IMPROVED TESTS OF EXPONENTIALITY IN SINGLE-AND MULTI-SAMPLE SITUATIONS

BY

M. L. TIKU* AND M. V. TAMHANKAR

McMaster University, Hamilton, Ontario, Canada

(Received : March, 1979)

1. INTRODUCTION

Tiku [14], [15], [16] and Tiku, Rai & Mead [18] developed new statistics, quite different from the well-known Kolmogorov-Smirnov and Shapiro-Wilk statistics, for testing an assumed distribution and, in particular, for testing uniformity, normality, exponentiality and log-normality. The rationale behind these statistics is as follows :

Consider testing the null distribution of the type

$$H_0 : \frac{1}{\sigma} f_0 \left(\frac{x-\mu}{\sigma}\right); \qquad \ldots(1.1)$$

the location and scale parameters μ and σ are not known but the functional form f_0 is completely specified.

Let $$x_1, x_2, \ldots, x_n \qquad \ldots(1.2)$$

be a random sample supposedly from (1.1) and

$$x_1, x_2 \ldots, x_n \qquad \ldots(1.3)$$

be the ordered sample obtained by arranging (1.2) in ascending order of magnitude.

Let $$X_{r_1+1}, X_{r_1+2}, \ldots\ldots, X_{n-r_2} \qquad \ldots(1.4)$$

be the censored sample obtained by censoring (removing) r_1 smallest and r_2 largest observations from (1.3). Let σ_c be the maximum likelihood estimator, or an estimator which is identical, at least asymptotically, to the maximum likelihood estimator (modified maximum likelihood estimator for example, Tiku, [11], [12], [13] and Tiku & Stewart, [19], of σ calculated from the censored sample (1.4),

---

* Most of this work was done when the senior author was on his sabbatical leave at the I.A.R.S. (I.C.A.R.), New Delhi.

and let $\hat{\sigma}$ be the maximum likelihood estimator of σ calculated from the complete sample (1.3), equivalently (1.2). The statistic for testing H_0 is defined as

$$T = h(\sigma_c/\hat{\sigma}), \qquad \ldots(1.5)$$

where h is a constant chosen to make $E(T) \cong 1$. It is easy to show that for fixed $q_1 = r_1/n$ and $q_2 = r_2/n$, the asymptotic null distribution of T is normal (Tiku, [16]). Since $\hat{\sigma}$ is more sensitive to departures from the null distribution H_0 than σ_c due to the presence of the dominating end-observations in $\hat{\sigma}$, the statistic T is sensitive to changes, particularly in the tails, in H_0. The statistic T can, particularly, be used for testing suspected outliers; see Tiku [17].

For testing exponentiality, Tiku, Rai & Mead [18] choose $r_1 = 0$ and $r_2 = [.5 + \frac{1}{2}n]$ which gives T_E, among all the other choices of r_1 and r_2 (*a*) almost the maximum power against a wide class of non-exponential distributions and (*b*) makes its null distribution converge to normality rapidly (effectively for $n > 15$). In this paper, we give a method for improving the statistic T_E for testing exponentiality. The improved statistic is not only more powerful than T_E but converges to normality much faster.

2. Testing Exponentiality

For testing the exponential distribution

$$H_0 : (1/\sigma) \exp\{-(x-\theta)/\sigma\},\ \theta < x < \infty, \qquad \ldots(2.1)$$

the statistic (Tiku, Rai & Mead, 1974)

$$T_E = \sigma_c(r)/\hat{\sigma},\ r = [\cdot 5 + \tfrac{1}{2}n], \qquad \ldots(2.2)$$

where

$$\sigma_c(r) = \left(\sum_{i=1}^{n-r} X_i + rX_{n-r} - nX_1\right) \Big/ (n-r-1),$$

and

$$\hat{\sigma} = \left(\sum_{i=1}^{n} X_i - nX_1\right) \Big/ (n-1),\ \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} x_i\,; \qquad \ldots(2.3)$$

