

Maximum Likelihood Estimation of Parameters in Linkage. Comparative Numerical Iterative Methods

Rodica SOBOLU¹⁾, Dana PUSTA²⁾, Ioana POP¹⁾

¹⁾Faculty of Horticulture, University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, 3-5 Manastur Street, 400372, Romania, Tel: +0040264-593792

²⁾Faculty of Veterinary Medicine, University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, 3-5 Manastur Street, 400372, Romania, Tel: +0040264-593792

*¹⁾Corresponding author, e-mail: rodicasobolu@yahoo.com, danapusta@yahoo.co.uk

BulletinUASVM Horticulture 72(2) / 2015

Print ISSN 1843-5254, Electronic ISSN 1843-5394

DOI:10.15835/buasvmcn-hort: 11418

Abstract

Taking into account the advantages of Maximum Likelihood Method (most precise estimation), the statistical properties of MLEs (unbiasedness, consistency, efficiency, invariance, asymptotic normality) this paper aim is to present MLE in the context of estimate the recombination fraction r in linkage analysis. Maximum Likelihood Method follows some steps: specifies the likelihood function; takes derivatives of likelihood with respect to the parameters; sets the derivatives equal to zero and finally generates a *likelihood equation*, that maximized provides the most precise estimation of the recombination fraction. Generally, it is solved by iterative procedures, if no, closed form solution exists for *likelihood equation*. In this work we discuss comparatively two iterative optimization methods useful in computing MLE of the recombination fraction: Newton-Raphson method and Fisher's Method of Scoring. We implemented these two methods in Maple application and we illustrated them by an example: the estimation of the recombination fraction in the case of the Morgan (1909) experiment on fruit flies. The Maple code for these two methods connected with the Morgan example is given in the appendix. We can not guarantee which of the two presented methods give us an optimal maximum.

Keywords: recombination fraction, MLE estimation, linkage analysis.

Introduction

Most traits of biological and agricultural importance are under the control of genes which are in interaction, each with a small effect, and of environmental factors. Linkage analysis is a powerful tool to study the structure of the genome and detect the location of genes responsible for same quantitative trait of interest within the pedigrees. The genes are identified by mapping their positions relative to known marker loci. The degree of linkage between any two markers is expressed in terms of the recombination fraction r . The maximum likelihood estimate (MLE) is a statistical inference procedure that provides precise estimators for some parameters. Fisher

published first the method of maximum likelihood in 1912 in term of "inverse probability" and outlined the term "likelihood" in 1921. The full development of the method appears in articles of 1922 and 1925. MLE is the value of some parameter(s) that maximizes the likelihood of the observed samples, i.e. makes the observed data most likely to occur. We present MLE in the context of optimization methods specialized in statistics and we discuss some numerical methods useful for solving statistical optimization problems. Devore et al. 2012 reveal some properties of maximum likelihood estimators. In this study we take certain known properties of these types of estimators and adapt them in the context of numerical

optimization methods for likelihood function. We assume only one unknown parameter θ .

2. Preliminaries

2.1 Linkage and recombination fraction

Linkage is the tendency for genes to be inherited together because they are located near one another on the same chromosome. A measure of the linkage is described in terms of the number of recombination fraction r between two loci. The r value quantifies the amount of linkage between

two genes, $r = \frac{R}{N}$ where R is the number of

recombinant gametes and N is the number of gametes in a sample of meioses. The recombination fraction has biological significance in the case when $r \in [0, 0.5]$. The value $r=0$ shows complete linkage and the value $r=0.5$ indicates the independence of the two loci. A.H. Sturtevant (1913) first used recombination fraction to order, i.e. map genes along fruit fly chromosomes.

