Genética de Poblaciones Humanas

PAULO ALBERTO OTTO

EDITORIAL UNIVERSITARIA
UNIVERSIDAD NACIONAL DE MISIONES

HUMAN POPULATION GENETICS

(GENÉTICA DE POBLACIONES HUMANAS)

PAULO A. OTTO

Departamento de Genética e Biologia Evolutiva
Instituto de Biociências
Universidade de São Paulo
Caixa Postal 11461
05422-970 São Paulo SP

Curso Teórico Práctico de Post-Grado
8 al 14 de Septiembre de 2006
Departamento de Genética
Laboratorio de Citogenética y Genética Humana
Facultad de Ciencias Exactas Químicas y Naturales
Universidad Nacional de Misiones
Posadas, Misiones, República Argentina

I - Teoría

EDITORIAL UNIVERSITARIA DE MISIONES

San Luis 1870
Posadas - Misiones – Tel-Fax: (03752) 428601

Correos electrónicos:
edunam-admini@arnet.com.ar
edunam-direccion@arnet.com.ar
edunam-produccion@arnet.com.ar
edunam-ventas@arnet.com.ar
edunam-prensa@arnet.com.ar

Otto, Paulo Alberto
Genética de poblaciones humanas. - 1a ed. - Posadas : EdUNaM - Editorial Universitaria de la Universidad Nacional de Misiones, 2008.
209 p.; 28x22 cm.

ISBN 978-950-579-113-2

1. Genética de Poblaciones. 2. Genética Humana. I. Título
CDD 616.042

Fecha de catalogación: 15/10/08

ISBN: 978-950-579-113-2
Impreso en Argentina

PAG.

HARDY-WEINBERG EQUILIBRIUM 6

HARDY-WEINBERG EQUILIBRIUM WITH OVERLAPPING GENERATIONS 21

FISHER'S PRINCIPLE ON EQUILIBRIUM POPULATIONS 25

SAMPLE ESTIMATES OF GENE FREQUENCIES 27

MAXIMUM LIKELIHOOD ESTIMATE FOR THE FREQUENCY OF DOMINANT AUTOSOMAL ALLELES 29

GENETIC EQUILIBRIUM IN RELATION TO A PAIR OF LOCI 33

CALCULATION OF HAPLOTYPE FREQUENCIES AND OF LINKAGE DISEQUILIBRIUM VALUES FOR LINKED GENE COMPLEXES 41

LINKAGE DISEQUILIBRIUM CALCULATIONS 45

GENETIC VARIABILITY AND ITS ASSESSMENT 54

INBREEDING 56

DISTRIBUTION OF GENOTYPES IN PAIRS OF RELATIVES 74

HIERARCHICAL STRUCTURE OF POPULATIONS: ISOLATE EFFECT (WAHLUND'S EFFECT) 77

MIGRATION 83

RACE ADMIXTURE CALCULATIONS 88

PROBABILITY OF EXTINCTION OF A NEUTRAL MUTANT GENE 91

GENETIC DRIFT 96

SELECTION 103

FUNDAMENTAL THEOREM OF NATURAL SELECTION 124

GENETIC LOAD 127

SELECTION WITH INBREEDING 128

EVOLUTION OF 1:1 SEX-RATIO 132

MUTATION-SELECTION BALANCE 135

IDENTIFICATION AND FORENSIC APPLICATIONS 144

A COLLECTION OF BASIC FORMULAE COMMONLY USED IN THE THEORY OF POPULATION GENETICS 175

DERIVATIVES (SUMMARY) 183

GENÉTICA DE POBLACIONES HUMANAS - EJERCICIOS EN CLASE 187

HARDY-WEINBERG EQUILIBRIUM

Let us consider a population of infinite size, consisting of diploid, sexually-reproducing individuals. In relation to a given autosomal locus where 2 alleles (A and a) are segregating, these individuals will belong to the genotypic classes AA, Aa and aa. Let us suppose that in a given generation the frequencies of these three genotypes, among individuals of both sexes, are d, h, and r respectively and that all matings occur entirely at random. Under this assumption, the probabilities of any individual of the population choosing a mate that is AA, Aa or aa are respectively d, h and r. Since d, h and r are also the probabibilities of the first individual being AA, Aa or aa, the various matings occurring in the population will be given by the cross-products shown in the matrix below:

		males		
		AA	Aa	aa
females	AA	d^2	dh	dr
	Aa	dh	h^2	hr
	aa	dr	hr	r^2

If the generations are discrete and the effects of selection, mutation and migration are considered negligible, that is, each mating pair contributes on average to the next generation with the same offspring number as all other couples, no gene A is transformed by mutation into a and vice-versa, and there is no exchange of genes among individuals belonging to this population and individuals from other populational aggregates, then we obtain the following results:

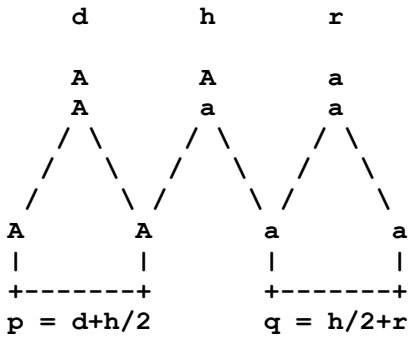
cross. (gen. n)			offspr. genot. frequencies (gen. n+1)		
mal.	fem.	frequencies	AA	Aa	aa
AA	AA	d^2	d^2	0	0
AA	Aa	dh	dh/2	dh/2	0
AA	aa	dr	0	dr	0
Aa	AA	dh	dh/2	dh/2	0
Aa	Aa	h^2	$h^2/4$	$h^2/2$	$h^2/4$
Aa	aa	hr	0	hr/2	hr/2
aa	AA	dr	0	dr	0
aa	Aa	hr	0	hr/2	hr/2
aa	aa	r^2	0	0	r^2

Since the probabilities of a given progeny do not depend upon the sex of its parents (for example, the expected proportions of AA and Aa progeny from crossings AA_m x Aa_f and AA_f x Aa_m are exactly the same), the table above can be simplified to:

cross. (gen. n)		frequencies	offspr. genot. frequencies (gen. n+1) AA	Aa	aa
AA	AA	d^2	d^2	0	0
AA	Aa	2dh	dh	dh	0
AA	aa	2dr	0	2dr	0
Aa	Aa	h^2	$h^2/4$	$h^2/2$	$h^2/4$
Aa	aa	2hr	0	hr	hr
aa	aa	r^2	0	0	r^2

Therefore, the frequency of AA individuals in the following generation is $d^2 + dh + h^2/4 = (d+h/2)^2$,
that of Aa is $dh + 2dr + hr + h^2/2 = 2(d+h/2)(h/2+r)$
and that of aa individuals is $h^2/4 + hr + r^2 = (h/2+r)^2$.

The quantities d+h/2 and h/2+r are respectively the frequencies of the alleles A and a, since each AA individual is represented by two A genes, each heterozygote by one A and one a and each aa homozygote is represented by two a genes:



In fact, if the numbers (or absolute frequencies) of genotypes AA, Aa and aa are D, H and R respectively (a mnemonics to dominant, heterozygote and recessive respectively, in spite of a not being necessarily recessive in relation to A), the numbers of A and a genes are respectively N(A) = 2D + H and N(a) = H + 2R, since each homozygote carries two identical copies of the same gene and a heterozygote has one copy of each allele. Since there are (2D + H) + (H + 2R) = 2D + 2H + 2R = 2N(A) + 2N(a) = 2N genes in the population, the frequencies of the two alleles are given respectively by

P(A) = (2D + H)/2N = 2D/2N + H/2N = D/N + ½ H/N = d + h/2 = p and

P(a) = (H + 2R)/2N = H/2N + 2R/2n = ½ H/N + R/N = h/2 + r = q .

Therefore, if we have a population of infinite size where the frequencies of genotypes AA, Aa and aa are d, h and r respectively and if matings occur at random ('panmixia'), individuals with genotypes AA, Aa and aa will occur after the proportions p^2, 2pq and q^2, where p =

d+h/2 and q = 1-p = h/2+r are the frequencies of the A gene and its allele a in the parental generation.

Obviously after one more generation of random matings the population will still present the same genotypic ratios p^2 : 2pq : q^2, as the following table shows :

cross.(gen. n+1)		frequencies	offspr. genot. frequencies (gen. n+2) AA	Aa	aa
AA	AA	p^4	p^4	0	0
AA	Aa	$4p^3q$	$2p^3q$	$2p^3q$	0
AA	aa	$2p^2q^2$	0	$2p^2q^2$	0
Aa	Aa	$4p^2q^2$	p^2q^2	$2p^2q^2$	p^2q^2
Aa	aa	$4pq^3$	0	$2pq^3$	$2pq^3$
aa	aa	q^4	0	0	q^4

The frequencies of AA, Aa and aa individuals in the generation n+2 are therefore

$$P(AA) = p^4 + 2p^3q + p^2q^2 = p^2(p^2 + 2pq + q^2) = p^2$$

$$P(Aa) = 2p^3q + 4p^2q^2 + 2pq^3 = 2pq(p^2 + 2pq + q^2) = 2pq$$

$$P(aa) = p^2q^2 + 2pq^3 + q^4 = q^2(p^2 + 2pq + q^2) = q^2 .$$

The main conclusion from the analyses shown above is that after one single generation of panmixia, the genotypic frequencies P(AA), P(Aa) and P(aa) are in the ratios p^2, 2pq and q^2, where p and q are the frequencies of a mutually exclusive pair of alleles segregating at an autosomal locus in a breeding population of infinite size. This is the principle, theorem or law of Hardy - Weinberg, named after the two authors who described it quite independently in 1908.

The Hardy-Weinberg principle can be demonstrated straightforwardly using the following argument: the individuals born to random mating pairs result obviously from fertilizations that occur also randomly among gametes produced by male and female individuals from the parental generation. Since the allelic pair under consideration is an autosomal one, among males as well as females from the population genotypes AA, Aa and aa are in the same ratios d, h and r; and males and females will produce gametes A and a in the ratios p = d+h/2 : q = h/2+r respectively. Random union of these gametes result in the offspring genotypes, AA, Aa and aa, that will occur in the ratios p^2 : 2pq : q^2 respectively:

		feminine gametes	
		A	a
		p	q
	A	AA	Aa
masculine	p	p^2	pq
gametes	a	Aa	aa
	q	pq	q^2

Since individuals AA, Aa and aa are now in the ratios p^2, 2pq and q^2, it comes out that the gametes A and a produced by males as well as females from this generation will occur respectively in the frequencies $p^2 + 2pq/2 = p(p+q) = p$

$2pq/2 + q^2 = q(p+q) = q;$

algebraically, all the above is equivalent to the binomial expansion

$[(p^2+pq)+(pq+q^2)]^2 = (p+q)^2 = p^2 + 2pq + q^2.$

Of course a population with Hardy-Weinberg equilibrium has a genotypic distribution $p^2 : 2pq : q^2$, but the inverse is not true (Stark, personal communication, 1983; Li, 1988): it is possible to show that some populations with no panmixia at all have the marginal genotypic distribution $p^2 : 2pq : q^2$.

The evolutionary importance of this simple principle is obvious: in the absence of factors such as mutation, random genetic drift and migration, there exists at the population level a static force that tends to keep genotypic ratios in the proportions $p^2 : 2pq : q^2$, therefore maintaining the population variability throughout time.

The table below (generated by the BASIC code that follows) shows the frequencies of AA, Aa and aa genotypes as functions of the frequency p of the A allele (or q of the a allele).

p	q	P(AA) = p^2	P(Aa) = 2pq	P(aa) = q^2
0.0000	1.0000	0.0000	0.0000	1.0000
0.0500	0.9500	0.0025	0.0950	0.9025
0.1000	0.9000	0.0100	0.1800	0.8100
0.1500	0.8500	0.0225	0.2550	0.7225
0.2000	0.8000	0.0400	0.3200	0.6400
0.2500	0.7500	0.0625	0.3750	0.5625
0.3000	0.7000	0.0900	0.4200	0.4900
0.3500	0.6500	0.1225	0.4550	0.4225
0.4000	0.6000	0.1600	0.4800	0.3600
0.4500	0.5500	0.2025	0.4950	0.3025
0.5000	0.5000	0.2500	0.5000	0.2500
0.5500	0.4500	0.3025	0.4950	0.2025
0.6000	0.4000	0.3600	0.4800	0.1600
0.6500	0.3500	0.4225	0.4550	0.1225
0.7000	0.3000	0.4900	0.4200	0.0900
0.7500	0.2500	0.5625	0.3750	0.0625
0.8000	0.2000	0.6400	0.3200	0.0400
0.8500	0.1500	0.7225	0.2550	0.0225
0.9000	0.1000	0.8100	0.1800	0.0100
0.9500	0.0500	0.9025	0.0950	0.0025
1.0000	0.0000	1.0000	0.0000	0.0000

```
REM PROGRAM FILENAME HARDYWE1.BAS
CLS : DEFDBL A-Z
PRINT "  p            q         P(AA) = p^2  P(Aa) = 2pq  P(aa) = q^2"
PRINT "-------------------------------------------------------------------"
FOR I = 0 TO 20: P = I / 20
PRINT USING " #.####      "; P; 1 - P; P ^ 2; 2 * P * (1 - P); (1 - P) ^ 2
NEXT I
PRINT "-------------------------------------------------------------------"
DO: LOOP WHILE INKEY$ <> " "
```

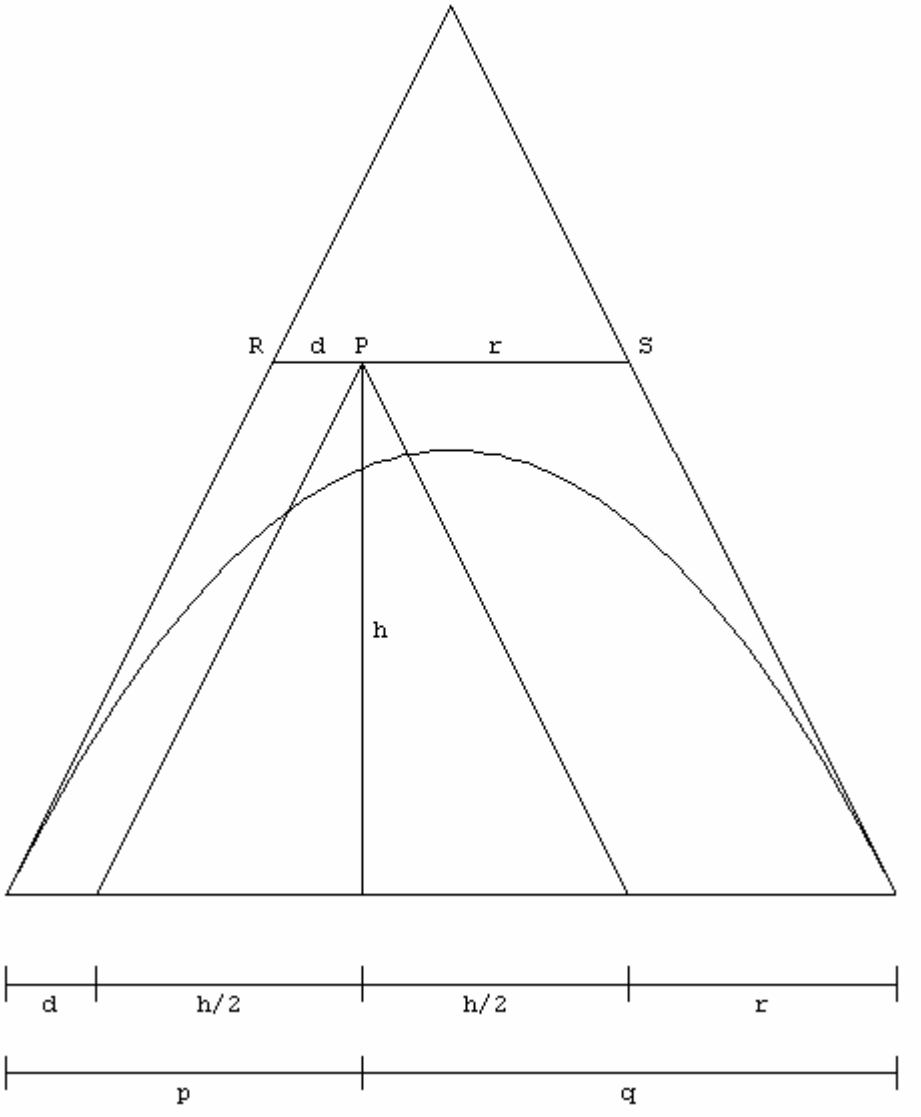
One important property of panmictic populations is that $h^2 = 4dr$. In fact, in these populations, $P(AA) = p^2$, $P(Aa) = 2pq$, $P(aa) = q^2$ and therefore $h^2 = (2pq)^2 = 4p^2q^2 = 4dr = 4.p^2.q^2$.

Another property is that the maximum possible frequency of heterozygotes is 0.5. In fact, if we differentiate $2pq = 2p(1-p) = 2p-2p^2$ in relation to the argument p, we obtain $d[2p(1-p)]/dp = 2-4p$; equating this result to zero, we obtain $2-4p = 0$ and hence $p = 2/4 = 0.5$. This is therefore the value of p that maximizes the function $f(p) = 2p(1-p)$; and for $p = 0.5$ the value of $f(p)$ is also 0.5.

This last property is intuitive: since for $0 < p < 1$ $h = f(p) = 2p(1-p)$ has equal values for complementary values of p adding up to unity and the value of the function is zero for $p = 0$ or $p = 1$, its maximum value takes place when $p = 1-p = 0.5$. And for this value of p the function $2p(1-p)$ has value 0.5. Also, going back to the gamete model we used to demonstrate Hardy-Weinberg equilibrium, it is obvious that the probability of drawing two different gametes (one A from the masculine pool and one a from the feminine one or vice-versa) is at a maximum when the two types of gametes occur within the respective gamete or gene pools with exactly equal frequencies. Therefore, the maximum frequency of heterozygotes in panmictic populations cannot exceed 0.5 or 50%. For

instance, inspecting the sample {AA : 100; Aa : 695; aa : 205} we can assure that the genotypic frquencies are not in Hardy-Weinberg ratios without making any statistical tests, since 695/1000 = 0.695 > 0.5 and this cannot be ascribed to chance fluctuations in a sample of this magnitude.

For the graphical representation of genotypic frequencies one commonly uses a system of triangular coordinates. A very simple system is the isosceles triangle coordinate system (Otto & Benedetti, J. Heredity 1995), the use of which is shown below for the case of a population point P with coordinates d = P(AA) = 0.10, h = P(Aa) = 0.70 and r = P(aa) = 0.20. The perpendicular distance h inside the isosceles triangle of unitary height and basis divides the latter in 2 segments in the proportions p : q, with p + q = 1. This constitutes a clear advantage in relation to the classical representations (Cartesian and equilateral [homogeneous] coordinate systems). Also shown inside the triangle is the Hardy - Weinberg or De Finetti parabola, which represents the set of population points such that d : h : r :: p^2 : 2pq : q^2 .



The figure above was generated by the following Mathematica code:

```
(* TRICOOR2.MA
   Isosceles triang. repres. of genotype freq. *)
Show[
    Plot[2*x*(1 - x), {x, 0, 1},
        Axes -> None,
        DisplayFunction -> Identity],
    Graphics[{
        Line[{{0, 0}, {.5, 1}}],
        Line[{{0, 0}, {1, 0}}],
        Line[{{.5, 1}, {1, 0}}],
        Line[{{0, -.1}, {1, -.1}}],
        Line[{{0, -.08}, {0, -.12}}],
        Line[{{1, -.08}, {1, -.12}}],
        Line[{{.4, -.08}, {.4, -.12}}],
        Line[{{.1, -.08}, {.1, -.12}}],
        Line[{{.7, -.08}, {.7, -.12}}],
        Line[{{0, -.2}, {1, -.2}}],
        Line[{{0, -.18}, {0, -.22}}],
        Line[{{1, -.18}, {1, -.22}}],
        Line[{{.4, -.18}, {.4, -.22}}],
        Line[{{.3, .6}, {.7, .6}}],
        Line[{{.4, .6}, {.4, 0}}],
        Line[{{.4, .6}, {.1, 0}}],
        Line[{{.4, .6}, {.7, 0}}],
        Text["P",{.4, .62}],
        Text["R",{.28, .62}],
        Text["S",{.72, .62}],
        Text["h",{.42, .3}],
        Text["d",{.35, .62}],
        Text["r",{.55, .62}],
        Text["d",{.05, -.12}],
        Text["h/2",{.24, -.12}],
        Text["h/2",{.54, -.12}],
        Text["r",{.85, -.12}],
        Text["p",{.2, -.22}],
        Text["q",{.7, -.22}],
        }],
        DisplayFunction -> $DisplayFunction,
        AspectRatio -> Automatic];
```

For the case of hereditary characteristics determined by autosomal codominant alleles it is possible to test whether the sample drawn from a population is consistent with the Hardy-Weinberg proportions (why authors insist so much on this is a quite different and mysterious problem). This is accomplished using the chi-squared test, the use of which in a real situation is shown below.

In a sample of 230 negroid, unrelated individuals from the city of Rio de Janeiro, Fragoso & Otto (Rev. Med. Est. Guanab. 34 : 59-62 , 1967) determined the haptoglobin types and found the following results:

phenotypes	abs. frequencies
Hp(1-1)	63
Hp(2-1)	117
Hp(2-2)	50

The three phenotypes detected through electrophoresis correspond respectively to the genotypes Hp^1/Hp^1, Hp^1/Hp^2 and Hp^2/Hp^2 determined by the combinations of the two autosomal codominant alleles Hp^1 and Hp^2.

The frequencies of the two alleles in the sample are estimated by direct counting. In fact, the sample above, consisting of 230 individuals, is equivalent to a sample of 2x230 = 460 genes. Since each Hp^1/Hp^1 individual carries two Hp1 genes, each heterozygote carries one Hp^1 gene and one Hp2 gene, and each Hp^2/Hp^2 individual carries two Hp^2 genes, the total number of Hp1 genes in the sample is simply $N(Hp^1)$ = 2x63+117 = 243; and $N(Hp^2)$ is equal to 117 + 2x50 = 217. Therefore the estimate of $p = P(Hp^1)$ is $N(Hp^1)/[N(Hp^1)+N(Hp^2)]$= 243/460= 0.528; the estimate of $q = P(Hp^2)$ is $N(Hp^2)/[N(Hp^1)+N(Hp^2)]$ = 217/460 = 1-p = 0.472.

Of course we cannot know the true frequency of the allele p in the population from which the above sample was drawn. This is not possible even with the sampling of the whole population, that is changing dynamically with time and has its exact genotypic composition submitted to small chance fluctuations varying with time. That is the reason why it is important to calculate the statistical error (standard error) of the estimate p (or q), and that is given in the case of autosomal codominant alleles by the simple formula $s.e.(p) = s.e.(q) = \sqrt{(pq/2N)}$, where N, as before, is the number of sampled individuals. In the above example, s.e.(p) has value 0.023. Since the binomial estimates obtained from samples of the same population will be normally distributed with mean p = 0.528 and s.e. = 0.023, we know, for instance, that the 95% confidence interval of p is given by $0.528 \pm 1.96 \times 0.023$, with limits therefore of 0.483 and 0.573, which permits us to say that the true value of the gene frequency lies between 0.483 and 0.573 with a probability of approximately 95% (that is, we know now that the error we are making when we state this is approximately 5%). More formally, this means that if we take a large number of samples of same size N from the same population, 95% of the confidence intervals thus constructed (i.e., using the parameters obtained from each sample) will contain the true gene frequency.

The expected absolute frequencies E(11), E(12) and E(22) of Hp(1-1), Hp(2-1) and Hp(2-2) phenotypes under the hypothesis of Hardy-Weinberg equilibrium are:

$E(11) = 230xp^2 = 64.18$
$E(12) = 230x2pq = 114.63$
$E(22) = 230xq^2 = 51.18$.

Then we contrast these expectations with the observed quantities O(11) = 63, O(12) = 117 and O(22) = 50 using the usual chi-squared statistics:

	Hp(1-1)	Hp(2-1)	Hp(2-2)	total
obs. abs. freq. O_i	63	117	50	230
obs. exp. freq. E_i	64.18	114.63	51.18	230
$(O_i-E_i)^2/E_i$	0.022	0.049	0.027	0.098

The formula we just used for obtaining the χ^2 figure is important because through it we can inspect the individual contributions for the total value of the chi-squared statistics and locate the class responsible for the largest deviation contributing to final figure of the statistics. In the case we are not interested in this, the formula above can be simplified to $\chi^2 = \Sigma(O_i-E_i)^2/E_i = \Sigma(O_i^2-2O_iE_i+E_i^2)/E_i = \Sigma(O_i^2/E_i) - 2\Sigma O_i + \Sigma E_i = \Sigma(O_i^2/E_i) - N$, since $\Sigma O_i = \Sigma E_i = N$. This simplified formula is often used in computer programs for calculating the value of the statistics and avoids rounding errors generated by the complete formula $\chi^2 = \Sigma(O_i-E_i)^2/E_i$.

For one degree of freedom (d.f), the chi-squared figure of 0.098 corresponds to a probability between 0.75 and 0.90 favoring the hypothesis just tested. Hence we conclude that the collected data are in accordance with Hardy-Weinberg proportions.

The chi-squared test just performed has 1 d.f. because in order to calculate the expected quantities E(11), E(12) and E(22) necessary to perform the test we used 2 sample parameters: the total number 230 and one gene frequency (p or q). If one is not satisfied with this formal definition of degrees of freedom of a chi-squared statistics for testing Hardy-Weinberg equilibrium, we can show the following: since p and N are used for obtaining the expected values, any single expected value we determine fixes automatically the values of the other two. For instance, if we calculate E(11) as being $Np^2 = 64.18$, the expected number of heterozygotes x is given by $E(12) = 2 \times (64.18 - Np) = 114.63$, because p = frequency of homozygotes + frequency of heterozygotes/2.

In the case of two autosomal codominant alleles (A, a) the usual formula for obtaining the chi-squared value, $\chi^2 = \Sigma[(O_i-E_i)^2/E_i]$, can be simplified using the following algebraic acrobatics.

For the two-allele case the expected numbers of AA , Aa and aa individuals, under the null hypothesis of Hardy-Weinberg equilibrium, are $N(AA) = Np^2$, $N(Aa) = 2Npq$, and $N(aa) = Nq^2$, where p and q are the sample estimates of the frequencies of the gene A and its allele a; these estimates, which actually coincide with the ones obtained through the maximum likelihood method, are obtained by simply counting the total genes of the respective types and then by expressing the counts as the proportions of the total of 2N genes counted: $p = (2D+H)/2N$ and $q = (H+2R)/2N$, where D, H, and R are the numbers of AA, Aa and aa individuals observed among the N sampled ones. Therefore we have:

$$\text{CHI-SQUARED (1 d.f.)} = \Sigma[(O_i-E_i)^2/E_i] = \Sigma(O_i^2/E_i) - N =$$
$$= D^2/Np^2 + H^2/2Npq + R^2/Nq^2 - N =$$
$$= [4ND^2(H+2R)^2+2NH^2(2D+H)(H+2R)+4NR^2(2D+R)^2 -$$
$$- N(2D+H)^2(H+2R)^2]/[(2D+H)^2(H+2R)^2] =$$
$$= N(H^4-8DH^2R+16D^2R^2)/[(2D+H)^2(H+2R)^2] =$$
$$= N\{(H^2-4DR)/[(2D+H)(H+2R)]\}^2 .$$

For D = 63, H = 117, R = 50 and N = 230 (numerical example worked above),

$$\text{CHI-SQUARED (1 d.f.)} = 230x[(13689-12600)/(243x217)]^2 =$$
$$= 230x1185921/2780558361 =$$
$$= 272761830/2780558361 =$$
$$= 0.098 .$$

Hardy-Weinberg law can be generalized in almost all its properties to a series of any number of alleles segregating at an autosomal locus:

$$(p + q + ... + z)^2 = p^2 + 2pq + q^2 + ... + z^2 .$$

For example, let a hypothetical hereditary characteristic be determined by three autosomal alleles A, B, and C. If the frequencies of genotypes AA, AB, AC, BB, BC and CC are respectively a, b, c, d, e and f at generation 0, then the allele frequencies P(A), P(B) and P(C) are given respectively by

$$p = (2a+b+c)/2 , q = (b+2d+e)/2 \text{ and } r = (c+e+2f)/2 .$$

Under the assumption of random matings, the individuals belonging to the next generation will occur in the frequencies

genotypes	frequencies
AA	p^2
AB	$2pq$
AC	$2pr$
BB	q^2
BC	$2qr$
CC	r^2

and the allele frequencies in this population continue to be

$$p^2 + 2pq/2 + 2pr/2 = p(p+q+r) = p$$
$$2pq/2 + q^2 + 2qr/2 = q(p+q+r) = q$$
$$2pr/2 + 2qr/2 + r^2 = r(p+q+r) = r .$$

If we denote by p_i and p_j the frequencies of any two alleles segregating at an autosomal locus, it comes out that the frequency of any genotype, under the assumption of panmixia, is given by

$P(A_iA_j) = (2-\delta_{ij}).p_i.p_j$, where δ_{ij} (Kronecker's delta) is an operator with the property $\delta_{ij} = 1$ if i=j , $\delta_{ij} = 0$ otherwise. Therefore,

$$P(A_iA_i) = (2-1).p_i.p_i = p_i^2$$
$$P(A_iA_j) = (2-0).p_i.p_j = 2p_ip_j.$$

As we commented before, it is intuitive that the chance of a heterozygous individual being produced in a panmictic population is at a maximum when gene frequencies are equal for all the n alleles segregating at an autosomal locus. If there exist n alleles, then under this assumption the frequency of each allele is obviously $p_i = 1/n$ and the frequency of each type of heterozygote is $P(a_i a_j) = 2p_i p_j = 2.1/n.1/n = 2/n^2$. When the number of alleles is n, there exist n(n-1)/2 different types of heterozygotes, and the maximum possible frequency of heterozygotes in a panmictic population is $2/n^2$ x n(n-1)/2 = (n-1)/n. The table below shows the values this frequency takes when n = 2, 3, ..., inf.:

n	1/n	$2/n^2$	n(n-1)/2	(n-1)/n
2	1/2	1/2	1	1/2
3	1/3	2/9	3	2/3
4	1/4	2/16	6	3/4
5	1/5	2/25	10	4/5
...	...	...	...	...
inf.	0	0	inf.	1

It is easy to infere that as the number of alleles increases within a given locus the proportion of heterozygotes in the population also increases.

If in the initial (0) generation a same allele has different frequencies among males and females, in the next generation, under panmixia, males and females will have the same gene frequency, and this has as value the arithmetic mean between parental gene frequencies, since males and females contribute equally to their offspring. In fact, if p' is the allele frequency among males and p" among females at generation 0, it comes out that in the first generation the genotypic distribution among males as well as females will be

P(AA) = p'.p"
P(Aa) = p'.q" + p".q' = p'(1-p")+p"(1-p') = p'+p"-2p'.p"
P(aa) = q'.q" = (1-p').(1-p") = 1-p'-p"+p'.p" ;

since p' ≠ p", then it comes out that P(AA) ≠ p^2, P(Aa) ≠ 2pq and P(aa) ≠ q^2.

Gene frequencies in this first generation are determined as usually:

P(A) = P(AA) + P(Aa)/2 = p'.p" + (p'+p")/2 - p'.p" = (p'+p")/2
P(a) = = (q'+q")/2 .

Therefore we can conclude that different allele frequencies among males and females determine a delay of one generation in the approach of Hardy-Weinberg equilibrium. This property is important to derive the approach to equilibrium in the case of sex-linked genes that we discuss in the lines below.

Let f_n and m_n be the frequencies of a same allele a from the X chromosome among females and males respectively, in a generic generation n. Under the assumption of panmixia, the following recurrence relations are obtained:

(1) $f_{n+1} = (m_n+f_n)/2$
(2) $m_{n+1} = f_n$.

Equation (1) results from the fact that each female receives one X chromosome from her mother and the other from her father. Equation (2) means that the only X chromosome present in males derive from their mothers.

From (1) and (2) we obtain also

$$(3)\quad f_{n+1} - m_{n+1} = (m_n + f_n)/2 - f_n = (f_n - m_n).(-1/2) = (f_n - m_n).r \ , \ r = -1/2 \ .$$

This last equation has the general solution

$$f_n - m_n = (f_0 - m_0).r^n = (f_0 - m_0).(-1/2)^n \ ,$$

which shows that each generation of panmixia halves the absolute value of the initial difference $f_0 - m_0$. Of course when n tends to infinity this difference tends to zero, so that at equilibrium gene frequencies will be the same among females and males: f = m = q.

Equation $f_n - m_n = (f_0 - m_0).r^n$ can be rewritten as

$$f_n = m_n + (f_0 - m_0).r^n = f_{n-1} + (f_0 - m_0).r^n \ .$$

It is easy to verify that

$$f_1 = f_0 + (f_0 - m_0).r$$
$$f_2 = f_1 + (f_0 - m_0).r^2 = f_0 + (f_0-m_0).r + (f_0-m_0).r^2$$
$$f_3 = f_2 + (f_0 - m0).r^3 = f_0 + (f_0-m_0).r + (f_0-m_0).r^2 + (f_0-m_0).r^3$$

and so on. Therefore,

$$f_n = f_0 + (f_0 - m_0).(r^1 + r^2 + r^3 + \ldots + r^n).$$

In the expression above, $r^1 + r^2 + r^3 + \ldots + r^n$ is the sum of the terms of a geometric series with ratio r = -1/2, the solution of which is given by the formula

$$r^1 + \ldots + r^n = (r-r^{n+1})/(1-r) = r(1-r^n)/(3/2) = 2r(1-r^n)/3 = -(1-r^n)/3 \ .$$

Therefore, the general solution of fn is given by

$$f_n = f_0 - (f_0-m_0).(1-r^n)/3 = f_0 - (f_0-m_0)/3 + (f_0-m_0).r^n/3 = (2f_0 + m_0)/3 + (f_0-m_0).(-1/2)^n/3 \ .$$

The limit of this expression, as n tends to infinity, is clearly

$$q = f = m = (2f_0 + m_0)/3.$$

The quantity above is a constant quantity :

$(2f_{n+1} + m_{n+1})/2 = q_{n+1} = [2(m_n+f_n)/2 + f_n]/2 = (2f_n+m_n)/2 = q_n = ... = q$,

representing the average (weighed) gene frequency in the whole population, in any generation. In fact, since 1/3 of all X chromosomes are in males and 2/3 in females, given that in the population there exist equal numbers of males and females, the average (weighed) frequency of the allele in the whole population is

$q_n = 2/3.f_n + 1/3.m_n = q = f = m$,

and this quantity must be a constant given the assumptions of absence of selection, mutation and differential migration.

The above results can be obtained straightforwardly, using the following more formal procedure:

Writing the recurrence equations $f_1 = (f_0+m_0)/2$ and $m_1 = f_0$ in matrix compressed form

$$\begin{pmatrix} f_1 \\ m_1 \end{pmatrix} = WQ_0 = \begin{pmatrix} 1/2 & 1/2 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} f_0 \\ m_0 \end{pmatrix} = RW_dR^{-1}Q_0 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1/2 \end{pmatrix}\begin{pmatrix} 4/3 & 2/3 \\ 2/3 & -2/3 \end{pmatrix}\begin{pmatrix} f_0 \\ m_0 \end{pmatrix}$$

the general solution $Q_n = RW_d^nR^{-1}Q_0$ is obtained immediately, from which we get

$f_n = (2f_0 + m_0)/3 + (f_0-m_0).(-1/2)^n/3$ and

$m_n = (2f_0 + m_0)/3 - 2(f_0-m_0).(-1/2)^n/3$.

As before, the limit of both expressions, as n tends to infinity, is clearly $q = f = m = (2f_0 + m_0)/3$.

The equilibrium condition for a sex-linked locus is that all its alleles have the same frequencies in males and females. This takes place asymptotically, in an oscillatory manner, since $m_{n+1}-f_{n+1} = -(m_n-f_n)/2$. At equilibrium genotypes are distributed after

genotypes	frequencies
Ay	p
ay	q
AA	p^2
Aa	2pq
aa	q^2

,

that is, the male genotypes (hemizygotes A and a) occur in gene frequencies while the female genotypes AA, Aa and aa follow a typical H-W distribution $p^2 : 2pq : q^2$.

As a numerical example to appreciate the approach to equilibrium, let us consider the following initial population :

$P_0(Ay) = 1.00$
$P_0(ay) = 0.00$
$P_0(AA) = 0.00$
$P_0(Aa) = 0.00$
$P_0(aa) = 1.00$.

From the data above, it comes out that the initial frequencies of the a gene in males and females are respectively $m_0 = 0$ and $f_0 = 1$. Under panmixia, the genotypes in the following generation will occur in the frequencies

$P_1(Ay) = 1-f_0 = 0.00$
$P_1(ay) = f_0 = 1.00$
$P_1(AA) = (1-m_0).(1-f_0) = 0.00$
$P_1(Aa) = (1-m_0).f_0 + m_0.(1-f_0) = 1.00$
$P_1(aa) = m_0.f_0 = 0.00$;

in this first generation gene frequencies are

$m_1 = f_0 = 1.00$ and $f_1 = (m_0+f_0)/2 = 0.5$.

Applying recursively the equations

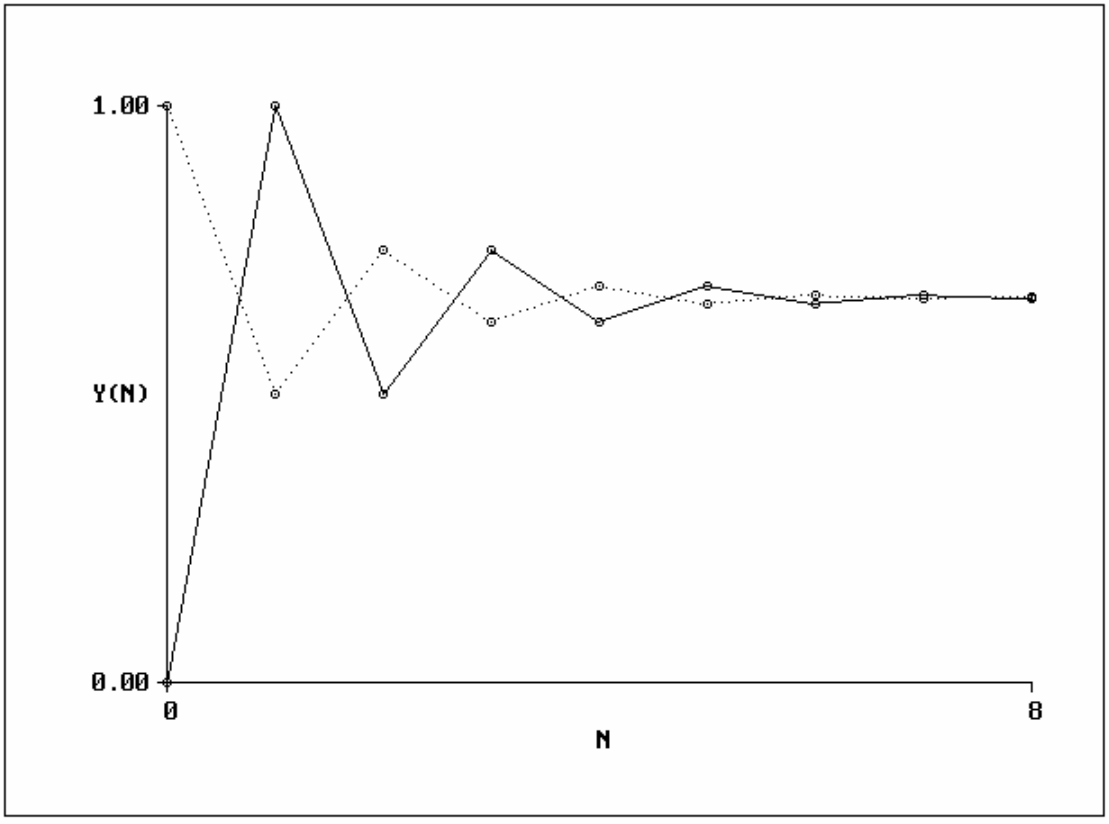
$m_n = P_n(ay)$
$f_n = P_n(Aa)/2 + P_n(aa)$
$P_{n+1}(Ay) = 1-f_n$
$P_{n+1}(ay) = f_n$
$P_{n+1}(AA) = (1-m_n).(1-f_n)$
$P_{n+1}(Aa) = (1-m_n).f_n + m_n.(1-f_n)$
$P_{n+1}(aa) = m_n.f_n$

the values corresponding to other generations are obtained and shown in the table below (followed by the respective BASIC code that generated it):

n	mn	fn	|mn-fn|	Pn(Ay)	Pn(ay)	Pn(AA)	Pn(Aa)	Pn(aa)
0	0.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	1.0000
1	1.0000	0.5000	0.5000	0.0000	1.0000	0.0000	1.0000	0.0000
2	0.5000	0.7500	0.2500	0.5000	0.5000	0.0000	0.5000	0.5000
3	0.7500	0.6250	0.1250	0.2500	0.7500	0.1250	0.5000	0.3750
4	0.6250	0.6875	0.0625	0.3750	0.6250	0.0938	0.4375	0.4688
5	0.6875	0.6563	0.0313	0.3125	0.6875	0.1172	0.4531	0.4297
6	0.6563	0.6719	0.0156	0.3438	0.6563	0.1074	0.4414	0.4512
7	0.6719	0.6641	0.0078	0.3281	0.6719	0.1128	0.4463	0.4409
8	0.6641	0.6680	0.0039	0.3359	0.6641	0.1102	0.4436	0.4462
9	0.6680	0.6660	0.0020	0.3320	0.6680	0.1115	0.4449	0.4436
10	0.6660	0.6670	0.0010	0.3340	0.6660	0.1109	0.4442	0.4449
11	0.6670	0.6665	0.0005	0.3330	0.6670	0.1112	0.4446	0.4442
12	0.6665	0.6667	0.0002	0.3335	0.6665	0.1111	0.4444	0.4446
13	0.6667	0.6666	0.0001	0.3333	0.6667	0.1111	0.4445	0.4444
14	0.6666	0.6667	0.0001	0.3334	0.6666	0.1111	0.4444	0.4445
15	0.6667	0.6667	0.0000	0.3333	0.6667	0.1111	0.4445	0.4444
16	0.6667	0.6667	0.0000	0.3333	0.6667	0.1111	0.4444	0.4445

```
REM PROGRAM FILENAME HWSEXL01.BAS
DEFDBL A-Z: DEFINT I
INPUT "P0(Ay) = "; PAY
INPUT "P0(ay) = "; PBY
INPUT "P0(AA) = "; PAA
INPUT "P0(Aa) = "; PAB
INPUT "P0(aa) = "; PBB
M = PBY: F = PAB / 2 + PBB
PRINT "-------------------------------------------------------------------"
PRINT " n    mn      fn    |mn-fn|  Pn(Ay)  Pn(ay)  Pn(AA)  Pn(Aa)  Pn(aa)"
PRINT "-------------------------------------------------------------------"
FOR I = 0 TO 16
PRINT USING "##  "; I;
PRINT USING "#.####  "; M; F; ABS(M - F); PAY; PBY; PAA; PAB; PBB
PAY = 1 - F: PBY = F: PAA = (1 - M) * (1 - F)
PAB = (1 - M) * F + M * (1 - F): PBB = M * F
M = PBY: F = PAB / 2 + PBB
NEXT I
PRINT "-------------------------------------------------------------------"
```

The approach to equilibrium can be appreciated by the following graph, where the dashed line indicates the allele frequencies among females and the continuous one the allele frequencies among males, for $m_0 = 0$ and $f_0 = 1$.



HARDY-WEINBERG EQUILIBRIUM WITH OVERLAPPING GENERATIONS

In the lines that follow the reasoning used by Moran (The statistical processes of evolutionary theory, Oxford University Press, Oxford, 1962, pp. 23-24) is adopted.

Let P(t) = frequency of AA individuals at time t
R(t) = frequency of Aa individuals at time t
Q(t) = frequency of aa individuals at time t ;

assuming that in the time interval dt a fraction dt of the population dies and is replaced by a new fraction dt produced by random mating, the equations that follow are obtained :

$$P(t+dt) = P(t) - P(t).dt + [P(t) + R(t)/2]^2.dt =$$
$$= P(t)(1-dt) + [P(t) + R(t)/2]^2.dt$$

$$R(t+dt) = R(t) - R(t).dt + 2[P(t) + R(t)/2][R(t)/2 + Q(t)].dt =$$
$$= R(t)(1-dt) + 2[P(t) + R(t)/2][R(t)/2 + Q(t)].dt$$

$$Q(t+dt) = Q(t) - Q(t).dt + [R(t)/2 + Q(t)]^2.dt =$$
$$= Q(t)(1-dt) + [R(t)/2 + Q(t)]^2.dt.$$

Rearranging the first of the above expressions, we obtain

$$P(t+dt)-P(t) = -P(t).dt + [P(t) + R(t)/2]^2.dt$$

and

$$[P(t+dt)-P(t)]/dt = P[(t+dt)-P(t)]/[(t+dt)-t]$$
$$= -P(t) + [P(t) + R(t)/2]^2;$$

the limit of this expression, as dt tends to zero, is

$$dP(t)/dt = -P(t) + [P(t) + R(t)/2]^2 ;$$

similarly, we obtain

$$dR(t)/dt = -R(t) +2[P(t) + R(t)/2][R(t)/2 + Q(t)]$$
$$dQ(t)/dt = -Q(t) + [R(t)/2 + Q(t)]^2 .$$

If we define p(t) = P(t) + R(t)/2 ,

it comes out that

$$dp(t)/dt = dP(t)/dt + 1/2.dR(t)/dt =$$
$$-[P(t) + R(t)/2] + [P(t) + R(t)/2]^2$$
$$+ [P(t) + R(t)/2][R(t)/2 + Q(t)]$$
$$= -[P(t) + R(t)/2] + [P(t) + R(t)/2][P(t) + R(t) + Q(t)]$$
$$= -[P(t) + R(t)/2] + [P(t) + R(t)/2] = 0 .$$

Therefore, p(t) and q(t) = 1 - p(t) are constant values (p, q).

Replacing these values in the equations for dP(t)/dt, dR(t)/dt and dQ(t)/dt, we obtain

$dP(t)/dt = -P(t) + p^2$
$dR(t)/dt = -R(t) + 2pq$
$dQ(t)/dt = -Q(t) + q^2$.

The solution for $dP(t)/dt = -P(t) + p^2$ is obtained in the lines below.

From $dP(t)/dt = -P(t) + p^2$ we have :

$dP(t)/[P(t)-p^2] = d \ln|P(t)-p^2| = -dt.$

Integrating both sides of $d \ln|P(t)-p^2| = -dt$, that is,

$\int d \ln|P(t)-p^2| = -\int dt$,

we obtain successively

$\ln|P(t)-p^2| = -t + C = -t + \ln C_1$
$\ln[|P(t)-p^2|/C_1] = -t$
$[P(t)-p^2]/C_1 = e^{-t}$
$P(t) = p^2 + C_1.e^{-t}$.

For t= 0 it comes out that

$P(0) = p^2 + C_1.e^0 = p^2 + C_1$

and

$C_1 = P(0)-p^2$.

Therefore, the complete solution of the equation

$dP(t)/dt = -P(t) + p^2$ is

$P(t) = p^2 + [P(0)-p^2].e^{-t}$.

Similarly, we obtain

$R(t) = 2pq + [R(0)-2pq].e^{-t}$
$Q(t) = q^2 + [Q(0)-q^2].e^{-t}$,

where $p = P(0) + R(0)/2$ and $q = 1-p$.

The limits of the above expressions, as t tends to infinity, are clearly

$P = p^2$
$R = 2pq$
$Q = q^2$.

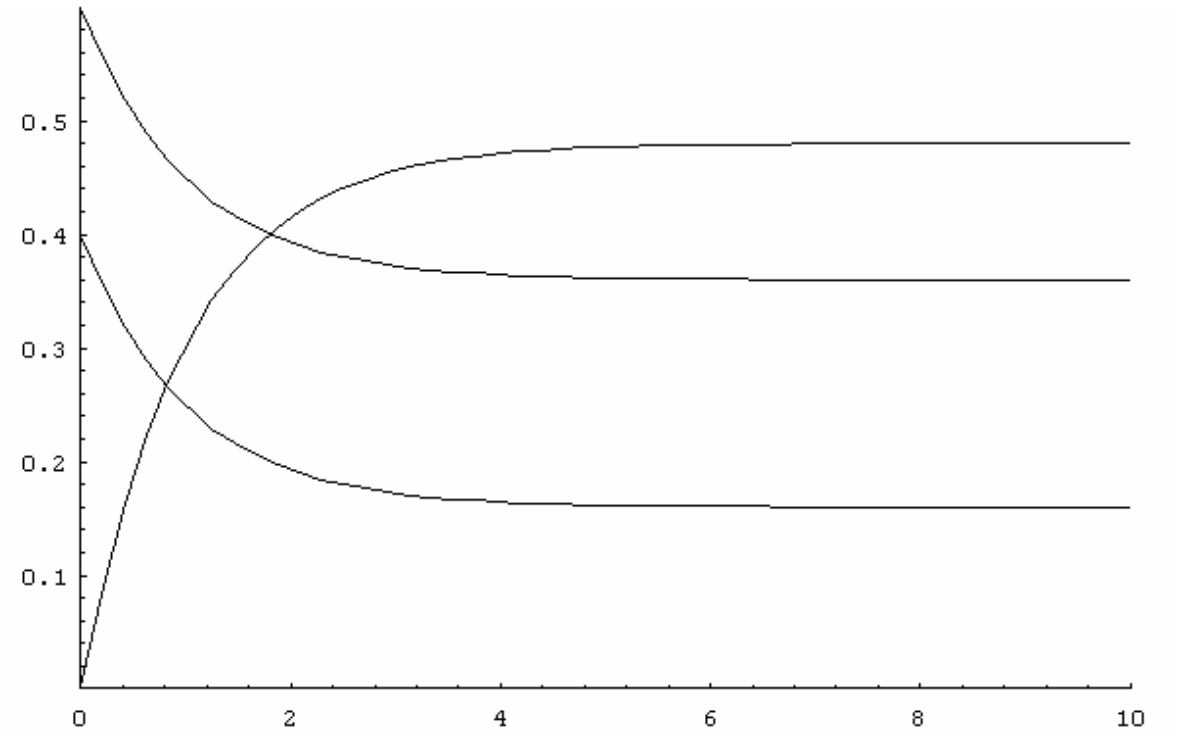
A numerical example of convergence is shown in the table below, followed by the Basic code used for generating it.

```
t            P(t)              R(t)            Q(t)
-----------------------------------------------------
 0         0.40000000       0.00000000       0.60000000
 1         0.24829107       0.30341787       0.44829107
 2         0.19248047       0.41503906       0.39248047
 3         0.17194890       0.45610221       0.37194890
 4         0.16439575       0.47120849       0.36439575
 5         0.19161711       0.47676579       0.36161711
 6         0.16059490       0.47881020       0.36059490
 7         0.16021885       0.47956230       0.36021885
 8         0.16008051       0.47983898       0.36008051
 9         0.16002962       0.47994076       0.36002962
10         0.16001090       0.47997821       0.36001090
11         0.16000401       0.47999198       0.36000401
12         0.16000147       0.47999705       0.36000147
13         0.16000054       0.47999892       0.36000054
14         0.16000020       0.47999960       0.36000020
15         0.16000007       0.47999985       0.36000007
16         0.16000003       0.47999995       0.36000003
17         0.16000001       0.47999998       0.36000001
18         0.16000000       0.47999999       0.36000000
19         0.16000000       0.48000000       0.36000000
20         0.16000000       0.48000000       0.36000000
-----------------------------------------------------
```

```
REM PROGRAM FILENAME HWEQCON1
PRINT "   t          P(t )          R(t)            Q(t)"
PRINT "-----------------------------------------------"
P=.4 : Q=.6 : D0=.40 : H0=0 : R0=.6
FOR T=O TO 20
    Dl=P*P+(D0-P*P)*EXP(-T)
    Hl=2*P*Q+(H0-2*P*Q)*EXP(-T)
    Rl=Q*Q+(R0-Q*Q)*EXP(-T)
    PRINT USING "####";T;
    PRINT USING "     #.#########";Dl;Hl;Rl
NEXT T
    PRINT "-----------------------------------------------"
```

The graph below, generated by the following Matematica code, shows the convergence to equilibrium for AA, Aa and aa genotype frequencies, using the initial numerical values given at the top of the table above.

```
(* hwoverla.ma *)
p = 0.4;
P0 = 0.4;
R0 = 0;
Q0 = 0.6;
P = p^2 + (P0 - p^2) * Exp[-t];
R = 2 * p * (1-p) + (R0 - 2 * p * (1-p)) * Exp[-t];
Q = (1-p)^2 + (Q0 - (1-p)^2) * Exp[-t];
graph = Plot[{P,Q,R},{t,0,10}];
Show[graph, PlotRange -> {0, 0.60},
          AxesOrigin -> {0, 0}]
```


0.5
0.4
0.3
0.2
0.1
0
2
4
6
8
10

FISHER'S PRINCIPLE ON EQUILIBRIUM POPULATIONS

Out of the six possible mating types that occur in a population, four of them (namely, AA × AA, AA × Aa, Aa × aa and aa × aa) reproduce in the offspring exactly the couple genotypic ratios (that is, AA × AA → 1AA:1AA, AA × Aa → 1AA:1Aa, Aa × aa → 1Aa:1aa and aa × aa → 1aa:1aa). Therefore, only the behavior of two crossings (namely AA × aa → 1Aa:1Aa and Aa × Aa → 1AA:2Aa:1aa) has to be analyzed in order to infer any possible equilibrium condition. We start by building the population mating matrix:

	$\mathbf{AA_f}$	$\mathbf{Aa_f}$	$\mathbf{aa_f}$
$\mathbf{AA_m}$	$P(AA_m \times AA_f)$	$P(AA_m \times Aa_f)$	$P(AA_m \times aa_f)$
$\mathbf{Aa_m}$	$P(Aa_m \times AA_f)$	$P(Aa_m \times Aa_f)$	$P(Aa_m \times aa_f)$
$\mathbf{aa_m}$	$P(aa_m \times AA_f)$	$P(aa_m \times Aa_f)$	$P(aa_m \times aa_f)$

The frequency of a given offspring genotype at generation n+1 (v.g., AA) is obtained from the sum of contributions of crossings occurring at generation n, as usual:

$$P_{n+1}(AA) = 1.P_n(AA_m \times AA_f) + \tfrac{1}{2}.[P_n(AA_m \times Aa_f) + P_n(Aa_m \times AA_f)] + \tfrac{1}{4}.P_n(Aa_m \times Aa_f) = P_n(AA \times AA) + P_n(AA \times Aa)/2 + P_n(Aa \times Aa)/4;$$

$$\text{at equilibrium, } P(AA) = 1.P(AA_m \times AA_f) + \tfrac{1}{2}.[P(AA_m \times Aa_f) + P(Aa_m \times AA_f)] + \tfrac{1}{4}.P(Aa_m \times Aa_f) = P(AA \times AA) + P(AA \times Aa)/2 + P(Aa \times Aa)/4 .$$

The frequency of a given parental genotype at generation n (v.g., AA_m) is obtained from the sum of probabilities of crossings in which it participates:

$$P_n(AA_m) = P_n(AA) = P_n(AA_m \times AA_f) + P_n(AA_m \times Aa_f) + P_n(AA_m \times aa_f)$$

$$= P_n(AA \times AA) + P_n(AA \times Aa)/2 + P_n(AA \times aa)/2 .$$

Therefore, a second equilibrium equation for the frequency of the genotype AA can be obtained by adding the elements of the corresponding column or row of the above mating matrix :

$$P(AA) = P(AA \times AA) + P(AA \times Aa)/2 + P(AA \times aa)/2 ,$$

which is equivalent to drop off the subscripts in the equation for $P_n(AA_m)$.

Equating the right sides of the equations

$P(AA) = P(AA\times AA) + P(AA\times Aa)/2 + P(Aa\times Aa)/4$ and
$P(AA) = P(AA\times AA) + P(AA\times Aa)/2 + P(AA\times aa)/2$,

we obtain

$$P(AA\times AA) + P(AA\times Aa)/2 + P(Aa\times Aa)/4 = P(AA\times AA) + P(AA\times Aa)/2 + P(AA\times aa)/2;$$

therefore, the equilibrium condition for any possible population is

$$\mathbf{P(Aa\times Aa) = 2.P(AA\times aa)}$$

The property above (Fisher, 1918) can be used as a short-cut method for determining straightforwardly equilibrium genotype frequencies, avoiding thus the application of tedious algebraic techniques that arise in complicated situations.

SAMPLE ESTIMATES OF GENE FREQUENCIES

1) Two autosomal codominant alleles

$N(AA) = n_1$, $N(Aa) = n_2$, $N(aa) = n_3$, $n_1+n_2+n_3 = N$

Likelihood function: $P = (K/2^{n_2}).(p^2)^{n_1}.(2pq)^{n_2}.(q^2)^{n_3}$
$= K.p^{2n_1+n_2}.q^{n_2+2n_3}$, $K = 2^{n_2}.N!/(n_1!n_2!n_3!)$
$L = \log(P) = (2n_1+n_2).\log(1-q) + (n_2+2n_3).\log q + k$

Max. lik. estimates: $\mathbf{p = (2n_1+n_2)/2N}$, $\mathbf{q = 1-p = (n_2+2n_3)/2N}$
$\mathbf{I(q) = 2N/pq = 2N/[q(1-q)]}$
$\mathbf{var(p) = var(q) = 1/I(q) = pq/2N = q(1-q)/2N}$

2) Two autosomal dominant alleles, A dominant over a

$N(A-) = N(AA)+N(Aa) = n_1$, $N(aa) = n_2$, $n_1+n_2 = N$

Likelihood function: $P = K.(1-q^2)^{n_1}.(q^2)^{n_2}$
$= K.(1-q^2)^{n_1}.q^{2n_2}$, $K = N!/(n_1!n_2!)$
$L = \log(P) = n_1.\log(1-q^2) + 2n_2.\log q + k$

Max. lik. estimates: $\mathbf{q = \sqrt{(n_2/N)}}$, $\mathbf{p = 1-q = 1-\sqrt{(n_2/N)}}$
$\mathbf{I(q) = 4N/(1-q^2)}$
$var(q^2) = q^2(1-q^2)/N = var(q).(dq^2/dq)^2 = 4q^2var(q)$
$\mathbf{var(q) = 1/I(q) = var(q^2)/4q^2 = (1-q^2)/4N}$
$= p^2/4N + pq/2N > pq/2N$

3) Two X-linked codominant alleles

$N(A) = n_1$, $N(a) = n_2$, $n_1+n_2 = N_m$
$N(AA) = n_3$, $N(Aa) = n_4$, $N(aa) = n_5$, $n_3+n_4+n_5 = N_f$

3.1) male sample

Likelihood function: $P = K.p^{n_1}.q^{n_2}$, $K = N_m!/(n_1!n_2!)$
$L = \log(P) = n_1.\log(1-q) + n_2.\log q + k$

Max. lik. estimates: $\mathbf{q = q_m = n_2/N_m}$, $\mathbf{p = p_m = 1-q_m = n_1/N_m}$
$\mathbf{I(q_m) = N_m/[q_m(1-q_m)]}$
$\mathbf{var(q_m) = 1/I(q_m) = q_m(1-q_m)/N_m}$

3.2) female sample

Likelihood function: $P = (K/2^{n_4}).(p^2)^{n_3}.(2pq)^{n_4}.(q^2)^{n_5}$
$= K.p^{2n_3+n_4}.q^{n_4+2n_5}$, $K = 2^{n_4}.N_f!/(n_3!n_4!n_5!)$
$L = \log(P) = (2n_3+n_4).\log(1-q) + (n_4+2n_5).\log q + k$

Max. lik. estimates: $\mathbf{q = q_f = (n_4+2n_5)/2N_f}$,
$\mathbf{p = p_f = 1-q_f = (2n_3+n_4)/2N_f}$
$\mathbf{I(q_f) = 2N_f/[q_f(1-q_f)]}$
$\mathbf{var(q_f) = 1/I(q_f) = q_f(1-q_f)/2N_f}$

3.3) total sample

Likelihood function: $P = (K/2^{n_4}).p^{n_1}.q^{n_2}.(p^2)^{n_3}.(2pq)^{n_4}.(q^2)^{n_5}$
$= K.p^{n_1+2n_3+n_4}.q^{n_2+n_4+2n_5}$,
$K = 2^{n_4}.N_m!N_f!/(n_1!n_2!n_3!n_4!n_5!)$
$L = \log(P) = (n_1+2n_3+n_4).\log(1-q)$
$+ (n_2+n_4+2n_5).\log q + k$

Max. lik. estimates: $q = (n_2+n_4+2n_5)/(N_m+2N_f)$,
$p = 1-q = (n_1+2n_3+n_4)/(N_m+2N_f)$
$I(q) = I(q_m) + I(q_f) = (N_m+2N_f)/[q(1-q)]$
$q \approx [q_m.I(q_m)+q_f.I(q_f)]/[I(q_m)+I(q_f)]$
$var(q) = 1/I(q) = 1/[I(q_m)+I(q_f)] = q(1-q)/(N_m+2N_f)$

4) Two X-linked alleles, A dominant over a

$N(A) = n_1$, $N(a) = n_2$, $n_1+n_2 = N_m$
$N(A-) = N(AA) + N(Aa) = n_3$, $N(aa) = n_4$, $n_3+n_4 = N_f$

4.1) male sample

Likelihood function: $P = K.p^{n1}.q^{n2}$, $K = N_m!/(n_1!n_2!)$
$L = \log(P) = n_1.\log(1-q) + n_2.\log q + k$
Max. lik. estimates: $q = q_m = n_2/N_m$, $p = p_m = 1-q_m = n_1/N_m$
$I(q_m) = N_m/[q_m(1-q_m)]$
$var(q_m) = 1/I(q_m) = q_m(1-q_m)/N_m$

4.2) female sample

Likelihood function: $P = K.(1-q^2)^{n3}.(q^2)^{n4}$
$= K.(1-q^2)^{n3}.q^{2n4}$, $K = N_f!/(n_3!n_4!)$
$L = \log(P) = n_3.\log(1-q^2) + 2n_4.\log q + k$
Max. lik. estimates: $q = q_f = \sqrt{(n_4/N_f)}$, $p = p_f = 1-q_f = 1-\sqrt{(n_4/N_f)}$
$I(q_f) = 4N_f/(1-q_f^2)$
$var(q_f) = 1/I(q_f) = (1-q_f^2)/4N_f$

4.3) total sample

Likelihood function: $P = K.p^{n1}.q^{n2}.(1-q^2)^{n3}.(q^2)^{n4}$
$= K.(1-q)^{n1}.q^{n2+2n4}.(1-q^2)^{n3}$,
$K = N_m!N_f!/(n_1!n_2!n_3!n_4!)$
$L = \log(P) = n_1.\log(1-q) + (n_2+2n_4).\log q$
$+ n_3.\log(1-q^2) + k$
Max. lik. estimates: $q = \{-n_1 + \sqrt{[n_1^2+4(N_m+2N_f)(n_2+2n_4)]}\}/2(N_m+2N_f)$,
$p = 1 - q$
$I(q) = I(q_m)+I(q_f) = [N_m+q(N_m+4N_f)]/[q(1-q^2)]$
$q \approx [q_m.I(q_m)+q_f.I(q_f)]/[I(q_m)+I(q_f)]$
$var(q) = 1/I(q) = 1/[I(q_m)+I(q_f)]$
$= q(1-q^2)/[N_m+q(N_m+4N_f)]$

MAXIMUM LIKELIHOOD ESTIMATE FOR THE FREQUENCY OF DOMINANT AUTOSOMAL ALLELES

Let D and R be the observed numbers of dominant (AA+Aa) and recessive (aa) individuals in a randon sample of G individuals. Assuming panmixia, it comes out that the probability of such a result is given by

$P(D,R) = G![P(1)]^D.[P(2)]^R/(D!R!)$, where $P(1) = 1-q^2$ and $P(2) = q^2$.

Putting $L = \ln P = \text{const.} + D.\ln(1-q^2) + 2R.\ln q$ and $dL/dq = 0$, it comes out that

$$dL/dq = 0 = -2Dq/(1-q^2) + 2R/q$$
$$2Dq^2 = 2R(1-q^2)$$
$$(2D+2R)q^2 = 2R$$
$$q = \sqrt{(R/G)}$$

and

$$d^2L/dq^2 = d(dL/dq)/dq = -2D(1+q^2)/(1-q^2)^2 - 2R/q^2 ;$$

since $R = Gq^2$ and $D = G(1-q^2)$,
at the estimation point $q = \sqrt{(R/G)}$ the second derivative has the numerical value

$$d^2L/dq^2 = - [2Dq^2(1+q^2)+2R(1-q^2)^2]/[q^2(1-q^2)^2] =$$
$$= - [2Gq^2(1-q^2)(1+q^2)+2Gq^2(1-q^2)^2]/[q^2(1-q^2)^2] =$$
$$= - [2G(1+q^2)+2G(1-q^2)]/(1-q^2) = - 4G/(1-q^2);$$

therefore, $\text{var}(q) = - 1/(d^2L/dq^2)$
$= (1-q^2)/4G$.

The result just obtained can be straightforwardly derived using the principle of functional invariance. This is shown in the lines that follow.
Putting $y = q^2$ and $1-y = 1-q^2$, it comes out that
$\text{var}(y) = \text{var}(q^2) = y(1-y)/G$, that is the usual formula for binomial variance); using the property
$\text{var}(y) = (dy/dq)^2 . \text{var}(q)$,
where $dy/dq = 2q$ [and therefore $(dy/dq)^2 = 4q^2$],
we get $\text{var}(y) = q^2(1-q^2)/G = (dy/dq)^2 . \text{var}(q) = 4q^2.\text{var}(q)$;
therefore, $\text{var}(q) = q^2(1-q^2)/4Gq^2 = (1-q^2)/4G$.

We note that $(1-q^2)/4G = (p^2+2pq)/4G = pq/2G + p^2/4G > pq/2G$, as expected.

Numerical example: in a sample of G = 18 randomly collected individuals D = 16 had the dominant phenotype (A- = AA or Aa), while R = 2 exhibited the recessive phenotype corresponding to genotype aa.

Under the ancillary hypothesis of panmixia the expected numbers of dominant individuals are respectively $G(1-q^2)$ and Gq^2, as shown below.

genotypes	expected frequencies	observed numbers	expected numbers
AA + Aa	$p^2 + 2pq = 1 - q^2$	D = 16	$G(1-q^2)$
aa	q^2	R = 2	Gq^2
total	1	G = 18	

The likelihood function is then given by

$P = 153.q^4.(1-q^2)^{16}$ or by $L = \ln(P) = \ln(153) + 4.\ln(q) + 16.\ln(1-q^2)$.

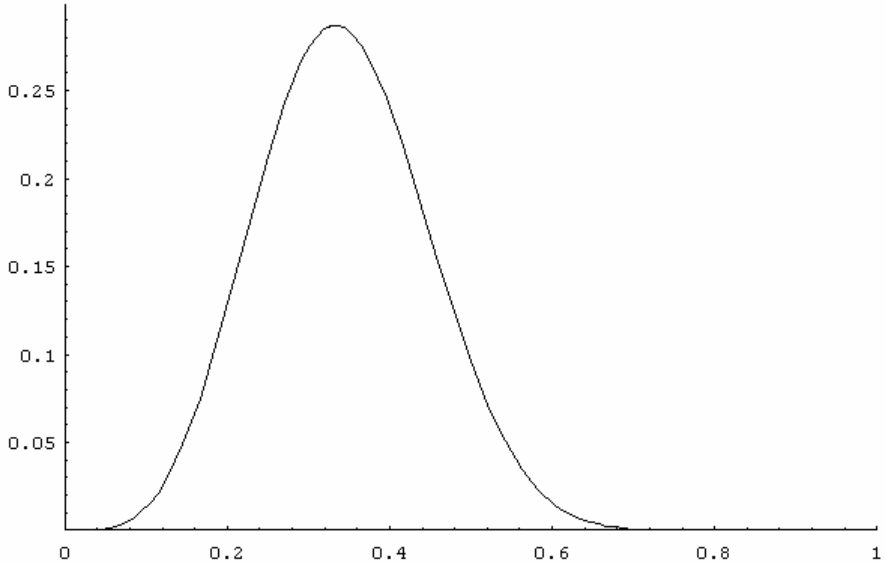
The values P and L take as q varies from 0 to 1 are shown in the table below.

q	P	L
0.000	0.000000	-inf
0.050	0.000919	-6.992541
0.100	0.013027	-4.340708
0.150	0.053818	-2.922154
0.200	0.127395	-2.060466
0.250	0.212810	-1.547356
0.300	0.274056	-1.294424
0.331	0.286860	-1.248761
0.332	0.286903	-1.248S12
0.333	0.286922	-1.248544
0.334	0.286918	-1.248558
0.335	0.286891	-1.248652
0.336	0.286841	-1.248827
0.337	0.286768	-1.249083
0.338	0.286671	1.249420
0.339	0.286552	1.249837
0.350	0.283738	-1.259704
0.400	0.240658	-1.424379
0.450	0.167970	-1.783968
0.500	0.095841	-2.345064
0.550	0.043939	-3.124954
0.600	0.015710	-4.153458
0.650	0.004180	~5.477443
0.700	0.000769	-7.169775
0.750	0.000087	-9.347148
0.800	0.000005	-12.208556
0.850	0.000000	-16.130587
0.900	0.000000	-21.962703
0.950	0.000000	-32.421182
1.000	0.000000	-inf

The maximum value P or L take occurs when $q = \sqrt{(2/18)} = 1/3$; and this is precisely the maximum likelihood estimate of q.

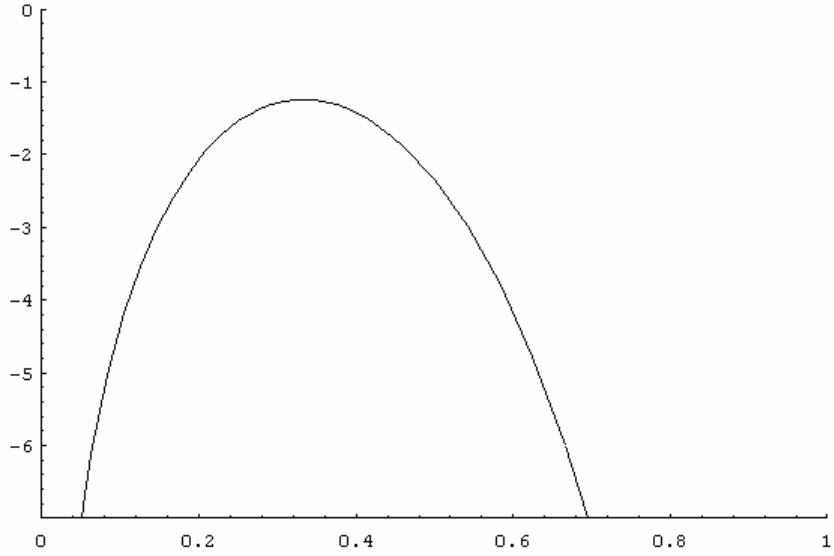
The graphs that follow show the values of P, L and var(q) as functions of q, for D = 16, R = 2 and G = 18. In the last graph the values of var(q) are compared to those obtained using the formula for binomial variance, pq/36. For any q < 1, (1-q^2)/72 > q(1-q)/36, as already stated. The Mathematica codes that generated the graphs are listed below the corresponding figures.

