

An Introduction to Bayesian Data Analysis

Lecture 2

D. S. Sivia

St. John's College
Oxford, England

April 29, 2012

Outline	2
2-Parameter estimation	3
The Poisson distribution	4
Amplitude and background	5
Amplitude and background (1)	6
Amplitude and background (2)	7
Amplitude and background (3)	8
Amplitude and background (4)	9
Nuisance parameters	10
Marginalisation	11
Amplitude and background (5)	12
Amplitude and background (6)	13
Amplitude and background (7)	14
Amplitude and background (8)	15
Summarising the inference	16
Covariance and correlation	17
Marginal versus conditional error-bars	18
Asymmetric posterior pdfs.	19
Multimodal posterior pdfs	20
Gaussian noise revisited (1)	21
Gaussian noise revisited (2)	22
“Systematic errors”	23
“Systematic errors”	24
Algorithms: a numerical interlude	25
Brute force and ignorance	26
The joys of linearity	27
Iterative linearisation	28
Hard problems	29
The phaseless Fourier problem	30
The phaseless Fourier problem	31

Outline

- The basics
- Parameter estimation I
- Parameter estimation II
- Model selection

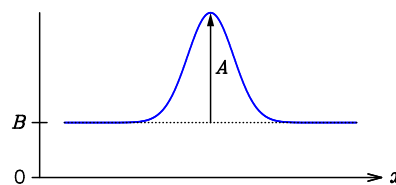
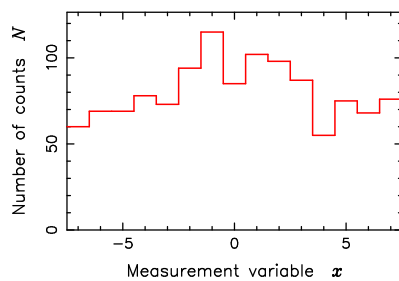
- Assigning probabilities
- Non-parametric estimation
- Experimental design
- Least-squares extensions*
- Nested sampling
- Quantification

D.S. Sivia (1996), Data analysis: a Bayesian tutorial, O.U.P.; 2nd edition with J. Skilling (2006).

IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

2 / 31

2-Parameter estimation



■ **Model:** $D_k = n_o \left[A e^{-(x_k - x_o)^2 / 2w^2} + B \right]$

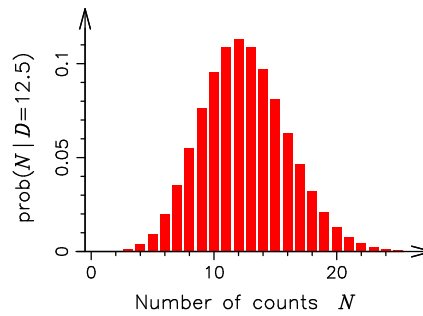
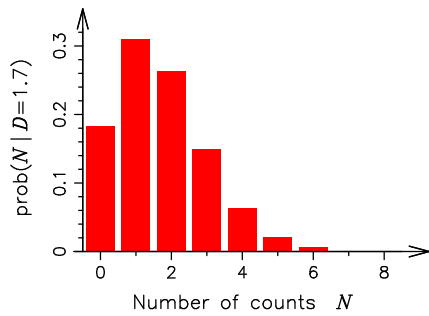
■ **Poisson data:** $\text{prob}(N_k | A, B, I) = \frac{D_k^{N_k} e^{-D_k}}{N_k!}$

IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

3 / 31

The Poisson distribution

$$\text{prob}(N|D) = \frac{D^N e^{-D}}{N!}$$



$$\langle N \rangle = \sum_{N=0}^{\infty} N \text{prob}(N|D) = D \quad (\text{mean})$$

$$\langle (N-D)^2 \rangle = \sum_{N=0}^{\infty} (N-D)^2 \text{prob}(N|D) = D \quad (\text{variance})$$

