5-1969

A Study of the Exponential Distirbutions and their Applications

Michael Chang-yu Wang
Utah State University

A STUDY OF THE EXPONENTIAL DISTRIBUTIONS AND THEIR APPLICATIONS

by

Michael Chang-yu Wang

A report submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Applied Statistics

Plan B

UTAH STATE UNIVERSITY
Logan, Utah

1969

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The exponential distribution is a widely known distribution in statistical theory. It can be regarded as the continuous analogue of the Poisson distribution, discussed by S. D. Poisson in 1837. The Poisson is a limiting form of the Binomial distribution which can be traced back as early as 1700, discussed by James Bernoulli. A paper by Marsden and Barratt (1911) on the radioactive disintegration of thorium gives a typical frequency distribution which follows the exponential law (8, p. 89). The exponential distribution has achieved importance recently in connection with the theory of stochastic process and has found a wide variety of applications in the fields of Physics, Biology, and Engineering. For instance, in the study of "Markov Processes" in continuous time, we notice that a very simple type of the process is the distribution of the time interval between any two successive events which follows the negative exponential distribution (1, p. 66-69).

Bulmer and Parzen have defined the exponential distribution in their books as a law of waiting times or as a law of time to failure such that any numerical valued random phenomena whose occurrences are random in time and independent of, what Bailey called, the past, present, and future state of the system may distribute exponentially (39, p. 262). Many physical, biölogical situations can be approximated by the exponential distribution, such as radioactive disintegration, telephone calls, mutant genes, infectious persons, the life of an electron tube, the time intervals between successive breakdowns of an electronic system, the time intervals between accidents, such as explosions in mines, etc. (34, 168-180). As an example, the numerical and graphical presentation of the time intervals in days between explosions in mines, involving more than 10 men killed, from December, 1875, to May, 1951, taken from Pearson (34, p. 168-180) are shown in Table 1 and Figure 1. It follows approximately the exponential distribution with mean time interval equal to 241 days.

Table 1. Time intervals in days between explosions in mines, involving more than 10 men killed, from December 6, 1875, to May 28, 1951

378	286	871	66
36	114	448	291
15	108	123	4
31	188	457	369
215	233	498	338
11	28	49	336
137	22	131	19
4	61	182	329
15	78	255	330
72	99	195	312
96	326	224	171
124	275	566	145
50	54	390	75
120	217	72	364
203	113	228	37
176	32	271	19
55	23	208	156
93	151	517	47
59	361	1613	129
315	312	54	1630
59	354	326	29
61	58	1312	217
1	275	348	7
13	78	745	18
189	17	217	1357
345	1205	120	
20	644	275	(Complete interval
81	467	20	to May 29, 1951)

Mean time interval = 241 days

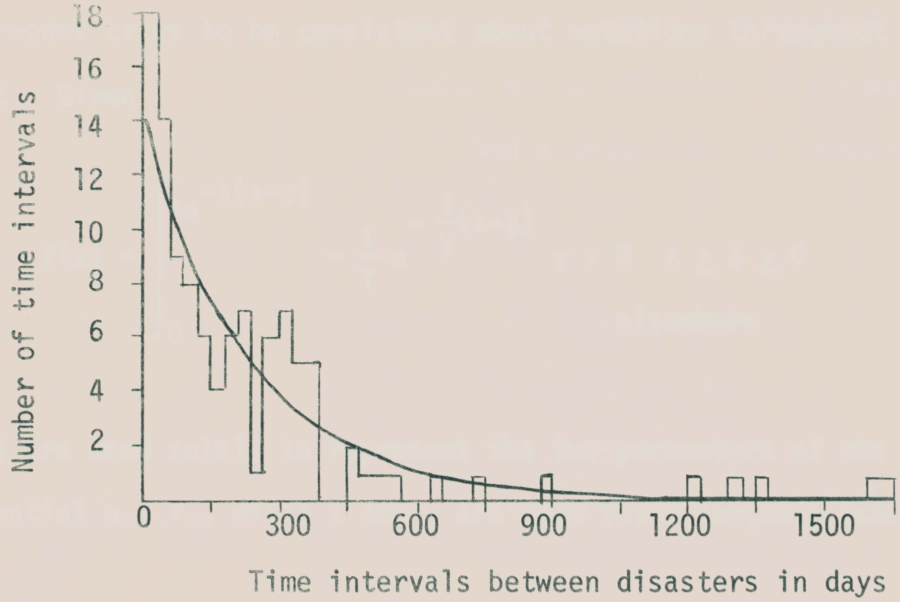

Figure 1. Histogram of time intervals (in days) plotted against frequencies.

