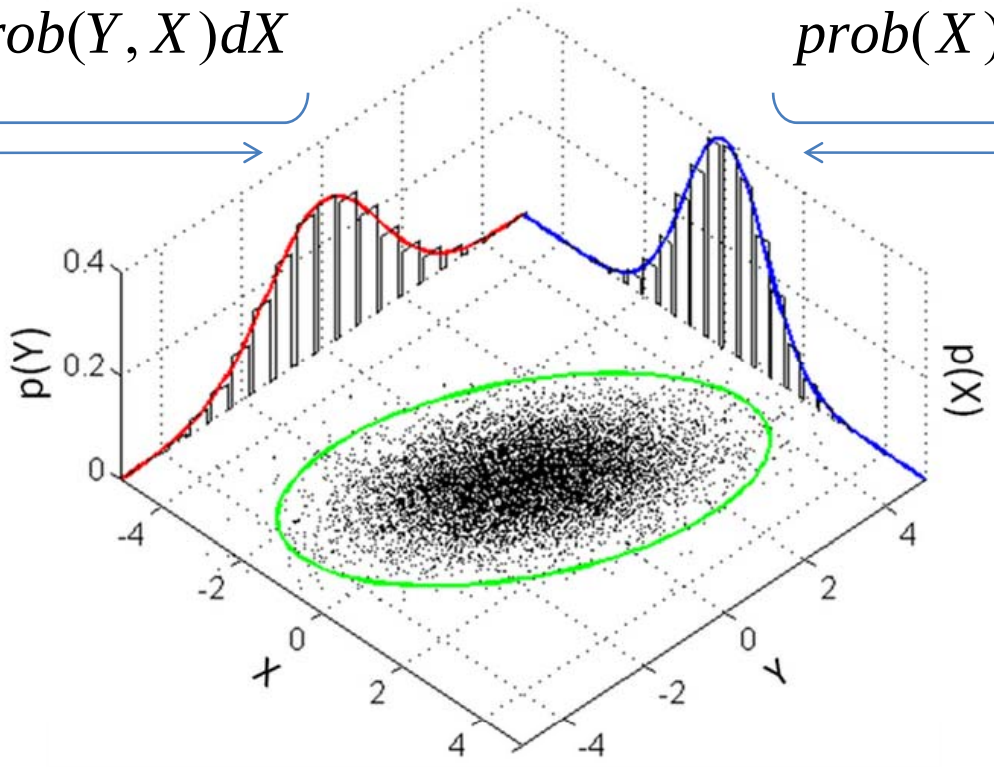
What can be do with Probability

MARGINALIZATION

We want to know something about X, no matter what happnes to Y
in other words Y represents a nuisance parameter... in the sense we are not interested in now!

$$prob(Y) = \int_{-\infty}^{\infty} prob(Y, X)dX$$

$$prob(X) = \int_{-\infty}^{\infty} prob(X, Y)dY$$



It allows to “project” a N-Dimensional PDF in the variable we are interested in

What can be do with Probability

MARGINALIZATION

We want to know something about X, no matter what happnes to Y
in other words Y represents a nuisance parameter... in the sense we are not interested in now!

$$prob(Y) = \int_{-\infty}^{\infty} prob(Y, X)dX = \int_{-\infty}^{\infty} prob(Y \mid X)\, prob(X)dX$$

$$prob(X \mid Y) = \frac{prob(X, Y)}{prob(Y)}$$

... for example t-student distribution can be obtained from the marginalization of σ
let D be the data

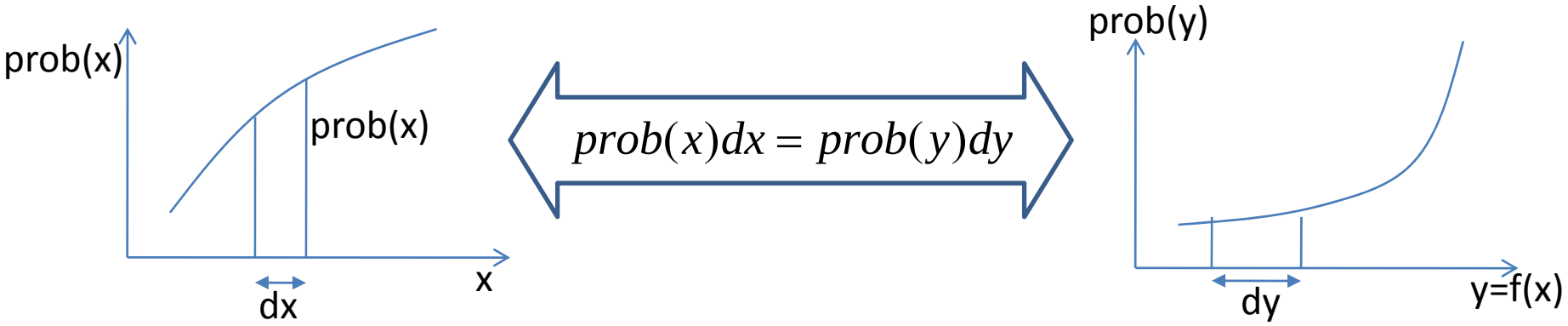
$$prob(\mu \mid D) = \int_{-\infty}^{\infty} prob(\mu, \sigma^2 \mid D)d\sigma^2 = \int_{-\infty}^{\infty} prob(\mu \mid D, \sigma^2)\, prob(\sigma^2 \mid D)d\sigma^2$$

What can be do with Probability

CHANGE OF VARIABLES

We want to know the PDF from a function of the original variable.

GOAL: after the change of variables the probabilty MUST NOT change



$$prob(x)dx = prob(y)dy$$

$$prob(x) = prob(y)\frac{dy}{dx}$$

What can be do with Probability

CHANGE OF VARIABLES

We want to know the PDF from a function of the original variable.

EXAMPLE

We have no idea about the “length scale” L of a problem .
It is reasonable that a maximum ignorance PDF is constant in log scale

$$prob(\ln(L)) = cte$$

What is the probability of L

$$prob(L) = \frac{d\left(\ln(L)\right)}{dL}\, prob(\ln(L)) \propto \frac{1}{L}$$

This is known as a Jeffrey’s PDF (prior as we will see shortly)

What can be do with Probability

CHANGE OF VARIABLES

We want to know the PDF from a function of the original variable.

EXAMPLE

We have no idea about the “length scale” L of a problem .
It is reasonable that a maximum ignorance PDF is constant in log scale

$$prob(\ln(L)) = cte$$

What is the probability of L

$$prob(L) = \frac{d\left(\ln(L)\right)}{dL}\, prob(\ln(L)) \propto \frac{1}{L}$$

This is known as a Jeffrey’s PDF (prior as we will see shortly)

What can be do with Probability

CHANGE OF VARIABLES

We want to know the PDF from a function of the original variable.

$$prob(x) = prob(y)\frac{dy}{dx} \quad \Rightarrow \quad prob(\{X_i\}) = prob(\{Y_i\})\left|\frac{\partial(Y_1, Y_2, \ldots Y_n)}{\partial(X_1, X_2, \ldots X_n)}\right|$$

Jacobian

What can be do with Probability

CHANGE OF VARIABLES

EXAMPLE: Changing from cartesian to polar coordinates

We want to express a bivariate gaussian $prob(x,y)=\frac{1}{2\pi\sigma^2}\exp\left[-\frac{x^2+y^2}{2\sigma^2}\right]$ in polar coordinates

We calculate the Jacobian $\left|\frac{\partial(x,y)}{\partial(R,\theta)}\right| = \begin{vmatrix}\cos\theta & -R\sin\theta \\ \sin\theta & R\cos\theta\end{vmatrix} = R\left[\cos^2\theta+\sin^2\theta\right] = R$

$$prob(R,\theta)=\frac{R}{2\pi\sigma^2}\exp\left[-\frac{R^2}{2\sigma^2}\right]$$

Moreover, we can now extract prob(R) by marginalization!!!

$$prob(R)=\int_0^{2\pi}\frac{R}{2\pi\sigma^2}\exp\left[-\frac{R^2}{2\sigma^2}\right]d\theta=\frac{R}{\sigma^2}\exp\left[-\frac{R^2}{2\sigma^2}\right]$$

Let's think about the result...

$$prob(R) = 2\pi R \cdot \frac{A}{\sigma^2} \exp\left[-\frac{R^2}{2\sigma^2}\right] \propto R \cdot \exp\left[-\frac{R^2}{2\sigma^2}\right]$$

probability that R lies in a narrow δR

probabilty at a distance R

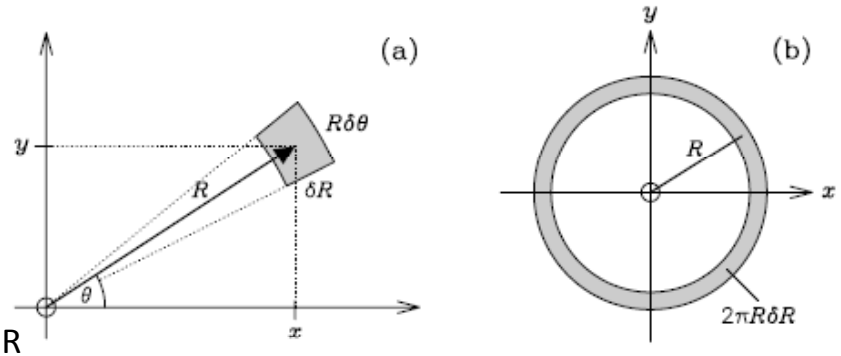


Fig. 3.14 Changing variables from Cartesian to polar coordinates.