Amplitude and background

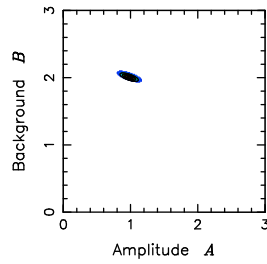
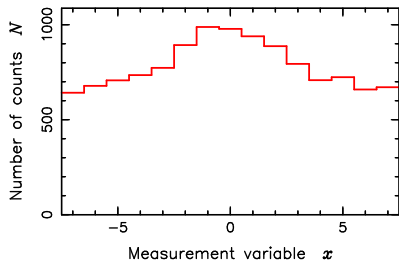
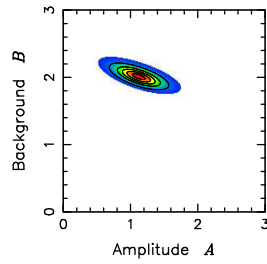
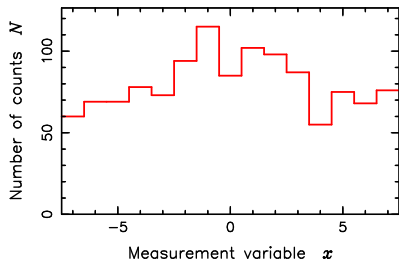
■ **Likelihood:** $\text{prob}(\{N_k\}|A, B, I) = \prod_{k=1}^M \frac{D_k^{N_k} e^{-D_k}}{N_k!}$ (Independence)

■ **Prior:** $\text{prob}(A, B|I) = \begin{cases} \text{constant} & \text{for } A \geq 0 \text{ and } B \geq 0 \\ 0 & \text{otherwise} \end{cases}$

■ **Bayes:** $\text{prob}(A, B|\{N_k\}, I) \propto \text{prob}(\{N_k\}|A, B, I) \times \text{prob}(A, B|I)$

$$\Rightarrow L = \ln[\text{prob}(A, B|\{N_k\}, I)] = \text{const} + \sum_{k=1}^M [N_k \ln(D_k) - D_k]$$

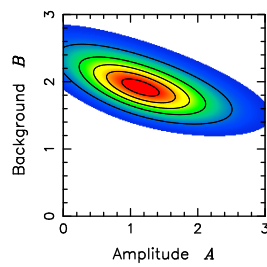
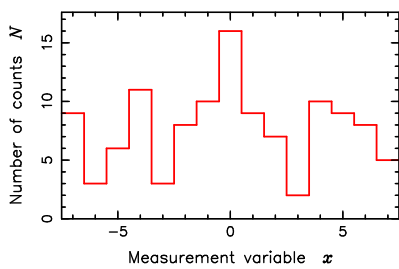
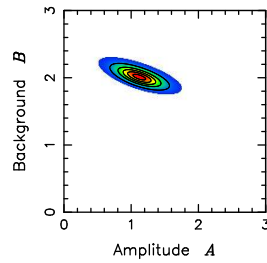
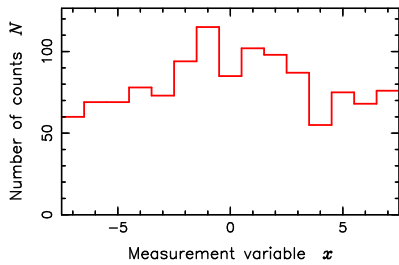
Amplitude and background (1)



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

6 / 31

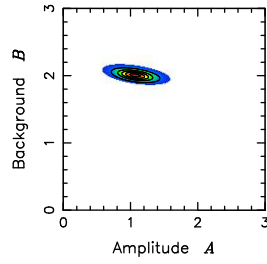
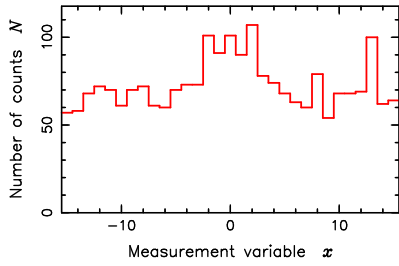
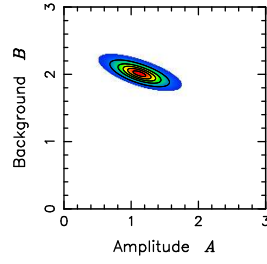
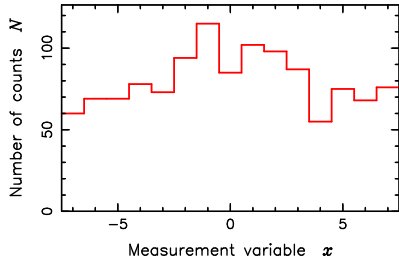
Amplitude and background (2)