This report is to provide a survey of the literatures which deal with topics such as the distribution function, estimation of distribution parameters, the validity test of the distribution, test of hypothesis of population parameters, and the application aspects (in examples) of the single, bivariate, and mixed exponential in complete, truncated, and censored cases whenever such literatures are available.

By truncated samples we mean that the population from which the samples are taken are truncated either to the right or the left, or at both ends. By censored samples we mean that the samples themselves are truncated. No observations greater or less (or both) than a certain preassigned value will be taken into consideration.

Since literatures taken into the survey are works of many authors, it is conceivable that different sets of notations are used. This report tries to be consistent about notations throughout the report. Given

$$f(x) = \begin{cases} \lambda e^{-\lambda(x-\theta)} = \frac{1}{\gamma} e^{-\frac{1}{\gamma}(x-\theta)} & \gamma > 0 \quad x \geq \theta \geq 0 \\ 0 & \text{elsewhere} \end{cases} \tag{1.1}$$

$\theta, \lambda(\frac{1}{\gamma})$ are used solely to represent the two parameters of the exponential p.d.f., while λ and γ are used interchangably according to:

$$\lambda = \frac{1}{\gamma} \tag{1.2}$$

$$\gamma = \frac{1}{\lambda} \tag{1.3}$$

This is done merely for the purpose of convenience of representation. Following is a list of the notations used quite consistently throughout the report.

Notations	Meanings
x_i, i=1, 2,...	observations from a complete sample.
y_i, i=1,2,...	ordered observations from truncated or censored sample.
U, V, W, Z	transformation of x_i or y_i.
δ	relation coefficient of any two random variable.
α,β	Type I and Type II error.

$\rho, L(x)$	likelihood-ratio and likelihood function.
$f(x)$, $F(x)$	density and cumulative function.
n, r	number of observations in a complete and a truncated (or censored) sample.

MATHEMATICAL DERIVATION

In this section, the mathematical derivations of the probability density function, the cumulative distribution function, and the moment generating function will be developed for one-parameter exponential distribution. The mean, variance, and higher moments will be derived from direct evaluation and from the moment generating function. The bivariate and mixed exponential will be considered in the latter part of this section.

2.1 One-parameter exponential

According to Bulmer (8, p. 89), the p.d.f., c.d.f., and m.g.f. are defined as follows:

Definition 2.1.1: A continuous, positive random variable is said to follow the exponential distribution if its p.d.f. is given by:

$$f(x) = \lambda e^{-\lambda x} \qquad x \geq 0$$

$$= 0 \qquad x < 0 \tag{2.1.1}$$

The distribution of the density function for several different values of λ is given in Figure 2.

Definition 2.1.2: The cumulative distribution function is given by:

$$F(x) = P[X \leq x] = 1-e^{-\lambda x} \tag{2.1.2}$$

Figure 2. (a) p.d.f. with $\lambda = 2$; (b) p.d.f. with $\lambda = .5$.

Definition 2.1.3: The moment generating function is given by:

$$m(x) = \frac{\lambda}{\lambda - x} \qquad (2.1.3)$$

To prove (2.1.1), two approaches will be considered.

<u>Bulmer and Bailey's approach</u>. It is assumed that in a short interval of time, Δx, the chance that an event will occur is $\lambda \Delta x$. If Δx approaches zero, the chance that no event occurs is $1-\lambda\Delta x$. Furthermore, it is assumed that the chance an event which occurs in Δx does not depend on how many events have already occurred. In other words, the events occur at random, or independent of one another.

