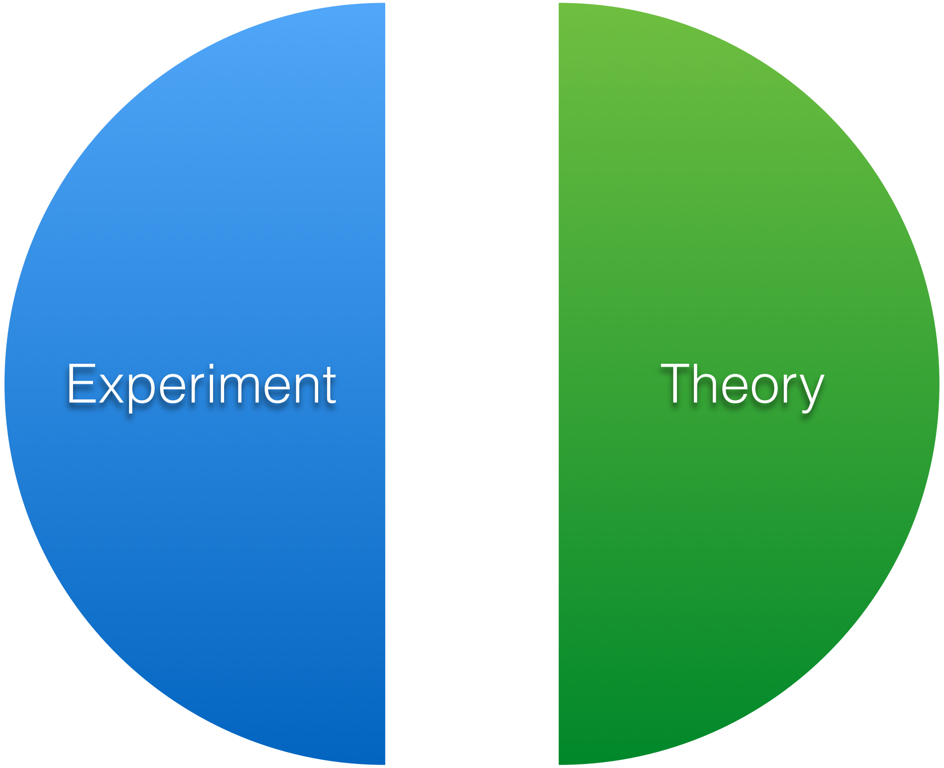
Statistics for Astronomy

Statistical Inference

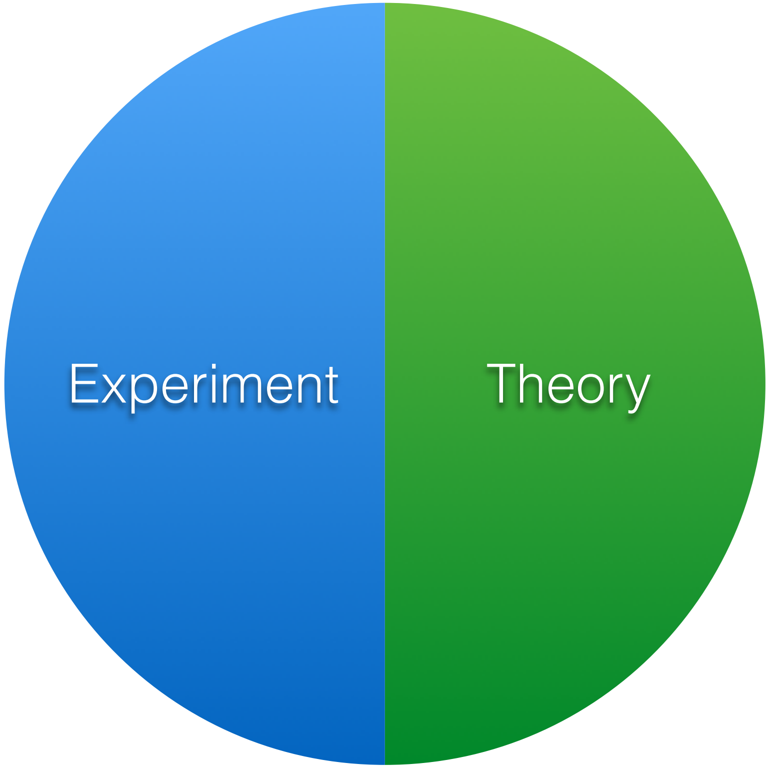
Adam Amara

ETH zürich

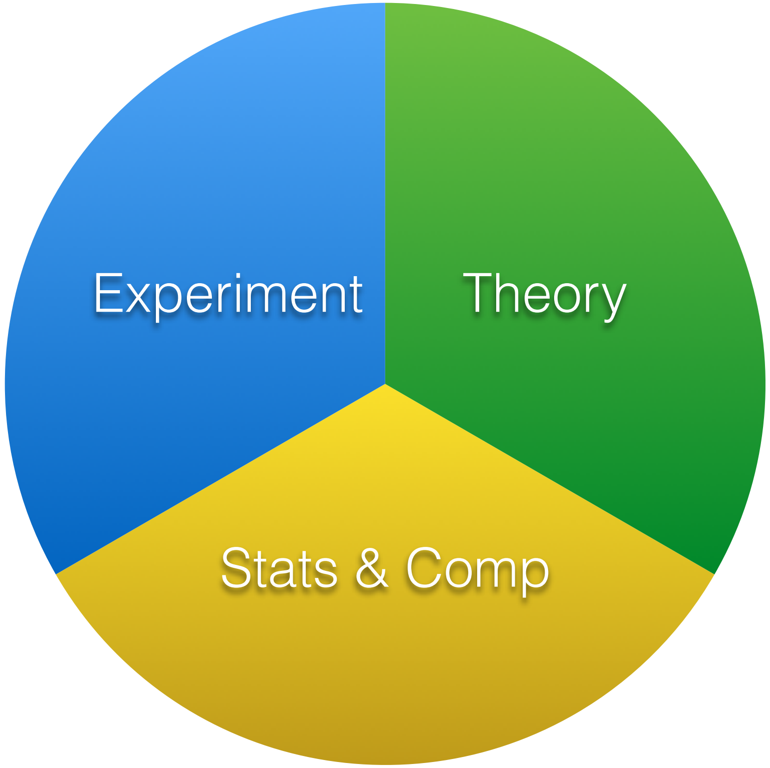
Domains in Physics



Domains in Astrophysics

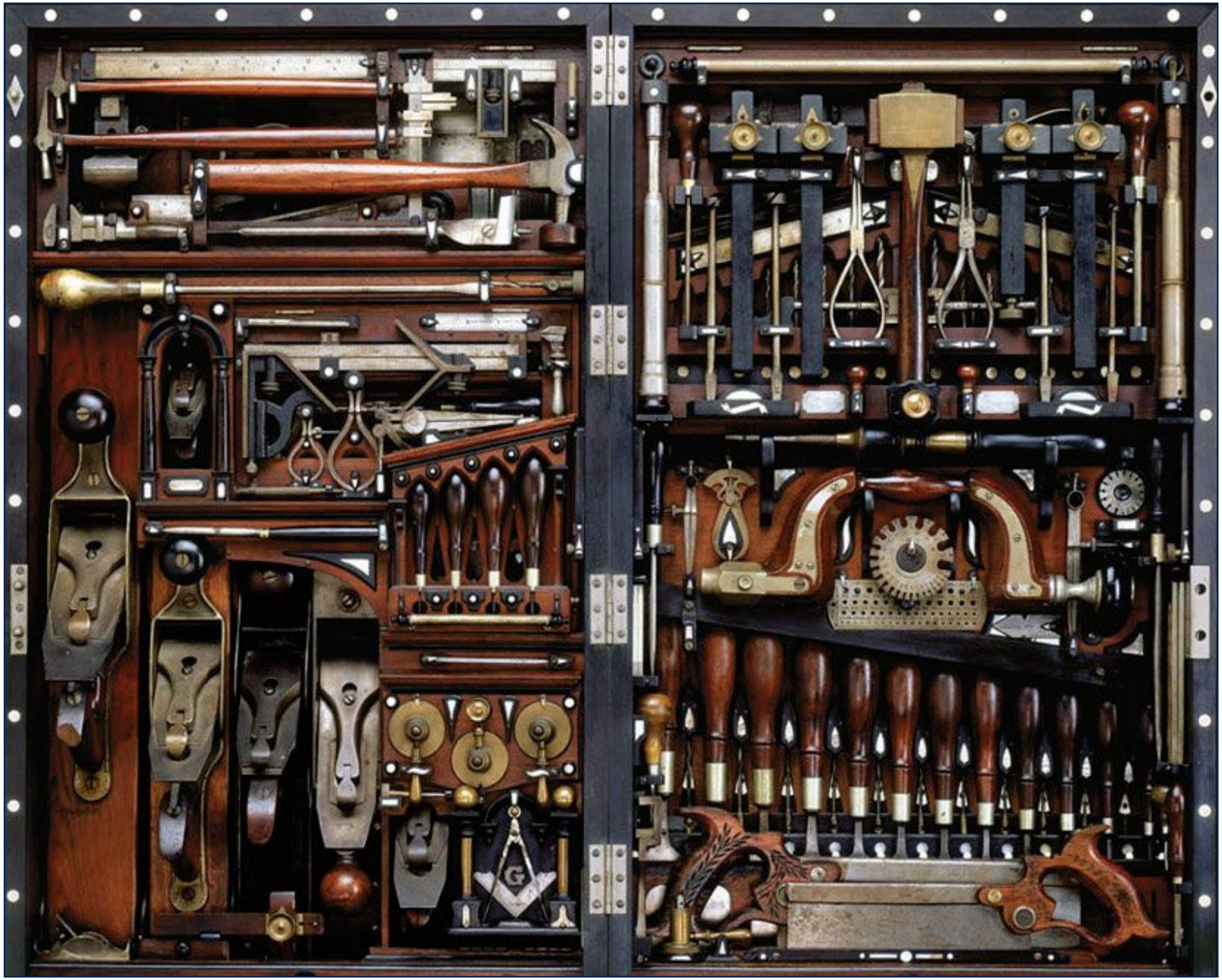


Domains in Astrophysics



The Aim

The Aim



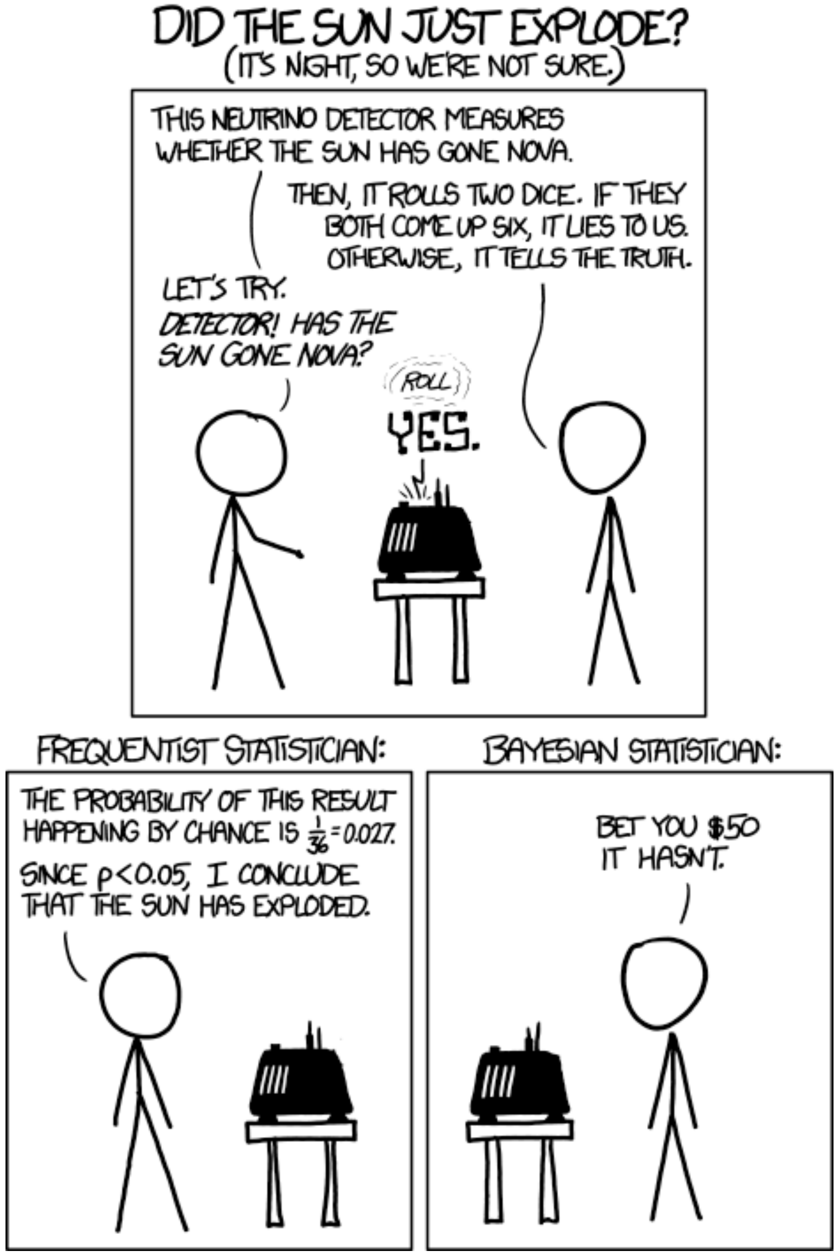
Bayesians, Frequentists and other

Probability

$$p(\theta|d) \propto p(d|\theta)p(\theta)$$

'prior'

'degree of belief'



Experimental errors and uncertainties should be treated probabilistically

