EFFECT OF TRUNCATION ON A TEST FOR THE SCALE PARAMETER OF THE EXPONENTIAL DISTRIBUTION

By A. P. Basu[1]

Rutgers—The State University and Bell Telephone Laboratories

1. Summary. In this paper we have investigated the effects on the operating characteristics of a test for the scale parameter of the exponential distribution on the assumption that the sample is from a "complete" exponential population when in reality it is known to have come from a truncated exponential population. The results derived are valid for samples of any size.

2. Introduction. In many physical situations, for example, in problems of life testing (see [2]), the exponential distribution is proposed as a useful statistical model. But often the assumption that a random variable X can take on values on $(0, \infty)$ is not a realistic one. In such cases our inference might be more usefully based on the model of a truncated exponential distribution. In this note we shall study the operating characteristics of a test appropriate for the scale parameter of the exponential distribution when in reality the sample actually comes from a truncated population.

In studying this we shall follow the same line as has been done by Aggarwal and Guttman [1] in case of the truncated normal distribution. In their paper Aggarwal and Guttman considered a "symmetrically truncated" normal distribution as the "real" model alternative to the usual normal distribution and studied the "loss in power" due to wrong selection of model.

The probability density function (p.d.f.) of the exponential distribution is

$$(1)\qquad \begin{aligned} f(x) &= \theta^{-1}e^{-x/\theta}, && \text{for } 0 \leqq X < \infty, \\ &= 0, && \text{otherwise, } \theta > 0. \end{aligned}$$

If the population is a truncated one, it is said to be truncated to the right with its terminus point x_0 if its p.d.f. is given by

$$(2)\qquad \begin{aligned} f^*(x) &= k\theta^{-1}e^{-x/\theta} && \text{for } 0 \leqq X \leqq x_0, \\ &= 0 && \text{otherwise,} \end{aligned}$$

where

$$(3)\qquad k^{-1} = (1 - e^{-x_0/\theta}).$$

3. Distribution of sample means. Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a population with the p.d.f. (2). We shall try to find the distribution of

$$\bar{X} = n^{-1}\sum_{j=1}^{n} X_j .$$

Received 4 February 1963; revised 10 September 1963.

[1] Now at the University of Minnesota.

Now the characteristic function of a random variable X having density $f^*(x)$ is given by

(4) $$\varphi^*(t; x) = k\{1 - \exp - (x_0/\theta)(1 - i\theta t)\}/(1 - i\theta t).$$

Therefore, the characteristic function $\varphi^*(t; y_n)$ of the sum $Y_n = X_1 + X_2 + \cdots + X_n$ is given by

(5) $$\varphi^*(t; y_n) = \sum_{r=0}^{n} b_r e^{itrx_0}(1 - i\theta t)^{-n},$$

where

$$b_r = (-1)^r \binom{n}{r} k^n e^{-rx_0/\theta}.$$

We are now to find the distribution $S^*(y_n)$ of Y_n. By the inversion formula

(6) $$\begin{aligned} S^*(y_n) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1 - e^{-ity_n}}{it} \varphi^*(t; y_n)\, dt \\ &= \sum_{r=0}^{n} b_r \int_{-\infty}^{\infty} \frac{1 - e^{-i\tau z_n}}{i\tau} e^{i\tau(rx_0/\theta)}(1 - i\tau)^{-n}\, d\tau, \end{aligned}$$

where $\tau = t\theta$ and $Z_n = Y_n/\theta$.

