4

Other Modelling Areas

One of the principal strengths of mathematical modelling is its generic nature; the same set of tools and methodology can be applied to a wide range of problems. This has been fully exploited at C-MMACS, applying mathematical modelling to such diverse areas like design of aircraft wing, gene sequence analysis and cryptography.

The work reported for 2001-2002 reflects this trend:

Inside

4. 1 A Symmetry Scheme for Amino Acid Codons

The genetic code uses four "letters" or the bases adenine(A), thymine(T), guanine(G) and cytosine(C) in the four nucleotides constituting the DNA (or uracil(U) instead of T in the corresponding RNA template) by reading them in groups of three. During protein synthesis, these triplets of three bases (or codons) encode for specific amino acids. There are thus $4^3 = 64$ possible codons. Of these, 61 code for the known 20 amino acids — the remaining three (UAG, UGA, and UAA) code for termination or "stop" codons. In many cases, more than one codon can code for a specific amino acid — for instance any one of six specific codons can code for serine, any one of four particular codons can code for valine, etc. Hence, the genetic code is degenerate, and even though there are 64 available codons, only 20 amino acids relevant to mammalian proteins actually occur in nature. A few years ago, it was discovered that one of the stop codons, UGA, translates under certain circumstances to a twenty first amino acid selenocysteine. A fundamental problem of molecular biology is why, despite the redundancy of the codons, the genetic code did not expand any further and stopped when the number of the amino acids is only about one third the number of available codons — does the genetic code have any mathematical property which gets optimised for these numbers? This question was addressed by adapting some standard group theoretical methods of particle physics to molecular biology. It was shown within a specific model that it is possible to account for the number of the naturally occurring amino acids.

The 64 codons were classified within a semi-empirical model which very closely resembles the decomposition of the Kronecker product 4 Ä 4 Ä 4 of the group SU(4). We looked at the hydrophilic and hydrophobic tendencies of the amino acid residues constituting the proteins, as they play a key role in determining the conformation of a protein and the way it folds. The codons fall into well-defined multiplets, and represent at their respective positions, the amino acids they code for — all the members of a multiplet share the same hydropathic property.

The model explains the existence of synonymous codons (thus explaining how twenty one amino acids (including selenocysteine) have been found to occur so far out of a possible sixty four). It also enables us to predict the possible existence of two more, as yet undiscovered amino acids : UAG having properties similar to histidine, and similarly, UAA having properties similar to lysine or arginine, even though these two codons, both differ from the codons for tyrosine only at the wobble position.

In general, the amino acids which have closer similarities between themselves, occur nearer to each other within each multiplet. The amino acids in each multiplet are also in close conformity with the standard suggested amino acid substitutions based on the Dayhoff matrix wherein amino acid residues which are near to each other in the Dayhoff plot are good candidates for mutual exchange for conservative mutations in proteins.

Earlier approaches pioneered by Hornos *et al*, for using group theoretical methods for studying the genetic code are different from ours, and do not account for the twenty first amino acid selenocysteine having properties similar to cysteine, and also presently lack the predictive power present in our model.

(Janaki Balakrishnan)

4. 2 Modelling of Growth Patterns in Road Transport in India

Transport infrastructure is vital for rapid socio-economic transformation of a traditionally agricultural society to an industrially advanced set-up. However, in India, even after seven five-year plans, the transport sector has steadily lagged behind requirements and has experienced bottlenecks and capacity shortages, with hardly any signs of resilience and robustness. This common perception underlines the need for creating capacity much ahead of manifestations of demand so that some cushion exists in the system to meet the expected peak loads and unexpected shifts in traffic patterns.

As the transport system grows, in a country like India, a stage is reached when the system becomes so complex as to limit the understanding and thereby control over the growth process. For adequate and proactive planning, mathematical modeling and computer simulation exercises are required.

In this work, we have analyzed the dominant long-term trends in independent India as regards the roads sector and discussed its implications for policy making.

The most striking feature of the growth of Indian road transport is the steep growth in two-wheelers. This is exemplified in the graph (Fig. 4.1) showing the changing proportions (percentages of particular categories with respect

to the total number of vehicles) of the various categories with time. From a modest proportion of approx. 10% in 1950, two-wheelers shot up to occupy a most dominant position at approx. 70% proportion by 1998; it's only around 1990's that its growth rate started to slow down. Such a situation creates much road congestion, increase in accident rates and pollution levels. As regards the four-wheelers like cars, jeeps etc., they were the most dominant segment in 1950's; they comprised more than 50% of all vehicles then. However, this segment has reduced in size to <15% in 1998; in fact, the nosedive started to slow down only in the late 1980's. Surprisingly, the percentage of buses also took a plunge: from >10% in 1950's to <2% in 1998!