$E(\sigma_c) = \sigma$ and $V(\sigma_c) = \sigma^2/(n-r-1)$. The null distribution of $y = \{(n-r-1)/(n-1)\}T_E$ is Beta distribution $B(a, b)$

$$\{1/\beta(a, b)\}\, y^{a-1}(1-y)^{b-1},\ 0 < y < 1, \qquad \ldots(2.4)$$

with $a = n-r-1$ and $b = r$; see Tiku, Rai & Mead [18]. The null distribution of T_E converges rapidly to a normal distribution with

mean $E(T_E) = 1$ and variance $V(T_E) = r/n(n-r-1)$. ...(2.5)

To improve upon the above statistic we take note of the fact that for each value of $r=1,2,3,\ldots\ldots$, T_E provides a test for exponentiality, although with different values of the power. We, therefore, consider defining a test-statistic in terms of an appropriate linear combination of the estimators $\sigma_c(r)$, $r=1,2,\ldots\ldots$, $n-2$, of σ despite the fact that they are not independent. Consider the linear combination with coefficients inversely proportional to the variances, *i.e.*,

$$Z=\sum_{r=1}^{n-2}(n-r-1)\sigma_c(r)=2\sum_{r=1}^{n-2}(n-r-1)(X_{r+1}-X_1)\ldots.(2.6)$$

Improved Statistics: The improved statistics for testing exponentiality are defined

$$U=\frac{2Z}{\left(\sum_{i=1}^{n}X_i-nX_1\right)(n-2)}=\frac{4\sum_{i=1}^{n-2}(n-i-1)(X_{i+1}-X_1)}{\left(\sum_{i=1}^{n}X_i-nX_1\right)^{n-2}}$$

(θ unknown)...(2·7)

$$U_0=\frac{2Z_0}{\left(\sum_{i=1}^{n}X_i\right)(n-1)}=\frac{4\sum_{i=1}^{n-1}(n-i)X_i}{\left(\sum_{i=1}^{n}X_i\right)(n-1)}\quad(\theta=0);\qquad\ldots(2.8)$$

small and large values indicate non-exponentiality. Note that U is location and scale invariant, and U_0 is scale invariant.

Theorem 1. The distribution of $u=\frac{1}{2}U$ is the same as the distribution of the mean of $(n-2)$ independently and identically distributed (*iid*) uniform (0,1) variates and is given by

$$p(u)=\frac{(n-2)^{n-2}}{(n-3)!}\sum_{r=0}^{z}(-1)^r\binom{n-2}{r}\left(u-\frac{r}{n-2}\right)^{n-3},\quad\ldots(2.9)$$

for $\frac{z}{n-2}\leqslant u\leqslant\frac{z+1}{n-2}$, $Z=0,1,\ldots,n-3$, Equivalently, the cumulative probability function

$$P(u\leqslant c)=\frac{(n-2)^{n-2}}{(n-2)!}\sum_{r=0}^{z}(-1)^{r}\binom{n-2}{r}\left(c-\frac{r}{n-2}\right)^{n-2},$$

$$\text{for } \frac{z}{n-2}\leqslant c\leqslant\frac{z+1}{n-2},\ z=0,1,\ldots\ldots,n-3. \qquad \ldots(2.10)$$

Proof *:* It is easy to show that

$$Z=\sum_{r=1}^{n-2}\sum_{i=1}^{n-r-1}D_i,\quad D_i=(n-i)(X_{i+1}-X_i)$$

$$=\sum_{i=1}^{n-2}(n-i-1)D_i \qquad \ldots(2.11)$$

Therefore,

$$Z\Big/\left(\sum_{i=1}^{n}X_i-nX_1\right)=\sum_{r=1}^{n-2}\sum_{i=1}^{n-r-1}D_i\Big/\sum_{i=1}^{n-1}D_i. \qquad \ldots(2.12)$$

It is well known that $u_i=D_i/\sum_{i=1}^{n-1}D_i$, $i=1,2,\ldots\ldots,n-1$, are jointly distributed as $(n-1)$ spacings generated by $(n-2)$ order statistics of a random sample from uniform (0,1) distribution, $f(u)=1$, $0<u<1$; see Lehmann [3], Seshadri, Csorgo & Stephens [6] or Karlin [1]. But then $\sum_{i=1}^{n-r-1}u_i$ is the $(n-r-1)$th ordered observation in a sample of size $(n-2)$ from the Uniform (0,1), and $\sum_{r=1}^{n-2}\left(\sum_{i=1}^{n-r-1}u_i\right)$ is the sum of the $(n-2)$ ordered observations in a sample of size $(n-2)$. This distribution is the same as the distribution of $(n-2)$ *iid* random variables from Uniform (0,1). The distribution of $u=Z/(\sum_{i=1}^{n}X_i-nX_1)(n-2)$ is, therefore, the same as the distribution of the mean of $(n-2)$ *iid* Uniform (0, 1) given by equation (2.9); see Kendall & Stuart [(2), p. 258].