1) Graph of $P = 153.q^4.(1-q^2)^{16}$



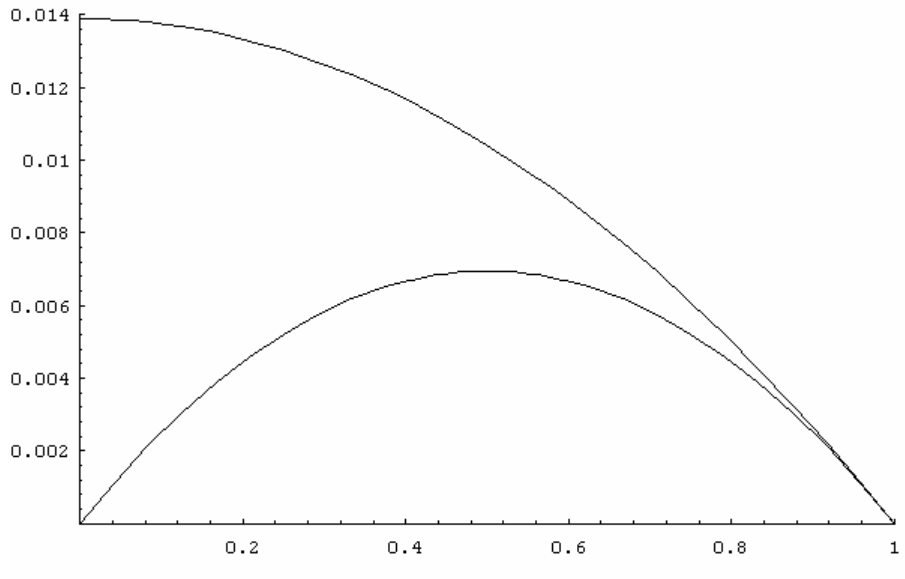

```
(* domlik01.ma *)
P = 153 * q^4 * (1-q^2)^16;
graph = Plot[P,{q,0,1}];
Show[graph, PlotRange -> {0,0.3}, AxesOrigin -> {0,0}]
```

2)Graph of $L = \log(153) + 4.\log(q) + 16.\log(1-q^2)$




```
(* domlik02.ma *)
P = 153 * q^4 * (1-q^2)^16; L = Log[P]; graph = Plot[L,{q,0,1}];
Show[graph, PlotRange -> {-7,0}, AxesOrigin -> {0,-7}]
```

3) Graph of $var'(q) = (1-q^2)/72$ and $var''(q) = q(1-q)/36$




```
(* domlik03.ma *)
var1 = q * (1-q)/36; var2 = (1-q^2)/72;
Plot[{var1,var2},{q,0,1}]
```

GENETIC EQUILIBRIUM IN RELATION TO A PAIR OF LOCI

Let $P_0(AB) = e$
$P_0(Ab) = f$
$P_0(aB) = g$
$P_0(ab) = h$

be the gamete (or haplotype, if the loci are syntenic) composition of a large diploid population at initial generation **0**. The recombination frequency between loci (**A**,**a**) and (**B**,**b**) has a value **r (0.5 ≥ r ≥ 0)**. This means that a coupling heterozygote **AB/ab** produces gametes **AB**, **Ab**, **aB** and **ab** with respective frequencies **(1-r)/2**, **r/2**, **r/2** and **(1-r)/2**, the combined frequency of recombinant gametes (**Ab** and **aB**) being r. Assuming panmixia, the following are the frequencies of possible genotypes in generation **1**:

	AB	Ab	aB	ab	
AB	e^2	ef	eg	eh	e
Ab	ef	f^2	fg	fh	f
aB	eg	fg	g^2	gh	g
ab	eh	fh	gh	h^2	h
	e	f	g	h	1

Therefore, the frequency of **AB** gametes in generation **1** is

$$\begin{aligned}P_1(AB) &= P_1(AB/AB) + P_1(AB/Ab)/2 + P_1(AB/aB)/2 + (1-r)P_1(AB/ab)/2 \\ &\quad + rP_1(Ab/aB)/2 = e^2 + ef + eg + (1-r)eh + rfg \\ &= e2 + ef + eg + eh - r(eh-fg) \\ &= e(e+f+g+h) - r(eh-fg) = e - r(eh-fg) \\ &= P_0(AB) - r[P_0(AB).P_0(ab)-P_0(Ab).P_0(aB)] \\ &= P_0(AB) - r.D_0 \\ &= e^2 + ef + eg + fg + (1-r)eh- fg + rfg \\ &= e(e+f) + g(e+f) + (1-r)(eh-fg) \\ &= (e+f)(e+g) + (1-r)(eh-fg) \\ &= P_0(A).P_0(B) + (1-r)[P_0(AB).P_0(ab)-P_0(Ab).P_0(aB)] \\ &= P_0(A).P_0(B) + (1-r).D_0\end{aligned}$$

Therefore we have

$$P_1(AB) = P_0(AB) - r.D_0 = P_0(A).P_0(B) + (1-r).D_0$$

and, by symmetry,

$$\begin{aligned}P_1(Ab) &= P_0(Ab) + r.D_0 = P_0(A).P_0(b) - (1-r).D_0 \\ P_1(aB) &= P_0(aB) + r.D0 = P_0(a).P_0(B) - (1-r).D_0 \\ P_1(ab) &= P_0(ab) - r.D0 = P_0(a).P_0(b) + (1-r).D_0\ .\end{aligned}$$

Since

$$\begin{aligned}P_1(A) &= P_1(AB) + P_1(Ab) = P_0(AB) + P_0(Ab) = P_0(A) = \ldots = P(A) \\ P_1(a) &= P_1(aB) + P_1(ab) = P_0(aB) + P_0(ab) = P_0(a) = \ldots = P(a) \\ P_1(B) &= P_1(AB) + P_1(aB) = P_0(AB) + P_0(aB) = P_0(B) = \ldots = P(B) \\ P_1(b) &= P_1(Ab) + P_1(ab) = P_0(Ab) + P_0(ab) = P_0(b) = \ldots = P(b)\end{aligned}$$

and

$$P_0(AB) - r.D_0 = P(A).P(B) + (1-r).D_0 = P(A).P(B) + D_0 - r.D_0 ,$$

it comes out that

$$P_0(AB) = P(A).P(B) + D_0 .$$

Comparing this equation with that for $P_1(AB)$,

$$P_1(AB) = P(A).P(B) + (1-r).D_0 ,$$

immediately we get the general solution

$$P_n(AB) = P(A).P(B) + (1-r)^n.D_0$$

and, again by symmetry,

$$P_n(Ab) = P(A).P(b) - (1-r)^n.D_0$$
$$P_n(aB) = P(a).P(B) - (1-r)^n.D_0$$
$$P_n(ab) = P(a).P(b) + (1-r)^n.D_0 .$$

Since at equilibrium, that is when **n** tends to infinity, since $(1-r)^n$ tends to zero as **n** increases,

$$P(AB) = P(A).P(B)$$
$$P(Ab) = P(A).P(b)$$
$$P(aB) = P(a).P(B)$$
$$P(ab) = P(a).P(b) .$$

So we deduce also that at equilibrium obviously the frequencies of the two possible types of double heterozygotes (in coupling and in repulsion) are the same. This is exactly the equilibrium condition, since with equal frequencies of the two possible types of double heterozygotes the production of gametes **AB**, **Ab**, **aB** and **ab** by the whole group of heterozygotes shall be of **1/4** for each gametic class, independent of **r** and as if the loci were unlinked:

	AB	Ab	aB	ab
AB/ab	(1-r)/2	r/2	r/2	(1-r)/2
Ab/aB	r/2	(1-r)/2	(1-r)/2	r/2
average	1/4	1/4	1/4	1/4

We can get the same results using an alternative reasoning, which is shown below. Let us consider again the difference equation

$$P_1(AB) = P(A).P(B) + (1-r).D_0 ;$$

making, in

$$P_2(AB) = P(A).P(B) + (1-r).D_1$$
$$= P(A).P(B) + (1-r).[P_1(AB).P_1(ab)-P_1(Ab).P_1(aB)]$$

the following substitutions

$$P_1(AB) = P_0(A).P_0(B) + (1-r).D_0$$
$$P_1(Ab) = P_0(A).P_0(b) - (1-r).D_0$$
$$P_1(aB) = P_0(a).P_0(B) - (1-r).D_0$$

$$P_1(ab) = P_0(a).P_0(b) + (1-r).D_0$$

we get

$$P_2(AB) = P(A).P(B) + (1-r).(1-r).D_0 = P(A).P(B) + (1-r)^2.D_0$$

and the general solution

$$P_n(AB) = P(A).P(B) + (1-r)^n.D_0$$

already found using the first method.

The method just shown is interesting because it demonstrates clearly that

$$D_1 = (1-r).D_0$$

and therefore that

$$D_n = (1-r)^n.D_0 \ ;$$

when **n** tends to infinity, $(1-r)^n$ tends to zero, so that deleting the subscripts, consistent with equilibrium, yields

$$D = P(AB).P(ab) - P(Ab).P(aB) = 0 \ ,$$

which is (again) the equilibrium condition.

There is a third manner to get the equations that describe the approach to equilibrium (Crow & Kimura, 1970, p.47-48). If we define:

a) $P_n(A_iB_j)$: frequency of the haplotype A_iB_j at generation **n**;
b) $P_{n+1}(A_iB_j)$: same frequency in the next generation;
c) $P_n(A_i) = P(A_i)$: frequency of the **i-th** allele of the **A** locus in any generation or, for sufficiently large populations, the probability of a given allele of the **A** locus being the **i-th** one;
d) $P_n(B_j) = P(B_j)$: frequency of the **j-th** allele of the **B** locus in any generation or, for sufficiently large populations, the probability of a given allele of the **B** locus being the **j-th** one;

we have immediately

$$\begin{aligned} P_{n+1}(A_iB_j) &= P_n(A_iB_j) + r.P(A_i).P(B_j) - r.P_n(A_iB_j) \\ &= (1-r).P_n(A_iB_j) + r.P(A_i).P(B_j); \end{aligned}$$

subtracting from both sides of the above equation the constant quantity $P(A_i).P(B_j)$, we obtain

$$P_{n+1}(A_iB_j) - P(A_i).P(B_j) = (1-r).[P_n(A_iB_j) - P(A_i).P(B_j)]$$

and, therefore, the general solution

$$P_n(A_iB_j) - P(A_i).P(B_j) = (1-r)^n.[(P_0(A_iB_j) - P(A_i).P(B_j)]$$

Since, as we have shown before,

$$P_0(A_iB_j) - rD_0 = P(A_i).P(B_j) + (1-r)D_O \ ,$$

it comes out that

$$P_0(A_iB_j) = P(A_i).P(B_j) + D_0$$

Substituting this in the general solution shown above, we obtain

$$P_n(A_iB_j) - P(A_i).P(B_j) = (1-r)^n.[(P_0(A_iB_j) - P(A_i).P(B_j)]$$
$$= (1-r)^n.[P(A_i).P(B_j) + D_0 - P(A_i).P(B_j)]$$

and

$$P_n(A_iB_j) = P(A_i).P(B_j) + (1-r)^n.D_0$$

which is the solution which we have obtained before.

In general, for any number of syntenic or linked loci (as well as for unlinked loci), at equilibrium

$$P(A_iB_jC_k...) = P(A_i).P(B_j).P(C_k). \ ... \ .$$

The quantity

$$\Delta(A_iB_jC_k...) = P(A_iB_jC_k...) - P(A_i).P(B_j).P(C_k). \ ...$$

is the so-called linkage disequilibrium value for the haplotype $A_iB_jC_k...$. This linkage disequilibrium value may arise as a result of the loci being very near [making thus recombination virtually impossible, as is the case of loci (C,c), (D,d) and (E,e) in Rh blood group system] or as a result of several other factors, such as differential viabilities (or adaptive values) acting on different haplotypes.

The important points to be kept in mind are the following:

1) it is impossible to ascertain linkage using population data, since at equilibrium the population distribution of possible genotypes is exactly the same one observed in relation of two independently inherited loci (that is, situated on different chromosomes); for both cases, this is given by

$$P(AABB) = P(A)^2.P(B)^2$$
$$P(AABb) = 2.P(A)^2.P(B).P(b)$$
$$P(AAbb) = P(A)^2.P(b)^2$$
$$P(AaBB) = 2.P(A).P(a).P(B)^2$$
$$P(AaBb) = P(AB/ab) + P(Ab/aB) = 4.P(A).P(a).P(B).P(b)$$
$$P(Aabb) = 2.P(A).P(a).P(b)^2$$
$$P(aaBB) = P(a)^2.P(B)^2$$
$$P(aaBb) = 2.P(a)^2.P(B).P(b)$$
$$P(aabb) = P(a)^2.P(b)^2 \ ;$$

2) when **r = 1/2**, both types of heterozygotes (**AB/ab** and **Ab/aB**) produce the four possible types of gametes **AB**, **Ab**, **aB** and **ab** with identical frequencies (each one equal to **1/4**); this case corresponds therefore to independent assortment; however, as in the case **r < 0.5**, the population is in an equilibrium state if and only if

$$P(AaBb) = 4.P(A).P(B).P(B).P(b).$$

3) if two loci are separated by a relatively large distance in the chromosome, it is quite probable that the recombination fraction value between the genes from these two loci will approach a value of **1/2,** rendering it difficult or even impossible to demonstrate linkage.

As a numerical example, let us consider the following population, whose gametic composition at generation **0** is the following one:

	B	**b**	
A	**0.2000**	**0.2500**	**0.4500**
a	**0.2000**	**0.3500**	**0.5500**
	0.4000	**0.6000**	**1.0000** ,

that is, $P_0(AB) = 0.20$, $P_0(Ab) = 0.25$, $P_0(aB) = 0.20$, $P_0(ab) = 0.35$, and

$P_0(A) = P(A) = P_0(AB) + P_0(Ab) = 0.45$,
$P0(a) = P(a) = P_0(aB) + P_0(ab) = 0.55$,
$P_0(B) = P(B) = P_0(AB) + P_0(aB) = 0.40$,
$P_0(b) = P(b) = P_0(Ab) + P_0(ab) = 0.60$.

Assuming panmixia and that the recombination frequency is **r = 0.5** (it is therefore irrelevant if the genes are syntenic or not) the following numerical values are obtained for genotype, haplotype and allele frequencies in generations **1 - 10**:

		BB	**Bb**	**bb**			**B**	**b**	
1	**AA**	**0.0400**	**0.1000**	**0.0625**	**0.2025**	**A**	**0.1900**	**0.2600**	**0.4500**
	Aa	**0.0800**	**0.2400**	**0.1750**	**0.4950**	**a**	**0.2100**	**0.3400**	**0.5500**
	aa	**0.0400**	**0.1400**	**0.1225**	**0.3025**		**0.4000**	**0.6000**	**1.0000**
		0.1600	**0.4800**	**0.3600**	**1.0000**				
		BB	**Bb**	**bb**			**B**	**b**	
2	**AA**	**0.0361**	**0.0988**	**0.0676**	**0.2025**	**A**	**0.1850**	**0.2650**	**0.4500**
	Aa	**0.0798**	**0.2384**	**0.1768**	**0.4950**	**a**	**0.2150**	**0.3350**	**0.5500**
	aa	**0.0441**	**0.1428**	**0.1156**	**0.3025**		**0.4000**	**0.6000**	**1.0000**
		0.1600	**0.4800**	**0.3600**	**1.0000**				
		BB	**Bb**	**bb**			**B**	**b**	
3	**AA**	**0.0342**	**0.0981**	**0.0702**	**0.2025**	**A**	**0.1825**	**0.2675**	**0.4500**
	Aa	**0.0796**	**0.2379**	**0.1775**	**0.4950**	**a**	**0.2175**	**0.3325**	**0.5500**
	aa	**0.0462**	**0.1440**	**0.1122**	**0.3025**		**0.4000**	**0.6000**	**1.0000**
		0.1600	**0.4800**	**0.3600**	**1.0000**				
		BB	**Bb**	**bb**			**B**	**b**	
4	**AA**	**0.0333**	**0.0976**	**0.0716**	**0.2025**	**A**	**0.1813**	**0.2687**	**0.4500**
	Aa	**0.0794**	**0.2377**	**0.1779**	**0.4950**	**a**	**0.2188**	**0.3312**	**0.5500**
	aa	**0.0473**	**0.1446**	**0.1106**	**0.3025**		**0.4000**	**0.6000**	**1.0000**
		0.1600	**0.4800**	**0.3600**	**1.0000**				
		BB	**Bb**	**bb**			**B**	**b**	
5	**AA**	**0.0329**	**0.0974**	**0.0722**	**0.2025**	**A**	**0.1806**	**0.2694**	**0.4500**
	Aa	**0.0793**	**0.2377**	**0.1780**	**0.4950**	**a**	**0.2194**	**0.3306**	**0.5500**
	aa	**0.0479**	**0.1449**	**0.1097**	**0.3025**		**0.4000**	**0.6000**	**1.0000**
		0.1600	**0.4800**	**0.3600**	**1.0000**				
		BB	**Bb**	**bb**			**B**	**b**	
6	**AA**	**0.0326**	**0.0973**	**0.0726**	**0.2025**	**A**	**0.1803**	**0.2697**	**0.4500**
	Aa	**0.0792**	**0.2376**	**0.1781**	**0.4950**	**a**	**0.2197**	**0.3303**	**0.5500**
	aa	**0.0481**	**0.1451**	**0.1093**	**0.3025**		**0.4000**	**0.6000**	**1.0000**
		0.1600	**0.4800**	**0.3600**	**1.0000**				

```
-----------------------------------------------------------------------
        BB      Bb      bb                      B       b
 7  AA  0.0325  0.0973  0.0727  0.2025      A   0.1802  0.2698  0.4500
    Aa  0.0792  0.2376  0.1782  0.4950      a   0.2198  0.3302  0.5500
    aa  0.0483  0.1451  0.1091  0.3025          0.4000  0.6000  1.0000
        0.1600  0.4800  0.3600  1.0000
-----------------------------------------------------------------------
        BB      Bb      bb                      B       b
 8  AA  0.0325  0.0972  0.0728  0.2025      A   0.1801  0.2699  0.4500
    Aa  0.0792  0.2376  0.1782  0.4950      a   0.2199  0.3301  0.5500
    aa  0.0483  0.1452  0.1090  0.3025          0.4000  0.6000  1.0000
        0.1600  0.4800  0.3600  1.0000
-----------------------------------------------------------------------
        BB      Bb      bb                      B       b
 9  AA  0.0324  0.0972  0.0729  0.2025      A   0.1800  0.2700  0.4500
    Aa  0.0792  0.2376  0.1782  0.4950      a   0.2200  0.3300  0.5500
    aa  0.0484  0.1452  0.1090  0.3025          0.4000  0.6000  1.0000
        0.1600  0.4800  0.3600  1.0000
-----------------------------------------------------------------------
        BB      Bb      bb                      B       b
10  AA  0.0324  0.0972  0.0729  0.2025      A   0.1800  0.2700  0.4500
    Aa  0.0792  0.2376  0.1782  0.4950      a   0.2200  0.3300  0.5500
    aa  0.0484  0.1452  0.1089  0.3025          0.4000  0.6000  1.0000
        0.1600  0.4800  0.3600  1.0000
-----------------------------------------------------------------------
```

The table above was generated by the following BASIC code:

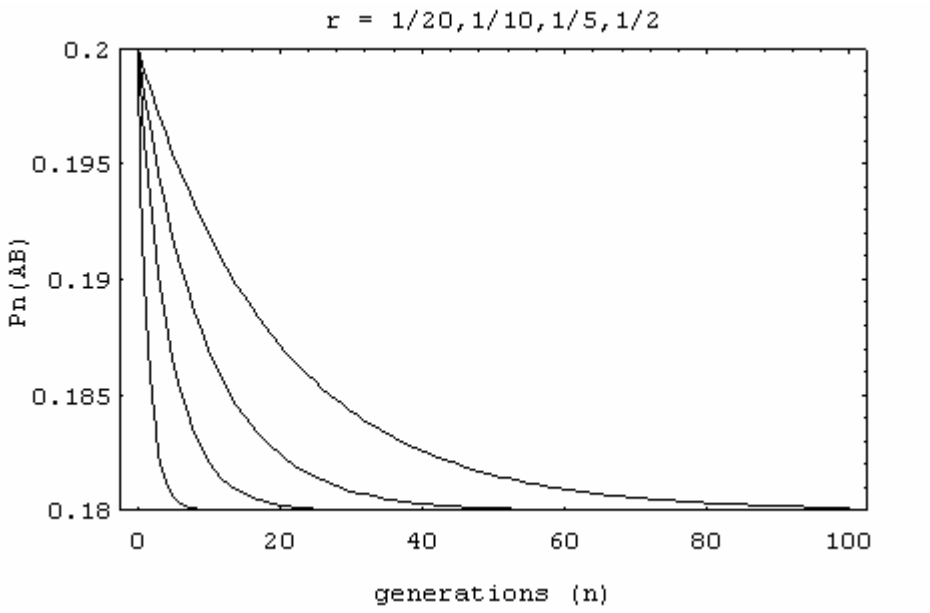
```
REM PROGRAM FILENAME LINKGE01.BAS
REM E = P(AB), F = P(Ab), G = P(aB), H = P(ab)
REM P = P(A) = E+F, Q = P(a) = G+H, S = P(B) = E+G, T = P(b) = F+H
CLS : DEFDBL A-Z
E(0) = .2: F(0) = .25: G(0) = .2: H(0) = .35
P = E(0) + F(0): Q = 1 - P: S = E(0) + G(0): T = 1 - S: R = .5
D = E(0) * H(0) - F(0) * G(0)
FOR I = 1 TO 10
GT1 = E(I - 1) * E(I - 1)'                                GT1 = P(AABB)
GT2 = 2 * E(I - 1) * F(I - 1)'                            GT2 = P(AABb)
GT3 = F(I - 1) * F(I - 1)'                                GT3 = P(AAbb)
GS1 = GT1 + GT2 + GT3'                                    GS1 = P(AA)
GT4 = 2 * E(I - 1) * G(I - 1)'                            GT4 = P(AaBB)
GT5 = 2 * E(I - 1) * H(I - 1) + 2 * F(I - 1) * G(I - 1)' GT5 = P(AaBb)
GT6 = 2 * F(I - 1) * H(I - 1)'                            GT6 = P(Aabb)
GS2 = GT4 + GT5 + GT6'                                    GS2 = P(Aa)
GT7 = G(I - 1) * G(I - 1)'                                GT7 = P(aaBB)
GT8 = 2 * G(I - 1) * H(I - 1)'                            GT8 = P(aaBb)
GT9 = H(I - 1) * H(I - 1)'                                GT9 = P(aabb)
GS3 = GT7 + GT8 + GT9'                                    GS3 = P(aa)
GS4 = GT1 + GT4 + GT7'                                    GS4 = P(BB)
GS5 = GT2 + GT5 + GT8'                                    GS5 = P(Bb)
GS6 = GT3 + GT6 + GT9'                                    GS6 = P(bb)
GST = GS1 + GS2 + GS3'                                    GST = 1
E(I) = P * S + (1 - R) ^ I * D
F(I) = P * T - (1 - R) ^ I * D
G(I) = Q * S - (1 - R) ^ I * D
H(I) = Q * T + (1 - R) ^ I * D
PRINT "-----------------------------------------------------------------------"
PRINT "        BB      Bb      bb                      B       b"
PRINT USING "##  AA  "; I; : PRINT USING "#.####  "; GT1; GT2; GT3; GS1;
PRINT "    A  "; : PRINT USING "#.####  "; E(I); F(I); P
PRINT "    Aa  "; : PRINT USING "#.####  "; GT4; GT5; GT6; GS2;
PRINT "    a  "; : PRINT USING "#.####  "; G(I); H(I); Q
PRINT "    aa  "; : PRINT USING "#.####  "; GT7; GT8; GT9; GS3;
```

```
PRINT "       "; : PRINT USING "#.####  "; S; T; S + T
PRINT "        "; : PRINT USING "#.####  "; GS4; GS5; GS6; GST
NEXT I
PRINT "-------------------------------------------------------------------------"
```

Inspection of the above table shows that Hardy-Weinberg proportions are attained for each locus separately already in the first generation of random mating, as shown by the marginal frequencies of the matrices for genotypic frequencies shown at left. The determinant of the gamete matrix represents the linkage disequilibrium value for haplotypes **AB** or **ab**, which is Δ **= 0.20 x 0.35 - O.20 x 0.25 = 0.07 - 0.05 = 0.02** at generation **0** and Δ **= 0.18 x 0.33 - 0.22 x 0.27 = 0** after an infinite number of generations (the rounded values, with four significant digits or an absolute error equal or less than 0.00005, obtained at generation 10 are already in this situation). The marginals of this last matrix indicate that gene frequencies do not suffer any alteration during the whole process.

In the example just worked, convergence to approximate equilibrium state was fast because **r** was assumed to be **0.5**, the highest value the recombination fraction can take. For other values (such as 0.2, 0.1 and 0.05), convergence takes place slowly, as the following Mathematica graph shows.

```
(* linkge01.ma
   Haplotype ab frequencies
   Pn = P(A).P(B) + (1-r)^n . D
   P0 = 0.20, D = 0.02, r = 0.5, 0.2, 0.1, 0.05
*)
F[n_,r_] := 0.18 + 0.02 * (1-r)^n;
Plot[{F[i,0.5],F[i,0.2],F[i,0.1],F[i,0.05]},{i,0,100},
         PlotPoints->101, Frame -> True,
         PlotLabel->"r = 1/20,1/10,1/5,1/2",
         FrameLabel->{"generations (n)","Pn(AB)"},
         PlotRange -> {0.18,0.20}, AxesOrigin -> {0,0.18}]
```



EXERCISES

1) The following are the frequencies of **Rh** haplotypes in England (Race et al., 1948, apud Race RR & Sanger R, "Blood Groups in Man", Blackwell Scientific Publications, Oxford, 1962):

CDE	**0.0024**
CDe	**0.4205**
CdE	**0.0000**
Cde	**0.0098**
cDE	**0.1411**
cDe	**0.0257**
cdE	**0.0119**
cde	**0.3886**
	1.0000

Estimate: a) the frequencies of alleles **C** and **c**, **D** and **d**, **E** and **e**, and their respective standard errors; b) the haplotype frequencies on the hypothesis of equilibrium; c) the linkage disequilibrium values for each one of the above haplotypes.

2) The following are the results of the testing of a sample of 1400 Hungarians (Rex-Kiss & Horvath, 1966) with 5 Rh anti-sera (anti-C, anti-c, anti-D, anti-E, and anti-e):

REACTION WITH ANTI-					
C	**c**	**D**	**E**	**e**	
+	**-**	**+**	**+**	**+**	**3**
+	**-**	**+**	**-**	**+**	**260**
+	**+**	**+**	**+**	**+**	**182**
+	**+**	**+**	**-**	**+**	**502**
+	**+**	**-**	**-**	**+**	**13**
-	**+**	**+**	**+**	**-**	**37**
-	**+**	**+**	**+**	**+**	**156**
-	**+**	**+**	**-**	**+**	**23**
-	**+**	**-**	**+**	**+**	**6**
-	**+**	**-**	**-**	**+**	**218**
					1400

Estimate the frequencies for the alleles (**C**, **c**), (**D** ,**d**), and (**E**, **e**), with respective standard errors. Estimate the frequencies of the eight haplotypes of Rh system (**CDE**, **CDe**, **CdE**, **Cde**, **cDE**, **cDe**, **cdE** and **cde**). In order to achieve this, you should first write a program based on information contained in pages 53-54 of Mourant, Kopec & Domaniewska-Sobczak's book. Estimate the linkage disequilibrium values for these haplotypes.

CALCULATION OF HAPLOTYPE FREQUENCIES AND OF LINKAGE DISEQUILIBRIUM VALUES FOR LINKED GENE COMPLEXES (E.G., HLA-SYSTEM, Rh-SYSTEM)

This is accomplished by determining, in a population sample, the frequencies of individuals **+/+**, **+/-**, **-/+** and **-/-** who react with two different anti-sera (e.g., anti-sera **anti-A_i** and **anti-B_j**, that is, anti-sera that detect the i-th antigen determined by the one of the alleles--the i-th one--belonging to the **A** locus and the j-th antigen determined by the j-th allele of the **B** locus. **A** and **B** afe **syntenic**, that is, are assumed to be located in the same chromosome). Let us suppose that the results among **N** sampled individuals were the following:

REACTION WITH ANTI-A_i	REACTION WITH ANTI-B_j	
+	+	n_1
+	-	n_2
-	+	n_3
-	-	n_4
		N

The above frequencies can be rearranged as the following contingency table:

		REACTION WITH ANTI-Ai: +	REACTION WITH ANTI-Ai: -	
REACTION WITH ANTI-Bj	+	n_1	n_3	n_1+n_3
	-	n_2	n_4	n_2+n_4
		n_1+n_2	n_3+n_4	N

The frequency of the "double-recessive" **ab/ab** is n_4/N. Since the sample is composed of **N** unrelated individuals, in order to proceed, we assume tacitly that the frequency of **ab/ab** individuals is the square of the **ab** frequency. Therefore the inferred frequency of the haplotype **ab** is

$$P(a\text{-}b) = \sqrt{(n_4/N)}.$$

Under the hypothesis that the linkage disequilibrium value is zero, the expected frequency for the **ab** haplotype is given by the expression

$$P'(a\text{-}b) = (1-p_i)(1-p_j) = q_i q_j ,$$

where $p_i = 1 - q_i$ is the frequency of the A_i allele in the **A** locus and $p_j = 1 - q_j$ the frequency of the B_j allele in locus **B**. The values q_i and q_j are easily estimated from the above contingency table:

$$q_i = \sqrt{[(n_3+n_4)/N]} \text{ and } q_j = \sqrt{[(n_2+n_4)/N]}.$$

If we define the linkage disequilibrium value as $\Delta(a\text{-}b) = P(a\text{-}b) - P'(a\text{-}b)$, it comes out that $\Delta(a\text{-}b) = \sqrt{(n_4/N)} - \sqrt{[(n_3+n_4)(n_2+n_4)]/N}$.

This is the required linkage disequilibrium value between the genes **a** and **b** of loci **A** and **B**.

Numerical exercise:

A sample of 1967 unrelated danes was typed as to antigens A_1 and B_8 of the HLA-system. The results were as follows (Svejgaard et al., 1975, ref. 115 apud Vogel F & Motulsky A, "Human Genetics", Springer Verlag, New York and Berlin, 1979):

REACTION WITH ANTI-A_1	REACTION WITH ANTI-B_8	
+	+	376
+	-	235
-	+	91
-	-	1265
		1967

1) Using a chi-squared test, determine if there is a significant deviation from linkage equilibrium (this can be done by verifying whether there is or is not association between the antigens).

2) What is the frequency, in this population, of the A_1 allele?

3) What is the frequency, in this population, of the B_8 allele?

4) What are the estimated frequencies of the four possible haplotypes (A_1-B_8, A_1-**b**, **a**-B_8, **a-b**) in this population?

5) Under equilibrium conditions, what should be the frequencies of the above haplotypes?

6) What is the value of the linkage disequilibrium between gene A_1 and B_8 from loci **A** and **B** of HLA-system?

Solution of exercise:

1) In the absence of association between the A_1 and B_8 antigens, the expected frequencies of A_1+B_8+, A_1+B_8-, A_1-B_8+, and A_1-B_8- individuals (shown here with the observed) are

	N(e)	N(o)
A_1+B_8+	$(n_1+n_2)(n_1+n_3)/N$ = 145.06	n_1 = 376
A_1+B_8-	$(n_1+n_2)(n_2+n_4)/N$ = 465.94	n_2 = 235
A_1-B_8+	$(n_3+n_4)(n_1+n_3)/N$ = 321.94	n_3 = 91
A_1-B_8-	$(n_3+n_4)(n_2+n_4)/N$ = 1034.06	n_4 = 1265

The value of the chi-squared test is

$$\chi^2 \text{ (1 d.f.)} = \Sigma_i\{[N_i(o)-N_i(e)]^2/N_i(e)\}$$
$$= \Sigma_i\{[(N_i(o)]^2/N_i(e)\} - N$$

$$= (n_1n_4-n_2n_3)^2N/[(n_1+n_2)(n_1+n_3)(n_2+n_4)(n_3+n_4)]$$
$$= 699.35 >>> 3.841 ,$$

indicating thus a very significant association between the two antigens.

2) The frequencies of the gene $\mathbf{A_1}$ and of its "allele" **a** are calculated as follows:

$$P(a) = q_1 = \sqrt{[(n_3+n_4)/N]} = 0.8303$$
$$P(A_1) = p_1 = 1 - q_1 = 1 - \sqrt{[(n_3+n_4)/N]} = 0.1697 .$$

3) And the frequencies of the gene $\mathbf{B_8}$ and of its "allele" **b** as:

$$P(b) = q_2 = \sqrt{[(n_2+n_4)/N]} = 0.8733$$
$$P(B_8) = p_2 = 1 - q_2 = 1 - \sqrt{[(n_2+n_4)/N]} = 0.1267 .$$

4) Since

$$P(a\text{-}B_8) + P(a\text{-}b) = q_1$$
$$P(A_1\text{-}b) + P(a\text{-}b) = q_2$$
$$P(A_1\text{-}B_8) = 1 - [P(A_1\text{-}b) + P(a\text{-}B_8) + P(a\text{-}b)] ,$$

it comes out that the inferred frequencies of the four possible haplotypes are

$$P(a\text{-}b) = \sqrt{(n_4/N)} = 0.8019$$
$$P(a\text{-}B_8) = q_1 - \sqrt{(n_4/N)} = 0.0283$$
$$P(A_1\text{-}b) = q_2 - \sqrt{(n_4/N)} = 0.0713$$
$$P(A_1\text{-}B_8) = 1 - q_1 - q_2 + \sqrt{(n_4/N)} = 0.0984 .$$

5) Under equilibrium conditions, the expected frequencies of these haplotypes should be

$$P'(a\text{-}b) = q_1q_2 = 0.7251$$
$$P'(a\text{-}B_8) = q_1(1 - q_2) = 0.1052$$
$$P'(A_1\text{-}b) = (1 - q_1)q_2 = 0.1482$$
$$P'(A_1\text{-}B_8) = (1 - q_1)(1 - q_2) = 0.0215 .$$

6) The linkage disequilibrium values of these four haplotypes are therefore

$$\Delta(a\text{-}b) = P(a\text{-}b) - P'(a\text{-}b) = +0.0769$$
$$\Delta(a\text{-}B_8) = P(a\text{-}B_8) - P'(a\text{-}B_8) = -0.0769$$
$$\Delta(A_1\text{-}b) = P(A_1\text{-}b) - P'(A_1\text{-}b) = -0.0769$$
$$\Delta(A_1\text{-}B_8) = P(A_1\text{-}B_8) - P(A_1\text{-}B_8) = +0.0769.$$

The following BASIC code performs all the calculations indicated above, as shown by the screen printout appended to it:

```
REM PROGRAM FILENAME HLAHAPL2
REM HLA SYSTEM HAPLOTYPE ESTIMATION
DEFDBL A-Z: CLS : LOCATE 10: C$ = "NO. OF INDIVIDUALS "
INPUT "FIRST  ANTIGEN IDENTIFICATION = "; A$
INPUT "SECOND ANTIGEN IDENTIFICATION = "; B$: PRINT
PRINT C$ + A$ + "(+)/" + B$ + "(+) = "; : INPUT "", N1
PRINT C$ + A$ + "(+)/" + B$ + "(-) = "; : INPUT "", N2
PRINT C$ + A$ + "(-)/" + B$ + "(+) = "; : INPUT "", N3
PRINT C$ + A$ + "(-)/" + B$ + "(-) = "; : INPUT "", N4: CLS
```

```
PRINT "   " + C$ + A$ + "(+)/" + B$ + "(+) = "; : PRINT USING "#####"; N1
PRINT "   " + C$ + A$ + "(+)/" + B$ + "(-) = "; : PRINT USING "#####"; N2
PRINT "   " + C$ + A$ + "(-)/" + B$ + "(+) = "; : PRINT USING "#####"; N3
PRINT "   " + C$ + A$ + "(-)/" + B$ + "(-) = "; : PRINT USING "#####"; N4
N = N1 + N2 + N3 + N4: PRINT "   " + C$ + "TESTED      = ";
PRINT USING "#####"; N: PRINT "GENE FREQUENCIES"
Q1 = SQR((N3 + N4) / N): P1 = 1 - Q1: Q2 = SQR((N2 + N4) / N): P2 = 1 - Q2
PRINT "   P(" + A$ + ") = "; : PRINT USING "#.####"; P1
PRINT "   P(" + B$ + ") = "; : PRINT USING "#.####"; P2
PA0B0 = SQR(N4 / N): PA0B1 = 1 - P1 - PA0B0: PA1B0 = 1 - P2 - PA0B0
PA1B1 = P1 + P2 + PA0B0 - 1: PRINT "INFERRED HAPLOTYPE FREQUENCIES"
PRINT "   P(" + A$ + "/" + B$ + ") = "; : PRINT USING "#.####"; PA1B1
PRINT "   P(" + A$ + "/ -) = "; : PRINT USING "#.####"; PA1B0
PRINT "   P(- /" + B$ + ") = "; : PRINT USING "#.####"; PA0B1
PRINT "   P(- / -) = "; : PRINT USING "#.####"; PA0B0
PRINT "EXPECTED HAPLOTYPE FREQUENCIES"
PRINT "   P(" + A$ + "/" + B$ + ") = "; : PRINT USING "#.####"; P1 * P2
PRINT "   P(" + A$ + "/ -) = "; : PRINT USING "#.####"; P1 * Q2
PRINT "   P(- /" + B$ + ") = "; : PRINT USING "#.####"; Q1 * P2
PRINT "   P(- / -) = "; : PRINT USING "#.####"; Q1 * Q2
PRINT "LINKAGE DISEQUILIBRIUM VALUES"
PRINT "   D(" + A$ + "/" + B$ + ") = "; : PRINT USING "#.####"; PA1B1 - P1 * P2
PRINT "   D(" + A$ + "/ -) = "; : PRINT USING "#.####"; PA1B0 - P1 * Q2
PRINT "   D(- /" + B$ + ") = "; : PRINT USING "#.####"; PA0B1 - Q1 * P2
PRINT "   D(- / -) = "; : PRINT USING "#.####"; PA0B0 - Q1 * Q2
```

```
   NO. OF INDIVIDUALS A1(+)/B8(+) =   376
   NO. OF INDIVIDUALS A1(+)/B8(-) =   235
   NO. OF INDIVIDUALS A1(-)/B8(+) =    91
   NO. OF INDIVIDUALS A1(-)/B8(-) =  1265
   NO. OF INDIVIDUALS TESTED      =  1967
GENE FREQUENCIES
   P(A1) = 0.1697
   P(B8) = 0.1267
INFERRED HAPLOTYPE FREQUENCIES
   P(A1/B8) = 0.0984
   P(A1/ -) = 0.0713
   P(- /B8) = 0.0283
   P(- / -) = 0.8019
EXPECTED HAPLOTYPE FREQUENCIES
   P(A1/B8) = 0.0215
   P(A1/ -) = 0.1482
   P(- /B8) = 0.1052
   P(- / -) = 0.7251
LINKAGE DISEQUILIBRIUM VALUES
   D(A1/B8) = 0.0769
   D(A1/ -) = -.0769
   D(- /B8) = -.0769
   D(- / -) = 0.0769
```

LINKAGE DISEQUILIBRIUM CALCULATIONS

In the lines that follow the notation used by Hill (Hill WG. Estimation of linkage disequilibrium in randomly mating populations. **Heredity 33 :** 229-239, 1974) and Weir & Cockerham (Weir BS & Cockerham CC. Estimation of linkage disequilibrium in randomly mating populations. **Heredity 42 :** 105-111, 1979) is retained whenever possible.

1) ABSENCE OF DOMINANCE

We will begin with the situation in which there is no dominance so that the three possible genotypes given by the alleles of each of the two loci [**(AA, Aa, aa)** and **(BB, Bb, bb)**] are easily distinguishable. Let N_{11}, N_{12}, etc be the observed numbers of genotypes **AB/AB**, **AB/Ab**, etc as shown below:

	BB	Bb	bb	
AA	N_{11}	N_{12}	N_{13}	$N_{1.}$
Aa	N_{21}	N_{22}	N_{23}	$N_{2.}$
aa	N_{31}	N_{32}	N_{33}	$N_{3.}$
	$N_{.1}$	$N_{.2}$	$N_{.3}$	N

That is,

N(AABB) = N(AB/AB) = N_{11}
N(AABb) = N(AB/Ab) = N_{12}
N(AAbb) = N(Ab/Ab) = N_{13}
N(AaBB) = N(AB/aB) = N_{21}
N(AaBb) = N(AB/ab) + N(Ab/aB) = $N_{22} = N'_{22} + N''_{22}$
N(Aabb) = N(Ab/ab) = N_{23}
N(aaBB) = N(aB/aB) = N_{31}
N(aaBb) = N(aB/ab) = N_{32}
N(aabb) = N(ab/ab) = N_{33}

Under panmictic equilibrium, the expected genotype frequencies are

	BB	Bb	bb	
AA	f_{11}^2	$2f_{11}f_{12}$	f_{12}^2	p^2
Aa	$2f_{11}f_{21}$	$2f_{11}f_{22} + 2f_{12}f_{21}$	$2f_{12}f_{22}$	$2p(1-p)$
aa	f_{21}^2	$2f_{21}f_{22}$	f_{22}^2	$(1-p)^2$
	q^2	$2q(1-q)$	$(1-q)^2$	1

where

$f_{11} = f(AB) = f(AABB) + f(AABb)/2 + f(AaBB)/2 + f(AB/ab)/2$
$f_{12} = f(Ab) = f(AAbb) + f(AABb)/2 + f(Aabb)/2 + f(Ab/aB)/2$
$f_{21} = f(aB) = f(aaBB) + f(AaBB)/2 + f(aaBb)/2 + f(Ab/aB)/2$
$f_{22} = f(ab) = f(aabb) + f(Aabb)/2 + f(aaBb)/2 + f(AB/ab)/2$

are the haplotype frequencies to be estimated from the data set and

$p = f(A) = (2N_{11} + 2N_{12} + 2N_{13} + N_{21} + N_{22} + N_{23})/2N$
$= (2N_{1.} + N_{2.})/2N = f_{11} + f_{12}$
$1-p = f(a) = 1 - f(A) = f_{21} + f_{22}$

$q = f(B) = (2N_{11} + 2N_{21} + 2N_{31} + N_{12} + N_{22} + N_{32})/2N$
$= (2N_{.1} + N_{.2})/2N = f_{11} + f_{21}$
$1-q = f(b) = 1 - f(B) = f_{12} + f_{22}.$

Since $N(AaBb) = N(AB/ab) + N(Ab/aB) = N'_{22} + N''_{22} = N_{22}$, N'_{22} can take any value from 0 to N_{22} while N''_{22} varies from N_{22} to 0. Therefore, the lower limit for **f(AB)** is necessarily

$f_1(AB) = (2N_{11}+N_{12}+N_{21})/2N$

while its upper limit is given (also necessarily) by

$f_u(AB) = (2N_{11}+N_{12}+N_{21}+N_{22})/2N.$

In the absence of linkage disequilibrium between the genes from loci **(A,a)** and **(B,b)**, the estimate of **f(AB)** is given simply by

$f_0(AB) = (2N_{11}+N_{12}+N_{21}+N_{22}/2)/2N.$

Since the coefficient of linkage disequilibrium is defined as

$\Delta(AB) = f(AB) - f(A).f(B) = f_{11} - pq,$

it comes out that

$\Delta(AB) = f_{11} - (f_{11}+f_{12})(f_{11}+f_{21})$
$= f_{11}(f_{11}+f_{12}+f_{21}+f_{22}) - (f_{11}+f_{12})(f_{11}+f_{21})$
$= f_{11}.f_{22}-f_{12}.f_{21}$
$= N'_{22}/N - N''_{22}/2N = (N'_{22}-N''_{22})/2N.$

Assuming that the marginal frequencies for both one-locus genoypes [**(AA, Aa, aa)** and **(BB, Bb, bb)**] are in Hardy-Weinberg proportions, the likelihood function is given by

$P = N!/(N_{11}!...N_{33}!).(f_{11}^2)^{N11}.(2f_{11}f_{12})^{N12}.(f_{12}^2)^{N13}$
$. (2f_{11}f_{21})^{N21}.(2f_{11}f_{22}+2f_{12}f_{21})^{N22}.(2f_{12}f_{22})^{N23}$
$. (f_{21}^2)^{N31}.(2f_{21}f_{22})^{N32}.(f_{22}^2)^{N33},$

so that the frequencies f_{11}, f_{12} and f_{21} can be estimated by maximizing the likelihood function in logarithmic form

$$L = \ln P = \text{const.} + \sum X_{ij}.\ln f_{ij} + N_{22}.\ln(f_{11}.f_{22}-f_{12}.f_{21})$$
$$= \text{const.} + X_{11}.\ln f_{11} + X_{12}.\ln f_{12} + X_{21}.\ln f_{21}$$
$$+ X_{22}.\ln f_{22} + N_{22}.\ln(f_{11}.f_{22}-f_{12}.f_{21})$$
$$= \text{const.} + X_{11}.\ln f_{11} + X_{12}.\ln f_{12} + X_{21}.\ln f_{21}$$
$$+ X_{22}.\ln (1-f_{11}-f_{12}-f_{21})$$
$$+ N_{22}.\ln[f_{11}(1-f_{11}-f_{12}-f_{21})-f_{12}.f_{21}],$$

where $X_{11} = 2N_{11} + N_{12} + N_{21}$
$X_{12} = 2N_{13} + N_{12} + N_{23}$
$X_{21} = 2N_{31} + N_{21} + N_{32}$
$X_{22} = 2N_{33} + N_{23} + N_{32}$.

The partial derivatives $\partial L/\partial f_{11}$, $\partial L/\partial f_{12}$ and $\partial L/\partial f_{21}$ are

$$\partial L/\partial f_{11} = X_{11}/f_{11} - X_{22}/(1-f_{11}-f_{12}-f_{21})$$
$$+ N_{22}(1-2f_{11}-f_{12}-f_{21})/[f_{11}(1-f_{11}-f_{12}-f_{21})+f_{12}f_{21}]$$

$$\partial L/\partial f_{12} = X_{12}/f_{12} - X_{22}/(1-f_{11}-f_{12}-f_{21})$$
$$+ N_{22}(f_{21}-f_{11})/[f_{11}(1-f_{11}-f_{12}-f_{21})+f_{12}f_{21}]$$

$$\partial L/\partial f_{21} = X_{21}/f_{21} - X_{22}/(1-f_{11}-f_{12}-f_{21})$$
$$+ N_{22}(f_{12}-f_{11})/[f_{11}(1-f_{11}-f_{12}-f_{21})+f_{12}f_{21}].$$

The estimates f_{11}, f_{12} and f_{21} are obtained by maximizing the function L, that is, they are the solutions of the set of linearly independent equations

$$\{\partial L/\partial f_{11} = 0,\ \partial L/\partial f_{12} = 0,\ \partial L/\partial f_{21} = 0\}.$$

Since it is not possible to obtain explicit solutions for this set of equations, a numerical method as the generalized Newton-Raphson iterative procedure is used:

$$(f_{ij})_{n+1} = (f_{ij})_n + ((-\partial(\partial L/\partial f_{ij})/\partial f_{ij})^{-1}.(\partial L/\partial f_{ij}))_n$$
$$= (f_{ij})_n + ((-\partial^2 L/\partial f_{ij}^2)^{-1}.(\partial L/\partial f_{ij}))_n$$
$$= (f_{ij})_n + ((V_{ij}).(\partial L/\partial f_{ij}))_n,$$

where $(f_{ij})_n$ is the column vector (at the nth iteration)

$(f_{11}, f_{12}, f_{21})^T$,

$(\partial L/\partial f_{ij})_n$ is the column vector, at iteration n, of partial derivatives

$(\partial L/\partial f_{11}, \partial L/\partial f_{12}, \partial L/\partial f_{21})^T$ and

$(-\partial^2 L/\partial f_{ij}2)_n$ is the variance-covariance matrix (also at iteration n)

$$\begin{matrix} V_{11} & V_{12} & V_{13} \\ (V_{21} & V_{22} & V_{23}) \\ V_{31} & V_{32} & V_{33} \end{matrix} = \begin{matrix} \text{VAR}(f_{11}) & \text{COV}(f_{11},f_{12}) & \text{COV}(f_{11},f_{21}) \\ (\text{COV}(f_{12},f_{11}) & \text{VAR}(f_{12}) & \text{COV}(f_{12},f_{21})) \\ \text{COV}(f_{21},f_{11}) & \text{COV}(f_{21},f_{12}) & \text{VAR}(f_{21}) \end{matrix}$$

$$= \begin{matrix} \text{VAR}(f_{11}) & \text{COV}(f_{11},f_{12}) & \text{COV}(f_{11},f_{21}) \\ (\text{COV}(f_{11},f_{12}) & \text{VAR}(f_{12}) & \text{COV}(f_{12},f_{21})) \\ \text{COV}(f_{11},f_{21}) & \text{COV}(f_{12},f_{21}) & \text{VAR}(f_{21}) \end{matrix}$$

$$= \begin{pmatrix} -\partial^2 L/\partial f_{11}^2 & -\partial^2 L/\partial f_{11}\partial f_{12} & -\partial^2 L/\partial f_{11}\partial f_{21} \\ -\partial^2 L/\partial f_{11}\partial f_{12} & -\partial^2 L/\partial f_{12}^2 & -\partial^2 L/\partial f_{12}\partial f_{21} \\ -\partial^2 L/\partial f_{11} df_{21} & -\partial^2 L/\partial f_{12}\partial f_{21} & -\partial^2 L/\partial f_{21}^2 \end{pmatrix}^{-1}$$

The literal values of the second derivatives are:

$$\partial^2 L/\partial f_{11}^2 = - x_{11}/f_{11}^2 - x_{22}/f_{22}^2 - N_{22}(f_{11}^2+2f_{12}f_{21}+f_{22}^2)/(f_{11}f_{22}+f_{12}f_{21})^2$$
$$\partial^2 L/\partial f_{12}^2 = - x_{12}/f_{12}^2 - x_{22}/f_{22}^2 - N_{22}(f_{21}-f_{11})^2/(f_{11}f_{22}+f_{12}f_{21})^2$$
$$\partial^2 L/\partial f_{21}^2 = - x_{21}/f_{21}^2 - x_{22}/f_{22}^2 - N_{22}(f_{12}-f_{11})^2/(f_{11}f_{22}+f_{12}f_{21})^2$$
$$\partial^2 L/\partial f_{11}\partial f_{12} = - x_{22}/f_{22}^2 - N_{22}(f_{11}^2-f_{11}f_{21}+f_{12}f_{21}+f_{21}f_{22})/(f_{11}f_{22}+f_{12}f_{21})^2$$
$$\partial^2 L/\partial f_{11}\partial f_{21} = - x_{22}/f_{22}^2 - N_{22}(f_{11}^2-f_{11}f_{12}+f_{12}f_{21}+f_{12}f_{22})/(f_{11}f_{22}+f_{12}f_{21})^2$$
$$\partial^2 L/\partial f_{12}\partial f_{21} = - x_{22}/f_{22}^2 - N_{22}(2f_{11}^2-f_{11})/(f_{11}f_{22}+f_{12}f_{21})^2$$
$$\partial^2 L/\partial f_{12} df_{11} = \partial^2 L/\partial f_{11}\partial f_{12}$$
$$\partial^2 L/\partial f_{21}\partial f_{11} = \partial^2 L/\partial f_{11}\partial f_{21}$$
$$\partial^2 L/\partial f_{21}\partial f_{12} = \partial^2 L/\partial f_{12}\partial f_{21} ,$$

where $\mathbf{f_{22} = 1 - f_{11} - f_{12} - f_{21}}$.

Since, at equilibrium, all double heterozygotes combined (**AB/ab** and **Ab/aB**) produce all types of gametes (**AB**, **Ab**, **aB** and **ab**) in exactly equal proportions, the following trial values of **f11**, **f12** and **f21** are used for the initial evaluation of the matrices $\mathbf{(\partial L/\partial f_{ij})}$ and $\mathbf{(-\partial^2 L/\partial f_{ij}^2)^{-1}}$ at the beginning of the iteration process:

$$\mathbf{f_{11} = (2N_{11}+N_{12}+N_{21}+N_{22}/2)/2N = (2X_{11}+N_{22})/4N}$$
$$\mathbf{f_{12} = (N_{12}+2N_{13}+N_{23}+N_{22}/2)/2N = (2X_{12}+N_{22})/4N}$$
$$\mathbf{f_{21} = (N_{21}+2N_{31}+N_{32}+N_{22}/2)/2N = (2X_{21}+N_{22})/4N}$$

After convergence has occurred to the final estimates $\mathbf{f_{11}}$, $\mathbf{f_{12}}$ and $\mathbf{f_{21}}$, the value of the estimate $\mathbf{f_{22}}$ is then directly obtained from $\mathbf{f_{22} = 1-f_{11}-f_{12}-f_{21}}$. The variances of the estimates $\mathbf{f_{11}}$, $\mathbf{f_{12}}$ and $\mathbf{f_{21}}$ are taken straightforwardly from the variance-covariance matrix at the final evaluation points. The variance of $\mathbf{f_{22}}$ is then calculated after

$$\mathbf{VAR(f_{22}) = VAR(f_{11}) + 2COV(f_{11},f_{12}) + 2COV(f_{11},f_{21})}$$
$$\mathbf{+ VAR(f_{12}) + 2COV(f_{12},f_{21}) + VAR(f_{21}).}$$

Since $\mathbf{f(A) = p = f(AB) + f(Ab) = f_{11} + f_{12}}$
$\mathbf{f(a) = 1-p = f(aB) + f(ab) = f_{21} + f_{22}}$
$\mathbf{f(B) = q = f(AB) + f(aB) = f_{11} + f_{21}}$ and
$\mathbf{f(b) = 1-q = f(Ab) + f(ab) = f_{12} + f_{22}}$,

the consistency of estimates can be tested by verifying the following property discovered by Fisher: the variance of **f(A)**, **VAR(p)** and that of **f(B)**, **VAR(q)**, are the ordinary binomial gene frequency variances

VAR(p) = VAR(1-p) = p(1-p)/2N and
VAR(q) = VAR(1-q) = q(1-q)/2N.

Should the estimates be consistent, then the numeric values thus obtained should match the quantities

$$\mathbf{VAR(p) = VAR(1-p) = VAR(f_{11}+f_{12})}$$
$$\mathbf{= VAR(f_{11}) + 2COV(f_{11},f_{12}) + VAR(f_{12})}$$

and

$$VAR(q) = VAR(1-q) = VAR(f_{11}+f_{21})$$
$$= VAR(f_{11}) + 2COV(f_{11},f_{21}) + VAR(f_{21})$$

taken from the variance-covariance matrix at the final evaluation point.

The linkage disequilibrium value is finally estimated from

$$\Delta(AB) = f_{11} - pq.$$

The logarithmic likelihood function

$$L = \ln P = \text{const.} + \sum X_{ij}.\ln f_{ij} + N_{22}.\ln(f_{11}.f_{22}-f_{12}.f_{21})$$
$$= \text{const.} + X_{11}.\ln f_{11} + X_{12}.\ln f_{12} + X_{21}.\ln f_{21}$$
$$+ X_{22}.\ln f_{22} + N_{22}.\ln(f_{11}.f_{22}-f_{12}.f_{21})$$

can also be expressed as a function of a single variable (one of the haplotype frequencies, v.g. f_{11}), since $f_{12} = p-f_{11}$, $f_{21} = q-f_{11}$ and $f_{22} = 1-p-q+f_{11}$:

$$L = \ln P = \text{const.} + X_{11}.\ln f_{11} + X_{12}.\ln(p-f_{11})$$
$$+ X_{21}.\ln(q-f_{11}) + X_{22}.\ln(1-p-q+f_{11})$$
$$+ N_{22}.\ln[f_{11}(1-p-q+f_{11})+(p-f_{11})(q-f_{11})].$$

The estimate f_{11} is then the solution of the equation obtained by putting $dL/df_{11} = 0$. Hill (1974), using a 'counting method,' found that the estimate f_{11} is the solution of the cubic equation

$$f_{11} = \{X_{11} + N_{22}.f_{11}(1-p-q+f_{11})/[f_{11}(1-p-q+f_{11})+(p-f_{11})(q-f_{11})]\}/2N.$$

As before, the estimate of the linkage disequilibrium value is obtained straightforwardly from

$$\Delta(AB) = f(AB) - f(A).f(B) = f_{11} - pq.$$

Instead of determining the value of $\Delta(AB)$ after estimating the haplotype frequencies, we can get it directly if we remember that under linkage disequilibrium the frequencies of the four haplotypes **AB**, **Ab**, **aB** and ab can be all expressed as a function of Δ and the constants **p** and **q**:

$$f_{11} = pq + \Delta$$
$$f_{12} = p(1-q) - \Delta$$
$$f_{21} = (1-p)q - \Delta$$
$$f_{22} = (1-p)(1-q) + \Delta,$$

where Δ is the linkage disequilibrium value of haplotypes **AB** or **ab** and **p**, **1-p**, **q** and **1-q** are the frequencies of the pairs of alleles **A**,**a** and **B**,**b**:

	A	a	
B	pq + Δ	(1-p)q - Δ	q
b	p(1-q) - Δ	(1-p)(1-q)+Δ	1-q
	p	1-p	1

If the observed absolute frequencies of the genotypes **AB/AB**, ..., **ab/ab** are respectively N_{11}, ..., N_{33} in a total of **N** sampled individuals, under the assumption of panmixia the expected quantities are:

GENOTYPE	OBS.ABS.FREQ.	EXP.ABS.FREQ.
AB/AB	N_{11}	$N(pq+\Delta)^2$
AB/Ab	N_{12}	$2N(pq+\Delta)[p(1-q)-\Delta]$
Ab/Ab	N_{13}	$N[p(1-q)-\Delta]^2$
AB/aB	N_{21}	$2N(pq+\Delta)[(1-p)q-\Delta]$
AB/ab + Ab/aB	N_{22}	$2N\{(pq+\Delta)[(1-p)(1-q)+\Delta]$ $+[p(1-q)-\Delta][(1-p)q-\Delta]\}$
Ab/ab	N_{23}	$2N[p(1-q)-\Delta][(1-p)(1-q)+\Delta]$
aB/aB	N_{31}	$N[(1-p)q-\Delta]^2$
aB/ab	N_{32}	$2N[(1-p)q-\Delta][(1-p)(1-q)+\Delta]$
ab/ab	N_{33}	$N[(1-p)(1-q)+\Delta]^2$

The likelihood function **L = ln P** is clearly

$$L = \text{const.} + X_{11}.\ln(pq+\Delta) + X_{12}.\ln[p(1-q)-\Delta] + X_{21}.\ln[(1-p)q-\Delta] + X_{22}.\ln[(1-p)(1-q)+\Delta] + N_{22}.\ln\{(pq+\Delta)[(1-p)(1-q)+\Delta] + [p(1-q)-\Delta][(1-p)q-\Delta]\},$$

where X_{11}, X_{12}, X_{21} and X_{22} are the summary measures already defined. The allelic frequencies can be treated as constants, and they are easily estimated by an independent direct counting method:

$$p = (X_{11}+X_{12}+N_{22})/2N,\ 1-p = (X_{21}+X_{22}+N_{22})/2N$$
$$q = (X_{11}+X_{21}+N_{22})/2N,\ 1-q = (X_{12}+X_{22}+N_{22})/2N\ .$$

The first derivative $dL/d\Delta$ has literal value

$$dL/d\Delta = X_{11}/(pq+\Delta) - X_{12}/[p(1-q)-\Delta] - X_{21}/[(1-p)q-\Delta] + X_{22}/[(1-p)(1-q)+\Delta] + N_{22}[4\Delta+(1-2p)(1-2q)]/[2\Delta^2+\Delta(1-2q)(1-2p)+2pq(1-p)(1-q)]$$

whereas the second derivative takes value

$$d^2L/d\Delta^2 = - X_{11}/(pq+\Delta)^2 - X_{12}/[p(1-q)-\Delta]^2 - X_{21}/[(1-p)q-\Delta]^2 - X_{22}/[(1-p)(1-q)+\Delta]^2 - N_{22}[8pq(1-p)(1-q)+1-4p(1-q)-4(1-p)q] / [2\Delta^2+\Delta(1-2q)(1-2p)+2pq(1-p)(1-q)].$$

The estimate Δ is the solution of the equation $dL/d\Delta = 0$. Since this equation has no explicit solution, a numerical method such as the Newton-Raphson procedure is used to obtain it:

$$\Delta_{n+1} = \Delta_n - f(\Delta)_n/f'(\Delta)_n = \Delta_n + (dL/d\Delta)_n.[-(d^2L/d\Delta^2)_n]^{-1} = \Delta_n + (dL/d\Delta)_n.VAR(\Delta)_n.$$

Hill (1974) showed that a suitable starting value for iteration is given by

$$f_{11} = \Delta_0 + pq = (X_{11}-X_{12}-X_{21}+X_{22})/4N + 1/2 - (1-p)(1-q)$$

and therefore

$$\Delta_0 = (X_{11}-X_{12}-X_{21}+X_{22})/4N + 1/2 - (1-p)(1-q) - pq.$$

Now, let **Fo** be the observed numbers and **Fe'** and **Fe"** respectively the expected values under the assumptions of $\Delta = \Delta(AB) = 0$ and $\Delta = \Delta(AB) \neq 0$ (estimated after any of the methods just delineated) as follows:

Fo	Fe'	Fe"
N_{11}	$N(pq)^2$	$N(pq+\Delta)^2$
N_{12}	$2Np^2q(1-q)$	$2N(pq+\Delta)[p(1-q)-\Delta]$
N_{13}	$N[p(1-q)]^2$	$N[p(1-q)-\Delta]^2$
N_{21}	$2Np(1-p)q^2$	$2N(pq+\Delta)[(1-p)q-\Delta]$
N_{22}	$4Np(1-p)q(1-q)$	$2N\{(pq+\Delta)[(1-p)(1-q)+\Delta]$ $+[p(1-q)-\Delta][(1-p)q-\Delta]\}$
N_{23}	$2Np(1-p)(1-q)^2$	$2N[p(1-q)-\Delta][(1-p)(1-q)+\Delta]$
N_{31}	$N[(1-p)q]^2$	$N[(1-p)q-\Delta]^2$
N_{32}	$2N(1-p)^2q(1-q)$	$2N[(1-p)q-\Delta][(1-p)(1-q)+\Delta]$
N_{33}	$N[(1-p)(1-q)]^2$	$N[(1-p)(1-q)+\Delta]^2$

For testing if $\Delta \neq 0$ the following G difference test is then used:

$$G = 2\sum\{Fo.\ln(Fo/Fe')\} - 2\sum\{Fo.\ln(Fo/Fe")\}$$
$$= 2\sum\{Fo[\ln(Fo/Fe')-\ln(Fo/Fe")]\}$$
$$= 2\sum[Fo.\ln(Fe"/Fe')]$$

$$= 2X_{11}.\ln(1+\Delta/pq) + 2X_{12}.\ln[1-\Delta/p(1-q)]$$
$$+ 2X_{21}.\ln[1-\Delta/(1-p)q] + 2X_{22}.\ln[1+\Delta/(1-p)(1-q)]$$
$$+ 2N_{22}.\ln\{(1+\Delta/pq)[1+\Delta/(1-p)(1-q)]$$
$$+ [1-\Delta/p(1-q)][1-\Delta/(1-p)q]\} .$$

This statistics has a chi-squared distribution with 1 d.f. The usual statistics (that asymptotically has also a chi-squared distribution with 1 d.f.) is

$$N\Delta^2/p(1-p)q(1-q).$$

2) DOMINANCE

If there is dominance in both linked loci **A,a** and **B,b**, it comes out that, in a panmictic sample of **N** individuals tested with anti-A and anti-B sera

$$
\begin{aligned}
f(A+B+) &= p^2 + 2pq + 2pr + 2qr + 2ps = 2p - p^2 + 2qr &&= f_1\\
f(A+B-) &= q^2 + 2qs &&= 2q - 2pq - q^2 - 2qr = f_2\\
f(A-B+) &= r^2 + 2rs &&= 2r - 2pr - 2qr - r^2 = f_3\\
f(A-B-) &= s^2 &&= (1-p-q-r)^2 = f_4
\end{aligned}
$$

where **p**, **q**, **r** and **s** are the frequencies of haplotypes **AB**, **Ab**, **aB** and **ab**.

If the observed numbers of **A+B+**, **A+B-**, **A-B+** and **A-B-** individuals are respectively N_1, N_2, N_3 and N_4 then the estimates **p**, **q**, **r** are the solutions of the set of equations

$\{\partial L/\partial p = 0,\ \partial L/\partial q = 0,\ \partial L/\partial r = 0\}$, where

$$
\begin{aligned}
L &= \sum N_i.\ln f_i\\
&= N_1.\ln f_1 + N_2.\ln f_2 + N_3.\ln f_3 + N_4.\ln f_4\\
&= N_1.\ln(2p-p^2+2qr) + N_2.\ln q + N_2.\ln(2-2p-q-2r)\\
&\quad + N_3.\ln r + N_3.\ln(2-2p-2q-r) + 2N_4.\ln(1-p-q-r).
\end{aligned}
$$

The solutions of the set of equations

$$
\begin{aligned}
\partial L/\partial p &= 2N_1(1-p)/(2p-p^2+2qr) - 2N_2/(q+2s) - 2N_3/(r+2s) - 2N_4/s = 0\\
\partial L/\partial q &= 2N_1 r/(2p-p^2+2qr) + 2N_2 s/(q^2+2qs) - 2N_3/(r+2s) - 2N_4/s = 0\\
\partial L/\partial r &= 2N_1 q/(2p-p^2+2qr) - 2N_2/(q+2s) + 2N_3 s/(r^2+2rs) - 2N_4/s = 0,
\end{aligned}
$$

where $s = 1-p-q-r$, are the obvious ones

$$
\begin{aligned}
p &= f(A) + f(B) + \sqrt{(N_4/N)} - 1\\
&= 1 + \sqrt{(N_4/N)} - \sqrt{[(N_3+N_4)/N]} - \sqrt{[(N_2+N_4)/N]} = 1-q-r-s\\
q &= 1 - f(B) - \sqrt{(N_4/N)}\\
&= \sqrt{[(N_2+N_4)/N]} - \sqrt{(N_4/N)} = \sqrt{[(q+s)^2]} - \sqrt{(s^2)}\\
r &= 1 - f(A) - \sqrt{(N_4/N)}\\
&= \sqrt{[(N_3+N_4)/N]} - \sqrt{(N_4/N)} = \sqrt{[(r+s)^2]} - \sqrt{(s^2)}\\
s &= \sqrt{(N_4/N)} = \sqrt{(s^2)}.
\end{aligned}
$$

The linkage disequilibrium value estimate Δ is obtained directly from

$$
\begin{aligned}
\Delta &= f(AB) - f(A).f(B)\\
&= f(ab) - f(a).f(b)\\
&= \sqrt{(N_4/N)} - \sqrt{[(N_2+N_4)(N_3+N_4)]/N}.
\end{aligned}
$$

For testing the hypothesis $\Delta = 0$ the following chi-squared statistics (with 1 d.f.) is used:

$$
\begin{aligned}
\chi^2 &= N_1^2/[N(1-Q_4^2)(1-Q_3^2)] + N_2^2/[NQ_4^2(1-Q_3^2)]\\
&\quad + N_3^2/[NQ_3^2(1-Q_4^2)] + N_4^2/(NQ_4^2Q_3^2) - N\\
&= N_1^2.N/[(N_1+N_3)(N_1+N_2)] + N_2^2.N/[(N_1+N_3)(N_3+N_4)]\\
&= N_3^2.N/[(N_2+N_4)(N_1+N_2)] + N_4^2.N/[(N_2+N_4)(N_3+N_4)] - N\\
&= (N_1N_4-N_2N_3)^2.N/[(N_1+N_2)(N_1+N_3)(N_2+N_4)(N_3+N_4)],
\end{aligned}
$$

where $Q_3 = 1-f(A) = \sqrt{[(N_3+N_4)/N]}$

$Q_4 = 1-f(B) = \sqrt{[(N_2+N_4)/N]}$.

Therefore, the statistics for testing $\Delta = 0$ is equivalent to test absence of association between antigens **A** and **B** in a 2x2 contingency table. Of course the usual continuity correction can be introduced in the above formula, that then takes the form

$$X^2 = [ABS(N_1N_4-N_2N_3)-N/2]^2.N/[(N_1+N_2)(N_1+N_3)(N_2+N_4)(N_3+N_4)].$$

Alternatively, a **G** test (log-likelihood ratio) can be used (and should be preferred since often numbers occurring in some cells of the table are small):

$$\begin{aligned} G \approx \chi^2 &= 4ND^2/[P_3(2-P_3)P_4(2-P_4)] \\ &= 4N\{\sqrt{(N_4/N)}-\sqrt{[(N_3+N_4)(N_2+N_4)]/N}\}^2/[(1-Q_3^2)(1-Q_4^2)] \\ &= 4N\{\sqrt{(N.N_4)}-\sqrt{[(N_3+N_4)(N_2+N_4)]}\}^2/[(N_1+N_2)(N_1+N_3)]. \end{aligned}$$

If **f(AB) = 0** then it comes out that

$$\begin{aligned} f(A+B+) &= 2qr &&= f_1 \\ f(A+B-) &= q^2 + 2qs &&= f_2 \\ f(A-B+) &= r^2 + 2rs &&= f_3 \\ f(A-B-) &= s^2 &&= f_4, \end{aligned}$$

where **q**, **r** and **s** are the frequencies of haplotypes **Ab**, **aB** and **ab**. The estimates **q**, **r**, **s** of haplotype frequencies **f(Ab)**, **f(aB)** and **f(ab)** are then obtained using the standard ABO blood group system estimation method.

GENETIC VARIABILITY AND ITS ASSESSMENT

Population genetics describes the genetical composition of populations and tries to explain its findings through grossly simplified mathematical models. The unit of measure of population genetics is the "gene" or "allele" frequency, defined as

$$p_i = P(a_ia_i) + \tfrac{1}{2} \Sigma_{j>i} P(a_ia_j) \ ,$$

a parameter with approximate binomial variance $var(p_i) = p_i(1-p_i)/2n$ (which takes place exactly when genotype proportions are in Hardy-Weinberg ratios [$P(a_ia_i) = p_i^2$, $P(a_ia_j) = 2p_ip_j$].

The variance, linearized by the square root transformation, can be used to construct approximate confidence intervals (v.g., 95% c.i.) for the "true" population frequency: $i.c.95\%\ p_i : p_i \pm 1.96 \sqrt{[var(p_i)]} = 1.96\ s.e.(p_i)$.

The mensuration of genetic variability is problematic, since organisms have **4,000 - 50,000** structural loci. After some authors, this problem can be circumvented through "random" samples, but what is a random sample of 4,000 - 50,000 loci?

Several indexes have been proposed to describe genetic variability. One of such indexes is simply the **number of alleles** that segregate in a given locus, with the obvious inconvenience that **k** (the number of detectable alleles) is proportional to **n** (the sample size) : $k \propto n$. The probability of detecting in the population a genotype that constains a rare gene is very small, as shown by the following table (adapted from Evett & Weir, 1998), that lists the required sample sizes (**N**) to detect, with a probability of 95%, genotypes with population frequencies (**P**):

P	n
1	1
0.1	30
0.01	300
0.001	3000
0.0001	30000
...	
P	3 / P

In any case, the number of alleles segregating at a given autosomal locus is an important provider of variability per se. Letting k be the number of such alleles and assuming that all alleles occur with the same frequency, $p_i = ... = p_j = 1/k$, it comes out that $P(a_ia_j) = 2p_ip_j = 2.1/k.1/k = 2/k^2$; since the number of different types of possible heterozygotes is given by $k(k-1)/2$, it follows that the probability of an individual being a heterozygote in such a population is given by the expression $P(het) = 2/k^2 \times k(k-1)/2 = (k-1)/k$. As the following table shows, the value of $(k-1)/k$ converges rapidly to 1.

k	1/k	$2/k^2$	k(k-1)/2	(k-1)/k
2	1/2	1/2	1	1/2
3	1/3	2/9	3	2/3
4	1/4	2/16	6	3/4
5	1/5	2/25	10	4/5
...	...	...	...	...
inf.	0	0	inf.	1

Another useful diversity parameter is the so-called "proportion of polymorphic loci." Polymorphic loci are arbitrarily defined as loci that contain at least two polymorphic alleles (alleles with frequency between 0.99 and 0.01, or between 0.95 and 0.05); genes with frequency larger than 0.99 (or 0.95) are known as monomorphic, in contrast with those with a frequency smaller than 0.01 (or 0.05), known as idiomorphic. The detection of polymorphisms suffers from the restraints associated with the probability of genotype detection commented above.

Another diversity parameter -- this a very important one -- is the index known alternatively as gene diversity or heterozigosity (h, H):

$$h = 1 - \Sigma p_i^2 \rightarrow 2n(1-\Sigma p_i^2)/(2n-1)$$
$$p_i = P(a_i a_i) + \tfrac{1}{2} \Sigma_{j>i} P(a_i a_j)$$
$$H = \Sigma h_j / r \ , \ var(H) = var(h)/r$$
$$var(h) = \Sigma (h_j - H)^2/(r-1)$$

The following table shows the overall results obtained with the analysis of 31 enzymatic loci in the fruit fly ***D. willistoni*** (Ayala et al. 1974) and with 11 proteic loci in the rodent ***S. douglasii*** (Smith & Coss 1984):

	No. of sampled loci	Ave. no. of alleles	proportion of polymorph. loci 5%	proportion of polymorph. loci 1%	heterozygosity H
Dw	31	5.4	14/31	24/31	0.177
Sd	11	2.8	4/11	6/11	0.045

Given the problems mentioned above, there exists a copious literature on the methodology necessary to circumvent them all. Since more than 50% of the loci in most species are monomorphic, one expects to find a large variance between loci using any variability index. This suggests the strategy of surveying a large number of loci instead of a large number of individuals in order to obtain more reliable estimates of H; but of course a reasonable number of individuals analyzed per locus makes the variance within loci smaller and the variance between loci more homogeneous. Mutation, selection, migration, and drift, on the other hand, have an opposite effect, making the variance between loci larger than within loci.