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

7 / 31

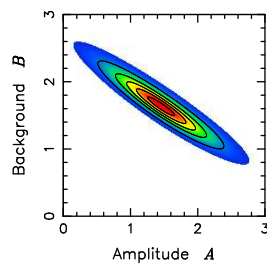
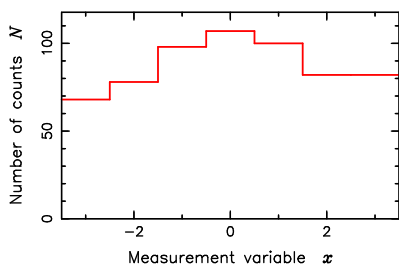
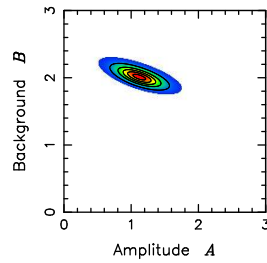
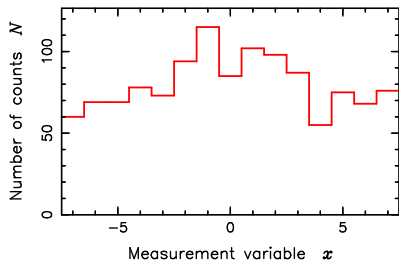
Amplitude and background (3)



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

8 / 31

Amplitude and background (4)



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

9 / 31

Nuisance parameters

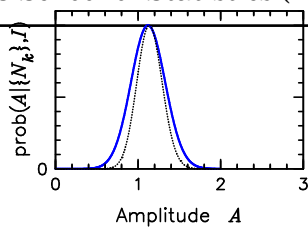
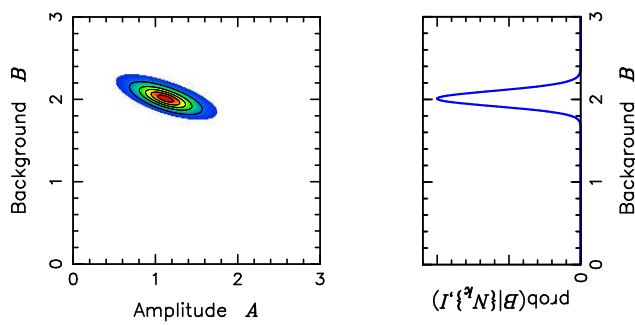
- Amplitude:

$$\text{prob}(A|\{N_k\}, I) = \int_0^\infty \text{prob}(A, B|\{N_k\}, I) dB$$

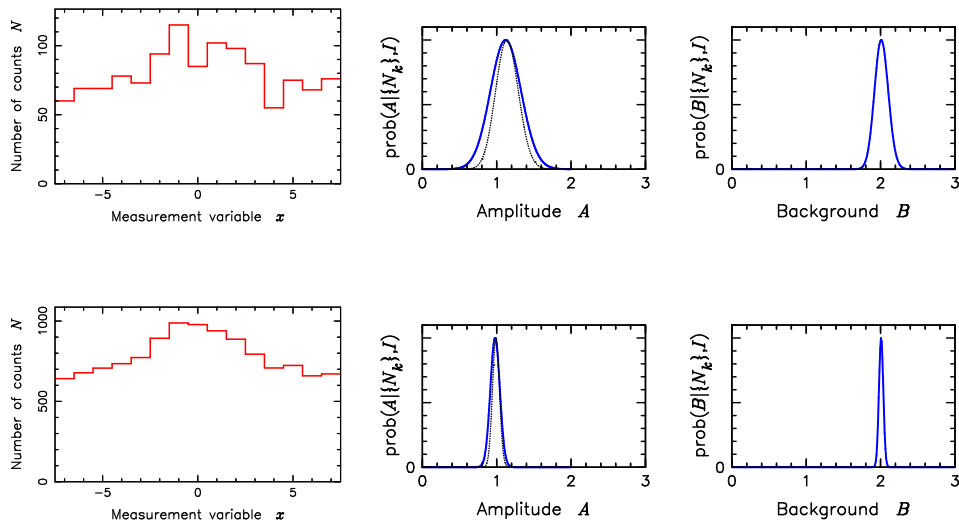
- Background:

