Statistical Genetics I: STAT/BIOST 550
Spring Quarter, 2014

Elizabeth Thompson
University of Washington
Seattle, WA, USA

MWF 8:30-9:20; THO 211

Web page: www.stat.washington.edu/∼thompson/Stat550/

Overview

Chapter 1: Genes in Populations, across Chromosomes

Chapter 2: Genes in relatives: at a locus

Chapter 3: Genes on Chromosomes: Two-locus linkage analysis

Chapter 4: Multilocus analysis methods

This is StatGen-I (STAT/BIOST 550). It does provide a self-contained course, but it is only the beginning.

StatGen-II (BIOST/STAT 551) continues to analysis of quantitatve genetic variation, while StatGen-III (BIOST/STAT 552) brings it all together with design and analysis of genetic studies.

Chapter 1: Genes in Populations

1.1.1 GENETIC TERMINOLOGY:

Genome – the complete set of DNA base pairs (bp) of an organism

Chromosome— long string of double-strand DNA (part of the genome)

Locus— position on a chromosome, or DNA at that position, or the piece of DNA coding for a trait.

Allele— type of the DNA at a particular locus on chromosome

Haplotype — the sequence of alleles along a chromosome

Diploids— organisms with two copies of the genome

Genotype— (unordered) pair of alleles at a particular locus in a particular individual.
Homozygote– a genotype with two like alleles.
Heterozygote – a genotype with two unlike alleles.

Phenotype— observable characteristics of an individual

1.1.2 EXAMPLE: ABO blood types:

The human (haploid) genome is 3×10^9 base-pairs of DNA.
Chromosomes range from 250 Mbp down to about 50 Mbp.

Humans are diploid. Cell nucleus — has 46 chromosomes
(22 pairs of autosomes, and 2 sex chromosomes, X,Y)

The ABO locus is on chromosome 9
The (main) alleles at the locus are A, B, and O
The 6 genotypes are AA, AO, BB, BO, AB and OO
Homozygotes are AA,BB,OO.
Heterozygotes are AO,BO and AB.
The 4 phenotypes are blood types A, B, AB and O

O allele is recessive to A and to B; A and B are each dominant to O
AO and AA are blood type A; BB and BO are blood type B.
A and B are codominant: AA, AB and BB are distinguishable.

What is a gene?? – the chunk of DNA coding for a functional protein.
Not a locus. Not an allele.

1.1.3 SEGREGATION OF ALLELES:

• One copy of our genome is maternal, the other is paternal. At any given locus, each individual segregates (or copies) a randomly chosen one of its two genes independently to each offspring.

• Alleles have frequencies in the population: the population allele frequency is the probability that the DNA segregating from a random member of the population is of this allelic type. Suppose alleles A, and B have frequencies p, $q = (1 - p)$.

• If the types of the maternal and paternal allele are independent, then the (Hardy-Weinberg) frequencies of the three genotypes AA, AB and BB are p^2, $2pq$ and q^2.

Punnett Square		A	B
		p	q
A	p	p^2	pq
B	q	pq	q^2

• **Example:** A heterozygote marries a random individual, what are the probabilities for their child's type: $p/2$, $(p + q)/2 = 1/2$, and $q/2$ for AA, AB and BB.

Punnett Square		A	B
		$\frac{1}{2}$	$\frac{1}{2}$
A	p	$p/2$	$p/2$
B	q	$q/2$	$q/2$

1.1.4 MENDEL's LAWS (Mendel, 1866):

1. At any given locus, each individual has two genes, one maternal and the other paternal. Each individual segregates a randomly chosen one of its two genes to each offspring, independently to each offspring, independently of gene segregated by the spouse, independently of gene segregated from parent.

