Maximum Likelihood

Maximum Likelihood in X-ray Crystallography

Kevin Cowtan
cowtan@ysbl.york.ac.uk

Maximum Likelihood

Inspired by Airlie McCoy's lectures.
http://www-structmed.cimr.cam.ac.uk/phaser/publications.html

Kevin Cowtan, cowtan@ysbl.york.ac.uk Sienna/Maximum Likelihood

Maximum Likelihood

- Science involves the creation of hypothesis (or theories), and the testing of those theories by comparing their predictions with experimental observations.
- In many cases, the conclusions of an experiment are obvious – the theory is supported or disproven.
- In other cases, results are much more marginal. e.g. How big a sample size do we need to distinguish a successful drug from placebo effect?

Kevin Cowtan, cowtan@ysbl.york.ac.uk Sienna/Maximum Likelihood

Maximum Likelihood

- In order to correctly understand the impact of our experiments on our theories, we need some knowledge of statistics.
- This is especially necessary in crystallography, since we have:
  - a very weak signal:
    (the observed magnitudes)
  - a great deal of noise:
    (the missing phases + measurement errors)
  - from which we are trying to test a very detailed hypothesis:
    (the position of every atom in the unit cell)

Kevin Cowtan, cowtan@ysbl.york.ac.uk Sienna/Maximum Likelihood

Maximum Likelihood

- Given the uncertainties of the data, we cannot usually determine whether a hypothesis is right or wrong – only how likely it is to be right:
  The probability of the hypothesis.
- In order to do this, our hypothesis must be detailed enough for us to work out how likely we would have been to get the results we observe, assuming that the hypotheses is true.
- We then use Bayes' theorem to determine the probability.

Kevin Cowtan, cowtan@ysbl.york.ac.uk Sienna/Maximum Likelihood

Maximum Likelihood

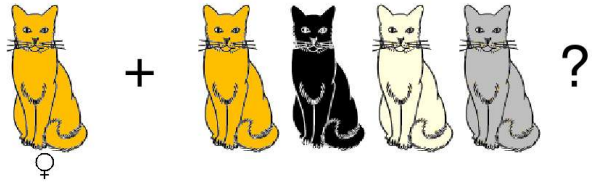
- Examples:
  - Hypothesis: The observed X-ray data arises from a molecule which is roughly homologous to this known molecule in this orientation in the cell.
    (Molecular replacement – how probable is a given model)
  - Hypothesis: The position of this atom (and its neighbors better explains the X-ray data when moved in this direction.
    (Refinement – what is the relative probability of two very similar models. Includes heavy-atom refinement.)
- In fact all problems come down to the comparison of different hypotheses.

Kevin Cowtan, cowtan@ysbl.york.ac.uk Sienna/Maximum Likelihood

Understanding Likelihood

Example:

- We have a cat. She has a kitten. We don't know who the father is, but there are four possibilities in the neighborhood.
- What can we say about the color of the father from the color of the kitten?

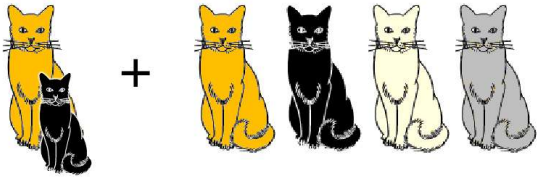


Understanding Likelihood

Cat genetics is complex: we will simplify.

Understanding Likelihood

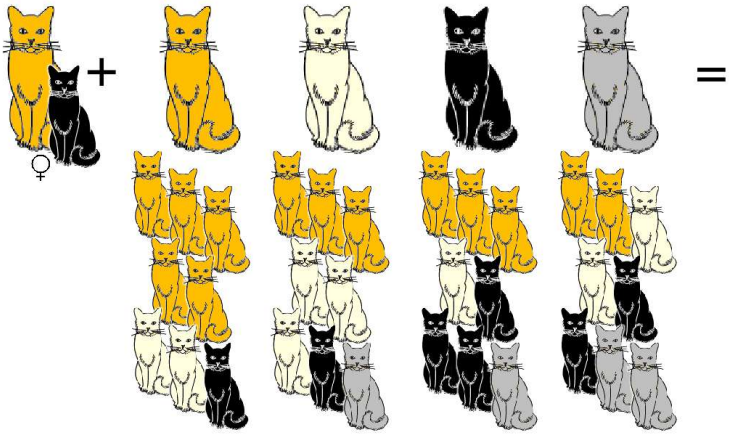
One black kitten: which is the father?