$$\text{prob}(B|\{N_k\}, I) = \int_0^\infty \text{prob}(A, B|\{N_k\}, I) dA$$

Marginalisation



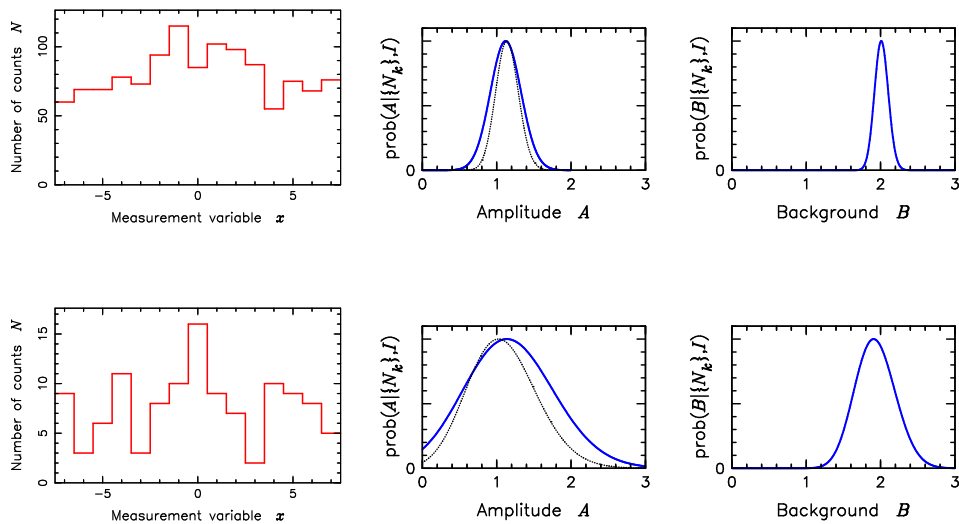
Amplitude and background (5)



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

12 / 31

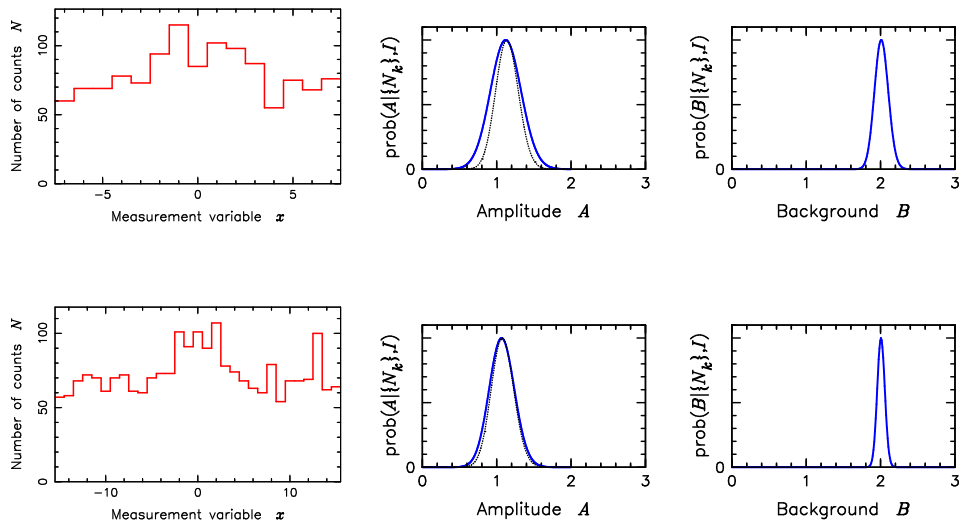
Amplitude and background (6)



