International Statistical Genetics Workshop • Boulder, Colorado • IBG

Likelihood

Baptiste Couvy-Duchesne

Mike Hunter

David Evans

2024

https://github.com/baptisteCD/Likelihood

https://baptistecd.github.io/Likelihood/

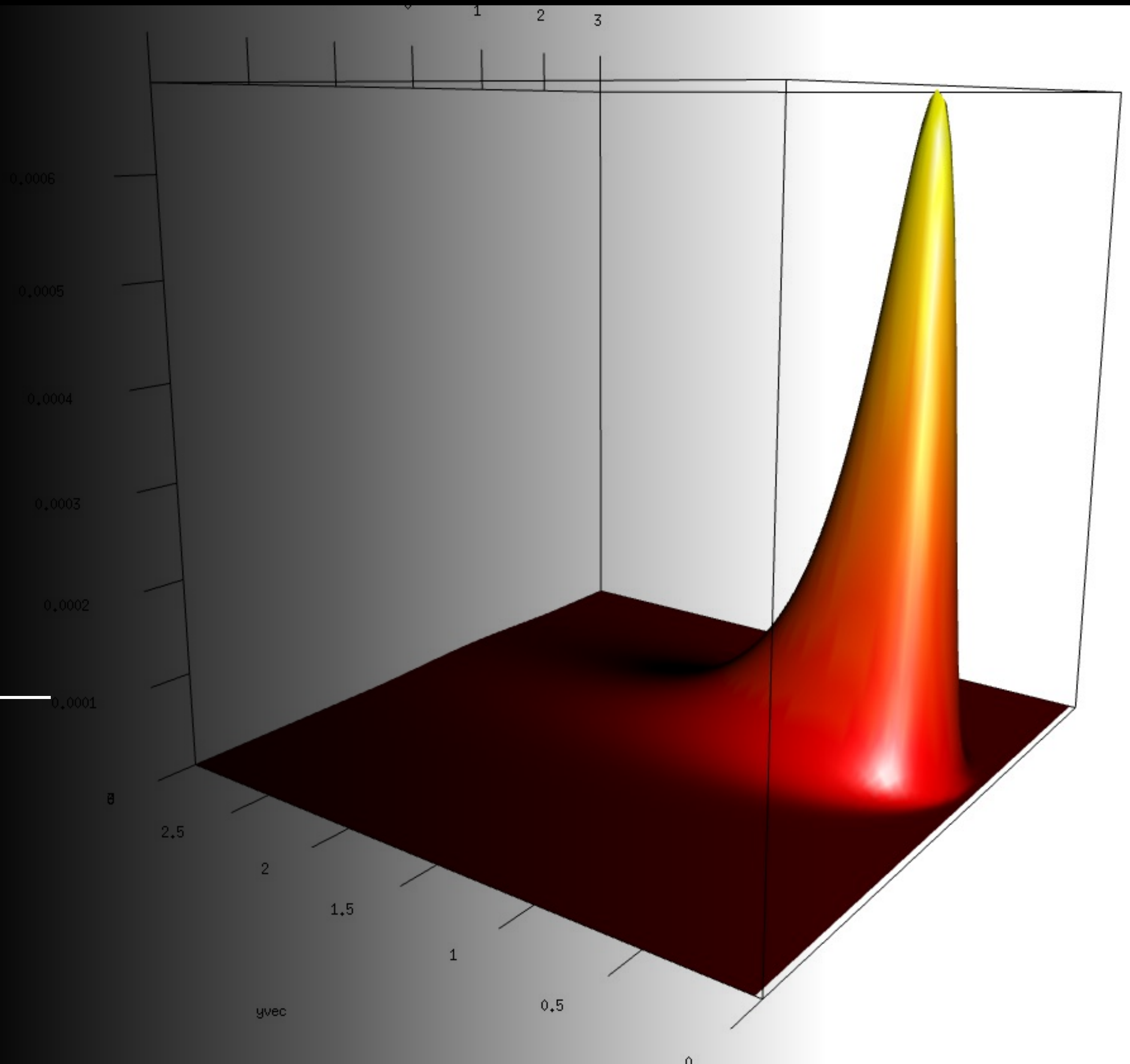
Overview

Likelihood is a central concept/object in statistics
Used across linear models, mixed models, SEM…

<u>It can be used to</u>

- Estimate model parameters
- Test significance of parameters
- Quantify goodness of fit of a model
- Estimate confidence intervals

Likelihood function



Likelihood (function)

The likelihood function (often simply called the likelihood) is the **joint probability** of the **observed data** viewed as a **function of the parameters of a statistical model.**

$$\mathcal{L}(\theta \mid x) = \prod_{j=1}^{N} P_\theta (x_j)$$

Assuming observations are i.i.d

It is not a probability density over the parameter θ
It is not the posterior probability of θ given the data x

Likelihood (function) – example

θ: *probability of heads*

$$\mathcal{L}(\theta \mid x) = C.\prod_{j=1}^{N} P_\theta(x_j) = C.P_\theta(head).P_\theta(head).P_\theta(tails)$$

$$= C.\theta\,.\,\theta\,.\,(1\text{-}\,\theta)$$

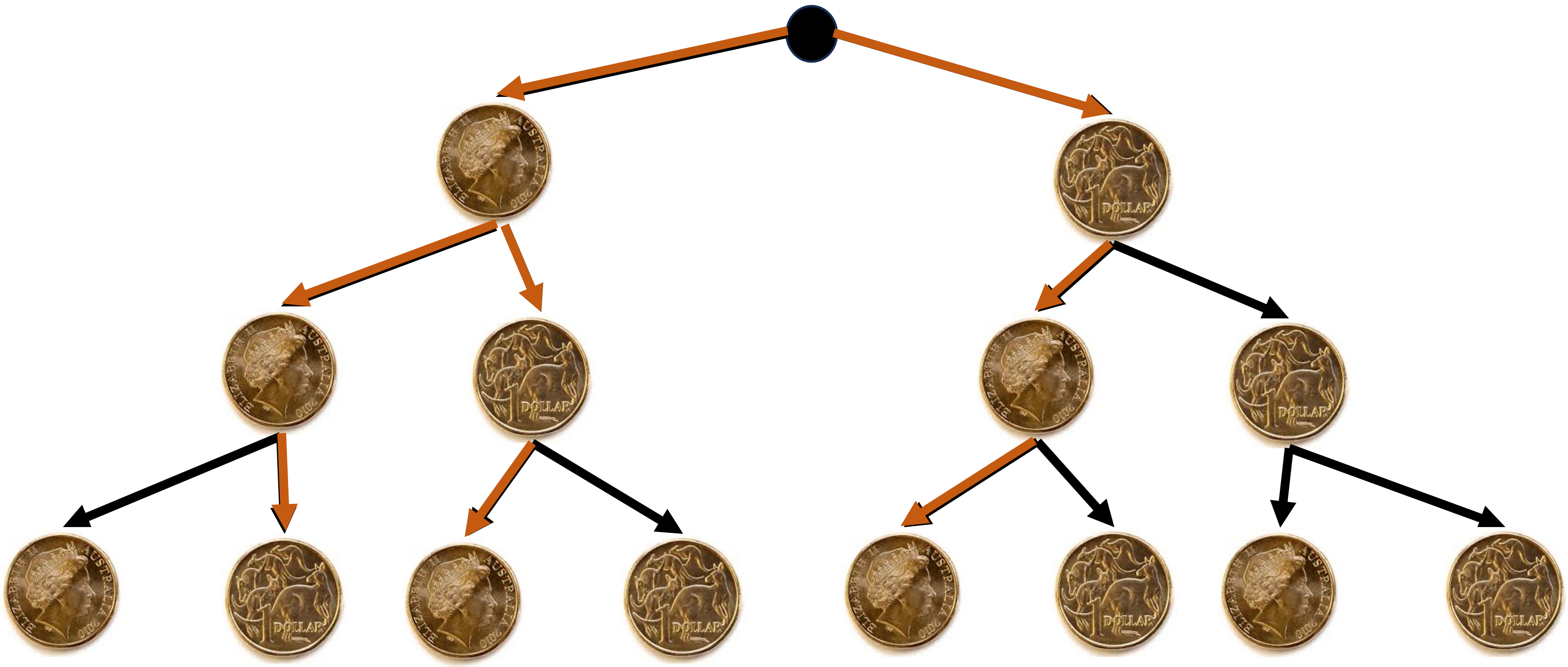
Combination - C

$$C = \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

n choose k

Number of k (heads) from n (coin flips)

Combination – C – with trees



$$\binom{3}{2} = \frac{3!}{2!\,(1)!} = \frac{1\times2\times3}{1\times2\times1} = 3$$

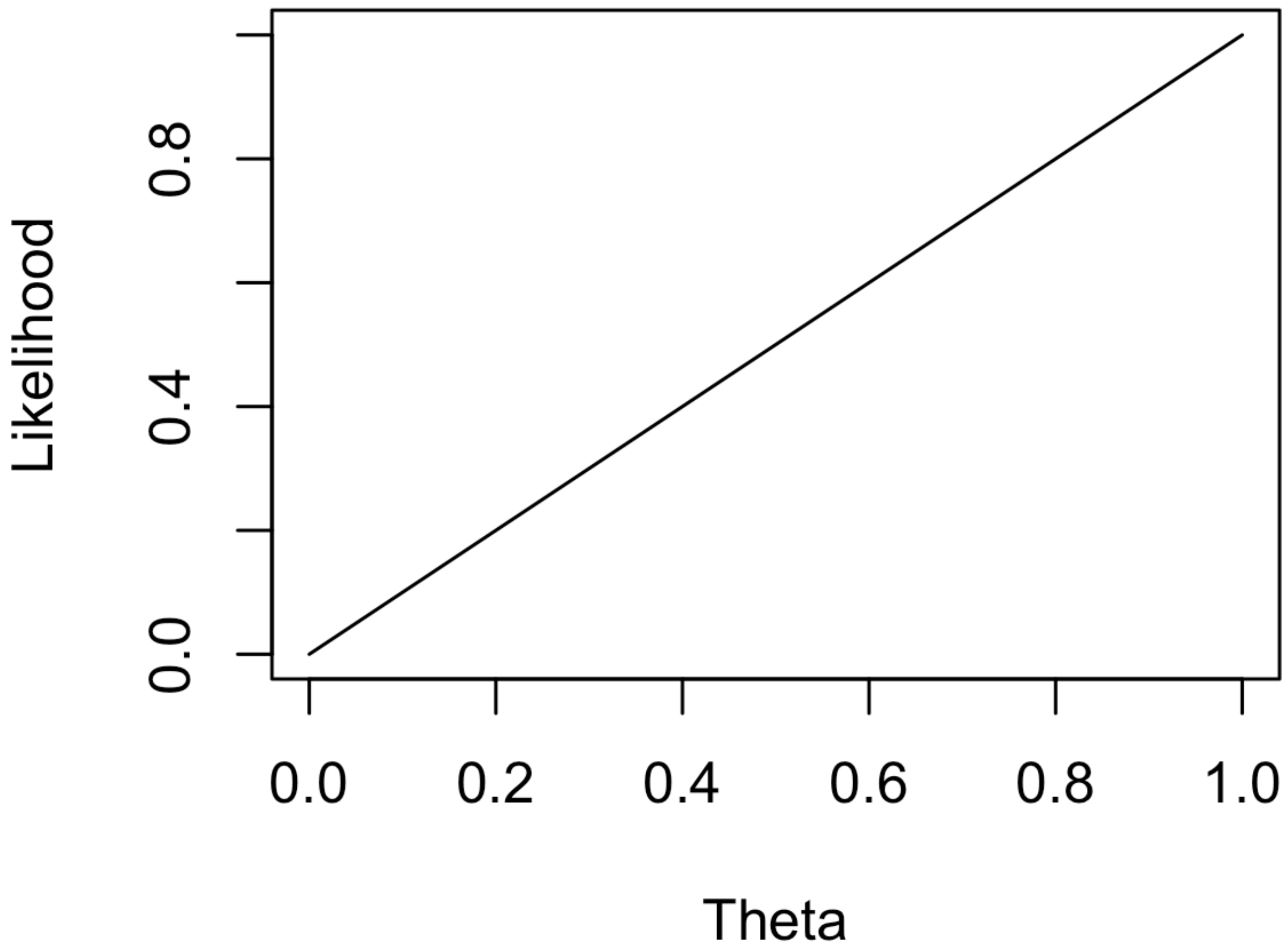
Number of branches with 2 heads out of 3 coin flips