2. Independently for different loci. (Not true; segregation of genes at loci on the same chromosome are dependent)

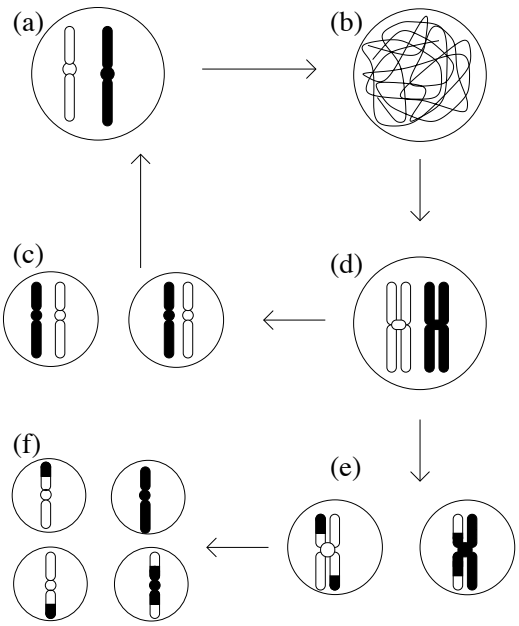
Gamete: The sperm or egg cell (or the DNA therein) that will join to form the offspring individual. (Note gametes are haploid – they carry a single copy of the 3×10^9 bp genome of double-stranded DNA.)

Mitosis Is the process of cell division, in which the diploid DNA is copied to each daughter cell.

Meiosis is the biological process of offspring gamete formation; the diploid parent cell produces haploid gametes.

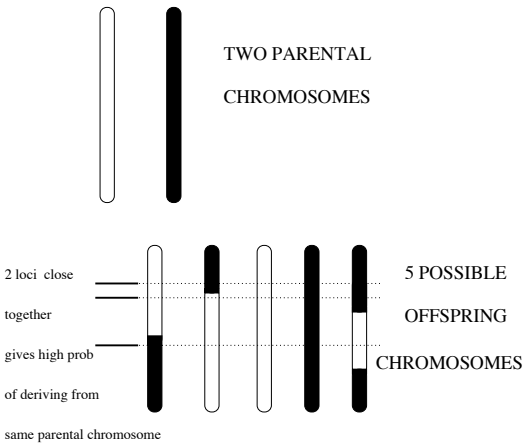
Mendel's first law says all meioses are independent.

1.1.5 MITOSIS and MEIOSIS:



Mitosis: The pair of chromosomes (a) expand into long filaments (2 metres!!), and replicate (b), and re-concentrate as a tetrad (d), and divide (c), to produce two identical diploid daughter cells. (Somatic changes can occur: "errors" in this process.)
Meiosis: (a), (b), (d) as in mitosis. Followed by the first meiotic division (e), in which chromosomes align tightly, so that one goes to each daughter cell. At this stage, exchanges of material can occur. The second meiotic division creates 4 haploid gametes. In males, all become sperm. In females, 1 becomes egg, other discarded.

1.1.6 INHERITANCE OF CHROMOSOMES:



- Offspring chromosomes consist of alternating segments of the two parental chromosomes.
- Chromosomes are inherited in large chunks, $\sim 10^8$ bp.
- Assume, no genetic interference.
- Then, in any meiosis, crossovers occur as a Poisson process along the chromosome. (i.e. "randomly and independently" at some constant rate)
- Between any two positions (loci), in any meiosis, there is recombination if the DNA at those positions derives from different parental chromosomes (i.e. odd number of crossovers).
- Probability of recombination, ρ, increases with genetic distance.
- At small distances, there is high dependence: $\rho \approx 0$.
- At large distances, even and odd have equal probability; $\rho \approx 1/2$ (independent inheritance).

1.1.7 RECOMBINATION AND MATHER'S FORMULA:

- Transmission from a parent at two loci at recombination ρ:

	Locus 1		
Locus-2	mat.	pat.	
mat.	$(1-\rho)/2$	$\rho/2$	1/2
pat.	$\rho/2$	$(1-\rho)/2$	1/2
	1/2	1/2	1

- Note that $\rho = 1/2$ corresponds to independent segregation.

- Recall the 4 chromosome copies (8 strands) of 1.1.5 (d) is the tetrad. An exchange point on the tetrad is a chiasma (pl. chiasmata).