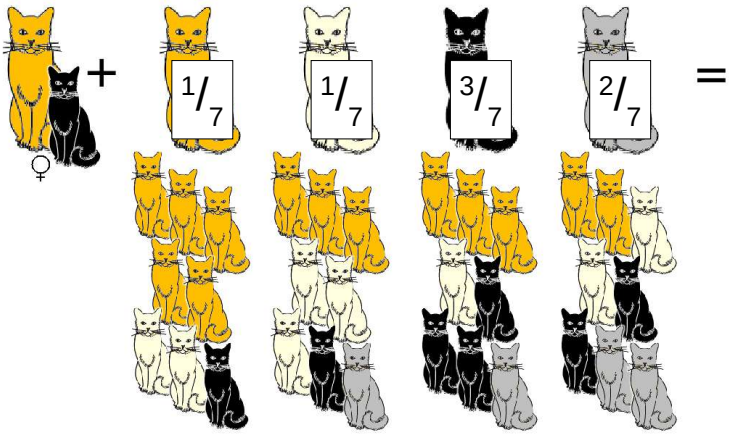
Actually, it could be any one, since they may all carry the appropriate genes. But they are not all equally probable.

We need some more information.

Understanding Likelihood



Understanding Likelihood



Understanding Likelihood

An extremely clever transformation has occurred:

- I gave you the probability of a kitten being a particular color, given that we know the colors of the father.

  P(*kitten color* | *father color*)

- You gave me the probability of the father being a particular color, given that we know the color of the kitten.

  P(*father color* | *kitten color*)

- **P(*x* | *y*): The probability of *x*, given *y*.**

Understanding Likelihood

This is a simple experiment:

- The result of the experiment is our observed data: the color of the kitten.
- The hypothesis is concerning the color of the cat. We make 4 hypotheses about the father (orange, black, cream, grey) and calculate the probability of each.
- We can work out P(data | hypothesis)
- We want to know P(hypothesis | data)

Understanding Likelihood

How do we do this sort of maths for a general problem?

- Use Bayes' theorem:

$$P(x|y)=\frac{P(y|x)P(x)}{P(y)}$$

- Proof: The probability of x *and* y is the probability of x *given* y multiplied by the probability of y.

$$P(x,y)=P(x|y)P(y)$$
$$P(y,x)=P(y|x)P(x)$$
$$P(x,y)=P(y,x)$$

Assumes x and y independent!

Understanding Likelihood

How do we do this sort of maths for a general problem?

- Use Bayes' theorem:

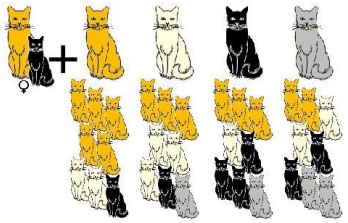
$$P(x|y)=\frac{P(y|x)P(x)}{P(y)}$$

- Therefore:

$$P(hypothesis|data)=\frac{P(data|hypothesis)P(hypothesis)}{P(data)}$$

Understanding Likelihood

$$P(x|y)=\frac{P(y|x)P(x)}{P(y)}$$



Therefore:

$$P(\text{father}_{color}|\text{kitten}_{color})=\frac{P(\text{kitten}_{color}|\text{father}_{color})P(\text{father}_{color})}{P(\text{kitten}_{color})}$$

