

NEGENTROPY AND POPULATION MEAN FITNESS IN  
GENETICAL SYSTEMS

V. ARUNACHALAM

Division of Genetics, Indian Agricultural Research Institute,  
New Delhi - 110012, India.

Brillouin (1962) defined 'Information' on a system in which  $n$  outcomes are possible with the  $j$ th outcome having a prior probability  $p_j$  as

$$I = -K \sum_{j=1}^n p_j \ln p_j \text{ where } K \text{ is a constant.}$$

Every constraint and every additional condition imposed on the possible freedom of choice of the outcomes immediately results in a decrease of information. In other words, constraints represent a certain advance information  $I_c$  so that the new information  $I' = I - I_c$  is less than  $I$ . Denoting  $I = -S$ ,  $S$  can be spoken of as entropy of the system, and  $I$  as negentropy (Brillouin, 1962). The second law of thermodynamics implies that entropy  $S$  must always increase or at least remain constant.

There are situations in the population dynamics of one or two loci in which an attempt can be made to relate changes in the population mean fitness under selection to those in negentropy. However, the analogy between mean fitness and negentropy is not as close as could be desired. Mean fitness is perhaps more analogous to negative potential energy, since, as shown by Mandel (1959) and Mulholland and Smith (1959), a population equilibrium state in a one gene - many alleles system is stable if and only if the mean fitness is at a maximum. Also, the effect of selection is usually regarded as conferring greater 'order' on the population and it should, therefore, diminish any quantity truly analogous to thermodynamic or to information entropy. It is, therefore, of interest to note how entropy functions may be defined for Mendelian populations. Attempts to define such functions and to bring out their relationship with changes in mean fitness under selection in one-locus and two-loci systems are reported in this paper.

ONE LOCUS

We consider a population with one locus and alleles  $H_1, H_2, \dots, H_n$  segregating with non-zero allelic frequencies  $p_1, p_2, \dots, p_n$  so that  $\sum_{i=1}^n p_i = 1$ .

Arunachalam and Owen (1971) gave a general formulation to partition genotypic variance into its components by adopting a generalised Fisherian approach in one- and two- loci diallelic systems.

Let  $w_{ij}$  be the fitness of the genotype  $H_i H_j$ , with  $w_{ij} = w_{ji}$ . We define by  $W$  the fitness matrix,

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & \dots & w_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ w_{i1} & w_{i2} & \dots & \dots & w_{in} \\ \dots & \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & \dots & w_{nn} \end{bmatrix}$$

Let  $w_i$  be the marginal fitness of the gamete  $i$  given by

$$w_i = \sum_{j=1}^n w_{ij} p_j$$

If  $(p-)$  is the row vector,  $(p_1, p_2, \dots, p_n)$  and  $(p)$  the corresponding column vector, the population mean fitness is given by  $V = (p-) W (p)$ . Li (1969) has shown that the change in mean fitness under natural selection would be given exactly by,

$$dV = G/V + Q \text{ where}$$

$$G = 2 \sum_{i=1}^n p_i (w_i - V)^2 = \text{additive genetic variance,}$$

$$Q = (dp-) W (dp) \text{ where } (dp-) \text{ is the row vector } (dp_1, \dots, dp_n)$$

denoting changes in gene frequencies and  $(dp)$  the corresponding column vector.

In the case of a diallelic locus,  $Q$  would represent the product of dominance variance and  $(dp)^2$ , the square of the change in gene frequency of  $H$ . This analogue would also extend to the multiple allelic case. Fisher (1958) postulated (under restrictions, see Arunachalam and Owen loc. cit.) that, under slow selection in a large random mating structure, the rate of change in mean fitness  $V$  would

essentially equal the additive genetic variance. A simple argument for this in the above case would be that  $V_i$  under slow selection, is approximately equal to unity and that  $Q$  involving  $(dp)^2$  being of second order smallness is negligible. In clonal systems of reproduction, the rate of change in mean fitness equals the genotypic variance.

Let now  $P_{ij}$  be the frequency of the 'ordered' genotype,  $H_i H_j$  (where  $H_i$  is the maternal and  $H_j$  the paternal gamete). For mathematical convenience, we distinguish the genotype  $H_i H_j$  from  $H_j H_i$ . The frequency of any of these genotypes by random association of gametes under random mating is  $p_i p_j$ . Thus it is easy to see that  $p_i = \sum_{j=1}^n P_{ij}$  (I)

The information  $I$ , as stated earlier, can be defined as

$I = \sum_i \sum_j P_{ij} \ln P_{ij}$  which is not entirely suited to our context since it contains the element of disorder. A corrected function can, however, be constructed to provide the appropriate negentropy function as

$$I = -S = \sum_i \sum_j P_{ij} \ln P_{ij} - \sum_i p_i \ln p_i - \sum_j p_j \ln p_j$$

where the correction terms correspond to the negentropy of the gametes involved in the parents.

$$\text{Thus } I = \sum_i \sum_j P_{ij} \ln P_{ij} - \sum_i \sum_j P_{ij} \ln p_i - \sum_j \sum_i P_{ij} \ln p_j$$

from (I)

$$= \sum_i \sum_j P_{ij} \ln(P_{ij}/p_i p_j)$$

$$= \sum_i \sum_j P_{ij} \ln \Theta_{ij} \text{ where } \Theta_{ij} = P_{ij}/p_i p_j$$

introduced by Kimura (1957) to indicate a departure from random association of alleles. Thus  $I$  is a measure of order in the sense of association of alleles, notions of 'order' based on gene frequency distribution having been eliminated.