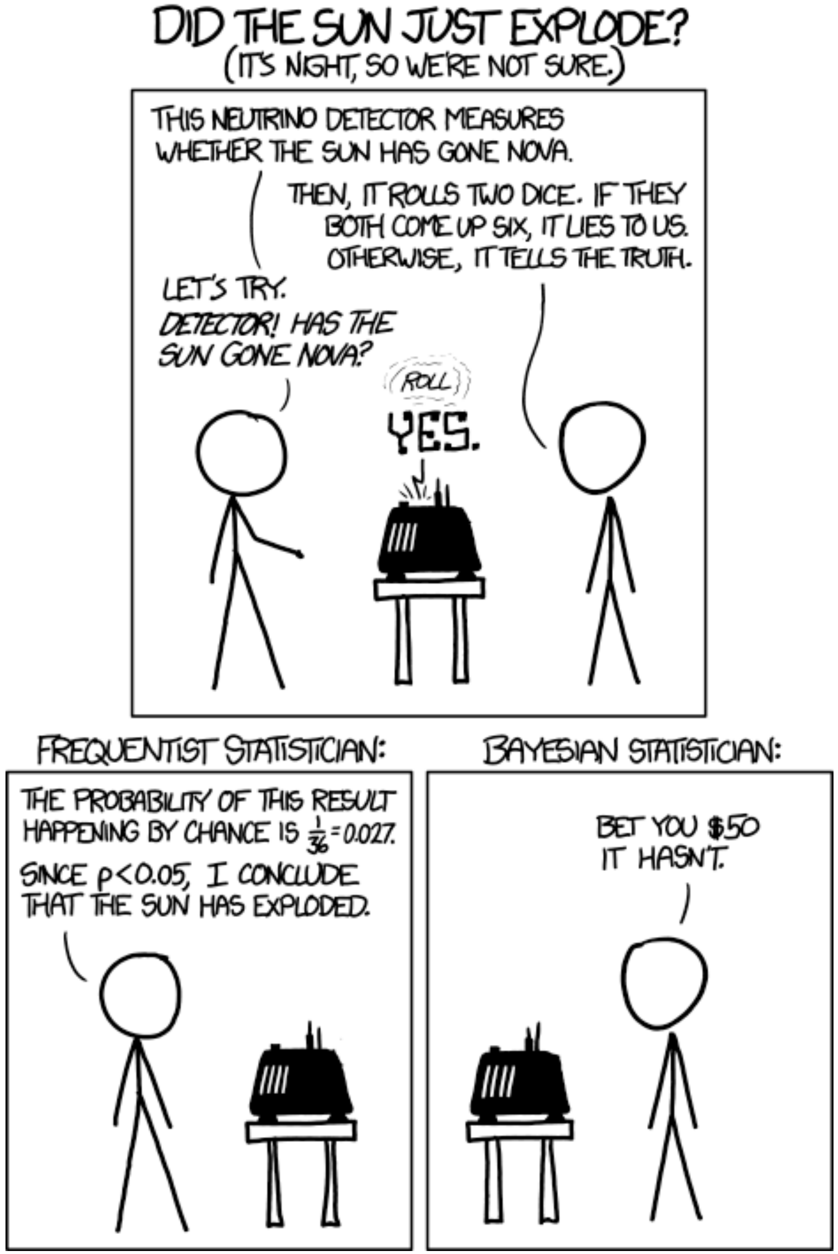
Treatment of probabilities should be mathematically sounds

Interpretation of probability (subjective vs objective) depends on the situation

Bayesians, Frequentists and other

It is possible to do garbage statistics as a 'Bayesian' or 'Frequentist': i.e. just because you use the word 'prior', does not mean you know what you're doing

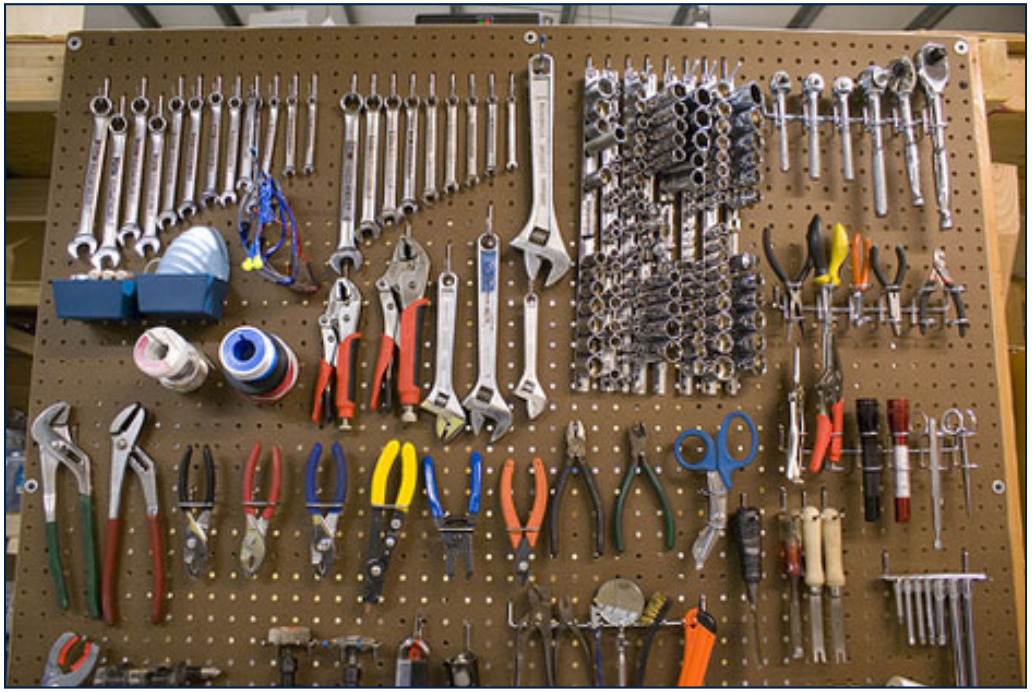
Best is to understand the structure of your problem (e.g. your error model) and to use mathematically sound probabilities.