INBREEDING

1) Regular systems of inbreeding

The main effect of inbreeding is an increase in the frequency of homozygotes in the population, with a corresponding decrease in heterozygosity. When inbreeding takes place systematically and exclusively among individuals with a close degree of biological relationship, it leads to the distribution of homozygotes in the gene frequencies and therefore to a complete loss of population heterozygosity. These effects can be appreciated easily when we consider a population of plants with self-fertilization. If we define

$$P_0(AA) = d_0$$
$$P_0(Aa) = h_0$$
$$P_0(aa) = r_0$$

as being the initial frequencies of the three possible genotypes determined by a pair of autosomal alleles **A**, **a** , it comes out that in next generation

$$P_1(AA) = d_1 = d_0 + h_0/4$$
$$P_1(Aa) = h_1 = h_0/2$$
$$P_1(aa) = r_1 = r_0 + h_0/4 \ .$$

Exact general solutions in simple analytical form are easily obtained for these first-order difference equations:

$$P_n(AA) = d_n = d_0 + h_0/2 - h_0/2^{n+1} = p - h_0/2^{n+1}$$

$$P_n(Aa) = h_n = h_0/2^{n+1}$$

$$P_n(aa) = r_n = r_0 + h_0/2 - h_0/2^{n+1} = q - h_0/2^{n+1} \ .$$

The limits (as **n** tends to infinity) of the above expressions are clearly

$$P(AA) = d = d_0 + h_0/4 + h_0/8 + h_0/16 + h_0/32 + \ldots = d_0 + h_0/2 = p$$
$$P(Aa) = h = h_0 - h_0/2 - h_0/4 - h_0/8 - h_0/16 - \ldots = h_0 - h_0 = 0$$
$$P(aa) = r = r_0 + h_0/4 + h_0/8 + h_0/16 + h_0/32 + \ldots = r_0 + h_0/2 = q \ .$$

The frequencies **p** and **q** are constant quantities (therefore independent from **n**), as we show below:

$$p_1 = d_1 + h_1/2 = (d_0+h_0/4) + (h_0/2)/2 = d_0 + h_0/2 = p_0 = \ldots = p$$

and therefore

$$q_1 = 1 - p_1 = q_0 = 1 - p_0 = \ldots = q \ .$$

After a large number of generations (that is, when **n** tends to infinity), the population tends to equilibrium. The process takes place without alterations in gene frequencies and with the heterozygote frequency being halved each generation of self-fertilization.

For other systems of continued and exclusive inbreeding among close relatives (full sibs, double first cousins, quadruple second cousins and octuple third cousins) the population heterozygosity decreases after

$h_{n+2} = h_{n+1}/2 + h_n/4$
$h_{n+3} = h_{n+2}/2 + h_{n+1}/4 + h_n/8$
$h_{n+4} = h_{n+3}/2 + h_{n+2}/4 + h_{n+1}/8 + h_n/16$
$h_{n+5} = h_{n+4}/2 + h_{n+3}/4 + h_{n+2}/8 + h_{n+1}/16 + h_n/32$.

In all these systems the equilibrium frequency of heterozygotes is zero. When crossings occur exclusively among individuals with a biological relationship more distant than that presented by first cousins, the decrease in the population heterozygosity takes place very slowly, and at equilibrium the frequency of heterozygotes tends to a limit different from zero but in all instances smaller than **2pq**, the expected frequency of heterozygotes under a random mating system.

The derivation of the recursion relations shown above is quite cumbersome. In the lines below we show just the derivation of the formula for the heterozygote frequency in a system of matings exclusively among full sibs.

Six different types of matings occur in any population, if we are considering an autosomal locus with two alleles:

a) **AA x AA** matings, whose only progeny is of type **AA**;
b) **AA x Aa** matings, that yield progeny **AA + Aa (1:1)**;
c) **AA x aa** matings, whose only progeny is of type **Aa**;
d) **Aa x Aa** matings, that yield the three possible genotyypes **AA + Aa + aa** in the proportions **1:2:1**;
e) **Aa x aa** matings, that yield progeny **Aa + aa (1:1)**;
f) **aa x aa** matings, whose only progeny is of type **aa**.

If matings are permitted to occur just within sibships, it is not difficult to determine the recursion relations between the matings in two successive generations, using the table shown below:

Matings (n)	Sibships (n+1)	Matings (n+1)	Frequencies
AA x AA	AA (1)	AA x AA	1
AA x Aa	AA + Aa (1:1)	AA x AA	1/4
		AA x Aa	1/2
		Aa x Aa	1/4
AA x aa	Aa (1)	Aa x Aa	1
Aa x Aa	AA + Aa + aa (1:2:1)	AA x AA	1/16
		AA x Aa	1/4
		AA x aa	1/8
		Aa x Aa	1/4
		Aa x aa	1/4
		aa x aa	1/16
Aa x aa	Aa + aa (1:1)	Aa x Aa	1/4
		Aa x aa	1/2
		aa x aa	1/4
aa x aa	aa (1)	aa x aa	1

If we call u_n, v_n, w_n, x_n, y_n, z_n the respective frequencies of **AA x AA**, **AA x Aa**, **AA x aa**, **Aa x Aa**, **Aa x aa**, and **aa x aa** matings in generation n, inspection of the above table shows clearly that

$$u_{n+1} = u_n + v_n/4 + x_n/16$$
$$vn+1 = v_n/2 + x_n/4$$
$$w_{n+1} = x_n/8$$
$$x_{n+1} = v_n/4 + w_n + x_n/4 + y_n/4$$
$$y_{n+1} = x_n/4 + y_n/2$$
$$z_{n+1} = x_n/16 + y_n/4 + z_n \ ,$$

or, in matrix form,

$$\begin{pmatrix} u_{n+1} \\ v_{n+1} \\ w_{n+1} \\ x_{n+1} \\ y_{n+1} \\ z_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & 1/4 & 0 & 1/16 & 0 & 0 \\ 0 & 1/2 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/8 & 0 & 0 \\ 0 & 1/4 & 1 & 1/4 & 1/4 & 0 \\ 0 & 0 & 0 & 1/4 & 1/2 & 0 \\ 0 & 0 & 0 & 1/16 & 1/4 & 1 \end{pmatrix} . \begin{pmatrix} u_n \\ v_n \\ w_n \\ x_n \\ y_n \\ z_n \end{pmatrix}$$

The frequency of heterozygotes in generation **n+1** is obviously

$$h_{n+1} = v_n/2 + w_n + x_n/2 + y_n/2 \ ;$$

and in generations **n+2** and **n+3** ,

$$h_{n+2} = v_{n+1}/2 + w_{n+1} + x_{n+1}/2 + y_{n+1}/2$$
$$= 3v_n/8 + w_n/2 + x_n/2 + 3y_n/8$$

and

$$h_{n+3} = v_{n+2}/2 + w_{n+2} + x_{n+2}/2 + y_{n+2}/2$$
$$= 3v_{n+1}/8 + w_{n+1}/2 + x_{n+1}/2 + 3y_{n+1}/8 =$$
$$= 5v_n/16 + w_n/2 + 3x_n/8 + 5y_n/16 \ ,$$

respectively. Comparing the expressions above we get immediately

$$h_{n+3} = h_{n+2}/2 + h_{n+1}/4 \ .$$

Therefore the recursion equation for the frequency of heterozygotes is

$$h_{n+2} = h_{n+1}/2 + h_n/4.$$

For large values of **n** the heterozygosity of the population in a given generation is **80.9%** of that of the previous generation (in contrast with the rate of **50%** for self-fertilization systems). We obtain this value dividing both sides of the recursion equation

$h_n = h_{n-1}/2 + h_{n-2}/4$ by h_{n-1} ;

we get then

$$h_n/h_{n-1} = 1/2 + h_{n-2}/4h_{n-1} \ ;$$

calling **r** the limit of h_n/h_{n-1} as **n** tends to infinity, it comes out that, for sufficiently large values of **n**,

$r = 1/2 + 1/4r$ or $4r^2 - 2r - 1 = 0$.

The positive root of the above quadratic equation (which is the characteristic equation of the recurrence equation $h_{n+2} - h_{n+1}/2 - h_n/4 = 0$) is

$r = (1+\sqrt{5})/4 = 0.809$.

The other possible solution of the above equation,

$r' = (1-\sqrt{5})/4 = -0.309$,

is non admissible, since **r** is the limit of h_n/h_{n-1} as **n** tends to infinity and h_n is equal to or greater than zero for any value **n** might take; **r** would therefore never have a negative sign.

A numerical example of what happens to the frequency of heterozygotes and to the ratio h_n/h_{n-1} in a population where matings occur only between sibs is shown below (followed by the BASIC code used for generating the table values), taking $h_0 = 1$ and $h_1 = 0.5$ as initial conditions:

n	h_n	h_n/h_{n-1}
0	1.00000	-
1	0.50000	0.50000
2	0.50000	1.00000
3	0.37500	0.75000
4	0.31250	0.83333
5	0.25000	0.80000
6	0.20313	0.81250
7	0.16406	0.80769
8	0.13281	0.80952
9	0.10742	0.80882
10	0.08691	0.80909
11	0.07031	0.80899
12	0.05688	0.80903
13	0.04602	0.80901
14	0.03723	0.80902
15	0.03012	0.80902
16	0.02437	0.80902
17	0.01971	0.80902
18	0.01595	0.80902
19	0.01290	0.80902
20	0.01044	0.80902
...	.....	.....
inf.	0.00000	0.80902

```
REM PROGRAM FILENAME INBREE02.BAS
REM LIMIT OF HN+1/HN IN A SIB MATING SYSTEM
DEFDBL A-Z: CLS : DIM H(20): H(0) = 1: H(1) = .5
PRINT USING "###   "; 0; : PRINT USING "#.#####   "; H(I); : PRINT "   -"
I = 1: GOSUB PRINTOUT
FOR I = 2 TO 20
 H(I) = H(I - 1) / 2 + H(I - 2) / 4
GOSUB PRINTOUT: NEXT I
PRINT "inf.  "; : PRINT USING "#.#####   "; 0; :
PRINT USING "#.#####   "; (1 + SQR(5)) / 4: END
```

```
PRINTOUT:
PRINT USING "###   "; I; : PRINT USING "#.#####   "; H(I); H(I) / H(I - 1)
RETURN
```

The recurrence relation $h_n = h_{n-1}/2 + h_{n-2}/4$ is linear and admits therefore an exact general solution in simple analytical form. This general solution has the form

$$h_n = C_1.r_1^n + C_2.r_2^n ,$$

where

$$r_1 = (1+\sqrt{5})/4 = 0.809$$

$$r_2 = (1-\sqrt{5})/4 = -0.309$$

$$C_1 = (h_1-h_0.r_2)/(r_1-r_2)$$

$$C_2 = (h_0.r_1-h_1)/(r_1-r_2).$$

Since in modulus both r_1 and r_2 are less than unity, it comes out that for large values of n r_1^n and r_2^n tend to zero, and therefore at equilibrium $h = 0$. Since in modulus r_1 is greater than r_2, r_2^n approaches zero faster than r_1^n; consequently, for large n,

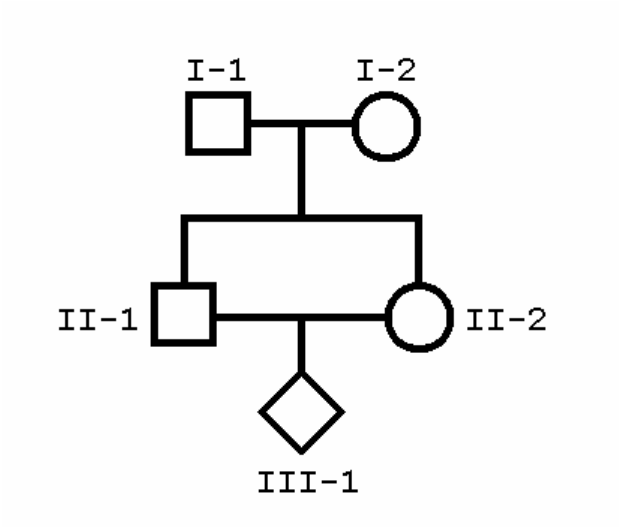
$h_n = C_1.r_1^n$ approximately;

therefore the limit of expression h_n/h_{n-1} as n tends to infinity is really $r_1 = 0.809$:

$$r_1 = \lim_{n\to\infty}(h_n/h_{n-1}) = (C_1/C_1).r_1^n/r_1^{n-1} = r_1.$$

2) Probability of homozygosis for the offspring of consanguineous parents and of sharing of identical genes by two relatives (derivation of the coefficients of inbreeding and common identity)

In panmictic populations, the frequencies of **AA**, **Aa**, **aa** individuals (where **A** and **a** are two alleles segregating at an autosomal locus) are respectively p^2, $2pq$, q^2. These are therefore the probabilities for a child born to unrelated parents of having a genotype respectively **AA**, **Aa**, **aa**. We shall determine now what are the corresponding probabilities for **AA**, **Aa**, **aa** individuals born to relatives. In order to achieve that, we shall consider, for the sake of simplicity, the offspring of a sib mating:



The child (**III-1**) of the above brother (**II-1**) - sister (**II-2**) mating can be homozygous as to the alleles of a given locus by receiving each one from each one of the two different grandparents (homozygosis by independent union of gametes); however, he can also be homozygous by receiving, through both parents (**II-1** and **II-2**), the same gene present in the grandparent **I-1** or **1-2**. This second type of homozygosis is called homozygosis by common descent or autozygosis (this last term having been coined by Cotterman).

In relation to the alleles of a given locus, the probability of autozygosis is, for children born to sibs, **1/4**. In fact, the grandparent **I-1** and the grandparent **I-2** present, each one, two alleles at the locus (total of 4 genes) and the probability of autozygosis for individual **III-I** is **1/16** for each one of these genes. The figure of 1/4 is obtained multiplying 1/16 times 4 : **F = 4 x 1/16 = 1/4**.

The probability of **III-1** being autozygous **AA** is **p/4** and that of being autozygous **aa** is **q/4**. In fact, grandparent **I-1** (the same reasoning is valid for grandparent **I-2**) can be **AA**, **Aa** or **aa** with probabilities p^2, **2pq** and q^2, respectively. If the individual **I-1** were **AA** (let us denote his genotype by A_1A_2 in order to differentiate between the gene A_1 received from his father and the gene A_2 received from his mother), the probability that **III-1** is born A_1A_1 is **1/16** and that he is A_2A_2 is also **1/16**; therefore the probability of **III-1** being autozygous **AA** given that the grandparent **I-1** is **AA** is **1/8**. If the grandparent **I-1** is **Aa**, the probabilities of the child **III-1** being born **AA** and **aa** have each the same value of **1/16** (this value of **1/16** arises of course from **1/2 x 1/2 x 1/2 x 1/2**, that is the probability of any gene being transmitted by **I-1** to both **II-1** and **II-2**, and from these to **III-1**). If the grandparent **I-1** is **aa**, the probability of **III-1** being autozygote for any one of these two genes is also **1/8** (as in the case of the **A** allele). Since the probabilities of **I-1** being **AA**, **Aa** or **aa** are p^2, **2pq** and q^2, it comes out that **III-1** has a probability $p^2/8 + 2pq/16 = p/8$ of being autozygote **AA** and a probability $q^2/8 + 2pq/16 = q/8$ of being autozygous **aa**.

Since the grandchild **III-1** can be autozygous **AA** or **aa** by receiving the alleles from the grandparent **I-2**, who has the same probabilities p^2, **2pq** and q^2 of being **AA**, **Aa** or **aa**, it is clear that the probabilities for a

child born to sibs to be an autozygote **AA** or an autozygote **aa** are respectively **p/4** and **q/4**.

We have just verified that the probability of a given locus of the offspring of sibs being autozygous is **p/4 + q/4 = 1/4**. There is, therefore, a probability of **3/4** of not being so; that is, in 3 out of 4 times the locus shall have a constitution **AA** or **aa** (homozygosis by independent union of gametes) or **Aa** (heterozygosis). Therefore the probabilities associated with the genotypes **AA**, **Aa** and **aa** by independent union of gametes are $\mathbf{3p^2/4}$, **3pq/2**, $\mathbf{3q^2/4}$.

Of course we can get these figures using the same reasoning shown some paragraphs above for calculating the chances of autozygosity. In order to differentiate between all genes present in grandparents **I-1** and **I-2**, let $\mathbf{A_1A_2}$, $\mathbf{A_3a_1}$, $\mathbf{a_2a_3}$ be the possible genotypes of individual **I-1**; and $\mathbf{A_4A_5}$, $\mathbf{A_6a_4}$, $\mathbf{a_5a_6}$ the corresponding ones of individual **I-2**.

Individual **III-1** has therefore the following probability of being homozygote by independent union of gametes:

$$\mathbf{P(III\text{-}1 = A_iA_j)}$$
$$\mathbf{= P(A_1A_2) + P(A_4A_5) + P(A_1A_4) + P(A_1A_5) + P(A_1A_6) + P(A_2A_4)}$$
$$\mathbf{+ P(A_2A_5) + P(A_2A_6) + P(A_3A_4) + P(A_3A_5) + P(A_3A_6)}$$

$$= 2.p^2/16 + 2.p^2/16 + 2.p^4/16 + + 2.p^4/16$$
$$+ 2.2.p^3.q/16 + 2.p^4/16 + 2.p^4/16 + 2.2.p^3.q/16$$
$$+ 2.2.p^3.q/16 + 2.2.p^3.q/16 + 2.4.p^2.q^2/16$$
$$= p^2/4 + p^4/2 + p^3.q + p^2.q^2/2$$
$$= p^2/4 + p^2(p^2 + 2pq + q^2)/2$$
$$= p^2/4 + P^2/2 = \mathbf{3.p^2/4}\ .$$

The probabilities shown above are of trivial determination. The only point that deserves some explanation is the factor 2 that multiplies each one of the partial expressions: it arises from the fact that individual **III-1** can be of genotype $\mathbf{A_iA_j}$ through two different paths: the gene $\mathbf{A_i}$ passes through individual **II-1** and the gene $\mathbf{A_j}$ through **II-2** or vice-versa.

By symmetry, the probability of **III-1** being a homozygote aa by distinct origins is

$$\mathbf{P(III\text{-}1 = a_ia_j) = 3.q^2/4}\ .$$

Subbtracting $3.p^2/4 + 3.q^2/4$ from 3/4 we obtain finally the probability of the individual III-1 being heterozygous :

$$\mathbf{P(III\text{-}1 = Aa) = 3.pq/2}\ .$$

The probability of autozygosis for any inbred individual, which takes the value of 1/4 when the parents are brother and sister, is the so-called **coefficient of inbreeding F**. Below we list some values of **F**, given the biological relationship between parents :

Mating	**F**
self-fertilization	**1/2**
parent-child	**1/4**
brother-sister	**1/4**

uncle-niece	**1/8**
double first-cousins	**1/8**
first cousins	**1/16**
first cousins once removed	**1/32**
second cousins	**1/64**
second cousins once removed	**1/128**

Generalizing the situation for any degree of biological relationship between the parents, the probabilities of an inbred individual being **AA**, **Aa**, and **aa** are respectively

$$P(AA) = pF + (1-F)p^2$$
$$P(Aa) = 2(1-F)pq$$
$$P(aa) = qF + (1-F)q^2 \ .$$

The factors **F** and **1-F** in the above formulae can be understood as the partition, among homozygotes, of autozygosis and homozygosis by independent union of gametes or allozygosis (another useful term coined by Cotterman).

The above formulae can be easily rearranged as

$$P(AA) = pF + (1-F)p^2 = p^2 + Fp - Fp^2 = p^2 + Fp(1-p)$$
$$= p^2 + Fpq$$
$$P(Aa) = 2pq - 2Fpq$$
$$P(aa) = q^2 + Fpq \ .$$

These latter representations are useful since they show directly the excess of homozygosis (or alternatively the decrease in the frequency of heterozygotes) in relation to the one existing in panmixia: this has a value of **Fpq**.

We have just defined a useful parameter in population genetics - the coefficient of inbreeding. It must be stressed (again) that this parameter is the probability of autozygosis of a given locus for an inbred individual. Of course it can be understood also as the fraction of genes of an inbred individual that are autozygous: for example, a child born to a couple of first cousins has an inbreeding coefficient of **F = 1/16**; this means that 1/16 of all his or her genes are in autozygotic state.

Another useful inbreeding parameter is the coefficient of common identity **R**. It represents the probability of one randomly chosen gene from one individual being identical by descent to a gene in a second person (if this second person is not biologically related to the individual this probability is zero). Of course this probability means also exactly the total amount of genes which are shared by two related individuals. In the literature this coefficient is sometimes called the coefficient of relationship, but we shall use the name "coefficient of common identity" in order to avoid a confusion that still persists in the specialized literature about the probabilistic meaning of the coefficient of relationship. Perhaps the name "coefficient of relationship" should be used only in the exact and restricted meaning Wright associated with it, that is the (zygotic) coefficient of genetic correlation between two individuals, which can be determined, for example, by the application of rules of path coefficients.

The table below shows some values of **R**, together with the corresponding **F** values for the children of individuals shown in table. It is quite obvious that **R = 2F** for any numerical value.

Relatives	**R**	**F**
Parent-child	**1/2**	**1/4**
Brother-sister	**1/2**	**1/4**
Uncle-niece	**1/4**	**1/8**
Double first cousins	**1/4**	**1/8**
First cousins	**1/8**	**1/16**
1st cousins once rem.	**1/16**	**1/32**
Second cousins	**1/32**	**1/64**
2nd cousins once rem.	**1/64**	**1/128**

The identity **R = 2F** arises from the following: let us choose, in the first individual, a given locus : it hosts, for example, alleles A_i and A_j. The probability that the relative of this individual has the allele A_i is by definition the coefficient of common identity **R**. The same figure (**R**) is true also for the allele $\mathbf{A_j}$. Therefore the chance that a child born to this couple of individuals is $\mathbf{A_iA_i}$ or $\mathbf{A_jA_j}$ (probability of autozygosis or **F**) is

$$\mathbf{F = P(A_iA_i) + P(A_jA_j) = 1.R.1/4 + 1.R.1/4 = R/2 .}$$

3) Applications of F and R in situations of genetic counseling

The probability that an inbred individual has the genotype **aa** is

$$\mathbf{P(aa|F>0) = q^2 + Fpq} \text{ ;}$$

since the frequency of the **aa** genotype among non-inbred individuals is

$$\mathbf{P(aa|F=0) = q^2} \text{ ,}$$

we may define a new parameter, that we shall call relative risk, as the ratio of the two proportions shown above:

$$\mathbf{RR = P(aa|F>0)/P(aa|F=0) = 1 + Fp/q = 1 + F(1-q)/q} \text{ .}$$

For the case of phenylketonuria, for example, **q = 0.008** (since the frequency of affected children, in the offspring of non-consanguineous spouses is $\mathbf{q^2 = 1/15,000}$); if **F = 1/16** (offspring of first cousins), the risk is

$$\mathbf{RR = 8.5} \text{ ,}$$

that is, the frequency of children affected by phenylketonuria is 8.5 times greater among children born to first cousins than to children of unrelated parents.

Frequently consanguineous couples seek genetic counsel in order to learn the risks for their offspring. In the case of first cousins with no record of genetic diseases in their families, the following reasoning can

be used : if all autosomal recessive diseases were produced by pathological alleles with an average frequency of **0.01** (this figure is of course imprecise but is also reasonable), the value of **RR** for any of these diseases should be about **7**. The frequency of recessive diseases at birth among non-inbred children can be estimated roughly in **0.01**. Therefore we deduce that the probability of a child born to a couple of first cousins spouses being affected by any recessive disorder is about **7%**. Other types of diseases (that is, non-recessive conditions) affect children belonging to both groups (consanguineous and non-consanguineous couples) with the same chance, and account for a proportion of about 2%. Therefore the risks for any disease present at birth are of 3% and 9%, respectively for children born to non-consanguineous and to first cousin couples.

One must keep in mind that the above estimates refer only to physical defects and do not include mental retardation. The frequency of this conditionn in the general population has been estimated to be about 1%; a 3 to 4-fold increase of this frequency was observed among children born to first cousin relatives. Including these figures in the risk estimates shown at the end of the last paragraph, we obtain risks of 4% for children of unrelated couples and of 13% for the offspring of first cousin unions.

The table below shows similar risk estimates for children born to several types of consanguineous couples. R_2 is the risk estimate that includes mental retardation and this is the one that should be used for genetic counseling purposes.

Marriage	R_1	R_2
brother-sister	**0.280**	**0.400**
uncle-niece	**0.140**	**0.220**
first cousins	**0.090**	**0.130**
lst cousins cnce removed	**0.060**	**0.085**
second cousins	**0.045**	**0.060**
unrelated persons	**0.030**	**0.040**

Of course the above method contains several simplified assumptions, but it is important because it enables one to calculate genetic risks as a function of **F**.

The coefficient of common identity can also be used in situations of genetic counseling, as we show below.

If we had an estimate of the average number of recessive pathological genes per individual, we could calculate easily the offspring risks. For example, let us suppose that on average each individual has one pathologic recessive gene in heterozygous state. Since the coefficient of common identity has a value **R = 1/8** for first cousins, the risk for their offspring could then be evaluated as being

$$P(aa) = R/4 = 1/32 \approx 3\% ,$$

since 1 is the probability of the first individual having the pathologic gene (we have just stated hypothetically that on average each individual has one pathologic recessive gene in heterozygous state), R is the

probability of this same gene being present in his or her cousin and 1/4 is the compound probability of two heterozygous partners transmitting the same allele to their offspring).

Unfortunately we do not have such estimates. We do have estimates of the average number of lethal equivalents and often in the literature this estimate has been confounded with an estimate of the average number of deleterious genes per person and then used unappropriately (in the manner we have just shown) in genetic counseling of consanguineous couples.

The coefficient of common identity **R** has however some useful applications in the genetic counseling of consanguineous couples, in the case of recorded diseases occurring in the family of the couple. Let us consider, for example, the following case: one albino (and oculo-cutaneous albinism is known to be an autosomal recessive disorder) and his normally pigmented cousin want to know the risk that a child they intend to have will be affected by the disease. Since **R** can be interpreted as the probability of a gene at a given locus in one person being identical by descent at the same locus in a second person and, given that the albino has a genotype $\mathbf{a_1a_2}$ (the subscripts are just to differentiate between the two genes), the probability that his cousin has one of the genes ($\mathbf{a_1}$ or $\mathbf{a_2}$) is **2R = 1/4**. So the risk for a child born to the couple of affection by the disease is **1 x 1/4 x 1/2** (these figures represent respectively the probabilities of the albino transmitting the recessive gene, of the woman being a heterozygote for this gene and, if a heterozygote, of transmitting the same gene). The final figure is **P(aa) = 1/8** or **12.5%**.

4) Average inbreeding coefficient of the population

The average inbreeding coefficient **f** of a population can be understood as the mean value of **F** in a given population:

$$\mathbf{f = \Sigma x_i F_i},$$

where $\mathbf{x_i}$ is the frequency of the class $\mathbf{F_i}$.

For example, let us suppose that in a given population 1000 couples have been randomly sampled, 952 of which were non-consanguineous ones; 32 couples were first degree cousins; and 16 were uncle-niece unions. This situation is summarized in the following table:

F_i	N_i	x_i	x_iF_i
0	952	0.952	0.000
1/16	32	0.032	0.002
1/8	16	0.016	0.002
-	1000	1.000	0.004

It is generally impossible to estimate directly the value of **f** from a population through the determination of the deviations of genotype frequencies from Hardy-Weinberg proportions, because consanguineous marriages occur with a very low frequency in most human populations. However we have the simple and practical method just shown, for which one just need to ascertain the frequencies, in the population, of the different classes of consanguineous matings.

If in a population the frequencies of the different classes of consanguineous matings remain constant from generation to generation

(i.e., if there exists in the population a regular system of inbreeding) the **f** value of the population tends to a constant value and the population is then said to be in an equilibrium state. If consanguineous marriages take place at a low rate as in the numerical example above, the equilibrium inbreeding coefficient will not differ significantly from the average inbreeding coefficient.

Taking as a first example a system of exclusive and continued self-fertilization, the chance of any individual being an autozygote after one generation is

$$f_1 = 1/2 \ ;$$

after two generations, f_n takes the value

$$f_2 = f_1 + (1-f_1)/2 = 3/4 \ ;$$

after three generations,

$$f_3 = f_2 + (1-f_2)/2 = 7/8 \ ;$$

the recursion relation is clearly

$$f_{n+1} = (f_n + 1)/2 \ .$$

Subtracting from the quantity **1** both sides of the equation above we get

$$1 - f_{n+1} = 1 - (f_n + 1)/2 = (1 - f_n).(1/2),$$

which general solution is given by

$$f_n = 1 - (1 - f_0).(1/2)^n.$$

The limiting value f_n takes is clearly **1**, as **n** tends to infinity. This means that after a great number of generations, the population tends to complete autozygosity. A numerical example (with appended BASIC code) is shown in the table below, where initial frequencies of **0.36**, **0.48** and **0.16** have been assumed for the genotypes **AA**, **Aa** and **aa**.

n	d_n	r_n	h_n	p_n	f_n
0	0.36000	0.48000	0.16000	0.60000	0.00000
1	0.48000	0.24000	0.28000	0.60000	0.50000
2	0.54000	0.12000	0.34000	0.60000	0.75000
3	0.57000	0.06000	0.37000	0.60000	0.87500
4	0.58500	0.03000	0.38500	0.60000	0.93750
5	0.59250	0.01500	0.39250	0.60000	0.96875
6	0.59625	0.00750	0.39625	0.60000	0.98438
7	0.59813	0.00375	0.39812	0.60000	0.99219
8	0.59906	0.00187	0.39906	0.60000	0.99609
9	0.59953	0.00094	0.39953	0.60000	0.99805
10	0.59977	0.00047	0.39977	0.60000	0.99902
11	0.59988	0.00023	0.39988	0.60000	0.99951
12	0.59994	0.00012	0.39994	0.60000	0.99976
13	0.59997	0.00006	0.39997	0.60000	0.99988
14	0.59999	0.00003	0.39999	0.60000	0.99994
15	0.59999	0.00001	0.39999	0.60000	0.99997

```
16   0.60000   0.00001   0.40000   0.60000   0.99998
17   0.60000   0.00000   0.40000   0.60000   0.99999
18   0.60000   0.00000   0.40000   0.60000   1.00000
19   0.60000   0.00000   0.40000   0.60000   1.00000
20   0.60000   0.00000   0.40000   0.60000   1.00000
-------------------------------------------------
```

```
REM PROGRAM FILENAME INBREE03.BAS
REM SELF-FERTILIZATION
DEFDBL A-Z: CLS : DIM D(20), H(20), R(20), P(20), F(20)
P = .6: Q = 1 - P: D(0) = P * P: H(0) = 2 * P * Q: R(0) = Q * Q: F(0) = 0
FOR I = 0 TO 20
 D(I) = P - H(0) / 2 ^ (I + 1): H(I) = H(0) / 2 ^ I
 R(I) = Q - H(0) / 2 ^ (I + 1)
 P(I) = D(I) + H(I) / 2: F(I) = 1 - H(I) / H(0)
 PRINT USING "###   "; I;
 PRINT USING "#.#####   "; D(I); H(I); R(I); P(I); F(I)
NEXT I
```

If we begin with a random-mating population, after one generation the frequencies of autozygous and allozygous AA homozygotes are given respectively by

$P'_1(AA) = p^2/2 + 2pq/4 = p/2$ and $P''_1(AA) = p^2/2$,
so that
$P_1(AA) = P'_1(AA) + P''_1(AA) = p/2 + p^2/2$;

and, in the second generation, by

$P'_2(AA) = p/2 + p^2/4 + 2pq/8 = p/2 + p/4 = 3p/4$ and $P''_2(AA) = p^2/4$,
so that
$P_2(AA) = P'_2(AA) + P''_2(AA) = 3p/4 + p^2/4$;

since in generations 0, 1, and 2 the values f_n takes are respectively 0, 1/2, and 3/4, it is easy to see that the equation above can be written as

$$P_n(AA) = P'_n(AA) + P''_n(AA) = f_n.p + (1-f_n).p^2 \ ;$$

evidently,

$$P_n(aa) = P'_n(aa) + P''_n(aa) = f_n.q + (1-f_n).q^2$$

and

$$P_n(Aa) = 2.(1-f_n).pq \ .$$

At equilibrium, **f** tends to a constant value (that is **1** in a self-fertilizing population) and the above equations become

$$P(AA) = f.p + (1-f).p^2$$
$$P(Aa) = 2.(1-f).pq$$
$$P(aa) = f.q + (1-f).q^2 \ .$$

Let us now consider as a second example a system of admixture of self-fertilization and random matings. Let **x** (for example **0.40**) be the fraction of the population that reproduces through self-fertilization and **1-x = 0.60** the fraction that reproduces sexually through panmixia. **x** can be interpreted also, in the above formulation, as being the constant probability that an individual, chosen at random from the population,

reproduces by self-fertilization. For the sake of simplicity (and also for comparing the results), let us consider the same initial population composition seen in the previous example :

$P_0(AA) = d_0 = 0.36$
$P_0(Aa) = h_0 = 0.48$
$P_0(aa) = r_0 = 0.16$.

The frequencies of genotypes **AA**, **Aa** and **aa** among individuals resulting from self-fertilization (fraction **x = 0.40** of the population) shall be, in the first generation,

$P'_1(AA) = d'_1 = d_0 + h_0/4 = 0.48$
$P'_1(Aa) = h'_1 = h_0/2 = 0.24$
$P'_1(aa) = r'_1 = r_0 + h_0/4 = 0.28$

and the frequencies of the same genotypes among the individuals resulting from random matings (fraction **1-x = O.60** of the population) shall be

$P"_1(AA) = d"_1 = (d_0 + h_0/2)^2 = p^2 = 0.36$
$P"_1(Aa) = h"_1 = 2(d_0 + h_0/2)(r_0 + h_0/2) = 2pq = 0.48$
$P"_1(aa) = r"_1 = (r_0 + h_0/2)^2 = q^2 = 0.16$

Therefore, the frequencies of **AA**, **Aa** and **aa** individuals in the population considered as a whole shall be, after one generation,

$P_1(AA) = d_1 = d'_1.x + d"_1.(1-x) = 0.408$
$P_1(Aa) = h_1 = h'_1.x + h"_1.(1-x) = 0.384$
$P_1(aa) = r_1 = r'_1.x + r"_1.(1-x) = 0.208$.

The frequencies of alleles **A** and **a**, evidently, remained the same:

$p1 = d1 + h1/2$
$= (d_0 + h_0/4).x + p_0^2.(1-x) + h_0.x/4 + 2p_0q_0.(1-x)/2$
$= (d_0 + h_0/2).x + (p_0^2 + p_0q_0).(1-x)$
$= p_0.x + p_0.(1-x) = p_0$

By applying the same recursion relations once more, we get at the second generation :

$P_2(AA) = d_2 = d'_2.x + d"_2.(1-x) = 0.4176$
$P_2(Aa) = h_2 = h'_2.x + h"_2.(1-x) = 0.3648$
$P_2(aa) = r_2 = r'_2.x + r"_2.(1-x) = 0.2176$

and so on.

In any generation, the frequencies of the genotypes **AA**, **Aa** and **aa** can be represented as

$P_n(AA) = d_n = p.f_n + p^2.(1-f_n)$,
$P_n(Aa) = h_n = 2pq.(1-f_n)$,
$P_n(aa) = r_n = q.f_n + q^2.(1-f_n)$,

in which f_n is the average inbreeding coefficient of the population in generation **n**. It is easy to show that, as **n** increases f_n tends to an equilibrium value situated between 1 and 0. In the present example, this value is of trivial determination: since at equilibrium

$h_{n+1} = h_n = h,$

if we take out the subscripts from the equation

$h_{n+1} = h_n.x/2 + 2pq.(1-x)$

we obtain immediately

$h = 4pq(1-x)/(2-x)$;

if we substitute this value in

$f = (2pq-h)/2pq = 1 - h/2pq$

we get straightforwardly

$f = x/(2-x)$.

This last result can be obtained direcly from the recursion equation for the frequency of autozygous homozygotes (f_n):
since $f_{n+1} = f_n.x + (1-f_n).x/2,$
if we put $f_{n+1} = f_n = f,$
it comes out that
$f = f.x + (1-f).x/2 = f.x/2 + x/2$ and $f = x/(2-x).$

There is a general exact solution in simple analytical form for the recurrence equation

$h_{n+1} = h_n.x/2 + 2pq.(1-x)$

and its determination is as follows : since $h = 4pq(1-x)/(2-x),$ it comes out that

$2pq(1-x) = h(2-x)/2 = h - h.x/2$;

substituting this in the above recurrence equation and subtracting from both sides of it the constant quantity $h,$ we obtain

$$h_{n+1}-h = h_n.x/2 + h - h.x/2 - h$$
$$= (h_n-h).x/2$$

and therefore

$h_n - h = (h_0 - h).(x/2)^n$

and

$h_n = h + (h_0 - h).(x/2)^n.$

In the numerical example the equilibrium values of h_n and f_n are $h = 0.36$ and $f = 0.25.$

The constancy of f corresponds to equilibrium in which the genotypic frequencies are

$P(AA) = d = pf + p^2.(1-f)$
$P(Aa) = h = 2pq.(1-f)$
$P(aa) = r = qf + q^2.(1-f)$.

In fact, for **p = 0.60, q =0.40, f = 0.25, x = 0.40** and **1-x = 0.60** we have

P_n(AA) = d_n = O.60 x 0.25 + 0.36 x 0.75 = 0.42
P_n(Aa) = h_n = 2 x 0.40 x O.60 x 0.75 = 0.36
P_n(aa) = r_n = O.40 x 0.25 + 0.16 x 0.75 = 0.22

and

P_{n+1}(AA) = d_{n+1} = O.40 x (0.42 + 0.09) + 0.60 x 0.36 = 0.42 = d_n
P_{n+1}(Aa) = h_{n+1} = 0.40 x O.18 + 0.60 x 0.48 = 0.36 = h_n
P_{n+1}(aa) = r_{n+1} = O.40 x (0.22 + 0.09) + 0.60 x 0.16 = 0.22 = r_n

The table below (generated by the appended BASIC code) shows the numerical values of **d_n, h_n, r_n, p_n,** and **f_n** for several generations:

n	d_n	h_n	r_n	p_n	f_n
0	0.360000	0.480000	0.160000	0.600000	0.000000
1	0.408000	0.384000	0.208000	0.600000	0.200000
2	0.417600	0.364800	0.217600	0.600000	0.240000
3	0.419520	0.360960	0.219520	0.600000	0.248000
4	0.419904	0.360192	0.219904	0.600000	0.249600
5	0.419981	0.360038	0.219981	0.600000	0.249920
6	0.419996	0.360008	0.219996	0.600000	0.249984
7	0.419999	0.360002	0.219999	0.600000	0.249997
8	0.420000	0.360000	0.220000	0.600000	0.249999
9	0.420000	0.360000	0.220000	0.600000	0.250000
10	0.420000	0.360000	0.220000	0.600000	0.250000

```
REM PROGRAM FILENAME INBREE01.BAS
REM ADMIXTURE OF SELF-FERTILIZATION AND PANMIXIA
DEFDBL A-Z: CLS
X = .4: P = .6: Q = 1 - P: F(0) = 0: P(0) = P
D(0) = P * P: H(0) = 2 * P * Q: R(0) = Q * Q
I = 0: GOSUB PRINTOUT
FOR I = 1 TO 10
 D(I) = (D(I - 1) + H(I - 1) / 4) * X + P * P * (1 - X)
 R(I) = (R(I - 1) + H(I - 1) / 4) * X + Q * Q * (1 - X)
 H(I) = H(I - 1) / 2 * X + 2 * P * Q * (1 - X)
 P(I) = D(I) + H(I) / 2: F(I) = 1 - H(I) / (2 * P * Q)
GOSUB PRINTOUT: NEXT I: END
PRINTOUT:
PRINT USING "###   "; I;
PRINT USING "#.######   "; D(I); H(I); R(I); P(I); F(I)
RETURN
```

The equilibrium

P(AA) = d = pf + p^2.(1-f)
P(Aa) = h = 2pq.(1-f)
P(aa) = r = qf + q^2.(1-f)

can be attained through various regular or irregular inbreeding mating systems, with or without admixture to panmixia.

This equilibrium is known as Wright's equilibrium; Hardy-Weinberg equilibrium,

$d = p^2$
$h = 2pq$
$r = q^2$

can be considered a special case of Wright's equilibrium for **f = 0.**

We should also note that Wright's equilibrium can be written in three algebraically equivalent, different manners :

	1	2	3
P(AA)	$d = p^2 + fpq$	$= p^2.(1-f) + pf$	$= p - (1-f).pq$
P(Aa)	$h = 2pq - 2fpq$	$= 2pq.(1-f) + 0$	$= 0 + 2(1-f).pq$
P(aa)	$r = q^2 + fpq$	$= q^2.(1-f) + qf$	$= q - (1-f).pq$

The algebraic transformations used to obtain the various formulations are quite obvious; in fact, since **p+q = 1**, it comes out that the expression **$pf + p^2.(1-f)$**, for example, is equal to both **$p^2 + fpq$** and **p - (1-f).pq** :

$pf + p^2.(1-f) = pf + p^2 - p^2.f = p^2 + p(1-p)f = p^2 + fpq$

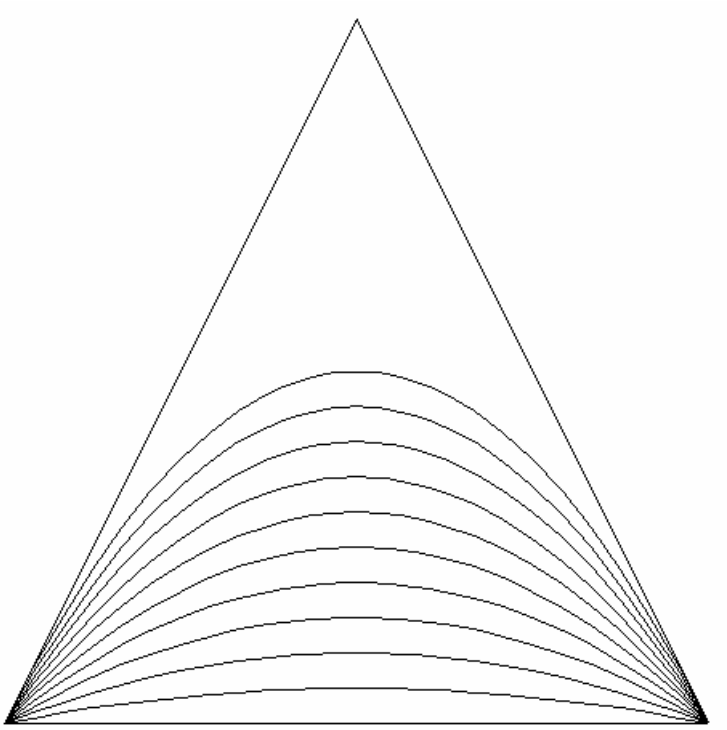
and

$p2 + fpq = p(1-q) + fpq = p - pq + fpq = p - (1-f)pq$.

The different formulations of Wright's equilibrium measure (see table above): under (1), the deviation from panmixia; under (2), the panmictic and fixed components of the population (for this reason the average inbreeding coefficient **f** is also know as the population fixation index or coefficient and **1-f** as the population panmictic index or coefficient); under (3), the deviation from complete fixation.

The Mathematica graph below shows, in a triangular diagram, the curves that represent the sets of equilibrium points **{d[←] = $pf + p^2.(1-f)$, h[↓] = $2pq.(1-f)$, r[→] = $qf + q^2.(1-f)$}** for inbred populations with **f = 0** (Hardy-Weinberg parabola), **f = 0.1, f = 0.2, ..., f = 1.0** (fixed population represented by the triangle base).

```
(* TRICOOR4.MA
   Isosceles triang. repres. of genotype frequencies
   in equilibrium inbred populations  with  f = 0.0,
   0.1, 0.2, ..., 0.9, 1.0
 *)
Show[
    Plot[{2.0*x*(1 - x), 1.8*x*(1 - x), 1.6*x*(1 - x),
          1.4*x*(1 - x), 1.2*x*(1 - x), 1.0*x*(1 - x),
          0.8*x*(1 - x), 0.6*x*(1 - x), 0.4*x*(1 - x),
          0.2*x*(1 - x), 0.0*x*(1 - x)}, {x, 0, 1},
          Axes -> None, DisplayFunction -> Identity],
    Graphics[{Line[{{0, 0}, {.5, 1}}],
              Line[{{.5, 1}, {1, 0}}]}],
    DisplayFunction -> $DisplayFunction,
    AspectRatio -> Automatic];
```



DISTRIBUTION OF GENOTYPES IN PAIRS OF RELATIVES

The joint distribution of genotypes in pairs of relatives can be straightforwardly derived by means of the following method, devised by Li & Sachs (Biometrics 10 : 347-360 , 1954).

Given that the genotype of the first relative **(1)** is **AA**, **Aa** or **aa**, and given that the pair shares **2**, **1** and **0** genes identical by descent, the chances that the second relative **(2)** has genotype **AA**, **Aa** or **aa** are

		(2)		
		AA	**Aa**	**aa**
	AA	1	0	0
(1)	**Aa**	0	1	0
	aa	0	0	1

if the pair shares two genes identical by descent, as monozygotic twins always do;

		(2)		
		AA	**Aa**	**aa**
	AA	p	q	0
(1)	**Aa**	p/2	1/2	q/2
	aa	0	p	q

if the pair shares one gene identical by descent, as mother-child pairs always do; and

		(2)		
		AA	**Aa**	**aa**
	AA	p^2	2pq	q^2
(1)	**Aa**	p^2	2pq	q^2
	aa	p^2	2pq	q^2

if they do not share any gene identical by descent at all, as pairs of unrelated individuals don't.

These transitional matrices are called respectively **I**, **T** and **O**.

The transitional matrix for parent-offspring (or mother-child) pairs is simply **T**. To obtain the population frequencies of pairs of mother-child combinations one just has to multiply the elements of the first line of the matrix **T** by p^2, the elements of the second line by **2pq** and the elements of the third line by q^2:

$$\begin{pmatrix} p^3 & p^2q & 0 \\ p^2q & pq & pq^2 \\ 0 & pq^2 & q^3 \end{pmatrix} .$$

When the two relatives are grandparent and grandchild, the respective transition matrix is given obviously by T^2, where **T** is the parent-offspring transition matrix. Since of all possible grandparent-grandchild pairs half of them share one gene identical by descent and the other half none, it comes out that $T^2 = 1/2.T + 1/2.O$, a result that evidently is also valid for half-sibs and for uncle-niece pairs:

$$T^2 = 1/2\begin{pmatrix} p & q & 0 \\ p/2 & 1/2 & q/2 \\ 0 & p & q \end{pmatrix} + 1/2\begin{pmatrix} p^2 & 2pq & q^2 \\ p^2 & 2pq & q^2 \\ p^2 & 2pq & q^2 \end{pmatrix} =$$

$$= \begin{pmatrix} p(1+p)/2 & q(1+2p)/2 & q^2/2 \\ p(1+2p)/4 & (1+4pq)/4 & q(1+2q)/4 \\ p^2/2 & p(1+2q)/2 & q(1+q)/2 \end{pmatrix} .$$

Multiplying the elements of the first, second and third rows by p^2, $2pq$ and q^2 respectively, we obtain the population frequencies of possible pairs of grandparents and grandchildren:

$$= \begin{pmatrix} p^3(1+p)/2 & p^2q(1+2p)/2 & p^2q^2/2 \\ p^2q(1+2p)/2 & pq(1+4pq)/2 & pq^2(1+2q)/2 \\ p^2q^2/2 & pq^2(1+2q)/2 & q^3(1+q)/2 \end{pmatrix} .$$

In the case of full sibs, since $1/4$ of them have 2 genes identical by descent, $1/2$ one and $1/4$ none, the conditional probabilities for the genotype of one sib when that of the other is known is given by $S = 1/4.I + 1/2.T + 1/4.O$:

$$S = 1/4\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + 1/2\begin{pmatrix} p & q & 0 \\ p/2 & 1/2 & q/2 \\ 0 & p & q \end{pmatrix} + 1/4\begin{pmatrix} p^2 & 2pq & q^2 \\ p^2 & 2pq & q^2 \\ p^2 & 2pq & q^2 \end{pmatrix} =$$

$$= \begin{pmatrix} (1+p)^2/4 & q(1+p)/2 & q^2/4 \\ p(1+p)/4 & (1+pq)/2 & q(1+q)/4 \\ p^2/4 & p(1+q)/2 & (1+q)^2/4 \end{pmatrix} .$$

Multiplying the elements of the first, second and third rows by p^2, $2pq$ and q^2 respectively, we obtain the population frequencies of possible pairs of full sibs:

$$\begin{pmatrix} p^2(1+p)^2/4 & p^2q(1+p)/2 & p^2q^2/4 \\ p^2q(1+p)/2 & pq(1+pq) & pq^2(1+q)/2 \\ p^2q^2/4 & pq^2(1+q)/2 & q^2(1+q)^2/4 \end{pmatrix} .$$

Frequencies of other types of pairs of relatives can be derived using the above matrices **I**, **T**, **O** and **S**.

In the case of uncle-niece pairs, the transition matrix is given by $TS = ST$, since a niece is a daughter of an individual's full sib. We have already seen that this same matrix has also the form T^2. Therefore, it follows that

$$T^2 = TS = ST = 1/2.T + 1/2.O .$$

In the case of first cousins, when the genotype of one of them is given, the probabilities for the possible genotypes of the other cousin are given by

$TST = T^3 = 1/4.T + 3/4.O$:

$$TST = \begin{pmatrix} p(1+3p)/4 & q(1+6p)/4 & 3q^2/4 \\ p(1+6p)/8 & (1+12pq)/8 & q(1+6q)/8 \\ 3p^2/4 & p(1+6q)/4 & q(1+3q)/4 \end{pmatrix} ,$$

so that the population frequencies are given by

$$\begin{pmatrix} p^3(1+3p)/4 & p^2q(1+6p)/4 & 3p^2q^2/4 \\ p^2q(1+6p)/4 & pq(1+12pq)/4 & pq^2(1+6q)/4 \\ 3p^2q^2/4 & pq^2(1+6q)/4 & q^3(1+3q)/4 \end{pmatrix}.$$

The first expression, **TST**, is obtained straightforwardly if we remember that first cousins are the offspring of two full sibs. The last formula, **1/4.T + 3/4.O**, expresses the plain fact that, among all first cousins, **1/4** of them share **1** gene identical by descent and **3/4** none.

The middle expression T^3 is obtained from $TST = TS.T = T^2.T = T^3$. The formulae $T^2 = 1/2.T + 1/2.O$ and $T^3 = 1/4.T + 3/4.O$ are just special cases, for $n = 2$ and 3 respectively, of the following general relationship, which holds for any value of $n \geq 1$:

$$T^n = 1/2^{n+1}.T + (1-1/2^{n+1}).O$$
$$= 1/2^{n-1}.(T-O) + O.$$

The formula above is valid for transitional matrices of unilineal relatives. In the case of bilineal relatives the pertinent cases are full sibs and double first cousins. The transitional matrix for the former case is given, as we have already seen, by

$$S = 1/4.I + 1/2.T + 1/4.O.$$

Of all possible pairs of double first cousins, **1/16** of them share two genes identical by descent, **6/16** one and **9/16** none. Also, double first cousins are the sibs of two brothers married to two sisters. Therefore, the transitional matrix for double first cousins has the form

$$S^2 = 1/16.I + 6/16.T + 9/16.O :$$

$$S^2 = \begin{pmatrix} (1+3p)^2/16 & 6q(1+3p)/16 & 9q^2/16 \\ 3p(1+3p)/16 & (4+18pq)/16 & 3q(1+3q)/16 \\ 9p^2/16 & 6p(1+3q)/16 & (1+3q)^2/16 \end{pmatrix}.$$

The matrix of population frequencies is given in this case by :

$$\begin{pmatrix} p^2(1+3p)^2/16 & 6p^2q(1+3p)/16 & 9p^2q^2/16 \\ 6p^2q(1+3p)/16 & 2pq(4+18pq)/16 & 6pq^2(1+3q)/16 \\ 9p^2q^2/16 & 6pq^2(1+3q)/16 & q^2(1+3q)^2/16 \end{pmatrix}.$$

HIERARCHICAL STRUCTURE OF POPULATIONS: ISOLATE EFFECT (WAHLUND'S EFFECT)

Let us consider a population subdivided into n isolates of equal size where a pair of alleles (A, a) is segregating at an autosomal locus in each one of the isolates. Assuming that each one of these subpopulations is in Hardy-Weinberg equilibrium, we have therefore:

$p_1, p_2, \ldots, p_n; q_1, q_2, \ldots, q_n; p_i + q_i = 1,$

and, evidently,

$q = \Sigma q_i/n \ , \ p = \Sigma p_i/n = 1-q$

$var(q) = var(1-p) = var(-p) = var(p) = \Sigma(q_i-q)^2/n$

$= (\Sigma q_i^2 + nq^2 - 2q\Sigma q_i)/n = \Sigma q_i^2/n - q^2,$

since $2q\Sigma q_i = 2q.nq = 2nq^2$.

For n isolates of different sizes $X_1, X_2, \ldots, X_n$, we have

$q = \Sigma x_i q_i \ , \ p = \Sigma x_i p_i = 1-q$ and

$var(q) = var(p) = \Sigma x_i q_i^2 - q^2 = \Sigma x_i p_i^2 - p^2,$

where $x_i = X_i/\Sigma X_i$.

In the total population (without isolate breakdown),

$P(AA) \neq p^2 \ , \ P(Aa) \neq 2pq \ , \ P(aa) \neq q^2;$

Hardy-Weinberg proportions will be found only if there is a breakdown of isolates with random matings among individuals of all subpopulations. The distribution of genotypes AA, Aa and aa in the total population (without isolate breakdown) is given by

$P(AA) = \Sigma p_i^2/n \ , \ P(Aa) = 2\Sigma p_i q_i/n \ , \ P(aa) = \Sigma q_i^2/n$ (in the case of n isolates with equal sizes) or by

$P(AA) = \Sigma x_i p_i^2 \ , \ P(Aa) = 2\Sigma x_i p_i q_i \ , \ P(aa) = \Sigma x_i q_i^2$ (in the case of n isolates with different sizes).

From the above formulae it comes out that, for isolates of equal sizes,

$var(p) = \Sigma p_i^2/n - p^2$ and hence $\Sigma p_i^2/n = p^2 + var(p)$;

$var(q) = \Sigma q_i^2/n - q^2$ and hence $\Sigma q_i^2/n = q^2 + var(q)$;

and, for isolates of different sizes,

$var(p) = \Sigma x_i p_i^2 - p^2$ and hence $\Sigma x_i p_i^2 = p^2 + var(p)$;

$var(q) = \Sigma x_i q_i^2 - q^2$ and hence $\Sigma x_i q_i^2 = q^2 + var(q)$.
Therefore,

$P(AA) = \Sigma p_i^2/n = \Sigma x_i p_i^2 = p^2 + var(p)$

$P(aa) = \Sigma q_i^2/n = \Sigma x_i q_i^2 = q^2 + var(q)$

$P(Aa) = 2\Sigma p_i q_i/n = 2\Sigma x_i p_i q_i = 2pq - 2.var(p) = 2pq - 2.var(q)$.

As a numerical example, let us consider the following 11 isolates, all with the same size:

isolate	p_i	q_i	p_i^2	$2p_iq_i$	q_i^2
1	0.00	1.00	0.00	0.00	1.00
2	0.10	0.90	0.01	0.18	0.81
3	0.20	0.80	0.04	0.32	0.64
4	0.30	0.70	0.09	0.42	0.49
5	0.40	0.60	0.16	0.48	0.36
6	0.50	0.50	0.25	0.50	0.25
7	0.60	0.40	0.36	0.48	0.16
8	0.70	0.30	0.49	0.42	0.09
9	0.80	0.20	0.64	0.32	0.04
10	0.90	0.10	0.81	0.18	0.01
11	1.00	0.00	1.00	0.00	0.00
A : total before isolate breakdown			0.35	0.30	0.35
B : total after compl. isol. brkd.			0.25	0.50	0.25
A-B			0.10	-0.20	0.10

Comparing

$P(AA) = p^2 + var(p)$, $P(Aa) = 2pq - 2.var(p)$, $P(aa) = q^2 + var(p)$

to the values found in Wright's equilibrium,

$P(AA) = p^2 + fpq$, $P(Aa) = 2pq - 2fpq$, $P(aa) = q^2 + fpq$,

where f is the average population inbreeding coefficient, it comes out that

$q^2 + fpq = q^2 + var(q)$,
$fpq = var(q)$ and
$f = var(p)/pq$.

In the example shown above, the effect of isolation (population subdivision) is equivalent to an average population inbreeding coefficient of

$f = 0.10/0.25 = 0.40$,

in spite of each subpopulation having its $f_i = 0$.

A special case takes place when there are only two isolates of equal size and matings occurring immediately after isolate breakdown take place only between individuals from different

populations. This situation was studied with some detail by Crow and Kimura (Introduction to population genetics theory, Harper & Row, New York, 1970). In this case the formulae for p, var(p) and P(AA) reduce to

$$p = (p_1 + p_2)/2$$

$$\begin{aligned} var(p) &= (p_1^2 + p_2^2)/2 - p^2 \\ &= (p_1^2 + p_2^2)/2 - (p_1 + p_2)^2/4 \\ &= (p_1^2 - p_2^2)/4 \end{aligned}$$

and

$$\begin{aligned} P(AA) &= p^2 + var(p) \\ &= (p_1 + p_2)^2/4 + (p_1^2 - p_2^2)/4 \\ &= (p_1^2 + p_2^2)/2 \ . \end{aligned}$$

It is also simple to verify that

$$\begin{aligned} p^2 - var(p) &= (p_1 + p_2)^2/4 - (p_1^2 - p_2^2)/4 \\ &= p_1p_2 \ , \end{aligned}$$

and this is precisely the frequency of AA individuals in the hybrid F_1 population.

Therefore, the frequency of AA individuals is:

$P(AA) = p^2 + var(p)$ before isolate breakdown;

$P(AA) = p_1p_2 = p^2 - var(p)$ in the first generation;

$P(AA) = p^2$ in the generations that follow,

being thus the arithmetic mean of the values in the two preceding generations.

Considering the generalized situation of k different isolates, within each of which the genotype frequencies are given by

$$\begin{aligned} P_k(AA) &= p_k^2 + F_kp_kq_k = F_kp_k + (1-F_k)p_k^2, \\ P_k(Aa) &= 2p_kq_k(1-F_k), \text{ and} \\ P_k(aa) &= q_k^2 + F_kp_kq_k = F_kq_k + (1-F_k)q_k^2; \end{aligned}$$

the genotype frequencies are, in the total population:

$$\begin{aligned} P(AA) &= \Sigma x_i[p_i^2+F_ip_iq_i], \\ P(Aa) &= 2\Sigma x_ip_iq_i(1-F_i), \text{ and} \\ P(aa) &= \Sigma x_i[q_i^2+F_ip_iq_i], \end{aligned}$$

where p_k, q_k and F_k are the allele frequencies and the fixation index of the k-th subpopulation and $x_k = N_k/\Sigma N_i$ is the contribution in size of the k-th subpopulation to the total population.

In the total population allele frequencies are calculated after

$p = \Sigma x_i p_i$ and $q = 1-p = \Sigma x_i q_i$;

and the variance of gene frequencies among subpopulations (isolates) by

$$var(p) = \Sigma x_i (p_i - p)^2 = \Sigma x_i p_i^2 - p^2$$
$$= var(q) = \Sigma x_i (q_i - q)^2 = \Sigma x_i q_i^2 - q^2.$$

The correlation between random gametes within subpopulations relative to gametes of the total population, that is, the fixation index generated by population subdivision or Wahlund's effect (F_{ST}) is calculated, as in the case of panmictic subpopulations, after

$$F_{ST} = var(p)/pq = [\Sigma x_i p_i^2 - (\Sigma x_i p_i)^2]/(\Sigma x_i p_i . \Sigma x_i q_i)$$
$$= [\Sigma x_i p_i^2 - (\Sigma x_i p_i)(1-\Sigma x_i q_i)]/(\Sigma x_i p_i . \Sigma x_i q_i)$$
$$= [\Sigma x_i p_i^2 - \Sigma x_i p_i + \Sigma x_i p_i . \Sigma x_i q_i)]/(\Sigma x_i p_i . \Sigma x_i q_i)$$
$$= [\Sigma x_i p_i . \Sigma x_i q_i - \Sigma x_i p_i (1-p_i)]/(\Sigma x_i p_i . \Sigma x_i q_i)$$
$$= 1 - \Sigma x_i p_i q_i/(\Sigma x_i p_i . \Sigma x_i q_i) = 1 - 2\Sigma x_i p_i q_i/2pq$$

The correlation between uniting gametes relative to gametes in the total population (F_{IT}), that is the fixation index in the total population due to both population subdivision and inbreeding occurring within subpopulations, that for the case when there is no inbreeding within populations takes value $F_{IT} = F_{ST}$, is obtained directly from

$$F_{IT} = 1 - \Sigma x_i P_i(Aa)/2pq$$
$$= 1 - P(Aa)/2pq$$
$$= 1 - 2\Sigma x_i p_i q_i (1-F_i)/2pq$$

The value of F_{IS}, the fixation index due to inbreeding within subpopulations, is taken from $F_{IS} = (F_{IT}-F_{ST})/(1-F_{ST})$, because $F_{IT} = F_{ST} + F_{IS} - F_{IS}.F_{ST}$; the last equation arises from the fact that the total probability of an individual being heterozygous relative to random mating ($1-F_{IT}$) is the product $(1-F_{IS})(1-F_{ST})$, that is, for being heterozygous this individual should not be homozygous neither by inbreeding within populations nor by Wahlund's effect. In fact, from $(1-F_{IT}) = (1-F_{IS})(1-F_{ST})$ we obtain successively

$$1 - F_{IT} = 1 - F_{IS} - F_{ST} + F_{IS}.F_{ST} ,$$
$$F_{IT} = F_{ST} + F_{IS}(1-F_{ST})$$
and

$F_{IS} = (F_{IT}-F_{ST})/(1-F_{ST})$.

Since $F_{IT} = 1 - P(Aa)/2pq$ and $F_{ST} = var(p)/pq$, it comes out that $F_{IS} = 1 - P(Aa)/\{2[pq-var(p)]\}$. But $P(Aa) = 2\Sigma x_i p_i q_i(1-F_i)$ and $pq - var(p) = pq - \Sigma x_i p_i^2 + p^2 = p - \Sigma x_i p_i^2 = \Sigma x_i p_i - \Sigma x_i p_i^2 = \Sigma x_i p_i q_i$. Therefore,

$F_{IS} = 1 - \Sigma x_i p_i q_i(1-F_i)/\Sigma x_i p_i q_i = 1 - 2\Sigma x_i p_i q_i(1-F_i)/2\Sigma x_i p_i q_i$,
$1 - F_{IS} = \Sigma x_i p_i q_i(1-F_i)/\Sigma x_i p_i q_i$, and
$F_{IS} = \Sigma x_i p_i q_i F_i/\Sigma x_i p_i q_i$;

therefore, the average fixation index due to inbreeding within each subpopulation can be obtained directly from the weighed mean shown above. We note that, if we (erroneously) put $F_{IS} = \Sigma x_i F_i$, as many authors do, the relationships among the three F's do not hold except for some particular cases without general interest.

Summary of formulae

Symbol	Definiton
F_{ST}	Fixation index due to population subdivision
F_{IT}	Fixation index in the total population
F_{IS}	Fixation index due to inbreeding within subpopulations
F_k	Fixation index of the k-th subpopulation

F	$f(F_{ij})$	f[P(Aa),p,q,var(p)]	$f[x_i,F_i,p_i,q_i]$
F_i			$1 - p_i(Aa)/2p_iq_i = 1 - h_i/2p_iq_i$
F_{ST}	$(F_{IT}-F_{IS})/(1-F_{IS})$	var(p)/pq	$[\Sigma x_i p_i^2-(\Sigma x_i p_i)^2]/(\Sigma x_i p_i.\Sigma x_i q_i)$
			$= 1 - \Sigma x_i p_i q_i/(\Sigma x_i p_i.\Sigma x_i q_i)$
			$= 1 - 2\Sigma x_i p_i q_i/2pq$
F_{IT}	$F_{ST}+F_{IS}-F_{IS}.F_{ST}$	1-P(Aa)/2pq	$1 - \Sigma x_i p_i q_i(1-F_i)/(\Sigma x_i p_i.\Sigma x_i q_i)$
			$= 1 - 2\Sigma x_i p_i q_i(1-F_i)/2pq$
F_{IS}	$(F_{IT}-F_{ST})/(1-F_{ST})$	1-P(Aa)/2[pq-var(p)]	$1 - \Sigma x_i p_i q_i(1-F_i)/\Sigma x_i p_i q_i =$
			$= 1 - 2\Sigma x_i p_i q_i(1-F_i)/2\Sigma x_i p_i q_i$

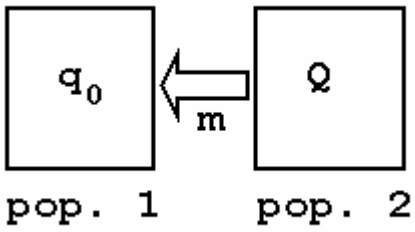
Numerical example: let us consider two isolates of approximately equal sizes, with the following genotype frequencies: $P_1(AA) = 0.28$, $P_1(Aa) = 0.24$, $P_1(aa) = 0.48$; and $P_2(AA) = 0.104$, $P_2(Aa) = 0.192$, $P_2(aa) = 0.704$. For the first isolate we obtain $p_1 = 0.4$, $q_1 = 0.6$ and $F_1 = 1 - P_1(Aa)/2pq = 0.5$; and for the second, $p_2 = 0.2$, $q_2 = 0.8$ and $F_2 = 1 - P_2(Aa)/2pq = 0.4$, so that in the total population $P(AA) = [P_1(AA)+P_2(AA)]/2 = 0.192$, $P(Aa) = [P_1(Aa)+P_2(Aa)]/2 = 0.216$, $P(aa) = [P_1(aa)+P_2(aa)]/2 = 0.592$, $p =$

$(p_1+p_2)/2 = 0.3$, $q = (q_1+q_2)/2 = 0.7$ and $F_{IS} = (p_1q_1F_1+p_2q_2F_2)/(p_1q_1+p_2q_2) = 0.46$. The variance of gene frequencies between isolates takes value $var(p) = [(0.4-0.3)^2+(0.2-0.3)^2]/2 = var(q) = [(0.6-0.7)^2+(0.8-0.7)^2]/2 = 0.01$. Therefore, we have

$F_{ST} = var(p)/pq$	$= 0.01/0.21$	$= 0.047619$
$F_{IS} = (p_1q_1F_1+p_2q_2F_2)/(p_1q_1+p_2q_2)$	$= 0.184/0.4$	
$= 1 - P(Aa)/\{2[pq-var(p)]\}$	$= 1 - 0.216/0.4$	$= 0.460000$
$F_{IT} = 1 - P(Aa)/2pq$	$= 1 - 0.216/0.42$	
$= F_{ST} + F_{IS}(1-F_{ST})$	$= 0.102/0.21$	$= 0.485714$

MIGRATION

Let us first consider the following simple migration model, proposed by Glass and Li (Amer. J. Hum. Genet. 5: 1-20, 1953):



where q_0 and Q are the frequencies of a given allele in populations 1 and 2 respectively. Assuming that a fraction m of the gene pool of population 1 is replaced by genes from population 2 per generation and that there is no migration from population 1 towards population 2, we obtain the following first order difference equation

$q_1 = q_0(1-m) + Qm$ or, in general,

$q_{n+1} = q_n(1-m) + Qm$

since the value of Q remains unchanged as generations go by, at equilibrium (that is, when n tends to infinity)

$q = q(1-m) + Qm$ and therefore $q = Q$.

Subtrating the quantity Q from both sides of $q_1 = q_0(1-m) + Qm$ we obtain

$$\begin{aligned} q_1 - Q &= q_0(1-m) + Qm \\ &= q_0 - q_0m + Qm - Q \\ &= (q_0 - Q)(1-m) \ ; \end{aligned}$$

therefore, the general solution for $q_n - Q$ is

$q_n - Q = (q_0 - Q)(1-m)^n$;

and that for q_n is

$q_n = Q + (q_0 - Q)(1-m)^n$.

Rearranging the expression above we obtain

$(1-m)^n = (q_n - Q)/(q_0 - Q)$.

The quantity $(q_n - Q)/(q_0 - Q)$ is the proportion, in population 1, of the genes originally contained in it before the migration process started, and as such is a good measure of racial or population admixture. In fact, the model above delineated was applied to data on American negroes by its authors and showed that American negroes have on average 30% of genes of white origin. Using the following data on R^0 (from Rh series) gene frequency: $q_0 = 0.630$ (among present black populations from Africa), Q =

0.028 (among American Caucasoids) and q_n = 0.446 (among American negroes), Glass and Li obtained the figure of

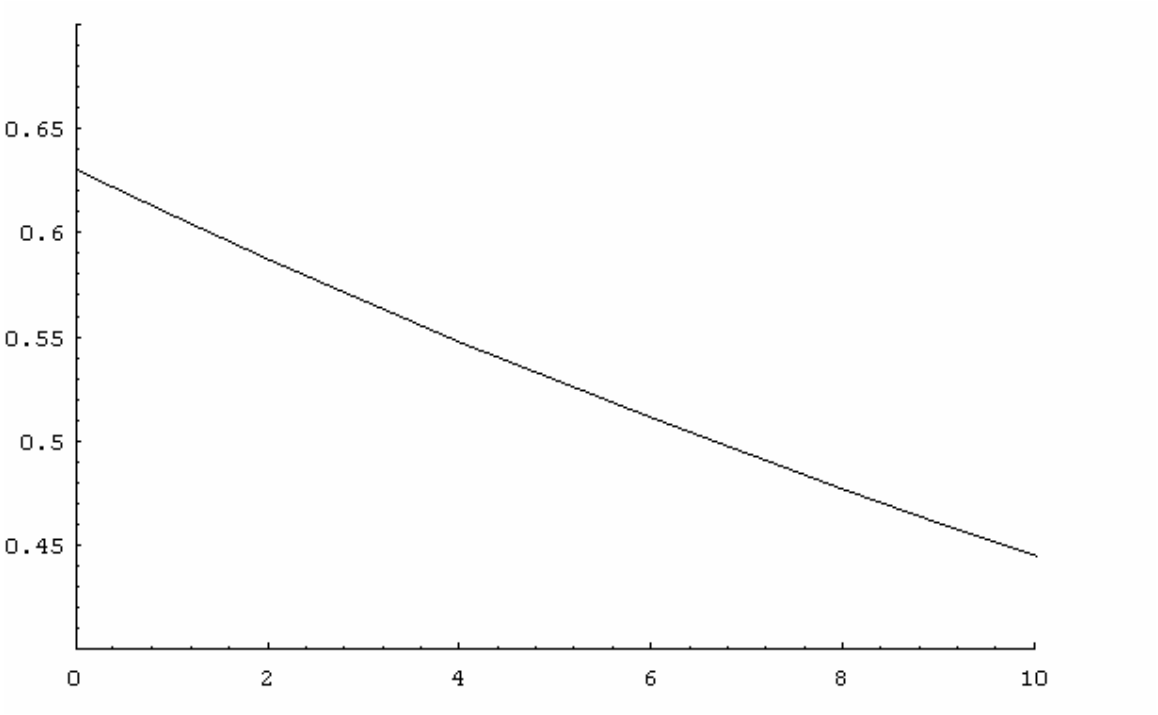
$(q_n - Q)/(q_0 - Q)$ = (0.446 - 0.028)/(0.630 - 0.028) = 0.69435 for the proportion of African genes in the American negro. (They have therefore around 30% of European genes.)

