Proc. Indian natn. Sci. Acad. **B48** No. 6 pp 748–754 (1982)

A Possible Genetic Interpretation of the Auto-regulatory Mechanism in Models for Protein Deficiency

P V SUKHATME and PREM NARAIN*
Department of Biometry, Maharashtra Association for the Cultivation of Science, Pune 411004
Indian Agricultural Statistics Research Institute, New Delhi 110012

(Received 23 April 1982; after revision 10 October 1982)

The inter- and intra-individual variability with serially correlated observations, as for instance with nitrogen balance on successive days for a number of individuals, were studied theoretically with particular reference to component of variance due to interaction between the genetic constitution of the individual and the environment provided by day-to-day food intake. The intra-individual variability was enhanced as a result of the interaction and it was related characteristically with changes in the serial correlation coefficient and the length of the sequence of day-to-day observations. Based on these considerations, it is argued that the strength of the interaction can be measured in terms of the serial correlation coefficient signifying the degree of autoregulatory mechanism in nutrition studies.

Key Words: **Genetic interpretation, Auto-regulatory mechanism, Protein deficiency**

Introduction

The main emphasis in adopting a sound nutritional policy over the years has been on feeding programmes particularly for school going children based on the widespread notion that our diets lack good quality protein. So acute is the problem that as many as a third of the children are estimated to eat at levels below the recommended intake for protein and run the risk of protein deficiency. Underlying this assessment is the assumption that protein requirement remains fixed in an individual in the sense that variation from day to day is negligible relative to variation between individuals of the same age, sex and body weight. In a series of papers, Sukhatme (19) and Sukhat-

*Request for Reprint: Dr Prem Narain, Director, Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi 110012

and Margen (1978) have shown that this assumption has no support in the available data. The principle that as the intake decreases below the recommended one, the risk of deficiency increases also, has been questioned by these authors. Drawing upon the evidence of carefully controlled experiments in men on fixed intake of protein and maintenance of body weight, they have shown that N-balance in man varies from day-to-day and that the variation persists even when the data on weekly averages are considered. Further, these authors have shown that N-balance follows a stochastic stationary distribution. They have concluded that as man advances in time, he becomes a different individual. Apparently the body interacts with the environment to keep the variance constant.

What is this constant variance due to? Is it possible that the genetic constitution of man and hence some genetic components interact with the environment provided by day-to-day food intake to produce the stationary variance mentioned above? This is the basic issue which we attempt to investigate in this paper.

Intra-Individual Variability with Random Observations

In order to understand the problem, let us begin by considering the simple situation where day-to-day observations on N-balance of a given individual are random and not serially correlated. Suppose we have such data on N-balance for k subjects randomly chosen and studied over n consecutive days. If we subject this data to analysis of variance, we will have two components of variance, one between individuals which we may denote by σ^2_b and the other within individuals which we may denote by σ^2_w. In other words, we set up a model

$$y_{it}=\mu+b_i+e_{it} \qquad ...(1)$$

where y_{it} is the N-balance of the i-th individual on the t-th day, μ is the overall mean, b_i is the random effect peculiar to the i-th individual with variance σ^2_b and e_{it}'s are the effects due to the day-to-day random variations of the same individual with variance σ^2_w. Each such individual inherits from his parents certain genetic potential to cope up with the protein intake; hence we may say that the effect b_i reflects the genetic component and its variance will contain the component of genetic variance which we may denote by V_G. Also, certain environmental effects such as intra-uterine and the external environment, are permanently associated with the individual's development. Such environmental effects provide a common environmental component of variance which we may denote V_{E_g}. The component, σ^2_w is, however, due purely to local environmental effects, varying from day-to-day for a given individual, that may be denoted by V_{E_s}. We thus have the following relations between the analysis of variance and the genetic/environmental components:

$$\sigma^2_b=V_G+V_{E_g} \qquad ...(2)$$

$$\sigma^2_w=V_{E_s} \qquad ...(3)$$

An alternative interpretation of the analysis of variance components can be provided in terms of the correlation between the different values of the same individual over time. If the total variability is fixed, the higher the correlation between different values of the same individual, the lesser would be σ^2_w and the greater would be σ^2_b. Such a correlation is known as intra-class correlation and can easily be measured by the ratio of σ^2_b to the sum of the σ^2_b and σ^2_w. Thus

$$r=\frac{\sigma^2_b}{\sigma^2_b+\sigma^2_w} \qquad ...(4)$$

In terms of the genetic/environmental components,

$$r = \frac{V_G+V_{Eg}}{V_G+V_{Eg}+V_{Es}}$$

$$= \frac{V_G+V_{Eg}}{V_P} \qquad ...(5)$$

where V_P is the total phenotypic variance. r then represents the fraction of phenotypic variability due to genetic and permanent environmental causes. In genetic literature, this is the 'repeatability' of the character, and sets an upper limit to the 'heritability', the genetic parameter found to be useful in plant and animal breeding for effecting genetic improvement.

