Maximum Likelihood Estimation

Week 2 - PSYC 5316

September 4, 2017

Recall

Last time, we gave a formal definition for a probability function. An example was the *binomial* distribution for N independent Bernoulli trials (e.g., coin flips):

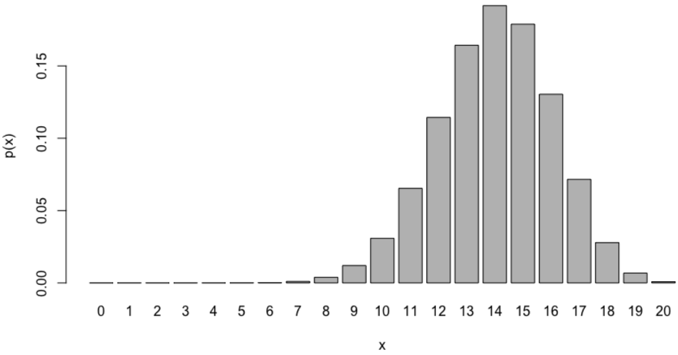
$$f(x \mid \theta) = \binom{N}{x}\theta^x(1-\theta)^{N-x}$$

where $x = \#$ of successes, and $\theta =$ probability of success.

Probability function

Suppose $N = 20$ and $\theta = 0.7$.

```
barplot(dbinom(0:20,size=20,prob=0.7),
        names.arg=0:20,
        ylab="p(x)",
        xlab="x")
```



Data and parameters

$$f(x \mid \theta) = \binom{N}{x} \theta^x (1 - \theta)^{N-x}$$

This function gives us the probability of data, *given* a specific parameter

Data and parameters

What if we switched these?

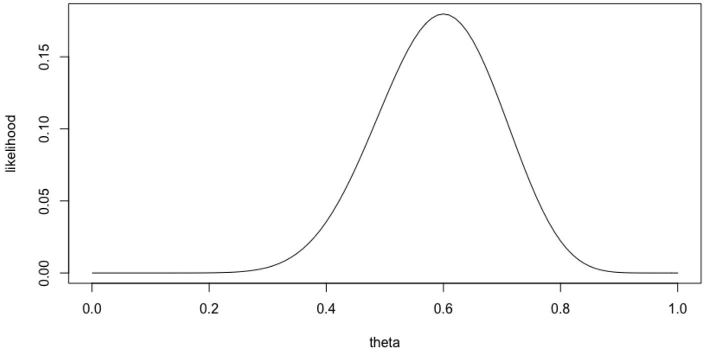
$$f(\theta \mid x) = \binom{N}{x} \theta^x (1-\theta)^{N-x}$$

This function then gives us the likelihood of a range of parameters, *given* a specific data point

Likelihood function

Suppose we observed 12 successes in 20 trials:

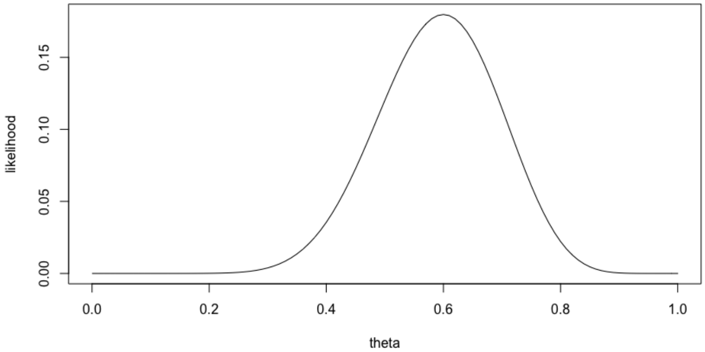
```
theta=seq(from=0, to=1, by=0.01)
plot(theta, dbinom(x=12, size=20, prob=theta),
     type="l",ylab="likelihood")
```



Likelihood function

Suppose we observed 12 successes in 20 trials:

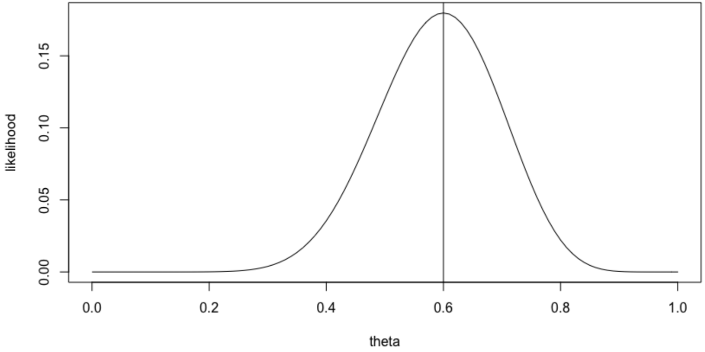
Natural question – what value of θ is most likely, given the data?



Likelihood function

Suppose we observed 12 successes in 20 trials:

Natural question – what value of θ is most likely, given the data?
Answer: $\theta = 0.6$



Maximum likelihood estimation

A key problem in statistical inference is how to infer from sample data to population parameters.

Maximum likelihood estimation is one solution to this problem

Maximum likelihood estimation

ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Mathematical Psychology 47 (2003) 90–100

Journal of Mathematical Psychology

http://www.elsevier.com/locate/jmp

Tutorial

Tutorial on maximum likelihood estimation

In Jae Myung*

Department of Psychology, Ohio State University, 1885 Neil Avenue Mall, Columbus, OH 43210-1222, USA

Received 30 November 2001; revised 16 October 2002

Abstract

In this paper, I provide a tutorial exposition on maximum likelihood estimation (MLE). The intended audience of this tutorial are researchers who practice mathematical modeling of cognition but are unfamiliar with the estimation method. Unlike least-squares estimation which is primarily a descriptive tool, MLE is a preferred method of parameter estimation in statistics and is an indispensable tool for many statistical modeling techniques, in particular in non-linear modeling with non-normal data. The purpose of this paper is to provide a good conceptual explanation of the method with illustrative examples so the reader can have a grasp of some of the basic principles.

Maximum likelihood estimation

Basic workflow:

1. collect data
2. decide on a "model" for the data (e.g., binomial, normal, etc.)
3. define a likelihood function based on the underlying model
4. find the parameter value(s) that maximize the likelihood function

Three examples

We will do three examples of MLE:

1. binomial model
2. normal model
3. ex-Gaussian model

Example 1

Suppose that in a sequence of 20 coin flips, we observe 12 successes ("heads"). What is the maximum likelihood estimate for θ?

Step 1 – collect data

- done: we observed $x = 12$ successes

Example 1

Suppose that in a sequence of 20 coin flips, we observe 12 successes ("heads"). What is the maximum likelihood estimate for θ?

Step 2 – choose a model

- assume a binomial model

$$x \sim \text{Binomial}(\theta, n = 20)$$

Example 1

Suppose that in a sequence of 20 coin flips, we observe 12 successes ("heads"). What is the maximum likelihood estimate for θ?

Step 3 – define likelihood function

- Note: we will actually define the "negative log-likelihood" function

```
nll.binom <- function(data,par){
  return(-log(dbinom(data, size = 20, prob = par)))
}
```

Example 1

Suppose that in a sequence of 20 coin flips, we observe 12 successes ("heads"). What is the maximum likelihood estimate for θ?

Step 4 – find parameter that maximizes the likelihood function

- Note: we will actually minimize the negative log-likelihood

```
optim(par=0.5, fn=nll.binom, data=12)
```

Example 1

Suppose that in a sequence of 20 coin flips, we observe 12 successes ("heads"). What is the maximum likelihood estimate for θ?

```
> optim(par=0.5, fn=nll.binom, data=12) # look at "par" output..this is the MLE
$par
[1] 0.6

$value
[1] 1.716434

$counts
function gradient
      26       NA

$convergence
[1] 0

$message
NULL
```

MLE

Example 2

The following command will load a data set into R, consisting of 1000 observations. Fit the data with a normal model and find MLEs for mean and standard deviation.