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

13 / 31

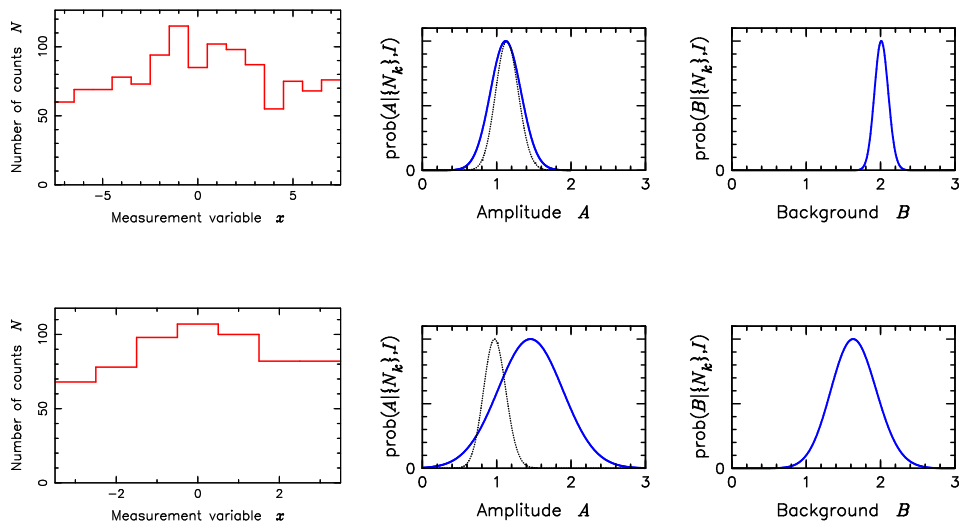
Amplitude and background (7)



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

14 / 31

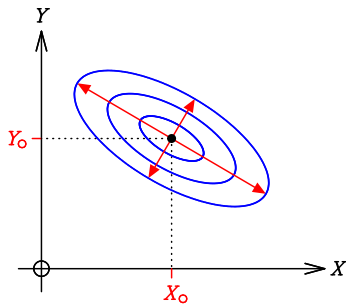
Amplitude and background (8)



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

15 / 31

Summarising the inference



Letting $L = \log_e[\text{prob}(\mathbf{X}|\{\text{data}\}, I)]$,

where $\nabla L(\mathbf{X}_0) = 0$,

Taylor: $L(\mathbf{X}) = L(\mathbf{X}_0) + \frac{1}{2}(\mathbf{X} - \mathbf{X}_0)^T \nabla \nabla L(\mathbf{X}_0) (\mathbf{X} - \mathbf{X}_0) + \dots$

$$\Rightarrow [\sigma^2]_{ij} = \langle (X_i - X_{0i})(X_j - X_{0j}) \rangle = -[\{\nabla \nabla L(\mathbf{X}_0)\}^{-1}]_{ij}$$

Covariance matrix = $-(\text{Hessian matrix})^{-1}$

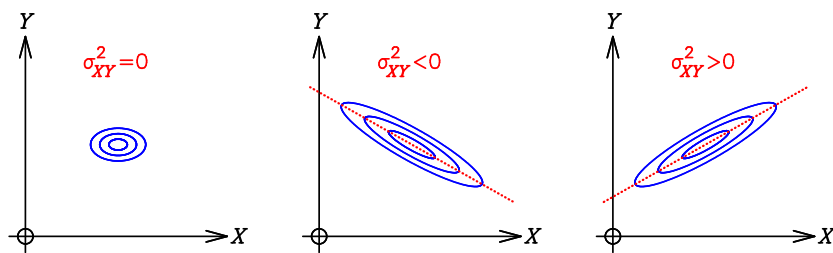
IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

16 / 31

Covariance and correlation

$$\begin{pmatrix} \sigma_x^2 & \sigma_{xY}^2 \\ \sigma_{xY}^2 & \sigma_Y^2 \end{pmatrix} = - \begin{pmatrix} \partial^2 L / \partial X^2 & \partial^2 L / \partial X \partial Y \\ \partial^2 L / \partial X \partial Y & \partial^2 L / \partial Y^2 \end{pmatrix}^{-1}$$

where $\sigma_x^2 = \langle (X - X_0)^2 \rangle$, $\sigma_Y^2 = \langle (Y - Y_0)^2 \rangle$, $\sigma_{xY}^2 = \langle (X - X_0)(Y - Y_0) \rangle$