Good overview of available tools

Well organised and maintained

Know which tool for which job



Relative Entropy

Hierarchical Models

K-S test

Maximum Likelihood

p-value

Joint Probability

Minimum 'χ^2'

Hypothesis testing

Graphical Models

Bayes Theorem

Host of Classical Tests

Student t-test

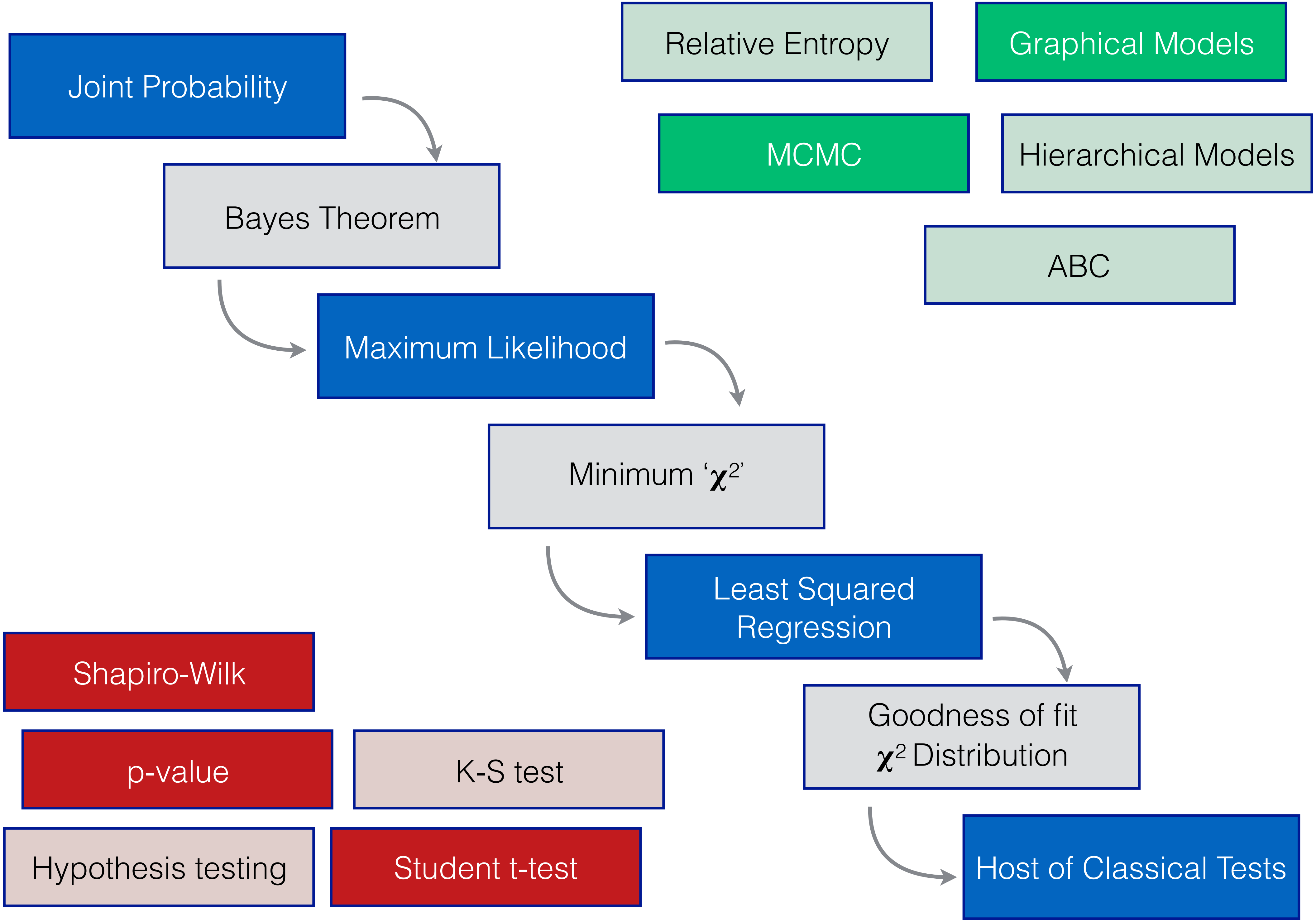
MCMC

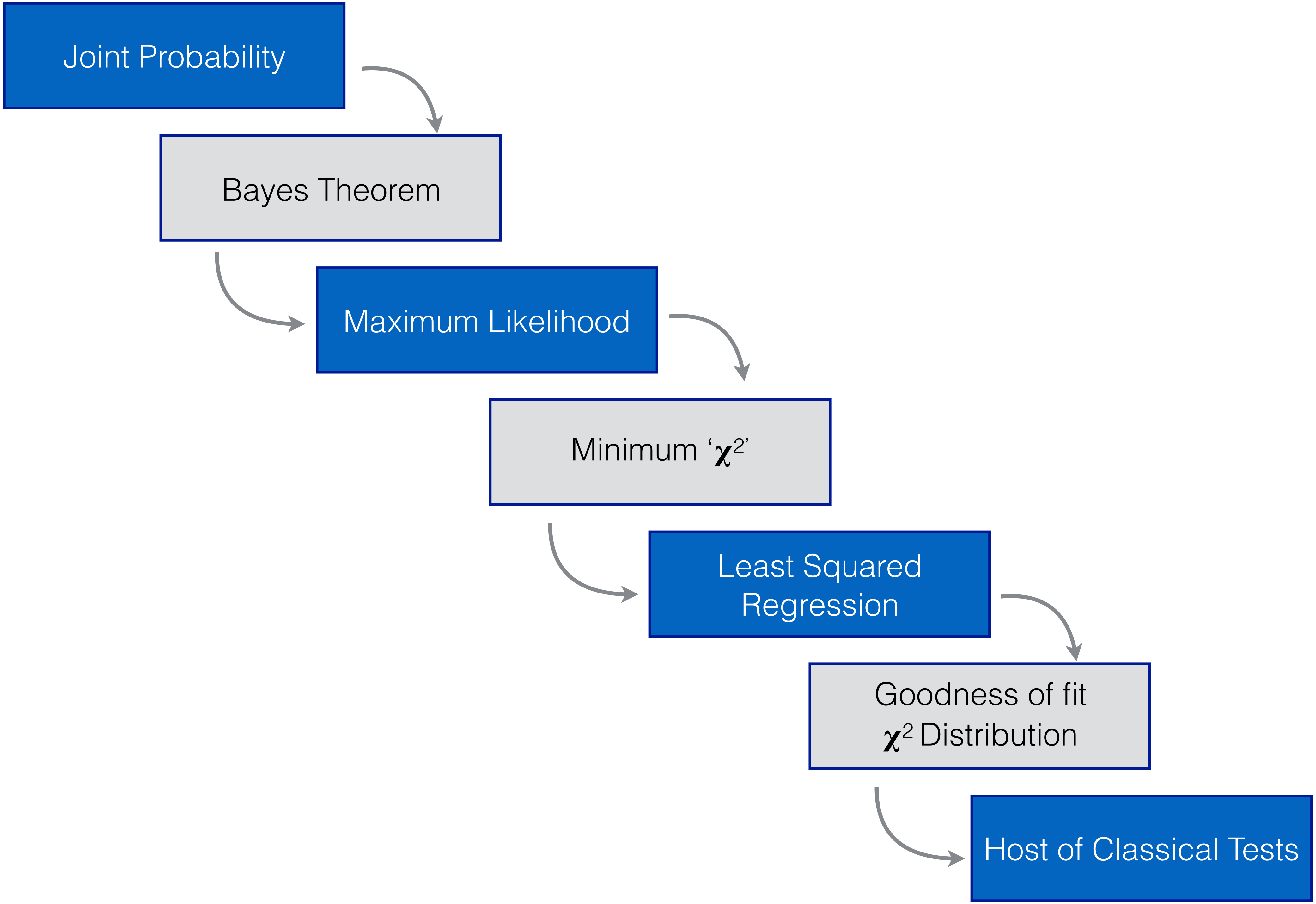
Shapiro-Wilk

Goodness of fit
χ^2 Distribution

Least Squared
Regression

ABC


Relative Entropy
Graphical Models
Joint Probability
MCMC
Hierarchical Models
Bayes Theorem
ABC
Maximum Likelihood
Minimum 'χ^2'
Least Squared Regression
Shapiro-Wilk
Goodness of fit
χ^2 Distribution
p-value
K-S test
Hypothesis testing
Student t-test
Host of Classical Tests


Joint Probability
Bayes Theorem
Maximum Likelihood
Minimum ‘χ^2’
Least Squared Regression
Goodness of fit χ^2 Distribution
Host of Classical Tests

