STATISTICAL ISSUES IN QTL MAPPING AND MAS FOR PLANT AND ANIMAL GENETICS IMPROVEMENT PROGRAMME

P Narain
New Delhi

The advent of the methods of molecular genetics has made significant impact on the statistico-genetic principles for the genetic improvement of domesticaed plants and animals. One of them in *Restiction Fragment Length Polymorphism* (RFLP) for detaction and analysis of *quantitative trait loci*(QTL). Attempts are made to locate the position of the QTL as well as to estimate its effect by studying the regression ofrthe trait value on the QTL genotype given the marker genotypes. However, the independent variable representing the QTL genotype is not known with certainity but takes values with probabilities which depend on the recombination probabiblities between the marker genotypes and the specified QTL genotypes. This raises statistical issues like method of maximum likelihood with incomplete data, two types of errors in detecting the QTL and the number of individuals required so that the *lod score* exceed the threshold. While such studies are conducted for every trait separately, evidence from the plant and animal genetic researches indicate that some genomic segments indeed have pleiotropic effects affacting several traits, some of them being discrete valued with an underlying continuous variable,thereby leading to correlated trait complexes. This gives rise to statistical problems involving more complicated structures than hitherto attempted. The methods of molecular genetics can be integrated with those of the artificial selection on individual and/or collateral basis by applying what has come to be known as *marker assisted selection*(MAS). With information availabe on several auxiliary traits and on the molecular scores corresponding to these traits, statistical considerations involve developing optimum selection indices which maximize genetic improvement in the main trait.

Some of the statistical issue cited above will be discussed.

P NARAIN
B– 3/27 A, LAWRENCE ROAD
KESHAV PURAM
NEW DELHI-110035

STOCHASTIC CREATION PROCESS AND LARGE SCALE STRUCTURE IN COSMOLOGY

Jayant V. Narlikar
Inter-University Centre for Astronomy and Astrophysics,
Pune 411 007, India

One of the most challenging features of the universe to understand and interpret today is its large scale structure. In the early days of modern cosmology theoreticians made the assumption that if the galaxy, typically of mass 10^{11} times the solar mass and containing as many stars, is taken as a unit then these are distributed homogeneously in the universe. Already in the mid-1930s this view was being questioned by Fritz Zwicky who argued that there was a larger unit, namely a cluster of galaxies, that emerged when one looked at the distribution of galaxies. Later in the late 1950s, when the distribution of galaxies was further examined by Abell, de Vaucouleurs and a few others, evidence for second order clustering began to emerge. However, it was not until the early 1980s that the concept of superclusters containing $10^{14} - 10^{15}$ solar masses in the form of galaxies and extending to sizes of the order of 50-100 Mpc (1 Megaparsec $\sim$ 3 million light years), was established. It is now known that superclusters are distributed in apparently filamentary structures separated by giant voids of size $\sim$ 100 Mpc.

With the completion of more and more surveys of galaxies with their redshifts, it has become possible to carry out statistical analysis of their distribution and clustering on different scales. One useful parameter to calculate is the two-point correlation function, which shows a scale-invariant power-law form:

$$\xi(r) \propto r^{-1.8} \tag{1}$$

where the number density of galaxies from any point to distance r from it varies as

$$n(r) = n_0[1 + \xi(r)]. \tag{2}$$

Here n_0 is the average number density of the population. Theoreticians have therefore to find out why and how such structural hierarchy evolved and why does it show the observed two-point correlation function.

It is against such a background that a toy model of computer simulations was tried by Nayeri et al. (1999) for a new cosmology, known as the *quasi-steady state cosmology* (QSSC in brief) which seems to hold promise in the sense that it seems to reproduce the observed distribution as outlined above. We first describe the essential features of the QSSC.