2.2 Likelihood-Estimation

The likelihood function is a theoretical concept that underlies many statistical inferences made in genetics. This function is the probability of the observed data, considered as a function of the parameter. Let be $y = (y_1, y_2, \dots, y_n)$ an iid sample and θ a scalar parameter or a parameter vector $\theta = (\theta_1, \theta_2, \dots, \theta_k)$, $k \leq n$. The likelihood function is

$$L(\theta | y) = f(y|\theta) \quad (1), \quad \text{where}$$

$f(y|\theta) = \prod_{i=1}^n f(y_i|\theta)$ is the sample density

function. Assume that $L(\theta | y)$ is twice-differentiable with respect to θ . The Maximum Likelihood Estimation method (MLE) chooses the estimated parameter $\hat{\theta}$, as the value that maximizes the log-likelihood function:

$$\log L(\theta | y) = \sum_{i=1}^n \log f_i(y_i | \theta) \quad (2).$$

The MLE estimator is the value $\hat{\theta}$ such that $\log L(\hat{\theta} | y) \geq \log L(\theta | y)$ for all θ . MLE method generates an equation, that maximized provides the most precise estimate.

In what follows we consider the Fisher's score function

$$S(\theta) = \frac{\partial \ln L(\theta | y)}{\partial \theta} \quad (3).$$

If θ is a parameter vector than $S(\theta)$ is the score vector of first partial derivatives, one for each component of θ . The first derivative of the score function multiplied by -1 is called the observed information

$$I(\theta) = -\frac{\partial S(\theta)}{\partial \theta} \quad (4).$$

The quantity $I(\theta)$ sets the sharpness of the peak in the likelihood function around its maximum. The expected information is defined as

$$J(\theta) = \text{Var}(S(\theta)) = E[I(\theta)] \quad (5).$$

Note that for large i.i.d. samples it holds approximately that $\hat{\theta} \sim N(\theta, J(\theta)^{-1})$. The score function has some interesting statistical properties: the expectation of $S(\theta)$ with respect to y is equal to zero $E[S(\theta)] = 0$ and the inverse of the information value of a MLE is its variance.

$$E(S(\theta)) = \int_{-\infty}^{+\infty} \frac{\partial L(\theta | y)}{\partial \theta} L(\theta | y) dy = \int_{-\infty}^{+\infty} \frac{\partial L(\theta | y)}{\partial \theta} L(\theta | y) dy = \frac{\partial}{\partial \theta} \int_{-\infty}^{+\infty} L(\theta | y) dy = 0 \quad (6).$$

$$\text{Var}(S(\theta)) = E \left[\left[\frac{\partial \ln L(\theta | y)}{\partial \theta} \right]^2 | y \right] = \int_{-\infty}^{+\infty} \left(\frac{\partial \ln L(\theta | y)}{\partial \theta} \right)^2 L(\theta | y) dy = I(\theta) \quad (7).$$

$$\text{Var}(\hat{\theta}) = I^{-1}(\hat{\theta}) = \left[-E \left(\frac{\partial^2 S(\theta)}{\partial \theta \partial \theta^T} \right) \right]^{-1} \quad (8).$$

2.3 Numerical Methods for solving equations

Numerical analysis combines computer science and mathematics in order to create and implement algorithms for solving equations or

systems of equations that are not linear. There are many iterative methods which approximate the solution of these types of equations: Newton's method, Bisection method, Secant method.

Newton's method is widely used for function optimization. It calls an iterative formula for finding a maximum of $\log L(\theta)$ and thus for solving the equation $S(\theta) = 0$ (10) when the direct solutions of this do not exist.

Newton's method (also known as the Newton-Raphson method) was proposed by the most influential scientists of all times, Isaac Newton (1642-1727). Generally, this method facilitates obtaining iteratively better approximations to the roots of the equation of type $f(x) = 0$ (11), where f is a real-valued function. Newton's method uses Taylor's theorem to approximate the equation (11).

We present below the Newton method in one variable (quoted by Ram 2010, Micula et. al, 2008): given a function f defined over the set of real numbers and its first derivative f' , we start with an initial guess x_0 for a root of the function f . A better approximation x_1 is expressed by the recurrence

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \quad (12),$$

where $(x_1, 0)$ is the intersection with the x -axis of a line tangent to f at the point $(x_0, f(x_0))$. The process is repeated as

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad (13)$$

until the increment becomes small, so $|x_{k+1} - x_k| < \varepsilon$ (14). The iterative process is graphically shown in Figure 1.

