Jour. Ind. Soc. Ag. Statistics
51(2&3), 1998 : 397-406

Sukhatme and the Exponential Distribution

S.R. Adke
University of Pune, Pune-411007

SUMMARY

This paper reviews the original contributions of Professor P.V. Sukhatme to the problems of statistical inference for the two parameter exponential distribution. The present state of the analysis connected with the k-sample problem is described. Some comments on the directions for future work are made.

Key Words: Two parameter exponential distribution, Tests of hypotheses, Single sample and multiple samples, Cox process, Poisson process.

1. Introduction

Professor Sukhatme's contributions to agricultural statistics, biometry, nutrition and sample survey methodology are well-known and are widely appreciated. However, his two papers, Sukhatme [15], [16], published at the beginning of his statistical career, dealing with problems of statistical inference for the two parameter exponential distribution have not received the acknowledgement, appreciation and credit they deserve. In this paper, I, therefore plan to discuss the contents of Sukhatme [15], [16], summarise the subsequent work done in this area and indicate some possible directions of future work in the problems whose solution was initiated by Professor Sukhatme more than sixty years ago.

In Section 2, I shall discuss the results of Sukhatme [16], which mainly deals with the one sample problem for the two parameter exponential distribution. The contents of Sukhatme [15], which is concerned with the k sample problem $k \geq 2$, are discussed in Section 3. The subsequent work on the k-sample problem is reviewed in Section 4. the last section, Section 5, describes some possible directions of future work.

2. The One Sample Problem

Professor Sukhatme opens his [4] paper by saying that while the normal distribution and the associated techniques of statistical inference are "adequate to problems of agricultural research, it is not so far non-normal situations, e.g., the intervals between random events distributed in an exponential law of variation - a non-normal curve, totally different in shape and form from the normal". Nevertheless, the emphasis of the paper is on bringing out the similarity between the "Normal theory" and the "Exponential theory", as well as of the applicability of the then available statistical tables like those of percentage points of the chi-square, Student's - t and Fisher's z distributions.

Let $Y_1, \ldots, Y_n$ be a random sample of size n, $n \geq 2$, from the exponential distribution Exp (β, θ), with location parameter β, $-\infty < \beta < \infty$, scale parameter $\theta, 0 < \theta < \infty$, and the probability density function

$$f(x; \beta, \theta) = \theta^{-1} \exp\{-(x-\beta)/\theta\}, \quad x > \beta$$

Let $X_1 < X_2 < \ldots < X_n$ be the corresponding order statistics. Sukhatme [16] shows that

$$\hat{\beta} = X_1 \text{ and } \hat{\theta} = \sum_{i=2}^{n} (X_i - X_1)/n$$

are the maximum likelihood estimators (m.l.e.'s) of β and θ respectively. He corrects them for the bias and claims that

$$\tilde{\beta} = \hat{\beta} + \hat{\theta}/(n-1) \text{ and } \tilde{\theta} = n\hat{\theta}/(n-1)$$

are their "best unbiased" estimators. The statistics X_1 and $T = \sum_{i=2}^{n} (X_i - X_1)/(n-1)$ are identified as sufficient statistics. They are shown to be independently distributed. While X_1 has Exp $(\beta, \theta/n)$ distribution, $2(n-1)T/\theta$ has chi-square distribution with $2(n-1)$ degrees of freedom. In the course of this investigation, Sukhatme [16] discovers that

$$\xi_r = (n-r+1)\{X_r - X_{r-1}\}, \qquad r = 1, \ldots, n$$

with $X_0 \equiv 0$, are independent and identically distributed random variables. This fact or its minor modifications have been repeatedly re-discovered, see e.g. Reyni [13] and Epstein and Sobel [7].

An immediate consequence of the above results is that Sukhatme [16] is able to specify tests of the following composite hypothesis:

(a) $\beta = \beta_0$ against $\beta \neq \beta_0$, θ a nuisance parameter;

(b) $\theta = \theta_0$ against $\theta \neq \theta_0$, β a nuisance parameter.

He emphasizes the utility of the then available statistical tables for "normal" theory in carrying out the tests proposed by him.