- Mather (1938) showed that, provided each chiasma ends up as a crossover in each gamete independently with probability 1/2, then

$$\rho = (1/2)\Pr(N > 0) = (1/2)(1 - \Pr(N = 0))$$

where N is the number of chiasmata on the tetrad in the interval between the two loci in question. (See Homework 1 #5).

- Thus ρ increases as we consider larger intervals, and $\rho \leq 1/2$.

1.2.1 A SAMPLE OF GENES:

• Consider a single genetic locus, with two codominant alleles A and B. Suppose each independent gene has allelic type A with probability q. We say q is the (population) allele frequency of allele A.

• For a random sample of $2n$ genes from the population: the number of A alleles is $T \sim Bin(2n, q)$. The MLE of q is $T/2n$, which is unbiased since $\mathbf{E}(T/2n) = 2nq/2n = q$.

The variance of the MLE is $q(1-q)/2n$ which is the smallest possible variance for any unbiased estimator.

• As an example, we go through the likelihood mechanics:
Since $\Pr(T = t) \propto q^t(1-q)^{2n-t}$ the log-likelihood is

$$\ell \;=\; t\log(q) \;+\; (2n-t)\log(1-q)$$

1.2.1 ctd: MLE of allele frequencies:

So differentiating the log-likelihood

$$\frac{\partial \ell}{\partial q} \;=\; \frac{t}{q} - \frac{2n-t}{1-q} \;=\; \frac{2n}{q(1-q)}(\frac{t}{2n} - q)$$

So the MLE of q is $t/2n$. Also

$$\frac{\partial^2 \ell(q;t)}{\partial q^2} \;=\; -\frac{t}{q^2} - \frac{2n-t}{(1-q)^2}$$

$$\mathbf{E}\left(-\frac{\partial^2 \ell(q;T)}{\partial q^2}\right) \;=\; \frac{2n}{q} + \frac{2n}{(1-q)} \;=\; \frac{2n}{q(1-q)}$$

So the Fisher information is $2n/q(1-q)$ and the (large-sample) variance of the MLE is $q(1-q)/2n$. In this example, it is the variance for any sample size. For large $2n$, MLE's are approx unbiased, and have approx the smallest possible variance.

1.2.2 A SAMPLE OF INDIVIDUALS:

Suppose we sample n individuals, and that n_0 have genotype AA, n_1 have genotype AB and n_2 have genotype BB). ($n_0+n_1+n_2 = n$). Then we have $(2n_0 + n_1)$ genes of allelic type A, in a sample of size $2n$. We can estimate q by $(2n_0 + n_1)/2n$, but properties of the estimator depend on the genotype frequencies.

$$\begin{aligned}\ell &= \log L \\ &= n_0 \log(\Pr(AA)) + n_1 \log(\Pr(AB)) + n_2 \log(\Pr(BB))\end{aligned}$$

In the absence of constraints:

$$\widehat{\Pr}(AA) = n_0/n, \;\; \widehat{\Pr}(AB) = n_1/n, \;\; \widehat{\Pr}(BB) = n_2/n$$

Extreme case: Complete positive dependence

Suppose there are no AB individuals in the population ($n_1 = 0$). The two homologous genes in an individual are of the same allelic type. The estimator is n_0/n and in effect we have a sample of n genes.

1.2.3 HARDY-WEINBERG EQUILIBRIUM (Hardy, 1908):

Other extreme: Hardy-Weinberg equilibrium (HWE);
There is independence of the allelic types of the two homologous genes within an individual. So $\Pr(AA) = q^2$, $\Pr(AB) = 2q(1-q)$ and $\Pr(BB) = (1-q)^2$

$$\begin{aligned}\ell &= n_0 \log(q^2) + n_1 \log(2q(1-q)) + n_2 \log((1-q)^2) \\ &= (2n_0 + n_1)log(q) + (n_1 + 2n_2) \log(1-q) + n_1 \log 2\end{aligned}$$

The MLE of q is $(2n_0+n_1)/2n$. If $T = 2n_0+n_1$, $T \sim Bin(2n, q)$. $\mathrm{Var}(T/2n) = q(1-q)/2n$ — back to binomial sampling.