Step 1: collect data

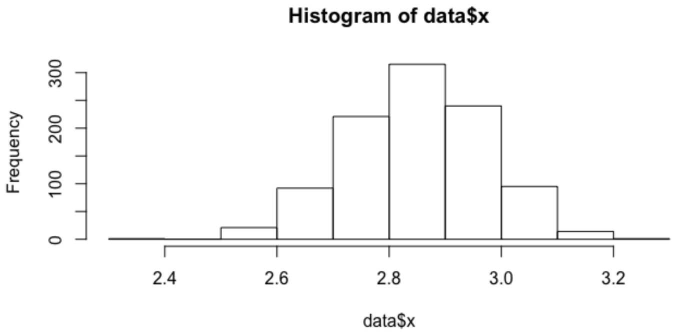
```
data <- read.csv("https://git.io/v58i8")
```

Example 2

The following command will load a data set into R, consisting of 1000 observations. Fit the data with a normal model and find MLEs for mean and standard deviation.

Step 1.5: before choosing a model, look at the distribution

```
hist(data$x)
```

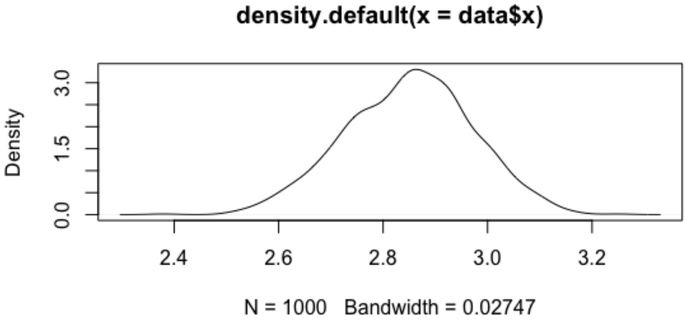


Example 2

The following command will load a data set into R, consisting of 1000 observations. Fit the data with a normal model and find MLEs for mean and standard deviation.

Step 1.5: before choosing a model, look at the distribution

```
plot(density(data$x))
```



Example 2

The following command will load a data set into R, consisting of 1000 observations. Fit the data with a normal model and find MLEs for mean and standard deviation.

Step 2: choose a model

- assume a normal model

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

Example 2

The following command will load a data set into R, consisting of 1000 observations. Fit the data with a normal model and find MLEs for mean and standard deviation.

Step 3: define likelihood function

```
nll.normal <- function(data,par){
  return(-sum(log(dnorm(data, mean=par[1], sd=par[2]))))
}
```

Example 2

The following command will load a data set into R, consisting of 1000 observations. Fit the data with a normal model and find MLEs for mean and standard deviation.

Step 4: find parameters that maximize the likelihood function

```
optim(par=c(1,0.1), fn=nll.normal, data=data$x)
```

Example 2

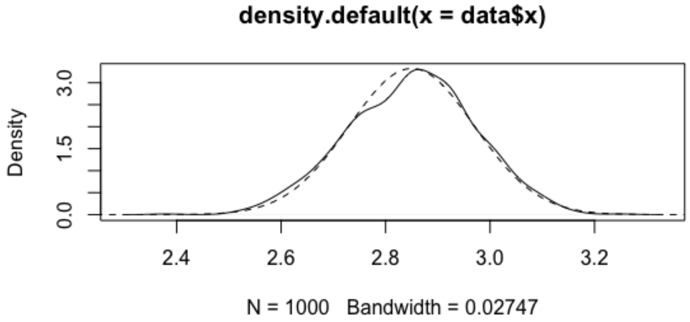
The following command will load a data set into R, consisting of 1000 observations. Fit the data with a normal model and find MLEs for mean and standard deviation.

```
> optim(par=c(1,0.1), fn=nll.normal, data=data$x) # look at par output; first number is mean, second
is sd
$par
[1] 2.8494538 0.1214792

$value
[1] -689.2001

$counts
function gradient
      85       NA

$convergence
[1] 0

$message
NULL
```

