A Synopsis of the thesis entitled

Some unbiased estimation problems in inlier(s) prone models

For the degree of

Doctor of Philosophy

Submitted to

The Maharaja Sayajirao University of Baroda

By
Bavagosai Pratima

Under the guidance of
Prof. K. Muralidharan

**Department of Statistics, Faculty of Science,
The Maharaja Sayajirao University of Baroda,
Vadodara-390002, India.**

August, 2018

Some unbiased estimation problems in inlier(s) prone models

1. Introduction

Efforts have been made consistently over the years to develop classes of "standard" distributions and their methodologies to suit a wide range of phenomena concerning nature, life and human activities. There are plethora of examples where the data do not conform to these "standard" distributions. In earlier years the efforts have also been made to develop probability models which are in fact mixtures of standard distributions of similar types. Of late the literature become available that provides and deals with "nonstandard" mixtures which are actually mixtures of a degenerate and a standard distribution which may be a discrete or continuous one.

Inliers are observations (or group of observations) sufficiently small relative to the rest of the observations, which appears to be inconsistent with the remaining data set (Kale and Muralidharan, 2000). They are either the resultant of instantaneous failures, or early failures or both, experienced in life testing experiments, clinical trials, weather predictions, geographic information systems, athlete performance analysis and many other such applications. The test items that fail at time 0 are called *instantaneous failures* and the test items that fail prematurely are called *early failures*. These occurrences may be due to inferior quality of a product or service, or faulty construction or alignment of events/objects, or due to no response to the treatments. Such failures usually discard the assumption of a single mode distribution and hence the usual method of modeling and inference procedures may not be accurate in practice.

Now there are many practical contexts, where inliers can be natural occurrences of the specific situations involved and degeneracy can happen at discrete points and a positive distribution for the remaining life times. Some of the situations are as follows:

1. In an audit sample, we have two pieces of information, namely, the book amount (recorded) and the audited amount (correct). The difference between the two is called the error amount. Here some population elements contain no error, whereas other population elements contain error of varying amounts. The distribution of errors can therefore be viewed as two distinguishable distributions, one with a discrete probability mass at 'zero'

and the other a continuous distribution of non-zero positive and/or negative error amounts. The data here can be modeled using a nonstandard distinguishable mixture.

2. In the mass production of technological components of hardware, intended to function over a period of time, some components may fail on installation and therefore have zero life lengths. A component that does not fail on installation will have a life length that is a positive random variable whose distribution may take different forms. Thus the overall distribution of lifetime is a nonstandard distinguishable mixture.

3. Time until remission is of interest in studies of drug effectiveness for treatment of certain diseases. Some patients respond and some do not respond to the treatment. This is an example for a distribution having a mixture of mass point at 0 which corresponds to instantaneous remission and a nontrivial continuous distribution having positive remission times.

4. In a study of tooth decay, the numbers of surfaces in a mouth which are filled, missing or decayed are scored to produce a decay index. Healthy teeth are scored 0 for no evidence of decay and are therefore a mixture of a mass point at 0 and a nontrivial continuous distribution of decay score.

5. The rainfall measurement at a place recorded during a season is modeled as a continuous distribution with a nonsingular distribution at zero, where zero measures those days having no rainfall etc.

6. Machines and software's are tested for its correctness and perfectness or reliability. Bugs in such situation are important to assess the durability and credibility of machines and programs. Zero defects or zero bugs are considered to be good in such situations. If there are bugs, then it can be measured in terms of some discrete measurements.

For similar such examples see Statistical models and analysis in Auditing: Panel on nonstandard mixtures of distributions, Statistical science (1989) and the details contained therein.