Note 1: One generation of random mating establishes HWE, since, by definition, the two genes in an individual are copies of independently sampled parental genes. (Assuming allele frequencies in the parental population are the same for males and females.)

Note 2: See 1.2.4 for a model of intermediate dependence.

1.2.4 POPULATION STRUCTURE:

Suppose populations i, each in HWE, with q_{ij} the freq of allele A_j in population i, and α_i the proportion of population i.
So $\Pr(A_j) = q_{\cdot j} = \sum_i \alpha_i q_{ij}$

$$\begin{aligned}
\Pr(A_jA_j) - (\Pr(A_j))^2 &= \sum_i \alpha_i q_{ij}^2 - q_{\cdot j}^2 \\
&= \sum_i \alpha_i (q_{ij} - q_{\cdot j})^2 \geq 0 \\
\Pr(A_jA_l) - 2\Pr(A_j)\Pr(A_l) &= 2(\sum_i \alpha_i q_{ij} q_{il} - q_{\cdot j} q_{\cdot l}) \\
&= 2\sum_i \alpha_i (q_{ij} - q_{\cdot j})(q_{il} - q_{\cdot l})
\end{aligned}$$

Thus, population subdivision results in homozygote excess relative to HWE. This excess is known as the Wahlund (1928) variance.
In total, we therefore have heterozygote deficiency, but NOT necessarily for each heterozygote.

1.2.5 CASE OF A RECESSIVE ALLELE A:

$t = n_0$ of type AA, and $n - t$ not of type AA.
Assuming HWE, $\Pr(AA) = q^2$, and $T \sim Bin(n, q^2)$. So

$$\ell = \text{const} + t\log(q^2) + (n-t)\log(1-q^2)$$

Differentiating

$$\begin{aligned}
\frac{\partial \ell}{\partial q} &= \frac{2t}{q} - \frac{(n-t)2q}{1-q^2} \\
&= \frac{2}{q(1-q^2)}(t - nq^2)
\end{aligned}$$

So the MLE of q is $\sqrt{t/n}$. Why should this be expected?

Suppose $\beta = q^2$, then

$$\frac{\partial \ell}{\partial q} = \frac{\partial \ell}{\partial \beta}\frac{d\beta}{dq} = 2q\frac{\partial \ell}{\partial \beta}$$

Fisher information and MLE Variance:

Now $T \sim Bin(n, q^2)$, but how can we find the variance of this MLE?

$$\frac{\partial^2 \ell}{\partial q^2} = -\frac{2t}{q^2} - \frac{2(n-t)}{(1-q^2)} - \frac{(n-t)4q^2}{(1-q^2)^2}$$
$$\mathbf{E}(-\frac{\partial^2 \ell}{\partial q^2}) = 2n + 2n + \frac{4q^2 n}{(1-q^2)} = \frac{4n}{1-q^2}$$

The variance of the MLE of q is approx. $(1-q^2)/4n$.
Note this is larger than $q(1-q)/2n$.

We have to make assumptions (HWE)
Variance of the estimator is larger; information $=$ $1/$variance.
We can measure the relative information;

$$I_1 = 2n/q(1-q), \quad I_2 = 4n/(1-q^2)$$
$$I_1/I_2 = (1+q)/2q > 0 \text{ and is large when } q \approx 0$$

1.3.1 TESTING Hardy-Weinberg PROPORTIONS (HWE):

Consider the following three samples, each of 100 individuals. Each has 120 A alleles, so the MLE of q is 0.6, but different genotypic counts n_c in genotype class c.

n	AA	AB	BB	$\widehat{\ell}$	$\widehat{q}$	$\tilde{\ell}$	$2(\widehat{\ell} - \tilde{\ell})$
100	36	48	16	-101.33	0.6	-101.33	0
100	30	60	10	-89.79	0.6	-93.01	6.5
100	45	30	25	-106.71	0.6	-113.81	14.2

With probability p_c for class c,

$$\ell = \text{const} + \sum_c n_c \log(p_c) \text{ with } \sum_c p_c = 1$$

With no constraints, MLE of p_c is n_c/n, and maximized value of the

log-likelihood is

$$\widehat{\ell} \;=\; \sum_c n_c \log(n_c/n) \;\;=\;\; \sum_c n_c \log(n_c) \;-\; n\log(n)$$