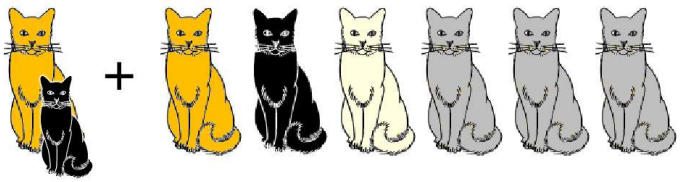
$$P(\text{father}_{cream}|\text{kitten}_{black}) = \frac{1/8\times1/4}{7/32} = \frac{1}{7}$$

$$P(\text{father}_{black}|\text{kitten}_{black}) = \frac{3/8\times1/4}{7/32} = \frac{3}{7}$$

$$P(\text{father}_{grey}|\text{kitten}_{black}) = \frac{2/8\times1/4}{7/32} = \frac{2}{7}$$

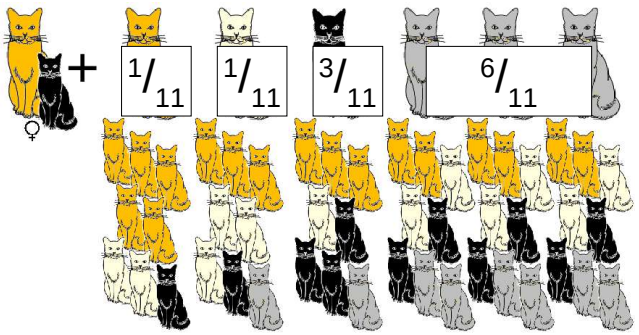
Understanding Likelihood

- What if the population of male cats is non-uniform?



i.e. $P(\text{father}_{color})$ is non-uniform

Understanding Likelihood



Understanding Likelihood

$$P(\text{father}_{\text{color}}|\text{kitten}_{\text{color}}) = \frac{P(\text{kitten}_{\text{color}}|\text{father}_{\text{color}})P(\text{father}_{\text{color}})}{P(\text{kitten}_{\text{color}})}$$

$$P(\text{father}_{\text{orange}}|\text{kitten}_{\text{black}}) = \frac{1/8\times 1/6}{11/40} = \frac{1}{11}$$

$$P(\text{father}_{\text{cream}}|\text{kitten}_{\text{black}}) = \frac{1/8\times 1/6}{11/40} = \frac{1}{11}$$

$$P(\text{father}_{\text{black}}|\text{kitten}_{\text{black}}) = \frac{3/8\times 1/6}{11/40} = \frac{3}{11}$$

$$P(\text{father}_{\text{grey}}|\text{kitten}_{\text{black}}) = \frac{2/8\times 3/6}{11/40} = \frac{6}{11}$$

Understanding Likelihood

- The elements of Bayes' theorem have names:

Likelihood: The probability of getting the observed data, given the current hypothesis.

Prior: The probability of the current hypothesis, before we even do an experiment.

$$P(hypothesis|data) = \frac{P(data|hypothesis)P(hypothesis)}{P(data)}$$

Posterior: Our final measure of confidence in the hypothesis having tested it.

Normalization: So that all the possibilities add up to 1. Independent of hypotheses. Usually ignore.

Understanding Likelihood

- As applied to the cats:

Likelihood: If we suppose that a given cat is the father, we can work out the probable kitten colors.

Prior: We know the distribution of male cats before the kitten is born.

$$P(\text{father}_{\text{color}}|\text{kitten}_{\text{color}}) = \frac{P(\text{kitten}_{\text{color}}|\text{father}_{\text{color}})P(\text{father}_{\text{color}})}{P(\text{kitten}_{\text{color}})}$$

Posterior: We can determine the probably father color from the results.

Normalization: Since we only want to compare relative probabilities, this doesn't matter.

Likelihood and Crystallography

- How do we apply this to crystallography?
  - Hypothesis: The observed X-ray data arises from a molecule which is roughly homologous to this known molecule in this orientation in the cell.
    (Molecular replacement – how probable is a given model)
  - Hypothesis: The position of this atom (and its neighbors better explains the X-ray data when moved in this direction.
    (Refinement – what is the relative probability of two very similar models. Includes heavy-atom refinement.)
- Each hypothesis leads to a set of predicted structure factors: $\underline{F}_c(\mathbf{h})$. How well these explain the observed $|\underline{F}_{obs}(\mathbf{h})|$ determines the likelihood.