Probabilities

Joint

$$p(W, H)$$

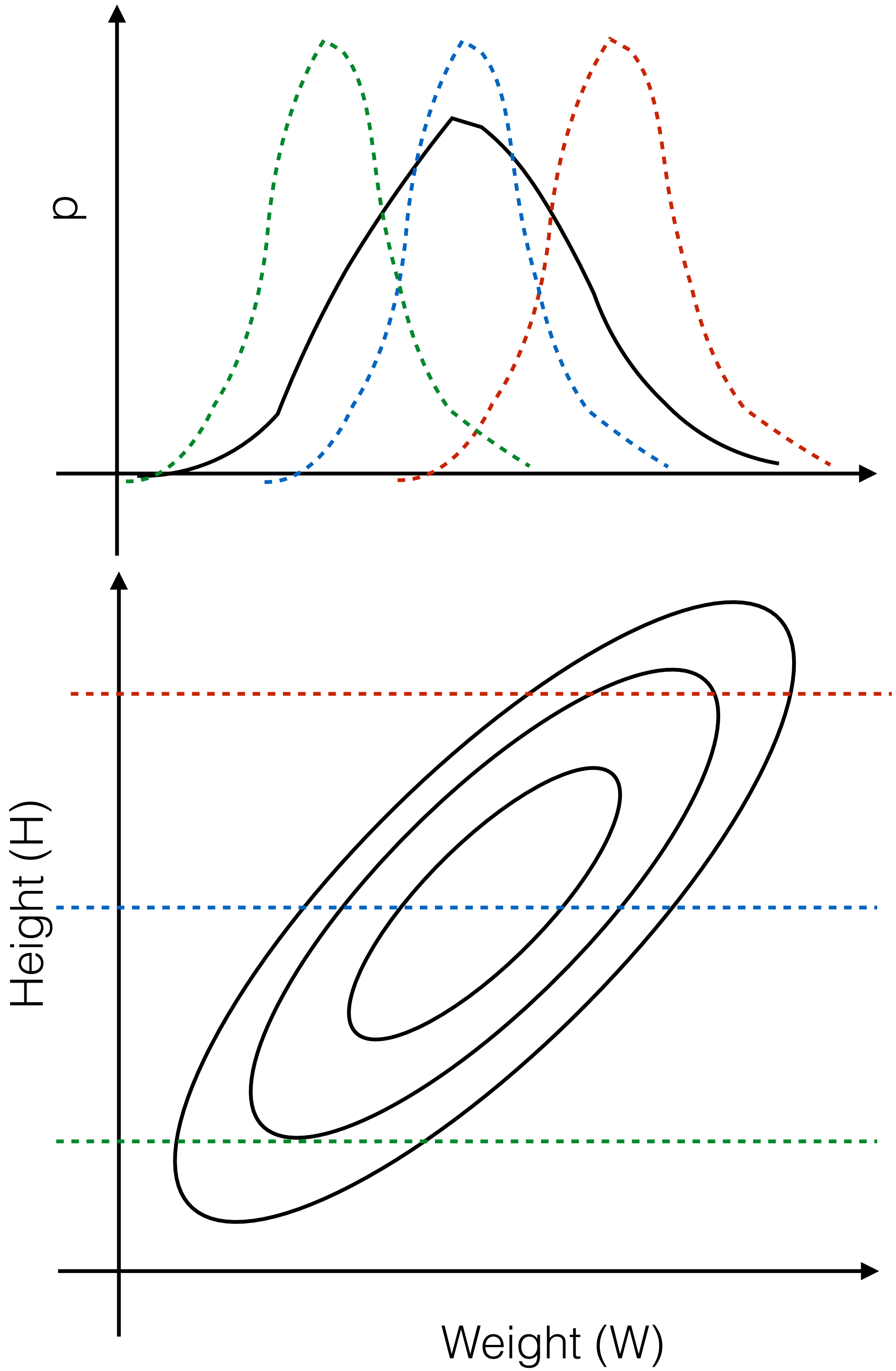
Marginalised

$$p(W) = \int p(W, H) dH$$

Conditioned

$$p(W|H)$$

$$\begin{aligned} p(W, H) &= p(H)p(W|H) \\ &= p(W)p(H|W) \end{aligned}$$

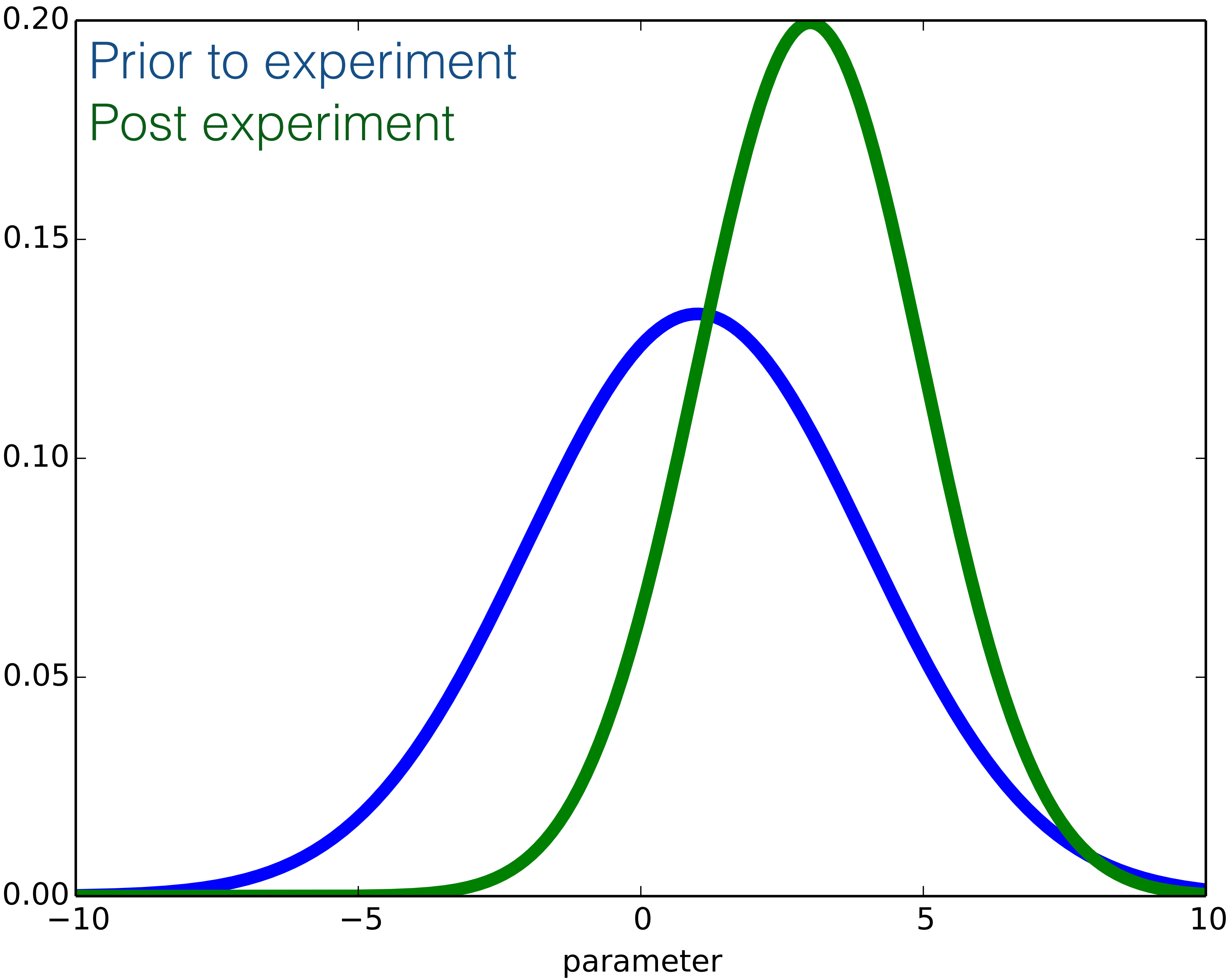


Bayes Theorem

Statistical inference problem:
Given some data (D) we wish to infer the parameters of a model (M)

$$p(M, D) = p(M|D)p(D) = p(D|M)p(M)$$

$$\underbrace{p(M|D)}_{\text{Posterior}} \propto \underbrace{p(D|M)}_{\text{Likelihood}}\underbrace{p(M)}_{\text{Prior}}$$

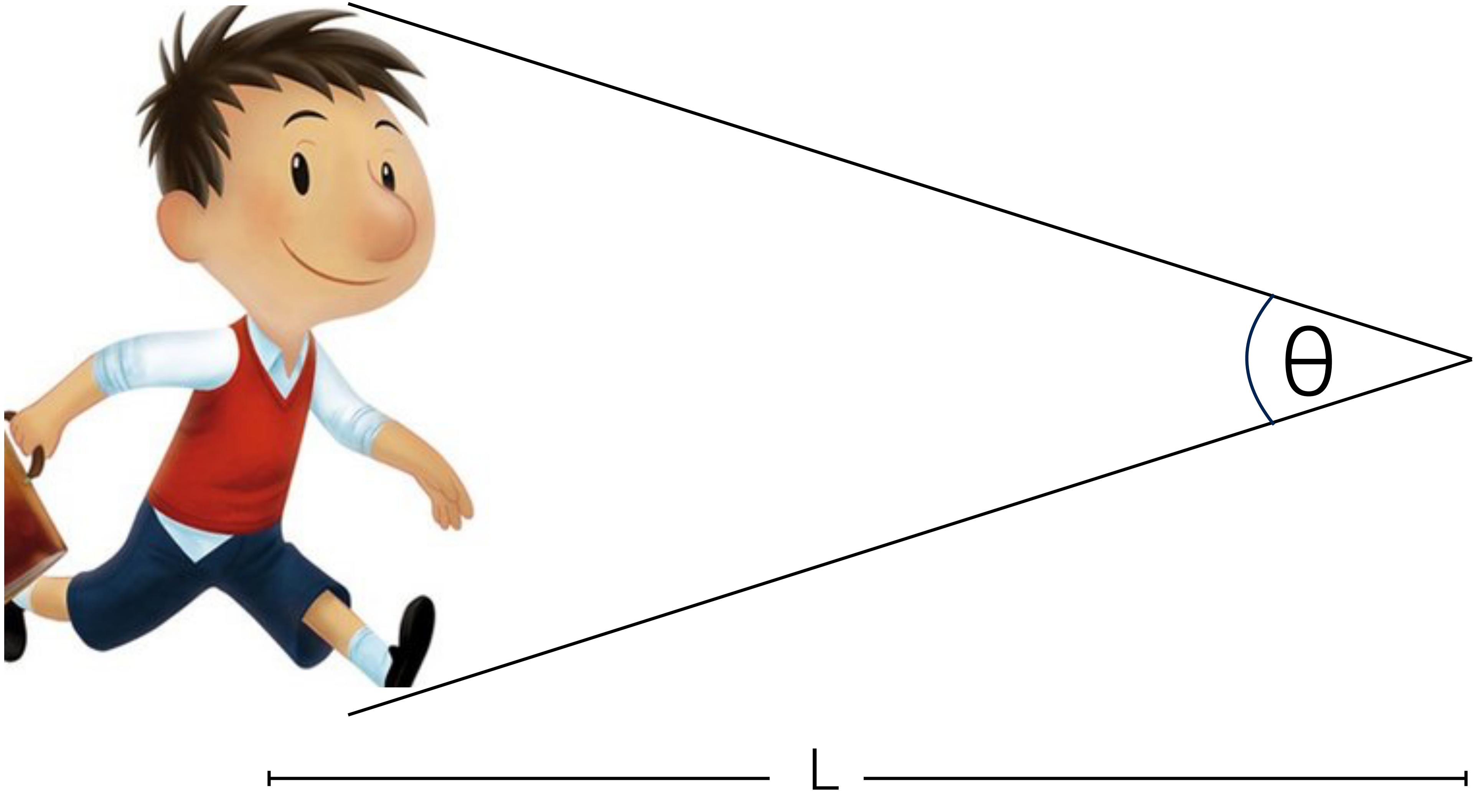


Prior to experiment
Post experiment
0.20
0.15
0.10
0.05
0.00
−10
−5
0
5
10
parameter

