An Introduction to
Bayesian Data Analysis
Lecture 2

D. S. Sivia
St. John's College
Oxford, England

April 29, 2012

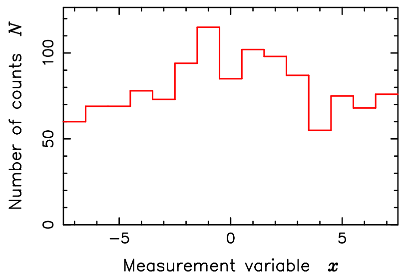
Outline

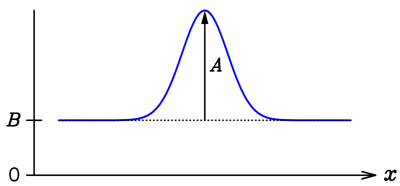
- The basics
- Parameter estimation I
- Parameter estimation II
- Model selection

- 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).

2-Parameter estimation



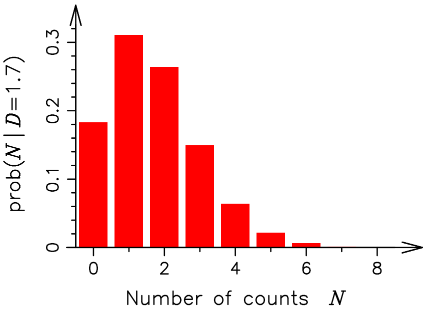


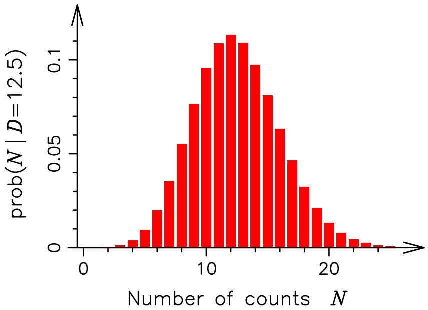
- **Model:** $D_k = n_o \left[A\, e^{-(x_k - x_o)^2/2w^2} + B \right]$

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

The Poisson distribution

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





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

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

Amplitude and background

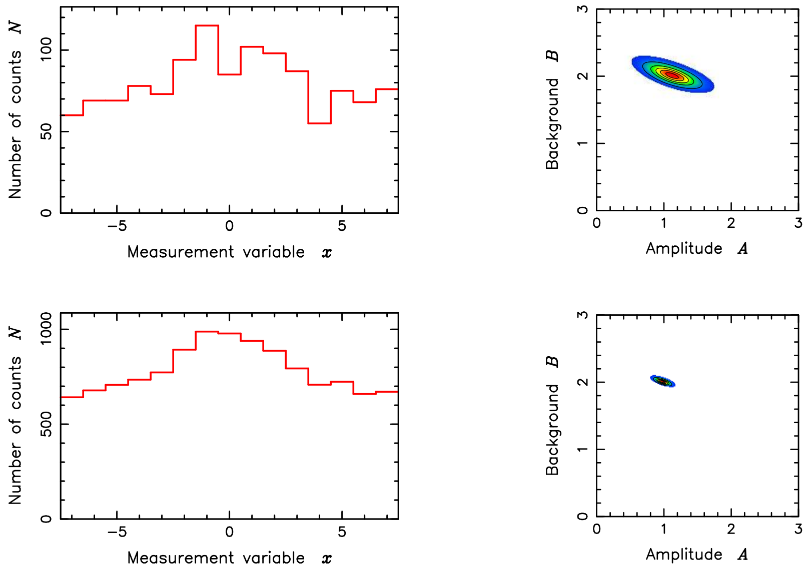
- Likelihood: $\mathrm{prob}(\{N_k\}|A,B,I) = \prod_{k=1}^{M} \frac{D_k^{N_k}\,\mathrm{e}^{-D_k}}{N_k!}$ **(Independence)**

