6. An Advanced Bayesian Toolbox – Part One

Course Programme

Lectures 6 to 10

6. **An Advanced Toolbox for Bayesian Inference**

7. **An Advanced Toolbox for Bayesian Inference**

8. **Bayesian Model Selection**

9. **Monte Carlo Simulation Methods**

10. **Fourier Methods**

Parameter estimation:

$$\underbrace{p(\text{model} \mid \text{data}, I)}_{\text{Posterior}} \propto \underbrace{p(\text{data} \mid \text{model}, I)}_{\text{Likelihood}} \times \underbrace{p(\text{model} \mid I)}_{\text{Prior}}$$

Parameter estimation: the Gaussian approximation

$$\boxed{p(\theta \mid \text{data}, I) \quad \propto \quad p(\text{data} \mid \theta, I) \times p(\theta \mid I)}$$

'Best' estimator: $\left.\dfrac{\partial p(\theta \mid \text{data}, I)}{\partial \theta}\right|_{\theta=\theta_0} = 0$ ← Maximise posterior

Equivalently, we can define $\ell = \log p(\theta \mid \text{data}, I)$ and compute $\left.\dfrac{\partial \ell}{\partial \theta}\right|_{\theta=\theta_0} = 0$

Parameter estimation: the Gaussian approximation

$$\boxed{p(\theta \mid \text{data}, \ I) \quad \propto \quad p(\text{data} \ \mid \theta, I) \times p(\theta \mid I)}$$

'Best' estimator: $\left.\dfrac{\partial p(\theta \mid \text{data}, I)}{\partial \theta}\right|_{\theta=\theta_0} = 0$ ← Maximise posterior

Equivalently, we can define $\ell = \log p(\theta \mid \text{data}, I)$ and compute $\left.\dfrac{\partial \ell}{\partial \theta}\right|_{\theta=\theta_0} = 0$

Taylor expand $\ell(\theta)$ around $\theta = \theta_0$:

$$\boxed{\ell(\theta) \quad = \quad \ell(\theta_0) + \left.\frac{\partial \ell}{\partial \theta}\right|_{\theta=\theta_0} (\theta - \theta_0) + \frac{1}{2} \left.\frac{\partial^2 \ell}{\partial \theta^2}\right|_{\theta=\theta_0} (\theta - \theta_0)^2 + \ldots}$$

Parameter estimation: the Gaussian approximation

$$\boxed{p(\theta \mid \text{data}, \ I) \quad \propto \quad p(\text{data} \mid \theta, I) \times p(\theta \mid I)}$$

'Best' estimator: $\left.\dfrac{\partial p(\theta \mid \text{data}, I)}{\partial \theta}\right|_{\theta=\theta_0} = 0$ $\longleftarrow$ Maximise posterior

Equivalently, we can define $\ell = \log p(\theta \mid \text{data}, I)$ and compute $\left.\dfrac{\partial \ell}{\partial \theta}\right|_{\theta=\theta_0} = 0$

Taylor expand $\ell(\theta)$ around $\theta=\theta_0$:

$$\boxed{\ell(\theta) \quad = \quad \ell(\theta_0) + \cancel{\left.\frac{\partial \ell}{\partial \theta}\right|_{\theta=\theta_0} (\theta-\theta_0)} + \frac{1}{2}\left.\frac{\partial^2 \ell}{\partial \theta^2}\right|_{\theta=\theta_0} (\theta-\theta_0)^2 + \ldots}$$

Parameter estimation: the Gaussian approximation

$$p(\theta \mid \text{data}, \; I) \quad = \exp\left[\ell(\theta)\right]$$

Neglecting higher order terms in $\ell(\theta)$

$$p(\theta \mid \text{data}, \; I) \quad \propto \quad \exp\left(-\frac{A}{2}(\theta - \theta_0)^2 \right)$$

where $A = -\left.\dfrac{\partial^2 \ell}{\partial \theta^2}\right|_{\theta=\theta_0}$

This is equivalent to a normal distribution, with $\sigma^{-2} = A = -\left.\dfrac{\partial^2 \ell}{\partial \theta^2}\right|_{\theta=\theta_0}$

Parameter estimation: the Gaussian approximation

$$p(\theta \mid \text{data}, \; I) \quad = \exp\left[\ell(\theta)\right]$$

Neglecting higher order terms in $\ell(\theta)$ ⟵ Gaussian approximation

$$p(\theta \mid \text{data}, \; I) \quad \propto \quad \exp\left(-\frac{A}{2}(\theta - \theta_0)^2 \right)$$

where $A = -\left.\dfrac{\partial^2 \ell}{\partial \theta^2}\right|_{\theta=\theta_0}$

This is equivalent to a normal distribution, with $\sigma^{-2} = A = -\left.\dfrac{\partial^2 \ell}{\partial \theta^2}\right|_{\theta=\theta_0}$

Parameter estimation: the Gaussian approximation

$$p(\theta \mid \text{data}, \ I) \quad = \exp\left[\ell(\theta)\right]$$

Neglecting higher order terms in $\ell(\theta)$ ← Gaussian approximation

$$p(\theta \mid \text{data}, \ I) \quad \propto \quad \exp\left(-\frac{A}{2}(\theta - \theta_0)^2 \right)$$

where $A = -\left.\frac{\partial^2 \ell}{\partial \theta^2}\right|_{\theta=\theta_0}$

This is equivalent to a normal distribution, with $\sigma^{-2} = A = -\left.\frac{\partial^2 \ell}{\partial \theta^2}\right|_{\theta=\theta_0}$

Can summarise inference from posterior by

$$\boxed{\theta = \theta_0 \pm \sigma}$$

University of Glasgow

SUPA

Question 13: Neglecting the higher order terms in the log posterior expansion produces a posterior which can be written as a normal pdf because

A The higher order moments of a Gaussian are all zero

B The Gaussian pdf is uniquely specified by its mean and variance

C The logarithm of a Gaussian pdf can be written in the form of a quadratic

D All of the above

Parameter estimation: 2-D case

Recall our definition of *variance*

$$\text{var}[x] \;=\; \int_{-\infty}^{\infty} \left(x - \langle x \rangle\right)^2 p(x \mid I)\, dx$$

Extends to 2 variables – *covariance*

$$\text{cov}[x, y] \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(x - \langle x \rangle\right)\left(y - \langle y \rangle\right) p(x, y \mid I)\, dx dy$$

Parameter estimation: 2-D case

Recall our definition of *variance*

$$\mathrm{var}[x] \;=\; \int_{-\infty}^{\infty} \left(x - \langle x \rangle\right)^2 p(x \mid I)\,dx$$

Extends to 2 variables – *covariance*

$$\mathrm{cov}[x,y] \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(x - \langle x \rangle\right)\left(y - \langle y \rangle\right) p(x,y \mid I)\,dx\,dy$$

If x and y are independent, $\mathrm{cov}[x,y] \;=\; 0$

This is because $p(x,y \mid I) = p(x \mid I)\,p(y \mid I)$

Parameter estimation: 2-D case

$$\boxed{p(\theta_1, \theta_2 \mid \text{data}, \; I) \quad \propto \quad p(\text{data} \; \mid \theta_1, \theta_2, I) \times p(\theta_1, \theta_2 \mid I)}$$

'Best' estimator: $$\left.\frac{\partial p(\theta_1, \theta_2 \mid \text{data}, I)}{\partial \theta_j}\right|_{\theta_j = \theta_{0j}} = 0$$

Compute $$\left.\frac{\partial \ell}{\partial \theta_j}\right|_{\theta_j = \theta_{0j}} = 0$$ where $$\ell = \log p(\theta_1, \theta_2 \mid \text{data}, I)$$

Parameter estimation: 2-D case

$$\boxed{p(\theta_1, \theta_2 \mid \text{data}, \ I) \quad \propto \quad p(\text{data} \ \mid \theta_1, \theta_2, I) \times p(\theta_1, \theta_2 \mid I)}$$

'Best' estimator: $\left.\dfrac{\partial p(\theta_1, \theta_2 \mid \text{data}, I)}{\partial \theta_j}\right|_{\theta_j = \theta_{0j}} = 0$

Compute $\left.\dfrac{\partial \ell}{\partial \theta_j}\right|_{\theta_j = \theta_{0j}} = 0$ where $\ell = \log p(\theta_1, \theta_2 \mid \text{data}, I)$

Taylor expand $\ell(\theta_1, \theta_2)$ around θ_{0j} :

Parameter estimation: 2-D case

Taylor expand $\ell(\theta_1, \theta_2)$ around θ_{0j} :

$$\ell(\theta_1,\theta_2) = \ell(\theta_{01},\theta_{02}) + \left.\frac{\partial \ell}{\partial \theta_1}\right|_{\theta_j=\theta_{0j}} (\theta_1-\theta_{01}) + \left.\frac{\partial \ell}{\partial \theta_2}\right|_{\theta_j=\theta_{0j}} (\theta_2-\theta_{02}) +$$

$$\frac{1}{2}\left[\left.\frac{\partial^2 \ell}{\partial \theta_1^{\,2}}\right|_{\theta_j=\theta_{0j}} (\theta_1-\theta_{01})^2 + \left.\frac{\partial^2 \ell}{\partial \theta_2^{\,2}}\right|_{\theta_j=\theta_{0j}} (\theta_2-\theta_{02})^2 + 2\left.\frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2}\right|_{\theta_j=\theta_{0j}} (\theta_1-\theta_{01})(\theta_2-\theta_{02}) \right] + \ldots$$