Likelihood (function) – visual example

θ: *probability of heads*

x : *head*

$\mathcal{L}(\theta \mid x) = C.\theta \propto \theta$

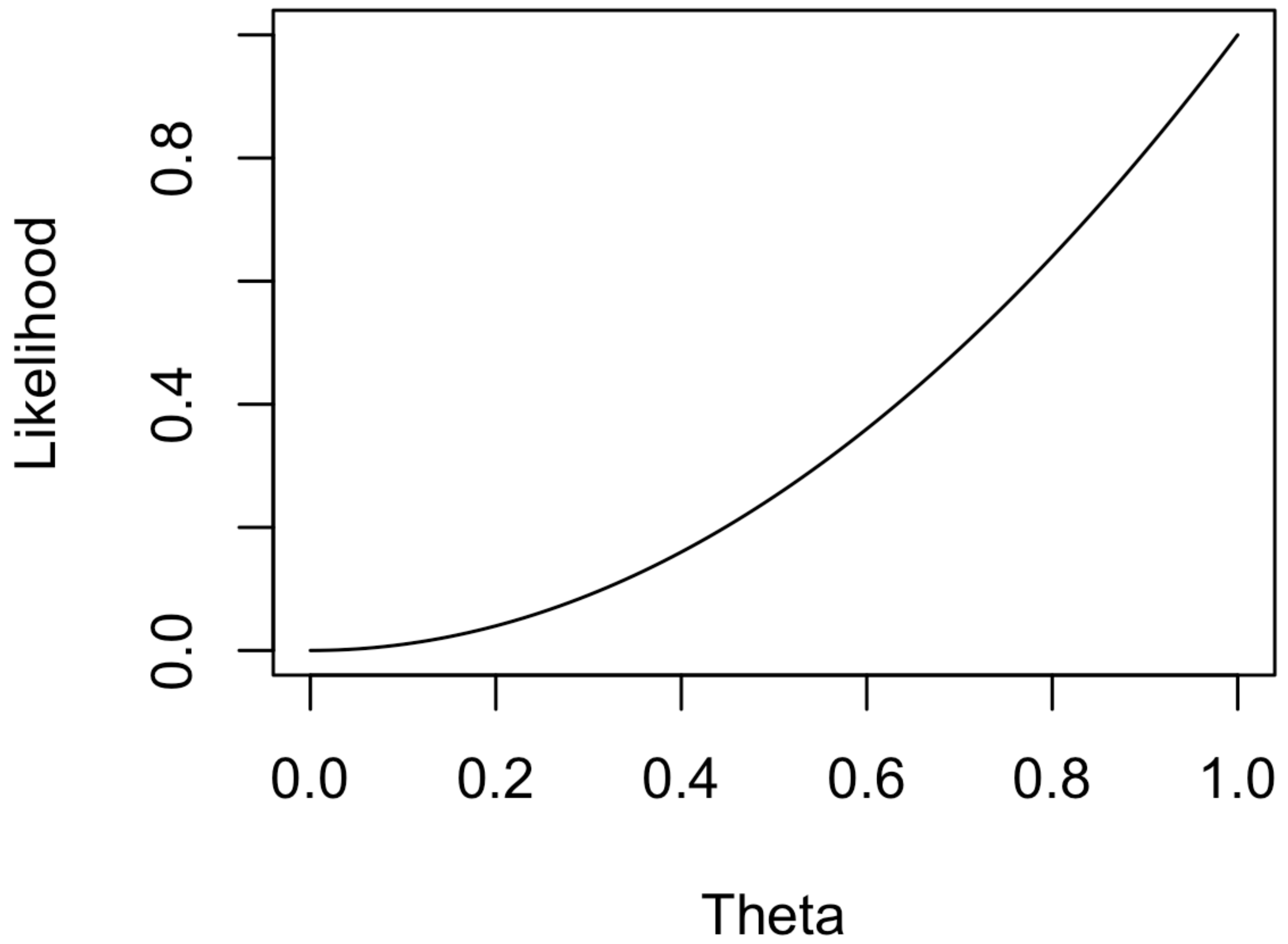


Likelihood (function) – visual example

θ: *probability of heads*

x : *heads, heads*

$\mathcal{L}(\theta \mid x) \propto \theta . \theta$

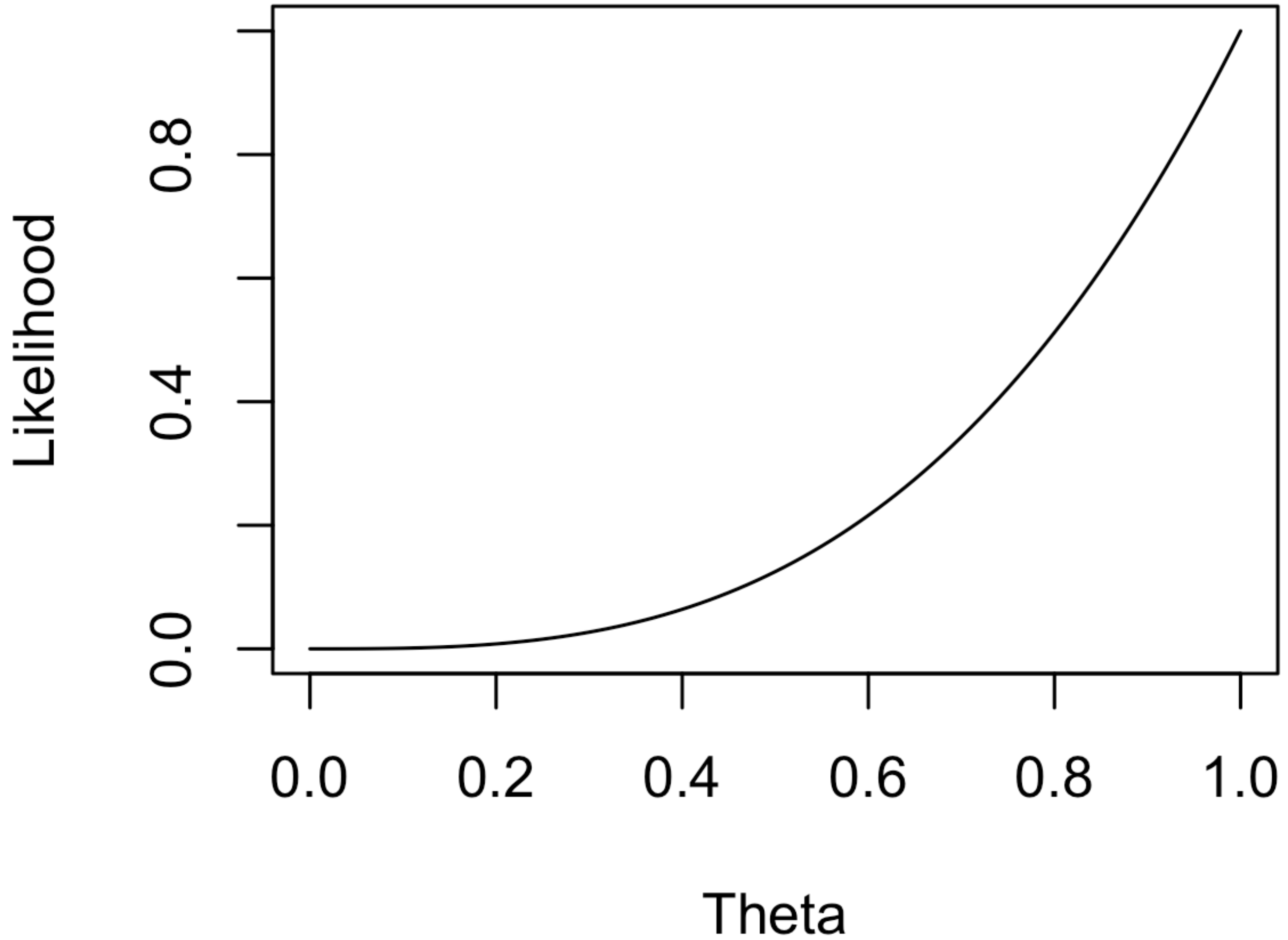


Likelihood (function) – visual example

$\theta : probability\ of\ heads$

$x : heads, heads, heads$

$\mathcal{L}(\theta \mid x) \propto \theta . \theta . \theta$

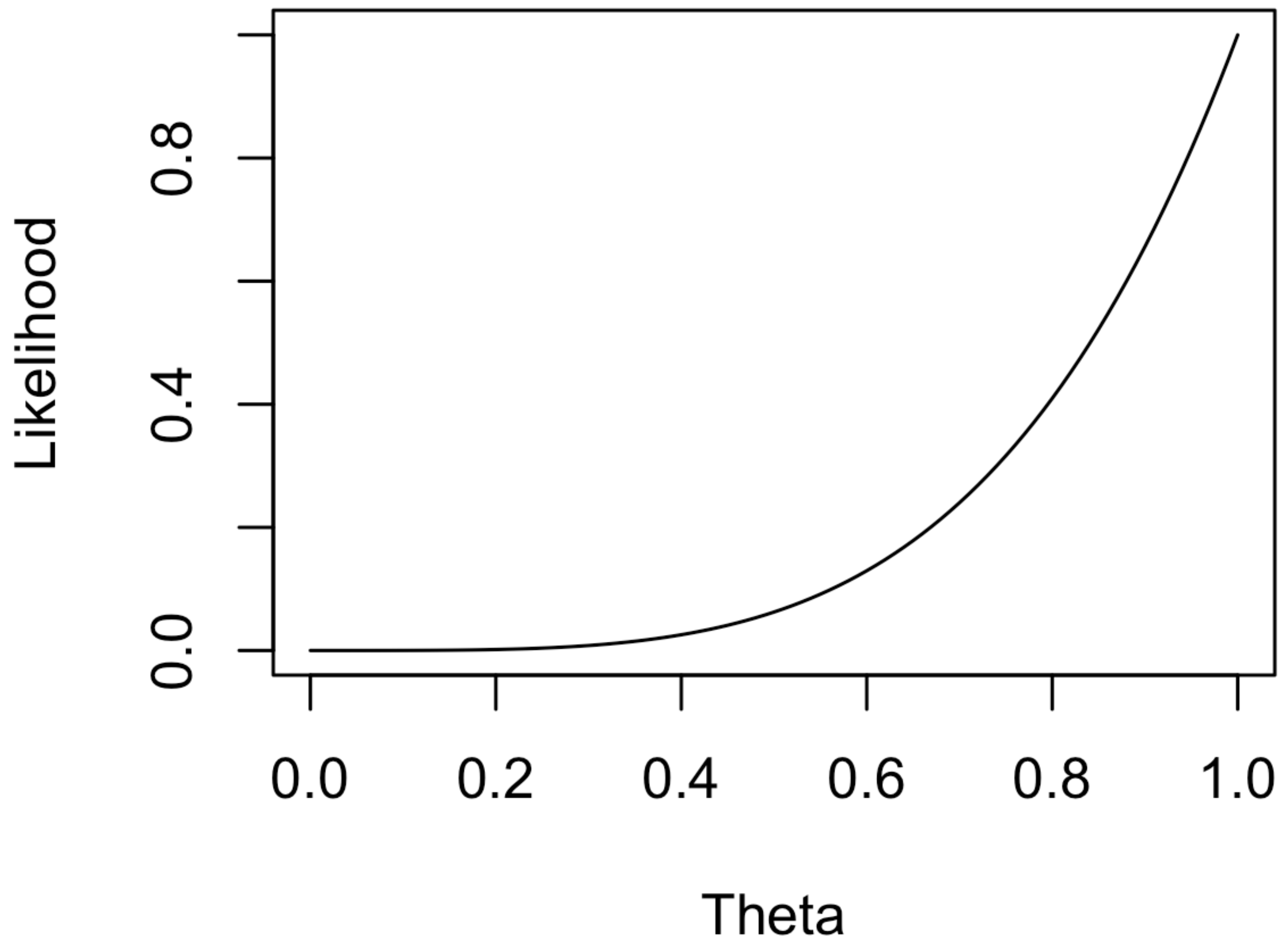


Likelihood (function) – visual example

θ: *probability of heads*

x : *heads, heads, heads, heads*

$\mathcal{L}(\theta \mid x) \propto \theta.\theta.\theta.\theta$

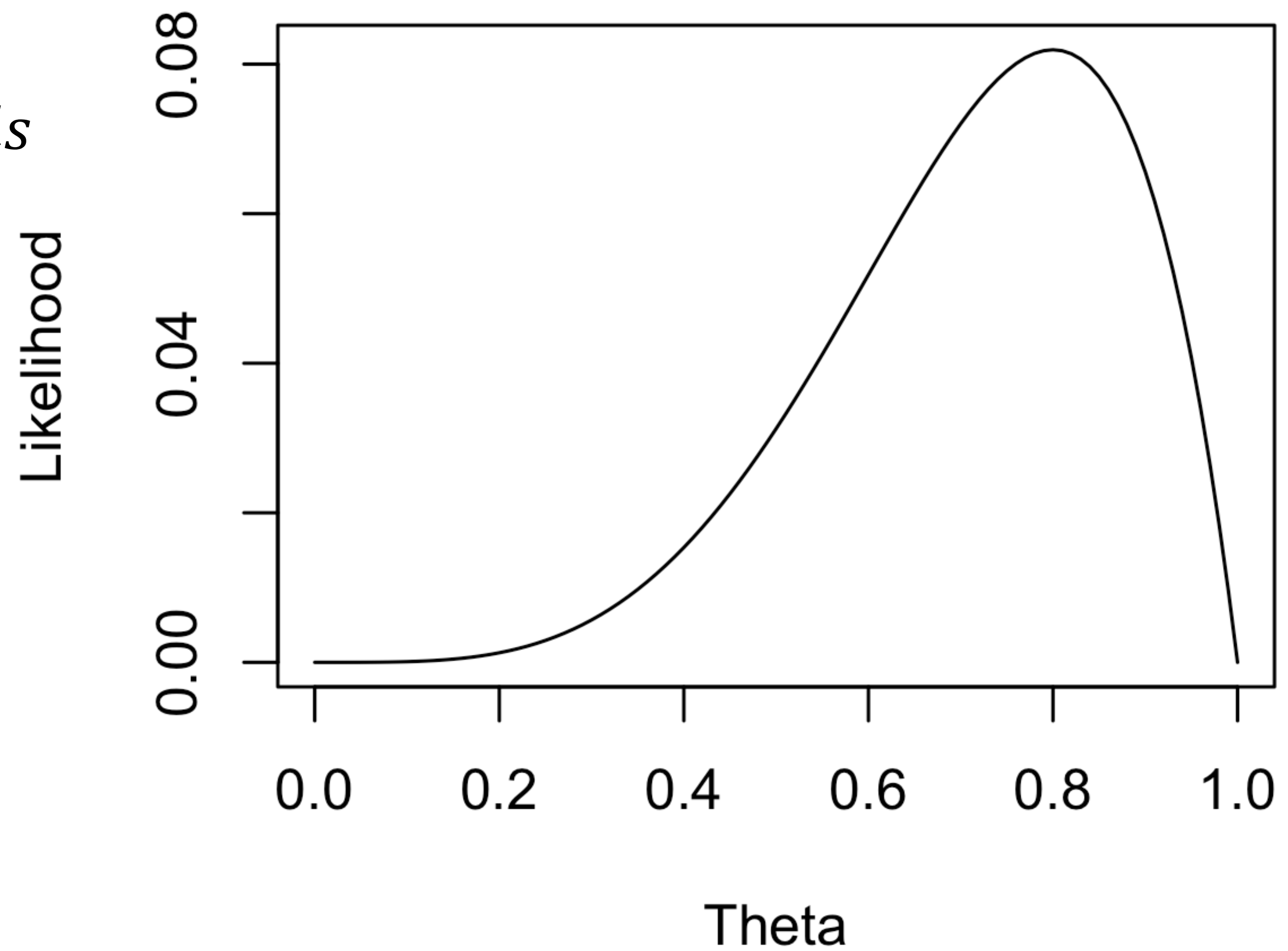


Likelihood (function) – visual example

θ: *probability of heads*

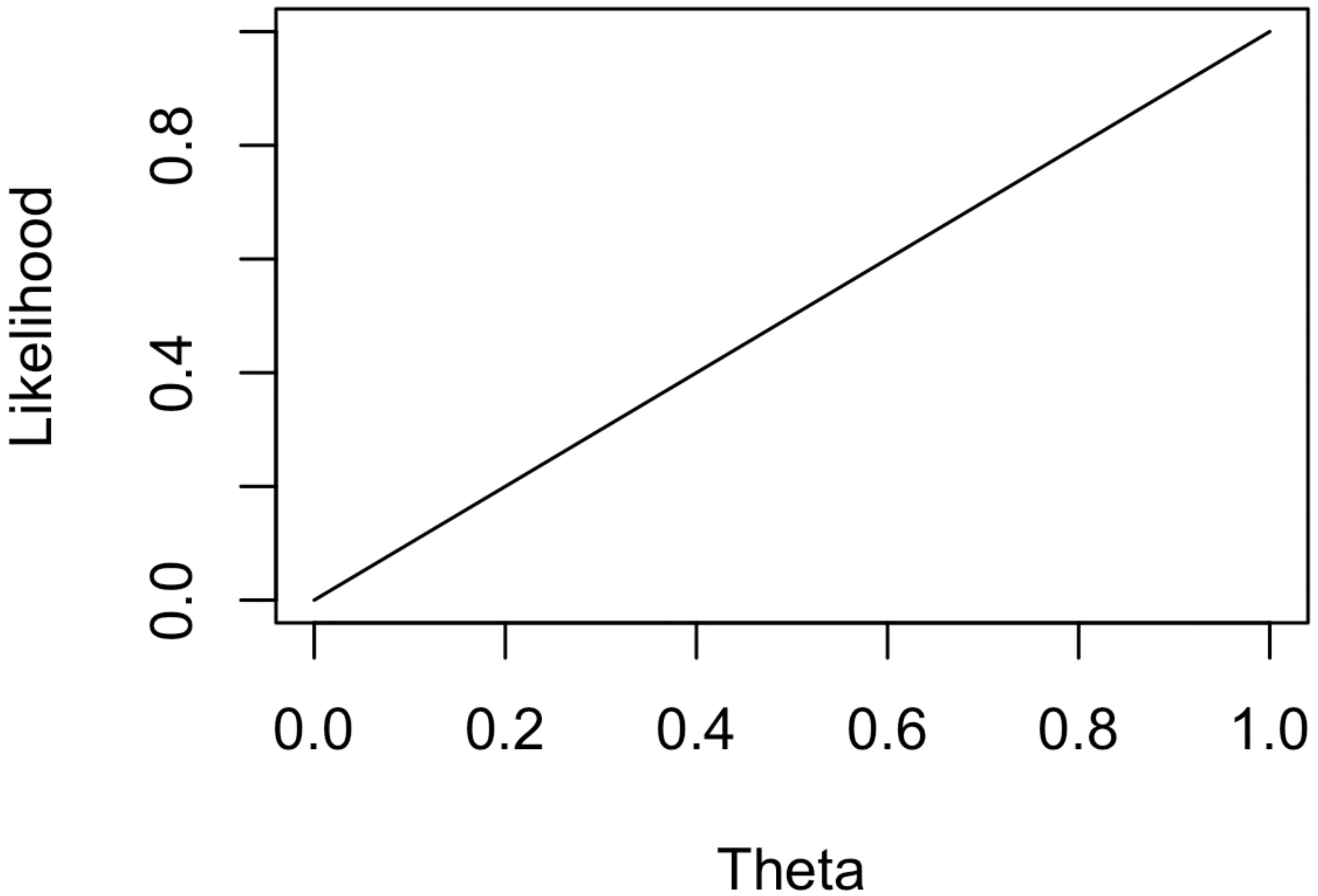
x : *heads, heads, heads, heads, tails*

$$\mathcal{L}(\theta \mid x) \propto \theta.\theta.\theta.\theta.(1-\theta)$$



Likelihood (function) – visual example

Evloution of Likelihood function as you add observations (1 to 100 coin flips)