Glass and Li calculated also the value of m in the expression $(1-m)^n = (q_n - Q)/(q_0 - Q)$ = 0.69435, assuming that the rate has been constant during 10 generations [n estimated as 10 generations occurring in the interval 1675 to 1950, having therefore the value of (1950-1675)/10 = 27.5 years]: since $(1-m)^n = (1-m)^{10}$ = 0.69435, it comes out that 10.ln(1-m) = ln 0.69435 and m = 0.036. Therefore, the flow of genes from the white populations takes place at a rate of 3.6% per generation. The alteration in gene frequency that occurred in the black american population is shown in the table below:

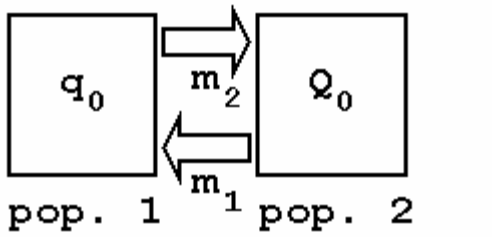
generation	freq. R^0	prop. of white genes
0	0.63	0.00
1	0.61	0.04
2	0.59	0.07
3	0.57	0.10
4	0.55	0.14
5	0.53	0.17
6	0.51	0.20
7	0.49	0.23
8	0.48	0.26
9	0.46	0.28
10	0.45	0.30

The data on R^0 gene frequency are plotted in the graph below, generated by the following Mathematica code:

```
(* migrat01.ma *)
q0 = 0.630; Q  = 0.028;
q[n_] := Q + 0.964^n * (q0-Q);
freq = Table[q[i], {i, 0, 10}];
ListPlot[freq, PlotJoined -> True,
          PlotRange -> {0.4,0.7},
          AxesOrigin -> {0,0.4}]
```



Let us now consider the case of two populations where the frequency of a given allele is q_0 (in population 1) and Q_0 (in population 2).



Assuming the constant migration rates m_1 (from population 2 to population 1) and m_2 (from population 1 to population 2) per generation (that is, m_1 is the proportion, in population 1, of genes that come from population 2 per generation and m_2 is the proportion, in population 2, of genes that come from population 1), we get the following system of first order difference equations:

$$q_1 = (1-m_1)q_0 + m_1Q_0$$
$$Q_1 = m_2q_0 + (1-m_2)Q_0 ,$$

or, in matrix compressed form,

$$\begin{pmatrix} q_1 \\ Q_1 \end{pmatrix} = \begin{pmatrix} 1-m_1 & m_1 \\ m_2 & 1-m_2 \end{pmatrix} \begin{pmatrix} q_0 \\ Q_0 \end{pmatrix} ;$$

subtrating the second of the equations above from the first one we get

$$q_1-Q_1 = (1-m_1-m_2)q_0 - (1-m_1-m_2)Q_0 = (1-m_1-m_2)(q_0-Q_0);$$

therefore,

$q_n - Q_n = (1-m_1-m_2)^n(q_0-Q_0)$;

since $0 < m_1, m_2 < 1$, it comes out that $0 < |1-m_1-m_2| < 1$. Therefore, at equilibrium (that is, when n tends to infinity), $q = Q$.

Rearranging equation $q_1 = (1-m_1)q_0 + m_1Q_0$, we obtain

$q_1 = q_0 - m_1(q_0-Q_0)$

and therefore

$$q_2 = q_1 - m_1(q_1-Q_1)$$
$$= q_0 - m_1(q_0-Q_0) - m_1(1-m_1-m_2)(q_0-Q_0)$$

$$q_3 = q_2 - m_1(q_2-Q_2)$$
$$= q_0 - m_1(q_0-Q_0) - m_1(1-m_1-m_2)(q_0-Q_0) - m_1(1-m_1-m_2)^2(q_0-Q_0)$$

...

$$q_n = q_0 - m_1(q_0-Q_0)[(1-m_1-m_2)^0 + (1-m_1-m_2)^1 + ...+ (1-m_1-m_2)^{n-1}]$$
$$= q_0 - m_1(q_0-Q_0)[1 - (1-m_1-m_2)^n]/(m_1 + m_2)$$
$$= q_0 - m_1(q_0-Q_0)/(m_1+m_2) + m_1(q_0-Q_0)(1-m_1-m_2)^n/(m_1+m_2),$$

which is the general solution for q_n . As n increases, $(1-m_1-m_2)^n$ tends to zero, so that at equilibrium

$$q = q_0 - m_1(q_0-Q_0)/(m_1+m_2)$$
$$= [q_0(m_1+m_2) - m_1(q_0-Q_0)]/(m_1+m_2)$$
$$= (q_0m_2+Q_0m_1)/(m_1+m_2) .$$

The general solution for Q_n is taken from $Q_n = q_n - (1-m_1-m_2)^n(q_0-Q_0)$ and has the form

$Q_n = q_0 - m_1(q_0-Q_0)/(m_1+m_2) - m_2(q_0-Q_0)(1-m_1-m_2)^n/(m_1+m_2)$;

therefore, at equilibrium, $Q = q = (q_0m_2+Q_0m_1)/(m_1+m_2)$, as already stated.

Special cases:

(1) $m_1 = m_2 = m$

In this case the general solutions are given by

$q_n = (q_0 + Q_0)/2 + (q_0-Q_0)(1-2m)^n/2$

$Qn = (q_0 + Q_0)/2 - (q_0-Q_0)(1-2m)^n/2$;

at equilibrium, therefore, $q = Q = (q_0 + Q_0)/2$.

(2) $m_1 = 1-m_2$

In this case, the equilibrium is attained in one single generation, having the form

$q = q_1 = Q = Q_1 = q_0 - m_1(q_0 - Q_0)$.

(3) $m_1 = m$ and $m_2 = 0$

The first case we examined (model of Glass and Li) turns out to be a special case of this model; in fact, when $m_1 = m$ and $m_2 = 0$, the system of difference equations takes the form

$$\begin{pmatrix} q_1 \\ Q_1 \end{pmatrix} = \begin{pmatrix} 1-m & m \\ 0 & 1 \end{pmatrix} \begin{pmatrix} q_0 \\ Q_0 \end{pmatrix},$$

from which we get the results

$q_1 = q_0(1-m) + mQ_0$ and $Q_1 = Q_0$ already discussed.

RACE ADMIXTURE CALCULATIONS

Let $\{h_1, a_1, b_1; h_2, a_2, b_2; \ldots ; h_n, a_n, b_n\}$ be the frequencies of n different genes in the hybrid Brazilian population, original European stock and original African stock respectively. If the unknown quantities that we want to determine are **x** and **y** (respective contributions of European and African genes to the hybrid population) then all we have to do is to obtain their best estimates from the set of n equations

$$\{h_1 = a_1x + b_1y \ , \ h_2 = a_2x + b_2y \ , \ \ldots \ , \ \ h_n = a_nx + b_ny \ \} \ .$$

This can be achieved through several reliable methods published in the literature. In the lines below we use the least squares method proposed by Roberts & Hiorns (**Amer. J. Hum. Genet. 14:** 261-267,1962; **Hum. Biol. 37:** 38-43, 1965), starting by rewriting the set of equations listed above in matrix condensed form:

$$\begin{pmatrix} h_1 \\ h_2 \\ . \\ . \\ h_n \end{pmatrix} = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ . & . \\ . & . \\ a_n & b_n \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Multiplying both sides of the above equation by the transpose of $(a_i \ , \ b_i)$, namely $(a_i \ , \ b_i)^T$, we obtain successively

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ . & . \\ . & . \\ a_n & b_n \end{pmatrix}^T \begin{pmatrix} h_1 \\ h_2 \\ . \\ . \\ h_n \end{pmatrix} = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ . & . \\ . & . \\ a_n & b_n \end{pmatrix}^T \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ . & . \\ . & . \\ a_n & b_n \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

and

$$\begin{pmatrix} \Sigma a_i h_i \\ \Sigma b_i h_i \end{pmatrix} = \begin{pmatrix} \Sigma a_i^2 & \Sigma a_i b_i \\ \Sigma a_i b_i & \Sigma b_i^2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Multiplying now both sides of this equation by the inverse of $[(a_i \ , \ b_i)^T(a_i \ , \ b_i)]$, namely $[(a_i \ , \ b_i)^T(a_i \ , \ b_i)]^{-1}$, we obtain successively

$$\begin{pmatrix} \Sigma a_i^2 & \Sigma a_i b_i \\ \Sigma a_i b_i & \Sigma b_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \Sigma a_i h_i \\ \Sigma b_i h_i \end{pmatrix} = \begin{pmatrix} \Sigma a_i^2 & \Sigma a_i b_i \\ \Sigma a_i b_i & \Sigma b_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \Sigma a_i^2 & \Sigma a_i b_i \\ \Sigma a_i b_i & \Sigma b_i^2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

$$1/(\Sigma a_i^2 \times \Sigma b_i^2 - \Sigma a_i b_i \times \Sigma a_i b_i) \ . \ \begin{pmatrix} \Sigma b_i^2 & -\Sigma a_i b_i \\ -\Sigma a_i b_i & \Sigma a_i^2 \end{pmatrix} \begin{pmatrix} \Sigma a_i h_i \\ \Sigma b_i h_i \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma a_i^2 & \Sigma a_i b_i \\ \Sigma a_i b_i & \Sigma b_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \Sigma a_i^2 & \Sigma a_i b_i \\ \Sigma a_i b_i & \Sigma b_i^2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$$

and

$$x = (\Sigma b_i^2 \times \Sigma a_i h_i - \Sigma a_i b_i \times \Sigma b_i h_i)/(\Sigma a_i^2 \times \Sigma b_i^2 - \Sigma a_i b_i \times \Sigma a_i b_i)$$
$$y = (\Sigma a_i^2 \times \Sigma b_i h_i - \Sigma a_i b_i \times \Sigma a_i h_i)/(\Sigma a_i^2 \times \Sigma b_i^2 - \Sigma a_i b_i \times \Sigma a_i b_i) \ .$$

Below we summarize the results obtained by applying these methods to concrete data (data from P. A. Otto, L. A. Praxedes, N. N. Salaru, S. Wendel, M. G. Aravechia, **ALLELE AND HAPLOTYPE FREQUENCIES FROM MNSs, KELL-CELLANO, Rh, ABO, DUFFY, KIDD AND SUTTER SYSTEMS IN BRAZILIAN CAUCASOIDS AND NEGROIDS FROM SOUTHERN BRAZIL AND ESTIMATES OF RACIAL ADMIXTURE**, in preparation since 1992).

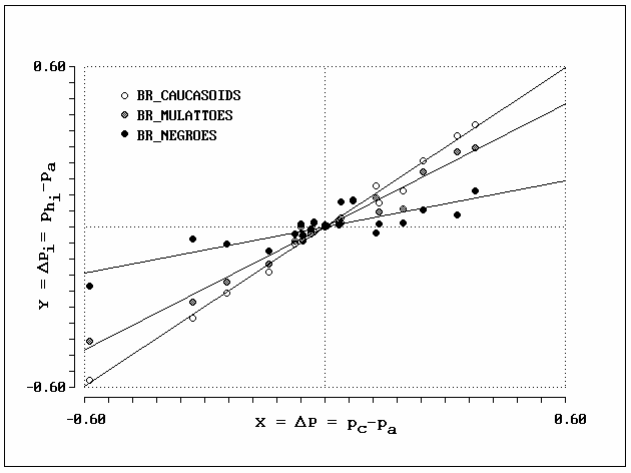
Allele and haplotype frequencies (± 1 s.e.) used in the calculations of racial admixture.

	af_negroes	eu_caucas.	br_caucas.	br_mulatt.	br_negroes
CDE	0.000±0.000	0.001±0.001	0.004±0.001	0.006±0.001	0.000±0.000
CDe	0.039±0.013	0.413±0.010	0.423±0.003	0.337±0.004	0.175±0.015
Cde	0.060±0.013	0.004±0.002	0.009±0.001	0.010±0.001	0.030±0.010
cDE	0.043±0.006	0.113±0.006	0.140±0.002	0.144±0.003	0.140±0.012
cDe	0.645±0.020	0.057±0.007	0.071±0.002	0.219±0.005	0.423±0.023
cdE	0.000±0.000	0.005±0.002	0.005±0.001	0.002±0.001	0.004±0.004
cde	0.213±0.019	0.408±0.011	0.348±0.003	0.282±0.006	0.228±0.021
A	0.157±0.004	0.293±0.006	0.247±0.003	0.212±0.004	0.169±0.012
B	0.132±0.003	0.057±0.003	0.068±0.001	0.079±0.002	0.107±0.010
O	0.711±0.005	0.650±0.006	0.685±0.003	0.710±0.004	0.724±0.015
Fy^a	0.039±0.004	0.369±0.022	0.382±0.006	0.321±0.005	0.084±0.024
Fy	0.961±0.004	0.631±0.022	0.618±0.006	0.679±0.005	0.916±0.024
Jk^a	0.762±0.014	0.517±0.025	0.513±0.008	0.555±0.007	0.698±0.072
Jk	0.238±0.014	0.483±0.025	0.487±0.008	0.445±0.007	0.302±0.072
MS	0.092±0.017	0.219±0.029	0.246±0.005	0.200±0.004	0.070±0.025
Ms	0.488±0.026	0.348±0.032	0.318±0.005	0.349±0.005	0.399±0.041
NS	0.044±0.014	0.084±0.022	0.079±0.004	0.059±0.003	0.137±0.031
Ns	0.376±0.026	0.348±0.031	0.357±0.005	0.391±0.005	0.394±0.042
K	0.003±0.001	0.037±0.008	0.031±0.001	0.019±0.002	0.010±0.004
k	0.997±0.001	0.963±0.008	0.969±0.001	0.981±0.002	0.990±0.004

Estimated proportions of european caucasoid genes ($b_C = \Sigma XY/\Sigma X^2$) and of african Bantu negro genes (b_a) in Brazilian caucasoids, mulattoes and negroes; $se(b_C) = \sqrt{var(b_C)}$, $var(b_C) = \{\Sigma Y^2-(\Sigma Y)^2/n - (\Sigma XY-\Sigma X.\Sigma Y/n)^2/[\Sigma X^2-(\Sigma X)^2/n]\}/\{(n-1)[\Sigma X^2-(\Sigma X)^2/n]\}$.

H_i	b_c	se(b_c)	95% CI(b_c)	r^2	t(19df)	P{b_c = 0}
Br.caucasoids	0.993	0.024	0.943 - 1.000	0.989	41.399	< 0.001
Br.mulattoes	0.766	0.033	0.697 - 0.835	0.966	23.231	< 0.001
Br.negroes	0.288	0.042	0.200 - 0.376	0.713	6.870	< 0.001

H_i	b_a	se(b_a)	95% CI(b_a)	r^2	t(19df)	P{b_a = 0}
Br.caucasoids	0.007	0.024	0.000 - 0.057	0.004	0.288	> 0.050
Br.mulattoes	0.234	0.033	0.165 - 0.303	0.725	7.079	< 0.001
Br.negroes	0.712	0.042	0.624 - 0.800	0.938	16.996	< 0.001



Estimated regressions $\mathbf{Y = \Delta P_i = (P_{hi} - p_a) = b_{ci}X = b_{ci}\Delta P = b_{ci}(P_c - P_a)}$, where $\mathbf{b_{ci}}$ stands for the proportions of caucasoid genes in Brazilians classified as "whites", "mulattoes" and "negroes".

PROBABILITY OF EXTINCTION OF A NEUTRAL MUTANT GENE

Let us consider a population with large but finite size **N**, in which just one individual carries a mutant gene **A** in heterozygous state. If this mutant individual has **k** children, the probability that the **A** gene is not transmitted to any of them is

$$L_k = (1/2)^k = 1/2^k \quad ,$$

since the probability of not transmitting the mutant allele to each child is 1/2 :

$$\mathbf{Aa \times aa \rightarrow 1/2\ Aa + 1/2\ aa} \ .$$

If we assume that the offspring number per couple follows the Poisson distribution, it comes out that the probability of a couple having **k** children is

$$P_k = e^{-m}\, m^k/k! \ ;$$

if the average number of children per couple is **m = 2** (hence the population size **N** will be kept constant as generations go by), it comes out that

$$P_k = e^{-2} 2^k/k! \ ;$$

if we assume that the mutation is neutral, that is, that the mutant individual **Aa** has the same probability as an **aa** individual of having **k** children, it comes out that the probability of the mutant **Aa** having **k** children is also

$$P_k = e^{-2} 2^k/k! \ ;$$

since, having **k** children, the probability that gene **A** is not transmitted to any of them is

$$L_k = 1/2^k \ ,$$

it comes out that the probability of loss of the gene in one generation is

$$E(1) = \sum_{k=0}^{inf} P_k L_k = \sum_{k=0}^{inf} e^{-2} 2^k/2^k k! = \sum_{k=0}^{inf} e^{-2}/k! = e^{-2} . \sum_{k=0}^{inf} 1/k! = e^{-2} . e = e^{-1}$$

$$= 0.3679 \ .$$

The table below shows the necessary calculations for getting the final value of **E(1)** using the preceding formula. Convergence to the exact value with eight decimal places (e^{-1} **= 0.36787944**) is very fast, already occurring when **k** takes the value 10.

k	P(k)	SP(k)	L(k)	P(k)L(k)	SP(k)L(k)
0	0.13533528	0.13533528	1.00000000	0.13533528	0.13533528
1	0.27067057	0.40600585	0.50000000	0.13533528	0.27067057
2	0.27067057	0.67667642	0.25000000	0.06766764	0.33833821
3	0.18044704	0.85712346	0.12500000	0.02255588	0.36089409
4	0.09022352	0.94734698	0.06250000	0.00563897	0.36653306
5	0.03608941	0.98343639	0.03125000	0.00112779	0.36766085
6	0.01202980	0.99546619	0.01562500	0.00018797	0.36784882
7	0.00343709	0.99890328	0.00781250	0.00002685	0.36787567
8	0.00085927	0.99976255	0.00390625	0.00000336	0.36787903
9	0.00019095	0.99995350	0.00195313	0.00000037	0.36787940
10	0.00003819	0.99999169	0.00097656	0.00000004	0.36787944
11	0.00000694	0.99999864	0.00048828	0.00000000	0.36787944
12	0.00000116	0.99999979	0.00024414	0.00000000	0.36787944
13	0.00000018	0.99999997	0.00012207	0.00000000	0.36787944
14	0.00000003	1.00000000	0.00006104	0.00000000	0.36787944
15	0.00000000	1.00000000	0.00003052	0.00000000	0.36787944
16	0.00000000	1.00000000	0.00001526	0.00000000	0.36787944
17	0.00000000	1.00000000	0.00000763	0.00000000	0.36787944
18	0.00000000	1.00000000	0.00000381	0.00000000	0.36787944
19	0.00000000	1.00000000	0.00000191	0.00000000	0.36787944
20	0.00000000	1.00000000	0.00000095	0.00000000	0.36787944
..	.....	.....	.....	.....	.....
inf.	0.00000000	1.00000000	0.00000000	0.00000000	0.36787944

This last table was generated by the following BASIC code:

```
REM PROGRAM FILENAME EXTINPR4.BAS
DEFDBL A-Z: CLS
PRINT "------------------------------------------------------------------------"
PRINT " k       P(k)          SP(k)        L(k)          P(k)L(k)     SP(k)L(k)"
PRINT "------------------------------------------------------------------------"
FOR I = 0 TO 20
  IF I = 0 THEN
    P = EXP(-2): L = 1: PL = P: SP = P: SPL = PL
  ELSE
P = 2 * P / I: L = 1 / 2 ^ I: PL = P * L: SP = SP + P: SPL = SPL + PL
  END IF
PRINT USING "##   "; I;
PRINT USING "  #.######## "; P; SP; L; PL; SPL
NEXT I
PRINT "..     .........   .........   .........   .........   ........."
PRINT "inf.   0.00000000   1.00000000   0.00000000   0.00000000   0.36787944"
PRINT "------------------------------------------------------------------------"
```

The value $\mathbf{E(1) = e^{-1} = 0.3679}$ can be straightforwardly obtained using the following reasoning: since in the population consisting of **N** individual there exists only one **Aa** heterozygote, it comes out that the probability of none out of the **2N** genes transmitted from one generation to the other being **A** is

$$\mathbf{(1-1/2N)^{2N}} \ ,$$

where **1/2N** is the probability that a randomly chosen gamete contains the **A** allele; since **N** is assumed to be large, it comes out that

$$(1-1/2N)^{2N} \cong e^{-2N/2N} = e^{-1} = 0.3679 = E(1) \ .$$

The following table shows that the approximation is good even for not so large values of **N**. This means that even for relatively small-sized populations (with 50 or more individuals) the probability of loss of the mutant allele after one generation is approximately 0.37 or 37%.

N	$(1-1/2N)^{2N}$	$(1-1/2N)^{2N}/e^{-1}$
10	0.358486	0.974466
50	0.366032	0.994979
100	0.366958	0.997495
150	0.367265	0.998331
200	0.367419	0.998749
250	0.367511	0.998999
300	0.367573	0.999166
350	0.367617	0.999285
400	0.367649	0.999375
450	0.367675	0.999444
500	0.367695	0.999500

Let us suppose now that gene **A** has been transmitted to **k** children of the mutant individual. For this, it is necessary that the **Aa** individual has an offspring number **n** such that $n \geq k$ and that he transmits gene **A** to **k** out of his **n** children. The probability of having **n** children is

$$P_n = e^{-2}2^n/n!$$

and the probability of transmitting gene **A** to **k** among his **n** children is

$$R_{n,k} = n!/[k!(n-k)!].(1/2)^k.(1/2)^{n-k} = n!/[k!(n-k)!].(1/2)^n \ ;$$

therefore, the probability that gene **A** is transmitted to **k** children of the mutant individual is

$$Q_k = \sum_{n=k}^{inf} P_n R_{n,k} = \sum_{n=k}^{inf} e^{-2}2^n/n!.n!/[k!(n-k)!].(1/2)^n = e^{-2}/k! \sum_{n=k}^{inf} 1/(n-k)!$$

$$= e^{-2}/k! \ . \ e = e^{-1}/k! \ .$$

Each time one gene is transmitted to generation 2, the probability that this gene is the mutant is **k/2N**, where **N** is the population size and **k** is the number of **A** genes transmitted to generation 2 by the single mutant individual of the initial population. The probability that not any **A** gene is transmitted to generation 2 is therefore

$$E(2|k) = (1-k/2N)^{2N} \cong e^{-2Nk/2N} = e^{-k} \ .$$

The probability that not any **A** gene reaches generation 2, no matter which is the number **k** of these genes in generation 1, is therefore

$$E(2) = \sum_{k}^{inf} Q_k E(2|k) = \sum_{k}^{inf} e^{-1}e^{-k}/k! = e^{-1} \sum_{k}^{inf} e^{-k}/k! = e^{-1} \sum_{k}^{inf} (e^{-1})^k/k!$$

$$= e^{-1}e^{(e-1)} = e^{E(1)-1} = e^{(0.3679-1)} = e^{-0.6321} = 0.5315 \ .$$

The value **E(2) = 0.5315** can be straightforwardly obtained using the following argument: since **E(1) = 0.3679,** it comes out that the probability of persistence of the gene after one generation is

$$1-E(1) = 1-e^{-1} = 0.6321 \ ;$$

hence, the probability that one randomly chosen gene transmitted to the following generation is the mutant one is

$$(1-e^{-1})/2N = 0.6321/2N \ ;$$

therefore, the probability that not any of the **2N** genes transmitted from generation 1 to generation 2 is **A** is

$$E(2) = (1-0.6321/2N)^{2N} \cong e^{-1.2642N/2N} = e^{-0.6321} = 0.5315 \ .$$

From

$$E(2) = e^{E(1)-1}$$

we determine the general recursion relation

$$E(t+1) = e^{E(t)-1} \ ;$$

evidently, at equilibriun (that is, when t tends to infinity) we have

$$E = e^{E-1} \ , \ lnE = E-1 \ , \ E = 1 \ .$$

This means that after a large number of generations the mutant gene is inexorably eliminated from the population.

The table below was developed by using the recursion relation

$$E(t+1) = e^{E(t)-1}$$

in the BASIC code

```
REM PROGRAM FILENAME EXTINPR5.BAS
DEFDBL A-Z: CLS
PRINT " ------------------------------------------------------------------------"
FOR I = 1 TO 5: PRINT "   t    E(t)   "; : NEXT I: PRINT
PRINT " ------------------------------------------------------------------------"
 P = 0
 FOR I = 1 TO 100
  J = I / 5: P = EXP(P - 1)
  PRINT USING " ###"; I; : PRINT USING "  #.###### "; P;
  IF J - INT(J) = 0 THEN PRINT
NEXT I
PRINT " ------------------------------------------------------------------------"
```

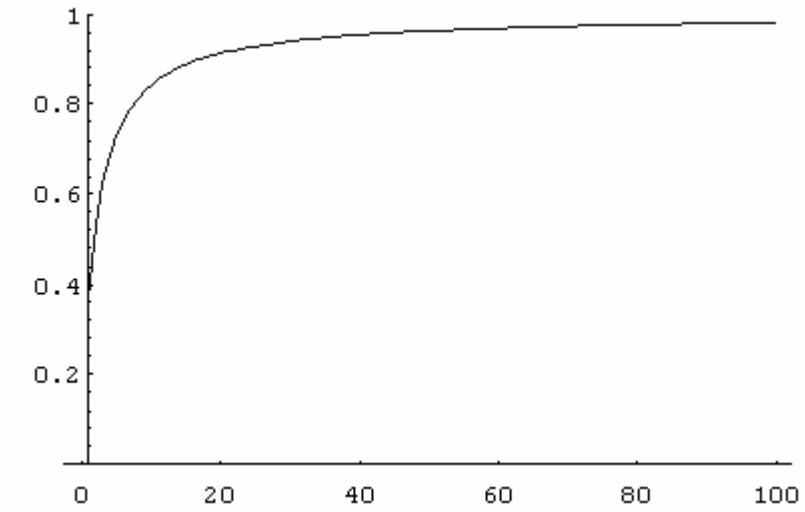
and shows the values of the probabilities of gene extinction [**E(t)**] for several values of **t.**

t	E(t)	t	E(t)	t	E(t)	t	E(t)	t	E(t)
1	0.367879	2	0.531464	3	0.625918	4	0.687920	5	0.731923
6	0.764849	7	0.790452	8	0.810950	9	0.827745	10	0.841765
11	0.853649	12	0.863854	13	0.872716	14	0.880483	15	0.887349
16	0.893463	17	0.898941	18	0.903880	19	0.908355	20	0.912429
21	0.916154	22	0.919573	23	0.922722	24	0.925632	25	0.928330
26	0.930838	27	0.933176	28	0.935360	29	0.937405	30	0.939323
31	0.941128	32	0.942827	33	0.944431	34	0.945947	35	0.947381
36	0.948742	37	0.950033	38	0.951261	39	0.952430	40	0.953544
41	0.954606	42	0.955621	43	0.956591	44	0.957520	45	0.958410
46	0.959263	47	0.960081	48	0.960868	49	0.961623	50	0.962350
51	0.963050	52	0.963725	53	0.964375	54	0.965002	55	0.965607
56	0.966192	57	0.966757	58	0.967303	59	0.967832	60	0.968344
61	0.968840	62	0.969320	63	0.969786	64	0.970238	65	0.970677
66	0.971102	67	0.971516	68	0.971918	69	0.972308	70	0.972688
71	0.973058	72	0.973418	73	0.973768	74	0.974109	75	0.974441
76	0.974765	77	0.975081	78	0.975389	79	0.975689	80	0.975982
81	0.976268	82	0.976548	83	0.976821	84	0.977087	85	0.977348
86	0.977602	87	0.977851	88	0.978095	89	0.978333	90	0.978566
91	0.978794	92	0.979017	93	0.979236	94	0.979450	95	0.979660
96	0.979865	97	0.980067	98	0.980264	99	0.980457	100	0.980647

The graph below, generated by the Mathematica code

```
(* extprob2.ma
   Extinction probability of a mutant allele
   F(1) = e^(-1) = 0.3679
   F(i+1) = e^[F(i)-1]
*)
F[n_] := Exp[F[n-1]-1]; F[1] = N[Exp[-1]];
extinct = Table[F[i], {i,1,100}];
ListPlot[extinct, PlotJoined -> True,
         PlotRange -> {0,1}, AxesOrigin -> {1,0}]
```

shows the probabilities of extinction as function of the number of generations (**t**).



GENETIC DRIFT

Let us consider a population of finite size **N**, kept constant as generations go by; let p_0 and q_0 be the frequencies of a pair of alleles (**A**, **a**) segregating at an autosomal locus in the initial generation 0. Since the population size is constant, individuals belonging to the first generation are produced by random union of **2N** gametes produced by individuals from generation 0, that is,

$$[P_1(aa)+P_1(Aa)+P_1(aa)] = (p_0+q_0)^{2N} \ ;$$

therefore, q_1 can take any of the (**2N+1**) following values :

$$0 = 0/2N \ , \ 1/2N, \ 2/2N, \ ..., \ j/2N, \ .... \ (2N-2)/2N, \ (2N-1)/2N,$$
$$1 = 2N/2N \ .$$

The probability that q_1 takes the particular value $q_1 = j/2N$ is

$$C(2N,j).p^{2N-j}q^{j} = C(2N,2Nq).(1-q)^{2N(1-q)}q^{2Nq}.$$

If, for example, $N = 2$, $2N = 4$, $p_0 = q_0 = 1/2$, it comes out that the possible gene frequencies and population states (**j**) in any subsequent generation will be

p	**1**	**3/4**	**1/2**	**1/4**	**0**
q	**0**	**1/4**	**1/2**	**3/4**	**1**
j	**0**	**1**	**2**	**3**	**4** .

It is easy to see what one means by 'state' : the number of genes a present in the population. Therefore, the probabilities that the population is in states **j = 0, 1, 2, 3** and **4** in the next generation are respectively:

$$0 : (1/2)^4 = 1/16$$
$$1 : 4(1/2)^3(1/2) = 1/4$$
$$2 : 6(1/2)^2(1/2)^2 = 3/8$$
$$3 : 4(1/2)(1/2)^3 = 1/4$$
$$4 : (1/2)^4 = 1/16 \ ,$$

which define the line vector of state probabilities

$$Q(1,j) = (Q(1,0), Q(1,1), Q(1,2), Q(1,3), Q(1,4))$$
$$= (1/16, 1/4, 3/8, 1/4, 1/16).$$

If the population is in state $j = 0$ ($p_1 = 1$, $q_1 = 0$), which takes place with probability **1/16**, the probabilities that the population in next generation is in states **j = 0, 1, 2, 3** and **4** are respectively

$$0 : 1$$
$$1 : 0$$
$$2 : 0$$
$$3 : 0$$
$$4 : 0 \ .$$

If the population is in state **j = 1 (p_1 = 3/4, q_1 = 1/4)**, which takes place with probability **1/4**, the probabilities that the population in next generation is in states **j = 0, 1, 2, 3** and **4** are respectively

$$
\begin{aligned}
&0 : (3/4)^4 &&= 81/256\\
&1 : 4(3/4)^3(1/4) &&= 27/64\\
&2 : 6(3/4)^2(1/4)^2 &&= 27/128\\
&3 : 4(3/4)(1/4)^3 &&= 3/64\\
&4 : (1/4)^4 &&= 1/256 .
\end{aligned}
$$

If the population is in state **j = 2 (p_1 = 1/2, q_1 = 1/2)**, which takes place with probability **3/8**, the probabilities that the population in next generation is in states **j = 0, 1, 2, 3** and **4** are respectively

$$
\begin{aligned}
&0 : (1/2)^4 &&= 1/16\\
&1 : 4(1/2)^3(1/2) &&= 1/4\\
&2 : 6(1/2)^2(1/2)^2 &&= 3/8\\
&3 : 4(1/2)(1/2)^3 &&= 1/4\\
&4 : (1/2)^4 &&= 1/16 .
\end{aligned}
$$

If the population is in state **j = 3 (p_1 = 1/4, q_1 = 3/4)**, which takes place with probability **1/4**, the probabilities that the population in next generation is in states **j = 0, 1, 2, 3** and **4** are respectively

$$
\begin{aligned}
&0 : (1/4)^4 &&= 1/256\\
&1 : 4(1/4)^3(3/4) &&= 3/64\\
&2 : 6(1/4)^2(3/4)^2 &&= 27/128\\
&3 : 4(1/4)(3/4)^3 &&= 27/64\\
&4 : (3/4)^4 &&= 81/256 .
\end{aligned}
$$

If the population is in state **j = 4 (p_1 = 0 , q_1 = 1)**, which takes place with probability **1/16**, the probabilities that the population in the next generation is in states **j = 0, 1, 2, 3** and **4** are respectively

$$
\begin{aligned}
&0 : 0\\
&1 : 0\\
&2 : 0\\
&3 : 0\\
&4 : 1 .
\end{aligned}
$$

Therefore, the probabilities that the population is in states **j = 0, 1, 2, 3, 4** in the second generation are respectively:

$$
\begin{aligned}
Q(2,0) &= Q(1,0).P(2,0|1,0) + Q(1,1).P(2,0|1,1) + Q(1,2).P(2,0|1,2)\\
&\quad + Q(1,3).P(2,0|1,3) + Q(1,4).P(2,0|1,4)\\
&= 1/16 \times 1 + 1/4 \times 81/258 + 3/8 \times 1/16 + 1/4 \times 1/256\\
&\quad + 1/16 \times 0\\
&= 85/512 = 0.166016 ;
\end{aligned}
$$

Q(2,1) = Q(1,0).P(2,1|1,0) + Q(1,1).P(2,1|1,1) + Q(1,2).P(2,1|1,2)
+ Q(1,3).P(2,1|1,3) + Q(1,4).P(2,1|1,4)
= 1/16 × 0 + 1/4 × 27/64 + 3/8 × 1/4 + 1/4 × 3/64
+ 1/16 × 0
= 27/128 = 0.210938 ;

Q(2,2) = Q(1,0).P(2,2|1,0) + Q(1,1).P(2,2|1,1) + Q(1,2).P(2,2|1,2)
+ Q(1,3).P(2,2|1,3) + Q(1,4).P(2,2|1,4)
= 1/16 × 0 + 1/4 × 27/128 + 3/8 × 1/4 + 1/4 × 27/128
+ 1/16 × 0
= 63/256 = 0.246094 ;

Q(2,3) = Q(1,0).P(2,3|1,0) + Q(1,1).P(2,3|1,1) + Q(1,2).P(2,3|1,2)
+ Q(1,3).P(2,3|1,3) + Q(1,4).P(2,3|1,4)
= 1/16 × 0 + 1/4 × 3/64 + 3/8 × 1/4 + 1/4 × Z7/64
+ 1/16 × 0
= 27/128 = 0.210938 ;

Q(2,4) = Q(1,0).P(2,4|1,0) + Q(1,1).P(2,4|1,1) + Q(1,2).P(2,4|1,2)
+ Q(1,3).P(2,4|1,3) + Q(1,4).P(2,4|1,4)
= 1/16 × 0 + 1/4 × 1/256 + 3/8 × 1/16 + 1/4 × 81/256
+ 1/16 × 1
= 85/512 = 0.166016 ,

which define the line vector of state probabilities

Q(2,j) = (Q(2,0) , Q(2,1) , Q(2,2) , Q(2,3) , Q(2,4))
= (85/512 , 27/128 , 63/256 , 27/128 , 85/512) .

In matrix form, the calculations just performed can be rewritten as

$$(1/16,1/4,3/8,1/4,1/16).\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 81/256 & 27/64 & 27/128 & 3/64 & 1/256 \\ 1/16 & 1/4 & 3/8 & 1/4 & 1/16 \\ 1/256 & 3/64 & 27/128 & 27/64 & 81/256 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

= (85/512,27/128,63/256,27/128,85/512) .

In compressed form, the recursion equation shown above is

Q(1,i).P(2,j|1,i) = Q(2,j) , {j,i = 0,1,2,3,4}

where **P(2,j|1,i)** is a transition matrix of conditional probabilities. Each element of this matrix is to be understood as the probability of the population being in state **j (j = 0,1,2,3,4)** in generation 2 given that the population was in state **i (i = 0,1,2,3,4)** in the previous generation. The result above can be generalized:

Q(n+1,j) = Q(n,i).P(n+1,j|n,i)

and the vectors **Q(3,j), Q(4,j), ..., Q(n,j)** obtained through recursive application of the formula. The table that follows shows the values of the elements of the line vectors **Q(n,j)**, obtained by applying the recursion relation derived above, for **n = 0** to **75**, initial gene frequencies $\mathbf{p_0 = q_0 = 1/2}$ and constant population size **N = 2**.

n	Q(n, 0)	Q(n, 1)	Q(n, 2)	Q(n, 3)	Q(n, 4)
0	0.0000000000	0.0000000000	1.0000000000	0.0000000000	0.0000000000
1	0.0625000000	0.2500000000	0.3750000000	0.2500000000	0.0625000000
2	0.1660156250	0.2109375000	0.2460937500	0.2109375000	0.1660156250
3	0.2489624023	0.1604003906	0.1812744141	0.1604003906	0.2489624023
4	0.3116703033	0.1205062866	0.1356468201	0.1205062866	0.3116703033
5	0.3587478995	0.0903990269	0.1017061472	0.0903990269	0.3587478995
6	0.3940604720	0.0678010806	0.0762768947	0.0678010806	0.3940604720
7	0.4205453116	0.0508509802	0.0572074164	0.0508509802	0.4205453116
8	0.4404089797	0.0381382511	0.0429055384	0.0381382511	0.4404089797
9	0.4553067344	0.0286036898	0.0321791516	0.0286036898	0.4553067344
10	0.4664800508	0.0214527675	0.0241343635	0.0214527675	0.4664800508
11	0.4748600381	0.0160895756	0.0181007726	0.0160895756	0.4748600381
12	0.4811450286	0.0120671817	0.0135755794	0.0120671817	0.4811450286
13	0.4858587714	0.0090503863	0.0101816846	0.0090503863	0.4858587714
14	0.4893940786	0.0067877897	0.0076362634	0.0067877897	0.4893940786
15	0.4920455589	0.0050908423	0.0057271976	0.0050908423	0.4920455589
16	0.4940341692	0.0038181317	0.0042953982	0.0038181317	0.4940341692
17	0.4955256269	0.0028635988	0.0032215486	0.0028635988	0.4955256269
18	0.4966442202	0.0021476991	0.0024161615	0.0021476991	0.4966442202
19	0.4974831651	0.0016107743	0.0018121211	0.0016107743	0.4974831651
20	0.4981123738	0.0012080807	0.0013590908	0.0012080807	0.4981123738
21	0.4985842804	0.0009060606	0.0010193181	0.0009060606	0.4985842804
22	0.4989382103	0.0006795454	0.0007644886	0.0006795454	0.4989382103
23	0.4992036577	0.0005096591	0.0005733664	0.0005096591	0.4992036577
24	0.4994027433	0.0003822443	0.0004300248	0.0003822443	0.4994027433
25	0.4995520575	0.0002866832	0.0003225186	0.0002866832	0.4995520575
26	0.4996640431	0.0002150124	0.0002418890	0.0002150124	0.4996640431
27	0.4997480323	0.0001612593	0.0001814167	0.0001612593	0.4997480323
28	0.4998110242	0.0001209445	0.0001360625	0.0001209445	0.4998110242
29	0.4998582682	0.0000907084	0.0001020469	0.0000907084	0.4998582682
30	0.4998937011	0.0000680313	0.0000765352	0.0000680313	0.4998937011
31	0.4999202759	0.0000510235	0.0000574014	0.0000510235	0.4999202759
32	0.4999402069	0.0000382676	0.0000430510	0.0000382676	0.4999402069
33	0.4999551552	0.0000287007	0.0000322883	0.0000287007	0.4999551552
34	0.4999663664	0.0000215255	0.0000242162	0.0000215255	0.4999663664
35	0.4999747748	0.0000161441	0.0000181622	0.0000161441	0.4999747748
36	0.4999810811	0.0000121081	0.0000136216	0.0000121081	0.4999810811
37	0.4999858108	0.0000090811	0.0000102162	0.0000090811	0.4999858108
38	0.4999893581	0.0000068108	0.0000076622	0.0000068108	0.4999893581
39	0.4999920186	0.0000051081	0.0000057466	0.0000051081	0.4999920186
40	0.4999940139	0.0000038311	0.0000043100	0.0000038311	0.4999940139
41	0.4999955105	0.0000028733	0.0000032325	0.0000028733	0.4999955105
42	0.4999966328	0.0000021550	0.0000024244	0.0000021550	0.4999966328
43	0.4999974746	0.0000016162	0.0000018183	0.0000016162	0.4999974746
44	0.4999981060	0.0000012122	0.0000013637	0.0000012122	0.4999981060
45	0.4999985795	0.0000009091	0.0000010228	0.0000009091	0.4999985795
46	0.4999989346	0.0000006818	0.0000007671	0.0000006818	0.4999989346
47	0.4999992010	0.0000005114	0.0000005753	0.0000005114	0.4999992010
48	0.4999994007	0.0000003835	0.0000004315	0.0000003835	0.4999994007
49	0.4999995505	0.0000002877	0.0000003236	0.0000002877	0.4999995505
50	0.4999996629	0.0000002157	0.0000002427	0.0000002157	0.4999996629
51	0.4999997472	0.0000001618	0.0000001820	0.0000001618	0.4999997472
52	0.4999998104	0.0000001214	0.0000001365	0.0000001214	0.4999998104
53	0.4999998578	0.0000000910	0.0000001024	0.0000000910	0.4999998578
54	0.4999998933	0.0000000683	0.0000000768	0.0000000683	0.4999998933
55	0.4999999200	0.0000000512	0.0000000576	0.0000000512	0.4999999200

56	0.4999999400	0.0000000384	0.0000000432	0.0000000384	0.4999999400
57	0.4999999550	0.0000000288	0.0000000324	0.0000000288	0.4999999550
58	0.4999999663	0.0000000216	0.0000000243	0.0000000216	0.4999999663
59	0.4999999747	0.0000000162	0.0000000182	0.0000000162	0.4999999747
60	0.4999999810	0.0000000121	0.0000000137	0.0000000121	0.4999999810
61	0.4999999858	0.0000000091	0.0000000103	0.0000000091	0.4999999858
62	0.4999999893	0.0000000068	0.0000000077	0.0000000068	0.4999999893
63	0.4999999920	0.0000000051	0.0000000058	0.0000000051	0.4999999920
64	0.4999999940	0.0000000038	0.0000000043	0.0000000038	0.4999999940
65	0.4999999955	0.0000000029	0.0000000032	0.0000000029	0.4999999955
66	0.4999999966	0.0000000022	0.0000000024	0.0000000022	0.4999999966
67	0.4999999975	0.0000000016	0.0000000018	0.0000000016	0.4999999975
68	0.4999999981	0.0000000012	0.0000000014	0.0000000012	0.4999999981
69	0.4999999986	0.0000000009	0.0000000010	0.0000000009	0.4999999986
70	0.4999999989	0.0000000007	0.0000000008	0.0000000007	0.4999999989
71	0.4999999992	0.0000000005	0.0000000006	0.0000000005	0.4999999992
72	0.4999999994	0.0000000004	0.0000000004	0.0000000004	0.4999999994
73	0.4999999995	0.0000000003	0.0000000003	0.0000000003	0.4999999995
74	0.4999999997	0.0000000002	0.0000000002	0.0000000002	0.4999999997
75	0.4999999997	0.0000000002	0.0000000002	0.0000000002	0.4999999997

For generating this table the following BASIC code (that can perform the calculations for any population size, any initial gene frequencies, and any number of generations) was used:

```
REM GENDRIF3.BAS
CLS : DEFDBL A-Z: DEFINT I-L
INPUT "POPULATION SIZE = "; N: K = 2 * N: L = K + 1
INPUT "NUMBER OF GENERATIONS = "; NGEN
DIM Q1(L), Q2(L), A(L, L)
PRINT "INITIAL VECTOR OF POPULATION FREQUENCIES AT GENERATION 0"
FOR I = 1 TO L
  PRINT USING "STATE J = ## : q = "; I - 1;
  PRINT USING "#.####"; (I - 1) / K
  PRINT USING "PROBABILITY OF POPULATION BEING AT STATE ## = "; I - 1;
  INPUT Q1(I)
NEXT I: PRINT
A(1, 1) = 1: A(L, L) = 1
FOR J = 2 TO K
  Q = (J - 1) / K: P = 1 - Q: A(J, 1) = P ^ K: A(J, L) = Q ^ K
    FOR I = 2 TO L - 1
      A(J, I) = (L - (I - 1)) * Q * A(J, I - 1) / ((I - 1) * P)
    NEXT I
NEXT J
PRINT "-------------------------------------------------------------------------"
PRINT "  n   ";
FOR I = 1 TO L
  PRINT USING " Q( n,##)     "; I - 1;
NEXT I: PRINT
PRINT "-------------------------------------------------------------------------"
PRINT USING "### "; 0;
FOR I = 1 TO L
  PRINT USING " #.########## "; Q1(I);
NEXT I: PRINT
FOR I1 = 2 TO NGEN + 1
  FOR J = 1 TO L
    FOR I = 1 TO L
      Q2(J) = Q2(J) + Q1(I) * A(I, J)
    NEXT I
  NEXT J
  FOR I = 1 TO L
    Q1(I) = Q2(I): Q2(I) = 0
  NEXT I
  PRINT USING "### "; I1 - 1;
  FOR I = 1 TO L
    PRINT USING " #.########## "; Q1(I);
  NEXT I: PRINT
  DO: LOOP WHILE INKEY$ <> " "
```

```
NEXT I1
PRINT "------------------------------------------------------------------------"
```

The average gene frequency, that is, the mean value of all possible gene frequencies in a given generation, is a constant quantity (**q**):

$$q_1 = 1/16 \times 0 + 1/4 \times 1/4 + 3/8 \times 1/2 + 1/4 \times 3/4 + 1/16 \times 1$$
$$= 8/16 = q_0$$
$$q_2 = 85/512 \times 0 + 27/128 \times 1/4 + 63/256 \times 1/2 + 27/128 \times 3/4$$
$$+ 85/512 \times 1 = 256/512 = q_1 = q_0 = q$$

whereas its variance increases as generations go by:

$$V(q_0) = (0-q_0)^2.0 + (1/4-q_0)^2.0 + (1/2-q_0)^2.1 + (3/4-q_0)^2.0$$
$$+ (1-q_0)^2.0 = 0$$
$$V(q_1) = (0-q_0)^2.1/16 + (1/4-q_0)^2.1/4 + (1/2-q_0)^2.3/8$$
$$+ (3/4-q_0)^2.1/4 + (1-q_0)^2.1/16 = 1/16$$
$$= (1/2 \times 1/2)/4 = q_0(1-q_0)/2N$$
$$V(q_{inf}) = (0-q_0)^2.1/2 + (1/4-q_0)^2.0 + (1/2-q_0)^2.0 + (3/4-q_0)^2.0$$
$$+ (1-q_0)^2.1/2 = 1/4 = 1/2 \times 1/2 = q_0(1-q_0) \ .$$

It is not difficult to derive a formula for the variance as a function of **n**. In fact, we are searching a function that is **0** when **n** is **0** and $q_0(1-q_0)$ when **n** tends to infinity. A simple function with this property is $f(n) = (1-x^n).q_0(1-q_0)$, $0 < x < 1$; all we have to do now is to equate this solution to the particular known value

$$V(q_1) = q_0(1-q_0)/2N = (1-x^1).q_0(1-q_0) \ ;$$

from this equation it comes out that

$$1-x = 1/2N \ , \ x = 1 - 1/2N \ \text{ and}$$
$$V(q_n) = q_0(1-q_0)[1-(1-1/2N)^n] \ .$$

In the lines that follow we present a formal derivation for this formula. The formula for the variance of the gene frequency among populations in the first generation (q_1) is obviously the usual formula for the binomial variance of the mean,

$$V(q_1) = q_0(1-q_0)/2N \ ;$$

since $E(q_1) = q_0 = q$, it comes out that

$$V(q_1) = E(q_1)-q^2 = E(q_1^2)-q_0^2 \ ,$$
$$E(q_1^2) = V(q_1) + q_0^2 = q_0(1-q_0)/2N + q_0^2 \ ;$$

the heterozygote frequency in generation 1 has expected value

$$H_1 = E(2p_1q_1) = 2E(q_1)-2E(q_1^2) = 2q_0-2q_0^2-2q_0(1-q_0)/2N$$
$$= 2q_0(1-q_0)(1-1/2N) = H_0(1-1/2N) \ , \ H_n = H_0(1-1/2N)^n \ .$$

The formula for the variance in generation **n** is

$V(q_n) = E(q_n^2) - q_0^2$;

since

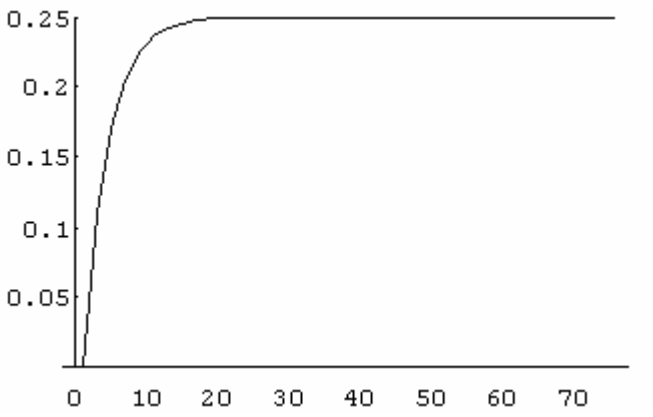
$E(2p_nq_n) = H_n = 2E(q_n) - 2E(q_n^2) = 2q_0 - 2E(q_n^2)$,

it comes out that

$E(q_n^2) = (2q_0 - H_n)/2$ and
$V(q_n) = q_0 - H_n/2 - q_0^2 = q_0(1-q_0)[1-(1-1/2N)^n]$.

The graph below (generated by the appended Mathematica code) shows, for the numerical values worked before, the values the variance takes as generations go by.

```
(*vardrift.ma
  v(q) = q(1-q)[1-(1-1/2N)^n]
*)
q = 0.5; ngen = 2;
F[t_] := q * (1-q) * (1-(1-1/(2 * ngen))^t);
drift = Table[F[i], {i,0,75}];
ListPlot[drift, PlotJoined -> True,
         PlotRange -> {0,0.25},
         AxesOrigin ->{0,0}]
```



SELECTION

Let W_1, W_2 and W_3 be the adaptive values associated respectively to genotypes **AA**, **Aa** and **aa** such that the frequencies of these genotypes in a given generation **n** are, before selection acts on a panmictic population,

$$P(AA) = p_n^2 = p^2$$
$$P(Aa) = 2p_nq_n = 2pq$$
$$P(aa) = q_n^2 = q^2 \ ;$$

and, after selection has acted (that is, frequencies of the same genotypes among the sexually adult individuals that participate in the mating pairs of the population, thus producing offspring for the next generation),

$$P'(AA) = p^2.W_1/W$$
$$P'(Aa) = 2pq.W_2/W$$
$$P'(aa) = q^2.W_3/W \ ,$$

where **W** (normalization factor that makes the sum of the three genotypic frequencies equal to unity) is the average or mean adaptive value of the population:

$$W = p^2.W_1 + 2pq.W_2 + q^2.W_3 =$$
$$= (1-q)^2.W_1 + 2q(1-q).W_2 + q^2.W_3$$
$$= q^2.(W_1-2W_2+W_3) - 2q.(W_1-W_2) + W1 \ .$$

Ignoring the non-relevant normalization factor, the adaptive values are therefore a measure of differential genotypic survival (which has several intrinsic components like viability and fertility) :

$$W_1 = p^2.W_1/p^2 = P'(AA)/P(AA)$$
$$W_2 = 2pq.W_2/2pq = P'(Aa)/P(Aa)$$
$$W_3 = q^2.W_3/q^2 = P'(aa)/P(aa).$$

Assuming that all gametes and embryos with different genotypes are equally viable, the frequencies of alleles **A** and **a** among individuals that could mate and produce offspring are obviously the gene frequencies of the population in the next generation at birth or adult eclosion, before selection acts. Therefore,

$$p' = p_{n+1}$$
$$= P'(AA) + P'(Aa)/2$$
$$= p(p.W_1 + q.W_2)/W$$

and

$$q' = q_{n+1}$$
$$= P'(Aa)/2 + P'(aa)$$
$$= q(p.W_2 + q.W_3)/W$$
$$= [q.W_2 + q^2.(W3-W2)]/[q^2.(W_1-2W_2+W_3)-2q.(W_1-W_2)+W_1]$$

Putting delta q $= \Delta q = q_{n+1}-q_n = q'- q$,

it comes out that

$$\Delta q = [q.W_2 + q^2.(W_3-W_2) - q.W]/W$$

$= q.[W_2 + q.(W_3-W_2) - W]/W$
$= q.[W_2 + q.(W_3-W_2) - q^2.(W_1-2W_2+W_3) + 2q.(W_1-W_2) - W_1]/W$
$= q.[(q-q^2).(W_1-2W_2+W_3)-(1-q).(W_1-W_2)]/W$
$= q(1-q).[q.(W_1-2W_2+W_3)-(W_1-W_2)]/W$.

Since, however, $dW/dq = W' = 2q(W_1-2W_2+W_3)-2(W_1-W_2)$,

it comes out also that

$\Delta q = q(1-q)W'/2W$.

As we shall show in the lines that follow, the above expression is sufficient for the study of all possible situations involving selection operation on genotypes determined by a pair of autosomal alleles.

At equilibium, that is, when no changes in gene frequencies take place when two consecutive generations are considered, by definition,

$\Delta q = 0$ or $q(1-q)W' = 0$.

The equation $q(1-q)W' = 0$ has three sets of solutions, two of them being of trivial determination :

$\{q^* = 0, p^* = 1\}$ and $\{q^* = 1, p^* = 0\}$.

The third set of solutions, valid for $p^* \neq 0$ and $q^* \neq 0$, is obtained from $W' = 0$:

$q^* = (W_1-W_2)/(W_1-2W_2+W_3) = (W_1-W_2)/[(W_1-W_2)+(W_3-W_2)]$,
$p^* = 1 - q^* =$
$= (W_3-W_2)/(W_1-2W_2+W_3) = (W_3-W_2)/[(W_1-W_2)+(W_3-W_2)]$.

Therefore, all possible situations of selection (that is, all possible combinations of W_1, W_2, W_3) lead either to the elimination of one of the two alleles (and consequently to the fixation of the other) or to the polymorphic equilibrium

$q^* = (W_1-W_2)/(W_1-2W_2+W_3)$, $p^* = 1 - q^*$,

in which case the equilibrium gene frequencies are solely determined by the adaptive values W_1, W_2 and W_3.

The inspection of the formula

$p^*/q^* = (W_3-W_2)/(W_1-W_2)$

shows that there is no possible equilibrium (unless the equilibrium frequencies are the trivial solutions $p^* = 0$ or $q^* = 0$) when the adaptive value of heterozygotes (W_2) is scalarly inside a range with limits given by the adaptive values of the two homozygotes. In fact, if $W_1 > W_2 > W_3$ it comes out that $W_1-W_2 > 0$ and $W_3-W_2 < 0$, and hence $p^*/q^* < 0$, what turns out to be an absurdity, since p^* as well as q^* are quantities greater than zero and p^*/q^* must always be greater than zero. A similar argumentation is used in the case $W_1 < W_2 < W_3$, which leads to the same result. Therefore, there only exists an equilibrium with both p^* and $q^* \neq 0$ if the adaptive value of heterozygotes is not scalarly between the ones of homozygotes. We have to consider therefore two cases:

a) $W_1 < W_2 > W_3$
b) $W_1 > W_2 < W_3$.

In both situations the polymorphic equilibrium with **p*** and **q* ≠ 0** is possible, because the quantity **p*/q*** is greater than zero.

We still have to determine the equilibrium stability conditions for the two above-mentioned conditions. This can be done using several elementary analytical methods; some of them are informaly summarized in the lines that follow and that consist in the analyses of the functions

$q_{n+1} = f(q_n)$,
$r_n = (q^*-q_{n+1})/(q^*-q_n)$ and
$\Delta q = q_{n+1}-q_n = q_n.(1-q_n).W'/2W$.

As numerical examples we shall use invariably, for the first case ($W_1 < W_2 > W_3$) the adaptive values $W_1 = 1/3$, $W_2 = 3/3 = 1$, $W_3 = 2/3$, and for the second ($W_1 > W_2 < W_3$) the respective values **2/3**, **1/3** and **3/3= 1**. The table below lists, for several values of q_n, the values of the three above-mentioned functions for the cases $W_1 < W_2 > W_3$ and $W_1 > W_2 < W_3$.

	W1=1/3 < W2=3/3 > W3=2/3			W1=2/3 > W2=1/3 < W3=3/3		
qn	qn+1	rn	Dq	qn+1	rn	Dq
0.00000	0.00000	1.00000	0.00000	0.00000	1.00000	0.00000
0.01667	0.04665	0.95387	0.02998	0.00875	1.02499	-.00791
0.03333	0.08751	0.91445	0.05418	0.01836	1.04991	-.01497
0.05000	0.12369	0.88050	0.07369	0.02883	1.07471	-.02117
0.06667	0.15603	0.85106	0.08936	0.04019	1.09929	-.02648
0.08333	0.18519	0.82540	0.10185	0.05243	1.12360	-.03090
0.10000	0.21168	0.80292	0.11168	0.06557	1.14754	-.03443
0.11667	0.23592	0.78317	0.11926	0.07961	1.17105	-.03706
0.13333	0.25826	0.76577	0.12492	0.09453	1.19403	-.03881
0.15000	0.27896	0.75041	0.12896	0.11033	1.21641	-.03967
0.16667	0.29825	0.73684	0.13158	0.12698	1.23810	-.03968
0.18333	0.31632	0.72486	0.13298	0.14448	1.25901	-.03885
0.20000	0.33333	0.71429	0.13333	0.16279	1.27907	-.03721
0.21667	0.34943	0.70497	0.13276	0.18188	1.29819	-.03479
0.23333	0.36472	0.69680	0.13139	0.20170	1.31631	-.03163
0.25000	0.37931	0.68966	0.12931	0.22222	1.33333	-.02778
0.26667	0.39329	0.68345	0.12662	0.24339	1.34921	-.02328
0.28333	0.40672	0.67812	0.12339	0.26514	1.36386	-.01819
0.30000	0.41969	0.67358	0.11969	0.28743	1.37725	-.01257
0.31667	0.43225	0.66978	0.11558	0.31018	1.38931	-.00649
0.33333	0.44444	0.66667	0.11111	0.33333	1.40000	0.00000
0.35000	0.45633	0.66421	0.10633	0.35682	1.40930	0.00682
0.36667	0.46796	0.66236	0.10129	0.38057	1.41717	0.01391
0.38333	0.47936	0.66109	0.09602	0.40451	1.42359	0.02118
0.40000	0.49057	0.66038	0.09057	0.42857	1.42857	0.02857
0.41667	0.50162	0.66019	0.08495	0.45267	1.43210	0.03601
0.43333	0.51254	0.66052	0.07921	0.47675	1.43418	0.04342
0.45000	0.52338	0.66135	0.07338	0.50073	1.43485	0.05073
0.46667	0.53414	0.66265	0.06747	0.52455	1.43411	0.05788
0.48333	0.54486	0.66443	0.06152	0.54813	1.43200	0.06480
0.50000	0.55556	0.66667	0.05556	0.57143	1.42857	0.07143
0.51667	0.56626	0.66936	0.04960	0.59437	1.42386	0.07771
0.53333	0.57700	0.67251	0.04366	0.61692	1.41791	0.08358

0.55000	0.58779	0.67612	0.03779	0.63900	1.41079	0.08900
0.56667	0.59865	0.68017	0.03198	0.66060	1.40255	0.09393
0.58333	0.60961	0.68468	0.02628	0.68165	1.39326	0.09831
0.60000	0.62069	0.68966	0.02069	0.70213	1.38298	0.10213
0.61667	0.63191	0.69509	0.01525	0.72200	1.37178	0.10534
0.63333	0.64330	0.70100	0.00997	0.74125	1.35972	0.10792
0.65000	0.65488	0.70740	0.00488	0.75985	1.34689	0.10985
0.66667	0.66667	0.71429	0.00000	0.77778	1.33333	0.11111
0.68333	0.67869	0.72169	-.00464	0.79503	1.31913	0.11170
0.70000	0.69099	0.72961	-.00901	0.81159	1.30435	0.11159
0.71667	0.70357	0.73809	-.01310	0.82747	1.28905	0.11080
0.73333	0.71648	0.74713	-.01686	0.84265	1.27329	0.10932
0.75000	0.72973	0.75676	-.02027	0.85714	1.25714	0.10714
0.76667	0.74337	0.76700	-.02330	0.87095	1.24066	0.10429
0.78333	0.75742	0.77790	-.02591	0.88408	1.22389	0.10075
0.80000	0.77193	0.78947	-.02807	0.89655	1.20690	0.09655
0.81667	0.78693	0.80177	-.02974	0.90837	1.18972	0.09170
0.83333	0.80247	0.81481	-.03086	0.91954	1.17241	0.08621
0.85000	0.81859	0.82867	-.03141	0.93009	1.15502	0.08009
0.86667	0.83534	0.84337	-.03133	0.94004	1.13757	0.07337
0.88333	0.85278	0.85899	-.03055	0.94939	1.12010	0.06606
0.90000	0.87097	0.87558	-.02903	0.95817	1.10266	0.05817
0.91667	0.88997	0.89320	-.02670	0.96641	1.08527	0.04974
0.93333	0.90985	0.91195	-.02348	0.97411	1.06796	0.04078
0.95000	0.93070	0.93190	-.01930	0.98130	1.05076	0.03130
0.96667	0.95261	0.95315	-.01405	0.98800	1.03368	0.02133
0.98333	0.97567	0.97581	-.00766	0.99423	1.01676	0.01089
1.00000	1.00000	1.00000	0.00000	1.00000	1.00000	0.00000

These values were generated by the following BASIC code:

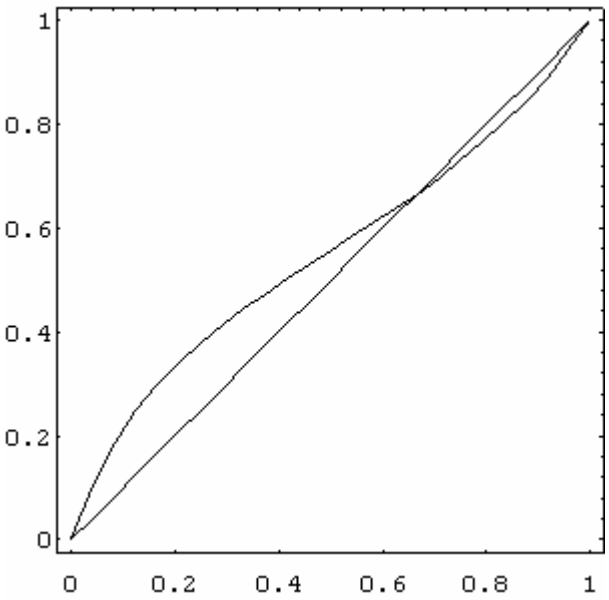
```
REM PROGRAM FILENAME SELECT01.BAS
CLS : DEFDBL A-Z
PRINT "-------------------------------------------------------------------"
PRINT "           W1=1/3  < W2=3/3 >  W3=2/3    W1=2/3  > W2=1/3 <  W3=3/3"
PRINT "   qn      ---------------------------------------------------------"
PRINT "            qn+1      rn        Dq        qn+1      rn        Dq"
PRINT "-------------------------------------------------------------------"
W11 = 1 / 3: W21 = 3 / 3: W31 = 2 / 3: W12 = 2 / 3: W22 = 1 / 3: W32 = 3 / 3
QE1 = (W11 - W21) / (W11 - 2 * W21 + W31)
QE2 = (W12 - W22) / (W12 - 2 * W22 + W32)
FOR I = 0 TO 60: Q = I / 60
Q1 = (Q * W21 + Q * Q * (W31 - W21))
Q1 = Q1 / (Q * Q * (W11 - 2 * W21 + W31) - 2 * Q * (W11 - W21) + W11)
Q2 = (Q * W22 + Q * Q * (W32 - W22))
Q2 = Q2 / (Q * Q * (W12 - 2 * W22 + W32) - 2 * Q * (W12 - W22) + W12)
IF QE1 = Q THEN R1 = (2 * W11 * W31 - W21 * (W11 + W31)) / (W11 * W31 - W21 * W21) ELSE R1 =
(QE1 - Q1) / (QE1 - Q)
IF QE2 = Q THEN R2 = (2 * W12 * W32 - W22 * (W12 + W32)) / (W12 * W32 - W22 * W22) ELSE R2 =
(QE2 - Q2) / (QE2 - Q)
D1 = Q1 - Q: D2 = Q2 - Q
PRINT USING "#.#####   "; Q; Q1; R1; D1; Q2; R2; D2
NEXT I
PRINT "-------------------------------------------------------------------"
```

1) Analysis of the function $q_{n+1} = f(q_n, W_1, W_2, W_3)$

If we plot (see numerical values displayed on table above) the values of q_{n+1} as functions of q_n, the equilibrium value **q*** will lie , for both situations, upon the diagonal uniting the lower left corner to the upper right one, since this diagonal represents the set of all possible loci $q_n = q_{n+1}$.

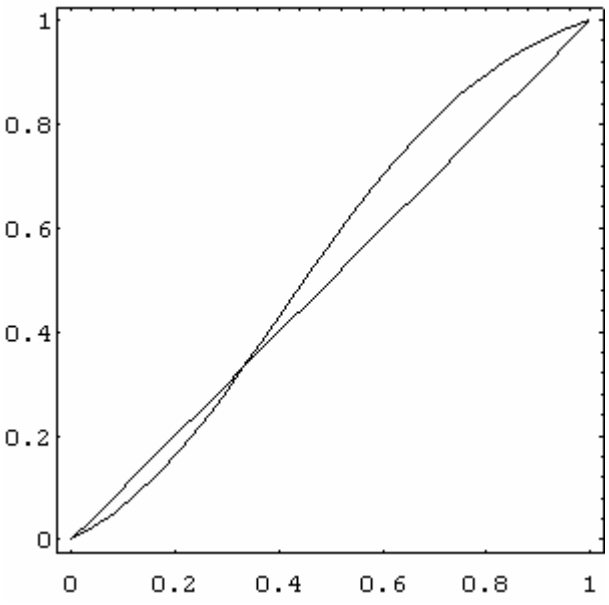
In the case of situation a ($W_1 < W_2 > W_3$), one notes that for any value of $q_n < q^*$, q_{n+1} is above the diagonal, indicating thus that $q_{n+1} > q_n$ and therefore that q_{n+1} is nearer to q^* than q_n. For any value $q_n > q^*$, inversely, $q_{n+1} < q_n$, but again q_{n+1} is nearer to q^* than q_n. This is shown in the graph depicted below (together with the Mathematica code used) and indicates clearly a **stable equilibrium** situation.

```
(*sel_01.ma
q(n+1)=f[q(n),w1,w2,w3]
      =[q(n).w2+q(n)^2.(w3-w2)]/[q(n)^2.(w1-2w2+w3)-2q(n).(w1-w2)+w1]
*)
w1 = 1/3; w2 = 1; w3 = 2/3;
q1 = (q*w2 + q^2*(w3 - w2))/(q^2*(w1 - 2*w2 + w3)-2*q*(w1 - w2) + w1);
q2 = q;
Plot[{q1, q2},{q, 0, 1}, AspectRatio -> 1, Frame-> True]
```



In the case of situation b ($W_1 < W_2 > W_3$), one notes that for any value of $q_n < q^*$, q_{n+1} is below the diagonal, indicating thus that $q_{n+1} < q_n$ and therefore that q_{n+1} is farther from q^* than q_n. For any value $q_n > q^*$, inversely, $q_{n+1} > q_n$, but again q_{n+1} is farther from q^* than q_n. This is shown in the graph depicted below (with the appended Mathematica code) and indicates clearly an **unstable equilibrium** situation.

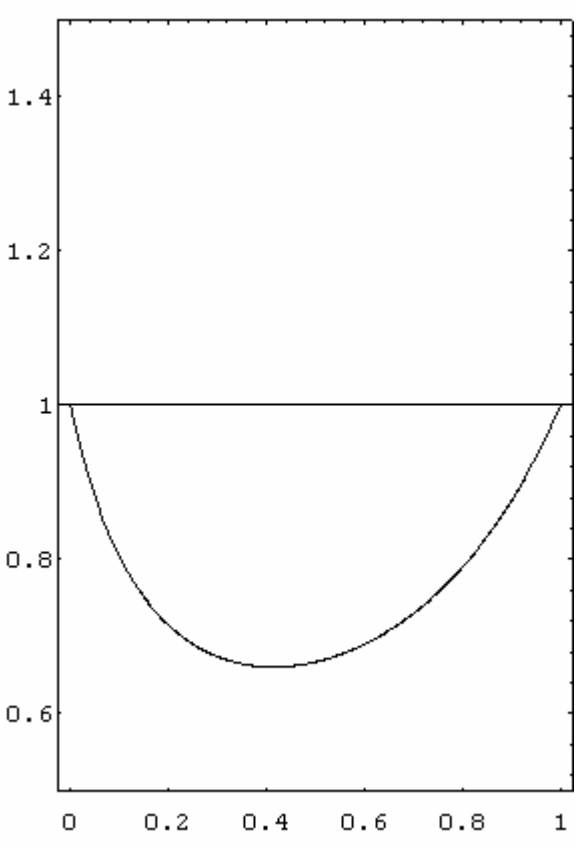
```
(*sel_02.ma
q(n+1)=f[q(n),w1,w2,w3]
      =[q(n).w2+q(n)^2.(w3-w2)]/[q(n)^2.(w1-2w2+w3)-2q(n).(w1-w2)+w1]
*)
w1 = 2/3; w2 = 1/3; w3 = 1;
q1 = (q*w2 + q^2*(w3 - w2))/(q^2*(w1 - 2*w2 + w3)-2*q*(w1 - w2) + w1);
q2=q;
Plot[{q1,q2},{q,0,1}, AspectRatio -> 1, Frame-> True]
```



2) Analysis of the function $r_n = (q^*-q_{n+1})/(q^*-q_n)$

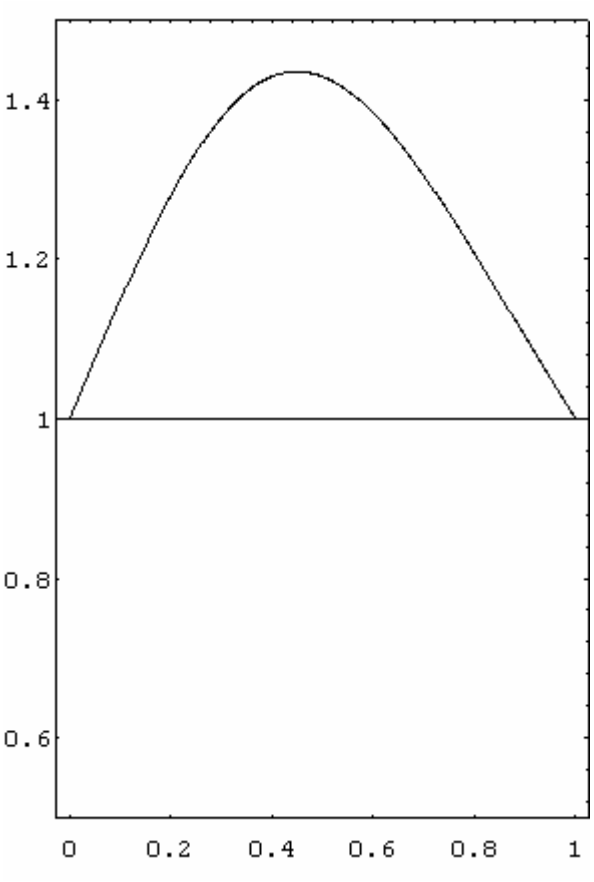
If **{p*, q*}** constitutes a stable equilibrium set, then for any value of q_n in the open interval $1 < q_n < 0$, r_n will always be, in modulus, less than unity. The reason is simple: if equilibrium is stable , in modulus the difference (q^*-q_{n+1}) is always smaller than (q^*-q_n), that is, for any q_n, q_{n+1} will be nearer to **q*** than q_n. This occurs in case **a** ($W_1 < W_2 > W_3$), as depicted in the graph below, together with its Mathematica code.

```
(*sel_03.ma
r(n) = [q-q(n+1)]/[q-q(n)]
q=(w1-w2)/(w1-2w2+w3)
q(n+1)=f[q(n),w1,w2,w3]
      =[q(n).w2+q(n)^2.(w3-w2)]/[q(n)^2.(w1-2w2+w3)-2q(n).(w1-w2)+w1]
*)
w1 = 1/3; w2 = 1; w3 = 2/3;
qe = (w1 - w2)/(w1 - 2*w2 + w3 );
q1 = (q*w2 + q^2*(w3 - w2))/(q^2*(w1 - 2*w2 + w3)-2*q*(w1 - w2) + w1);
r = (qe - q1)/(qe - q);
Plot[r,{q,0,1}, PlotPoints -> 500, Frame -> True,
                PlotRange -> {.5, 1.5},
                AspectRatio -> 1.5, AxesOrigin -> {0,1}]
```



If **{p*, q*}** constitutes, on the contrary, an unstable equilibrium set, then for any value of q_n in the open interval $1 < q_n < 0$, r_n will always be greater than unity. The reason is simple: if equilibrium is unstable then in modulus the difference (q^*-q_{n+1}) will always be greater than (q^*-q_n), that is, for any q_n, q_{n+1} will be farther from **q*** than q_n; and this takes place when $W_1 > W_2 < W_3$, as depicted in the Mathematica graph below.

```
(*sel_04.ma
r(n) = [q-q(n+1)]/[q-q(n)]
q=(w1-w2)/(w1-2w2+w3)
q(n+1)=f[q(n),w1,w2,w3]
      =[q(n).w2+q(n)^2.(w3-w2)]/[q(n)^2.(w1-2w2+w3)-2q(n).(w1-w2)+w1]
*)
w1 = 2/3; w2 = 1/3; w3 = 1;
qe = (w1 - w2)/(w1 - 2*w2 + w3 );
q1 = (q*w2 + q^2*(w3 - w2))/(q^2*(w1 - 2*w2 + w3)-2*q*(w1 - w2) + w1);
r = (qe - q1)/(qe - q);
Plot[r,{q,0,1}, PlotPoints -> 500, Frame -> True,
                PlotRange -> {.5, 1.5},
                AspectRatio -> 1.5, AxesOrigin -> {0,1}]
```



The limit, as **n** tends to infinity (or as q_n tends to **q***), of the expression r_n, is clearly

$$r = q'_{n+1} = dq_{n+1}/dq_n = df(q_n, W_1, W_2, W_3)/dq_n,$$

evaluated at equilibrium point $q^* = (W_1-W_2)/(W_1-2W_2+W_3)$.