Let the actual root of the equation (11) be α , $f(\alpha) = 0$. A sequence of iterates x_k that converges to α has order of convergence $p > 1$ if

$$\lim_{k \rightarrow \infty} \frac{|x_{k+1} - \alpha|}{|x_k - \alpha|^p} = c, \text{ where } 0 < c < 1 \text{ is the convergence}$$

factor. Convergence is quadratic if $p=2$, linear if $p=1$ and sub linear if $p < 1$. When we choose x_0 sufficiently near α Newton's method converges

very fast. In the case when the start value x_0 is far from α the methods presents lack of stability because it oscillates and not converges at all. Suppose that we solve the equation of type (11) on

the interval $[x_0, x_1]$. In this article, the starting point is chosen as the point for which the condition

$f(x_1) \cdot f''(x_1) > 0$ is satisfied. Also, we consider that the actual root α of the equation (11) is approximated with the error ε if the following condition is fulfilled

$$\frac{|f(x_n)|}{m} < \varepsilon, \quad m = \min_{x \in (x_n, \alpha)} f'(x) \quad \text{and} \quad |\alpha - x_n| < \frac{|f(x_n)|}{m} \quad (15).$$

2.4 Estimation of linkage

The phenomenon of variable linkage between characters or recombination is a powerful tool in the process of location of genes of interest. Mendel (1866) enumerated the offspring types of a hybrid cross. He made artificial self-fertilization and studied garden pea characters, each character having two phenotypes. For F2 self-fertilization with recessive phenotype Mendel found that the resulting offspring were all of recessive type. Likewise he noticed that for F2 self-fertilization with dominant phenotype 2/3 of the resulting offspring were both phenotypes in a dominant to recessive ratio 3:1 and 1/3 of the resulting offspring were dominant. Starting from this experiment Mendel formulated the First Law of heredity. Mendel (1866) also studied two characters at a time and he formulated the Second Law of independent segregation of characters. Mendel's work was published in 1866. Similar experiments were performed (Correns, 1900; Tschermak, 1900; Vries, 1900) in order to rediscover Mendel's results. The phenomenon of

complete linkage (complete gametic coupling - Correns, 1900) highlights that the alleles of two or more different characters are always inherited together. This theory contradicts the Mendel's second law, but still there is a generalization of this law. It states that in complete linkage, the genes are physically attached. The development of the theory of heredity (Sutton, 1903; Morgan, 1911; Sturtevant, 1913) also considered the genetic linkage understanding. It was analyzed the inheritance of characters in *Drosophila melanogaster* and represented in a genetic map the physical locations of several genes located on the X chromosome of *Drosophila melanogaster*.

2.5 Mating Design

Generally, linkage analysis is studied in backcross design and in F2 design. These mating designs cross two inbred lines which are different in both the phenotypic values of traits and allele frequencies of marker loci considered in the linkage map.

In what follows we analyze the F2 mating design and we take into account two markers in order to find the MLE estimation of the recombination fraction. Let r be the recombination fraction between loci A and B. Let P1 and P2 be the two parents that initiate the line cross. The hybrid progeny between P1 and P2 is denoted by F1. The intercross between two F1 hybrids is called F2 progeny.

Let n_{ij} the number of individuals combining the i^{th} gamete from one parent and the j^{th} gamete from other parent, $j=1,2,3,4$. The 16 possible genotypes are displayed in the gametic array from the table below.

We assume the existence of linkage phase for two loci with allele A and a respectively B and b , so A and B are on one homologous chromosome and a and b are on the other homologous chromosome. Let r be the recombination fraction between loci A and B. Considering coupling linkage, AB and ab are parental gametes ($AB//ABxab//ab$) respectively Ab and aB are recombinant gametes. Under repulsion linkage phase Ab and aB are parental gametes ($Ab//AbxB//aB$) and AB and ab are recombinant gametes.

We consider the case of coupling when r is the parental combination fraction, so the recombination fraction is $1-r$. We will estimate r in

F2 starting from the observed numbers of the four different F2 genotypes.

