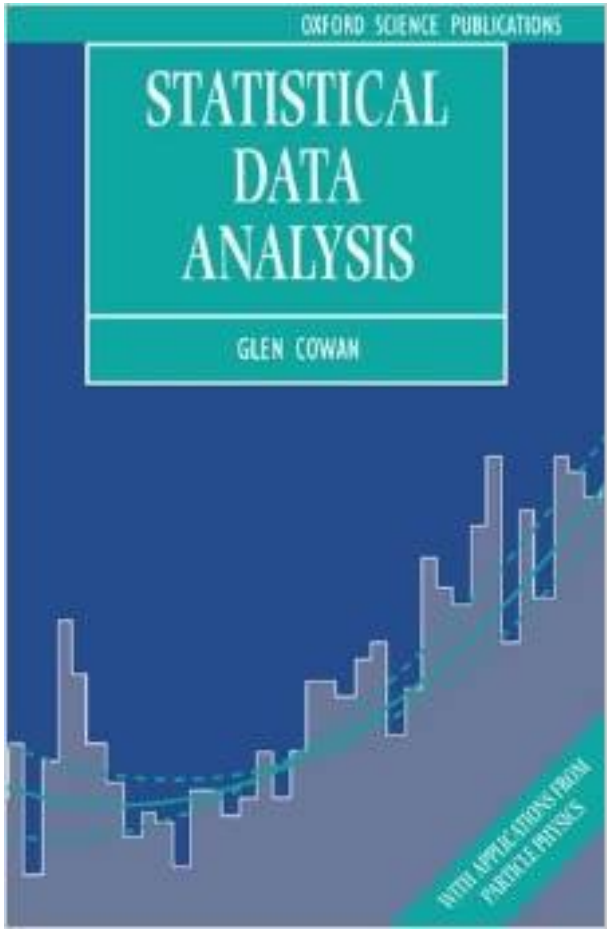
The Art of Data Fitting

Bibliography



Estimation

- Everyday life: It means a rough and imprecise procedure leading to a rough and imprecise result. You estimate when you cannot measure exactly

- Statistics: Technical term. It is a precise and accurate procedure, leading to a result which may be imprecise, but where at least the extent of the imprecision is known

Properties of Estimators

- An estimator is a procedure applied to the data sample which gives a numerical value for a property of the parent population or, as appropriate, a property or parameter of the parent distribution

- A "good" estimator is consistent, unbiased, and efficient

Consistency

- An estimator is consistent if it tends to the true value as the number of data values tend to infinity:

$$\lim_{N \to \infty} \hat{a} = a$$

- For a finite number N we cannot hope that for a particular sample data $\hat{a}$ will have the same value as true a. But we can require that chances of an overestimation balance those of an underestimate. Such estimator is *unbiased.*

- It is not usually difficult to find a consistent estimator, as the law of large numbers is on your side.

Bias

- An estimator is unbiased if its expectation value is equal to the true value

$$< \hat{a} >= a$$

Efficiency

- An estimator is efficient if its variance is small

Take-Home Message

- The choice of estimator to use in a particular application requires judgement
- There is no "ideal best estimator"
- Two reasons for this
  - Variance of the estimator depends on the distribution concerned
  - Detailed analysis may show that the most efficient estimator is biased. You have to weigh the relative merits of one estimator to another

Likelihood

- The data values x_i are drawn from some probability density function *P(x;a)* which depends on *a*

- The probability of a particular set of data $\{x_1,x_2,..,x_N\}$ is the product of the individual probabilities. This product is called likelihood:

$$L(x_1, x_2, ..., x_N; a) = \prod_{i=1}^{N} P(x_i; a)$$

Minimum Variance Bound (MVB)