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

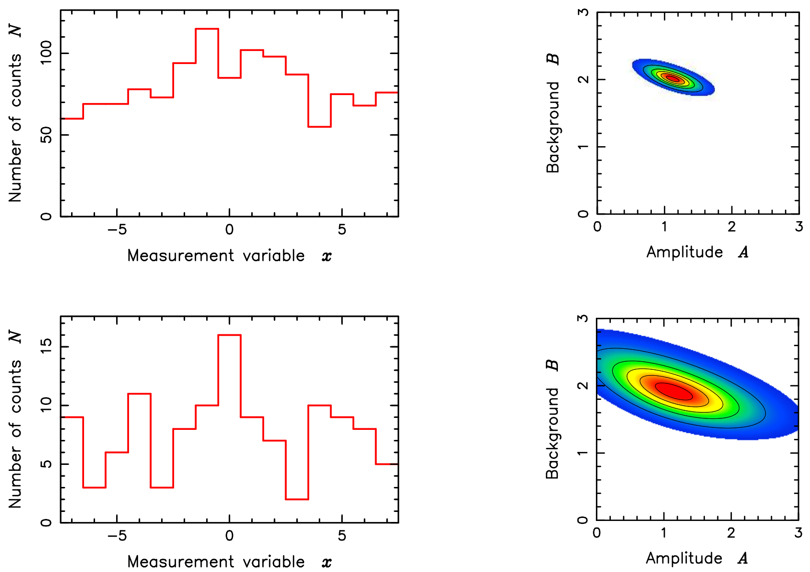
- Bayes: $\mathrm{prob}(A,B|\{N_k\},I) \propto \mathrm{prob}(\{N_k\}|A,B,I) \times \mathrm{prob}(A,B|I)$

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

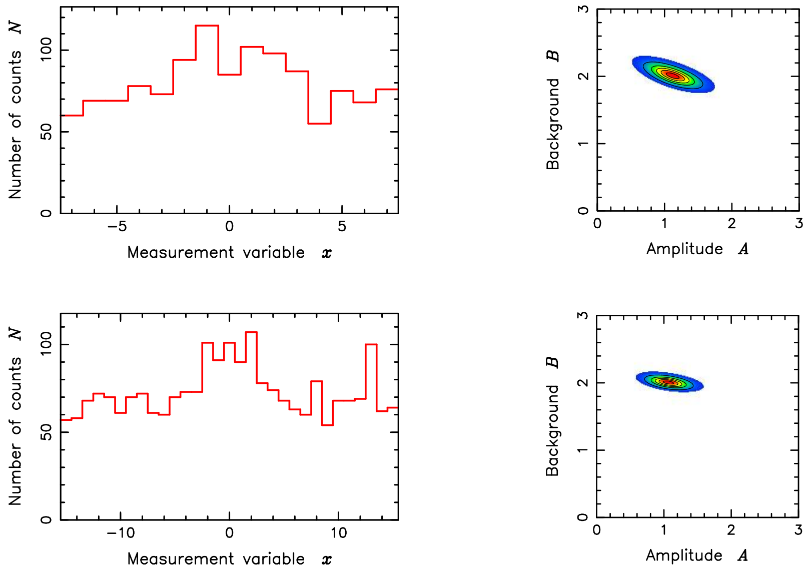
Amplitude and background (1)



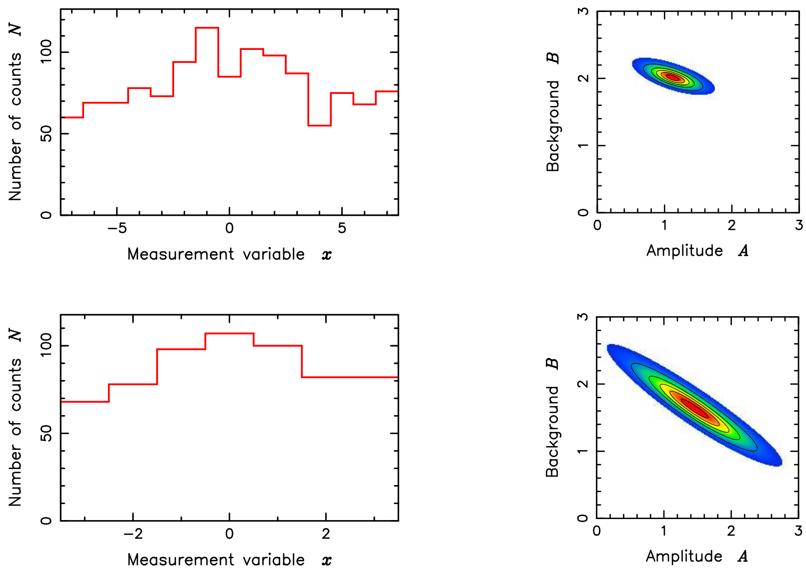
Amplitude and background (2)



