

Today

- Recap & questions from homework
- Coding likelihood intervals?

p-values

- Use constraint; prefer uncertainty intervals
- Key points
 - p-value is not the probability that: “null is true”, “data were generated by the null”, “by chance alone”
 - $p < 0.05$ does not mean “the null hypothesis is false”
 - small p-value does not mean “the effect was large or important”
 - $p > 0.05$ does not mean “there was no effect”, or “the null is true”, or “the effect was small”
 - if many replicated studies have $p > 0.05$ it does not mean “there was no effect”

Likelihood in data science

- This week: **pure** likelihood inference
 - Learning goal: understand likelihood
- Likelihood is also used in
 - Frequentist: as a sample statistic
 - Bayesian: part of the posterior
 - Information theory: e.g. AIC
 - likelihood + complexity penalty

The likelihood principle

All the **evidence** in an observation (data) about the parameters (model) is in the **likelihood function**

The likelihood function

Counts all the ways the data could have happened **for a given model or hypothesis**

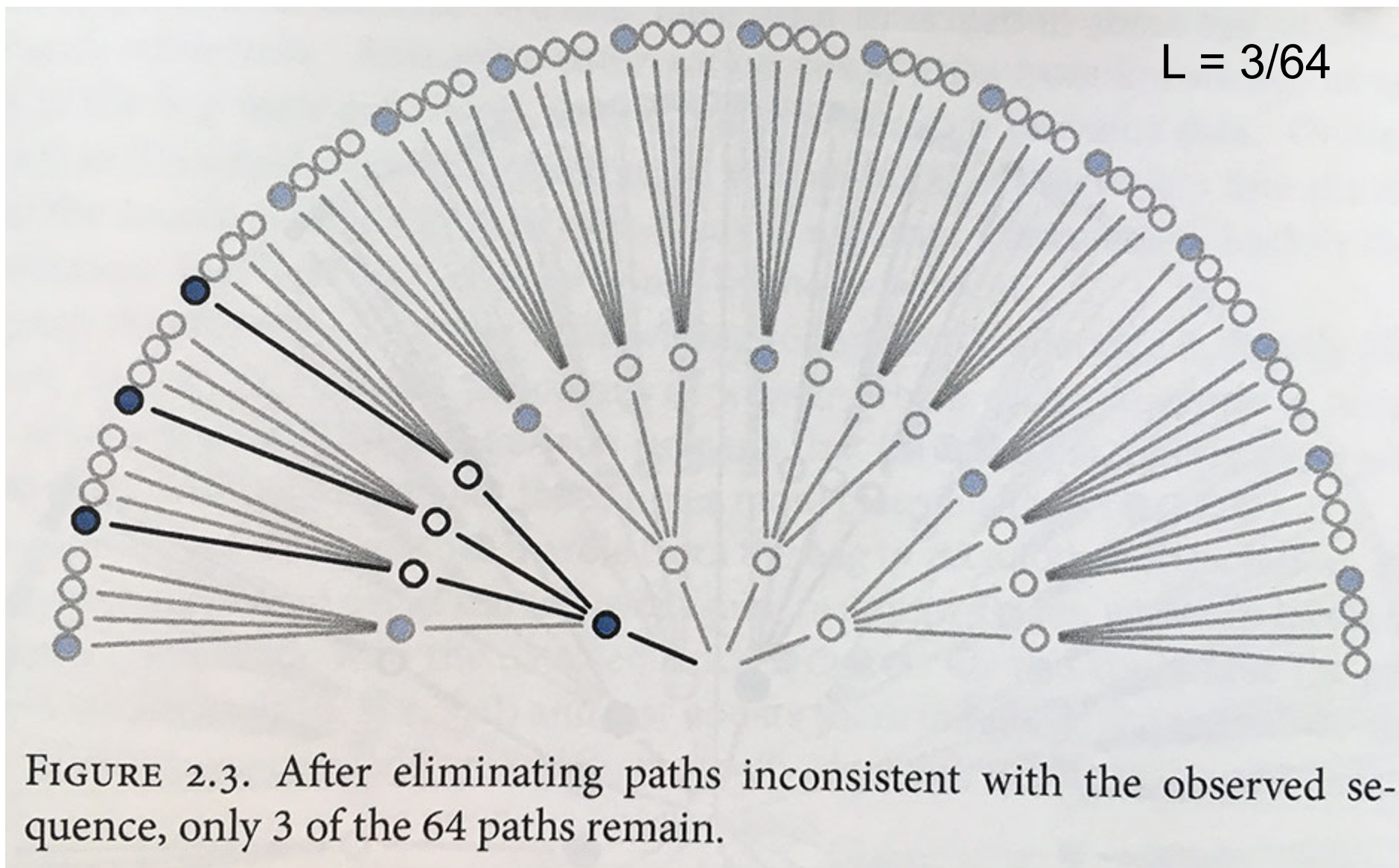
The likelihood

- Probability of the data **given a model**

$$L = P(\text{Data} \mid \text{Model 2})$$

$$P(y \mid \theta_2)$$

A conditional probability

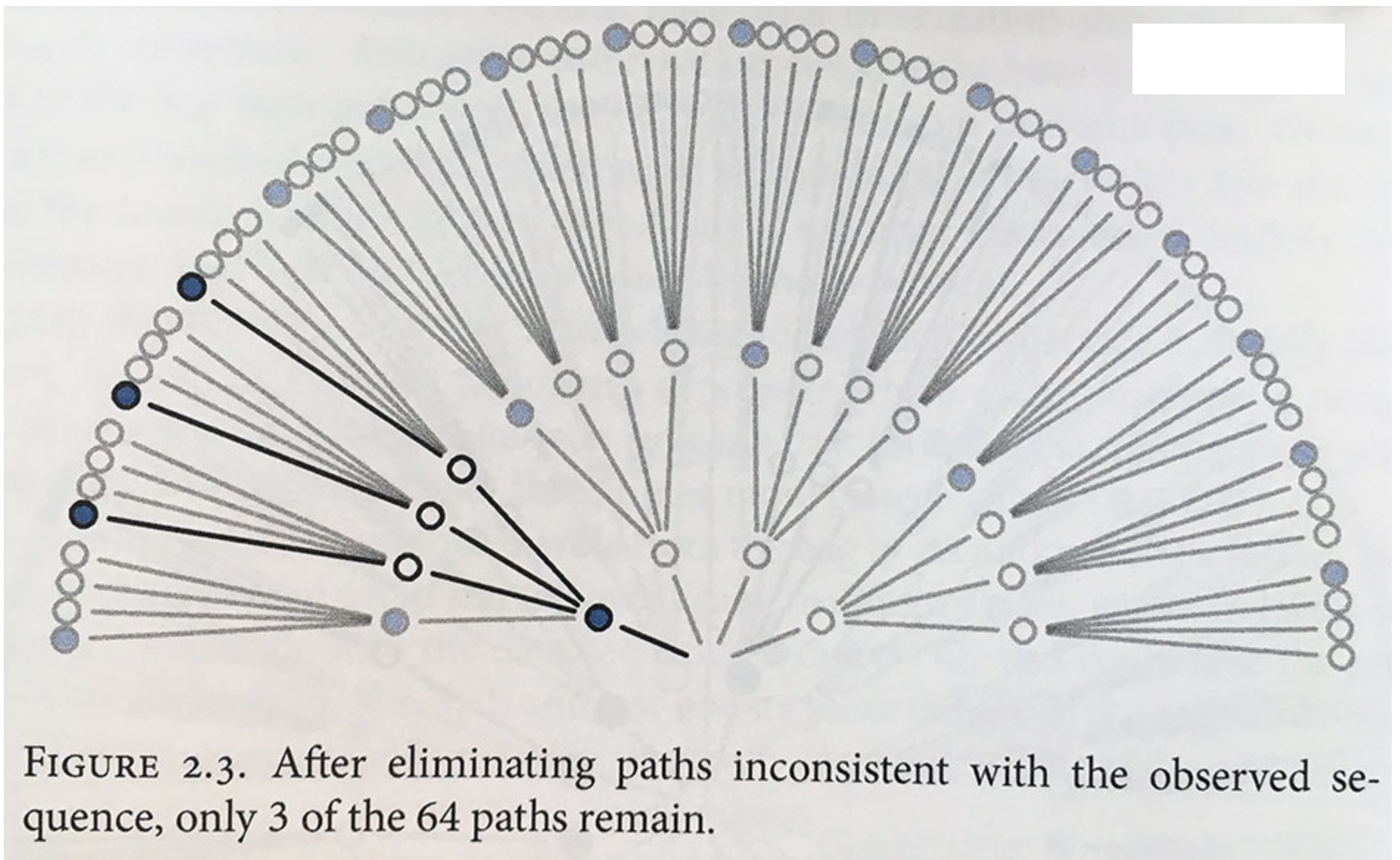


Paths for data



given M_2

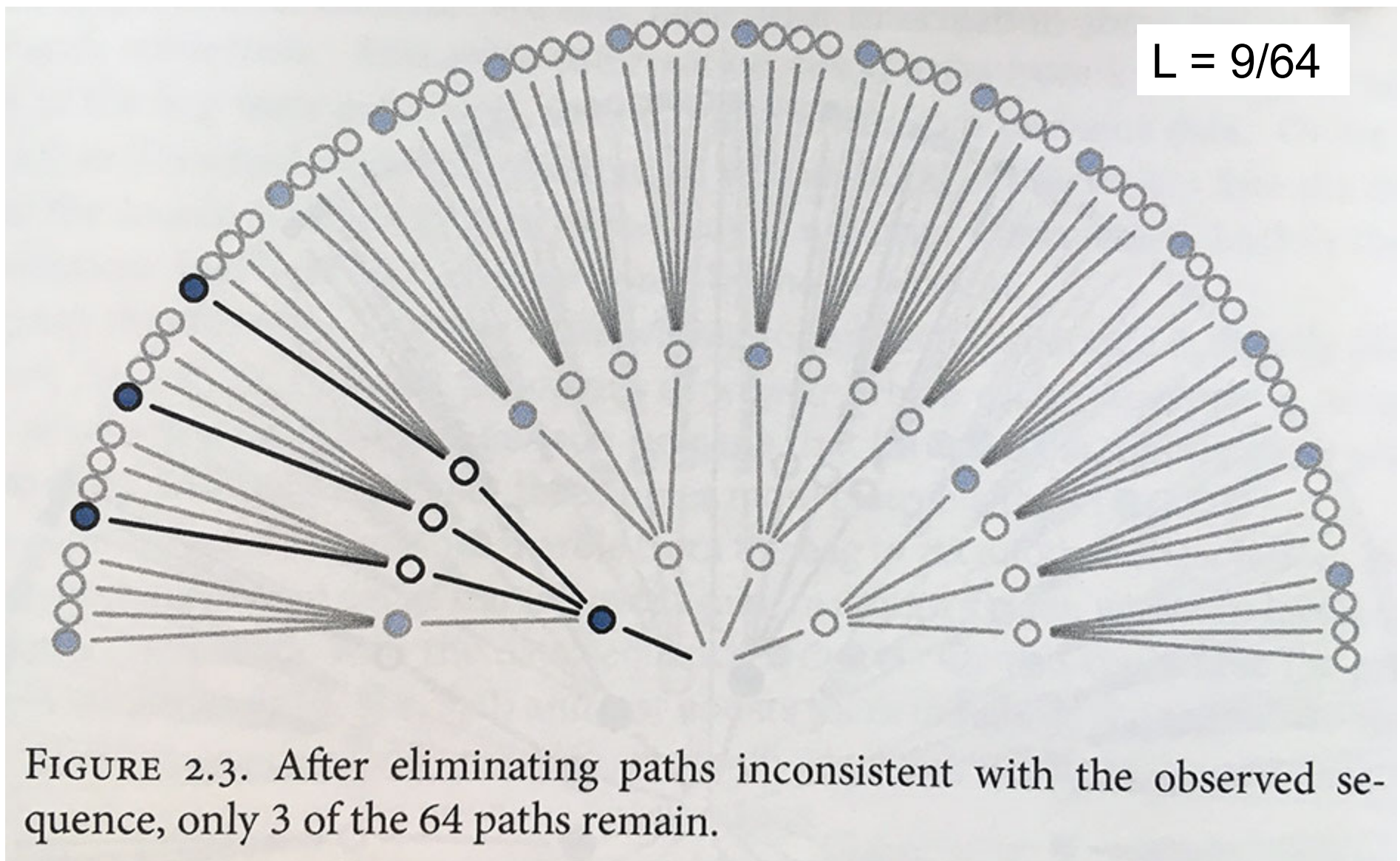




What is the likelihood for
2 blue + 1 white in any order?

given M_2

[●○○○]



What is the likelihood for
2 blue + 1 white in any order?

given M_2

[●○○○]

Likelihood inference for the linear model

Model algorithm

Writing down the model:

$$y \sim \text{Normal}(\mu, \sigma)$$

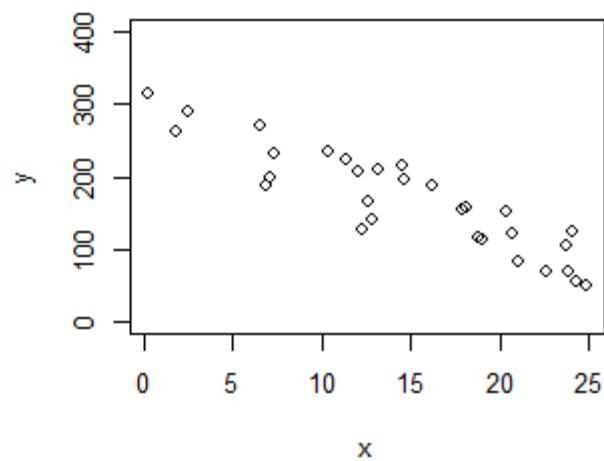
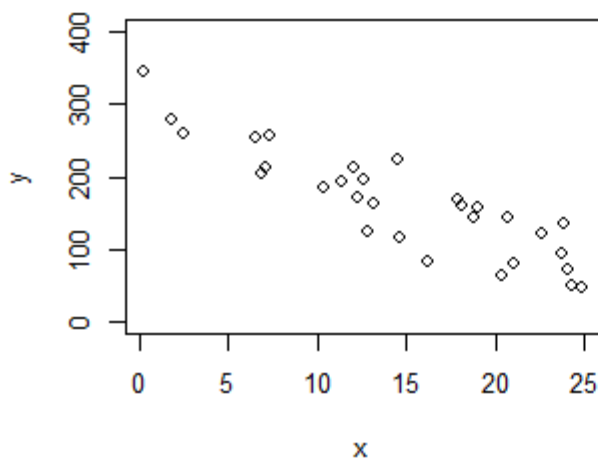
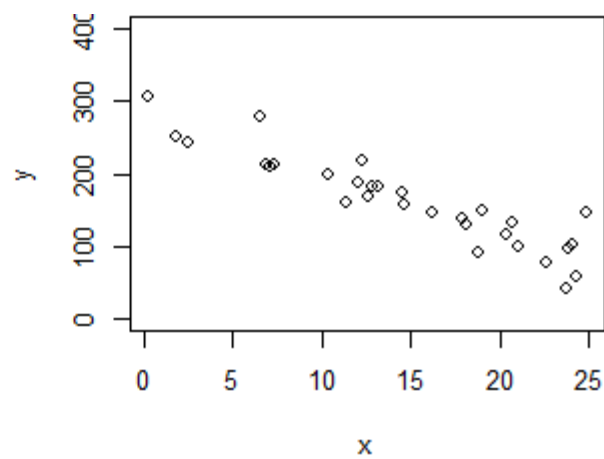
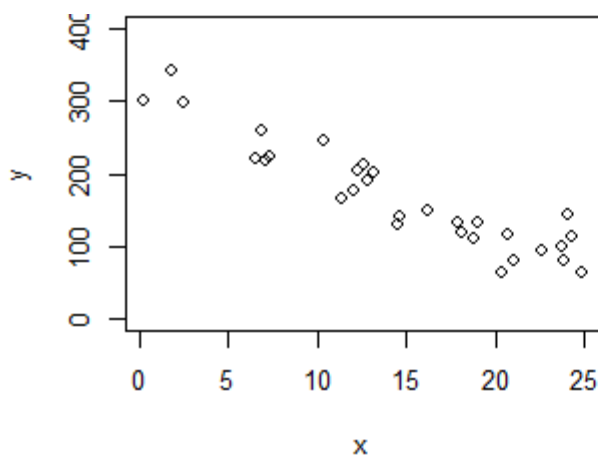
$$\mu = \beta_0 + \beta_1 x$$

Algorithm:

