ESTIMATION OF PARAMETERS OF TRUNCATED OR CENSORED EXPONENTIAL DISTRIBUTIONS

By Walter L. Deemer, Jr. and David F. Votaw, Jr.

United States Air Force and Yale University

1. Summary. This paper gives maximum likelihood estimators of parameters of truncated and censored exponential distributions, asymptotic variances of the estimators, and asymptotic confidence intervals for the parameters.

Applications to bombing accuracy studies and to life testing are pointed out. As regards bombing accuracy the parameter estimated is the reciprocal of the variance in a normal bivariate distribution having circular symmetry. The reciprocal is estimated because there is no maximum likelihood estimator of the variance and any estimator of the variance is badly biased (see Section 2).

Results of a synthetic sampling experiment are given to provide information on rapidity of convergence of the distributions of the estimators to their asymptotic distributions.

2. Introduction. In bombing accuracy studies and in other aiming accuracy studies, the assumption is often made that aiming errors (range and deflection errors in bombing; azimuth and elevation errors in gunnery) have a bivariate normal distribution with mean at the aiming point, zero correlation and equal variances.

Under these assumptions the radial error, or distance from the aiming point to the point of impact, is a chance quantity say R with probability density function

$$k(r) = r\sigma^{-2} \exp [-r^2/(2\sigma^2)], \qquad 0 < r < \infty. \tag{2.1}$$

Let $\frac{1}{2}R^2 = Z$, say, and denote σ^{-2} by c. The density, say $h(z)$, of Z is

$$h(z) = ce^{-cz}, \qquad 0 < z < \infty; c > 0; \tag{2.2}$$

thus Z has an exponential distribution.

In some situations values of Z greater than a fixed value cannot be observed. For example, in gun camera missions the view angle of the camera defines the maximum observable R (and thus the maximum observable Z). An example arises in life testing from an exponential distribution when the time of testing is fixed in advance (see [3], pp. 4–9). (Cases in which the time of testing is determined by a sample are treated in [1], [3], [4], and [6], p. 416.)

Before proceeding with the estimation in truncated and censored cases let us consider estimation[1] of c in (2.2) on the basis of a sample $Z_1, Z_2, \cdots, Z_N$

Received November 4, 1953; revised November 20, 1954.

[1] For estimation of $\sigma (= c^{-1/2})$ when the observations are grouped see [5].

of values of Z. The likelihood function, $L(c)$, of c is

$$L(c) = [ce^{-c\bar{z}}]^N \tag{2.3}$$

where $\bar{z}$ is the sample mean. The value, say $\hat{c}$, of c for which $L(c)$ assumes its maximum value is

$$\hat{c} = (\bar{z})^{-1}, \text{ the maximum likelihood estimator of } c. \tag{2.4}$$

The estimator $\hat{c}$ has a finite mean if $N \geqq 2$, and a finite variance if $N \geqq 3$.

It is well known that $2Nc/\hat{c}$ has a chi-square distribution with $2N$ degrees of freedom. Equation (2.4) is equivalent to the well-known result that the maximum likelihood estimator, say $\hat{\sigma}^2$, of σ^2 is

$$\hat{\sigma}^2 = \sum_{i=1}^{N} r_i^2/2N. \tag{2.5}$$

The asymptotic variance of $(N)^{1/2}$ $(\hat{c} - c)$ is

$$[- E(\partial^2 \log h(z)/\partial c^2)]^{-1}.$$

From (2.2) we have that this equals c^2; therefore, for large N

$$\text{Variance } [(N)^{1/2} (\hat{c} - c)] = c^2. \tag{2.6}$$

Derivations of the asymptotic variance of a maximum likelihood estimator are given in [6], pp. 208–212, and [7], pp. 136–139.

When the distribution is truncated or censored, we shall replace Z by X and denote by x_0 the maximum value of X that can be observed. It is assumed that x_0 is known in advance. The two cases will now be described.