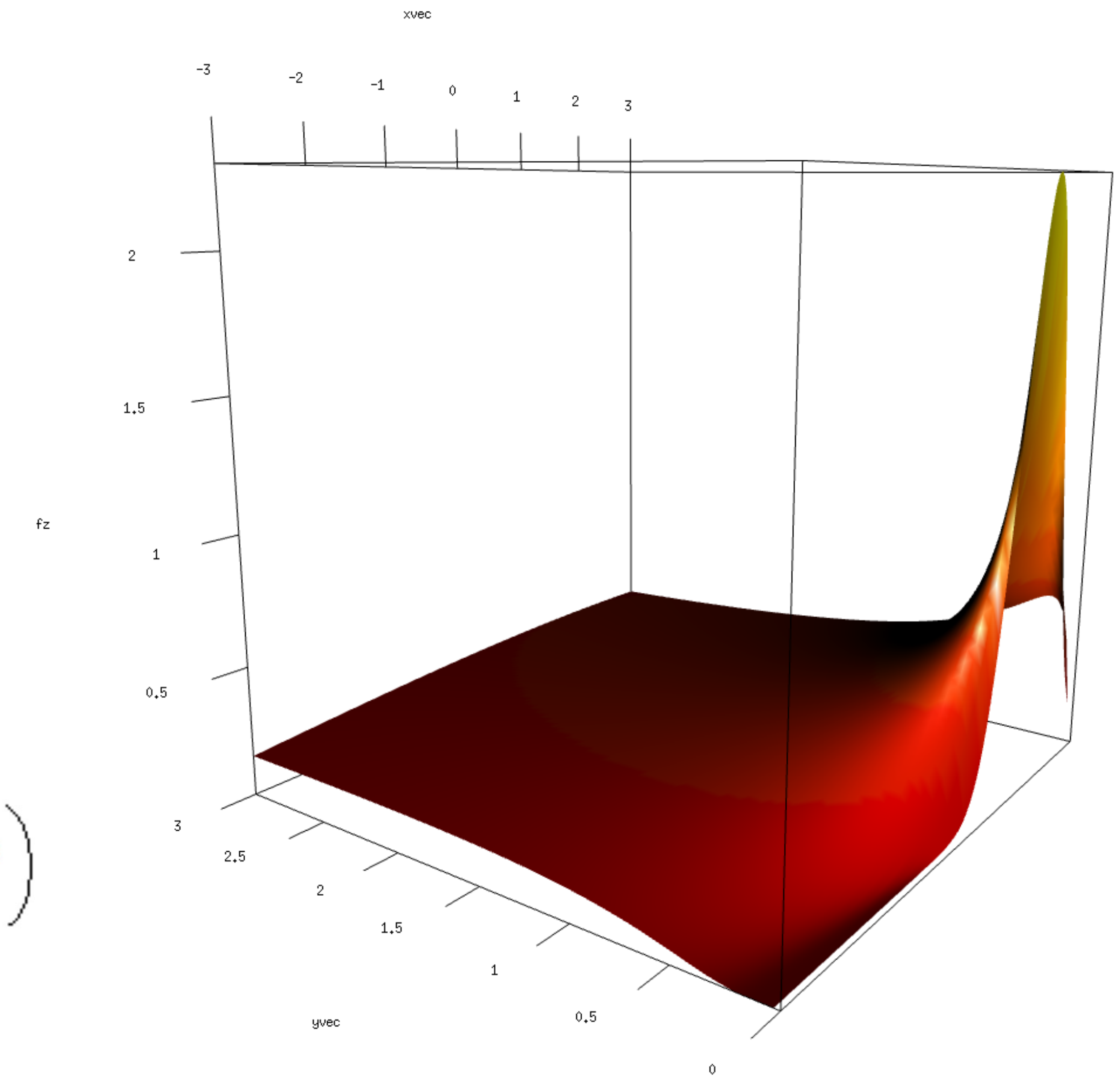


Likelihood (function) – of linear model

$$y = X\beta_0 + \varepsilon$$

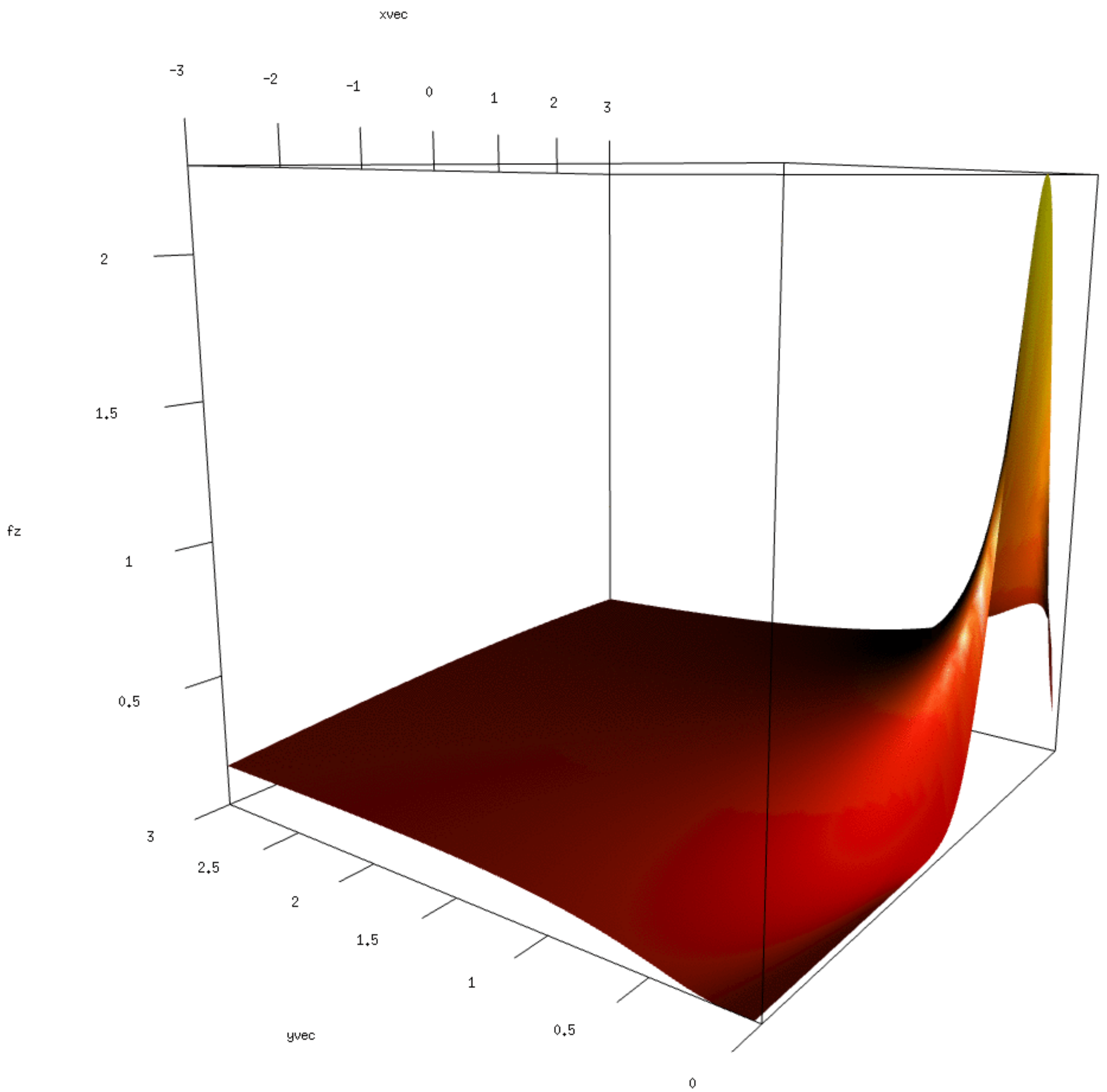
$$\sigma_0^2 = \mathrm{Var}[\varepsilon_i | X]$$

$$L(\beta, \sigma^2; y, X) = (2\pi\sigma^2)^{-N/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i\beta)^2\right)$$

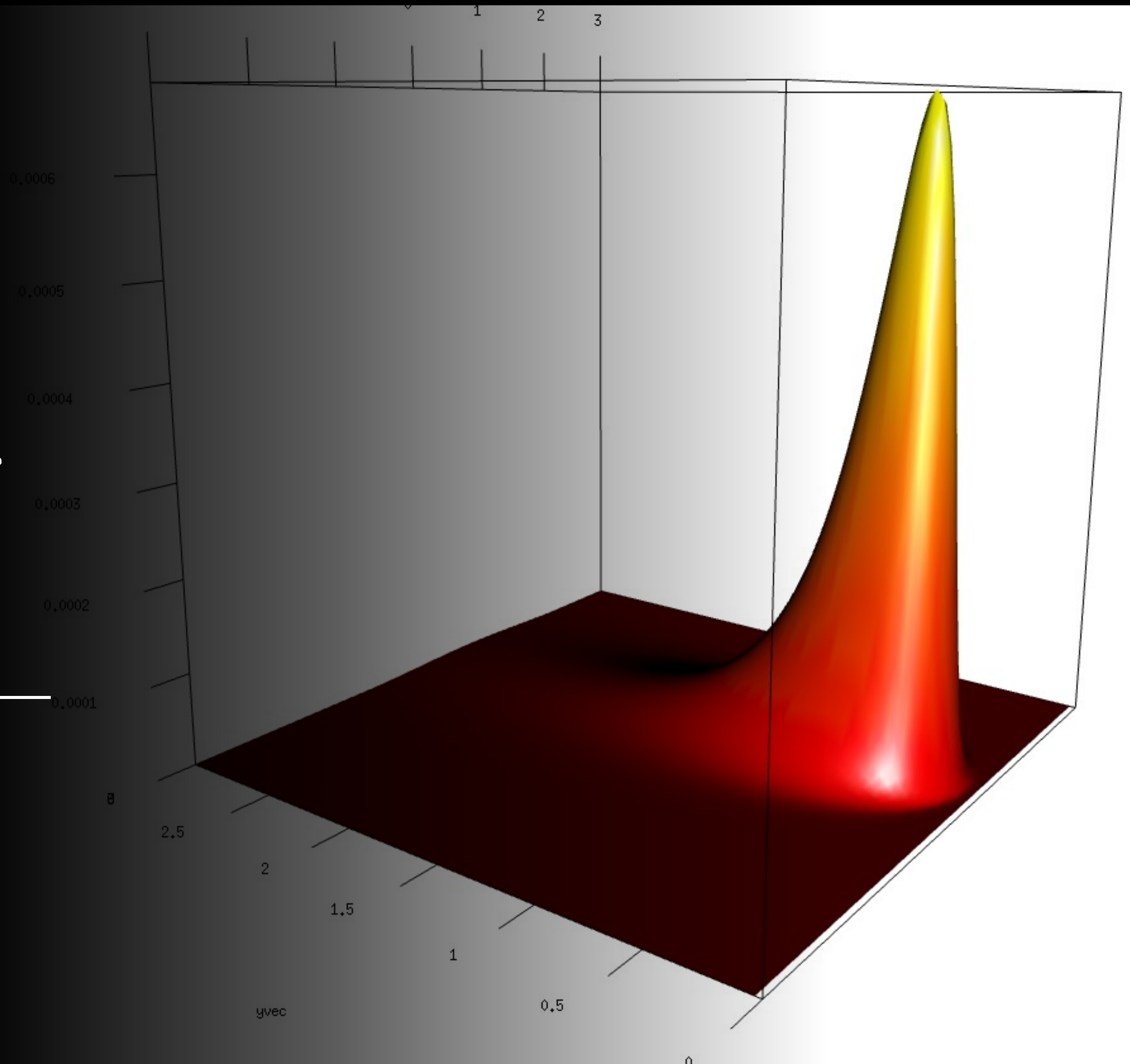


Likelihood (function) – of linear model

Evolution of the likelihood function as we add more data (from 1 to 30 observations)