Parameter estimation: 2-D case

Taylor expand $\ell(\theta_1,\theta_2)$ around θ_{0j} :

$$\ell(\theta_1,\theta_2) = \ell(\theta_{01},\theta_{02}) + \cancel{\left.\frac{\partial \ell}{\partial \theta_1}\right|_{\theta_j=\theta_{0j}}(\theta_1-\theta_{01})} + \cancel{\left.\frac{\partial \ell}{\partial \theta_2}\right|_{\theta_j=\theta_{0j}}(\theta_2-\theta_{02})} +$$

$$\frac{1}{2}\left[\left.\frac{\partial^2 \ell}{\partial \theta_1^2}\right|_{\theta_j=\theta_{0j}}(\theta_1-\theta_{01})^2 + \left.\frac{\partial^2 \ell}{\partial \theta_2^2}\right|_{\theta_j=\theta_{0j}}(\theta_2-\theta_{02})^2 + 2\left.\frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2}\right|_{\theta_j=\theta_{0j}}(\theta_1-\theta_{01})(\theta_2-\theta_{02})\right] + \dots$$

$$p(\theta_1,\theta_2 \mid \text{data}, I) \propto \exp\left[\ell(\theta_1,\theta_2)\right]$$

$$\propto \exp\left[-\tfrac{1}{2}Q\right]$$

Gaussian approximation

χ^2

Parameter estimation: 2-D case

Taylor expand $\ell(\theta_1,\theta_2)$ around θ_{0j} :

$$\boxed{\begin{aligned}\ell(\theta_1,\theta_2) \;=\; \ell(\theta_{01},\theta_{02}) \;+\; \cancel{\left.\frac{\partial \ell}{\partial \theta_1}\right|_{\theta_j=\theta_{0j}} (\theta_1-\theta_{01})} \;+\; \cancel{\left.\frac{\partial \ell}{\partial \theta_2}\right|_{\theta_j=\theta_{0j}} (\theta_2-\theta_{02})} \;+ \\ \frac{1}{2}\left[\left.\frac{\partial^2 \ell}{\partial \theta_1^{\,2}}\right|_{\theta_j=\theta_{0j}} (\theta_1-\theta_{01})^2 + \left.\frac{\partial^2 \ell}{\partial \theta_2^{\,2}}\right|_{\theta_j=\theta_{0j}} (\theta_2-\theta_{02})^2 + 2\left.\frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2}\right|_{\theta_j=\theta_{0j}} (\theta_1-\theta_{01})(\theta_2-\theta_{02})\right] \;+\ldots\end{aligned}}$$

$$p(\theta_1,\theta_2 \mid \text{data}, \; I) \;\propto\; \exp\left[\ell(\theta_1,\theta_2)\right]$$

$$\propto\; \exp\left[-\tfrac{1}{2}Q\right]$$

Gaussian approximation

χ^2

Maximising posterior

$\equiv$ Minimising χ^2

Parameter estimation: 2-D case

Taylor expand $\ell(\theta_1, \theta_2)$ around θ_{0j} :

$$\boxed{Q = \begin{pmatrix} \theta_1 - \theta_{10} & \theta_2 - \theta_{20} \end{pmatrix} \begin{bmatrix} A & C \\ C & B \end{bmatrix} \begin{pmatrix} \theta_1 - \theta_{10} \\ \theta_2 - \theta_{20} \end{pmatrix}}$$

Quadratic form

where $A = \left. \dfrac{\partial^2 \ell}{\partial \theta_1^{\,2}} \right|_{\theta_j = \theta_{0j}}$ $\quad B = \left. \dfrac{\partial^2 \ell}{\partial \theta_2^{\,2}} \right|_{\theta_j = \theta_{0j}}$ $\quad C = \left. \dfrac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2} \right|_{\theta_j = \theta_{0j}}$

Parameter estimation: 2-D case

Taylor expand $\ell(\theta_1, \theta_2)$ around θ_{0j} :

$$\boxed{Q = \begin{pmatrix} \theta_1 - \theta_{10} & \theta_2 - \theta_{20} \end{pmatrix} \begin{bmatrix} A & C \\ C & B \end{bmatrix} \begin{pmatrix} \theta_1 - \theta_{10} \\ \theta_2 - \theta_{20} \end{pmatrix}}$$

Quadratic form

where $A = \left.\dfrac{\partial^2 \ell}{\partial \theta_1^{\,2}}\right|_{\theta_j = \theta_{0j}}$ $\quad B = \left.\dfrac{\partial^2 \ell}{\partial \theta_2^{\,2}}\right|_{\theta_j = \theta_{0j}}$ $\quad C = \left.\dfrac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2}\right|_{\theta_j = \theta_{0j}}$

This is a bivariate normal distribution with covariance matrix

$$\sigma_{ij}^2 = \mathrm{cov}_{ij} = \left\langle (\theta_i - \theta_{i0})(\theta_j - \theta_{j0}) \right\rangle = \left[-\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right]^{-1}$$

Fisher information matrix

$$\mathbf{F} \equiv F_{ij} = \frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j}$$ is known as the **Fisher information matrix**

It provides a measure of how much information a given dataset can yield about the parameters of a model.

We can see this most easily in the case where the Fisher matrix is **diagonal**.

Then $$\mathbf{F} = -\mathrm{diag}\left(\sigma_1^{-2}, \ldots, \sigma_n^{-2}\right)$$

If the i^{th} element of the Fisher matrix is large (negative), the **variance** of parameter θ_i is small (and positive).

In general the Fisher matrix (and covariance matrix) will **not** be diagonal; the Fisher matrix then tells us which **combinations** of the parameters are well constrained by the data. (see later).

So if, for our model:

- the likelihood is Gaussian in shape (or if we can approximate it as Gaussian – i.e. if the higher order terms in the Taylor expansion of the log likelihood can be neglected);
- the parameters have broad, uniform priors;

then the posterior will also be Gaussian.

If we can evaluate the first and second partial derivatives of the log likelihood, we can:

- compute the **Fisher Information Matrix;**
- compute the **Covariance Matrix** of the posterior.

We can also compute **credible regions** for the parameters (in fact for this we don't need the derivatives – see Section 9)

We can write the log posterior as

$$\ell(\theta_1,\theta_2) \;=\; \text{const} \;-\tfrac{1}{2}\chi^2(\theta_1,\theta_2)$$

Now $\chi^2 = \chi^2_{\min}$ when $(\theta_1,\theta_2)=(\theta_{01},\theta_{02})$

Maximising posterior $\equiv$ Minimising χ^2

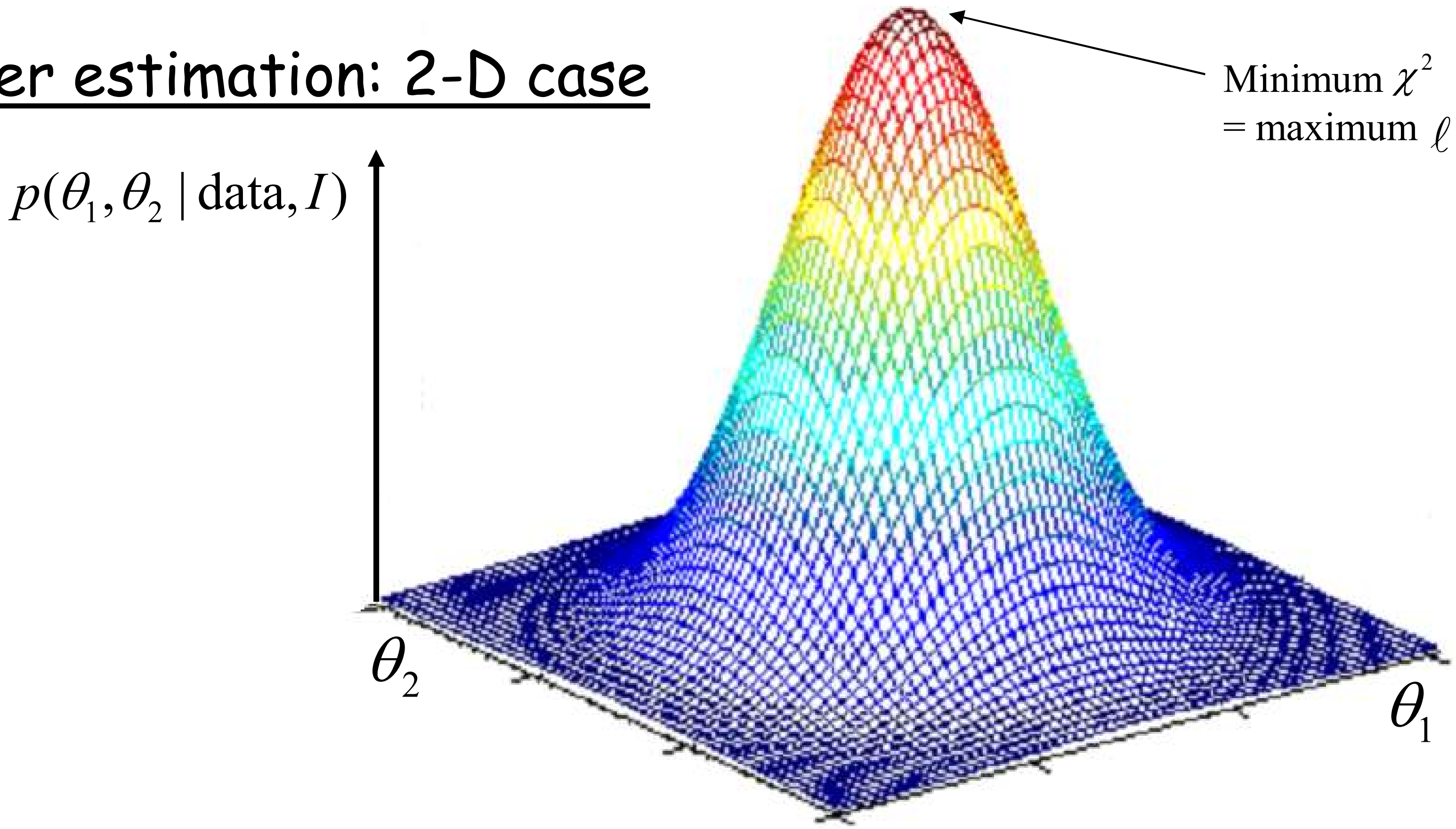
so we can write, for $\Delta\chi^2(\theta_1,\theta_2)=\chi^2(\theta_1,\theta_2)-\chi^2_{\min}$

$$\ell(\theta_1,\theta_2) \;=\; \ell(\theta_{01},\theta_{02}) \;-\tfrac{1}{2}\Delta\chi^2(\theta_1,\theta_2)$$

So that

$$p(\theta_1,\theta_2 \mid \text{data}, I) \;=\; \underbrace{p(\theta_{01},\theta_{02} \mid \text{data}, I)}_{\text{Maximum of the posterior}} \;\exp\left[-\tfrac{1}{2}\Delta\chi^2(\theta_1,\theta_2)\right]$$