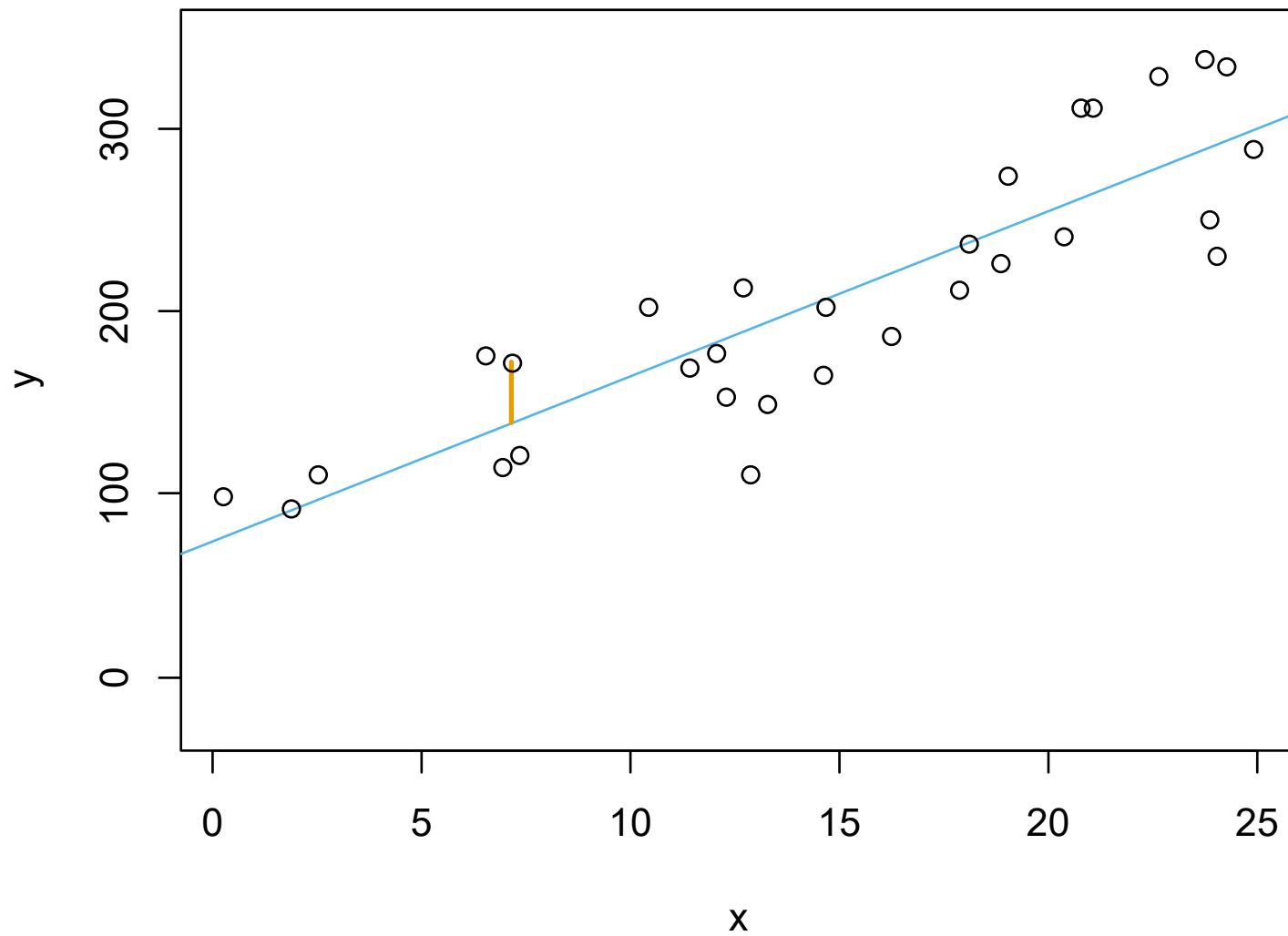
```
lmod <- function(b0, b1, x) {  
  return(b0 + b1 * x)  
}  
ystoch <- function(mu=lmod(b0, b1, x), sigma) {  
  return(rnorm(n=length(mu), mean=mu, sd=sigma))  
}
```