If $p_0(x)$ = the probability that no event occurs before time x, then $p_0(0) = 1$ (the initial condition), and $p_0(x + \Delta x) = p_0(x)\cdot(1-\lambda\Delta x)$. This is equivalent to

$$\frac{p_0(x + \Delta x) - p_0(x)}{\Delta x} = -\lambda p_0(x) \qquad (2.1.4)$$

By assumption, when $\Delta x \to 0$, (2.1.4) becomes

$$\lim_{\Delta x \to 0} \frac{p_0(x + \Delta x) - p_0(x)}{\Delta x} = \frac{d\, p_0(x)}{d(x)} = -\lambda p_0(x)$$

$$\int_0^X \frac{d\, p_0(x)}{p_0(x)} = -\lambda \int_0^X dx \quad \text{or} \quad \ln p_0(x) = -\lambda x$$

$$p_0(x) = e^{-\lambda x}$$

which is the probability that at time X no event has occurred. The cumulative distribution function of the arrival time x of the first event is

$$F(x) = 1 - e^{-\lambda x}$$

Parzen's approach. For $x \geq 0$, let $F_r(x)$ be the probability that the time of occurrence of the rth event $\leq X$. Then $1-F_r(x)$ will be the probability that the time of occurrence of the rth event $> X$, or the probabiliyt that the number of events which occur in the time from 0 to X is less than r. A density function which describes the random phenomena can be expressed as the waiting time to the rth event in a series of events happening in accordance with the Poisson probability law at the rate of λx per unit of time (or space). Consequently,

$$1-F_r(x) = \sum_{k=0}^{r-1} \frac{1}{k!} (\lambda x)^k e^{-\lambda x} \qquad (2.1.5)$$

By differentiating (2.1.5) with respect to x, it becomes

$$f(x) = \frac{\lambda}{(r-1)!} (\lambda x)^{r-1} e^{-\lambda x} \qquad x \geq 0 \tag{2.1.6}$$

When r = 1, the waiting time to the first event, (2.1.6) becomes

$$f(x) = \lambda e^{-\lambda x} \qquad x \geq 0$$

$$= 0 \qquad x < 0$$

For both approaches, the mean and variance can be found by the expected value

$$E(x) = \int_0^\infty x \cdot \lambda e^{-\lambda x} \, dx \tag{2.1.7}$$

Let $u = \lambda x$; (2.1.7) becomes

$$\gamma \int_0^\infty ue^{-u} \, du = \gamma$$

$$E(x^2) = \int_0^\infty x^2 \cdot \lambda e^{-\lambda x} \, dx \tag{2.1.8}$$

Let $u = x^2$, $dv = \lambda e^{-\lambda x} \, dx$

$du = 2x dx$, $v = -e^{-\lambda x}$; (2.1.8) becomes

$$\int_0^\infty 2xe^{-\lambda x} \, dx \tag{2.1.9}$$

Let $w = \lambda x$; (2.1.9) becomes

$$2\,\gamma\int_0^\infty we^{-w}\,dw = 2\gamma^2$$

$$\sigma^2 = E(x^2) - E^2(x) = \gamma^2$$

The cumulative distribution function is found by

$$F(x) = P[X \le x] = \int_0^X \lambda e^{-\lambda x}\,dx = 1 - e^{-\lambda x} \qquad x \ge 0$$

It is clear from the equation that as x increases from 0 to ∞, F(x) increases from 0 to 1. The scale of F(x) increases as λ decreases. The moment generating function is found by

$$M(t) = E(e^{tx}) = \int_0^\infty e^{-(\lambda-t)x}\,dx \qquad (2.1.10)$$

Let $u = (\lambda-t)x$; (2.1.10) becomes

$$\lambda\int_0^\infty e^{-u}\,\frac{du}{(\lambda - t)} = \frac{\lambda}{\lambda - t}$$

The mean, variance, and higher moments are found by differentiating M(t) with respect to t and set to to 0.

$$u = E(x) = M'(t) = \left.\frac{\lambda}{(\lambda - t)^2}\right|_{t=0} = \gamma$$

$$\sigma^2 = E(x^2) - E^2(x) = M''(t) - E^2(x) = \gamma^2$$

$$u_3 = E[(x-u)^3] = 2\gamma^3 \tag{2.1.11}$$

So skewness is +2. From Figure 2, it can be seen that it skews highly to the right.