An important feature of the above graph is that all vehicles show stabilization tendency at the present percentage levels; for nearly a decade, this leveling off can be noticed. Policy-wise, the most important result concerns that of the buses. It is imperative that their small percentage share be made to grow to significant proportions. Only then can the pollution, road congestion etc. be made to decrease and hygienic and safer cities be developed.

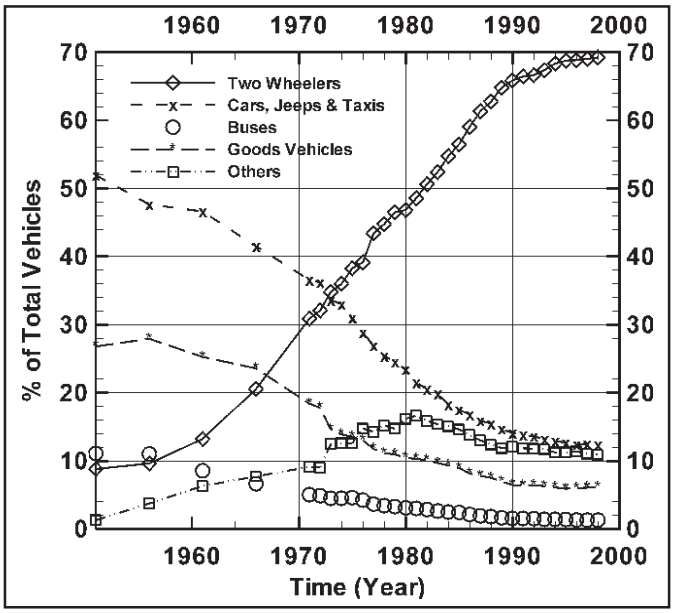

Fig. 4.1 Change in the percentage composition of vehicles in various categories with respect to the total number of vehicles.

In this connection, it is instructive to look at the growth in the proportion of two-wheelers and four-wheelers (more or less the personalized transport segment) to buses which is depicted Fig. 4.2. The tearing pace at which two-wheelers have outnumbered buses is easily evident in this graph. Starting from an approximately equal situation in 1951, they have grown to about 40 times the number of buses by late 1980's. While the growth of four-wheelers have only been at a marginally higher rate on the

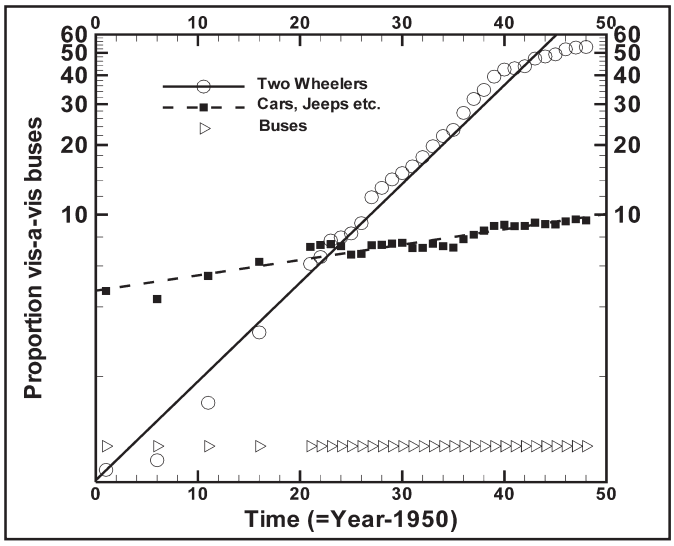

Fig. 4.2 Growth in the ratios of other vehicles to buses.

whole, it is the two-wheelers that have grown at such a pace as to leave the others virtually standing in the same place. However, the rate has come down in the 1990's and seems to be stabilizing at the present (1998) proportion of about 53 two-wheelers to one bus; it was 49.25 in 1995.

In an earlier study (1991-92; using data up to 1988), we had observed the alarming proportion of two-wheelers to buses and our forecasts of the number of vehicles had indicated that, around the turn of the century, there would be about 130 two-wheelers to a single bus. The slowing down in the growth of two-wheelers in 1990's have changed this scenario for the better and has indicated there would only be around 58 two-wheelers to a single bus in year 2000. Our earlier study had also drawn attention to the near-perfect exponential growth evident in the data, which indicated that the gap between supply and demand, on the inadequate side, would also amplify exponentially. It is heartening to note that some slowing down in the exponential growth is evident so that the chances of catching up with the demand has brightened.