We can calculate the expected proportions φ_i , $i=1,2,3,4$ directly from the gametic array. For

$$\text{example, } \varphi_1 = \frac{3r^2}{4} + \frac{2(1-r)^2}{4} + \frac{4r(1-r)}{4}.$$

2.6 MLE estimation of linkage

The MLE of r can be used to establish the degree of linkage between two given genes. If the two genes are completely linked i.e. $\hat{r} = 0$ then in F2 we have only non-recombinant (parental) gametes. If the linkage is absent i.e. $\hat{r} = 0.5$ then in F2 we have both recombinant and non-recombinant (parental) gametes.

The segregation data reveals the number of individuals observed in each genotypic class: AB, Abs, abs, ab. In what follows we use the multinomial distribution to model frequencies.

Let be n_i the number of individuals in the i^{th} genotypic class, k the number of genotypic classes and p_i the probability of choosing an individual

from l class. We let $p = (p_1, p_2, \dots, p_k)$, $\sum_{i=1}^k p_i = 1$. Probabilities p_i can be represented as functions of r using the gametic array. We also know the frequencies n_i from the observed data. The likelihood function that will provide the value of the unknown parameter r is

$$L(r) = \frac{n!}{n_1!n_2!\dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k} \quad (16).$$

In order to simplify calculations we find the logarithm of the expression (16). We obtain

$$\ln L(r) = c + n_1 \ln p_1 + n_2 \ln p_2 + \dots + n_k \ln p_k \quad (17).$$

We find MLE estimator of r maximizing the equation $\mathbf{h} L(r) = 0$ (18). Therefore we differentiate with respect to r and obtain the system

$$\frac{\partial \ln L(r)}{\partial r} = 0 \tag{19}$$

Which is equivalent to

$$n_1 \frac{\partial \ln(p_1)}{\partial r} + n_2 \frac{\partial \ln(p_2)}{\partial r} + \dots + n_k \frac{\partial \ln(p_k)}{\partial r} = 0 \tag{20}$$

So, the log likelihood in the case of multinomial distribution is

$$\ln L(r) = \sum_{i=1}^k n_i \ln p_i(r) \tag{21}$$

The corresponding score vector is

$$S(r) = \nabla \ln L(r) = \sum_{i=1}^k \frac{n_i}{p_i(r)} \nabla p_i(r) \tag{22}$$

and the observed information is

$$I(r) = d^2 \ln L(r) = \sum_{i=1}^k \left\{ -\frac{n_i}{p_i(r)^2} \nabla p_i(r) dp_i(r) + \frac{n_i}{p_i(r)} d^2 p_i(r) \right\} \tag{23}$$

We have $E(n_i) = np_i(r)$ which implies

$$I(r) \approx -\sum_{i=1}^k \frac{n_i}{p_i(r)} \nabla p_i(r) dp_i(r) \approx -n \sum_{i=1}^k \frac{1}{p_i(r)} \nabla p_i(r) dp_i(r) \tag{24}$$

The symbol E represents expectation of the data given to the parameter value. The equation $S(r) = 0$, which provides the MLE estimator of the parameter r can be written in the following simplified form

$$(n_1 + n_2 + n_3 + n_4)r^4 - (n_1 - 2n_2 - 2n_3 - n_4)r^2 - 2n_4 = 0 \tag{25}$$

Material and method

In this paper we estimate the recombination fraction r in F2 coupling using the MLE method. We consider Morgan (1909) experiment on fruit flies. He crossed two inbred lines with the following genotypic traits

Eye color	A: red	a: purple
Wing length	B: normal	b: vestigial

He obtained 2839 crosses AB//ABxab//ab. In F2 coupling the observed numbers of individuals for each genotypic class i is $i=1,2,3,4$ were

	AB	Ab	aB	ab	Total
Observed number in F2	1339	151	154	1195	2839

Starting from these data, equation (17) becomes

$$\ln L(r) = c + 1339 \ln \frac{2+r^2}{4} + 151 \ln \frac{1-r^2}{4} + 154 \ln \frac{1-r^2}{4} + 1195 \ln \frac{r^2}{4}$$

We differentiate with respect to r and then we find the MLE estimate of r solving the equation $1339r^2(1-r^2) - 305r^2(2+r^2) + 1195(2+r^2)(1-r^2) = 0$ (26). We denote $r^2 = t$ and we obtain the equation $2839t^2 + 566t - 2390 = 0$ (27). The solution biologically significant of equation (11) is $r = 0.907$. Since the data are in coupling, the recombination fraction is $r = 1 - 0.907$, i.e. 9.26%.