2.2 Bivariate exponential

The properties of bivariate distribution about the normal case have been studied intensively since Bravais and Karl Pearson. Yet, according to Gumbel (24, p. 698-707), none of the well known properties of the bivariate normal distribution are applicable to the bivariate exponential. He further noted that a bivariate distribution is not determined by the knowledge of the margins. Under different conditions, different bivariate exponentials can be derived. He, then, derived three bivariate exponentials all with exponential margins.

If $F_1(x)$, $F_2(y)$, $f_1(x)$, and $f_2(y)$ are the c.d.f. of x and y, and p.d.f. of x and y, then a bivariate probability function $F(x_1y)$ with these marginal distributions is monotonically increasing from 0 to unity and is subject to the following conditions:

$$F(\infty,y) = F(x_1-\infty) = 0;\ F(x_1\infty) = F_1(x);\ F(\infty,y) = F_2(y);$$

$$F(\infty,\infty) = 1 \tag{2.2.1}$$

$$P\{x_1 \leq X \leq x_2,\ y_1 \leq Y \leq y_2\} = F(x_2,y_2) - F(x_2,y_1)$$

$$- F(x_1,y_2) + F(x_1,y_1) \geq 0 \tag{2.2.2}$$

$$\frac{\partial^2 F}{\partial x \partial y} = f(x,y) \geq 0 \tag{2.2.3}$$

$$\int_{-\infty}^{\infty} f(x,y)\, dy = f_1(x); \qquad \int_{-\infty}^{\infty} f(x,y)\, dx = f_2(y) \tag{2.2.4}$$

Definition 2.2.1: Given x and y a two-dimentional random variable each has an exponential marginal probability density function. From (2.2.2), a bivariate c.d.f. is given by

$$F(x,y) = 1 - e^{-x} - e^{-y} + e^{-x-y-\delta xy} \quad x \geq 0,\ y \geq 0,\ 0 \leq \delta \leq 1 \tag{2.2.5}$$

When $\delta = 0$, the fact that $F(x,y) = (1-e^{-x})(1-e^{-y}) = F_1(x)\cdot F_2(y)$ leads to independence.

Proof: From (2.2.3), the density function is given by

$$f(x,y) = e^{-x(1+\delta y-y)}[(1+\delta x)(1+\delta y)-\delta] \tag{2.2.6}$$

with $f(\infty,y) = f(x,\infty) = 0$; $f(0,0) = 1 - \delta$

From $f(0,0) = 1 - \delta$, $f(\infty,\infty) = 1$ and the nonnegativity of a density function it follows $0 \leq \delta \leq 1$.

With restriction (2.2.6), the conditions (2.2.1), (2.2.3), and (2.2.4) are fulfilled. Therefore, (2.2.5) is a bivariate distribution with exponential margins.

The functional relatiionships, such as the conditional density function, etc., will be given below. For details, see Gumbel (24, p. 698-707). The conditional density function (39, p. 334)

$$f(x|y) = \frac{f(x,y)}{f_2(y)} = e^{-x(1+\delta y)}[(1+\delta x)(1+\delta y) - \delta] \tag{2.2.7}$$

The conditional expectation

$$E(x|y) = \int_{-\infty}^{\infty} xf(x|y)\, dx = \frac{1 + \delta + \delta y}{(1 + \delta y)^2} \tag{2.2.8}$$

The conditional second moment

$$E(x^2|y) = \int_{-\infty}^{\infty} x^2 f(x|y)\, dx = \frac{2}{(1 + \delta y)^2} + \frac{4\delta}{(1 + \delta y)^3} \tag{2.2.9}$$

The conditional variance $\sigma^2(x|y)$ of x as a function of y

$$\sigma(x|y) = E(x^2|y) - E^2(x|y) = \frac{1}{(1 + \delta y)^2} + \frac{2\delta}{(1 + \delta y)^2} + \frac{\delta^2}{(1 + \delta y)^4} \tag{2.2.10}$$