Simulating the model

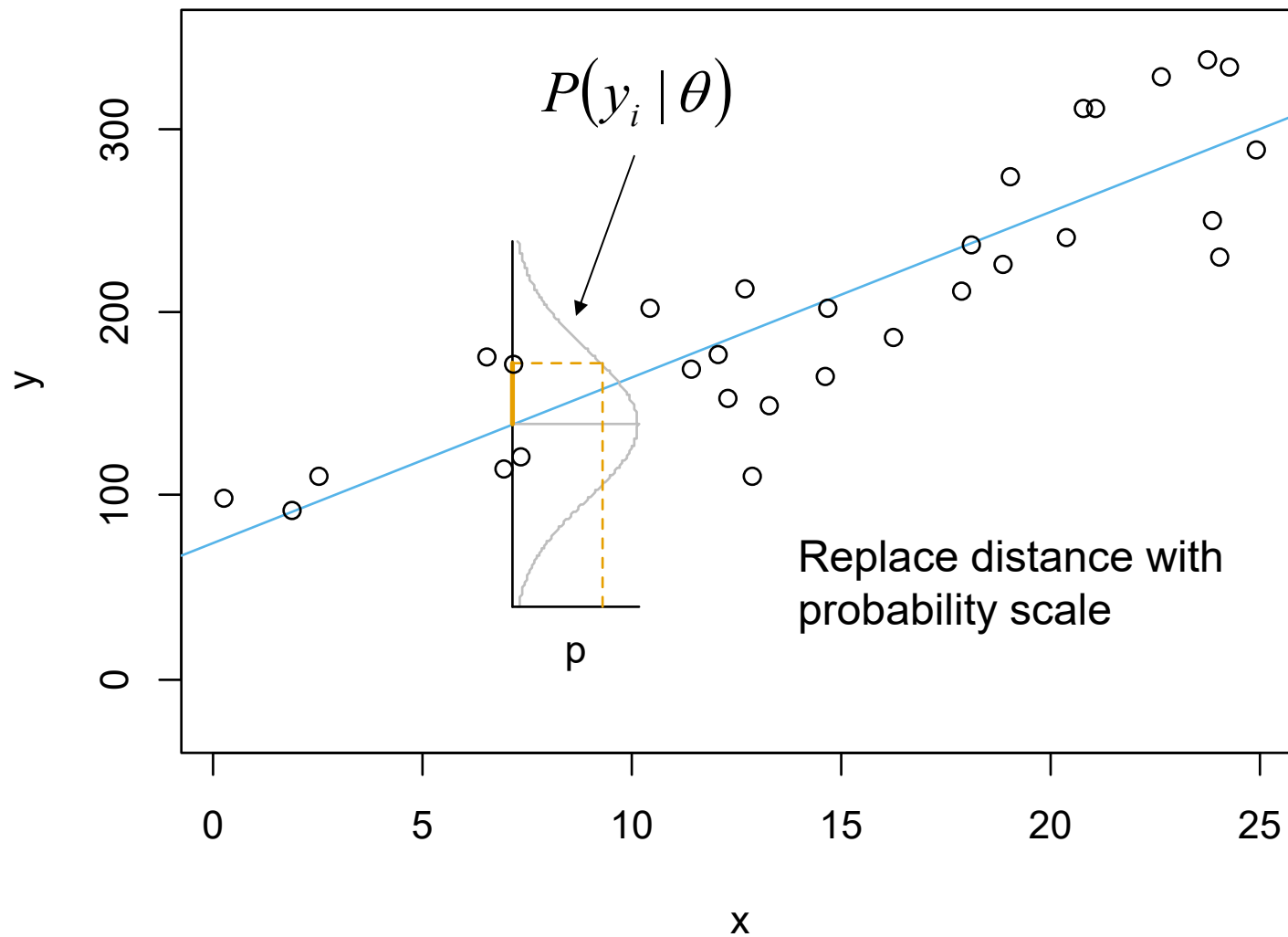
```
ystoch(mu=lmod(b0=300, b1=-9, x=x), sigma=30)
```



Likelihood (linear, Normal)



Likelihood (linear, Normal)



Likelihood (linear, Normal)

Likelihood for the model:

$$L(\theta) = P(y \mid \theta) = P(y \mid \beta_0, \beta_1, \sigma)$$

```
dnorm(y, mean = beta_0 + beta_1 * x, sd)
```