In the present case, $r = dq_{n+1}/dq_n$ has the value

$$r = \{[W_2+2q^*(W_3-W_2)].W-[q^*W_2+q^{*2}(W_3-W_2)].W'\}/W^2 \ ;$$

since at equilibrium point $W' = dW/dq = 0$,

the above expression simplifies to

$$r = [W_2+2q^*(W_3-W_2)]/W$$
$$= [2W_1.W_3 - W_2.(W_1+W_3)]/(W_1.W_3 - W_2^2).$$

The analysis of the above expression is sufficient for all conclusions regarding equilibrium conditions. Let us consider first the case $W_1 < W_2 > W_3$, that is when $W_1 = W_2-d_1$ and $W_3 = W_2-d_3$, where d_1 and d_3 are two positive quantities. Making the appropriate substitutions in the formula above, it can be put in the more suitable form

$$r = [2W_1.W_3 - W_2.(W_1+W_3)]/(W_1.W_3 - W_2^2)$$
$$= [(W_1.W_3-W_1.W_2)+(W_1.W_3-W_2.W_3)]/[W_1.W_3 - (W_1+d_1)(W_3+d_3)]$$
$$= [W_1.(W_3-W_2) + W_3(W_1-W_2)]/(W_1.W_3-W_1.W_3-W_1.d_3-W3.d_1-d_1.d_3)$$

$= (W_1.d_3+W_3.d_1)/(W_1.d_3+W_3.d_1+d_1.d_3);$

since $d_1 > 0$ and $d_3 > 0$, it comes out that, for any possible values of W_1, W_2 and W_3 subject to the constraint $W_1 < W_2 > W_3$, **r** is always smaller than unity. This is precisely the condition for the equilibrium set being stable.

Now, if $W_1 > W_2 < W_3$, we put $W_1 = W_2+d_1$ and $W_3 = W_2+d_3$, where, as before, d_1 and d_3 are positive quantities. Making the appropriate substitutions on the formula for **r** as a function of W_1, W_2 and W_3 we obtain

$$r = [2W_1.W_3 - W_2.(W_1+W_3)]/(W_1.W_3 - W_2^2)$$
$$= [(W_1.W_3-W_1.W_2)+(W_1.W_3-W_2.W_3)]/[W_1.W_3 - (W_1-d_1)(W_3-d_3)]$$
$$= [W_1.(W_3-W_2) + W_3(W_1-W_2)]/(W_1.W_3-W_1.W_3+W_1.d_3+W3.d_1-d_1.d_3)$$
$$= (W_1.d_3+W_3.d_1)/(W_1.d_3+W_3.d_1-d_1.d_3)$$

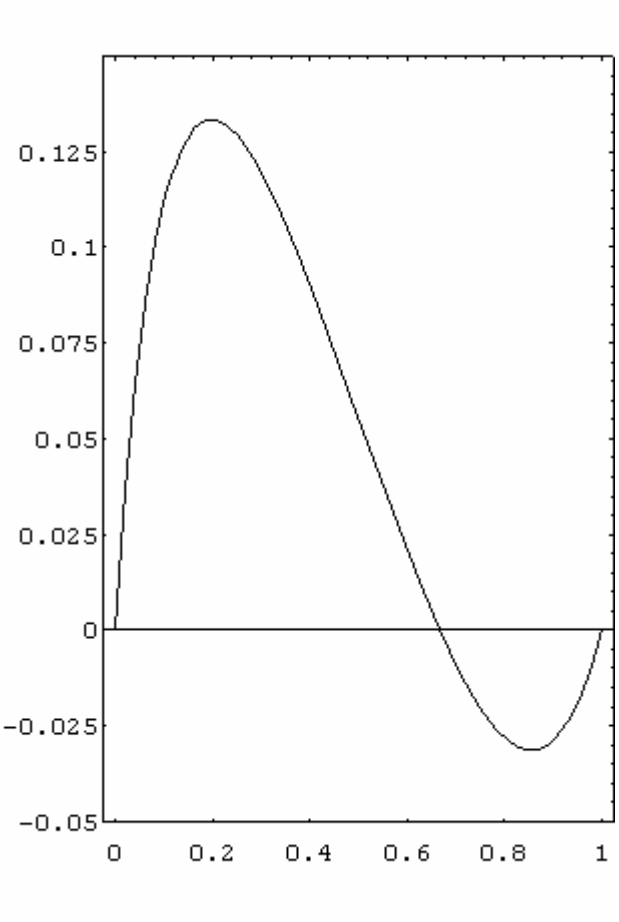
since $d_1 > 0$ and $d_3 > 0$, it comes out $W_1.d_3 + W_3.d_1$ is always greater than $W_1.d_3 + W_3.d_1 - d_1.d_3$ and therefore **r** is always greater than unity in the case $W_1 > W_2 < W_3$ and the equilibrium that can occur in this situation is unstable.

3) Analysis of the function $\Delta q = q_{n+1} - q_n$

If, for any value of $q_n < q^*$, Δq has a positive value and for any value of $q_n > q^*$, Δq is negative, the equilibrium $\{p^*, q^*\}$ is stable, since this implies that, in both cases, q_{n+1} is nearer to q^* than q_n. Inversely, if for any value of $q_n < q^*$, Δq has a negative sign and, for any value of $q_n > q^*$, Δq is positive, this implies that, in both cases, q_{n+1} is farther from q^* than q_n and, consequently, that the equilibrium is unstable.

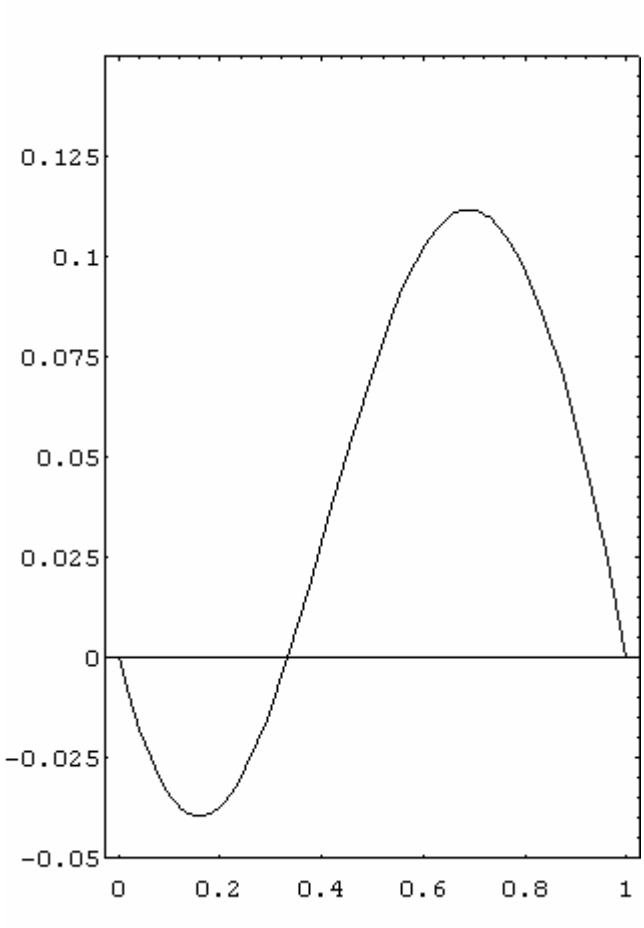
It is easy to verify that the first situation occurs, again, when $W_1 < W_2 > W_3$. This is depicted in the graph below.

```
(*sel_05.ma
dq = q(n+1) - q(n)
q(n+1)=f[q(n),w1,w2,w3]
      =[q(n).w2+q(n)^2.(w3-w2)]/[q(n)^2.(w1-2w2+w3)-2q(n).(w1-w2)+w1]
*)
w1 = 1/3; w2 = 1; w3 = 2/3;
q1 = (q*w2 + q^2*(w3 - w2))/(q^2*(w1 - 2*w2 + w3)-2*q*(w1 - w2) + w1);
dq = q1 - q;
Plot[dq,{q,0,1}, Frame -> True,
               PlotRange -> {-0.05, 0.15},
               AspectRatio -> 1.5, AxesOrigin -> {0,0}]
```



The unstable equilibrium case takes place, as expected, when $W_1 > W_2 < W_3$ and this is shown in the following Mathematica graph:

```
(*sel_06.ma
dq = q(n+1) - q(n)
q(n+1)=f[q(n),w1,w2,w3]
      =[q(n).w2+q(n)^2.(w3-w2)]/[q(n)^2.(w1-2w2+w3)-2q(n).(w1-w2)+w1]
*)
w1 = 2/3; w2 = 1/3; w3 = 1;
q1 = (q*w2 + q^2*(w3 - w2))/(q^2*(w1 - 2*w2 + w3)-2*q*(w1 - w2) + w1);
dq = q1 - q;
Plot[dq,{q,0,1}, Frame -> True,
               PlotRange -> {-0.05, 0.15},
               AspectRatio -> 1.5, AxesOrigin -> {0,0}]
```



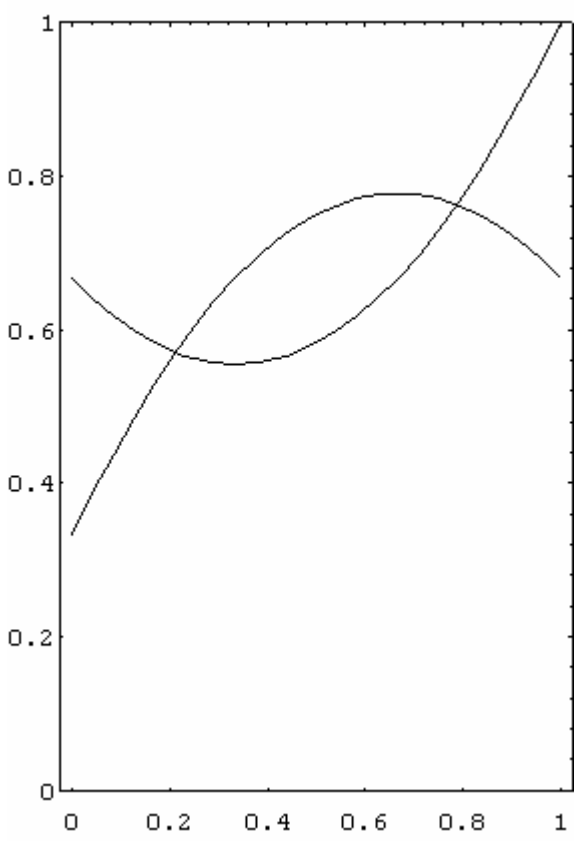
The analysis of the function $W = q^2.(W_1-2W_2+W_3) - 2q.(W_1-W_2) + W_1$ is also useful for determining if the equilibrium is stable or not. The action of selection is clearly to increase the average adaptive value of the population. Therefore, at equilibrium **W** should be at a maximum in the case of stable equilibrium. The equilibrium point **q*** was obtained by putting **dW/dq** equal to zero, therefore at **q*** **W** is necessarily an extremum point. To investigate if this extremum is a maximum or a minimum, we should observe the sign of the second derivative of **W**, $W'' = d^2W/dq^2 = d(dW/dq)/dq$, at point **q***. The second derivative **W"** has a value

$$W'' = 2(W_1-2W_2+W_3) \ ;$$

and **W"** is greater than zero if $W_1+W_3 > 2W_2$ or $(W_1+W_3)/2 > W_2$, smaller than zero if $W_1+W_3 < 2W_2$ or $(W_1+W_3)/2 < W_2$. The first situation occurs when $W_1 > W_2 < W_3$ and **W" > 0** implies that the extremum is a minimum (unstable equilibrium). When $W_1 < W_2 > W_3$, **W" < 0** and the extremum is a maximum (stable equilibrium). The Mathematica graph below, using the same numerical values as before, shows the variation of **W** as a function of **q** for the two cases just considered.

```
(*sel_07.ma
w = f[q(n),w1,w2,w3]
  = q(n)^2.(w1-2w2+w3)-2q(n).(w1-w2)+w1
*)
w11 = 1/3; w21 = 1; w31 = 2/3;
w1= q^2*(w11 - 2*w21 + w31)-2*q*(w11 - w21) + w11;
```

```
w12 = 2/3; w22 = 1/3; w32 = 1;
w2 = q^2*(w12 - 2*w22 + w32)-2*q*(w12 - w22) + w12;
Plot[{w1,w2}, {q,0,1}, Frame -> True,
                PlotRange -> {0,1}, AspectRatio -> 1.5]
```



In the lines that follow we consider the study of some particular cases. To characterize each case, we shall use a scale, such as

$$|----------|----------|----------|$$
$$0 \qquad W_1 \qquad W_2 \qquad W_3=1 \ ,$$

which in the example depicted represents the situation $W_1 < W_2 < W_3$, where the largest adaptive value is represented by unity. This is accomplished by dividing all adaptive values (expressed in any scale) by the largest of them. The resulting figures are the so-called relative adaptive values or simply "fitnesses" : $W_1 = W_1'/W_3'$, $W_2 = W_2'/W_3'$, $W_3 = W_3'/W_3' = 1$, in the example above. The complementary quantities $s_1 = 1-W_1$, $s_2 = 1-W_2$ and $s_3 = 1-W_3$ receive the name of coefficients of selection. It should be kept in mind that W_1, W_2 and W_3 are the adaptive values associated to genotypes **AA**, **Aa** and **aa** respectively.

A) Complete selection against recessive homozygotes

$$|--------------------------------|$$
$$W_3=0 \qquad\qquad W_1=W_2=1$$

This situation, characterized by the first-order linear difference equation

$q_{n+1} = q_n/(1+q_n)$,

admits the general solution

$q_n = q_0/(1 + n.q_0)$

or

$n = (q_0 - q_n)/(q_0.q_n) = 1/q_n - 1/q_0 = q_n^{-1} - q_0^{-1}$.

At equilibrium, **p* = 1** and **q* = 0**.

B) Partial selection against recessive homozygotes

```
|----------|---------------------|
0        W3=1-s                W1=W2=1
```

The recurrence equation is then

$q_{n+1} = (q_n - s.q_n^2)/(1-s.q_n^2)$,

which is non-linear and therefore admits no general solution in simple analytical form. For large values of n and small values of s, however, the equation $\Delta q = q_{n+1} - q_n$ can be substituted by the differential equation

$dq/dt = (q-s.q^2)/(1-s.q^2) - q = -s.q^2(1-q)/(1-s.q^2)$

that can be integrated from **t = 0** to **t = n** giving the result

$ns = (q_0-q_n)/(q_0.q_n) + \ln[q_0.(1-q_n)] - \ln[q_n.(1-q_0)]$.

Again, at equilibrium, **p* = 1** and **q* = 0**.

C) Complete selection against heterozygotes

```
|--------------------------------|
W2 = 0                           W1=W3=1
```

The recurrence equation is then

$q_{n+1} = q_n^2/[(1-q_n)^2 + q_n^2]$,

which, in spite of being non-linear, has the exact general solution in simple analytic form

$q_n = q_0^m/[(1-q_0)^m + q_0^m]$, $m = 2^n$.

There are 3 possible equilibrium sets **{p*,q*} : {1,0} , {1/2,1/2}** and **{0,1}**. The first and the last are stable ones while the second is unstable. Convergence occurs to the first set if **p(0) > q(0)**, to the second if **p(0) = q(0)** , to the third if **p(0) < q(0)**.

D) Complete selection against heterozygotes, partial against one of the homozygotes

```
|----------|---------------------|
W2 = 0   W1=w                  W3=1
```

The recurrence equation is given by

$$p_{n+1} = p_n^2/[p_n^2.w + (1-p_n)^2] ,$$

which, in spite of being non-linear, has the exact general solution in simple analytic form

$$p_n = (p_0.w)^m/[(p_0.w)^m + w.(1-p_0)^m], \; m = 2^n .$$

There are 3 possible equilibrium sets **{p*,q*} : {1,0} , {1/2,1/2}** and **{0,1}**. The first and the last are stable ones while the second is unstable. Convergence occurs to the first set if **w > 1**, to the second if **w = 1** and **p(0) = q(0)** , to the third if **w < 1**.

E) Partial selection against heterozygotes

```
|----------|---------------------|
0       W2=1-s                 W1=W3=1
```

The recurrence equation is

$$q_{n+1} = [q_n - s.q_n(1-q_n)]/[1-2s.q_n(1-q_n)] .$$

This equation admits no exact general solution in simple analytical form but the equilibrium sets are the same ones that occurred in the previous case, with convergence to each set depending again on the initial conditions **p(0), q(0).**

F) Selection favouring heterozygotes

```
|----------|----------|----------|
0       W1=1-s1   W3=1-s3       W2=1

|----------|----------|----------|
0       W3=1-s3   W1=1-s1       W2=1
```

The recurrence equation is

$$q_n = (q_n - s_3.q_n^2]/[1-s_1.(1-q_n)^2 - s_3.q_n^2]$$

and the equilibrium points are

q* = 0 or **1** (unstable points) and

$$q^* = s_1/(s_1+s_3)$$ (stable equilibrium point).

G) Selection favouring heterozygotes, complete against one of the homozygotes

```
|--------------------|----------|
W3=0               W1=1-s       W2=1
```

The recurrence equation is

$q_{n+1} = q_n/[1-s+q_n.(1+s)]$,

which turns out to be a fractional first-order difference equation with the exact general solution in simple analytical form:

$q_n = s.q_0/\{s(1-s)^n + q_0.(1+s)[1-(1-s)^n]\}$.

The equilibrium points (see previous case) are

$q^* = 0$, $q^* = 1$ and $q^* = s/(1+s)$.

H) Partial selection against dominants

```
|---------------------|----------|
0                  W1=W2=1-s      W3=1
```

The non-linear recurrence equation is

$q_{n+1} = [q_n - s.q_n.(1-q_n)]/[1-s(1-q_n^2)]$

and the only possible equilibrium set $\{p^*,q^*\}$ given by $p^* = 0$, $q^* = 1$.

I) Partial selection against heterozygotes, total against dominant homozygotes

```
|---------------------|----------|
W1=0                W2=1-s        W3=1
```

The recurrence equation for gene frequency is given by

$q_{n+1} = [1-s(1-q_n)]/[1+(1-q_n).(1-2s)]$.

Like some of the previous difference equations, this is a fractional first order recurrence equation which therefore admits an exact general solution in simple analytical form:

$p_n = 1-q_n = s.p_0.(1-s)^n/\{s + p_0.(1-2s).[1-(1-s)^n]\}$.

It is not difficult to show that the only stable equilibrium set is given by $\{p^* = 0 , q^* = 1\}$.

J) $W_1 : W_2 : W_3 :: w^2 : w^1 = w : w^0 = 1$ (adaptive values in geometric progression)

```
|---------------|---------------|---------------|
0         W1=w^2=(1-s)^2      W2=w=1-s          W3=w^0=1
```

The recurrence equation in this case is the fractional first-order difference equation

$q_{n+1} = q_n/[1 - s(1 - q_n)]$

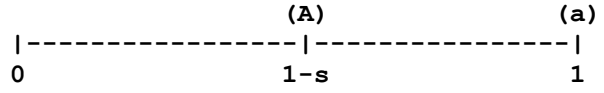
with exact general solution in simple analytic form

$q_n = q_0/[q_0 + w^n.(1-q_0)]$.

There are two stable equilibrium sets : $\{p^* = 1, q^* = 0\}$ and $\{p^* = 0, q^* = 1\}$. Convergence occurs to the first case if $w > 1$, to the second case

if $w < 1$. This takes place because in the first instance $W_1 > W_2 > W_3$ and in the second $W_1 < W_2 < W_3$.

K) Gametic selection



The recurrence equation is given by

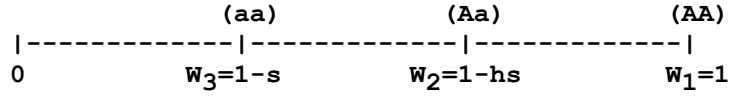
$$q_{n+1} = q_n.(1-s)/(1 - s.q_n) ,$$

which has the exact general solution in simple analytic form

$$q_n = q_0.(1-s)^n/\{1 - q(0)[1 - (1-s)^n]\} .$$

The only equilibrium set is given by $\{p^* = 1 , q^* = 0\}$.

All the cases thus far studied and even the general formulation using W_1, W_2, W_3 can be analyzed using the notation that follows:



The quantity h is of course to be understood as a dominance measurement in relation to adaptive values:

$$h = hs/s = (1-W_2)/(1-W_3) .$$

Using this notation, the average adaptive value of the population has value

$$W = p_n^2 + 2p_nq_n.(1-hs) + q_n^2.(1-s)$$
$$= 1 - 2hs.q_n - s(1-2h).q_n^2$$

and its first derivative has value

$$W' = dW/dq_n = -2hs - 2q_n.s(1-2h) .$$

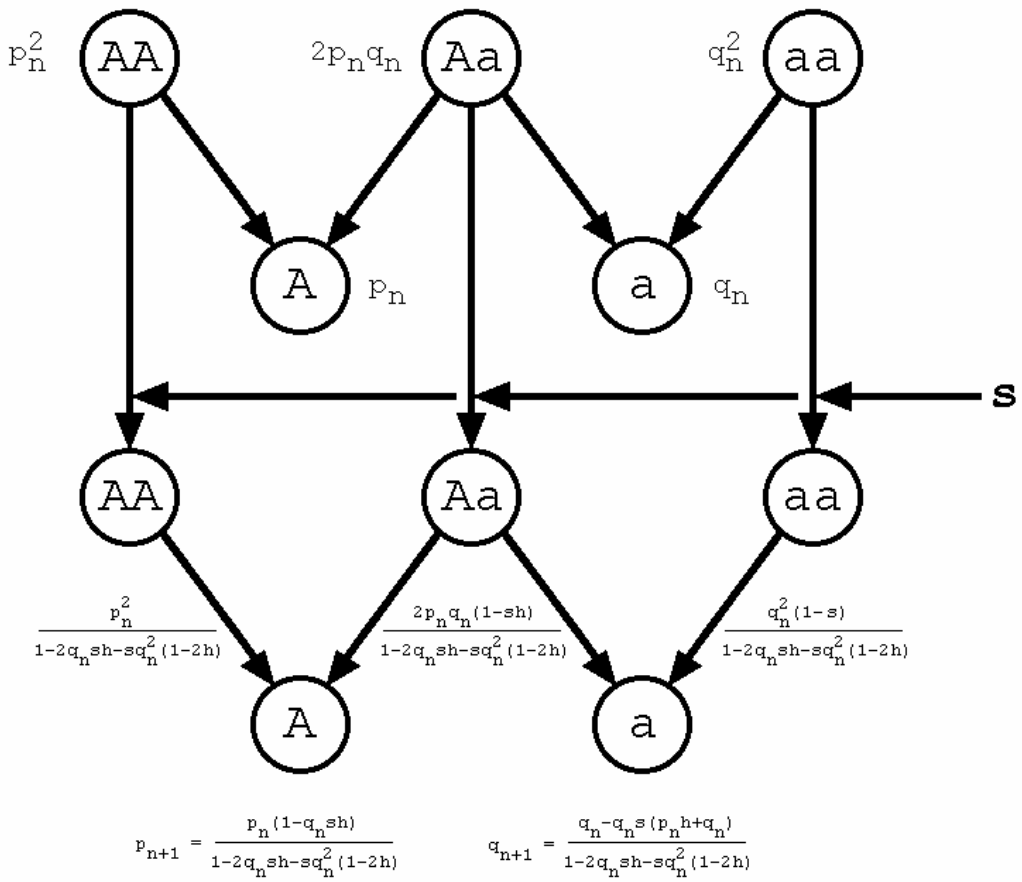
The recurrence equation is given by

$$q_{n+1} = q_n.[1 - hs - q_n.s(1-h)]/W$$

and $\Delta q = q_{n+1} - q_n$ by

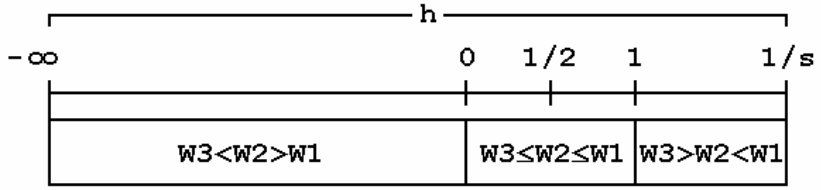
$$\Delta q = -q(1-q).[hs+qs(1-2h)]/[1-2hsq-s(1-2h)q^2]$$
$$= q(1-q).W'/2W ,$$

exactly like the result obtained in the general formulation using W_1, W_2, W_3 .



$$\Delta q = q_{n+1} - q_n = -\frac{q(1-q)[hs+qs(1-2h)]}{1-2qsh-sq^2(1-2h)}$$

$$\Delta q = 0 \quad \Rightarrow \quad q = 0 \qquad q = 1 \qquad q = h/(2h-1) \Rightarrow 0 > h > 1$$



The solutions of $\Delta q = 0$ are $q^* = 0$, $q^* = 1$ and $q^* = h/(2h-1)$.

The possible interior equilibrium frequency is given by $q^*=h/(2h-1)$. Since q^* must belong to the interval $(0,1)$, it comes out then that h must be either smaller than zero or greater than 1.

Polymorphic, stable equilibrium sets occur when **h** is smaller than zero **(-inf < h < 0)**. Unstable polymorphic equilibrium sets can occur if **h** is greater than unity **(1/s > h > 1)**. For any value of **h** in the closed interval **(0,1)**, there is no possible internal equilibrium and sets of stable, monomorphic equilibria **{p* = 1, q* = 0}** always take place. It is worth noting that all possible cases are included if adaptive values between homozygotes are interchanged, that is, if we put $W_1 = 1-s$ and $W_3 = 1$ instead of $W_1 = 1$ and $W_3=1-s$. This is shown in the table below, where the correspondence in each instance to the usual notation (W_1, W_2, W_3) is indicated. The value **s = 0.1** was used in all calculations and, for any given value of h, each adaptive value shown (relative adaptive value) was obtained dividing W_1, W_2 and W_3 by the largest of them, that is, $W_1' = W_1/W_{max}$, $W_2' = W_2/W_{max}$, $W_3' = W_3/W_{max}$, where $W_{max} = \max(W_1, W_2, W_3)$.

h	W1=1	W2=1-hs	W3=1-s	W1=1-s	W2=1-hs	W3=1
-inf.	0.00000	1.00000	0.00000	0.00000	1.00000	0.00000
-3.0	0.76923	1.00000	0.69231	0.69231	1.00000	0.76923
-2.8	0.78125	1.00000	0.70313	0.70313	1.00000	0.78125
-2.6	0.79365	1.00000	0.71429	0.71429	1.00000	0.79365
-2.4	0.80645	1.00000	0.72581	0.72581	1.00000	0.80645
-2.2	0.81967	1.00000	0.73770	0.73770	1.00000	0.81967
-2.0	0.83333	1.00000	0.75000	0.75000	1.00000	0.83333
-1.8	0.84746	1.00000	0.76271	0.76271	1.00000	0.84746
-1.6	0.86207	1.00000	0.77586	0.77586	1.00000	0.86207
-1.4	0.87719	1.00000	0.78947	0.78947	1.00000	0.87719
-1.2	0.89286	1.00000	0.80357	0.80357	1.00000	0.89286
-1.0	0.90909	1.00000	0.81818	0.81818	1.00000	0.90909
-0.8	0.92593	1.00000	0.83333	0.83333	1.00000	0.92593
-0.6	0.94340	1.00000	0.84906	0.84906	1.00000	0.94340
-0.4	0.96154	1.00000	0.86538	0.86538	1.00000	0.96154
-0.2	0.98039	1.00000	0.88235	0.88235	1.00000	0.98039
0.0	1.00000	1.00000	0.90000	0.90000	1.00000	1.00000
0.2	1.00000	0.98000	0.90000	0.90000	0.98000	1.00000
0.4	1.00000	0.96000	0.90000	0.90000	0.96000	1.00000
0.6	1.00000	0.94000	0.90000	0.90000	0.94000	1.00000
0.8	1.00000	0.92000	0.90000	0.90000	0.92000	1.00000
1.0	1.00000	0.90000	0.90000	0.90000	0.90000	1.00000
1.2	1.00000	0.88000	0.90000	0.90000	0.88000	1.00000
1.4	1.00000	0.86000	0.90000	0.90000	0.86000	1.00000
1.6	1.00000	0.84000	0.90000	0.90000	0.84000	1.00000
1.8	1.00000	0.82000	0.90000	0.90000	0.82000	1.00000
2.0	1.00000	0.80000	0.90000	0.90000	0.80000	1.00000
2.2	1.00000	0.78000	0.90000	0.90000	0.78000	1.00000
2.4	1.00000	0.76000	0.90000	0.90000	0.76000	1.00000
2.6	1.00000	0.74000	0.90000	0.90000	0.74000	1.00000
2.8	1.00000	0.72000	0.90000	0.90000	0.72000	1.00000
3.0	1.00000	0.70000	0.90000	0.90000	0.70000	1.00000
10.0	1.00000	0.00000	0.90000	0.90000	0.00000	1.00000

The above table was generated by the following BASIC code:

```
REM PROGRAM FILENAME SELECT02.BAS
CLS : DEFDBL A-Z
PRINT " ------------------------------------------------------------------"
PRINT "   h       W1=1     W2=1-hs    W3=1-s    W1=1-s   W2=1-hs    W3=1"
```

```
PRINT " ------------------------------------------------------------------"
PRINT "  -inf.   ";
PRINT USING "#.#####   "; 0; 1; 0; 0; 1; 0
S = 1 / 10: FOR H1 = -30 TO 30 STEP 2: H = H1 / 10
W11 = 1: W21 = 1 - H * S: W31 = 1 - S
IF W11 >= W21 AND W11 >= W31 THEN WMAX = W11
IF W21 >= W11 AND W21 >= W31 THEN WMAX = W21
IF W31 >= W11 AND W31 >= W21 THEN WMAX = W31
W11 = W11 / WMAX: W21 = W21 / WMAX: W31 = W31 / WMAX
W12 = 1 - S: W22 = 1 - H * S: W32 = 1
IF W12 >= W22 AND W12 >= W32 THEN WMAX = W12
IF W22 >= W12 AND W22 >= W32 THEN WMAX = W22
IF W32 >= W12 AND W32 >= W22 THEN WMAX = W32
W12 = W12 / WMAX: W22 = W22 / WMAX: W32 = W32 / WMAX
PRINT USING " ###.#    "; H;
PRINT USING "#.#####   "; W11; W21; W31; W12; W22; W32
DO: LOOP WHILE INKEY$ <> " "
NEXT H1
PRINT "  10.0    ";
PRINT USING "#.#####   "; 1; 0; .9; .9; 0; 1
PRINT " ------------------------------------------------------------------"
```

In the table below (see appended BASIC code) the values of **Δq** are shown as functions of **q** and of some selected values of **h.**

q	h								
	-inf.	-200	-2	-1	0	+0.5	+1	+2	+10
0.00000	0.50000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.01563	0.48438	0.18448	0.00294	0.00146	-.00002	-.00077	-.00152	-.00302	-.01540
0.03125	0.46875	0.25670	0.00552	0.00273	-.00009	-.00152	-.00295	-.00584	-.03031
0.04688	0.45313	0.29049	0.00775	0.00381	-.00021	-.00224	-.00430	-.00846	-.04470
0.06250	0.43750	0.30659	0.00966	0.00471	-.00037	-.00295	-.00556	-.01088	-.05852
0.07813	0.42188	0.31307	0.01127	0.00544	-.00056	-.00363	-.00674	-.01310	-.07170
0.09375	0.40625	0.31377	0.01259	0.00601	-.00080	-.00429	-.00784	-.01513	-.08421
0.10938	0.39063	0.31071	0.01364	0.00643	-.00107	-.00492	-.00886	-.01697	-.09598
0.12500	0.37500	0.30507	0.01443	0.00670	-.00137	-.00554	-.00980	-.01862	-.10696
0.14063	0.35938	0.29758	0.01498	0.00683	-.00170	-.00613	-.01066	-.02008	-.11709
0.15625	0.34375	0.28874	0.01530	0.00684	-.00206	-.00670	-.01145	-.02137	-.12631
0.17188	0.32813	0.27886	0.01540	0.00672	-.00245	-.00724	-.01217	-.02247	-.13455
0.18750	0.31250	0.26818	0.01531	0.00649	-.00287	-.00776	-.01281	-.02341	-.14176
0.20313	0.29688	0.25686	0.01502	0.00615	-.00330	-.00826	-.01339	-.02417	-.14788
0.21875	0.28125	0.24503	0.01456	0.00571	-.00376	-.00874	-.01389	-.02478	-.15284
0.23438	0.26563	0.23279	0.01394	0.00517	-.00423	-.00919	-.01433	-.02522	-.15660
0.25000	0.25000	0.22020	0.01316	0.00455	-.00472	-.00962	-.01471	-.02551	-.15909
0.26563	0.23438	0.20733	0.01224	0.00384	-.00522	-.01002	-.01502	-.02565	-.16028
0.28125	0.21875	0.19422	0.01119	0.00306	-.00573	-.01040	-.01527	-.02565	-.16013
0.29688	0.20313	0.18091	0.01002	0.00221	-.00625	-.01076	-.01546	-.02551	-.15861
0.31250	0.18750	0.16743	0.00873	0.00130	-.00678	-.01109	-.01559	-.02524	-.15571
0.32813	0.17188	0.15380	0.00735	0.00033	-.00731	-.01140	-.01567	-.02485	-.15140
0.34375	0.15625	0.14005	0.00588	-.00068	-.00785	-.01168	-.01570	-.02434	-.14571
0.35938	0.14063	0.12620	0.00433	-.00174	-.00838	-.01194	-.01567	-.02371	-.13866
0.37500	0.12500	0.11226	0.00271	-.00284	-.00891	-.01218	-.01560	-.02299	-.13029
0.39063	0.10938	0.09824	0.00103	-.00396	-.00944	-.01239	-.01548	-.02216	-.12065
0.40625	0.09375	0.08416	-.00070	-.00511	-.00996	-.01257	-.01531	-.02125	-.10982
0.42188	0.07813	0.07002	-.00247	-.00628	-.01048	-.01273	-.01511	-.02025	-.09789
0.43750	0.06250	0.05584	-.00428	-.00747	-.01098	-.01287	-.01486	-.01917	-.08498
0.45313	0.04688	0.04163	-.00610	-.00865	-.01146	-.01298	-.01457	-.01803	-.07122
0.46875	0.03125	0.02739	-.00794	-.00984	-.01194	-.01306	-.01425	-.01683	-.05675
0.48438	0.01563	0.01313	-.00979	-.01103	-.01239	-.01312	-.01390	-.01558	-.04172
0.50000	0.00000	-.00114	-.01163	-.01220	-.01282	-.01316	-.01351	-.01429	-.02632
0.51563	-.01563	-.01541	-.01345	-.01335	-.01323	-.01317	-.01310	-.01296	-.01071
0.53125	-.03125	-.02968	-.01525	-.01447	-.01361	-.01315	-.01266	-.01160	0.00493
0.54688	-.04688	-.04394	-.01702	-.01557	-.01397	-.01311	-.01220	-.01022	0.02040
0.56250	-.06250	-.05818	-.01874	-.01663	-.01430	-.01304	-.01171	-.00884	0.03553
0.57813	-.07813	-.07240	-.02041	-.01764	-.01459	-.01294	-.01121	-.00745	0.05014
0.59375	-.09375	-.08658	-.02202	-.01860	-.01485	-.01282	-.01069	-.00608	0.06408
0.60938	-.10938	-.10071	-.02355	-.01951	-.01506	-.01267	-.01016	-.00472	0.07717

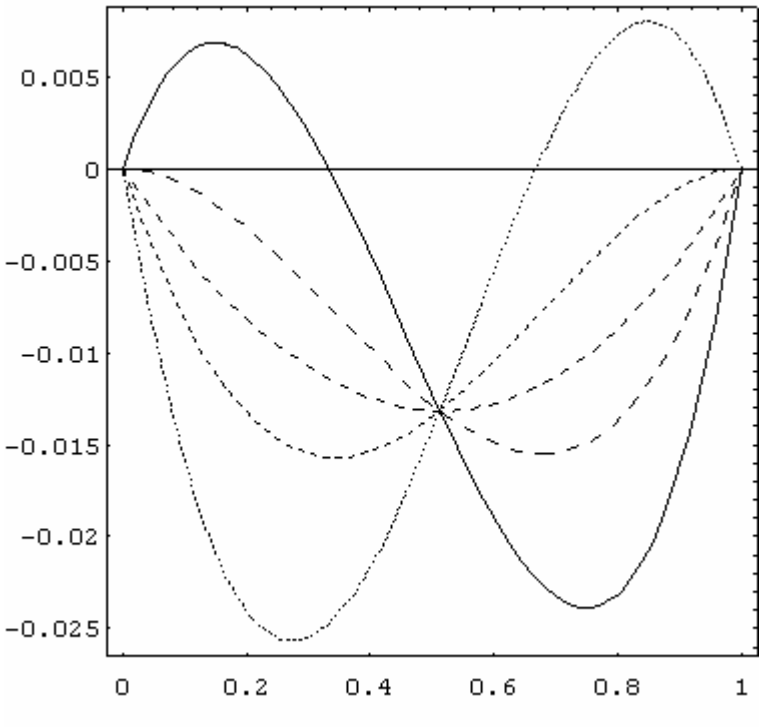
```
0.62500 -.12500 -.11480 -.02500 -.02035 -.01524 -.01250 -.00962 -.00338 0.08929
0.64063 -.14063 -.12881 -.02635 -.02112 -.01538 -.01230 -.00906 -.00207 0.10030
0.65625 -.15625 -.14275 -.02760 -.02181 -.01547 -.01207 -.00850 -.00081 0.11011
0.67188 -.17188 -.15660 -.02873 -.02241 -.01551 -.01182 -.00794 0.00040 0.11863
0.68750 -.18750 -.17033 -.02973 -.02293 -.01550 -.01154 -.00738 0.00155 0.12579
0.70313 -.20313 -.18394 -.03060 -.02334 -.01544 -.01123 -.00682 0.00263 0.13154
0.71875 -.21875 -.19740 -.03130 -.02364 -.01532 -.01089 -.00626 0.00364 0.13585
0.73438 -.23438 -.21067 -.03185 -.02382 -.01514 -.01053 -.00571 0.00456 0.13871
0.75000 -.25000 -.22372 -.03221 -.02389 -.01490 -.01014 -.00517 0.00540 0.14011
0.76563 -.26563 -.23652 -.03238 -.02381 -.01459 -.00972 -.00464 0.00613 0.14007
0.78125 -.28125 -.24900 -.03234 -.02360 -.01422 -.00927 -.00413 0.00675 0.13862
0.79688 -.29688 -.26110 -.03208 -.02323 -.01377 -.00879 -.00364 0.00725 0.13579
0.81250 -.31250 -.27273 -.03158 -.02271 -.01325 -.00829 -.00316 0.00763 0.13163
0.82813 -.32813 -.28377 -.03083 -.02201 -.01265 -.00776 -.00271 0.00788 0.12620
0.84375 -.34375 -.29407 -.02980 -.02113 -.01198 -.00720 -.00228 0.00799 0.11954
0.85938 -.35938 -.30340 -.02848 -.02007 -.01121 -.00661 -.00188 0.00796 0.11173
0.87500 -.37500 -.31145 -.02686 -.01880 -.01036 -.00599 -.00152 0.00777 0.10283
0.89063 -.39063 -.31777 -.02490 -.01732 -.00942 -.00535 -.00118 0.00742 0.09289
0.90625 -.40625 -.32164 -.02259 -.01562 -.00839 -.00467 -.00088 0.00691 0.08200
0.92188 -.42188 -.32193 -.01991 -.01368 -.00726 -.00397 -.00062 0.00622 0.07021
0.93750 -.43750 -.31662 -.01683 -.01150 -.00602 -.00323 -.00041 0.00536 0.05759
0.95313 -.45313 -.30191 -.01333 -.00905 -.00468 -.00247 -.00023 0.00431 0.04419
0.96875 -.46875 -.26950 -.00938 -.00633 -.00324 -.00168 -.00011 0.00307 0.03010
0.98438 -.48438 -.19727 -.00494 -.00332 -.00168 -.00085 -.00003 0.00163 0.01535
1.00000 -.50000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000
--------------------------------------------------------------------------------
```

```
REM PROGRAM FILENAME SELECT03.BAS
CLS : DEFDBL A-Z
DATA 999,-200,-2,-1,0,0.5,1,2,10
FOR I = 1 TO 9: READ H(I): NEXT I
PRINT "-------------------------------------------------------------------------";
PRINT "-----------"
PRINT "                                          h"
PRINT "    q      ";
PRINT "-------------------------------------------------------------------------";
PRINT "           -inf.  -200     -2      -1       0    +0.5      +1      +";
PRINT "2     +10"
PRINT "-------------------------------------------------------------------------";
PRINT "-----------"
S = .1: FOR Q1 = 0 TO 64: Q = Q1 / 64
PRINT USING "#.##### "; Q;
FOR I = 1 TO 9
IF I = 1 THEN
   DQ = 1 / 2 - Q
ELSE
DQ = -Q * (1 - Q) * (H(I) * S + Q * S * (1 - 2 * H(I)))
DQ = DQ / (1 - 2 * H(I) * S * Q - S * (1 - 2 * H(I)) * Q * Q)
END IF
PRINT USING "#.##### "; DQ;
NEXT I: PRINT
DO: LOOP WHILE INKEY$ <> " "
NEXT Q1
PRINT "-------------------------------------------------------------------------";
PRINT "-----------"
```

The Mathematica graph below shows, for **s** = 0.1 and the case $\{W_1 = 1/W_{max}, W_2 = (1 - hs)/W_{max}, W_3 = (1 - s)/W_{max}, W_{max} = max(W_1,W_2,W_3)\}$, the variation of Δq as function of **q** and the following values of **h**: -1, 0, 0.5, 1, and 2.).

```
(*
gensel01.ma
                (aa)       (Aa)       (AA)
|----------------|---------|---------|
0             W3=1-s    W2=1-hs    W1=1
```

```
h  = hs/s = (1-W2)/(1-W3)
dq = -q(1-q).[hs+qs(1-2h)]/[1-2hsq-s(1-2h)q^2]
*)
s = 0.1;
f[q_,h_]:= -q * (1-q) * (h * s + q * s * (1 - 2 * h))/
       (1 - 2 * h * s * q - s * (1 - 2 * h) * q^2);
Plot[{f[q,-1],f[q,0],f[q,0.5],f[q,1],f[q,2]},{q,0,1},
     PlotStyle->{{},{Dashing[{0.020}]},{Dashing[{0.015}]},
                 {Dashing[{0.010}]},{Dashing[{0.005}]}},
     Frame -> True, AspectRatio -> 1]
```




```
(*
_____ : h = -1.0 -> W3=0.9 < W2=1.10 > W1=1.0
_ _ _ : h =  0.0 -> W3=0.9 < W2=1.00 = W1=1.0
- - - : h =  0.5 -> W3=0.9 < W2=0.95 < W1=1.0
----- : h =  1.0 -> W3=0.9 = W2=0.90 < W1=1.0
..... : h =  2.0 -> W3=0.9 > W2=0.80 < W1=1.0
*)
```

FUNDAMENTAL THEOREM OF NATURAL SELECTION

Following a simple reasoning proposed by Wallace, we will begin with the analysis of a population of haploid individuals a_1 and a_2, occurring with frequencies p_1 and p_2. Letting w_1 and w_2 be the fitness (adaptive) values associated to genotypes a_1 and a_2, the average population fitness value is given by $w = p_1w_1 + p_2w_2$; the gene frequencies in the next generation are given by $p_1' = p_1w_1/w$ and $p_2' = p_2w_2/w$, when the average population fitness value becomes

$$w' = p_1'w_1 + p_2'w_2 = p_1w_1^2/w + p_2w_2^2/w = (p_1w_1^2 + p_2w_2^2)/w \ ;$$

therefore,

$$\Delta w = w' - w = (p_1w_1^2 + p_2w_2^2)/w - w = (p_1w_1^2 + p_2w_2^2 - w^2)/w$$
$$= [\Sigma p_iw_i^2 - (\Sigma p_iw_i)^2]/w = \Sigma p_i(w_i-w)^2/w$$

and

$$\Delta w/w = \Sigma p_i(w_i-w)^2/w^2 = var(w_i)/w^2 \propto var(w_i) \ ,$$

what demonstrates that the net increase of the average population fitness value is proportional to the genetic additive variance in fitness (Fisher's fundamental theorem). The simplification shown above avoids any mathematical complications at all, but nevertheless the principle can be easily generalized for the case of diploid individuals (as given originally by Fisher), as shown in the lines below. In fact, let

$$P(a_1a_1) = p_1^2$$
$$P(a_1a_2) = 2p_1p_2$$
$$P(a_2a_2) = p_2^2$$

be the frequencies of the three possible genotypes determined by a pair of alleles segregating at an autosomal locus; and w_{11}, w_{12} and w_{22} the adaptive values of genotypes a_1a_1, a_1a_2 and a_2a_2; the average fitness value of the population is given by

$$w = \Sigma[P(a_i a_j).w_{ij}] = \Sigma p_i p_j.w_{ij}$$
$$= p_1^2.w_{11} + 2p_1p_2.w_{12} + p_2^2.w_{22}$$
$$= p_1(p_1.w_{11} + p_2.w_{12}) + p_2(p_1.w_{12} + p_2.w_{22}) = p_1.w_1 + p_2.w_2 ,$$

where $w_1 = p_1.w_{11} + p_2.w_{12}$ and $w_2 = p_1.w_{12} + p_2.w_{22}$ are the average fitness values of all individuals having the alleles a_1 and a_2. These quantities are also known as the average excesses of alleles a_1 and a_2. Since after selection the allele frequencies are

$$p_1' = (p_1^2.w_{11} + p_1p_2.w_{12})/w = p_1(p_1.w_{11} + p_2.w_{12})/w = p_1.w_1/w \quad \text{and}$$
$$p_2' = (p_2^2.w_{22} + p_1p_2.w_{12})/w = p_2(p_1.w_{12} + p_2.w_{22})/w = p_2.w_2/w ,$$

it comes out that

$$\Delta p_1 = p_1' - p_1 = p_1.w_1/w - p_1 = p_1(w_1-w)/w \quad \text{and}$$
$$\Delta p_2 = p_2' - p_2 = p_2.w_2/w - p_2 = p_2(w_2-w)/w ;$$

since $p_1 = 1 - p_2$ and $p_1' = 1 - p_2'$, we have also

$$\Delta p_1 = p_1' - p_1 = 1 - p_2' - 1 + p_2 = p_2 - p_2' = - \Delta p_2 .$$

Therefore,

$$w' = p_1'^2.w_{11} + 2p_1'p_2'.w_{12} + p_2'^2.w_{22}$$
$$= (p_1+\Delta p_1)^2 w_{11} + 2(p_1+\Delta p_1)(p_2+\Delta p_2)w_{12} + (p_2+\Delta p_2)^2 w_{22}$$
$$\approx p_1^2 w_{11} + 2p_1p_2w_{12} + p_2^2 w_{22} + 2\Delta p_1(p_1w_{11}+p_2w_{12}) + 2\Delta p_2(p_1w_{12}+p_2w_{22})$$
$$= w + 2\Delta p_1.w_1 + 2\Delta p_2.w_2 = w + 2\Delta p_1(w_1-w_2)$$
$$= w + 2\Delta p_1[(w_1-w)-(w_2-w)] = w + 2\Delta p_1(w_1-w) + 2\Delta p_2(w_2-w)$$
$$= w + 2p_1(w_1-w)^2/w + 2p_2(w_2-w)^2/w ,$$

$$\Delta w = w' - w = 2p_1(w_1-w)^2/w + 2p_2(w_2-w)^2/w =$$
$$= 2[p_1(w_1-w)^2+p_2(w_2-w)^2]/w = 2\Sigma p_i(w_i-w)^2/w$$

and

$$\Delta w/w = (w' - w)/w = 2\Sigma p_i(w_i-w)^2/w^2 \propto \Sigma p_i(w_i-w)^2 \ .$$

Therefore, the increment rate of the average population fitness value per generation is proportional to the additive genetic variance of fitness values at that time.

GENETIC LOAD

Genetic load is the fraction by which the population mean fitness value (**W**) is changed as a consequence of the factor under consideration in comparison with an identical population in which the factor is missing and which fitness value is taken as $W_{max} = 1$:

$L = (W_{max} - W)/W_{max} = 1 - W$.

Genetic load can be due to mutation, segregation, incompatibility, meiotic drive, and other factors. In the lines that follow we detail the cases of mutation and segregation load.

1. Mutation load

If we put $W_1 = W_{AA} = 1$, $W_2 = W_{Aa} = 1 - sh$, $W_3 = W_{aa} = 1 - s$ and if μ is the mutation rate [$\mu = P(A \rightarrow a)$], then it comes out that

$$q_{n+1} = [\mu p_n(1-q_n sh)+q_n-p_n q_n sh-q_n^2 s]/(1-2p_n q_n sh-q_n^2 s) \ .$$

At equilibrium,

$$\mu = (q^2 s + qsh -2q^2 sh)/(1-qsh) \approx sq^2 + qsh$$

and

$$W \approx 1 - 2hspq -sq^2 \ ;$$

if $h = 0$, $\mu = sq^2$, $W = 1 - sq^2 = 1 - \mu$,
and $L = \mu$;

if $h = 1/2$, $\mu = sq/2$, $W = 1 - spq - sq^2 \approx 1 - sq = 1 - 2\mu$,
and $L = 2\mu$;

if $h = 1$, $\mu = sq$, $W = 1 - 2spq - sq^2 \approx 1 - 2sq = 1 - 2\mu$,
and $L = 2\mu$;

therefore, the mutation load **L** takes always values between μ and 2μ, for panmictic populations.

2. Segregation load

If there is overdominance and the mutation rate is small as compared to s_1 and s_3 (the coefficients of selection of genotypes **AA** and **aa**), at equilibrium

$$p = s_3/(s_1 + s_3),$$

$$q = s_1/(s_1 + s_3),$$

$$\begin{aligned} W &= 1 - s_1 p^2 - s_3 q^2 \\ &= 1 - s_1 s_3/(s_1 + s_3) \\ &= 1- qs_3 = 1 - ps_1 \ , \end{aligned}$$

and

$$L = 1 - W = s_1 p^2 + s_3 q^2 = s_1 s_3/(s_1 + s_3) = qs_3 = ps_1 \ .$$

SELECTION WITH INBREEDING

If the population is not inbred, the genotype frequencies before selection acts are in the ratios p^2, $2pq$ and q^2; letting W_1, W_2 and W_3 be the adaptive (fitness) values associated with genotypes **AA**, **Aa** and **aa**, it comes out that

$$p' = (p^2W_1 + pqW_2)/W = p(pW_1+qW_2)/W = (1-q)[(1-q)W_1+qW_2]/W$$
$$= f_2(q)/W$$

$$q' = (q^2W_3 + pqW_2)/W = q(qW_3+pW_2)/W = q[qW_3+(1-q)W_2]/W$$
$$= f_1(q)/W$$

$$W = f_1(q)/W + f_2(q)/W$$

and

$$\Delta q = q' - q = f_1(q)/W - q = [f_1(q)-qf_1(q)-qf_2(q)]/W$$
$$= [(1-q)f_1(q)-qf_2(q)]/W = q(1-q)[q(W_3-W_2)-(1-q)(W_1-W_2)];$$

since

$$df_1(q)/dq = 2q(W_3-W_2) + W_2 ,$$
$$df_2(q)/dq = -2(1-q)(W_1-W_2) - W_2 , \text{ and}$$
$$dW/dq = df_1(q)/dq + df_2(q)/dq = 2q(W_3-W_2)-2(1-q)(W_1-W_2) ,$$

we get the result

$$\Delta q = q(1-q)/2W \; . \; dW/dq \; .$$

In the lines that follow we analyze the overdominance case. If the population is inbred with a fixation index **F**, the frequencies of the genotypes **AA**, **Aa**, and **aa** before selection acts are respectively p^2+pqF, $2pq(1-F)$ and q^2+pqF; since the fitness values associated with these genotypes are $1-s_1$, 1 and $1-s_3$, gene frequencies after selection has acted are

$$p' = [p-s_1p(p+qF)]/w$$
$$q' = [q-s_3q(q+pF)]/w \text{ , where}$$

$$w = 1-s_1p(p+qF)-s_3q(q+pF),$$

so that

$$p'/q' = p/q \; . \; [1-s_1(p+qF)]/[1-s_3(q+pF)];$$

without taking into account mutation effects, **p'** and **q'** are the gene frequencies in next generation; at equilibrium, **p' = p** and **q' = q**; therefore,

$s_1(p+qF) = s_3(q+pF)$ and

$$q = (s_1-s_3F)/[(s_1+s_3)(1-F)].$$

The numerical analysis of $\mathbf{q' = [q-s_3q(q+pF)]/w}$ shows that if $\mathbf{s_1 = s_3}$ convergence occurs to $\mathbf{q = 0.5}$, for any value of **F**. If $\mathbf{s_1 > s_3}$ and $\mathbf{F \geq s_3/s_1}$ the equilibrium point is $\mathbf{q = 1}$; if $\mathbf{s_3 > s_1}$ and $\mathbf{F \geq s_1/s_3}$ the equilibrium point is $\mathbf{q = 0}$. The equilibrium point $\mathbf{q = (s_1-s_3F)/[(s_1+s_3)(1-F)]}$ is attained only if $\mathbf{F < s_1/s_3}$ when $\mathbf{s_1 < s_3}$ or if $\mathbf{F < s_3/s_1}$ when $\mathbf{s_3 < s_1}$.

For $\mathbf{F = 0}$, $\mathbf{Q = s_1/(s_1+s_3)}$ and $\mathbf{P = s_3/(s_1+s_3)}$. At the equilibrium point with inbreeding, therefore,

$$\mathbf{q-Q = (s_1-s_3F)/[(s_1+s_3)(1-F)] - s_1(1-F)/[(s_1+s_3)(1-F)]}$$
$$\mathbf{= F(s_1-s_3)/[(s_1+s_3)(1-F)] = F(Q-P)/(1-F) = F(2Q-1)/(1-F).}$$

Therefore,

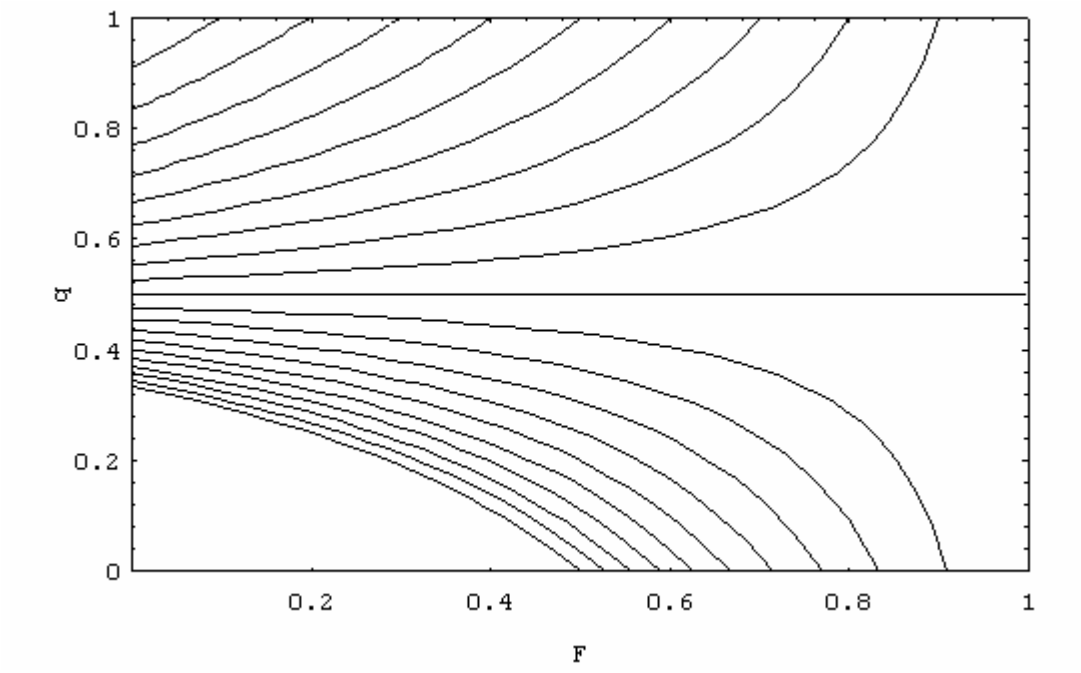
$$\mathbf{q = Q + F(2Q-1)/(1-F),}$$

that is, the effect of inbreeding in the polymorphic equilibrium point Q is to shift it to another equilibrium point q; q > Q if Q > 0.5 and q < Q if Q < 0.5, as the following numerical examples show:

S1	S3	F	q	Q	q-Q
0.100	0.400	0.000	0.2000	0.2000	0.0000
0.100	0.400	0.050	0.1684	0.2000	-.0316
0.100	0.400	0.100	0.1333	0.2000	-.0667
0.100	0.400	0.150	0.0941	0.2000	-.1059
0.100	0.400	0.200	0.0500	0.2000	-.1500
0.100	0.400	0.250	0.0000	0.2000	-.2000
.....	.....	.....	.....	.....	.....
0.400	0.100	0.000	0.8000	0.8000	0.0000
0.400	0.100	0.050	0.8316	0.8000	0.0316
0.400	0.100	0.100	0.8667	0.8000	0.0667
0.400	0.100	0.150	0.9059	0.8000	0.1059
0.400	0.100	0.200	0.9500	0.8000	0.1500
0.400	0.100	0.250	1.0000	0.8000	0.2000

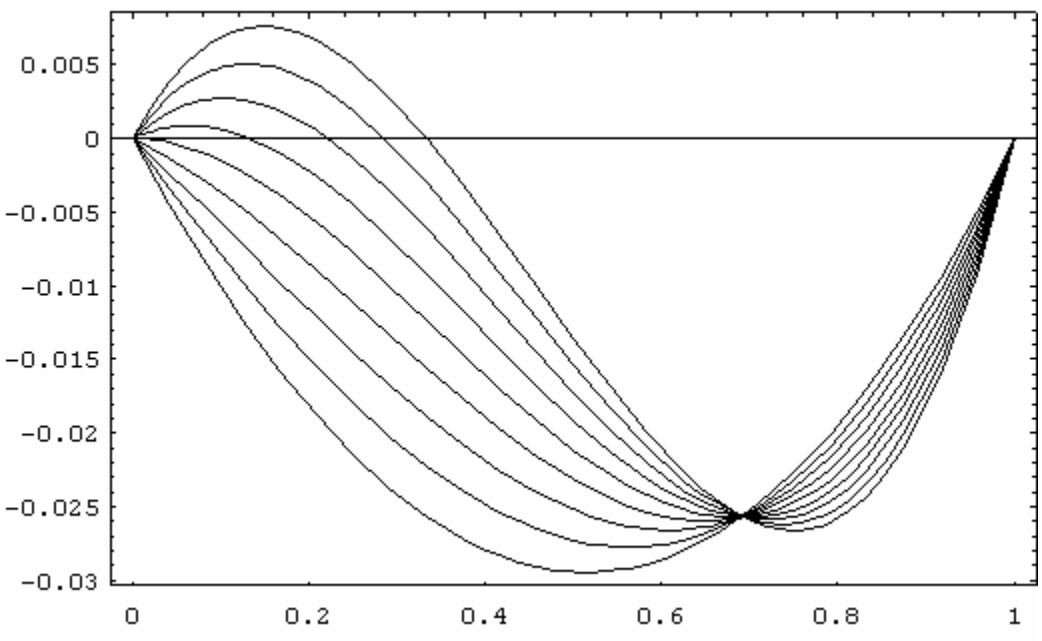
The graph that follows shows the equilibrium values **q** as function of **F**, with $\mathbf{s_1 = 0.5}$ and $\mathbf{s_3 = \{0.0, 0.1, 0.2, ..., 0.9, 1.0\}}$. The value $\mathbf{Q = s_1/(s_1+s_3)}$ corresponds always to $\mathbf{F = 0}$ for any combination of $\mathbf{s_1}$ and $\mathbf{s_3}$, and can be directly read at the intersection of the functions $\mathbf{q=f(F,s_1,s_3)}$ with the ordinate axis

```
(*
selmut06.ma
*)
s1=1/2;
For[i=1,i<=20,++i,
   F1[i_]:=N[i/20];
   F2[i_]:=Min[N[F1[i]/s1],N[s1/F1[i]]];
   F3[i_]:=(s1-F1[i]*f)/((s1+F1[i])*(1-f));
   Gr[i_]:=Plot[F3[i],{f,0,F2[i]},FrameLabel->{"F","q"},
               PlotRange->{{0,1},{0,1}}, Frame-> True,
               DisplayFunction->Identity];
   ];
Show[Gr[1],Gr[2],Gr[3],Gr[4],Gr[5],Gr[6],Gr[7],Gr[8],
     Gr[9],Gr[10],Gr[11],Gr[12],Gr[13],Gr[14],Gr[15],
     Gr[16],Gr[17],Gr[18],Gr[19],Gr[20],
     DisplayFunction->$DisplayFunction]
```



The important point to be kept is that if $s_1 > s_3$ and $F \geq s_3/s_1$ the equilibrium point is **q = 1**; if $s_3 > s_1$ and $F \geq s_1/s_3$ the equilibrium point is **q = 0** ; that is, under these conditions any polymorphic equilibrium **{0 << p = 1 - q << 1}** is disrupted. This is immediately clear from the Mathematica graph below, where the various **Δq** functions were calculated for s1 = 0.1, s2 = 0.2 and f = 0, 0.125, 0.25, ..., 1.

```
(*
selmut07.ma
*)
s1=.1;s3=.2;
deltaq[f_]:= (q-s3*q*(q+(1-q)*f))/(1-s1*(1-q)*(1-q+q*f)-s3*q*(q+(1-q)*f))-q;
Solve[deltaq[0]==0,q]
{{q -> 0.}, {q -> 0.333333}, {q -> 1.}}
Solve[deltaq[1/8]==0,q]
{{q -> 0.}, {q -> 0.285714}, {q -> 1.}}
Solve[deltaq[1/4]==0,q]
{{q -> 0.}, {q -> 0.222222}, {q -> 1.}}
Solve[deltaq[3/8]==0,q]
{{q -> 0.}, {q -> 0.133333}, {q -> 1.}}
Solve[deltaq[1/2]==0,q]
{{q -> 0.}, {q -> 0.}, {q -> 1.}}
Solve[deltaq[5/8]==0,q]
{{q -> -0.222222}, {q -> 0.}, {q -> 1.}}
Solve[deltaq[3/4]==0,q]
{{q -> -0.666667}, {q -> 0.}, {q -> 1.}}
Solve[deltaq[7/8]==0,q]
{{q -> -2.}, {q -> 0.}, {q -> 1.}}
Plot[{deltaq[0],deltaq[1/8],deltaq[1/4],deltaq[3/8],deltaq[1/2],deltaq[5/8],
      deltaq[3/4],deltaq[7/8],deltaq[1]},{q,0,1}, Frame->True]
```


0.005
0
-0.005
-0.01
-0.015
-0.02
-0.025
-0.03
0
0.2
0.4
0.6
0.8
1

EVOLUTION OF 1:1 SEX-RATIO

The intuitive argumentation is simple, but subtle since it requires the analysis of two generations. Let us suppose that in a given population there exists a surplus of males. Since each individual results from a fertilization in which the two gametes equally participate, females will have on average a larger offspring number than males. If in this population a mutation arises that makes its carriers poduce more females than males, this will result that these carriers will produce more individuals belonging to the sex that on average has a larger offspring. Therefore these individuals will have on average more grandchildren and the genes responsible for the increase of females will tend to increase in frequency. The inverse argumentation (population where initially there exists more females than males) is identical, that is, mutations that produce an excess of males tend to increase in frequency. We conclude therefore that the 1:1 sex ratio is evolutionary stable.

The following mathematical argumentation was adapted from Maynard-Smith (Maynard-Smith J. Evolutionary Genetics, Oxford University Press, Oxford, 1989). Let us consider the rare gene **M**, that has no expression in males and that makes females produce **m*** sons and **f*** daughters, contrarily to the offspring of other individuals, that have **m** sons and **f** daughters. It is important to stress that the presence of the gene **M** produces only a distortion in the sex ratio, in such a manner that **m + f = m* + f***. Let, now, **P** e **p** be the frequencies of **M/+** females and **M/+** males. Since the **M** is rare, the frequencies of **M/M** males and females can be taken as negligible ones; therefore, they will not be considered in the simplified calculations that follow.

crossings		frequencies	offspring			
			males		females	
fem.	males		M/+	+/+	M/+	+/+
M/+	+/+	P(1-p) = P-Pp ~ P	m*P/2	m*P/2	f*P/2	f*P/2
+/+	M/+	p(I-P) = p-Pp ~ p	mp/2	mp/2	fp/2	fp/2
+/+	+/+	(l-P)(l-p)=1-P-p+Pp~1-P-p	-	m(1-P-p)	-	f(1-P-p)
			m+P(m*-m) ~ m		f+P(f*-f) ~ f	

The frequencies **P'** and **p'** of **M/+** males and females in the following generation are taken directly from the table above, and take values

P' = (f*P/2+fp/2)/[f+P(f*-f)] ~ (f*P/2+fp/2)/f = f*P/2f + p/2
p' = (m*P/2+mp/2)/[m+P(m*-m)] ~ (m*P/2+mp/2)/m = m*P/2m + p/2
P'+p' = p + (f*P/f + m*P/m)/2 = p + P(f*/f + m*/m)/2
= p + P + P[(f*/f + m*/m)/2 - 1] = p + P + RP,
where **R = (f*/f + m*/m)/2 - 1 .**

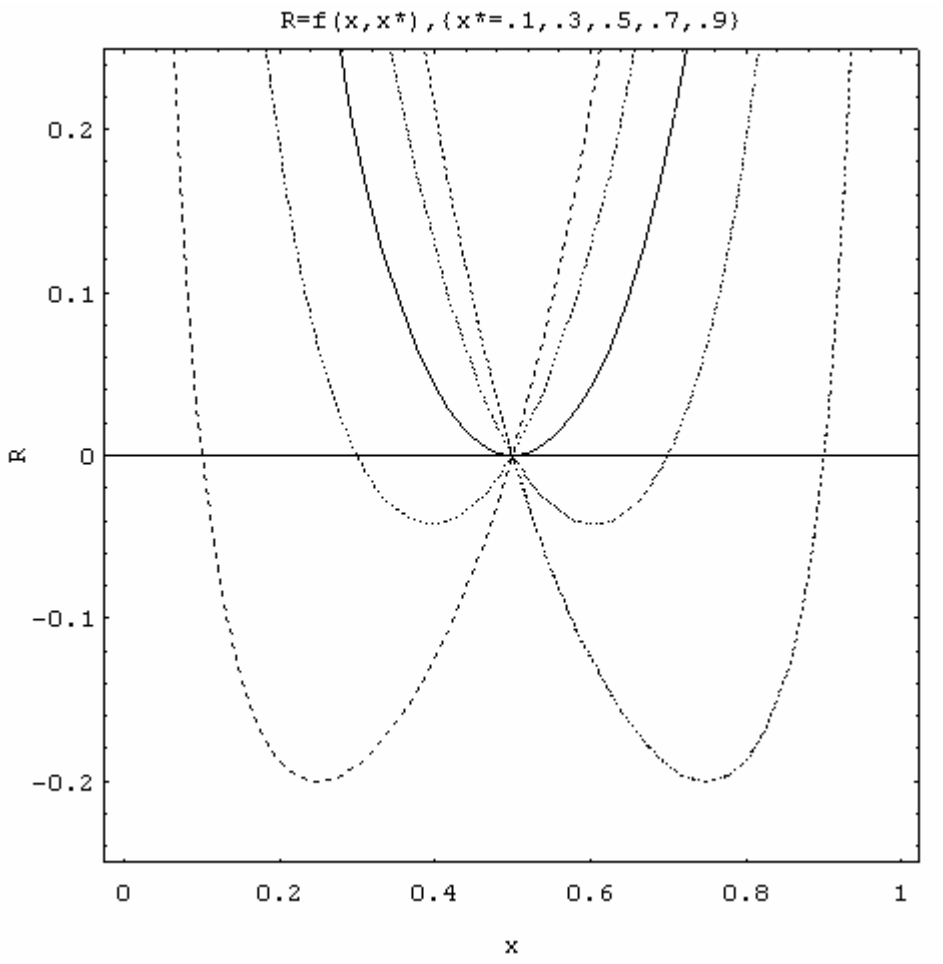
Since **m*+f* = m+f**, it comes out that **f*/f + m*/m = [f*/(m*+f*)]/[f/(m+f)] + [m*/(m*+f*)]/[m/(m+f)] = (1-x*)/(1-x) + x*/x** , where **x = proportion of males** and **1-x = proportion of females**. Therefore, **R = [(1-x*)/(1-x)+x*/x]/2 - 1 = [x(1-x*)+x*(1-x)-2x(1-x)]/[2x(1-x)] = (1-2x)(x*-x)/[2x(1-x)].**

P'+p' will be larger than **P+p** if **R>0**. **R** is larger than zero if **x<1/2 and x*>x** or if **x>1/2 and x>x***. Therefore, when **x<1/2**, the mutants that increase the sex-ratio increase their frequency; and when **x>1/2** the mutants that decrease the sex-ratio increase their frequency. Therefore,

the solution of the equation **R = 0 (x = 1/2)**, is an evolutionary stable sex ratio.

The figure below, generated by the enclosed *Mathematica* code, shows the values of **-0.25 <R < 0.25** as function of **x** in the interval (0,1) and of **x* = 0.1, 0.3, 0.5, 0.7** and **0.9**.

```
(* sexrat01.ma *)
R[x1_,x2_]:=(1-2*x1)*(x2-x1)/(2*x1*(1-x1))
Plot[{R[i,.1],R[i,.3],R[i,.5],R[i,.7],R[i,.9]},{i,0,1},
      PlotRange->{-.25,.25}, AspectRatio->1,
      PlotStyle->{{Dashing[{0.008}]},{Dashing[{0.004}]},
               {},{Dashing[{0.002}]},{Dashing[{0.001}]}},
      PlotLabel->"R=f(x,x*),{x*=.1,.3,.5,.7,.9}",
      FrameLabel->{"x","R"},
      Frame->True]
```



The argumentation discussed above can be easily generalized for the generic situation of parental expenditure suggested by Fisher. Let **m + kf = C** and **m* + kf* = C**, where **C** represents the total expenditure in the offspring, a female costing **k** times more than a male. From these two equations we obtain **f = (C-m)/k** and **f* = (C-m*)/k**. Replacing these values in **R = (f*/f + m*/m)/2 - 1**, we immediately obtain **R = (C-2m)(m*-m)/[2m(1-**

m)]. R will be greater than zero if **m < C/2 and m* > m** or if **m > C/2 and m > m***. Therefore, if **m < C/2,** mutants that increase the value of **m** will tend to increase their frequency; if **m > C/2,** mutants that decrease the value of **m** will tend to increase their frequency. Therefore the evolutionary stable sex-ratio is given by **m = C/2**. If **k = 1, C = m + f = m* + f*** and **x = 1/2** (as seen before).