It is interesting that while this paper settles most of the inferential issues relating to a single sample from a two-parameter Exp (β, θ) distribution, if treats the exponential distribution as a chi-square distribution with 2 degrees of freedom. Moreover, apart from repeatedly pointing out the analogy with the "Normal theory", Sukhatme, in contrast to his [15] paper, does not introduce any motivating "practical" problems.

I may also add that Sukhatme [16] has been wrongly cited as dealing with the problems of tests of hypothesis for $k, k \geq 2$, samples, see e.g. Bain and Englehardt ([3], p. 199); Vaughan and Tiku [18]. In fact, the correct reference is Sukhatme [15].

3. The k-sample Problem

As mentioned in the introduction, Sukhatme [15] develops tests of hypotheses about the equality of location and/or scale parameters of k independent, two-parameter exponential distributions. He motivates the study of these problems by citing data about lengths of intervals between the arrivals of successive calls at the Franklin Exchange of London Telephone Service during the busy period (11.00 a.m. to 01.00 p.m.) from each of seven different local exchanges. the questions posed by Professor Sukhatme are:

(a) Have the calls arrived at random in general?

(b) Could the records on different days or at different times on same day for a given exchange be combined without loss of homogeneity?

(c) Did the intensity of traffic at different exchanges differ significantly on the particular days in question?

Sukhatme [15] emphasizes that the then prevailing practice of employing the Poisson distribution to model "frequencies of these events occuring in fixed interval of time (or space)" was not suitable to answer the above mentioned questions. He therefore recommends the use of "interval analysis" as opposed to "count analysis". He also recognises that if one adopts the Poisson distribution for the "count data", one has to employ the exponential distribution for the "interval analysis". Sukhatme [15] does not specifically mention the relation between the Poisson process and the exponential distribution as the distribution of the inter-event durations. It is, however, easy to recognise that Sukhatme [15] poses problems relating to k independent Poisson processes and it is thus a pioneering contribution to the study of statistical inference for stochastic processes, a branch which flourished only after the late sixties.

Suppose then that a random sample of size n_i, $n_i \geq 2$, is available from the Exp (β_i, θ_i) distribution, $i = 1, \ldots, k$. Let $\omega_i = (\beta_i, \theta_i)$ and $\omega = (\omega_1, \ldots, \omega_k)$. The natural parameter space for the k-sample problem is

$$\Omega = \{\omega \mid -\infty < \beta_i < \infty, 0 < \theta_i < \infty; \quad i = 1, \ldots, k\}$$

The following three problems of testing of hypotheses are discussed in Sukhatme [15].

(i) The k Exp $(\beta_1, \theta_1), \ldots,$ Exp (β_k, θ_k) distributions are, in fact identical distributions. The null hypothesis

$$H_0 : \omega \in \Omega_0 = \{\omega \mid \beta_1 = \ldots = \beta_k; \theta_1 = \ldots = \theta_k\}$$

is to be tested against the alternative hypothesis

$$A_0 : \omega \in \Omega - \Omega_0$$

(ii) The k exponential distributions have the same scale parameter. The null hypothesis

$$H_1 : \omega \in \Omega_1 = \{\omega \mid \theta_1 = \ldots = \theta_k\}$$

is to be tested against the alternative

$$A_1 : \omega \in \Omega - \Omega_1$$

Here the location parameters $\beta_1, \ldots, \beta_k$ are nuisance parameters, both in H_1 and A_1.

(iii) Suppose the k exponential distributions have a common but unknown scale parameter θ. The hypothesis H_2 to be tested is that, within this sub-class of distributions, the k location parameters are equal against the alternative H_2 that is negation of H_2. Symbolically,

$$H_2 : \omega \in \Omega_0$$

and

$$A_2 : \omega \in \Omega_1 - \Omega_0$$

wherein θ is the nuisance parameter both under H_2 and A_2.

Sukhatme [15] develops likelihood ratio tests for the three problems (i), (ii) and (iii) described above. If λ_0, λ_1 and λ_2 denote the likelihood ratios for H_0, H_1 and H_2 respectively, then Sukhatme computes the moments of $L_0 = \lambda_0^{1/N}$ and $L_1 = \lambda_1^{1/N}$ and obtains the exact distribution of $L_2 = \lambda_2^{1/N}$ where

$N = \sum_{i=1}^{k} n_i$, under the respective null hypotheses. The moments of L_0 and L_1 are used to obtain the Pearsonian approximations to their distributions. It may be mentioned here that Sukhatme [16] provides an analogue of Analysis of Variance technique for testing H_2 against A_2.