Let's extend the result to N-dimensions...

$$prob(D \mid H) \propto \exp\left[-\frac{r_1^2 + r_2^2 + \ldots r_N^2}{2}\right] \quad \text{being} \quad r_i = \frac{D_i - H_i}{\sigma_i}$$

This is almost a χ^2 distribution after some massage. Following the previous reasoning

$$prob(R \mid H) \propto R^{N-1} \exp\left[-\frac{R^2}{2}\right] \quad \text{being} \quad R = \sqrt{\sum_{i=1}^{N} r_i^2}$$

Finally, taking into account $\chi^2 = R^2$ we get the probability for χ^2 (and not for R)

$$prob(\chi^2 \mid H) \propto R^{N-1} \exp\left[-\frac{R^2}{2}\right] \frac{dR}{d\chi^2} = \left(\sqrt{\chi^2}\right)^{N-1} \exp\left[-\frac{R^2}{2}\right] \frac{d\sqrt{\chi^2}}{d\chi^2} = \left(\chi^2\right)^{\frac{N}{2}-1} \exp\left(-\frac{\chi^2}{2}\right)$$

$$prob(\chi^2 \mid H) \propto \left(\chi^2\right)^{\frac{N}{2}-1} \exp\left(-\frac{\chi^2}{2}\right)$$

probabilty that χ at a distance δχ

probabilty at a distance χ

And finally: Bayes theorem

From the “multiplication” of probability:

$$prob(A, B) = prob(A) \cdot prob(B \mid A)$$

$$prob(A, B) = prob(B) \cdot prob(A \mid B)$$

dividing the two eqations we get:

Bayes theorem

$$prob(A \mid B) = \frac{prob(B \mid A)\, prob(A)}{prob(B)}$$

That help us to reverse probabilities...

Bayes theorem

$$prob(H \mid D) = \frac{prob(D \mid H)\, prob(H)}{prob(D)}$$

Posterior prob(H|D):
What you want to know is the probability that your hypothesis is true given the data

Likelihood (or L) prob(D|H):
What you know is yout hypothesis. And therefore you can calculate the probability that your data "gathers" around your hypothesis

Prior prob(H):
You might want to include any prioir information about your hypothesis

Evidence (E) prob(D):
Is "simply" a normalization factor... we will come back when doing model selection

Bayes theorem

Likelihood
Probability that our data describes the hypothesis

Prior
Our forehand knowledge

$$P(H \mid D) = \frac{P(D \mid H) \cdot P(H)}{P(D)}$$

T. Bayes.

Posterior
Probability that the hypothesis is true given the experimental data

Evidence
Normalization factor

In our case is very simple...

Bayes theorem

Likelihood
Probability that our data describes the hypothesis

Maximum ignorance Prior

$$P(H \mid D) \propto \frac{P(D \mid H) \cdot P(H)}{P(D)}$$

T. Bayes.

Posterior
Probability that the hypothesis is true given the experimental data

We only care about proportionality

Thus, assuming a maximum ignorance prior:

$$prob(H \mid D) \propto prob(D \mid H)$$

But what is exactly H, in practice?
Let's make clear some notation:

H is in our case a function of some parameters that describe your data

$$H_i(a,b) = a + bx_i$$

And we will find this written in many ways in the literature: argh!!

$$H \equiv H\left\{P_i\right\} \equiv \left\{P_i\right\} \equiv H, \vec{P} \equiv H, \vec{\alpha} \equiv H, \vec{\theta}$$

Remember our goals

1.- Parameter estimation

2.- Hypothesis testing (model selection)

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Remember our goals

1.- Parameter estimation

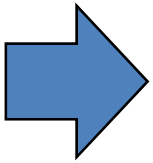
2.- Hypothesis testing (model selection)

Parameter estimation

We would like to estimate the best estimate of a quantity, given some data.
Let's define the posterior as:

$$prob\left(X \mid \{data\}\right) = P$$

The best estimate of X is given by
a maximum in its probability and thus



$$\left.\frac{dP}{dX}\right|_{X_0} = 0$$

We want now to expand P as a Taylor series,
since P varies too fast we take the logarithm of P

$$L = \ln(P) = \ln\left(prob\left(X \mid \{data\}\right)\right)$$

The Taylor expansion of L is:

$$L = L(X_0) + \frac{1}{2}\left.\frac{d^2L}{dX^2}\right|_{X_0}\left(X - X_0\right)^2 + \ldots$$

Taking the first term we get for P:

$$prob\left(X \mid \{data\}\right) \approx A\exp\left[\frac{1}{2}\left.\frac{d^2L}{dX^2}\right|_{X_0}\left(X - X_0\right)^2\right]$$

Let's compare with a gaussian

$$prob\left(X \mid \{data\}\right) \approx A\exp\left[\frac{\left(X - X_0\right)^2}{2\sigma^2}\right]$$

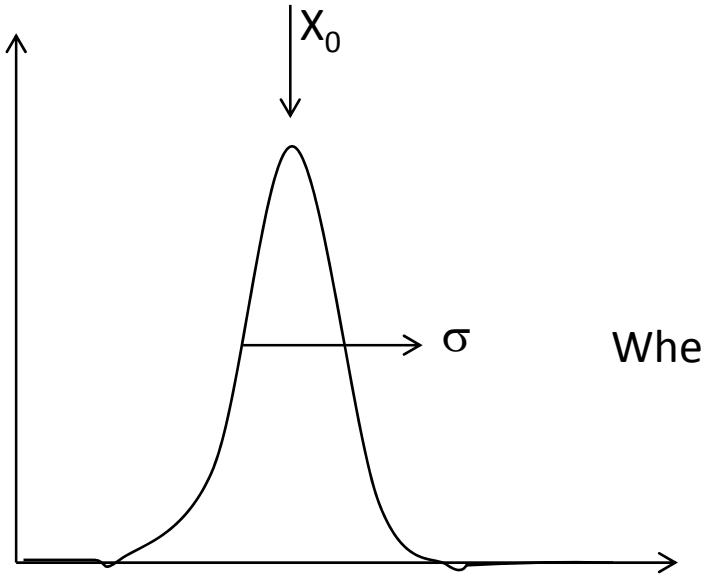
Comparing

Taking the first term we get for P:

$$prob\left(X \mid \{data\}\right) \approx A\exp\left[\frac{1}{2}\frac{d^2L}{dX^2}\bigg|_{X_0}\left(X - X_0\right)^2\right]$$

Let's compare with a gaussian

$$prob\left(X \mid \{data\}\right) \approx A\exp\left[-\frac{\left(X - X_0\right)^2}{2\sigma^2}\right]$$



$$\sigma = \left(-\frac{d^2L}{dX^2}\bigg|_{X_0}\right)^{-1/2}$$

Where $\frac{d^2L}{dX^2}\Big|_{X_0}$ is always negative, so: no problem ;-)

$$X = X_0 \pm \sigma$$

Value plus error comes from a Taylor series

(closing the eyes to the "real" shape... that sometimes is ok, and sometimes not)

Remember our goals

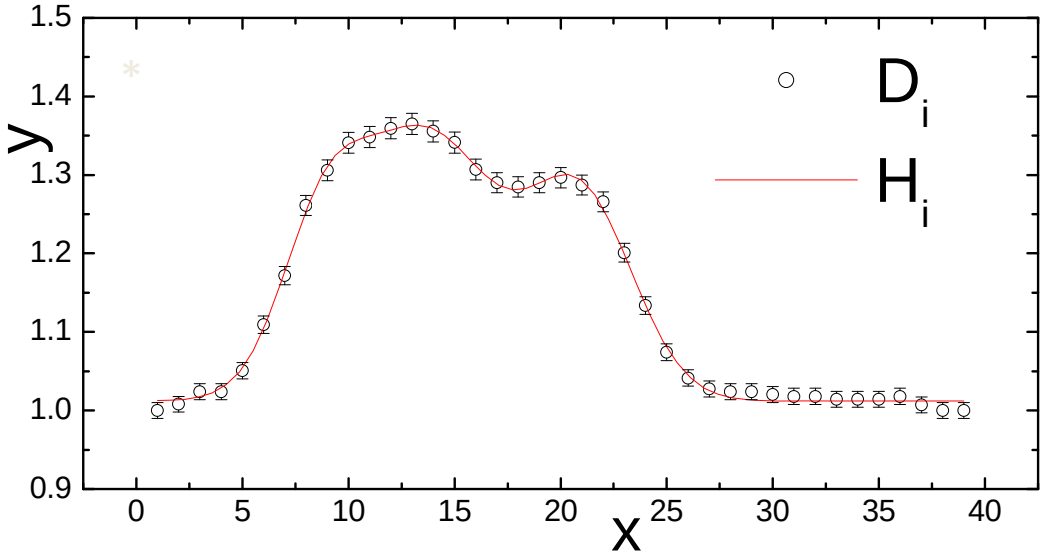
1.- Parameter estimation

2.- Hypothesis testing (model selection)

Bayes theorem

$$P(H \mid D) \propto P(D \mid H) \equiv L$$

D_i Data (i=1,n) $H_i\{P_l\}$ Hypothesis (i=1,n) using a parameter set $\{P_l\}$ (l=1,m)



$$P(H_i\{P_l\} \mid D_i) \propto P(D_i \mid H_i\{P_l\})$$

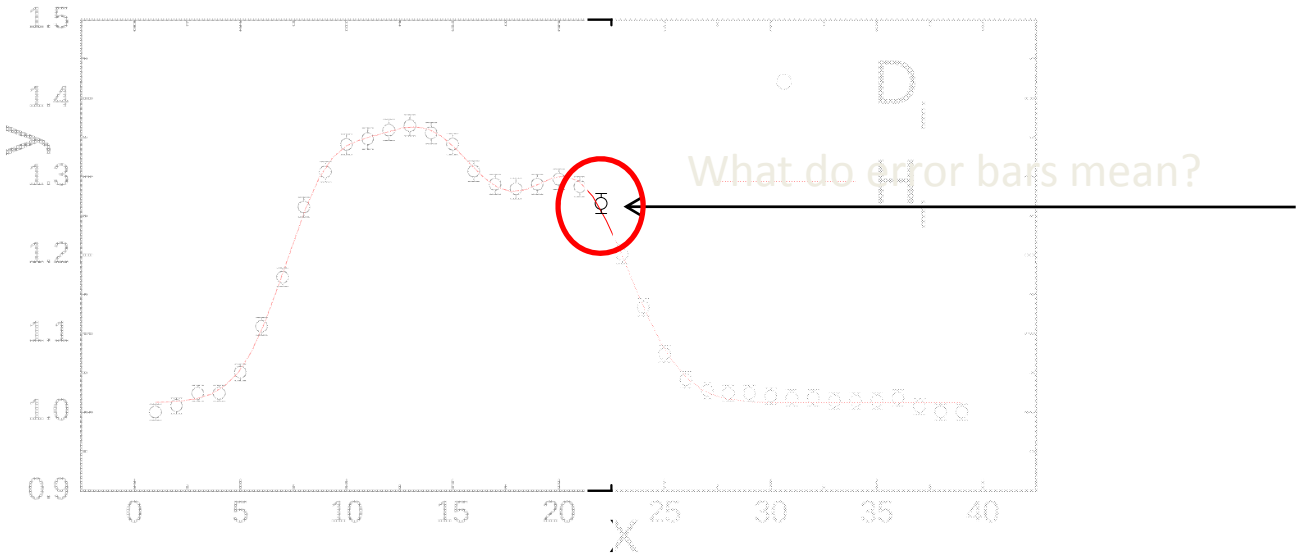
* Figure adapted from „Le petit prince“ A. Saint Exupery (1943)

