

A Brief Outline of the Project

1. Introduction:

A model is a mathematical expression whose structure and hence behavior corresponds in some sense to that of a particular reality or phenomenon.

Three approaches to modeling are

- i. Conceptual approach: based on a rational argument or on the knowledge.
- ii. Empirical approach: based on the data and it does not take an account of any other thing.
- iii. Eclectic approach: a combination of conceptual and empirical approaches.

Some important stages in modeling are

- a. Identification
- b. Estimation and Fitting
- c. Validation
- d. Application

The main objective of the project *"STATISTICAL MODELS DERIVED FROM WELL-KNOWN DISTRIBUTIONS"* is to introduce new statistical models essentially derived from well-known families of distributions. When the experiments are conducted under ideal conditions (i.e. in the absence of distorting factors), a distribution from a well-known model may fit well to the data. This fitted distribution is used to describe the random behavior of the outcomes of the experiment. But, in the presence of a distorting factor these standard models may not give a good fit, as a consequence random behavior of the outcomes of the experiment can not be properly described by using statistical models and as such some new models have to be derived from well-known distributions.

To develop new models we use the techniques of Truncation, Modification (of the form of the densities or distribution functions), Modified version of failure rate and Compounding. We develop some models by using typical stopping rules.

2. An outline of the work done :

By using the technique of truncation, an extended family of distributions can be defined. As in this case range of the variable is confined to a subset of the original, there would be a scope to enlarge the natural parameter space. Such models by using exponential and geometric distributions have been proposed. Distributional properties and related inference problems for the new models are studied. (Please refer to Unit 1 of the report.) On the proposed model, a seminar was presented in the Department of Statistics, Pune University. We are thankful to Professor B. K. Kale for visiting us and offering his valuable comments.

For the modification purpose, we consider $\mathcal{H}$, a certain family of non-decreasing functions from $[0, 1]$ to $[0, 1]$ and indexed by α . By using such a class $\mathcal{H}$, a given family of distributions can be extended. The classes of distributions proposed by Marshall and Olkin (1997), Mudholkar et al (1995) can be obtained by this technique. Extreme value distributions also can be obtained by using such functions. Various class properties of $\mathcal{H}$ have been defined. An article "Families of Distribution Functions" was prepared and presented in the International Conference at Allahabad. This article is enclosed as Unit 2 of the report. We thank to Professor Ratnaparakhi of Wright State University for fruitful discussions that we had in Pune during his visits to India.

Another technique for modification is to introduce a 'drift' parameter, ' a '. By using this technique a class of asymmetric distributions can be obtained based on a symmetric (or for that matter any) class of distributions. Such classes

obtained are different from those of Jones and Faddy (2003). In particular, we study drifted distributions obtained from a typical type of densities and in particular from Laplace and Normal distributions. The drifted Laplace distribution with θ , the location parameter and σ , the scale parameter has been considered. Maximum likelihood estimators of the α , θ and σ when some specified and none specified are obtained. As estimators do not have closed form expressions, a program in C language to compute MLE's is developed and is enclosed in the appendix of Unit 3. Two piece normal distribution is a particular example with location and scale parameter θ and σ respectively and ' α ' as the drift parameter. MLE of ' α ' and as well of (θ, σ) , when other parameters are known, are considered. These problems are different from those considered by John S. (1982), Nabar S. P. and Barpande S. P. (2001). Extensive simulation study has been carried out to study the performances of the proposed estimators. The results are reported in the tabular form and a C program developed for the purpose is given in the appendix of Unit 3. In the course, we are able to introduce a new class $\{T(\alpha) : \alpha \geq 0\}$ of estimator of location parameter. Well-known measures mean, median and mode turn out to be $T(1)$, $T(0)$ and $T(\infty)$ respectively. For details please refer to Unit 3 of the report.

In the literature, the models have been proposed based on failure rate $(r(t) = f(t)/(1-F(t))$ or $r(k) = P_k/(1 - F(k-1))$, according as the r.v. is positive continuous or positive integer valued. We propose models by specifying P_1 and the ratio

$$\frac{P_k}{P[X \geq k-1]},$$

where $P_k = P\{X = k\}$, particularly when the ratio is constant. For details, please refer to Unit 4 of the report. We are grateful to Professor M. Sreehari and Professor M. N. Deshpande for very fruitful discussions.

We introduce a class of bivariate r.v.s (X, Y) where X is a non-negative integer valued r.v. and Y is a sum of X i.i.d r.v.s. $Z_1, Z_2, \dots$. That is Y is a compounded r.v. Let θ_1, θ_2 be the parameters of the distributions of X and Z respectively and $\theta = (\theta_1, \theta_2)$. In general, the distribution of Y depends on θ . In many situations, it depends on a function of θ . We discuss the problem of retainment of identifiability of the r.v. Y . An article "Compound Bivariate Distributions" was presented in a National Conference at Vallabh Vidyanagar. By considering the families of marginal distributions of X and Y , we introduce 'change of order' relation and illustrative examples are given. For details, refer Unit 5 of the report.

