

Interface Among Statistics, Cybernetics and Genetics

Prem Narain [□]

I consider it a great honour and privilege to be chosen by the Indian Society of Agricultural Statistics this year to deliver the prestigious Dr. Rajendra Prasad Memorial Lecture — a lecture in the memory of the Founder President of the Society. I express my deep sense of gratitude to the Society for enabling me to share some of my thoughts with you on this solemn occasion.

The name of Dr. Rajendra Prasad is enshrined in the history of India as the one who played a pivotal role in the freedom struggle of our motherland, as one of the founding fathers of our Republic and as its first President. During my student days in Lucknow I read about him that he had a brilliant academic career and adopted a very simple life-style but with high thinking. In 1955 I came to Delhi and took up a statistical job in Government of India. I barely missed an opportunity of seeing him in person along with a group of my relations who had come from Bihar to meet him. Later on, I learned particularly from Dr. P.V. Sukhatme how, as the then Minister of Food and Agriculture, Government of India, he took keen interest in the problems of statistics in relation to agriculture and was instrumental in founding and nurturing this Society.

Rajendra Babu's concern about statistics and its profession in relation to agriculture can be best understood by reading his views in the Presidential Addresses which he used to deliver during the Annual Meetings of the Society and which are published in the issues of the Journal of the Society from time to time. For instance, in his Address delivered during the First Annual Meeting of the

-
- Dr. Rajendra Prasad memorial lecture delivered on 20 February, 1993 during the 46th Annual Conference of the Society at OUAT, Bhubaneswar.
 - Principal Scientist, IARI, Pusa, New Delhi and
Former Director, IASRI, New Delhi - 110 012

No. of Strata	<i>First Stage</i>	2000 approx.	<i>First Stage</i>	540 (approx.)	<i>First Stage</i>	1000-1200	<i>First Stage</i>	1000-1200
	<i>Second Stage</i>	None	<i>Second Stage</i>	4-5	<i>Second Stage</i>	No change	<i>Second Stage</i>	No change
Sample Size	<i>First Stage Units</i>	5000 approx.	<i>First Stage Units</i>	9,500	<i>First Stage Units</i>	No change	<i>First Stage Units</i>	No change
	<i>Second Stage Units</i>	4 per FSU	<i>Second Stage Units</i>	20-25 per FSU	<i>Second Stage Units</i>	No change	<i>Second Stage Units</i>	18-20 per FSU
Sample Allocation	Total no. of FSUs allotted to states on joint consideration of crop area and value of production. Within state, FSUs allotted to districts in proportion to gross cropped area and to tehsils in proportion to no. of villages with the restriction that each stratum has at least 2 FSUs.		FSUs in a state allotted to strata in proportion to the stratum population with the restriction that the no. of FSUs in a stratum is 4 or a multiple of 4.		No change upto district level. Allocation to strata within district in proportion to population. No. of FSUs per stratum at least 2.		No change except that no. of FSUs per stratum is 2 or a multiple of 2.	
Sample Selection	<i>First Stage Units</i>	Simple random sampling without replacement	<i>First Stage Units</i>	p.p.s. to population with replacement	<i>First Stage Units</i>	p.p.s. to population with replacement (as in NSS-SE)	<i>First Stage</i>	No change
	<i>Second Stage Units</i>	Simple random sampling without replacement	<i>Second Stage Units</i>	Circular systematic sampling with random start	<i>Second Stage Units</i>	No change	<i>Second Stage</i>	No change

Society held at New Delhi on the 11th December, 1947, he stressed the need of statistical knowledge and proper methods to acquire them for the purpose of agricultural planning. In this regard, he was not content with *any* method but with a method which is objective and based on personal knowledge. For instance, in regard to the method of aerial photography for determining crop production he had this to say :

“Another method mentioned by your Vice President, Mr. M.S. Randhawa, is the method of air photography. That may also be tried, but I do not know whether it will be feasible. I have no idea of its cost and of the techniques required for air photography of crops. While this method may give us information of the area under crops, assuming that different crops can be identified from the photographs, we shall still need information on yield, which, I do not think, this method can supply. As an example, last year we had a severe epidemic of rust on wheat in the country. Photographs of the field if taken would have given the impression that the crop was a bumper crop. What happened was that the crop *looked* to be quite prosperous. If you saw the ears you would imagine that a very rich crop was standing in the field. But when the ears were actually examined, not a single grain was found inside; it was all chaff. I am mentioning this case as an example of the dangers to be avoided in developing such methods”.

In spite of notable advances in statistical science and technology during the intervening period of four and a half decades, however, what Dr. Rajendra Prasad said about the objectivity of the statistical method then is still true and of vital importance to all of us. There is no substitute to the objective method of observation. The reason why I am stressing this point is that during the period of 30 years which I spent at the Indian Agricultural Statistics Research Institute (IASRI) at New Delhi, of which 10 years as its Director, I participated in numerous discussions both in India and abroad on this issue of objectivity in the conduct of Sample Surveys and witnessed a gradual decline in the attitude towards it in statistical projects and programmes. I feel therefore that there is a need to create a scientific temper in our thinking process on statistical issues and problems.

The topic of my lecture on the interface among Statistics, Cybernetics and Genetics is primarily aimed at illustrating the basic

universality of the interconnection among all things as against the alienated world view that objects are isolated until proven otherwise.

We live in a world where everything is related to everything else in an interacting manner. Only the interactions are often so feeble that we tend to neglect them and view things as isolated. In particular, it is the ignoring of the interconnections across scientific disciplinary boundaries in a complex field like agriculture that we are not often able to transfer laboratory findings to the farmers' field conditions. The usual *reductionist-mechanistic* approach in agricultural research which we have borrowed from basic sciences—the approach of understanding a complex system by reducing it into isolated parts—has led us to recommend to the farmers agronomic practices which they are not sure to take advantage of in their fields. A *dialectical* perspective of agricultural research seems to be essential for sustainable agricultural development particularly in a country like ours where small and marginal farmers are averse to taking risk due to poor resources available to them (Narain, 1992 a, b). In this lecture, however, I will limit myself to demonstrate how the practical issues in such diverse and seemingly independent disciplines like statistics, cybernetics and genetics are interconnected and can benefit immensely if we trace out the indirect or sometimes speculative connections among them and justify their interactions.