Case A (*Censored*[2] *Distribution*). Here the number of observations greater than x_0 is known. When $Z \leqq x_0$, $X = Z$; when $Z > x_0$, the only information obtained about X is simply that $X > x_0$. X can be regarded as having a density, say $g(x)$, when $X \leqq x_0$; thus

$$\begin{aligned} g(x) &= ce^{-cx}, \qquad 0 < x \leqq x_0, \\ \Pr(X > x_0) &= e^{-cx_0}. \end{aligned} \tag{2.7}$$

Case B (*Truncated Distribution*). Here the number of observations greater than x_0 is unknown. X has a density, say $f(x)$, which is the conditional density of Z given that $Z \leqq x_0$; thus

$$f(x) = ce^{-cx} (1 - e^{-cx_0})^{-1}, \qquad 0 < x \leqq x_0. \tag{2.8}$$

The maximum likelihood estimator of c will be derived for Case A and for Case B. It is noteworthy that in each case no maximum likelihood estimator of σ^2 $(= c^{-1})$ exists and the bias of any estimator of σ^2 tends to $-\infty$ as σ^2 tends to $+\infty$. For this reason the quantity c instead of c^{-1} is chosen as the parameter to be estimated.

---

[2] For further discussion of censored and truncated distributions see [2], p. 144.

3. Maximum-likelihood estimators. For Case A let n be the number of observations of X such that $X \leqq x_0$ and let m be the number of observations such that $X > x_0$. Let $N = m + n$. The likelihood function, say $L_A(c)$, of c is (see (2.7)),

$$L_A(c) = \begin{cases} N!\,[m!\,n!]^{-1}\, c^n \exp\left[-c \sum_1^n x_i - mcx_0\right], & n > 0 \\ e^{-Ncx_0}, & n = 0. \end{cases} \tag{3.1}$$

(It should be noted that this is the likelihood function of a chance quantity having the density given in (2.7) and a probability e^{-cx_0} of taking the value x_0. Halperin [3], pp. 4–9, has proved that the maximum likelihood estimator of c in this mixed continuous discrete case has the properties of consistency, asymptotic normality, and minimum asymptotic variance.)

The maximum-likelihood estimator, say $\hat{c}_A$, of c is

$$\hat{c}_A = n\left[mx_0 + \sum_1^n x_i\right]^{-1}. \tag{3.2}$$

$\hat{c}_A$ has a finite mean if $N \geqq 2$ and a finite variance if $N \geqq 3$.

For Case B let the sample be $X_1, \cdots, X_{n'}$. The likelihood function, say $L_B(c)$, is (see (2.8)),

$$\begin{aligned} L_B(c) &= c^{n'}(1 - e^{-cx_0})^{-n'} \exp\left[-c\sum_1^{n'} x_i\right] \\ &= [ce^{-c\bar{x}}(1 - e^{-cx_0})^{-1}]^{n'}, \end{aligned} \tag{3.3}$$

where $\bar{x}$ is the sample mean. It follows that

$$\partial \log L_B(c)/\partial c = n'[c^{-1} - x_0 e^{-cx_0}(1 - e^{-cx_0})^{-1} - \bar{x}]. \tag{3.4}$$

It can be shown that the function $c^{-1} - x_0e^{-cx_0}(1 - e^{-cx_0})^{-1}$ is monotonic decreasing in c; as c tends to 0 the function tends to $\frac{1}{2}x_0$, and as c tends to infinity the function tends to 0. When $0 < \bar{x} < \frac{1}{2}x_0$, there exists a solution, say c', of the equation formed by setting $\partial \log L_B(c)/\partial c$ equal to 0 (see (3.4)). Clearly c' is the maximum likelihood estimator of c when $0 < \bar{x} < \frac{1}{2}x_0$. When $\bar{x} \geqq \frac{1}{2}x_0$, the function $L_B(c)$ assumes its maximum value for $c = 0$. The maximum likelihood estimator, say $\hat{c}_B$, of c can be described as follows:

$$\hat{c}_B = \begin{cases} c', & \text{when } 0 < \bar{x} < \frac{1}{2}x_0 \\ 0, & \text{when } \bar{x} \geqq \frac{1}{2}x_0. \end{cases} \tag{3.5}$$

A table of $\bar{x}/x_0$ as a function of $c'x_0$ is given in Table 1.

The estimator $\hat{c}_B$ is less than $n'\left(\sum_1^{n'} x_i\right)^{-1}$, which is the estimator $\hat{c}$ when $n' = N$ (see (2.4)). This follows from the fact that when $n' = N$, $L_B(c) = L(c)(1 - e^{-cx_0})^{-n'}$ (see (2.3), (3.3)). The estimator $\hat{c}_B$, therefore, has finite mean for $n' \geqq 2$ and finite variance for $n' \geqq 3$.

