

Note: We won't observe the **haplotypes**.

Example 1

Data

	AA	Aa	aa
BB	58	9	0
Bb	8	95	14
bb	1	12	53

Probabilities

	AA	Aa	aa
BB	$\frac{1}{4} (1 - \theta)^2$	$\frac{1}{2} \theta(1 - \theta)$	$\frac{1}{4} \theta^2$
Bb	$\frac{1}{2} \theta(1 - \theta)$	$\frac{1}{2} [\theta^2 + (1 - \theta)^2]$	$\frac{1}{2} \theta(1 - \theta)$
bb	$\frac{1}{4} \theta^2$	$\frac{1}{2} \theta(1 - \theta)$	$\frac{1}{4} (1 - \theta)^2$

Question: Possible estimates of the recombination fraction, θ ?

Maximum likelihood estimation

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

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

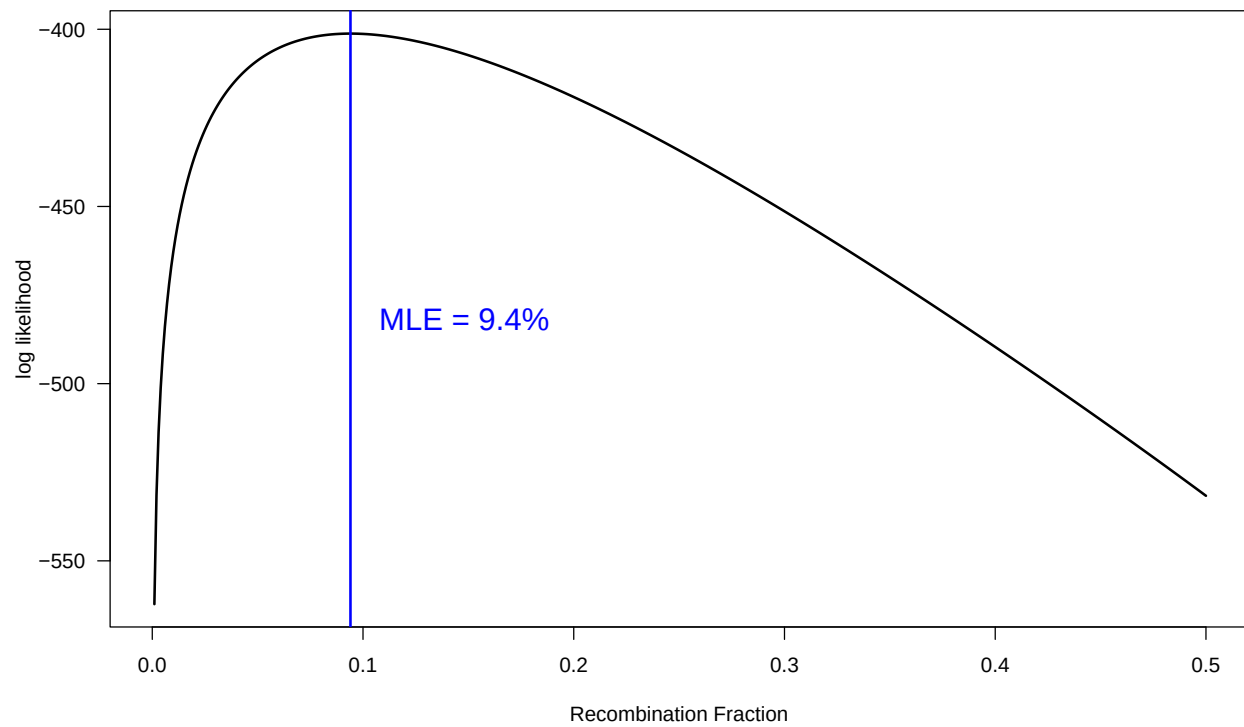
Maximum likelihood estimate:

Choose, as the estimate of θ , the value of θ for which $L(\theta)$ is **maximized**.

