The likelihood principle and parameter estimation

Day 2

Dominique Gravel

August 15th, 2017

UNIVERSITÉ DE SHERBROOKE

Content

- The likelihood principle
- Optimization methods
- Finding likelihood for a more complex problem

A standard statistical model

$$y_i = a + bx_i + \varepsilon_i$$

with

$$\varepsilon \sim N(0, \sigma)$$

and where α, β and σ are fitted parameters.

Classical “Frequentist” statis««««tics

What is the probability of observing a particular value of a predefined test statistic (e.g. the slope is zero), given an assumed hypothesis about the underlying scientific model (e.g. linear model), and assumptions about the probability model of the test statistic (e.g. parameters are estimated from a normal distribution).

Implicit assumption

The data are an approximate “sample” of an underlying “true” reality (the H_1 hypothesis).

Another way of formulating a statistical model

$$y \sim N(\mu, \sigma)$$

where

$$\mu = \alpha + \beta x$$

Which means that the probability of an observation x_i is distributed as $P(y_i|x_i, \alpha, \beta, \sigma)$

The “likelihood principle”

$$L(\theta|x) \propto P(x|\theta)$$

In words: The likelihood (L) of the set of parameters (q) (in the scientific model), given an observation (x), is proportional to the probability of observing the data given the parameters...

The likelihood approach flip the interpretation of a statistical model : the data is the reality, and we look at the “likelihood” of an explanation. Among several things, this formulation facilitates the comparison of candidate models (and sets of parameters).

Likelihood for a set of observations

For i = 1...n independent observations, and a vector of **X** observations (x_i):

$$L(\theta|\mathbf{X}) \propto \prod g(x_i|\theta)$$

A product of probabilities is a very small number, so often we prefer to use logarithms, such that the log-likelihood is :

$$\log[L(\theta|\mathbf{X})] \propto \sum \log[g(x_i|\theta)]$$

Estimating the mean of a set of values

Let consider the following set of numbers $X = \{4, 2, 10, 5, 8, 4\}$. It's very straightforward to compute the mean, but how to do it by maximum likelihood ?

A very basic example

Estimating the mean of a set of values

Let's try first by minimizing the sum of squares as we do with traditional linear models :

```
ss <- function(X, u) {
  res <- (X-u)*(X-u)
    return(sum(res))
}
X <- c(4, 2, 10, 5, 8, 4)
ss(X, 5)
```

```
## [1] 45
```

```
ss(X, 5.5)
```

```
## [1] 43.5
```

Now let's do it by maximizing the likelihood, assuming a normal distribution of residuals :

```
ll <- function(X, u, sd) {
    sum(dnorm(X, u, sd, log = TRUE))
}
ll(X, 5, 1)
```

```
## [1] -28.01363
```

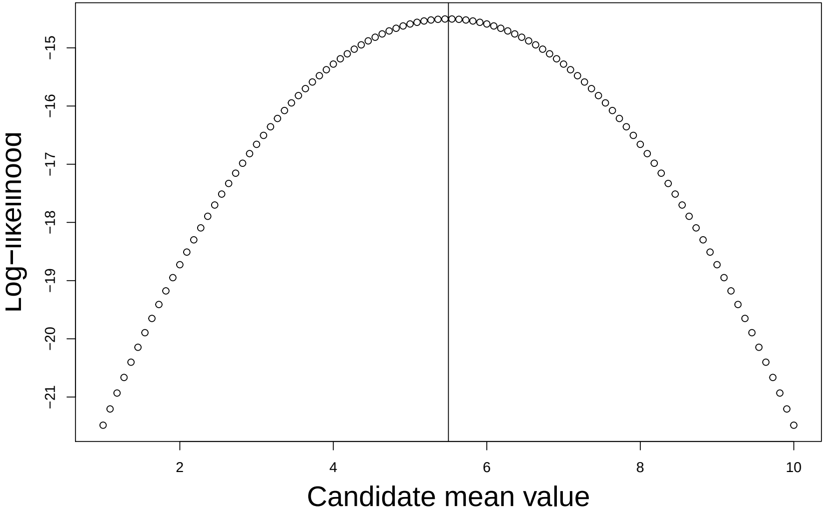
```
ll(X, 5.5, 1)
```

```
## [1] -27.26363
```

```
ll(X, 5.5, 2.95)
```

```
## [1] -14.50374
```

A very basic example



A different approach to statistics and model fitting

We believe that "truth" (full reality) in the biological sciences has essentially infinite dimension, and hence . . . cannot be revealed with only . . . finite data and a "model" of those data. . .