In situations above and many others, the random mechanism is modeled using a complete mixture of two or more distributions, or a nonstandard mixture of distributions. We consider a general case of a two-component univariate complete mixture model where one component's distribution function is F_1 and the other component's distribution function is F_2 while the mixing proportion p unknown. Such a model can be defined at its most general form as

$$G = p\,F_1 + (1-p)\,F_2,\, 0 \le p \le 1 \tag{1}$$

An inliers prone model can be viewed as follows: If the distribution F_1, is due to instantaneous or early failures, then F_1 may be called, the degenerate distribution at some point, where, the observation may be reported as either 'zero' or 'close to zero'; otherwise, the observation may be coming from a positive probability distribution F_2. That is, such problems lead us to consider a random variable X with the following properties: There is a non-zero probability p that $X = 0$ and probability $1 - p$ that X is non-zero. Sometimes the mixtures are distinguishable in the sense that one can tell which population an observation has come from, where as in some cases F_2 also permits 'zero' value with positive probability. Modeling these situations need special care and treatment.

The problem of outliers is well-known in statistics: an outlier is a value that is far from the general distribution of the other observed values, and can often perturb the results of a statistical analysis. Various procedures exist for identifying and studying outliers (Barnett and Lewis, 1994). An inlier, by contrast, is an observation lying within the general distribution of other observed values, generally does not perturb the results but is nevertheless non-conforming and unusual. For single variables, an inlier is practically impossible to identify, but in the multivariate case, thanks to interrelationships between variables, values can be identified that are observed to be more central in a distribution but would be expected, based on the other information in the data matrix, to be more outlying. In that sense, the lower outlier may be treated as inliers.

In literature some authors have defined inliers as those observations which are not outliers. While outliers are erroneous observations located farther away from the sample mean, inliers are erroneous observations located closer to the mean, (Akkaya and Tiku, 2005). According to UN publication (UNECE, 2000): An inlier is a data observation that lies in the interior of a data set and is unusual or an error. Because inliers are difficult to distinguish from the other data values, they are sometimes difficult to find and – if they are in error – to correct.

Aitchison (1955) was the first to discuss the inference problem of instantaneous failures in life testing. The author has provided the efficient estimation of parametric functions under various probability models. Kale and Muralidharan (2000) have introduced the term *inliers* in connection with the estimation of (p, θ) of early failure model with modified failure time

distribution (FTD) as exponential with mean θ assuming p known. See also Jayade and Prasad (1990), Shinde and Shanubhougue (2000) Muralidharan and Lathika (2004), Kale and Muralidharan (2000, 2008) and Muralidharan (2010), and Muralidharan and Pratima (2016b) for more studies on these models.

From the above examples, it is seen that the values including zeros and close to zeros are important as well as significant in most cases. For instance, zero errors in an audit report, zero tooth decay, and zero bugs in a computer program or electronic machine are all good to judge the prevailing situation, and hence they are significant. Similarly, zero life time, zero rainfall (dry day) etc. are all situationally bad but significant as per the conditions and situations. Thus inliers are more natural than outliers, where most of the time inliers are retained after the detection and considered for future analysis. As a consequence, the modeling of inliers distribution is more important than its detection and further treatment. Below we introduce various inliers prone models. The thesis will further delve into the theoretical treatment of these inliers models. Many practical problems are treated in various chapters in this thesis for estimation of parameters and testing of hypothesis.

1.1. Instantaneous failure models

Since, instantaneous failures are a natural occurrence and such failures usually discard the assumption of a unimodal distribution and hence the usual method of modeling and inference procedures may not be accurate in practice. To tackle these situations the model is represented as

$$G(x; p, \theta) = \begin{cases} 1 - p, & x = 0 \\ 1 - p + p\, F(x; \theta), & x > 0 \end{cases} \tag{2}$$

with respect to a measure μ which is the sum of Lebeasgue measure on $(0, \infty)$ and a singular unit measure at the origin; and $0 < p < 1$. Here the model $\mathcal{F} = \{F(x; \theta), x \geq 0, \theta \in \Theta\}$ where $F(x; \theta)$ is a continuous failure time distribution function with $F(0) = 0$ is to be suitably modified as a non-standard mixture of distribution by mixing a singular distribution at zero to accommodate instantaneous failures. Aitchison (1955) was first to study this model with underlying distribution $F(x; \theta)$ as exponential. This was further studied by Kleyle and Dahiya (1975), Jayade and Prasad (1990), Vannman (1991, 1995), Kale (1998, 2003), Muralidharan (1999), Muralidharan and Kale