TABLE 1

$$\frac{\bar{x}}{x_0} = \frac{1}{c'x_0} - \frac{1}{e^{c'x_0} - 1}$$

$c'x_0$	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.		.4916	.4832	.4750	.4668	.4584	.4504	.4422	.4340	.4260
1.	.4180	.4102	.4024	.3946	.3870	.3794	.3720	.3648	.3576	.3504
2.	.3434	.3366	.3300	.3234	.3168	.3106	.3044	.2984	.2924	.2866
3.	.2810	.2754	.2700	.2648	.2596	.2546	.2496	.2450	.2402	.2358
4.	.2314	.2270	.2228	.2188	.2148	.2110	.2072	.2036	.2000	.1966
5.	.1932	.1900	.1868	.1836	.1806	.1778	.1748	.1720	.1694	.1668
6.	.1642	.1616	.1592	.1568	.1546	.1524	.1502	.1480	.1460	.1440
7.	.1420	.1400	.1382	.1364	.1346	.1328	.1310	.1294	.1278	.1262
8.	.1246	.1232	.1216	.1202	.1188	.1174	.1160	.1148	.1134	.1122

4. Asymptotic variances of the estimators. With regard to Case A we have from results of Halperin [3], pp. 4–9, that the asymptotic variance of $(N)^{1/2}\cdot(\hat{c}_A - c)$ is the reciprocal of

$$\int_0^{x_0} \left[\frac{\partial \log g(x)}{\partial c}\right]^2 g(x)\,dx + q\left[\frac{\partial \log q}{\partial c}\right]^2, \tag{4.1}$$

where $q = \Pr(X > x_0) = e^{-cx_0}$ (see (2.7)).

The expression in (4.1) equals

$$c^{-2}(1 - e^{-cx_0}); \tag{4.2}$$

accordingly, for large N

$$\text{Variance } [(N)^{1/2}(\hat{c}_A - c)] = c^2(1 - e^{-cx_0})^{-1}. \tag{4.3}$$

Note that this is always greater than the asymptotic variance of $(N)^{1/2}(\hat{c} - c)$ (see (2.6)).

The asymptotic variance of $(n')^{1/2}(\hat{c}_B - c)$ is the reciprocal of

$$-E(\partial^2 \log f(x)/\partial c^2),$$

where $f(x)$ is given in (2.8). Thus for large n'

$$\text{Variance } [(n')^{1/2}(\hat{c}_B - c)] = [c^{-2} - x_0^2 e^{-cx_0}(1 - e^{-cx_0})^{-2}]^{-1}. \tag{4.4}$$

Having obtained the asymptotic variances of $\hat{c}_A$ and $\hat{c}_B$ let us compare them. The comparison will be made for $n' = N$, which is the most favorable situation for Case B. Let

$$R = \frac{\text{Variance } [(n')^{1/2}(\hat{c}_B - c)]}{\text{Variance } [(n')^{1/2}(\hat{c}_A - c)]}, \qquad n' \text{ large.} \tag{4.5}$$

From (4.3) and (4.4) it follows that

$$R = (1 - e^{-cx_0})/[1 - (cx_0)^2 e^{-cx_0}(1 - e^{-cx_0})^{-2}]. \tag{4.6}$$

TABLE 2

Ratio of the Variances of $\hat{c}_B$ and $\hat{c}_A$

$t = cx_0$	$R(t)$	$t = cx_0$	$R(t)$
0.01	1194	1.1	7.02
0.02	594	1.2	6.25
0.04	294	1.3	5.61
0.06	194	1.4	5.07
0.08	144	1.5	4.62
0.1	113	1.6	4.23
0.2	54.6	1.7	3.90
0.3	34.8	1.8	3.61
0.4	24.9	1.9	3.36
0.5	19.1	2.0	3.13
0.6	15.3	3.0	1.89
0.7	12.6	4.0	1.41
0.8	10.7	5.0	1.20
0.9	9.15	10.0	1.00
1.0	7.97		

R can be considered as a function of $cx_0 = t$, say. A table of R as a function of t is given in Table 2. ($R(t) > 1$ for $t > 0$, and $R(t) \to \infty$ as $t \to 0$.)