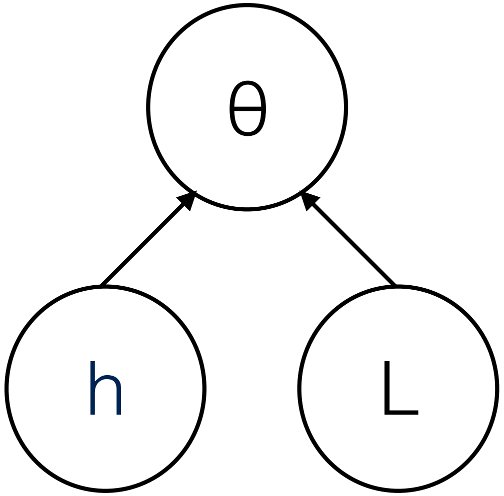
ETH zürich

Petit Nicola

$$p(h|\theta) \propto p(\theta|h)p(h)$$



Petit Nicola

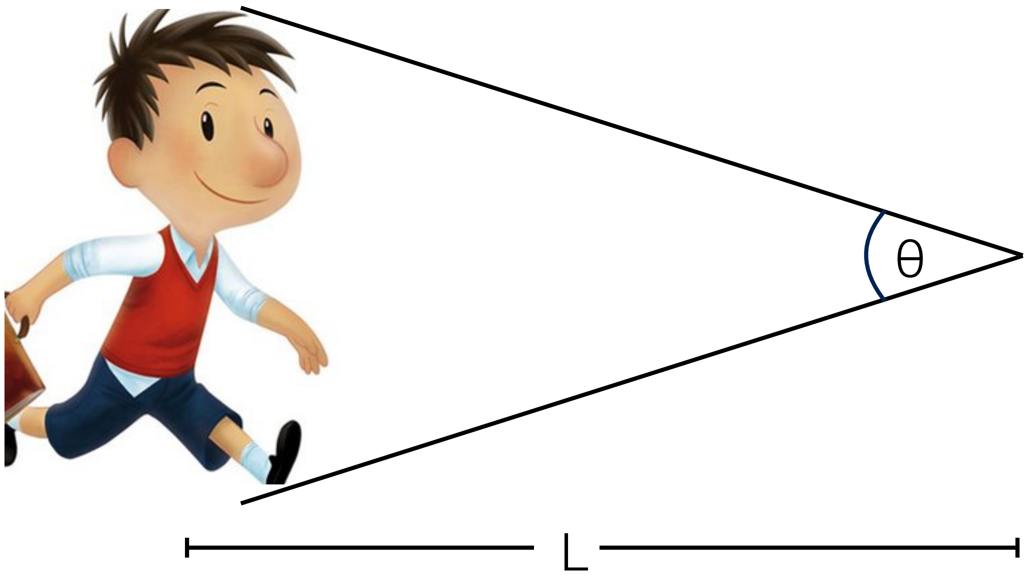


$$p(h|\theta) \propto p(\theta|h)p(h)$$

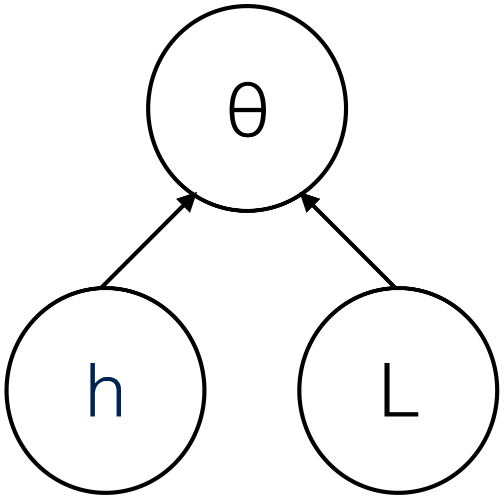
$$p(\theta|h) \propto \exp\left[-\frac{(h/L-\theta)^2}{2\sigma_\theta^2}\right]$$

$p(h) \longleftarrow$ what do we know before making a measure

$$\left(p(h) = \int p(h,\theta)d\theta\right)$$



Petit Nicola

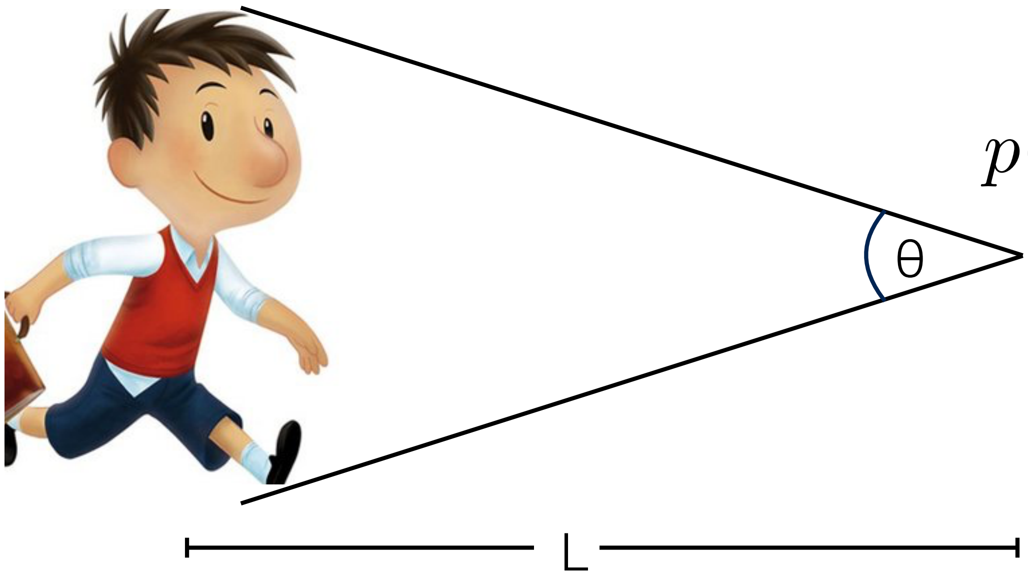


$$p(h|\theta)$$

$$p(h, L|\theta) \propto p(\theta|h, L)p(h)p(L)$$

$$p(h|\theta) = \int p(h, L|\theta)dL$$

$$p(\theta|h, L) \propto \exp\left[-\frac{(h/L-\theta)^2)}{2\sigma_\theta^2}\right]$$

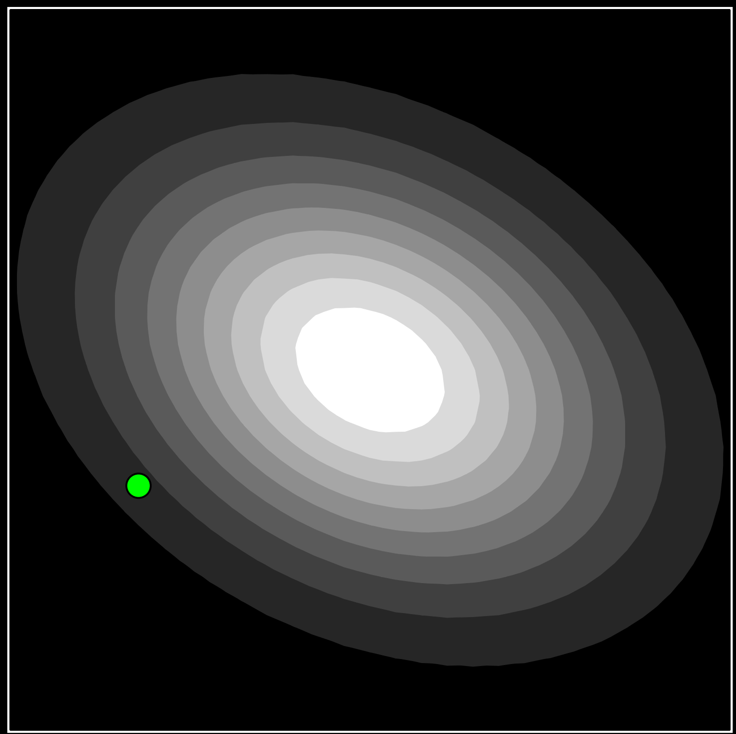


MCMC - Point sampling of pdf

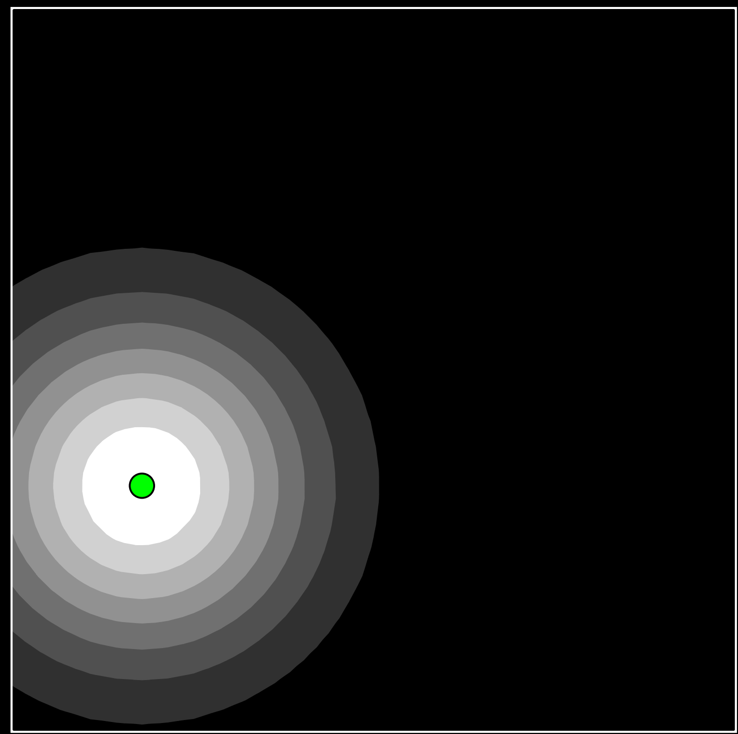
$$\langle x \rangle = \int x \ p(x) dx$$