When x and y are independent of each other, i.e., when $\delta = 0$, $\sigma^2(x|y) = 1$. The conditional standard deviation of x as a function of y

$$\sigma(x|y) = [(1 + y)^2 + 1 + 2y]^{\frac{1}{2}} (1 + y)^{-2} \tag{2.2.11}$$

The squared conditional coefficient of variation obtained from (2.2.8) and (2.2.10)

$$\frac{\sigma^2(x|y)}{E^2(x|y)} = \frac{(1 + \delta y)^2 + 2\delta\,(1 + \delta y) - \delta^2}{(1 + \delta y)^2 + 2\delta\,(1 + \delta y) + \delta^2} \tag{2.2.12}$$

When y increases, the preceding equation converges to unity. In notation,

$$\sigma(x|y) \simeq E(x|y)$$

Definition 2.2.2: In two previous papers of his, E. J. Gumbel (24, p. 707) shows that, given two c.d.f. $F_1(x)$ and $F_2(y)$, a bivariate distribution function is given by

$$F(x,y) = F_1(x)\ F_2(y)\ [1 + \delta\ (1 - F_1(x))(1 - F_2(y))]$$

$$-1 \leq \delta \leq 1 \qquad (2.2.13)$$

The bivariate p.d.f. is given by

$$f(x,y) = f_1(x)\ f_2(y)\ [1 + \delta\ (2F_1(x)-1)(2F_2(y)-1)] \qquad (2.2.14)$$

Definition 2.2.3: Given two exponential distribution functions $F_1(x)$ and $F_2(y)$, from Definition 2.2.2, their bivariate distribution function becomes

$$F(x_1 y) = (1 - e^{-x})(1 - e^{-y})[1 + \delta\ e^{-x-y}] \qquad x \geq 0;\ y \geq 0 \qquad (2.2.15)$$

Their bivariate p.d.f. is given by

$$f(x_1 y) = e^{-x-y}[1 + \delta\ (2e^{-x} - 1)(2e^{-y} - 1)] \qquad (2.2.16)$$

The proof of (2.2.15) and (2.2.16) are identical with those in Definition 2.2.1. As in Definition 2.2.1, the various functional relationships are given below. For reference, see Gumbel (24, p. 704). The conditional density function

$$f(x|y) = e^{-x}\,(1 + \delta - 2\delta\, e^{-y}) - 2\,\delta\, e^{-2x(1-2e^{-y})} \qquad (2.2.17)$$

The conditional expectation

$$E(x|y) = E_{\delta}\,(x|y) = 1 + \frac{\delta}{2}\delta e^{-y} \qquad (2.2.18)$$

The conditional variance $\sigma^2(x|y)$ of x as a function of y

$$\sigma^2(x|y) = E(x^2|y) - E^2(x|y) =$$

$$= 1 + \frac{\delta}{2}\,(1 - 2e^{-y}) - \frac{\delta^2}{4}\,(1 - 2e^{-y})^2 \qquad (2.2.19)$$

The squared conditional coefficient of variation

$$\frac{\sigma^2(x|y)}{E^2(x|y)} = \frac{(\frac{1}{4} + 2e^{-y} - e^{-2y})}{\frac{1}{4}\,(1 - 2e^{-y} + e^{-2y})} \qquad (2.2.20)$$

which converges, with increasing y, towards unity.

Definition 2.2.4: Given x and y a two-dimensional variable each followed an exponential distribution, a bivariate distribution function of x and y is given by

$$F(x,y) = 1 - e^{-x} - e^{-y} + P(x,y) \tag{2.2.21}$$

where

$$P(x,y) = e^{-(x^m + y^m)^{1/m}} \tag{2.2.22}$$

x and y are distributed independently if and only if $m = 1$.

The p.d.f. is given by

$$f(x,y) = P(x,y)\cdot(x^m + y^m)^{\frac{1}{m}-2}\; x^{m-1}\, [(x^m + y^m)^{\frac{1}{m}} + m^{-1}] \tag{2.2.23}$$

is nonnegative only if $m \geq 1$.