- There is a limit to the accuracy of an estimator
- For an unbiased estimator

$$V(\hat{a}) \geq \frac{1}{< d \log L / da >}$$

Maximum Likelihood

- The *principle* of maximum likelihood is a method for estimation

- For a data sample $\{x_1,\ldots,x_N\}$ the maximum likelihood estimator $\hat{a}$ is the value of a for which likelihood is *a* maximum

$$L(x_1, x_2, \ldots, x_N; a) = \prod_{i=1}^{N} P(x_i; a)$$

- Usually instead of the likelihood you maximize the logarithm of the likelihood

Example: Lifetime

$$P(t) = \frac{1}{\tau} e^{-t/\tau}$$

$$\log L = -\sum_i \log \frac{e^{-t_i/\tau}}{\tau} = -\sum_i \left[\frac{t_i}{\tau} + \log \tau \right]$$

$$\left[\frac{d \log L}{d\tau} \right]_{\tau = \hat{\tau}} = -\sum_i \left(\frac{t_i}{\hat{\tau}^2} - \frac{1}{\hat{\tau}} \right) = 0$$

$$\Rightarrow \hat{\tau} = \frac{1}{N} \sum_i t_i$$

Notice that this estimation employs events!

Notes on Maximum Likelihood

- ML is a sensible way of producing an estimator
- It estimates the value that makes your data "most likely"
- For large samples, $\hat{a}$ has a probability distribution that is unbiased and normally distributed about the true value a, with variance equal to the MVB, so in the large *N* limit it is the best estimator. For smaller samples this not necessarily true
- The method provides errors: 1σ are those where the log likelihood falls by *1/2*
- For small *N*, ML estimators are usually biased
- Be careful not to apply large *N* formulae to small *N*.
  - <u>You have been warned</u>
- You need to know the parent distribution function. If your assumptions about *P(x;a)* are wrong...

χ^2

If random variables are independent and Normal distributed

$$P(\{a\}) = \prod_{i=1}^{N} p(a_i) \approx \exp\left[-\sum_{i=1}^{N} \frac{a_i^2}{2\sigma_i^2}\right] = \exp\left[-\frac{1}{2}\chi^2\right]$$