Now the characteristic function of the Gamma distribution

$$G_n(x) = \int_0^x \frac{1}{\Gamma(n)} e^{-y} y^{n-1}\, dy$$

is given by $\psi_n(\tau) = (1 - i\tau)^{-n}$. The integrand of the rth term of (6) is seen to be

$$[(1 - e^{-i\tau z_n})/i\tau]\psi_n(\tau)e^{i\tau(rx_0/\theta)} = [(1 - e^{-i\tau z_n})/i\tau]\lambda_r(\tau),$$

where $\lambda_r(\tau) = e^{i\tau(rx_0/\theta)}\psi_n(\tau)$ is obviously the characteristic function corresponding to the distribution $G_n(z_n - rx_0/\theta)$. Hence the distribution of Y_n is given by

(7) $$S^*(y_n) = \sum_{r=0}^{n} b_r G_n((y_n - rx_0)/\theta),$$

where $G_n(y_n - rx_0) = 0$ for $y_n < rx_0$.
From (7) the distribution of $\bar{X}$ follows. Let it be $F_n^*(\bar{x} \mid \theta)$.

When the X_j's follow (1), $(j = 1, 2, \cdots, n)$, $n\bar{X}/\theta$ follows the gamma distribution with parameter n and the distribution of $\bar{X}$ is given by

(8) $$dF_n(\bar{x} \mid \theta) = f_n(\bar{x} \mid \theta)\, d\bar{x} = (1/\Gamma(n))e^{-n\bar{x}/\theta}(n\bar{x}/\theta)^{n-1}\, d(n\bar{x}/\theta).$$

It is clear that (8) gives the first term in the expansion (7) multiplied by k^{-n}.

4. Tests of hypothesis under truncation. Given a sample of size n from either (1) or (2), to test the one-sided hypothesis on the scale parameter

(9) $$H: \theta = \theta_0, \qquad \text{Alt } H: \theta < \theta_0$$

we consider the tests T_1, T_2 : Reject H if $\bar{X} < \omega_i$ $(i = 1, 2)$, accept otherwise; where

$$\int_0^{\omega_1} f_n(\bar{x} \mid \theta_0)\, d\bar{x} = \int_0^{\omega_2} f_n^*(\bar{x} \mid \theta_0)\, d\bar{x} = \alpha \tag{10}$$

and α is the stipulated size of the critical region. T_1 is the uniformly most power-

TABLE I

Values of α' and $L(\theta_0, x_0)$ expressed as a percentage of $\alpha = 0.05$

n	1		2		3		4		5	
	α'	$-L$	α'	$-L$	α'	$-L$	α'	$-L$	α'	$-L$
$x_0 = 1.0$, $\theta_0 = 0.5$	.0578	15.6	.0668	33.6	.0772	54.4	.0893	78.6	.1033	106.6
$x_0 = 5.0$, $\theta_0 = 3.0$	.0616	23.2	.0760	52.0	.0937	87.4	.1155	131.0	.1424	184.4

TABLE II

Lower 100α% points of $f_n^*(\bar{x}|\theta)$

x_0	θ_0	n	α = .005	.010	.025	.05	.10
1.0	0.5	1	.0023	.0045	.0114	.0227	.0454
		2	.0145	.0290	.0494	.0749	.1115
		3	.0444	.0608	.0870	.1139	.1513
		4	.0695	.0848	.1124	.1402	.1772
		5	.0903	.1056	.1342	.1603	.1945
5.0	3.0	1	.0128	.0256	.0639	.1278	.2557
		2	.0766	.1533	.2768	.4216	.6220
		3	.2451	.3457	.4774	.6270	.8372
		4	.3806	.4754	.6264	.7769	.9715
		5	.4933	.5834	.7344	.8805	1.0685
10.0	5.0	1	.0227	.0454	.1136	.2272	.4543
		2	.1452	.2903	.4942	.7494	1.1154
		3	.4440	.6077	.8701	1.1389	1.5126
		4	.6955	.8475	1.1235	1.4016	1.7716
		5	.9029	1.0561	1.3419	1.6034	1.9434
25.0	15.0	1	.0639	.1279	.3196	.6393	1.2785
		2	.3832	.7664	1.3839	2.1079	3.1098
		3	1.2255	1.7285	2.3868	3.1352	4.1859
		4	1.9032	2.3772	3.1318	3.8847	4.8577
		5	2.4666	2.9168	3.6719	4.4024	5.3424
50.0	30.0	1	.1279	.2557	.6393	1.2785	2.5570
		2	.7664	1.5328	2.7678	4.2158	6.2195
		3	2.4511	3.4570	4.7736	6.2704	8.3718
		4	3.8064	4.7544	6.2637	7.7694	9.7153
		5	4.9332	5.8337	7.3439	8.8047	10.6848