MUTATION-SELECTION BALANCE

1) Complete selection against recessive individuals

If μ is the mutation rate of the recessive gene **a** [$\mu = P(A \rightarrow a)$] per generation, then it comes out that

$$q_{n+1} = (\mu + q_n)/(1 + q_n) \ ;$$

this is a fractionary difference equation of first order, the general exact solution in simple analytic form of which is given by

$$q_n = \sqrt{\mu}[(q_0+\sqrt{\mu})(1+\sqrt{\mu})^n+(q_0-\sqrt{\mu})(1-\sqrt{\mu})^n]/[(q_0+\sqrt{\mu})(1+\sqrt{\mu})^n-(q_0-\sqrt{\mu})(1-\sqrt{\mu})^n] .$$

If we define

$$\Delta_1 = \mu/(1+q)$$

and

$$\Delta_2 = -q^2/(1+q) \ ,$$

then it comes out that the net change in gene frequency per generation is

$$\Delta q = \Delta_1 + \Delta_2 = (\mu-q^2)/(1+q) \ ;$$

at equilibrium (that is, when **n** tends to infinity),

$$\Delta q = 0 \ , \ \mu = q^2 \ , \ q = \sqrt{\mu} \ .$$

This last result can be obtained straightforwardly by taking the limit of the above general expression of q_n as n tends to infinity; when **n** tends to infinity, the expression $(1-\sqrt{\mu})^n$ tends to zero, since $0 < 1-\sqrt{\mu} < 1$; and the expression for q_n reduces to

$$\lim_{n\rightarrow\infty} q_n = q = \sqrt{\mu}[(q_0+\sqrt{\mu}).\lim_{n\rightarrow\infty}(1+\sqrt{\mu})^n]/[(q_0+\sqrt{\mu}).\lim_{n\rightarrow\infty}(1+\sqrt{\mu})^n]$$
$$= \sqrt{\mu}[(q_0+\sqrt{\mu})/(q_0+\sqrt{\mu})] = \sqrt{\mu} \ .$$

2) Partial selection against recessive individuals

If μ is the mutation rate [$\mu = P(A \rightarrow a)$] and **s** the selection coefficient of **aa** individuals, then it comes out that

$$q_{n+1} = [q_n(1-sq_n)+\mu(1-q_n)]/(1-sq_n^2) \ .$$

If we define, as before,

$$\Delta_1 = \mu(1-q)/(1-sq^2)$$

and

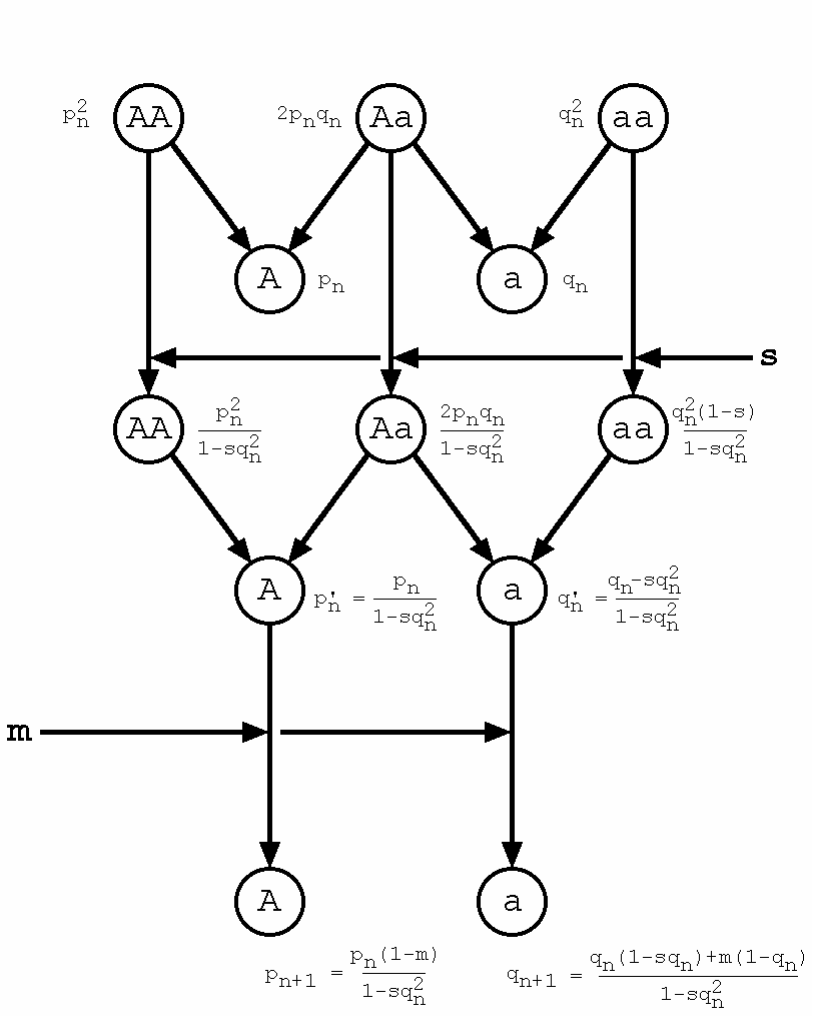
$$\Delta_2 = -sq^2(1-q)/(1-sq^2) \ ,$$

then it comes out that

$$\Delta q = \Delta_1 + \Delta_2 = (\mu-sq^2)(1-q)/(1-sq^2) \ ;$$

at equilibrium,

$\Delta q = 0$, $\mu = sq^2$, $q = \sqrt{(\mu/s)}$.



3) Selection against dominant individuals

If μ is the mutation rate of the dominant gene **A** [μ = **P(a -> A)**] per generation, **s** the selection coefficient of **Aa** individuals and **1** the selection coefficient of **AA** individuals, then it comes out that

$p_{n+1} = [p_n(1-s)+\mu(1-sp_n)]/[1+p_n(1-2s)]$.

If we define :

$\Delta_1 = \mu(1-sp)/[1+p(1-2s)]$

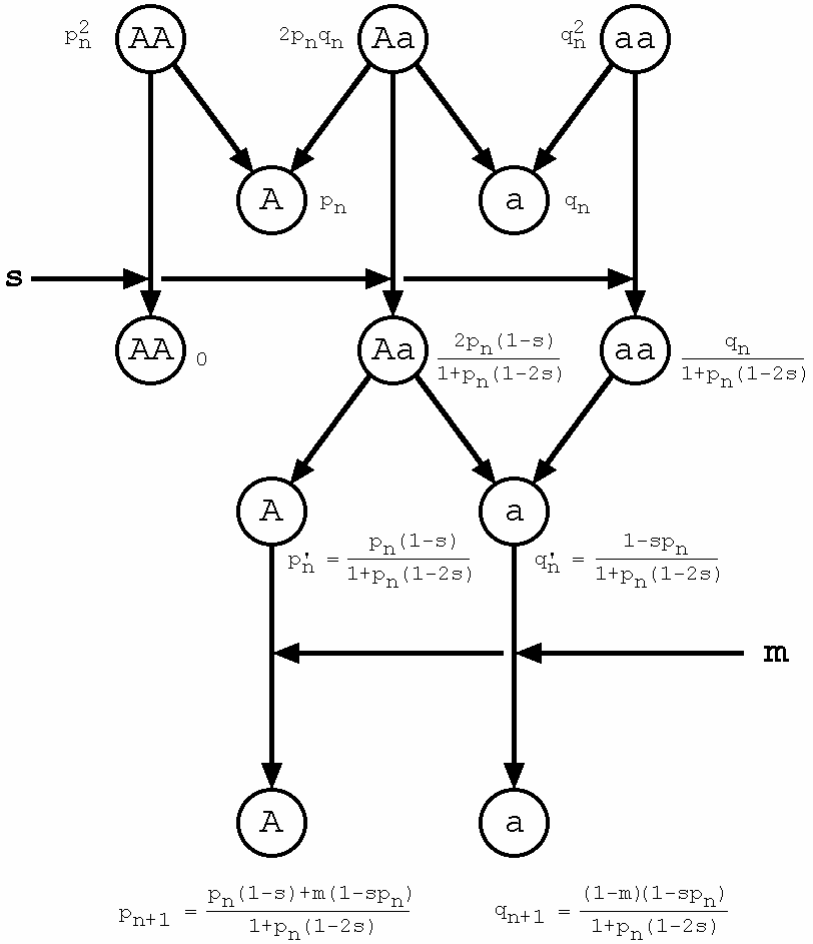
$\Delta_2 = -[p^2(1-2s)+sp]/[1+p(1-2s)]$ and

$\Delta p = \Delta_1 + \Delta_2 = [\mu-p^2(1-2s)-sp(1+\mu)]/[1+p(1-2s)]$,

at equilibrium, $\Delta p = 0$ and

$$\mu = [p^2(1-s)+sp(1-p)]/(1-sp) = [sp + p^2(1-2s)]/(1-sp)$$

$$\approx sp(1-p) \approx spq \approx sp \ .$$



4) General case

If we put $W_1 = W_{AA} = 1$, $W_2 = W_{Aa} = 1-s = 1-sh$, $W_3 = W_{aa} = 1-s$ and if μ is the mutation rate [$\mu = P(A \rightarrow a)$], then it comes out that

$$q_{n+1} = [\mu p_n(1-q_nsh)+q_n-p_nq_nsh-q_n^2s]/(1-2p_nq_nsh-q_n^2s) \ .$$

At equilibrium,

$$\mu = (q^2s + qsh - 2q^2sh)/(1-qsh)$$

$$\approx sq^2 + qsh \ .$$

If **h = 0** the situation is reduced to the case of partial selection against recessive individuals (**AA** = **Aa** = dominant, **aa** = recessive) : $\mu \approx sq^2$; if **h = 0.5**, there is no dominance, since then $W_1 = 1-0 = 1$, $W_2 = 1-(0+s)/2 = 1 - s/2$ and $W_3 = 1-s$; if **h = 1**, the situation is reduced to the case of partial selection against dominant individuals (with **AA** = recessive genotype , **Aa** = **aa** = dominant genotypes) : $\mu = sq^2 + qsh \approx qsh = qs$, since $sq^2 \approx 0$.

All formulae derived above can be straightforwardly obtained using the following intuitive argument : at equilibrium, the proportion of newly introduced genes in the population (and this proportion is obviously represented by the mutation rate μ) is exacly counterbalanced by the proportion of genes eliminated from the population per generation through selection. And this quantity equals sq^2 in the case of recessive genes and $2spq/2 = spq \approx sq$ (or **sp**) in the case of dominant genes. In fact, in the first case sq^2 represents the frequency of recessive individuals eliminated per generation (N_{aa} individuals of genotype **aa** in a population of size **N** : $sq^2 = N_{aa}/N$) and this is also the frequency of a genes eliminated per generation from the population gene pool, since each time an **aa** individual is eliminated from the population of size **N** two a genes are eliminated from the gene pool of size **2N** : $sq^2 = 2N_{aa}/2N = N_{aa}/N$. In relation to dominant genes, the proportion of dominant genes A eliminated per generation is simply the frequency $sp = s(2N_{AA}+N_{Aa})/2N$, where N_{AA} is the number of **AA** and N_{Aa} that of **Aa** individuals in the population. Since **p** is in this case very small, N_{AA}/N is a negligible quantity. The expression is then reduced to $sp = sN_{Aa}/2N$, where N_{Aa}/N is the population frequency of heterozygotes; and $sp = sN_{Aa}/N.1/2 = spq \approx p$, since $E(N_{Aa}/N) = 2pq$. Finally, in formula $\mu = sq^2 + qsh$, μ represents the frequency of **a** genes introduces in the population per generation, sq^2 the frequency of **a** genes eliminated throuqh homozygotes **aa**, and **sqh** the proportion of **a** genes eliminated through heterozygotes per generation. When the elimination of genes **a** is the same among homozygotes and heterozygotes, it comes out that $sq^2 = sqh$, $q = h$, $\mu = sq^2 + qsh = 2qsh = 2sq^2$ and $h = q = \sqrt{(\mu/2s)}$. Therefore, when **h > q** there is dominance (in relation to adaptive values), when **h < q** there is recessivity.

5) Practical application : calculation of mutation rates

The above notions enable us to estimate the mutation rates for deleterious recessive and dominant autosomal genes and for X-linked recessive alleles responsible for diseases.

5.1) Autosomal dominant genes

For the case of autosomal dominant genes with complete penetrance, the mutation rate μ can be determined by two different methods:

direct method : $\mu = x_1'/2$; and

indirect method : $\mu = sx_1"/2$,

where $x_1"$ is the frequency, at birth, of sporadic cases (attributable to newly arisen mutations and easily recognized when the penetrance is complete because both parents are normal); and $x_1"$ (in the second formula) is the overall frequency of affected individuals at birth, born to affected as well as non-affected parents. Since the dominant gene

present in any propositus has to be originated at some point of his or her genealogy, the heterozygosis probability can be expressed as a function of both **s** and **μ**, being the sum of the following terms:

$P(0) = 2\mu$ (the mutation occurred in one of the gametes that originated the propositus);

$P(1) = 2.2\mu.(1-s).1/2 = 2\mu(1-s)$: the mutation occurred in one of the gametes that originated the proband's parents and was transmitted to him or her with probability (1-s)/2 ; therefore,

$P(2) = 2\mu(1-s)^2$,..., **$P(n) = 2\mu(1-s)^n$** and

$$\Sigma P(i) = P(0) + P(1) + P(2) + \ldots$$
$$= 2\mu[(1-s)^0 + (1-s)^1 + (1-s)^2 + \ldots]$$
$$= 2\mu/s.$$

This is the frequency of heterozygotes expressed as a function of both **μ** and **s**. On the other hand, the frequency of heterozygotes as a function of gene frequencies is given by **$P(Aa) = 2pq = x_1"$**. Equating these two quantities we obtain **$2\mu/s = 2pq = x_1"$** and **$\mu = spq = sx_1"/2$**. If the coefficient of selection **s** has value **s = 1** (fully lethal dominant condition), only the first term **P(0)** is to be considered and **$P(0) = P(Aa) = 2\mu = 2pq$** , **$\mu = pq \approx p = x"_1/2$**.

5.2) Autosomal recessive genes

For the case of deleterious autosomal recessive genes, the formula becomes **$\mu = sx_2 = sq^2$**, where **x_2** is the frequency of affected individuals at birth, born to non-consanguineous parents. If the parents belong to an inbred population, then **x_2** takes value **$x_2 = q^2 + Fpq$** and **$\mu = sx_2 = s(q^2 + Fpq)$** , where **F** is the population inbreeding coefficient or fixation index.

5.3) X-linked recessive genes

For the case of deleterious X-linked recessive genes, the formula becomes **$\mu = sx_3/3$** , where **x_3** is the frequency of affected individuals among all males of the population. Since **$x_3 = q$** , the formula can also be written as **$\mu = sq/3$** . The first argument used to derive this formula is purely intuitive : since **$sq = sx_3$** is the frequency with which the gene is eliminated in the masculine population and since the males carry **1/3** of all X chromosomes of the whole population with sex ratio **1:1**, it comes out that **$\mu = sx_3/3$**.

Another way of obtaining the formula is described in the lines that follow. If we let **$P_n(Aa)$** be the frequency of (normal) heterozygous females and **$P_n(a)$** the frequency of affected males at generation **n**, we get the following system of recursion equations:

$$P_{n+1}(Aa) = 1/2.P_n(Aa) + (1-s).P_n(a) + 2\mu$$
$$P_{n+1}(a) = 1/2.P_n(Aa) + \mu .$$

At equilibrium, since

$P_{n+1}(Aa) = P_n(Aa) = P(Aa)$ and $P_{n+1}(a) = P_n(a) = P(a)$,

we obtain successively

$P(Aa) = 2.(1-s).P(a) + 4\mu$,
$P(a) = 1/2.P(Aa) + \mu$,
$P(Aa) = 2\mu(3-s)/s$;

therefore,

$P(a) = 1/2.P(Aa) + \mu = \mu(3-s)/s + \mu = 3\mu/s$

and

$\mu = s.P(a)/3 = sx_3/3 = sq/3$.

A third manner to obtain this result is similar to the one presented in the case of autosomal dominant genes : the probability of hemizygosis for the gene **a** for any male newborn is the sum of the following probabilities:

P(0) = μ : the child is affected because a mutation occurred in the X chromosome he inherited from his mother;

P(1) = μ : the child is affected because a mutation occurred in any of the two X chromosomes that originated his mother (probability **2μ**) and was then transmitted to him (probability **1/2**);

P(2) = μ(2-s)/2 : the child is affected because a mutation occurred in either the only X chromosome his maternal grandfather received (**μ**), was transmitted to his mother [**(1-s).1**] and then to him (**1/2**) or in one of the two X chromosomes that originated his maternal grandmother (**2μ**), was transmitted to his mother (**1/2**) and then to him (**1/2**): **μ(1-s)/2 + μ/2 = μ(2-s)/2**; in a similar way we obtain the successive terms

P(3) = μ(4-3s)/4 ,

P(4) = μ(8-9s+2s²)/8 , and so on.

It is not difficult to verify that the terms of this series satisfy the recurrence equation

P(n+2) = 1/2.P(n+1) + 1/2.(1-s).P(n) ,

which has the general solution

$P(n) = C_1 r_1^n + C_2 r_2^n$, where

$r_1 = [1 + \sqrt{(9-8s)}]/4$,

$r_2 = [1 - \sqrt{(9-8s)}]/4$,

$C_1 = [P(1) - P(0).r_2]/(r_1-r_2) = \mu(1-r_2)/(r_1-r_2)$, and

$C_2 = [P(0).r_1 - P(1)]/(r_1-r_2) = -\mu(1-r_1)/(r1-r2)$.

Since $|r_1, r_2| < 1$ for $0 < s < 1$, it comes out that the sum of all terms from zero to infinity is

$$\Sigma P(i) = P(0) + P(1) + P(2) + \ldots$$
$$= C_1/(1-r_1) + C_2/(1-r_2)$$
$$= \mu(1-r_2)/[(r_1-r_2)(1-r_1)] - \mu(1-r_1)/[(r_1-r_2)(1-r_2)]$$
$$= \mu[(1-r_2)^2 - (1-r_1)^2]/[(r_1-r_2)(1-r_1)(1-r_2)]$$
$$= \mu(2-r_1-r_2)/[(1-r_1)(1-r_2)]$$
$$= 3\mu/s \ .$$

Therefore, since $P(a) = P(0) + P(1) + \ldots$, it comes out that $P(a) = 3\mu/s$, as stated.

One of the most important causes of error in the estimation of mutation rates for autosomal recessive genes takes place when there is selection (even small) against **Aa** heterozygotes and one estimates the mutation rate solely based on the selective disadvantage of homozygotes **aa**. Let us suppose, for example, that individuals with a certain autosomal recessive disease have an adaptive value of 90%, occurring at birth in the offspring of non-related individuals with a frequency of 1/10000. If we estimate the mutation rate using the formula $\mu = sq^2$ we get the figure of $\mu = 10^{-5}$. Let us suppose, however, that heterozygotes Aa have a negligible selective disadvantage, say $W_2 = W_{Aa} = 0.99$. Taking this fact into account and using the formula $\mu = sq^2 + sqh$ with the values $s = 0.1$, $q = 0.01$ and $sh = 0.01$, we get the actual figure of $\mu = 10^{-5} + 10^{-4} = 11 \times 10\text{-}5$, that is, a value 11 times larger than the estimate obtained using the formula $\mu = sq^2$. The elimination of genes a through **Aa** heterozygotes (in spite of their small selective disadvantage) is in this example $10/1 = 10$ times larger than the elimination through **aa** homozygotes. As pointed out before, in order to determine the point in which the gene elimination is the same through heterozygotes and homozygotes, we put $qsh = sq^2$, obtaining thus the desired condition $h = q$.

6) Overdominance ($WAA = 1-s_1$, $WAa = 1$, $Waa = 1-s_3$)

In this case, mutation produces a negligible shifting from the equilibrium points [$P = s_3/(s_1+s_3)$ and $Q = s_1/(s_1+s_3)$] obtained when the mutation rate μ is assumed to be zero, as we show below.

In the overdominance model, after selection has occurred gene frequencies become

$p' = p(1-s_1p)/w$ and
$q' = q(1-s_3q)/w$, where
$w = 1 - s_1p^2 - s_3q^2$.

After mutation $[P(A \rightarrow a) = \mu]$ has taken place, the new gene frequencies are

$p'' = p(1-s_1p)(1-\mu)/w$ and
$q'' = [q(1-s_3q) + \mu p(1-s_1p)]/w$ so that
$q''/p'' = (q/p).(1-s_3q)/[(1-s_1p)(1-\mu)] + \mu/(1-\mu)$.

At equilibrium, $p'' = p$, $q'' = q$ and

$$q/p = (q/p).(1-s_3q)/[(1-s_1p)(1-\mu)] + \mu/(1-\mu) \ .$$

After a few algebraic manipulations we obtain

$$(q-\mu)/q = (1-s_3q)/(1-s_1p) \quad \text{and}$$

$$q^2(s_1+s_3) - s_1q(1+\mu) - \mu(1-s_1) = 0 \ .$$

The pertinent solution ($0 < q < 1$) of this quadratic equation is the equilibrium gene frequency q taking into account both selection and mutation:

$$q = \{s_1(1+\mu) + \sqrt{[s_1^2(1+\mu)^2+4\mu(1-s_1)(s_1+s_3)]}\}/2(s_1+s_3) \ .$$

This expression can be rewritten as

$$q = s_1(1+\mu)/2(s_1+s_3) + s_1(1+\mu)/2(s_1+s_3) \ . \ \sqrt{\{1+4\mu(1-s_1)(s_1+s_3)/[s_1^2(1+\mu)^2]\}} \ ;$$

since $\mu \to 0$, the leftmost expression ($\sqrt{\{...\}}$) can be replaced, without loss of accuracy, by $1 + 2\mu(1-s_1)(s_1+s_3)/[s_1^2(1+\mu)^2]$. After a few algebraic manipulations, we get

$$q^* = s_1/(s_1+s_3) + \mu[\ 1/(s_1+s_3) + (1-s_1)/s_1(1+\mu)\ ]$$
$$\approx s_1/(s_1+s_3) + \mu[\ 1/(s_1+s_3) + (1-s_1)/s_1\ ] \ .$$

Since $Q = s_1/(s_1+s_3)$ is the equilibrium gene frequency without taking into account the counterbalancing effects of mutation,

$$\Delta_1 = |q^*-Q| = \mu[\ 1/(s_1+s_3) + (1-s_1)/s_1\ ]$$

can be defined as the absolute difference between equilibrium gene frequencies taking or not taking into account the mutation pressure. As the following table, the values that $\Delta = |q-Q|$ and $\Delta_1 = |q^*-Q|$ (where Δ is the exact figure) take are always negligible as compared to the value of q and q^*.

μ	s_1	s_3	Q	q	q^*	Δ	Δ_1
0.000001	0.1000	0.1000	0.5000	0.5000	0.5000	0.000009	0.000014
0.000001	0.1000	0.3000	0.2500	0.2500	0.2500	0.000009	0.000011
0.000001	0.1000	0.5000	0.1667	0.1667	0.1667	0.000009	0.000011
0.000001	0.1000	0.7000	0.1250	0.1250	0.1250	0.000009	0.000010
0.000001	0.1000	0.9000	0.1000	0.1000	0.1000	0.000009	0.000010
0.000001	0.3000	0.1000	0.7500	0.7500	0.7500	0.000003	0.000005
0.000001	0.3000	0.3000	0.5000	0.5000	0.5000	0.000003	0.000004
0.000001	0.3000	0.5000	0.3750	0.3750	0.3750	0.000003	0.000004
0.000001	0.3000	0.7000	0.3000	0.3000	0.3000	0.000003	0.000003
0.000001	0.3000	0.9000	0.2500	0.2500	0.2500	0.000003	0.000003
0.000001	0.5000	0.1000	0.8333	0.8333	0.8333	0.000002	0.000003
0.000001	0.5000	0.3000	0.6250	0.6250	0.6250	0.000002	0.000002
0.000001	0.5000	0.5000	0.5000	0.5000	0.5000	0.000001	0.000002
0.000001	0.5000	0.7000	0.4167	0.4167	0.4167	0.000001	0.000002
0.000001	0.5000	0.9000	0.3571	0.3571	0.3571	0.000001	0.000002
0.000001	0.7000	0.1000	0.8750	0.8750	0.8750	0.000001	0.000002
0.000001	0.7000	0.3000	0.7000	0.7000	0.7000	0.000001	0.000001
0.000001	0.7000	0.5000	0.5833	0.5833	0.5833	0.000001	0.000001
0.000001	0.7000	0.7000	0.5000	0.5000	0.5000	0.000001	0.000001
0.000001	0.7000	0.9000	0.4375	0.4375	0.4375	0.000001	0.000001
0.000001	0.9000	0.1000	0.9000	0.9000	0.9000	0.000001	0.000001
0.000001	0.9000	0.3000	0.7500	0.7500	0.7500	0.000001	0.000001

```
0.000001  0.9000  0.5000  0.6429  0.6429  0.6429  0.000001  0.000001
0.000001  0.9000  0.7000  0.5625  0.5625  0.5625  0.000001  0.000001
0.000001  0.9000  0.9000  0.5000  0.5000  0.5000  0.000001  0.000001
0.000010  0.1000  0.1000  0.5000  0.5001  0.5001  0.000095  0.000140
0.000010  0.1000  0.3000  0.2500  0.2501  0.2501  0.000092  0.000115
0.000010  0.1000  0.5000  0.1667  0.1668  0.1668  0.000092  0.000107
0.000010  0.1000  0.7000  0.1250  0.1251  0.1251  0.000091  0.000103
0.000010  0.1000  0.9000  0.1000  0.1001  0.1001  0.000091  0.000100
0.000010  0.3000  0.1000  0.7500  0.7500  0.7500  0.000031  0.000048
0.000010  0.3000  0.3000  0.5000  0.5000  0.5000  0.000028  0.000040
0.000010  0.3000  0.5000  0.3750  0.3750  0.3750  0.000027  0.000036
0.000010  0.3000  0.7000  0.3000  0.3000  0.3000  0.000026  0.000033
0.000010  0.3000  0.9000  0.2500  0.2500  0.2500  0.000026  0.000032
0.000010  0.5000  0.1000  0.8333  0.8334  0.8334  0.000018  0.000027
0.000010  0.5000  0.3000  0.6250  0.6250  0.6250  0.000016  0.000023
0.000010  0.5000  0.5000  0.5000  0.5000  0.5000  0.000015  0.000020
0.000010  0.5000  0.7000  0.4167  0.4167  0.4167  0.000014  0.000018
0.000010  0.5000  0.9000  0.3571  0.3572  0.3572  0.000014  0.000017
0.000010  0.7000  0.1000  0.8750  0.8750  0.8750  0.000013  0.000017
0.000010  0.7000  0.3000  0.7000  0.7000  0.7000  0.000011  0.000014
0.000010  0.7000  0.5000  0.5833  0.5833  0.5833  0.000010  0.000013
0.000010  0.7000  0.7000  0.5000  0.5000  0.5000  0.000009  0.000011
0.000010  0.7000  0.9000  0.4375  0.4375  0.4375  0.000009  0.000011
0.000010  0.9000  0.1000  0.9000  0.9000  0.9000  0.000010  0.000011
0.000010  0.9000  0.3000  0.7500  0.7500  0.7500  0.000009  0.000009
0.000010  0.9000  0.5000  0.6429  0.6429  0.6429  0.000008  0.000008
0.000010  0.9000  0.7000  0.5625  0.5625  0.5625  0.000007  0.000007
0.000010  0.9000  0.9000  0.5000  0.5000  0.5000  0.000006  0.000007
0.000100  0.1000  0.1000  0.5000  0.5009  0.5014  0.000948  0.001400
0.000100  0.1000  0.3000  0.2500  0.2509  0.2512  0.000922  0.001150
0.000100  0.1000  0.5000  0.1667  0.1676  0.1677  0.000912  0.001067
0.000100  0.1000  0.7000  0.1250  0.1259  0.1260  0.000906  0.001025
0.000100  0.1000  0.9000  0.1000  0.1009  0.1010  0.000902  0.001000
0.000100  0.3000  0.1000  0.7500  0.7503  0.7505  0.000308  0.000483
0.000100  0.3000  0.3000  0.5000  0.5003  0.5004  0.000283  0.000400
0.000100  0.3000  0.5000  0.3750  0.3753  0.3754  0.000271  0.000358
0.000100  0.3000  0.7000  0.3000  0.3003  0.3003  0.000263  0.000333
0.000100  0.3000  0.9000  0.2500  0.2503  0.2503  0.000258  0.000317
0.000100  0.5000  0.1000  0.8333  0.8335  0.8336  0.000183  0.000267
0.000100  0.5000  0.3000  0.6250  0.6252  0.6252  0.000162  0.000225
0.000100  0.5000  0.5000  0.5000  0.5001  0.5002  0.000150  0.000200
0.000100  0.5000  0.7000  0.4167  0.4168  0.4169  0.000142  0.000183
0.000100  0.5000  0.9000  0.3571  0.3573  0.3573  0.000136  0.000171
0.000100  0.7000  0.1000  0.8750  0.8751  0.8752  0.000130  0.000168
0.000100  0.7000  0.3000  0.7000  0.7001  0.7001  0.000113  0.000143
0.000100  0.7000  0.5000  0.5833  0.5834  0.5835  0.000101  0.000126
0.000100  0.7000  0.7000  0.5000  0.5001  0.5001  0.000093  0.000114
0.000100  0.7000  0.9000  0.4375  0.4376  0.4376  0.000087  0.000105
0.000100  0.9000  0.1000  0.9000  0.9001  0.9001  0.000101  0.000111
0.000100  0.9000  0.3000  0.7500  0.7501  0.7501  0.000086  0.000094
0.000100  0.9000  0.5000  0.6429  0.6429  0.6429  0.000075  0.000083
0.000100  0.9000  0.7000  0.5625  0.5626  0.5626  0.000067  0.000074
0.000100  0.9000  0.9000  0.5000  0.5001  0.5001  0.000061  0.000067
-------------------------------------------------------------------------
```

This table was generated by the BASIC code

```
REM PROGRAM FILENAME OVERMUSE.BAS
DEFDBL A-Z: CLS : U = 1 / 1000000
LOOPHERE: FOR V = 1 TO 9 STEP 2: FOR W = 1 TO 9 STEP 2
S1 = V / 10: S3 = W / 10: Q = S1 / (S1 + S3)
Q1 = (SQR((S1 * (1 + U)) ^ 2 + 4 * U * (1 - S1) * (S1 + S3)) + S1 * (1 + U))
Q1 = Q1 / (2 * (S1 + S3)): Q2 = Q + U * (1 / (S1 + S3) + (1 - S1) / S1)
DELTA1 = ABS(Q - Q1): DELTA2 = ABS(Q - Q2)
PRINT USING "#.######  "; U;PRINT USING "#.####  "; S1; S3;
PRINT USING "#.####  "; Q; Q1; Q2;PRINT USING "#.######  "; DELTA1; DELTA2
DO: LOOP WHILE INKEY$ <> " "
NEXT W, V: U = U * 10: IF U < .0001 THEN GOTO LOOPHERE ELSE END
```

IDENTIFICATION AND FORENSIC APPLICATIONS

1) Identity

1.a) Probability of Identity Exclusion [$P(E_1)$]

Let $\{a_i\}$ be the i-th codominant allele out of the possible n segregating at an autosomal locus in a panmictic population; then, $P(a_ia_i) = P^2(a_i) = p_i^2$, $P(a_ia_j) = 2P(a_i)P(a_j) = 2p_ip_j$, $j \neq i$.

Then, given that an individual is falsely accused of a crime and that there is no biological relationship between him and the real perpretator,

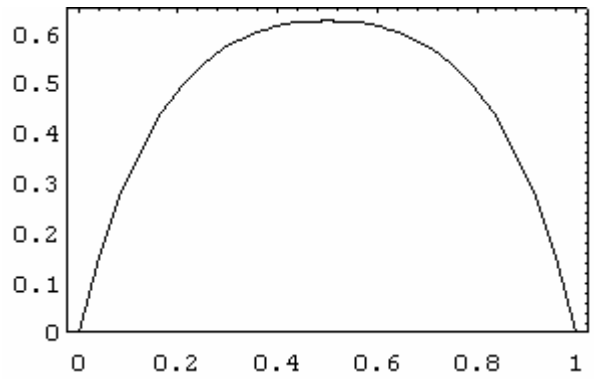
$$P(E_1) = P(a_1a_1)[1-P(a_1a_1)] + \ldots + P(a_na_n)[1-P(a_na_n)]$$
$$= 1 - \Sigma p_i^4 - 2\Sigma\Sigma p_i^2 p_j^2$$
$$= 1 - 2(\Sigma p_i^2)^2 + \Sigma p_i^4 \ .$$

For the special case n = 2, $p_1 = p$, $p_2 = q = 1-p$, and the above expression simplifies to

$$PE(_1) = 2\theta(2-3\theta) \ , \ \theta = pq \ .$$

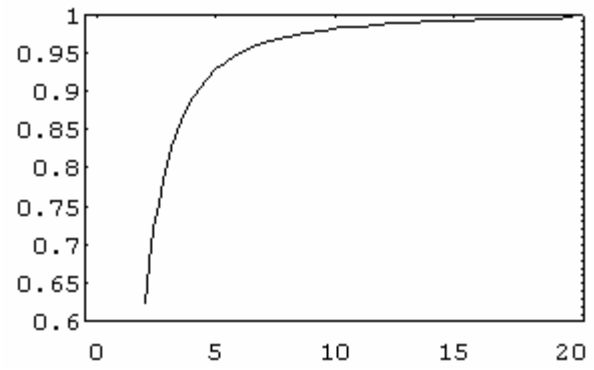
The graph below, generated by the enclosed Mathematica code, shows the values $P(E_1) = 2\theta(2-3\theta) = 2q(1-q)[2-3q(1-q)]$ as function of the argument q in the interval (0,1). The maximum value of $P(E_1)$ takes place when p = q = 0.5, since $dP(E_1)/dq = dP(E_1)/d\theta \ . \ d\theta/dq = (4-12\theta).(1-2q) = (1-3q+3q^2).(1-2q)$ and the only real root of the equation $4(1-3q+3q^2)(1-2q) = 0$ is q = 0.5. For p = q = 0.5, $P(E_1)$ takes the value 5/8 = 0.625.

```
Plot[2*q*(1-q)*(2-3*q*(1-q)),{q,0,1},
     PlotRange->{0,0.65},
     AxesOrigin->{0,0},
     Frame->True]
```



For the general case with n alleles, the probability of exclusion is at a maximum when all allelic frequencies are equal: $p_1 = ... = p_n = 1/n$. Then, $\mathbf{P_{max}(E_1) = 1 - 2(\Sigma p_i^2)^2 + \Sigma p_i^4 = 1 - 2(n.1/n^2)^2 + n.1/n^4 = 1 - (2n-1)/n^3}$. For n = 2 this expression has value 1 - 3/8 = 5/8 = 0.625 as expected. The figure that follows shows the values of $\mathbf{P_{max}(E_1) = 1 - (2n-1)/n^3}$ as function of n.

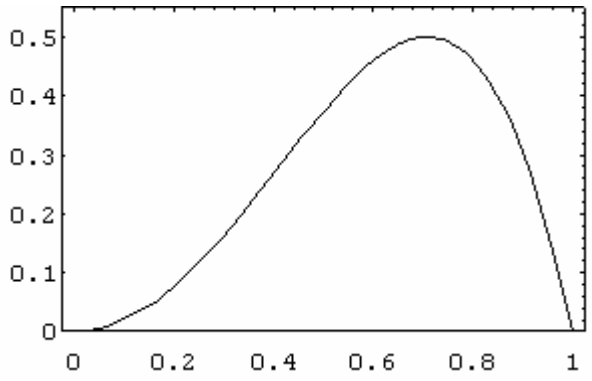
```
Plot[1-(2*k-1)/k^3,{k,2,20},
     PlotRange->{0.6,1},
     AxesOrigin->{0,0},
     Frame->True]
```



For the special case of a pair of alleles with dominance, $\mathbf{P(E_1)}$ takes the literal value $\mathbf{P(E_1) = q^2(1-q^2) + (1-q^2)q^2 = 2q^2(1-q^2) = 2pq^2(1+q)}$, where $\mathbf{q = 1-p}$ is the frequency of the recessive allele. The value of $\mathbf{q}$ that maximizes $\mathbf{P(E_1)}$ is taken straightforwardly from $\mathbf{dP(E_1)/dq = 4q(1-2q^2) = 0}$ $\mathbf{\therefore\ q = \sqrt{(1/2)} = 0.7071}$; for this value of q, $\mathbf{P_{max}(E_1) = 2.1/2.(1-1/2) = 0.5}$.

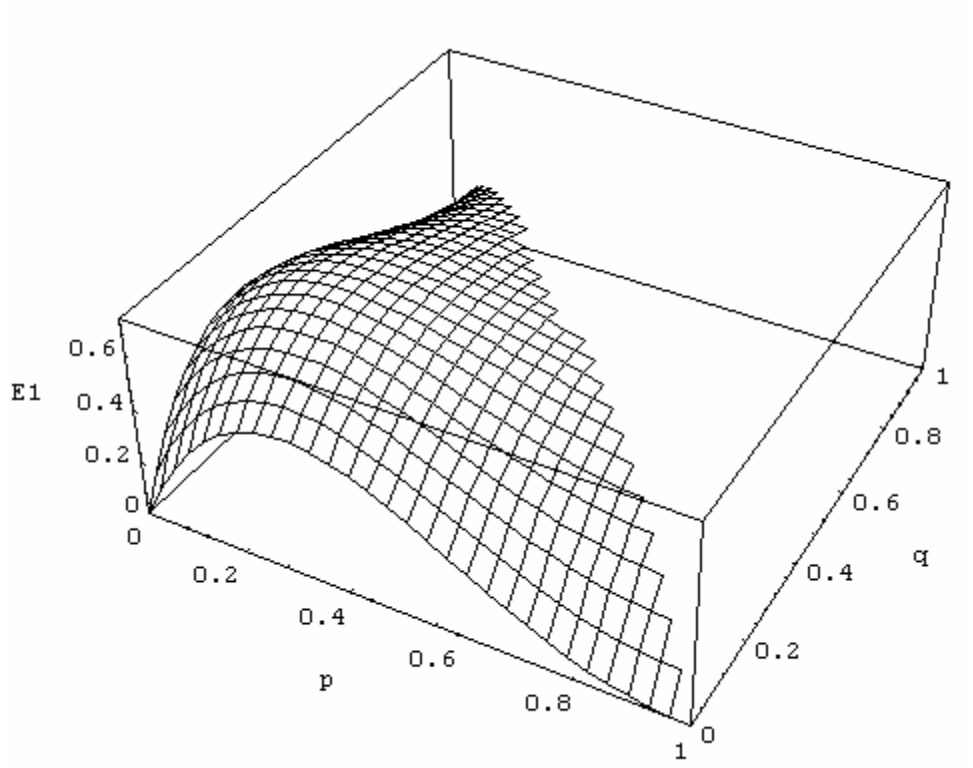
The figure below shows the values of $\mathbf{P(E_1)}$ as function of q for the case of two autosomal alleles with dominance.

```
Plot[2*q^2*(1-q^2),{q,0,1},
     PlotRange->{0,0.51},
     AxesOrigin->{0,0},
     Frame->True]
```



Another special case is given by the ABO blood-group system; then, $\mathbf{P(E_1) = P(A)[1-P(A)]+P(B)[1-P(B)]+P(AB)[1-P(AB)]+P(O)[1-P(O)] = (p^2+2pr)[1-(p^2+2pr)] + (q^2+2qr)[1-(q^2+2qr)] + 2pq(1-2pq) + r^2(1-r^2) = 2pq(2-3pq) + 2r^2(1-r^2) + 4pqr}$, where **p**, **q**, and $\mathbf{r = 1-p-q}$ are the frequencies of the **A**, **B** and **O** alleles. The graph below, originated by the enclosed Mathematica code, shows the values of $\mathbf{P(E_1)}$ as function of ($\mathbf{0 < p < 1}$) and ($\mathbf{0 < q < 1}$), with the restriction $\mathbf{p + q \leq 1}$.

```
f[p_,q_]:= 2 * p * q * (2 - 3 * p * q) +
           2 * (1 - p - q)^2 * (1 - (1 - p - q)^2) +
           4 * p * q * (1 - p - q) /; p+q <=1
Plot3D[f[p,q], {p, 0, 1}, {q, 0, 1},
       PlotRange -> {0, 0.74}, AxesLabel -> {"p  ", "  q", "E1  "},
       Shading -> False, PlotPoints -> 30]
```

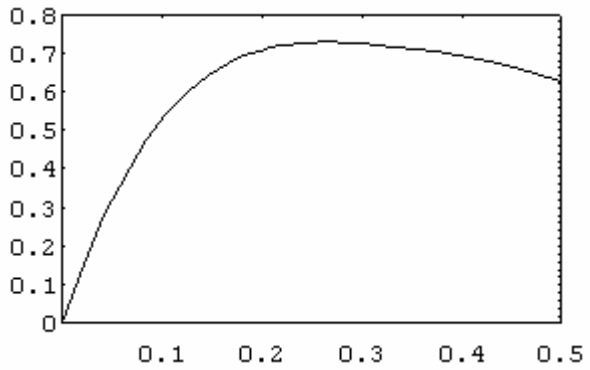


Putting $\partial P(E_1)/\partial p = 0$ and $\partial P(E_1)/\partial q = 0$ and solving this set of equations we obtain the values **p** and **q** that maximize $P(E_1)$. Since it is intuitive that both frequencies should then be equal, we can make, in $P(E_1) = 2pq(2-3pq) + 2r^2(1-r^2) + 4pqr$, p = q and r = 1-p-q = 1-2q, obtaining thus $P(E_1|p=q,r=1-2q) = 8q - 32q^2 + 56q^3 - 38q^4$. The only real root of the equation $dP(E_1|p=q,r=1-2q)/dq = 0$ is $q = 0.26632$, as shown by the following Mathematica code:

```
dedq = D[8*q - 32q^2 + 56*q^3 - 38*q^4,q];
N[Solve[dedq==0]]
{{q -> 0.26632}, {q -> 0.419471 + 0.147203 i}, {q -> 0.419471 - 0.147203 i}}
```

For $p = q = 0.26632$ and $r = 1-p-q = 1-2q = 0.46736$, $P_{max}(E_1) = 0.7275$, which is the maximum value $P(E_1)$ can take, as shown by the graph of the function $P(E_1|p=q,r=1-2q) = 8q - 32q^2 + 56q^3 - 38q^4$.

```
Plot[8*q - 32*q^2 + 56*q^3 - 38*q^4, {q,0,0.5},
     PlotRange->{{0,0.5},{0,0.8}},
     AxesOrigin->{0,0}, Frame->True]
```



Using N different genetic systems, the joint probability of exclusion is given by $\mathbf{P_{tot}(E_1) = 1 - \Pi_i[1-P_i(E_1)]}$. This result is also valid for any problem of biological relationship exclusion, taking in general the form $\mathbf{P_{tot}(E_j) = 1 - \Pi_i[1-P_i(E_j)]}$, where j denotes the exclusion situation.

1.b) Probability of True Identity for Individuals not Excluded

The accused individual has genotype $\mathbf{a_i a_j}$; given that the sampled material has the same genotype, the conditional probabilities **P(T)** and **P(F)** that the material belongs to him or her (true identity) or not (false identity) are in the ratios $\mathbf{P(T) : P(F) :: 1 : P(a_i a_j)}$, that is, $\mathbf{P(T) = 1/[1+P(a_i a_j)] = 1/(1+p_i^2)}$ if $\mathbf{i = j}$ and $\mathbf{P(T) = 1/(1+2p_i p_j)}$ if $\mathbf{i \neq j}$; $\mathbf{P(F) = P(a_i a_j)/[1+P(a_i a_j)] = p_i^2/(1+p_i^2)}$ **if** $\mathbf{i = j}$ and $\mathbf{P(F) = 2p_i p_j/(1+2p_i p_j)}$ if $\mathbf{i \neq j}$. Since $\mathbf{p_i \leq 1}$, $\mathbf{P(T) \geq P(F)}$. If we use the logical operator δ_{ij} {δ_{ij} = 1 if i = j, 0 otherwise} , both formulae reduce to $\mathbf{P(T) = 1/[1+(2-\delta_{ij})p_i p_j]}$. If all the n codominant alleles segregating at this locus occur with equal frequencies, p_i = 1/n, $\mathbf{P(T) = n^2/(1+n^2)}$ if he or she is a homozygote and $\mathbf{P(T) = n^2/(2+n^2)}$ if he or she is a heterozygote.

If two systems (v.g., $\{a_i\},\{b_i\}$) are used and the identity is not excluded, it comes out that $\mathbf{P_a(T) = 1/[1+(2-\delta_{ij})p_i p_j]}$ and $\mathbf{P_b(T) = 1/[1+(2-\delta_{ij})q_i q_j]}$. By applying Bayes' theorem to these results we obtain

	T	F
a	1	$(2-\delta_{ij})p_i p_j$
b	1	$(2-\delta_{ij})q_i q_j$
total	1	$(2-\delta_{ij})p_i p_j . (2-\delta_{ij})q_i q_j$

so that $P_{ab}(T) = 1/[1+(2-\delta_{ij})p_ip_j.(2-\delta_{ij})q_iq_j]$

$= 1/\{1 + [P_a(F).P_b(F)]/[P_a(T).P_b(T)]\}$

$= 1/[1 + P_a(F)/P_a(T) \ . \ P_b(F)/P_b(T)]$

$= 1/\{1 + [1 - P_a(T)]/P_a(T) \ . \ [1-P_b(T)]/P_b(T)\}$

$= P_a(T).P_b(T)/\{P_a(T).P_b(T) + [1-P_a(T)].[1-P_b(T)]\}$.

The above formula can be generalized easily for the case of N different systems used simultaneously:

$$P_{1...N}(T) = 1/\{1 + \prod_i[P_i(F)/P_i(T)]\} = \{1 + \prod_i[P_i(F)/P_i(T)]\}^{-1}.$$

The ratio $P_0(F)/P_0(T)$ of prior probabilities can also be introduced in the product $\prod_i[P_i(F)/P_i(T)]$ to give the final probabilities favoring false and true biological relationship.

2) Monozygosity

2.a) - Probability of Monozygosity Exclusion for Dizygotic Twins [$P(E_2)$]

The table below shows the probabilities of the six possible pairs of genotypes of dizygotic twins in the generalized case of n autosomal alleles , where the subscripts **i**, **j** and **k** indicate different alleles and gene frequencies ($j \neq i$, $k \neq i,j$):

genotypes	P
a_ia_i,a_ia_i	$p_i^2(1+p_i)^2/4$
a_ia_i,a_ja_j	$p_i^2p_j^2/2$
a_ia_i,a_ia_j	$p_i^2p_j(1+p_i)$
a_ia_i,a_ja_k	$2p_i^2p_ip_k$
a_ia_j,a_ia_j	$p_ip_j(1+p_i+p_j+2p_ip_j)/2$
a_ia_j,a_ia_k	$p_ip_jp_k(1+2p_i)$
a_ia_j,a_ka_l	$2p_ip_jp_kp_l$

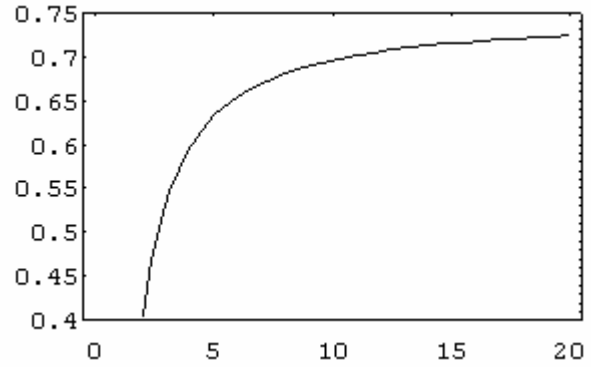
Monozygosity is excluded when the pair is genotypically discordant and this takes place with probability

$$P(E_2) = 1 - \Sigma p_i^2(1+p_i)^2/4 - \Sigma\Sigma p_ip_j(1+p_i+p_j+2p_ip_j)/4$$

$$= 1/4.[4 - (\Sigma p_i^2 + 2\Sigma p_i^3 + \Sigma p_i^4 + \Sigma\Sigma p_i p_j + \Sigma\Sigma p_i^2 p_j + \Sigma\Sigma p_i p_j^2 + 2\Sigma\Sigma p_i^2 p_j^2)]$$
$$= 3/4 - 1/2.\Sigma p_i^2 - 1/2.(\Sigma p_i^2)^2 + 1/4.\Sigma p_i^4 \ .$$

As in the previous item, the maximum value $P(E_2)$ can take occurs when all allelic frequencies are equal. Then, the expression for $P(E_2)$ takes the form $P_{max}(E_2) = 3/4 - 1/2.n.(1/n)^2 - 1/2.[n.(1/n)^2]^2 + 1/4.n.(1/n)^4 = 3/4 - (2n^2+2n-1)/4n^3$. As $n \to \infty$, $P(E_2)$ tends to **3/4** and the chance of two dizygotic twins (or sibs) having the same genotype is **1/4**, as expected, since then both parents are surely different heterozygotes (v.g., **AB** and **CD**) and the children will have the same genotype if and only if both receive the same allele combination from them: $P = P(AC,AC)+P(AD,AD)+P(BC,BC)+P(BD,BD) = 4.(1/2 \times 1/2)^2 = 1/4$. The figure that follows shows the values of $P_{max}(E_2) = 3/4 - (2n^2+2n-1)/4n^3$ as function of **n**.

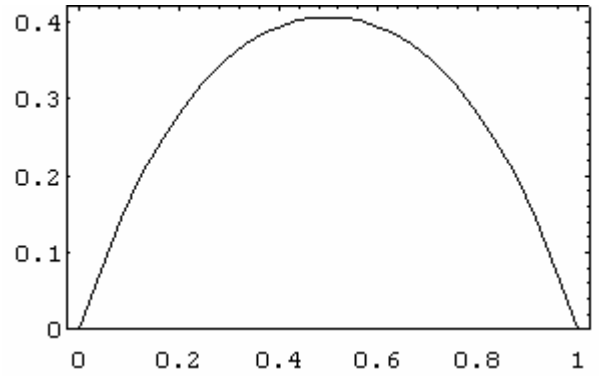
```
Plot[3/4 - (2*k^2 + 2k - 1)/(4*k^3),{k,2,20},
     PlotRange->{0.4,0.75},
     AxesOrigin->{0,0},
     Frame->True]
```



For the case of two autosomal alleles without dominance,
$P(E_2) = 2pq - 3p^2q^2/2 = 2\theta - 3\theta^2/2$, $\theta = pq$.

The graph below, generated by the enclosed Mathematica code, shows all possible values of $P(E_2)$ as function of q in the domain (0,1). The maximum value of $P(E_2)$ is 0.40625, taking place when p = q = 0.5, because q = 0.5 is the only real root of the equation $dP(E_2)/d\theta.d\theta/dq = (2-3\theta).(1-2q) = (2-3q+3q^2)(1-2q) = 0$.

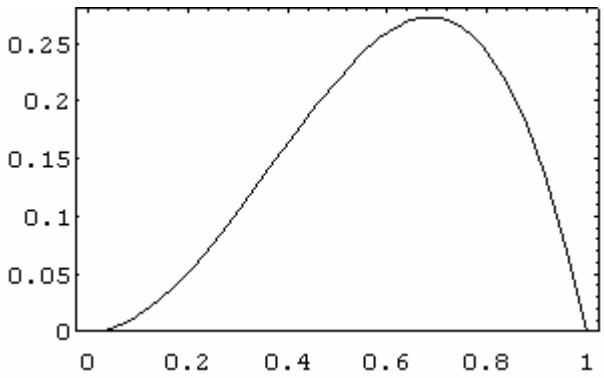
```
Plot[2*(1-q)*q-3*(1-q)^2*q^2/2,{q,0,1},
     PlotRange->{0,0.420},
     AxesOrigin->{0,0},
     Frame->True]
```



For the special case of a pair of alleles with dominance, exclusion of monozygosity occurs when one of the twins has the dominant phenotype and the other the recessive one; then, $\mathbf{P(E_2)}$ takes the literal value $\mathbf{P(E_2) = q^2(1-q)(3+q)/2}$, where $\mathbf{q = 1-p}$ is the frequency of the recessive allele. The value of $\mathbf{q}$ that maximizes $\mathbf{P(E_2)}$ is the pertinent root $(0 < q < 1)$ of the equation $\mathbf{dP(E_2)/dq = q(3-3q-2q^2) = 0 \therefore q = 0.6861}$; for this value of q, $\mathbf{P_{max}(E_2) = 0.2723}$.

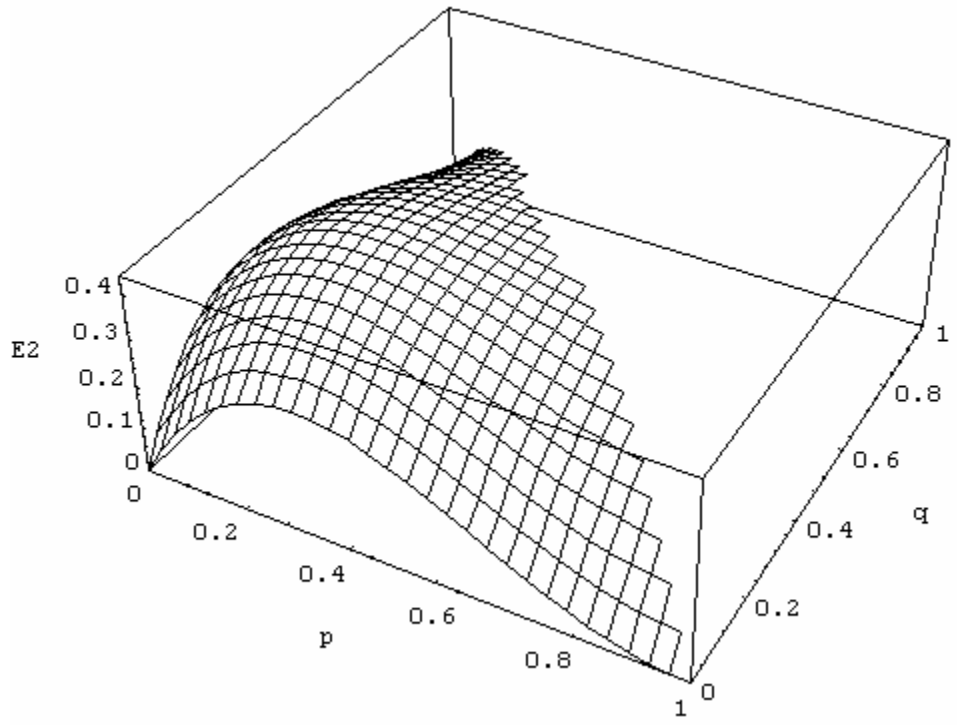
The figure below shows the values of $\mathbf{P(E_2)}$ as function of q for the case of two autosomal alleles with dominance.

```
Plot[q^2*(1-q)*(3+q)/2,{q,0,1},
     PlotRange->{0,0.28},
     AxesOrigin->{0,0},
     Frame->True]
```



Another special case is given by the ABO blood-group system; then, $\mathbf{P(E_2) = P[(A\text{-}B) + P(A\text{-}AB) + P(A\text{-}O) + P(B\text{-}AB) + P(B\text{-}O) + P(AB\text{-}O) = pq(pq+4r+2r^2)/2 + pq(p+r+p^2+2pr) + pr^2(2+p+2r)/2 + pq(q+r+q^2+2qr) + qr^2(2+q+2r)/2 + pqr^2 = pq(4+6r+2r^2-3pq)/2 + r2(2+p^2+q^2-2r^2)/2}$, where **p**, **q**, and **r = 1-p-q** are the frequencies of alleles **A**, **B** and **O**. The graph below, originated by the enclosed Mathematica code, shows the values of $\mathbf{P(E_2)}$ as function of (**0 < p < 1**) and (**0 < q < 1**), with the restriction **p + q ≤ 1**.

```
f[p_,q_]:= p * q * (4 + 6 * (1 - p - q) + 2 * (1 - p - q)^2 -
             3 * p * q)/2 + (1 - p - q)^2 * (2 + p^2 + q^2 -
             2 * (1 - p - q)^2)/2 /; p+q <=1
   Plot3D[f[p,q], {p, 0, 1}, {q, 0, 1},
          PlotRange -> {0, 0.44}, AxesLabel -> {"p  ", "  q", "E2  "},
          Shading -> False, PlotPoints -> 30]
```



Putting $\partial\mathbf{P(E_1)/\partial p = 0}$ and $\partial\mathbf{P(E_1)/\partial q = 0}$ and solving this set of equations we obtain the values **p** and **q** that maximize $\mathbf{P(E_2)}$. Since it is intuitive that both frequencies should be equal, we can make, in $\mathbf{P(E_2) = pq(4+6r+2r^2-3pq)/2 + r^2(2+p^2+q^2-2r^2)/2}$, p = q and r = 1-p-q = 1-2q, obtaining thus $\mathbf{P(E_2|p=q,r=1-2q) = 4q - 13q^2 + 18q^3 - 19q^4/2}$. The only

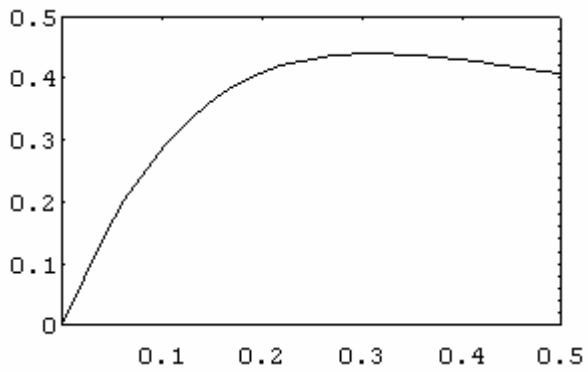
real root of the equation $\mathbf{dP(E_2|p=q,r=1-2q)/dq = 0}$ is $\mathbf{q = 0.3092}$, as shown by the following Mathematica code:

```
dedq = D[4*q - 13*q^2 + 18*q^3 - 19*q^4/2,q];
N[Solve[dedq==0]]
```

```
{{q -> 0.309238}, {q -> 0.555907+0.177094 i}, {q -> 0.555907-0.177094 i}}
```

For $\mathbf{p = q = 0.3092}$ and $\mathbf{r = 1-p-q = 1-2q = 0.3816}$, $\mathbf{P_{max}(E_2) = 0.4392}$, which is the maximum value $\mathbf{P(E_2)}$ can take, as shown by the graph of the function $\mathbf{P(E_2|p=q,r=1-2q) = 4q - 13q^2 + 18q^3 - 19q^4/2}$.

```
Plot[4*q - 13*q^2 + 18*q^3 - 19*q^4/2, {q,0,0.5},
     PlotRange->{{0,0.5},{0,0.5}},
     AxesOrigin->{0,0}, Frame->True]
```



2.b) Probability of Dizygosity for Twins not Excluded from Monozygosity

Given that the twins have the same genotype, the conditional probabilities favoring dizygosity (DZ) and monozygosity (MZ) are in the ratios (DZ/MZ):

genotypes	DZ	MZ	DZ/MZ
a_ia_i,a_ia_i	$p_i^2(1+p_i)^2/4$	p_i^2	$(1+p_i)^2/4$
a_ia_j,a_ia_j	$p_ip_j(1+p_i+p_j+2p_ip_j)/2$	$2p_ip_j$	$(1+p_i+p_j+2p_ip_j)/4$

When the number of codominant alleles is two (v.g., **M** and **N**, with frequencies **p** and **q**), the conditional probabilities shown above reduce to

genotypes	DZ	MZ	DZ/MZ
MM,MM	$p^2(1+p)^2/4$	p^2	$(1+p)^2/4$
MN,MN	$pq(1+pq)$	$2pq$	$(1+pq)/2$
NN,NN	$q^2(1+q)^2/4$	q^2	$(1+q)^2/4$

In the special case of a pair of autosomal alleles with dominance, the conditional probabilities shown above become

genotypes	DZ	MZ	DZ/MZ
D-,D-	$p(1+pq+pq^2/4)$	$p^2+2pq = p(1+q)$	$1-q^2(3+q)/4(1+q)$
dd,dd	$q^2(1+q)^2/4$	q^2	$(1+q)^2/4$

In the special case of the ABO blood group system, the conditional probabilities favoring dizygosity (DZ) and monozygosity (MZ) are in the ratios (DZ/MZ) shown in the following table, where, as before, **p**, **q**, and **r** stand for the frequencies of alleles **A**, **B** and **O**:

genotypes	DZ	MZ	DZ/MZ
A,*A*	$pr(1+2p)(2-q)/2+p^2(1+p)^2/4$	p^2+2pr	$[2r(1+2p)(2-q)+p(1+p)^2]/4(p+2r)$
B,*B*	$qr(1+2q)(2-p)/2+q^2(1+q)^2/4$	q^2+2qr	$[2r(1+2q)(2-p)+q(1+q)^2]/4(q+2r)$
AB,*AB*	$pq(2pq+2-r)/2$	$2pq$	$(2pq+2-r)/4$
O,*O*	$r^2(1+r)^2/4$	r^2	$(1+r)^2/4$

Usually in problems involving the testing of twins the genotypes of the parents are also determined, and the above probabilities calculated taking conditionally to the parents' genotypes:

parents	twins	DZ	MZ	DZ/MZ
a_ia_i,a_ia_i	a_ia_i	1	1	1
a_ia_i,a_ia_j	a_ia_i	1/2.1/2 = 1/4	1/2.1 = 1	1/2
a_ia_i,a_ia_j	a_ia_j	1/2.1/2 = 1/4	1/2.1 = 1	1/2
a_ia_i,a_ja_k	a_ia_j	1/2.1/2 = 1/4	1/2.1 = 1/2	1/2
a_ia_i,a_ja_k	a_ia_k	1/2.1/2 = 1/4	1/2.1 = 1/2	1/2
a_ia_j,a_ia_j	a_ia_i	1/4.1/4 = 1/16	1/4.1 = 1/4	1/4
a_ia_j,a_ia_j	a_ia_j	1/2.1/2 = 1/4	1/2.1 = 1/4	1/2
a_ia_j,a_ia_j	a_ja_j	1/4.1/4 = 1/16	1/4.1 = 1/4	1/4
a_ia_j,a_ia_k	a_ia_i	1/4.1/4 = 1/16	1/4.1 = 1/4	1/4
a_ia_j,a_ia_k	a_ia_k	1/4.1/4 = 1/16	1/4.1 = 1/4	1/4
a_ia_j,a_ia_k	a_ia_j	1/4.1/4 = 1/16	1/4.1 = 1/4	1/4
a_ia_j,a_ia_k	a_ja_k	1/4.1/4 = 1/16	1/4.1 = 1/4	1/4
a_ia_j,a_ka_l	a_ia_k	1/4.1/4 = 1/16	1/4.1 = 1/4	1/4

a_ia_j, a_ka_l	a_ja_k	1/4.1/4 = 1/16 1/4.1 = 1/4	1/4
a_ia_j, a_ka_l	a_ia_l	1/4.1/4 = 1/16 1/4.1 = 1/4	1/4
a_ia_j, a_ka_l	a_ja_l	1/4.1/4 = 1/16 1/4.1 = 1/4	1/4

The following table shows a numerical example, where the blood groups of ABO, MN and Rh systems have been determined in a pair of twins having the same sex and in their parents:

father	mother	twins	DZ	MZ	DZ/MZ
AB	O	A	1/2.1/2 = 1/4	1/2.1 = 1/2	1/2
MM	MN	MM	1/2.1/2 = 1/4	1/2.1 = 1/2	1/2
D-	dd	dd	1/2.1/2 = 1/4	1/2.1 = 1/2	1/2

Since dizygotic twin births are about two times more frequent than monozygotic twin births and taking into account that the conditional probabilities of a twin pair having the same sex is **1** under the MZ hypothesis and **1/2** under the DZ hypothesis, we get the following results:

Probabilities	DZ	MZ	
prior	2/3	1/3	(or 2:1)
conditional			
same sex	1/2	1	(or 1:2)
same blood groups|parents	1/8	1	(or 1:8)
joint	1/24	1/3	(or 1:8)

The final probability favoring the hypothesis of monozygosity is therefore **P(MZ) = 8/(1+8) = 8/9 = 0.8889.**

3) Maternity

3.a) Probability of Maternity Exclusion [P(E_3)]

Using the same notation as before, it comes out that, under the hypothesis that the mother is false and there is no biological relationship between her and the alleged child, **P(E_3)** can be taken straightforwardly from the probabilities associated with the events shown in the table below:

woman	child
a_1a_1	a_2a_2, a_2a_3,..., a_3a_3, a_3a_4,..., a_na_n
a_1a_2	a_3a_3, a_3a_4, a_4a_4, a_4a_5,..., a_na_n
a_2a_2	a_1a_1, a_1a_3, a_3a_3, a_3a_4,..., a_na_n
	...

It is not difficult to see that the probability $\mathbf{P(E_3)}$ takes value

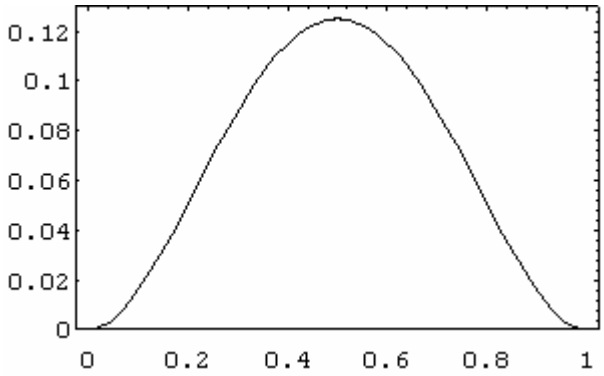
$$\mathbf{P(E_3)} = p_1^2(1-p_1)^2 + \ldots + p_4^2(1-p_4)^2 + \ldots$$
$$+ 2p_1p_2(1-p_1-p_2)^2 + \ldots + 2p_3p_4(1-p_3-p_4)^2 + \ldots$$
$$\mathbf{= \Sigma p_i^2(1-p_i)^2 + \Sigma\Sigma p_ip_j(1-p_i-p_j)^2}$$
$$\mathbf{= 1 - 4\Sigma p_i^2 + 4\Sigma p_i^3 - 3\Sigma p_i^4 + 2(\Sigma p_i^2)^2} \ .$$

For the case $\mathbf{n = 2}$, $\mathbf{P(E_3)}$ takes, as expected, value

$$\mathbf{P(E_3) = 2p^2q^2 - 2pq(1-p-q) = 2p^2q^2 = 2\theta^2,\ \theta = pq.}$$

The graph below, generated by the enclosed Mathematica code, shows all possible values of $\mathbf{P(E_3)}$ as function of q in the domain (0,1). The maximum value of $\mathbf{P(E_3)}$ is 0.125, taking place when p = q = 0.5, because q = 0.5 is the only pertinent root (0 < q < 1) of the equation $\mathbf{dP(E_3)/d\theta.d\theta/dq = 4q(1-q)(1-2q) = 0.}$

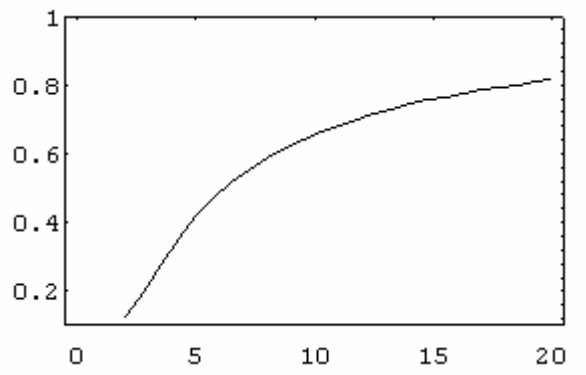
```
Plot[2*q^2*(1-q)^2,{q,0,1},
     PlotRange->{0,0.130},
     AxesOrigin->{0,0},
     Frame->True]
```



In the general case of n alleles, $\mathbf{P(E_3)}$ is at a maximum when all $\mathbf{p_i}$ in

$P(E_3) = \Sigma p_i^2(1-p_i)^2 + \Sigma\Sigma p_i p_j(1-p_i-p_j)^2$ are equal to 1/n; the value that $P(E_3)$ takes reduces then to $P_{max}(E_3) = n.(1/n)^2(1-1/n)^2+n(n-1).(1/n)^2.(1-2/n)^2 = [(n-1)^2+(n-1)(n-2)^2]/n^3$. For n = 2 this expression has value 1/8 = 0.125 as expected. The figure that follows shows the values of $P_{max}(E_3) = [(n-1)^2+(n-1)(n-2)^2]/n^3$ as function of n.

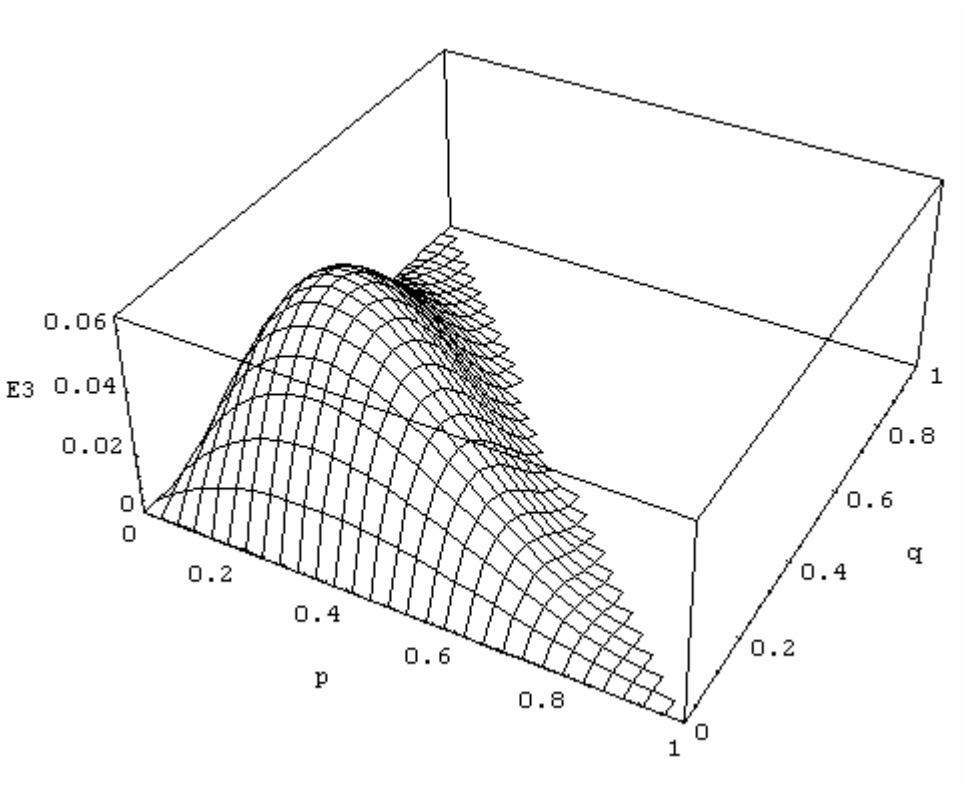
```
Plot[((k-1)^2+(k-1)*(k-2)^2)/k^3,{k,2,20},
     PlotRange->{0.1,1},
     AxesOrigin->{0,0},
     Frame->True]
```



For the case of a pair of alleles with dominance, maternity exclusion is impossible.