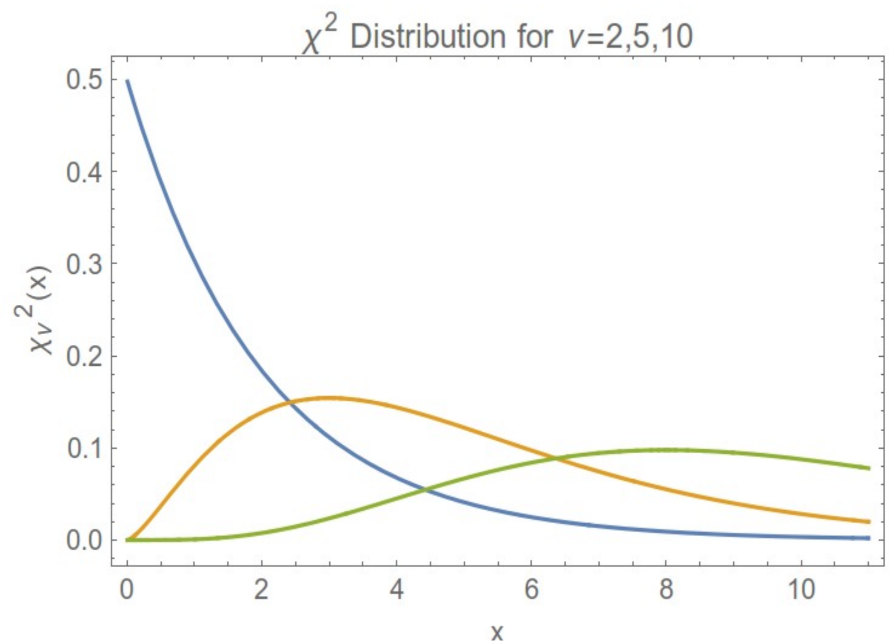


Fig. from M. Döring

χ^2

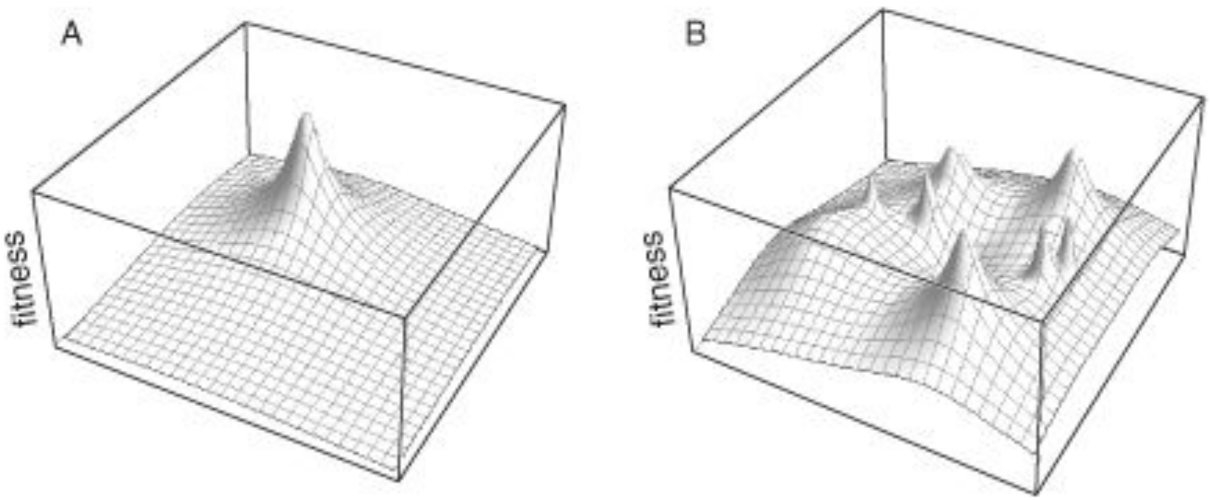
$$\chi^2/\mathrm{dof} = \frac{1}{N-p}\sum_{i=1}^{N}\left[\frac{f(\{x\};\{a\}) - \mathrm{Exp}_i(\{x\})}{\Delta\mathrm{Exp}_i(\{x\})}\right]^2$$

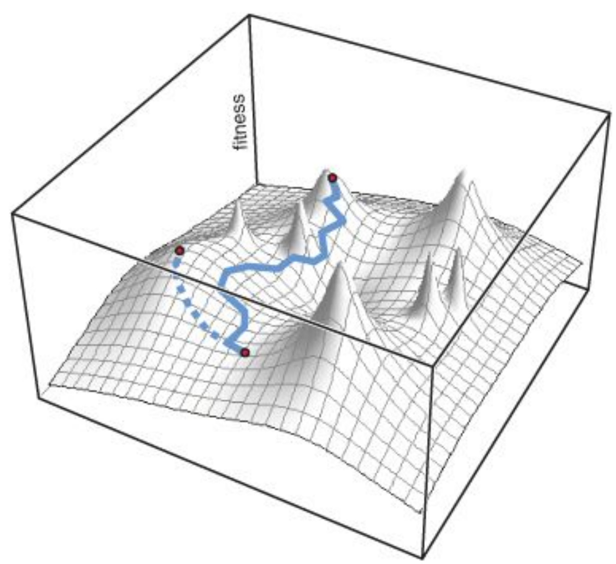
Expected value = 1

Optimization 101

- Calculus based methods
  - Direct (hill-climbing, gradient-based)
  - Indirect
- Enumerative
- Stochastic
  - Stochastic hillclimbing
  - Simulated annealing
  - Genetic algorithms

Optimization 101





How to avoid local minima

- Gradient-based
  - ☑ MC + hope for the best
- Stochastic hillclimbing
  - ☑ MC + hope for the best
- Simulated annealing
  - ☑ Guaranteed for $T \rightarrow \infty$
- Genetic algorithms
  - ☑ Guaranteed for $T \rightarrow \infty$

What is a good fit?

