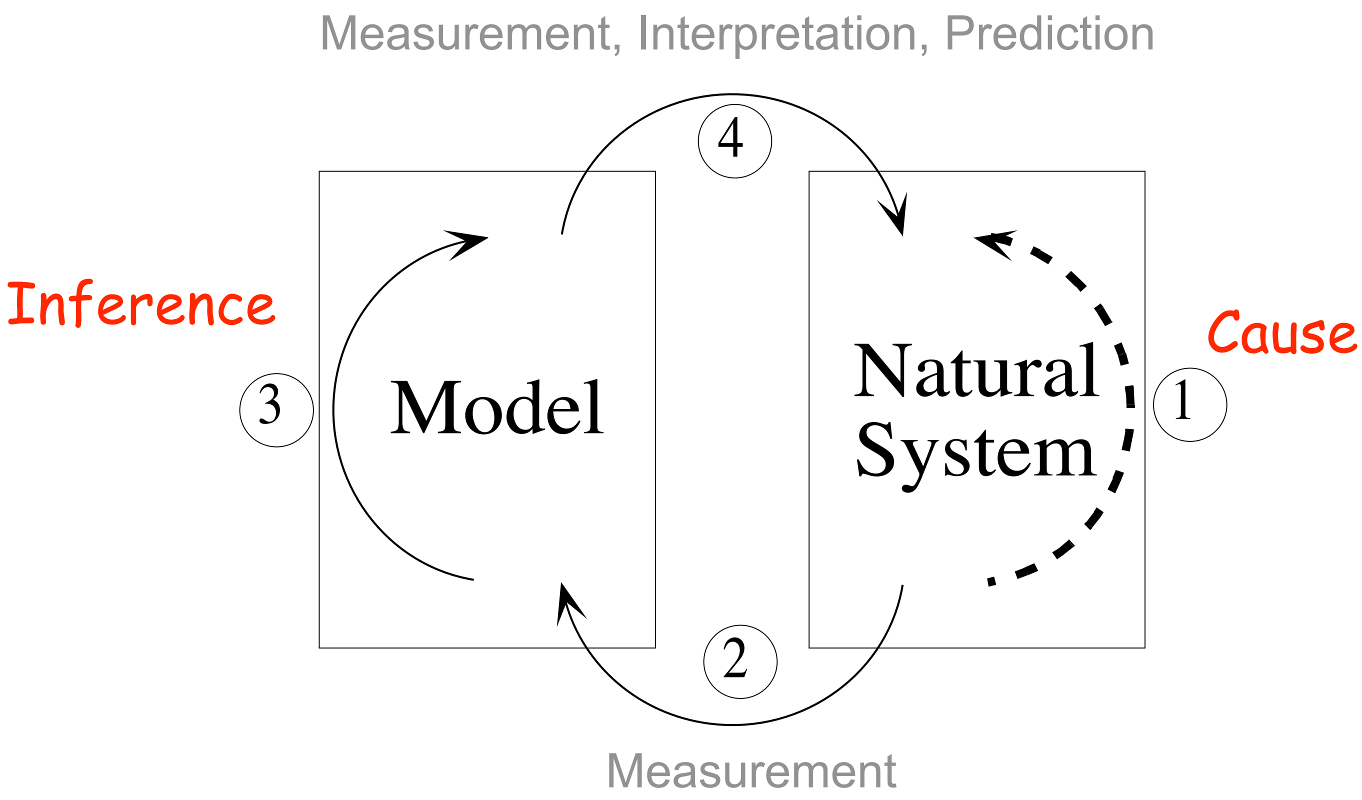
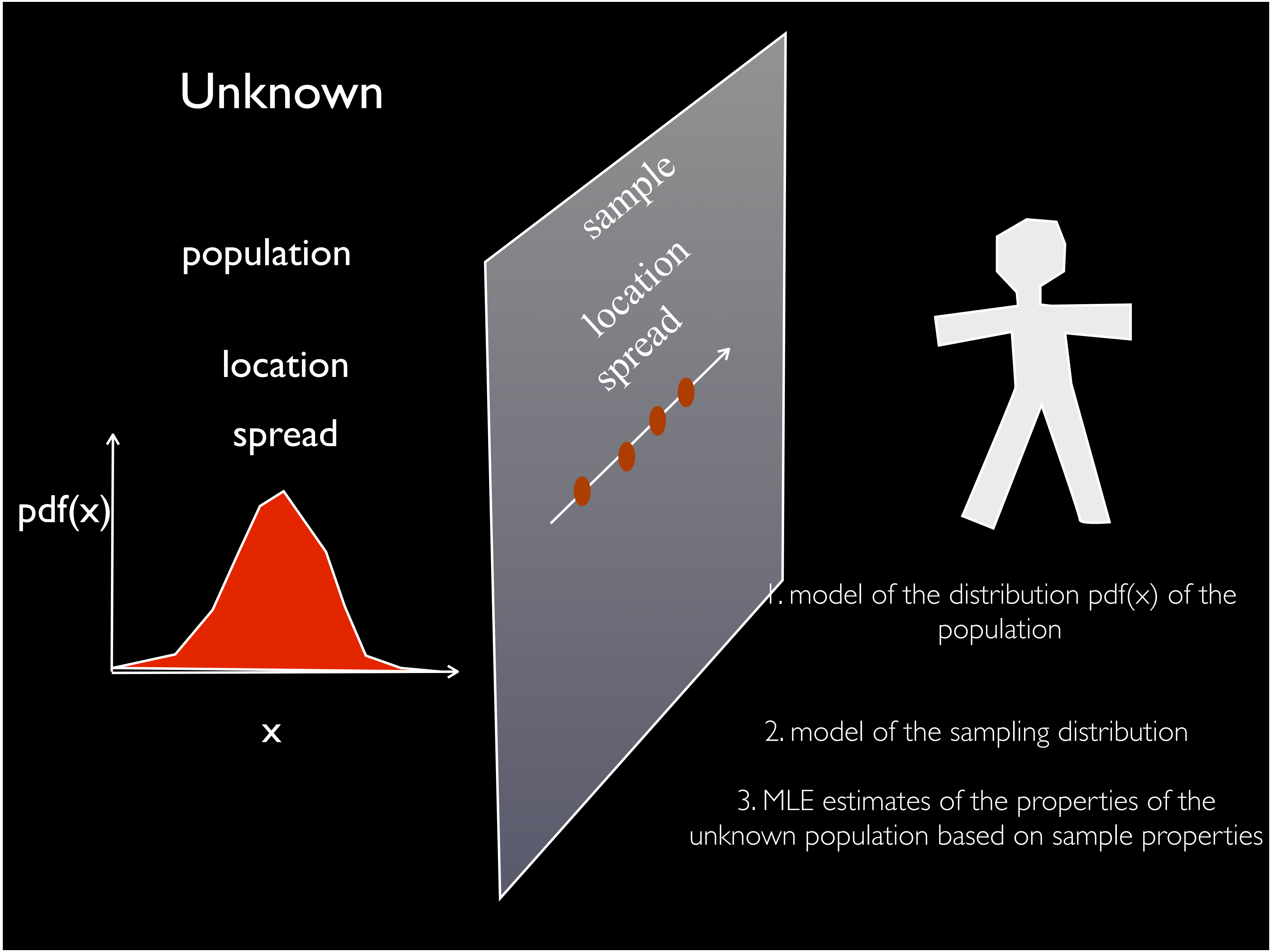
How to control your fear of the unknown: measurement theory, estimation and error propagation

The Modelling Relation



[After: Rosen (1991) Life Itself. Columbia University Press. New York]

Inferential structure is mapped to causality structure


Unknown
population
location
spread
pdf(x)
x
sample
location
spread
1. model of the distribution pdf(x) of the population
2. model of the sampling distribution
3. MLE estimates of the properties of the unknown population based on sample properties

