Maximum Likelihood Estimation

Estimation

Goal: Estimate a population parameter θ.

Data: $X_1, X_2, \ldots, X_n \sim$ iid with distribution depending on θ.

If one has many estimators to choose from, pick

- That with the smallest SE, among all unbiased estimators.
- That with the smallest RMS error, even if biased.

$\longrightarrow$ Sometimes it is not clear how to form even one good estimator.

Maximum likelihood estimation

Likelihood function: $L(\theta) = Pr(\text{data} \mid \theta)$

Log likelihood: $l(\theta) = \log Pr(\text{data} \mid \theta)$

Maximum likelihood estimate:

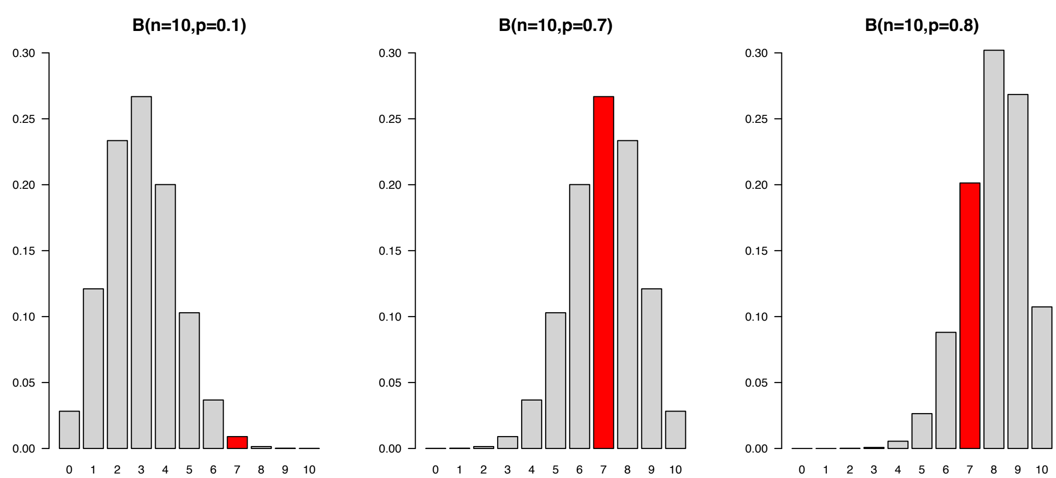
$\longrightarrow$ Choose, as the estimate of θ, the value of θ for which the likelihood function $L(\theta)$ (or equivalently, the log likelihood function) is maximized.

You need to solve these equations analytically or numerically!

Likelihood

Imagine we have a Binomial(n=10,p) and observe 7 successes.

$L(p) = Pr(7 \text{ successes out of 10 trials} \mid p)$



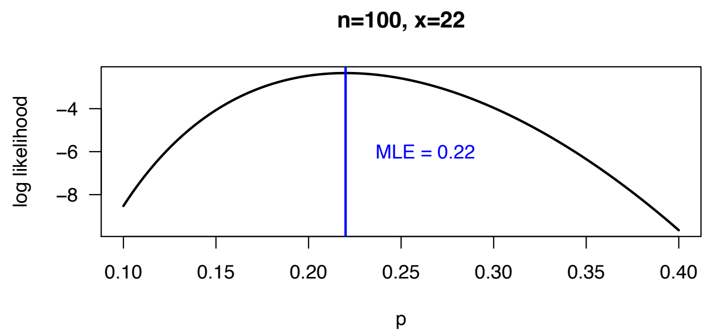
Example 1

Suppose $X \sim \text{Binomial}(n, p)$.

log likelihood function: $l(p) = \log\left\{\binom{n}{x} p^x (1-p)^{(n-x)}\right\}$

$$= x \log(p) + (n-x) \log(1-p) + \text{constant}$$

MLE: the obvious thing: $\hat{p} = x/n$



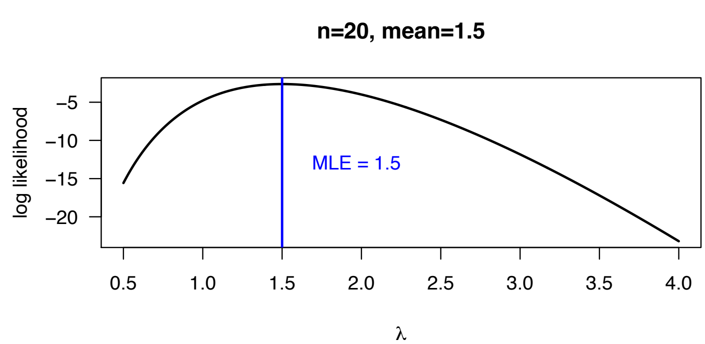
Example 2

Suppose $X_1, \ldots, X_{20} \sim$ iid Poisson(λ).