(T R Krishna Mohan and K S Yajnik)

4. 3 Cryptography: An Interactive Algorithm for Classical Key Distribution

Quantum key distribution claims to have overcome the threat posed by quantum computers to the modern cryptographic methods. Though the basic laws of physics prove its credibility as a secure way of communication, the physical realization is still far from reality. It may so happen

that one-day quantum computer will be a reality, but still we will be quite far away from a fully operational quantum cryptographic method. This emphasizes the need of a parallel research in classical cryptographic methods so that we don't run out of solutions at any situation. In addition to this, a quantum–classical combined method may provide better solution in future.

The method basically describes an interactive algorithm between Alice and Bob to generate a random symmetric key with nearly zero knowledge to eve. In this method prior to communication, Alice and Bob generate completely independent set of random binary number string. Then they inform each other about their bit positions in a psudo-encrypted way. The communication is carried out in a three-phase algorithm as follows.

a. Estimation of Error

Alice and Bob estimate the error rate R in a publicly selected and agreed upon random sample of the binary string by publicly comparing the bits. The revealed bits are discarded from the binary string. If R exceeds a certain threshold then the random number generation process is repeated.

b. Extraction of the key

Alice and Bob publicly agree upon a random permutation and apply to the remaining binary string.

· Alice and Bob partition the binary string into blocks of length L.

· For each of these blocks, Alice and Bob publicly compare overall parity checks, making sure each time to discard the last bit of the compared block.

· Each time an overall parity check does not agree, Alice and Bob initiate a binary search, i.e., bisecting the block into two sub blocks, publicly comparing the parities of each sub blocks, discarding the right most bit of each sub block. The bisective search continues on the sub block for which their parities are not in agreement. This is carried out for all blocks.

· The above steps are repeated again and again until all the blocks parity agree between Alice and Bob even after many numbers of random permutations.

c. Key verification

Alice and Bob verify the equalness of the remaining key by publicly selecting a random sample of them and comparing publicly. If it doesn't match 100% then the whole process is repeated again. If it matches exactly then the revealed bits are discarded from the binary string and the rest of the bits or a subset of it are used as the symmetric key.

The advantage of this encryption method is that it is not based on any mathematical computational problem. The key generated is so random in nature that, even the same set of initial binary string will generate two different keys both in size and value at two different instants of time.

(G K Patra)

4.4 A More Robust Approach to the Analysis of Spatial Data

Standard spatial statistics involves exploratory data analysis (EDA) and the computation of a semi-variogram from spatial data. However, the uncertainties in the computation of the most of the EDA and semi-variogram parameters cannot be estimated. Standard EDA allows the computation of the uncertainties associated with only the sample mean. The standard Bootstrap methodology was extended to spatial statistics computations using finite strain data from the Sheeprock thrust sheet, Sevier fold-and-thrust belt, western USA. The Bootstrapped EDA is found to be more robust than the standard EDA because it allows uncertainties to be computed for all EDA parameters. Thus, using Bootstrapped EDA, # Mean for the Sheeprock strain data was recomputed to be 1.2834 ± 0.0006 instead of 1.2832 ± 0.0127, Median 1.2866 ± 0.0006 instead of 1.2845, Mode 1.2929 ± 0.0013 instead of 1.2870, Skewness 0.5746 ± 0.0243 instead of 0.4726 and Peakedness 4.8335 ± 0.0503 instead of 4.8335. Bootstrapped spatial statistics allows the computation of a better and a more robust semi-variogram than standard spatial statistics as the uncertainties associated with the semi-variogram parameters can be ascertained. Thus, the Range (**a**) = 750 m and Sill (**c**) = 0.008 values associated with the exponential semi-variogram

$$\gamma(|h|) = c\,[1 - e^{\frac{-3|h|}{a}}]$$

computed for the Sheeprock strain data was recomputed, using Bootstrapped Spatial Statistics, to be 719±32.2048 and 0.0097±0.0004 respectively. Kriging estimates computed using the Bootstrapped semi-variogram indicate that the results are practically insensitive to the uncertainty associated with the estimation of parameters of the semi-variogram used.

(Malay Mukul, Debjani Roy, S Satpathy, Anil Kumar)