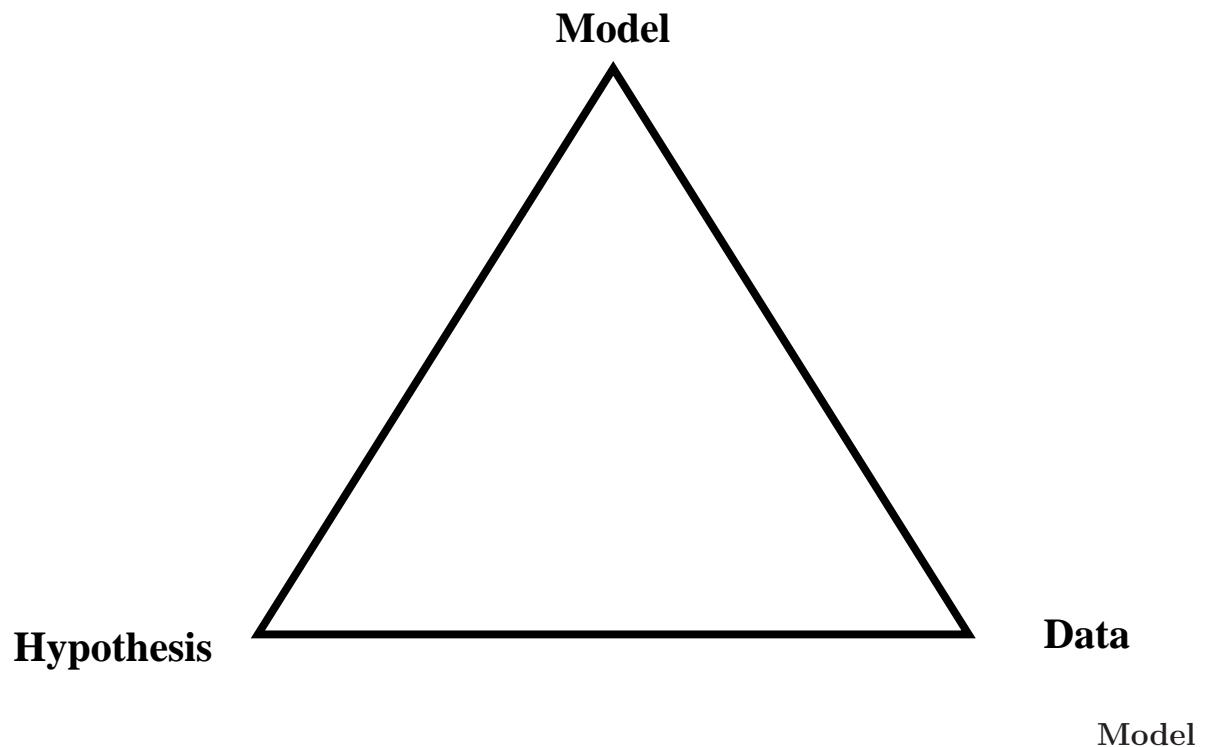


- So far we have taken a hypothesis testing approach
 - compare distribution to expected and quantify the difference with a test statistic
 - Compare test-statistic to a sampling distribution
- Distribution fitting can also be considered an estimation procedure
 - If one knows that the data come from a particular probability distribution, distribution fitting can be used to *estimate* the parameters of that distribution (location parameters and shape parameters)

Estimation Triangle



The model refers to a probability model

- common pdf
 - discrete (binomial, multinomial, Poisson)
 - continuous (gaussian, F , χ^2)

- Maybe more interestingly, complex distributions (mixtures of multiple distributions, conditional distributions)

Hypothesis

The hypothesis refers to the actual parameter values that specify the form of the model

- μ, σ^2, λ

Data

The data are observations that could have come from this model under these parameter values

What is an estimator?