(2002) and Muralidharan and Lathika (2005, 2006). Kale and Muralidharan (2007, 2008), Adlouni et al. (2011), Muralidharan and Pratima (2016$_c$) have also considered other commonly used parametric models such as Lognormal, Gamma, Weibull, Pareto, Lindley and other distributions.

1.2. Early failure model-1

If it is assumed that $\lambda(x) = \lambda = \frac{1}{\theta}$ for all x from exponential distribution, then the assumption of an exponential density is equivalent to the assumption of a constant failure rate. Under this set up, Miller (1960), proposes the early failure model as

$$\lambda(x) = \begin{cases} \lambda_1, & 0 \leq x < T_0 \\ \lambda_2, & T_0 \leq x \end{cases} \tag{3}$$

where $\lambda_1 > \lambda_2$. The probability density correspond to this failure rate is

$$f_X(x; \lambda_1, \lambda_2) = \begin{cases} \lambda_1 e^{-\lambda_1 x}, & 0 \leq x < T_0 \\ \lambda_2 e^{-\lambda_1 T_0 - \lambda_2 (x - T_0)}, & T_0 \leq x \end{cases} \tag{4}$$

The justification follows from the fact that when a component is put on test, it is not known whether it is an 'early failure' or a 'standard' item. Since some will be early failures, the failure rate on the average will be high at the start, but if an item has survived for a certain period of time T_0, then it cannot be an 'early failure' so its failure rate will be lower for the succeeding time period. The above model can also be viewed as a model for shift in hazard function of exponential distribution.

1.3. Early failure mode-2

To accommodate early failures, the family $\mathcal{F}$ is modified to $\mathcal{G}_1 = \{G(x; p, \theta), x \geq 0, \theta \in \Theta, 0 < p < 1\}$ where the CDF corresponding to $g_1 \in \mathcal{G}_1$ is given by $G_1(x; p, \theta) = (1 - p) H(x) + p F(x; \theta)$. Here $H(x)$ is a CDF with $H(\delta)$ for δ sufficiently small, assumed known and specified in advance. Then the modified family $\mathcal{G}_1$ has a PDF with respect to measure μ, which is sum of Lebesgue measure on (δ, ∞) and a singular measure at δ as

$$G_1(x; p, \theta) = \begin{cases} 0, & x < \delta \\ 1 - p + p\,F(\delta; \theta), & x = \delta \\ 1 - p + p\,F(x; \theta), & x > \delta \end{cases} \tag{5}$$

Some of the references which treats early failure analysis with exponential distributions are Kale and Muralidharan (2000, 2007, 2008), Kale (2003), Muralidharan and Lathika (2006), Muralidharan and Arti (2008), Muralidharan (2010), Muralidharan and Arti (2013), Muralidharan and pratima (2016[a]) and the references contained therein. These authors treated early failures as inliers using the sample configurations from other parametric models including Weibull, Pareto, Normal and Gompertz distributions.

The models in (2) and (5) can be combined to form the CDF as

$$G(x; p, \theta) = \begin{cases} 0, & x < d \\ (1 - p) + pF(x; \theta), & x \geq d \end{cases} \tag{6}$$

with the corresponding probability density function (pdf) as:

$$g(x; p, \theta) = \begin{cases} 0, & x < d \\ 1 - p + pF(d; \theta), & x = d \\ pf(x; \theta), & x > d \end{cases} \tag{7}$$

If $d = 0$ the model reduces to the instantaneous failures case and if $d > 0$, it reduces to the case of early failures. The thesis address the inlier estimation for the above models under various probability models. One may also see Lai et al. (2007) for a complete mixture model, where they have treated the instantaneous part based on Direct delta function and a probability distribution for the positive distribution.