Equation (20) which provides the MLE estimation of recombination fraction r does not solve easily than in some particular cases. In other cases we call iterative methods such as Newton Method, Fisher's Scoring Method, EM method.

3.1 Newton's method

For the first we choose Newton's method described in recurrence relation (13) in order to solve the equation $S(r) = 0$, where S is the score function. Let $r^{(0)}$ be the initial value of r . We start with an initial MLE estimation at (k) step and we obtain an improved estimation of MLE at step $(k+1)$. Combining the relations (3), (4) and (13) the recursive relation becomes

$$\hat{r}^{(k+1)} = \hat{r}^{(k)} - \frac{S(\hat{r}^{(k)})}{I(\hat{r}^{(k)})} \tag{28}$$

Relation (28) can be written as follow

$$\hat{r}^{(k+1)} = \hat{r}^{(k)} - \frac{\left[\frac{\partial \ln l(\hat{r})}{\partial(\hat{r})} \right]_{r=\hat{r}^{(k)}}}{\left[-\frac{\partial^2 \ln l(\hat{r})}{\partial \hat{r}^2} \right]_{r=\hat{r}^{(k)}}} \tag{29}$$

The MLE estimation \hat{r} is obtained when $|\hat{r}^{(k+1)} - \hat{r}^{(k)}| < \varepsilon$, where ε is the error of estimation. In order to obtain an approximation

error equal with ε is enough to be carried out the inequality expressed by (15).

We implement the *Newton's method* algorithm in Maple application. We solve the equation $S(r) = 0$ running the procedure called *Newton* with actual parameters: $n_p, n_2, n_3, n_4, r_0, r_1, \varepsilon, m$. We set the approximation error $\varepsilon = 10^{-2}$ and we obtain $r = 0.9168$ after three iterations. So, the recombination fraction in F2 coupling obtained with Newton's method in F2 coupling is $\hat{r} = 0.08317$, i.e. 8.31%.

According to relation (8), the variance

of the MLE estimation \hat{r} is $Var(\hat{r}) = \frac{1}{I(\hat{r})}$ and $SE(\hat{r}) = \sqrt{Var(\hat{r})}$. In our example we get $Var(\hat{r}) = 0.0117$ and $SE(\hat{r}) = 0.1094$. In conclusion, in the case of Newton method, the MLE estimation of the recombination fraction r between two loci is $8.31 \pm 10.94\%$.

3.2 Fisher's method of scoring

The observed information $I(r)$ can be approximated in several ways. In many situations $J(r)$ requires less computation to evaluate than $I(r)$. Because of this we usually use in relation (28) the expected information $J(r) = E(-d^2 \ln L(r))$ instead of $I(r)$ and we

have $\hat{r}^{(k+1)} = \hat{r}^{(k)} + \frac{S(\hat{r}^{(k)})}{J(\hat{r}^{(k)})}$ (30). Using the

expected information $J(r)$ which is positive-definite (see (5)) we eliminate the possible non-convergence problems of the Newton's method. In order to solve the equation $S(r) = 0$ by *Fisher's scoring method* we run the procedure *Fisher* implemented in Maple application. This procedure has the following actual parameters $n_p, n_2, n_3, n_4, r_0, r_1, \varepsilon, m$. We set the approximation error $\varepsilon = 10^{-2}$ and we obtain $r = 0.9235$ after 20 iterations. Since the data are in coupling the recombination value is 0.0764, i.e. 7.64%.

In this situation, too we measure the deviation of the estimate from the true parameter by the standard error of the estimate. We report the

estimation in a form like $\hat{r} \pm SE(\hat{r})$. The estimation error is a function of recombination fraction r ,

with the maximum found at $r = \frac{1}{2}$ when the two loci are unlinked.