Depends on what you are doing…

And that is the best answer I have

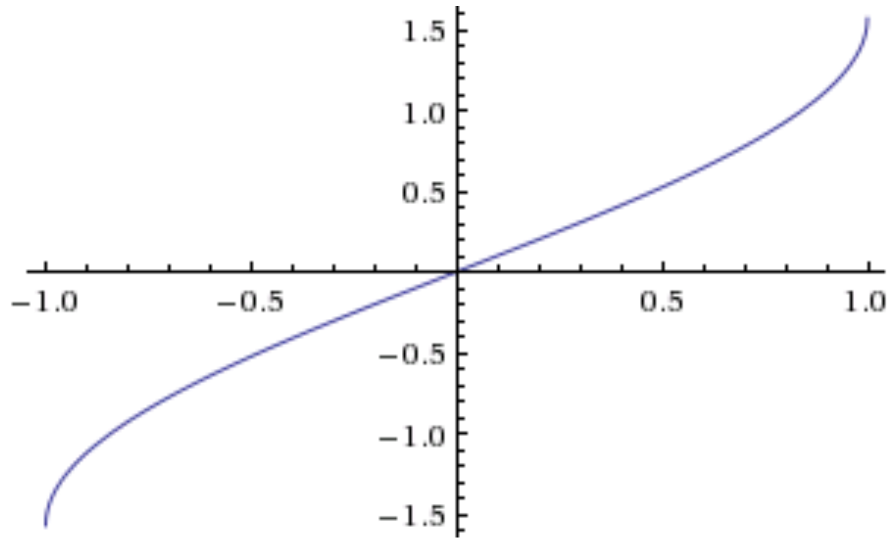
You need to use your best knowledge to separate the signal from the noise

Risky business: putting limits in the parameters

$$y \in [a, b] \quad , \quad x \in (-\infty, \infty)$$

$$x = \arcsin\left[\frac{2(y-a}{b-a} + 1\right]$$

$$y = a + \frac{b-a}{2}\left(1 + \sin x\right)$$



Beyond the χ^2: Run test

- Wald-Wolfowitz run test
- Tests the hypothesis that each element is independent of the next
- χ^2 tests distance, Run Test tests the distribution of the data

$$N = N_+ + N_-$$

$$\mu = \frac{2\,N_+\,N_-}{N} + 1$$

$$\sigma^2 = \frac{2N_+N_(2N_+N_- - N)}{N^2(N-1)} = \frac{(\mu-1)(\mu-2)}{N-1}$$

Another options to improve the analysis are:

- ☑ Kolmogorov-Smirnov test
- ☑ Wilcoxon signed rank test

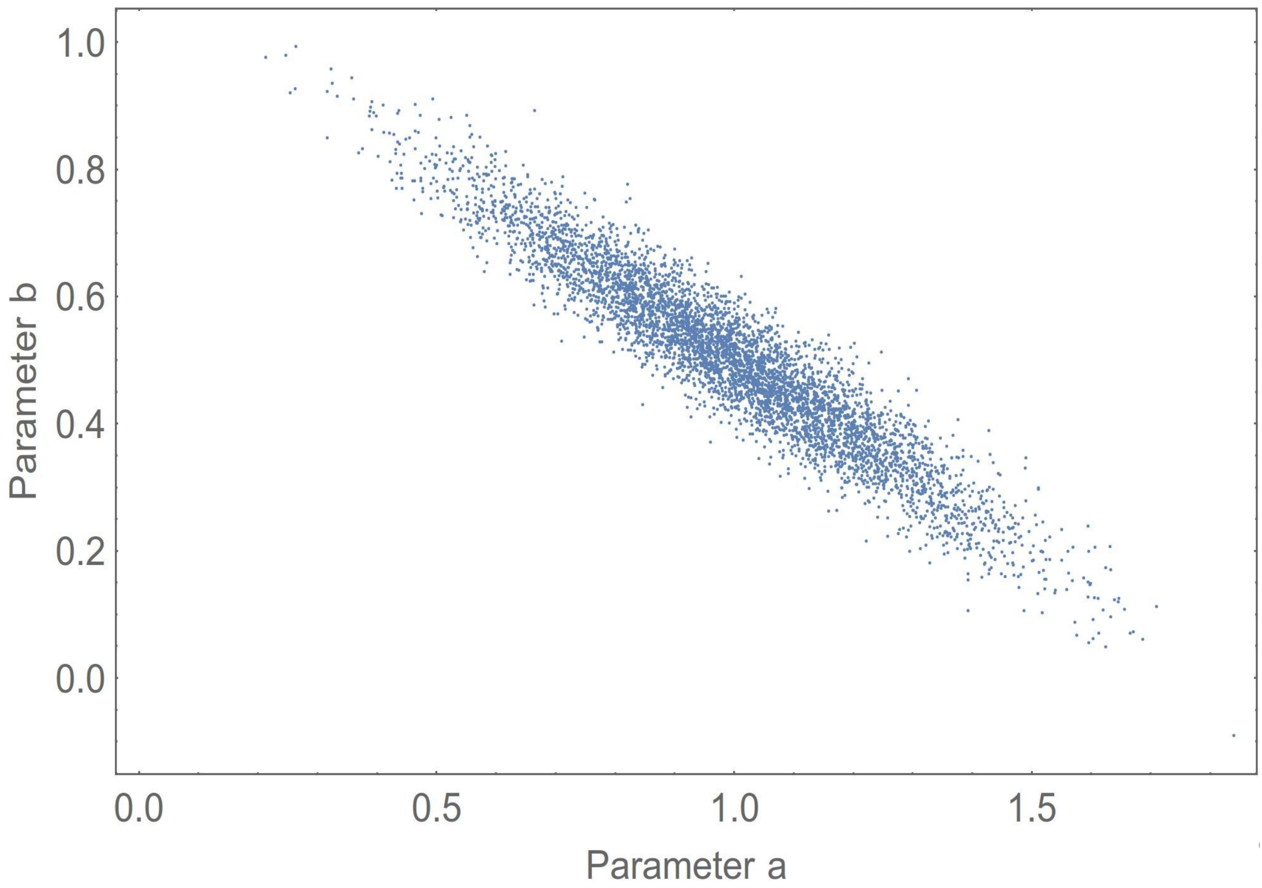
Confidence level and uncertainties

Once we have the parameters we need a method to assess uncertainty

Look at the whiteboard

Recommendation: Take a look at Cowan's book chapter 9

Error ellipse



Hessian method

$$\chi^2(\{a\}) - \chi^2_{min}(\{a\}) = \frac{1}{2}\sum_{i,j}\frac{\partial\chi^2(\{a\})}{\partial a_i}\frac{\partial\chi^2(\{a\})}{\partial a_j}\Delta a_i \Delta a_j + O(3)$$

$$= \frac{1}{2}(\Delta a)^T H(\{a\})(\Delta a) + O(3)$$

Compute eigendirection
Compute eigenvalues
Rotate
Compute $\chi^2_{min} + 1$

Bootstrap

(This slide intentionally left blank)

Bootstrap

The place where I have found the best explanation on bootstrap:

W. H. Press, S. A. Teukolsky, W.T. Vetterling, and B. P. Flannery, Numerical Recipes: The Art of Scientific Computing (Cambridge University Press, 1992)

and talking to Michael Döring from George Washington University

Take home points

- Be practical: get the best possible fit
  - Beware of local minima
- Estimator:
  - Maximum Likelihood for events
  - χ^2 for binned data
- χ^2 is not everything (but it certainly matters)
- Errors are not one-dimensional quantities
  - If (code=fast) bootstrap else hessian

Hands on

- Written in Python2
- You need to have installed: check your e-mail
- Download code from:
  - https://github.com/nobuosato/HUGS2015/tree/master/

- If you do not have software installed and codes downloaded **DO IT NOW** during the break