Example 2

As a check, we can see how well our model "fits" the data:

```
x=seq(-3,4,0.01)
plot(density(data$x))
lines(x,dnorm(x,mean=2.85, sd=0.12),lty=2)
```

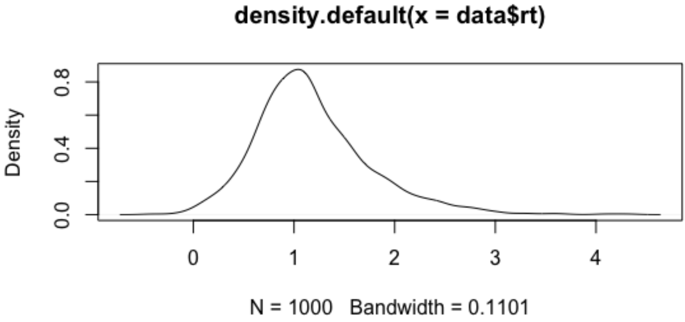


Example 3

Step 1: collect data

Here is a data set of 1000 response times (RTs) in a mental arithmetic task, followed by a plot of the data.

```
data <- read.csv("https://git.io/v58yI")
plot(density(data$rt))
```



Example 3

Step 2: choose a model

- distributions with long rightward tails can be modeled by an ex-Gaussian distribution, which is essentially a combination of the Gaussian (normal) distribution with an exponential distribution

Behavior Research Methods, Instruments, & Computers
1988, 20 (1), 54-57

Fitting the ex-Gaussian equation to reaction time distributions

MICHAEL R. W. DAWSON
University of Alberta, Edmonton, Alberta, Canada

Two programs that can be used to determine the probability distributions of reaction times are detailed. The first program takes rank-ordered reaction times as input and outputs a file of quantized data. The second program uses a simplex procedure to estimate the parameters of the ex-Gaussian equation that provides the best description of the quantized data. The advantages of this type of data analysis are also discussed.

Example 3

Step 2: choose a model

- distributions with long rightward tails can be modeled by an ex-Gaussian distribution, which is essentially a combination of the Gaussian (normal) distribution with an exponential distribution

$$rt \sim \text{exGaussian}(\mu, \sigma, \tau)$$

- μ = mean of normal component
- σ = sd of normal component
- τ = rate of exponential component

Example 3

Step 3: define likelihood function

- ExGaussian is not already a function in R. So, we'll build it ourselves.
- note: just copy from the week2.R script!

```
dexg <- function(x, mu, sigma, tau){
  return((1/tau)*exp((sigma^2/(2*tau^2)) -
  (x-mu)/tau)*pnorm((x-mu)/sigma-(sigma/tau)))
}
```

Example 3

Step 3: define likelihood function

```
nll.exg <- function(data,par){
  return(-sum(log(dexg(data,
  mu=par[1],
  sigma=par[2],
  tau=par[3]))))
}
```

Example 3

Step 4: find parameters that maximize likelihood function

```
optim(par=c(0,0.1,0.1), fn=nll.exg, data = data$rt)
```

```
> optim(par=c(0,0.1,0.1), fn=nll.exg, data = data$rt)
$par
[1] 0.7151246 0.3363003 0.4646291

$value
[1] 794.2717

$counts
function gradient
     150       NA

$convergence
[1] 0

$message
NULL
```

Example 3

As before, let's check our model fit

```
x=seq(0,4,0.1)
plot(density(data$rt))
lines(x,dexg(x,mu=0.715, sigma=0.336, tau=0.465),lty=2)
```