STATISTICS

The subject of statistics as we know deals with random phenomenon and uncertainty. Its foundations are based on mathematics and probability and has developed into a body of knowledge which is known as theoretical statistics. As a science, however, it deals with measuring and minimising uncertainty surrounding the data generated in a given discipline like biology, agriculture, engineering etc. It involves collection, analysis and interpretation of data. In modern times we use the term data analysis for this purpose. It helps in understanding the complexity of the problem, in predicting the future and in planning of the investigation. The driving force behind the scientific aspects of statistics is in fact the need to solve practical problems by drawing valid inferences from data. This is amply demonstrated in this country by the development of research and training in the specialisation *agricultural statistics* due to the efforts of this Society and the IASRI, New Delhi, under the leadership of eminent

statisticians like P.V. Sukhatme, V.G. Panse, G.R. Seth, D. Singh and several others.

CYBERNETICS

The foundation of cybernetics was laid in 1948 by the American mathematician Norbert Wiener with the publication of his work *Cybernetics*. He coined this term from the Greek word *Kybernetes* which means steers-man. Originally Wiener defined it as the *science of control and communication in animals and machines*. But subsequently it found applications in several other fields involving optimal control over complicated processes and operations which take place either in the living organism or in the human society or in the industry. It therefore cuts across several disciplines like biology, physiology, genetics, psychology, neurology, systems analysis, sociology, economics, engineering etc. In a way it is the science of sciences. In general, cybernetics is concerned with developing a comprehensive view of complexities, uncertainties, and controlled mechanism for a related dynamic system by formulating and handling feed-back and feed-forward mechanisms.

GENETICS

The fascinating biological science of genetics tells us the manner in which the characteristics are inherited from one generation to another. The *genotype* of an individual, the sum total of hereditary material received by him/her from his/her parents, interacts with the environment to produce its *phenotype*, the outward appearance of the individual, being the sum total of its expressed peculiarities of form, size, colour, behaviour etc. Probably the greatest of all the scientific accomplishments of the twentieth century has been the discovery of this science. No other scientific inquiry has made so startling an impact on all aspects of life, including man, as the science of genetics. With the rediscovery of Mendel's principles of inheritance in 1900 to the recent splicing of genes, this science has witnessed a series of breath-taking discoveries. With the recent advent of *recombinant DNA technology*, the destiny of mankind now seems to have come within the grasp of man himself.

STATISTICS AND CYBERNETICS

The statistical science is primarily inductive – we go from the particular to the general, the process of inference. This introduces a certain amount of arbitrariness which the statistical procedure attempts to minimise. Mathematics helps us in developing a deductive logic for this purpose. The development of this science since 1920s and 1930s has led to two broad approaches on the fundamental rules of statistical inference – the *Bayesian subjectivist* and the *sampling theory frequentist*. It is unfortunate that the protagonists of the two approaches have been, for ages, fiercely critical of each others' viewpoint. There have been several attempts by eminent statisticians to reconcile the two lines of thought and give a unified theory but in vain. In the context of applied statistical problem-solving, as in agricultural statistics discipline, such controversies simply do not matter. What matters to us, the agricultural statisticians, are issues like what data should be collected and subjected to analysis, what variables to be included, how finely the observations should be stratified, what could be concluded from the data and how far the conclusions could be trusted etc. In fact, more often we are concerned with a data-analytic focus rather than with the optimality of the rules of statistical inference. I do not therefore propose to go into the latter but concentrate on the former and wish to state that by adopting cybernetic concepts in data analysis, a good deal of subjectivity can be dispensed with.

In the initial stages of my career, soon after I took my M.Sc. in Mathematical Statistics from Lucknow University, I was employed in the Indian Council of Agricultural Research (ICAR) scheme on 'Cost of cultivation of Sugarcane in U.P.' I was quite enthusiastic, along with my colleagues in the scheme, to apply my university knowledge to the problem at hand as per technical advice of Late Dr. V.G. Panse and Dr. K. Kishen. There was no occasion to question any of the statistical methods used. In the later years, I joined Indian Veterinary Research Institute (IVRI) as Assistant Research Officer (Quantitative Inheritance), again in an ICAR scheme dealing with a problem on genetics. I then came in contact with animal experimentation and had to interact with numerous scientists in the interpretation of data. I felt often concerned with the replicability of the experiments. When a null hypothesis is rejected at 5 per cent level of significance on the basis of one experiment, does the same thing happen when we repeat the experiment? On the other hand, the experimenter wishes to arrive at the same result and therefore

he repeats it several times, as for instance did Mendel in experimentation with garden peas. The probabilities that we work with are not *natural* probabilities fixed by the experiment under study. By changing the sampling design or the statistical model set up we can change the probability structure for the same problem. This is what leads to subjectivity in scientific inference and may take one away from the truth which he is seeking for. How do we circumvent it? For this purpose, let us look at the characteristics of cybernetics.

Like applied statistics, cybernetics is also concerned with practical problem-solving in a given system. However, the classical method of statistical data analysis allows one to study only individual elements of a system at a given time with no scope to extend the methodology to an interacting system with several elements with feed-back and feed-forward relationships. In other words, there is a lack of *systems approach* in the classical statistical method. This is clearly a disadvantage. In fact when a shift from reductionist to systematic or dialectical thinking in a systems approach takes place, it requires us to change our way of looking at the concept of relationship between elements. When we adopt the machine world view, the relationship is understood in terms of interactions between two and only two elements. A complex system has therefore necessarily to be broken down and looked at in terms of pairs. In the systems approach, on the other hand, multifarious interactions between all the elements have to be recognised as such. The system is studied in a holistic manner. The most important characteristic of such a system is the property of *synergy*. Because of the synergistic property, the whole is *more* than the sum of its parts.

The function of cybernetics, as already stated, is to develop a comprehensive view of the complexities, uncertainties and control mechanism in the related dynamic system which is not the case with statistics where mostly we are concerned with a static situation. By interfacing therefore with cybernetic concepts and adopting a systems approach, the statistical problem-solving can be made more realistic and suited to the needs of an experimental scientist. At the same time, in the estimation of complex relation structures with uncertainties, statistical methods can be fruitfully made use of.