Parameter estimation: 2-D case

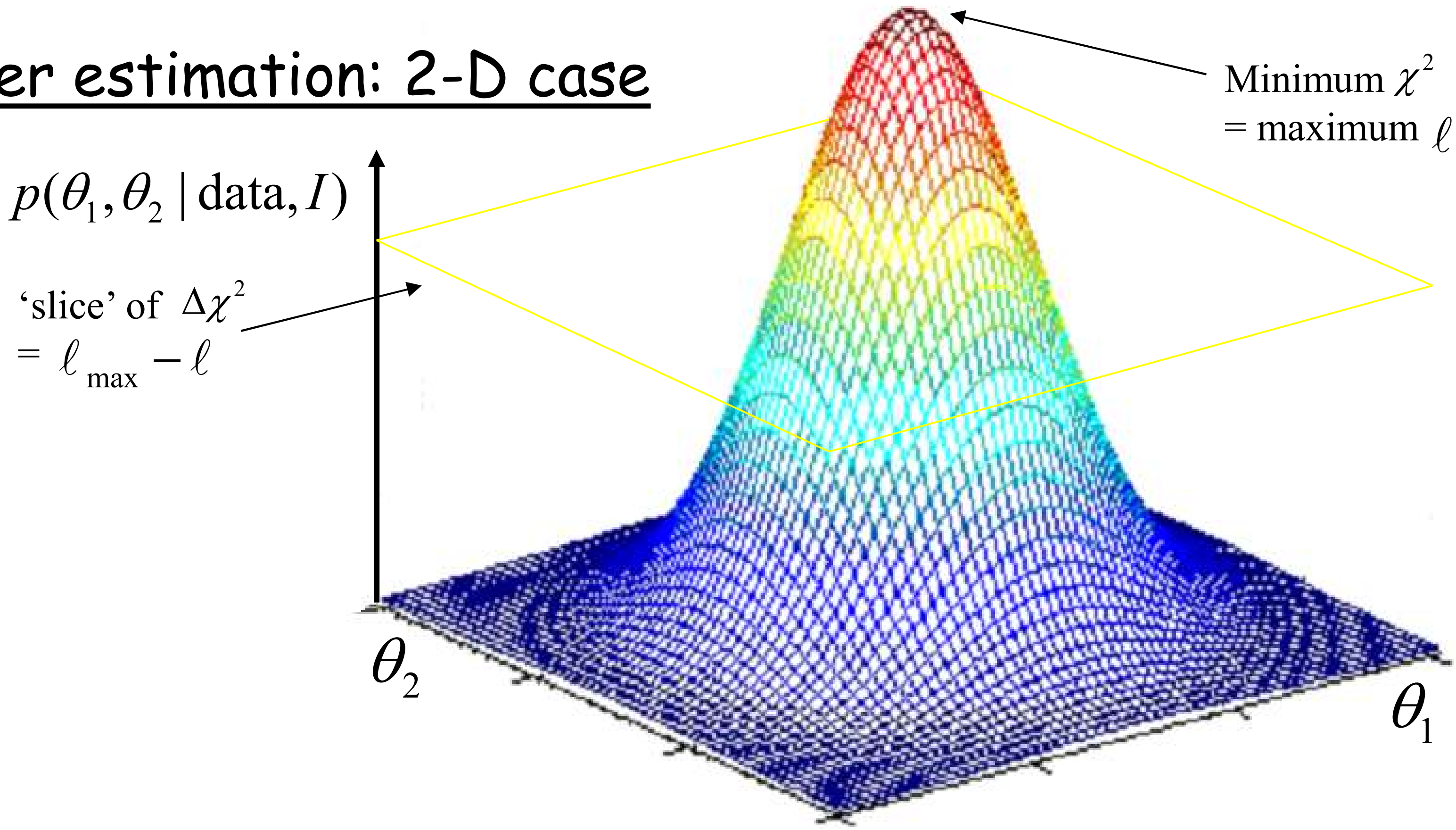


This is a bivariate normal distribution with covariance matrix

$$\sigma_{ij}^2 = \text{cov}_{ij} = \left\langle (\theta_i - \theta_{i0})(\theta_j - \theta_{j0}) \right\rangle = \left[-\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right]^{-1}$$

Fisher information matrix

Parameter estimation: 2-D case

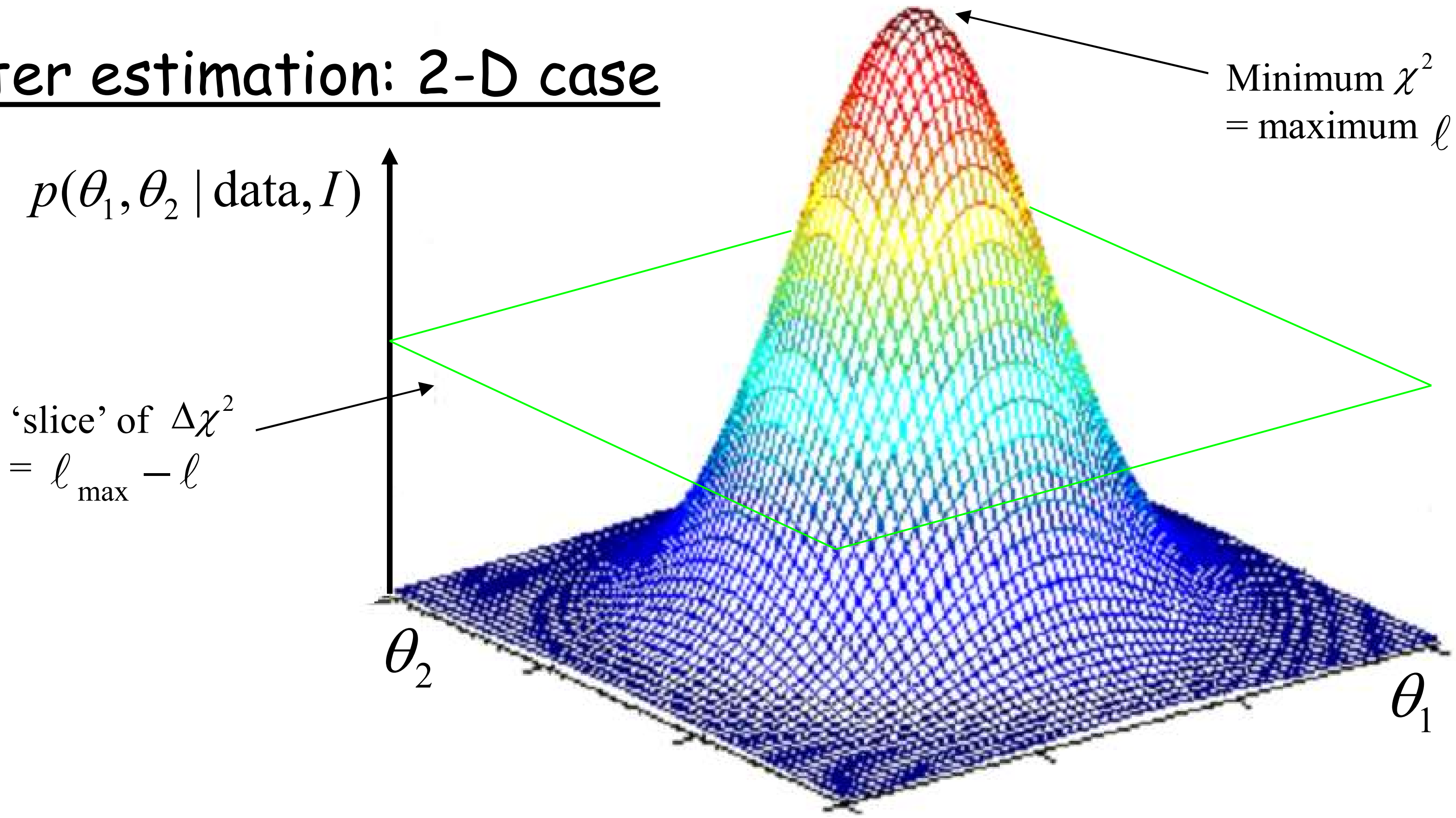


This is a bivariate normal distribution with covariance matrix

$$\sigma_{ij}^2 \quad = \quad \mathrm{cov}_{ij} \quad = \quad \left\langle (\theta_i - \theta_{i0})(\theta_j - \theta_{j0}) \right\rangle \quad = \quad \left[-\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right]^{-1}$$

Fisher information matrix

Parameter estimation: 2-D case



This is a bivariate normal distribution with covariance matrix

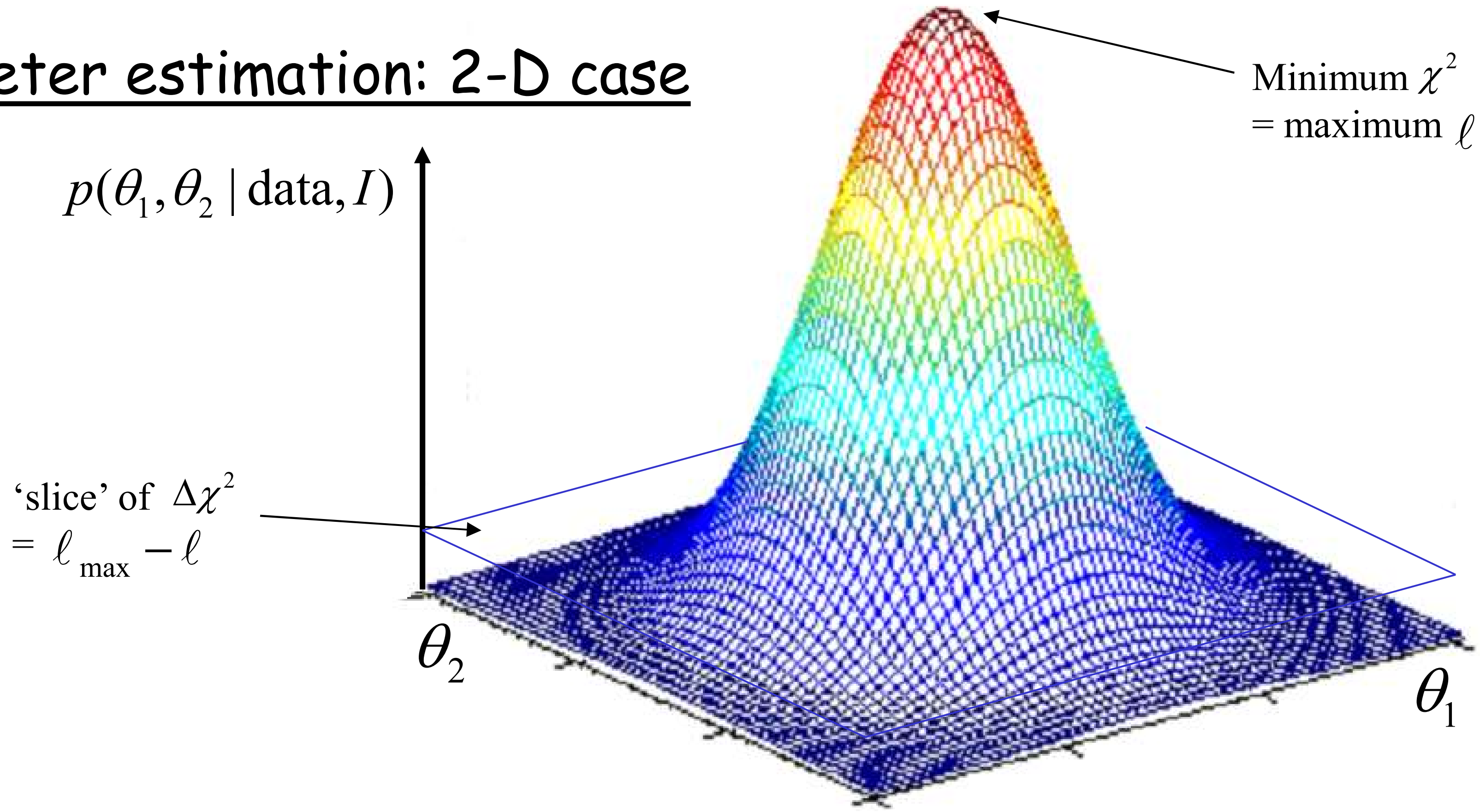
$$\sigma_{ij}^2 \quad = \quad \mathrm{cov}_{ij} \quad = \quad \left\langle (\theta_i - \theta_{i0})(\theta_j - \theta_{j0}) \right\rangle \quad = \quad \left[-\frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j} \right]^{-1}$$