The data on N-balance, when subjected to analysis of variance, provides the two mean squares, viz. S^2_b and S^2_w with degrees of freedom as $(k-1)$ and $k(n-1)$, respectively. The expected values of these mean squares are given by

$$E(S^2_b) = \sigma^2_w + n\sigma^2_b \qquad ...(6)$$

$$E(S^2_w) = \sigma^2_w \qquad ...(7)$$

which can also be expressed in terms of the repeatability of the character as well as the variability, i.e.

$$E(S^2_b) = V_P[1+(n-1)r] \qquad ...(8)$$

$$E(S^2_w) = V_P[1-r] \qquad ...(9)$$

The variance of the mean of the individual when averaged over n different days can then be expressed in several ways, as

$$V_{P(n)} = \sigma^2_b + \frac{1}{n}\sigma^2_w$$

$$= V_G + V_{Eg} + \frac{1}{n}V_{Es}$$

$$= V_P[r + \frac{1}{n}(1-r)] \qquad ...(10)$$

This shows that as we increase the number of days for which the average is taken, the variance of the mean is decreased. In fact, if we take an infinitely large number of observations for the same individual, the variance of the mean approaches the variance component which is the sum of the genetic variance and the variance due to common environmental causes. In case the genetic causes influencing the daily N-balance interact with the intake, as one expects this to happen in a living system, the variation within individuals would not purely be environmental. An individual component of variance will appear in the within-variance component so that when the observations are averaged on several days, it does not bring about a reduction in the variance of the mean of the individuals to the extent it would do if the genetical and physiological processes of N-metabolism had been the same on each day. We can call such an additional component as the variance due to the interaction between the genetic constitution of an individual and the micro-environment provided by the food intake on different days. If we denote this component by V_{GE_s}, we will have to write the variance of mean of the individuals as

$$V^*_{P(n)} = V_G + V_{Eg} + \frac{1}{n}V_{Es} + V_{GE_s} \qquad ...(11)$$

In the limit, when n approaches infinity, the variance of the mean approaches the variance component which is now the sum of the genetic variance, the variance due to permanent environmental causes and the variance due to the interaction between genetic and environmental causes. It may be noted that we cannot now express $V^*_{P(n)}$ alternatively in terms of V_P, r and n as we did earlier.

Intra-Individual Variability with Serially Correlated Observations

Now we consider the situation where the day-to-day observations are serially correlated. We postulate a model involving serially correlated errors,

$$y_{it} = \mu + b_i + w_{it} \qquad ...(12)$$

where w_{it} is related to $w_{i(t-1)}$ by the relation

$$w_{it} = \rho w_{i(t-1)} + e^*_{it} \qquad ...(13)$$

ρ being the serial correlation of order one. In other words, we say that a part of the variability in the energy balance on a given day is explained by its value on the immediately preceding day. The remaining variability not explained by this auto-regression is attributable to the error of measurement denoted by e^*_{it}, to distinguish it from e_{it} of the uncorrelated case. The nature of the serial correlation is such that it declines with time. Thus, the correlation between observations separated by an interval m days would be ρ^m. Let us assume, for a moment, that we are considering only a given subject so that there is no inter-individual variability, all the variability being only due to intra-individual differences which we may denote by σ^{*2}_w to distinguish it from σ^2_w of the uncorrelated case.