Amplitude and background (3)



Amplitude and background (4)



Nuisance parameters

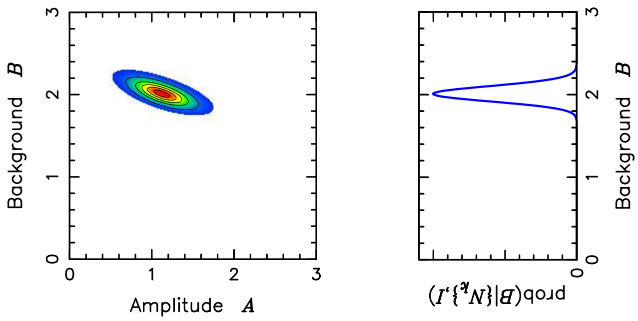
- Amplitude:

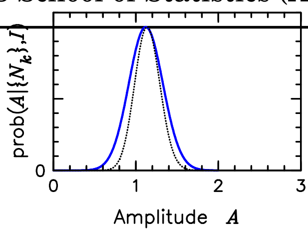
$$\mathrm{prob}(A|\{N_k\},I) = \int_0^\infty \mathrm{prob}(A,B|\{N_k\},I)\,\mathrm{d}B$$

- Background:

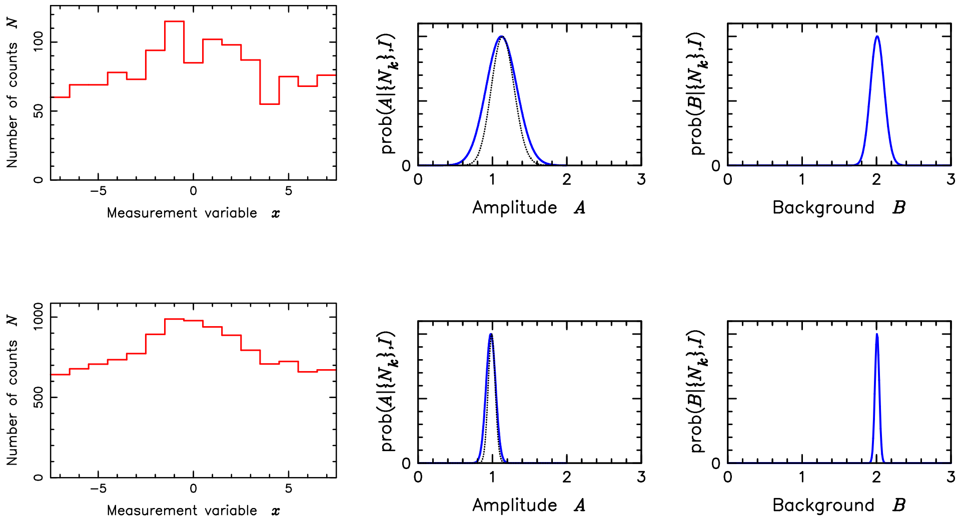
$$\mathrm{prob}(B|\{N_k\},I) = \int_0^\infty \mathrm{prob}(A,B|\{N_k\},I)\,\mathrm{d}A$$