To estimate a feature of a population distribution, one uses a rule or *estimator*
A good estimator:

- is consistent
- is unbiased
- exhibits low variance

$$\bar{x} = \frac{\sum_i^n x_i}{n}$$

Methods of estimation

- Least-squares
- Method of Moments
- Maximum likelihood

Maximum likelihood

Maximum likelihood estimators depend on a *Likelihood function*. The likelihood function provides the probability of an observation or set of observations given a probability model.

$$L = p(D|M, \theta)$$

Where D is data, M is a model and θ represents the unknown parameters
The probability model includes the unknown parameter one is attempting to estimate.

Characteristics of MLE

The values of the parameter(s) that result in the highest probability of observing the data are the maximum likelihood estimates of those population parameters

Maximum likelihood estimators are usually (always?) consistent and are often unbiased and exhibit low variance

Example: Binomial

Remember the binomial distribution:

The parameter of this model is p . So the likelihood of p given some data is:

$$L(p|Y, k) = \binom{k}{Y} p^Y q^{k-Y} = \frac{k!}{Y!(k-Y)!} p^Y q^{k-Y}$$

Notice that the parameter p does not depend on the binomial coefficient. That implies for Maximum likelihood estimation, we can just use:

$$L(p|Y, k) = p^Y q^{k-Y}$$

Analytical solution

It is possible to estimate p directly from the data in this case (just a little calculus):

$$p = \frac{Y}{k + Y}$$

This is the value of p that maximizes the quantity in the previous slide.

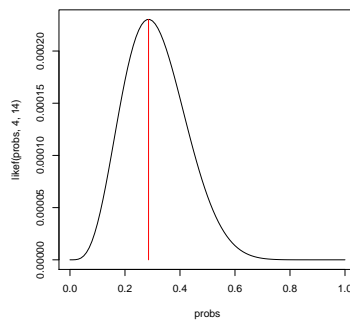
Numerical approach

A numerical solution requires a search across lots of different parameter values to find the one that maximizes the likelihood function

Figure

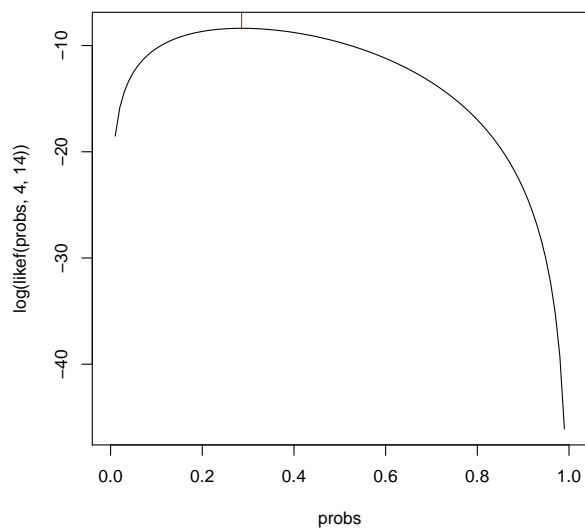
```
> likef <- function(p, suc, tot) {  
+   ((p^suc) * ((1 - p)^(tot - suc)))  
+ }  
> probs <- seq(0, 1, 0.01)  
  
> plot(likef(probs, 4, 14) ~ probs, type = "l")  
> points(c((4/14), (4/14)), c(0, likef((4/14), 4, 14)), col = "red",  
+   type = "l")  
> 4/14
```

[1] 0.2857143



Logarithm version

```
> plot(log(likef(probs, 4, 14)) ~ probs, type = "l")
> points(c((4/14), (4/14)), c(0, log(likef((4/14), 4, 14))), col = "red",
+       type = "l")
```



finding maxima

`optim()` is used to maximize functions in R. It requires a function to optimize with parameters. `likef()` shows the form with a vector `p` as the first parameter and other required data as subsequent parameters.

```

> llikef <- function(p, suc, tot) {
+   if ((p <= 1) & (p >= 0)) {
+     -log(((p^suc) * ((1 - p)^(tot - suc))))
+   }
+   else {
+     Inf
+   }
+ }
> opt <- optim(c(0.5), llikef, suc = 4, tot = 14)
> opt$par

[1] 0.2857422