Let the average correlation between observations of a given individual be $\bar{r}$, then the variance of the mean of n observations, in repeated sampling of the same individual, would be

$$\sigma^{*2}_w \left[\bar{r} + \frac{(1-\bar{r})}{n} \right] \qquad ...(14)$$

whereas the intra-individual variance would be

$$\sigma^{*2}_w [1-\bar{r}] \qquad ...(15)$$

Since there are $(n-m)$ pairs with interval of m days within a sequence of n days (n being at least 2), we have

$$\bar{r} = \frac{\sum_{m=1}^{n-1} (n-m)\,\rho^m}{\sum_{m=1}^{n-1} (n-m)} \qquad ...(16)$$

But

$$\sum_{m=1}^{n-1} (n-m) = n(n-1)/2 \qquad ...(17)$$

$$\sum_{m=1}^{n-1} n\rho^m = \frac{n(\rho-\rho^n)}{(1-\rho)} \qquad ...(18)$$

$$\sum_{m=1}^{n-1} m\rho^m = \frac{\rho - n\rho^n + (n-1)\rho^{n+1}}{(1-\rho)^2} \qquad ...(19)$$

Therefore $\bar{r}$ becomes

$$\bar{r} = \frac{2\rho}{(n-1)(1-\rho)} \left[1 - \frac{1-\rho^n}{n(1-\rho)} \right] \qquad ...(20)$$

The variance of the estimated mean of n days given in (14) would then be

$$\frac{\sigma^{*2}_w}{n} \left[\left(\frac{1+\rho}{1-\rho} \right) - \frac{2\rho(1-\rho^n)}{n(1-\rho)^2} \right] \qquad ...(21)$$

which is identical with the expression given by Sukhatme and Margen (1978). Thus when $\rho=0$, we have the familiar result for the variance of the estimated mean as σ^2_w/n, replacing σ^{*2}_w by σ^2_w for the uncorrelated case. Also, when we take an infinitely large sequence of days, this variance tends to zero.

Inter- and Intra-Individual Variability

Now let us introduce the inter-individual variability by considering a number of k subjects simultaneously. This variability measures the variance of the true mean of the individual based on an infinitely large sequence of days. It contains the genetic component as well as component due to permanent environmental causes and would not be affected by the presence of the serial correlation so that we can still denote it by σ^2_b as before. Since intra-individual variability for an infinitely large sequence is σ^{*2}_w, the total variance would now become

$$V_P^* = \sigma_b^2 + \sigma_w^{*2} \qquad \ldots(22)$$

When we consider finite number n of serially correlated days, the variance of the estimated mean would involve the inter-individual variability as well. Thus

$$V_{P(n)}^* = \sigma_b^2 + \frac{\sigma_w^{*2}}{n}\left[\left(\frac{1+\rho}{1-\rho}\right) - \frac{2\rho(1-\rho^n)}{n(1-\rho)^2}\right] \qquad \ldots(23)$$

In terms of the genetic/environmental components, using (2), this becomes

$$V_{P(n)}^* = V_G + V_{Eg} + \frac{\sigma_w^{*2}}{n}\left[\left(\frac{1+\rho}{1-\rho}\right) - \frac{2\rho(1-\rho^n)}{n(1-\rho)^2}\right] \qquad \ldots(24)$$

We are not expressing σ_w^{*2} in terms of the environmental component V_{Es} as we did for σ_w^2 by relation (3) since it is this variability which is affected by interaction component in the correlated case as we shall see presently. The variance of mean given by (24) would become that given by (10) when $\rho=0$. The variance of the estimated mean is thus greater than its value when $\rho=0$.

We have already seen that this enhanced variability could be accounted for by invoking an interaction component V_{GEs} between the individual's genotype and the micro-environment provided by the day-to-day food intake as in (11). But such a component cannot be measured. It may, however, result in exhibiting a serial correlation of the first order of a Markov process. We may, therefore, equate the two increases shown by (11) and (24) and study the nature of interaction component in terms of ρ. For this purpose, we express $V_{P(n)}^*$ in terms of $\bar{r}$, using (14) as

$$V_{P(n)}^* = V_G + V_{Eg} + \bar{r}\,\sigma_w^{*2} + \frac{(1-\bar{r})\,\sigma_w^{*2}}{n} \qquad \ldots(25)$$

Comparing this with the expression given in (11), we find that we may express

$$\bar{r}\,\sigma_w^{*2} = V_{GEs} \qquad \ldots(26)$$

and

$$(1-\bar{r})\,\sigma_w^{*2} = V_{Es} \qquad \ldots(27)$$

This immediately reveals the nature of intra-individual variability since it gives,

$$\sigma_w^{*2} = V_{Es} + V_{GEs} = \sigma_w^2 + V_{GEs} \qquad \ldots(28)$$

It shows that in the correlated case the intra-individual variability is enhanced by the presenee of interaction component. Further,

$$\bar{r} = \frac{V_{GEs}}{V_{Es} + V_{GEs}} \qquad \ldots(29)$$

It is interesting to observe the similarity of (29) with (5) which gives the repeatability of the character. The average correlation $\bar{r}$ between the observations of a given individual expresses the fraction of the intra-individual variability which is due to interaction between the genotype and environment. Its minimum value, i.e., zero would correspond to the absence of such interactions whereas its maximum value of unity would mean that the environmental component due to errors of measurement is zero.