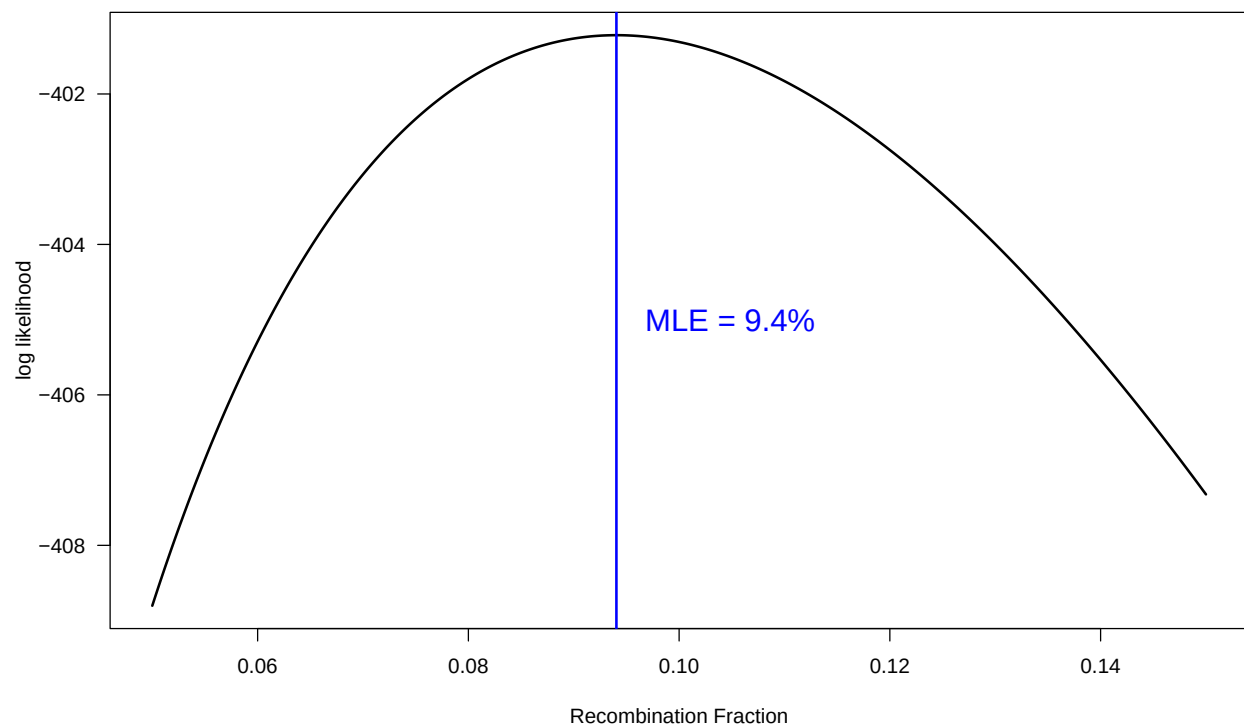
For the example,

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

Example 1: Log likelihood function



A closer view



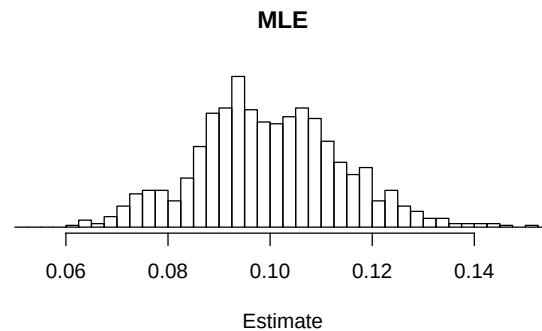
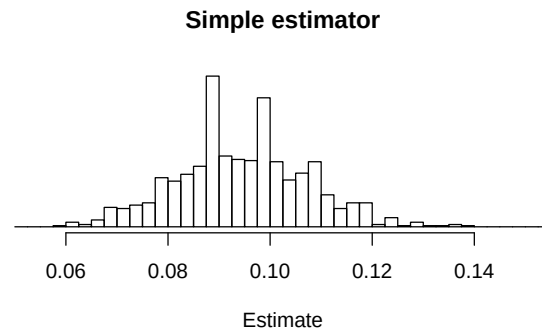
A comparison of two estimators

A simple estimator:

Assume double-heterozygotes are non-recombinant

For the case $n = 250$, $\theta = 0.10$; the results of 1000 simulations:

	Simple	MLE
Bias	-0.005	0.000
SE	0.013	0.014
RMSE	0.014	0.014

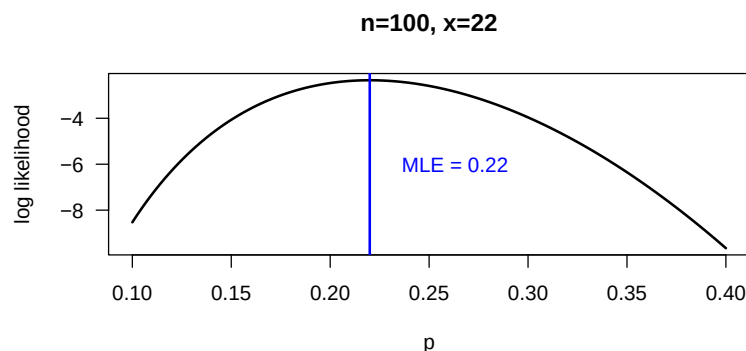


Example 2

Suppose $x \sim \text{binomial}(n, p)$.

$$\begin{aligned}\text{log likelihood function: } l(p) &= \log \left\{ \binom{n}{x} p^x (1-p)^{(n-x)} \right\} \\ &= x \log(p) + (n-x) \log(1-p) + \text{constant}\end{aligned}$$