Relevant to the problem concerned, the cybernetic approach is first to build an information system regarding elements, sub-systems, their links, feed-back and feed-forward relationships and possible goals regarding forecasting, self-organisation and

prescription. However, processing of information through a designed data base, identification of relationship, inference over specific periods of time etc. have to be done using statistical tools. The essential feature of cybernetic modelling is that it admits multi-models and multi-solutions. This feature resembles the approach of *Exploratory Data Analysis* (EDA) introduced by Tukey (1977) for practical problem-solving in statistics. This is so because it advocates, for a given set of data, the exploration of several models, their solutions and inferences without any assumptions regarding the distribution of characters in the hypothetical population from which a sample of a given size has been drawn. The subsequent confirmatory data analysis gives a clue as to whether the exploration has been done correctly. If not, it gives a feed-back to correct the model assumption and then proceed again. There is a continual forward-and-backward interaction between cybernetic modelling and statistical processing of data. It is this interface or interaction which enables one to understand a complex system (Ghosal 1978, 1986).

Cybernetic Regression Analysis

In regression analysis, cybernetic concepts of feed-back mechanism can be introduced by relating the input (x_t) and output (y_t) at time t in successive steps. First y_t is regressed on x_t through a function $f(x_t)$. The y_t will give an impulse to the input, giving another $x_{(t+1)}$ from the regression of the type $x_{(t+1)} = g(y_t) + e_{x_{(t+1)}}$ which will give the expected $y_{(t+1)}$ as $f[x_{(t+1)}]$ and so on. Thus (x_t, y_t) interact in a closed loop. A more realistic modelling would be to introduce non-linear transfer function models as in Box and Jenkins (1970).

Non-parametric Statistical Data Modelling

Parzen (1979) has recently developed a unified theory of statistical science in which EDA, parametric statistical inference, robust statistical inference, non-parametric statistical inference, goodness-of-fit procedures etc. are used simultaneously. He introduced new types of keys for EDA based on estimating the *quantile function* $Q(u)$ and *density-quantile function* ($f[Q(u)]$) of a random variable X of which the data X_1, \dots, X_n are independent or dependent observations.

In order to reveal statistical regularities or patterns in the data at hand, it is customary to model the distribution function

$F(x) = P_r[X \leq x]$ and the probability density function $f(x) = F'(x)$. The quantile function, $Q(u)$, is defined as the inverse of $F(x)$ i.e.

$$Q(u) = \inf. [x : F(x) \geq u]$$

In the given sample, the empirical quantile function (EQF) is a piecewise constant function whose values are the order statistics.

$$X_{(1)} < X_{(2)} < \dots < X_{(n)}$$

The density-quantile function, $f[Q(u)]$, is defined as the reciprocal of quantile-density function $q(u) = Q'(u)$.

To estimate Q , H_0 is specified as

$$Q(u) = \mu + \sigma Q_0(u) \text{ with } Q_0(u) \text{ as known.}$$

By estimating Q and q from the sample and examining how well various theoretical Q -functions match or fit the sample values, one can explore the patterns in the data. The density-quantile function $f[Q(u)]$ is then estimated on a raw basis either from the plotting of histogram-quantile function or from the plotting of a density-quantile function depending respectively on whether the data are reported as a histogram (density) or as a sample quantile function \hat{Q} . The function fQ is smoothed by density estimation procedures. One then fits and estimates Q by a parametric smoothing approach as

$$\hat{Q}(u) = \hat{\mu} + \hat{\sigma} Q_0(u)$$

The parameters μ and σ are estimated by the regression analysis of the sample quantile function $\hat{Q}(u)$, $0 \leq u \leq 1$.

Essentially the above procedure is a *density estimation* approach to goodness-of-fit tests which also yields estimates of Q . The statistical estimation and testing problems can therefore be viewed as problems of density estimation and testing of white noise (i.e. the density is constant). Such an approach provides a firm probability basis for the EDA. It can be seen that the approach of Parzen which is both firstly non-parametric and secondly parametric with a feed-back between the two with the following steps is essentially cybernetic in character:

- i. Explore ;
- ii. Try models and transformations we know ;

- iii. Explore again ;
- iv. Build a model from the data ;
- v. Explore once again.

I was motivated to study the methodology of non-parametric statistical data modelling during the course of an investigation (Narain *et al.*, 1985; Garg *et al.*, 1991) on the statistical aspects of Comprehensive Crop Insurance Scheme of the Government of India, which is in operation since 1985 and is based on *homogeneous area approach* notified for each crop in each state by the Union Ministry of Agriculture through General Insurance Corporation of India. The average yield per hectare of the insured crop in the notified area is determined on the basis of an adequate number of crop-cutting experiments. The indemnity is calculated on the basis of the shortfall in the actual yield from the threshold yield per hectare of the insured crop which is usually fixed at 80 per cent of the average normal yield of the crop over the last 5 to 10 years based on the crop-cutting data for the defined area. The estimation of actuarial premium rate, as the expected value of indemnity, depends on the distribution of crop yield over the years. When this is assumed to be normal, premia are found to be proportional to both, the coefficient of variation of the crop yield as well as the level of coverage. However, in actual practice, the data do not confirm to normal distribution. We can then invoke non-parametric density estimation procedures as discussed above to determine premium rates. In particular, Mittal (1989) showed that Kernel method of density estimation provides better estimates of the rates as compared to those on the basis of normal or log-normal distribution.

Prequential Forecasting System

In statistical science we always regard the past as a guide for the future. The pattern of law which we derive from historical data, usually of large size, is taken as valid for the future. Such an attitude can only serve our purpose if the world is static. But in a dynamic world it is not necessarily so—things change through time with unpredictable directions of change. There may be changes in the estimates of parameters even if the model is the same or else there may be an entirely changed model. In addition, future may be guided by the nature of human intervention in the form of a policy or major shift in the process. It is therefore necessary to introduce a *sequential* approach in the task of forecasting the future. Today we can make the forecast of tomorrow. Come tomorrow, we can observe

the outcomes of the events which were the subject of today's forecast and with this additional experience to draw on, formulate our new forecast for the following day. And so on, day after day, drawing on the accumulating experience. Dawid (1984) has called such an approach as *prequential forecasting system* which, in essence, combines probability forecasting with sequential prediction.

Given any statistical model depending on the unknown parameter θ , we can form, at time n , its estimate $\hat{\theta}_n$, based on the data $X^{(n)} = (X_1, X_2 \dots X_n)$ and use it as a true value (known as plug-in estimate) in the predictive distribution of the next observation X_{n+1} . This amounts to changing the estimate of the parameter at each stage of forecasting, taking into account the accumulated experience. For instance, let $X_i = 1$ with probability θ and $X_i = 0$ with probability $(1 - \theta)$, independently. After observing r 1's on n trials, the maximum likelihood plug-in forecast probability for $X_{n+1} = 1$ is r/n . The forecast probability for 3-steps ahead such as $(X_{n+1} = 1, X_{n+2} = 0, X_{n+3} = 1)$ would be $(r/n) [\frac{n-r}{n+1}] [\frac{r+1}{n+2}]$ instead of the usual result based on static world concept as $(r/n) [\frac{n-r}{n}] (r/n)$.