Theorem 2. The distribution of $u_0=\frac{1}{2}U_0$ is the same as the distribution of the mean of $(n-1)$ *iid* Uniform (0,1) variates.

Proof : The proof follows exactly on the same lines as above. The distribution of u_0 and its cumulative probability function are given by (2.9) and (2.10), respectively, with n replaced by $n+1$.

It is well known that the distribution of the mean of *iid* Uniform (0,1) random variables tends to normality very rapidly. The distribution of U is therefore normal, effectively for $n>5$, with

$$E(U)=1 \text{ and } V(U)=1/3(n-2), \qquad \text{...(2.13)}$$

and the distribution of U_0 is normal, effectively for $n\geqslant 5$, with

$$E(U_0)=1 \text{ and } V(U_0)=1/3(n-1). \qquad \text{...(2.14)}$$

For example for U_0 we have the following values of the exact and approximate values of the percentage points :

Upper % points	Sample Size n							
	5		7		11		21	
	Exact	Approx	Exact	Approx	Exact	Approx.	Exact	Approx.
10	1.376	1.370	1.306	1.302	1.236	1.234	1.166	1.165
5	1.477	1.475	1.389	1.388	1.301	1.300	1.213	1.212
1	1.650	1.672	1.536	1.548	1.419	1.425	1.299	1.300

It is clear that the normal approximation is adequate for $n\geqslant 5$.

Power Comparison : The values of the power of U for several non-exponential distributions were compared with the corresponding values for the statistic T_E (Tiku, Rai & Mead, [8]), and Shapiro & Wilk [8] statistic

$$W=n(\bar{x}-X_1)^2/(n-1)\sum_{i=1}^{n}(x_i-\bar{x})^2. \qquad \text{...(2.15)}$$

The distribution of W is not known but Shapiro & Wilk [8] give simulated percentage points.

The estimated values (based on 2000 Monte Carlo runs) of the power are given in Table I. The sum of the two entries (lower and upper) in the table determines the power of a two-tailed test of size $2\alpha\%$. The statistic U is generally more powerful than W against distributions with skewness $\sqrt{\beta_1}>2$ and slightly less powerful against distributions with $\sqrt{\beta_1}<2$ (values of $\sqrt{\beta_1}$ are given in Shapiro, Wilk & Chen, 1968). However, U is more powerful than T_E.

The statistic U_o is more powerful than T_{E_o} (Tiku, Rai and Mead, [18]) and, on the whole, slightly more powerful than Kolmogorov-Smirnov type statistics $\tilde{D}$ and $\hat{D}$; see Lilliefors [4], Srinivasan [9] and Schafer, Finkelstein & Collins [5]. For example, we have the following values of the power for a two-tailed test of size 5% :

Alternative Distribution	$n=10$				$n=20$			
	$\overleftarrow{D}_n$	$\hat{D}_n$	T_{E_o}	U_o	$\overleftarrow{D}_n$	$\hat{D}_n$	TE_o	U_o
Log normal, $\sigma=\cdot4$	0.90	0.95	0.64	0.94	1.00	0.98	0.96	1.00
$\sigma=.8$	.12	.16	.09	.17	.25	.30	.11	.26
$\sigma=2.0$	.67	.61	.64	.69	.90	.89	.91	.93
$\sigma=2.4$	.81	.77	.80	.83	.97	.97	.97	.98
Chi-square ($\nu=1$)	.30	.25	.31	.33	.48	.44	.49	.57