Fisher information matrix

University of Glasgow

SUPA

Parameter estimation: 2-D case

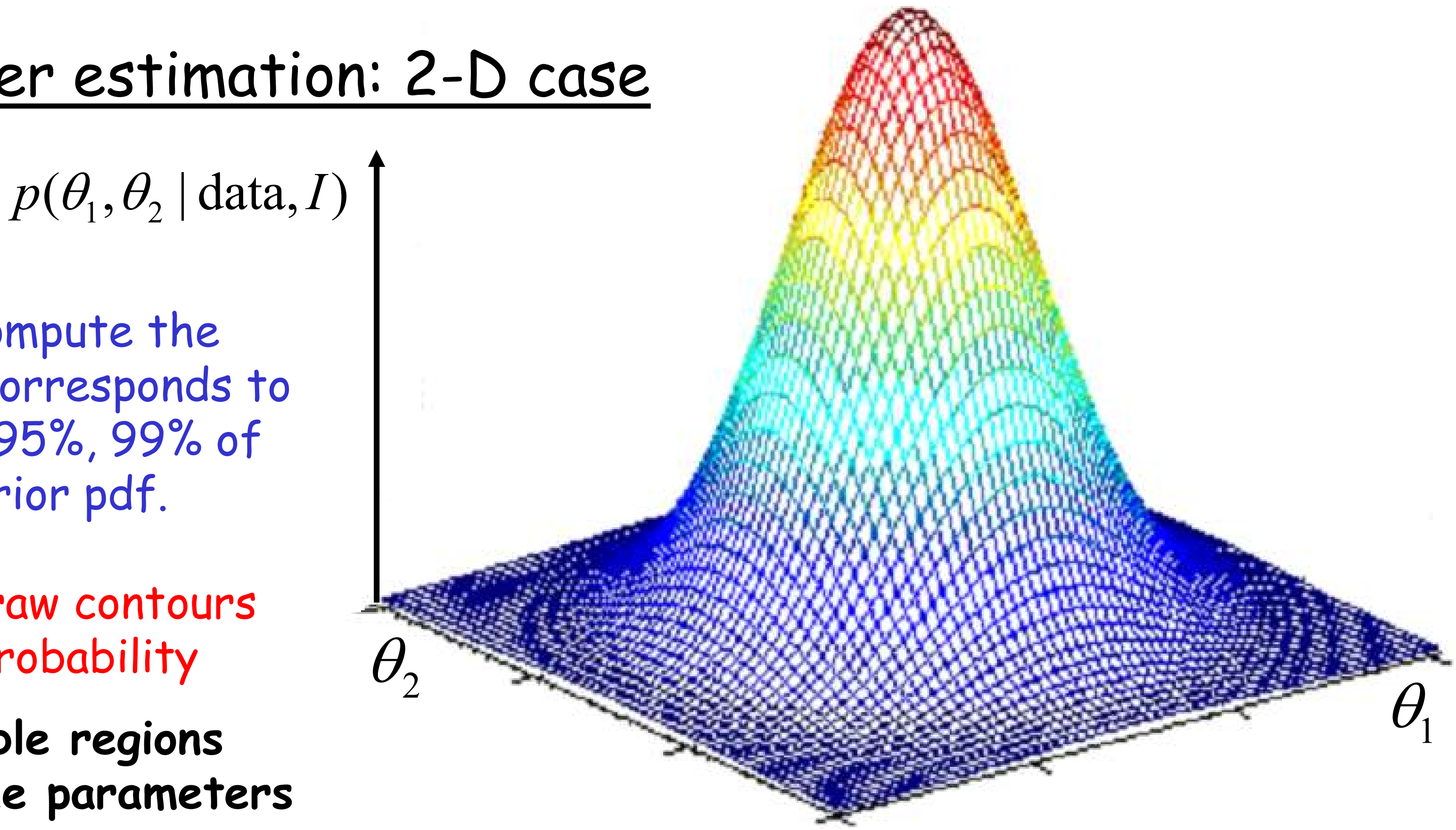


This is a bivariate normal distribution with covariance matrix

$$\sigma_{ij}^2 \quad = \quad \mathrm{cov}_{ij} \quad = \quad \left\langle (\theta_i - \theta_{i0})(\theta_j - \theta_{j0}) \right\rangle \quad = \quad \left[-\frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j} \right]^{-1}$$

Fisher information matrix

Parameter estimation: 2-D case



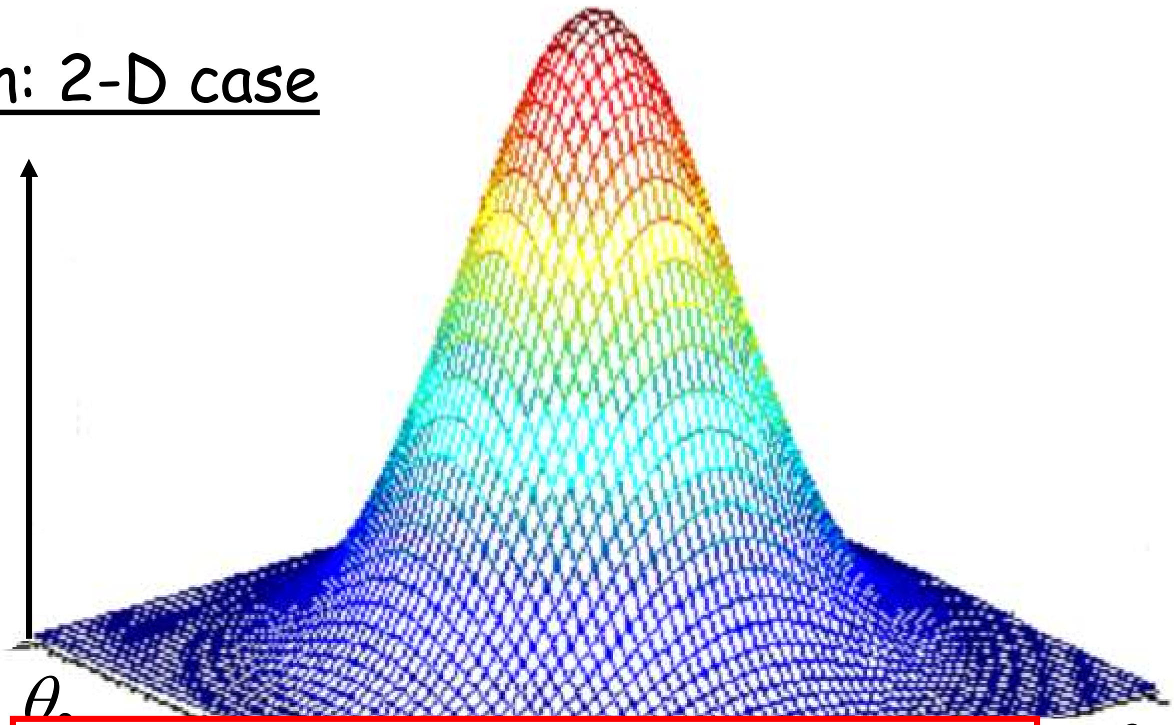
We can compute the $\Delta\chi^2$ that corresponds to e.g. 68%, 95%, 99% of the posterior pdf.

We can draw contours of equal probability

$\Rightarrow$ **Credible regions for the parameters**

Extends easily to N parameters – or *degrees of freedom*

Parameter estimation: 2-D case



We can compute the $\Delta\chi^2$ that corresponds to e.g. 68%, 95%, 99% of the posterior pdf.

We can draw contours of equal probability

$\Rightarrow$ **Credible regions for the parameters**

Extends easily to N parameters – or *degrees of freedom*

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

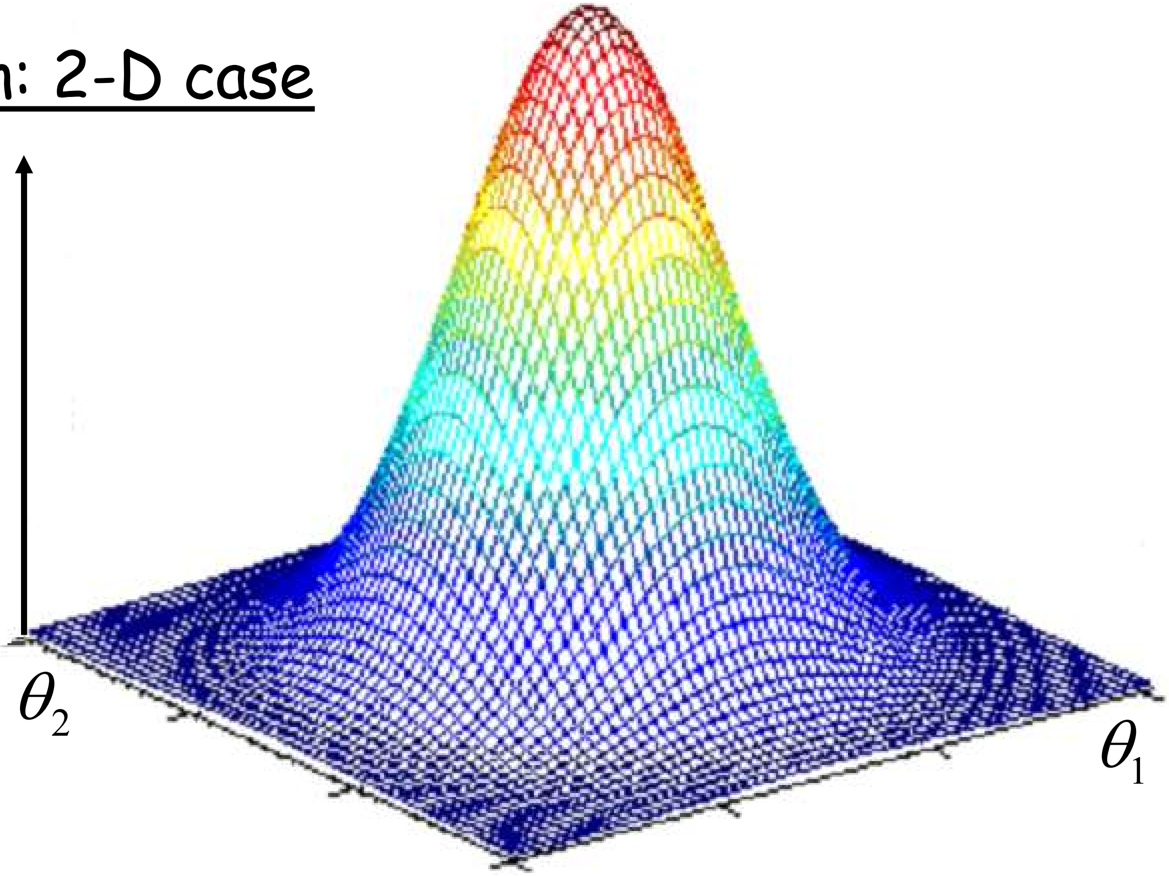
	ν					
p	1	2	3	4	5	6
68.3%	1.00	2.30	3.53	4.72	5.89	7.04
90%	2.71	4.61	6.25	7.78	9.24	10.6
95.4%	4.00	6.17	8.02	9.70	11.3	12.8
99%	6.63	9.21	11.3	13.3	15.1	16.8
99.73%	9.00	11.8	14.2	16.3	18.2	20.1
99.99%	15.1	18.4	21.1	23.5	25.7	27.8

From Numerical Recipes

University of Glasgow

SUPA

Parameter estimation: 2-D case



Contours of constant probability are ellipses.