The prequential forecasting, however, can be done in several ways. In the place of plug-in approach described above, we can have a Bayesian or a Fiducial predictive distribution. We then have the problem of comparing these different ways and choosing the best amongst them.

Apart from forecasting for the future, there is a need for taking into account any intervention which might occur due to policy change or other forces. Bhattacharya (1982) developed a *bivariate feed-back intervention model*. The sample period $(0, T)$ is split into two parts; pre-intervention $(0, T_1)$ and (T_1, T) where intervention occurred at time T_1 . In the former case, there are two time series, Y and X whereas in the post-intervention period (T_1, T) there are three time series, Y , X and Z where (Y_t, X_t) is a bivariate time series on which observations are available at equidistant points and Z is an intervention variable affecting variable Y_t only. The following methodology can then be adopted:

- i. Develop a pre-intervention bivariate feed-back model connecting y and x ;
- ii. Specify an intervention model from *a priori* knowledge of the case under study;
- iii. Fit the models to the series, pre-and post-intervention;
- iv. Interpret results.

Computer-Intensive Methods

For a successful interface between statistics and cybernetics, it is obvious from the above discussion that there should be tremendous computational power in one's hand. This is happily provided by the advent of computer technology—both in hardware as well as software. In fact, the revolution in the field of electronic computation has led to the development of new statistical theory and methods in the 1980s, particularly the upsurge of computer-intensive statistical methodology such as *bootstrap methods*, *non-parametric regression*, *generalized additive models* etc. The most significant point about these methods is that fewer distributional assumptions are required for their application than the corresponding classical statistical methods. The replicability of the results of the experimental scientists can therefore be better ensured. For instance, in most of the cases of the classical methods, normal distribution is invariably assumed whereas in practice it need not necessarily be true so that conclusions obtained from such methods are often approximate or even get vitiated unless these methods are robust. No wonder then that on repeating the experiment, we get different results.

Bootstrap methods are basically simulation methods conducted on high-speed computers and aimed at generating new data sets from the observed original data set. When the data set happens to come from a long-tailed probability distribution, a *trimmed mean*, where a certain percentage of observations are deleted from the lower and upper ends of the ordered data set, provides a much more accurate statistic than the usual mean. Since we do not know *a priori* whether the true distribution is long-tailed or not, it is preferable to use trimmed means. For instance, Efron and Tibshirani (1991) illustrate the use of trimmed mean with the help of data on cholesterol reduction scores of 9 men after taking cholestyramine. The following are the scores in an ordered random sample from the scores of 164 men:

-21.0, 3.25, 10.75, 13.75, 32.50, 39.50, 41.75, 56.75, 80.0

The ordinary mean is $\bar{x} = 28.58$ with an estimated standard error of 10.13. The estimated confidence interval is then $28.58 \pm 10.13z$ where $z = 1.645, 1.960, 2.576$ relating to the areas under the normal curve for the respective 90, 95 and 99 per cent confidence intervals. When we take 25 per cent trimmed mean, $\bar{x}(0.25)$ is obtained as 27.81. For obtaining the standard error of this estimate, we do not have any algebraic formula like in the case of ordinary mean. A

number of bootstrap samples of size 9 are therefore drawn independently with replacement from the original data set using random number generator on computer, the 25 per cent trimmed mean is calculated for each sample and an empirical standard deviation of these sample means is worked out to obtain the bootstrap estimate of the standard error for \bar{x} (0.25). For different number of bootstrap samples, we get

Number of samples	Standard error
25	12.44
50	9.71
100	11.50
200	10.70
400	10.48

The randomness in the bootstrap standard errors comes from the use of a finite number of samples. It has been found that this is negligible when the number of samples is greater than 200. The setting up of confidence intervals with bootstrap is, however, not that simple.

The above simple algorithm of bootstrap sampling can be applied in several other cases like regression, time series, stratified sampling, multivariate structure etc. The algorithm is perfectly general. For instance, in the case of non-parametric regression between two variables y and x , *loess* – a computer-based fitting method – instead of usual least squares is used. In this method, no model is assumed over the entire range of the independent variable x . Instead, a series of *local* regression curves for different values of x (the target point) is fitted, in each case using only data points near the target x -value of interest (say a moving window of 20 per cent of the points). The process is repeated for all possible target points to obtain a non-parametric regression. The bootstrap algorithm is applied as before to obtain the bootstrap *loess* curves giving the standard errors of the predicted values at each of the target points.

STATISTICS AND GENETICS

The mechanism of random phenomenon which is the basis of statistics permeates the whole science of genetics and hence

statistical concepts and methodology interface extensively with genetics (Narain, 1987). For instance, in diploid organisms, when the sperm and egg, the sex cells or gametes from the two parents, unite to form a zygote, the beginning of a new individual, the biological process of fertilization faces all the risks one can think of. If the fertilization is successful, the union of sperm and egg bridges the generations and ensures the survival of the species. If not, the energy and labour invested in the production of gametes is wasted. It is, therefore, due to the role played by random phenomenon during the process of fertilization that the process of inheritance takes place. Similar examples of random phenomenon can be cited in almost every aspect of genetics and breeding. This random mechanism is handled through the calculus of probability which provides the theoretical base to statistics. When the principles of genetics and breeding are applied to domesticated plants and animals, data get generated. The methodology of statistics then provides with a most powerful means for efficient planning of observational and experimental genetics and breeding programmes as well as an unambiguous interpretation of the resulting data for drawing objective inferences. Efficient methods of estimation of genetic parameters and tests of hypotheses for appropriate inferences about these parameters are made use of extensively in the application of statistical genetic principles to problems in genetics and breeding. Narain (1990, 1991) give details of the interface between statistics and genetics in a comprehensive manner.

The random phenomenon requiring the use of statistics is encountered both when we study inheritance at the individual level as well as when we consider inheritance at the populational level. In the latter case, *population genetics* to be precise, particularly when the population size is finite, random process plays a predominant role in the form of a stochastic process of gene frequency which has been exploited recently in relation to evolution at the molecular level. The features of statistics become of paramount importance in the context of the inheritance of quantitative characters, *quantitative genetics* to be precise, which is basic to the improvement of domesticated plants and animals by breeding.