To maintain the quality of the units produced, it is essential to monitor the production process and it is achieved by using appropriate control charts. By using these charts, a unit or a group of units is categorized as of acceptable quality or not. As outcomes being binary, the set up can be described by using a sequence of Bernoulli trials. The time to get a signal for process being out of control can be viewed as a stopping time corresponding to certain stopping rule based on a sequence of Bernoulli trials. In some of these situations, essentially we need to solve certain difference equations. By using generating functions one can obtain the probabilities. We define certain stopping rules and derive expressions for average time to stop by using different methods. These expressions can be used to develop new efficient control charts. These results are reported in Unit 6 of the report.

In process control techniques the production process is stopped upon getting 'enough' evidence for a shift in the process level. To obtain more efficient procedures of desired precision, various types of rules are being proposed. Consider the following rule described in terms of sequence of iid Bernoulli r.v.s.

'Start with coin A, if tail occurs stop, else continue with coin B till a tail is obtained on B and then restart with coin A'.

For such a type of rule based on the run lengths average time to stop can be derived. An article 'Average Time to Signal in a Sequence of Independent Trials' has been prepared jointly written with M.P. Gadre. For the details please refer to Unit 7 of the report. Further based on this result a paper 'A Unit and Group Runs Based Control Chart to Identify Increases in Fraction Non-conforming' (jointly written with M. P. Gadre) was presented in an International conference, Kolkata.

3. Visits:

A) Reference work and Discussions:

1. Indian Statistical Institute, Delhi : 17. 01 2002 to 19. 01.2002
2. Indian Statistical Institute, Bangalore and National Institute of Mental Health and Science (NIMHS), Bangalore: 25 02.2003 to 02.03.2003
3. Indian Statistical Institute, Kolkata: December 15-17 2003.
4. Department of Statistics, University of Pune
5. Professor B. K. Kale from Pune visited the Department under the project and we had very fruitful discussions.

B) Paper presentations at seminars / conferences:

1. International conference on Statistics Combinatorics and Related Areas, Allahabad December 21-23, 2002
2. Departmental Seminar, Pune
3. International Symposium on Emerging Strategies for Improving Productivity, Quality and Reliability, Kolkata December 12-14, 2003
4. The UGC sponsored National level Seminar : Vallabh Vidyanagar January 29-31, 2004

4. Data collection :

- To study goodness of fit of some proposed models, the data related to
- (i) the behavior of the arrival times of persons to the place of work, board a train and to a Cinema theatre
 - (ii) the tendency of persons to pay the electricity bill and
 - (iii) weights of some packages
- were collected. Few M. Sc. students were involved in the fieldwork.

5. Purchase of Equipments and Books_:

A. Equipments : Computer and Printer (Rs. 58, 500/-)

B. Books :

Sr. No.	Title	Author	Price
1	Continuous Multivariate Distributions Vol. I, Ed. II	Johnson N.L.& Kotz S.	Rs. 5290/-
2	Modeling Survival Data	Therneau, Grambsch	Rs. 2790/-
3	Continuous Univariate Distributions Vol. I	Johnson N.L.& Kotz S.	Rs. 254/-
4	Continuous Univariate Distributions Vol. II	Johnson N.L.& Kotz S.	Rs. 254/-
5	Modeling Survival Data in Medical Research	Collett D.	Rs. 2182/-
6	Reliability Modeling: A Statistical Approach	Wolstenholme L. C.	Rs. 1745/-
7	Nonlinear Models in Medical Statistics	Lindsey J. K.	Rs. 2276/-
8	In All Likelihood	Pawitan Yudi	Rs. 2601/-
9	Analysis of Failure and Survival Data	Smith Peter J.	Rs. 2170/-
10	Fitting Statistical Distributions	Karian Z.A.& Dudewicz E.J.	Rs. 3646/-
11	Statistical Models	Davison A. C.	Rs. 1657/-
Total			Rs. 24, 865/-

6. Appointment of Project Fellow(s) :

Ms. Arati Erandole joined as a P.F. on 2nd July, 2001 and resigned on 11th June 2003. Ms. Reshma R. Jamadar joined as a P.F. on 26th August, 2003 and worked up to 31st May 2004, the last date of the project. The work carried out by both of them is satisfactory. They assisted in data collection and analysis. They developed computer programs to obtain the estimators and carried out extensive simulation studies to study the performance of the proposed estimators.