log likelihood function: $l(\lambda) = \log\left\{\prod_i e^{-\lambda} \lambda^{x_i}/x_i!\right\}$

$$= \ldots = -20\lambda + \left(\sum x_i\right) \log\lambda + \text{ constant}$$

MLE: the obvious thing: $\hat{\lambda} = \bar{x}$



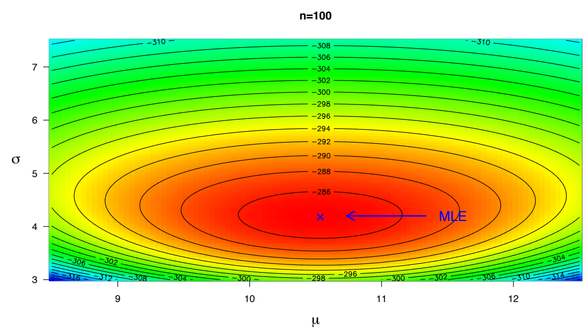
Example 3

Suppose $X_1, \ldots, X_n \sim$ iid $N(\mu, \sigma)$

log likelihood function: $l(\mu, \sigma) = \log \left\{ \prod_i \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2 \right] \right\}$

MLEs: almost the obvious things:

$$\hat{\mu} = \bar{x} \qquad \hat{\sigma} = \sqrt{\sum (x_i - \bar{x})^2 / n}$$



About MLEs

Maximum likelihood estimation is a general procedure for finding a reasonable estimator

- In many cases, the MLE turns out to be the obvious thing.
- MLEs are often very good (but not necessarily the best) possible estimators:
  - ○ Unbiased or nearly unbiased.
  - ○ Small standard errors.
- Sometimes obtaining the MLEs requires hefty computation!

Example 4: ABO blood groups

Phenotype	Genotype	Frequency
O	OO	p_O^2
A	AA or AO	$p_A^2 + 2p_Ap_O$
B	BB or BO	$p_B^2 + 2p_Bp_O$
AB	AB	$2p_Ap_B$

Frequencies under the assumption of Hardy-Weinberg equilibrium.

Example 4: Data

Phenotype	No. subjects	% subjects
O	117	46.8%
A	98	39.2%
B	29	11.6%
AB	6	2.4%
Total	**250**	**100%**

$\longrightarrow$ What are the estimates of p_A, p_B, p_O?

Example 4: Estimates

Simple estimates:

$\longrightarrow$ $\tilde{p}_O = \sqrt{0.468} = 0.684$

$\longrightarrow$ $\tilde{p}_A^2 + 2\tilde{p}_A 0.684 = 0.392 \longrightarrow \tilde{p}_A = 0.243$

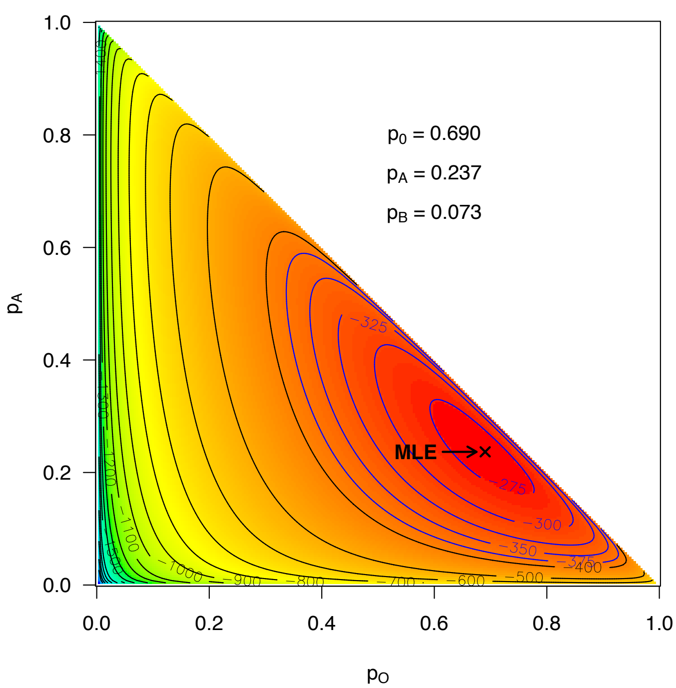
$\longrightarrow$ $\tilde{p}_B = 0.024/(2\tilde{p}_A) = 0.072$

Log likelihood Remember the Multinomial distribution function!