Assuming HWE,

$$\begin{aligned} \tilde{p}_c \;&=\; (\widehat{q}^2, 2\widehat{q}(1-\widehat{q}), (1-\widehat{q})^2) \;\;=\;\; (0.36, 0.48, 0.16) \\ \tilde{\ell} \;&=\; \sum_c n_c \log(\tilde{p}_c) \end{aligned}$$

Now, if HWE is true, $2\log\Lambda \;=\; 2(\widehat{\ell} - \tilde{\ell})$ is approximately χ^2_1, and larger otherwise. In our three examples, the values are 0, 6.5 and 14.2. What do we conclude?

Case 1: Someone fudged the data– which I did.

Case 2: $\Pr(\chi^2_1 > 6.5) \approx 0.01$ – marginally OK ?

Case 3: $\Pr(\chi^2_1 > 14.2) \approx 0.0002$ – Reject the null hypothesis!!

1.3.2 TESTING THE ABO BLOOD GROUP MODEL:

	factor freq.		phenotype frequencies			
	A	B	A	B	AB	0
Data			0.422	0.206	0.078	0.294
H_1 model	p	q	$p(1-q)$	$(1-p)q$	pq	$(1-p)(1-q)$
H_1 fitted	0.500	0.284	0.358	0.142	0.142	0.358
H_2 model	p	q	p^2+2pr	q^2+2qr	$2pq$	r^2
H_2 fitted	0.295	0.155	0.411	0.194	0.091	0.303

Bernstein (1925)
reported *ABO* blood types on a sample of 502 individuals:
42.2% type A, 20.6% type B, 7.8% type AB and 29.4% type O.
(Did he drop 2 individuals?)

The popular model at that time was:
H_1: A and B are independently inherited factors
Frequency of individuals having the factor A is 0.500 and of B is 0.284. Independence of the factors would give an AB frequency of $0.500 \times 0.284 = 0.142$ much larger than the 0.078 observed.

1.3.3 TESTING GOODNESS OF FIT: MODEL H_1:

For the general model

$$\begin{aligned}\widehat{\ell} &= 502(.422\log .422 + .206\log .206 + \\ &\qquad .078\log .078 + .294\log .294) \\ &= -626.71\end{aligned}$$

Under H_1 the estimated frequencies are as shown in Table, and the log-likelihood is

$$\begin{aligned}\ell_1 &= 502(.422\log .358 + .206\log .142 + \\ &\qquad .078\log .142 + .294\log .358) \\ &= -647.50\end{aligned}$$

Twice the log-likelihood difference is 41.58, and would be the value of a χ^2_1 random variable if H_1 were true. Clearly, H_1 is rejected.

1.3.4 TESTING MODEL H_2:

H_2: A and B are the two non-null alleles of a single system. Assuming HWE, if the three alleles A, B and O have frequencies p, q and r ($p+q+r=1$), then the frequencies of the four blood types are p^2+2pr, q^2+2qr, $2pq$ and r^2.
Bernstein pointed out that the sum of the A and O blood type frequencies is $(p+r)^2$, or one minus the square root of this frequency is $(1-p-r)=q$. Similarly one minus the square root of the sum of the B and O blood type frequencies is p, and the square root of the O blood type frequency is r. The sum of these three numbers should be one. For his data

$$\begin{aligned}&(1-\sqrt{0.422+0.294})+(1-\sqrt{0.206+0.294})+\sqrt{0.294} \\ &\qquad = 0.99\end{aligned}$$

which is close to one, suggesting a good fit.

1.3.5 LIKELIHOOD RATIO TEST FOR H_2:

More formally, we may perform a likelihood ratio test. Finding the MLEs of the parameters p, q and r is not simple; in fact, we shall see later that these MLEs are $\widehat{p} = 0.2945$ and $\widehat{q} = 0.1547$, with the resulting fitted frequencies given in the table. The log-likelihood is

$$\ell_2 = 502(.422 \log .4114 + .206 \log .1942 + .078 \log .0911 + .294 \log .3033) = -627.52$$

Twice the log-likelihood difference between this and the general alternative is now only 1.62. Again, this is the value of a χ^2_1 random variable if H_2 is true: $\Pr(\chi^2_1 > 1.62) \approx 0.2$. H_2 is not rejected.