We never have access to

the true probability distribution of population

neither

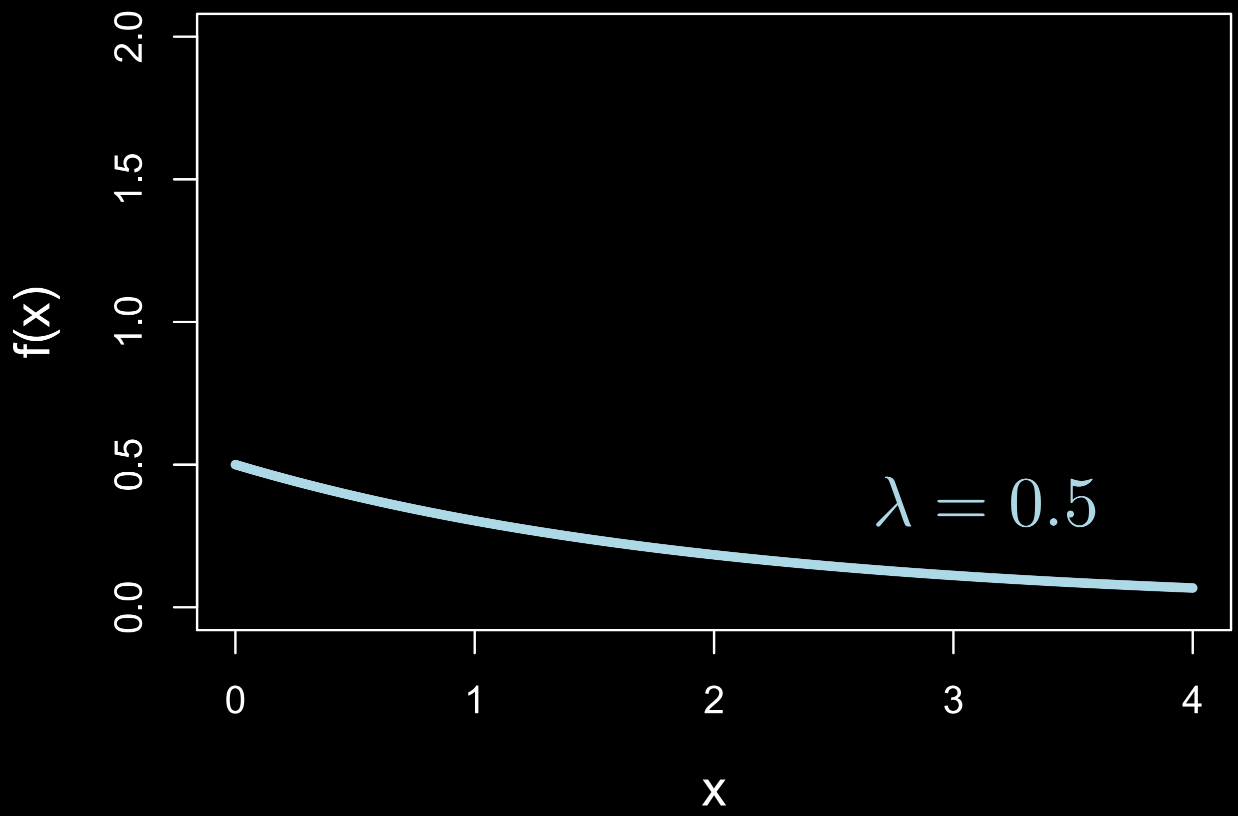
the true sampling probability distribution

We can only get estimates !

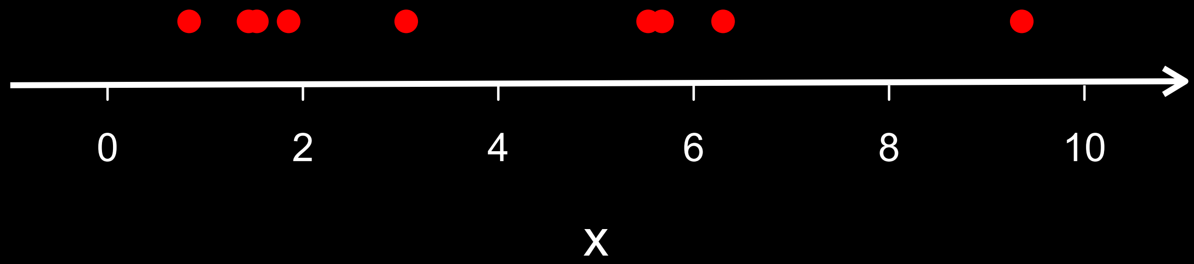
Maximum likelihood estimation

$$f(x|\lambda) = \lambda e^{-\lambda x}$$

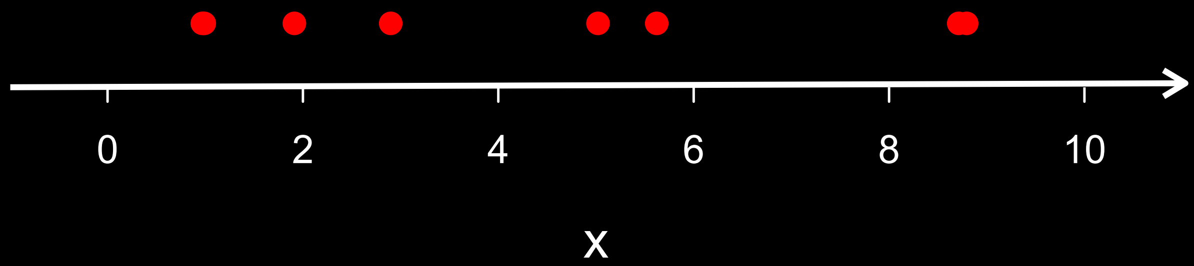
$$f(x|\lambda) = \lambda e^{-\lambda x}$$