Stochastic Process in Evolutionary Genetics

Perhaps the most important and advanced application of statistics in genetics is in relation to the

mathematical-cum-statistical theory of population genetics and evolution (Narain, 1983). Fisher, Haldane and Wright contributed significantly, by developing deterministic and stochastic models in this context. Since characters are transmitted from one generation to another by genetic factors, evolutionary changes are expressed in terms of replacement of a given factor by its counterpart in a given population. In other words, it is the gene substitution at the population level which is reflected in evolution. For describing quantitatively the process of evolution and measuring its rate, we study the population dynamics of gene substitution. This involves population genetic principles and particularly the stochastic process of gene frequency (proportion of various *alleles* in the population) when the population is considered as finite. But the evolution occurs at the phenotypic level. Till recently, we did not have any means to connect the gene frequency change with phenotypic evolution. Mathematical and statistical models of population genetics were used to infer about the manner in which evolution might have occurred. With the advent of molecular genetics, this difficulty has now been overcome. It is now possible to compare in related organisms, say man and mouse, RNA molecules (direct product of genes) and proteins (ultimate product of genes) so as to estimate the rate at which allelic genes are substituted in molecular evolution. By synthesis of protein chemistry and palaeontological (fossil) studies, we can better understand the mechanics of evolution.

Random change in gene frequency due to finite population size is given the name *Sewell Wright effect* as Wright (1921, 1931) was the first to point it out. It is actually random genetic drift involving random fluctuation of gene frequency in a finite population over time caused by random sampling of gametes in reproduction. In any sexually reproducing species, the total number of individuals is usually a random sample chosen from a much larger gene pool of male and female gametes produced by the parental generation. Suppose we consider a model population consisting of 4 individuals each with two alleles, both of them being equally distributed in the ratio of 1:1. The population gene frequency is thus 50 per cent. If random sampling follows a binomial law, it may happen that out of 8 gametes we may find 3 of one kind and 5 of another kind in the random sample ultimately selected. This will give rise to the second generation of individuals in which the gene frequency will now be 37.5 per cent. This drift in gene frequency from 50 to 37.5 per cent is purely by chance. The average change in the frequency of a gene

in one generation is, however, zero but the variance follows the binomial law.

In considering the population dynamics of mutant substitution, we need information regarding the probability of gene fixation as well as the average length of time involved for each gene substitution. For getting quantitative estimates of time to fixation, a general theory based on diffusion approximation as well as Markov chain methods was developed by Narain (1974, 1977). This involves conditioning the process for the contingency of eventual fixation. In the diffusion approximation approach it leads to conditioned forward as well as backward diffusion equations with modified drift coefficient but with the same diffusion coefficient as in the case of the unconditional process. Using the backward form of the conditioned diffusion equation, one develops differential equations for obtaining the various moments of the distribution of time until fixation of a gene. In the case of pure random drift, we are able to obtain an explicit formula, showing that the average time to fixation depends on the effective population size as well as the initial gene frequency. For the case of a mutant introduced singly in the population in the initial stage, this average time tends to a limit $4N_e$, where N_e is the effective population size.

Quantitative Genetics

For several characters which show differences of degree rather than of kind, characters such as yield of a crop or milk yield of a cow showing continuous variation, it is important to determine, by means of appropriate statistical methods, the extent of differences between the individuals due to genetic factors as opposed to non-genetic ones. For the inheritance of such quantitative characters, the group of individuals or population is studied in terms of statistical parameters which summarize the population characteristics. Quantitative genetics is thus essentially statistical in nature and this is the reason why the statistical science has played an important role in the genetic improvement of economic characters in domesticated plants and animals which has ultimately led to the improvement of agriculture for the benefit of mankind.

A statistical measure of the variability of a quantitative character in a given population, is provided by the phenotypic variance which is compounded of genetic and environmental variability. The ratio of the genetic to total variance, which has been found to be very useful, is known as *heritability* of the character. The heritability of a character in a narrow sense (h^2) enables us to predict the breeding

value of a plant or an animal for the given character. It is in fact the regression coefficient (b_{AP}) of the breeding value (A) on the phenotypic values (P) since it can be shown that the covariance between the two is equal to the variance due to additive effects of genes or in other words the genic variance σ_A^2 . The variance due to phenotypic values σ_P^2 , being the total variance, the correlation (r_{AP}) between A and P is equal to h . The heritability therefore expresses the reliability of the phenotypic value as a guide to the breeding value and enters almost every formula connected with the breeding method. Many decisions of the breeders depend on its magnitude. It must, however, be noted that heritability is a property not only of a character, but also of the population and of environmental conditions under which individuals are observed and measured. Any change in the genetic or environmental components of the total variance can bring about a change in its value. The genetic component is affected by gene frequencies and therefore may differ from one population to another according to its past history. For example, a small population maintained for a very long time, may have genes which are virtually fixed and therefore may indicate low heritability than a large population. Similarly if environmental conditions are variable due to differences in management, feeding etc., the value of heritability will be reduced. Hence, any estimate of the heritability of a character must be referred to a particular population under particular environmental conditions.

When expressed in the narrow sense, the parameter (h^2) is particularly useful for prediction of the gains to be expected from selection practised by breeders of domestic plants and animals for the improvement of economic characteristics. Assuming that the character is normally distributed, let the mean of the parent population be \bar{X}_p with standard deviation σ_p . If individuals having values greater than a certain value c are selected as progenitors of the next generation and the mean of this selected group is say \bar{X}_s , then the selection differential to which the population is subject to, is $S = \bar{X}_s - \bar{X}_p$. Expressed in units of σ_p , this is the intensity of selection $i = (S/\sigma_p)$ which is equal to (z/p) where p is the proportion of the population selected and z is the ordinate of the standard normal distribution at the truncation point c such that the area to the right of it is p . The expected genetic gain in the character, or in other words, the response R to the selection applied is then

$$R = h^2 S = i h^2 \sigma_p = ih \sigma_A$$

Thus for a population with a certain variability expressed in terms of σ_p and at a given selection pressure expressed in terms of i , the response is more if h^2 is high and less if h^2 is low. This relationship can be extended to the multivariate case also (Narain, 1992).