The three cases, when $k = 2$, i.e., the two sample analogues of (i), (ii) and (iii), are studied in considerable detail.

The paper ends with a discussion of the application of the tests developed therein to the telephone data mentioned at the beginning of this section and to a data-set relating to accidents to industrial workers.

Sukhatme [15] assumes that all the n_i observations in the i-th sample are completely obtained and that there is no censoring. However, in the industrial context, it was apparent by the early fifties, that one may be forced to deal with censored data. Thus the subsequent work, which I review in the next section, is mainly concerned with statistical inference based on censored samples from the two parameter exponential distribution.

4. Current Status

The results in Sukhatme [15], [16] naturally extend to censored samples of Type II. In most of the life-testing experiments, when n items are put on test, the data naturally arise in an ordered manner; the smallest observation X_1 is available first, followed by the second order statistics and so on. In type II censored samples, we have usually observations on the first r order statistics $X_1, \ldots, X_r$ and the remaining $(n - r)$, $r < n$, observations are known to be greater than X_r. Under the assumption that these observations are available from the two-parameter exponential distribution (2.1), it is well known that X_1 and $S = \sum_{i=1}^{r} X_i + (n - r) X_r - n X_1$ are jointly sufficient for (β, θ). They are independently distributed, X_1 has Exp $(\beta, \theta/r)$ distribution and $2(r - 1) S / \theta$ has chi-square distribution with $2(r - 1)$ degrees of freedom. Thus it is obvious that the distributional properties of the sufficient statistics in the type II censored samples are the same as those in the uncensored case except that role of the sample size n is taken over by the censoring size r.

Suppose one has k independent, censored samples, the censoring size for the i-th sample being r_i; with n_i denoting the sample size, $i = 1, \ldots, k$. Adke [1] considered the three testing situations (i), (ii) and (iii) described in Section 3 for k censored samples. He showed that Sukhatme's [15] results about the distributional properties of the likelihood ratio criteria, continue to remain valid in the type II censored situation with n_i replaced by r_i, $i = 1, \ldots, k$. The argument in the previous paragraph demonstrates the rational of this observation by Adke [1]. In fact this rational also enables one to assert that it is needless to consider censored and uncensored case separately.

It is admitedly difficult to obtain the sampling distributions of the likelihood ratio criteria L_0 and L_1. Several authors have therefore suggested approximations and alternative to them.

Hogg and Tannis [8] suggested an iterative procedure to test H_0 against A_0, which essentially made repeated use of the test suggested by Epstein and Tsao [6] for the two sample problem ($k = 2$).

On the other hand Singh [14] suggests a simple and asymptotically optimal test of the hypothesis H_0 against A_0. His procedure consists of carrying out an F test for the hypothesis H_2 against A_2. This is followed by an analogue of the Bartlett test (cf. Bartlett [4]) for homogeneity of k normal variances, to test the hypothesis H_1 against A_1. These two tests are shown to be independent tests. Since $\Omega_0 \cap \Omega_1 = \Omega_0$, Singh suggests that the Fisher technique of combining two independent tests can be used to test H_0 against A_0. This simple test can be shown to be asymptotically optimal. It may be mentioned here that Singh [14] does not impose any restriction on r_i's and n_i's and that the power performance of this test is satisfactory.

Elwa *et al* [5] use the moments of L_0 computed by Sukhatme [15] to obtain its Mellin transform, which they invert to obtain a infinite series expression for the distribution function of L_0. This infinite series can be used to compute the percentage points of the distribution of L_0.

I have already noted that the test of H_1 against A_1 is similar to testing the homogeneity of k normal variances. This fact suggests the use of the Bartlett test for homogeneity of variances in the present context also. However, the Bartlett statistic has only approximately a chi-square distribution. Nagarsenker and Nagarsenker [11], therefore, adopt the moments of L_1 and invert its Mellin transform to obtain the distribution function of L_1 in a closed form when the sample sizes are all equal.