Estimate parameters of the model



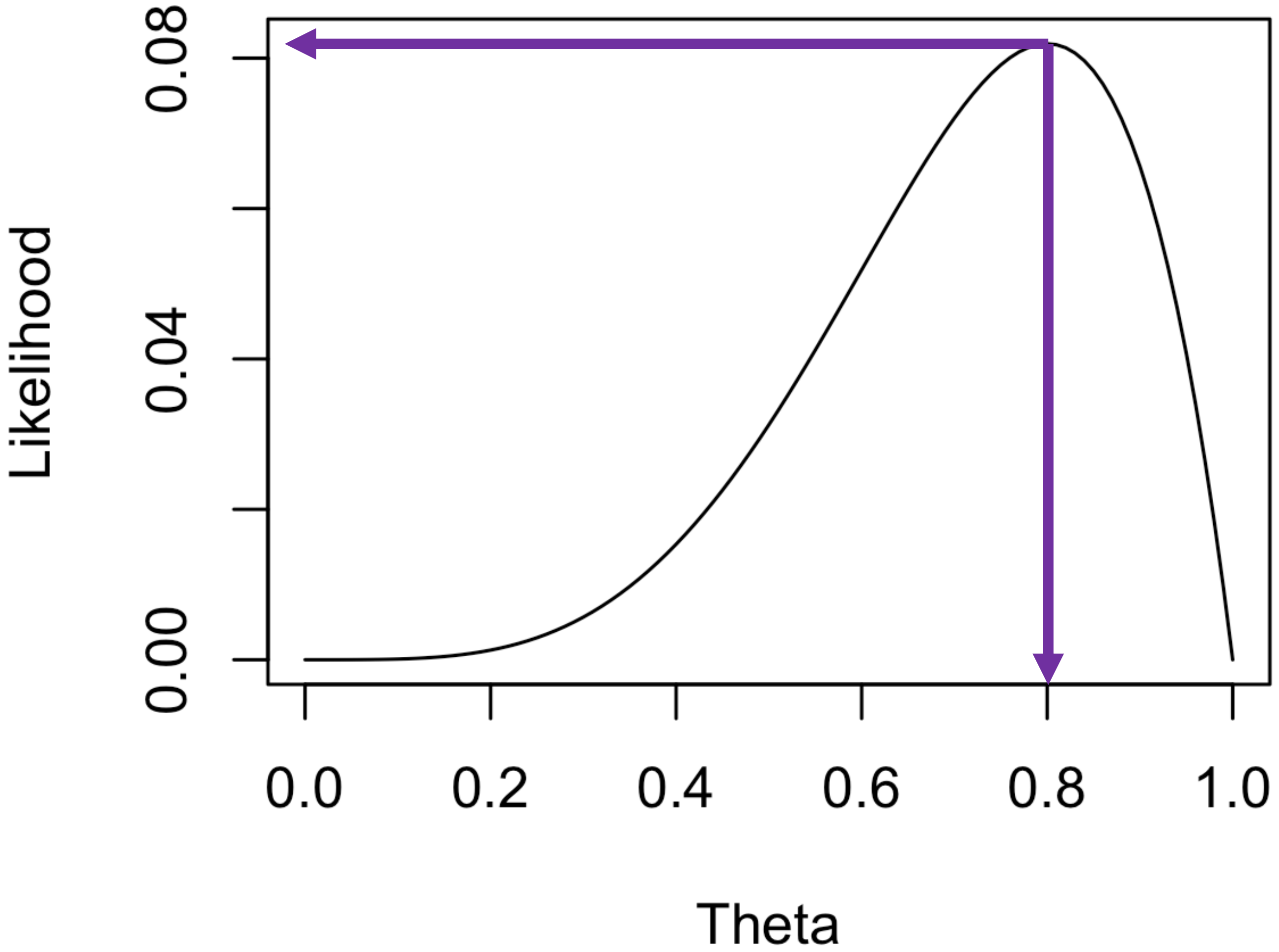
Maximum Likelihood estimate

Parameter that maximises the probability of the observed data

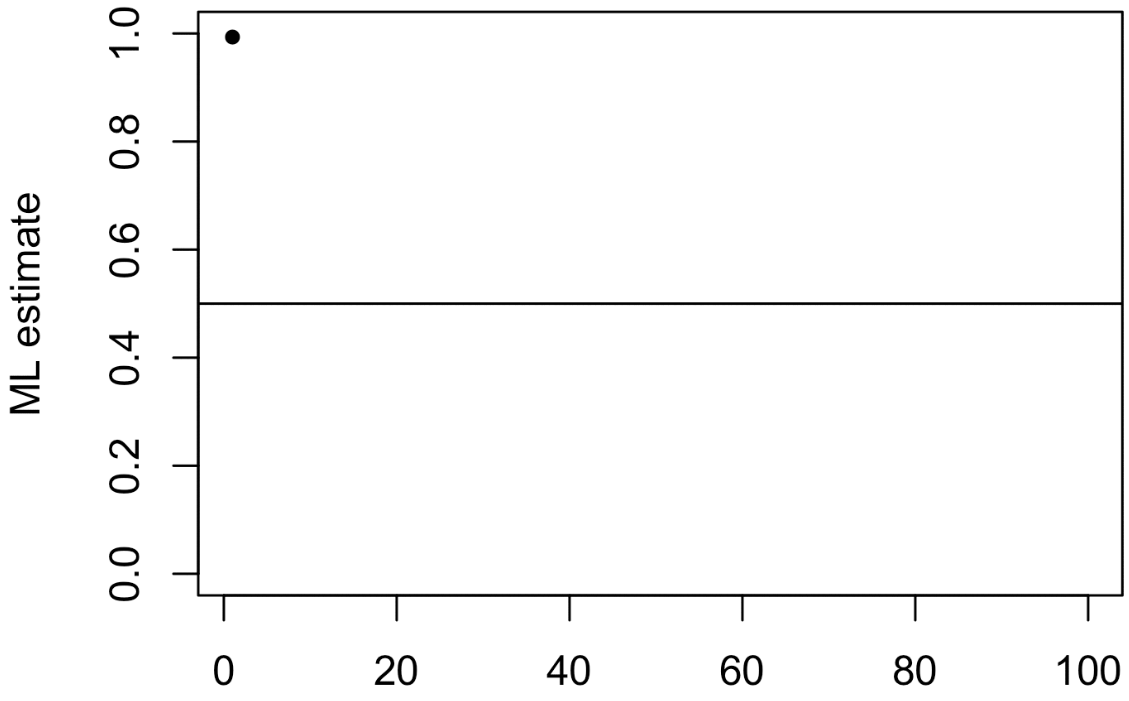
$\theta: probability\ of\ heads$

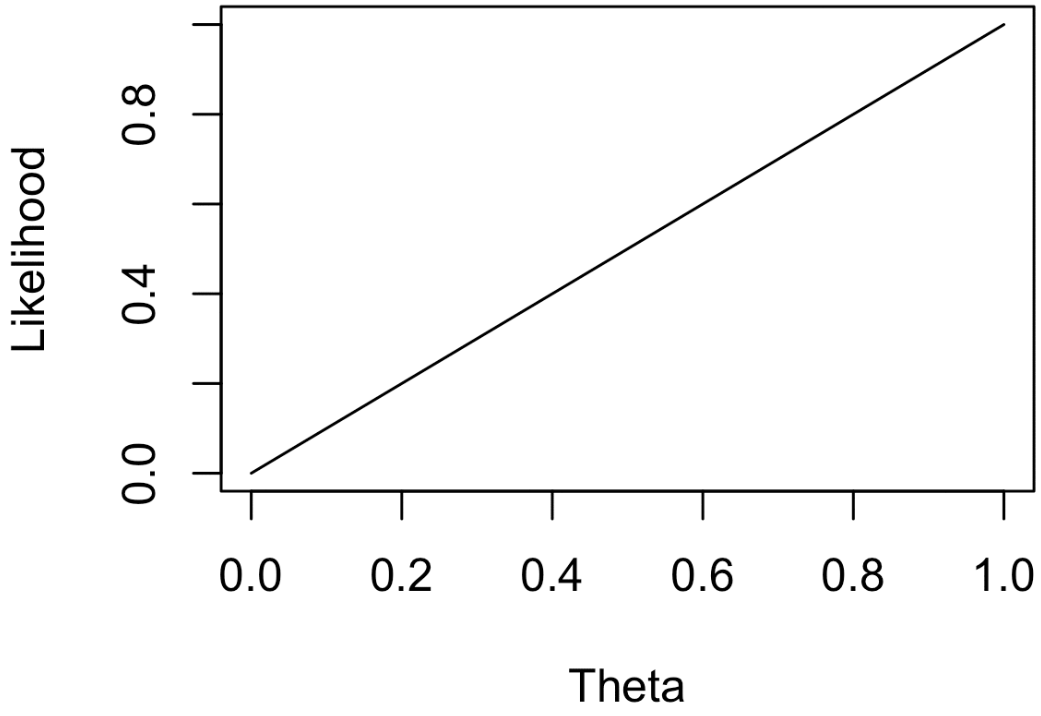
$x : heads, heads, heads, heads, tails$

$\mathcal{L}(\theta \mid x) = \theta.\, \theta.\, \theta.\theta\, .(1-\theta)$



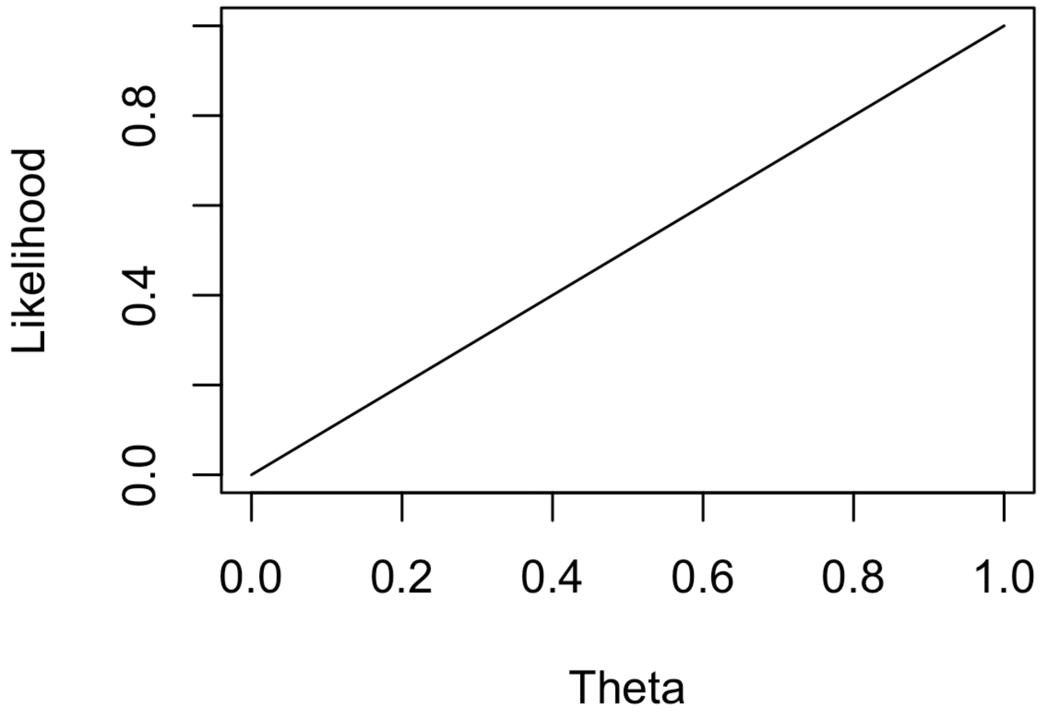
Maximum Likelihood estimate

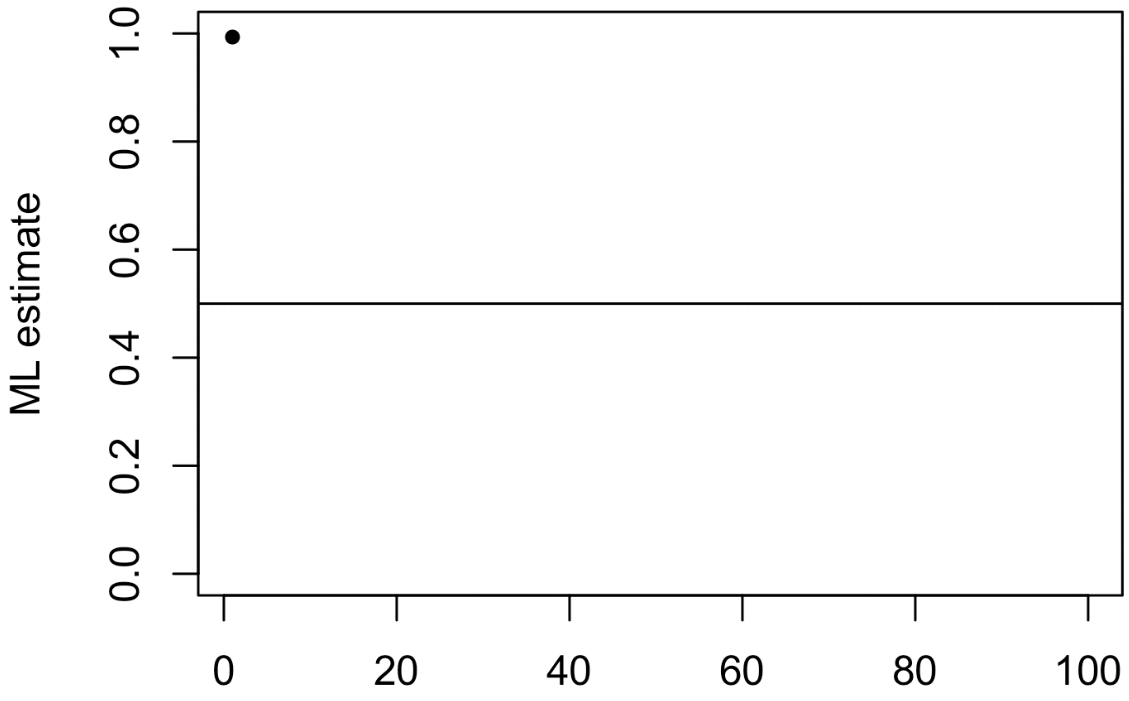




For each new data point
The likelihood function gets updated
And the ML estimate gets updated

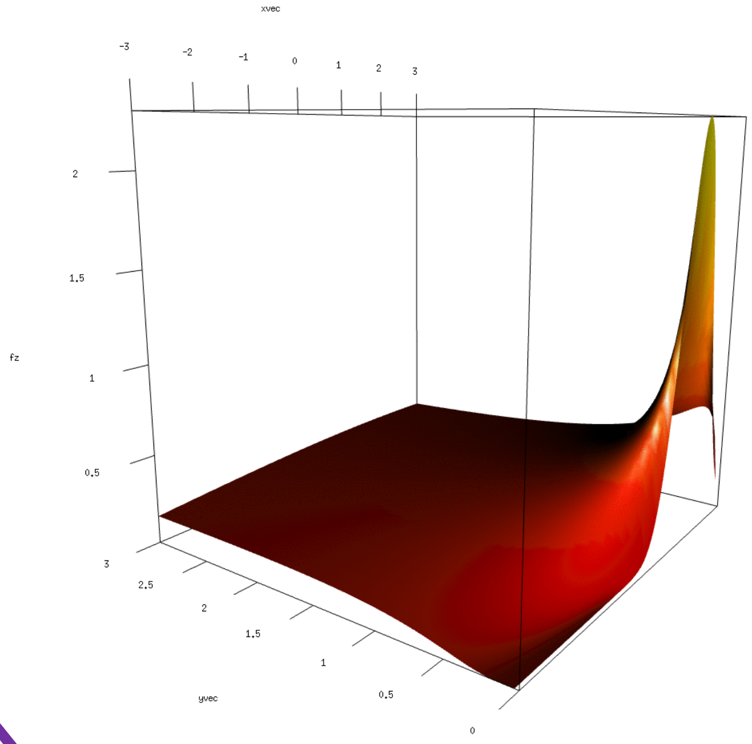
Maximum Likelihood estimate

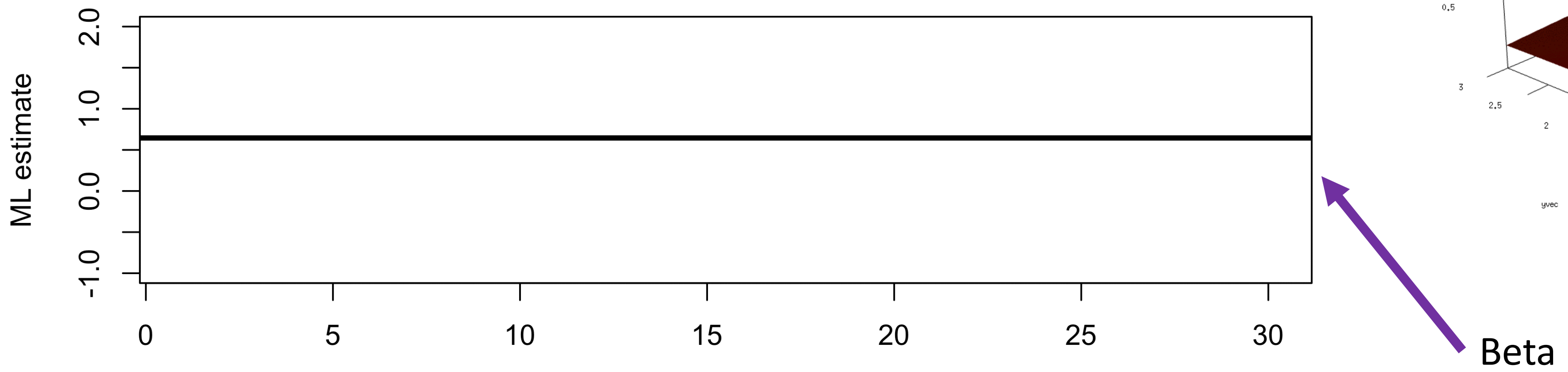


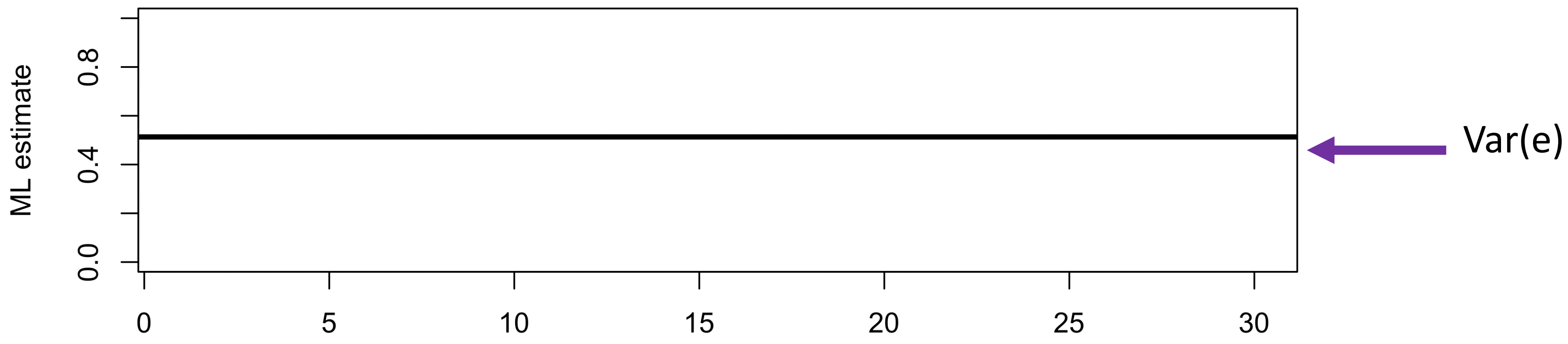


- <u>Asymptotically</u> unbiased
- Consistent
- Efficient
- Scale Invariant
- Sampling distribution of estimates is asymptotically normal

Maximum Likelihood estimate







Optimization

Maximum likelihood estimates can sometimes be solved in closed form

MLE of coin toss = Number of heads / number of tosses

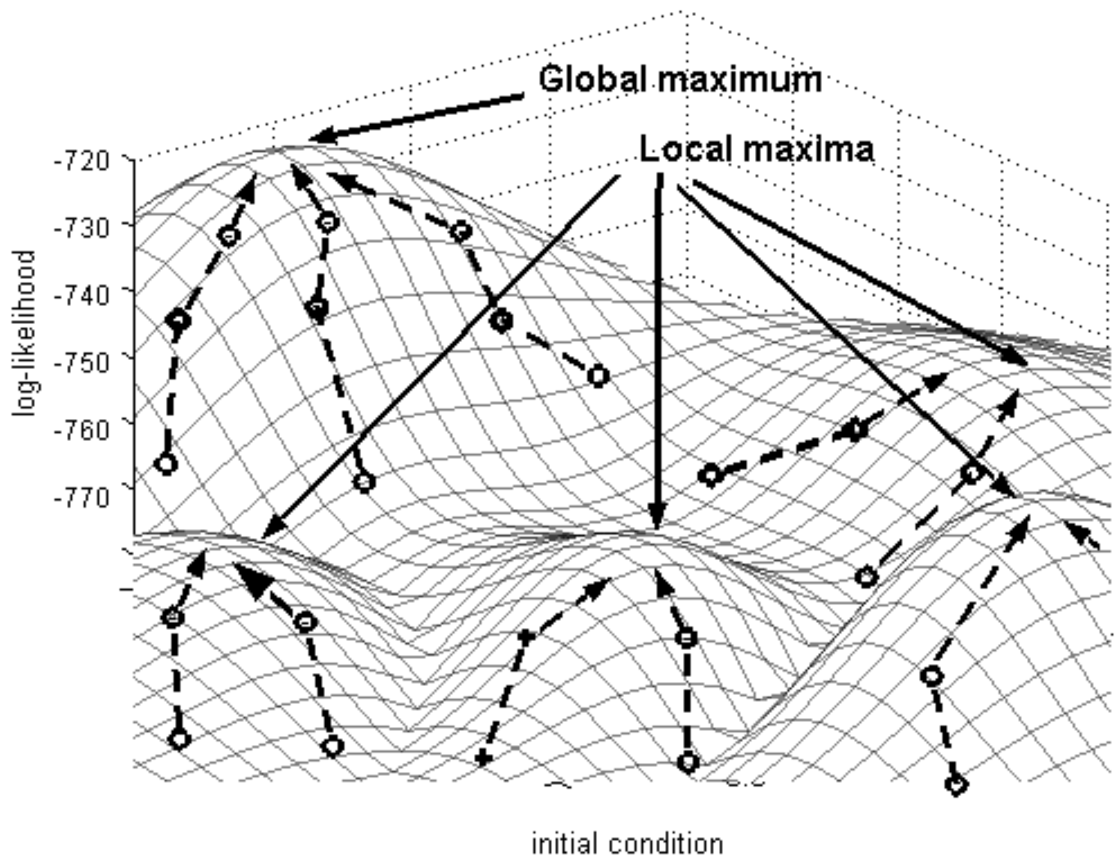
MLE of linear regression :

$$\hat{\beta}_N = (X^\top X)^{-1} X^\top y$$

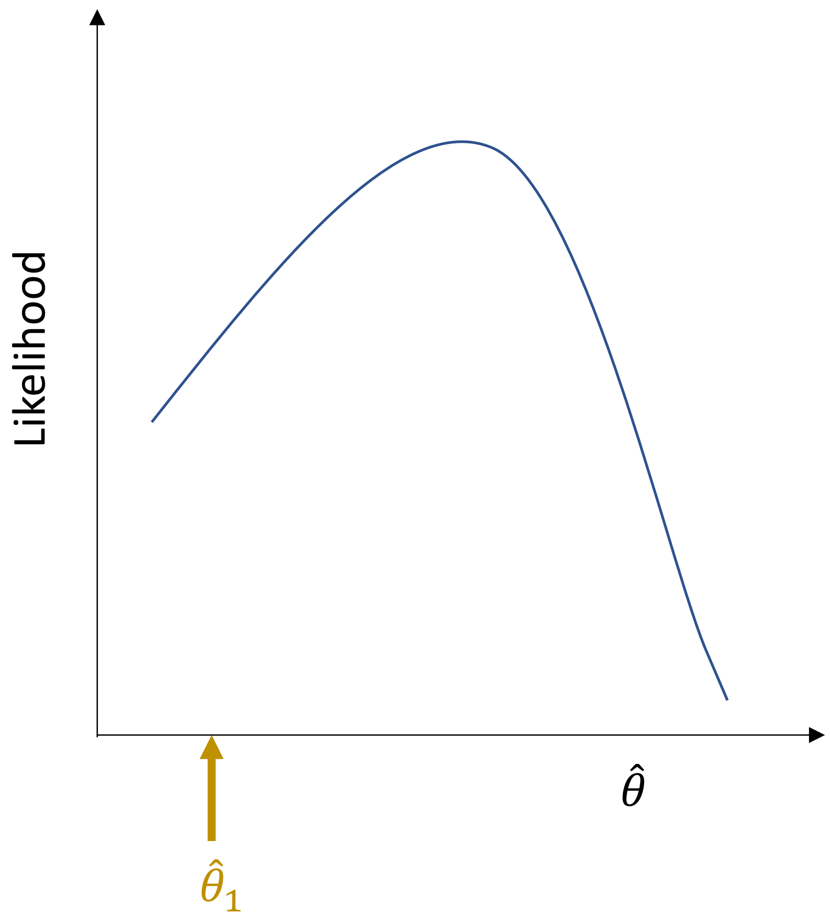
$$\hat{\sigma}_N^2 = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - x_i \hat{\beta}_N \right)^2$$