CYBERNETICS AND GENETICS

As already stated earlier, cybernetics is the science of control and communication in animals and machines. In the former case, cybernetics has interfaces with several biological sciences particularly genetics and evolution. The concept of control in cybernetics implies auto-regulation – a property possessed by all biological organisms – known as *homeostasis*. It originated with the concept of physiological homeostasis developed and elaborated by Walter B. Cannon in the field of physiology. It refers to the property of the organism to adjust itself in response to fluctuating inner and outer environments. It is this property which enables organism to adapt, survive and evolve. The idea of homeostasis has been extended to an inter-breeding group or more precisely to Mendelian populations. Self-regulation at the level of such populations is referred to as *genetic homeostasis* (Lerner, 1954).

Due to the work of Cannon (1932), the property of homeostasis has been reduced in terms of interactions between constituent parts and functions. In the context of evolution, the origin of biological feed-back mechanism – the cybernetic concept – is in itself a question of central interest. Simple genetic models normally assume the absence of interaction between genotype and environment and therefore do not consider this inevitable property of self-regulation. For a more realistic modelling of living organism, we must take such interactions into account so as to throw light on the genetic-cum-environmental homeostasis. However, these interactions lead to regulatory mechanism by controls from within the individual. There has to be sufficient feed-back for these controls to be effective. How do the internal controls arrange for survival and evolution can only be understood by invoking the principles of genetics. This can be made clear by looking into the nature and control of a physiological characteristics like feed intake in man.

In regard to food intake, we have necessarily to consider three factors, viz., the appetite, the ability of individual to digest the food that he takes and his ability to convert the absorbed nutrients into energy for the purpose of the activity in which he is engaged. The

amount of nutrients absorbed by the individual depends on the rate of intake and on the digestive efficiency. These nutrients are used for the maintenance and activity of the individual (in the case of children, for growth also). Individuals having large appetite will tend to have a larger food intake leading thereby to some set-back in their digestive efficiency but they would also have a higher quantity of absorbed nutrients available for the maintenance and work. No doubt any superfluous nutrients would be diverted towards deposition of fat instead of protein and the activity would be somewhat impaired. The genetic differences between individuals are therefore mostly due not to the differences in the inherent efficiency of the separate bio-chemical processes but in the way in which the individuals direct the absorbed nutrients to different functions and activities of the body.

Although the chromosomal DNA in the cells of man controls his all activities, it alone is not sufficient. The living body is engaged in active exchange with the environment and DNA regulates these exchanges. It functions as the information store having a permanent record of the individual's potential. Any particular activity such as the utilisation of the energy produced by intake at any given time depends upon which sections of the information encoded in the DNA are being expressed and also the relative rates at which this information is being transcribed. The information transfer from the genes is the basic source of the regulatory system of the living creature by virtue of which individuals maintain homeostasis. The genetic controls are inherited from the parents and passed on to the offspring. Using them, other polynucleotides are produced including messengers to control production of appropriate enzymes for physiological controls, metabolic systems and effectors which are all inter-dependent, the required energy for these reactions being drawn from the metabolic system. The whole system takes actions that oppose tendencies in the environment to disintegrate them and a steady state of living organism is maintained (Waterman, 1968)

Models of Protein Deficiency

Sukhatme and Margen (1978) developed the concept of protein requirement of individuals in a dynamic sense and indicated the method by which it can be extended to those of populations. Their approach is to take into account the intra-individual variability in requirement, not as a random noise due to measurement errors but in a manner represented by an auto-regressive 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

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

where w_{it} is the balance on the t -th day for the i -th individual, ρ is the serial correlation of order one between w_{it} and $w_{i(t-1)}$ and e_{it} is a random variable with mean zero and variance σ_e^2 . This model represents an auto-regressive Markov process, comprising 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 $\rho = 0$. The expected value of w_{it} is zero with variance $\sigma_e^2/(1-\rho^2)$ which is independent of t and therefore remains constant. Such a process is known as stationary stochastic process. The observed value of balance on any given day will then be distributed around mean zero within limits $\pm 2\sigma_e/\sqrt{1-\rho^2}$ which are known as homeostatic limits.

The data relating to daily N-balance on fixed intake of Calloway and Margen (1971) as well as from experiments conducted at the Department of Nutritional Sciences, University of California, Berkeley, USA, were examined by Sukhatme and Margen (1978) for the nature and degree of intra-individual variation. 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 implies that the daily N-balance is regulated. This regulation is produced by 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. At high levels of protein, metabolism is altered, becoming more rapid than before, body weight increases, and the frequency and amplitude of oscillations become larger and irregular. At low levels of protein, body weight decreases, and the frequency and amplitude of the oscillations increase and decrease respectively. In either case, the organism is under stress. It was shown that the magnitude of stationary variance is 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 conclude that protein deficiency must be defined as a failure of the process to be in statistical control, and not defined in the manner that assumes requirement to be fixed whereby if an individual consumes protein below this level, he suffers from protein deficiency.

Genetic Significance of Auto-Regulatory Mechanism

Sukhatme and Narain (1982a,b) 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 handled by invoking a model for the one-way analysis of variance with correlated errors.

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

$$Y_{it} = \mu + b_i + w_{it}$$

where Y_{it} is the corresponding response with μ as overall mean, b_i 's are independently and identically distributed with variance σ_b^2 , independently of w_{it} and w_{it} 's for the same individual, are n consecutive random variables following an auto-regressive process of order one as described earlier. They have shown that the variance of the mean of the individual when averaged over n different days can be expressed as

$$V_{P(n)} = \sigma_b^2 + \bar{r} \sigma_e^2 + \left(\frac{1 - \bar{r}}{n} \right) \sigma_e^2$$

where \bar{r} is the average correlation between observations of a given individual. It is related to ρ , the serial correlation of the auto-regressive process.

The effect b_i in the above model reflects genetic effects of the i -th 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_{Eg}) so that

$$\sigma_b^2 = V_G + V_{Eg}$$

In so far as σ_e^2 is concerned, it reflects 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 in the within-individual component so that when the observations are averaged for several days, it does not bring about the reduction in the variance of the mean of the individual to the extent it would do if the genetical-physiological process of calorie or protein metabolism had been the same on each day. We, therefore, get

$$\bar{r} = \frac{V_{GE_s}}{V_{E_s} + V_{GE_s}}$$

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 as already discussed. 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 x 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.