$$\langle h(x) \rangle = \int h(x) \ p(x) dx$$

Metropolis-Hastings

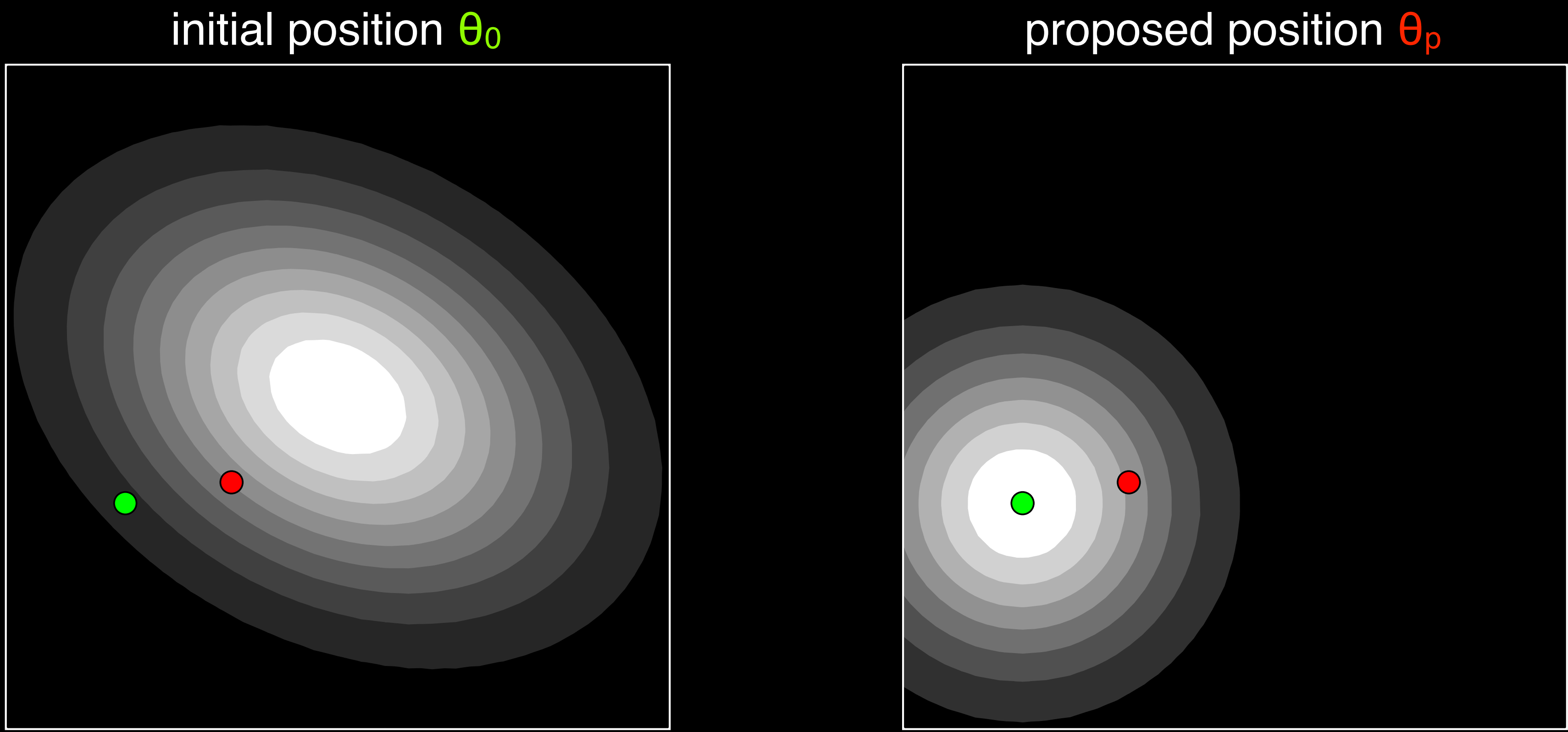


target distribution $P(\theta)$



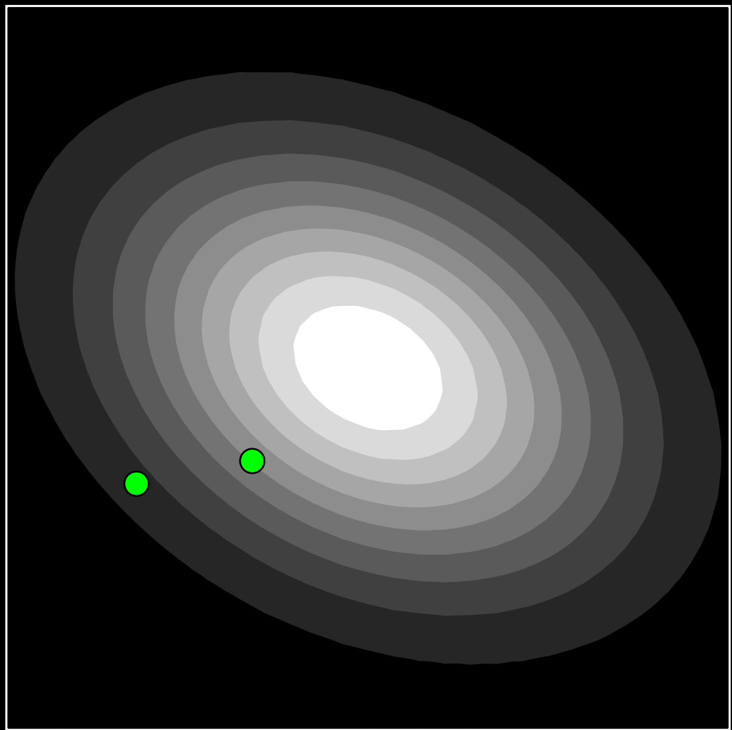
proposal distribution $Q(\theta)$

Metropolis-Hastings

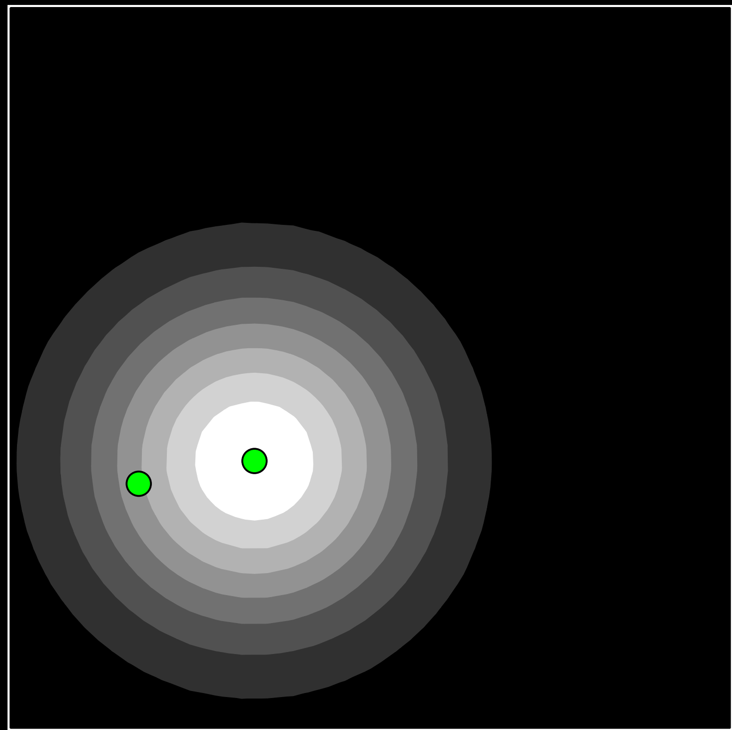


accepted with probability: $\min\left(1, P(\theta_p)/P(\theta_0)\right)$

Metropolis-Hastings

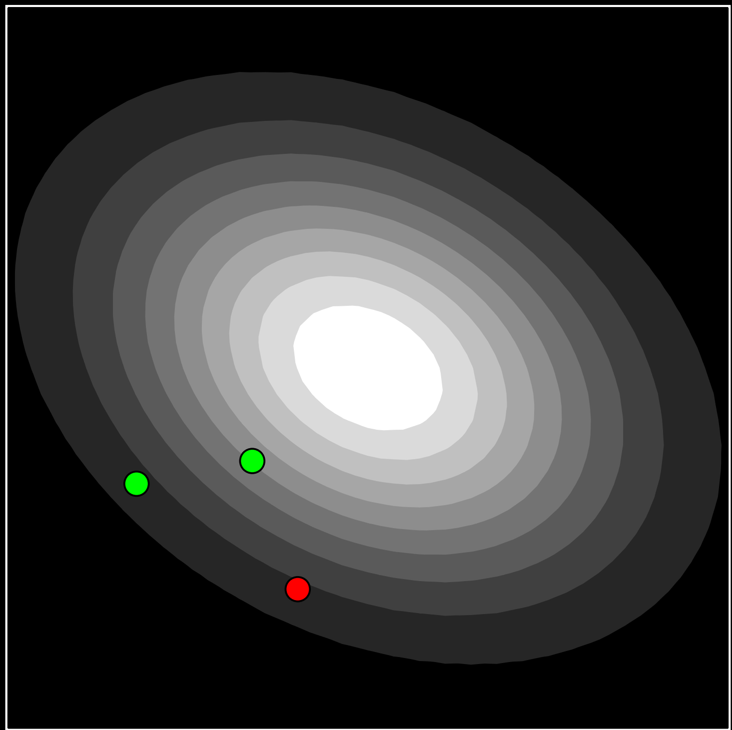


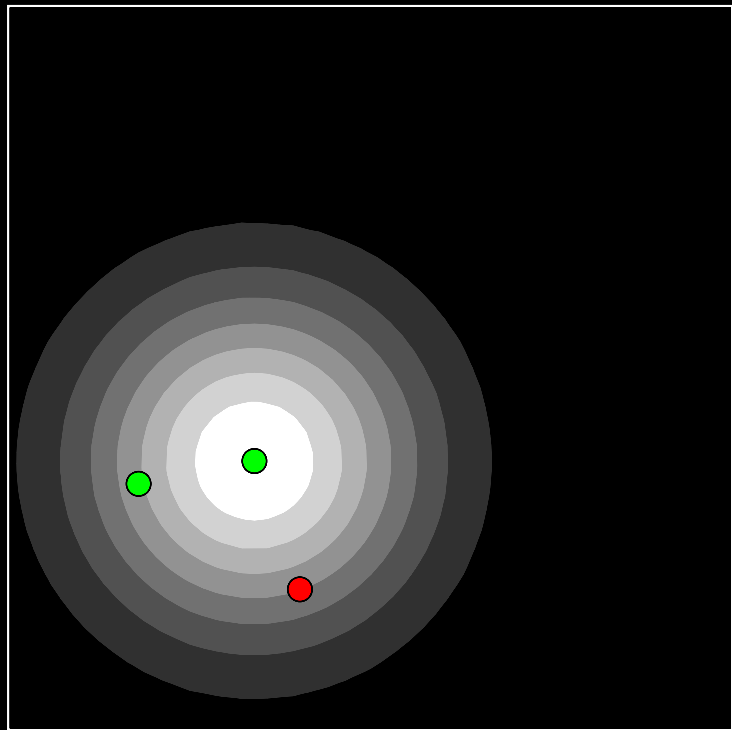
target



proposal

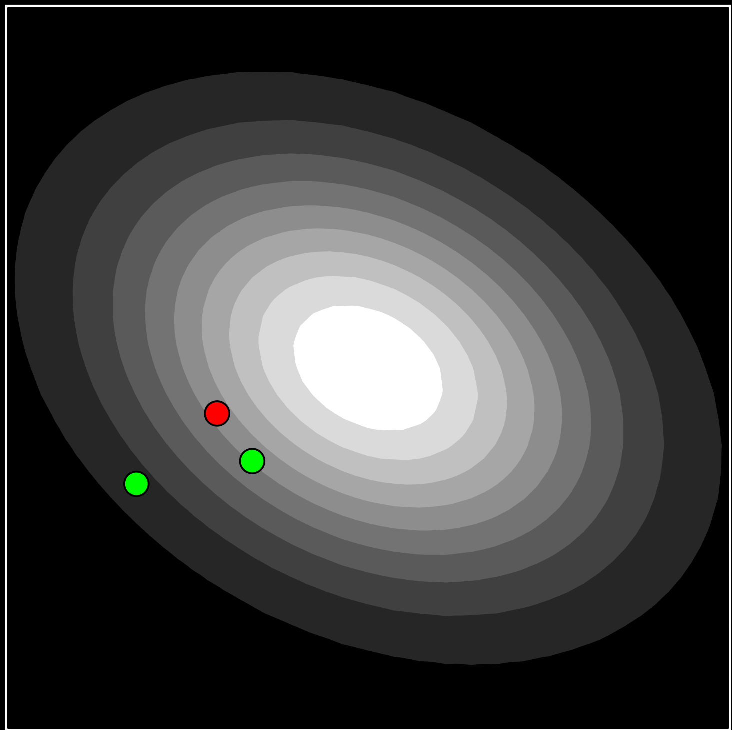
Metropolis-Hastings


target

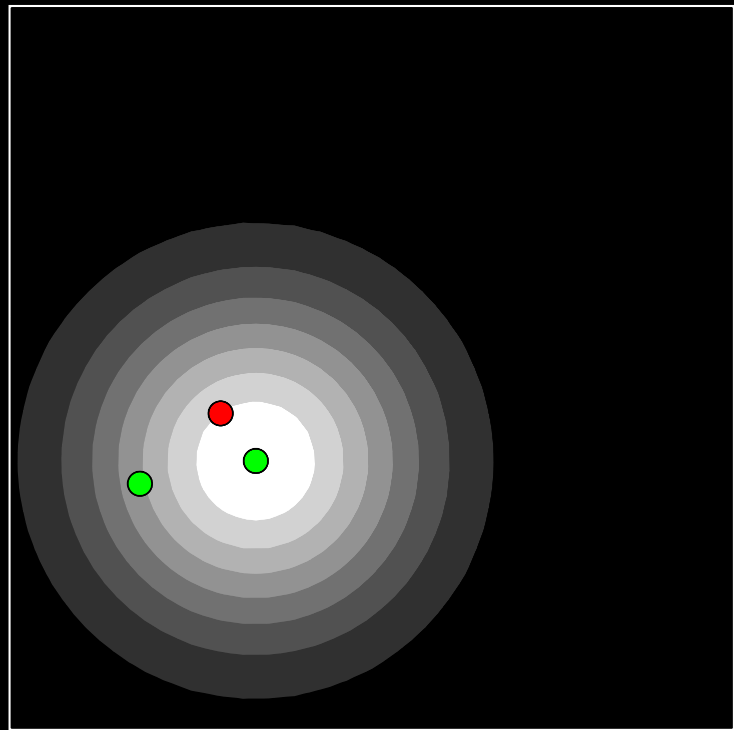

proposal

ETH zürich

Metropolis-Hastings