If we use (20), we can write

$$V_{GEs} = \bar{r}\,\sigma_w^{*2} = \frac{2\rho}{(n-1)(1-\rho)}\left\{1 - \frac{1-\rho^n}{n(1-\rho)}\right\}\sigma_w^{*2} \qquad \ldots(30)$$

which can be helpful in determining the variability due to interaction effects from the observed data. It also shows how, for a given value of σ_w^{*2}, the variance due to interaction behaves with changes in ρ, the serial correlation coefficient and n, the length

of the sequence of day-to-day observations. When $\rho=0$, the interaction component would disappear. This would be the situation when the individual is outside the homoeostatic limits. The intra-individual variability would drop down to the level of V_{E_S}, the variation due to errors of measurement and its component in the variance of the estimated mean can be brought down to zero by averaging the daily observations over several days. If we let ρ tend to unity and take the limit of (30), we find that $\bar{r}$ tends to unity and V_{GE_S} equals σ_w^{*2}. This means the entire intra-individual variability is due to interaction component and the variance of the estimated mean cannot be reduced even if we take an infinitely large sequence of daily observations. It tends to become stationary at a value equal to total variability V_P^* given by (22). In practice, we would have ρ between *0* and *1* and the behaviour of V_{GE_S} would depend upon the length of the sequence of days considered. It is obvious that we have to consider n necessarily greater than 1. The least value that n can take then is 2. The behaviour of V_{GE_S} per unit of σ_w^{*2} i.e. of $\bar{r}$ with variation in n for specific values of ρ are shown in figure 1.