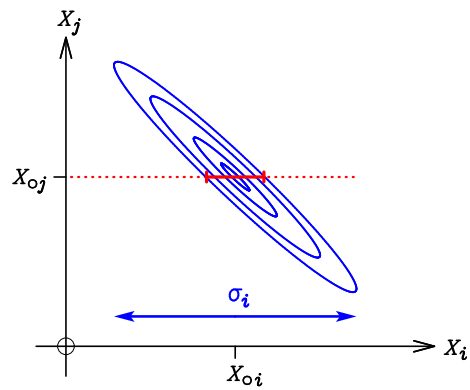


$$-1 \leq \frac{\sigma_{xY}^2}{\sqrt{\sigma_x^2 \sigma_Y^2}} \leq 1$$

IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

17 / 31

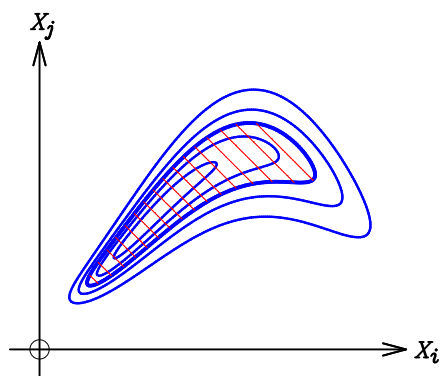
Marginal versus conditional error-bars



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

18 / 31

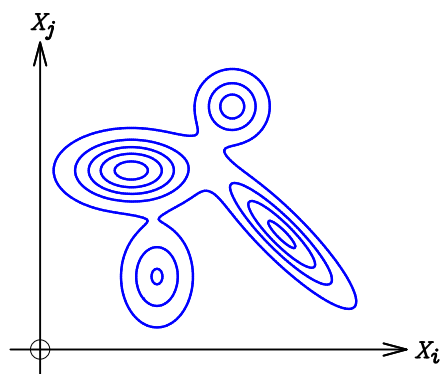
Asymmetric posterior pdfs



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

19 / 31

Multimodal posterior pdfs



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

20 / 31

Gaussian noise revisited (1)

Given N independent measurements of a quantity μ , $\{d_k\}$, all subject to *unknown* Gaussian noise σ , what can we say about the value of μ ?

■ **Marginalise:** $\text{prob}(\mu | \{d_k\}, I) = \int_0^\infty \text{prob}(\mu, \sigma | \{d_k\}, I) d\sigma$

◆ **Bayes:** $\text{prob}(\mu, \sigma | \{d_k\}, I) \propto \text{prob}(\{d_k\} | \mu, \sigma, I) \times \text{prob}(\mu, \sigma | I)$

◆ **Likelihood:**

$$\text{prob}(\{d_k\} | \mu, \sigma, I) = (\sigma \sqrt{2\pi})^{-N} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^N (d_k - \mu)^2\right]$$

◆ **Prior:** $\text{prob}(\mu, \sigma | I) = \text{constant}$ (for $\sigma > 0$)

IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

21 / 31

Gaussian noise revisited (2)

$$\Rightarrow L = \log_e [\text{prob}(\mu | \{d_k\}, I)] = \text{const} - \frac{(N-1)}{2} \log_e \left[\sum_{k=1}^N (d_k - \mu)^2 \right]$$

$$\left. \frac{dL}{d\mu} \right|_{\mu_0} = \frac{(N-1) \sum (d_k - \mu_0)}{\sum (d_k - \mu_0)^2} = 0 \quad \Rightarrow \quad \mu_0 = \frac{1}{N} \sum_{k=1}^N d_k$$