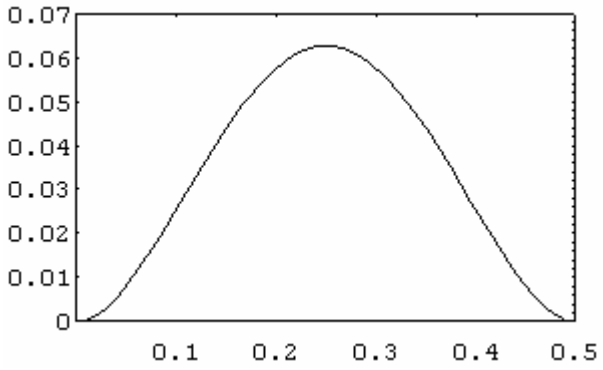
For the special case of ABO blood-group system, maternity is excluded only when the woman belongs to **AB** blood group and the child is **O** or vice-versa: $P(E_3) = P(AB).P(O) + P(O).P(AB) = 2pq.r^2 + 2pq.r^2 = 4pqr^2$, where **p**, **q**, and **r** are the frequencies of alleles A, B, and O. The graph below, originated by the enclosed Mathematica code, shows the values of $P(E_3)$ as function of ($0 < p < 1$) and ($0 < q < 1$), with the restriction $p + q \leq 1$.

```
f[p_,q_]:= 4 *p * q * (1 - p - q)^2 /; p+q <=1
Plot3D[f[p,q], {p, 0, 1}, {q, 0, 1},
       PlotRange -> {0, 0.064}, AxesLabel -> {"p  ", "  q", "E3  "},
       Shading -> False, PlotPoints -> 30]
```



As before, $\mathbf{P(E_3)}$ has its maximum when p = q and r = 1-2q; making these substitutions in the expression for $\mathbf{P(E_3)}$ we obtain $\mathbf{P(E_3|p=q,r=1-2q) = 4q^2(1-2q)^2}$. Putting $\mathbf{dP(E_3|p=q,r=1-2q)/dq = 0}$ we get $\mathbf{p = q = 0.25}$ and $\mathbf{r = 1-2q = 0.50}$; for these values, $\mathbf{P(E_3) = 1/16 = 0.0625}$, which is the maximum value it can take $\mathbf{[P_{max}(E_3)]}$, as shown by the graph of the function $\mathbf{P(E_3|p=q,r=1-2q) = 4q^2(1-2q)^2}$.

```
Plot[4*q^2*(1-2*q)^2, {q,0,0.5},
     PlotRange->{{0,0.5},{0,0.07}},
     AxesOrigin->{0,0}, Frame->True]
```



3.b) Probability of True Maternity for Individuals not Excluded

For computing this probability we begin by deriving the distribution of mother-child pairs in a panmictic population in the case of n alleles, where the subscripts i, j and k indicate different alleles and frequencies ($j \neq i$ and $k \neq i,j$):

mother	child	P
a_ia_i	a_ia_i	p_i^3
a_ia_i	a_ia_j	$p_i^2p_j$
a_ia_j	a_ia_i	$p_i^2p_j$
a_ia_j	a_ia_j	$p_ip_j(p_i+p_j)$
a_ia_j	a_ia_k	$p_ip_jp_k$

Therefore, given that the pair is genetically compatible, the conditional probabilities of false (F) or true (T) motherhood are in the ratios (F/T):

woman	child	F	T	F/T
a_ia_i	a_ia_i	p_i^4	p_i^3	p_i
a_ia_i	a_ia_j	$2p_i^3p_j$	$p_i^2p_j$	$2p_i$
a_ia_j	a_ia_i	$2p_i^3p_j$	$p_i^2p_j$	$2p_i$
a_ia_j	a_ia_j	$4p_i^2p_j^2$	$p_ip_j(p_i+p_j)$	$4p_ip_j/(p_i+p_j)$
a_ia_j	a_ia_k	$4p_i^2p_jp_k$	$p_ip_jp_k$	$4p_i$

The conditional probabilities P(T) and P(F) favouring the hypotheses of true and false motherhood are straightforwardly taken, in each situation, from $P(T) = 1/(1 + F/T) = T/(F + T)$ and $P(F) = 1 - P(T) = 1/(1 + T/F) = F/(F + T)$:

$$P(T|m = a_ia_i,\ c = a_ia_i) = 1/(1 + p_i)$$

$$P(T|m = a_ia_i,\ c = a_ia_j) = 1/(1 + 2p_i)$$

$$P(T|m = a_ia_j,\ c = a_ia_i) = 1/(1 + 2p_i)$$

$$P(T|m = a_ia_j,\ c = a_ia_j) = (p_i + p_j)/(p_i + p_j + 4p_ip_j)$$

$P(T|m = a_i a_j,\ c = a_i a_k) = 1/(1 + 4p_i)$

and

$P(F|m = a_i a_i,\ c = a_i a_i) = p_i/(1 + p_i)$
$P(F|m = a_i a_i,\ c = a_i a_j) = 2p_i/(1 + 2p_i)$
$P(F|m = a_i a_j,\ c = a_i a_i) = 2p_i/(1 + 2p_i)$
$P(F|m = a_i a_j,\ c = a_i a_j) = 4p_i p_j/(p_i + p_j + 4p_i p_j)$
$P(F|m = a_i a_j,\ c = a_i a_k) = 4p_i/(1 + 4p_i)$.

4) Paternity

4.a) Probability of Paternity Exclusion [$P(E_4)$]

Using the same notation as in previous sections, it comes out that, under the hypothesis that the father is false and there is true biological relationship between the woman and the child, $P(E_4)$ can be taken straightforwardly from the probabilities associated with the events shown in the table below. A falsely accused individual is excluded when his genotype is incompatible with the genotype of the child given the genotype of the woman (who is assumed to be the true mother); for instance, in the case $m(a_i a_i)$-$c(a_i a_i)$, the falsely accused man is excluded if he does not have the allele a_i. Since the frequency of these individuals in the population is $(1-p_i)^2$, the probability of the event takes form $p_i^3(1-p_i)^2$, where p_i^3 is the probability of the pair $m(a_i a_i)$-$c(a_i a_i)$. In the table below, the subscripts i, j and k indicate different alleles and frequencies ($j \neq i$ and $k \neq i,j$).

mother	child	accused man		P
$a_i a_i$	$a_i a_i$	$a_l a_m$	$l,m \neq i$	$p_i^3(1-p_i)^2$
$a_i a_i$	$a_i a_j$	$a_l a_m$	$l,m \neq j$	$p_i^2 p_j(1-p_j)^2$
$a_i a_j$	$a_i a_i$	$a_l a_m$	$l,m \neq i$	$p_i^2 p_j(1-p_i)^2$
$a_i a_j$	$a_j a_j$	$a_l a_m$	$l,m \neq j$	$p_i p_j^2(1-p_j)^2$
$a_i a_j$	$a_i a_j$	$a_l a_m$	$l,m \neq i,j$	$p_i p_j(p_i+p_j)(1-p_i-p_j)^2$
$a_i a_j$	$a_i a_k$	$a_l a_m$	$l,m \neq k$	$p_i p_j p_k(1-p_k)^2$
$a_i a_j$	$a_j a_k$	$a_l a_m$	$l,m \neq k$	$p_i p_j p_k(1-p_k)^2$

Summing all the expressions in each line over all possible alleles and adding all these results we obtain the expression for $P(E_4)$:

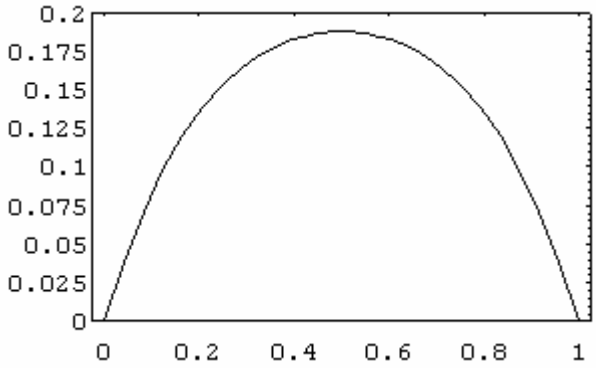
$$P(E_4) = \Sigma p_i^3(1-p_i)^2 + \Sigma\Sigma p_i^2 p_j(1-p_j)^2$$
$$+ 1/2.\Sigma\Sigma[p_i^2 p_j(1-p_i)^2 + p_i p_j^2(1-p_j)^2]$$
$$+ 1/2.\Sigma\Sigma p_i p_j(p_i+p_j)(1-p_i-p_j)^2$$
$$+ 1/2.\Sigma\Sigma\Sigma[p_i p_j p_k(1-p_k)^2 + p_i p_j p_k(1-p_k)^2]$$
$$= 1 - 2\Sigma p_i^2 + \Sigma p_i^3 - 2(\Sigma p_i^2)^2 + 2\Sigma p_i^4$$
$$+ 3(\Sigma p_i^2)(\Sigma p_i^3) - 3\Sigma p_i^5 .$$

For the case of two codominant alleles, $P(E_4)$ takes value

$P(E_4) = pq(1-pq) = \theta(1-\theta)$, $\theta = pq$.

As the graph below shows, the maximal probability of paternity exclusion [$P_{max}(E_4) = 0.1875$] takes place when $p = q = 0.5$; in fact, $q = 0.5$ is the only real root of equation $dP(E_4)/d\theta.d\theta/dq = (1-2\theta).(1-2q) = (1-2q+2q^2)(1-2q) = 0$.

```
Plot[q*(1-q)*(1-q*(1-q)),{q,0,1},
     PlotRange->{0,0.20},
     AxesOrigin->{0,0},
     Frame->True]
```

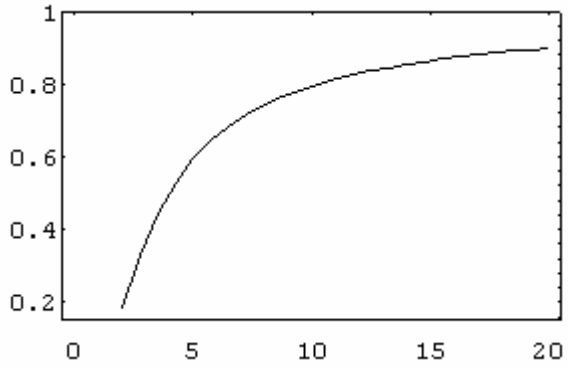


In the general case of n codominant alleles segregating at an autosomal locus, this maximum occurs when $p_1 = ... = p_n = 1/n$; then, $P_{max}(E_4)$ takes the form

$$P_{max}(E_4) = n[1/n.(1-1/n)^2] - 1/2.n(n-1).(1/n)^4.(4-6/n)$$
$$= 1 - (2n^3+n^2-5n+3)/n^4 .$$

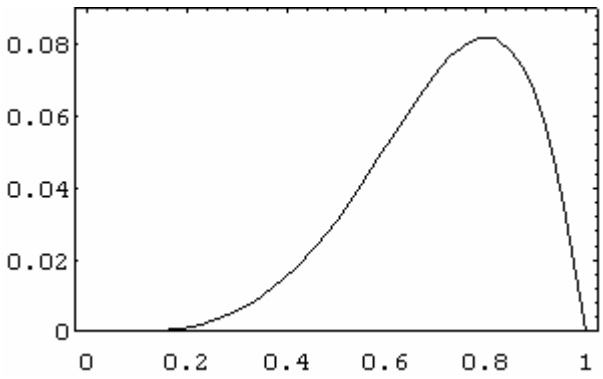
The figure that follows, generated by the enclosed Mathematica code, shows the values of $P_{max}(E_4) = 1 - (2n^3+n^2-5n+3)/n^4$ as function of n.

```
Plot[1-(2*k^3+k^2-5*k+3)/k^4,{k,2,20},
     PlotRange->{0.15,1},
     AxesOrigin->{0,0},
     Frame->True]
```



In the special case of two autosomal alleles with dominance, paternity is excluded only when the mother is recessive, the child is dominant and the accused individual is recessive. The probability of occurrence of this event is $P(E_4) = pq^2.q^2 = pq^4 = q^4-q^5$, where pq^2 is the frequency of rec-dom mother-child pairs and q^2 is the frequency of recessive individuals. As the following graph shows, the maximum value $P(E_4)$ takes is 0.0819 when q = 0.8, because this is the only pertinent root $(0 < q < 1)$ of the equation $d(q^4-q^5)/dq = q^3(4-5q) = 0$.

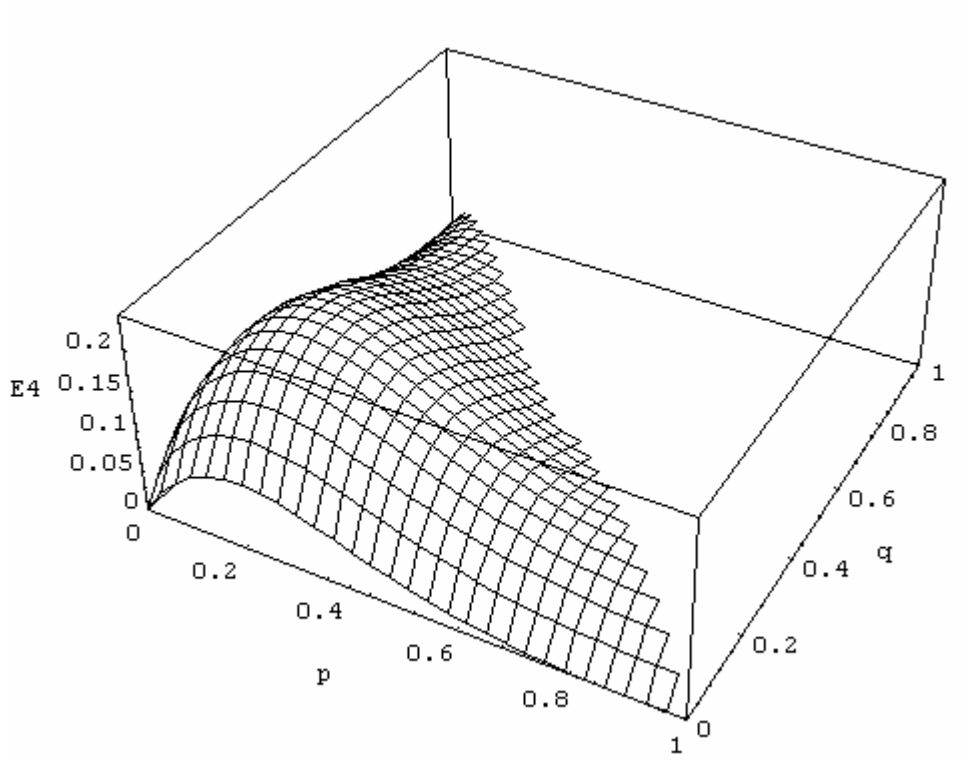
```
Plot[q^4*(1-q),{q,0,1},
     PlotRange->{0,0.09},
     AxesOrigin->{0,0},
     Frame->True]
```



For the special case of the ABO blood-group system,

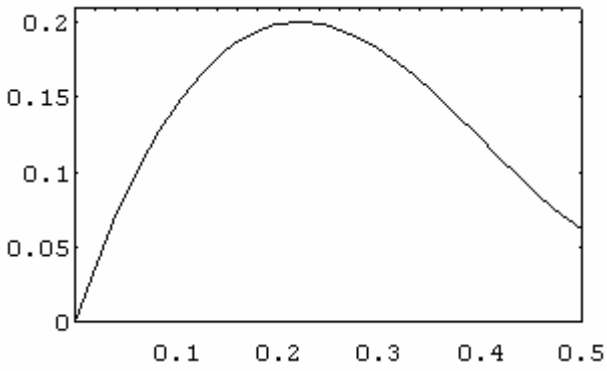
$P(E_4) = P(A+O).\{P[(A).(B)]+P[(A).(AB)]+P[(O).(B)]\}$
$+ P(B+O).\{P[(B).(A)]+P[(B).(AB)]+P[(O).(A)]\}$
$+ P(AB).\{P[(A).(O)]+P[(B).(O)]+P[(O).(O)]\}$
$+ P(O).P[(AB).(AB)]$
$= pqr^2(2+p+q)+p(q+r)^4+q(p+r)^4$, where **p**, **q**, and **r** are the frequencies of alleles **A**, **B**, and **O**. The graph below, originated by the enclosed Mathematica code, shows the values of $P(E_4)$ as function of ($0 < p < 1$) and ($0 < q < 1$), with the restriction $p + q \leq 1$.

```
f[p_,q_]:= p * q * (1 - p - q)^2 * (2 + p + q ) +
           p * (1 - p)^4 + q * (1 - q)^4 /; p+q <=1
   Plot3D[f[p,q], {p, 0, 1}, {q, 0, 1},
        PlotRange -> {0, 0.24}, AxesLabel -> {"p  ", "  q", "E4  "},
        Shading -> False, PlotPoints -> 30]
```



As before, $P(E_4)$ has its maximum when p = q and r = 1-2q; making these substitutions in the expression for $P(E_4)$ we obtain $P(E_4|p=q,r=1-2q) = 2q^2(1-2q)^2(1+q)+2q(1-q)^4$. Putting $dP(E_4|p=q,r=1-2q)/dq = 0$ we get $p = q = 0.2212$ and $r = 1-2q = 0.5576$; for these values, $P(E_4) = 0.1999$, which is the maximum value it can take [$P_{max}(E_4)$], as shown by the graph of the function $P(E_4|p=q,r=1-2q) = 2q^2(1-2q)^2(1+q)+2q(1-q)^4$.

```
Plot[2*q^2*(1-2*q)^2*(1+q)+2*q*(1-q)^4, {q,0,0.5},
     PlotRange->{{0,0.5},{0,0.21}},
     AxesOrigin->{0,0}, Frame->True]
```



4.b) Probability of True Paternity for Individuals not Excluded

If the individual is not excluded (for example, when the genotypes of the accused individual, woman and her child are respectively AB, AA, and AA), the conditional probabilities of true and false paternity are in the ratios **P(AB.AA.AA):P(AB).P(AA.AA) = T/F**, where **P(AB.AA.AA)** is the probability of occurrence of a couple (m x f) **AB.AA** with a child **AA**, **P(AB)** is the population frequency of **AB** individuals and **P(AA.AA)** is the population frequency of **AA.AA** mother-child pairs. If **p** and **q** are the frequencies of the alleles **A** and **B**, $P(AB.AA.AA) = 2pq.p^2.1/2 = p^3q$, $P(AB) = 2pq$, $P(AA.AA) = p^3$, $P(AB).P(AA.AA) = 2p^4q$, and $T/F = p^3q/2p^4q = 1/2p$. The following table summarizes all the possible results in the generalized case of n autosomal alleles, where p_i is the frequency of the i-th allele (a_i) segregating at a locus. As before, the subscripts i, j , k, and l indicate different alleles and frequencies ($j \neq i$, $k \neq i,j$, $l \neq i,j,k$).

ac.ind.	moth.	child	T	F	T/F
a_ia_i	a_ia_i	a_ia_i	p_i^4	p_i^5	$1/p_i$
a_ia_j	a_ia_i	a_ia_i	$p_i^3p_j$	$2p_i^4p_j$	$1/2p_i$
a_ia_i	a_ia_j	a_ia_i	$p_i^3p_j$	$p_i^4p_j$	$1/p_i$
a_ia_j	a_ia_j	a_ia_i	$p_i^2p_j^2$	$2p_i^3p_j^2$	$1/2p_i$
a_ja_j	a_ia_i	a_ia_j	$p_i^2p_j^2$	$p_i^2p_j^3$	$1/p_j$
a_ja_k	a_ia_i	a_ia_j	$p_i^2p_jp_k$	$2p_i^2p_j^2p_k$	$1/2p_j$
a_ia_i	a_ia_j	a_ia_j	$p_i^3p_j$	$p_i^3p_j(p_i+p_j)$	$1/(p_i+p_j)$
a_ia_j	a_ia_j	a_ia_j	$2p_i^2p_j^2$	$2p_i^2p_j^2(p_i+p_j)$	$1/(p_i+p_j)$
a_ja_k	a_ia_j	a_ia_j	$p_ip_j^2p_k$	$2p_ip_j^3p_k(p_i+p_j)$	$1/2(p_i+p_j)$
a_ja_j	a_ia_k	a_ia_j	$p_ip_j^2p_k$	$p_ip_j^3p_k$	$1/p_j$
a_ja_l	a_ia_k	a_ia_j	$p_ip_jp_kp_l$	$2p_ip_j^2p_kp_l$	$1/2p_j$

The same results are obtained if, instead of comparing the chances of occurrence of the trio under the two hypotheses, we compare directly the gamete contributions of both parents to the observed genotype of the child under the two alternative hypotheses [T: P(father gametic contribution) × P(mother gamete contribution); F: P(random male gamete contribution) × P(mother gamete contribution)]. If the individual is not excluded (for example, when the genotypes of the accused individual, woman and her child are, as before, respectively AB, AA, and AA), the conditional probabilities of true and false paternity are in the ratios **1/2 . 1 : p . 1 = 1/2p = T/F**. In fact, **1/2** is the probability of the accused individual (that is a heterozygote), being the true father, transmitting the **A** gene to the **AA** child and **1** is the homozygote **AA** mother's corresponding probability; under the alternative hypothesis in which the individual is not the father, the probability of the child being **AA** is **p . 1 = p**, where **p** is the probability of the child receiving an **A** gene from a male of the population and **1** the corresponding probability of receiving the **A** allele from his mother, who happens to be **AA**. The table below summarizes all possible results.

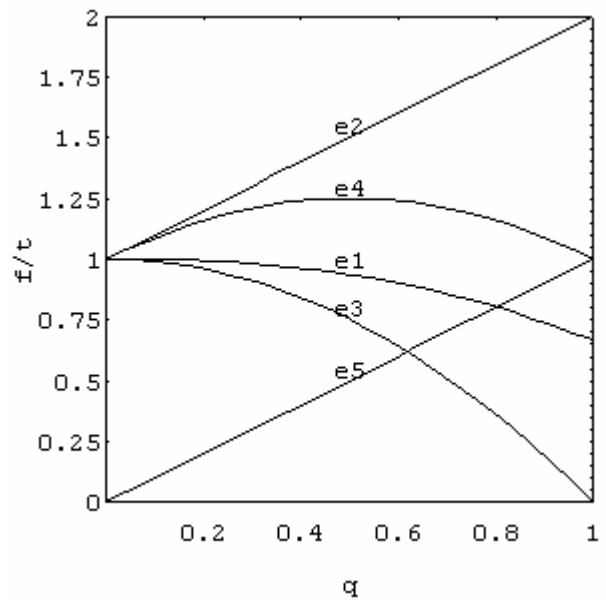
ac.ind.	moth.	child	T	F	T/F
$a_i a_i$	$a_i a_i$	$a_i a_i$	1.1 = 1	$1.p_i = p_i$	$1/p_i$
$a_i a_j$	$a_i a_i$	$a_i a_i$	1.1/2 = 1/2	$1.p_i = p_i$	$1/2p_i$
$a_i a_i$	$a_i a_j$	$a_i a_i$	1/2.1 = 1/2	$1/2.p_i = p_i/2$	$1/p_i$
$a_i a_j$	$a_i a_j$	$a_i a_i$	1/2.1/2 = 1/4	$1/2.p_i = p_i/2$	$1/2p_i$
$a_j a_j$	$a_i a_i$	$a_i a_j$	1.1 = 1	$1.p_j = p_j$	$1/p_j$
$a_j a_k$	$a_i a_i$	$a_i a_j$	1.1/2 = 1/2	$1.p_j = p_j$	$1/2p_j$
$a_i a_i$	$a_i a_j$	$a_i a_j$	1/2.1 = 1/2	$1/2.(p_i+p_j)$	$1/(p_i+p_j)$
$a_i a_j$	$a_i a_j$	$a_i a_j$	1/2.(1/2+1/2)	$1/2.(p_i+p_j)$	$1/(p_i+p_j)$
$a_j a_k$	$a_i a_j$	$a_i a_j$	1/2.1/2 = 1/4	$1/2.(p_i+p_j)$	$1/2(p_i+p_j)$
$a_j a_j$	$a_i a_k$	$a_i a_j$	1/2.1 = 1/2	$1/2.p_j = p_j/2$	$1/p_j$
$a_j a_l$	$a_i a_k$	$a_i a_j$	1/2.1/2 = 1/4	$1/2.p_j = p_j/2$	$1/2p_j$

The tables and graphs (generated by the enclosed Mathematica codes) that follow show the values of **F/T** for the special cases of two autosomal

alleles with dominance (**D-, dd**) and without dominance (**MM, MN, NN**) and of the **ABO** blood group system.

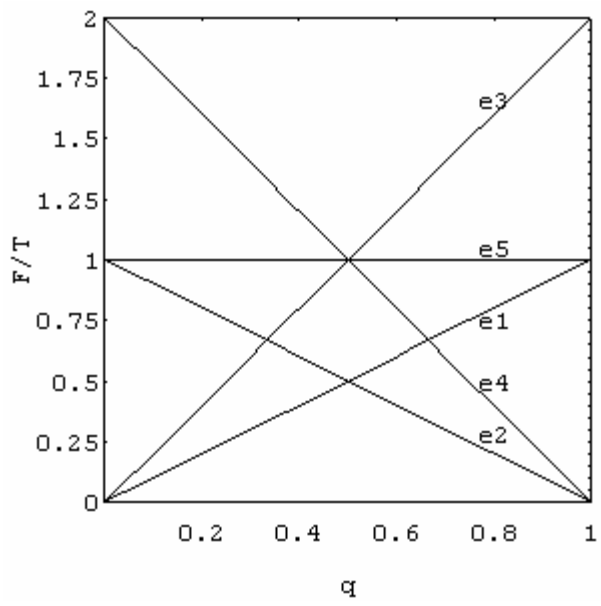
ac.ind.	moth.	child	F/T	
D-	D-	D-	$(1+q)(1+q-q^2)/(1+2q)$	e1
D-	D-	dd	$1+q$	e2
D-	dd	D-	$1-q^2$	e3
D-	dd	dd	$1+q$	e2
dd	D-	D-	$1+q-q^2$	e4
dd	D-	dd	q	e5
dd	dd	dd	q	e5

```
e5 = q; e3 = 1-e5^2; e4 = e3+e5; e2 = 1+e5; e1 = e2*e4/(e2+e5);
Show[Plot[{e1,e2,e3,e4,e5},{q,0,1}, PlotRange -> {{0,1},{0,2}},
          Frame -> True, FrameLabel -> {"q","F/T"},
          AspectRatio -> 1, DisplayFunction -> Identity],
     Graphics[{Text["e2",{0.5,1.55}], Text["e4",{0.5,1.3}],
               Text["e1",{0.5,1.0}], Text["e3",{0.5,0.80}],
               Text["e5",{0.5,0.55}]}],
     DisplayFunction -> $DisplayFunction]
```



ac.ind.	moth.	child	F/T	
MM	MM	MM	1-q	e2
MM	MN	MM	1-q	e2
MM	MN	MN	1	e5
MM	NN	MN	1-q	e2
MN	MM	MM	2(1-q)	e4
MN	MM	MN	2q	e3
MN	MN	MM	2(1-q)	e4
MN	MN	MN	1	e5
MN	MN	NN	2q	e3
MN	NN	MN	2(1-q)	e4
MN	NN	NN	2q	e3
NN	MM	MN	q	e1
NN	MN	MN	1	e5
NN	MN	NN	q	e1
NN	NN	NN	q	e1

```
e1 = q; e2 = 1-e1; e3 = 2*e1; e4 = 2*e2; e5 = 1;
Show[Plot[{e1,e2,e3,e4,e5},{q,0,1}, PlotRange -> {{0,1},{0,2}},
          Frame -> True, FrameLabel -> {"q","F/T"},
          AspectRatio -> 1,DisplayFunction -> Identity],
     Graphics[{Text["e2",{0.8,0.28}], Text["e4",{0.8,0.50}],
               Text["e1",{0.8,0.75}], Text["e3",{0.8,1.66}],
               Text["e5",{0.8,1.05}]}],
     DisplayFunction -> $DisplayFunction]
```



ac.ind.	moth.	child	F/T	
A	A	A	$(p+2r)(p^2+3pr+r^2)/(p+r)(p+3r)$	e1
A	A	O	$p+2r$	e2
A	B	A	$p(p+2r)/(p+r)$	e3
A	B	B	$(p+2r)(q^2+3qr+r^2)/r(q+r)$	e4
A	B	AB	$p(p+2r)/(p+r)$	e3
A	B	O	$p+2r$	e2
A	AB	A	$p+r$	e5
A	AB	B	$(p+2r)(q+r)/r$	e6
A	AB	AB	$(p+2r)(p+q)/(p+r)$	e7
A	O	A	$p(p+2r)/(p+r)$	e3
A	O	O	$p+2r$	e2
B	A	A	$(q+2r)(p^2+3pr+r^2)/r(p+r)$	e8
B	A	B	$q(q+2r)/(q+r)$	e9
B	A	AB	$q(q+2r)/(q+r)$	e9
B	A	O	$q+2r$	e10
B	B	B	$(q+2r)(q^2+3qr+r^2)/(q+r)(q+3r)$	e11
B	B	O	$q+2r$	e10
B	AB	A	$(q+2r)(p+r)/r$	e12
B	AB	B	$q+r$	e13
B	AB	AB	$(q+2r)(p+q)/(q+r)$	e14
B	O	B	$q(q+2r)/(q+r)$	e9
B	O	O	$q+2r$	e10
AB	A	A	$2(p^2+3pr+r^2)/(p+2r)$	e15
AB	A	B	$2q$	e16
AB	A	AB	$2q$	e16
AB	B	A	$2p$	e17
AB	B	B	$2(q^2+3qr+r^2)/(q+2r)$	e18
AB	B	AB	$2p$	e17
AB	AB	A	$2(p+r)$	e19
AB	AB	B	$2(q+r)$	e20
AB	AB	AB	$p+q$	e21
AB	O	A	$2p$	e17
AB	O	B	$2q$	e16
O	A	A	$2(q^2+3qr+r^2)/(q+2r)$	e22
O	A	O	r	e23
O	B	B	$(q^2+3qr+r^2)/(q+r)$	e24
O	B	O	r	e23
O	AB	A	$p+r$	e5
O	AB	B	$q+r$	e13
O	O	O	r	e23

5 - Joint Parentage

5.a) Probability of Joint Parentage Exclusion [$P(E_5)$]

The cells of the following matrix give the probabilities of a random child from the population being compatible with a non related couple $\mathbf{a_i a_j} \times \mathbf{a_k a_l}$ which genotypes are shown in the margins of the table. The total compatibility probability is obtained multiplying the each cell probability by its corresponding marginal probabilities and adding all

these products: $\mathbf{1 - P(E_5) = p_1^2.p_1^2.p_1^2 + p_1^2.2p_1p_2.(p_1^2+2p_1p_2) + ... + p_n^2.p_n^2.p_n^2}$. Putting, as before i, j, k, and l as being the subscripts used for identifying the n alleles segregating at an autosomal locus, and letting $\mathbf{j \neq i}$, $\mathbf{k \neq i,j}$ and $\mathbf{l \neq i,j,k}$, we obtain the following compatibility probabilities, arranged according the possible marginal combinations of unrelated individuals:

	a_1a_1 p_1^2	**a_1a_2** $2p_1p_2$	**a_1a_3** $2p_1p_3$	**a_1a_4** $2p_1p_4$	**a_2a_2** p_2^2
a_1a_1 p_1^2	p_1^2	$p_1^2+2p_1p_2$	$p_1^2+2p_1p_3$	$p_1^2+2p_1p_4$	$2p_1p_2$
a_1a_2 $2p_1p_2$	$p_1^2+2p_1p_2$	$(p_1+p_2)^2$	$p_1^2+2p_2p_3+$ $2p_1(p_2+p_3)$	$p_1^2+2p_2p_4+$ $2p_1(p_2+p_4)$	$p_2^2+2p_1p_2$
a_1a_3 $2p_1p_3$	$p_1^2+2p_1p_3$	$p_1^2+2p_2p_3+$ $2p_1(p_2+p_3)$	$(p_1+p_3)^2$	$p_1^2+2p_3p_4+$ $2p_1(p_3+p_4)$	$2p_2(p_1+p_3)$
a_1a_4 $2p_1p_4$	$p_1^2+2p_1p_4$	$p_1^2+2p_2p_4+$ $2p_1(p_2+p_4)$	$p_1^2+2p_3p_4+$ $2p_1(p_3+p_4)$	$(p_1+p_4)^2$	$2p_2(p_1+p_4)$
a_2a_2 p_2^2	$2p_1p_2$	$p_2^2+2p_1p_2$	$2p_2(p_1+p_3)$	$2p_2(p_1+p_4)$	$2p_1p_2$
a_2a_3 $2p_2p_3$	$2p_1(p_2+p_3)$	$p_2^2+2p_1p_3+$ $2p_2(p_1+p_3)$	$p_3^2+2p_1p_2+$ $2p_3(p_1+p_2)$	$2(p_1+p_4).$ (p_2+p_3)	$p_2^2+2p_2p_3$
a_2a_4 $2p_2p_4$	$2p_1(p_2+p_4)$	$p_2^2+2p_1p_4+$ $2p_2(p_1+p_4)$	$2(p_1+p_3).$ (p_2+p_4)	$p_4^2+2p_1p_2+$ $2p_4(p_1+p_2)$	$p_2^2+2p_2p_4$
a_3a_3 p_3^2	$2p_1p_3$	$2p_3(p_1+p_2)$	$p_3^2+2p_1p_3$	$2p_3(p_1+p_4)$	$2p_2p_3$
a_3a_4 $2p_3p_4$	$2p_1(p_3+p_4)$	$2(p_1+p_2).$ (p_3+p_4)	$p_3^2+2p_1p_4+$ $2p_3(p_1+p_4)$	$p_4^2+2p_1p_3+$ $2p_4(p_1+p_3)$	$2p_2(p_3+p_4)$
a_4a_4 p_4^2	$2p_1p_4$	$2p_4(p_1+p_2)$	$2p_4(p_1+p_3)$	$p_4^2+2p_1p_4$	$2p_2p_4$

	a_2a_3 $2p_2p_3$	**a_2a_4** $2p_2p_4$	**a_3a_3** p_3^2	**a_3a_4** $2p_3p_4$	**a_4a_4** p_4^2
a_1a_1 p_1^2	$2p_1(p_2+p_3)$	$2p_1(p_2+p_4)$	$2p_1p_3$	$2p_1(p_3+p_4)$	$2p_1p_4$
a_1a_2 $2p_1p_2$	$p_2^2+2p_1p_3+$ $2p_2(p_1+p_3)$	$p_2^2+2p_1p_4+$ $2p_2(p_1+p_4)$	$2p_3(p_1+p_2)$	$2(p_1+p_2).$ (p_3+p_4)	$2p_4(p_1+p_2)$
a_1a_3 $2p_1p_3$	$p_3^2+2p_1p_2+$ $2p_3(p_1+p_2)$	$2(p_1+p_3).$ (p_2+p_4)	$p_3^2+2p_1p_3$	$p_3^2+2p_1p_4+$ $2p_3(p_1+p_4)$	$2p_4(p_1+p_3)$
a_1a_4 $2p_1p_4$	$2(p_1+p_4).$ (p_2+p_3)	$p_4^2+2p_1p_2+$ $2p_4(p_1+p_2)$	$2p_3(p_1+p_4)$	$p_4^2+2p_1p_3+$ $2p_4(p_1+p_3)$	$p_4^2+2p_1p_4$
a_2a_2 p_2^2	$p_2^2+2p_2p_3$	$p_2^2+2p_2p_4$	$2p_2p_3$	$2p_2(p_3+p_4)$	$2p_2p_4$
a_2a_3 $2p_2p_3$	$(p_2+p_3)^2$	$p_2^2+2p_3p_4+$ $2p_2(p_3+p_4)$	$p_3^2+2p_2p_3$	$p_3^2+2p_2p_4+$ $2p_3(p_2+p_4)$	$2p_4(p_2+p_3)$
a_2a_4 $2p_2p_4$	$p_2^2+2p_3p_4+$ $2p_2(p_3+p_4)$	$(p_2+p_4)^2$	$2p_3(p_2+p_4)$	$p_4^2+2p_2p_3+$ $2p_4(p_2+p_3)$	$p_4^2+2p_2p_4$
a_3a_3 p_3^2	$p_3^2+2p_2p_3$	$2p_3(p_2+p_4)$	p_3^2	$p_3^2+2p_3p_4$	$2p_3p_4$
a_3a_4 $2p_3p_4$	$p_3^2+2p_2p_4+$ $2p_3(p_2+p_4)$	$p_4^2+2p_2p_3+$ $2p_4(p_2+p_3)$	$p_3^2+2p_3p_4$	$(p_3+p_4)^2$	$p_4^2+2p_3p_4$
a_4a_4 p_4^2	$2p_4(p_2+p_3)$	$p_4^2+2p_2p_4$	$2p_3p_4$	$p_4^2+2p_3p_4$	p_4^2

$a_ia_i \times a_ia_i$: $P_1 = \Sigma p_i^6$

$a_ia_j \times a_ia_j$: $P_2 = 2\Sigma\Sigma p_i^2 p_j^2 (p_i+p_j)^2$

$a_ia_i \times a_ia_j$: $P_3 = 4\Sigma\Sigma p_i^3 p_j (p_i^2+2p_ip_j)$

$a_ia_i \times a_ja_j$: $P_4 = 2\Sigma\Sigma p_i^3 p_j^3$

$a_ia_i \times a_ja_k$: $P_5 = 2\Sigma\Sigma\Sigma p_i^3 p_j p_k (p_j+p_k)$

$a_ia_j \times a_ia_k$: $P_6 = 4\Sigma\Sigma\Sigma p_i^2 p_j p_k [p_i^2+2p_jp_k+2p_i(p_j+p_k)]$

$a_ia_j \times a_ka_l$: $P_7 = 16\Sigma\Sigma\Sigma\Sigma p_ip_jp_kp_l [(p_i+p_j)(p_k+p_l) + (p_i+p_k)(p_j+p_l) + (p_i+p_l)(p_j+p_k)]$, $i<j<k<l$

Therefore,

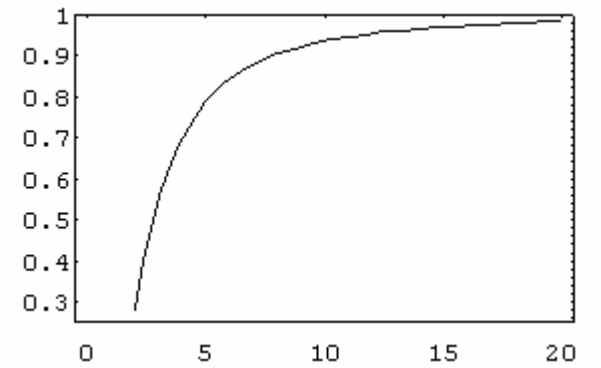
$$P(E_5) = 1 - (P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7)$$
$$= 1 + 5\Sigma p_i^6 - 4(\Sigma p_i^2)(\Sigma p_i^4) - 2(\Sigma p_i^3)^2 + 12(\Sigma p_i^2)(\Sigma p_i^3) - 8\Sigma p_i^5 + 4\Sigma p_i^4 - 8(\Sigma p_i^2)^2$$

As in previous cases, the maximal probability of exclusion takes place when all allelic frequencies are equal: $p_i = p_j = \ldots = 1/n$, where n is the number of alleles segregating at the autosomal locus. Under this assumption, the expression for $P(E_5)$ becomes

$$P_{max}(E_5) = 1 - (8n^3-16n^2+14n-5)/n^5.$$

The figure that follows, generated by the enclosed Mathematica code, shows the values of $P_{max}(E_5) = 1 - (8n^3-16n^2+14n-5)/n^5$ as function of n.

```
Plot[1-(8*k^3-16*k^2+14*k-5)/k^5,{k,2,20},
     PlotRange->{0.25,1},
     AxesOrigin->{0,0},
     Frame->True]
```

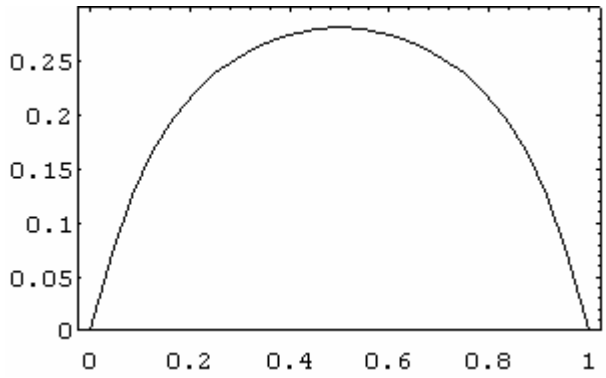


For the special case of n = 2 alleles,

$\mathbf{P(E_5) = pq(2-5pq+6p^2q^2) = \theta(2 - 5\theta + 6\theta^2)}$, $\theta = pq$.

The graph below shows the variation of $\mathbf{P(E_5)}$ as function of $\mathbf{0 < q < 1}$.

```
Plot[q*(1-q)*(2-5*q*(1-q)+6*q^2*(1-q)^2),{q,0,1},
     PlotRange->{0,0.30},
     AxesOrigin->{0,0},
     Frame->True]
```



The maximal probability of joint parentage exclusion [$\mathbf{P_{max}(E_5)} = 0.28125$] takes place when $p = q = 0.5$; in fact, $q = 0.5$ is the only real root of equation $\mathbf{dP(E_5)/d\theta.d\theta/dq = 2(1-5\theta+9\theta^2).(1-2q) = 2(1-5q+14q^2-18q^3+9q^4)(1-2q) = 0}$, as the following Mathematica code shows.

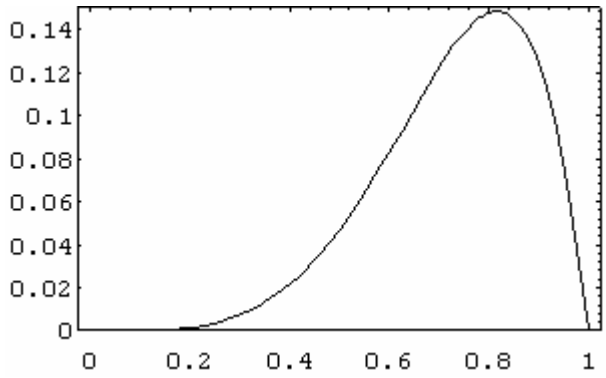
```
Solve[2*(1-5*q+14*q^2-18*q^3+9*q^4)*(1-2*q)==0]
{{q -> 1/2},
      18 - Sqrt[324 - 72 (5 - i Sqrt[11])]
{q -> ------------------------------------},
                     36
      18 + Sqrt[324 - 72 (5 - i Sqrt[11])]
{q -> ------------------------------------},
                     36
       18 - Sqrt[324 - 72 (5 + i Sqrt[11])]
{q -> ------------------------------------},
                     36
       18 + Sqrt[324 - 72 (5 + i Sqrt[11])]
{q -> ------------------------------------}}
                     36
```

In the general case of n codominant alleles segregating at an autosomal locus, this maximum occurs when $\mathbf{p_1 = ... = p_n = 1/n}$.

In the special case of two autosomal alleles with dominance, joint parentage is excluded only when both man and woman are recessive and the child is dominant. The probability of occurrence of this event is $\mathbf{P(E_5) = q^2.q^2.(1-q^2) = q^4-q^6}$. As the following graph shows, the maximum value

$P(E_5)$ takes is 0.1481 when q = 0.8165, because this is the only pertinent root **$(0 < q < 1)$** of the equation **$d(q^4-q^6)/dq = 2q^3(2-3q^2) = 0$.**

```
Plot[q^4*(1-q^2),{q,0,1},
     PlotRange->{0,0.15},
     AxesOrigin->{0,0},
     Frame->True]
```



For the special case of the ABO blood-group system,

$P(E_5) = P(B+AB).\{P[(A).(A)]+P[(A).(O)]\}$

$+ P(A+AB).\{P[(B).(B)]+P[(B).(O)]\}$

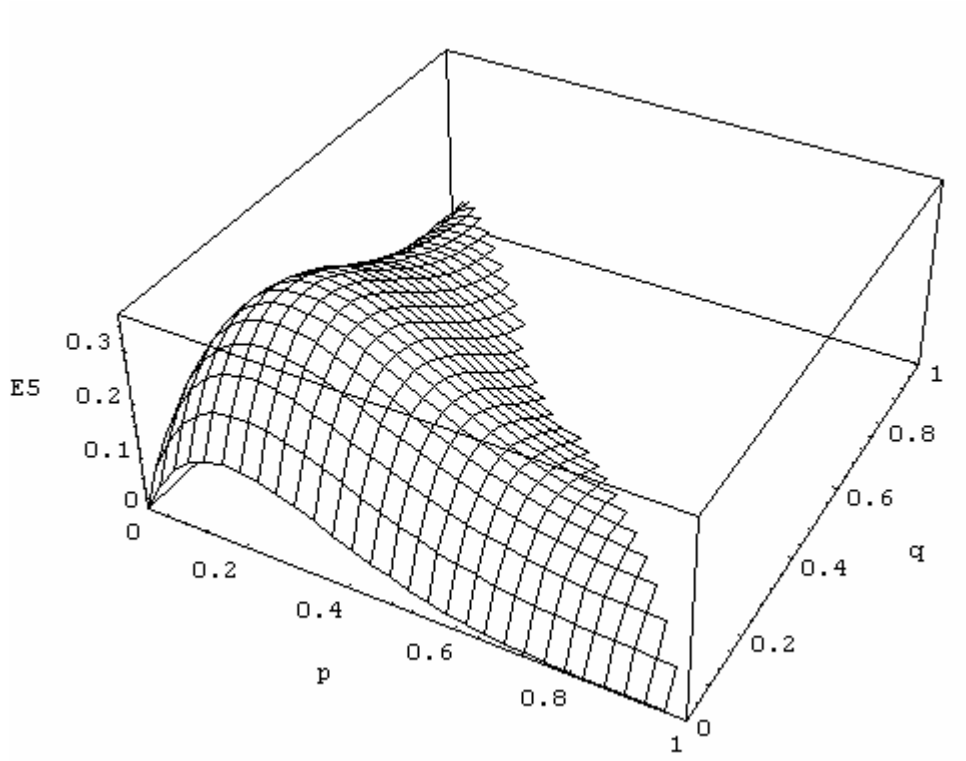
$+ P(O).\{P[(A).(AB)]+P[(B).(AB)]+P[(AB).(AB)]\}$

$+ P(AB+O).P[(AB).(O)]$

$+ P(A+B+AB).P[(O).(O)]$

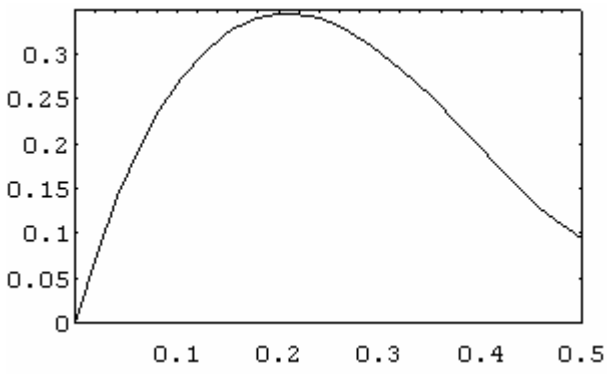
$= (1-q)^4 - (1-q)^6 + (1-p)^4 - (1-p)^6 + 4pqr^2(1+pq) - 2pqr^4$, where **p**, **q**, and **r** are the frequencies of alleles **A**, **B**, and **O**. The graph below, originated by the enclosed Mathematica code, shows the values of **$P(E_5)$** as function of (**$0 < p < 1$**) and (**$0 < q < 1$**), with the restriction **$p + q \leq 1$.**

```
f[p_,q_]:= (1-q)^4 - (1-q)^6 + (1-p)^4 - (1-p)^6 +
           4*p*q*(1-p-q)^2*(1+p*q)-2*p*q*(1-p-q)^4 /; p+q <=1
   Plot3D[f[p,q], {p, 0, 1}, {q, 0, 1},
        PlotRange -> {0, 0.36}, AxesLabel -> {"p  ", "  q", "E5  "},
        Shading -> False, PlotPoints -> 30]
```



As before, $\mathbf{P(E_5)}$ has its maximum when p = q and r = 1-2q; making these substitutions in the expression $\mathbf{P(E_5) = (1-q)^4-(1-q)^6+(1-p)^4-(1-p)^6+4pqr^2(1+pq)-2pqr^4}$ we obtain $\mathbf{P(E_5|p=q,r=1-2q) = 2(1-q)^4-2(1-q)^6+4q^2(1-2q)^2(1+q^2)-2q^2(1-2q)^4}$. Putting $\mathbf{dP(E_5|p=q,r=1-2q)/dq = 0}$ we get $\mathbf{p = q = 0.2081}$ and $\mathbf{r = 1-2q = 0.5838}$; for these values, $\mathbf{P(E_5) = 0.3448}$, which is the maximum value it can take $\mathbf{[P_{max}(E_5)]}$, as shown by the graph of the function $\mathbf{P(E_5|p=q,r=1-2q) = 2(1-q)^4-2(1-q)^6+4q^2(1-2q)^2(1+q^2)-2q^2(1-2q)^4}$.

```
Plot[2*(1-q)^4-2*(1-q)^6+4*q^2*(1-2*q)^2*(1+q^2)-2*q^2*(1-2*q)^4,
     {q,0,0.5},
     PlotRange->{{0,0.5},{0,0.35}},
     AxesOrigin->{0,0}, Frame->True]
```



5.b) Probability of True Joint Parentage for Couples not Excluded

If the couple is not excluded (for example, when the genotypes of the couple and the child are respectively **AB.AA**, and **A**A), the conditional probabilities of true and false joint parentage are in the ratios **P(AB.AA.AA):P(AB.AA).P(AA) = T/F**, where **P(AB.AA.AA)** is the probability of occurrence of a couple **AB.AA** with a child **AA**, **P(AB.AA)** is the frequency of matings **AB.AA** and **P(AA)** is the population frequency of children **AA**. If **p** and **q** are the frequencies of the alleles **A** and **B**, $\mathbf{P(AB.AA.AA) = 4pq.p^2.1/2 = 4p^3q.1/2 = 2p^3q}$, $\mathbf{P(AB.AA) = 4pq.p^2 = 4p^3q}$, $\mathbf{P(AA) = p^2}$, $\mathbf{P(AB.AA).P(AA) = 4p^3q.p^2 = 4p^5q}$, and $\mathbf{T/F = 2p^3q/4p^5q = 1/2p^2}$. Since the probability of mating ($\mathbf{4p^3q}$ in the example) is common to both expressions, $\mathbf{T:F :: 1/2 : p^2}$, so that we get immediately $\mathbf{T/F = 1/2p^2}$, that is, **T/F** is equal to the probability of the couple with the observed genotypes having such a child (**1/2** in the example) divided by the probability of a couple from the population having a child with the observed genotype ($\mathbf{p^2}$). The following table summarizes all the possible results in the generalized case of n autosomal alleles, where $\mathbf{p_i}$ is the frequency of the i-th allele ($\mathbf{a_i}$) segregating at a locus. As before, the subscripts i, j , k, and l indicate different alleles and frequencies ($\mathbf{j \neq i}$, $\mathbf{k \neq i,j}$, $\mathbf{l \neq i,j,k}$).

couple	child	T	F	T/F
$a_ia_i \times a_ia_i$	a_ia_i	1	p_i^2	$1/p_i^2$
$a_ia_i \times a_ia_j$	a_ia_i	1/2	p_i^2	$1/2p_i^2$
$a_ia_i \times a_ia_j$	a_ia_j	1/2	$2p_ip_j$	$1/4p_ip_j$
$a_ia_j \times a_ia_j$	a_ia_i	1/4	p_i^2	$1/4p_i^2$
$a_ia_j \times a_ia_j$	a_ia_j	1/2	$2p_ip_j$	$1/2p_ip_j$
$a_ia_j \times a_ia_j$	a_ja_j	1/4	p_j^2	$1/4p_j^2$
$a_ia_j \times a_ia_k$	a_ia_i	1/4	p_i^2	$1/4p_i^2$
$a_ia_j \times a_ia_k$	a_ia_j	1/4	$2p_ip_j$	$1/8p_ip_j$
$a_ia_j \times a_ia_k$	a_ia_k	1/4	$2p_ip_k$	$1/8p_ip_k$
$a_ia_j \times a_ia_k$	a_ja_k	1/4	$2p_jp_k$	$1/8p_jp_k$
$a_ia_j \times a_ka_l$	a_ia_k	1/4	$2p_ip_k$	$1/8p_ip_k$
$a_ia_j \times a_ka_l$	a_ia_l	1/4	$2p_ip_l$	$1/8p_ip_l$
$a_ia_j \times a_ka_l$	a_ja_k	1/4	$2p_jp_k$	$1/8p_jp_k$
$a_ia_j \times a_ka_l$	a_ja_l	1/4	$2p_jp_l$	$1/8p_jp_l$

A COLLECTION OF BASIC FORMULAE COMMONLY USED IN THE THEORY OF POPULATION GENETICS

1) Gene (allele) and genotype frequencies

Genotypes	AA	Aa	aa	total
Absolute frequencies (obs. nos.)	**D**	**H**	**R**	**N**
(Relative) frequencies	**d = D/N**	**h = H/N**	**r = R/N**	**1**

Allele (gene) frequency estimates

P(A) = p = No. of **A** alleles/Total number of genes
= **(2D + H)/2N = D/N + 1/2 . H/N = d + h/2**
P(a) = q = No. of **a** alleles/Total number of genes =
= **(H + 2R)/2N = 1/2 . H/N + R/N =**
= **h/2 + r** = **1 - p**

Standard deviation (s. error) of estimate: $\mathbf{se(p) = se(q) = \sqrt{(pq/2N)}}$

2) Hardy-Weinberg equilibrium (HWE)

Genotypes	AA	Aa	aa	total
Frequencies	$\mathbf{p^2}$	**2pq**	$\mathbf{q^2}$	**1**

Properties of **HWE**: 1) $\mathbf{h \leq 1/2}$
2) $\mathbf{[(2pq)^2 = 4.p^2.q^2] \Rightarrow h^2 = 4dr}$
3) $\mathbf{\{d, h, r\} \longrightarrow \{p^2, 2pq, q^2\}}$
↑
1 single generation of panmixia (discr. gener.)

Test: chi-square ($\mathbf{\chi^2 = \sum(o_i - e_i)^2/e_i = \sum o_i^2/e_i - N)}$, **d.f. = 1**
o_i : obs.values : $o_1 = \mathbf{D}$ e_i : exp. values.: $e_1 = \mathbf{Np^2}$
$o_2 = \mathbf{H}$ $e_2 = \mathbf{2Np(1-p)}$
$o_3 = \mathbf{R}$ $e_3 = \mathbf{N(1-p)^2}$

Critical values (at α **= 0.05**) of the chi-squared distribution : **3.84** (1 d.f.), **5.99** (2 d.f); in the 2-allele case (as above), the chi-squared test formula simplifies algebraically to

$$\chi^2 = (H^2-4DR)^2.N/[(2D+H)^2.(H+2R)^2]$$

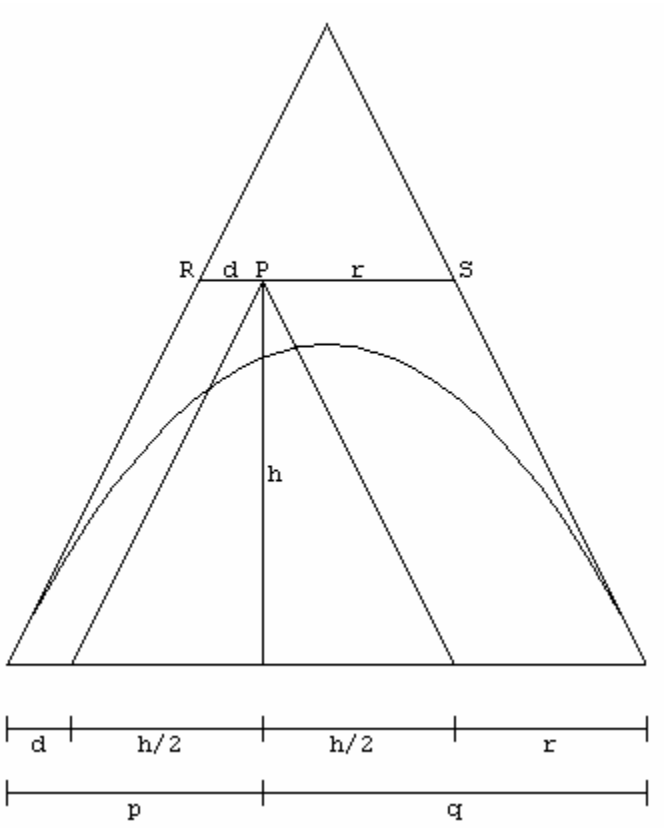
Generalization (case of any number of alleles):

A B C AA AB AC BB BC CC
$$(p + q + r + ...)^2 = p^2 + 2pq + 2pr + q^2 + 2qr + r^2 + ...$$

Special case of initial gene frequencies different among males (p_0') and females (p_0''):

$$p_1 = (p_0' + p_0'')/2 \ , \ h_1 = p_0'(1-p_0'') + p_0''(1-p_0') \neq 2pq$$

Graphical representation:



In the picture above, the height and the base of the isosceles triangle are unitary and the parabola depicted in its interior represents the set of all populations in Hardy-Weinberg proportions. It is easy to see that **d + h + r = 1**, where **d**, **h**, and **r** are the coordinates of the population point P. The projection of the population point on the base of the triangle divides it in two segments with values **p = d + h/2** and **q = 1 - p = h/2 + r**, with **p + q = 1**.

Dominance : **P(a) = q = $\sqrt{(R/N)} = \sqrt{r}$**
se(q) = $\sqrt{[(1-q^2)/4N]}$

X-chromosome (= sex-linked) alleles

|<-------- males ---------->|<--------------- females--------------->|

Ay	ay	Total	AA	Aa	aa	Total
S	**T**	**N_m**	**D**	**H**	**R**	**N_f**

Allele A frequency estimates:

a) among males : **$p_m = S/N_m = s$**
b) among females : **$p_f = (2D+H)/2N_f = d + h/2$**
c) in the total population : **$p = (S+2D+H)/(N_m+2N_f)$**

If **$N_m = N_f = N$** : **$p = (p_m + 2p_f)/3$**

Equilibrium dynamics for X-linked alles:

Let **m** and **f** be the frequencies , respectively among males and females, of a given allele (v.g., **a**); under **panmixia**, it comes out that:

(1) $m_{n+1} = f_n$
(2) $f_{n+1} = (m_n+f_n)/2;$
(3) $f_{n+1} - m_{n+1} = -(1/2).(f_n - m_n)$
(4) $f_n - m_n = -(1/2)^n.(f_0 - m_0);$

at equilibrium, that is, when the number of generations tends to infinity,

(5) $f = m = q$
(6) $q = (m_n + 2f_n)/3 = (m_{n+1}+2f_{n+1})/3 = (m_0+2f_0)/3$.

Case of two autosomal loci

Let a_ib_j be a given haplotype (haploid genotype corresponding to a pair of syntenic genes), with a recombination fraction r ($0.5 \geq r \geq 0$) between loci. The case r = 0 corresponds to complete linkage; when r = 0.5, independent segregation takes place.

$$P_{n+1}(a_ib_j) = P_n(a_ib_j) - r.P_n(a_iB_j) + r.P_n(a_i).P_n(b_j)$$
$$= (1-r).P_n(a_ib_j) + r.P(a_i).P(b_j)$$
$$P_{n+1}(a_ib_j) - P(a_i).P(b_j) = (1-r).[P_n(a_ib_j) - P(a_i).P(b_j)]$$
$$P_n(a_ib_j) - P(a_i).P(b_j) = (1-r)^n.[P_0(a_ib_j) - P(a_i).P(b_j)]$$

As **n** tends to infinity, $P_n(a_ib_j)$ clearly takes the value

$P(a_ib_j) = P(a_i).P(b_j)$, therefore independent from r. When

$P(a_ib_j) \neq P(a_i).P(b_j)$ and therefore $P(a_ib_j) - P(a_i).P(b_j) \neq 0$

we define $\Delta_{ij} = P(a_ib_j) - P(a_i).P(b_j)$ as being the *coefficient of linkage disequilibrium*, a misnomer since there exists no linkage equilibrium.

4) Deviations from panmixia

Self-fertilization : $h_{n+1} = h_n/2$, $h_n = h_0/2^n$

Wright's equilibrium

	AA	Aa	aa
(1)	$Fp + (1-F)p2$	$0 + 2pq(1-F)$	$Fq + (1-F)q2$
(2)	$p^2 + Fpq$	$2pq - 2Fpq$	$q^2 + Fpq$
(3)	$p - (1-F)pq$	$0 + 2pq(1-F)$	$q - (1-F)pq$,

where **F** is the **fixation index** or the average population inbreeding coefficient. Within nuclear families in genealogies, the inbreeding coefficient is the probability of autozygosis (homozygosis by common descent) for a given autosomal locus in the offspring of a consanguineous couple with a coefficient of relationship **r = 2F**.

Estimation of F : $h = 2pq(1-F) = 2pq - 2pqF$
$2pqF = 2pq - h$
$F = (2pq - h)/2pq = 1 - h/2pq$.

The parameter thus estimated can be used in the formula $\chi^2 = NF^2$ for testing the null hypothesis of HW equilibrium (equivalently, of testing $F = 0$, that is the case of Wright's equilibrium that corresponds to HW proportions).

5) Probability of extinction (PE) of a neutral mutation

$$P(E) = \lim_{n\to\infty} (1 - 1/2N)^{2N} = e^{-1} = 0.3679 ,$$

where $e = 2.71828...$

$$= \lim_{n\to\infty} (1 + 1/n)^n = \lim_{z\to 0} (1 + z)^{1/z} = \sum_{i=0}^{\infty} 1/i!$$

and

$$e^x = \lim_{n\to\infty} (1 + x/n)^n = \sum_{i=0}^{\infty} x^i/i!$$ (exponential function)

6) Recurrent mutations

μ = mutation rate (per locus or gamete and per generation)
$O(\mu) = 10^{-4}$ a 10^{-8} .

$p_{n+1} = p_n - \mu p_n = (1-\mu)p_n$

$p_n = (1-\mu)^n p_0 \rightarrow p_n = p_0 . e^{-n\mu}$

$p_n/p_0 = e^{-n\mu}$, $\ln(p_n/p_0) = -n\mu$;

$n = [\ln(p_0/p_n)]/\mu$ = number of generations required for the initial gene frequency p_0 to decrease to p_n.

Effect of the reverse mutation rate (ν) :

$p_{n+1} = p_n - \mu p_n + \nu q_n = p_n - \mu p_n + \nu(1-p_n)$;

at equilibrium, that is when n tends to infinity:

$p = p - \mu p + \nu(1-p)$, $p = \nu/(\mu+\nu)$, $q = \mu/(\mu+\nu)$.

7) Migration (gene flow)

Let m be the unidirectionl rate of gene flow (proportion of genes in a receptor population that are replaced, in one generation, by genes from a donor population; Q , the frequency of a given allele in the donor population; q_0 , the initial gene frequency in the receptor population; and q_n , the frequency of the same allele in the receptor population after n generations of constant gene flow. It comes out that

$q_1 = (1-m)q_0 + mQ \rightarrow q_1 - Q = (1-m)q_0 + mQ - Q = (1-m)(q_0-Q)$
$q_2 - Q = (1-m)(q_1-Q) = (1-m)^2(q_0-Q)$
...
$q_n - Q = (1-m)^n(q_0-Q)$, $(1-m)^n = (q_n-Q)/(q_0-Q)$.

8) Estimation of rates of racial admixture

Let $\mathbf{p = p_1x_1 + p_2x_2 + p_3x_3 + ...}$ be the gene frequency in the hybrid population, $\mathbf{p_i}$ the corresponding gene frequency in the i-th population stock that formed the hybrid population and $\mathbf{x_i}$ the contribution of this i-th racial stock. In the case of a hybrid population formed by only two racial stocks, $\mathbf{p = p_1x_1 + p_2x_2 = p_1x_1 + p_2(1-x_1) \Rightarrow x_1 = (p-p_2)/(p_1-p_2)}$. This last formula enables the calculation of the contribution of stock 1 to the hybrid population from known allelic frequencies in the hybrid population and in its two original racial stocks (**1** and **2**).

9) Genetic drift

Population size : **N**
Possible $\mathbf{q_1}$ values: **0/2N = 0, 1/2N, ..., (2N-1)/2N, 2N/2N = 1.**
Probability that q_1 takes the particular value $\mathbf{q_1 = j/2N}$:
$\mathbf{P(q_1 = j/2N) = C(2N,j).p^{2N-j}q^j = \{(2N)!/[(2N-j)!j!]\}.(1-q)^{2N-j}q^j}$.

In the case **N = 2** individuals (or **2N = 4** genes), the possible gene frequencies for any population are

j	**q**
0	**0**
1	**1/4**
2	**1/2**
3	**3/4**
4	**1**

[**j** : population state (= number of a genes present in the population; obviously, $\mathbf{q_j = j/2N}$]

and the conditional (*transition*) probabilities of a given state **i** at generation **t** becoming **j** in the next geration **t+1** [**P(i|t->j|t+1)**] are

		(t+1) j=0	j=1	j=2	j=3	j=4
	i=0	**1**	**0**	**0**	**0**	**0**
	i=1	**81/256**	**27/64**	**27/128**	**3/64**	**1/256**
(t)	i=2	**1/16**	**1/4**	**3/8**	**1/4**	**1/16**
	i=3	**1/256**	**3/64**	**27/128**	**27/64**	**81/256**
	i=4	**0**	**0**	**0**	**0**	**1**

$\mathbf{q_1 = q_0}$ [average gene frequency over all populations]
$\mathbf{V(q_0) = 0}$ [variance of gene frequencies among populat.]
$\mathbf{V(q_1) = p_0q_0/2N}$
$\mathbf{V(q_{inf}) = p_0q_0}$
$\mathbf{V(q_n) = q_0(1-q_0)[1-(1-1/2N)^n]}$

10) Selection

Genotypes	**Adaptive [=fitness] values**
AA	$\mathbf{W_1}$
Aa	$\mathbf{W_2}$
aa	$\mathbf{W_3}$

W = average population fitness = $\mathbf{p_n^2W_1 + 2p_nq_nW_2 + q_n^2W_3}$

$$q_{n+1} = q_n(p_nW_2+q_nW_3)/W$$
$$= [q_nW_2+q_n^2(W_3-W_2)]/[q_n^2(W_1-2W_2+W_3)-2q_n(W_1-W_2)+W_1]$$
$$\Delta q = q_{n+1}-q_n = q' - q = q(1-q)[q(W_1-2W_2+W_3)-(W_1-W_2)]/W$$
$$W' = dW/dq = 2q(W_1-2W_2+W_3)-2(W_1-W_2)$$
$$\Delta q = q(1-q)W'/(2W).$$

At equilibrium, $\Delta q = 0$ and the possible solutions of the equation $q(1-q)W' = 0$ are:

$q_1 = 0$, $q_2 = 1$ and $q_3 = (W_1-W_2)/(W_1-2W_2+W_3)$.

Any possible selection scheme leads therefore either to the extinction of the A gene (or of its allele a) or to a point q_3 interior to the interval (0, 1). Since the expression $q_3/p_3 = (W_1-W_2)/(W_3-W_2)$ can only take positive values, we conclude that the last equilibrium takes place only when either $W_1 \geq W_2 \leq W_3$ or $W_1 \leq W_2 \geq W_3$, that is, when the fitness value of heterozygotes is smaller or larger than those of both homozygotes. In the first case the equilibrium is unstable and in the second one stable. In this last case, if we make $W_2 = 1$ and replace W_1 and W_3 respectively with $1-s_1$ and $1-s_3$, we obtain $q = s_1/(s_1+s_3)$.

11) Hierarchical structure of populations

Case of any number of subpopulations, each in HWE:

$$P(AA) = \Sigma p_i^2/n = \Sigma x_i p_i^2 = p^2 + var(p)$$
$$P(aa) = \Sigma q_i^2/n = \Sigma x_i q_i^2 = q^2 + var(q)$$
$$P(Aa) = 2\Sigma p_i q_i/n = 2\Sigma x_i p_i q_i = 2pq - 2.var(p) = 2pq - 2.var(q).$$
$$var(p) = \Sigma x_i p_i^2 - p^2 = var(1-p) = var(q) = \Sigma x_i q_i^2 - q^2$$
$$FST = var(p)/pq \text{ (Wahlund's effect)} ,$$

where $P(AA)$, $P(Aa)$ and $P(aa)$ are the genotype frequencies in the whole population formed by n isolates with different sizes X_i and $x_i = X_i/\Sigma X_i$ is the contribution of the i-th isolate to the whole population.

Generic case with inbreeding within each subpopulation:

$$P_k(AA) = p_k^2 + F_k p_k q_k = F_k p_k + (1-F_k)p_k^2,$$
$$P_k(Aa) = 2p_k q_k(1-F_k), \text{ and}$$
$$P_k(aa) = q_k^2 + F_k p_k q_k = F_k q_k + (1-F_k)q_k^2;$$
$$P(AA) = \Sigma x_i[p_i^2+F_i p_i q_i],$$
$$P(Aa) = 2\Sigma x_i p_i q_i(1-F_i), \text{ and}$$
$$P(aa) = \Sigma x_i[q_i^2+F_i p_i q_i]$$

As before, $var(p) = var(q) = \Sigma x_i p_i^2 - p^2 = \Sigma x_i q_i^2 - q^2$ and the fixation index generated by population subdivision or Wahlund's effect (F_{ST}) is calculated, as in the case of panmictic subpopulations, after

$$F_{ST} = var(p)/pq = 1 - 2\Sigma x_i p_i q_i/2pq .$$

The fixation index in the total population due to both population subdivision and inbreeding occurring within subpopulations, that for the case when there is no inbreeding within populations takes value $F_{IT} = F_{ST}$, is obtained directly from

$$F_{IT} = 1 - \Sigma x_i P_i(Aa)/2pq = 1 - 2\Sigma x_i p_i q_i (1-F_i)/2pq \ .$$

The fixation index due to inbreeding within subpopulations is taken from

$$F_{IS} = (F_{IT}-F_{ST})/(1-F_{ST}) = 1 - 2\Sigma x_i p_i q_i (1-F_i)/2\Sigma x_i p_i q_i .$$

12) Effective population size (Ne)

Let **Nm** and **Nf** be the respective numbers of males and females in a population of diploid individuals. Whatever the population sex-ratio, each individual results from a fertilization that took place between a male and a female gamete. Therefore, the probability that any gene randomly drawn from a population at generation **t** was transmitted by a male belonging to generation **t-1** is **1/2**; the probability associated with two such genes is therefore **1/4** and the probability of randomly drawing two genes that originated in the *same male* of generation **t-1** is **1/4Nm**; clearly the probability of randomly drawing two genes originated in the *same female* of generation **t-1** is (by simmetry) **1/4Nf**; and the probability that these two genes originated in the same individual of generation **t-1** is evidently **1/4Nm + 1/4Nf**. Making this result equal to **1/Ne** we obtain

$1/Ne = 1/4Nm + 1/4Nf = (Nm + Nf)/4NmNf$
$Ne = 4NmNf/(Nm + Nf)$
If **$Nm = Nf = N$, $Ne = 4N^2/2N = 2N = Nm+Nf$**

The quantity **Ne** is known as the *effective (inbreeding) number* and corresponds to the size of a population with equal numbers of males and females that would generate the same amount of genetic drift produced in the actual population by the observed sex-ratio distortion.

DERIVATIVES (SUMMARY)

Let $\mathbf{y = f(x)}$ be a continuous function of $\mathbf{x}$ in a given interval; and $\mathbf{y_1 = f(x_1)}$ and $\mathbf{y_2 = f(x_2)}$ the values that the function $\mathbf{y}$ takes when the corresponding values of the argument $\mathbf{x}$ are $\mathbf{x_1}$ and $\mathbf{x_2}$. Let also:

$$\mathbf{\Delta x = x_2 - x_1} \ ,$$
$$\mathbf{\Delta y = y_2 - y_1 = f(x_2) - f(x_1) = f(x_1 + \Delta x) - f(x_1)} \ ,$$

where $\mathbf{\Delta x}$ and $\mathbf{\Delta y}$ are the increments of $\mathbf{x}$ and $\mathbf{y}$ at point $\mathbf{\{x_1, y_1\}}$.