For small deviations of 'order',  $P_{ij} = p_i p_j + x$  where  $x$  is of first order smallness. Then  $\ln \Theta_{ij} = \ln(1+(x/p_i p_j)) = x/p_i p_j$  to first order.

It then easily follows that  $I = \sum_i \sum_j (P_{ij} - p_i p_j)^2 / p_i p_j$ . By Lagrange's method, it can be easily seen that the negentropy has a unique stationary value (which is clearly zero) attained when  $P_{ij} =$

$p_i p_j$ . Thus if  $\chi^2$  is the chi-squared value to test whether a sample of  $N$  genotypes conforms to Hardy-Weinberg expectations, then  $\chi^2/N$  is an approximate estimate of the negentropy.

The negentropy function under slow selection is of interest. Let  $w_{ij} = 1 + m_{ij}$  where  $m_{ij}$  is of first order smallness. Write  $T = \sum_i \sum_j m_{ij} p_i p_j$ . Then  $V = 1 + T$ . We shall simplify all expressions to 1st order smallness. If  $P'_{ij}$  is the frequency of the genotype  $H_i H_j$  after selection,

$$\begin{aligned} P'_{ij} &= (1 + m_{ij}) p_i p_j / (1 + T) = p_i p_j (1 + m_{ij}) (1 - T) \\ &= p_i p_j (1 + m_{ij} - T) \end{aligned}$$

Let  $m_i = \sum_{j=1}^n (m_{ij} - T) p_j$ . Then  $p'_i$ , the gene frequency of  $H_i$  after selection, is given by

$$\begin{aligned} p'_i &= \sum_{j=1}^n (1 + m_{ij}) p_i p_j / (1 + T) = \sum_{j=1}^n p_i p_j (1 + m_{ij} - T) \\ &= p_i + p_i m_i \end{aligned}$$

At an equilibrium,  $p'_i = p_i$ , giving  $m_i = 0$  as the condition for a non-degenerate equilibrium.

$$\text{Now } P'_{ij} - p_i p_j = p_i p_j (m_{ij} - m_i - m_j - T) = \delta_{ij} p_i p_j$$

where  $\delta_{ij}$  is the dominance deviation.

Thus, negentropy  $I = \sum_i \sum_j p_i p_j \delta_{ij}^2$  to first order smallness  
= dominance variance

When an equilibrium is reached,  $I$  accounts for the whole genotypic variance, since the additive genetic variance vanishes at an equilibrium given by  $m_i = 0$ . Otherwise, under slow selection we have the equation, Genotypic variance =  $\Delta V + I$

where  $\Delta V$  is the change in mean fitness per unit of time. This provides an interesting supplement to Fisher's fundamental theorem of natural selection in one-locus systems with discrete generations, random mating, large populations and slow selection.

TWO LOCI

Let  $H_1, H_2$  be the two loci with alleles  $H_1, h_1; H_2, h_2$ . Let  $P_i (i=1, 4)$  be the gametic frequencies of  $H_1H_2, h_1h_2, H_1h_2$  and  $h_1H_2$  and  $p_1, p_2$  the gene frequencies of  $H_1$  and  $H_2$  respectively ( $\sum_{i=1}^4 P_i = 1; p_1+q_1 = p_2+q_2 = 1$ ). Let  $w_{ij}$  be the fitness of the genotype  $H_iH_j$  so that  $W = ((w_{ij}))$  is the fitness matrix with  $w_{ij} = w_{ji}$  and  $w_{12} = w_{34}$ .

The negentropy in a single locus system involved  $P'_{ij}$ , the frequency of the genotype  $H_iH_j$  after selection. A simplified function, was, however, available for negentropy in the single locus case. In the two-loci system, the corresponding function for negentropy,  $I$  is given by

$$I = \sum_{i,j=1}^4 (w_{ij} p_i p_j / V) \ln (w_{ij} p_i p_j / V) \text{ where } V \text{ is the mean}$$

$$\text{fitness} = \sum_{i,j} w_{ij} P_i P_j$$

Again an interesting case is provided by slow selection. We consider therefore, only slow selection hereafter and expand expressions to first order smallness only.

