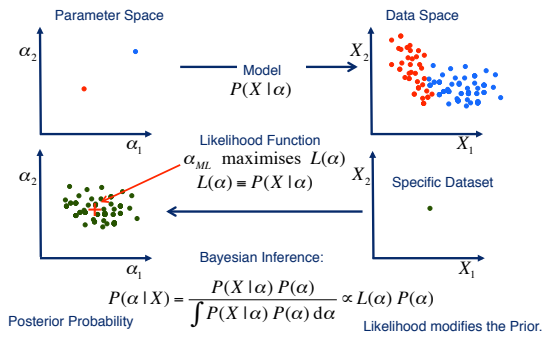
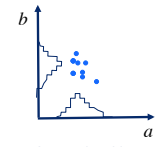


Review: Max Likelihood and Bayesian Inference



Monte-Carlo Error Propagation

1. Create **fake datasets**.
 - 1a. "Jiggle" the data points (using Gaussian random numbers).
 - * Requires good error bars.
 - 1b. (and/or) **"Bootstrap"** samples:
 - Pick N data points at random, with replacement (some points omitted, some repeated).
 - * Requires more data than parameters ($N > M$).
 - * Works with no error bars available.
2. Fit the model to each "fake" dataset. $\langle X_i \rangle = a_i + b$
3. Observe how the best-fit parameter values "dance".
4. Accumulate histograms approximating the parameter probability distributions.
5. Compute mean, median, variance, etc. of the parameters, or **any function of the parameters**.



Confidence interval on a single parameter (1-parameter, k-sigma confidence interval)

The **1- σ confidence interval** on α includes 68% of the area under the likelihood function:

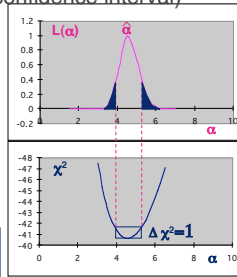
$$L(\alpha) \equiv P(X|\alpha) \propto \frac{e^{-\chi^2/2}}{\prod_i \sigma_i}$$

or posterior probability distribution, for non-uniform prior $P(\alpha)$:

$$P(\alpha|X) \propto L(\alpha)P(\alpha)$$

For a **k- σ (1-parameter) confidence interval**, use $\Delta\chi^2 = k^2$,

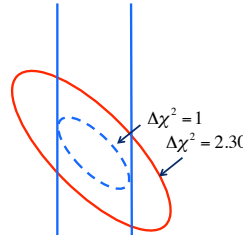
- $\Delta\chi^2 = 1$ for 1- σ , 68% probability
- $\Delta\chi^2 = 4$ for 2- σ , 95.4% probability
- $\Delta\chi^2 = 9$ for 3- σ , 99.73% probability ...



Generalise:

$$\chi^2 \Rightarrow -2 \ln(L(\alpha)P(\alpha))$$

2-parameter 1-sigma Confidence Region



If Y is a "nuisance parameter", use the **1-parameter 1-sigma confidence interval** in X , tangent to the $\Delta\chi^2 = 1$ contour in (X, Y) . The interval encloses 68% probability.

If both X and Y are of interest, use the **2-parameter 1-sigma confidence region**, the $\Delta\chi^2 = 2.30$ contour in (X, Y) . The contour encloses 68% probability.

Use $-2 \ln(L \times P)$ instead of χ^2 , if needed.

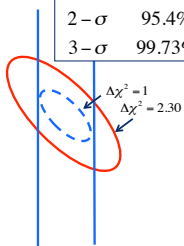
Note: Contour enclosing 68% probability must be **wider** than the 1-parameter confidence interval. Why?

$$L(\alpha) \equiv P(X|\alpha) \propto \frac{e^{-\chi^2/2}}{\prod_i \sigma_i}$$

M-parameter k-sigma Confidence Regions

$\Delta\chi^2$ thresholds for M-parameter k-sigma Confidence Regions

| Prob | M=1 | 2 | 3 | 4 | |
|-------------|--------|---|------|------|------|
| 1- σ | 68% | 1 | 2.30 | 3.53 | 4.72 |
| 2- σ | 95.4% | 4 | 6.17 | 8.02 | 9.70 |
| 3- σ | 99.73% | 9 | 11.8 | 14.2 | 16.3 |



The **M-parameter confidence region** is enclosed by the $\Delta\chi^2$ surface enclosing the desired probability. All **nuisance parameters must be re-fitted** (or integrated over) for each set of fixed values for the M parameters in the sub-space of interest. The $\Delta\chi^2$ in the M -parameter sub-space has a χ^2_M distribution.