Likelihood and Crystallography

- For most purposes, we treat each reflection as an independent observation. Therefore we can consult each reflection separately to determine how well it agrees with the model. Then, we multiply all the resulting likelihoods together.
- Problem: the product of 10,000s of small numbers gives an underflow on a computer.
- Solution: Take the log of all the likelihoods and sum them.

$$\sum_i \log(x_i) = \log\left(\prod_i x_i\right)$$

Usually minimize -log(likelihood) because it is +ve.

Likelihood and Crystallography

- Each hypothesis leads to a set of predicted structure factors: $\underline{F}_c(\mathbf{h})$. How well these explain the observed $|\underline{F}_{obs}(\mathbf{h})|$ determines the likelihood.
- But: we have a continuum of hypotheses. We can rotate an MR model or move a refinement atom continuously to improve the model.
- We refine the parameters of the model (e.g. rotation of MR model, position of refinement atoms) in order to best explain the observed data, i.e. to give the highest value of the likelihood, hence:

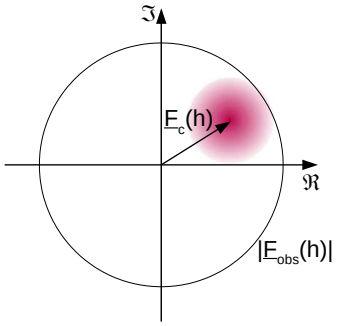
Maximum Likelihood

Likelihood and Crystallography

- But actually we want to maximize the posterior. e.g. in refinement:
  - Prior gives the probability of the model on the basis of its agreement with stereo-chemical restraints.
  - Likelihood gives the probability of the model on the basis of the observed X-ray data.
- If we just maximize the likelihood, we get lousy geometry.
- But people call it 'maximum likelihood' anyway.

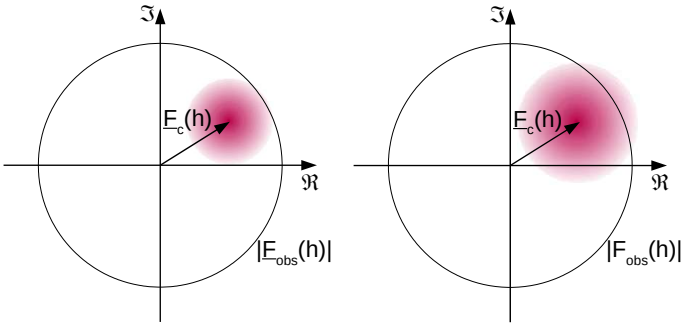
Likelihood and Crystallography

- Each hypothesis leads to a set of predicted structure factors: $\underline{F}_c(\mathbf{h})$. How well these explain the observed $|\underline{F}_{obs}(\mathbf{h})|$ determines the likelihood.
- Note: To calculate a probability we must also estimate the error associated with the $\underline{F}_c(\mathbf{h})$.
- The error estimation is a vital part of the model or hypothesis.



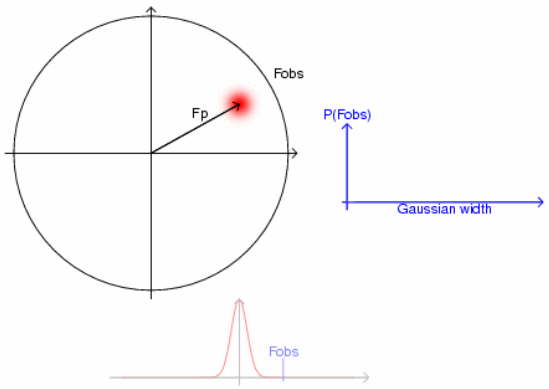
Likelihood and Crystallography

- How do we estimate the errors? Surely as the error estimate increases, the model always becomes a better description of the data?