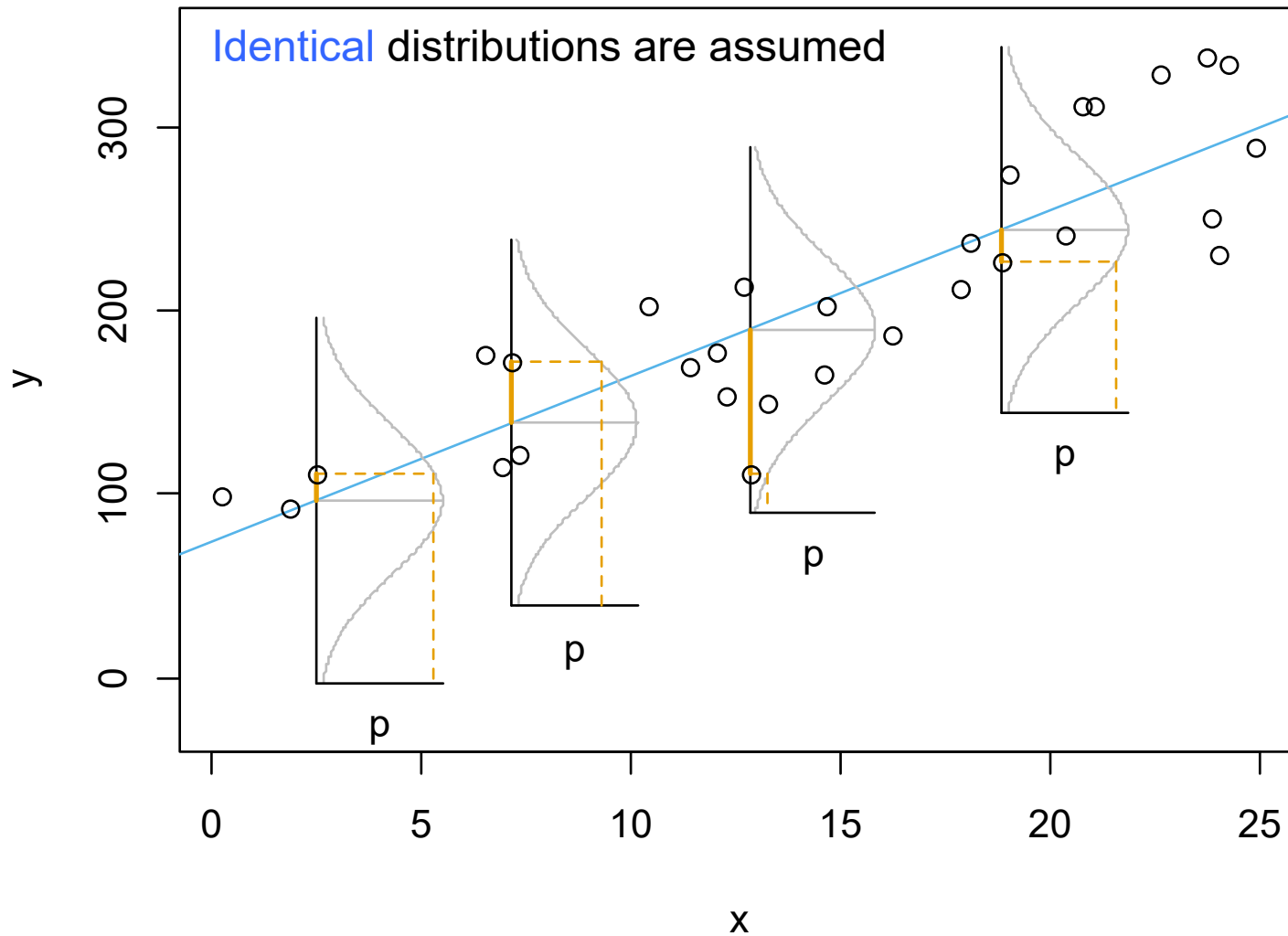
Total likelihood for a data set

One data point: $P(y_1 | \theta)$

All data points: $\prod_i^n P(y_i | \theta)$

because probabilities multiply together to give total probability (n is the number of datapoints). **Independence** is assumed.

Likelihood (linear, Normal)



Support function

The log likelihood:

$$\sum_i^n \ln P(y_i | \theta)$$

Instead of multiplying small probabilities, it is more accurate and convenient to sum their logs.

```
sum(dnorm(y, mean=beta_0 + beta_1 * x, sd, log=TRUE))
```

Training algorithm: Maximum likelihood

The **values of the parameters** that **maximize the likelihood**. In other words, the model that maximizes the probability of the data.

An **optimization** problem.

In practice: minimize the negative log likelihood. The model with the most support, has the smallest negative log likelihood.

Training algorithm

```
lmod <- function(b0, b1, x) {  
  return(b0 + b1 * x)  
}  
lm_nll <- function(p, y, x) {  
  mu <- lmod(b0=p[1], b1=p[2], x=x)  
  nll <- -sum(dnorm(y, mean=mu, sd=p[3], log=TRUE))  
  return(nll)  
}  
fitlm <- optim(p=start_pars, lm_nll, y=y, x=x)  
fitlm
```

call the linear model

Find maximum likelihood estimates (MLE) for all 3 parameters

minus sum log likelihoods

Training general approach

- 1) process model function
 - the biology
 - > `biomod(parameters)`
- 2) nll function (error distribution)
 - the stochasticity or error
 - > `-sum(ddist(data, dpar=biomod, error_parameters, log=T))`
- 3) optimize
 - find biology parameters and error parameters
 - > `optim(parameters, nllfunc, data)`
- This recipe is the same no matter how complicated the process model or what the error distribution is

Inference algorithm

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

Bayes rule to the rescue:

$$\frac{P(\theta_2|y)}{P(\theta_1|y)} = \frac{kP(y|\theta_2)}{kP(y|\theta_1)} = \frac{P(y|\theta_2)}{P(y|\theta_1)} = LR$$

for each pair of models in a set
calculate likelihood ratio
judge the relative evidence for the models

$$\frac{P(y|\beta_{1i})}{P(y|\beta_{1MLE})} \quad \text{Compare } \beta_1 \text{ values for model } i \text{ against MLE model}$$