The ubiquitous χ^2

Bayes theorem

$$P(H \mid D) \propto P(D \mid H)$$

D_i Data (i=1,n) $H_i\{P_l\}$ Hypothesis (i=1,n) using a parameter set $\{P_l\}$ (l=1,m)

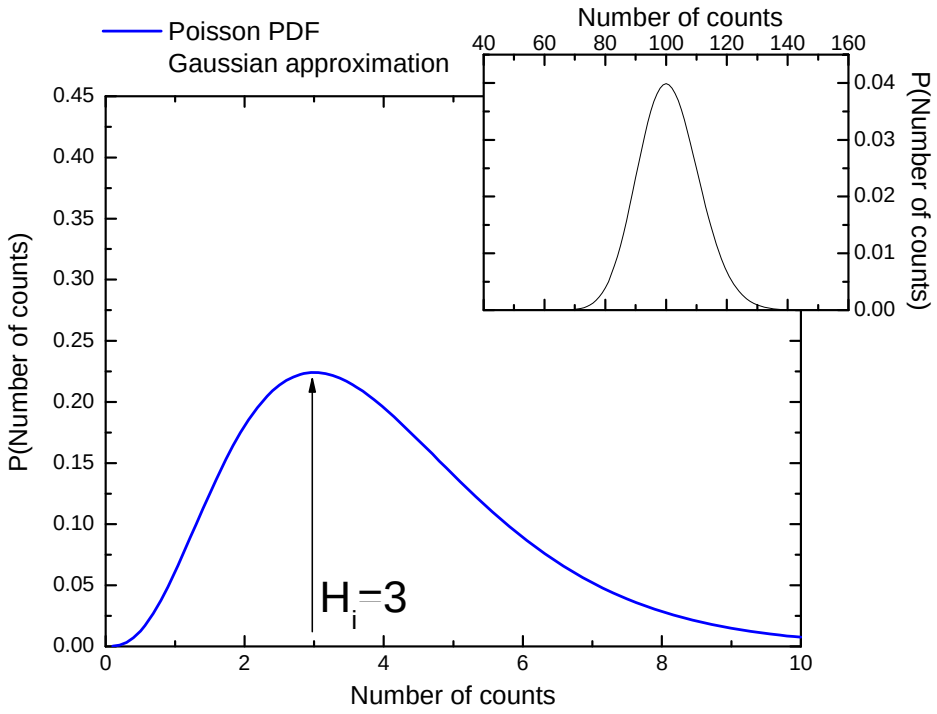


$$P(H_i\{P_l\} \mid D_i) \propto P(D_i \mid H_i\{P_l\})$$

The ubiquitous χ^2

In a counting experiment, if the expected value is H_i measured values will follow a Poisson statistics

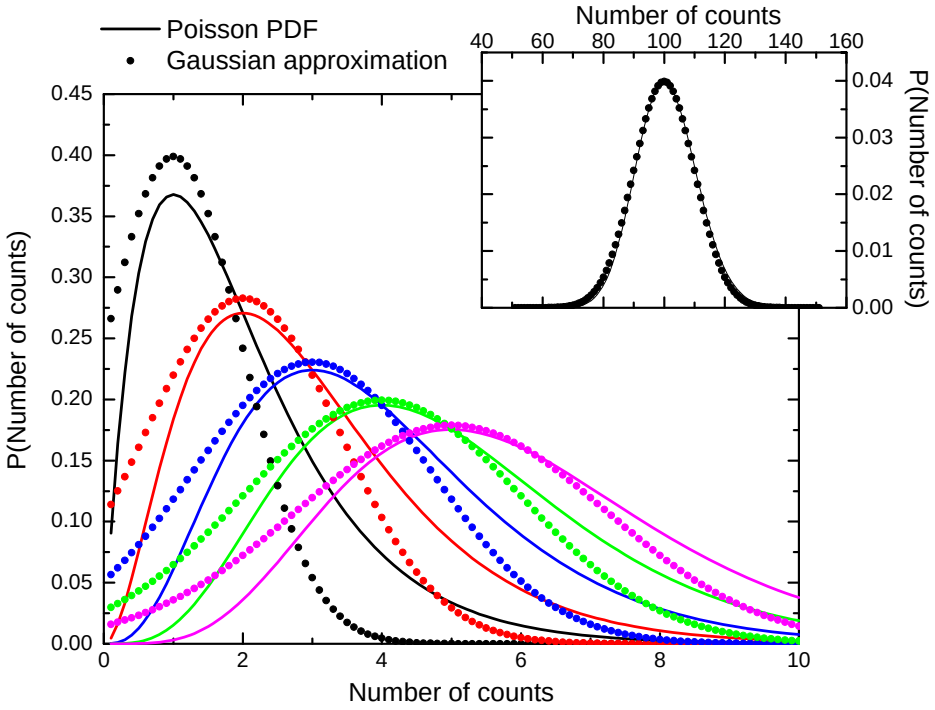
$$P(D_{i=k} \mid H_{i=k}) \propto H_i^{Di=k} e^{-Hi=k}$$



The ubiquitous χ^2

In a counting experiment, if the expected value is H_i measured values will follow a Poisson statistics

$$P(D_{i=k} \mid H_{i=k}) \propto H_i^{Di=k} e^{-Hi=k}$$



If the number of counts is high enough poisson statistics =normal distribution with $\sigma_i = \sqrt{D_i}$

$$P(D_{i=k} \mid H_{i=k}) \propto \exp - \frac{(H_{i=k} - D_{i=k})^2}{2\sigma_{i=k}^2}$$

The error is the square root of the variance $\varepsilon_i = \sigma_i$

Bayes theorem

$$P(H_i\{P_l\} \mid D_i) \propto P(D_i \mid H_i\{P_l\})$$

We now consider all the points i=1,2,... n

$$L = P(D_i \mid H_i\{P_l\}) \propto \prod_{i=1}^{n} \exp - \frac{(H_i - D_i)^2}{2\sigma_i^2}$$

$$= \exp \sum_{i=1}^{n} - \frac{(H_i - D_i)^2}{2\sigma_i^2} = \exp - \frac{\chi^2}{2}$$

Therefore χ^2 is related to the likelihood:

$$\chi^2 \propto -2 \cdot \ln L$$

So, we got the exact meaning on probability bases of χ^2

Let's use it!

The ubiquitous χ^2

Least squares method therefore:

$$L = prob\left(D \mid H, P_i\right) \propto \exp\left(-\chi^2/2\right)$$

Maximize the probability that the Data is described by your hypothesis

$$prob\left(H_i, P_i \mid D_i\right) \propto \exp\left(-\chi^2/2\right)$$

And therefore to maximize the probability of your hypothesis!!!

Hypothesis testing

Does our model describe our data correctly?

$$\chi^2 \propto -2 \cdot \ln L$$

the logarithm of the likelihood answers the question χ^2 follows a χ^2 distribution...

... so we can go back to P values and company. But...

did we made all this way to go back to P values???

Model selection

Imagine two competing models or Hypothesis H_1 and H_2

$$P(H_1|D)=\frac{P(D|H_1)P(H_1)}{P(D)} \qquad P(H_2|D)=\frac{P(D|H_2)P(H_2)}{P(D)}$$

$$\underbrace{\frac{P(H_1|D)}{P(H_2|D)}}_{\text{posterior ratio}}=\underbrace{\frac{P(D|H_1)}{P(D|H_2)}}_{\text{Bayes factor: B}}\cdot\underbrace{\frac{P(H_1)}{P(H_2)}}_{\text{prior for theories (usually 1)}}$$

P(D|H):To assign probabilities we take the hypothesis **INDEPENDENT** of the particular parameter we obtain in the fitting: we marginalize the parameters:

$$P(D|H)=\int prob(D,\vec{\alpha}|H)\,d\vec{\alpha}=\int prob(D|\vec{\alpha},H)\,prob(\vec{\alpha}|H)\,d\vec{\alpha}$$

Therefore

$$\mathrm{B}=\frac{P(D|H_1)}{P(D|H_2)}=\frac{\int prob(D|\vec{\alpha},H_1)\,prob(\vec{\alpha}|H_1)\,d\vec{\alpha}}{\int prob(D|\vec{\beta},H_2)\,prob(\vec{\beta}|H_2)\,d\vec{\beta}}$$

Model selection

Imagine two competing models or Hypothesis H_1 and H_2

$$\mathrm{B} = \frac{P(D \mid H_1)}{P(D \mid H_2)} = \frac{\int prob(D \mid \vec{\alpha}, H_1)\, prob(\vec{\alpha} \mid H_1)\, d\vec{\alpha}}{\int prob(D \mid \vec{\beta}, H_2)\, prob(\vec{\beta} \mid H_2)\, d\vec{\beta}}$$

Probability of the data around best fit α_i^*

$$prob(D \mid \alpha_i, H) = prob(D \mid \alpha_i^*) \cdot \exp\left[-\frac{(\alpha_i - \alpha_i^*)^2}{2\delta\alpha^2} \right]$$

$$L = A \exp\left(-\frac{\hat{\chi}^2}{2} \right)$$

prior from out parameter maximum ignorance:

location

$$prob(\alpha_i) = \frac{1}{\alpha_{\max} - \alpha_{\min}}$$

scale factor

$$prob(\alpha_i) = \frac{1/\alpha_i}{\int_{\alpha_{\min}}^{\alpha_{\max}} 1/\alpha_i \, d\alpha_i} = \frac{1}{\alpha_i \ln\left(\frac{\alpha_{\max}}{\alpha_{\min}} \right)}$$