```

Mult paramters, continuous

So let's fit some data to a normal distribution:

```

> fitdat <- rnorm(100, 3, 2)
> llike.n <- function(p, dat) {
+   if (p[2] > 0) {
+     -sum(dnorm(dat, mean = p[1], sd = p[2], log = T))
+   }
+   else {
+     Inf
+   }
+ }
> opt.n <- optim(c(0, 1), llike.n, dat = fitdat)
> opt.n

$par
[1] 2.950750 2.009609

$value
[1] 211.6833

$counts
function gradient
      75      NA

$convergence
[1] 0

$message
NULL

```

```
> mean(fitdat)
```

```
[1] 2.950929
```

```
> sd(fitdat)
```

```
[1] 2.019641
```

Slick R implementation

This works well for many continuous distributions, provides the estimate, but not the likelihood of the estimate

```
> library(MASS)
```

```
> fitdistr(fitdat, "normal")
```

mean	sd
2.9509285	2.0095178
(0.2009518)	(0.1420944)

Likelihood ratio tests

It is possible to estimate the likelihood of

- multiple hypotheses (as seen above)
- multiple models

As long as the data are unchanged among estimates

Since the likelihoods are proportional to probabilities, the ratio of likelihoods is the ratio of probabilities of the two models. (Again, as long as the data do not change.)

Furthermore, it has been shown that twice the log of the likelihood ratio is χ^2 distributed, so significance tests can be constructed when models differ in parameters.

Modeling a distribution

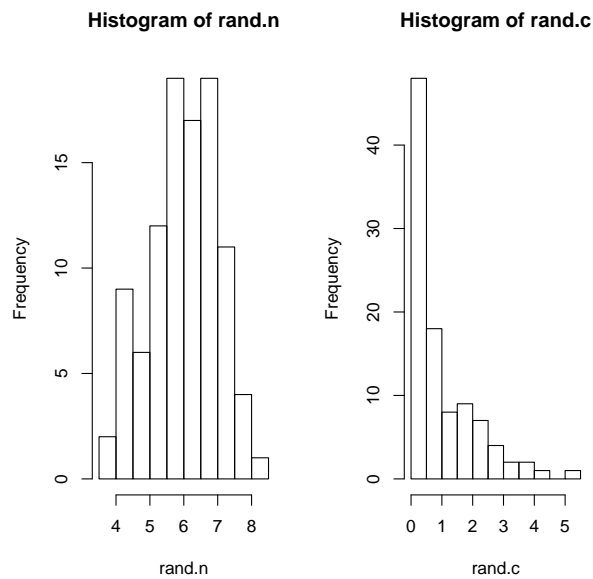
```
> rand.n <- rnorm(100, 6, 1)
```

```
> rand.c <- rchisq(100, 1)
```

```
> par(mfrow = c(1, 2))
```

```
> hist(rand.n)
```

```
> hist(rand.c)
```



cont

```
> llike.n <- function(p, dat) {
+   if (p[2] > 0) {
+     -sum(dnorm(dat, mean = p[1], sd = p[2], log = T))
+   }
+   else {
+     Inf
+   }
+ }
> llike.c <- function(p, dat) {
+   if (p[1] > 0) {
+     -sum(dchisq(dat, df = p[1], log = T))
+   }
+   else {
+     Inf
+   }
+ }
```

Likelihood of χ^2

```
> opt.c.c <- optim(c(1), llike.c, dat = rand.c)
> opt.n.c <- optim(c(1), llike.c, dat = rand.n)
> opt.c.c$value
```

```
[1] 65.95452  
  
> opt.c.c$par  
  
[1] 0.9304687  
  
> opt.n.c$value  
  
[1] 218.7717  
  
> opt.n.c$par  
  
[1] 6.946875
```

Likelihood of the normal dist

```
> opt.c.n <- optim(c(1, 1), llike.n, dat = rand.c)  
> opt.n.n <- optim(c(1, 1), llike.n, dat = rand.n)  
> opt.c.n$value  
  
[1] 200.2113  
  
> opt.c.n$par  
  
[1] 0.9404736 1.7916551  
  
> opt.n.n$value  
  
[1] 143.0327  
  
> opt.n.n$par  
  
[1] 6.060230 1.011655
```