We calculate the standard error of \hat{r} starting from the information. In the considered multinomial case we calculate $I(r)$ starting from the relation (24) and we get

$$I(r) = -n \left[\frac{4}{2+r^2} \left(\frac{2r}{4} \right)^2 + \frac{8}{1-r^2} \left(\frac{-2r}{4} \right)^2 + \frac{4}{r^2} \left(\frac{2r}{4} \right)^2 \right]$$

(31). So the MLE estimation of r in F2 design based on the multinomial distribution provides the

information in the general form $I(\hat{r}) = n \frac{2(1+2r^2)}{(1-r^2)(2+r^2)}$

(32). In conclusion, in the case of Morgan experiment we obtain

$$I(\hat{r}) = 2839 \frac{2(1+0.0764^2)}{(2+0.0764^2)(1-0.0764^2)} = 2899.1935$$

. The variance is $Var(\hat{r}) = 0.000349$ and the standard error is $SE(\hat{r}) = 0.01857$.

Consequently, the Fisher's scoring method provides a MLE estimation of the recombination fraction r between two loci in the case of Morgan experiment in F2 design as $7.64\% \pm 1.85\%$.

4. Results and discussions

In this study we applied comparatively two iterative methods, Newton's method and Fisher's scoring method in order to solve the equation $S(r) = 0$ which provides the MLE estimation \hat{r} of the recombination fraction. We started from the same initial value $r^{(0)} = 0.999$ and we considered

Table 3. Newton's method iterations

Iteration k	$r^{(k)}$
0	0.9990000
1	0.9253260
2	0.9161624

Table 4. Fisher's scoring iterations

Iteration k	$r^{(k)}$	Iteration k	$r^{(k)}$	Iteration k	$r^{(k)}$	Iteration k	$r^{(k)}$
0	0.9990000	5	0.9932048	10	0.9653089	16	0.9256905
1	0.9985231	6	0.9902200	11	0.9560277	17	0.9220478
2	0.9978234	7	0.9861110	12	0.9466286	18	0.9196825
3	0.9968022	8	0.9806370	13	0.9380397	19	0.9182095
4	0.9953231	9	0.9736666	14	0.9309718	20	0.9173169

the same approximation error $\varepsilon = 10^{-2}$ in the both methods. We run these algorithms with

convergence criterion expressed as $\frac{|S(\hat{r}^{(k)})|}{m} < \varepsilon$, where $m = \min I(r)$ under considered starting range. We got the MLE estimation \hat{r} after three iterations in the case of Newton's method respectively after twenty-one iterations in the case of Fisher's scoring method. We rediscover the observation who noted that Newton Method algorithm generally converges more quickly while Fisher scoring is more robust and will converge when Newton method doesn't (Knight's, 2000).

Using the formula (15): (Micula et. al, 2008) we obtained the estimation value $\hat{r} = 0.9167949$.

Using the formula (15): (Micula et. al, 2008) we obtained the estimation value $\hat{r} = 0.9235331$.

We note that if the initial value $r^{(0)}$ is sufficiently far from \hat{r} then Newton's method is unstable. In our example, for $r^{(0)} = 0.2$ we obtain $r^{(1)} = 7.647$. So the methods oscillates and not converges at all.

Newton's method represents the basis of many modern optimization algorithms (quasi-Newton method, BFGS method) and it has a high standard in terms of speed of convergence. In many cases log-likelihood is close to the quadratic functions around their maximum points and so the approximation \hat{r} becomes a very good approximation near the maximum point.

Fisher's scoring method has the advantage that $J(r)$ is positive-definite (see (5)) and thus we eliminate the probable non-convergence problems that may arise in the case of Newton's method.

Fisher's scoring provides a quick calculation of standard deviations of the parameters estimates starting from the relation (32).

The algorithms implemented in Maple application (Newton's method and Fisher's Scoring method) run for any number n_i of individuals distributed in the i genotypic class, $i=1,2,3,4$. The algorithms can be run with different starting values. The results of the algorithms, i. e. recombination fraction approximation will be seen by executing the worksheets *Newton* respectively *Scoring*.

The method used in this paper is based on maximum likelihood estimation principle and it will be developed in order to study the linkage analysis between others characters in F2 design.