Model selection

Imagine two competing models or Hypothesis H_1 and H_2

$$\mathrm{B} = \frac{P(D \mid H_1)}{P(D \mid H_2)} = \frac{\int prob(D \mid \vec{\alpha}, H_1)\, prob(\vec{\alpha} \mid H_1)\, d\vec{\alpha}}{\int prob(D \mid \vec{\beta}, H_2)\, prob(\vec{\beta} \mid H_2)\, d\vec{\beta}}$$

Let's assume that each fit has only one parameter:

$$\mathrm{B} = \underbrace{\frac{\exp\left(-\chi^2_{H_1}/2\right)}{\exp\left(-\chi^2_{H_2}/2\right)}}_{\text{favours best fit}} \underbrace{\frac{\delta\alpha}{\alpha_{\max} - \alpha_{\min}} \frac{\beta_{\max} - \beta_{\min}}{\delta\beta}}_{\text{penalizes additional parameters}}$$

Ockham factor

$$\alpha_{\max} - \alpha_{\min} >> \delta\alpha$$

Model selection

Akaike and Bayesian information criteria

Akaike information criteria

gausssian distributed points

$$AIC = -2\ln L_{\max} + 2k = \chi^2 + 2k$$

Bayesian Information Criteria

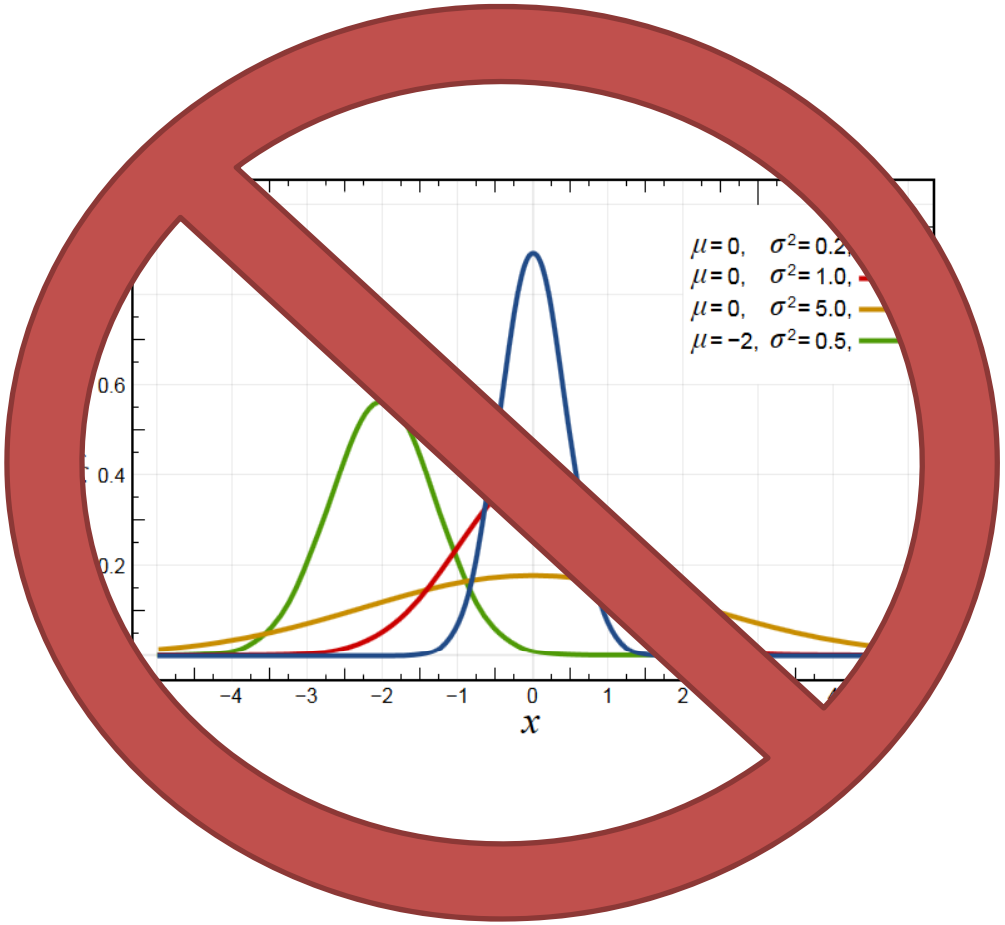
$$P(D \mid H) = \int prob(D, \vec{\alpha} \mid H)\, d\vec{\alpha} = \int prob(D \mid \vec{\alpha}, H)\, prob(\vec{\alpha} \mid H)\, d\vec{\alpha}$$

Can be approximated by:

$$P(D \mid H) \approx BIC = -2\ln L_{\max} + 2k = \chi^2 + k\ln N$$

where k is the number of parameters, and N the number of points.
We should then choos the one with minimum AIC or BIC

Beyond the Normal distribution



There is a lot still based on normal distribution: is there a method to avoid this????

$$L = P(D_i \mid H_i\{P_l\}) \propto \prod_{i=1}^{n} \exp - \frac{(H_i - D_i)^2}{2\sigma_i^2}$$

You cannot avoid to assign a probability to each point, that might be gaussian as here...
or can be a poisson distribution:

$$\log L = \sum_{i=1}^{i=Ndata} \log\left[P(H|Data_i)\right] = \sum_{i=1}^{i=Ndata} \log\left[\frac{e^{-H_i} H_i^{D_i}}{D_i!}\right] \propto \sum_{i=1}^{i=Ndata} \left[-H_i + D_i \cdot \log H_i\right]$$

or any other distribution...

There is a lot still based on normal distribution: is there a method to avoid this????

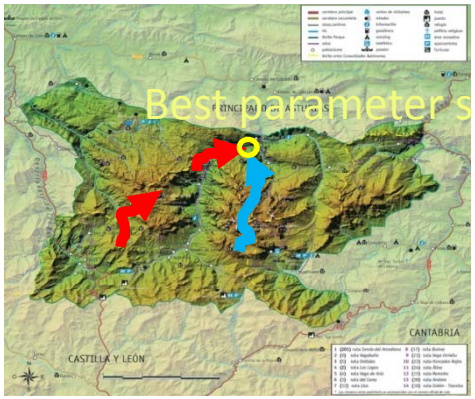
...also, gaussianity comes from exploring around the minimum:
no matter if the minimum is flat, or multimodal...

$$L = L(X_0) + \frac{1}{2}\frac{d^2L}{dX^2}\bigg|_{X_0}\left(X - X_0\right)^2 + \ldots$$

Can we simply get L??

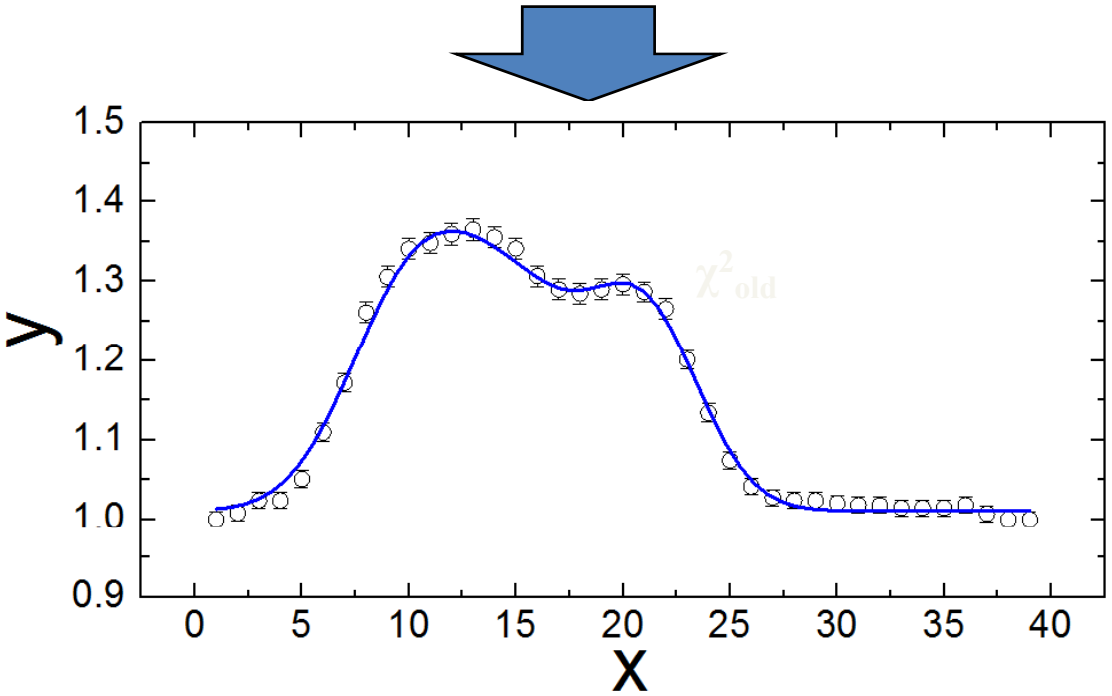
Markov Chain Montecarlo method
we can explore the likelihood, better its logarithm χ^2

Let's define a parameter space $\chi^2\{Pi\}$



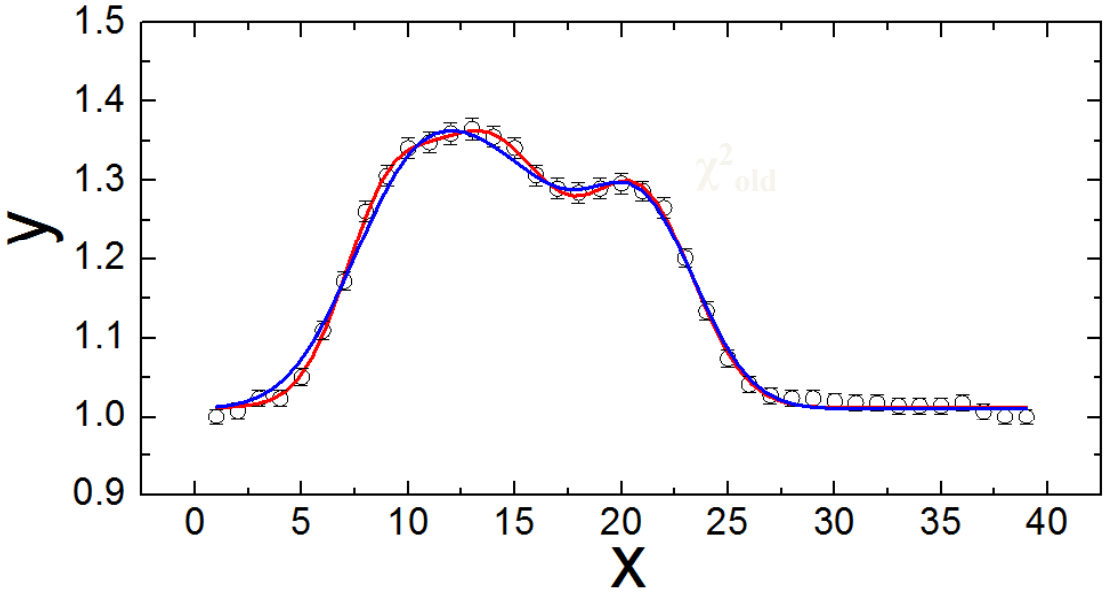
Bayesian tool to explore χ^2 landscape

Step	χ^2	P_1	P_2	P_3
1	$1.368 \cdot 10^4$	1.341	0.00223	10.223
2				
3				
4				
…	…	…	…	…



Bayesian tool to explore χ^2 landscape

Step	χ^2	P_1	P_2	P_3
1	$1.368 \cdot 10^4$	1.341	0.00223	10.223
2	$1.360 \cdot 10^4$	**1.342**	0.00223	10.223
3				
4				
...	...	...	...	...



