Probability rules.

- Two approaches
  - Frequentists look at the frequency of occurrence of an event given a number of trials. Probabilities are associated to the limit of infinite trials
  - Bayesian view: look at the degree of confidence in a hypothesis: Use prior information, probability theory, **judgement.**

Probability rules.

- Two approaches
  - Frequentists look at the frequency of occurrence of an event given a number of trials. Probabilities are associated to the limit of infinite trials
  - Bayesian view: look at the degree of confidence in a hypothesis: Use prior information, probability theory, **judgement.**
- **Probability rules**
  - **P(A|B)** degree for which a proposition B implies A to be true.
  - Non-negative, normalized, linear P(a) +P(b) = P(a+b) for exclusive events a and b
  - P(a,b) = P(a|b) P(b)   Product rule: Probability of a and b.
  - $P(a) = \sum_i P(a, b_i)$ Sum → Normalization
  - Bayes theorem stems from P(a,b) = P(b,a)

Limiting the problem: Parameter fitting.

- Bayesian theorem

priors likelihood

$$\text{posterior} \quad \mathcal{P}(H|D) = \frac{\mathcal{P}(H)\mathcal{P}(D|H)}{\mathcal{P}(D)}$$

evidence

From

$$\mathcal{P}(a,b) = \mathcal{P}(a)\mathcal{P}(b|a) \quad ; \qquad \mathcal{P}(b,a) = \mathcal{P}(b)\mathcal{P}(a|b)$$

- Priors: State of knowledge about parameters before data.
- Hypothesis in the form of vector of θ parameters
- Likelyhood (Hypothesis): what you can ask to a *”figure of merit”* is ”Given a set of parameters, what is the probability that this data set could have occurred?” In practice, a normal distribution when errors are Gaussian. Exp of the χ-square distribution.
- Evidence: Can be ignored for parameter inference (central for model comparison)

Limiting the problem: Parameter fitting.

- Bayesian theorem

$$\mathcal{P}(H|D) = \frac{\mathcal{P}(H)\mathcal{P}(D|H)}{\mathcal{P}(D)}$$

From $\mathcal{P}(a,b) = \mathcal{P}(a)\mathcal{P}(b|a)$; $\mathcal{P}(b,a) = \mathcal{P}(b)\mathcal{P}(a|b)$

- The fundamental issue. We are looking at **statistical inference.**

INFERENCE:

P is a degree of belief, what we want to determine is $P(H|D)$ and use Bayes theory. Probabilities are conditional to Prior information I and a model M that produces the hypothesis

$$P(H|D) = P(H|D, I, M)$$

So $P(H) = P(H|M)$

For inference we can ignore evidence (Probability of the data)

Limiting the problem: Parameter fitting.

- Performing **statistical inference:** adjusting the parameters to maximize the agreement one obtains the *best fit parameters*.
- a fitting procedure should provide
  - *a)* best fit parameters
  - *b)* error estimates on the parameters
  - *c)* possibly a statistical measure of the goodness of fit. (need priors)

The bayesian theory approach

- Likelyhood (Hypothesis): what you can ask to a *”figure of merit”* is ”Given a set of parameters, what is the probability that this data set could have occurred?”

- Chi-Square method: Model fitting and parameter estimation. The formula is given by

$$\chi^2 = \sum_i V_i \left[y(x_i) - Y(x_i, \bar{\theta})\right]^2$$

y is the model predictions with params θ for each event x_i . Y_i = data points for event x_i.
Minimum variance weights $V_i = 1/\sigma^2$

In many cases, we rely on the central limit theorem and the Likelyhood can be approximated by multivariate Gaussian

$$\mathcal{P}(y|\bar{\theta}) = \mathcal{L}(y|\bar{\theta}) = \frac{1}{(2\pi)^n |\mathrm{det} C|^{1/2}} \exp\left[-\frac{1}{2}\sum_{ij}(y(x_i) - Y(x_i, \bar{\theta}))C_{ij}(y(x_j) - Y(x_j, \bar{\theta}))\right]$$

Where *C* is the covariance matrix.

The maximum likelyhood

- If we set P(D) = 1 (Have all the data) so that Evidence = 1, and ignore the priors (Degree of believe in your model), then from Bayes theorem, by maximising likelihood, we find the most likely hypothesis, or the most likely parameters of a given model.
- Ignoring priors: Technique does not provide a goodness of fit.
- Ignoring evidence: we can only compute relative probabilities. This proba is not invariant among models. Relevant in model selection.