DEGREES OF FREEDOM: Total number of categories = 4

Lose 1 degree of freedom for fixed total: 4-1 = 3
Lose 1 for each parameter estimated : 3-2 = 1
(Under each of H_1 and H_2 we estimate p and q:
Under H_2, $r = 1 - p - q$.)
Note we test each of H_1,H_2 against general model, not H_1 vs H_2.

1.4.1 GENE COUNTING: CASE OF RECESSIVE TRAIT:

Observe n_0 of recessive phenotype AA, and n_1 of dominant type.

current q	current $2q/(1+q)$	recessive phenotype $t_0 = 36$ AA	dominant phenotype $t_1 + t_2 = 64$ AB	BB	new $q = (2t_0 + t_1)/2n$
0.5	0.667	36	42.67	21.33	0.573
0.573	0.729	36	46.64	17.36	0.593
0.593	0.745	36	47.66	16.34	0.598
0.598	0.749	36	47.91	16.09	0.600
0.600	0.750	36	48.00	16.00	0.600

The three genotypes are AA, AB and BB, with counts say t_i, $(i = 0, 1, 2)$. Now, $n_0 = t_0$ is observed, but the counts of AB and BB are unobservable since B is dominant to A. However $n_1 = t_1 + t_2$ is known.

The counting algorithm:

If counts, t_1 and t_2, were known, then the number of A alleles is $m_1 = 2t_0 + t_1$, and the MLE of q would be $(2t_0 + t_1)/2n$. Further,

$$\Pr(AB \mid AB \text{ or } BB) \;=\; \frac{2q(1-q)}{1-q^2} \;=\; \frac{2q}{1+q}$$

so

$$\mathbf{E}_q(t_1 \mid n_1 = t_1 + t_2 = 64) = 64\frac{2q}{1+q}.$$

The EM-algorithm implements the sequence of iterates shown.
Starting from an arbitrary initial value $q = 0.5$,
• **E-step:** the proportion $2q/(1+q)$ is computed, and the 64 individuals of dominant phenotype divided into the expected numbers t_1 and t_2 that are AB and BB, respectively.
• **M-step:** a new value of q is estimated as $(2t_0 + t_1)/2n$.
Continue alternating E-steps and M-steps until convergence.

1.4.2 EM ALGORITHM FOR MULTINOMIAL DATA:

In latent variable problems, suppose the actual data are $\mathbf{Y}$, and the ideal data that would make the problem easy are $(\mathbf{Y}, \mathbf{X})$. The complete-data log-likelihood is

$$\ell^* \;=\; \log \Pr((\mathbf{Y}, \mathbf{X}) = (\mathbf{y}, \mathbf{x})).$$

The actual log-likelihood to be maximized is

$$\ell \;=\; \log \Pr(\mathbf{Y} = \mathbf{y}) \;=\; \log \left(\sum_{\mathbf{x}} \Pr((\mathbf{Y}, \mathbf{X}) = (\mathbf{y}, \mathbf{x})) \right).$$

E-step (expectation):
At the current estimate θ^* compute ECDLL

$$H_{\mathbf{y}}(\theta; \theta^*) \;=\; \mathbf{E}_{\theta^*}(\log P_\theta(\mathbf{X}, \mathbf{Y}) \mid \mathbf{Y} = \mathbf{y})$$

M-step (maximization):
Maximize $H_{\mathbf{y}}(\theta; \theta^*)$ w.r.t. θ to obtain a new estimate $\tilde{\theta}$.

Theoretical result: $\ell(\tilde{\theta}) \geq \ell(\theta^*)$.

Thus the EM algorithm for finding MLEs alternates E-steps and M-steps. The likelihood is non-decreasing over the process. Where the likelihood surface is unimodal, convergence to the MLE is assured, although it may be slow. Where computable, evaluate the (log)-likelihood to assess convergence.