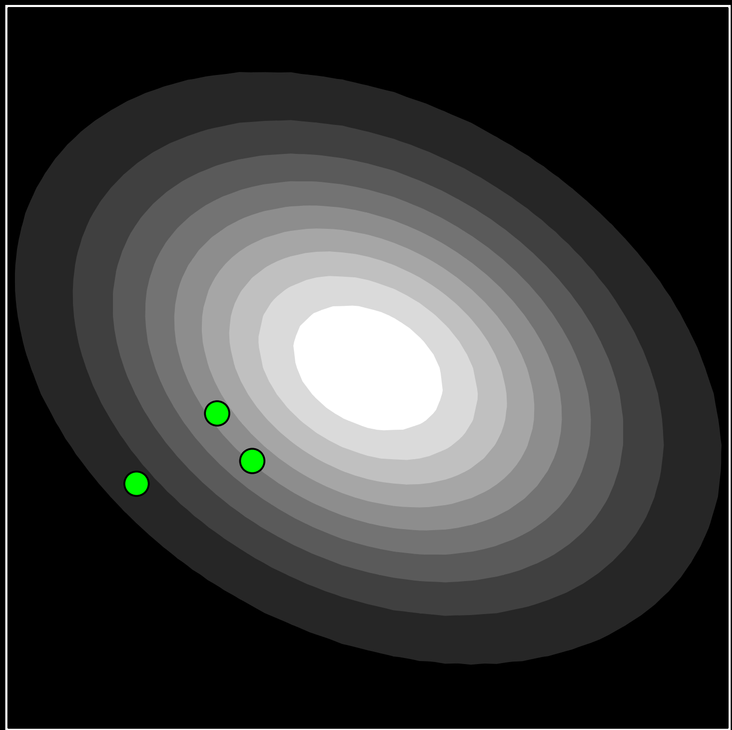


target

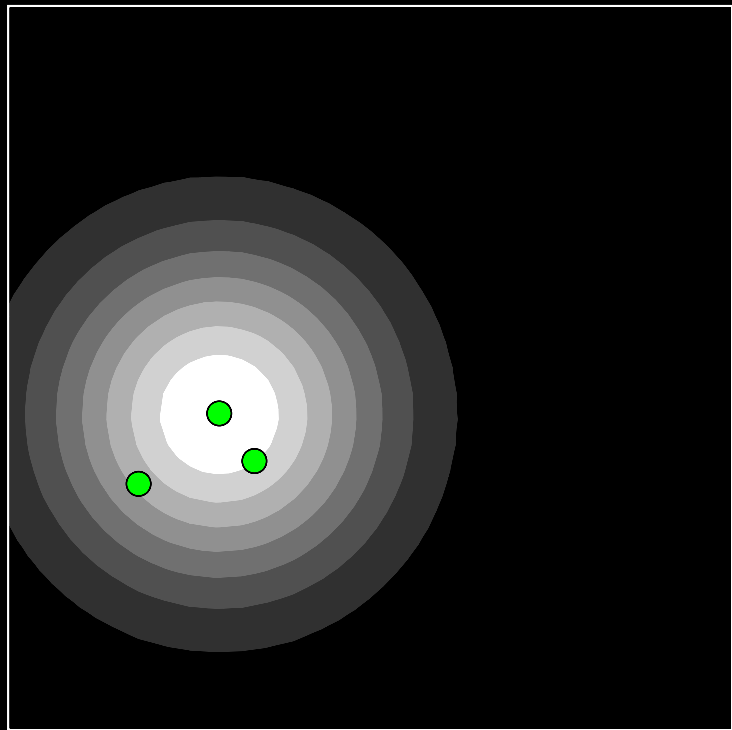


proposal

Metropolis-Hastings

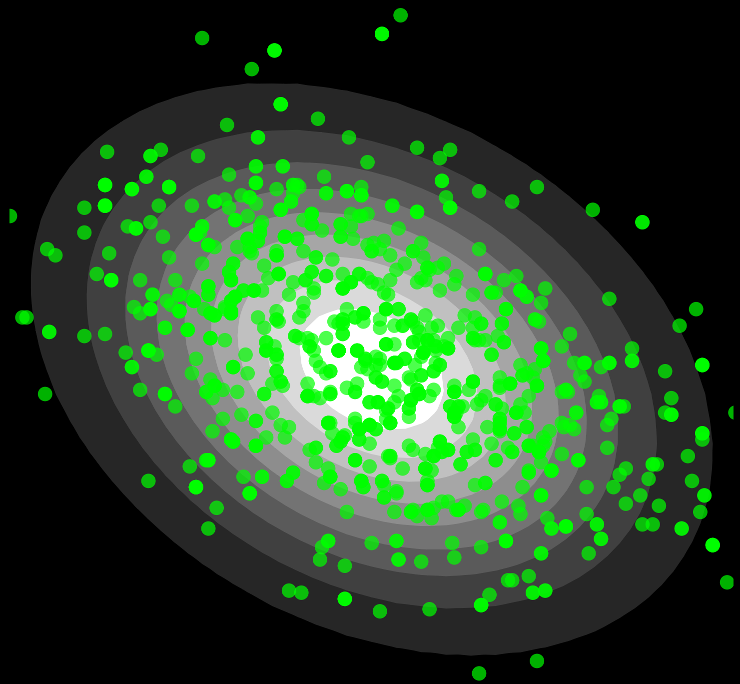


target



proposal

Metropolis-Hastings



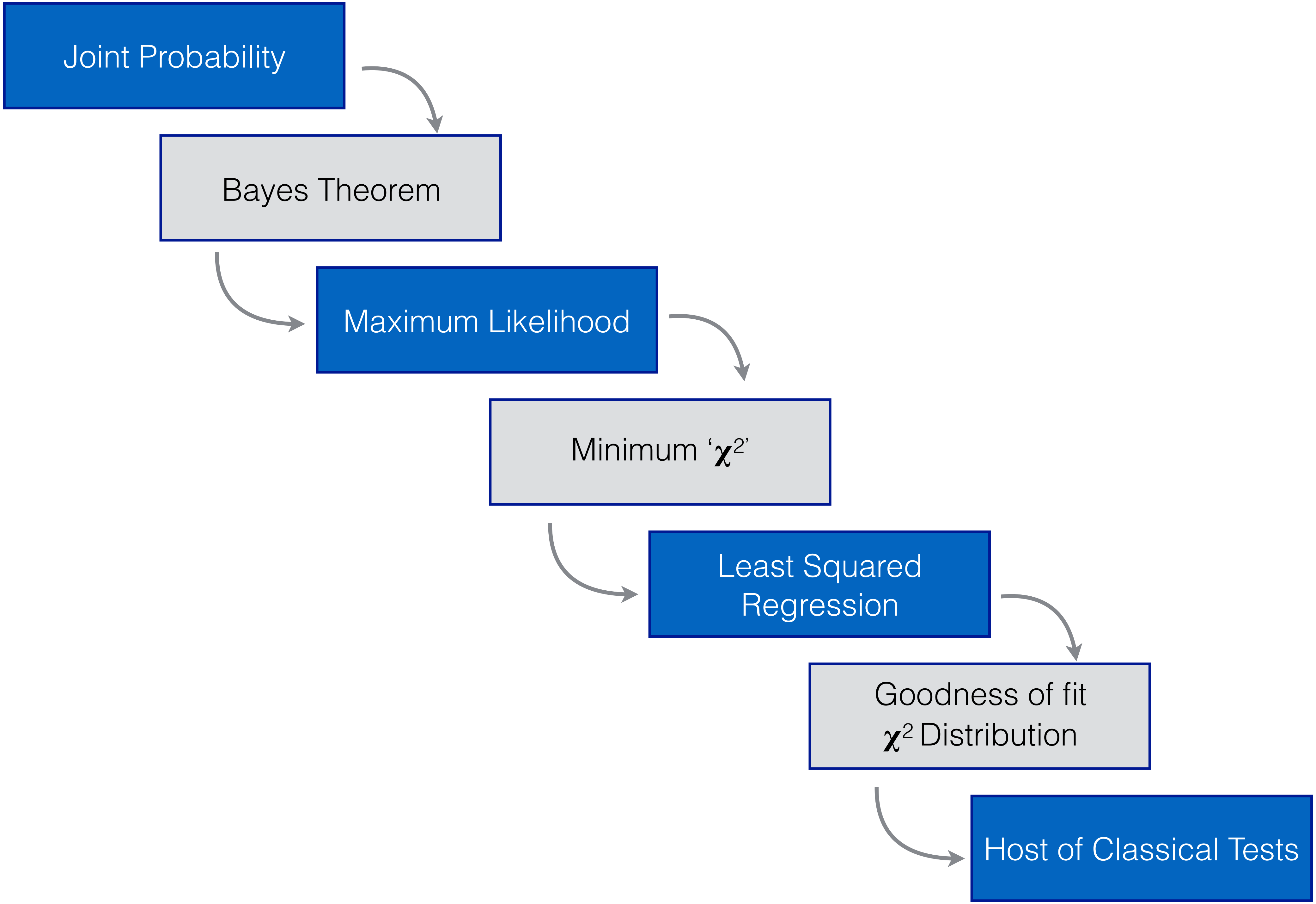
after 1,000 iterations

MCMC - Point sampling of pdf

$$\langle x \rangle = \int x \ p(x) dx$$

$$\langle h(x) \rangle = \int h(x) \ p(x) dx$$

Other measures can also be extracted from the p(x) e.g. x for max probability


Joint Probability
Bayes Theorem
Maximum Likelihood
Minimum 'χ^2'
Least Squared Regression
Goodness of fit χ^2 Distribution
Host of Classical Tests