$$P(D|M) \neq P(D|M')$$

$\Delta\chi^2$ as a Function of Confidence Level and Degrees of Freedom

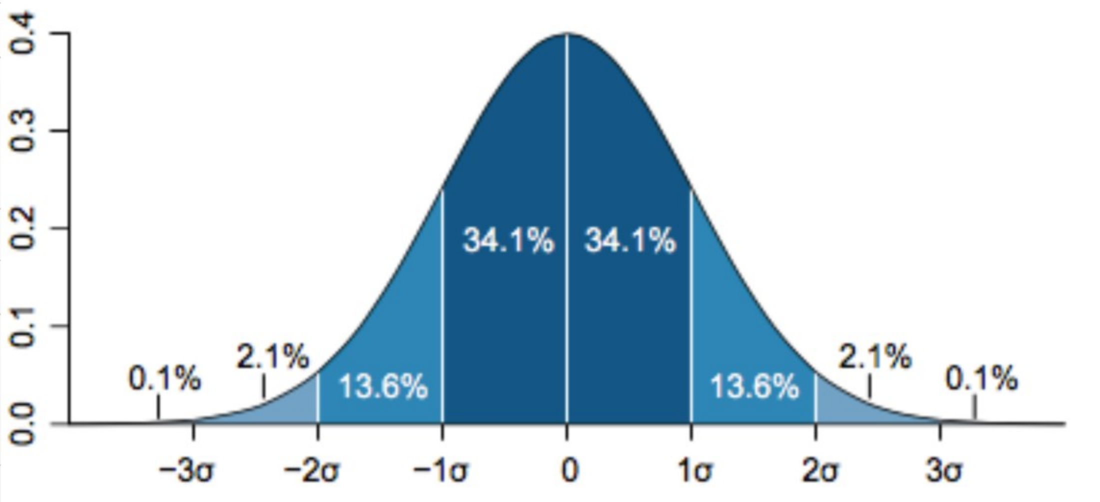
	ν					
p	1	2	3	4	5	6
68.3%	1.00	2.30	3.53	4.72	5.89	7.04
90%	2.71	4.61	6.25	7.78	9.24	10.6
95.4%	4.00	6.17	8.02	9.70	11.3	12.8
99%	6.63	9.21	11.3	13.3	15.1	16.8
99.73%	9.00	11.8	14.2	16.3	18.2	20.1
99.99%	15.1	18.4	21.1	23.5	25.7	27.8

Confidence regions

- For Bayesian statistics, integrated probability in regions of the parameter space are **Confidence Regions R**

$$\int_R \mathcal{P}(\vec{\alpha}|D)d\vec{\alpha}$$

- **HOW DO WE INTEGRATE THEM?**



The Metropolis Hastings for SNe

- What we did was to look at the likelyhood (log-like), of the one parameter set $\{\Omega_\Lambda\}$. Now look at an acceptable value and errors of this inference by sampling the parameter space.

- Bayesian statistics yields the better way to do this by associating probabilities to parameter values rather than finding the best value to fit the data.

- Other set of parameters $\{\Omega_\Lambda, w\}$ or other theories $\{R^n, \Omega_\Lambda\}$ can be tested

The Metropolis Hastings for SNe

Repeat this a thousand times!

Step 1: Pick a point to start the walk $(\Omega_{\Lambda 0})$
Step 2: Pick a second point in the neigh

$$h_1 = h_0 + \Delta \cdot GaussRand(0,1)$$

Step 3: Compute the likelihood of the new point and compare it to the previous point

$$\text{Ratio} = \log\left({}^{\mathcal{L}(Y|\Theta_1)}/_{\mathcal{L}(Y|\Theta_0)}\right) = \chi^2(\Theta_0) - \chi^2(\Theta_1)$$

if Ratio > 0 then accept the new point as the next step
If Ratio < 0 then
  if exp(Ratio) $> Rand(0,1)$ then accept the point
  if exp(Ratio) $< Rand(0,1)$ then keep the old point as the new point

Plot your results

- Create a mesh of N bins per side and also divide the your data (maxima and minima of the set of parameters $\Theta = (\Omega_\Lambda)$) in N intervals Δ_Θ.

$$graph = \mathrm{np.zeros}\big[[\mathrm{N}]\big]$$
$$inter_\Theta = (max_\Theta - min_\Theta)/N$$

- Populate the bins: for each step of the Markov chain compute, for each parameter,

$$K_\Theta = \text{Integer}\Big[(\Theta_i - \mathrm{min}_\Theta)/\Delta_\Theta\Big]$$