The limit of the increment rate $\mathbf{\Delta y/\Delta x}$ as $\mathbf{\Delta x}$ tends to zero,

$$\lim_{\Delta x \to 0} \mathbf{\Delta y/\Delta x = \lim_{\Delta x \to 0} [f(x+\Delta x) - f(x)]/\Delta x} \ ,$$

is the *derivative with respect to* $\boldsymbol{x}$ *of* $\boldsymbol{y = f(x)}$

$$\mathbf{y' = f'(x) = \lim_{\Delta x \to 0} \Delta y/\Delta x = dy/dx} \ .$$

General differentiation rule	**Example**
$y = f(x)$	$y = 4x^2$
$y + \Delta y = f(x+\Delta x)$	$y + \Delta y = 4(x + \Delta x)^2 = 4x^2 + 8x.\Delta x + 4(\Delta x)^2$
$\Delta y = f(x+\Delta x) - y = f(x+\Delta x) - f(x)$	$\Delta y = 8x.\Delta x + 4(\Delta x)^2$
$\Delta y/\Delta x = [f(x+\Delta x) - f(x)]/\Delta x$	$\Delta y/\Delta x = 8x + 4\Delta x$
$dy/dx = \lim_{\Delta x \to 0} [f(x+\Delta x) - f(x)]/\Delta x$	$dy/dx = 8x + 4.0 = 8x$

FUNCTION y = f(x)	**DERIVATIVE y' = dy/dx = df(x)/dx**
$y = k$, k = constant	$y' = 0$
$y = x$	$y' = 1$
$y = k.x$, k = constant	$y' = k$
$y = f(u)$, $u = f(x)$	$y' = dy/du.du/dx$
$y = f(x)$, $x = f(y)$	$y' = 1/(dx/dy)$
$y = u \pm v$, $u = f_1(x)$, $v = f_2(x)$	$y' = u' \pm v'$
$y = uv$, $u = f_1(x)$, $v = f_2(x)$	$y' = u'.v + u.v'$
$y = x^n$	$y' = n.x^{n-1}$
$y = k.x^n$, k = constant	$y' = n.k.x^{n-1}$
$y = u^n$, $u = f(x)$	$y' = n.u^{n-1}.u'$
$y = u^m v^n$, $u = f_1(x)$, $v = f_2(x)$	$y' = m.u^{m-1}.u'.v^n + n.v^{n-1}.v'.u^m$
$y = u/v$, $u = f_1(x)$, $v = f_2(x)$	$y' = (u'.v - u.v')/v^2$
$y = u/v^n$, $u = f_1(x)$, $v = f_2(x)$	$y' = (u'.v - n.u.v')/v^{n+1}$
$y = u^{1/n}$, $u = f(x)$	$y' = u'/[n.(u^{n-1})^{1/n}]$
$y = \sqrt{u} = u^{1/2}$, $u = f(x)$	$y' = u'/[2u^{1/2}] = u'/[2\sqrt{u}]$
$y = \ln x = \log_e x$	$y' = 1/x$
$y = \ln u$, $u = f(x)$	$y' = u'/u$
$y = \log x$	$y' = \log e/x$
$y = e^x$, $e = \lim_{n \to \infty} (1+1/n)^n$	$y' = e^x$
$y = e^u$, $u = f(x)$	$y' = e^u.u'$
$y = a^u$, $u = f(x)$	$y' = a^u.u'/\log e$

$y = u^v$, $u = f_1(x)$, $v = f_2(x)$	$y' = u^v.v'.\ln u + u^{v-1}.u'.v$
$y = \sin x = (e^{ix}-e^{-ix})/2i$	$y' = \cos x$
$y = \cos x = (e^{ix}+e^{-ix})/2$	$y' = -\sin x$
$y = \tan x = \sin x/\cos x$	$y' = \sec^2 x$
$y = \cot x = 1/\tan x$	$y' = -\csc^2 x$
$y = \sec x = 1/\cos x$	$y' = \sec x \,.\, \tan x$
$y = \csc x = 1/\sin x$	$y' = -\csc x \,.\, \cot x$
$y = \sin u$, $u = f(x)$	$y' = \cos u \,.\, u'$
$y = \cos u$, $u = f(x)$	$y' = -\sin u \,.\, u'$
$y = \tan u$, $u = f(x)$	$y' = \sec^2 u \,.\, u'$
$y = \cot u$, $u = f(x)$	$y' = -\csc^2 u \,.\, u'$
$y = \sec u$, $u = f(x)$	$y' = \sec u.\ \tan u \,.\, u'$
$y = \csc u$, $u = f(x)$	$y' = -\csc u.\ \cot u.\ u'$
$y = \sinh x = (e^{x}-e^{-x})/2$	$y' = \cosh x$
$y = \cosh x = (e^{x}+e^{-x})/2$	$y' = \sinh x$
$y = \tanh x = \text{senh } x/\cosh x$	$y' = \text{sech}^2 x$
$y = \coth x = 1/\tanh x$	$y' = -\text{csch}^2 x$
$y = \text{sech } x = 1/\cosh x$	$y' = -\text{sech } x \,.\, \tanh x$
$y = \text{csch } x = 1/\text{senh } x$	$y' = -\text{csch } x \,.\, \coth x$
$y = \text{arc sin } x = \sin^{-1}x$	$y' = 1/\sqrt{(1-x^2)}$
$y = \text{arc cos } x = \cos^{-1}x$	$y' = -1/\sqrt{(1-x^2)}$
$y = \text{arc tan } x = \tan^{-1}x$	$y' = 1/(1+x^2)$
$y = \text{arc cot } x = \cot^{-1}x$	$y' = -1/(1+x^2)$
$y = \text{arc sin } u$, $u = f(x)$	$y' = u'/\sqrt{(1-u^2)}$
$y = \text{arc cos } u$, $u = f(x)$	$y' = -u'/\sqrt{(1-u^2)}$
$y = \text{arc tan } u$, $u = f(x)$	$y' = u'/(1+u^2)$
$y = \text{arc cot } u$, $u = f(x)$	$y' = -u'/(1+u^2)$

Application: determination of maximal and minimal values of a function in a given interval

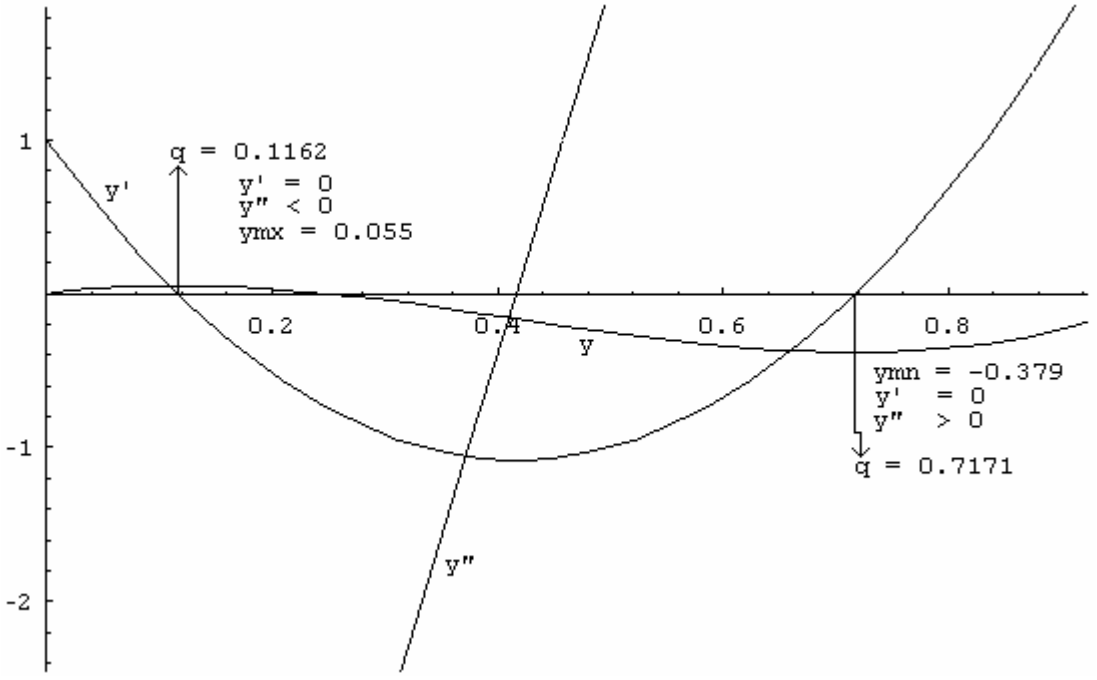
Let $\mathbf{y = 4q^3-5q^2+q}$ be a function defined in the interval (0,1). The first and second derivatives of y are

$\mathbf{y' = dy/dq = 12q^2-10q+1}$ and
$\mathbf{y" = d(dy/dq)/dq = d^2y/dq^2 = 24q-10}$.

The table below lists some values y, y' and y" take as function of q (varying from 0 to 1 in intervals of 0.05):

q	y	y'	y"
0.0500	+0.038	+0.53	-8.8
0.1000	+0.054	+0.12	-7.6
0.1162	+0.055	0.00	-7.2
0.1500	+0.051	-0.23	-6.4
0.2000	+0.032	-0.52	-5.2
0.2500	0.000	-0.75	-4.0
0.3000	-0.042	-0.92	-2.8
0.3500	-0.091	-1.03	-1.6
0.4000	-0.144	-1.08	-0.4
0.4500	-0.198	-1.07	+0.8
0.5000	-0.250	-1.00	+2.0
0.5500	-0.297	-0.87	+3.2
0.6000	-0.336	-0.68	+4.4
0.6500	-0.364	-0.43	+5.6
0.7000	-0.378	-0.12	+6.8
0.7171	-0.379	0.00	+7.2
0.7500	-0.375	+0.25	+8.0
0.8000	-0.352	+0.68	+9.2
0.8500	-0.306	+1.17	+10.4
0.9000	-0.234	+1.72	+11.6
0.9500	-0.133	+2.33	+12.8
1.0000	0.000	+3.00	+14.0

The graph below shows the same values of **y, y'** e **y"** as function of **q** in the interval (0, 1).



The graph shows clearly that the function $y = f(q) = 4q^3 - 5q^2 + q$ takes, in the interval (0, 1), a maximum (**ymx**) as well as a minimum (**ymn**). The maximum and minimum values of the function correspond respectively to the values $q_2 = 0.116$ and $q_1 = 0.717$; these two extremum values are the roots that make the first derivative of **y**, $y' = f'(q) = 12q^2 - 10q + 1$, equal to zero; in fact, if $y' = 12q^2 - 10q - 1 = 0$, it comes out that

$q_1 = (10 + \sqrt{52})/24 = 0.717$ and
$q_2 = (10 - \sqrt{52})/24 = 0.116$.

The value $y = f(q_2) = f(0.116) = 0.0549632$ is really the maximum value that **y** can take, as we show by simply verifying the values that the function takes in the immediate neighborhood of this point: in fact, if we make $\Delta q = 0.001$, it comes out that $y_0 = f(q_2) = f(0.116) = 0.0549632$, $y_1 = f(q_2+\Delta q) = f(0.117) = 0.0549614 < f(q_2)$ and $y_2 = f(q_2-\Delta q) = f(0.115) = 0.0549582 < f(q_2)$.

The value $y = f(q_1) = f(0.717) = -0.3790378$, on the other hand, is really the minimum value that **y** can take: in fact, if we make (again) $\Delta q = 0.001$, it comes out that $y_0 = f(q_1) = f(0.717) = -0.3790378$, $y_1 = f(q_1+\Delta q) = f(0.718) = -0.3790352 > f(q_1)$ and $y_2 = f(q_1-\Delta q) = f(0.716) = -0.3790336 > f(q_1)$.

If $y = f(q_2)$ is the maximum value the function can take inside the interval we are considering, its derivative **y'** evaluated at this point is zero; **q** values less than q_2 correspond to a derivative **y'** larger than zero, and values larger than q_2 correspond to negative values of the derivative **y'**: let $\Delta q = 0.001$; it follows that $y_0' = f'(q_2) = f'(0.116) = 0.0000$, $y_1' = f'(q_2+\Delta q) = f'(0.117) = -0.0057 < 0$ and $y_2' = f'(q_2-\Delta q) = f'(0.115) = +0.0087 > 0$; therefore, in the region around q_2, **y'** is a decreasing function of **q** and its derivative $y'' = f''(q_2)$ is smaller than zero.

If $y = f(q_1)$, on the contrary, is the minimum value the function can achieve in the interval under consideration, its derivative **y'** evaluated at this point is also zero; **q** values less than q_1 correspond to a derivative **y'** less than zero, and values larger than q_1 correspond to positive values of the derivative **y'**: let $\Delta q = 0.001$; it comes out that $y_0' = f'(q_1) = f'(0.717) = 0.0000$, $y_1' = f'(q_1+\Delta q) = f'(0.718) = +0.0063 > 0$ and $y_2' = f'(q_1-\Delta q) = f'(0.116) = -0.0081 < 0$; therefore, in the region around q_1, **y'** is an increasing function of **q** and its derivative $y'' = f''(q_1)$ is larger than zero.

The observations above suggest the following practical rule for investigating the nature of any extremum of a function:

1) determine the first derivative of **y**, $y' = f'(q)$;
2) solve the equation $y' = f'(q) = 0$; let q_1 be a root of this equation;
3) evaluate $y'' = f''(q_1)$;
4) if $f''(q_1) < 0$, then $y = f(q_1)$ is a maximum;
5) if $f''(q_1) > 0$, then $y = f(q_1)$ is a minimum.

HUMAN POPULATION GENETICS

(GENÉTICA DE POBLACIONES HUMANAS)

PAULO A. OTTO

Departamento de Genética e Biologia Evolutiva
Instituto de Biociências
Universidade de São Paulo
Caixa Postal 11461
05422-970 São Paulo SP

Curso Teórico Práctico de Post-Grado
8 al 14 de Septiembre de 2006
Departamento de Genética
Laboratorio de Citogenética y Genética Humana
Facultad de Ciencias Exactas Químicas y Naturales
Universidad Nacional de Misiones
Posadas, Misiones, República Argentina

II – Ejercicios en Clase 1-12

EJERCICIO EN CLASE 01

En una muestra poblacional cuidadosamente colectada y constituida por 556 individuos, se determinaron los grupos sanguíneos del sistema MN mediante el uso de sueros anti-M y anti-N. Los resultados figuran en la siguiente tabla:

reacción con suero		**número de individuos**
anti-M	**anti-N**	
+	**-**	**167**
+	**+**	**280**
-	**+**	**109**

a) ¿Qué se entiende, en genética, por "muestra cuidadosamente colectada"? b) ¿Por qué no se muestrea la población entera, ya que lo que se desea es minimizar las imprecisiones? c) ¿Cuál es la frecuencia p del gen M en ese muestreo? d) ¿Cuál es el error estándar de esa estimativa? e) ¿Qué se puede decir sobre la verdadera frecuencia (desconocida) del gen M en la población de la cual fue obtenida la muestra? f) ¿Es esa muestra representativa de una población panmíctica (en equilibrio de Hardy-Weinberg)? ¿Por qué? g) ¿Por qué el test de chi-cuadrado usado para responder a esa pregunta tiene un grado de libertad? h) En el estudio de la misma muestra de arriba, supongamos que los investigadores solamente dispondrán de suero anti-M. Construya una nueva tabla con los resultados encontrados. i) Calcule, a partir de esos datos, la frecuencia p del alelo M. ¿Qué es necesario admitir para que eso pueda ser realizado? ¿Por qué? ¿En que redunda eso? j) Calcule el valor del error estándar de p. k) ¿Por qué ese error estándar es mayor que el obtenido anteriormente? l) Compare el valor obtenido con el del cálculo anterior. ¿Cuál es el cálculo mas preciso? ¿Por qué en el caso del ejemplo ellos son, en tanto, parecidos?

En una muestra de 150 hombres y 300 mujeres negroides, Tönz y Rossi (1964) verificaron la siguiente distribución de genotipos sobre la deficiencia de G6PD (característica condicionada por un par de alelos codominantes ligados al cromosoma X):

genotipo	número de indivíduos
A	137
a	13
AA	247
Aa	50
aa	3

a) Calcule la frecuencia del alelo en la población masculina, en la femenina y en la muestra total. b) Verifique si la distribución observada de genotipos está de acuerdo con la ley de Hardy-Weinberg empleando un test de chi-cuadrado.

En una investigación sobre la distribución de dos grupos sanguíneos Xg en una muestra de 2082 individuos caucasoides, fueron observados los siguientes resultados:

sexo	gr. sang.	genotipo	n de indiv.
masculino	Xg(a+)	Xg^a	667
	Xg(a-)	Xg	346
femenino	Xg(a+)	Xg^aXg^a, Xg^aXg	967
	Xg(a-)	Xg Xg	102

a) ¿Cuál es la frecuencia del alelo Xg en la muestra masculina?
b) ¿Cuál es el error estándar de esa estimación?
c) ¿Cuál es la frecuencia del alelo Xg en la muestra femenina?
d) ¿Cuál es el error estándar de esa estimación?
e) ¿De qué manera las dos estimativas de arriba pueden ser combinadas con la finalidad de obtener una única estimación de la frecuencia génica en la muestra total?

N = n(MM) + n(MN) + n(NN) = D + H + R
p = P(M) = [2n(MM)+n(MN)]/[2n(MM)+2n(MN)+2n(NN)]
= (2D+H)/2N = d + h/2
var(p) = var(d+h/2) = var(d) + var(h/2) + 2cov(d,h/2)
= d(1-d)/N + h(1-h)/4N - dh/N
= d/N + h/4N - (d+h/2)^2/N = d/2N + (d+h/2)/2N - 2p^2/2N
= (p + d - 2p^2)/2N ≈ p(1-p)/2N = pq/2N se d ≈ p^2
se(p) = se(q) = √[var(p)] ≈ √[(pq)/2N]
ic95%(p) ≈ p ± 1.96 se(p)
χ^2 = Σi[(oi-ei)^2/ei] = Σi(oi^2/ei) - N
= (H^2-4DR)^2.N/[(2D+H)^2.(H+2R)^2]
q^2 = r ↔ q = √r = √(R/N) = √[n(NN)/N]
var(q^2) = q^2(1-q^2)/N = var(q).(dq^2/dq)^2 = var(q).4q^2
var(q) = var(q^2)/4q^2 = (1-q^2)/4N
se(q) ≈ √[(1-r)/4N] ≈ √[(1-q^2)/4N]

$N(A) = n_1$, $N(a) = n_2$, $n_1+n_2 = N_m$
$N(AA) = n_3$, $N(Aa) = n_4$, $N(aa) = n_5$, $n_3+n_4+n_5 = N_f$
$q_m = n_2/N_m$, $p_m = 1-q_m = n_1/N_m$
$var(q_m) = q_m(1-q_m)/N_m$
$q_f = (n_4+2n_5)/2N_f$, $p_f = 1-q_f = (2n_3+n_4)/2N_f$
$var(q_f) = q_f(1-q_f)/2N_f$
$q = (n_2+n_4+2n_5)/(N_m+2N_f)$, $p = 1-q = (n_1+2n_3+n_4)/(N_m+2N_f)$
$I(q_m) = 1/var(q_m)$, $I(q_f) = 1/var(q_f)$
$q \approx [q_m.I(q_m)+q_f.I(q_f)]/[I(q_m)+I(q_f)]$
$var(q) = 1/I(q) = 1/[I(q_m)+I(q_f)] = q(1-q)/(N_m+2N_f)$

$N(A) = n_1$, $N(a) = n_2$, $n_1+n_2 = N_m$
$N(A-) = N(AA) + N(Aa) = n_3$, $N(aa) = n_4$, $n_3+n_4 = N_f$
$q_m = n_2/N_m$, $p_m = 1-q_m = n_1/N_m$
$var(q_m) = q_m(1-q_m)/N_m$
$q_f = \sqrt{(n_4/N_f)}$, $p_f = 1-q_f = 1-\sqrt{(n_4/N_f)}$
$var(q_f) = (1-q_f^2)/4N_f$
$q = \{-n_1 + \sqrt{[n_1^2+4(N_m+2N_f)(n_2+2n_4)]}\}/2(N_m+2N_f)$, $p = 1 - q$
$I(q_m) = 1/var(q_m)$, $I(q_f) = 1/var(q_f)$
$q \approx [q_m.I(q_m)+q_f.I(q_f)]/[I(q_m)+I(q_f)]$
$var(q) = 1/I(q) = 1/[I(q_m)+I(q_f)] = q(1-q^2)/[N_m+q(N_m+4N_f)]$

```
REM PROGRAM FILENAME EXAULA01.BAS
DEFDBL A-Z: CLS
REM INPUT "N(MM),N(MN),N(NN) = "; D, H, R
DATA 167, 280, 109
READ D, H, R: N = D + H + R
P = (2 * D + H) / (2 * N): Q = 1 - P: VP = P * Q / (2 * N): SEP = SQR(VP)
CHI2 = ((H ^ 2 - 4 * D * R) / ((2 * D + H) * (H + 2 * R))) ^ 2 * N
Q1 = SQR(R / N): P1 = 1 - Q1: SEP1 = SQR((1 - Q1 ^ 2) / (4 * N))
PRINT "Gene frequency estimate with s.e. (codominance)"
PRINT USING "p = P(M) = (2D+H)/(2N) = #.####"; P
PRINT USING "se(p)    = sqr[pq/(2N)]= #.####"; SEP: PRINT
PRINT "testing of Hardy-Weinberg proportions [p^2:2pq:q^2]"
PRINT "GENOTYPE       OBS.NO.      EXP.NO.    CONTR. TO CHI-SQ."
PRINT "-------------------------------------------------------"
PRINT "  MM     "; : PRINT USING "      ###   "; D;
PRINT USING "       ###.##    "; N * P ^ 2;
PRINT USING "   #####.##"; (D - N * P ^ 2) ^ 2 / (N * P ^ 2)
PRINT "  MN     "; : PRINT USING "      ###   "; H;
PRINT USING "       ###.##    "; 2 * N * P * Q;
PRINT USING "   #####.##"; (H - 2 * N * P * Q) ^ 2 / (2 * N * P * Q)
PRINT "  NN     "; : PRINT USING "      ###   "; R;
PRINT USING "       ###.##    "; N * Q ^ 2;
PRINT USING "   #####.##"; (R - N * Q ^ 2) ^ 2 / (N * Q ^ 2)
PRINT "-------------------------------------------------------"
PRINT " total   "; : PRINT USING "      ###   "; N;
PRINT USING "       ###.##    "; N;: PRINT USING "   #####.##"; CHI2: PRINT
PRINT "Gene frequency estimate with s.e. (dominance)"
PRINT USING "p = P(M) = 1 - q = 1 - sqr(R/N) = #.####"; P1
PRINT USING "s.e.(p or q)  = sqr[(1-q^2)/4N] = #.####"; SEP1
```

```
Gene frequency estimate with s.e. (codominance)
p = P(M) = (2D+H)/(2N) = 0.5522
se(p)    = sqr[pq/(2N)]= 0.0149

testing of Hardy-Weinberg proportions [p^2:2pq:q^2]
GENOTYPE       OBS.NO.      EXP.NO.    CONTR. TO CHI-SQ.
-------------------------------------------------------
  MM           167          169.51           0.04
  MN           280          274.97           0.09
  NN           109          111.51           0.06
-------------------------------------------------------
 total         556          556.00           0.19

Gene frequency estimate with s.e. (dominance)
p = P(M) = 1 - q = 1 - sqr(R/N) = 0.5572
s.e.(p or q)  = sqr[(1-q^2)/4N] = 0.0190
```

```
REM PROGRAM FILENAME XLINK.BAS
CLS : DEFDBL A-Z: INPUT "N(A-,aa,A,a) = "; D, R, S, T: NF = D + R: NM = S + T
F = SQR(R / NF): VF = (1 - F * F) / (4 * NF): M = T / NM: VM = M * (1 - M) / NM
PRINT USING "q(f) = #.#### , se[q(f)] = "; F; : PRINT USING "#.####"; SQR(VF)
PRINT USING "q(m) = #.#### , se[q(m)] = "; M; : PRINT USING "#.####"; SQR(VM)
PRINT USING "qmf  = #.####"; (F / VF + M / VM) / (1 / VF + 1 / VM)
Q = (-S + SQR(S ^ 2 + 4 * (2 * R + T) * (2 * NF + NM))) / (2 * (2 * NF + NM))
VQ = Q * (1 - Q ^ 2) / (M + Q * (4 * NF + NM))
PRINT USING "q    = #.#### , se( q  ) = "; Q; : PRINT USING "#.####"; SQR(VQ)
CHISQ = D ^ 2 / (NF * (1 - Q ^ 2)) + R ^ 2 / (NF * Q ^ 2)
CHISQ = CHISQ + S ^ 2 / (NM * (1 - Q)) + T ^ 2 / (NM * Q) - (NM + NF)
PRINT USING "chi-square (1 d.f.) = ##.##"; CHISQ
```

```
N(A-,aa,A,a) = ? 967,102,667,346
q(f) = 0.3089 , se[q(f)] = 0.0145
q(m) = 0.3416 , se[q(m)] = 0.0149
qmf  = 0.3248
q    = 0.3251 , se( q  ) = 0.0130
chi-square (1 d.f.) =  2.44
```

EJERCICIO EN CLASE 02

1. Las frecuencias de haplotipos AB, ab, Ab y aB son respectivamente 0.35, 0.35, 0.15 y 0.15. Sabiendo que la tasa de recombinación r entre los *loci* (A,a) y (B,b) es 0.2, calcular: a) la frecuencia de los haplotipos en la generación siguiente; b) la frecuencia de los haplotipos en equilibrio; c) la frecuencia del genotipo AB/ab en equilibrio; d) la frecuencia de individuos AaBb en equilibrio.

2. En una población panmíctica las frecuencias de los complejos génicos Rh (según la notación de Fisher) son las siguientes:

CDE 0.01	cDE 0.15
CDe 0.40	cDe 0.05
CdE 0	cdE 0.01
Cde 0.02	cde 0.36

Calcular el valor del desequilibrio de ligación para todos esos haplotipos.

3. El gen A1 es uno de los muchos alelos de la serie A y el gen B8 uno de los muchos alelos de la serie B del sistema HLA. Los locus A (A1,...) y B (B1,...) están situados próximos uno del otro en el cromosoma 6. En una muestra poblacional colectada al azar y constituida por 1967 dinamarqueses no emparentados, usando los anti-sueros A1 y B8 fueron obtenidos los siguientes resultados:

anti-A1	anti-B8	nº de indivíduos
+	+	376
+	-	235
-	+	91
-	-	1265

a) ¿Cuáles son las frecuencias de los alelos A1 y B8 del sistema HLA en esa muestra? b) ¿Existe o no asociación entre los antígenos A1 y B8? c) ¿Cuál es la frecuencia esperada, en esa población, del haplotipo A1B8, bajo la hipótesis de equilibrio? d) ¿Cuál es la frecuencia observada, en esa muestra, del haplotipo A1B8? e) ¿Cuál es el valor del desequilibrio de ligación en relación a ese haplotipo? f) Suponiendo que apenas esos dos antígenos (A1 y B8) fueran importantes para la aceptación de transplantes, cuál es la probabilidad de que un transplante sea exitoso en condiciones de emergencia (o sea, que no fuera posible determinar el tipo de donador y el receptor sobre el sistema HLA, sabiendo apenas que ellos no son emparentados)?

EJERCÍCIO EN CLASE 03

1. Un individuo es falsamente acusado de haber cometido un robo. El verdadero ladrón, al forzar el cofre, se hirió la mano, dejando algunas gotas de sangre en el lugar del robo. Eso permitió determinar que el ladrón pertenecía a los grupos sanguíneos M del sistema MN y O del sistema ABO. a) ¿Cuál es la posibilidad de que un individuo acusado falsamente sea excluido de la acusación antes de que se determinen los grupos sanguíneos MN del acusado y del ladrón? b) ¿Cuál es la posibilidad de que, después de determinados los grupos sanguíneos de los sistemas MN y ABO del individuo falsamente acusado, él sea excluido de la acusación? Se sabe que el individuo acusado no tiene ningún grado de parentesco con el verdadero ladrón, que la frecuencia del gen M es 0.55 y que la frecuencia del gen O es 0.65.

2. Sabiendo que el sistema de grupos sanguíneos MN es determinado por un par de alelos autosómicos codominantes (M y N) de frecuencias p y q respectivamente, calcular: a) la probabilidad de exclusión de identidad para un individuo acusado falsamente de haber cometido un robo; b) la probabilidad de exclusión de monocigosidad para un par de gemelos del mismo sexo que en realidad son dicigóticos; c) la probabilidad de exclusión de maternidad para una mujer que falsamente alega que un niño es suyo; d) la probabilidad de exclusión de paternidad para un individuo acusado falsamente por una mujer de ser el padre de un niño que es realmente de ella.

3) Un individuo de esa población es acusado por una mujer de ser el padre de un niño que ella tuvo. Fueron determinados los grupos sanguíneos del trío, con los siguientes resultados:

individuo	madre	niño
M	MN	M

Esos resultados obviamente no excluyen al individuo acusado de ser el padre del niño. Tomando en cuenta apenas los resultados detallados arriba, cuál es la probabilidad de que él sea realmente el padre del niño?

EJERCICIO EN CLASE 04

1) ¿Qué sistemas de cruzamientos son posibles, además de la panmixia?

2) Defina de la forma mas precisa posible los cruzamientos preferenciales y los cruzamientos endogámicos.

3) La distinción entre los dos tipos de sistemas, que parece obvia y completa, es igualmente ambigua. ¿Por qué?

4) ¿Cómo podemos distinguir operacionalmente los dos sistemas?

5) Evolución de la estructura de una población sometida a autofecundación:

generación	**0 : d**	**h**	**r**
	1 : d+h/4	**h/2**	**r+h/4**
	2 : d+h/4+h/8	**h/4**	**r+h/4+h/8**
	3 : d+h/4+h/8+h/16	**h/8**	**r+h/4+h/8+h/16**
	n : d+h/2 = p	**0**	**r+h/2 = q**

5.1) Suponiendo que la probabilidad de homocigosis por origen común (autocigosis) es cero en la generación inicial, cuáles son los valores que F_t toma para t = 1, 2, 3, t y ∞?

5.2) Partiendo de una población inicial panmíctica **{{t = 0} : {d = p^2, h = 2pq, r = q^2}}**, muestre que en cualquier generación las frecuencias genotípicas pueden ser expresadas como **{{pF_t+p^2(1-F_t)}, {2pq(1-F_t)}, {qF_t+q^2(1-F_t)}}** y, en equilibrio, por **{{pF+p^2(1-F)}, {2pq(1-F)}, {qF+q^2(1-F)}}**.

5.3) La distribución de los genotipos en una población con un sistema endogámico de cruzamientos puede ser descripta alternativamente por las formas (1), (2) y (3) abajo, dadas por fórmulas algebráicamente equivalentes. ¿Qué describe particularmente cada conjunto de fórmulas?

	(1)	**(2)**	**(3)**
P(AA)	**pF + p^2(1-F)**	**p^2 + pqF**	**p - pq(1-F)**
P(Aa)	**0 + 2pq(1-F)**	**2pq - 2pqF**	**0 + 2pq(1-F)**
P(aa)	**qF + q^2(1-F)**	**q^2 + pqF**	**q - pq(1-F)**

6) ¿Cuál es el coeficiente medio de endocruzamiento (índice de fijación) de la muestra **{N(AA) = 672, N(Aa) = 256, N(aa) = 72, N = 1000}?**

7) ¿Es ese valor de **F** significativamente diferente de cero? (Note que eso es equivalente a testar si la muestra es panmíctica).

8) ¿Cuál es la relación entre el valor de chi-cuadrado obtenido y el valor estimado de **F**?

$$\chi^2 = (o_i-e_i)^2/e_i = [N(p^2+pqF)-Np^2]^2/(Np^2) + [2Npq(1-F)-2Npq]^2/(2Npq) + [N(q^2+pqF)-Nq^2]^2/(Nq^2) = Nq^2F^2 + 2NpqF^2 + Np^2F^2 = NF^2 \therefore F = (\chi^2/N) .$$

9) En una población sometida a un régimen de auto-fecundación exclusiva la heterocigosis cae a la mitad en cada generación: $\mathbf{H_{t+1} = Ht/_2}$. En un sistema de cruzamientos exclusivos entre hermanos (como los que son realizados para el mantenimiento de las líneas de animales de laboratorio), la heterocigosis de la población se reduce según la relación de recurrencia $\mathbf{H_{t+2} = H_{t+1}/2 + H_t/4}$. Después de un cierto número de generaciones, la heterocigosis de la población deberá caer por generación a qué tasa fija? ¿Por qué?

10) En una determinada región las condiciones eólicas, las visitas de insectos polinizadores y la estructura anatómica de los órganos reproductivos de una población de plantas son tales que el 60% de las fecundaciones ocurren entre gametas producidas por individuos diferentes y 40% entre gametas producidas por el mismo individuo. ¿Cuál es el valor en equilibrio del coeficiente medio de endocruzamiento (**F**) de esa población?

11) La frecuencia de una determinada enfermedad recesiva es 15 veces mayor en la prole de primos en primero grado de que en la prole de matrimonios no consanguíneos. ¿Cuál es la frecuencia del gen que determina la enfermedad?

12) Un antropólogo verificó que, en una población de indios con un complicado sistema de casamientos consanguíneos, 17 de un total de 800 personas eran homocigotas para el alelo A del sistema ABO. Mas tarde el sistema de casamientos que prevalecía en la población desapareció, los individuos pasaron a casarse en un régimen de panmixia y la frecuencia de esos homocigotas cayó a 1/100. ¿Cuál era el valor del coeficiente medio de endocruzamiento (índice de fijación) en la vigencia del antiguo sistema de casamientos?

13) 1000 individuos de una población fueron tipificados para tres *loci* autosómicos independientes, obteniéndose los resultados mostrados abajo:

	D	**H**	**R**
Locus (A,a)	**208**	**384**	**408**
Locus (B,b)	**352**	**396**	**252**
Locus (C,c)	**72**	**256**	**672**

¿Cuál es el sistema de cruzamientos vigente en la población (panmixia, cruzamientos preferenciales o endogamia)?. Justifique su respuesta, indicando el valor de F si se trata de endogamia y el *locus* relacionado si se trata de cruzamientos preferenciales.

EJERCICIO EN CLASE 05

1) Existe una isla, habitada por 990.000 nativos, todos con ojos castaños (genotipo **CC**), llegan 10.000 hombres escandinavos, todos con ojos azules (**cc**). Estos son bien recibidos por los habitantes de la isla y se integran inmediatamente a la población, teniendo hijos con las mujeres jóvenes locales. Se pregunta: ¿Cuáles son las frecuencias **p'** y **q'** de los genes **C** y **c** entre los nativos? ¿Cuáles son los valores de **p"** y **q"** de esos mismos genes entre los escandinavos?. Suponiéndose un régimen de cruzamientos al azar, cuáles serán las frecuencias, en la generación siguiente de la llegada de los escandinavos, de los genotipos **CC**, **Cc** y **cc** en la población de la isla? ¿Cuáles serán las frecuencias **p** y **q** de los genes **C** y **c**? ¿Están las frecuencias genotípicas en las proporciones $\mathbf{p^2:2pq:q^2}$ de Hardy-Weinberg? ¿Por qué? ¿Cuáles serán las frecuencias génicas y genotípicas de la segunda generación y las siguientes, admitiéndose que siempre ocurran casamientos al azar?

2. Los datos mostrados abajo fueron obtenidos a partir de muestreos realizados en tres poblaciones (1, 2, 3), una de las cuales se sabe que fue formada a partir de individuos emigrantes de las dos restantes.

	1	**2**	**3**
AA	**92**	**251**	**127**
Aa	**216**	**138**	**130**
aa	**192**	**11**	**43**

¿Cuál es la frecuencia del gen A en la población 1? ¿Qué se puede decir al respecto de la verdadera frecuencia del gen **A** en la población 2? En qué población el coeficiente medio de endocruzamiento **F** es significativamente diferente de cero? ¿Cuál es su valor? ¿Cuál es la contribución genética, para la formación de la población híbrida, de cada una de las otras dos poblaciones?

3) Entre negroides americanos, la frecuencia de un determinado alelo del sistema Rh es **0.446;** el mismo alelo tiene una frecuencia de **0.630** en poblaciones africanas actuales y **0.028** en poblaciones caucasoides americanas o norte-europeas. ¿Qué porcentajes de genes de origen africano y europeo poseen los negroides norte-americanos? Sabiendo que los africanos fueron introducidos en los EUA a partir de 1675 y que la investigación antes mencionada fue realizada en 1950, calcule la tasa/generación de introducción de genes europeos en el pool génico de la población de origen africana, admitiendo una duración media de 27.5 años por generación.

Generación 0:
Nativos: p', q', N'
Escandinavos: p", q", N"
x' = N'/(N'+2N"), x" = N"/(N'+2N")
pf0 = p', qf0 = q'
pm0 = x'p' + x"p", qm0 = x'q' + x"q"

Generación 1:
P1(AA) = pm0.pf0
P1(Aa) = pm0.qf0 + pf0.qm0
P1(aa) = qm0.qf0
p = pm1 = pf1 = P1(AA) + P1(Aa)/2 = (pm0+pf0)/2 = [(1+x')p' + x"p"]/2
q = qm1 = qf1 = P1(aa) + P1(Aa)/2 = (qm0+qf0)/2 = [(1+x')q' + x"q"]/2

Generación 2 y las siguientes:
P(A) = p, P(a) = q
P(AA) = p^2
P(Aa) = 2pq
P(aa) = q^2

Por los datos numéricos, x' = 495.000/(495.000+10.000) = 0.980198019802 y x" = 10.000/(495.000+10.000) = 0.019801980198, que redondeamos a 0.98 y 0.02; los valores exactos, no redondeados, son x' = 99/101 y x" = 2/101. Aplicando esos valores en p = [(1+x')p' + x"p"]/2, obtenemos p = [(200/101).1 + (2/101).0]/2 = 200/202 = 0.990099009901. Ese resultado es igual a N'/(N'+N") = 990.000/(990.000+10.000) = 0.99 apenas cuando redondeamos, lo que estaba correcto, mientras tanto, pues 0.990099009900... no difiere en orden de magnitud de 0.99.

x = prop. de genes afric. en la muestra
$y = 1 - x$ = prop. de genes europ. en la muestra
p_a = freq. del gen A en la pobl. afric. parental
p_e = freq. del gen A en la pobl. europ. parental
p_h = freq. del gen A en la pobl. híbrida

$$p_h = x.p_a + y.p_e = x.p_a + (1-x).p_e$$
$$x = (p_h - p_e)/(p_a - p_e)$$

$$p_{h1} = p_{h0}.(1-m) + m.p_e$$
$$\lim_{t\to\infty}(p_{ht}) = p_e$$
$$p_{h1} - p_e = p_{h0}.(1-m) + m.p_e - p_e = p_{h0}.(1-m) - p_e.(1-m)$$
$$= (p_{h0} - p_e)(1-m)$$
$$p_{ht} - p_e = (p_{h0} - p_e).(1-m)^t \therefore p_h - p_e = (p_a - p_e).(1-m)^t$$
$$(p_h - p_e)/(p_a - p_e) = (1-m)^t \therefore x = (1-m)^t$$
$$\log(x) = t.\log(1-m) \therefore \log(1-m) = [\log(x)]/t$$
$$1-m = 10^{\{[\log(x)]/t\}}$$
$$m = 1 - 10^{\{[\log(x)]/t\}}$$

EJERCICIO EN CLASE 06

$\mu = P(A \rightarrow a) = 10^{-6}$

t	$e^{-\mu t}$	$(1-\mu)^t$
1	0.9999990	0.9999990
10	0.9999900	0.9999900
100	0.9999000	0.9999000
1000	0.9990005	0.9990005
10000	0.9900498	0.9900498
100000	0.9048374	0.9048374

n	$a = (1 - 1/2n)^{2n}$	$b = e^{-1}$	$\lvert b-a \rvert / b$
1	0.2500000	0.3678794	0.3204295
10	0.3584859	0.3678794	0.0255342
100	0.3669578	0.3678794	0.0025052
1000	0.3677875	0.3678794	0.0002498
10000	0.3678702	0.3678794	0.0000250
100000	0.3678785	0.3678794	0.0000024
1000000	0.3678793	0.3678794	0.0000003

1) La tasa de mutación $\mu = P(A \rightarrow a)$ es de $1/10^6 = 10^{-6}$/generación. Admitiéndose un tratamiento puramente determinístico, ¿cuántas generaciones son necesarias para que la frecuencia del gen **A** se reduzca a 1/3 de la actual?

2) Un *locus* con dos alelos (**B**,**b**) admite dos tasas de mutación, una principal $[\mu = P(B \rightarrow b)]$ y otra secundaria (reversa) $[\nu = P(b \rightarrow B)]$. ¿Cuál es la frecuencia en equilibrio del gen **B**, sabiendo que la tasa de mutación principal es 19 veces superior a la tasa de mutación reversa?

3) En una población razonablemente grande con N individuos diploides y tamaño constante a lo largo de las generaciones, nace un individuo portador de una mutación A absolutamente neutra. ¿Cuál es la probabilidad de que ocurra eliminación de esa mutación al cabo de una generación?

(1) $p_{t+1} = p_t - p_t.\mu = p_t(1-\mu) \rightarrow p_t = p_0(1-\mu)^t \approx p_0.e^{-\mu t}$
$p_t/p_0 \approx e^{-\mu t} = 1/e^{\mu t} \therefore p_0/p_t = e^{\mu t} \therefore t = [\ln(p_0/p_t)]/\mu$

(2) $q_{t+1} = q_t - q_t.\nu + p_t.\mu = q_t - q_t.\nu + (1-q_t)\mu = \mu + q_t(1-\mu-\nu)$
$t \rightarrow \infty \Rightarrow q.\nu = p.\mu = (1-q)\mu \therefore q = \mu/(\mu+\nu)$.

(3) $P(Aa) = 1/N$, $P(A) = 1/2N$, $P(a) = 1-1/2N$
$P(\text{elim } A) = (1-1/2N)^{2N}$
$\lim_{N\rightarrow\infty}[(1-1/2N)^{2N}] = \lim_{x\rightarrow\infty}[(1-1/x)^x]$
$= e^{-1} = \Sigma_{\{i=0,\infty\}}[(-1)^i/i!]$

EJERCICIO EN CLASE 07

1) Una población tiene un tamaño de N = 4 individuos diploides y una frecuencia del alelo A igual a 1/4 [$\mathbf{p_0 = 0.25}$]. Suponiendo que el tamaño de la población es constante a lo largo de las generaciones, se pregunta: (a) ¿Cuántos genes (A o a) son sorteados por generación para constituir la población de la generación siguiente? (b) ¿Cuántas frecuencias génicas (p del alelo A o q del alelo a son posibles en la generación 1 o en cualquier otra? (c) ¿Cuál es la probabilidad de que en la generación 1 la frecuencia génica no oscile en relación a la frecuencia original? (d) ¿Cuál es la probabilidad de que en la generación 1 la frecuencia génica p sea menor que 0.25? (e) y mayor que 0.25? (f) ¿Cuál es la probabilidad de que, después de un número infinitamente grande de generaciones, la frecuencia génica p sea menor o mayor que 0.25? ¿Por qué?

2) Una población tiene un tamaño constante **N = 2** individuos y frecuencia inicial del alelo A igual a 1/2. (a) En la generación siguiente, ¿cuáles serán las probabilidades de que la población se encuentre en los estados j = 0, 1, 2, 3 y 4 (frecuencias génicas 0/4 = 0, 1/4, 2/4 = 1/2, 3/4 y 4/4 = 1)? (b) determine todas las probabilidades de transición de un estado i en la generación t para un estado j en la generación t+1:

i	j	$P(i_t \rightarrow j_{t+1})$	i	j	$P(i_t \rightarrow j_{t+1})$
0	0	1	1	0	
0	1	0	1	1	
0	2	0	1	2	
0	3	0	1	3	
0	4	0	1	4	
2	0		3	0	
2	1		3	1	
2	2		3	2	
2	3		3	3	
2	4		3	4	
4	0	0			
4	1	0			
4	2	0			
4	3	0			
4	4	1			

3) ¿Cómo deberemos proceder para hallar las probabilidades de que la población, en la generación 2 y siguientes, se encuentre en los estados j = 0, 1, 2, 3, 4?

4) Imaginando que lo que acabamos de describir corresponde a un sorteo de un número infinito de poblaciones de tamaño fijo N, ¿qué debe suceder con la frecuencia génica media p o q, a lo largo de las generaciones, de todas esas poblaciones? Y con la variancia de frecuencias génicas entre todas esas poblaciones?

5) Establezca una analogía directa entre el modelo de autofecundación con el destino de una población de tamaño N = 1 individuo (o 2N = 2 genes).

```
REM BINPROB2.BAS
CLS : LOCATE 5: DEFDBL A-Z: DEFINT I-N
INPUT "N, q = "; N, Q
I = 0: PROB = (1 - Q) ^ N: PROBC = PROB
PRINT USING "N = #####"; N: PRINT USING "q = #.###"; Q: PRINT
PRINT "  X    p(X)      P(X)"
PRINT "------------------------"
PRINT USING "###"; I; : PRINT USING "    #.####"; PROB; PROBC
FOR I = 1 TO N
  PROB = (N + 1 - I) * Q * PROB / (I * (1 - Q))
  PROBC = PROBC + PROB
  PRINT USING "###"; I; : PRINT USING "    #.####"; PROB; PROBC
NEXT I
PRINT "------------------------"
```

```
N, q = ? 8,.25
N =     8
q = 0.250
```

X	p(X)	P(X)
0	0.1001	0.1001
1	0.2670	0.3671
2	0.3115	0.6785
3	0.2076	0.8862
4	0.0865	0.9727
5	0.0231	0.9958
6	0.0038	0.9996
7	0.0004	1.0000
8	0.0000	1.0000

1a) 8
1b) 9 (genéricamente, 2N+1, o sea, desde q = 0/2N hasta q = 2N/2N, donde N es el tamaño de la población o número de individuos diploides)
1c) $P(q_1 = q_0)$ = C(8,2).(1/4)^2.(3/4)^6 = 8!/(2!6!).(1/16).(729/4096)
= 28.(729/65536) = 0.3115
1d) $P(q_1 < q_0)$ = C(8,0).(1/4)^0.(3/4)^8 + C(8,1).(1/4)^1.(3/4)^7
= 0.1001 + 0.2670 = 0.3671
1e) $P(q_1 > q_0)$ = 1 - $P(q_1 = q_0)$ - $P(q_1 < q_0)$
= 0.3214
1f) P = 1. Porque después de un número infinito de generaciones la población tendrá una frecuencia génica q = 0 o q = 1.

2a)

P(q=0) = C(4,0).(1/2)^4.(1/2)^0 = 1.(1/2)^4 = 1/16
P(q=1/4) = C(4,1).(1/2)^3.(1/2)^1 = 4.(1/2)^4 = 1/4
P(q=1/2) = C(4,2).(1/2)^2.(1/2)^2 = 6.(1/2)^4 = 3/8
P(q=3/4) = C(4,3).(1/2)^1.(1/2)^3 = 4.(1/2)^4 = 1/4
P(q=1) = C(4,4).(1/2)^0.(1/2)^4 = 1.(1/2)^4 = 1/16

2b)

$P(i_t \to j_{t+1})$: i_t	j_{t+1} = 0	1	2	3	4
0	1	0	0	0	0
1	81/256	27/64	27/128	3/64	1/256
2	1/16	1/4	3/8	1/4	1/16
3	1/256	3/64	27/128	27/64	81/256
4	0	0	0	0	1

3)

$$(1/16,1/4,3/8,1/4,1/16).\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 81/256 & 27/64 & 27/128 & 3/64 & 1/256 \\ 1/16 & 1/4 & 3/8 & 1/4 & 1/16 \\ 1/256 & 3/64 & 27/128 & 27/64 & 81/256 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

= (85/512,27/128,63/256,27/128,85/512),

etc.

4) q_t permanece constante ($q_t = \ldots = q_0$); $v(q_t) = 0$ para $t = 0$, $v(q_t) = p_0q_0/2N$ para $t = 1$ y $v(q_t) = p_0q_0$ para $t = \infty$; y, genéricamente, $v(q_t) = p_0q_0\{1 - [1-(1/2N)^t]\}$

5) El modelo de autofecundación corresponde exactamente a una población de tamaño n subdividida en n subpoblaciones de tamaño **1.** Se concluye por lo tanto que la heterocigosis decae en las dos situaciones según la tasa **1 - 1/2N** por generación.

EJERCICIO EN CLASE 08

1. Existiendo 3 aislamientos de tamaño muy grande; consideramos un par de alelos autosómicos (A,a) segregando en cada uno de ellos. Se supone que los aislamientos tienen el mismo tamaño, que cada uno de ellos está en equilibrio de Hardy-Weinberg y que $\mathbf{p_1 = 1-q_1 = 0.5}$; $\mathbf{p_2 = 1-q_2 = 0.6}$; $\mathbf{p_3 = 1-q_3 = 0.7}$, se pregunta: a) ¿Cuál es el valor de **q** (media de las frecuencias génicas)? b) ¿Cuál es el valor de la varianza interpoblacional de q? c) ¿Cuál es el valor de **P(AA)**, **P(Aa)** y **P(aa)** antes y después del quiebre total de los aislamientos?

2. Siendo n aislamientos de igual tamaño; consideramos un par de alelos autosómicos (A,a) segregando en cada uno de ellos. Se supone que cada aislamiento está en equilibrio de Hardy-Weinberg, se pregunta: a) ¿Cuál es el valor de q (frecuencia media del alelo a en el conjunto de los aislamientos)? b) ¿Cuál es el valor de la varianza interpoblacional de q? c) considerando el conjunto de los aislamientos, ¿Cuál es el valor de **P(AA)**, **P(Aa)** y **P(aa)**? d) habiendo quiebre total de los aislamientos, cuál es el valor de **P(AA)**, **P(Aa)** y **P(aa)**? e) compare los valores determinados en el ítem c con los valores encontrados en el equilibrio de Wright y determine el valor del coeficiente medio de endocruzamiento de la población como función de **Var(q)**, **p** y **q**. ¿Qué es lo que Ud. deduce de eso?

3. Deduzca las fórmulas solicitadas en los ítems (a) hasta (e) arriba mencionados para el caso genérico de n aislamientos con tamaños diferentes.

4. Dos muestras de poblaciones próximas y aisladas, formadas por 500 individuos cada una, son testadas con anti-sueros anti-M y anti-N. Los resultados fueron los siguientes:

Reacción con anti-suero		**Población**	
anti-M	**anti-N**	**1**	**2**
+	**-**	**120**	**198**
+	**+**	**260**	**204**
-	**+**	**120**	**98**

a) ¿Cuál es la frecuencia del alelo M en las poblaciones 1 y 2? b) ¿en cuál de las poblaciones el coeficiente medio de endocruzamiento es estadísticamente diferente de cero? ¿por qué? ¿cuál es su valor? c) ¿cuál es el valor numérico del coeficiente de endocruzamiento F_{ST} generado por el aislamiento reproductivo de las poblaciones 1 y 2 (efecto Wahlund)? g) ¿Cuáles son las frecuencias genotípicas de MM, MN y NN en una población formada por individuos emigrantes de las poblaciones 1 y 2, sabiendo que la contribución de la población 1 para la formación de esa población fue de 25% y que el coeficiente medio de endocruzamiento en esa población híbrida es F = 0.10?

5. Dos aislamientos de tamaños aproximadamente iguales fueron analizados para el mismo *locus* autosómico (A,a):

	AA	Aa	aa
1	0.846	0.108	0.046
2	0.174	0.252	0.574
1+2	0.510	0.180	0.310

Determine los valores de los índices de fijación (**F**) debidos a la endogamia dentro de las subpoblaciones ($\mathbf{F_{IS}}$) y al efecto Wahlund ($\mathbf{F_{ST}}$). ¿Cuál es el índice de fijación de la población total ($\mathbf{F_{IT}}$)? ¿Cuál es la relación entre $\mathbf{F_{IS}}$, $\mathbf{F_{ST}}$ y $\mathbf{F_{IT}}$?

EJERCICIO EN CLASE 09

1) Una isla, poblada en la víspera por 80 náufragos, de los cuales 5 son BB, 30 son Bb y 45 son bb, es devastada por terremotos y plagas, de modo que solamente sobreviven dos parejas. Como los recursos de la isla son ahora escasos, las parejas resuelven, cada una, tener apenas dos hijos, un hábito reproductivo que es pasado a las futuras generaciones sin grandes dificultades, ya que entre esos individuos todos los embarazos redundan en partos gemelares dicigóticos de sexos diferentes. ¿Cuál es la probabilidad de que, en la generación siguiente, las frecuencias de los genes B y b sean iguales a las de la generación anterior? ¿Cuál es la probabilidad de que, después de un número infinitamente grande de generaciones, todos los individuos de la población sean BB?

2) El tamaño efectivo de una población (**Ne**) es obtenido de **1/Ne = 1/4M + 1/4F**, donde **M** es el número de machos y **F** el de hembras. Explicar el por qué de las cantidades que ocurren en la fórmula y lo que significa (o a qué corresponde) el tamaño efectivo de una población. Suponga una manada muy grande de bovinos donde todos los machos son castrados para engorde y corte y apenas un toro es mantenido para reproducción. Si ese sistema persistiera por un número muy grande de generaciones, que acabará fatalmente sucediendo con la población? ¿Por qué? En niveles de pérdida de heterocigosis, a que sistema de endogamia o a que número poblacional (de poblaciones formadas por machos y hembras en proporciones iguales) generando deriva eso corresponde? En términos prácticos, cuál es el valor del tamaño efectivo de esa población?

3) Explicar porque la razón sexual **1:1** es evolutivamente estable.

4) Explicar lo que significan los parámetros calculados por el programa siguiente.

```
DATA 004,032,064
DATA 024,032,144
DATA 272,096,032
FOR I = 1 TO 3
  READ D(I), H(I), R(I)
  N(I) = D(I) + H(I) + R(I)
  N = N + N(I): D = D + D(I): H = H + H(I): R = R + R(I)
  P(I) = (2 * D(I) + H(I)) / (2 * N(I))
  Q(I) = 1 - P(I)
  F(I) = 1 - (H(I) / N(I)) / (2 * P(I) * Q(I))
PRINT "p("; I; ") = "; : PRINT USING "#.####"; P(I)
PRINT "F("; I; ") = "; : PRINT USING "#.####"; F(I)
NEXT I
FOR I = 1 TO 3
  X(I) = N(I) / N: P = P + X(I) * P(I): VP = VP + X(I) * P(I) * P(I)
NEXT I
VP = VP - P * P
Q = 1 - P: FIT = 1 - (H / N) / (2 * P * Q)
FST = VP / (P * Q)
FIS = (FIT - FST) / (1 - FST)
PRINT "p      = "; : PRINT USING "#.####"; P
PRINT "var(p) = "; : PRINT USING "#.####"; VP
PRINT "FIT    = "; : PRINT USING "#.####"; FIT
PRINT "FST    = "; : PRINT USING "#.####"; FST
PRINT "FIS    = "; : PRINT USING "#.####"; FIS
```

p(1) = 0.2000
F(1) = 0.0000
p(2) = 0.2000
F(2) = 0.5000
p(3) = 0.8000
F(3) = 0.2500
p   = 0.5429
var(p) = 0.0882
FIT   = 0.5395
FST   = 0.3553
FIS   = 0.2857

5) ¿Qué son los **genes polimorfos**, los ***loci* polimórficos** y los **polimorfismos genéticos**?

Una muestra de 100 individuos es tipificada sobre los productos de 5 *loci* autosómicos independientes, obteniéndose los siguientes resultados:

locus	D	H	R
A,a	100	0	0
B,b	97	3	0
C,c	87	12	1
E,e	20	40	40
F,f	26	48	26

Describir la variabilidad genética de la población en términos de tasa de *loci* polimórficos y de heterocigosis observada y esperada (índice de diversidad génica).

```
REM PROGRAM FILENAME DIVERSI1.BAS
DEFDBL A-Z: CLS
DATA  A,100,  0,  0
DATA  B, 97,  3,  0
DATA  C, 87, 12,  1
DATA  D, 20, 40, 40
DATA  E, 26, 48, 26
FOR I = 1 TO 5
READ ALLELE$, D, H, R
N = D + H + R: P = (2 * D + H) / (2 * N): Q = 1 - P
IF P >= .01 AND Q >= .01 THEN P1 = P1 + 1
IF P >= .05 AND Q >= .05 THEN P5 = P5 + 1
HE(I) = P * P + Q * Q: SHE = SHE + HE(I): HO(I) = H / N: SHO = SHO + HO(I)
PRINT "LOCUS (" + ALLELE$ + "," + LCASE$(ALLELE$) + ")   ";
PRINT USING "h(e) = #.###   "; 1 - HE(I);
PRINT USING "h(o) = #.###   "; HO(I);
IF P >= .01 AND Q >= .01 THEN PRINT " +    ";  ELSE PRINT " -    ";
IF P >= .05 AND Q >= .05 THEN PRINT " +    " ELSE PRINT " -    "
NEXT I: PRINT
HEM = 1 - SHE / 5: HOM = SHO / 5
FOR I = 1 TO 5
VHEM = VHEM + (1 - HE(I) - HEM) ^ 2: VHOM = VHOM + (HO(I) - HOM) ^ 2
NEXT I
VHEM = VHEM / 20: VHOM = VHOM / 20
PRINT USING "H(exp) = #.### +/- "; HEM; : PRINT USING "#.###"; SQR(VHEM)
PRINT USING "H(obs) = #.### +/- "; HOM; : PRINT USING "#.###"; SQR(VHOM)
PRINT USING "PPL(1%) = #.###"; P1 / 5
PRINT USING "PPL(5%) = #.###"; P5 / 5
```

LOCUS (A,a) h(e) = 0.000 h(o) = 0.000 - -
LOCUS (B,b) h(e) = 0.030 h(o) = 0.030 + -
LOCUS (C,c) h(e) = 0.130 h(o) = 0.120 + +
LOCUS (D,d) h(e) = 0.480 h(o) = 0.400 + +
LOCUS (E,e) h(e) = 0.500 h(o) = 0.480 + +
H(exp) = 0.228 +/- 0.109
H(obs) = 0.206 +/- 0.098
PPL(1%) = 0.800
PPL(5%) = 0.600

EJERCICIO EN CLASE 10

1. 1500 huevos resultantes de cruzamientos entre heterocigotas (**Aa × Aa**) de *Drosophila melanogaster* son colocados en una caja de poblaciones. Todos los individuos son verificados inmediatamente después de su eclosión del pupario y algunos días después, cuando todos ya están en activa fase de reproducción. 1500 huevos colocados por las hembras son transferidos para una nueva caja de poblaciones y el procedimiento de tipificación fenotípica es repetido. Fueron verificados los siguientes resultados:

		AA	Aa	aa
generación 0	**- emergencia**	**255**	**510**	**255**
	madurez	**207**	**414**	**0**
generación 1	**- emergencia**	**432**	**432**	**108**
	madurez	**321**	**321**	**0**

¿Está ocurriendo selección? ¿En que fase? ¿Cuáles son los valores adaptativos de los tres genotipos? ¿Cuáles son los coeficientes de selección de los tres genotipos? En cada generación, ¿cuáles son las frecuencias génicas antes y después de que actúe la selección? ¿Qué deberá suceder después de un número grande de generaciones? ¿Cuál es el número de generaciones necesarias para que la frecuencia del gen **a** caiga a 1/5 de la frecuencia inicial?

2. Dadas las poblaciones descriptas debajo, cuáles son las frecuencias, en equilibrio, de los alelos A y a, sabiendo que los valores adaptativos de los genotipos **AA**, **Aa y aa** son respectivamente **W1**, **W2 y W3**?

población	W1	W2	W3
1	1,00	1,00	0,00
2	1,00	1,00	0,80
3	0,00	1,00	0,00
4	0,00	1,00	1,00
5	0,80	1,00	1,00

3. En una región de África ecuatorial endémica para la malaria producida por el *Plasmodium falciparum*, a cada generación por vuelta del nacimiento la frecuencia de individuos **SS** es 1/100. La malaria es bastante antigua en esa región, de manera que podemos considerar que esa población está en equilibrio. Los eritrocitos con hemoglobina **S** son parasitados con baja probabilidad por el *Plasmodium falciparum.* Los individuos que presentan apenas hemoglobina **A** (normal del adulto) presentan una alta tasa de mortalidad por causa de la malaria. Los individuos **SS** (que presentan en sus eritrocitos apenas hemoglobina **S**) exhiben una tasa de mortalidad prácticamente de 100% debido a la anemia falciforme. ¿Cuál es la frecuencia, en equilibrio, del alelo **S**? ¿Cuáles son las frecuencias, al nacer y a la edad reproductiva, de individuos **AA**, **AS** y **SS** en cada generación? Suponga que la malaria sea erradicada milagrosamente en una única generación de esas regiones. ¿Qué sucederá con ese polimorfismo genético, al cabo de un número grande de generaciones? Después de un número infinitamente grande de generaciones, suponga que nacen, en cada generación, 4 individuos **SS** dentro de 1.000.000 recién nacidos de ambos sexos. ¿Cuál es la tasa de mutación del alelo **S**?

EJERCICIO EN CLASE 11

1) En regiones de África ecuatorial endémicas para la malaria producida por el *Plasmodium falciparum* los valores adaptativos de los individuos **AA** (hemoglobina normal en el adulto), **AS** (heterocigotas con el trazo siclémico y que son aparentemente normales en cualquier ambiente) y **SS** (homocigotas afectados por anemia falciforme) son respectivamente **0.5**, **1.0** y **0.1**. Se sabe que la malaria es bastante antigua en esas regiones (de modo que podemos considerar a las poblaciones de esas áreas en equilibrio) y que los eritrocitos conteniendo hemoglobina **S** son parasitados con baja probabilidad por el *Plasmodium falciparum* y que, además de eso, los individuos que presentan en sus eritrocitos apenas hemoglobina **A** (normal del adulto) presentan una alta mortalidad por causa de la malaria y que los individuos **SS** (que presentan en sus eritrocitos apenas hemoglobina **S**) presentan un altísimo grado de mortalidad debido a la anemia falciforme (esto tanto en los ambientes malarígenos como en los ambientes donde no prevalece la enfermedad). Se pregunta: a) ¿Cuál es la frecuencia, en equilibrio, del alelo **S**? b) ¿Cuál es la frecuencia, al nacer, de individuos **AA**, **AS** y **SS** en cada generación? c) ¿Cuál es la frecuencia, en la población adulta que se reproduce, de individuos **AA**, **AS** y **SS** en cada generación?

2) En la incompatibilidad Rh materno-fetal existe selección en contra de individuos Rh(+) nacidos de mujeres Rh(-). ¿Por qué? Bajo la hipótesis de panmixia, cuáles son las frecuencias de individuos heterocigotas nacidos de mujeres DD, Dd y dd? Suponiéndose que el valor adaptativo de los heterocigotas nacidos de mujeres dd sea **w = 1-s** y que el valor de los demás heterocigotas sea 1, cuál es el valor adaptativo medio **w' = 1-s'** de los heterocigotas Dd? ¿Cuál es el valor adaptativo medio de la población? ¿Cuándo ese valor adaptativo medio deberá ser un máximo? ¿Cuál es la relación de recurrencia entre la frecuencia del gen d en dos generaciones consecutivas (t y t+1)? Haciéndose $\Delta q = q_{t+1} - q_t = 0$, ¿cuáles son los valores posibles de equilibrio para la frecuencia q del gen d? ¿En qué condiciones ocurren cada una de las formas posibles de equilibrio?

	AA	**AS**	**SS**
1) frec. al nacimiento	**p^2**	**2pq**	**q^2**
2) valores adaptativos	**1/2**	**1**	**1/10**
3) frec. en madurez	**5p^2/w**	**20pq/w**	**q^2/w**

w = 5p^2+20pq+q^2

p' = (5p^2+10pq)/w
q' = (10pq+q^2)/w

en equilibrio,

p'/q' = p/q = (5p^2+10pq)/(10pq+q^2) = p/q . (5p+10q)/(10p+q)
10p+q = 9p+1 = 5p+10q = 5+5q = 10-5p
14p = 9, p = 9/14 , q = 5/14

o (diretamente)

p = s(SS)/[s(AA)+s(SS)] = (9/10)/(1/2+9/10) = 9/14
q = s(AA)/[s(AA)+s(SS)] = (1/2)/(1/2+9/10) = 5/14

por lo tanto:

al nacimiento

P(AA) = 81/196 = 0.413265
P(AS) = 90/196 = 0.459184
P(SS) = 25/196 = 0.127551
P(A) = P(AA) + P(AS)/2 = (81+45)/196 = 126/196
= 9/14 = 0.642857
P(S) = P(AS)/2 + P(SS) = (45+25)/196 = 70/196
= 5/14 = 0.357143

en la madurez

P'(AA) 81/196 . 1/2
= 81/392 81/266 = 0.304511
P'(AS) 90/196 . 2/2
= 180/392 180/266 = 0.676692
P'(SS) 25/196 . 1/10
= 5/392 5/266 = 0.018797

w = 266/392 1.000000

P'(A) = P'(AA) + P'(AS)/2 = (81+90)/266
= 171/266 = 9*19/14*19 = 9/14
P'(S) = 1 - P'(A) = 5/14

P(hijo Dd, madre DD) = qp^2 , w(Dd,DD) = 1
P(hijo Dd, madre Dd) = 2pq(p/2+q/2) = pq , w(Dd,Dd) = 1
P(hijo Dd, madre dd) = pq^2 , w(Dd,dd) = 1-s
w(Dd) = [qp^2 . 1 + pq . 1 + pq^2 . (1-s)]/2pq
= (2pq - spq^2)/2pq = 1 - sq/2
w = p^2 . 1 + 2pq . (1 - sq/2) + q^2 . 1 = 1 - spq^2

q' = [q^2 + qp^2/2 + pq/2 + pq^2(1-s)/2]/(1-spq^2)
= [q^2 + qp^2/2 + pq/2 + pq^2/2 - pq^2/2 + pq^2(1-s)/2]/(1-spq^2)
= [q^2 + pq - pq^2/2(1-1+s)]/(1-spq^2)
= [q - spq^2/2]/(1-spq^2)

Δq = q'- q = [q-spq^2/2-q(1-spq^2)]/(1-spq^2)
= (spq^3-spq^2/2)/(1-spq^2)
= spq^2(q-1/2)/(1-spq^2)

EJERCICIO EN CLASE 12

1. Para genes autosómicos recesivos, la tasa de mutación del gen [**$\mu = P(A \rightarrow a)$**] es calculada según **$\mu = sq^2$** ¿Por qué? ¿Cuál es la frecuencia de individuos **aa** eliminados por generación? ¿Cuál es la frecuencia de genes **a** eliminados por cada generación? ¿Por qué? Dado un afectado aa cualquiera, cuáles son las probabilidades que favorecen las posibilidades de que el caso sea heredado o sea el resultado de uma mutación nueva?

2. Para genes autosómicos dominantes, la tasa de mutación del gen [**$\mu = P(b \rightarrow B)$**] es calculada según **$\mu = spq \approx sp$**. ¿Por qué? ¿Cuál es la frecuencia de afectados (heterocigotas) **Bb** eliminados por generación? ¿Cuál es la frecuencia de genes **B** eliminados en cada generación? ¿Por qué? Dado un afectado **Bb** cualquiera, cuáles son las probabilidades que favorecen las posibilidades de que el caso sea heredado o sea el resultado de una mutación nueva?

3. Para genes recesivos ligados al cromosoma X, la tasa de mutación del gen [**$\mu = P(b \rightarrow B)$**] es calculada según **$\mu = sq/3$**. ¿Por qué? ¿Cuál es la frecuencia de afectados (hemicigotos) **c** eliminados en cada generación? ¿Cuál es la frecuencia de genes **c** eliminados en cada generación? ¿Por qué? Dado un afectado **c** cualquiera, cuáles son las probabilidades que favorecen las posibilidades de que el caso sea heredado o que sea el resultado de uma mutación nueva? Dado una mujer heterocigota normal **Cc** cualquiera, cuáles son las probabilidades favoreciendo las posibilidades de que ella haya heredado el gen o de que sea el resultado de una mutación nueva?

4. Escribir las fórmulas que permiten la estimativa de las tasas de mutación para las tres situaciones de arriba usando las cantidades **μ** (tasa de mutación), **s** (coeficiente de selección) y **x** (frecuencia de afectados al nacer).

5. De entre 100.000 niños nacidos en maternidades de Dinamarca, se verificó que 10 eran afectados por acondroplasia; dos de estos 10 niños tenían uno de los progenitores también afectado; los 8 restantes eran casos aislados. Sabiendo que la acondroplasia es determinada por un gen autosómico dominante de penetrancia completa, que los afectados son heterocigotas para ese gen y que el valor adaptativo de los acondroplásicos fue calculado en 20%, estimar la tasa de mutación del gen de la acondroplasia: a) por el método directo; b) por el método indirecto.

6. La distrofia muscular progresiva de tipo Duchenne es una enfermedad producida por un gen letal recesivo ligado al cromosoma X. La tasa de mutación de ese gen fue estimada en **$\mu = 0{,}000025$**. Se pregunta a) ¿Cuál es la probabilidad de que una mujer cualquiera sea heterocigota para ese gen? b) ¿Cuál es la probabilidad de que una mujer con un hijo afectado por la enfermedad sea heterocigota para ese gen? c) ¿Cuál es la probabilidad de que una mujer con dos hijos afectados por la enfermedad sea heterocigota para ese gen? d) ¿Cuál es la frecuencia, al nacer, de niños de sexo masculino afectados por la distrofia muscular progresiva de tipo Duchenne?

7. Un joven presenta distrofia muscular progresiva de tipo Becker. Sabiendo que el valor adaptativo de los afectados es 0.7, que riesgo de recurrencia Ud. daría para un próximo hermano de ese afectado? Calcule ese riesgo bajo las hipótesis de que el caso sea aparentemente aislado, con y sin informaciones sobre la normalidad de sus ascendentes directos.

8. Un matrimonio normal tiene un hijo afectado por fenilcetonuria, enfermedad condicionada por un mecanismo autosómico recesivo (la frecuencia del gen que condiciona la enfermedad es del orden de 1%). Explicar por que el riesgo de recurrencia de la enfermedad para un próximo hijo es 1/4.

9. Cierto gen dominante tiene una penetrancia de 0.7. Un matrimonio normal tiene un hijo afectado por la enfermedad producida por ese gen en heterocigosis, que es el único caso en la familia. ¿Cuál es el riesgo para el próximo hijo del matrimonio? Suponga que no sea conocida la penetrancia del gen. ¿Qué riesgo Ud. daría para el matrimonio?

10. La frecuencia del albinismo es de 1/10.000. El valor adaptativo de los afectados (homocigotos recesivos) es de 90% y el valor adaptativo de los heterocigotas fue estimado en 99%. Se pregunta: a) ¿Cuál es la frecuencia del gen del albinismo? b) ¿Cuál es el coeficiente de selección de los afectados? c) Siendo **t** el coeficiente de selección de los homocigotas y **ht** el de los heterocigotas, ¿Cuál es el valor de **h**? d) ¿Cuál es la eliminación, por generación, de genes recesivos debido a la selección en contra de los homocigotas recesivos? e) ¿Cuál es la eliminación, por generación, de genes recesivos debido a la selección contra los heterocigotas? f) Estime la tasa de mutación basándose apenas en la eliminación de genes de los homocigotas. g) Estime la tasa de mutación basándose apenas en la eliminación de genes de los heterocigotas. h) Estime la tasa de mutación basándose en la eliminación de genes a través de homocigotas **aa** y de heterocigotas **Aa**. i) Para $\mathbf{q = 0.01}$ y $\mathbf{t = 0.1}$, ¿Cuál es el valor de **h** que torna la eliminación de genes a igual entre homocigotas y heterocigotas? j) En ese caso (pregunta i), ¿Cuál seria la tasa de mutación?

11. La tasa de mutación para genes deletéreos es calculada según $\mu = s_2.pq + s_3.q^2$. ¿Qué significan, en esa fórmula, s_2, s_3, pq, q^2, $s_2.pq$ y $s_3.q^2$? ¿Por qué es siempre importante verificar si existe alguna selección contra los heterocigotas, aunque su intensidad sea de um orden de magnitud bastante inferior a la observada contra los homocigotas?

12. La tasa de mutación μ **[P(A→a)]** tiene un valor de 1/1000 y la tasa reversa ν **[P(a→A)]** vale 10% de μ. ¿Cuál es el valor numérico de la frecuencia **q** en equilibrio del gen **a**? Si partimos de una población con frecuencias alélicas p_0 y q_0 iguales, cuáles son los valores de esas frecuencias en la generación siguiente? Y si μ fuera igual a ν?

13. En una población **A**, el coeficiente medio de endocruzamiento es cero, mientras que en otra población B el coeficiente medio de endocruzamiento es $\mathbf{F = 0.3}$. Una determinada enfermedad autosómica recesiva ocurre con frecuencia de 1/10.000 en la población A. Sabiendo que en ambas poblaciones los números medios de hijos de afectados y no afectados son respectivamente **1.40** y **1.75**, se pregunta: a) ¿Cuál es el coeficiente

selectivo (**s = 1-W**) de la enfermedad? b) ¿Cuál es la tasa de mutación del gen que determina la enfermedad? c) ¿Cuál es la frecuencia de afectados en la población **B**? d) ¿Cuál es la frecuencia del gen en la población **A**? e) ¿Cuál es la frecuencia del gen en la población **B**? f) Interprete el porqué de la diferencia, si esta existe.

Genética de Poblaciones Humanas

ISBN 978-950-579-113-2

Editorial Universitaria
Universidad Nacional de Misiones