Thiagarajah and Paul [17] have investigated the problem of testing H_1 against A_1 from the perspective of different methods of constructing the test criteria. They allow arbitrary sample size and differing censoring sizes in the k samples. The following four procedures are adopted by them:

(a) Modified marginal likelihood ratio statistic (MLB)

(b) the quadratic statistics (Qu),

(c) the Neyman C (α) statistic from the profile likelihood function (CPL)

(d) ratio of extremal scale parameter estimates (ESP)

In case (a), Thiagarajah and Paul [17] employ the likelihood ratio principle using the marginal likelihood for $\theta_1, \ldots, \theta_k$. This results in the Bartlett statistic to which they apply the Bartlett correction, thus obtaining the MLB statistic. Let $\hat{\theta}_1, \ldots, \hat{\theta}_k$ be the mle's of $\theta_1, \ldots, \theta_k$ and let $\hat{\theta}$ be the mle of the common scale parameter θ under the null hypothesis H_1.

The Nelson quadratic test (cf. Nelson [12]), employs the statistic

$$Qu = \sum_{i=1}^{k} (r_i - 1) \{(\hat{\theta}_i - \hat{\theta}) / \hat{\theta}\}^2$$

The C (α) statistic obtained from the profile likelihood is

$$CPL = \sum_{i=1}^{k} r_i \{\hat{\theta}_i - \hat{\theta} / \hat{\theta}\}^2$$

All the three statistics described above have asymptotically, under the null hypothesis, the chi-square distribution with $(k - 1)$ degrees of freedom.

The ESP statistic

$$ESP = \max \{\hat{\theta}_1, \ldots, \hat{\theta}_k\} / \min \{\hat{\theta}_1, \ldots, \hat{\theta}_k\}$$

has intuitive appeal but very little is known about its distribution. Based on their simulation studies of the attained level and power of the different tests, Thiagarajah and Paul [17] recommend the MLB test.

The problem of testing the equality of location parameters $\beta_1, \ldots, \beta_k$ when no restrictions are imposed on the scale parameters is a difficult one, being similar to the Beheren-Fisher problem in the context of the Normal distribution. Sukhatme [15], [16] discusses this problem under the restriction that the scale parameters are equal, both under H_2 and A_2. Various authors have suggested alternative procedures. Hsieh [9] modified the likelihood ratio statistic to get rid of the nuisance parameters. More specifically, he shows that

$$-2 \sum_{i=1}^{k} (r_i - 1) \log [\, r_i \hat{\beta}_i / \{r_i \hat{\theta}_i + W_i\}]$$

where $\hat{\theta}_i$ is the mle of θ_i, $\hat{\beta}$ is mle of the common location parameter under the null hypothesis and

$$W_i = n_i \{X_{1i} - \hat{\beta}\}$$

X_{1i} being the smallest observation in the i-th sample $i = 1, \ldots, k$.

Vaughan and Tiku [18] revert to the original problem discussed by Sukhatme [15] of testing H_2 against A_2, but assume that there is left censoring, as well as right censoring. More specifically, they assume that the i-th sample consists of the order statistics

$$X_{ir_i+1}, \ldots, X_{in_i - s_i}$$

the smallest r_i and the largest s_i observations being censored. Since the mle's of the location and scale parameters are not explicitly obtainable, Vaughan and Tiku [18] employ a modification of the likelihood function which yields explicit modified mle's of the parameters. They then adopt the likelihood ratio principle to construct a test statistic which is shown to have asymptotically the chi-square distribution with (k – 1) degrees of freedom. The numerical studies of the power indicate that it is superior to the test proposed by Kambo and Awad [10].

5. *Concluding Remarks*

The exponential distribution occurs as a waiting time distribution in association with a continuous time, discrete state-space Markov process. Thus, for example, in a Markovian M/M/1 queue, the inter-arrival durations of the customers and their service times are both exponentially distributed. The sojourn times in different states of a discrete Markov process are also exponentially distributed. Thus the "interval analysis" propounded by Sukhatme [15], as opposed to the "count analysis", can be effectively employed in many of the standard models in Applied Probability.