3. Multi-Sample Generalised Statistics

Let

$$x_{j1},\ x_{j2},\ldots\ldots,x_{jn_j}\ ,\ j=1,\ 2,\ldots\ldots\ ,k,$$

be k independent random samples, and

$$X_{j1},\ X_{j2},\ldots\ldots,X_{jn_j}\ ,\ j=1,\ 2,\ldots\ldots,\ k, \qquad \ldots(3.1)$$

be the corresponding k ordered samples. We want to test that these samples come from k exponential populations

$$(1/\sigma_j)\ \exp\{-(x-\theta_j)/\sigma_j\},\ \theta_j<x<\infty\ (j=1,\ 2,\ldots\ldots,\ k). \qquad \ldots(3.2)$$

Case I : Por θ_j's not known and σ_j's not equal, the generalized statistic is

$$U^{**}=\{1/(N-2k)\}\sum_{j=1}^{k}(n_i-2)\ U_j,\ N=\sum_{i=1}^{k}n_i, \qquad \ldots(3.3)$$

where U_j is the statistic U (equation, 2.7) calculated from the jth sample, $j=1, 2,\ldots, k$. The null distribution of $\frac{1}{2}U^{**}$ is given by (2.9) with $(n-2)$ replaced by $(N-2k)$. For $n_i\geqslant5$ and $k\geqslant2$, this distribution is very closely approximated by a normal distribution with

$$E(U^{**})=1 \text{ and } V(U^{**})=1/3(N-2k). \qquad \ldots(3.4)$$

The corresponding generalized T_E statistic is of the type

$$T_E^{**}=(1/k)\sum_{j=1}^{k}T_E(j),$$

but the exact distribution of T_E^{**} is not known ; see Tiku [15].

TABLE I

Values of the Power of W, T_E and U for Lower and Upper α Percentage Points

Alternative Distribution	α	n	*W*		T_E		*U*	
			Lower	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*
Chi-square (ν=1)	2.5	10	0.17	0.00	0.22	0.00	0.22	0.00
		20	.28	.00	.39	.00	.49	.00
	5	10	.26	.01	.31	.01	.31	.01
		20	.39	.00	.49	.00	.59	.00
Chi-square (ν=3)	2.5	10	.01	.07	.01	.05	.01	.06
		20	.00	.10	.00	.08	.00	.10
	5	10	.01	.12	.02	.09	.02	.11
		20	.01	.18	.01	.14	.01	.19
Chi-square (ν=4)	2.5	10	.00	.12	.01	.07	.00	.10
		20	.00	.22	.00	.14	.00	.20
	5	10	.01	.17	.01	.12	.01	.17
		20	.00	.31	.00	.23	.00	.34
Weibull (c=½)	2.5	10	.43	.00	.49	.00	.53	.00
		20	.73	.00	.80	.00	.86	.00
	5	10	.54	.00	.60	.00	.63	.00
		20	.82	.00	.87	.00	.91	.00
Weibull (c=2)	2.5	10	.00	.26	.00	.16	.00	.24
		20	.00	.63	.00	.49	.00	.63
	5	10	.00	.38	.00	.27	.00	.36
		20	.00	.75	.00	.61	.00	.78
Beta (2,1)	2.5	10	.00	.72	.00	.56	.00	.67
		20	.00	.98	.00	.93	.00	.98
	5	10	.00	.82	.00	.70	.00	.78
		20	.00	1.00	.00	.97	.00	.99
Half-normal	2.5	10	.00	.11	.00	.08	.00	.10
		20	.00	.21	.00	.15	.00	.17
	5	10	.00	.18	.01	.14	.01	.17
		20	.00	.33	.00	.25	.00	.30
Half-Cauchy	2.5	10	.40	.00	.30	.00	.41	.00
		20	.68	.00	.56	.00	.67	.00
	5	10	.47	.01	.39	.01	.48	.01
		20	.74	.00	.64	.00	.74	.00
Log-normal (σ=2.4)	2.5	10	.67	.00	.72	.00	.74	.00
		20	.93	.00	.96	.00	.98	.00
	5	10	.77	.00	.81	.00	.81	.00
		20	.96	.00	.97	.00	.99	.00