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

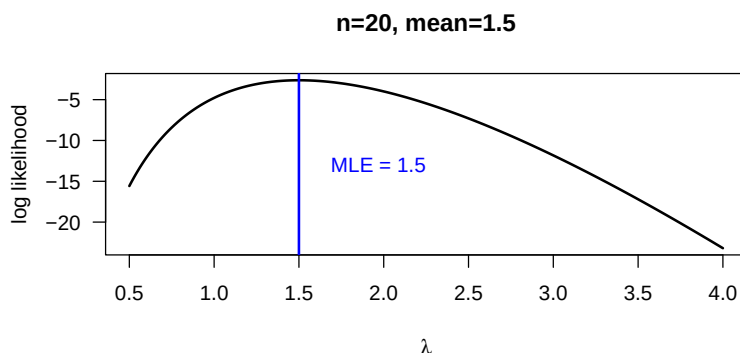


Example 3

Suppose $x_1, \dots, x_{20} \sim \text{iid Poisson}(\lambda)$.

$$\begin{aligned} \text{log likelihood function: } l(\lambda) &= \log \left\{ \prod_i e^{-\lambda} \lambda^{x_i} / x_i! \right\} \\ &= \dots = -20\lambda + (\sum x_i) \log \lambda + \text{constant} \end{aligned}$$

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



Example 4

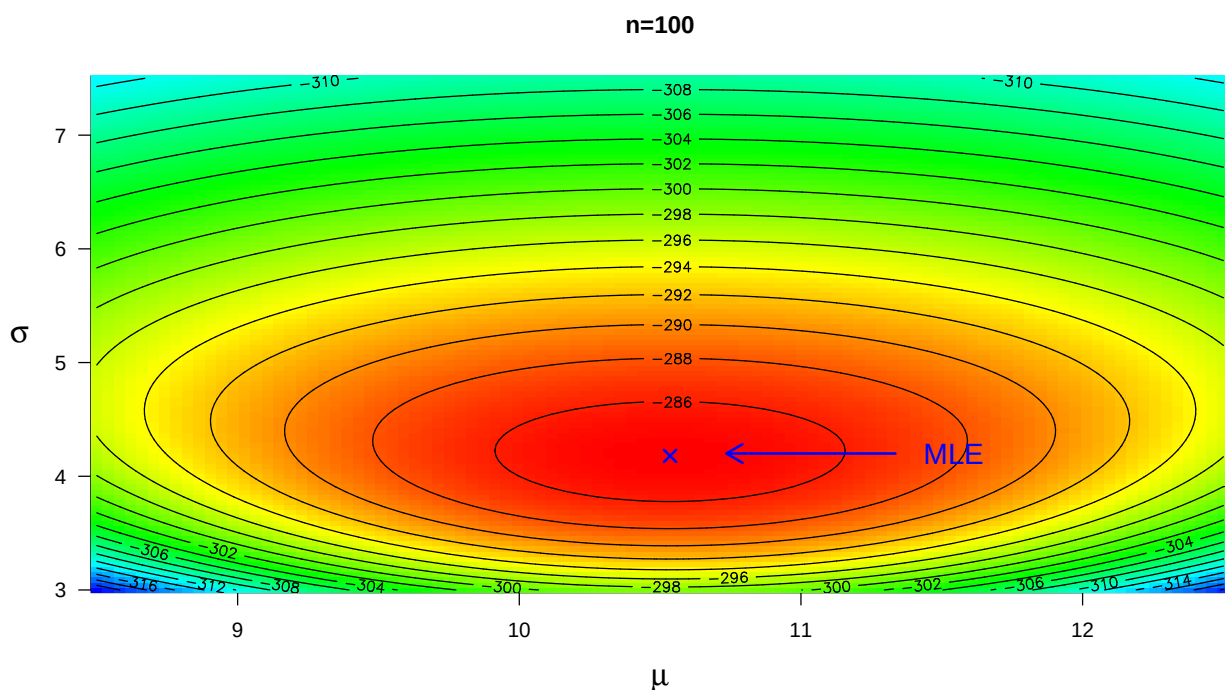
Suppose $x_1, \dots, x_n \sim \text{iid } N(\mu, \sigma)$

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

MLEs: almost the obvious things:

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

Example 4: the log likelihood surface



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 good (but not necessarily the best) estimators
 - Nearly unbiased
 - small SE
- Sometimes obtaining the MLEs requires hefty computation

Example 5: 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 5: Data

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

Question: Estimates of p_A , p_B , p_O ?

Example 5: Estimates

Simple estimates

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

$$\tilde{p}_A \text{ to solve } \tilde{p}_A^2 + 2\tilde{p}_A 0.684 = 0.392$$

$$\implies \tilde{p}_A = 0.243$$

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

Log likelihood:

$$l(p_O, p_A, p_B) =$$

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

Example 5: log likelihood