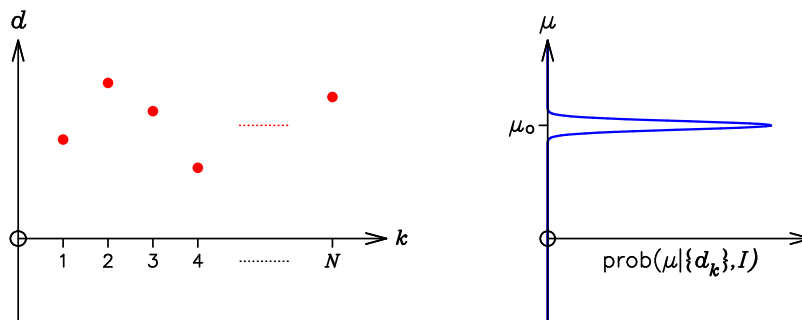
$$\left. \frac{d^2L}{d\mu^2} \right|_{\mu_0} = - \frac{N(N-1)}{\sum (d_k - \mu_0)^2} \quad \Rightarrow \quad \mu = \mu_0 \pm \frac{S}{\sqrt{N}}$$

$$\text{where } S^2 = \frac{1}{N-1} \sum_{k=1}^N (d_k - \mu_0)^2$$

IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

22 / 31

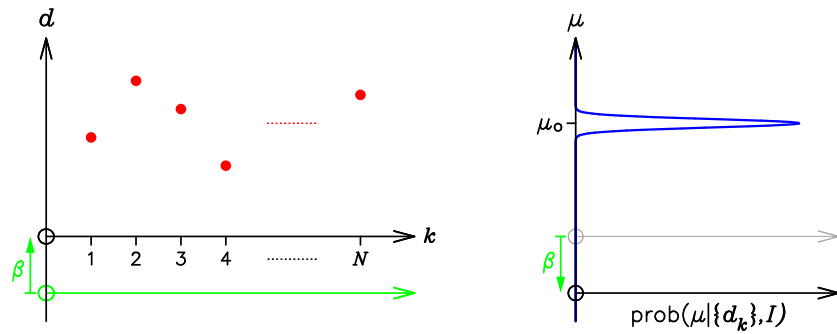
“Systematic errors”



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

23 / 31

“Systematic errors”



$$\blacksquare \quad \text{prob}(\mu|\{d_k\}, I) = \int \text{prob}(\mu|\{d_k\}, \beta, I) \underbrace{\text{prob}(\beta|\{d_k\}, I)}_{\text{prob}(\beta|I)} d\beta$$

$$\blacklozenge \quad \text{If } \beta = \beta_0 \pm \epsilon, \quad \mu = \beta_0 + \frac{1}{N} \sum_{k=1}^N d_k \pm \sqrt{\frac{S^2}{N} + \epsilon^2}$$

Algorithms: a numerical interlude

Numerical recipes

W. H. Press et. al. (1986), Cambridge University Press

Practical optimization

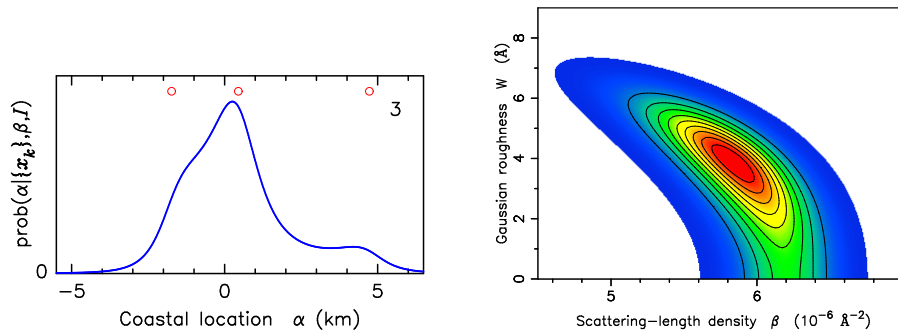
P. E. Gill, W. Murray and M.H. Wright (1981), Academic Press

Numerical method that usually work

F. S. Acton (1970), Harper and Row

Brute force and ignorance

Ideal for *visualisation*, and *robust*, for one or two-parameter problems; rapidly becomes impractical for more than a *few* parameters.