$l(p_O, p_A, p_B) =$

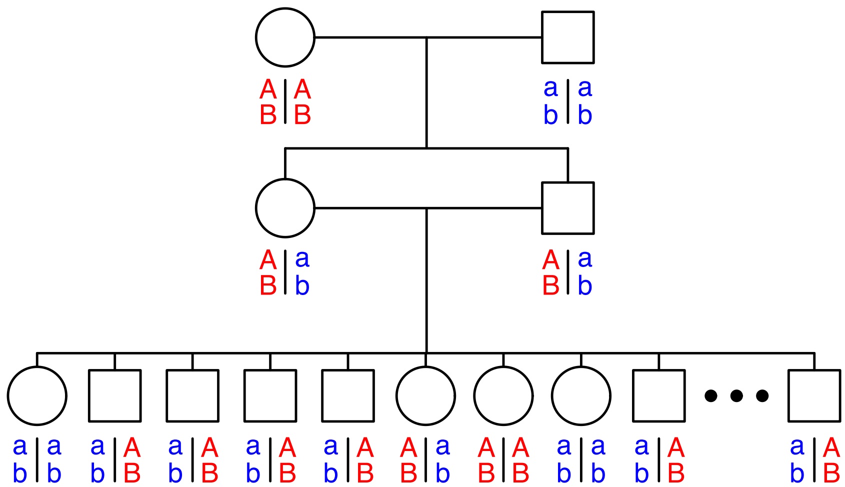
$117 \log(p_O^2) + 98 \log(p_A^2 + 2p_A p_O) + 29 \log(p_B^2 + 2p_B p_O) + 6 \log(2p_A p_B)$

Example 4: log likelihood



Example 5

Consider the problem of estimating the recombination fraction (call that parameter θ) between two genetic markers in an intercross.



$\longrightarrow$ Note: We won't observe the haplotypes.

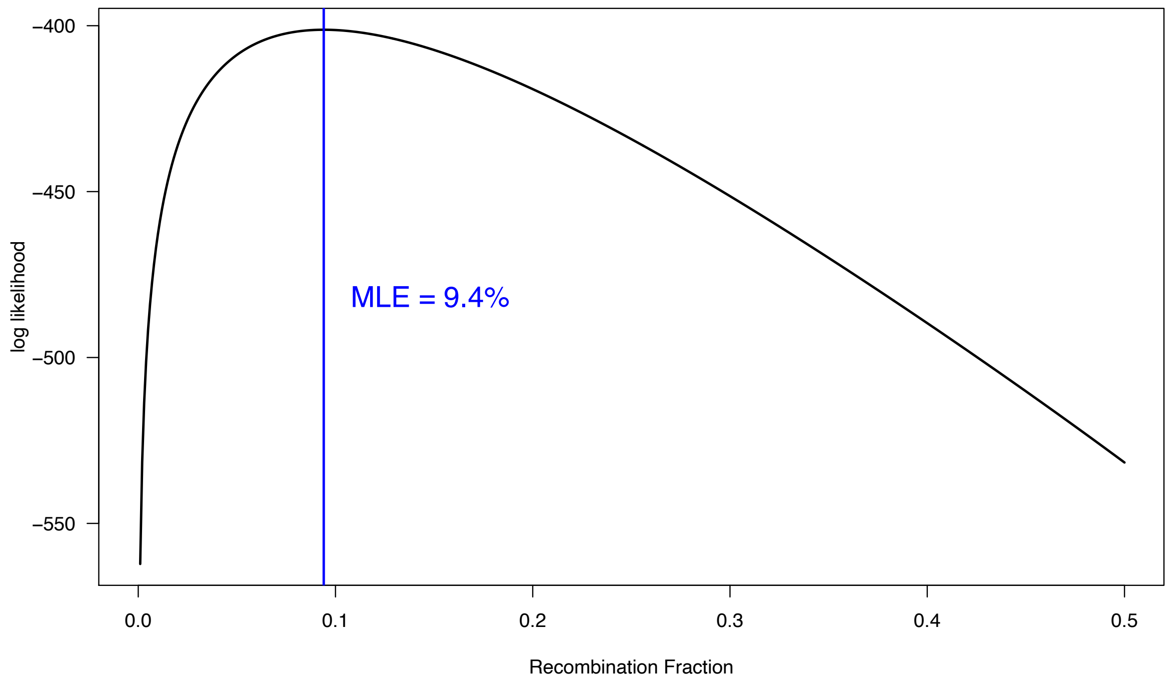
Example 5

	Data			Probabilities		
	AA	Aa	aa	AA	Aa	aa
BB	58	9	0	$\frac{1}{4}(1-\theta)^2$	$\frac{1}{2}\theta(1-\theta)$	$\frac{1}{4}\theta^2$
Bb	8	95	14	$\frac{1}{2}\theta(1-\theta)$	$\frac{1}{2}[\theta^2+(1-\theta)^2]$	$\frac{1}{2}\theta(1-\theta)$
bb	1	12	53	$\frac{1}{4}\theta^2$	$\frac{1}{2}\theta(1-\theta)$	$\frac{1}{4}(1-\theta)^2$

$\longrightarrow$ Possible estimates of the recombination fraction, θ?

$$L(\theta) \propto \left\{\tfrac{1}{4}(1-\theta)^2\right\}^{(58+53)} \times \left\{\tfrac{1}{2}\theta(1-\theta)\right\}^{(9+8+14+12)} \times \left\{\tfrac{1}{4}\theta^2\right\}^{(1+0)} \times \left\{\tfrac{1}{2}[\theta^2+(1-\theta)^2]\right\}^{95}$$

Example 5: log likelihood function



A closer view