- Add an integer (add 1) to the K_Θ bin in each axis in the mesh plot

$$graph\big[K_{\Omega_\Lambda}\big] \mathrel{+}= 1$$

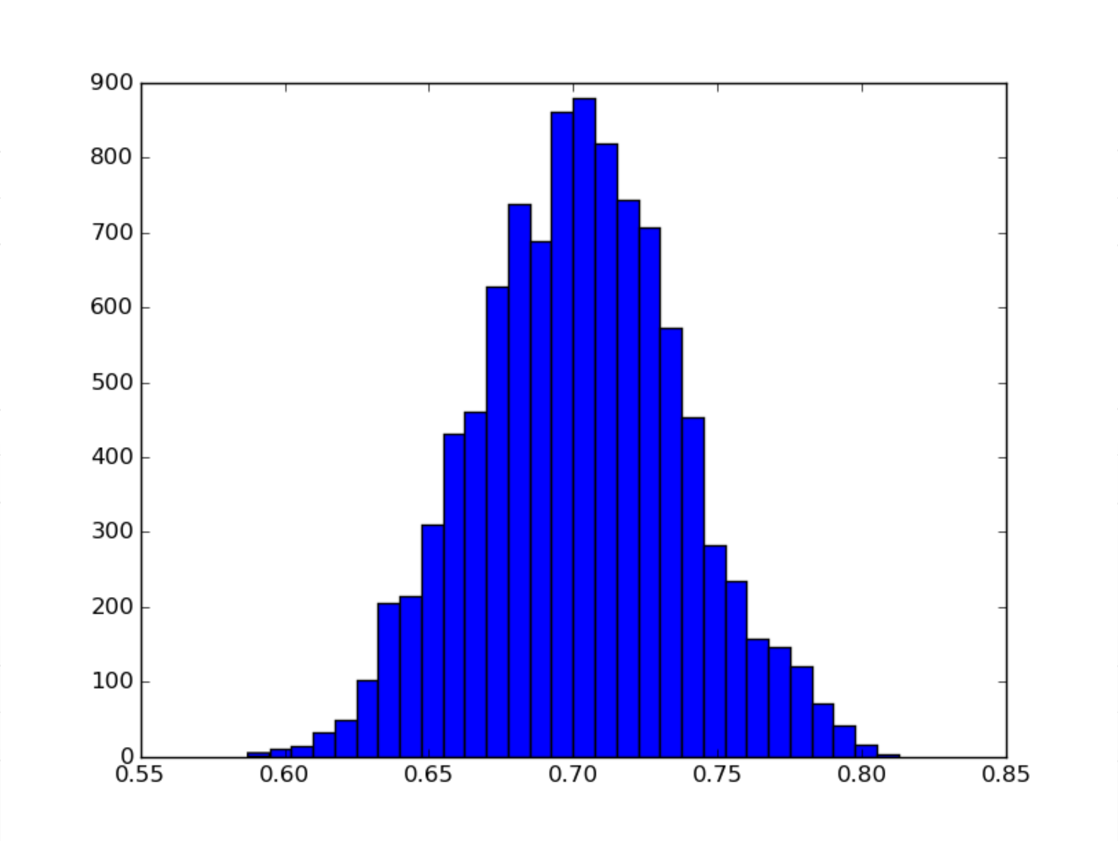
- Plot the mesh

$$plt.colormesh(inter_h, inter_{\Omega_\Lambda}, graph)$$

- **How do YOU pick confidence regions for THIS one parameter??**

Plot your results

- Make a plot for the Ω_Λ parameter



The Metropolis Hastings for SNe (2 params)

- What we did was to look at the likelyhood (log-like), of the one parameter set $\{\Omega_\Lambda\}$. Now look at an acceptable value and errors of this inference by sampling the parameter space

- Bayesian statistics yields the better way to do this by associating probabilities to parameter values rather than finding the best value to fit the data.

- Other set of parameters $\{\Omega_{\mathrm{m}}, \Omega_\Lambda\}$ or other theories $\{R^n, \Omega_\Lambda\}$ can be tested

The Metropolis Hastings for 2-param SNe

Step 1: Pick a point to start the walk $(\boldsymbol{\Omega_{\Lambda 0}}, \boldsymbol{w})$

Step 2: Pick a second **oriented** point in the neigh

$$\overrightarrow{h_1} = \overrightarrow{h_o} + GaussRand(0,1) * \overleftrightarrow{CH_{ij}}$$

Step 3: Compute the likelihood of the new point and compare it to the previous point

$$\text{Ratio} = \log\left({}^{\mathcal{L}(Y|\Theta_1)}\big/_{\mathcal{L}(Y|\Theta_0)}\right) = \chi^2(\Theta_0) - \chi^2(\Theta_1)$$

if Ratio > 0 then accept the new point as the next step

If Ratio < 0 then

if exp(Ratio) $> Rand(0,1)$ then accept the point

if exp(Ratio) $< Rand(0,1)$ then keep the old point as the new point

Repeat this a thousand times!

Plot your results

- Create a mesh of N bins per side and also divide the your data (maxima and minima of the set of parameters $\Theta = (\Omega_\Lambda, w)$ in N intervals Δ_Θ.

$$graph = \mathrm{np.zeros}\big[[\mathrm{N}], [\mathrm{N}]\big]$$
$$inter_\Theta = (max_\Theta - min_\Theta)/N$$

- Populate the bins: for each step of the Markov chain compute, for each parameter,

$$K_\Theta = \text{Integer}\Big[(\Theta_i - \mathrm{min}_\Theta)/\Delta_\Theta\Big]$$

- Add an integer (add 1) to the K_Θ bin in each axis in the mesh plot

$$graph\big[K_{\Omega_\Lambda}\big] += 1$$

- Plot the mesh

$$plt.colormesh(inter_h, inter_{\Omega_\Lambda}, graph)$$

- **pick confidence regions for THIS Two params**