Example: Estimate both μ and σ

$$L(\mu, \sigma) = P(X|\mu, \sigma) = \frac{e^{-\chi^2/2}}{(2\pi)^{N/2} \sigma^N}$$

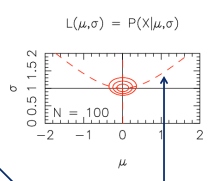
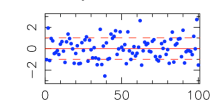
$$-2 \ln L = \sum_{i=1}^N \left(\frac{X_i - \mu}{\sigma} \right)^2 + 2N \ln \sigma + \text{const}$$

$$0 = \frac{\partial}{\partial \mu} [-2 \ln L] = -2 \sum_{i=1}^N \frac{X_i - \mu}{\sigma^2}$$

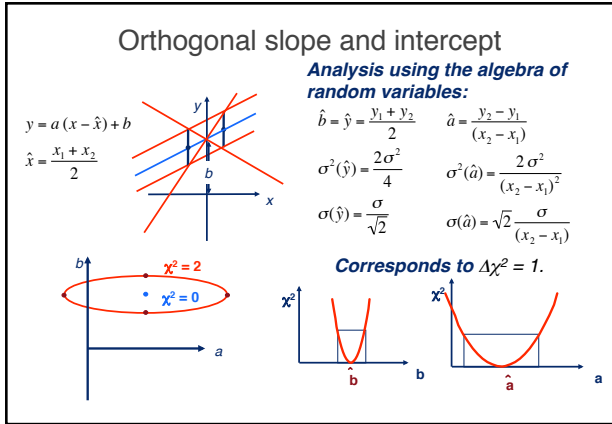
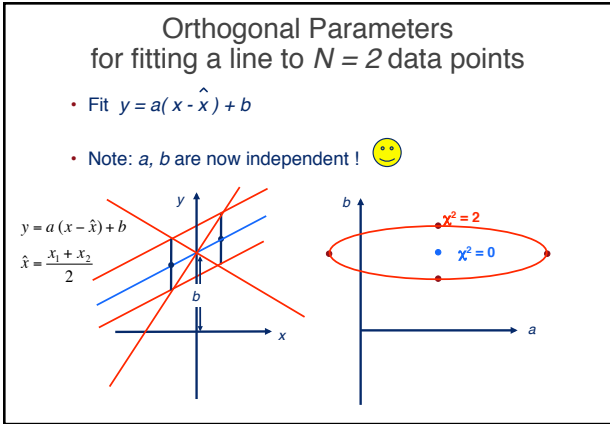
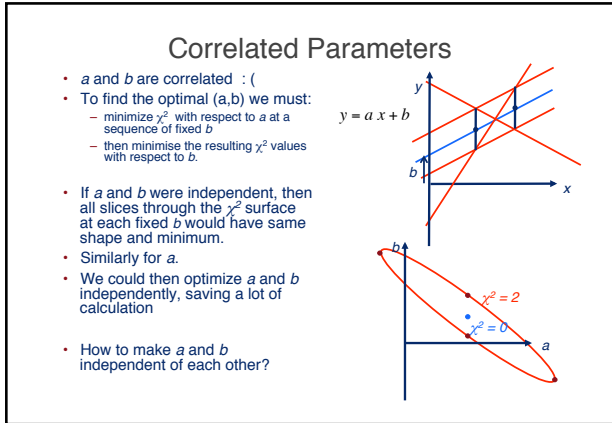
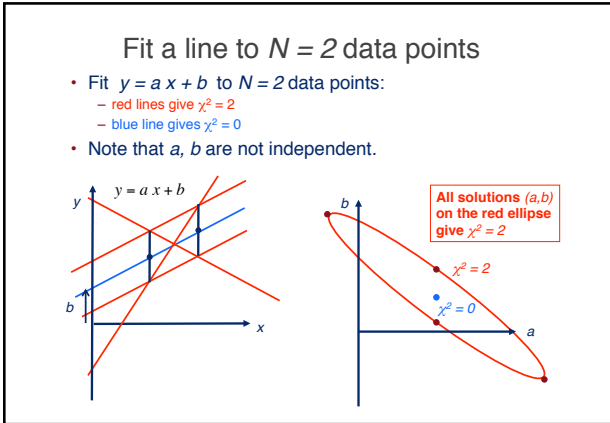
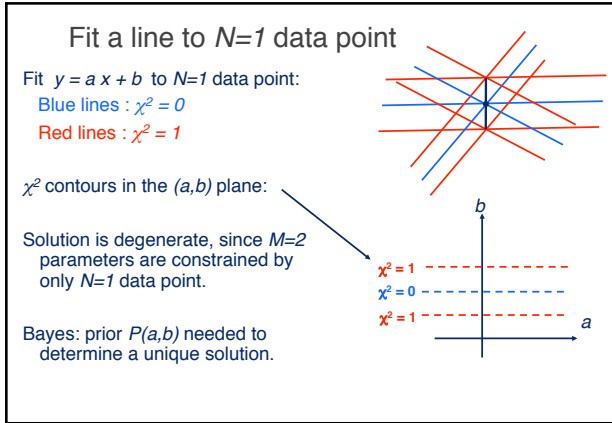
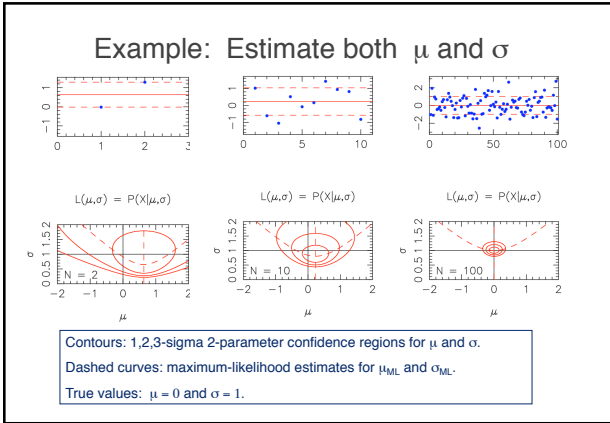
$$0 = \frac{\partial}{\partial \sigma} [-2 \ln L] = -2 \sum_{i=1}^N \frac{(X_i - \mu)^2}{\sigma^3} + \frac{2N}{\sigma}$$

