Lecture 2 – Intro to Maximum Likelihood Estimation (MLE)

MLE = technique for estimating model parameters.

↳ has some advantages over least-squares estimation (i.e, the "lm" function in R)

To understand MLE, need to discuss the difference between probability and likelihood

Simple example: modeling sequences of trials with only two possible outcomes (success, failure).

Typical question: if probability of success on each trial is only 20% (i.e, 0.20), what is the probability of getting 3 successes in 10 trials?

Ans: use "binomial distribution"

$$P = \binom{10}{3} (0.2)^3 \cdot (0.8)^7 =$$

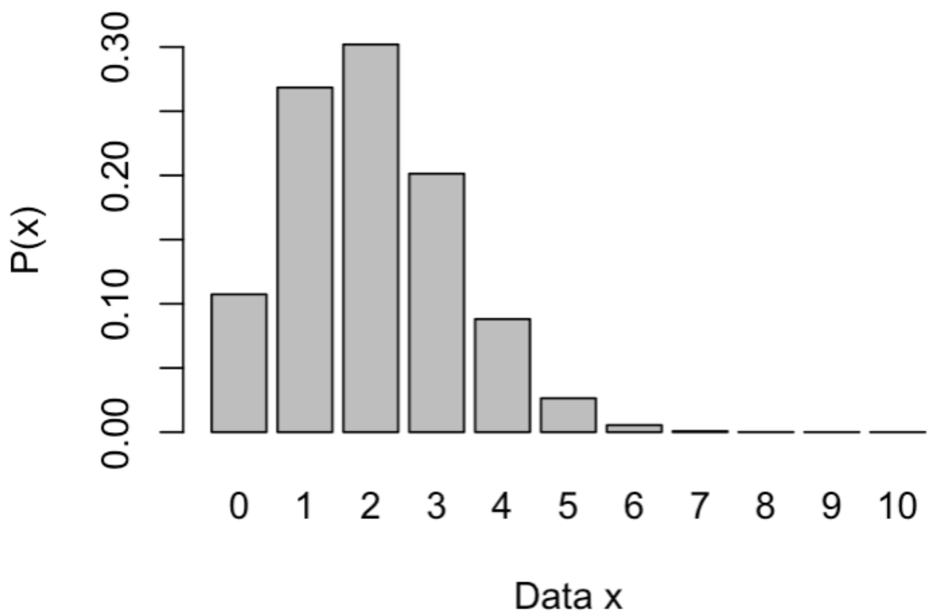
```
# Example: w = 0.2, N = 10 trials, x = 3,4 successes
choose(10,3) * (0.2)^3 * (0.8)^7
```

In R, we can use the "dbinom" function.

```r
# R has a built-in function for this
dbinom(x=3, size=10, prob=0.2)
```

If we plot the probabilities $P(x)$ for <u>all</u> possible observed data x, we get a probability distribution

```r
# plot "probability density function"
N = 10
w = 0.2 # change to 0.7 and see effect
x = seq(from=0, to=N, by=1)
barplot(dbinom(x, size=N, prob=w),
        names.arg = 0:N,
        xlab="Data x",
        ylab="P(x)")
```

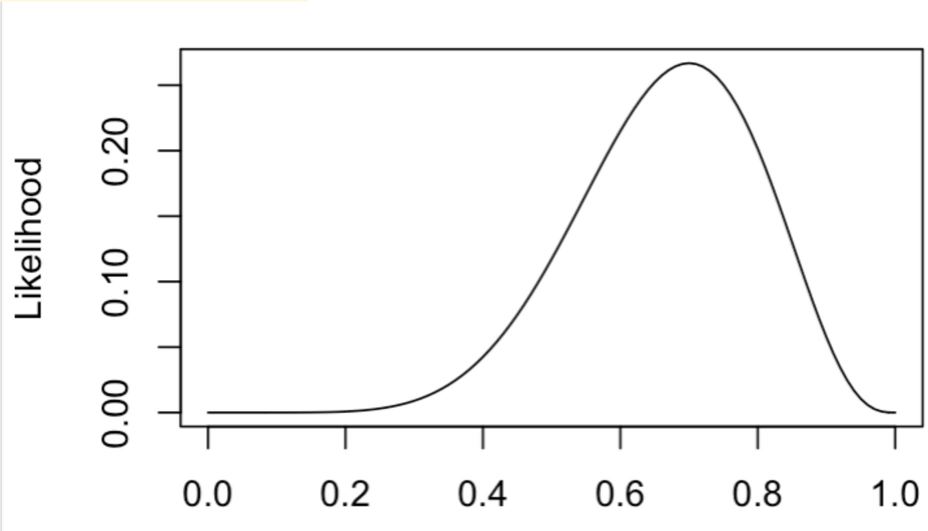


Note: this graph is for a <u>fixed</u> parameter value (i.e, $w = 0.2$).

What if we are given the data x, but NOT the parameter w?