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

26 / 31

The joys of linearity

■ **Want:** $\nabla L(\mathbf{X}_0) = 0$

■ **Linear:** $\nabla L = \mathbf{H}\mathbf{X} + \mathbf{C}$

$$\therefore \mathbf{X}_0 = -\mathbf{H}^{-1}\mathbf{C}$$

$$\nabla \nabla L = \mathbf{H} \quad \Rightarrow \quad \sigma^2 = -\mathbf{H}^{-1} \quad (\text{Covariance})$$

IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

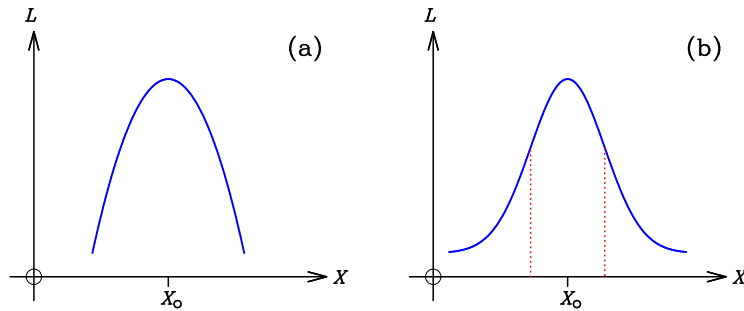
27 / 31

Iterative linearisation

■ Taylor: $\nabla L = \nabla L(\mathbf{X}_1) + \nabla \nabla L(\mathbf{X}_1)(\mathbf{X} - \mathbf{X}_1) + \dots$

$$\therefore \nabla L(\mathbf{X}_0) = 0 \Rightarrow \mathbf{X}_0 \approx \mathbf{X}_1 - [\nabla \nabla L(\mathbf{X}_1)]^{-1} \nabla L(\mathbf{X}_1)$$

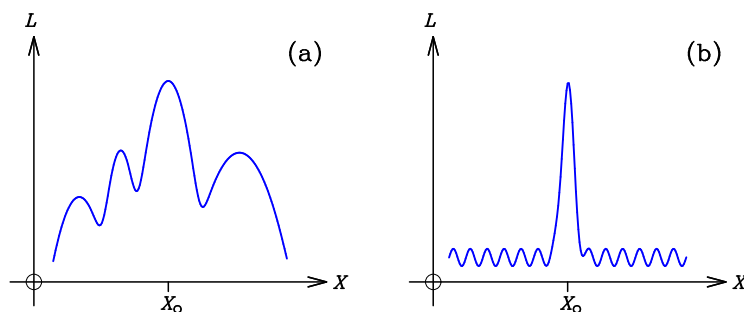
■ Newton: $\mathbf{X}_{N+1} = \mathbf{X}_N - [\nabla \nabla L(\mathbf{X}_N) + c\mathbf{I}]^{-1} \nabla L(\mathbf{X}_N)$



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

28 / 31

Hard problems



■ Markov Chain Monte Carlo (MCMC): $\mathbf{X}_{N+1} = \mathbf{X}_N + \Delta \mathbf{X}$

- ◆ inspired trial-and-error
- ◆ simulated annealing, genetic algorithms, ...

IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

29 / 31

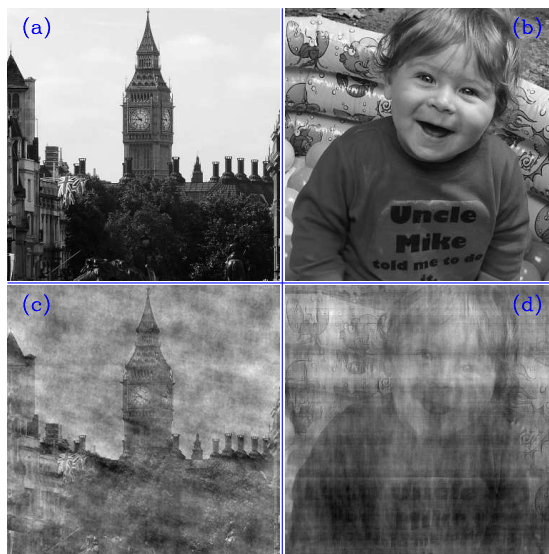
The phaseless Fourier problem



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

30 / 31

The phaseless Fourier problem



IN2P3/CNRS School of Statistics (Autrans 2012): Bayesian Lecture 2

31 / 31