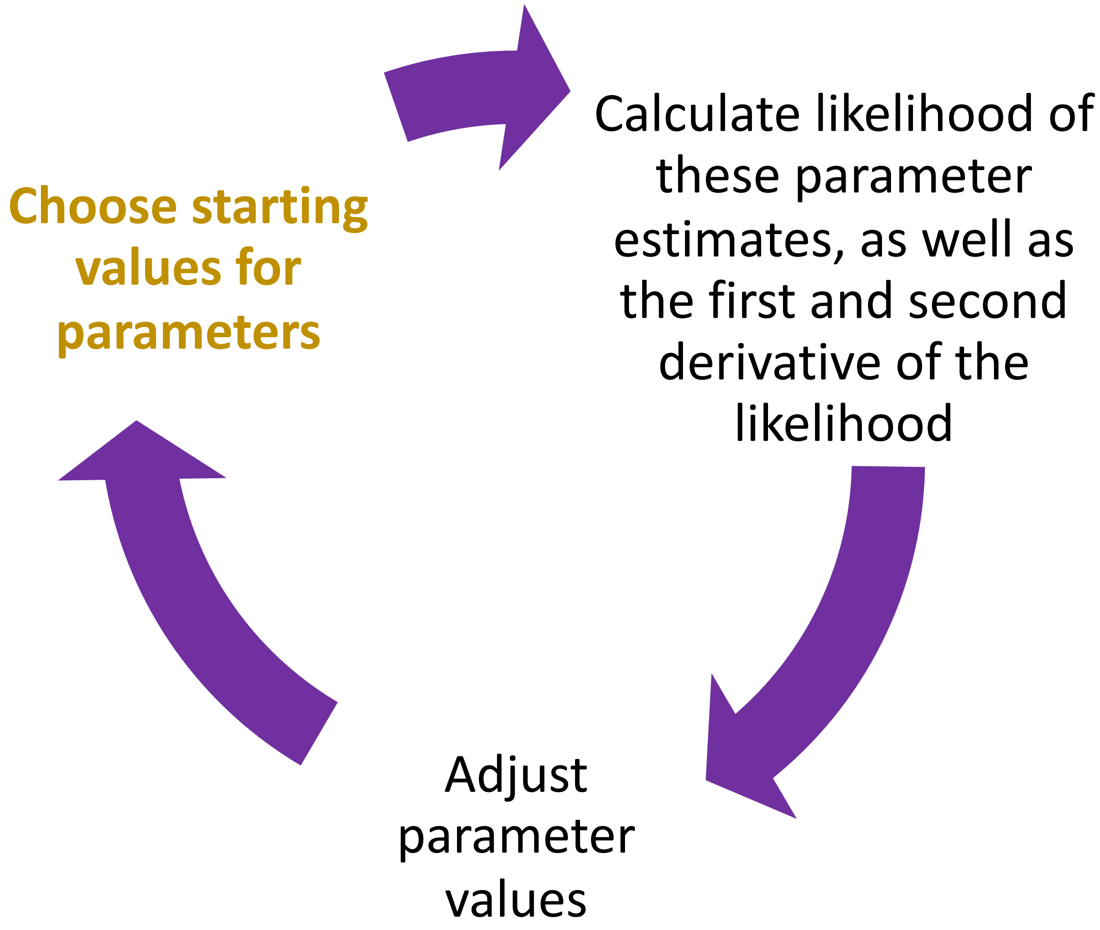
Optimization

For more complex models solutions can rarely be solved in closed form - rather iterative optimization procedures are commonly needed