Marginalisation

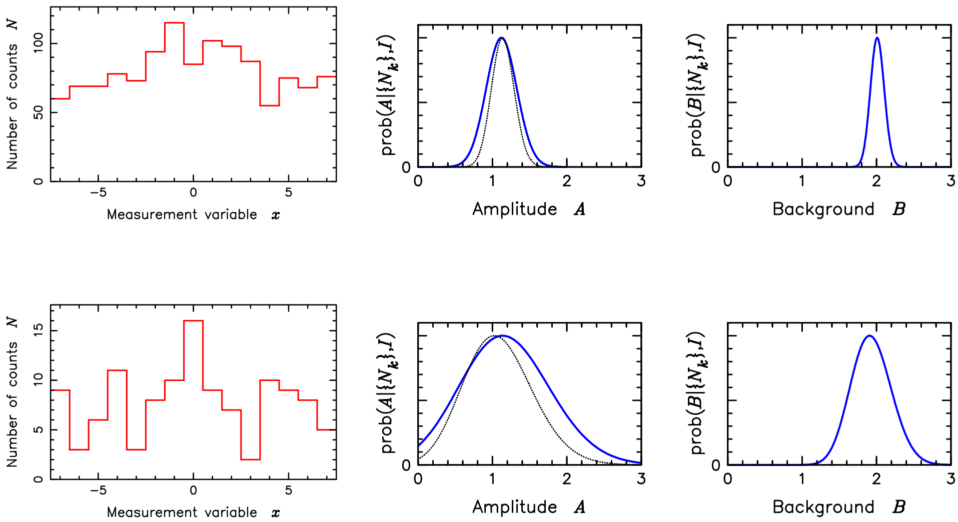




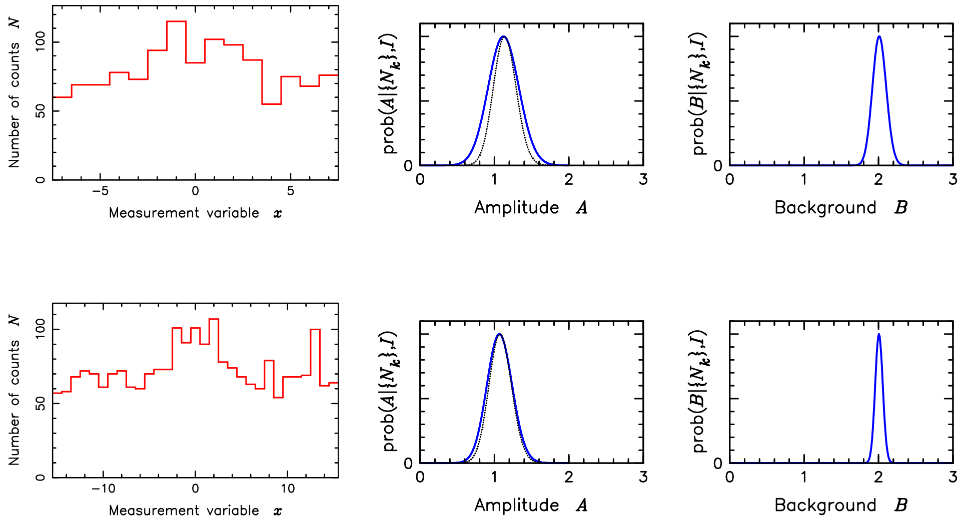
Amplitude and background (5)



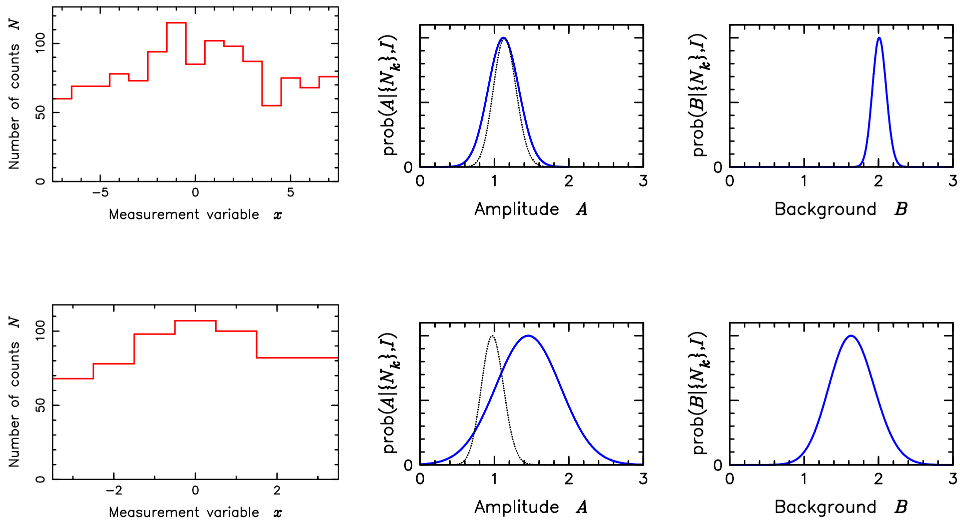
Amplitude and background (6)