I have pointed out in Section 3, that Sukhatme's principal motivation for his [15] paper was the analysis of a series of events. Although he had used the telephone data for purposes of illustration, he had implicitely refered to Poisson process "in space" as opposed to a series of events progressing in time. Thus the natural sequel to the work initiated by Sukhatme would be to investigate the various problems of statistical inference connected with the different generalizations of the homogeneous Poisson process like the non-homogeneous Poisson process, the Poisson process in a general space and the Cox processes. An expository account of the recent literature in this area can be found in Adke [2]. However, as pointed out in that paper, no answer is available to the following easy to formulate questions:

(i) Does the Poisson process (on a general space, or a Cox process) provide an adequate model for the point process under observation?

(ii) How should one compare two or more generally k, k > 2, independent Poisson or Cox processes?

It appears that the solutions provided by Sukhatme in his [15] and [4] papers have generated newer problems which still await solution. This, in my opinion, is the true indication of the pioneering nature of the work of Late Professor P.V. Sukhatme.

ACKNOWLEDGEMENT

I am thankful to the editors of this volume for the opportunity they have given me to pay my tribute to Professor Sukhatme.

REFERENCES

[1] Adke, S.R., 1958. Analysis of k censored samples from exponential population. *J. Univ. Poona, Sc. and Tech. Section*, **12**, 112-122.

[2] Adke, S.R., 1996. Statistical inference for Poisson process and its modifications. In: *Stochastic Processes and Statistical Inference*, Ed. Rao, B.L.S.P. and Bhat, B.R., New Age International Publishers, New Delhi, 1-19.

[3] Bain, L.J. and Englehardt, M., 1991. *Statistical Analysis of Reliability and Life testing Models*. Second Edition, M. Dekker, New York.

[4] Bartlett, M.S., 1937. Properties of sufficiency and statistical tests. *Proceedings of Royal Society London*, **A160**, 268-282.

[5] Elwa, S.A., Nagarsenker, B.N. and Nagarsenker, P.B., 1992. On a test of equality of exponential populations based on censored samples. *Commun. Statist.-Simula.*, **21**, 909-917.

[6] Epstein, B. and Tsao, C.K., 1953. Some tests based on ordered observations from two exponential populations. *Ann. Math. Statist.*, **24**, 458-466.

[7] Epstein, B. and Sobel, M., 1954. Some theorems relevant to life testing from an exponential population. *Ann. Math. Statist.*, **25**, 373-381.

[8] Hogg, R.V. and Tannis, E.A., 1963. An iterated procedure for testing equality of several exponential distributions. *J. Amer. Statist. Assoc.*, **58**, 435-443.

[9] Hsieh, H.K., 1980. An exact test for comparing location parameters of k exponential distributions with unequal scales based on type II censored data. *Tecnometrics*, **28**, 157-164.

[10] Kambo, N.S. and Awad, A.M., 1985. Testing equality of location parameters of k exponential distributions. *Commun. Statist. - Theory and Methods*, **14**, 567-583.

[11] Nagarsenker, P.B. and Nagarsenker, B.N., 1986. On a test of equality of scale parameters of several exponential distributions. *Commun. Statist-Simula.*, **15**, 497-504.

[12] Nelson, W., 1982. *Applied Life Data Analysis*. John Wiley and Sons, New York.

[13] Renyi, A., 1953. On the theory of order statistics. *Acta Math. Acad. Scient. Hungaricae*, **4**, 191-232.

[14] Singh, N., 1985. A simple and asymptotically optimal test for the equality of k ($\geq$ 2) exponential distributions based on type II censored samples. *Commun. Statist.–Theory and Methods*, **14**, 1615-1625.

[15] Sukhatme, P.V., 1936. On the analysis of k samples from exponential populations with especial reference to the problem of random interval. *Statist. Research Memories*, **1**, 94-112.

[16] Sukhatme, P.V., 1937. Tests of significance for samples of the χ^2- population with two degrees of freedom. *Ann. Eugenics*, **8**, 52-56.

[17] Thiagarajah, K. and Paul, S.R., 1990. Testing for quality of scale parameters of k ($\geq$ 2) exponential population based on complete and type II censored samples. *Commun. Statist- Simula.*, **19**, 891-902.

[18] Vaughan, D.C. and Tiku, M.L., 1993. Testing the equality of location parameters of exponential population from censored samples. *Commun. Statist. -Theory and Methods*, **22**, 2567-2581.