5. Interval estimation of c. Approximate $100(1 - q)$ per cent confidence limits for c in (2.7) can be obtained by means of the following approximation when the sample size is large:

$$\Pr(-y_q < y < y_q) = q, \tag{5.1}$$

where y_q is the $100(1 - \frac{1}{2}q)$ per cent point of the standard normal distribution and

$$y = N^{1/2}(\hat{c}_A - c)/[c(1 - e^{-cx_0})^{-1/2}]. \tag{5.2}$$

Similarly, when the sample size is large, $100(1 - q)$ per cent confidence limits for c in (2.8) can be obtained by means of (5.1), where

$$y = (n')^{1/2}(\hat{c}_B - c)[c^{-2} - x_0^2 e^{-cx_0}(1 - e^{-cx_0})^{-2}]^{1/2}. \tag{5.3}$$

The procedure given in [6], Section 11.7, for constructing confidence limits could be used in the cases discussed above.

6. Synthetic sampling experiment. To throw some light on the rapidity of approach of the distributions of $\hat{c}_A$ and $\hat{c}_B$ to their limiting normal distributions we have carried out a synthetic sampling experiment. With regard to $\hat{c}$ the rapidity of approach can be determined by analytic methods since the exact distribution of $\hat{c}$ is known (see Section 2).

A random sample of 140,000 cases was drawn from a rectangular distribu-

TABLE 3

Synthetic sampling experiment

P is the probability that values of χ^2 as large or larger than that obtained would have been obtained under the null hypothesis.*

Serial No. of set of 20,000 cases	Number of cases per sample	Number of samples	$\hat{c}$		$\hat{c}_A$		$\hat{c}_B$	
			χ^2	P	χ^2	P	χ^2	P
1	100	200	20.2	.38	15.8	.67	18.0	.52
2	100	200	21.6	.31	24.2	.19	29.2	.063
3	100	200	25.6	.14	21.4	.32	14.0	.78
4	200	100	17.6	.55	20.0	.39	20.4	.32
5	200	100	19.2	.44	46.8	.00040	33.2	.023
6	200	100	24.4	.18	26.4	.12	27.2	.10
7	200	100	21.2	.33	12.8	.85	15.2	.71

* Equi-probability intervals (.05) were used throughout; thus there are 19 degrees of freedom.

tion and randomly divided into seven sets of 20,000 cases each. Three of these seven sets were divided into 200 samples of 100 cases each; the other four sets were divided into 100 samples of 200 cases each. The variable with the rectangular distribution was then converted (a) to a variable with density function as given in (2.2) with $c = 1$, and (b) to a variable with density function as given in (2.8) with $c = 1$ and $x_0 = 1$. The variable of (a) was used to calculate $\hat{c}$ for each sample (600 samples of 100 cases each; 400 samples of 200 cases each); this distribution was then censored at $x_0 = 1$ and $\hat{c}_A$ was calculated for each of the 1000 samples. The variable of (b) was used to calculate $\hat{c}_B$ for each of the 1000 samples. The goodness of fit of the limiting normal distributions to the observed distributions of $\hat{c}_A$ and $\hat{c}_B$ was tested by chi-square. The goodness of fit of the exact distribution of $\hat{c}$ to the observed distribution was tested similarly. The chi-square probabilities are given in Table 3. Each of the seven lines of Table 3 represents one of the seven independent sets of 20,000 cases. The three values of the chi-square probability, P, on a given line are not independent because they are based on the same samples.

The results suggest that when cx_0 is as small as 1 and the sample size is as small as 100, the distributions of the estimators are fairly well approximated by the limiting distributions. With less severe limitations (i.e., $cx_0 > 1$) the approximation would be better.

REFERENCES

[1] B. Epstein and M. Sobel, "Some theorems relevant to life testing from an exponential distribution," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 373–381.

[2] A. Hald, *Statistical Theory with Engineering Applications*, John Wiley and Sons, New York, 1952.

[3] Max Halperin, "Empirical study of restriction of range; estimation of parameters in

truncated processes," Report No. 4, USAF School of Aviation Medicine, October 1950.

[4] MAX HALPERIN, "Maximum likelihood estimation in truncated samples," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 226–238.

[5] H. M. HUGHES, "Estimation of the variance of the bivariate normal distribution," *University of California Publications in Statistics*, Vol. 1, No. 4, University of California Press, 1949, pp. 37–52.

[6] A. M. MOOD, *Introduction to the Theory of Statistics*, McGraw-Hill Book Co., New York, 1950.

[7] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, Princeton, 1943.