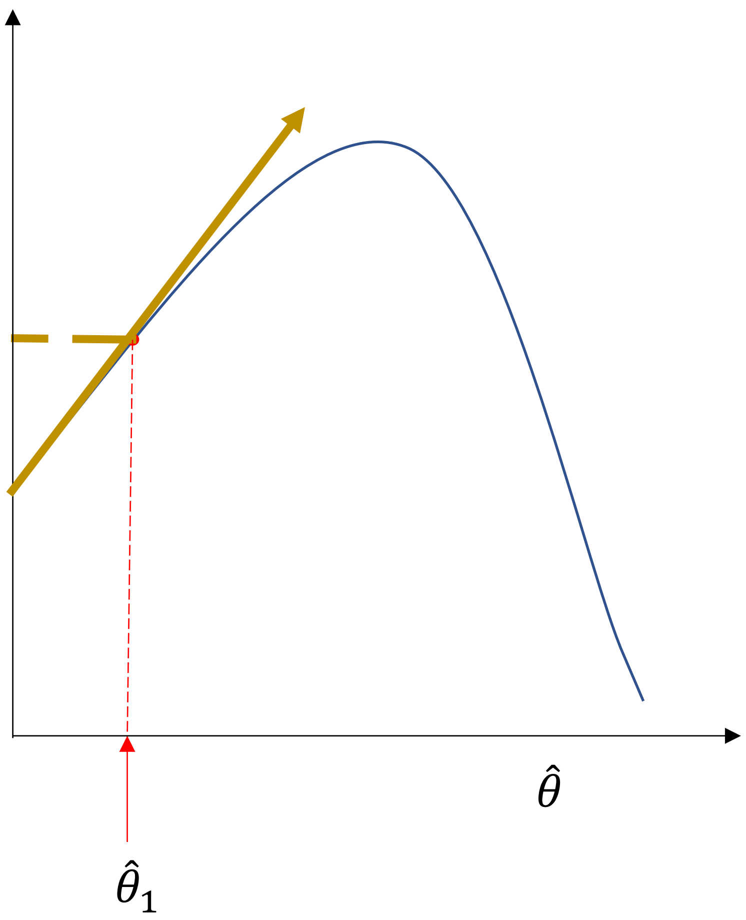


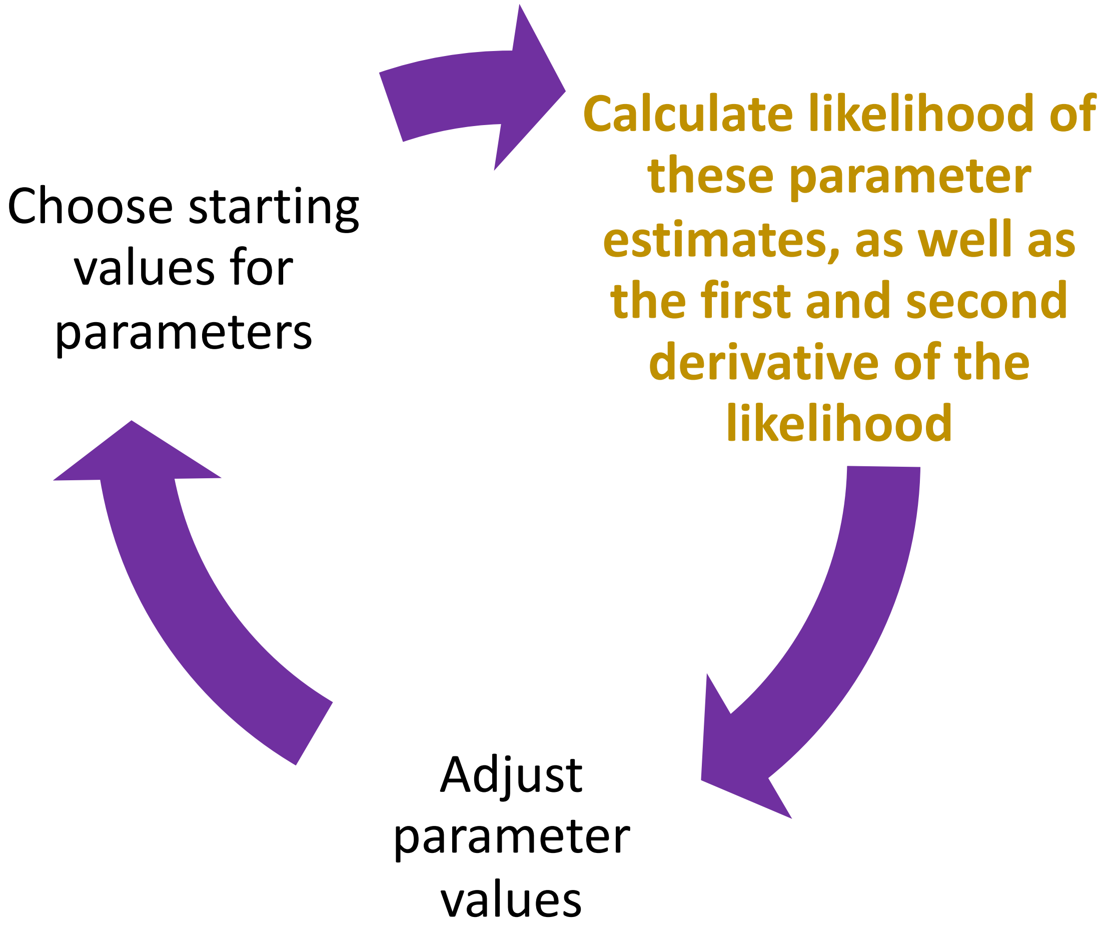
Optimization



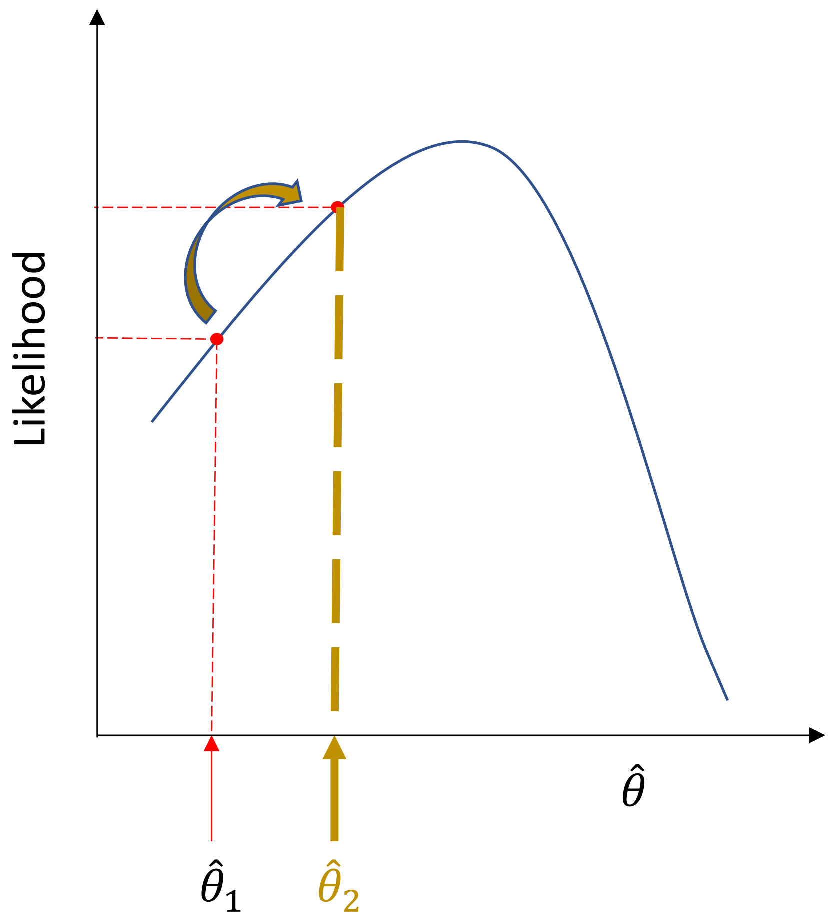


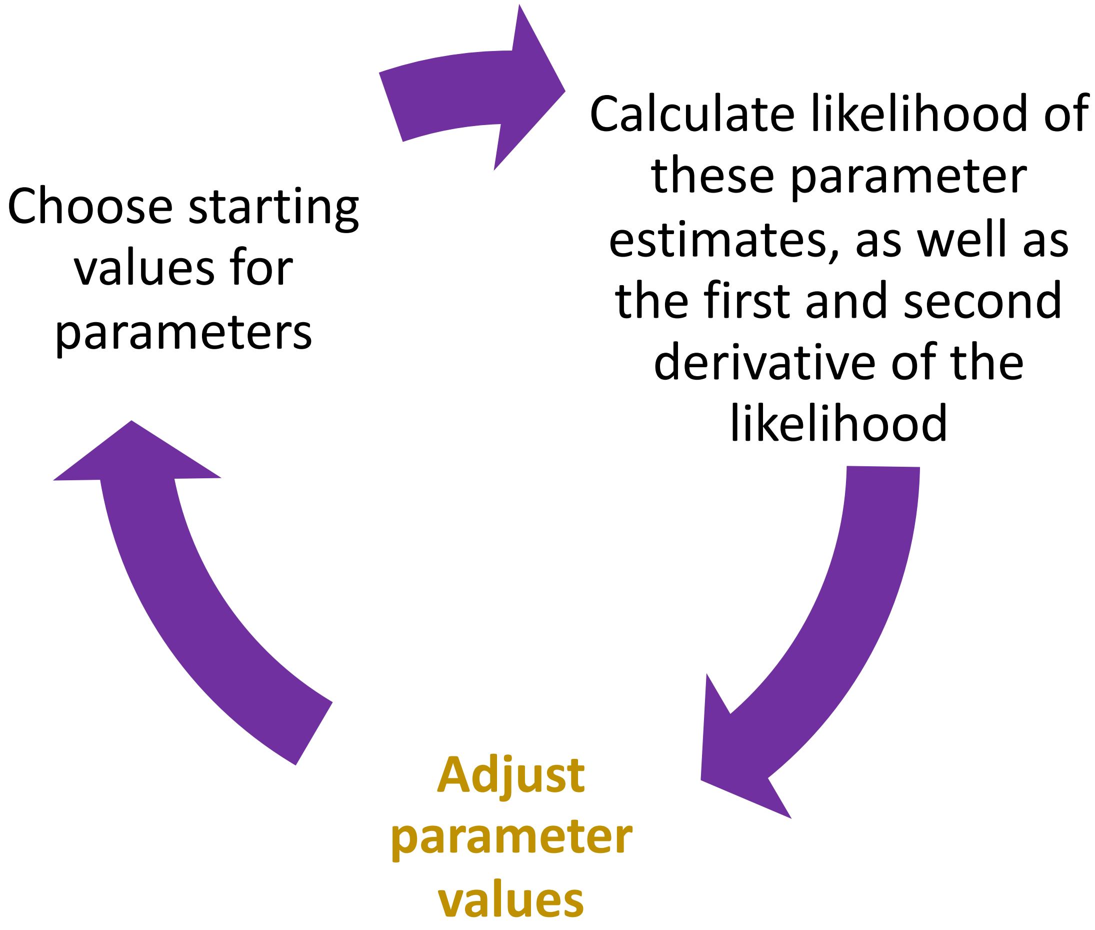
Optimization



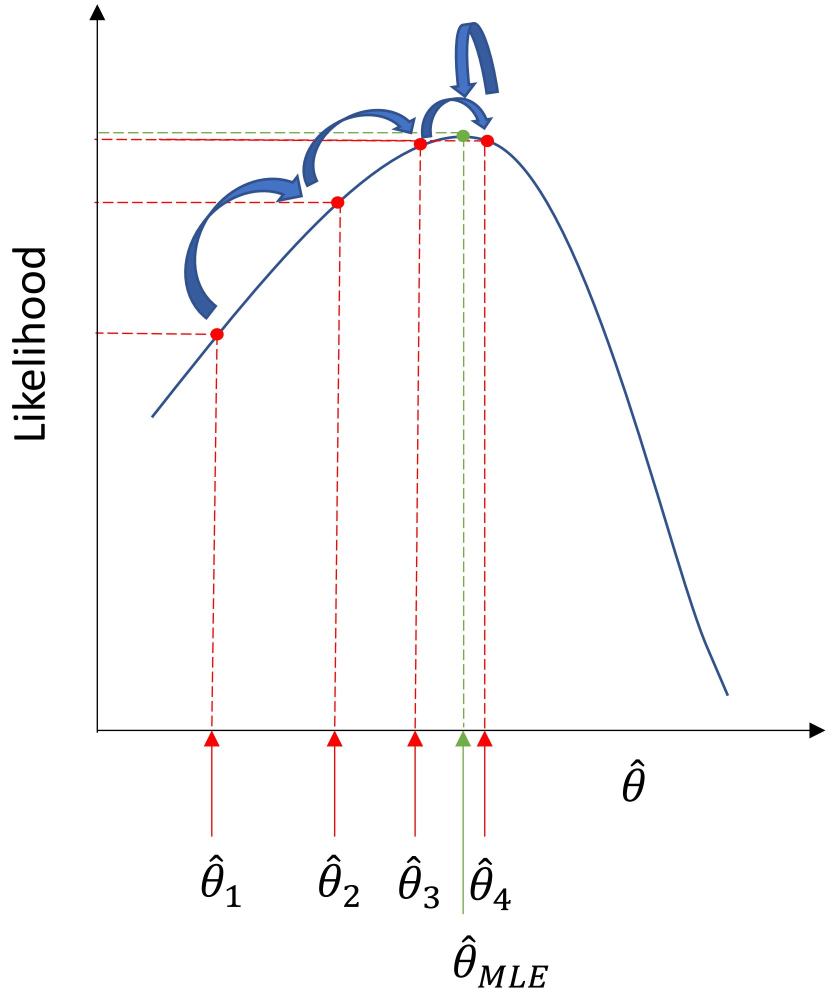


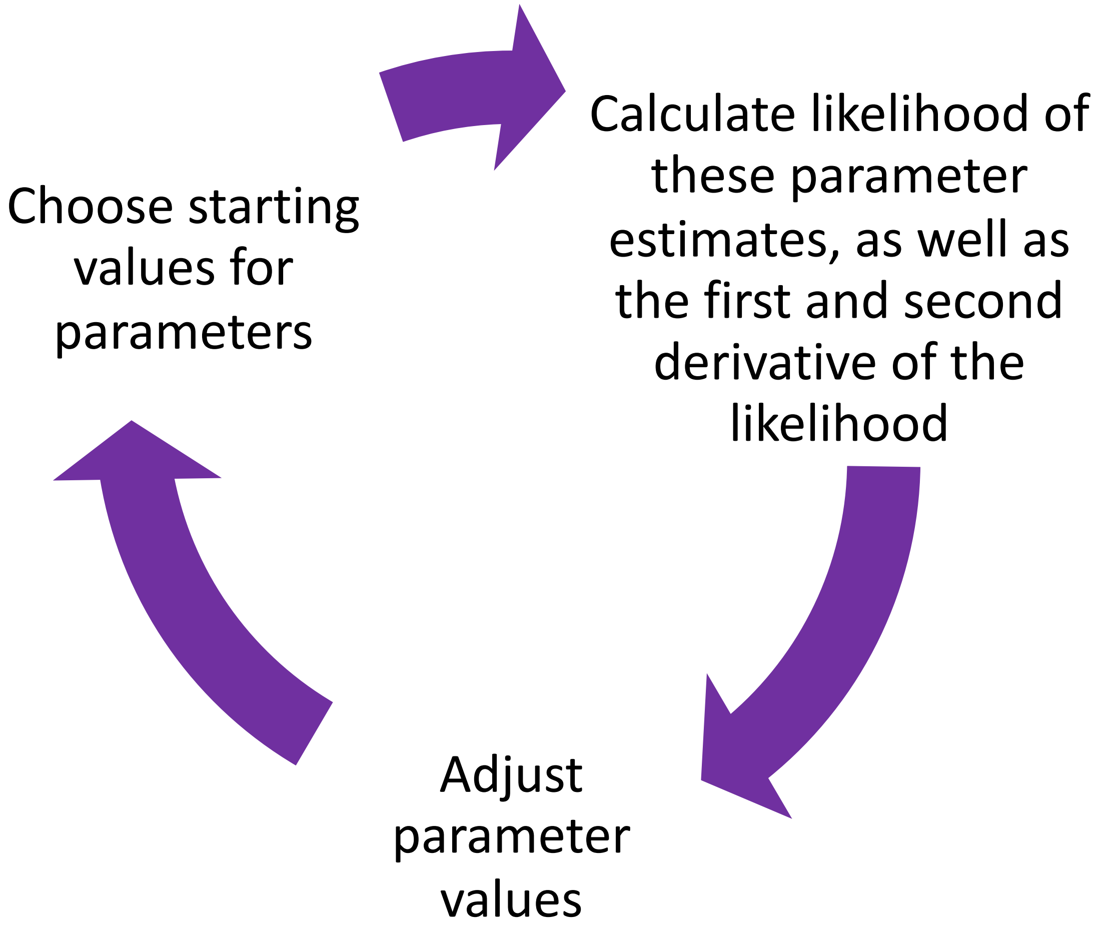
Optimization





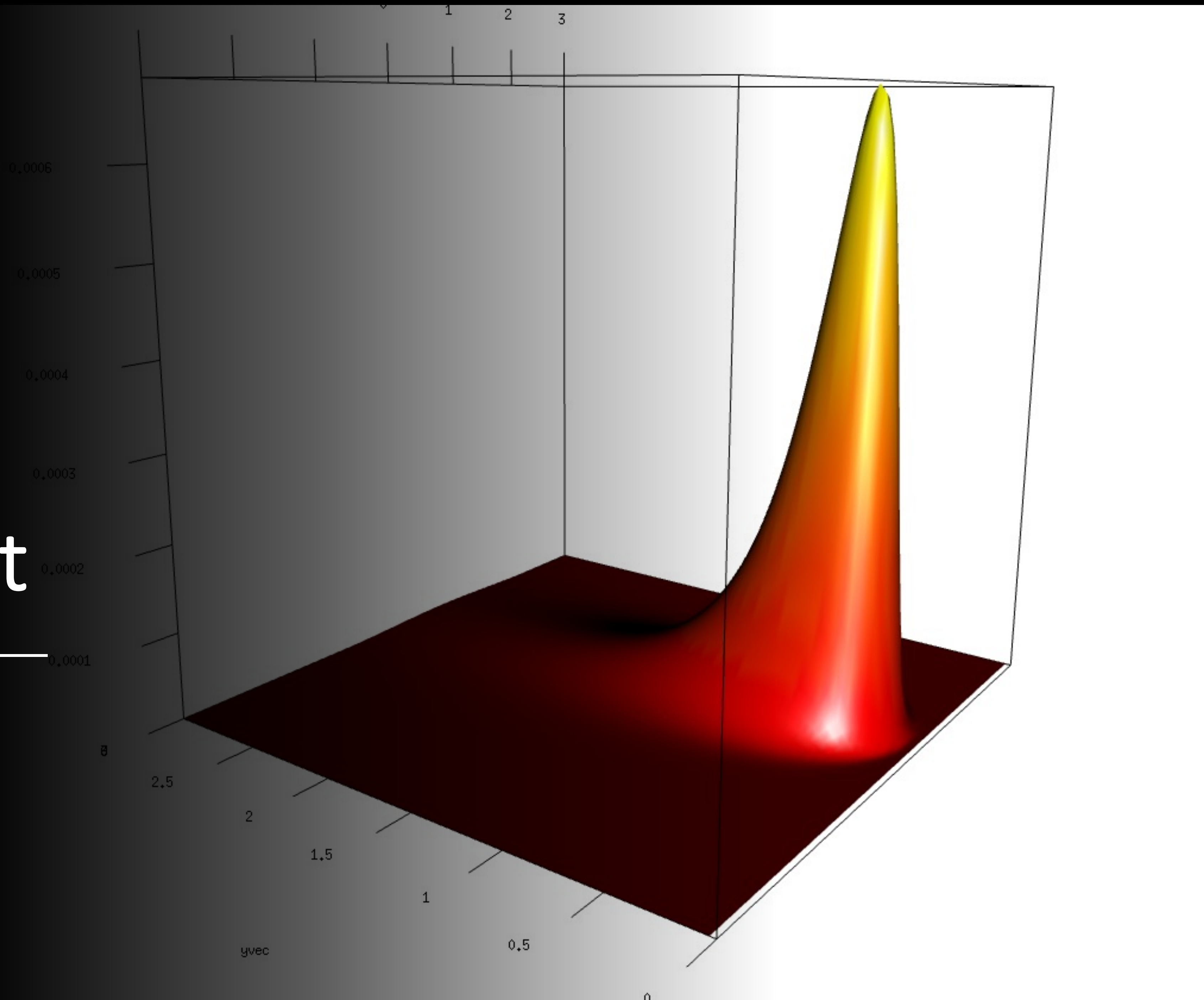
Optimization





Repeat process until stopping criterion is reached

Significance testing & goodness of fit



Likelihood ratio test

- Twice the difference in log-likelihood between **nested** models is distributed as chi-square

$$\lambda_{\text{LR}} = -2\left[\,\ell(\theta_0) - \ell(\hat{\theta})\,\right]$$

e.g. Consider $\boldsymbol{\theta}_F$ = (a, c, e); $\boldsymbol{\theta}_R$ = (a, e, c=0)- twice the difference in log-likelihoods between the models would be distributed as χ^2_1

Model comparison

e.g. ACE vs. CE => significance test of heritability

Goodness of fit

$$AIC = -2ln(\mathcal{L}_{ML}) + 2k$$

Akaike information criterion

$$BIC = -2ln(\mathcal{L}_{ML}) + k\ ln(n)$$

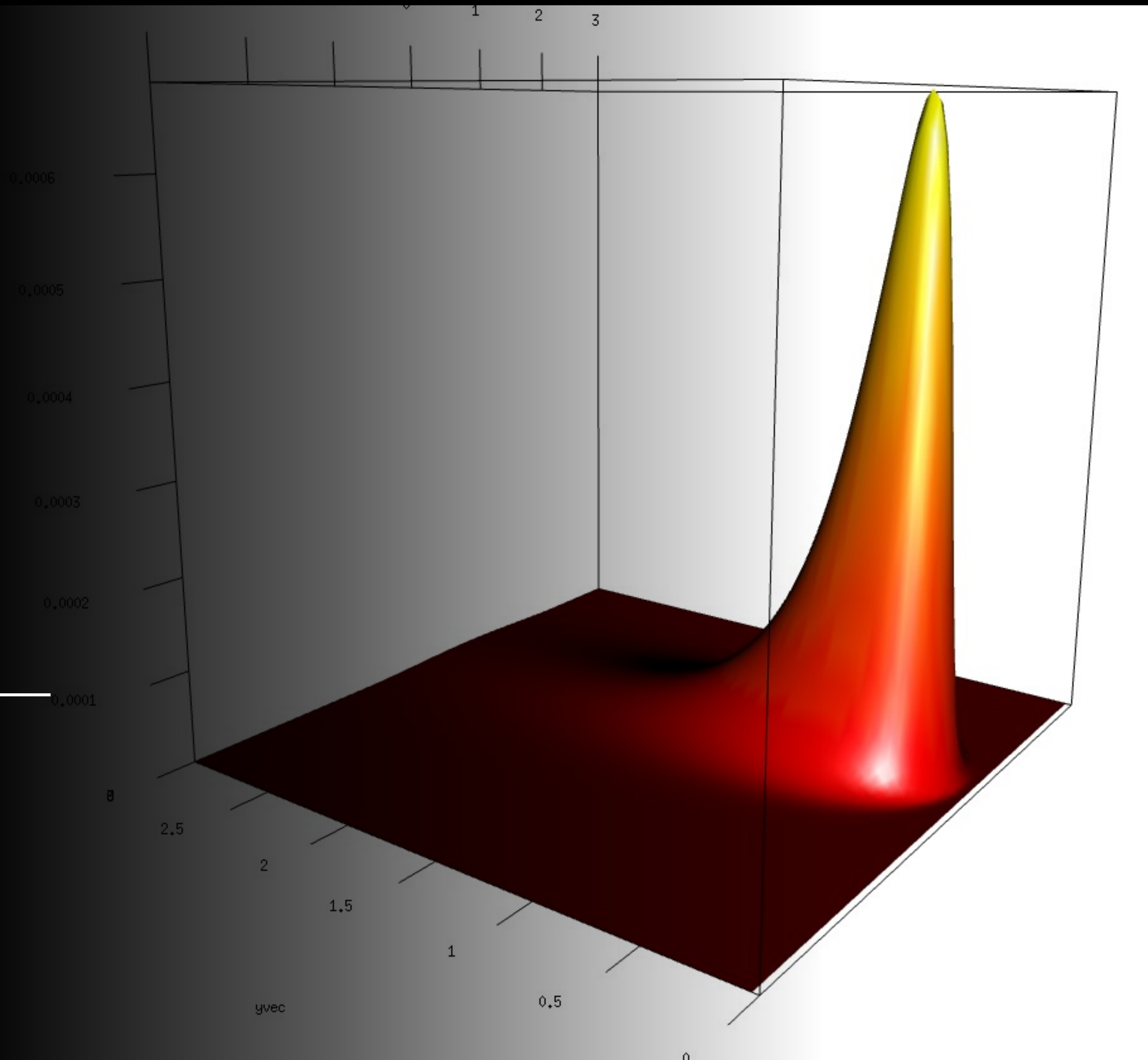
Bayesian Information Criterion

k: number of parameters estimated in the model
n: sample size

- Smaller values of AIC/BIC indicate better fit
- Penalty on number of parameter favours parsimonious models
- Often used to compare models when statistical testing not straighforward (e.g. models not nested ACE vs. ADE)

Precaution: models need to be fitted on the exact same data

Genetic model(s)



ACE model

Covariates (age, sex)

Genetic

Shared environment

Trait of interest

Residuals (incl. environment)

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{a} + \mathbf{c} + \mathbf{e}$$

$$\mathbf{a} \sim N(0, \mathbf{G} \,.\, Va)$$

$$\mathbf{c} \sim N(0, \mathbf{C} \,.\, Vc)$$

$$\mathbf{e} \sim N(0, \mathbf{I} \,.\, Vc)$$

$$Var(Y) = Va + Vc + Ve$$

$Va/(Va+Vc+Ve)$: heritability

$Vc/(Va+Vc+Ve)$: shared E

Main parameters of interest:

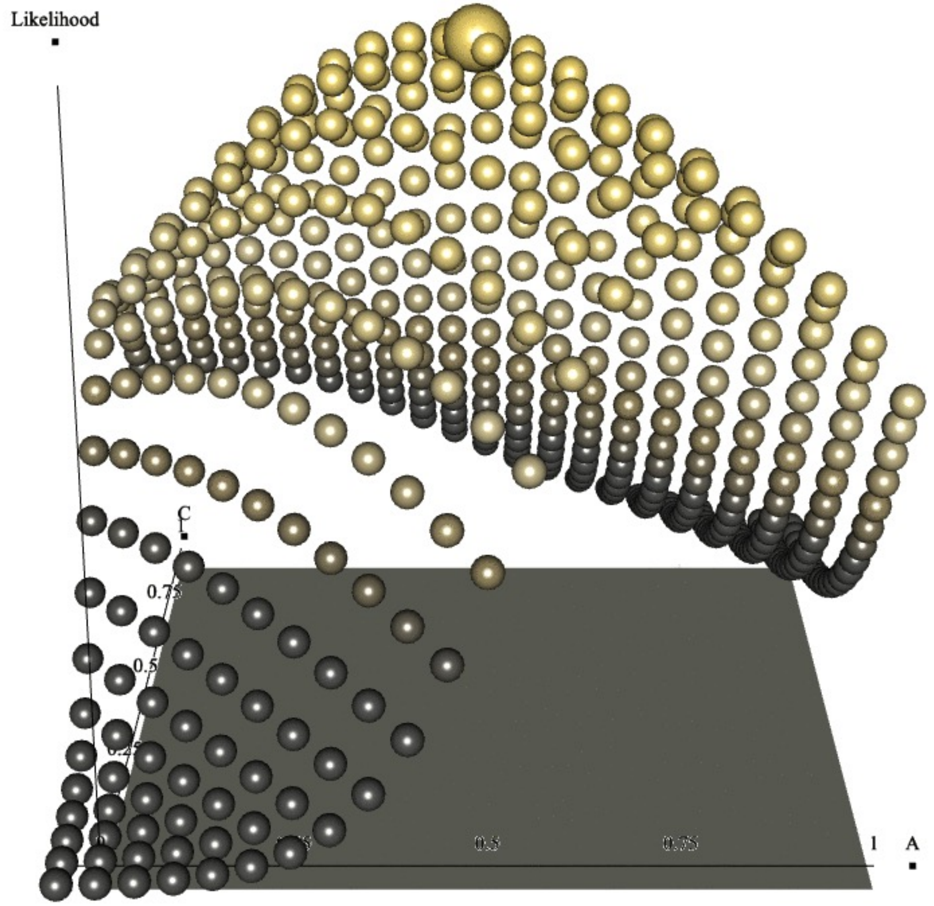
Va and Vc

Likelihood as a function of Va and Vc

"Real data" with 500 MZ + 500 DZ pairs
covariates

Fitted model in OpenMx
Estimated likelihood for a range of set Va and Vc values

```
> fitACE$output$estimate
     interC        betaS        betaA         VA11         VC11         VE11
-3.26259017   0.09401213   0.16099604   0.48645944   0.23707928   0.25814284
```