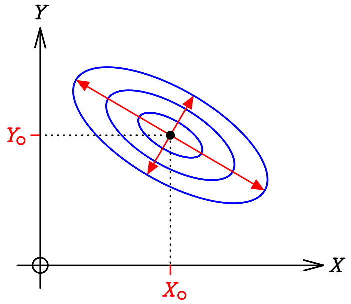
Amplitude and background (7)



Amplitude and background (8)



Summarising the inference



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

where $\boldsymbol{\nabla} L(\mathbf{X}_\mathrm{O}) = 0$,

Taylor: $L(\mathbf{X}) = L(\mathbf{X}_\mathrm{O}) + \frac{1}{2}\,(\mathbf{X}-\mathbf{X}_\mathrm{O})^\mathrm{T}\,\boldsymbol{\nabla\nabla} L(\mathbf{X}_\mathrm{O})\,(\mathbf{X}-\mathbf{X}_\mathrm{O}) + \cdots$

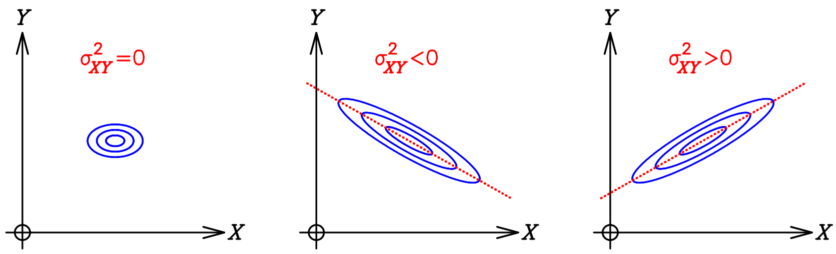
$$\Rightarrow \left[\boldsymbol{\sigma}^2\right]_{ij} = \Big\langle (X_i - X_{\mathrm{O}i})(X_j - X_{\mathrm{O}j}) \Big\rangle = -\left[\left\{\boldsymbol{\nabla\nabla} L(\mathbf{X}_\mathrm{O})\right\}^{-1}\right]_{ij}$$

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

Covariance and correlation

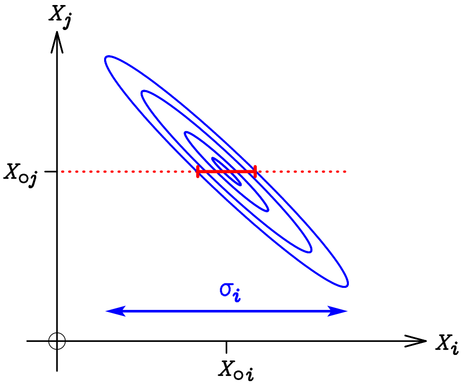
$$\begin{pmatrix} \sigma_X^2 & \sigma_{XY}^2 \\ \sigma_{XY}^2 & \sigma_Y^2 \end{pmatrix} = -\begin{pmatrix} \partial^2 L/\partial X^2 & \partial^2 L/\partial X\,\partial Y \\ \partial^2 L/\partial X\,\partial Y & \partial^2 L/\partial Y^2 \end{pmatrix}^{-1}$$

where $\sigma_X^2 = \Big\langle (X - X_\mathrm{O})^2 \Big\rangle$, $\sigma_Y^2 = \Big\langle (Y - Y_\mathrm{O})^2 \Big\rangle$, $\sigma_{XY}^2 = \Big\langle (X - X_\mathrm{O})(Y - Y_\mathrm{O}) \Big\rangle$