$$f(x|\lambda) = \lambda e^{-\lambda x}$$



$$f(x|\lambda) = \lambda e^{-\lambda x}$$



$$f(x|\lambda) = \lambda e^{-\lambda x}$$

Maximum likelihood estimation of λ

$$f(x|\lambda) = \lambda e^{-\lambda x}$$

$$X = (x_1, x_2, x_3, ..., x_n)$$

$$f(X|\lambda) = f(x_1|\lambda)f(x_2|\lambda)f(x_3|\lambda)...f(x_n|\lambda)$$

$$f(X|\lambda) = \prod_{i=1}^{n} f(x_i|\lambda)$$

$$L(\lambda|X) = f(X|\lambda)$$

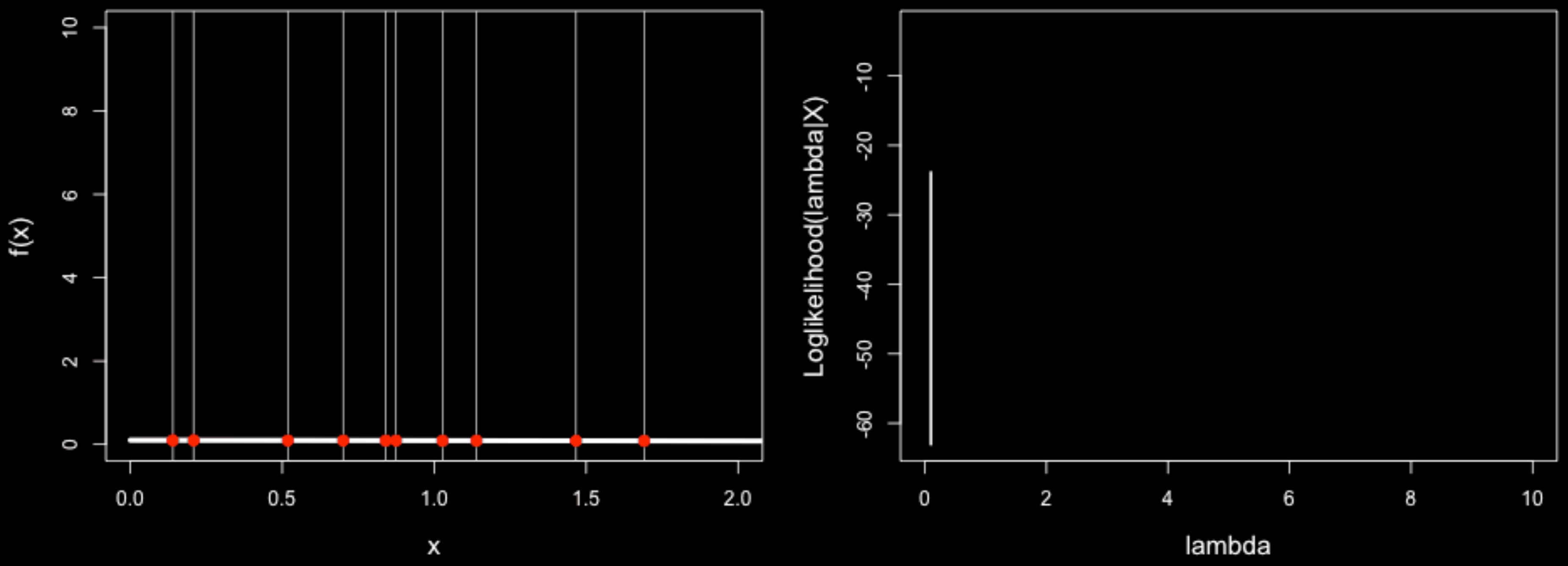
$$ln(L(\lambda|X)) = ln(f(X|\lambda))$$

$$f(x|\lambda) = \lambda e^{-\lambda x}$$

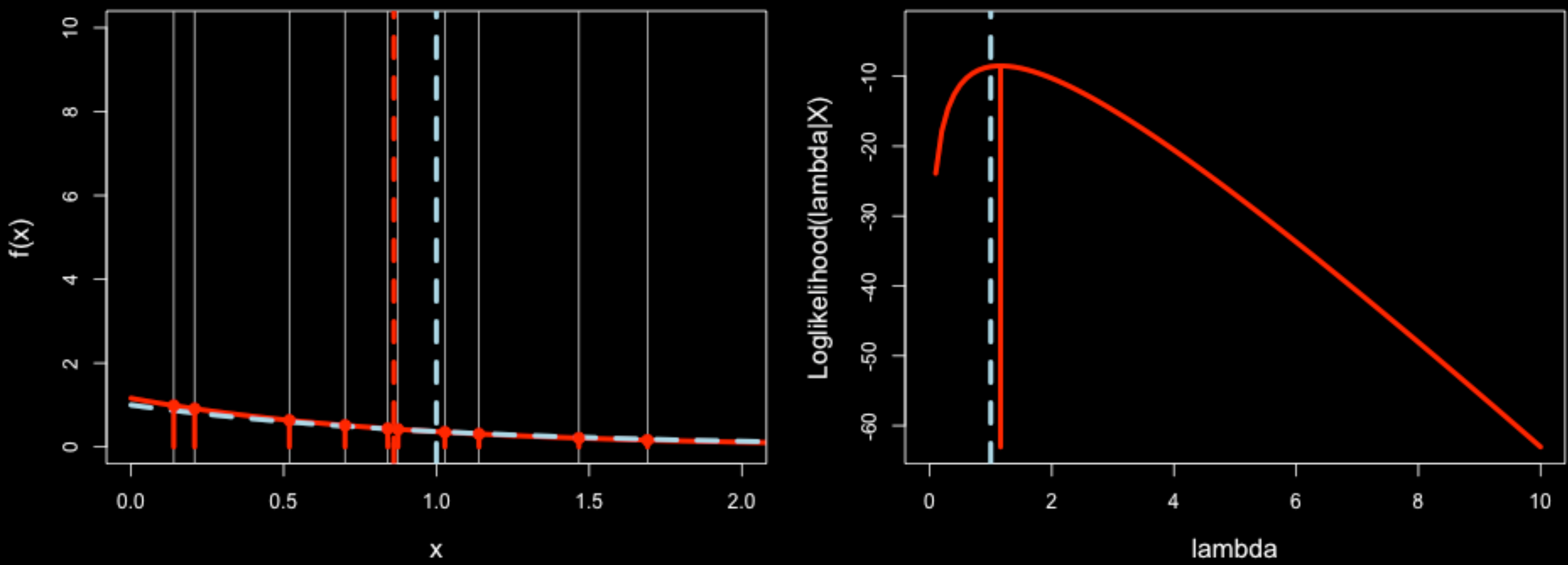
$$ln(L(\lambda|X)) = ln(f(X|\lambda))$$

$$ln(f(X|\lambda)) = ln(\prod_{i=1}^{n} f(x_i|\lambda))$$

$$ln(f(X|\lambda)) = \sum_{i=1}^{n} ln(f(x_i|\lambda))$$



f(x)
x
0.0
0.5
1.0
1.5
2.0
0
2
4
6
8
10
Loglikelihood(lambda|X)
lambda
-10
-20
-30
-40
-50
-60



X=0.1401877,1.6909198,
1.1393720,0.8407607,
1.0274900,0.7008068,
0.2084256,0.5186256 ,
0.8743341,1.4665110

estimated lambda=1.16

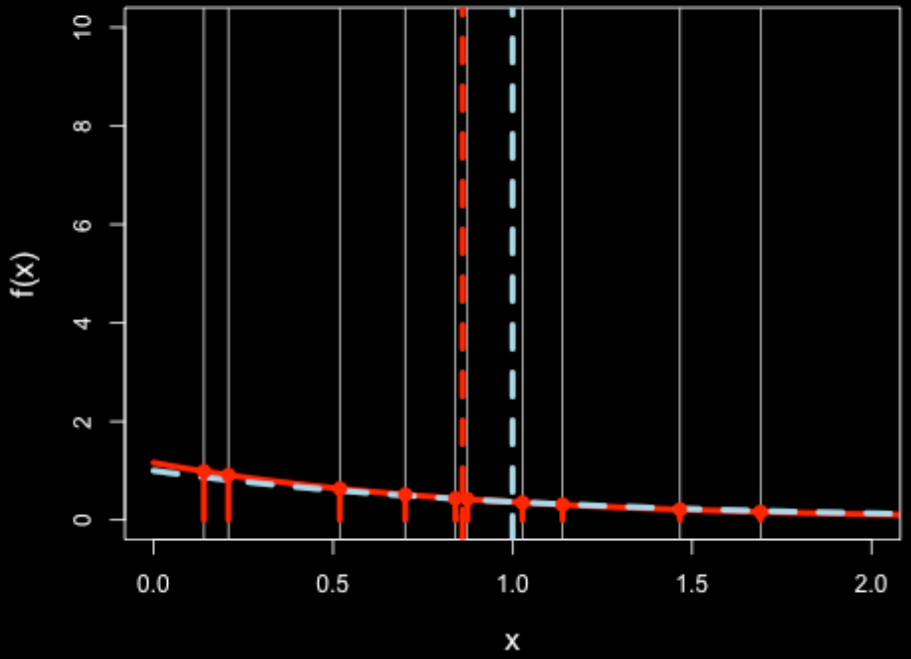
$$f(x|\lambda) = \lambda e^{-\lambda x}$$

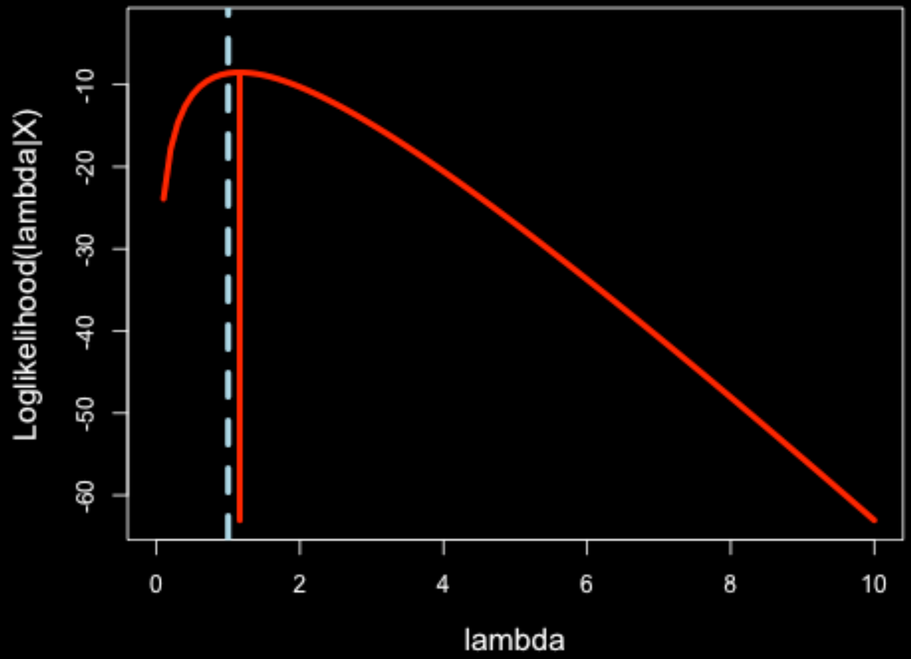
$$\begin{aligned} ln(L(\lambda|X)) &= ln(f(X|\lambda)) \\ &= \sum_{i=1}^{n} ln(\lambda) - \lambda x_i \\ &= n ln(\lambda) - \sum_{i=1}^{n} \lambda x_i \end{aligned}$$