The above considerations can be extended to consider covariance between relatives so as to reveal how genetic relationship controls the auto-regulatory mechanism measured in terms of ρ . 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

$$\bar{r}^* = \frac{r_g V_{GE_s}}{V_{E_s} + r_g V_{GE_s}}$$

We can then relate the two as

$$\bar{r}^* = \frac{r_g \bar{r}}{1 - (1 - r_g) \bar{r}}$$

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

STATISTICS, CYBERNETICS AND GENETICS

It is thus apparent from the bilateral interfaces discussed in the foregoing that we expect to have interface among statistics, cybernetics and genetics. As a matter of fact, cybernetics helps us to understand genetic systems using statistical principles. In order to illustrate this, we present the analysis conducted by Sukhatme and Narain (1983) on the data collected by Edholm *et al* (1970) regarding food intake and energy expenditure of army recruits. The data collected by Edholm *et al* (1970) involved simultaneous observations on intake and expenditure on 35 young army recruits during the 2nd, 5th and 8th week of a nine-week period of training at 6 different Depot Centres in UK, using the most sophisticated

Table 1. Analysis of daily energy intake and expenditure in kcal/kg of body weight in Depot Centre D
(period : Monday to Friday)

	d.f.	Mean Square		Estimates of True Variance	
		Intake	Expenditure	Intake	Expenditure
Between subjects	5	788	37	37	—
Between weeks within subjects	12	230	58	22	4
Between days within weeks	72	122	36	122	36
σ_w		12		6	
Mean		61		60	
% C.V.		20		11	
Var. of mean		46		11	
% S.E.		11		6	

techniques of measurement. During weekdays, the subjects were engaged in similar activities of group nature such as drill, lectures, meals and games. The intake of the subjects was measured using the weigh as you eat method from a common table and expenditure was calculated by timing activities and from energy expended during each activity as estimated from the amount of oxygen consumed using Integrated Motor Pneumotachograph (IMP) devised for the purpose. The examination of data showed that intake varies much more than expenditure from day to day at the various centres. The cumulative energy balance at the end of week (period) was very large and varied considerably from one period to another. Table I gives analysis of variance of the data on daily intake and expenditure at Depot Centre D for the period of five days from Monday through Friday. The intake and expenditure are expressed on per kilogram body weight basis.

Let σ_b^2 stand for the true variance between subjects, σ_p^2 for true variance between periods within subjects and σ_d^2 for true variance between days within period within subjects. Then the expected values of the three Mean Squares in the hierarchical model of ANOVA would be :

$$E(\text{between subjects within Depots}) = 15\sigma_b^2 + 5\sigma_p^2 + \sigma_d^2$$

$$E(\text{between weeks within subjects}) = 5\sigma_p^2 + \sigma_d^2$$

$$E(\text{between days within weeks}) = \sigma_d^2$$

where the numbers preceding σ_b^2 and σ_p^2 represent the number of days observed for each subject and during given period. It follows that the variance of a single observation for any subject is given by

$$\sigma_w^2 = \sigma_b^2 + \sigma_p^2 + \sigma_d^2$$

and that of the mean intake and expenditure per kg of body weight over p days by

$$\sigma_p^2 + \frac{\sigma_d^2}{p}$$

The Table also shows estimates of true variance as also of σ_w and of the variance of mean intake and expenditure per kg of body weight together with the respective coefficients of variation. It will

Table II. Variance of individual's mean energy intake, expenditure and balance based on p successive days as proportion of unit variance for $p=1$. Depot centre D

Period (days)	Observed			Expected	
	Intake	Expenditure	Balance	$\rho = 0.00$	$\rho = 0.30$
1	1.00	1.00	1.00	1.00	1.00
3	0.43	0.63	0.55	0.33	0.49
4	0.27	0.37	0.31	0.25	0.39
5	0.28	0.29	0.29	0.20	0.32
6	0.21	0.32	0.25	0.16	0.27

be seen that the mean square between periods is significantly larger than mean square between days within periods. This indicates that the variance of the mean of p values stabilizes at σ_p^2 . The CV of daily intake is seen to be 20 per cent. If day to day variations were random, the CV of mean daily values over time in the same individual would be smaller. The data show that it is not.

Unlike intake, the variation in the case of energy expenditure is smaller. The CV for daily expenditure is 11 per cent and that of the mean daily expenditure is 6 per cent. The smaller magnitude of the CV in the case of energy expenditure is to be expected since the army recruits were engaged in fixed tasks from day to day, but they had choice of intake on all days. This analysis refutes the assumption that expenditure in an individual maintaining body weight and engaged in similar activities from day to day is constant and is equal to individual's habitual intake. This suggests that the body regulates its energy balance on a range of intakes by varying the efficiency of intake utilisation.

The series of Edholm *et al* (1970) referred to earlier relate to 3 non-continuous weeks and therefore does not permit a direct study of auto-correlation to prove that energy requirement is self-regulated over a range of intakes much in the same manner as the series of daily N-balance in man living on fixed level of N and maintaining body weight studied in Sukhatme and Margan (1978) indicate. The mathematical model for such an auto-correlation, however, can be indirectly verified by computing variances of the mean balance.

when daily balance is averaged over 2, 3 or more successive days. The results of this exercise are given in Table II. It is found that the variance of the mean balance does not vary inversely as the length of the period, but that it decreases slowly, thus confirming that successive values are serially correlated. Further, the hypothesis that the daily balance is distributed in a stochastic stationary manner of the same Markovian type with serial correlation of the first order equal to 0.3 seems perfectly plausible. Edholm's data must, therefore, be interpreted to mean that although intake may not be equal to expenditure even when averaged over a week, man is in balance every day in a probabilistic sense with varying intervals between peaks and troughs and varying amplitudes in daily balance.

The estimates of intra-class correlation (\bar{r}) and of serial correlation of order one (ρ) in daily intake and expenditure in kcals/kg of body weight in Depot Centre D are :

	Intake	Expenditure
\bar{r}	0.15	0.10
ρ	0.23	0.17

The heritable portion of the variance arising from interaction with environment is admittedly small being of the order of 0.15 for intake and 0.10 for expenditure. Although small, the resulting value of the auto-correlation is consistent with the hypothesis of auto-regressive model of order one advanced earlier. The small size need not detract from its value in explaining the cause of the persistent nature of intra-individual variation. In the nature of things \bar{r} and with it ρ , are expected to be small when the main concern is to study the influence of day to day change on the performance of the next day within the frame-work of ontogenic growth. In other words, while the normal range of \bar{r} will be $-1/(p-1)$ to 1, and the distribution skew, it will be even narrower and the distribution skewer in the short term phenomenon that is studied here.