Covariance matrix is *not* in general diagonal

$\Rightarrow$ What we infer about θ_1 and θ_1 is *not* independent

Parameter estimation: 2-D case

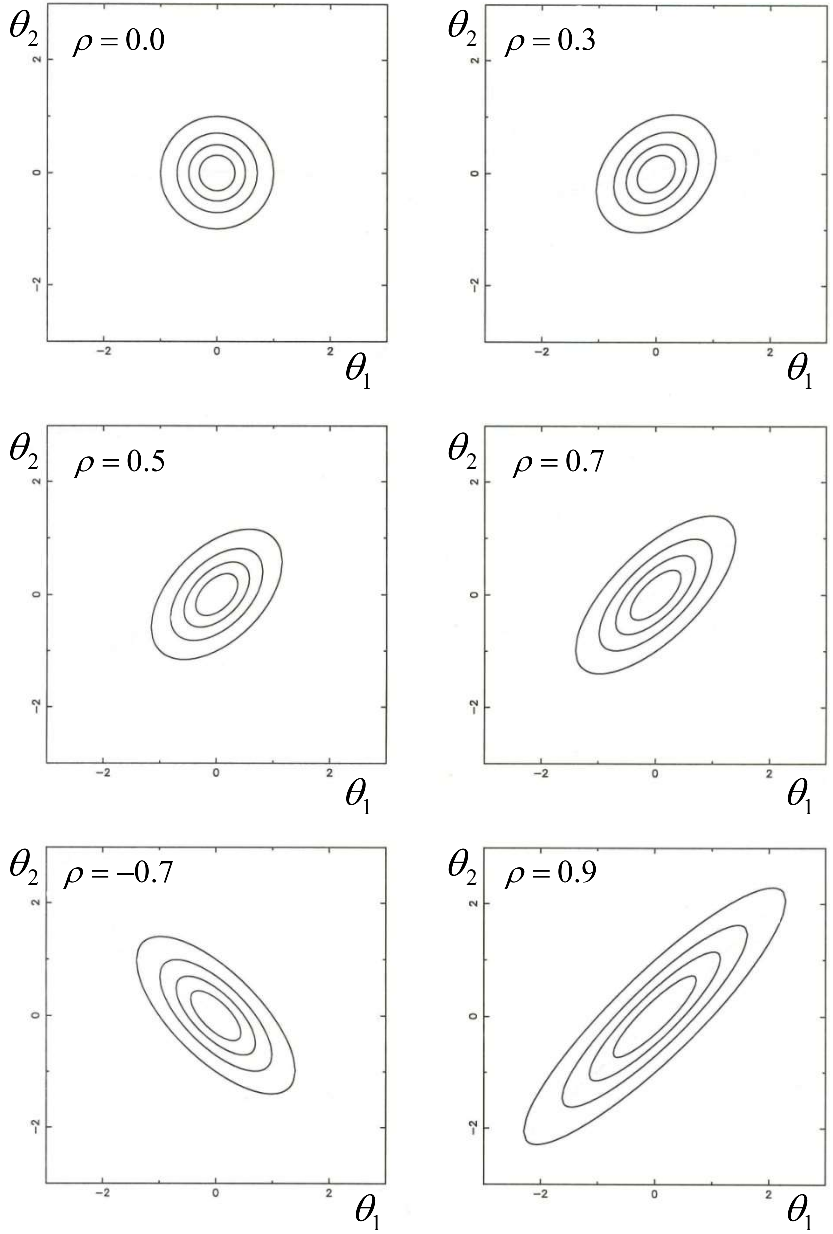
Can define *correlation coefficient*

$$\rho = \frac{\mathrm{cov}[\theta_1, \theta_2]}{\sqrt{\mathrm{var}[\theta_1]}\sqrt{\mathrm{var}[\theta_2]}} \qquad -1 \le \rho \le 1$$

Covariance matrix becomes less diagonal

$\Rightarrow$ $|\rho|$ increases

$\Rightarrow$ isoprobability contours elongate



Parameter estimation: 2-D case

Can define *correlation coefficient*

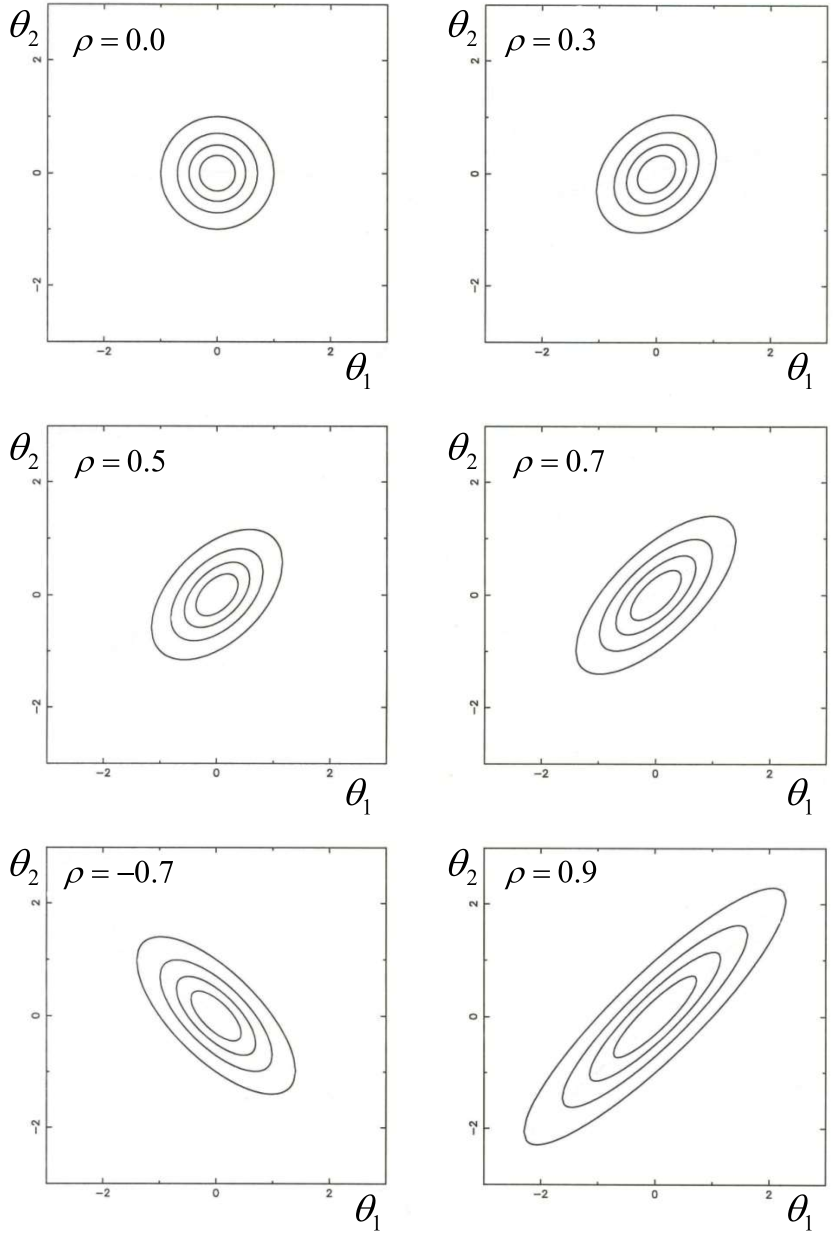
$$\rho = \frac{\mathrm{cov}[\theta_1, \theta_2]}{\sqrt{\mathrm{var}[\theta_1]}\sqrt{\mathrm{var}[\theta_2]}} \qquad -1 \le \rho \le 1$$