References

- Devore, J.L., Berk, K.N. (2012). Modern Mathematical Statistics with Applications. Springer. New York Heidelberg London.
- Ram B. (2010). Numerical Methods. Pearson. Delhi-Chennai-Chandigarh.
- Micula, S., Sobolu, R., Micula, M. (2008). Analiză numerică cu MAPLE, Ed. AcademicPress. Cluj-Napoca. Romania.
- Carlton, A., Devore, J.L., Berk, N.K. (2007). Modern mathematical statistics with applications. Belmont. CA: Thomson Higher Education.
- Thornley, J.H.M., J. France (2007). Mathematical Models in Agriculture. Cromwell Press. Trowbridge.
- Wu, R., C.X. Ma, G. Casella (2007). Statistical Genetics of Quantitative Traits. Springer.
- Bailey, N.T. (1961). Mathematical Theory of Genetic Linkage. The Oxford University Press. London. England.
- Zhao, W., Y.Q. Chen, G. Casella, J.M Cheverud, R.L.Wu (2005). A nonstationary model for functional mapping of longitudinal quantitative traits. *Bioinformatics*, 21, 2469-2477.
- Wu, R. L., C.X. Ma, I. Painte, Z.B. Zeng (2002a). Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. *Theoretical Population Biology*, 61, 349-363.
- Casella, G. and R. L. Berger (2001). Statistical Inference. Second Edition. Brooks-Cole.

11. Kao, C.-H. (2000). On the differences between maximum likelihood and regression interval mapping in the analysis of quantitative trait loci. *Genetics*, 156, 855–865.
12. Knight, K. (2000). *Mathematical Statistics*. Chapman & Hall/CRC. Boca Raton.
13. Elizabeth A Thompson (2000). *Statistical inference from genetic data on pedigrees*. CBMS, IMS.
14. Lange, K. (1999). *Numerical Analysis for Statisticians*. Statistics and Computing. Springer Verlag.
15. Terwilliger, J.D., Ott, J. (1999). *Handbook of human genetic linkage*. Johns Hopkins University Press.
16. Ott, J. (1999). *Analysis of human genetic linkage*. Johns Hopkins University Press.
17. Aldrich, J. (2000). Fisher's "Inverse Probability" of 1930. *International Statistical Review*, 68 (2), 155-172.
18. Aldrich, J. (1997). R.A. Fisher and the Making of Maximum Likelihood 1912-1922. *Statistical Science*, 12 (3), 162-176.
19. Lange, K., (1997). *Mathematical and statistical methods for genetic analysis*. Springer Verlag New York Berlin Heidelberg.
20. Rao, C.R. (1992). R.A. Fisher: The Founder of Modern Statistics. *Statistical Science*, 7, 34-48.
21. Basu, D. (1988). *Statistical Information and Likelihood: A Collection of Critical Essays*. *Lecture Notes in Statist*, 45. Springer, New York.
22. Haseman, J.K., Elston, R.C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics*, 1, 2-19.
23. De Vries, H. (1950). Concerning the law of segregation of hybrids. *Genetics*, 35, 35-38.
24. Fisher, R.A. (1948). A quantitative theory of genetic recombination and chiasma formation. *Biometrics*, 4, 1-9.
25. Fisher, R.A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd. Edinburgh.
26. Fisher, R.A. (1925). Theory of Statistical Estimation. *Proceeding of the Cambridge Philosophical Society*, 22, 700-725.
27. Fisher, R.A. (1922a). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A*, 222, 309-368.
28. Fisher, R.A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*, 1, 3-32.
29. Sturtevant, A. H. (1913). The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *Journal of Experimental Zoology*, 14: 43-59.
30. Fisher, R.A. (1912). On an absolute criterion for fitting frequency curves. *Messenger of Mathematics*, 41, 155-160.
31. Morgan, T.H. (1911a). The origin of nine wing mutations in *Drosophila*. *Science*, 33, 496.
32. Morgan, T.H., (1909). What are "Factors" in Mendelian Explanations? *American Breeders Association Reports*, 5:365-368.
33. Sutton, W.S. (1903). The Chromosomes in Heredity. *Biological Bulletin*, 4, 231-251.
34. Correns, C.E. (1900a). G. Mendel's law on the behavior of progeny of variable hybrids. *Berichte der Deutschen Botanischen Gesellschaft*, 8, 156-168.
35. Tschermak, E. (1900). Ueber kunstliche kreuzung bei *Pisum sativum*. *Berichte der Deutschen Botanischen Gesellschaft*, 18, 232-249.
36. De Vries, H. (1900a). Sur la loi de disjunction des hybrids. *Comptes Rendus de l'Academie des Sciences*, 130, 845-7.
37. www.netspace.org/MendelWeb