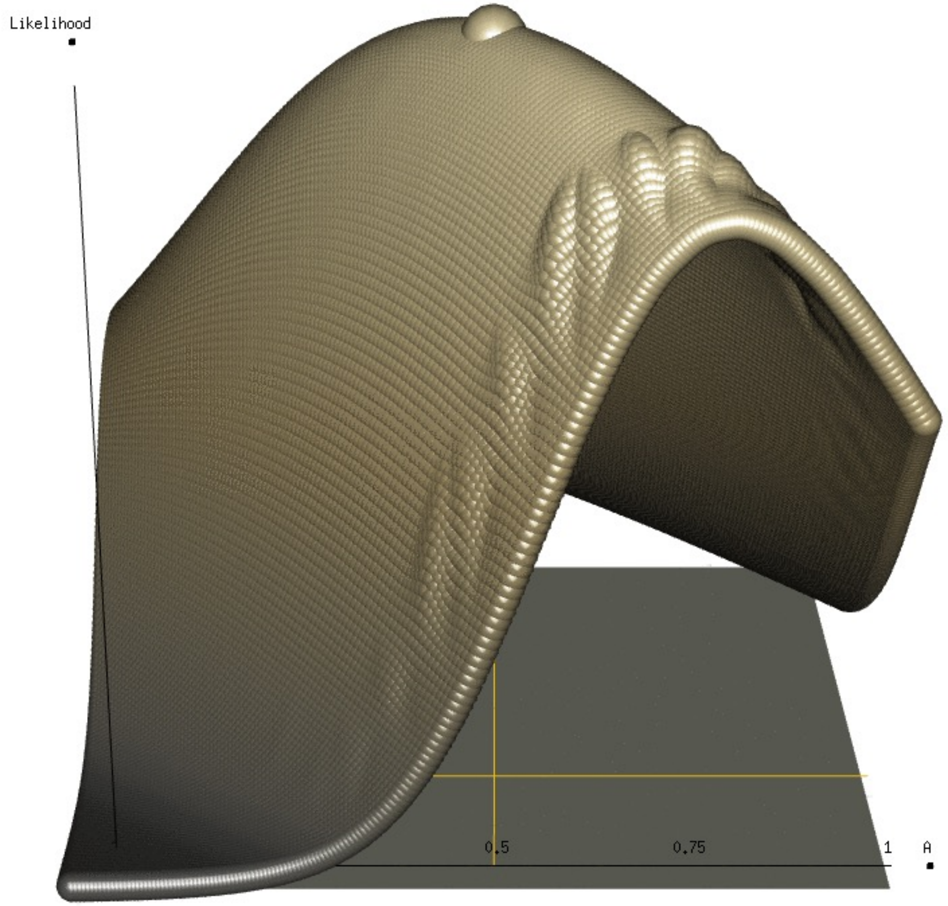


Likelihood (interpolated)

ML estimates :
Va =0.48
Vc = 0.24

Likelihood can be estimated for Va, Vc < 0. But note what happens near boudary of parameter space

```
> fitACE$output$estimate
     interC       betaS       betaA        VA11        VC11        VE11
-3.26259017  0.09401213  0.16099604  0.48645944  0.23707928  0.25814284
```

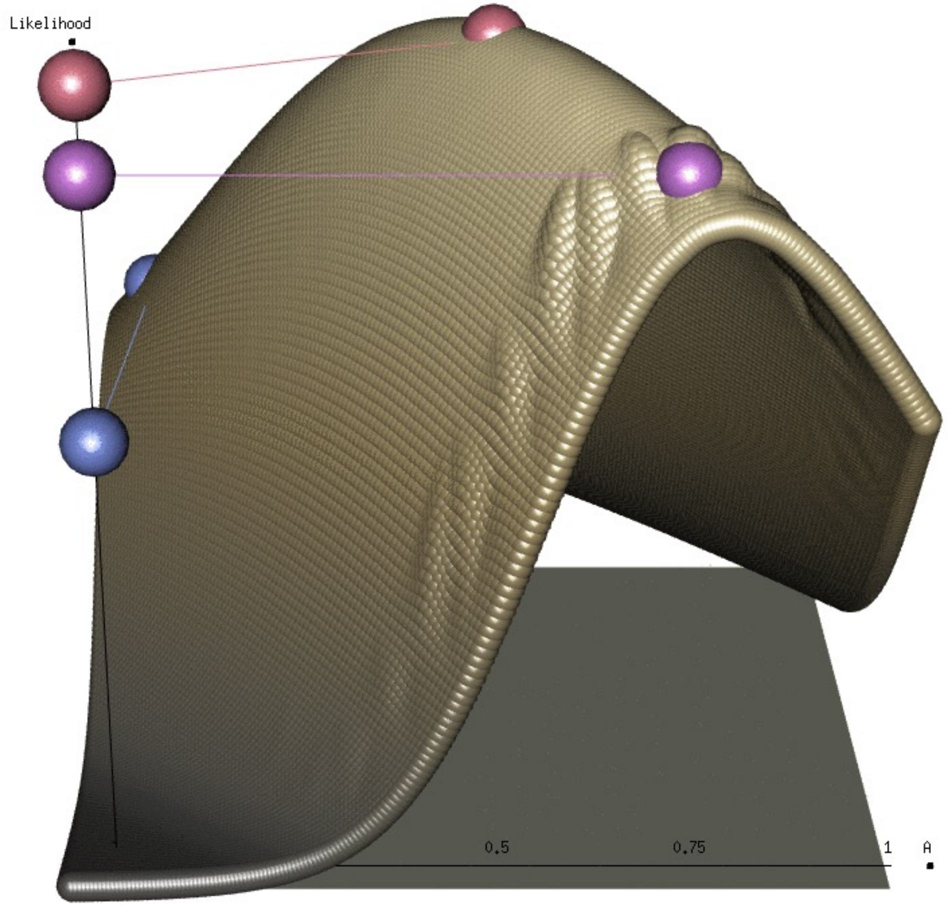


Likelihood ratio test

ACE model
AE model
CE model

Test statistic : twice the difference of log-likelihoods

```
> mxCompare(fitACE, c(fitAE, fitCE))
   base comparison ep  minus2LL   df       AIC     diffLL diffdf            p
1 ACEvc       <NA>  6 10220.367 3994 10232.367         NA     NA           NA
2 ACEvc         AE  5 10242.711 3995 10252.711  22.343713      1 2.2795805e-06
3 ACEvc         CE  5 10334.400 3995 10344.400 114.032618      1 1.2818252e-26
```

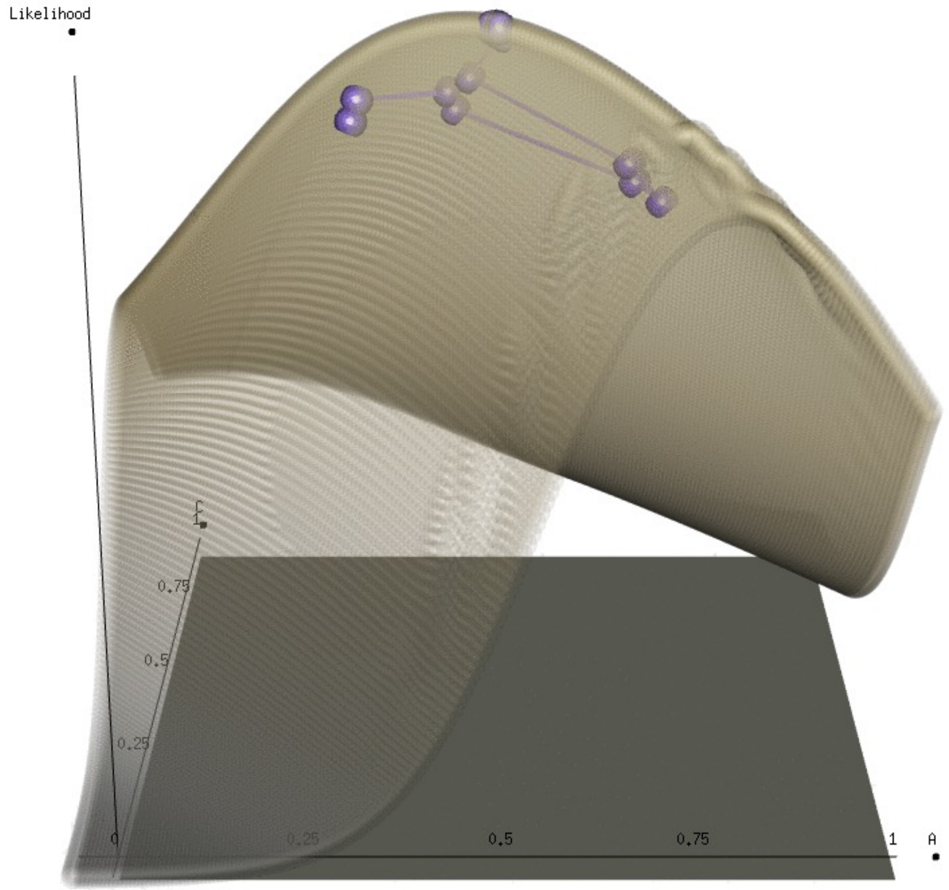


Optimization

SLSQP optimizer

Started at
Vc=Va=0.3

Found ML in 18
interations



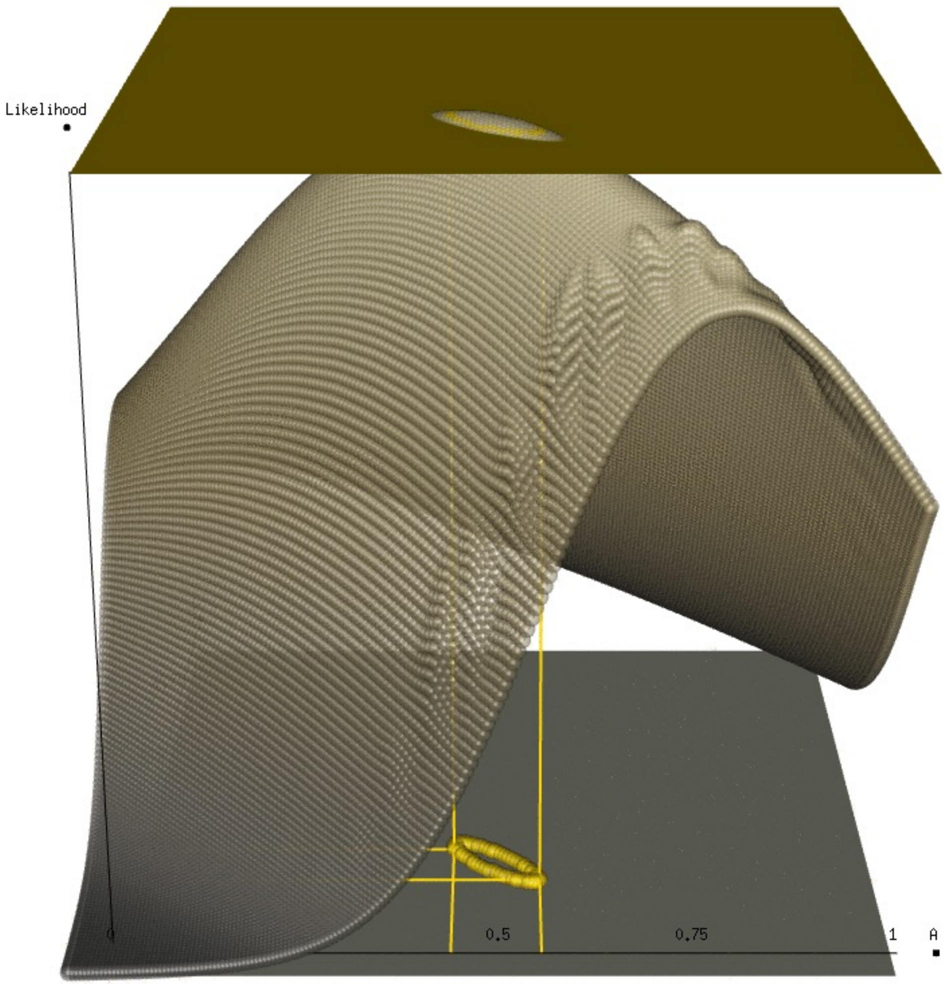
Confidence intervals

Start from maximum likelihood

Degrade (lower) the likelihood so that difference is significant (chi2 test) at 1-CI

For 95% CI : $chi^2 = 3.84$ ⬄ pvalue=0.05

```
> fitACE$output$confidenceIntervals
                    lbound  estimate    ubound
ACEvc.VarC[1,4] 0.4005072 0.4955369 0.5960494
ACEvc.VarC[1,5] 0.1466367 0.2415032 0.3290960
```

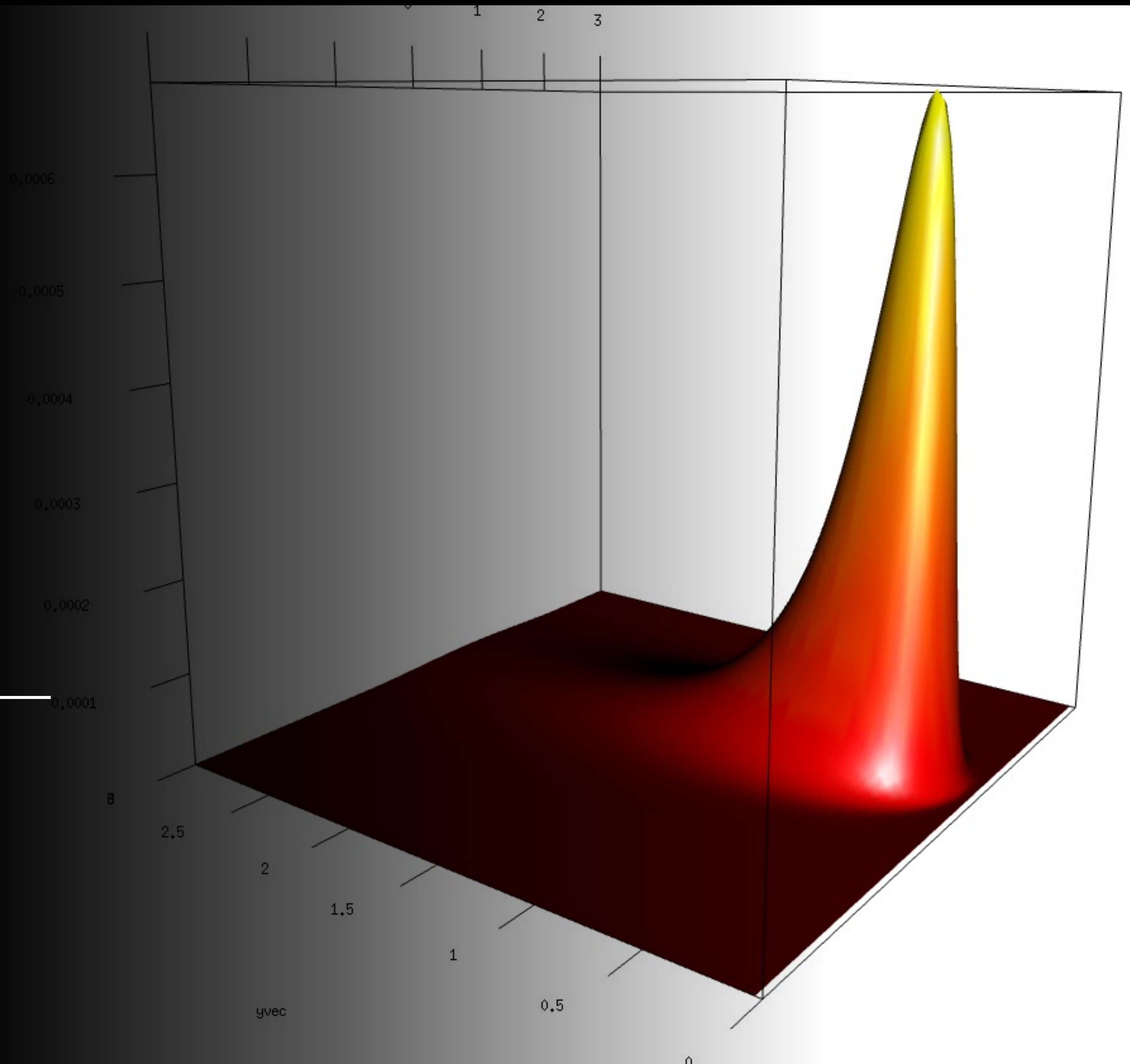


ML, FIML, REML

ML: Maximum likelihood	Fine for fixed effect models
FIML: Full Information Maximum Likelihood	Handles missing values
REML: Restricted Maximum Likelihood	Minimises bias in variance estimation of mixed models

Also pseudo likelihood, or quasi-likelihood..

