Today

- Training algorithm code
- Likelihood inference
- Coding likelihood intervals
- Common models for data generating process

Independent project

- Data generating process
- System or dataset of your choice
- Submit .py or .R
- Short presentation in exam week
- Due end of semester

Training algorithm code

- Walk through
- `likelihood_inference.md`

Compared to SSQ training algo

- Likelihood with a Normal distribution

Likelihood for a dataset

pdf of the Normal distribution

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\cdot\frac{(y_i-\mu_i)^2}{\sigma^2}}$$

y_i are the data points

μ_i is the mean relationship (det skel)

σ^2 is the variance

Negative log likelihood

$$-\ln\big(L(\theta)\big) = \underbrace{n\left[\ln(\sigma) + \frac{1}{2}\ln(2\pi)\right]}_{} + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu_i)^2$$

This is the SSQ!

Constant w.r.t μ

So, minimizing the nll is the same as minimizing the SSQ

Inference algorithm

$$\frac{P(y|\theta_2)}{P(y|\theta_1)}$$ Likelihood ratio

Algorithm

for each pair of models in a set

    calculate likelihood ratio

judge the relative evidence for the models

Calibrating likelihood ratio

- Calibrate strength of evidence
- How strong do you think it is?
- Two bags with many marbles; which bag?
  - Bag 1: half white, half blue
  - Bag 2: all white
- 3 whites LR = 2^3 = 8

$$\frac{P(3\text{ white} \mid \text{bag2})}{P(3\text{white} \mid \text{bag 1})} = \frac{1}{(\frac{1}{2})^3} = 2^3 = 8$$

- 5 whites LR = 2^5 = 32
- 10 whites LR = 2^{10} = 1024

Likelihood profiles

$$\frac{P(y|\beta_{1i})}{P(y|\beta_{1MLE})}$$

Compare β_1 values for model i against MLE model

Algorithm (no different than previous)