ful (UMP) test of H if X follows (1) and will be called the usual test. T_2 is the UMP test of H if X follows (2) and ω_2 is a function of x_0, the point of truncation.

It is useful to define three power functions,

$$P_u(\theta, x_0) = \int_0^{\omega_1} f_n(\bar{x} \mid \theta)\, d\bar{x} \tag{11}$$
$$= \text{Power of usual test } (T_1) \text{ when } X \text{ is untruncated,}$$

$$P(\theta, x_0) = \int_0^{\omega_1} f_n^*(\bar{x} \mid \theta)\, d\bar{x} \tag{12}$$
$$= \text{Power of usual test } (T_1) \text{ when } X \text{ truncated,}$$

and

$$P_c(\theta, x_0) = \int_0^{\omega_2} f_n^*(\bar{x} \mid \theta)\, d\bar{x} \tag{13}$$
$$= \text{Power of test } (T_2) \text{ when } X \text{ is truncated.}$$

If in fact X is truncated but the experimenter does not know this and uses the usual test, the actual size of the test differs from α. It is, of course, $P(\theta_0, x_0)$ which may be labelled α'.

Let

$$E(\theta, x_0) = P_u(\theta, x_0) - P(\theta, x_0). \tag{14}$$

$E(\theta, x_0)$ is the "error" incurred if the usual procedure is followed while sampling is actually from a truncated distribution. For $\theta = \theta_0$, $E(\theta_0, x_0) = \alpha - \alpha'$. In

TABLE III

Values of $P_u(\theta, x_0)$, $P(\theta, x_0)$, $P_c(\theta, x_0)$ for $\alpha = 0.05$ and $\theta < \theta_0$

n	(x_0, θ_0)	$x_0 = 1.0, \theta_0 = 0.5$		$x_0 = 5.0, \theta_0 = 3.0$				$x_0 = 10.0, \theta_0 = 5.0$		$x_0 = 25.0, \theta_0 = 15.0$		$x_0 = 50.0, \theta_0 = 30.0$
	θ	0.2	0.4	0.5	1.0	1.5	2.0	1.0	3.0	2.0	13.0	25.0
1	P_u	.1222	.0625	.2699	.1448	.0995	.0750	.2301	.0833	.3255	.0577	.0600
	P	.1230	.0681	.2699	.1458	.1032	.0817	.2301	.0864	.3255	.0676	.0694
	P_c	.1075	.0588	.2247	.1199	.0841	.0662	.2024	.0747	.2730	.0548	.0566
2	P_u	.2192	.0724	.6214	.2834	.1564	.0988	.5230	.1171	.7382	.0633	.0673
	P	.2222	.0859	.6215	.2873	.1682	.1173	.5230	.1259	.7382	.0868	.0900
	P_c	.1758	.0663	.5024	.2096	.1180	.0801	.4415	.0976	.6223	.0595	.0625
3	P_u	.3321	.0836	.8646	.4404	.2236	.1248	.7710	.1560	.9422	.0700	.0765
	P	.3389	.1081	.8647	.4494	.2493	.1614	.7711	.1740	.9422	.1124	.1183
	P_c	.2501	.0720	.7248	.2972	.1481	.0910	.6633	.1205	.8479	.0602	.0644
4	P_u	.4431	.0942	.9677	.5833	.2916	.1512	.9079	.1951	.9912	.0753	.0841
	P	.4552	.1327	.9679	.5993	.3372	.2130	.9081	.2256	.9912	.1417	.1504
	P_c	.3173	.0759	.8672	.3872	.1808	.1028	.8101	.1392	.9505	.0628	.0680
5	P_u	.5444	.1032	.9912	.7008	.3579	.1764	.9673	.2334	.9990	.0804	.0913
	P	.5631	.1584	.9914	.7248	.4291	.2705	.9675	.2799	.9990	.1771	.1887
	P_c	.3391	.0815	.9382	.4646	.2088	.1119	.9014	.1596	.9849	.0654	.0701

