Computational Cognitive Science

Lecture 5: Maximum Likelihood Estimation; Parameter Uncertainty

Chris Lucas
(Slides adapted from Frank Keller's)

School of Informatics
University of Edinburgh
clucas2@inf.ed.ac.uk

3 October, 2017

Finding the Maximum Likelihood

- Optimizing log Likelihood
- Fitting Individuals or Groups
- Properties of Maximum Likelihood Estimation

Error on Maximum Likelihood Estimates

- Sources of Variability in Modeling
- Standard Errors

Reading: Chapters 4 and 5 of L&F.

Maximum Likelihood Estimate

Idea behind maximum likelihood estimation: determine parameter values such that they maximize the likelihood of the data.

The *maximum likelihood estimate* (MLE) $\hat{\theta}$ of a parameter θ is:

$$\hat{\theta} = \arg\max_{\theta} L(\theta|y) = \arg\max_{\theta} P(y|\theta)$$

We discussed this in the previous lecture; now we are interested in how the MLE can be obtained.

Finding the Maximum Likelihood

In the general case, we will have not just one data point, but many, and a vector of parameters. We compute the *joint likelihood* of the data points (assumed to be independent):

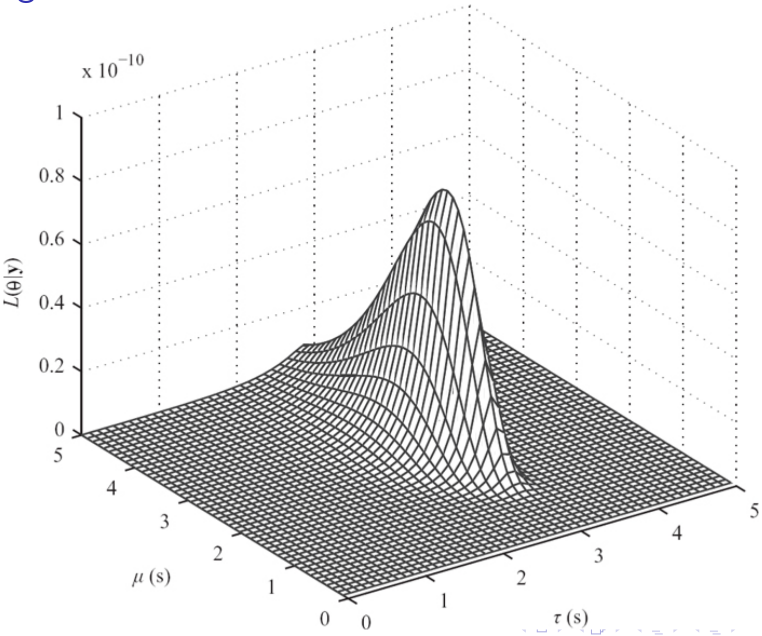
$$L(\theta|y) = \prod^{k} L(\theta|y_k)$$

Example: ex-Gaussian model: assume the data vector $y = [3\ 4\ 4\ 4\ 4\ 5\ 5\ 6\ 6\ 7\ 8\ 9]$ and plot:

$$L(\theta|y) = L(\mu, \tau, \sigma = 0.1|y)$$

We want to find the *mode* of the likelihood function: it gives us the *maximum likelihood estimate* of the parameters.

Finding the Maximum Likelihood



Optimizing log Likelihood

To find the mode of $L(\theta|y)$, we can use the optimization methods discussed in lecture 3 (Nelder-Mead, simulated annealing), among others. Note however that:

- we are now maximizing rather than minimizing;
- sometimes there is an analytic solution (i.e., we can differentiate $L(\theta|y)$ and get the maximum that way);
- typically we maximize the *logarithm of the likelihood;* this makes the maths easier and avoids numeric underflow.

$$\ln L(\theta|y) = \sum_{k=1}^{K} \ln L(\theta|y_k)$$

Fitting Individuals or Groups

We can use maximum likelihood estimation to fit a separate set of parameters for each participant. Then we can:

- plot histograms of parameter values;
- report mean parameter values and estimate their variability;
- sum the log likelihoods of the models for all participants to get a measure of fit for the whole data set;
- compare the optimal parameter values (and log likelihoods) for subgroups of participants.

