*Jour. Ind. Soc. Ag. Statistics*
*51(2&3), 1998 : 129-134*

# Genetic Relationship Underlying the Intra-Individual Variation in Nutrition Studies

P. Narain
*Indian Agricultural Research Institute, New Delhi*

SUMMARY

The role of intra-individual variation in nutrition studies was highlighted by Prof. P.V. Sukhatme both theoretically as well as by analysis of actual live data on intake and expenditure collected over time in a number of individuals. The genetic significance of this intra-individual variation has been discussed in this paper drawing upon Professor Sukhatme's contributions.

*Keywords*: Autoregressive Markov process model, Homeostatic limits, Variability, Intra-individual repeatability, Serial correlation coefficient, Auto-regulatory mechanism.

## *1. Introduction*

Prof. P.V. Sukhatme was a multi-facet personality who made everlasting contributions at the national and international levels in the fields of theoretical statistics as well as statistics applied to agriculture, biology and nutrition. The most important outcome of his deliberations in the field of nutrition is the demonstration how the science of statistics can be of great assistance in understanding the nutritional concepts and in helping thereby to tackle serious and important issues like malnutrition, undernutrition and health. A process view of nutrition, emphasising the role of intra-individual variation, was very dear to him which he advocated in several problems connected with health and nutrition. I pay my humble tribute to his memory by discussing the genetic significance of this variation in nutrition studies.

## *2. Auto-Regressive (AR) Model for Protein Deficiency*

Sukhatme and Margen [3] developed the concept of protein requirement of individuals and indicated the method by which it can be extended to those of populations. According to joint FAO/WHO Ad-hoc Expert Committee on Energy and Protein Requirements, the safe level of protein intake is defined as the average requirement plus twice the standard deviation. According to them, an individual eating below this level, though not malnourished, runs the risk

of developing protein deficiency and the risk increases as the intake falls below the safe level.

Several models based on bivariate normal distribution of intake and requirements were postulated for calculating the proportion of population below the safe level. All along it was assumed that the requirements remain constant in an individual. Sukhatme's approach was to take into account the intra-individual variability in requirement, not as a random noise due to measurement but in a manner represented by an auto-regressive (AR) stochastic process.

When we have time series data on daily N-balance in man maintaining body weight on fixed intake and on the assumption that energy intake is not a limiting factor in the diet, we can represent the series as

$$w_t = \rho w_{t-1} + e_t \tag{1}$$

where $w_t$ is the balance on the t-th day, $\rho$ is the serial correlation coefficient (of order one) between $w_t$ and $w_{t-1}$ and $e_t$ is a random variable with mean zero and variance $\sigma^2$. This autoregressive Markov process model consists of two components—one a short-term component arising from the current value of the process at the previous time point and the other a long-term component in the form of errors of measurement. In such a process, the errors get incorporated into the motion of the process to determine the balance on any given day and are not cancelled out as they would do in a purely random process with $\rho = 0$. Since

$$E(w_t) = 0 \tag{2}$$

$$V(w_t) = \sigma_e^2 / (1 - \rho^2), \text{ independent of t}$$

the observed value of balance on any given day is distributed around zero mean within limits $\pm 2\sigma_e / \sqrt{1-\rho^2}$ which are known as homeostatic limits.

The nature and degree of intra-individual variation were studied by Sukhatme and Margen [3] by analysing the data relating to daily N-balance on fixed intake reported by Calloway and Margen [1] as well as the data on experiments conducted at the Department of Nutritional Sciences, University of California, Berkeley, USA. It was found that for intakes in the range of 3.5 to 12 gms N/day, the day to day fluctuations in N-balance were not random but were serially correlated in an auto-regressive process. This implied that the daily N-balance, like energy balance, is regulated according to a probabilistic generating mechanism which remains constant through time. At very high or negligible N-intake, this regulation is shown to break down i.e. homeostasis can no longer be maintained and the organism is under stress. The magnitude

of stationary variance was found to be comparable with the variation between individuals. The result was found to hold even when the daily requirement was averaged over several days. Sukhatme and Margen [3] concluded that protein deficiency must be defined as the failure of the process to be in statistical control and not defined in the manner that assumes requirements to be fixed whereby if an individual consumes protein below this level, he suffers from protein deficiency.