Appendix

```

> Scoring := proc(n1, n2, n3, n4, x0, x1, eps, min)
  local xn, xp, sxp, jxp, s, fd1, fd2, r, var, se, ptan, alpha, t, m, J;
  s := x → (n1 + n2 + n3 + n4) · x4 - (n1 - 2 · n2 - 2 · n3 - n4) · x2 - 2 · n4;
  J := x →  $\frac{(n1 + n2 + n3 + n4) \cdot 2 \cdot (1 + x^2)}{(2 + x^2) \cdot (1 - x^2)}$ ;
  fd1 := x → diff(s(x), x);
  fd2 := x → diff(fd1(x), x);
  xp := x0;
  xn := x1;
  if evalf(s(xp)) · evalf(eval(fd2(x), x = xp)) > 0 then
    ptan := xp;
  else
    ptan := xn;
  fi;
  xp := ptan;

  while  $\frac{\text{abs}(\text{evalf}(s(xn)))}{\text{abs}(\text{min})} \geq \text{eps}$  do
    xp := xn;
    sxp := evalf(s(xp));
    jxp := evalf(J(xp));
    xn := xp -  $\frac{\text{sxp}}{\text{jxp}}$ ;
  print(cat('xn = ', convert(xn, string)));
  print(cat('r = ', convert(r, string), ', ', convert(xp, string), ' '));
od;
r := xn +  $\frac{\text{abs}(\text{evalf}(s(xn)))}{\text{min}}$ ;
print(cat('r = ', convert(r, string)));
print(cat('recombination = ', convert(1 - r, string)));
var :=  $\frac{1}{\text{eval}(J(x), x = 1 - r)}$ ;
se := sqrt(var);
print(cat('Variance = ', convert(var, string)));
print(cat('StandardError = ', convert(se, string)));
end proc;

> Newton := proc(n1, n2, n3, n4, x0, x1, eps, min)
  local xn, xp, s, fd1, fd2, r, var, se, ptan, alpha, t, m;
  s := x → (n1 + n2 + n3 + n4) · x4 - (n1 - 2 · n2 - 2 · n3 - n4) · x2 - 2 · n4;
  fd1 := x → diff(s(x), x);
  fd2 := x → diff(fd1(x), x);
  xp := x0;
  xn := x1;
  if evalf(s(xp)) · evalf(eval(fd2(x), x = xp)) > 0 then
    ptan := xp;
  else
    ptan := xn;
  fi;
  xp := ptan;
  while  $\frac{\text{abs}(\text{evalf}(s(xn)))}{\text{abs}(\text{min})} \geq \text{eps}$  do
    xp := xn;
    print(cat('r = ', convert(xn, string), ', ', convert(xp, string), ' '));

    xn := xp -  $\frac{\text{evalf}(s(xp))}{\text{eval}(\text{eval}(fd1(x), x = xp))}$ ;
    print(cat('r = ', convert(xn, string), ', ', convert(xp, string), ' '));
  od;
  r := xn +  $\frac{\text{abs}(\text{evalf}(s(xn)))}{\text{min}}$ ;
  print(cat('r = ', convert(r, string)));
  print(cat('recombination = ', convert(1 - r, string)));
  var :=  $\frac{1}{\text{eval}(fd1(x), x = 1 - r)}$ ;
  se := sqrt(var);
  print(cat('Variance = ', convert(var, string)));
  print(cat('StandardError = ', convert(se, string)));
end proc;

```