$w_{ij} P_i P_j / V = (1+w_{ij}^*) (1-V^*) P_i P_j$  where the starred quantities are small and of first order.

$$\begin{aligned} &= (1+w_{ij}^* - V^*) P_i P_j \\ I &= \sum_{i,j} (1+w_{ij}^* - V^*) P_i P_j \ln(1+w_{ij}^* - V^*) \\ &\quad + 2 \sum_{i,j} P_i P_j (1+w_{ij}^* - V^*) \ln P_i \end{aligned}$$

The first term is equal to zero and the second term will reduce to  $\sum_i P_i (\ln P_i) (1+\alpha_i)$  after some simplification, where  $\alpha_i = w_i - V$  and

$$w_i = \sum_{j=1}^4 w_{ij} P_j \text{ as defined in Arunachalam and Owen (1971, p. 84)}$$

$$I = 2 \sum_i P_i (\ln P_i) (1+\alpha_i)$$

$$\Delta I / 2 = \sum_i \alpha_i \Delta P_i + \sum_i P_i (\ln P_i) \Delta (w_i - V)$$

Since  $2 \sum \alpha_i \Delta P_i = 2 \sum (w_i - V) \Delta P_i = 2 \sum w_i \Delta P_i = \Delta V$  to first order (Arunachalam and Owen, 1971, Chapter 4) and  $\sum P_i (\ln P_i) \Delta w_i = \sum w_i (\ln P_i) \Delta P_i$ ; after some algebra, we would find that

$$\Delta I = \Delta V + 2 \sum (\ln P_i) (w_i \Delta P_i - P_i \Delta V).$$

It can easily be seen that  $\Delta I$  will be equal to zero at a non-degenerate equilibrium. When the system reaches a degenerate equilibrium,  $\ln P_i$  will be undefined and hence  $\Delta I$  too.

Fitness Matrix

	$H_1H_2$	$h_1h_2$	$H_1h_2$	$h_1H_2$
$H_1H_2$	.97	1.00	.96	.96
$h_1h_2$	1.00	.97	.96	.96
$H_1h_2$	.96	.96	.98	1.00
$h_1H_2$	.96	.96	1.00	.98

Recombination Fraction = 0.04

Gen	$\Delta I \times 10$	$\Delta V \times 100$	$\Delta CF \times 10$	$P_1$	$P_2$	$P_3$	$P_4$
1	.0174	-.2520	-.7049	.201	.201	.299	.299
2	-.0177	-.0003	-.0165	.202	.202	.298	.298
3	-.0168	-.0003	-.0157	.203	.203	.297	.297
4	-.0159	-.0003	-.0149	.204	.204	.296	.296
5	-.0152	-.0003	-.0141	.205	.205	.295	.295
6	-.0144	-.0003	-.0134	.206	.206	.294	.294
7	-.0137	-.0003	-.0128	.207	.207	.293	.293
8	-.0130	-.0003	-.0122	.208	.208	.292	.292
9	-.0124	-.0003	-.0116	.209	.209	.291	.291
10	-.0118	-.0002	-.0110	.209	.209	.291	.291
50	-.0020	-.0001	-.0019	.229	.229	.271	.271
100	-.0003	-.0000	-.0003	.236	.236	.264	.264
400	0	0	0	.238	.238	.262	.262

Table 1: Changes in negentropy and mean fitness in a two locus system under slow selection ( $\Delta I = \Delta V + \Delta CF$  see text)

011212

To verify whether the derived relationship between  $\Delta I$  and  $\Delta V$  holds good, the example of slow selection given by Karlin and Feldman (1969) was considered and the selection process simulated in a computer until the system reached the non-degenerate equilibrium,  $P_1 = P_2 = 0.238$ ;  $P_3 = P_4 = 0.262$  (Table 1). It was found that the changes in  $I$  and  $V$  were very gradual and slow. But the changes in  $2 \sum (\ln P_i) (w_i \Delta P_i - P_i \Delta V)$  denoted by  $CF$  in table 1, were fast as would be expected and  $\Delta I$ ,  $\Delta V$  and  $\Delta CF$  reached the null values around generation 400. It would thus appear that the complications due to epistatic interactions and linkage would be reflected in the term  $CF$ , under natural selection.

The results clearly show that the concept of negentropy in fitness extends to two loci systems (excluding the neighbourhood of a degenerate equilibrium). How far it would extend to multi-loci systems is an open question.

#### ACKNOWLEDGEMENT

I express my sincere gratitude to Dr. A. R. G. Owen for initiating me into this problem.

#### REFERENCES

- ARUNACHALAM, V. and OWEN, A. R. G. (1971). Polymorphisms with linked loci. Chapman & Hall, London.
- BRILLOUIN, L. (1962). Science and information theory. Academic Press, New York, 2nd Edn.
- FISHER, R. A. (1958). The genetical theory of natural selection. Dover Publications, New York, 2nd Edn.
- KARLIN, S. and FELDMAN, M. W. (1969). Linkage and Selection: New equilibrium properties of the two locus symmetric viability model. Proc. Natl. Acad. Sci. U. S. 62, 70-74.
- KIMURA, M. (1958). On the change of population fitness by natural selection. Heredity 12, 145-167.
- LI, C. C. (1969). Increment of average fitness for multiple alleles. Proc. Natl. Acad. Sci. U. S. 62, 395-398.
- MANDEL, S. P. H. (1959). The stability of a multiple allelic system. Heredity 13, 289-302.
- MULHOLLAND, H. P. and SMITH, C. A. B. (1959). An inequality arising in genetical theory. Amer. Math. Monthly, 66, 673-683.