Covariance matrix becomes less diagonal

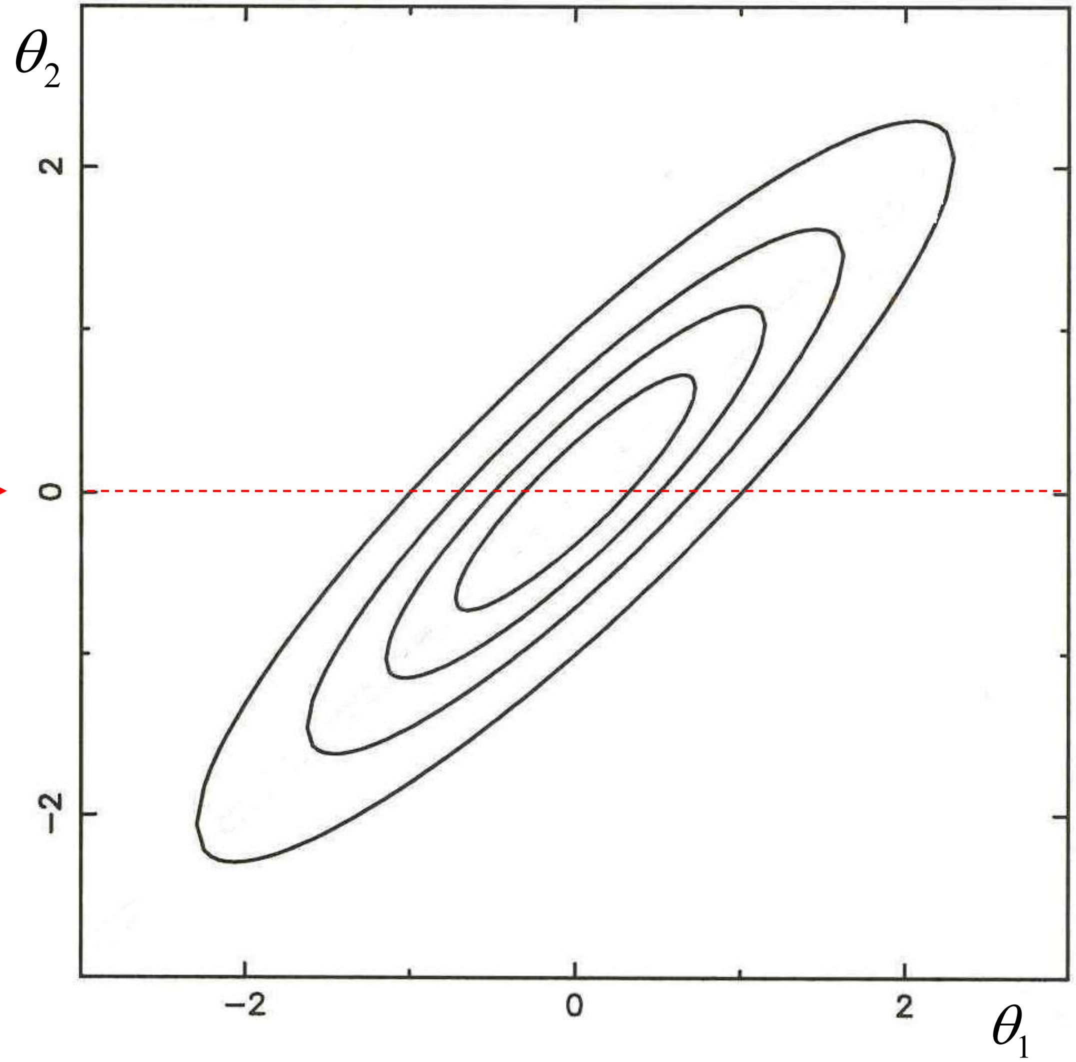
$\Rightarrow$ $|\rho|$ increases

$\Rightarrow$ isoprobability contours elongate

Very important if we are interested only in *one* parameter



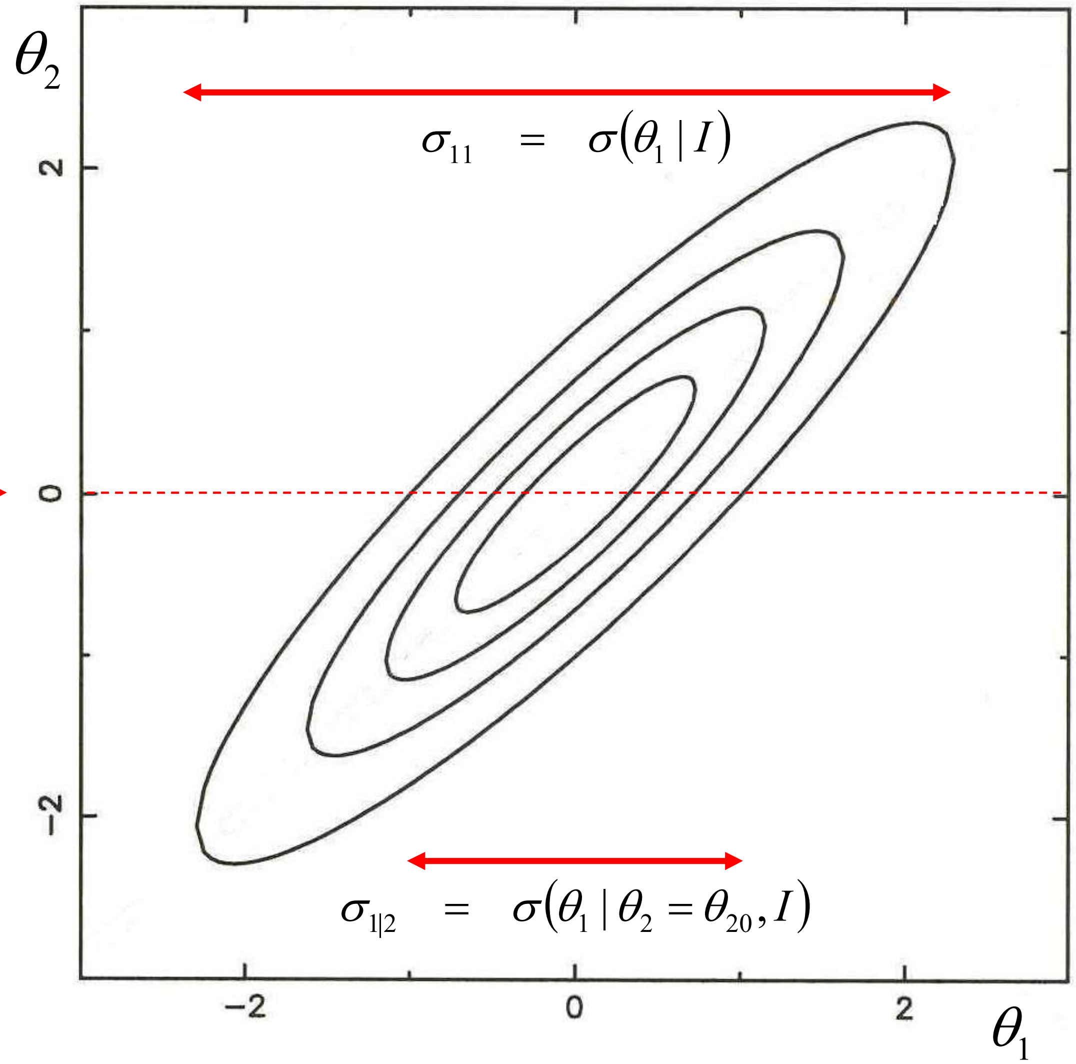
Parameter estimation: 2-D case



'Best-fit' value of θ_2 , found from

$$\left.\frac{\partial \ell}{\partial \theta_j}\right|_{\theta_j=\theta_{0j}} = 0$$

Parameter estimation: 2-D case



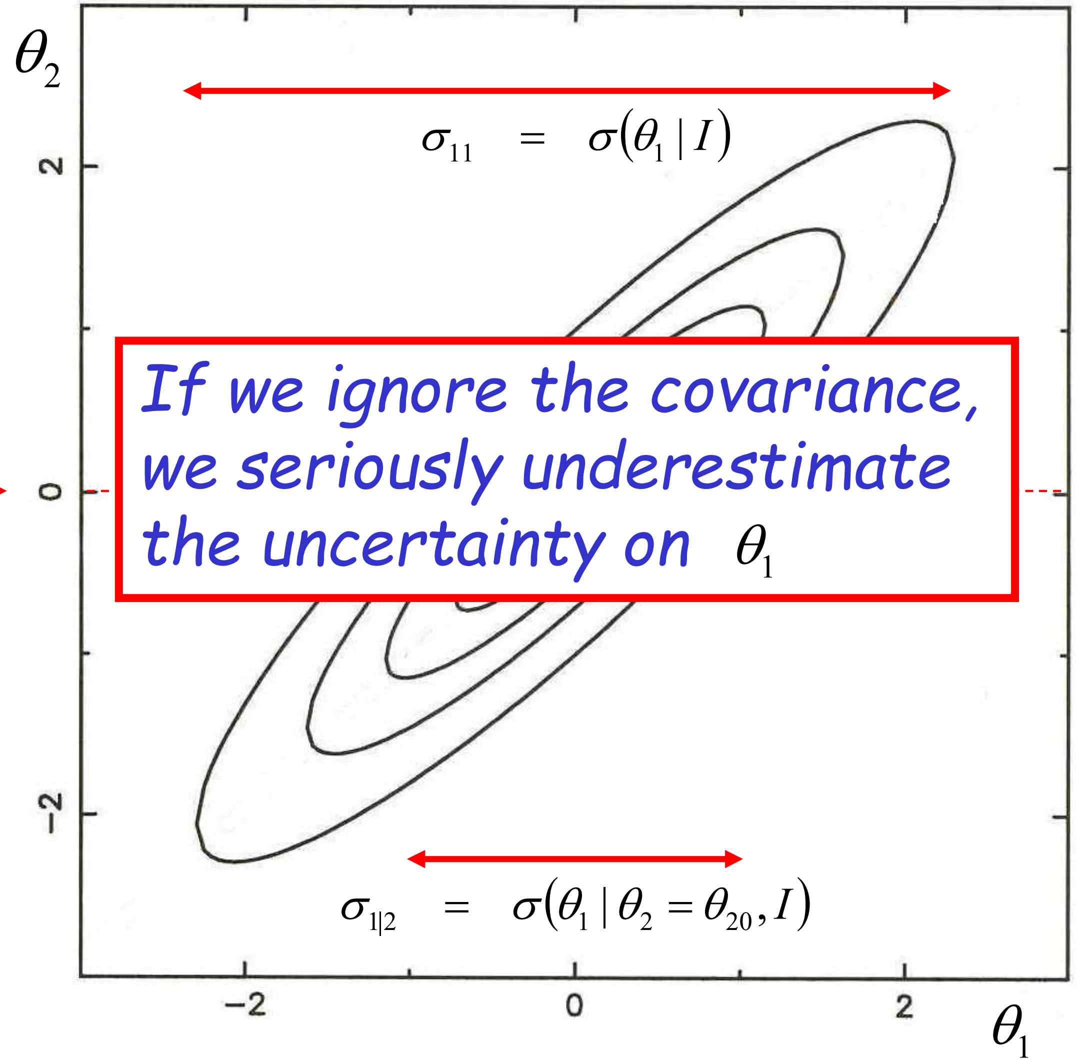
'Best-fit' value of θ_2, found from

$$\left.\frac{\partial \ell}{\partial \theta_j}\right|_{\theta_j = \theta_{0j}} = 0$$

Parameter estimation: 2-D case

'Best-fit' value of θ_2 , found from

$$\left.\frac{\partial \ell}{\partial \theta_j}\right|_{\theta_j = \theta_{0j}} = 0$$



Question 14: The marginal and conditional error bars on θ_1 will be equal provided