. . . We can only hope to identify a model that provides a good approximation to the data available.

Burnham and Anderson 2002, p. 20

Time to sharpen your pen. . .

I want to estimate the probability that the bee *Bembix oculata* visit the flower *Mentha pulegium* over a sampling time of 5 minutes. I do 10 transects, in which I record for each of them if an individual visits a flower and obtain the following sequence of events $X = \{0, 1, 1, 0, 1, 1, 0, 0, 0, 0\}$ representing if an interaction occurred ($X_i = 1$). What is the maximum likelihood estimate for this probability event ?

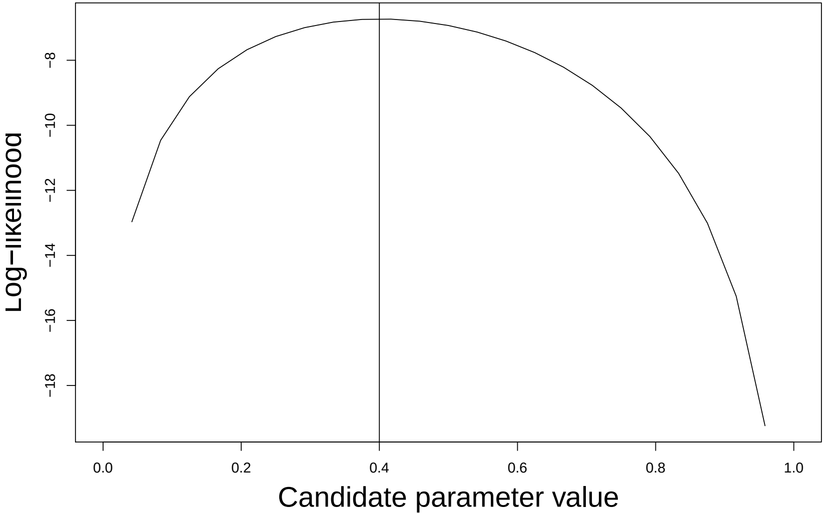
Solution

What we aim to describe is a binomial process with 10 trials. We seek to estimate the interaction probability p.

If we observe an interaction, the likelihood of that event is p, while if we do not observe the interaction the likelihood is $1 - p$.

As a result, we get the following likelihood for each observation $L = \{1 - p, p, p, 1 - p, p, p, 1 - p, 1 - p, 1 - p, 1 - p\}$.

Solution


Log−likelihood
−8
−10
−12
−14
−16
−18
0.0
0.2
0.4
0.6
0.8
1.0
Candidate parameter value

Another exercise

Consider the set of values X = $\{5.3, 3.9, 4.2, 6.8, 4.3, 0.1, 1.9, 2.5, 6.2, 4.4\}$ and Y = $\{5.3, 1.9, 4.8, 7.0, 5.6, 2.3, 0.5, 2.9, 8.2, 5.4\}$.

- Look at the data
- Propose a mathematical model for the variable Y as a function of X
- Pick your PDF
- Write down a likelihood function
- Try different parameter combinations by hand and estimate the most likely parameter values

Useful distributions to get familiar with

For discrete events:

- Binomial
- Multinomial
- Poisson
- Negative binomial

For continuous events:

- Normal
- Lognormal
- Gamma
- Beta

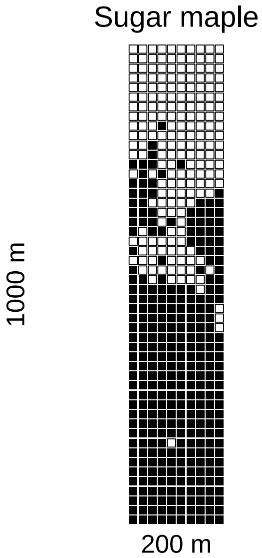
How to choose the right one

- Type of data
- Is the data bounded ?
- Look at the residuals
- Hypothesis / theory driven

Typical problems

- Log of null or negative probabilities
- Likelihood functions converging to 0 in the limit of $\theta \rightarrow \infty$
- Numerical errors (e.g. a normal distribution with a very small SD)
- Impossible predictions (e.g. negative probabilities)
- Underflow / Overflow (if you forget to log)
- Very rough likelihood surface

A species distribution problem



A species distribution problem

The model

$$y \sim f(E, \theta)$$

- f is a probability density function
- [y] is the abundance (in stems per quadrat) of sugar maple
- [E] is elevation
- [θ] is a set of parameters

Things to think about : What are the characteristics of the data ? What is the form of the function ? The probability density function ?