$$\frac{\partial ln(f(X|\lambda))}{\partial \lambda} = 0$$

$$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

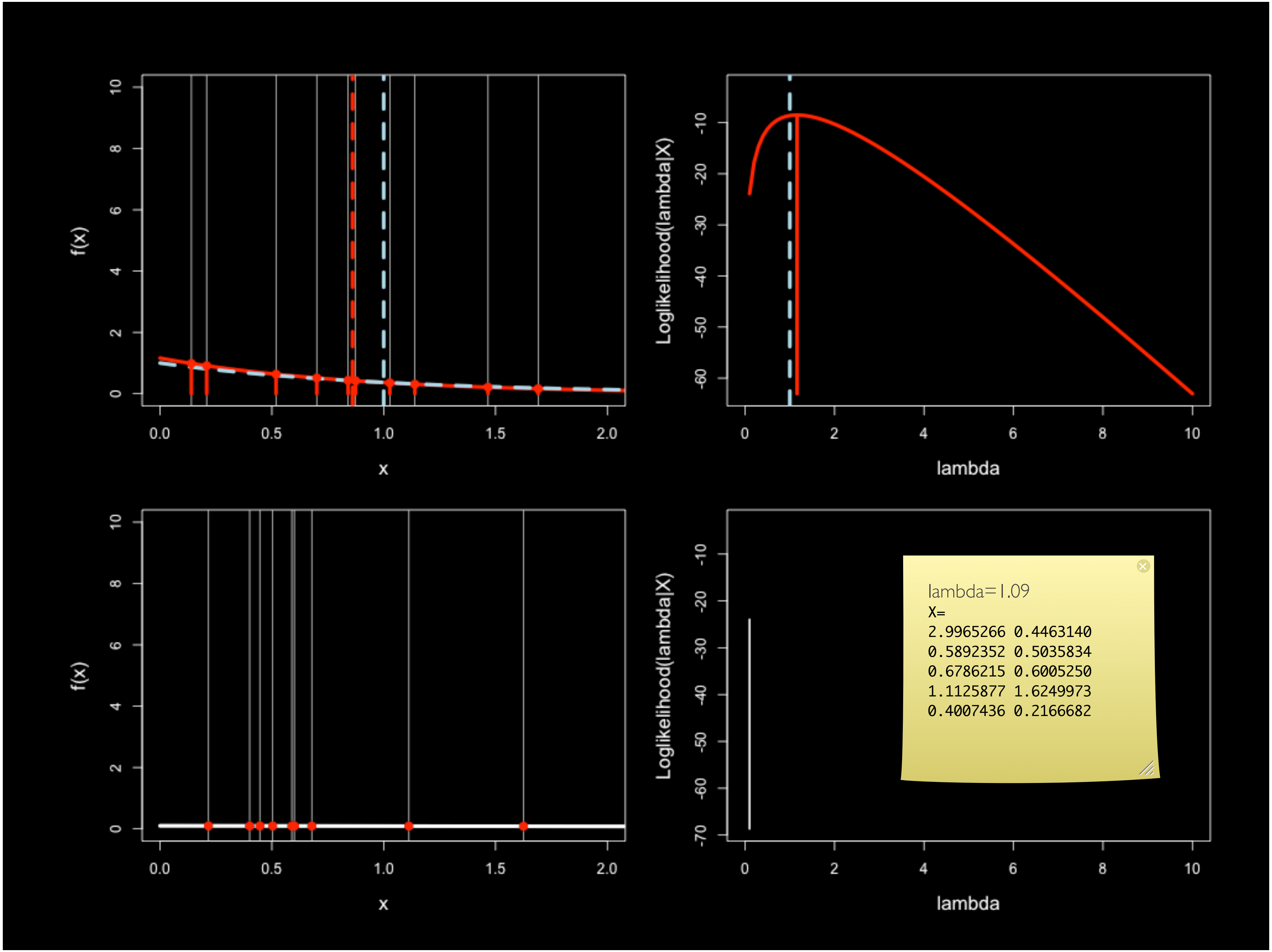
$$\lambda = \frac{n}{\sum_{i=1}^{n} x_i}$$