For multinomial data:
Let n_c be actual data-counts, and
m_{c^*} complete-data counts for idealized data:
Each class c may be divided into several classes c^*.
So $\ell^* = \sum_{c^*} m_{c^*} \log(p_{c^*})$: finding the ECDLL just means finding

$$\mathbf{E}(m_{c^*}|n_c) = n_c \Pr(c^*|c,\theta^*) = n_c \frac{p_{c^*}(\theta^*)}{\sum_{c^* \to c} p_{c^*}(\theta^*)}$$

In the multinomial case, computing the ECDLL just involves imputing the "hidden" counts, but only because ℓ^* is a linear function of these counts.

1.4.3 ABO log-likelihood for phenotypes and genotypes:

• Assume HWE, and allele frequencies p, q and r for alleles A, B and O ($p+q+r=1$).

• The actual data are the phenotype counts:
$Y = (n_A, n_B, n_{AB}, n_O)$ with log-likelihood

$$\ell = n_A \log(p^2+2pr) + n_B \log(q^2+2qr) + n_O \log(r^2) + n_{AB} \log(2pq)$$

Clearly this is hard to maximize directly, (but can be computed).

• The complete data are the genotype counts:
$X = (m_{AA}, m_{AO}, ...)$ with complete-data log-likeliood

$$\begin{aligned}\ell^* &= m_{AA}\log(p^2) + m_{AO}\log(2pr) + m_{BB}\log(q^2) \\ &\quad + m_{BO}\log(2qr) + m_{OO}\log(r^2) + m_{AB}\log(2pq) \\ &= \text{const} + (2m_{AA}+m_{AO}+m_{AB})\log p \\ &+ (2m_{BB}+m_{BO}+m_{AB})\log q + (2m_{OO}+m_{AO}+m_{BO})\log r\end{aligned}$$

1.4.4 ESTIMATION OF ABO ALLELE FREQUENCIES:

For the MLE of *ABO* blood group allele frequencies. the EM-algorithm is one of the easiest ways to find the MLEs. (see table, next page)
E-step: partition the A phenotypes into expected counts of AA and AO genotypes, and similarly B into BB and BO:

$$\Pr(AO \mid \text{type } A) = \frac{2pr}{p^2+2pr} = \frac{2r}{p+2r}$$
$$\Pr(BO \mid \text{type } B) = \frac{2qr}{q^2+2qr} = \frac{2r}{q+2r}.$$

M-step: Then $\tilde{p} = (2\Pr(AA) + \Pr(AO) + \Pr(AB))/2$,
and $\tilde{q} = (2\Pr(BB) + \Pr(BO) + \Pr(AB))/2$.
Note $\tilde{p}$ does not change monotonely, but $\tilde{\ell}$ does.
($\tilde{\ell}$ is the current value of ℓ, not of ℓ^*.)

- **Important:** Do not confuse ℓ and ℓ^*
  ℓ^* is just a tool that lets us maximize ℓ.

Results for Bernstein's data:

current values				phenotype A		phenotype B		...
p	q	$\frac{2r}{p+2r}$	$\frac{2r}{q+2r}$	$\Pr(A) = 0.422$		$\Pr(B) = 0.206$		...
				AA	AO	BB	BO	...
0.3	0.3	0.73	0.73	0.115	0.307	0.056	0.150	...
0.308	0.170	0.77	0.86	0.096	0.326	0.029	0.177	...
0.298	0.156	0.79	0.87	0.091	0.331	0.026	0.180	...
0.295	0.155	0.79	0.88	0.089	0.333	0.025	0.181	...

	phen AB	phen O	new values		
...	$\Pr(AB) =$	$\Pr(OO) =$	$\tilde{p}$	$\tilde{q}$	$\tilde{\ell}$
...	0.078	0.294			-687.12
...	0.078	0.294	0.308	0.170	-629.00
...	0.078	0.294	0.298	0.156	-627.57
...	0.078	0.294	0.295	0.155	-627.53
...	0.078	0.294	0.295	0.155	-627.52