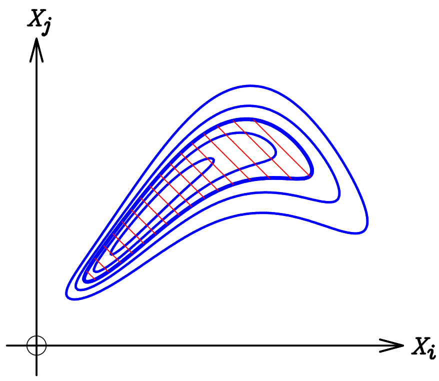


$$-1 \leqslant \frac{\sigma_{XY}^2}{\sqrt{\sigma_X^2\,\sigma_Y^2}} \leqslant 1$$

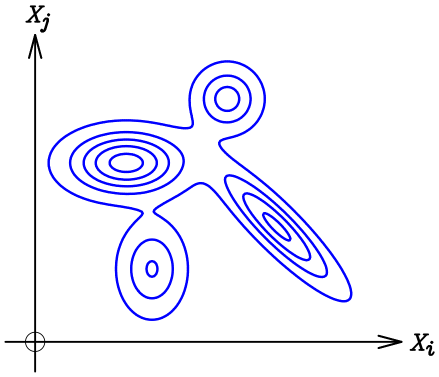
Marginal versus conditional error-bars



Asymmetric posterior pdfs



Multimodal posterior pdfs



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: $\text{prob}(\mu|\{d_k\},I) = \int_0^\infty \text{prob}(\mu,\sigma|\{d_k\},I)\,\mathrm{d}\sigma$
  - Bayes: $\text{prob}(\mu,\sigma|\{d_k\},I) \propto \text{prob}(\{d_k\}|\mu,\sigma,I) \times \text{prob}(\mu,\sigma|I)$
  - Likelihood:
    $$\text{prob}(\{d_k\}|\mu,\sigma,I) = \left(\sigma\sqrt{2\pi}\right)^{-N} \exp\left[-\frac{1}{2\sigma^2}\sum_{k=1}^{N}(d_k-\mu)^2\right]$$
  - Prior: $\text{prob}(\mu,\sigma|I) = \text{constant} \quad (\text{for } \sigma>0)$

Gaussian noise revisited (2)

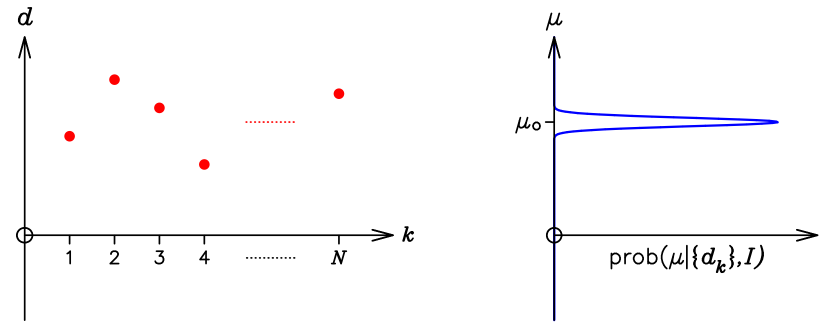
$$\Rightarrow \; L = \log_e\!\left[\,\text{prob}(\mu|\{d_k\},I)\,\right] = \text{const} - \frac{(N-1)}{2}\log_e\!\left[\sum_{k=1}^{N}(d_k-\mu)^2\right]$$

$$\left.\frac{\mathrm{d}L}{\mathrm{d}\mu}\right|_{\mu_o} = \frac{(N-1)\sum(d_k-\mu_o)}{\sum(d_k-\mu_o)^2} = 0 \qquad \Rightarrow \; \mu_o = \frac{1}{N}\sum_{k=1}^{N} d_k$$

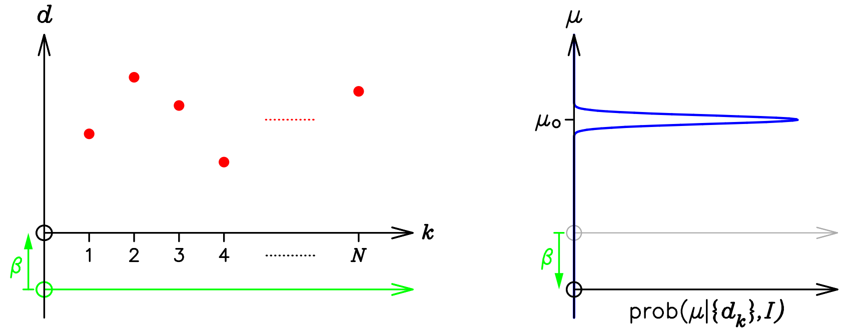
$$\left.\frac{\mathrm{d}^2L}{\mathrm{d}\mu^2}\right|_{\mu_o} = -\frac{N(N-1)}{\sum(d_k-\mu_o)^2} \qquad \Rightarrow \; \mu = \mu_o \pm \frac{S}{\sqrt{N}}$$