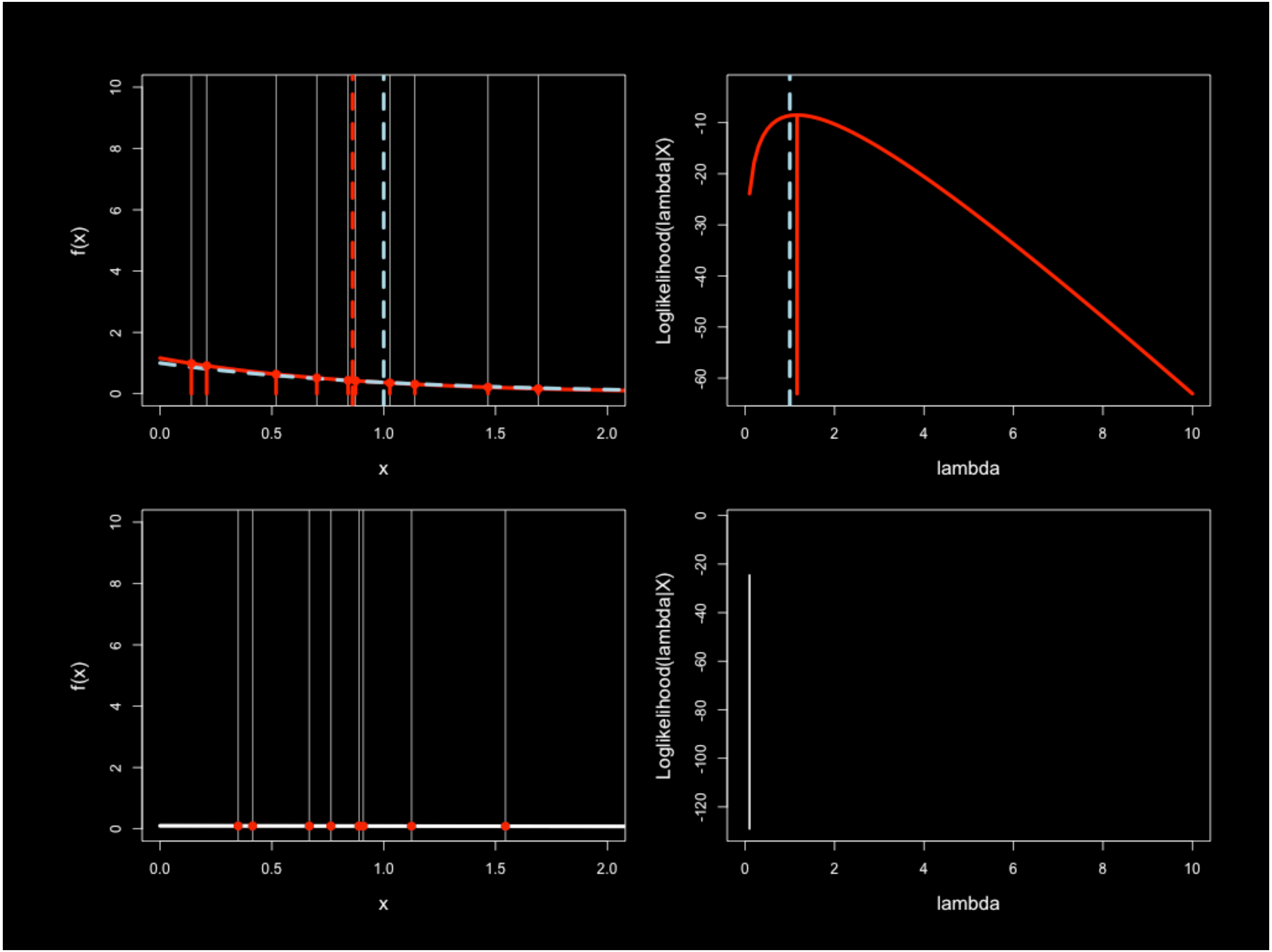


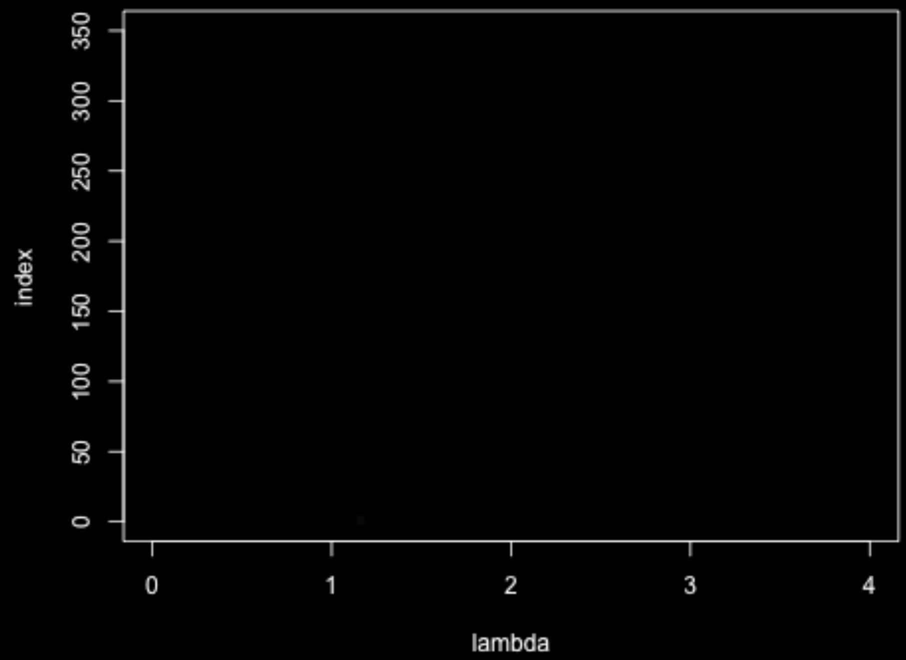


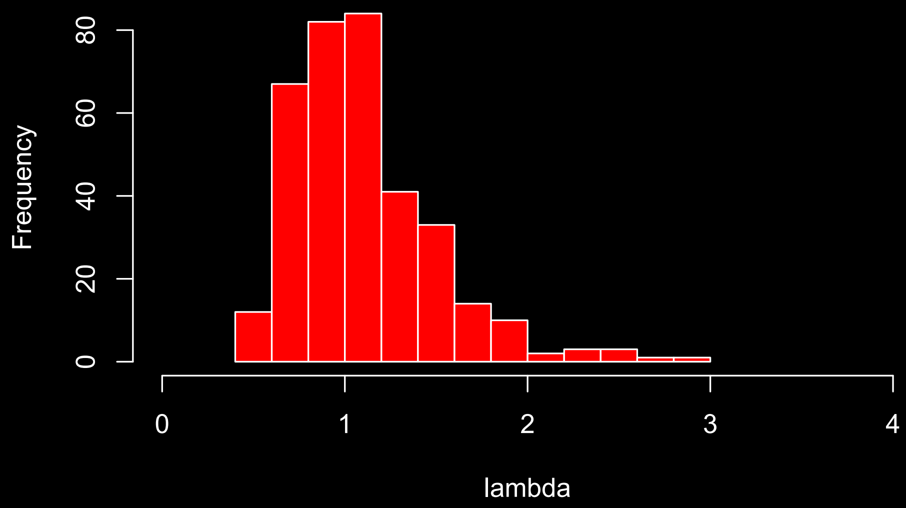
X=0.1401877,1.6909198,
1.1393720,0.8407607,
1.0274900,0.7008068,
0.2084256,0.5186256 ,
0.8743341,1.4665110

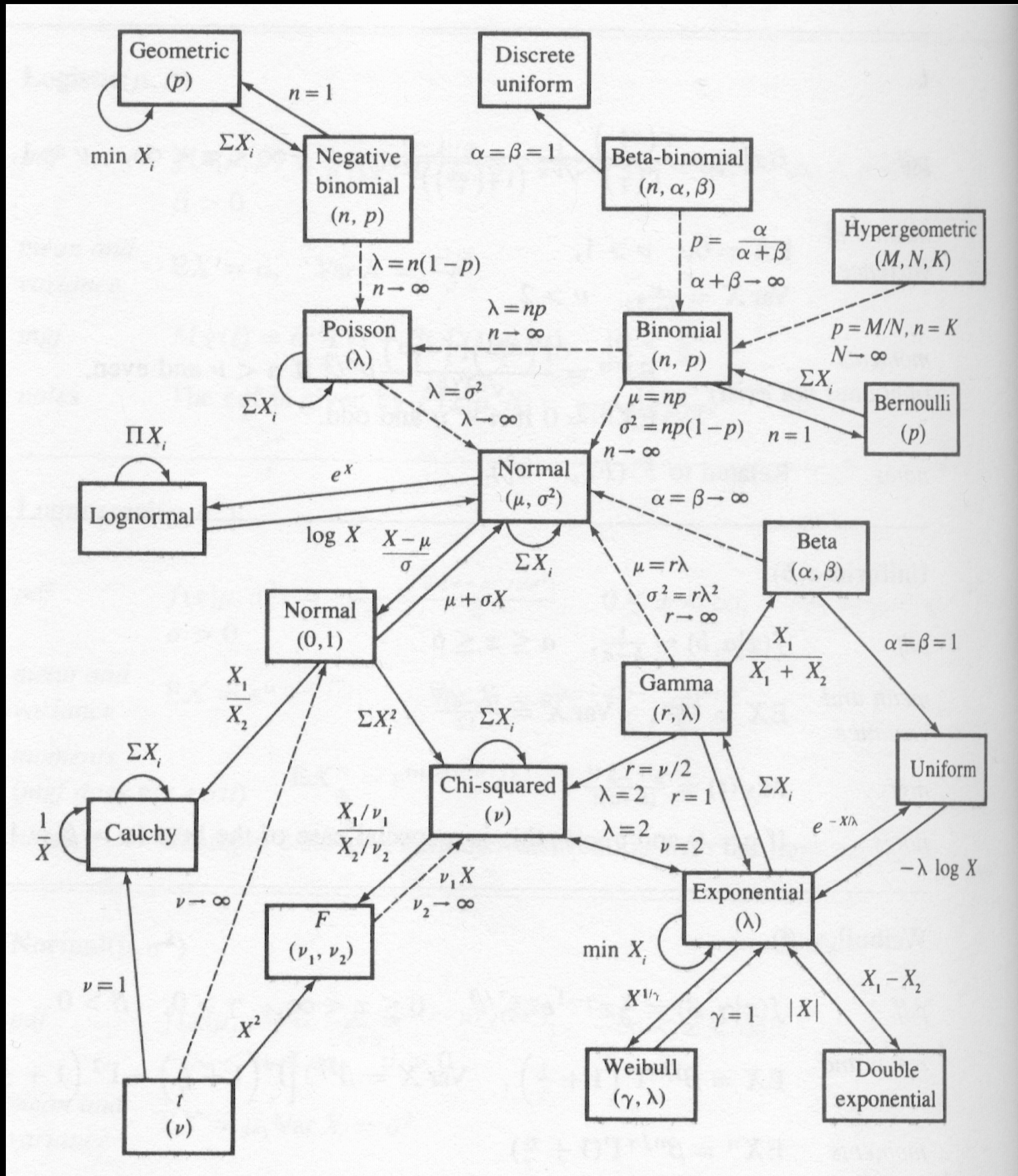
estimated lambda=1.16


f(x)
x
Loglikelihood(lambda|X)
lambda
f(x)
x
Loglikelihood(lambda|X)
lambda
lambda=1.09
X=
2.9965266 0.4463140
0.5892352 0.5035834
0.6786215 0.6005250
1.1125877 1.6249973
0.4007436 0.2166682