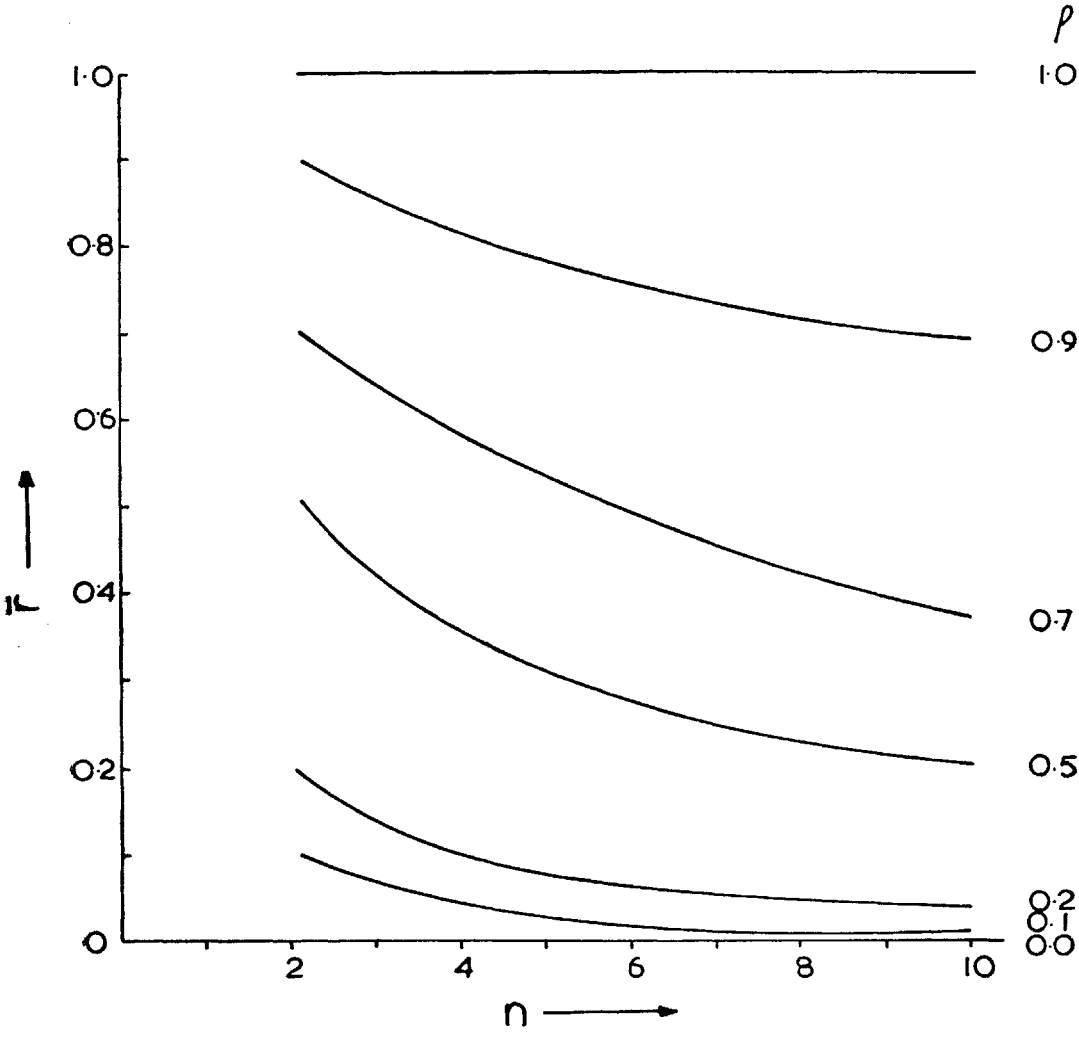

Figure 1 Relationship between auto-regulation and heritability in longitudinal studies

For $n=2,3$ and 4, however, V_{GE_S} becomes

$$V_{GE_S}=\rho\ \sigma_w^{*2},\ n=2 \qquad ...(31)$$

$$V_{GE_S}=\left(\frac{2}{3}\ \rho+\frac{1}{3}\ \rho^2\right)\sigma_w^{*2},\ n=3 \qquad ...(32)$$

$$V_{GE_S}=\left(\frac{1}{2}\ \rho+\frac{1}{3}\ \rho^2+\frac{1}{6}\ \rho^3\right)\sigma_w^{*2},\ n=4 \qquad ...(33)$$

For large n, however, we can approximate by ignoring higher order terms in n,

$$V_{GE_S}\simeq\frac{2\rho}{n(1-\rho)}\ \sigma_w^{*2} \qquad ...(34)$$

For large n, the variance of the estimated mean, however, becomes

$$V^*_{\bar{P}(n)}\simeq V_G+V_{E_g}+\frac{\sigma_w^{*2}}{n}\left(\frac{1+\rho}{1-\rho}\right) \qquad ...(35)$$

Discussion

We thus see that when we increase n, the length of the sequence of observations, the contribution of $(1-\bar{r})\ \sigma_w^{*2}=V_{E_S}$ to the variance of the estimated mean gets reduced but the contribution of $\bar{r}\ \sigma_w^{*2}=V_{GE_S}$ remains intact over and above the contribution of $\sigma_b^2=V_G+V_{E_g}$. It is this additional variability which enhances the standard error of the mean. This was also noted in the study of Sukhatme and Margen (1978). In their table 4, reproduced below, the variances of the mean of n values with unit variance when successive observations are correlated in the Markov fashion are given.

n	0.0	0.50	0.66	0.80
1	1.00	1.00	1.00	1.00
3	0.33	0.61	0.73	0.83
5	0.20	0.45	0.58	0.72
7	0.14	0.35	0.48	0.58

It can be seen from this table that when the serial correlation of the first order is having the value around 0.6 to 0.7, the variance of the estimated mean, even when based on 7 days, (and losing in the process six pieces of information) will be three times as much as when based on assumptions of independence between successive observations. This means that even when the average refers to 7 days, the standard error will be reduced only in the ratio of 1:0.7 and is therefore, not negligible.

As we have seen, additional variability can be interpreted in two ways. Firstly, it is the variation explained by the auto-regressive nature of the Markov process of the first order. Secondly, it is the variation due to interaction of the individual's genotype as he advances in time. The strength of this interaction can therefore, be measured in terms of the serial correlation coefficient signifying the degree of auto-regulatory mechanism in nutrition studies. We are thus able to explain the cause of the stationarity of variance in terms of a genetic parameter (heritability) and relate the genetic component due to interaction with the auto-regulatory mechanism.

References

Sukhatme P V (1974) The protein problem, its size and nature; *J. Roy. Statist. Soc.* **137**

Sukhatme P V and Margen S (1978) Models for protein deficiency; *Amer. J. Clin. Nutrition.* **31**