Table I we have tabulated a few values of α' for different sample sizes and $x_0 = 1.0, 5.0$ when $\alpha = .05$.

Table II gives the significance points for the test T_2 for different values of n, α and x_0. Thus, if sampling is from a truncated exponential distribution, T_2 gives the correct test of (9) and Table II gives the correct significance points for this problem. Next we prove the following lemma useful in reducing subsequent computational labor.

LEMMA. *Given n and triples (x_0, θ_0, θ), $\theta < \theta_0$,* (a) *$P(\theta, x_0)$ is a constant if (for each combination of x_0, θ_0 and θ) both θ_0/θ and x_0/θ are constants; and* (b) *$P_u(\theta, x_0)$ and $P_c(\theta, x_0)$ are constants if θ_0/θ is a constant.*

PROOF. (a) From (8) and (10), for fixed n, $n\omega_1/\theta_0 = m$ (say, where m is a constant). Therefore,

$$P(\theta, x_0) = \sum_{r=0}^{n} b_r G_n\left(\frac{n\omega_1 - rx_0}{\theta}\right) = \sum_{r=0}^{n} b_r G_n\left(\frac{m\theta_0}{\theta} - r\frac{x_0}{\theta}\right). \tag{15}$$

From (15) the result follows. (b) can be proved similarly.

A few typical values of $P_u(\theta, x_0)$, $P(\theta, x_0)$ and $P_c(\theta, x_0)$ have been computed for different values of $\theta < \theta_0$, and the five terminus points $x_0 = 1.0, 5.0, 10.0, 25.0$ and 50.0 in Table III for $\alpha = 0.05$ and $n = 1(1)5$. The quantities $P_u(\theta, x_0)$, $P(\theta, x_0)$ and $P_c(\theta, x_0)$ were computed systematically with the help of Pearson's Tables of the incomplete gamma function [3].

5. Conclusion. In Table I, it is found that $\alpha' > \alpha$ everywhere whereas in case of Aggarwal and Guttman [1] $\alpha' < \alpha$. In general, it is noted that

$$P(\theta, x_0) \geqq P_u(\theta, x_0) \geqq P_c(\theta, x_0).$$

This can also be shown directly. This difference between our case and that of Aggarwal and Guttman is because they consider alternatives to the right ($\theta > \theta_0$) while we consider alternatives to the left ($\theta < \theta_0$). We also note that serious errors occur both in the power and size of the tests especially when the sample size is large. So one must be very careful on deciding the type of truncation to be used in a particular situation.

The computational procedure, though tedious for large n, is straightforward and can be carried out systematically for any value of n. However, in this paper we consider only values of n up to five as the general tendency is quite evident even in such small samples.

6. Acknowledgments. The author is grateful to Dr. M. B. Wilk, Mr. J. K. Ghosh, Dr. R. S. Pinkham, Professor D. G. Chapman and a referee for their valuable suggestions.

REFERENCES

[1] AGGARWAL, O. P. and GUTTMAN, I. (1959). Truncation and tests of hypothesis. *Ann. Math. Statist.* **30** 230–238.

[2] EPSTEIN, B. and SOBEL, M. (1953). Life testing. *J. Amer. Statist. Assoc.* **48** 486–502.

[3] PEARSON, K. *Tables of the Incomplete Gamma Function.* Cambridge Univ. Press.