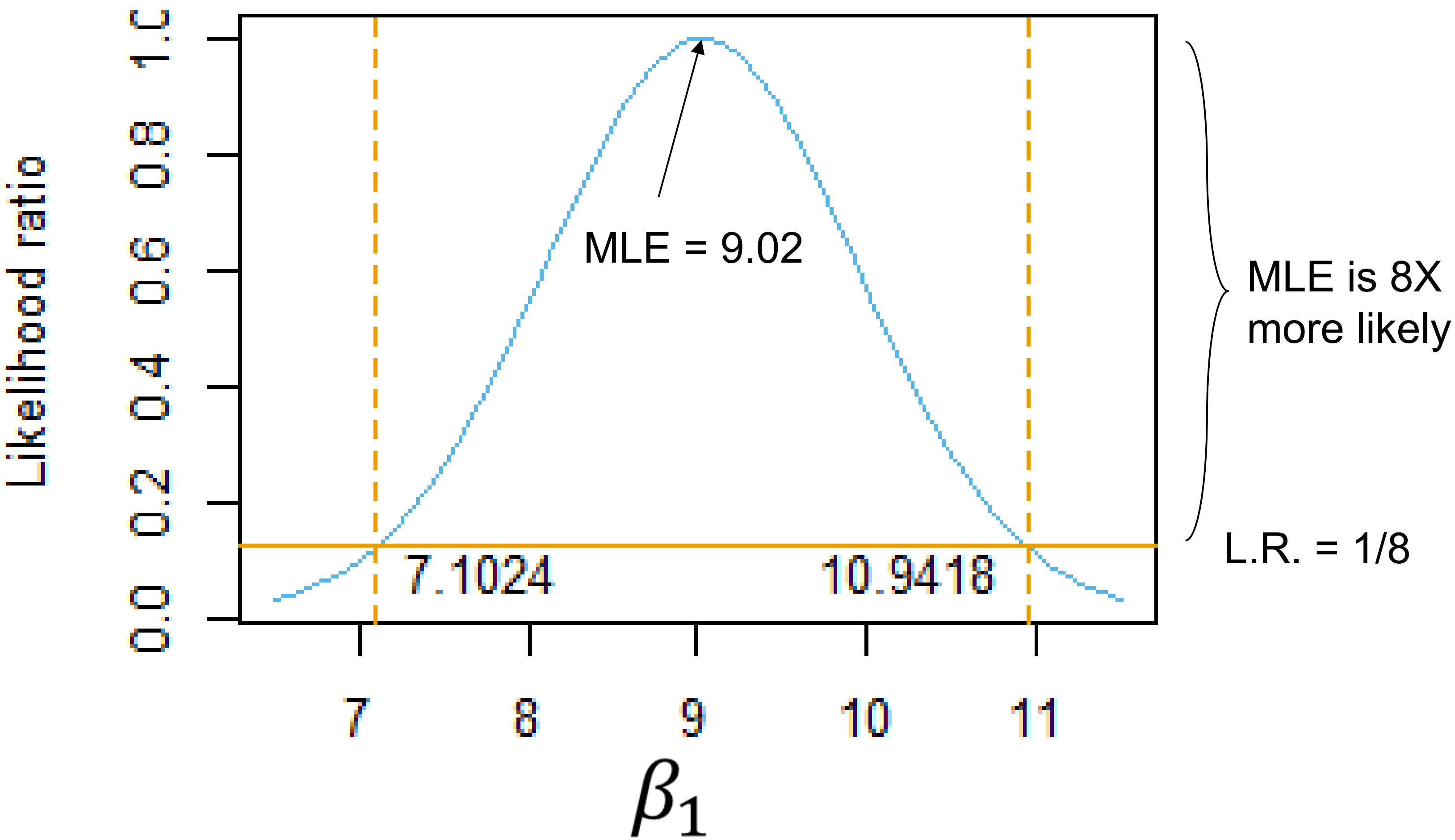
for each pair of models in a set

    calculate likelihood ratio

judge the relative evidence for the models

Likelihood profile & interval

Grid search β_1 while optimizing β_0



Coding likelihood intervals

- Code at end of: `likelihood_inference.Rmd`

Confidence interval	vs	Likelihood interval
95%		1/8
99%		1/32

Notation (equivalent variants)

$$L(\theta) = \mathrm{L} = P(y \mid \theta)$$

Probability of the data given the model parameters

“The likelihood of the model is the probability of the data given the model”

The following is equivalent:

$$L(y;\theta) = P(y \mid \theta)$$

Notice that we use a semicolon or comma here rather than a vertical bar

“The likelihood function is the probability of the data given the model”

Sometimes you may see it this way (e.g. Edwards 1992. "Likelihood".):

$$L(\theta \mid y) = P(y \mid \theta)$$

“The likelihood of the model given the data ...”

The vertical bar is the conditional operator.

Hilborn and Mangel (1997) and some other places in ecology/evolution:

$$L(y \mid \theta) = P(y \mid \theta)$$

This is probably not technically correct.

But DON'T read it thus “The likelihood of the data given the model ...”

Common DGP models

Normal	Poisson	Binomial
$y_i \sim \text{Normal}(\mu_i, \sigma)$	$y_i \sim \text{Poisson}(\mu_i)$	$y_i \sim \text{Binomial}(\mu_i, n)$

Key properties:

Normal	Poisson	Binomial
y: -∞ to ∞, continuous	y: 0 to ∞, discrete, integer	y: 0, 1, discrete, binary
μ: -∞ to ∞, continuous	μ: 0 to ∞, continuous	μ: 0 to 1, probability

Generalized linear models

Normal + Identity link	Poisson + Log link	Binomial + Logit link
$y_i \sim \text{Normal}(\mu_i, \sigma)$	$y_i \sim \text{Poisson}(\mu_i)$	$y_i \sim \text{Binomial}(\mu_i, n)$
$\mu_i = \beta_0 + \beta_1 x_i$	$\log(\mu_i) = \beta_0 + \beta_1 x_i$	$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 x_i$

Key properties:

y: -∞ to ∞, continuous *μ*: -∞ to ∞, continuous	*y*: 0 to ∞, discrete, integer *μ*: 0 to ∞, continuous	*y*: 0, 1, discrete, binary *μ*: 0 to 1, probability