1.4. Model with inliers at zero and one

In some of the examples discussed above, the observations 0 and 1 become natural occurrence with other positive observations. If these observations are treated as inliers, then, the distribution function of such models can be written as

$$H(x; p_1, p_2, \theta) = \begin{cases} 0, & x < 0 \\ p_1, & 0 \le x < 1 \\ p_1 + p_2, & x = 1 \\ p_1 + p_2 + (1 - p_1 - p_2)\frac{F(x; \underline{\alpha}) - F(1; \theta)}{1 - F(1; \theta)}, & x \ge 1 \end{cases} \qquad (8)$$

where p_1 and p_2 are the proportion of 0 and 1 observations. This model was first studied by Muralidharan and Pratima (2017, 2018) with $F(x; \theta)$ as exponential and Weibull. One can also use other probability models for $F(x; \theta)$. Chapter 7 address the issue of Pareto 0-1 inlier model.

2. Unbiased estimation

Suppose we have a random sample $X_1, X_2, \dots, X_n$ whose assumed probability distributions depended on some unknown parameter θ. Our primary goal will be to find a point estimator $U(X_1, X_2, \dots, X_n)$ such that $U(X_1, X_2, \dots, X_n)$ is a "good" point estimate of θ when $x_1, x_2, \dots, x_n$ are the real values of the random sample. For getting such an estimator, we used to consider the underlying characteristics of the sample and then study the properties of their estimators. The desirable properties of good estimator are unbiasedness, consistency, efficiency and sufficiency. Apart from this the other properties are minimum variance estimator (MVE), minimum variance unbiased estimator (MVUE), uniformly minimum variance unbiased estimator (UMVUE), best linear unbiased estimator (BLUE) etc. Among all the estimators, the UMVUE are found be a good estimator and is unique, if it exists.

Many authors have studied the problem of minimum variance unbiased estimation for different classes of distributions. Roy and Mitra (1957), Joshi and Park (1974), Charalambides (1974) have studied the estimation problem for power series distribution, Patil (1963[a]), Patil (1965[c]) has studied the same for generalized power series distribution, Jani (1977) and Gupta (1977) has studied for modified power series distribution. Patel (1978) has studied the UMVUE of parameters for the multivariate modified power series distribution. All these studies include discrete distributions only. Jani and Dave (1990) have studied the problem of MVU estimation in one parameter exponential family of distributions which includes power series distribution, modified power series distribution and univariate continuous distributions. Further a characterization property of power series distribution using one and two moment was given by

Khatri (1959). Jani (1993) extended this for one-parameter exponential family of distribution which includes all earlier cases. Jani and Singh (1995) have further studied MVU estimation in multi-parameter exponential family of distributions. Given these facts, we explore UMVU estimation with due importance, for each inlier models.

3. Objectives of the study

The main objective of our study is to explore different unbiased estimation procedures of parametric functions including the density and the survival function for newly developed non-standard mixture of distribution using Roy and Mitra (1957) and other techniques. The other area of study includes the following:

- To propose various tests for inliers
- To suggest unbiased estimates for distribution function, survival function, and hazard rate function
- To obtain estimators of parameters exploring various estimation procedures.
- Conduct a simulation to study the standard error and other properties.
- To apply the estimation procedure to real life data.

We now discuss the plan of the study as follows:

Chapter 1 is a detailed introduction of the study and its foundation. An exhaustive literature survey on study of inliers is presented. It also contains necessary prerequisites for other chapters. We discuss the problem of UMVU estimation of the density, survival function and other parametric functions of inlier prone models (6) under Type II censoring scheme. We discuss the problem of UMVU estimation using Roy and Mitra (1957) approach instead of the usual approach of taking conditional expectation of an unbiased estimator given complete sufficient statistic. It can be seen that the unified approach is much simpler and easier than the usual approach to find UMVUE's of the density, survival function and other parametric functions of the distributions in the class of one parameter exponential family. Also the distributional properties of the complete sufficient statistics are explored. Using these complete sufficient statistics, an easy method of constructing UMVU estimators of some parametric functions and those of the variances of the estimators in the family have been discussed under model (8) for Type II censoring scheme.