f(x)
x
Loglikelihood(lambda|X)
lambda
f(x)
x
Loglikelihood(lambda|X)
lambda


index
0
50
100
150
200
250
300
350
0
1
2
3
4
lambda


Frequency
0
20
40
60
80
0
1
2
3
4
lambda



Geometric (p)
$n = 1$
ΣX_i
min X_i
Discrete uniform
$\alpha = \beta = 1$
Negative binomial (n, p)
Beta-binomial (n, α, β)
Hypergeometric (M, N, K)
$\lambda = n(1-p)$
$n \to \infty$
$p = \frac{\alpha}{\alpha+\beta}$
$\alpha + \beta \to \infty$
$\lambda = np$
$n \to \infty$
Poisson (λ)
Binomial (n, p)
$p = M/N$, $n = K$
$N \to \infty$
ΣX_i
$\lambda = \sigma^2$
$\lambda \to \infty$
ΣX_i
Bernoulli (p)
$\mu = np$
$\sigma^2 = np(1-p)$
$n \to \infty$
$n = 1$
ΠX_i
e^X
Normal (μ, σ^2)
Lognormal
log X
$\alpha = \beta \to \infty$
$\frac{X-\mu}{\sigma}$
ΣX_i
Beta (α, β)
$\mu = r\lambda$
$\sigma^2 = r\lambda^2$
$r \to \infty$
$\mu + \sigma X$
Normal (0,1)
$\frac{X_1}{X_1 + X_2}$
$\alpha = \beta = 1$
Gamma (r, λ)
$\frac{X_1}{X_2}$
ΣX_i^2
ΣX_i
ΣX_i
Uniform
Chi-squared (ν)
$r = \nu/2$
$\lambda = 2$
$r = 1$
ΣX_i
$\frac{1}{X}$
Cauchy
$\frac{X_1/\nu_1}{X_2/\nu_2}$
$\lambda = 2$
$\nu = 2$
$e^{-X/\lambda}$
$-\lambda$ log X
$\nu_1 X$
$\nu_2 \to \infty$
$\nu \to \infty$
Exponential (λ)
F (ν_1, ν_2)
min X_i
$\nu = 1$
$X_1 - X_2$
$X^{1/\gamma}$
$\gamma = 1$
$|X|$
X^2
Weibull (γ, λ)
Double exponential
t (ν)

Interval of confidence

Interval of confidence

$$\hat{\lambda}\left(1-\frac{1.96}{\sqrt{n}}\right) < \lambda < \hat{\lambda}\left(1+\frac{1.96}{\sqrt{n}}\right).$$

(95% confidence Gaussian approximation)

Interval of confidence

$$\lambda = 1.09 \quad [0.41, 1.77]$$

Propagation of uncertainty

Propagation of uncertainty

$$z_1 \pm \epsilon_1$$

$$z_2 \pm \epsilon_2$$

...

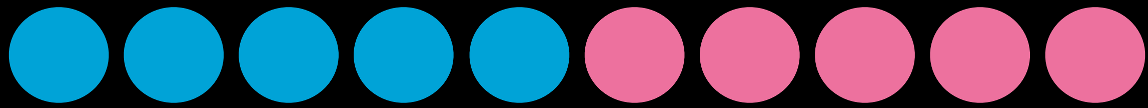
$$z_n \pm \epsilon_n$$

$$g(z_1, z_2, z_3, ...z_n)$$

$$g \pm \epsilon_g$$

$$\epsilon_g = \sqrt{\sum_{i=1}^{n} \frac{\partial g}{\partial z_i} \epsilon_i}$$

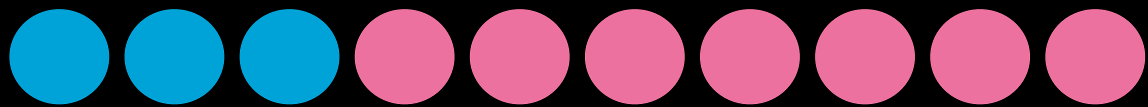
An exercice



fair

10x 

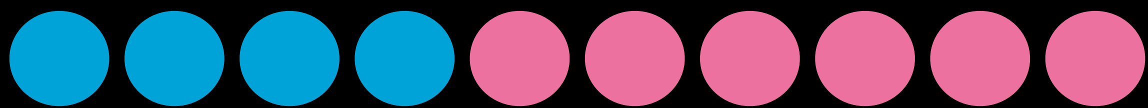
$$f(3|n=10, w=0.5) = \frac{10!}{3!(10-3)!}0.5^3(1-0.5)^{10-3}$$



biased

10x 

$$f(4|n=10, w=0.3) = \frac{10!}{4!(10-4)!}0.3^4(1-0.3)^{10-4}$$



Consider a data vector y of individual observations y_i:

$$y = (y_1, y_2, y_3, ..., y_m)$$

which is a random sample from an unknown population.

Let $f(y|w)$ be the *probability density function* (PDF) of the vector y given the parameter vector w

$$w = (w_1, w_2, ..., w_k)$$

If the individual observations are independent then, according to probability theory, the PDF for the data y can be expressed as a multiplication of the densities of individual observations:

$$f(y|w) = f(y = (y_1, y_2, ..., y_m)|w) = f_1(y_1|w) f_2(y_2|w) ... f_m(y_m|w)$$

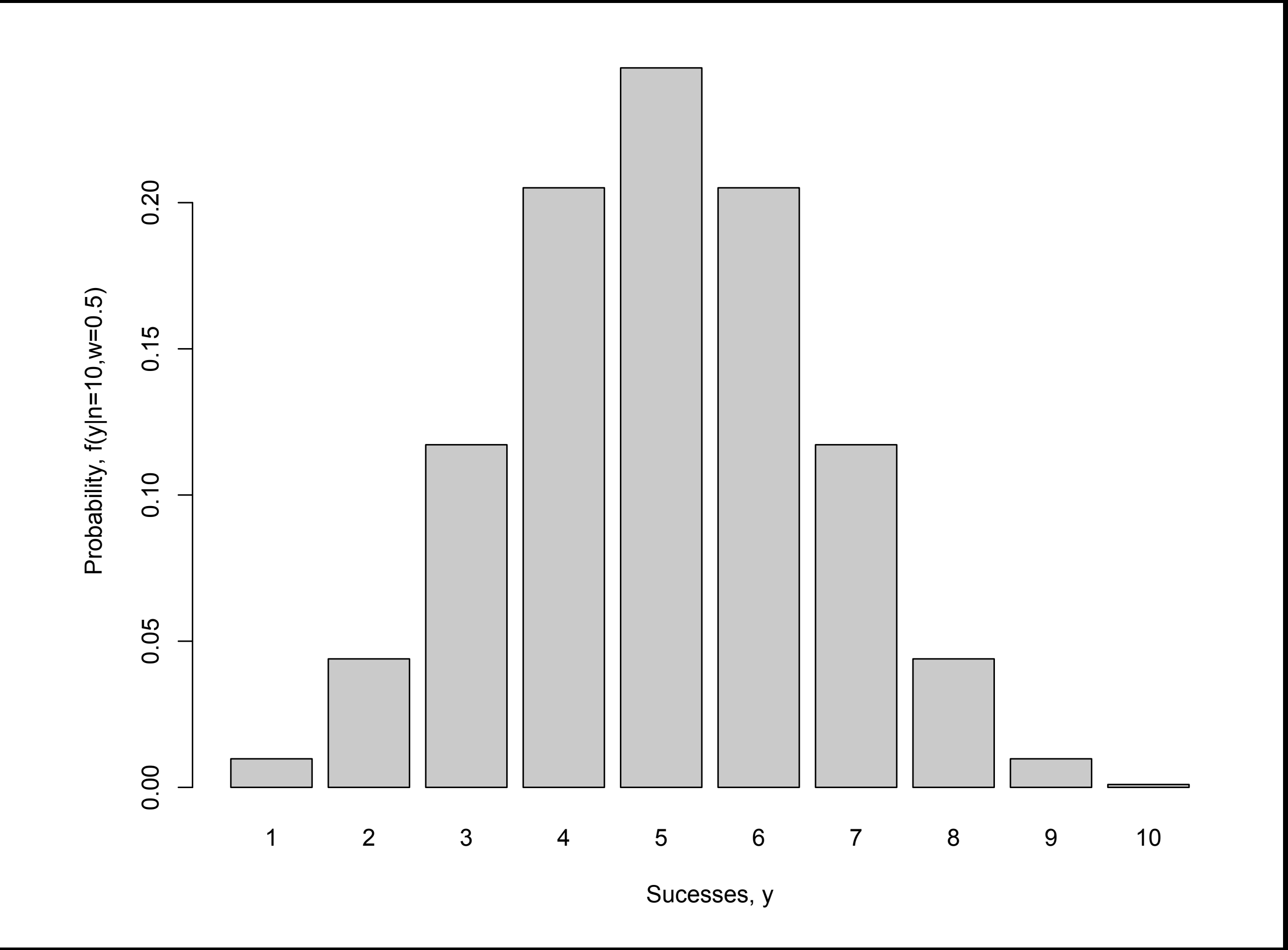
Tossing a 1 euro fair coin 10 times and success getting "1" face