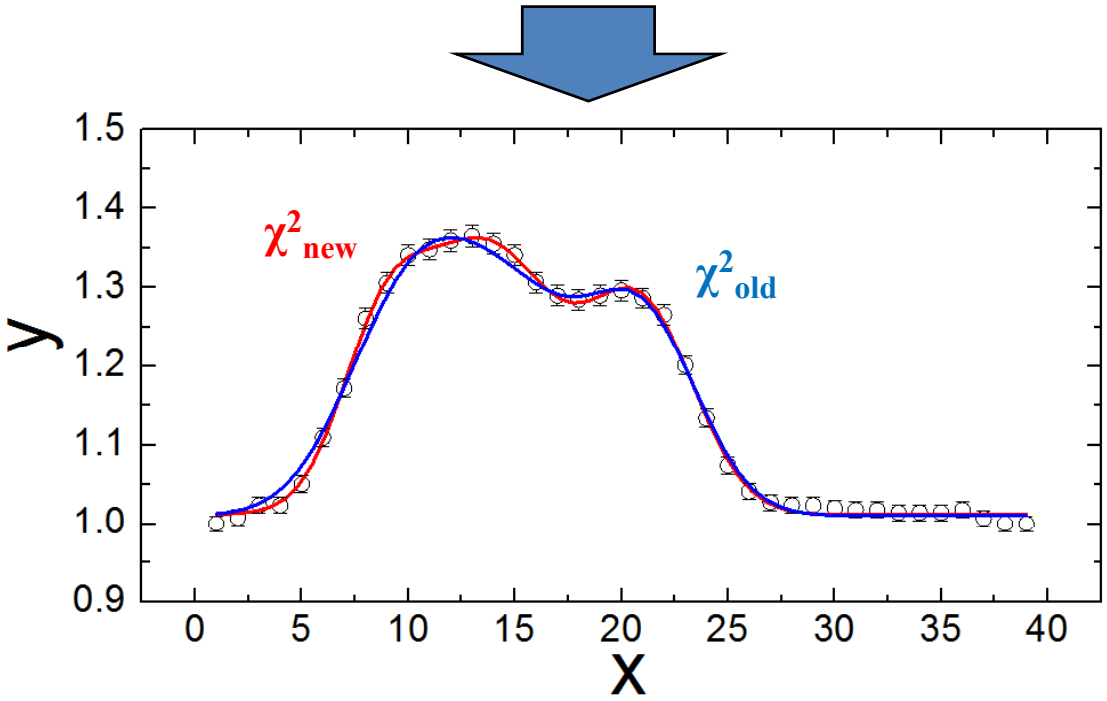
Acceptance rules:

- If better always
- If not, with probability

$$P_{old \to new} = \exp - \left(\frac{\chi^2_{new} - \chi^2_{old}}{2} \right)$$

Bayesian tool to explore χ^2 landscape

Step	χ^2	P_1	P_2	P_3
1	$1.368 \cdot 10^4$	1.341	0.00223	10.223
2	$1.360 \cdot 10^4$	**1.342**	0.00223	10.223
3	$1.362 \cdot 10^4$	1.342	0.00212	10.223
4	$1.348 \cdot 10^4$	1.342	0.00212	10.170
...	...	...	...	...



Acceptance rules:

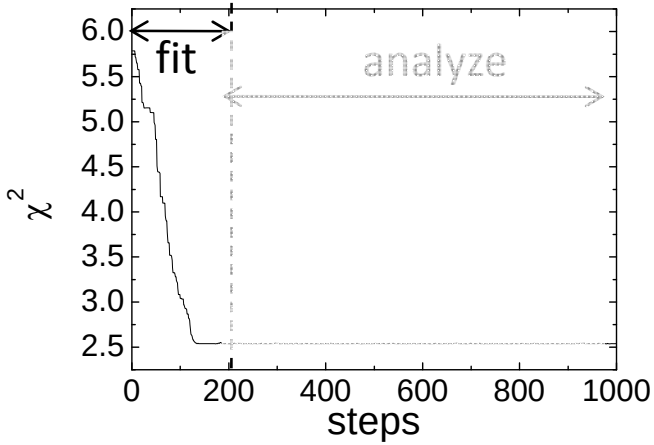
- If better always
- If not, with probability

$$P_{old \to new} = \exp - \left(\frac{\chi^2_{new} - \chi^2_{old}}{2} \right)$$

Bayesian tool to explore χ^2 landscape

- ❑ It does not get stuck

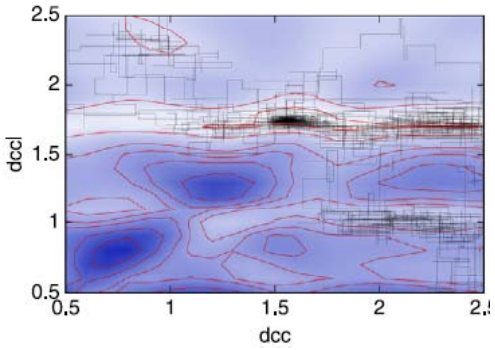
... and if it does
we increase the errors



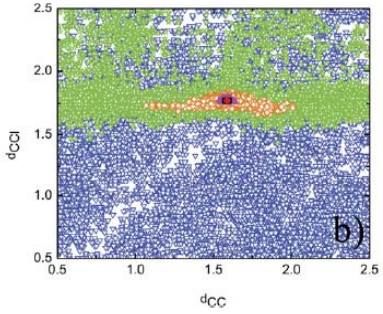
Simulated annealing

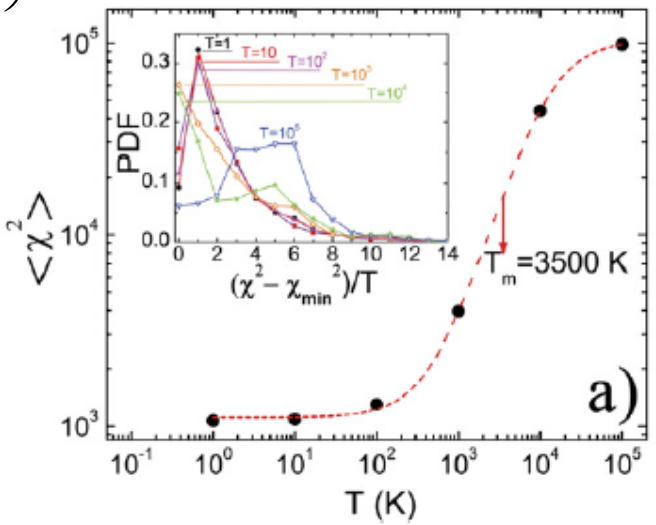
$$P_{old\to new} = \exp-\left(\frac{\chi^2_{new} - \chi^2_{old}}{2T}\right)$$

