

# Likelihood - Intro

Nathaniel Beck

Department of Politics, NYU, New York, NY 10012, [nathaniel.beck@nyu.edu](mailto:nathaniel.beck@nyu.edu)

QII - Week 1

1

## Intro

- Let  $\mathbf{y}$  be observations on  $y_i$  for individuals  $i = 1, \dots, N$ . ( $\mathbf{y}$  is stochastic.) Let  $\Theta$  be a vector of parameters of interest.
- Note: this is a very general setup than we had previously. Here the  $\mathbf{X}$ 's are just "stuff" and only the stochastic dependent variable is modelled. This will be clearer when we do OLS below.
- Note that if we have a simeq setup (or other setup with several stochastic variables,  $y_i$  is just a vector, so not innocuous as to whether something in  $x$  or  $y$ . For now we will work with a single  $y$ , what in regression we would call a single dependent variable. Likelihood theory will also give us a good way to see when we can treat  $x$  as stuff and when we need to model it (that is, when it is part of  $y$ ).
- By Bayes Theorem, the conditional density of  $\Theta|\mathbf{y}$ ,

$$\Lambda(\Theta|\mathbf{y}) = \frac{f(\mathbf{y}|\Theta)g(\Theta)}{h(\mathbf{y})} \quad (1)$$

2

- Note that the denominator is just a function of the data. It only makes sense to compare these conditional densities for the same data, so we can ignore the denominator.
- The second term in the numerator is the “prior” density of  $\Theta$  which is fixed before our observations and hence is invariant over our problem.
- WE WILL ALWAYS TAKE THE DATA AS GIVEN AND THINK OF ESTIMATION GIVEN SOME DATA (SAMPLE)
- If the data is given, we can write

$$\Lambda(\Theta|\mathbf{y}) \propto f(\mathbf{y}|\Theta)g(\Theta) \quad (2)$$

- $f$  is called the “likelihood”, that is,

$$L(\Theta|\mathbf{y}) = f(\mathbf{y}|\Theta) \quad (3)$$

- In words, the likelihood is the sample information that transforms a “prior” into a “posterior” density for  $\Theta$ .
- We then think of  $L$  as a function of  $\Theta$ , ( $\mathbf{y}$  is already observed, so everything is conditioned on that observation). The best estimator,  $\hat{\Theta}$  is then whatever value of  $\Theta$  maximizes

3

## Relative odds or Bayes factor interpretation

- We can think of the posterior density of the parameter ( $\Lambda$ ) as giving the relative odds of different parameter values (remembering that this is a continuous density of a parameter, not a set of observations, and so some would deny that one can do this, but we can ignore the theological arguments).
- Thus if we have two candidate values for the parameter,  $\theta_1$  and  $\theta_2$  (scalar), the “odds” on a draw of  $\theta_1$  vs  $\theta_2$  are  $\frac{\Lambda(\theta_1)}{\Lambda(\theta_2)}$ .
- Note ML chooses the value of  $\theta$  which has the highest odds against any other parameter.
- Given a set of data, we can write

$$\frac{\Lambda(\theta_1)}{\Lambda(\theta_2)} = \frac{f(\theta_1) g(\theta_1)}{f(\theta_2) g(\theta_2)} \quad (5)$$

- We can interpret the second fraction as the prior odds ratio (our guess about the relative odds of the two  $\theta$ 's before observing the data) and the ratio of the likelihoods tells us how those prior odds are transformed into posterior odds (having observed the data).

4

- Note that if *a priori* we have no particular reason to believe that  $\theta_1$  is more likely than  $\theta_2$  or vice versa, the likelihood ratio then gives us the posterior odds of  $\theta_1$  vs  $\theta_2$ .
- Either way, the likelihood ratio is called (by Bayesians) the “Bayes factor” and tells us how much the data favor one parameter value over another.
- Maximum likelihood then tells us to estimate  $\theta$  by choosing the value with the highest Bayes factor.
- Unfortunately to continue would get us involved in theology, so I will note that lots of seemingly sane people think that the Bayesian interpretation is the work of the devil, while others find it divinely inspired.

5

## Likelihood for Independent Observations

- IF the  $y_i$  are all independent (or conditionally independent, given  $x_i$ )

$$L = f(\mathbf{y}) = \prod_{i=1}^N f(y_i|\Theta) \quad (6)$$

or

$$\log L = \sum_{i=1}^N \log f(y_i|\Theta) \quad (7)$$

- Since likelihoods are all positive and the log function is monotone, the likelihood and its log have their maxima at the same place (simple theorem from calculus). It is almost always much simpler to work with the log of the likelihood since get to sum things up
- For this course, all logs are natural, NOT base 10. Sometimes you will see  $\ln$  used,  $\ln$  and  $\log$  are interchangeable.

6