A simple example



A coin toss experiment

p: *probability of heads*

x : *heads, heads, heads, heads, tails, tails, heads, heads, heads, heads*

Q1: can you write the likelihood function (as a function of p)?

$$\mathcal{L}(p \mid x) = \binom{10}{8} p^8 (1-p)^2$$

$$\binom{10}{8} = \frac{10!}{8!\,(2)!} = \frac{9 \times 10}{2} = 45$$

A coin toss

p: $probability\ of\ heads$

x : $heads, heads, heads, heads, tails, tails, heads, heads, heads, heads$

Q2: what is the maximum likelihood estimate of p?

$$\widehat{p_{ML}} = \frac{8}{10} = 0.8$$

A coin toss

p: $probability\ of\ heads$

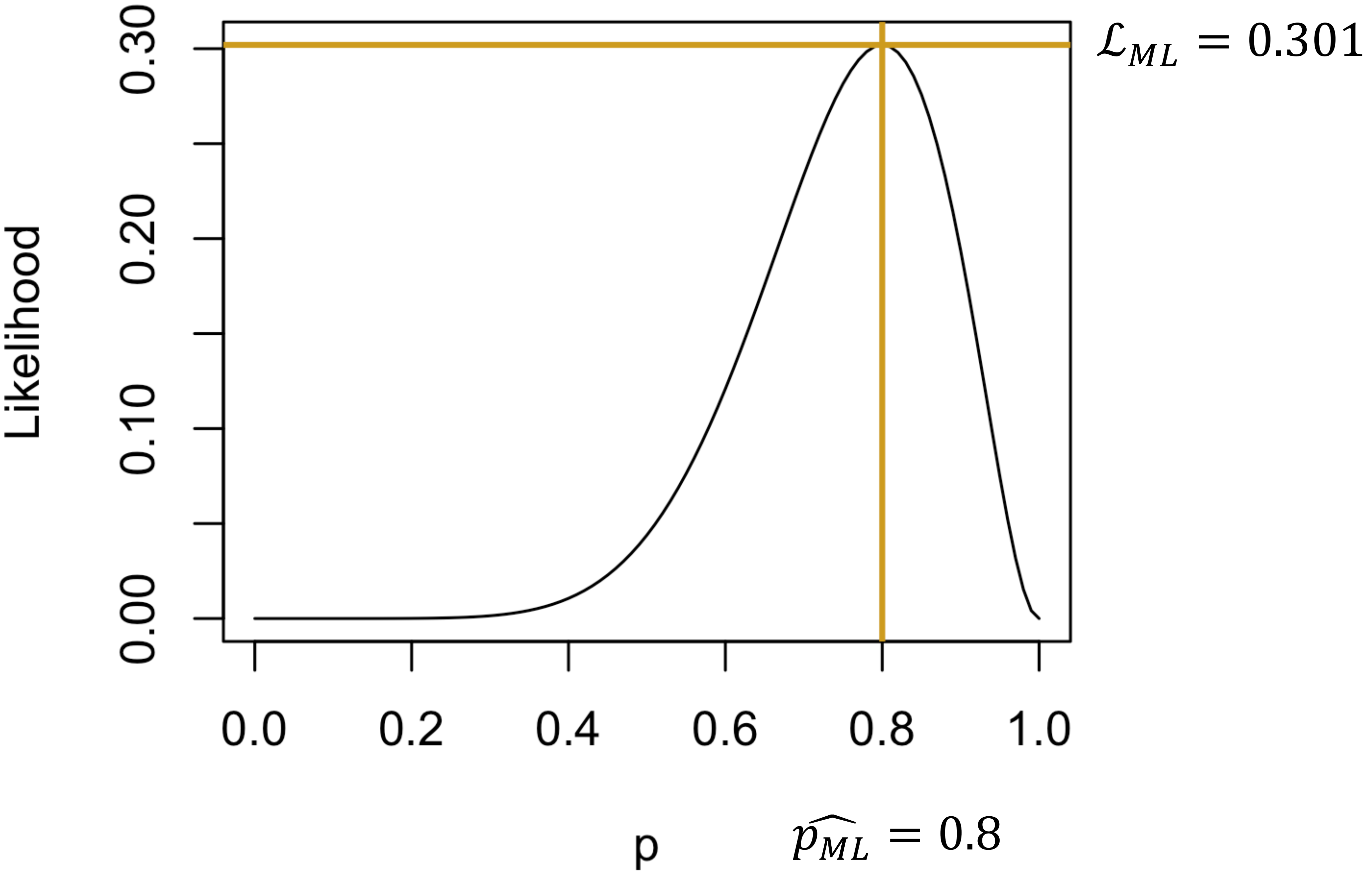
x : $heads, heads, heads, heads, tails, tails, heads, heads, heads, heads$

Q3: what is the maximum likelihood – i.e. value of the likelihood function at its maximum?

$$\mathcal{L}(p \mid x) = \binom{10}{8} p^8 (1-p)^2$$

$$\widehat{p_{ML}} = \frac{8}{10} = 0.8$$

$$\mathcal{L}_{ML} = 45\,.\,0.8^8.\,(1-0.8)^2 = 0.301$$



$\mathcal{L}_{ML} = 0.301$
Likelihood
0.00
0.10
0.20
0.30
0.0
0.2
0.4
0.6
0.8
1.0
p
$\widehat{p_{ML}} = 0.8$

A coin toss

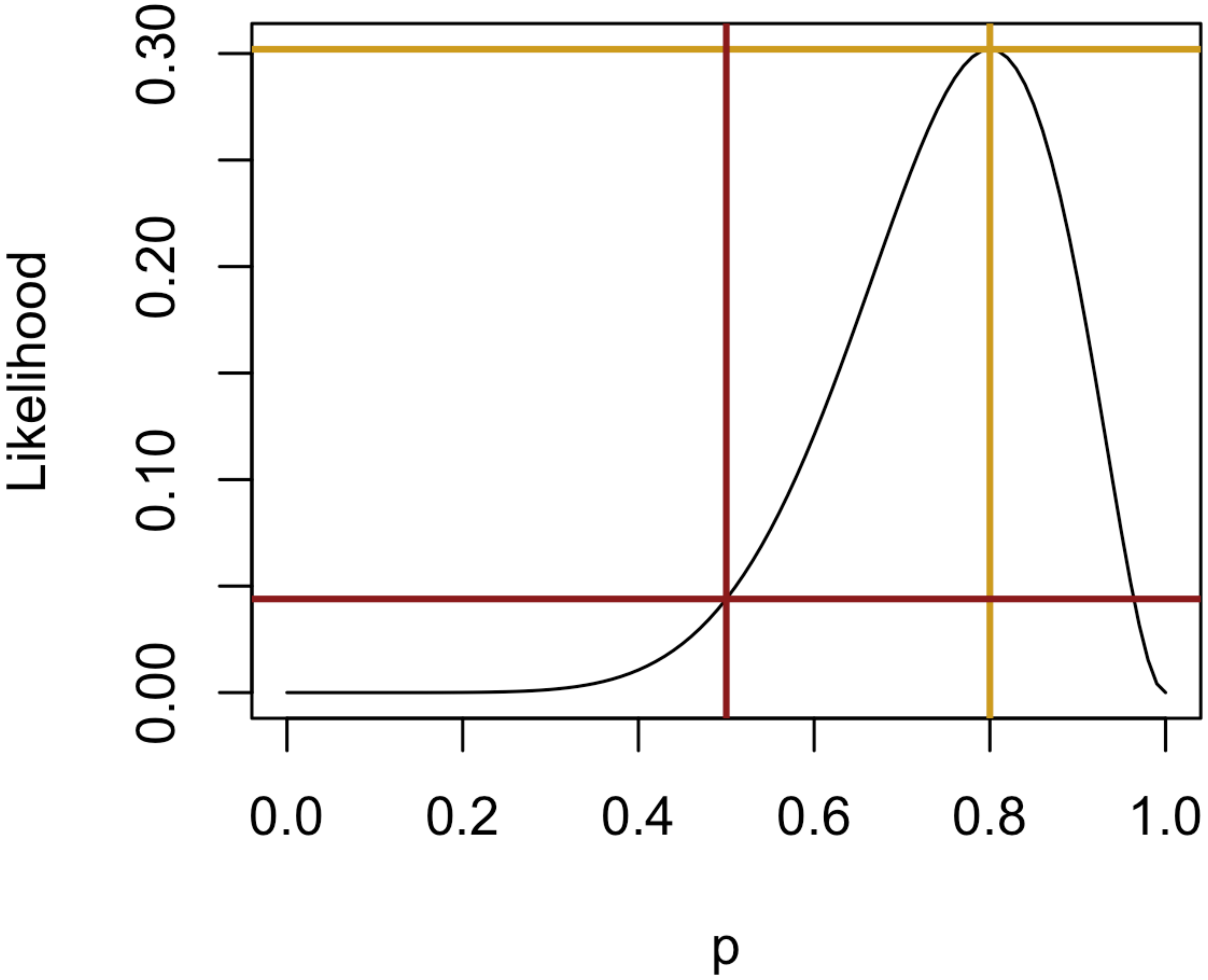
p: $probability\ of\ heads$

x : $heads, heads, heads, heads, tails, tails, heads, heads, heads, heads$

Q4: what is the likelihood estimate of the null model : "H0: coin is fair"

$$\mathcal{L}(p \mid x) = \binom{10}{8} p^8 (1-p)^2$$

$$\mathcal{L}_0 = 45\,.\,0.5^8.\,(1-0.5)^2 = 0.044$$



$\mathcal{L}_{ML} = 0.301$

$\mathcal{L}_{0} = 0.044$

A coin toss

Q5: what is the likelihood ratio test statistic, to test if coin probability is different from H0 (fair)?

$$\mathcal{L}_0 = 45\,.\,0.5^8.\,(1-0.5)^2 = 0.044$$

$$\mathcal{L}_{ML} = 45\,.\,0.8^8.\,(1-0.8)^2 = 0.301$$

$$\chi^2 = -2.\,(\ln(\mathcal{L}_0) - \ln(\mathcal{L}_{ML}))$$

A coin toss

Q6: Is the test significant ?

$$\chi^2 = -2.(\ln(\mathcal{L}_0) - \ln(\mathcal{L}_{ML})) = 3.85$$

```
> 1-pchisq(3.85, df = 1)
[1] 0.04974599
```

A coin toss

Q6: how much do you have to degrade likelihood to get 95% Cis?

$Q = -2.(\ln(\mathcal{L}_{degraded}) - \ln(\mathcal{L}_{ML}))$ with Q 95% quantile of chi2

$$ln(a) - \ln(b) = \ln\left(\frac{a}{b}\right)$$

$$c.\,ln(a) = \ln(a^c)$$

$$\exp(\ln(a)) = a$$

A coin toss

Q6: how much do you have to degrade likelihood to get 95% Cis?

$$Q = -2.(\ln(\mathcal{L}_{degraded}) - \ln(\mathcal{L}_{ML}))$$

with Q 95% quantile of chi2

```
> qchisq(p = 0.95, df = 1)
[1] 3.8414588
```

$$Q = \ln((\frac{\mathcal{L}_{degraded}}{\mathcal{L}_{ML}})^{-2})$$

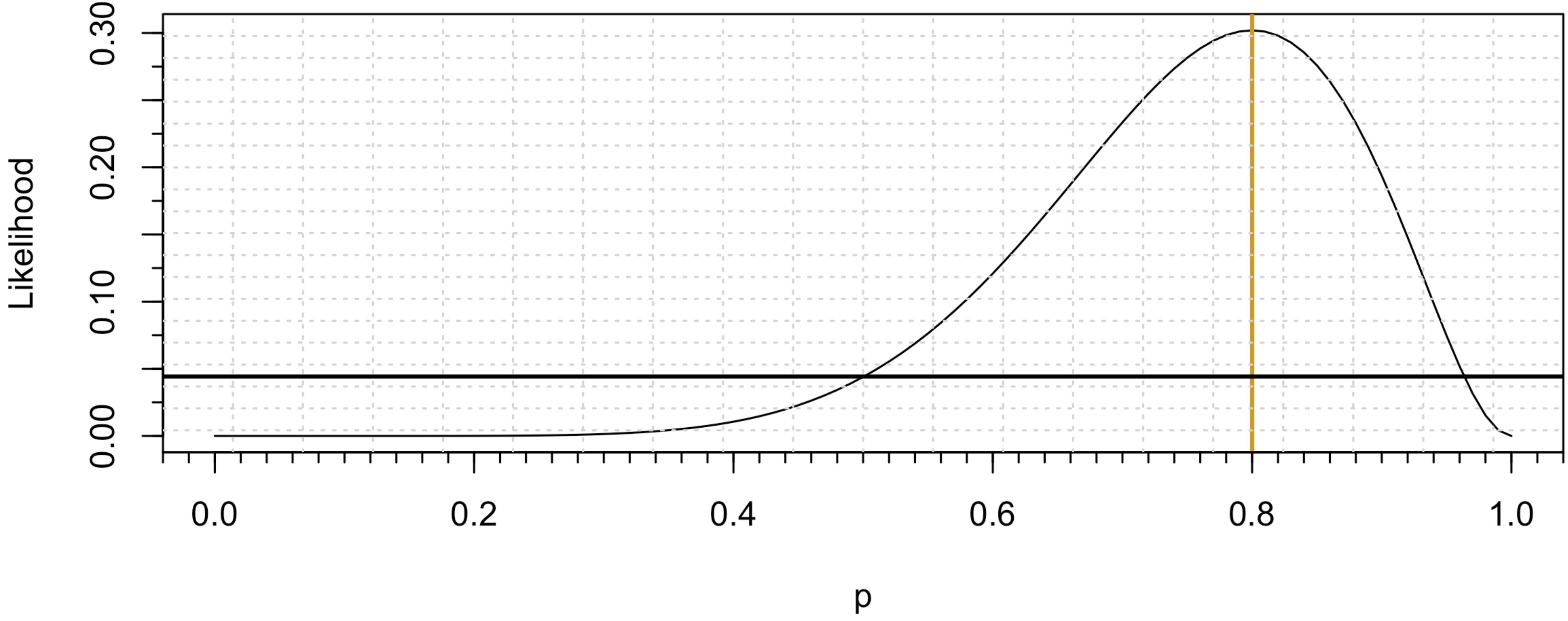
$$e^Q = (\frac{\mathcal{L}_{ML}}{\mathcal{L}_{degraded}})^2$$

$$\boldsymbol{\mathcal{L}_{degraded} = \frac{\mathcal{L}_{ML}}{e^{Q/2}} = \frac{\mathcal{L}_{ML}}{6.82}}$$

Q7: can you give a visual estimate of the 95% CI?

p_{ML}=0.8

95% CI ~ 0.5 – 0.95 NB: non-symmetric



Summary

p: *probability of heads*

x : 8 *heads*, 2 *tails*

$$\mathcal{L}(p \mid x) = \binom{10}{8} p^8 (1-p)^2$$

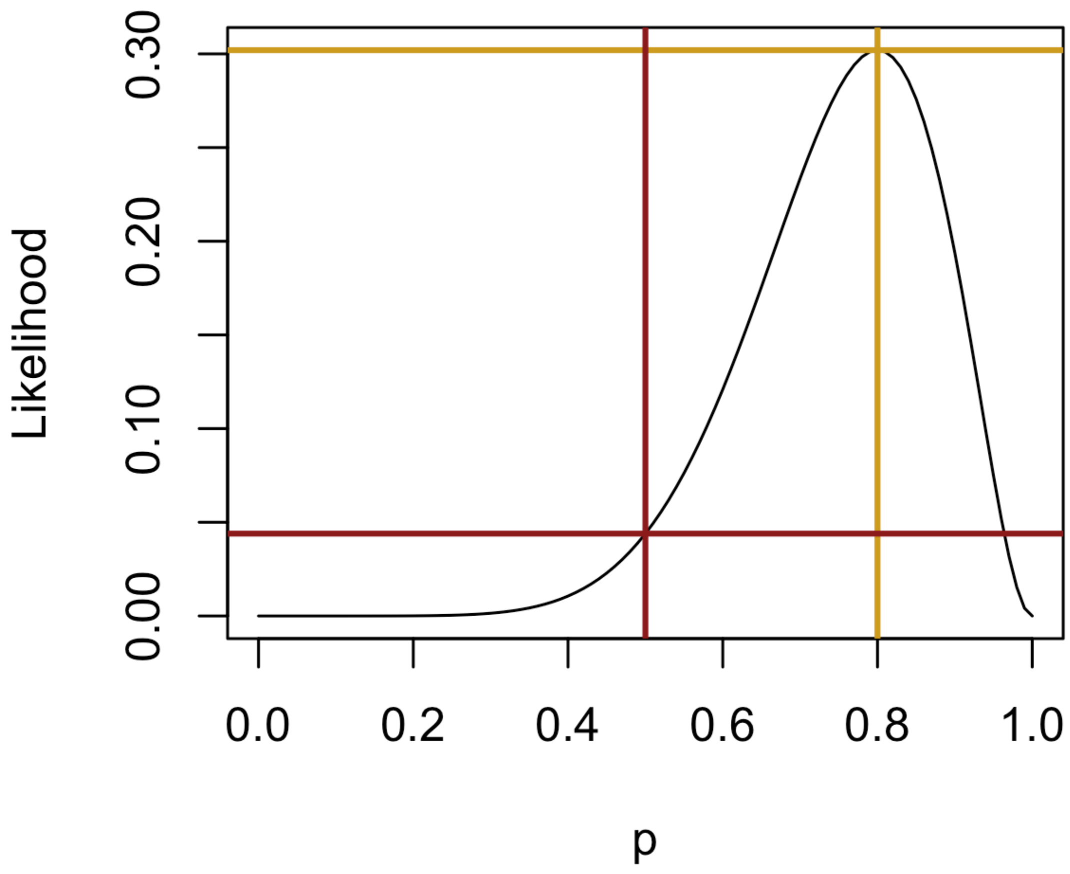
$$\widehat{p_{ML}} = \frac{8}{10} = 0.8 \qquad 95\% CI \sim 0.5 - 0.95$$

$$\mathcal{L}_{ML} = 45 \, . \, 0.8^8 . (1 - 0.8)^2 = 0.301$$

$$\mathcal{L}_0 = 45 \, . \, 0.5^8 . (1 - 0.5)^2 = 0.044$$

$$\chi^2 = -2. (\ln(\mathcal{L}_0) - \ln(\mathcal{L}_{ML})) = 3.85$$

```
> 1-pchisq(3.85, df = 1)
[1] 0.04974599
```



Conclusion

Likelihood is a central concept/object in statistics

It is the probability of the data, as function of the model parameters

Uses of likelihood include:

- Estimate model parameters (maximum likelihood)
- Test significance (likelihood ratio test)
- Quantify goodness of fit (e.g., AIC, BIC)
- Estimate confidence intervals

Thank you

- Mike Hunter
- PennState
- College of health and Human Development

- Dave Evans
- The University of Queensland
- Institute for Molecular Bioscience

See github/baptisteCD for the code used in this presentation