A 2D χ^2 landscape



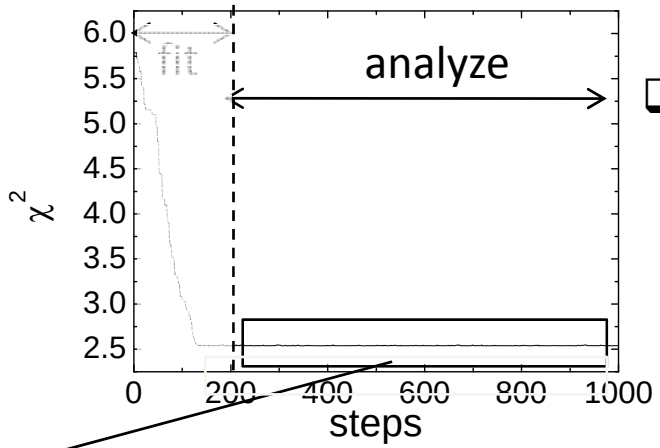
Warming up the fitting

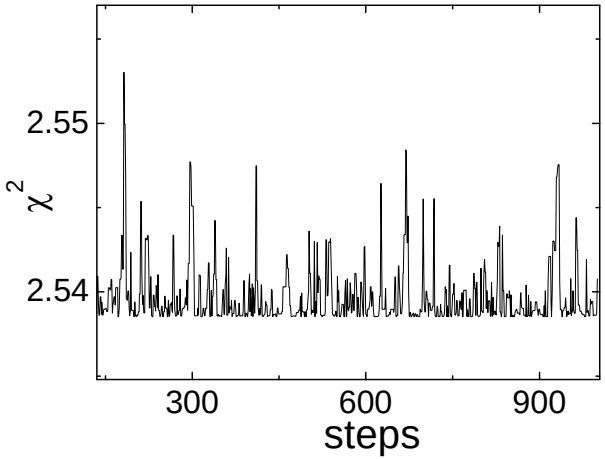


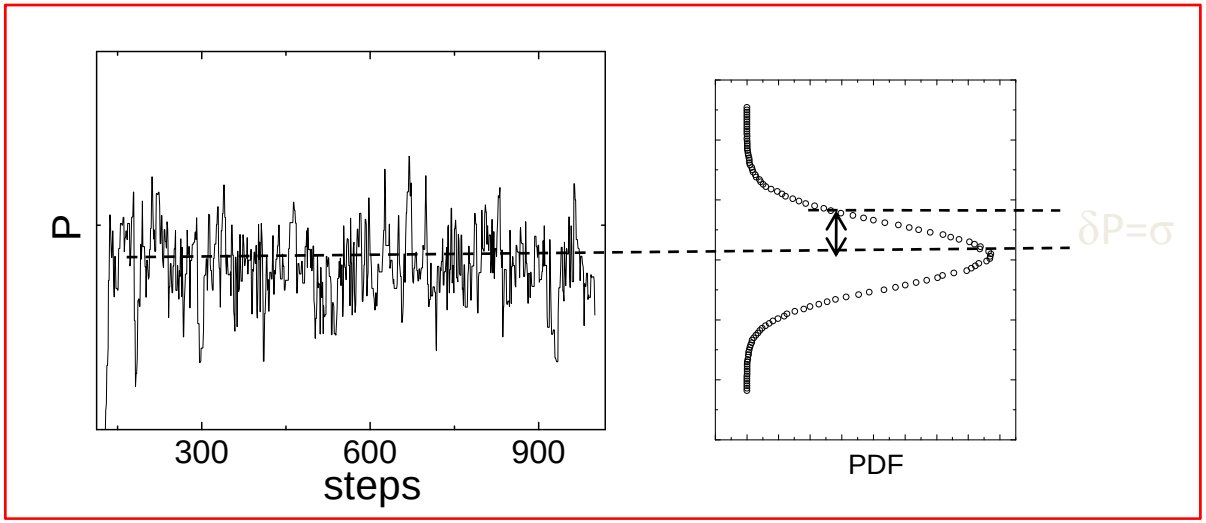


Bayesian tool to explore χ^2 landscape

- ❑ It does not get stuck
- ❑ Robust parameter estimation



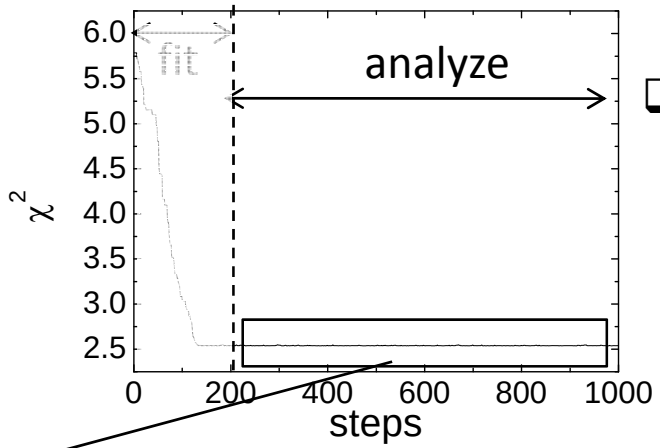


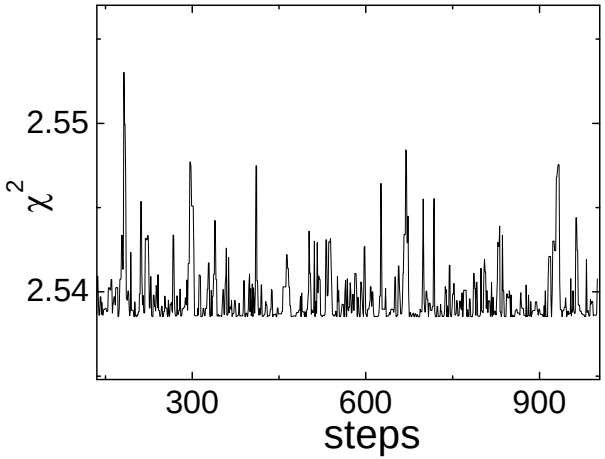


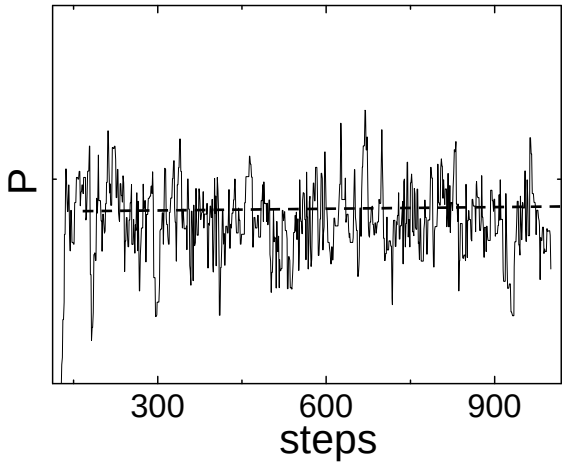
Parameter estimation:
Parameters are obtained as PDF's not as P±δP

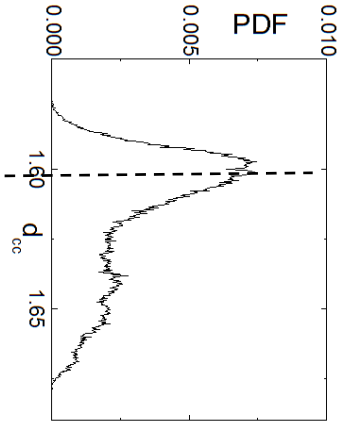
Bayesian tool to explore χ^2 landscape

- ❑ It does not get stuck
- ❑ Robust parameter estimation







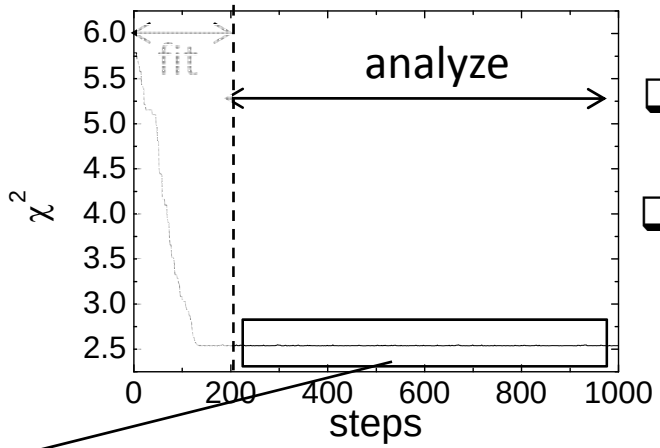


Parameter estimation: Parameters are obtained as PDF's not as P±δP

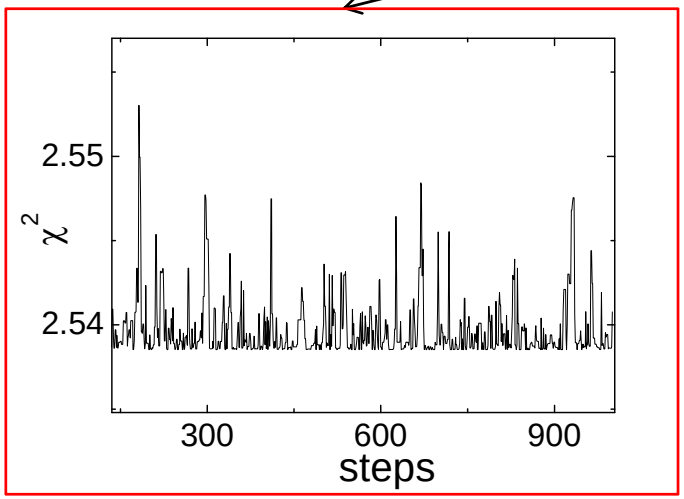
You are marginalizating all parameters except the one you are interested in

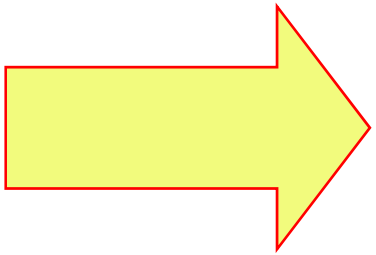
Bayesian tool to explore χ^2 landscape

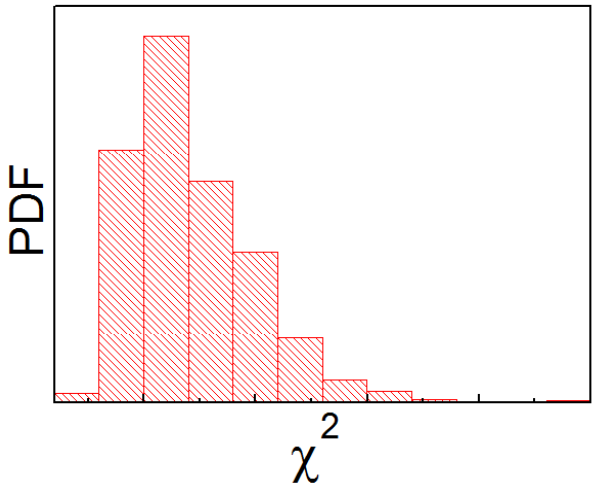
- ❑ It does not get stuck



- ❑ Robust parameter estimation
- ❑ Robust model selection







Model selection:
all possible combinations of parameters are investigated
and their χ^2 calculated

Molecular spectroscopy and Bayesian spectral analysis—how many lines are there?