Proposed in 1993 by F. Hoyle, G. Burbidge and J.V. Narlikar, this cosmology describes the universe as undergoing a long-term exponential expansion together with short term oscillations. The scale factor of the simplest QSSC model is described by

$$S(t) = \exp\,(t/P) \times [1 + \eta \cos 2\pi\tau(t)/Q], \tag{3}$$

with typically $P \gg Q$. The parameter η lies in the open interval (0,1), thus preventing $s(t)$ from becoming zero. The function $\tau(t)$ is very nearly like t. The mathematical derivation of this formula and other details of the QSSC see Hoyle et al. (1995) and Sachs et al. (1996). This cosmology has no instant marking a beginning of the universe. It has time axis from $-\infty$ to $+\infty$, with each cycle of oscillation physically the same as the previous one. Hence the adjective 'steady' in QSSC. The adjective 'quasi' emphasizes the oscillatory part with period Q. Had there been no oscillation, the model would have been the same as the steady state theory of Bondi, Gold and Hoyle proposed in 1948.

In the QSSC, matter creation takes place preferentially near collapsed massive objects, and the created mass may be expelled as particles and radiation, or as coherent masses. It is the latter aspect that was used by Nayeri et al (op.cit.) for simulation in the toy model. The theory envisages that most of the creation of new matter takes place at epochs close to the oscillatory minima. Thus if the density of matter at one minimum epoch were ρ, then by the time the next minimum were reached the density would diminish by the factor $\exp[-3Q/P]$. The shortfall to be made up by the new creation at this epoch is $f \times \rho$, where

$$f = 1 - \exp[-3Q/P] \tag{4}$$

With this background we now describe the toy-model of Nayeri et al (1999).

Consider a unit cube in which a random number generating programme is used to produce a large number $N \sim 10^5 - 10^6$ of points. Around each of a fraction f of these points *chosen at random*, create a neighbour placed randomly within a sphere of radius $\alpha \times N^{-1/3}$ where the fraction α is chosen to lie between 0 and 1. These are creation events. Next, the entire cube is homologously expanded by a linear factor $\exp[Q/P]$. The density of points in the expanded cube will now be the same as in the original cube when the N points were first produced. Take the concentric inner cube of unit side from this expanded cube by deleting the outer shell. This would leave on an average N points and we then repeat the exercise for the next 'creation event'. After a few such repetitions we begin to see the initial random distribution give way to clusters and voids. We can also see filaments, if we modify the creation algorithm slightly. However, for the purposes of this presentation, we may not consider this possibility.

The interesting result arises when we consider the two-point correlation function for the distribution. After a few iterations, the $[\log \xi(r) - \log r]$ plot closely approximates to a slope of -1.8 before falling down steeply telling us the characteristic distance up to which clustering continues. Evidently, the algorithm that is generating the cluster-void distribution in expanding space, uses the standard probability theory and stochastic processes. While astronomically it is satisfactory to see the algorithm producing the observed distribution, one would like to know: Why does the distribution tend to the -1.8 slope for the correlation function? To what extent is the conclusion robust? Although the toy model has α as a culstering parameter, to what extent is the final slope dependent on its value and on the ratio P/Q which determines how frequent are the creation events?

This unsolved problem is brought here to the attention of statisticians and probability theorists in the hope that they may have encountered this type of fractal-like structure elsewhere and have a theoretical understanding of it. For, an explanation at this level will be extremely useful in understanding the hierarchical structure found in the universe.

References

Hoyle, F., Burbidge, G and Narlikar, J.V. (1995), *Proc. R. Soc. A.*,**448**, 191

Nayeri, A., Engineer, S. Narlikar, J.V. and Hoyle, F. (1999), *Ap. J.*, **525**, 10

Sachs, R., Narlikar, J.V. and Hoyle, F.(1996), *A&A*, **313**, 703.

Jayant V. Narlikar
Inter-University Centre for Astronomy
and Astrophysics
Post Bag 4, Pune 411 007, India
Fax No.: 91-20-5656417
E-mail : jayant@iucaa.ernet.in

MAPPING QUANTITATIVE TRAIT LOCI THROUGH PRINCIPAL COMPONENTS REGRESSION

J. T. Gene Hwang
Cornell University

Dan Nettleton
Iowa State University