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

“Systematic errors”



“Systematic errors”



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

  - If $\beta = \beta_\mathrm{o} \pm \epsilon$, $\quad \mu = \beta_\mathrm{o} + \frac{1}{N}\sum_{k=1}^{N} d_k \pm \sqrt{\frac{S^2}{N} + \epsilon^2}$

Algorithms: a numerical interlude

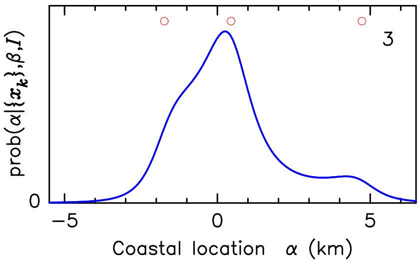
Numerical recipes
W. H. Press et. al. (1986), Cambridge University Press

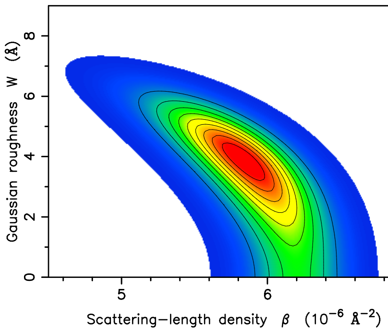
Practical optimization
P. E. Gill, W. Murray and M.H. Wright (1981), Academic Press

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.





The joys of linearity

- Want: $\nabla L(\mathbf{X}_{\mathrm{O}}) = 0$

- Linear: $\nabla L = \mathbf{H}\,\mathbf{X} + \boldsymbol{C}$

$$\therefore \quad \mathbf{X}_{\mathrm{O}} = -\mathbf{H}^{-1}\boldsymbol{C}$$

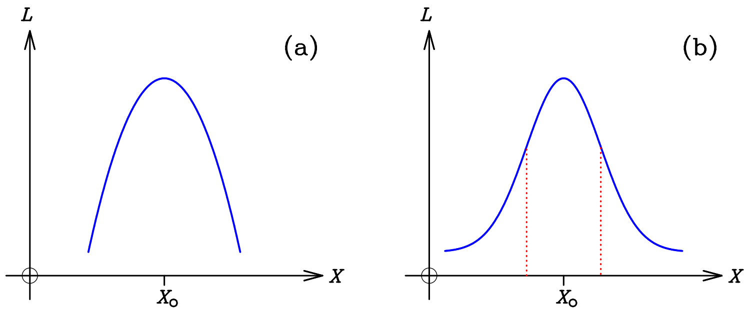
$$\nabla\nabla L = \mathbf{H} \quad \Rightarrow \quad \boldsymbol{\sigma}^2 = -\mathbf{H}^{-1} \quad \text{(Covariance)}$$

Iterative linearisation

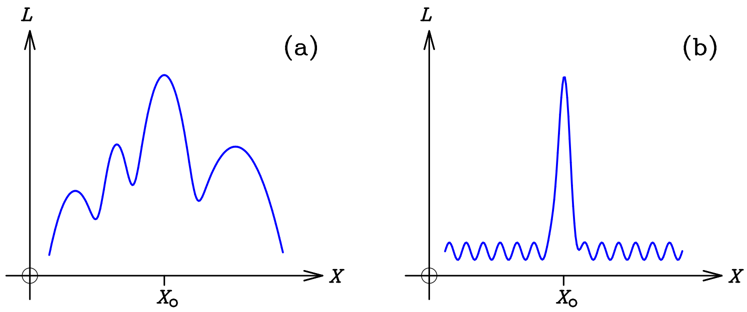
- Taylor: $\nabla L = \nabla L(\mathbf{X}_1) + \nabla\nabla L(\mathbf{X}_1)(\mathbf{X}-\mathbf{X}_1) + \cdots$

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

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



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, . . .

The phaseless Fourier problem

The phaseless Fourier problem