The PDF for y is given by a binomial distribution:

$$f(y|n = 10, w = 0.5) = \frac{10!}{y!(10-y)!} 0.5^y (1-0.5)^{10-y}$$

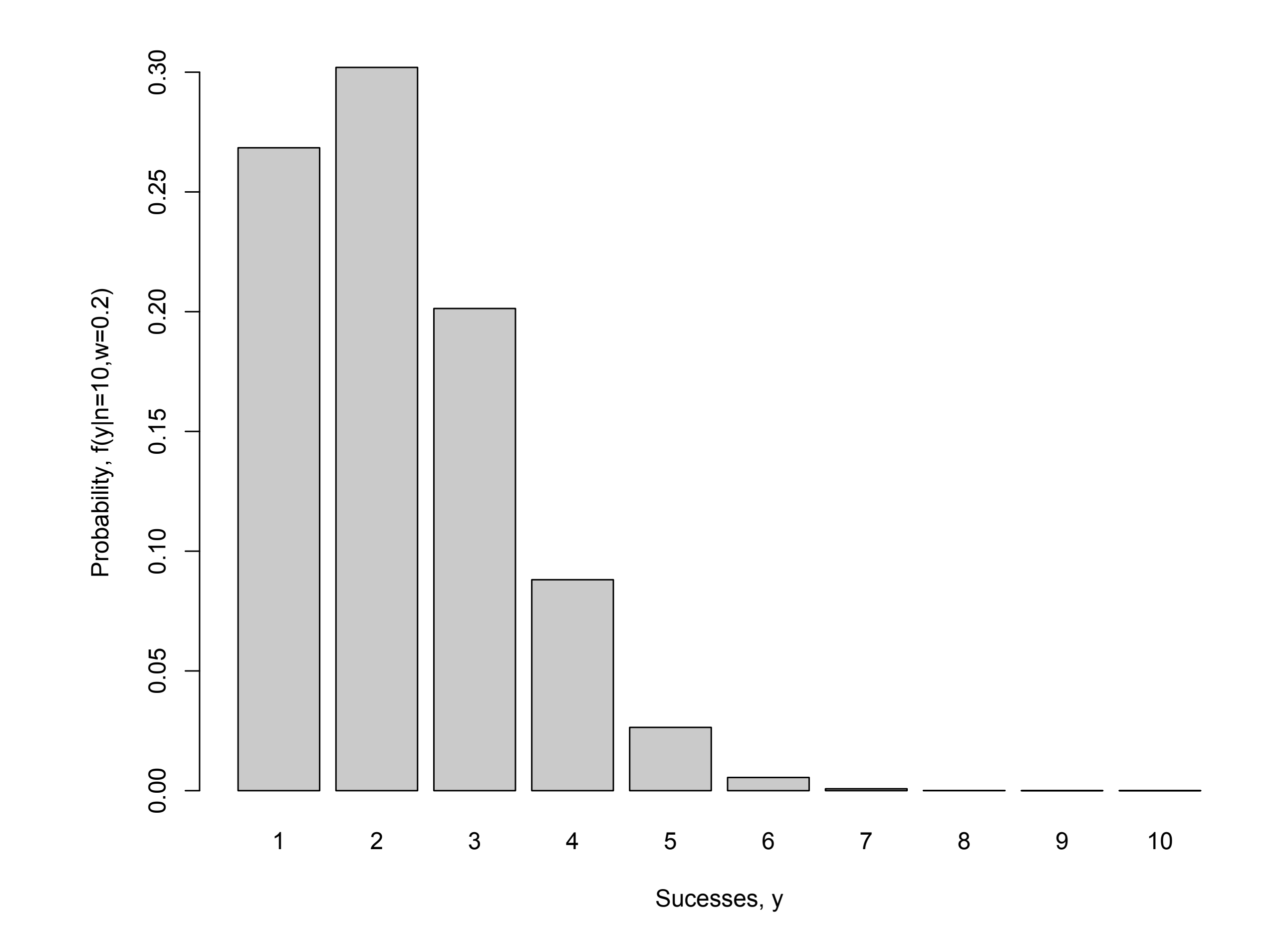
$$y \in \{1, 2, ..., 10\}$$

$$f(y|n=10,w=0.5)=\frac{10!}{y!(10-y)!}0.5^y(1-0.5)^{10-y}$$




```
n<-10
y<-1:n
w<-0.5
fy<-factorial(n)/(factorial(y)*factorial(n-y))*w^y*(1-w)^(n-y)
barplot(fy,names.arg=y,xlab="Sucesses, y",ylab="Probability, f(y|n=10,w=0.5)")
```

What if the coin is "unfair"?




```
n<-10
y<-1:n
w<-0.2
fy<-factorial(n)/(factorial(y)*factorial(n-y))*w^y*(1-w)^(n-y)
barplot(fy,names.arg=y,xlab="Sucesses, y",ylab="Probability, f(y|n=10,w=0.2)")
```

Supose we take a coin (fair or unfair) and toss it 10 times and get:

$$y=4$$

What is the best estimate for the value of w ?