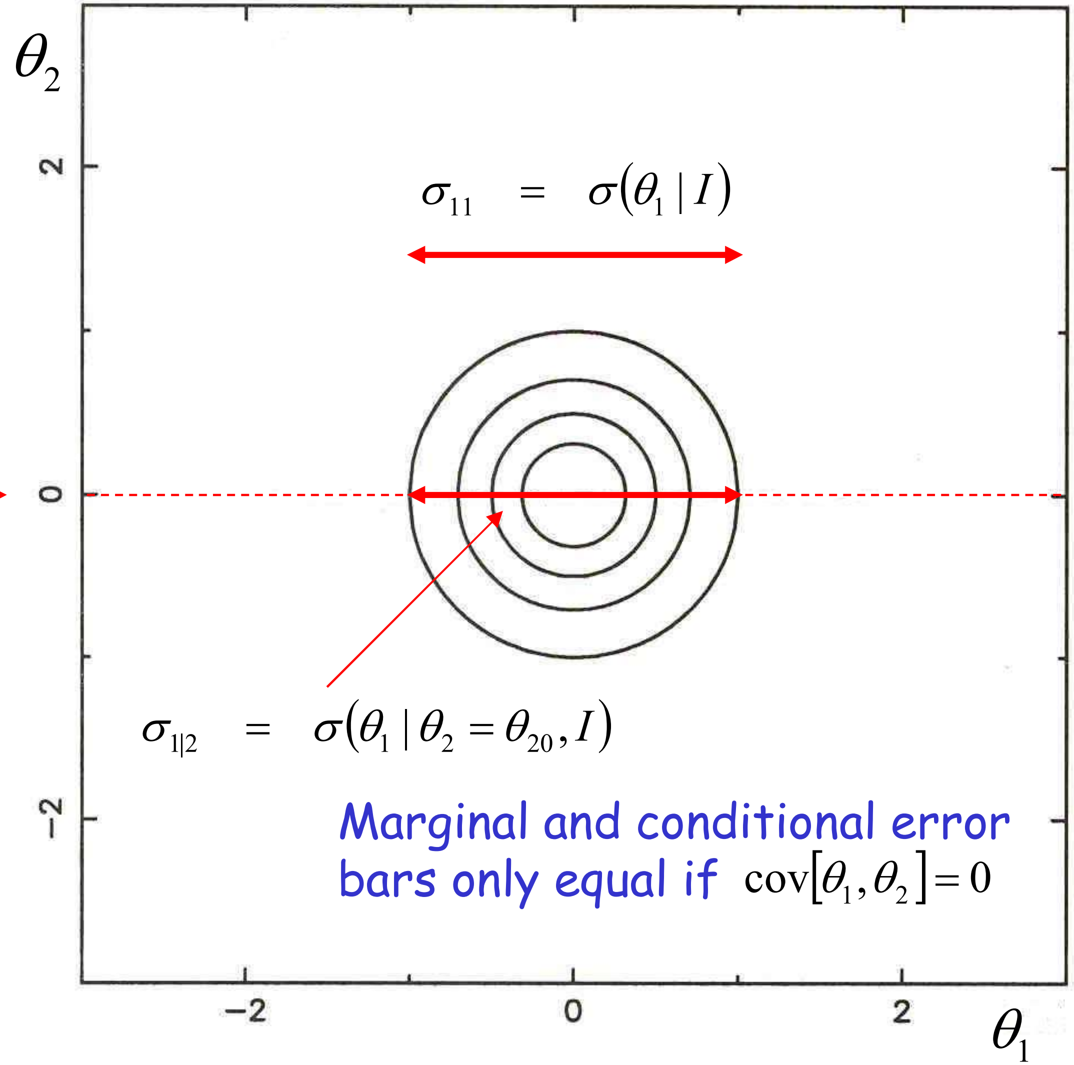
A $\mathrm{cov}[\theta_1, \theta_2] = 0$

B $\mathrm{cov}[\theta_1, \theta_2] = 1$

C $\mathrm{cov}[\theta_1, \theta_2] = -1$

D None of the above

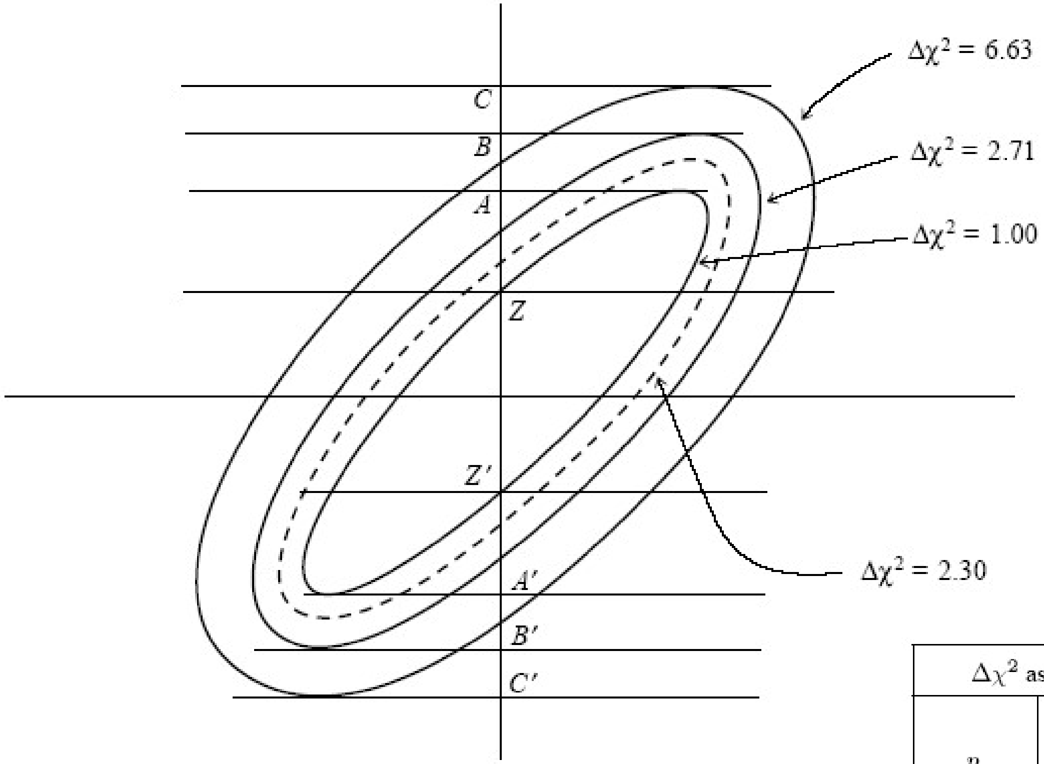
Parameter estimation: 2-D case



'Best-fit' value of θ_2, found from

$$\left.\frac{\partial \ell}{\partial \theta_j}\right|_{\theta_j = \theta_{0j}} = 0$$