### *3. Genetic Interpretation of Intra-Individual Variation*

Sukhatme and Narain ([4], [5], [6]) showed that the intra-individual variability in calorie or protein intake is enhanced due to interaction between the genotype of the individual and the environment as he advances in time. Theoretically, the situation is best illustrated by invoking a model of one-way ANOVA with correlated errors.

Suppose we have data on energy balance or protein intake for k-subjects recorded at n successive days. The model for the response of the i-th subject on the t-th day is given by

$$y_{it} = \mu + b_i + w_{it} \quad i = 1, 2, \ldots, k; t = 1, 2, \ldots, n \tag{3}$$

where $\mu$ is the overall mean, $b_i$ is the effect of ith subject with expectation zero and variance $\sigma_b^2$ and $w_{it}$'s, for the same i-th subject are n consecutive random variables with variance $\sigma_w^2$ following AR process given by

$$w_{it} = \rho\, w_{i(t-1)} + e_{it} \tag{4}$$

where $\rho$ is the serial correlation coefficient of order one and $e_{it}$'s, are independently distributed with mean value zero and variance $\sigma_e^2$. It may be noted that total phenotypic variance of y, denoted by $\sigma_p^2$, has two components $\sigma_b^2$ and $\sigma_w^2$ with

$$\sigma_b^2 = R\,\sigma_p^2 \qquad \text{and} \qquad \sigma_w^2 = (1 - R)\,\sigma_p^2 \tag{5}$$

where R is intra-class correlation, which is usually termed as repeatability in genetic literature. The intra-individual component $\sigma_w^2$ is further divisible into two components -one due to serial correlation between successive observations based on a Markov stationary process of order one, which we may denote by $\sigma_d^2$ and the other due to the remainder error term denoted by $\sigma_e^2$. The ratio of $\sigma_d^2$ and $\sigma_w^2$ is what we denote, in the sequel, by $\bar{r}$, giving

$$\sigma_d^2 = \bar{r}\,\sigma_w^2 \text{ and } \sigma_e^2 = (1-\bar{r})\,\sigma_w^2 \tag{6}$$

The expected value of the between subjects mean square ($S_b^2$) is then

$$E(S_b^2) = [n-(n-1)\,\theta]\,\sigma_w^2 + n\,\sigma_b^2 \tag{7}$$

where

$$\theta = \left[1-2\left(\frac{n+1}{n}\right)\rho+\left(\frac{n+1}{n-1}\right)\rho^2-\frac{2\rho\,(n+1)}{n\,(n-1)}\right](1-\rho)^{-2} \tag{8}$$

The variance of the mean of the individual when averaged over n different days is then

$$\begin{aligned} V_{P(n)} &= E(S_b^2)/n \\ &= \sigma_b^2+\left[\frac{1+(n-1)\,\bar{r}}{n}\right]\sigma_w^2 \\ &= R\,\sigma_p^2+\left[\frac{1+(n-1)\,\bar{r}}{n}\right](1-R)\,\sigma_p^2 \\ &= \sigma_p^2\left[R+\bar{r}\,(1-R)+\frac{(1-R)\,(1-\bar{r})}{n}\right] \end{aligned} \tag{9}$$

The $\bar{r}$, the average correlation between observations of a given individual is related to $\rho$, the serial correlation coefficient as

$$\begin{aligned} \bar{r} &= \frac{2\rho}{(n-1)\,(1-\rho)}\left[1-\frac{(1-\rho^n)}{n\,(1-\rho)}\right] \\ &\approx 2\rho/n\,(1-\rho) \end{aligned} \tag{10}$$

The effect $b_i$ in the model, pertaining to the i-th subject, reflect genetic effect of the individual as well as certain environmental effects permanently associated with the individual's development such as intra-uterine and external environment experienced by him. Its variance would therefore contain the genetic component of variance ($V_G$) as well as common environmental component of variance ($V_{E_g}$) i.e.

$$\sigma_b^2 = V_G + V_{E_g} \tag{11}$$

On the other hand, intra-individual component $\sigma_w^2$ would reflect only the variability due to local environmental effects ($V_{E_s}$) provided the genotype does not interact with the environment. If it is not so, another component of variance

due to the interaction ($V_{GE_s}$) would enter into the component so that when the observations are averaged for several days, it does not bring about the reduction in variance of the mean of the individual to the extent it would do if the genetico-physiological process of calorie or protein metabolism had been the same on each day.