Fitting Individuals or Groups

If we want to fit a single set of parameters for the whole data set:

- we can aggregate the data and fit averages (e.g., mean proportion correct, rather than individual proportion correct);
- this changes the interpretation of our model (it is now modeling average data);
- we may also need a new data model (e.g., Gaussian distribution instead of binomial distribution);
- note that for the Gaussian distribution, minimizing RMSD is the equivalent to finding the MLE.

Properties of Maximum Likelihood Estimation

Maximum likelihood estimation has attractive theoretical properties (assuming the likelihood function is continuous and the parameters are identifiable):

- *parametrization invariance:* finding the MLE of $g(\theta)$ is equivalent to finding the MLE of θ and then applying g. May not be useful if $g(\theta)$ is not invertible.
- *consistency:* as the sample size increases, the estimated parameter values converge to the true parameter values. Assumes that mode is not misspecified.
- *efficiency:* maximum likelihood estimation delivers the least variable parameters estimates (asymptotically).
- *asymptotic normality:* MLEs are asymptotically normally distributed.

Often we need to quantify the *uncertainty* of parameter estimates. Sources of variability and uncertainty:

- Uncertainty and variability in our model, including *nuisance parameters*: Additional parameters that we can't easily estimate and may not care about, but still have to represent, e.g., decay of individual words in the Baddeley model.
- Variability in our data: Individuals differ, as do responses by a single individual.

If we compute MLEs for two different data sets, we are likely to obtain different results.

The model itself can also be uncertain. *Model comparison* is a way of determining the best model for a given data set – stay tuned.

How can we know how good our estimates are, and what parameter values are plausible given our data?

- Bayesian inference is a principled approach to this problem, but it requires us to specify priors – more on this later.
- Alternately, compute standard errors for parameter estimates.

Standard Errors

If we fit a separate model for each participant, then we typically report the mean across participants for our parameter values.

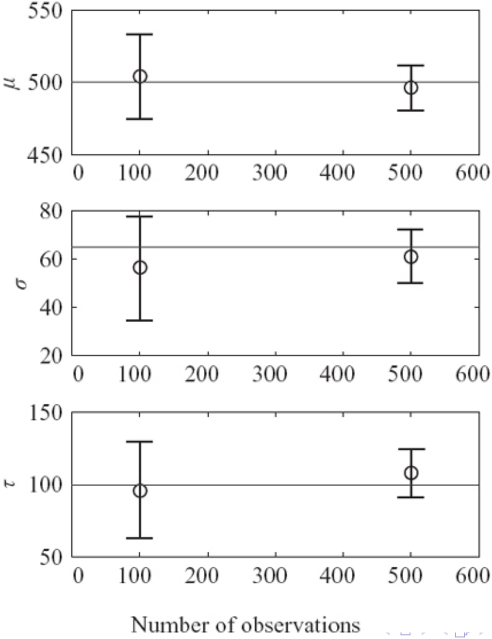
We can then compute the *standard error* $SE_{\bar{x}}$ around the mean $\bar{x}$ to quantify the uncertainty of the parameter estimates:

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

where s is the standard deviation of the variable, and n is the sample size (number of participants).

Standard statistical tests (t-test, Anova) can then be applied based on the standard errors.

Standard Errors



Summary

- Maximum likelihood estimation finds parameter values that maximize the likelihood function, given the data;
- It can be useful to visualize data likelihood as a function of parameters (where possible)
- MLE has attractive theoretical properties: parametrization invariance; consistency; efficiency; asymptotic normality (but these shouldn't be over-interpreted);
- if we fit a separate model for each participant, then we can compute standard errors for our parameter estimates;

References

Lewandowsky, S. & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. Thousand Oaks, CA: Sage.