Maximum Likelihood Estimation

MLE is one of the fundamental tools used in statistical inference.

It consists of modeling the process of interest with probabilistic statements and estimates the parameters as the ones that maximize the likelihood (probability of the data given the parameters).

Maximum Likelihood Estimate Converges to the True Parameter

$$\hat{\theta}_{\text{mle}} \rightarrow \theta_{\text{true}} + \frac{1}{\sqrt{n}} \mathcal{N}\left(0, \mathbb{F}^{-2}\right)$$

where $\mathbb{F}$ is the Fisher information matrix

$$\mathbb{F}_{jk} = E\left[-\frac{\partial^2 \log f_{\theta_{\text{true}}}(X)}{\partial\theta_j\, \partial\theta_k}\right]$$

2

The maximum likelihood estimates converges to the true parameter as the sample size *n* grows (under some reasonable regularity conditions).

Likelihood

$$P(\text{data} \,|\, \text{model parameters})$$

3

The likelihood is defined as the probability of the data given (conditional on) the model parameters.

Exercise: calculate the likelihood with known parameter

$$P(\text{data} \,|\, \text{model parameters})$$

- We have a coin that has probability 0.5 of landing head
- Calculate likelihood, P(data | prob H 0.5), if
  - we toss a coin once and get H
    - P(H | fair coin)
  - we toss a coin 10 times and get (H,H,T,T,H,H,H,H,H,H)

4

Exercise: calculate the likelihood with unknown parameter

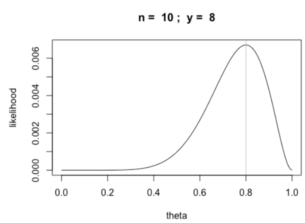
$$P(\text{data} \mid \text{model parameters})$$

- We have a coin that has probability θ of landing head
  - and (1-θ) of landing tail
- Calculate likelihood, P(data l prob H = θ), if
  - we toss a coin once and get H
  - we toss a coin 10 times and get (H,H,T,T,H,H,H,H,H,H)

Exercise: calculate maximum likelihood estimate of θ

likelihood $= \theta^8(1-\theta)^2$

which θ maximizes the likelihood?



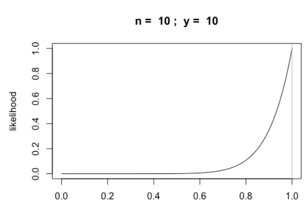
MLE is the value of the parameter that maximizes the likelihood of the data under the chosen model (i.e. parametrization). How would you calculate θmle?

If the likelihood is differentiable, we can use the fact that the derivative of the function is 0. What happens if the maximum were at the boundary of the parameter space?

Exercise: calculate maximum likelihood estimate of θ

likelihood $= \theta^{10}$

which θ maximizes the likelihood?



what happens if all tosses land on head?

what is the MLE?

Homework: calculate MLE θ for *n* tosses with *y* heads

likelihood $= \binom{n}{y}\theta^y(1-\theta)^{n-y}$ explain why

show that the maximum likelihood estimate of θ is

$$\hat{\theta}_{\text{mle}} = \frac{y}{n}$$

Why did we add the n choose y factor here?

Does it change the MLE?

Exercise: write likelihood of normal r.v.

$$X_i \sim N(\mu, \sigma^2)$$

We have 2 data points

iid: independent and identically distributed

$x_1 = 0.8$
$x_2 = -2.4$

Normal $\mathcal{N}(\mu, \sigma^2)$ $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$ $x \in (-\infty, \infty)$

Remarks

- We use log likelihood because they are more convenient
- Likelihood of iid rv. are products of single likelihoods
  - When log transforming, we get a sum
- how do the estimate differ when we calculate the maximum likelihood or the maximum log likelihood?

how do the estimates differ when we look for the parameter that maximizes the likelihood or the log likelihood? Why?