In **Chapter 2** we revisit various inlier prone models (2), (4) and (5) stated above and propose inferential procedures to deal with different life testing situations. We studied the likelihood estimator and its characteristics, also proposed UMVUE for various parametric functions, wherever possible keeping FTD as exponential distribution. Various estimators and characteristics are studied with the support of numerical examples, one on failure times (in weeks) of 50 items as considered by Murthy et al. (2004) and another one on Vannman (1991) data on drying of woods under schedule 1 of experiment 2.

Chapter 3 reviews the various type of parametric tests for parameters of inlier prone models. We propose most powerful (MP) test locally most powerful (LMP) sequential probability ratio test (SPRT) for parameters p and θ of exponential distribution. Various test procedures for testing hypothesis consists of single and multiple inliers are reviewed. Some of the existing tests are revisited and studied in detail. The outward procedure for detecting inliers, namely, the block test is constructed through identified inlier model and labeled slippage model. A discussion on data descriptions with inlier proness and comparative study of sequential procedure and block tests is also included in this Chapter. We present the masking effect in Dixon type test and Cochran type test for inliers. The performance of the test is studied based on powers and masking effects. An extensive Monte Carlo study is carried out to investigate the powers and the error probabilities for the effects of masking and swamping effect in outward test when the number of inliers are more than one. We illustrated the same for two real examples in this chapter.

Chapter 4 we studies the inferences of the models by considering the FTD as Lindley distribution. We provide the likelihood estimator of the parameters of model and UMVUE for various parametric functions, including pdf and reliability function along with the SE of estimates. Various estimators and characteristics are studied with two real data sets, one is NFHS-3 survey on child's age on death for Gujarat state (http://www.dhsprogram.com) and another one is Vannman (1991) data on drying of woods under schedule 1 of experiment 2.

Chapter 5 the inferential studies of inliers in Gompertz distribution with Type II censored data is carried out using the general form of inlier prone model given in equation (6). Apart from MLE, UMVUE, we also studied least squares, weighted least squares and percentile based estimation procedures for estimating the parameters. An application of inliers prone models is illustrated with a real data set of Vannman (1991) data on drying of woods under schedule 1 of experiment 2.

The estimation of parameters based on Type-II censored sample for model presented in equation (8) with FTD Weibull distribution is experimented in **Chapter 6**. The Maximum Likelihood Estimators (MLE) are developed for estimating the unknown parameters. The Fisher information matrix, as well as the asymptotic variance-covariance matrix of the MLEs are derived. UMVUE of model parameters as well as UMVUE of various parametric functions are obtained. The model is implemented on a real data of tumor size in invasive ductal breast carcinoma (IDC) of female patients (www.cbioportal.org). The particular case of exponential distribution is also included in the chapter as it has lot of practical significance. The proposed model is then applied on real data based on the NFHS-3 survey for Gujarat state (http://www.dhsprogram.com).

Chapter 7 is devoted to the Type-II censored sample from a Pareto distribution. The parameters are estimated through various techniques. The model is implemented on a real data set of loss ratios (yearly data in billions of dollars) for earthquake insurance in California from 1971 through 1993 (Embrechts et al., 1999).

We also provide some discussion on the importance of inlier proness in practical applications in **Chapter 8.** The limitations, future work and conclusion are also suggested in this chapter**.**

The output of this thesis are now in the form of articles. So far we have already published five articles. Another two are under preparation and communication.