Parameter estimation: 2-D case

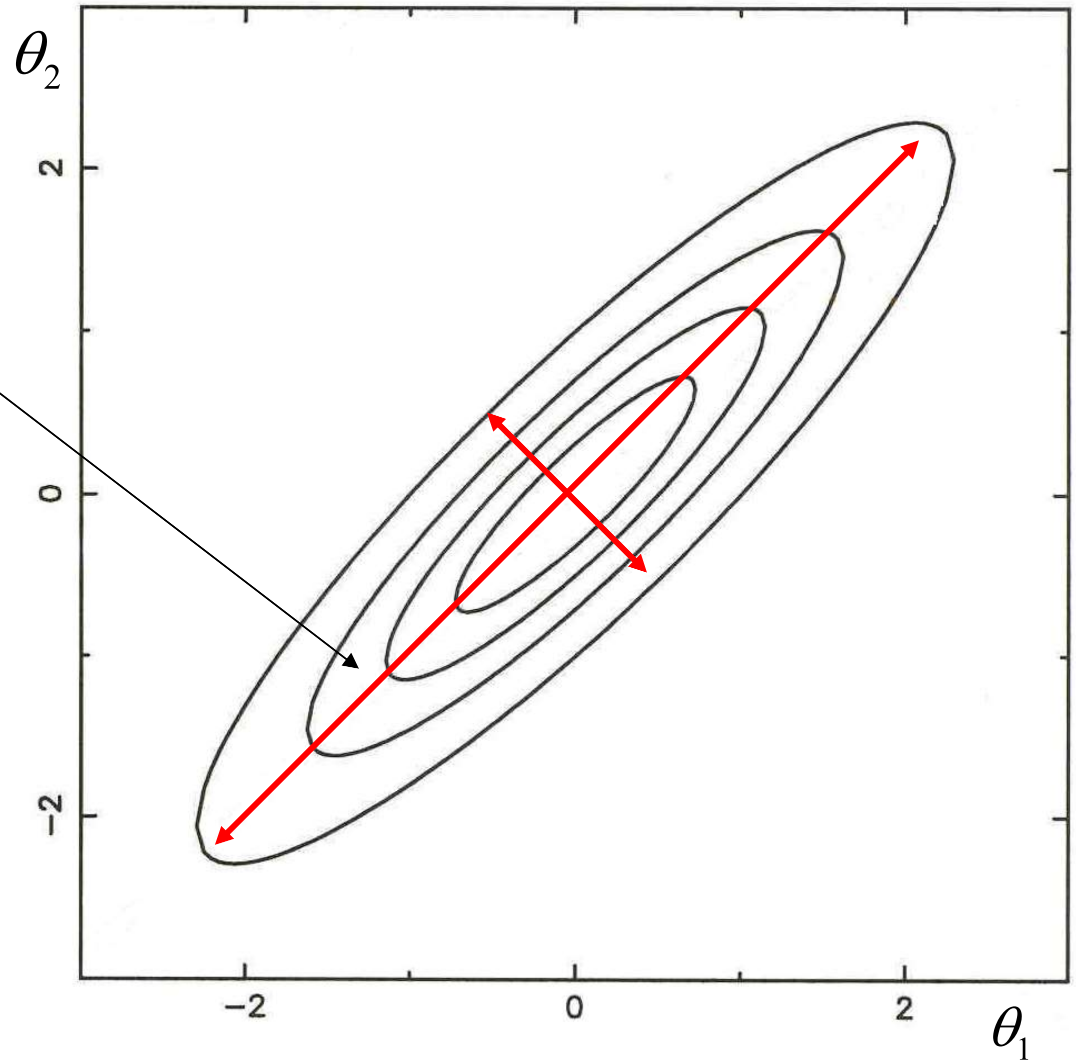


From Numerical Recipes

$\Delta\chi^2$ 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

Parameter estimation: 2-D case

Linear combination of θ_1 and θ_2 well constrained by data



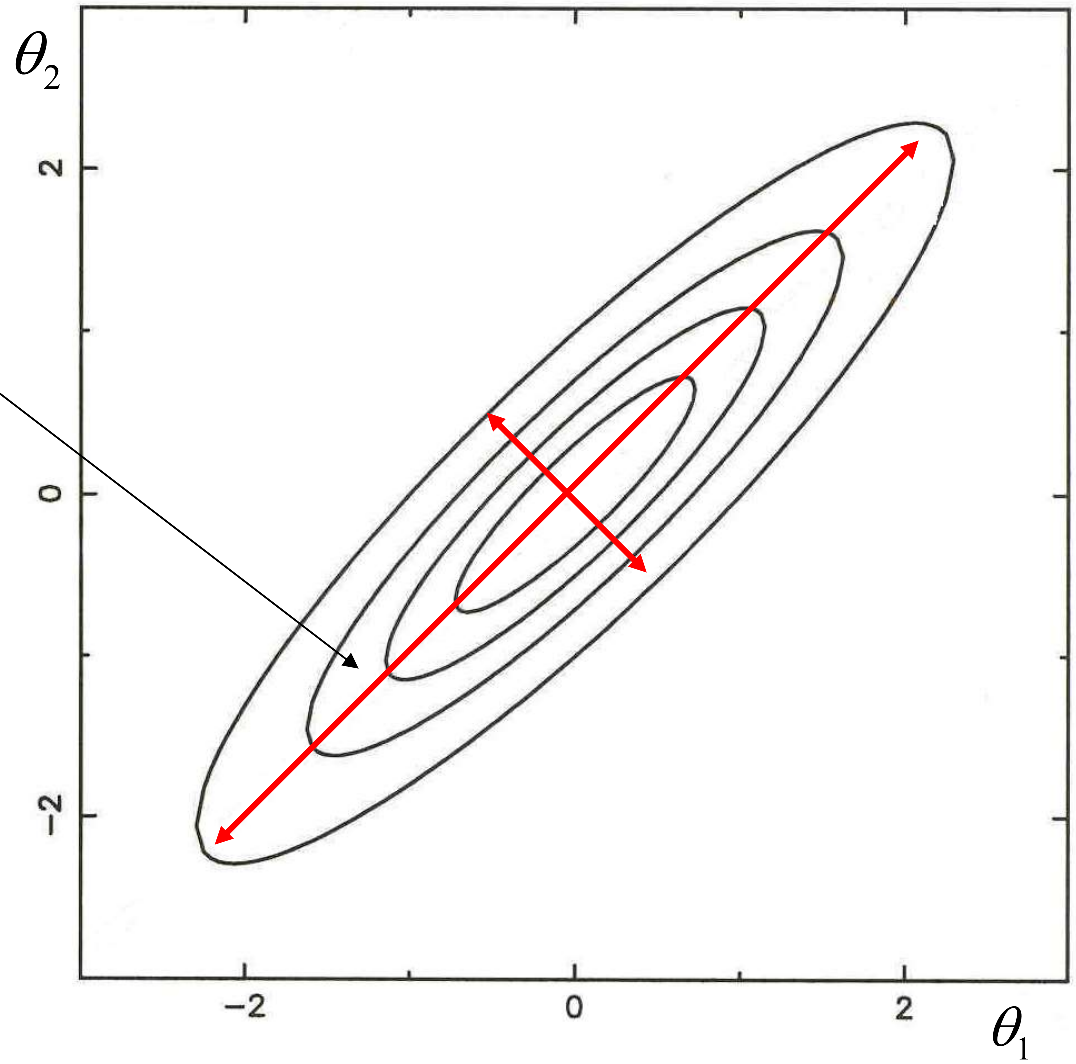
Parameter estimation: 2-D case

Linear combination of θ_1 and θ_2 well constrained by data

Length of axes determined by the **eigenvalues** of the Fisher information matrix

$$F_{ij} = \frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j} = \left[-\sigma_{ij}^2\right]^{-1}$$

$$\boxed{\boldsymbol{F\,\theta} = \lambda\,\boldsymbol{\theta}}$$



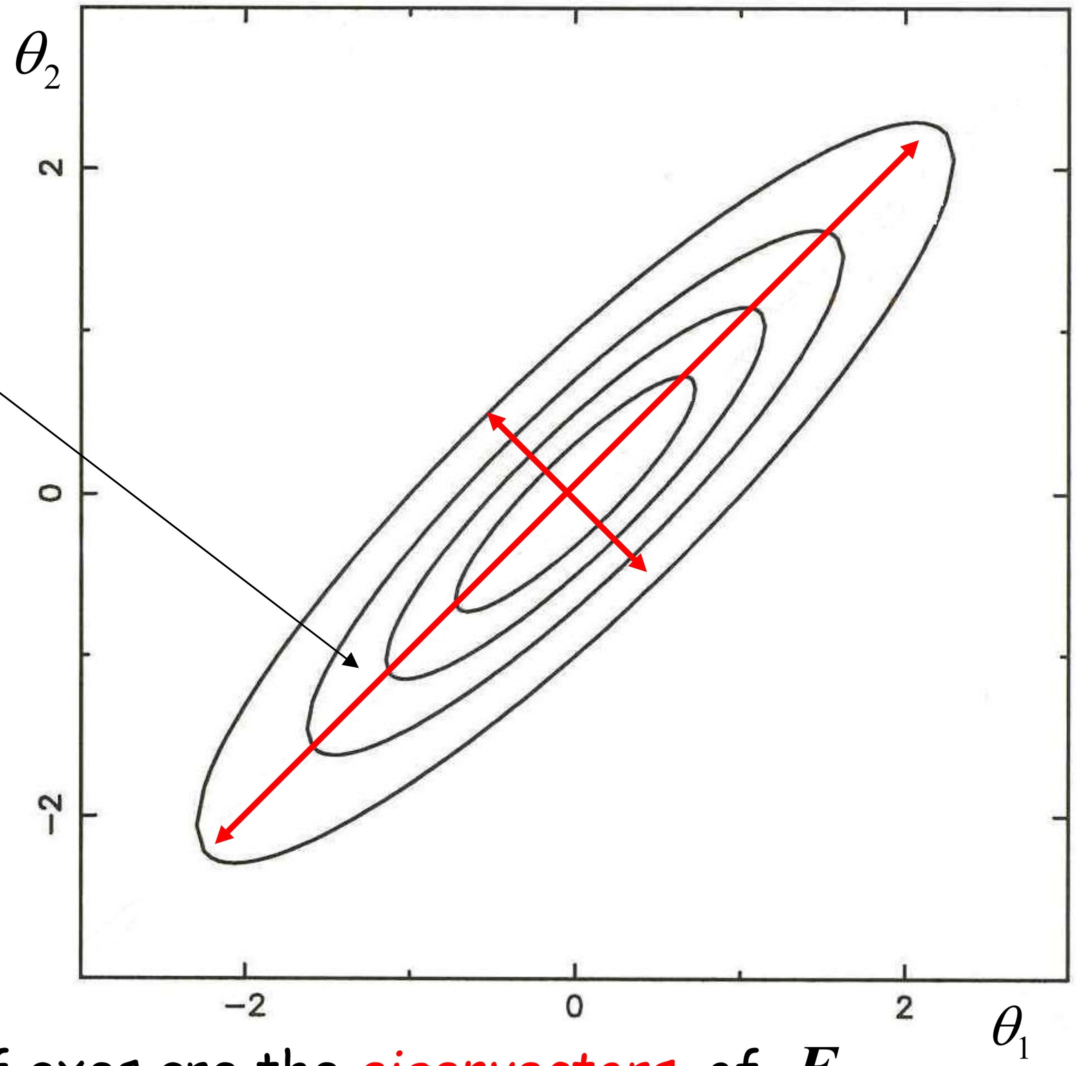
Parameter estimation: 2-D case

Linear combination of θ_1 and θ_2 well constrained by data

Length of axes determined by the **eigenvalues** of the Fisher information matrix

$$F_{ij} = \frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j} = \left[-\sigma_{ij}^2\right]^{-1}$$

$$\boxed{\boldsymbol{F\,\theta} = \lambda\,\boldsymbol{\theta}}$$



Direction of axes are the *eigenvectors* of $\boldsymbol{F}$

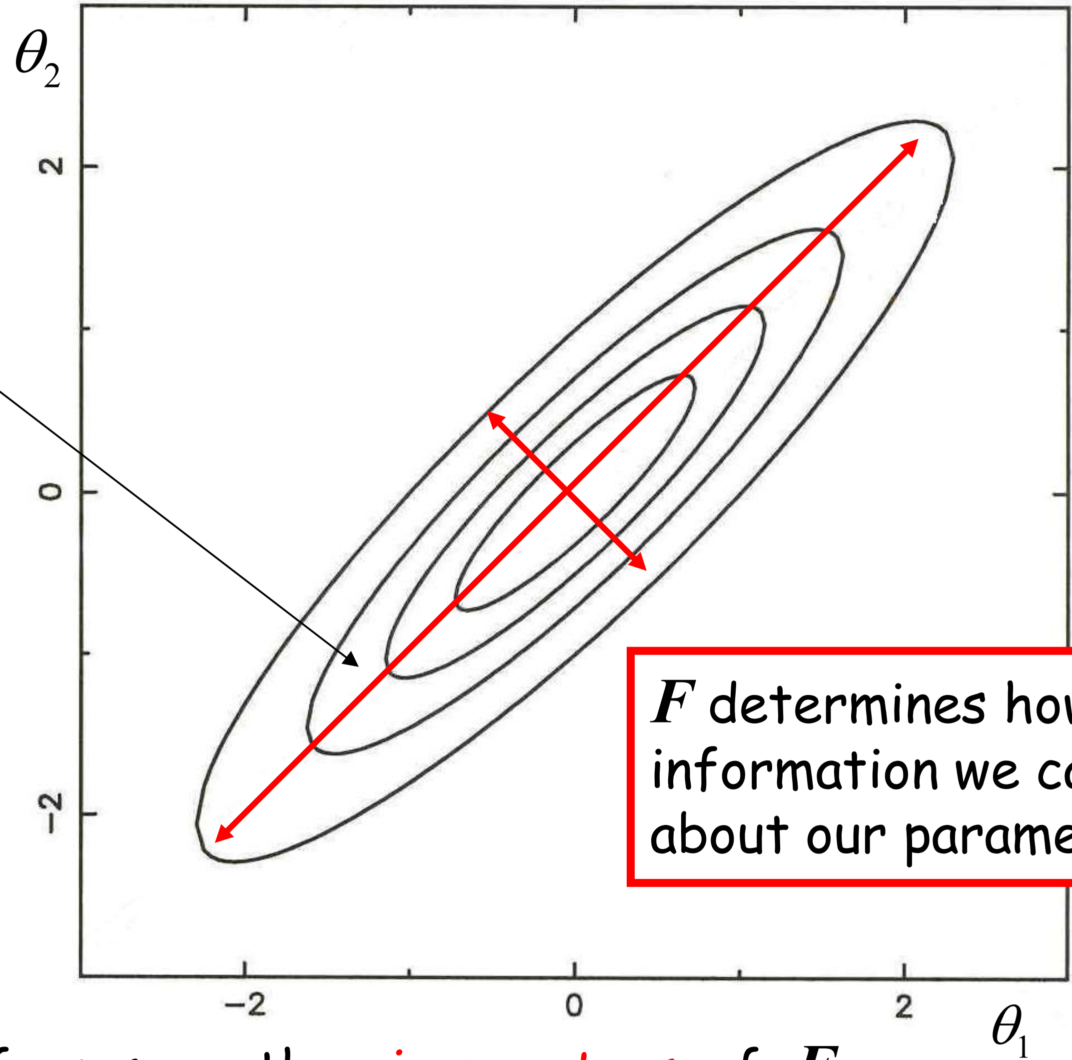
Parameter estimation: 2-D case

Linear combination of θ_1 and θ_2 well constrained by data

Length of axes determined by the **eigenvalues** of the Fisher information matrix

$$F_{ij} = \frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j} = \left[-\sigma_{ij}^2\right]^{-1}$$

$$\boxed{\boldsymbol{F}\,\boldsymbol{\theta} = \lambda\,\boldsymbol{\theta}}$$



F determines how much information we can learn about our parameters

Direction of axes are the *eigenvectors* of ***F***