We, therefore, get

$$\bar{r}\,\sigma_w^2 = V_{GE_s}$$

$$(1-\bar{r})\,\sigma_w^2 = V_{E_s} \tag{12}$$

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

$$\sigma_w^2 = V_{E_s} + V_{GE_s} = \sigma_e^2 + V_{GE_s} \tag{13}$$

showing that the intra-individual variability is enhanced by the presence of genotype × environment interaction. We then get

$$\bar{r} = V_{GE_s}\left[V_{E_s} + V_{GE_s}\right]^{-1}$$

Then

$$V_{P(n)} = \sigma_b^2 + \bar{r}\,\rho_w^2 + \frac{(1-\bar{r})}{n}\sigma_w^2$$

$$= V_G + V_{E_g} + V_{GE_s} + V_{E_s}/n \tag{15}$$

The average correlation $\bar{r}$ can then be given a genetic interpretation as heritability of the individual or *intra-individual heritability* in a manner similar to the concept of heritability used in quantitative genetics. It is the fraction of the total intra-individual variability which is due to interaction between the genotype and the environment and could take any value between 0 and 1. The existence of the genotype × environment interaction thus enhances the intra-individual variability with stabilisation of variance as we increase the period of time over which the data are collected. The strength of this interaction can be measured in terms of the serial correlation coefficient signifying the degree of auto-regulatory mechanism.

### *4. Effect of Genetic Relationship on the Intra-Individual Heritability*

The above considerations can be extended to determine the covariance between relatives so as to reveal how genetic relationship controls the auto-regulatory mechanism measured in terms of $\rho$ (Narain [2]). It can be shown

that compared to the average correlation in the given generation $\bar{r}$, the average correlation ($\bar{r}^*$) in the relatives with a genetic relationship $r_g$ takes the form

$$\bar{r}^* = \left(r_g V_{GE_s}\right)\left(V_{E_s} + r_g V_{GE_s}\right)^{-1} \tag{16}$$

We can then relate the two average correlation coefficients as

$$\bar{r}^* = (r_g \bar{r})\,[1 - (1 - r_g)\bar{r}\,]^{-1} \tag{17}$$

where $r_g$ is 1 for identical twins, (1/2) for parent-offspring or full-sibs and (1/4) for half-sibs.

Thus as $\bar{r}$ increases from 0 to 1, $\bar{r}^*$ also increases characteristically from 0 to 1 depending on the value of $r_g$. At a given value of $\bar{r}$, the value $\bar{r}^*$ is highest i.e. equal to $\bar{r}$ for identical twins and lowest for half-sibs.

## ACKNOWLEDGEMENTS

* The author gratefully acknowledges the financial support received from CSIR grants under their Emeritus Scientist Scheme for undertaking this work. He is also grateful to Dr. R.B. Singh, Director, I.A.R.I., New Delhi for providing the necessary facilities.

## REFERENCES

[1] Calloway, D.S. and Margen, S., 1971. Variation in endogenous nitrogen excretion and dietary nitrogen utilisation as determinants of human protein requirements. *J. Nutrition*, **101**, 205.

[2] Narain, P., 1993. Interface among statistics, cybernetics and genetics. Dr. Rajendra Prased Memorial Lecture. 46th Annual Conference of ISAS, Bhubaneswar, *J. Ind. Soc. Agril. Statis.*, **45**, 48-75.

[3] Sukhatme, P.V. and Margen, S., 1978. Models for protein deficiency. *Amer. J. Clinical Nutrition*, **31**, 1237-1256.

[4] Sukhatme, P.V. and Narain, P., 1982a. The genetic significance of intra-individual variation in energy requirement. In: Rao, P.S.R.S. and Sedransk, J. (eds.) W.G. Cochran's Impact on Statistics. pp 275-284. John Wiley and Sons, New York.

[5] Sukhatme, P.V. and Narain, P., 1982b. A possible genetic interpretation of the autoregulatory mechanism in models for protein deficiency. *Proceedings of Indian National Science Academy*, **B 48**, 748-754.

[6] Sukhatme, P.V. and Narain, P., 1983. Intra-individual variation in energy requirement and its implications. *Indian J. Med. Res.*, **78**, 857-865.