Flat prior -> Likelihood analysis

$$p(M|D) \propto p(D|M)p(M)$$

Posterior

Likelihood

Prior

Maximum Likelihood

Posterior calculations with a flat prior and the maximum probably point recover a statistical method called the Maximum Likelihood Estimator (MLE) approach.

Though it was developed almost 100 years ago (R.A. Fisher 1920s), it is still a useful method for many applications.

Maximum Likelihood: IID process

$$L(\theta) = \Pi f(X_i;\theta) = p(X_i|D)$$

$$\ell(\theta) = \ln[L(\theta)] = \sum lnf(X_i;\theta)$$

The MLE of θ ($\hat{\theta}_{MLE}$) is found by maximising the $L(\theta)$ or $\ell(\theta)$.
$\frac{\partial \ell}{\partial \theta} = 0$.

Maximum Likelihood: Gaussian example

Mean of a sample of size n from Gauss $N(\mu, \sigma^2)$

Minimum 'χ2' & the χ2 Distribution

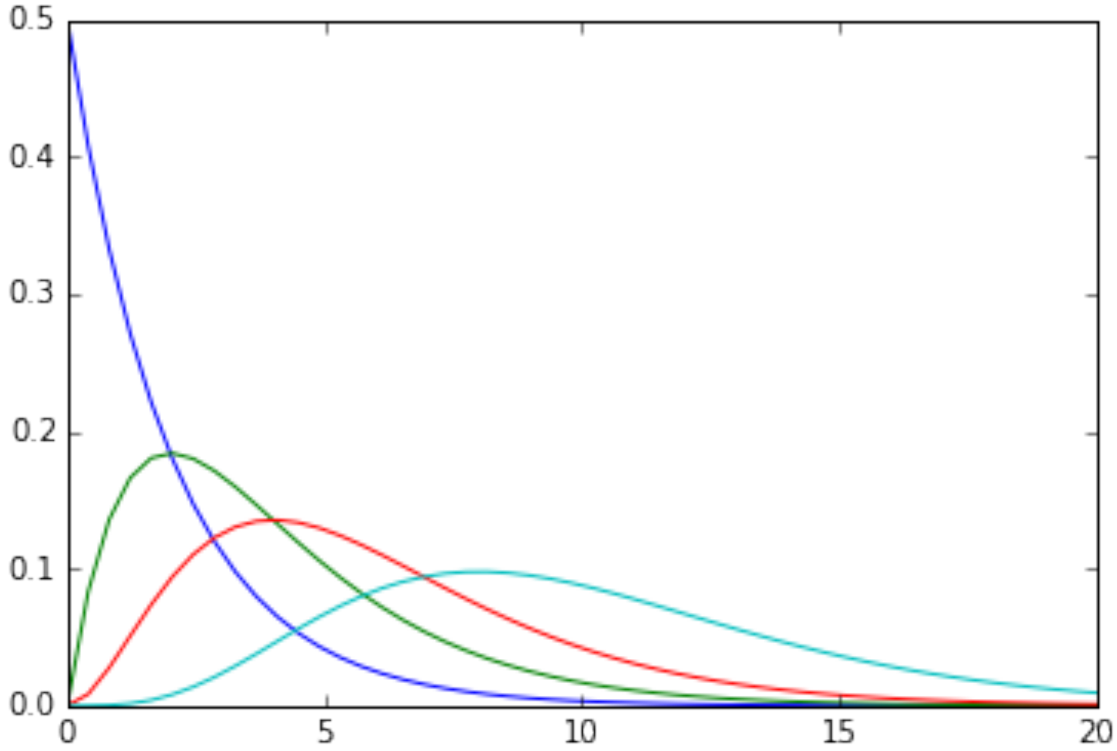
Draw numbers from a normal distribution:

$$\mathcal{N}(0,1)$$

$$\ell \propto -\sum \frac{(x_i-\mu)^2}{2\sigma^2}$$

Interested in variable that is the sum of squares:

$$Q_n = x_1^2 + x_2^2 + \ldots + x_n^2$$



Distribution of Qn follows a χ2 distribution.

$$\frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)}\, x^{\frac{k}{2}-1}e^{-\frac{x}{2}}$$

Least Squared Regression - MLE for linear models

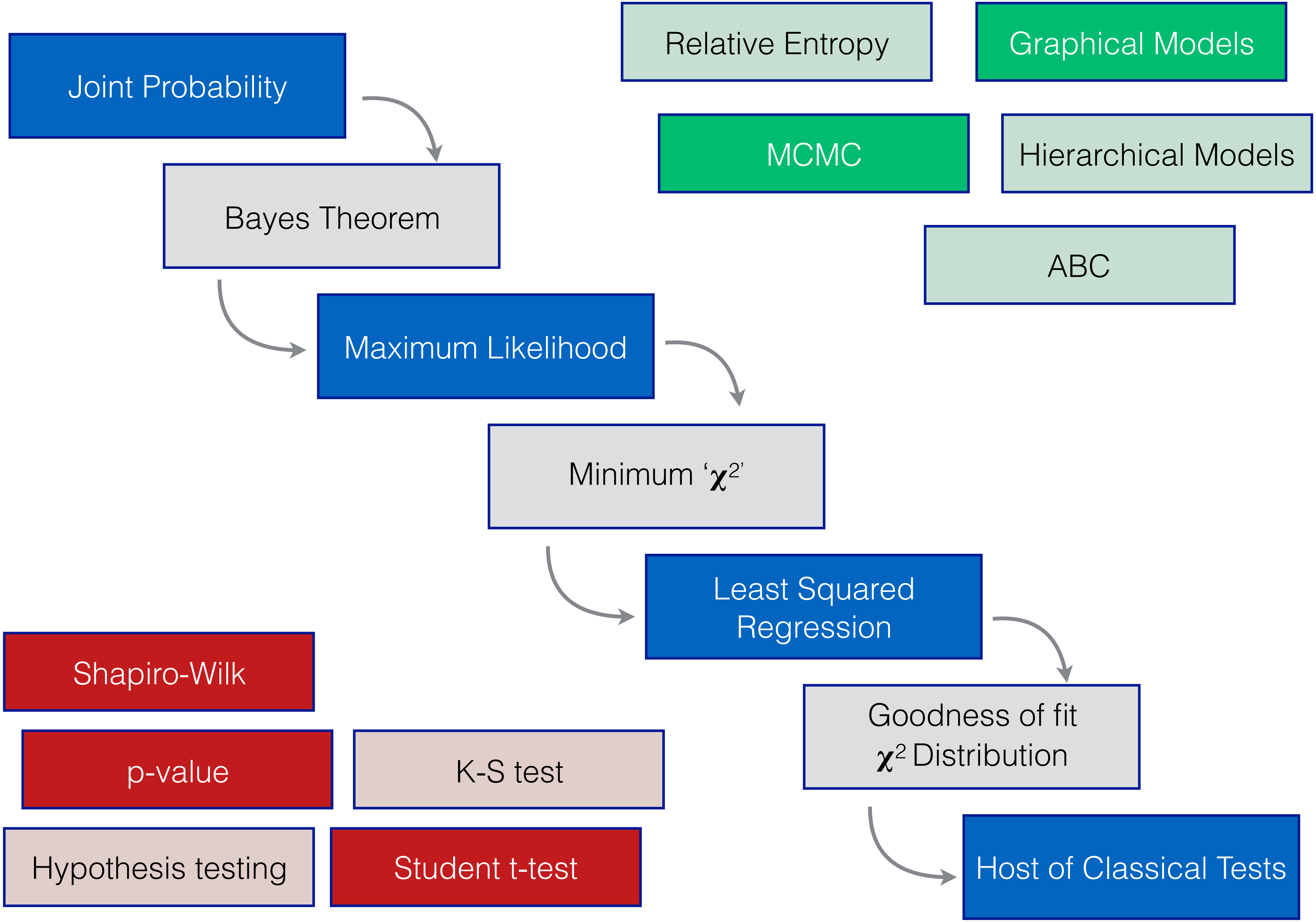
$$y = X\theta + \epsilon$$

$$\hat{y} = X\hat{\theta}$$

$S_g = (y - X\hat{\theta})^T C^{-1}(y - X\hat{\theta})$.
Where the known covariance, C, is
$C_{ij} = \langle (y_i - \langle y_i \rangle)(y_j - \langle y_j \rangle) \rangle$.

$$\frac{\partial S_g}{\partial \hat{\theta}} = 0$$

In this case the estimator is given by
$\hat{\theta}_g = (X^T C^{-1} X)^{-1} X^T C^{-1} y$,
with a covariance matrix of $(X^T C^{-1} X)^{-1}$


Relative Entropy
Graphical Models
Joint Probability
MCMC
Hierarchical Models
Bayes Theorem
ABC
Maximum Likelihood
Minimum 'χ^2'
Least Squared Regression
Shapiro-Wilk
Goodness of fit χ^2 Distribution
p-value
K-S test
Hypothesis testing
Student t-test
Host of Classical Tests

ETH zürich