Case II : For θ_j's known (say $\theta_j=0, j=1, 2, \; , k$) and σ_j's not equal, the generalized statistic is

$$U_0^{**}=\{1/(N-k)\} \sum_{j=1}^{k}(n_j-1)U_{j0},\ N= \sum_{i=1}^{k} n_i, \quad \ldots(3.5)$$

where U_{j_0} is the statistic U_0 (equation, 2.8) calculated from the jth ($j=1, 2,\ldots, k$) sample (3.1). The null distribution of $\frac{1}{2}U_0^{**}$ is given by (2.9) with $(n-1)$ replaced by $(N-k)$. For $n_j \geqslant 4$ and $k \geqslant 2$, this distribution is adequately represented by normal with

$$E(U_0^{**})=1 \text{ and } V(U_0^{**})=1/3(N-k). \quad \ldots(3.6)$$

The exact distribution of the corresponding T_E statistic, $T_{E_0}^{**}$ is not known.

The above results seem to be very interesting because U^{**} and U_0^{**} are the only generalized statistics known so far whose exact distributions are worked out ; see for example Wilk & Shapiro [20] and Tiku [15].

For the other two cases, that is, θ_j's unknown and σ_j's equal, and θ_j's known (=0 say) and σ_j's equal, the generalized U statistics involve ordering $N-k$ and N observations, respectively, that is, ordering $X_{ji}-X_{j1}$, and X_{ji}, $i=1, 2,\ldots,n_j, j=1,2\ldots, k$, and calculating statistics similar to U and U_0. Since N will, in practice, be large this prospect does not seem to be attractive. However, the corresponding generalized T statistics, T_E^* and $T_{E_0}^*$, do not involve such ordering and are easy to calculate, and their null distributions are Beta distributions.

The above generalized statistics have excellent power properties as is clear from the following table (Table II).

The T-statistic (eq. 1.5) for testing other distributions, in particular, testing uniformity, normality and log-normality, can be improved on the same lines as above. For example the improved T statistic (Tiku, 1975, p. 116) for testing uniformity $U(\theta_1,\theta_2)$ is

$$U=2 \sum_{i=1}^{n-2} (X_{i+1}-X_1)/(X_n-X_1)(n-2); \quad \ldots(3.7)$$

the distribution of $u=\frac{1}{2}U$ is the same as (2.9). However, the distributions of the improved statistics for testing normality and log normality are difficult to work out.

It seems posssible to define statistics similar to T_E^{**} and U^{**} to test bivariate exponentiality. The statistics T_E and U can also be generalized to test exponentiality from censored samples. This is under investigation at the present time.

TABLE 2

Values of the Power of U for Lower and Upper α Percentage Points**

Alternative Distribution	α	$K=2$				$k=4$	
		$n_1=n_2=10$		$n_1=n_2=20$		$n_1=n_2=n_3=n_4=10$	
		Lower	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*
Chi-square	2.5	0.34	0.00	0.70	0.00	0.57	0.00
($\nu=1$)	5	.45	.00	.79	.00	.68	.00
Chi-square	2.5	.00	.09	.00	.17	.00	.15
($\nu=3$)	5	.01	.15	.00	.27	.00	.23
Chi-square	2.5	.00	.16	.00	.39	.00	.30
($\nu=4$)	5	.00	.26	.00	.52	.00	.43
Weibull	2.5	.78	.00	.98	.00	.96	.00
($c=\frac{1}{2}$)	5	.84	.00	.99	.00	.97	.00
Weibull	2.5	.00	.45	.00	.92	.00	.78
($c=2$)	5	.00	.59	.00	.96	.00	.87
Log-Normal	2.5	.00	.30	.00	.76	.00	.55
($\sigma=.4$)	5	.00	.44	.00	.85	.00	.68
Log-normal	2.5	.84	.00	1.00	.00	.98	.00
($\sigma=2$)	5	.89	.00	1.00	.00	.99	.00
Log-normal	2.5	.93	.00	1.00	.00	1.00	.00
($\sigma=2.4$)	5	.96	.00	1.00	.00	1.00	.00
Beta (2,1)	2.5	.00	.94	.00	1.00	.00	1.00
	5	.00	.97	.00	1.00	.00	1.00