$$\mu_{ML} = \frac{1}{N} \sum_i X_i \quad (\sigma_{ML})^2 = \frac{1}{N} \sum_i (X_i - \mu_{ML})^2$$

Posterior \propto Likelihood \times Prior
 $P(\mu, \sigma|X) \propto L(\mu, \sigma)P(\mu, \sigma)$

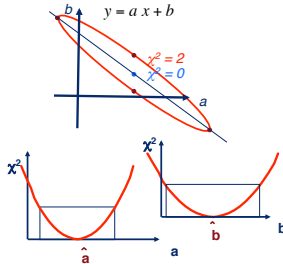


Note: ML gives biased estimate for σ .



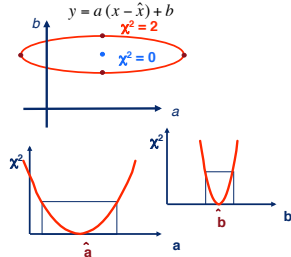
Orthogonal vs Correlated Parameters

Correlated Parameters:



For each a , a different b minimises χ^2 .
For each b , a different a minimises χ^2 .

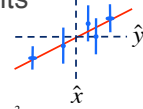
Orthogonal Parameters:



For any a , the same b minimises χ^2 .
For any b , the same a minimises χ^2 .

Fit a line to N data points

- If we use $y = ax + b$ then a, b are correlated.



- To make a, b orthogonal:

$$y = a(x - \hat{x}) + b \quad \hat{x} = \frac{\sum x_i / \sigma_i^2}{\sum 1 / \sigma_i^2}$$

- Intercept: **optimal average:**

$$\hat{b} = \hat{y} = \frac{\sum y_i / \sigma_i^2}{\sum 1 / \sigma_i^2}, \quad \text{Var}[\hat{b}] = \frac{1}{\sum 1 / \sigma_i^2}$$

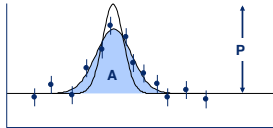
- Slope: **optimal scaling:**

$$\hat{a} = \frac{\sum (y_i - \hat{y})(x_i - \hat{x}) / \sigma_i^2}{\sum (x_i - \hat{x})^2 / \sigma_i^2}, \quad \text{Var}[\hat{a}] = \frac{1}{\sum (x_i - \hat{x})^2 / \sigma_i^2}$$

Choose Orthogonal Parameters

- Good practice (when possible).
- Results for any one parameter don't depend on values of other parameters.
- Example: fit a gaussian profile. 2 fit parameters:

- Width, w
- Area or peak value. Which is best?



Peak value depends on width – bad

$$f(x) = P e^{-\frac{1}{2} \left(\frac{x-x_0}{w} \right)^2}$$

$$g(x) = \frac{A}{w\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-x_0}{w} \right)^2}$$

Area is (more nearly) independent of width – good