Calibrating likelihood ratio

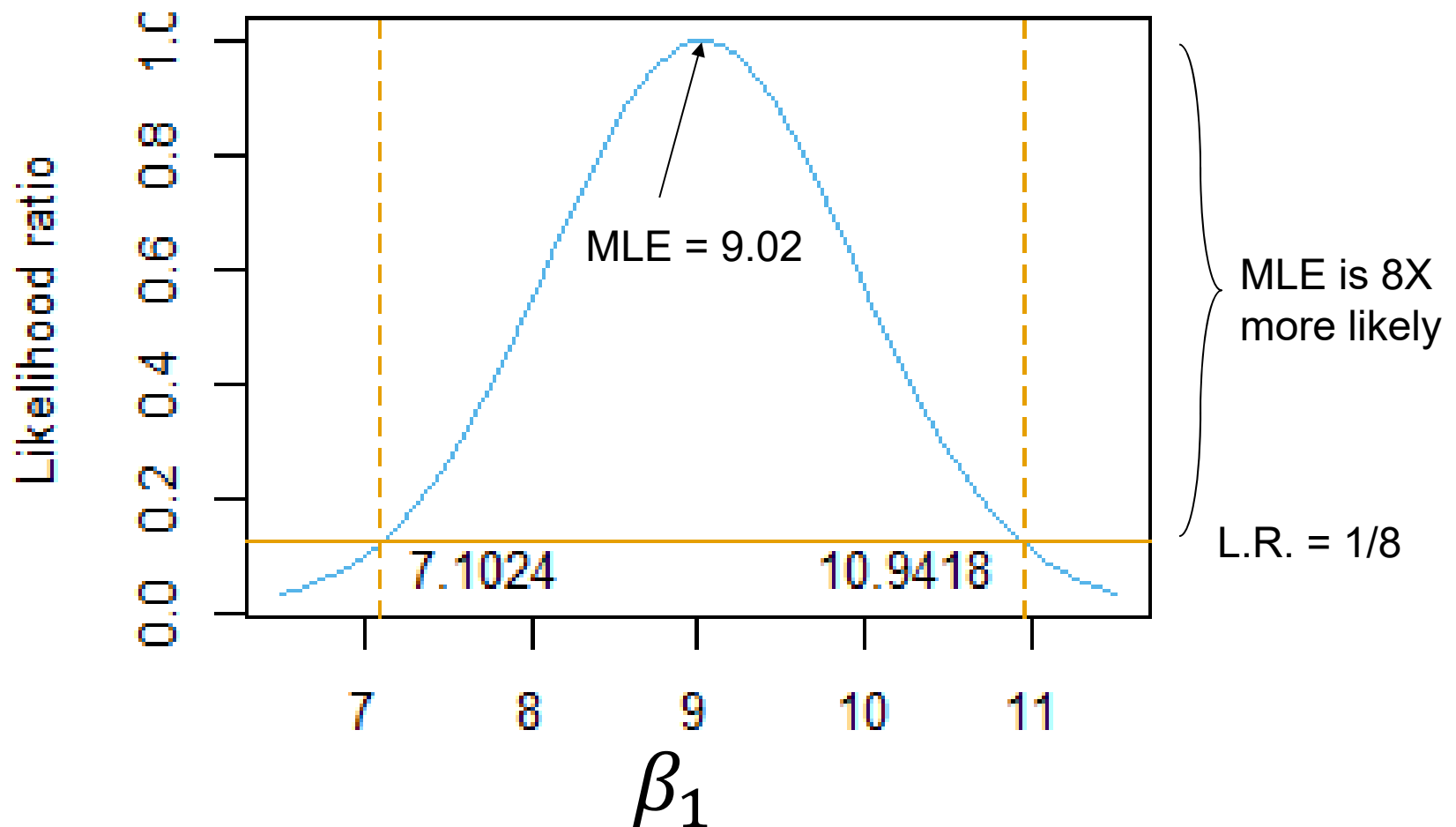
- Measure strength of evidence
- How strong do you think it is?
- Two bags with many marbles
 - Bag 1: half white, half blue
 - Bag 2: all white

- 3 whites $LR = 2^3 = 8$
- 5 whites $LR = 2^5 = 32$
- 10 whites $LR = 2^{10} = 1024$

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

Likelihood profile & interval

Grid search β_1 while optimizing β_0



Compared to SSQ

- Likelihood with a Normal distribution

Likelihood for a dataset

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

pdf of the Normal distribution

y_i are the data points
 μ_i is the mean relationship
 σ^2 is the variance

Negative log likelihood

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

This is the SSQ!

Constant w.r.t μ

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

Notation (equivalent variants)

$$L(\theta) = \mathcal{L} = P(y | \theta)$$

← 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:

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

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".):

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

The vertical bar is the conditional operator.

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

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

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

This is probably not technically correct.

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

Coding likelihood intervals

- Do it for your data
- Code at end of
06_3_likelihood_inference.Rmd