$$w=?$$

Maximum likelihood estimation of w

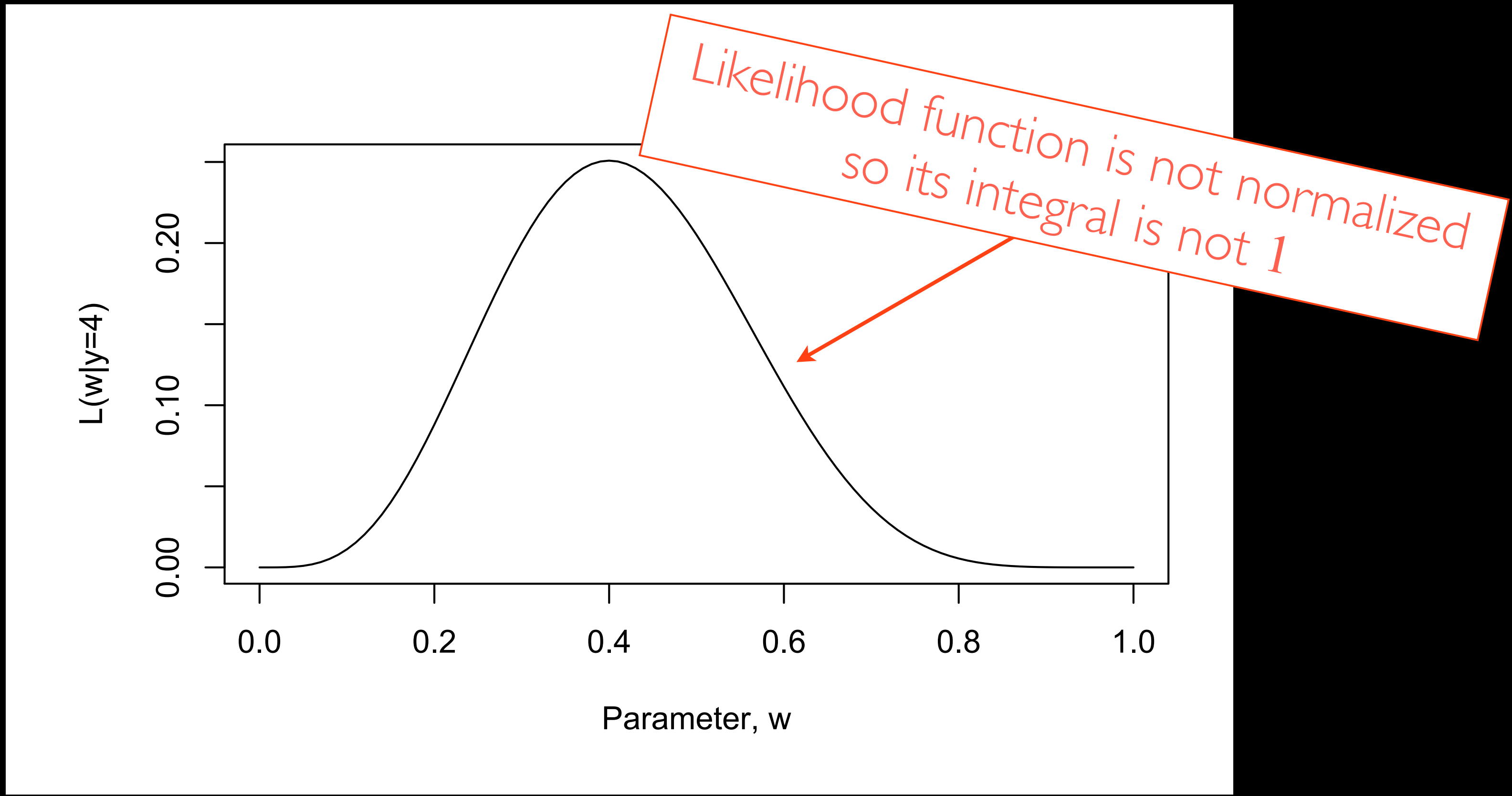
We define the likelihood function of the parameter w given the data y as:

$$L(w|y) = f(y|w)$$

For the one parameter binomial distribution this is:

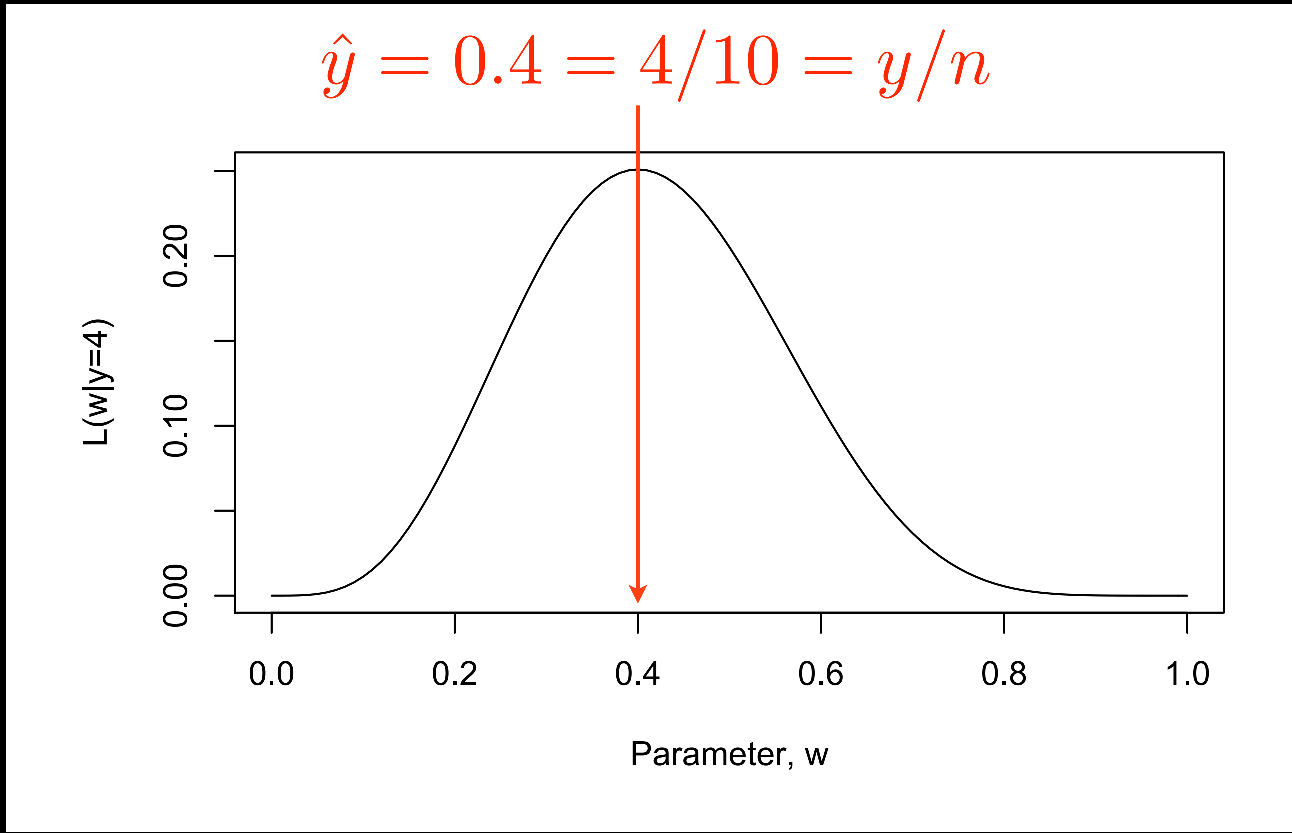
$$L(w|y=4) = f(y=4|n=10, w) = \frac{10!}{4!(10-4)!} w^4 (1-w)^{10-4}$$

How does the likelihood function look like?




```
n<-10
y<-4
w<-seq(0,1,by=0.01)
fy<-factorial(n)/(factorial(y)*factorial(n-y))*w^y*(1-w)^(n-y)
plot(w,fy,names.arg=y,xlab="Parameter, w",ylab="L(w|y=4)",type="l")
```

What is the maximum of the likelihood function ?



Maximum likelihood estimation

For convenience MLE are obtained by maximizing the log-likelihood function

$$lnL(w|y)$$

Maximum likelihood estimation

At the maximum of a function its first derivative vanishes ...

$$\frac{\partial lnL(w|y)}{\partial w} = 0$$

... and its second derivative is negative

$$\frac{\partial^2 lnL(w|y)}{\partial w^2} < 0$$

Maximum likelihood estimation

Setting y=4 in the log-likelihood yields:

$$lnL(w|y=4) = ln\left(\frac{10!}{4!(10-4)!}w^4(1-w)^{10-4}\right)$$

$$= ln\left(\frac{10!}{4!6!}\right) + 4ln(w) + 6ln(1-w)$$

Maximum likelihood estimation

The first derivative of the log-likelihood is

$$\frac{\partial lnL(w,|y=4)}{\partial w} = \frac{4}{w} - \frac{6}{1-w}$$

Setting this to zero...

$$\frac{4}{w} - \frac{6}{1-w} = 0$$

...and solving to *w* yields:

$$w = 0.4$$

Maximum likelihood estimation

The second derivative of the log-likelihood at $w = 0.4$ is negative

$$\frac{\partial ln^2 L(w|y=4)}{\partial w^2} = -\frac{4}{w^2} - \frac{6}{(1-w)^2} < 0$$