D. S. Sivia and C. J. Carlile
ISIS Pulsed Neutron Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, OX11 0QX, United Kingdom

(Received 24 July 1991; accepted 24 September 1991)

Models tested:

- Stretched exponential (KWW)
- Stretched exponential + gaussian distribution of lorentz lines
- Two gaussian distribution of lorentz lines
- Two lorentzians
- Three lorentzians

A. Vispa et al. PCCP 10.1039/C5CP05143F (2016)

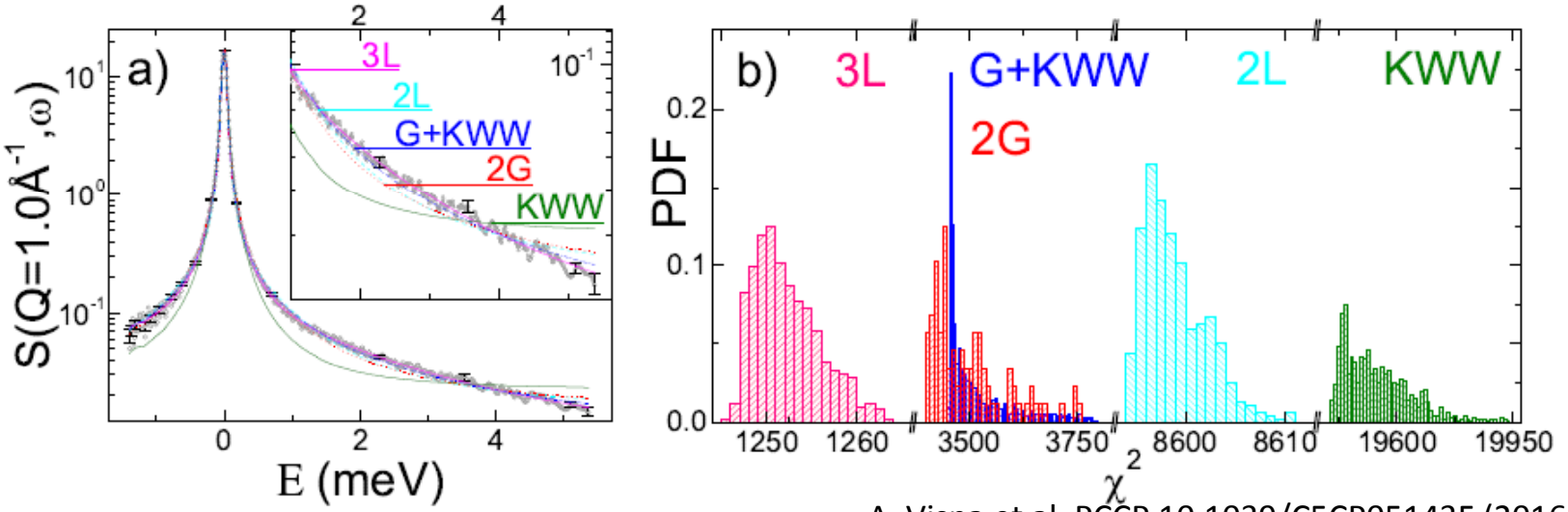
Molecular spectroscopy and Bayesian spectral analysis—how many lines are there?

D. S. Sivia and C. J. Carlile
ISIS Pulsed Neutron Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, OX11 0QX, United Kingdom

(Received 24 July 1991; accepted 24 September 1991)

Models tested:

- Stretched exponential (KWW)
- Stretched exponential + gaussian distribution of lorentz lines
- Two gaussian distribution of lorentz lines
- Two lorentzians
- Three lorentzians



A. Vispa et al. PCCP 10.1039/C5CP05143F (2016)

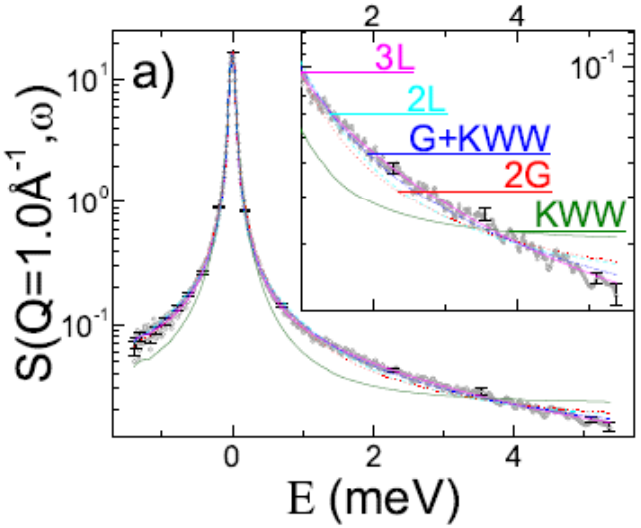
Molecular spectroscopy and Bayesian spectral analysis—how many lines are there?

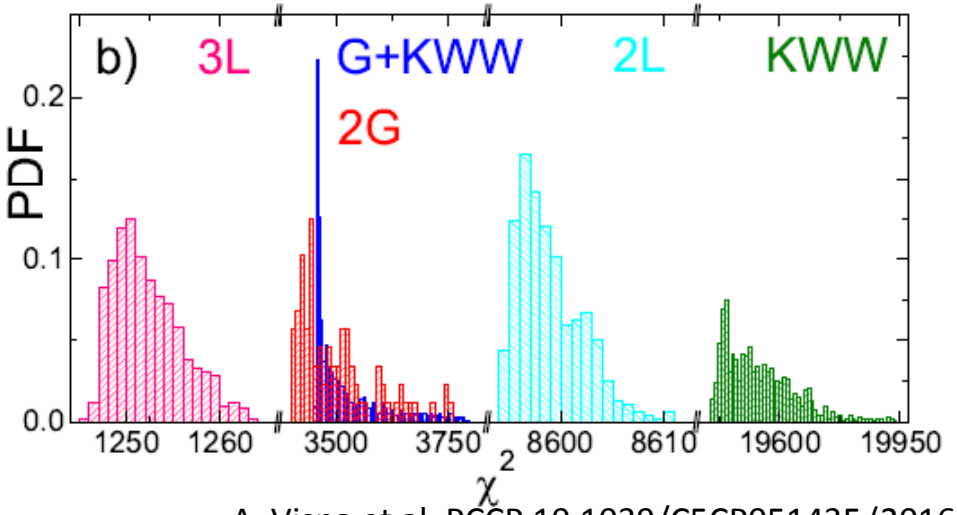
D. S. Sivia and C. J. Carlile
ISIS Pulsed Neutron Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, OX11 0QX, United Kingdom

(Received 24 July 1991; accepted 24 September 1991)

Models tested:

- Stretched exponential (KWW)
- Stretched exponential + gaussian distribution of lorentz lines
- Two gaussian distribution of lorentz lines
- Two lorentzians
- **Three lorentzians THE WINNER**



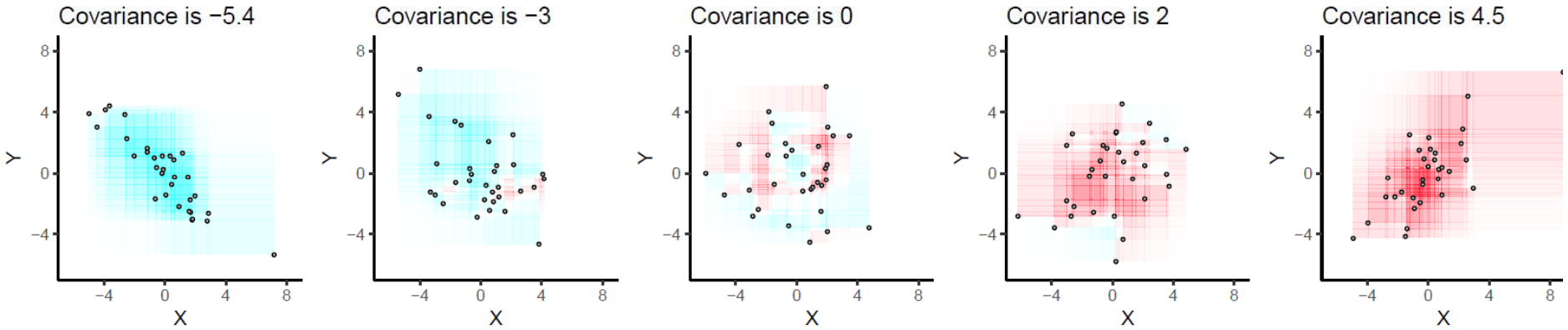


A. Vispa et al. PCCP 10.1039/C5CP05143F (2016)

PARAMETER CORRELATION

Usually only **linear** correlation, and **twobody** correlations: **covariance**

$$\sigma_{XY}^2 = \sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y})$$



Indeed, we can define a covariance matrix

$$\begin{pmatrix} \sigma_X^2 & \sigma_{XY}^2 \\ \sigma_{XY}^2 & \sigma_Y^2 \end{pmatrix}$$

PARAMETER CORRELATION

Usually only correlation around the minimum, and **twobody** correlations: **covariance**

The bayesian counterpart is (not assuming gaussianity, but only the shape of the minimum)

If you remember from parameter estimation:

$$L = \ln\left(prob\left(X \mid \{data\} \right) \right) = \ln\left(likelihood \right) \equiv \ln\left(\ell \right) \qquad \sigma_X^2 = \left(-\left. \frac{d^2 L}{dX^2} \right|_{X_0} \right)^{-1}$$

Bayesianilly speaking, we can thus define the matrix

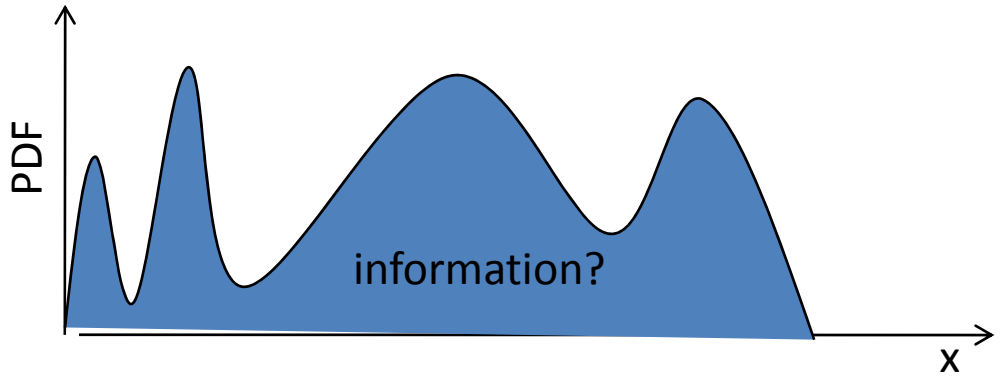
$$\begin{pmatrix} \dfrac{\partial^2 \ln \ell}{\partial X^2} & \dfrac{\partial^2 \ln \ell}{\partial X \partial Y} \\ \dfrac{\partial^2 \ln \ell}{\partial Y \partial X} & \dfrac{\partial^2 \ln \ell}{\partial Y^2} \end{pmatrix}$$