References

Adlouni, S. El., Ezzahid, E. and Mouatassim, Y. (2011). Mixed distributions for loss severity modeling with zero in operational risk losses. *International Journal of Applied Mathematics and Statistics*, 21(11): 14 pages.

Aitchison, J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of American Statistical Association*, 50: 901 – 908.

Akkaya, A. K. and Tiku, M. L. (2005). Robust estimation and hypothesis testing under short-tailedness and inlier, *Sociedad de Estadistica e Inveastigacion Operativa Test*, 14(1): 129-150.

Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data,* 3rd ed., John Wiley & Sons, New York.

Charalambides, C. H. (1974). Minimum variance unbiased estimation for a class of left truncated distributions. *Sankhya A*, 36: 392 – 418.

Dixit, V. U. (2003) Estimation of parameters of mixed failure time distribution based on an extended modified sampling scheme. *Communication in Statistical Theory Methods*, 32(10): 1911 – 1923.

Embrechts, P., Resnick, S. I. and Samorodnitsky, G (1999). Extreme value theory, *North American Actuarial Journal*, 3(2): 30 – 41.

Gupta, R. C. (1977). Minimum variance unbiased estimation in modified power series distribution and some of its applications. *Communication in Statistics*, 6: 977 – 991.

Jani, P.N. (1977). Minimum variance unbiased estimation for some left-truncated modified power series distributions. *Sankhya*, 39: 258 – 278.

Jani, P. N. (1993). A characterization of one-parameter exponential family of distributions. *Calcutta Statistical Association Bulletin*, 43 (3–4): 253 – 256.

Jani, P. N. and Singh, A. K. (1995). Minimum variance unbiased estimation in multi-parameter exponential family of distributions. *Metron*, 53: 93 – 106.

Joshi, S. W. and Park, C. J. (1974). Minimum variance unbiased estimation for truncated power series distributions. *Sankhya A*, 36: 305 – 314.

Kale, B. K. (1998). Optimal estimating equations for discrete data with higher frequencies at a point. *Journal of Indian Statistical Association*, 36: 125 – 136.

Kale, B. K. (2003). *Modified failure time distributions to accommodate instantaneous and early failures. Industrial Mathematics and Statistics*, Ed. J. C. Misra, Narosa Publishing House, New Delhi: 623 – 648.

Kale, B. K. and Muralidharan, K. (2000). Optimal estimating equations in mixture distributions accommodating instantaneous or early failures. *Journal of the Indian Statistical Association*, 38: 317 – 329.

Kale, B. K. and Muralidharan, K. (2007). Masking effect of inliers. *Journal of the Indian Statistical Association*, 45(1): 33 – 49.

Kale, B. K. and Muralidharan, K. (2008). Maximum Likelihood estimation in presence of inliers. *Journal of Indian Society for Probability and Statistics*, 10: 65 – 80.

Khatri, C. G. (1959). On certain properties of power series distributions. *Biometrica*, 46: 486 – 490.

Kleyle, R. M. and Dahiya, R. L. (1975). Estimation of parameters of mixed failure time distribution from censored data. *Communication in Statistics –Theory and Methods*, 4(9): 873 – 882.

Lai, C. D., Khoo, B. C., Muralidharan, K. and Xie, M. (2007). Weibull model allowing nearly instantaneous failures. *Journal of Applied Mathematics and Decision Sciences*, Article ID 90842, 11 pages.

Miller, R. G. (1960). Early failures in life testing. *Journal of American Statistical Association*, 48: 491 – 502.

Muralidharan, K. (1999). Tests for the mixing proportion in the mixture of a degenerate and exponential distribution. *Journal of the Indian Statistical Association*, 37(2): 105 – 119.

Muralidharan, K. (2010). Inlier prone models: A review. *ProbStat Forum*, 3: 38 – 51.

Muralidharan, K. and Arti, M. (2008). Analysis of instantaneous and early failures in Pareto distribution. *Journal of statistical theory and Applications,* 7: 187 – 204.

