Likelihoods

Definition: the joint density ("probability" if discrete) of the observed data, viewed as a function of the unknown parameters.

Example: coin flips

data: HHHT

Model: prob(Heads) = p
prob(Tails) = $(1-p)$

It's easy to write the probability of coin flips given p.

$p \longrightarrow$ H H H T is easy. If coin fair, then prob(data) $= \left(\frac{1}{2}\right)^4$

Statisticians are interested in the inverse problem:
Given data, what can we say about p?

the parameter of the data generative model

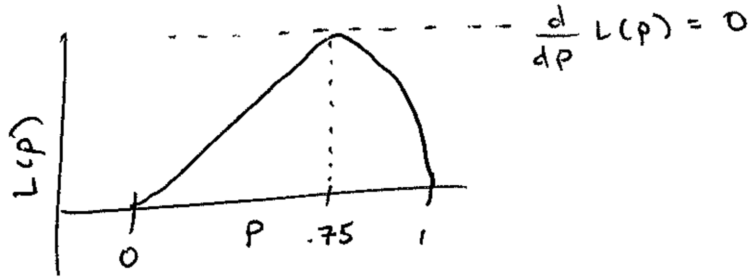
$p \xleftarrow{?}$ H H H T

Guess: $\hat{p} \stackrel{?}{=} \frac{3}{4}$

The likelihood: $L(p) = p^{3}(1-p)^{1}$

data

Plot $L(p) = p^3(1-p)$, $p \in [0,1]$:



How to find maximum of $L(p)$? i.e. how to find $\underset{p}{\operatorname{argmax}} L(p)$? Take derivative, set = 0.

But wait!

Take the log first for (1) numerical stability and (2) easier derivative.

This works because log is monotonic:

i.e. $a \le b \Rightarrow \log(a) \le \log(b)$

Easy to see since $\frac{d}{dx} \log(x) = \frac{1}{x} > 0$

it suffices to find where $\frac{d}{dp} \log L(p) = 0$

Let's do it!

$$\frac{d}{dp} \log L(p) = \frac{d}{dp} 3\log(p) + \log(1-p)$$

$$= \frac{3}{p} - \frac{1}{1-p} \quad \text{set = to } 0:$$

$$3 = 4\hat{p}$$

$$\boxed{\hat{p}_{MLE} = \frac{3}{4}}$$

The maximum likelihood estimate of p is $\frac{3}{4}$.

Exercise: normal likelihood

Let $y_i | x_i \overset{iid}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$.

What is the likelihood for the data $y_1, \ldots, y_n | x_1, \ldots, x_n$?

Solution: Note: $y \sim N(\mu, \sigma^2) \Rightarrow p(y) = (2\pi\sigma^2)^{-1/2} \exp\{\frac{-1}{2\sigma^2}(y-\mu)^2\}$

The likelihood is the joint density of the data viewed as a function of the parameters:

$$p(y_1, \ldots, y_n | x_1, \ldots, x_n) = \prod_{i=1}^{n} p(y_i | x_i) = L(\underbrace{\beta_0, \beta_1, \sigma^2}_{\text{unknown parameters}}).$$

$$\Rightarrow L(\beta_0, \beta_1, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\{\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2\}$$

What about the MLE?

$$\frac{d}{d\beta_0} \log L = \frac{d}{d\beta_0}\left(\cancel{\frac{-n}{2} \log 2\pi\sigma^2}\right) + \frac{d}{d\beta_0}\left(\frac{-1}{2\sigma^2} \underbrace{\sum (y_i - \beta_0 - \beta_1 x_i)^2}_{\boxed{SSR}\ !}\right)$$

0

set = to 0:

$$0 = \frac{\cancel{2}}{\cancel{2}\sigma^2} \Sigma (y_i - \beta_0 - \beta_1 x_i)$$

$$\Rightarrow n\bar{y} - n\beta_0 - n\bar{x}\beta_1 = 0$$

$$\Rightarrow \boxed{\hat{\beta}_{0_{MLE}} = \bar{y} - \bar{x}\beta_1}$$

... same as OLS, notice

Can do same w/ β_1 to find $\hat{\beta}_{1MLE}$.

" " " " σ^2 " " $\hat{\sigma}^2_{MLE}$.

$$\frac{d}{d\sigma^2} \log L = \frac{d}{d\sigma^2}\left(\frac{-n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2} SSR\right) = 0$$

$$= \frac{-n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} SSR = 0$$

$$\Rightarrow \frac{1}{(\sigma^2)^2} \cdot SSR = \frac{n}{\sigma^2}$$

$$\Rightarrow SSR = n\sigma^2$$

$$\Rightarrow \boxed{\frac{SSR}{n} = \hat{\sigma}^2_{MLE}} \approx \frac{RSS}{n-p} \equiv \hat{\sigma}^2 \quad \text{for large } n \gg p.$$

close to our usual estimator

Exercise: $y | X \sim MVN(X\beta, I\sigma^2)$

What's the likelihood of one vector y given X?

$$(2\pi)^{-n/2} \underbrace{\det(I\sigma^2)^{-1/2}}_{(\sigma^2)^{-n/2}} \exp\left\{ \frac{-1}{2}(y - X\beta)^T (I\sigma^2)^{-1}(y - X\beta)\right\}$$

Recall: determinant of diagonal matrix = product of diagonals!

Notice this is identical to simple linear regression likelihood above: $(y - X\beta)^T(y - X\beta) = \sum_{i=1}^{n} (y_i - x_i\beta)^2 = SSR$

Moreover:

$$\frac{d}{d\beta} \log L(\beta, \sigma^2) = \frac{d}{d\beta}\left(\frac{-1}{2\sigma^2}(y - X\beta)^T(y - X\beta)\right) = 0$$

$$\Rightarrow \frac{d}{d\beta} SSR(\beta) = 0$$

$$\Rightarrow \hat{\beta}_{MLE} = \hat{\beta}_{OLS} = (X^T X)^{-1} X^T y$$

MLEs have nice properties we will discuss next time...

Asymptotic
1. consistency
2. efficiency
3. normality