Fischer information matrix

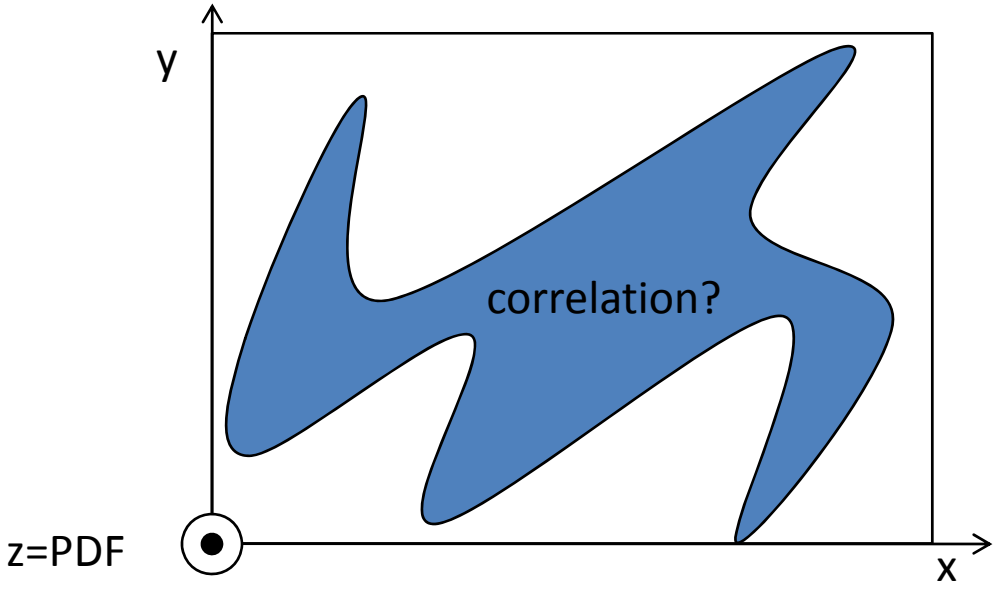
But we can go to information theory to avoid these problems, rejoyce!!!

Information theory tells us:

- How many information is carrying a PDF



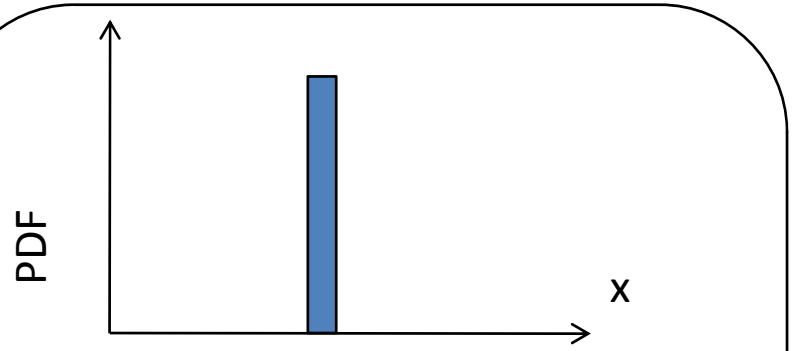
- How correlated is the information in a multidimensional PDF



Information theory tells us:

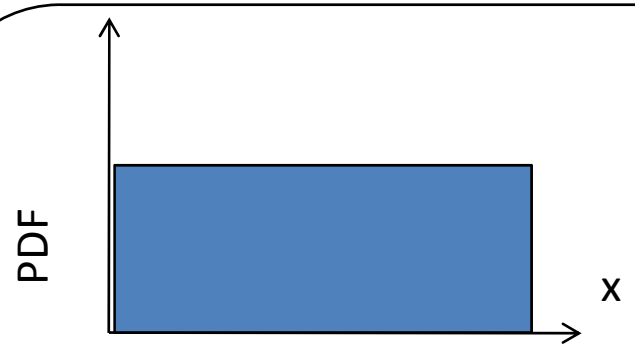
- How many information is carrying a PDF:

Shannon ENTROPY $$H(x) = -\sum_{i=1}^{n} P_i(x) \ln P_i(x) = 0$$



$$H_{\min} = -\sum_{i=1}^{n} 1 \ln 1 = 0$$

"SHARP" PDF



$$H_{\max} = -\sum_{i=1}^{n} \frac{1}{n} \ln \frac{1}{n} = \ln n$$

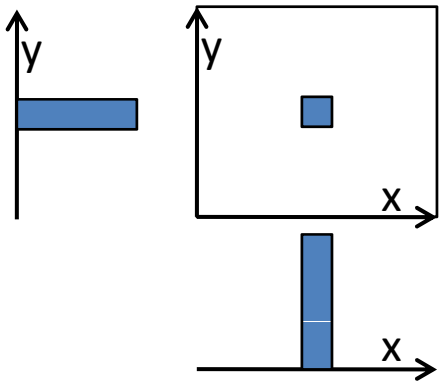
Non informative PDF

Information theory tells us:

- How correlated are x,y in a PDF

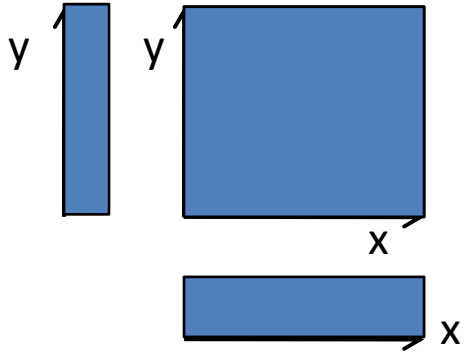
MUTUAL INFORMATION

$$I(x,y) = H(x) + H(y) - H(x,y)$$



$$I(x,y) = H(x) + H(y) - H(x,y)$$

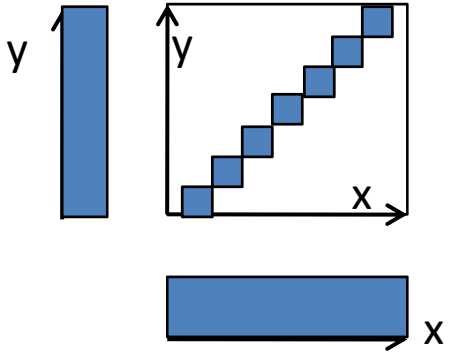
$$I_{\min}(x,y) = 0 + 0 - 0 = 0$$



$$I(x,y) = H(x) + H(y) - H(x,y)$$

$$I_{\min}(x,y) = n + n - 2n = 0$$

x and y NOT correlated



$$I(x,y) = H(x) + H(y) - H(x,y)$$

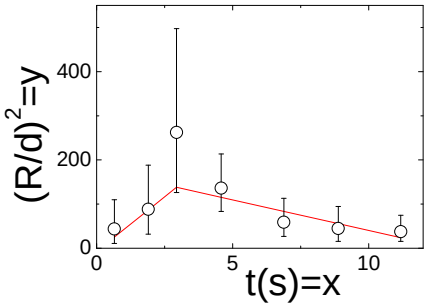
$$I_{\max}(x,y) = n + n - n = n$$

x and y strongly correlated

MUTUAL INFORMATION

$$I(x,y) = H(x) + H(y) - H(x,y)$$

model (astrophysics)

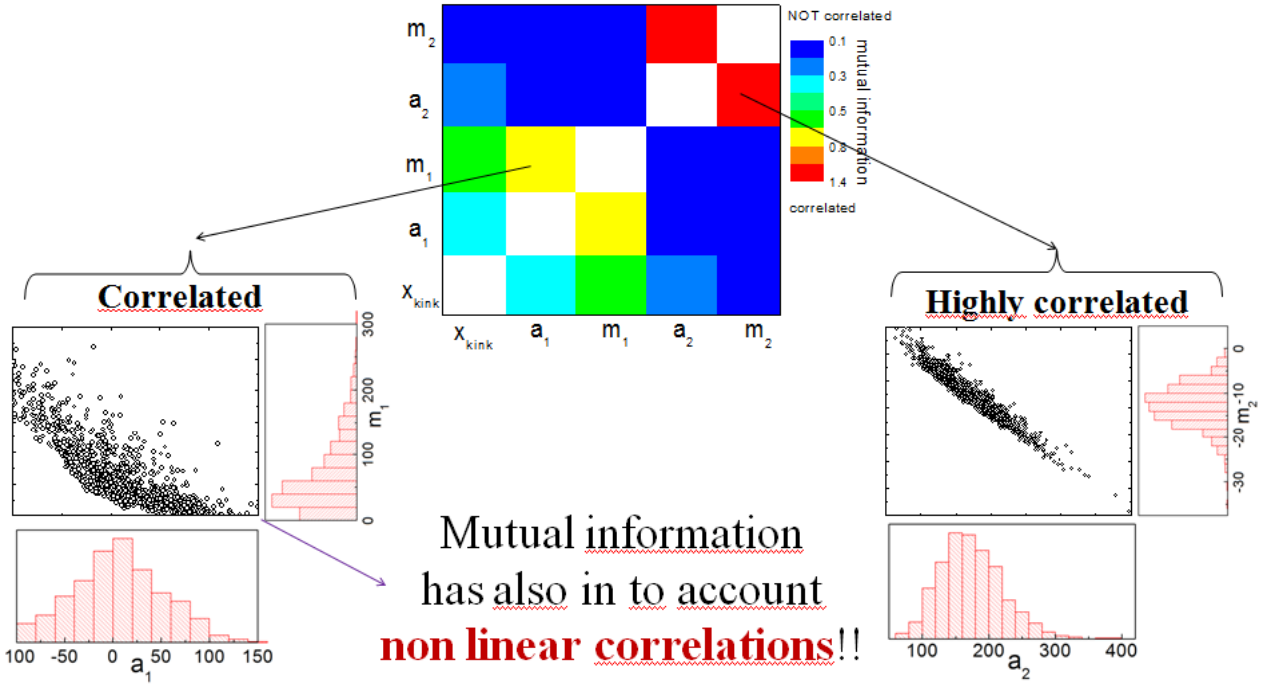


For $x<x_{kink}$ $\quad y_1 = a_1 + m_1 x$

For $x>x_{kink}$ $\quad y_2 = a_2 + m_2 x$

Continuity: $\quad a_2 = a_1 + (m_1 - m_2)\, x_{kink}$

We calculate MI for all combinations of two parameters



Information theory tells us:

- How many information is carrying a PDF

Shannon ENTROPY

$$H(x) = -\sum_{i=1}^{n} P_i(x) \ln P_i(x) = 0$$

- How correlated is the information in a multidimensional PDF

MUTUAL INFORMATION

$$I_n(A_1,\ldots A_n) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{i_1<\ldots<i_k} S(A_{i_1},\ldots A_{i_k})$$