It would thus appear that energy balance is under the physiological control of the body with different genes switched on and off on different days and periods of ontogenic development in response to specialised environment. It is because of this interaction that man is able to bring for himself a change in the rate of energy

flow, speeding it sometimes and slowing it down on other occasions, to maintain homeostasis. No one can change the genetic blue-print with which one is born, but it appears that one can influence its development under the impact of sustained specialised environment. Although accounting for only a small portion of the total variance, this possibility has obviously far reaching implications for social, economic and cultural changes.

CONCLUSIONS

Finally, I hope the interfaces which I have presented, would bring home the point that at least, in so far as statistical profession is concerned, we should look at it as a cross-disciplinary activity and develop a feel of the subject matter area in which it is applied. Such an approach is gaining momentum as can be seen from a Report of Olkin and Sacks (1988), Co-chairs of a Panel of the Institute of Mathematical Statistics and the setting up of an Institute of Statistical Sciences in North Carolina, USA based on their recommendations. I would like to end by quoting from their Report:

'The widespread influence of interaction between statistics and other disciplines and the very nature of statistics as the science of the meaning and use of data establish the statistical science as the discipline with the most central and complex cross-disciplinary activity'.

REFERENCES

- [1] Bhattacharya, M.N., 1982. Assessing sales response from intervening television advertisement by a firm. *SCIMA*, **11**, 33-45.
- [2] Box, G.E.P. and Jenkins, G.M., 1970. *Time Series Analysis. Forecasting and Control*, Holden-Dey, San Francisco.
- [3] 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.
- [4] Cannon, W.B., 1932. *The Wisdom of the Body*. Norton, New York.
- [5] Dawid, A.P., 1984. Statistical Theory. The prequential approach. Present position and potential developments : Some personal views. *J.R. Statist. Soc. A.*, **147**, 278-292.
- [6] Efron, B. and Tibshirani, R. 1991. Statistical data analysis in the computer age. *Science*. 253, 390-395.
- [7] Garg, J.N., Narain, P., Singh, S. and Kumar, M. 1991. Statistical studies in relation to crop insurance. IASRI Project Report New Delhi.
- [8] Ghosal, A. 1978. *Applied Cybernetics : Its Relevance to Operations Research*, Gordon and Breach, New York.
- [9] Ghosal, A. 1986. Interface between statistics and cybernetics. In : Ghosal, A., (ed.) *Pragmatic Cybernetics*. pp. 53-66. South Asian Publishers, New Delhi.
- [10] Lerner, I.M. 1954. *Genetic Homeostatis*. Oliver and Boyd, Edinburgh & London.
- [11] Mittal, P. 1989. Statistical aspects of crop insurance. Ph.D. Dissertation. IASRI, New Delhi.
- [12] Narain, P. 1974. The conditional diffusion equation and its use in population genetics. *J.R. Statist. Soc. B.* **36**, 258-266.
- [13] Narain, P. 1977. The conditional Markov chain in a genetic context. *J. Genetics*. **63**, 49-62.
- [14] Narain, P. 1983. Stochastic problems and methods in population genetics. Presidential Address. Section of Statistics, 70th session of Indian Science Congress held at Tirupati. Proc. Part II, pp. 1-32.
- [15] Narain, P. 1987. The role of random process in genetics. Presidential Address. Physical Sciences Section. 57th Annual Session of the National Academy of Sciences, India, Allahabad held at Bharathidasan University, Tiruchirappalli: pp. 1-14.
- [16] Narain, P. 1990. *Statistical Genetics*. John Wiley & Sons, New York and Wiley Eastern Limited, New Delhi.

- [17] Narain, P. 1991. Statistics in genetics. In : Srivastava U.S. (ed.) Glimpses of Science in India. pp. 131-146. Diamond Jubilee Volume, National Academy of Sciences, India, Allahabad.
- [18] Narain, P. 1992a. Presidential Address. National Seminar held at IARI, New Delhi in October, 1991. In : Sinha, M.N. and Rai, R.K. (eds.). Agriculture in Nineties : Challenges and Research Needs. pp. 1-5. Indian Society of Agricultural Sciences, New Delhi.
- [19] Narain, P. 1992b. A dialectical perspective of agricultural research for sustainable development. *J. Nuclear Agri. and Biol.*, **21**, 140.
- [20] Narain, P. 1992c. Population genetics of quantitative characters. In: Majumder, P.P (ed.). Modern Human Genetics : A Centennial Tribute to J.B.S. Haldane. Plenum Publishing Co., New York.
- [21] Narain, P., Singh, S., Garg, J.N. and Kumar, M. 1985. Statistical aspects of comprehensive crop insurance scheme. Symposium, 39th Conference, ISAS. *Jour. Ind. Soc. Ag. Statistics*, **38**, 105-108.
- [22] Olkin, I. and Sacks, J. 1988. Co-chairs, Cross-disciplinary researches in the statistical sciences. A Report of the Panel of the Institute of Mathematical Statistics, USA, National Science Foundation.
- [23] Parzen, E. 1979. Non-parametric statistical data modelling (with discussion). *JASA*, **74**, 105-131.
- [24] Sukhatme, P.V. and Margen, S. 1978. Models for protein deficiency. *Amer. J. Clinical Nutrition*, **31**, 1237-1256.
- [25] 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 & Sons, New York.
- [26] Sukhatme, P.V. and Narain, P. 1982b. A possible genetic interpretation of the auto-regulatory mechanism in models for protein deficiency. *Proc. Indian National Science Academy, Series B*, **48**, 748-754.
- [27] Sukhatme, P.V. and Narain, P. 1983. Intra-individual variation in energy requirement and its implications. *Indian J. Med. Res.*, **78**, 857-865.
- [28] Tukey, J.W. 1977. Exploratory Data Analysis. Addison-Wesley Publishing Co., Reading, Mass.
- [29] Waterman, T.H. 1968. Systems theory and Biology - view of a Biologist, Systems theory and Biology. Proc. III Systems Symp. Case Inst. Tech. (ed. M.D. Mesarovic), New York.
- [30] Wright, S. 1921. Systems of mating, I-V. *Genetics*, **6**, 111-178.
- [31] Wright, S. 1931. Evolution in Mendelian populations. *Genetics*, **16**, 97-159.