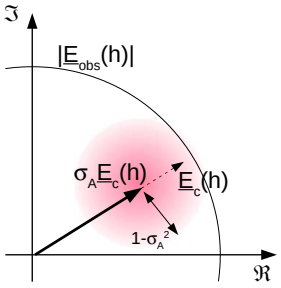
Likelihood and Crystallography

- No, the likelihood favors appropriate noise levels:



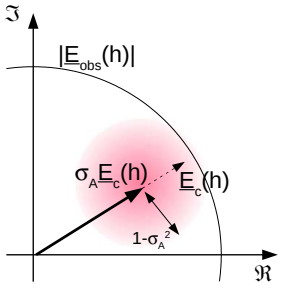
Likelihood and Crystallography

- Error estimation is in terms of a parameter σ_A, where σ_A is the fraction of the normalised structure factor $E_C(h)$ which is correct, and $(1-\sigma_A^2)$ is the variance of the noise signal.
- Typically estimated as a function of resolution.
- Read (1986) Acta Cryst A42, 140-149



Likelihood and Crystallography

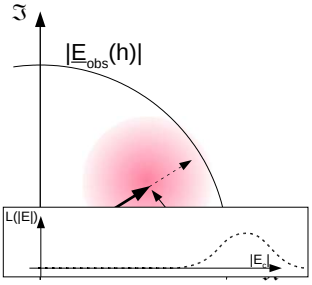
- We can calculate the Likelihood Function for $\underline{E}_{obs}$ given $\underline{E}_c$:



$$P(\underline{E}_{obs}|\underline{E}_c) \propto \exp\left(\frac{|\underline{E}_{obs} - \sigma_a \underline{E}_c|}{\epsilon(1-\sigma_a^2)}\right)$$

Likelihood and Crystallography

- But we don't know $\underline{E}_{obs}$!
  - The data are the observed magnitudes: $|\underline{E}_{obs}|$
- We want P(data|model)
- Therefore, sum (integrate) the likelihood over all the unknown phases: Rice fn
  (i.e. eliminate nuisance variable)



$$P(|\underline{E}_{obs}||\underline{E}_c) \propto \exp\left(\frac{|\underline{E}^2_{obs}|+\sigma_a^2|\underline{E}^2_c|}{\epsilon(1-\sigma_a^2)}\right) I_0\left(\frac{2|\underline{E}_{obs}|\sigma_a|\underline{E}_c|}{\epsilon(1-\sigma_a^2)}\right)$$

Likelihood and Crystallography

Steps:

- Construct a model, with some parameters: e.g.
  - MR: Rotation R, Translation T, Error term σ_A
  - Refinement: Coords x_i, Temp factors B_i, Error term σ_A
- Refine parameters R,T / x_i,B_i , σ_A to maximize the likelihood using the known **magnitudes**.
- Then use the resulting probability function for the **phases** to calculate an electron density map.
  - Programs will output ML map coefficients.

Likelihood and crystallography

Other details: Molecular replacement

- Programs will also use a likelihood function for un-positioned models to rank rotation function results.
- More complex likelihood functions allow combination of information from multiple fragments, even when relative position is unknown.
- See for example
  Read (2001), Acta Cryst. D57, 1373-1382.

Likelihood and crystallography

Other details: Refinement

- Programs may also perform anisotropy correction, TLS refinement, bulk solvent correction. ML parameter refinement may be used to refine all of these parameters.
- Heavy atom refinement is similar, but is applied against multiple data sets simultaneously.
- See for example
  Pannu, Murshudov, Dodson, Read, (1998) Acta Cryst. D54, 1285-1294.

Likelihood and crystallography

Summary:

- Likelihood provides a tool for establishing the probability of a hypothesis.
- When data is weak, this is vital for describing our current state of knowledge.
- Direct benefits include improved models and weighted maps.
- Employed in:
  - phasing, MR, refinement, phase improvement, map interpretation.