Example: suppose we observed $x = 7$ successes in 10 trials. What is the value of w (i.e, the prob of success on any one trial)?

```r
w = seq(from=0, to=1, by=0.01)
x = 7
plot(w, dbinom(x, size=N, prob=w),
      type="l",
      ylab="Likelihood")
```



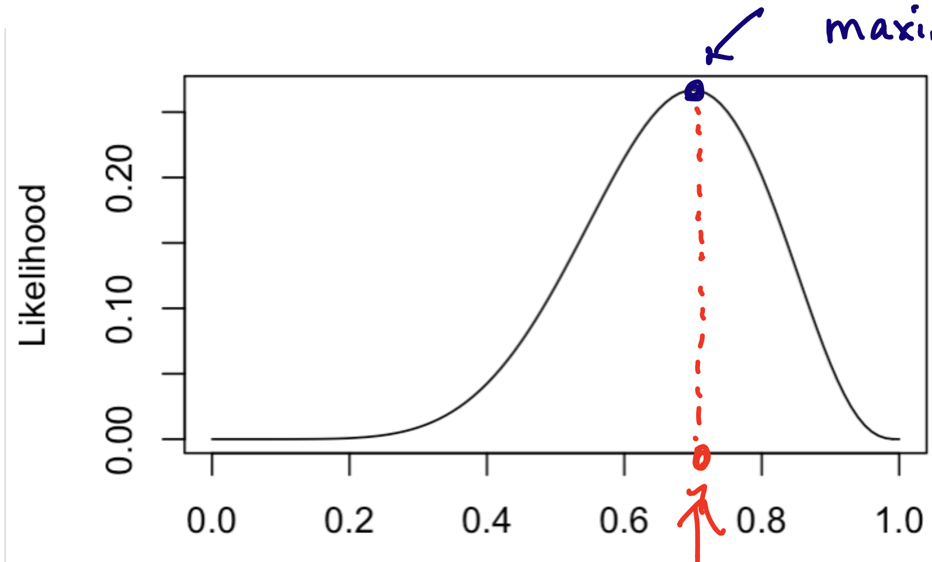
* this is a likelihood function
* the data is fixed
* Probability = given parameter, find $p(\text{data})$

  Likelihood = given data, find $l(\text{parameter})$

MLE = maximum likelihood estimation

↳ given observed data, find parameter value(s) that maximize the likelihood function.

"maximum likelihood"



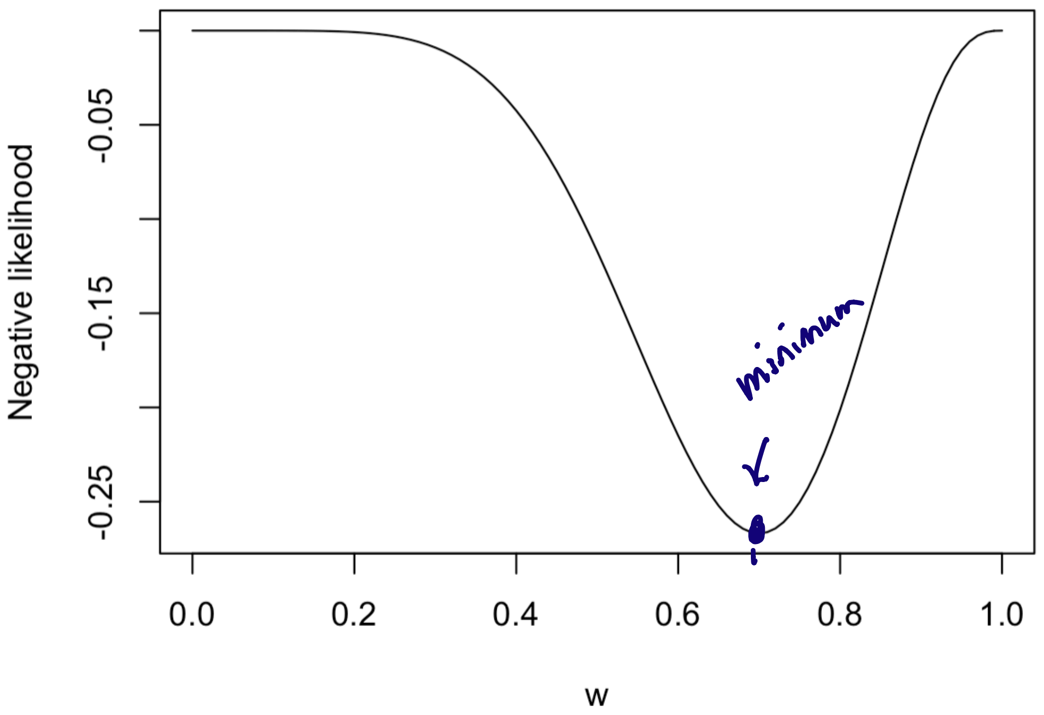
what is this value of ω?

In R, we use the "optim" function. It needs 3 things:

1. the observed data
2. a function to minimize
3. an initial guess

Note: maximizing likelihood is equivalent to minimizing the negative of likelihood

```r
plot(w, -dbinom(x=7, size=10, prob=w),
     type="l",
     ylab="Negative likelihood")
```



Define "objective" function (the thing to minimize)

```r
# define NL (function with two variables - "data" and "parameter values")
nl.binom = function(data, par){
  -dbinom(data, size=10, prob=par)
}
```

Perform optimization:

nl.binom

```
> optim(par=0.5, fn=nll.binom, data=7)
$par
[1] 0.7

$value
[1] -0.2668279

$counts
function gradient
      30       NA

$convergence
[1] 0

$message
NULL
```

MLE

Other outputs:

$value – the y-axis on the NL graph at MLE
(not really needed)

$counts – # of steps in minimization algorithm

$convergence – did optimizer converge?
↳ 0 ⇒ yes!
≠0 → see help file (?optim)