Muralidharan, K. and Arti, M. (2013). Inlier proness in Normal distribution. *Reliability: Theory and Applications*, 8(1): 86 – 99.

Muralidharan, K. and Bavagosai, P. (2016$_a$). Some inferential studies on inliers in Gompertz distribution. *Journal of Indian Society for Probability and Statistics*, 17: 35 – 55.

Muralidharan, K. and Bavagosai, P. (2016$_b$). A revisit to early failure analysis in life testing. *Journal of the Indian Statistical Association*, 54(1 & 2): 43 – 69.

Muralidharan, K. and Bavagosai, P. (2016$_c$). Some inferential study on inliers in Lindley distribution. *International Journal of Statistics and Reliability Engineering*, 3(2): 108 – 129.

Muralidharan, K. and Bavagosai, P. (2017). Analysis of lifetime model with discrete mass at zero and one. *Journal of Statistical Theory Practice*, 11(4): 670–692.

Muralidharan, K. and Bavagosai, P. (2018). A new Weibull model with inliers at zero and one based on type-II censored samples. *Journal of Indian Society of Probability and Statistics*, 19: 121 – 151.

Muralidharan, K. and Kale, B. K. (2002). Modified gamma distribution with singularity at zero. *Communications in Statistics– Simulation and computation*, 31(1): 143 – 158.

Muralidharan, K. and Lathika, P. (2004). The concept of inliers. *Proceeding of first Sino-International Symposium on Probability, Statistics and Quantitative Management*, Taiwan, October: 77 – 92.

Muralidharan, K. and Lathika, P. (2005). Statistical modeling of rainfall data using modified Weibull distribution. *Mausam*, 56(4): 765 – 770.

Muralidharan, K. and Lathika, P. (2006). Analysis of instantaneous and early failures in Weibull distribution. *Metrika*, 64(3): 305 – 316.

Murthy, D. N. P., Xie, M. and Jiang, R. (2004). *Weibull model*. New York: Wiley.

Patel, S. R. (1978). Minimum variance unbiased estimation of multivariate modified power series distribution. *Metrika*, 25: 155 – 161.

Patil, G. P. (1963$_a$). Minimum variance unbiased estimation and certain problem of additive number theory. *Annals of Mathematics and Statistics*, 34: 1050 – 1056.

Patil, G. P. (1965$_c$). On multivariate generalized power series distribution and its application to the multinomial and negative multinomial, Classical and Contagious discrete Distributions, (Edited by G. P. Patel), *Statistical Publishing Society*, Calcutta: 183 – 194.

Roy, J. and Mitra, S. K. (1957). Unbiased minimum variance estimation in a class of discrete distributions. *Sankhya*, 18: 371 – 378.

Shinde, R. L. and Shanubhogue, A. (2000). Estimation of parameters and the mean life of a mixed failure time distribution. *Communications in Statistics– Theory and Methods*, 29(1): 2621 – 2642.

Statistical methods and analysis in Auditing: Panel on nonstandard mixtures of distributions (1989). *Statistical Science*, 4(1): 2 – 33.

UNECE (2000). Glossary of Terms on Statistical Data Editing. United Nations Statistical Commission and Economic Commission for Europe. URL: http://www.unece.org/fileadmin/DAM/stats/publications/editingglossary.pdf, last accessed 28 May 2014.

Vannman, K. (1991). *Comparing samples from nonstandard mixtures of distributions with applications to quality comparison of wood*. Research report: 2, submitted to division of quality technology, Lulea University, Lulea, Sweden.

Vannman, K. (1995). On the distribution of the estimated mean from the nonstandard mixtures of distribution. *Communication in Statistics– Theory and Methods,* 24(6):1569 – 1584.

http://www.cbioportal.org/study?id=brca_metabric#clinical.

The DHS Program. India: Standard DHS, 2005–06. http://www.dhsprogram.com/data/dataset/ India_Standard-DHS_2006.cfm?flag=0