SUMMARY

A method of improving Tiku's [14] to [17] goodness-of-fit statistics is described. Improved statistics for testing exponentiality in single-and multi-sample situations are given explicitly and their exact distributions obtained. The improved statistics for testing exponentiality are shown to be more powerful than Tiku statistics and slightly more powerful than kolmogorov-Smirnov type statistics. For testing the exponentiality of a single sample, the improved statistic is shown to be generally more powerful than Shapiro-Wilk against distributions with skewness $\sqrt{\beta_1}>2$ and slightly less powerful against distributions with $\sqrt{\beta_1}<2$ ($\sqrt{\beta_1}=2$ for an exponential distribution). A very interesting feature of this paper is that generalized statistics for testing exponentiality of k independent samples with unequal population variances are defined and the exact distributions of these statistics are worked out. These exact distributions are shown to be the same as the distribution of the mean of independently and identically distributed Uniform (0, 1) variates.

ACKNOWLEDGEMENTS

Thanks are due to the referees for helpful comments. Thanks are also due to the National Research Council of Canada for a research grant and to Mrs. *E.* Moore for the typing.

REFERENCES

[1] Karlin, S. (1966) : *A First Course In Stochastic Processes*. New York : Academic Press.

[2] Kendall, M.G. & Stuart, A. (1958) : *The Advanced Theory of Statistics*, Vol. I. London : Charles Griffin and Company Ltd.

[3] Lehmann, E.L. (1966) : *Testing Statistical Hypotheses*. New York : John Wiley and Sons, Inc.

[4] Lilliefors, H.W. (1969) : On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. *J. Amer. Statist. Assoc.* **64**, 387-89.

[5] Schafer, R.E., Finkelstein, J. M. & Collins, J. (1972) : On a goodness-of-fit test for the exponential distribution with mean unknown. *Biometrika* 59, 222-4.

[6] Seshadri, V., Csorgo, M. & Stephens, M.A (1969) : Test for the exponential distribution using Kolmogorov-type statistics. *J.R. Statist. Soc. B*, **31**, 499-509.

[7] Shapiro, S.S., Wilk, M.B. & Chen, H.J. (1968) : A comparative study of various tests for normality. *J. Amer Statist. Assoc.* **63**, 1343-72.

[8] Shapiro, S.S. & Wilk, M.B. (1972) : An analysis of variance test for exponential distribution. *Technometrics* **14**, 355-70.

[9] Srinivasan, R. (1970) : An approach to testing the goodness-of-fit of incompletely specified distributions. *Blometrika* **57**, 605-11.

[10] Tiku, M.L. (1967) : Estimating the mean and standard deviation from a censored normal sample. *Biometrika* **54**, 155-65.

[11] Tiku, M.L. (1967) : A note on estimating the location and scale parameters of the exponential distribution from a censored sample. *Aust. J. Statist.* **9**, 49-54.

[12] Tiku, M.L. (1968) : Estimating the parameters of log-normal distribution from censored samples. *J. Amer Statist. Assoc.* **63**, 134-40.

[13] Tiku, M.L. (1973) : Testing group effects from type II censored normal samples in experimental design. *Biometrics*, **29**, 25-33.

[14] Tiku, M.L. (1974a) : A new statistic for testing for normality. *Commun. Statist.* **3** 3 ,223-32.

[15] Tiku, M.L. (1974b) : Testing normality and exponentiality in multi-sample situations. *Commun Statist.* 3 8, 777-94.

[16] Tiku, M L. (1975a) : A new statistic for testing an assumed distribution. *Model Building and Model Selection*, Vol.II of Statistical Distributions in Scientific Work, (Ed. G.P. Patil, S. Kotz and J.K. Ord). Boston : D. Reidel Publ. Co., pp. 113-24.

[17] Tiku, M.L. (1975b) : A new statistic for testing suspected outliers. *Commun. Statist.* 4 8, 737-52.

[18] Tiku, M.L., Rai, K. & Mead, E. (1974) : A new statistic for testing exponentiality. *Commun. Statist.* 3 5, 485-93.

[19] Tiku M.L. & Stewart, D.E. (1977) Estimating and testing group effects from type I censored normal samples in experimental design. *Comm. Statist.* A6 (15), 1485-1501.

[20] Wilk, M.B. & Shapiro, S S. (1968) : The joint assessment of normality of several independent samples. *Technometrics* **10**. 825-39.