By knowing that the marginal distributions are exponential, under different criteria, a unique bivariate exponential can be derived. As an example, J. E. Freund has derived, under his assumption, a bivariate exponential designed, in particular, for the life testing of two-component systems (22, p. 971-977). The p.d.f. is given by

$$f(x,y) = \begin{cases} \lambda_1\lambda_2' \, e^{-\lambda_2' y - (\lambda_1 + \lambda_2 - \lambda_2')x} & 0 < x < y \\ \lambda_2\lambda_1' \, e^{-\lambda_1' x - (\lambda_1 + \lambda_2 - \lambda_1')y} & 0 < y < x \end{cases} \tag{2.2.24}$$

which is quite different from Gumbel's. But, nevertheless, when $\lambda_1 = \lambda_1'$, and $\lambda_2 = \lambda_2'$, i.e., x and y are independent, (2.2.24) reduced to two marginal exponential p.d.f. of x and y.

$$f(x) = \lambda_1 e^{-\lambda_1 x} \qquad\qquad f(y) = \lambda_2 e^{-\lambda_2 y}$$

2.3 Mixed exponential

By mixed exponential we mean a population composed of $i = 1,2,\ldots.,r$ subpopulations, each distributed exponentially, mixed in proportion $p_1, p_2,\ldots.,p_r$, where $0 \leq p_i \leq 1$, and $\sum_{i=1}^{r} p_i = 1$. The simple case when $i = 2$ is given by P. R. Rider (42, p. 143-147) and W. Mendenhall (36, p. 504-505). For simplicity of notation, let $q = 1 - p$.

Given x a random variable which can be described by the exponential p.d.f. of the form

$$f(x) = \lambda e^{-\lambda x} \quad \gamma > 0, \; 0 \leq x \leq \infty \tag{2.3.1}$$

Suppose two populations of (2.3.1), with parameter γ_1 and γ_2 respectively are mixed in the unknown proportion p and q. The resulting p.d.f. is given by

$$f(x) = p \cdot f_1(x) + q \cdot f_2(x) = p\, \lambda_1 e^{-\lambda_1 x} + q\, \lambda_2 e^{-\lambda_2 x} \tag{2.3.2}$$

The c.d.f. is given by

$$F(x) = p \cdot F_1(x) + q \cdot F_2(x) \tag{2.3.3}$$

If the survival function is given as

$$G_i(x) = 1 - F_i(x) \qquad i = 1,2 \tag{2.3.4}$$

$$G(x) = 1 - F(x) \tag{2.3.5}$$

This is needed because in practice the relative proportion of each subpopulation is generally subject to change with time. At time t, the subpopulation would be mixed in the proportion $p(x):1-p(x)$. These are called the conditional mixture proportions. Consequently,

$$p(x) = \frac{p\,G_1(x)}{G(x)} \tag{2.3.6}$$

$$p(0) = p \tag{2.3.7}$$

TEST OF VALIDITY THAT THE UNDERLYING DISTRIBUTION IS REALLY EXPONENTIAL

To assure that the underlying distribution is exponential, B. Epstein (14, p. 83-101) has a number of graphical and analytical procedures for testing the assumption that the underlying distribution is really exponential.

3.1 A graphical procedure

This procedure is particularly useful if large samples are available. Suppose that the underlying cumulative distribution function is really exponential, given by

$$
\begin{aligned}
F(x) &= 1 - e^{-\lambda x} \qquad && x \geq 0,\ \gamma > 0 \\
&= 0 && x < 0 \qquad\qquad (3.1.1)
\end{aligned}
$$

Then $z = \log\left(\frac{1}{1-F(x)}\right) = \lambda x$ when plot against x is a straight line with slope λ.

Given a sample of size n and arranged in ascending order such that $y_1 \leq y_2 \leq \ldots \leq y_n$. If the assumption holds, the values

$$F(y_i) = \frac{i}{n+1} \qquad (3.1.2)$$

plotted against y should fit well by a straight line passing through the origin. In censored case, one expects a good linear fit up to y_r.

If the underlying distribution is the two-parameter exponential, i.e.,

$$F(y;\theta,\lambda) = 1 - e^{-\lambda(y-\theta)} \qquad y \geq \theta \geq 0$$

$$= 0 \qquad y < 0 \qquad (3.1.3)$$

then $\log\left(\frac{1}{1 - F(y;\theta,\lambda)}\right) = \lambda(y-\theta)$, when plotted against y following the one-parameter exponential case would be a straight line that cuts the y axis at the point $y = \theta$.

Example 3.1.1: Forty-nine items, such as electronic tubes, were placed on life test until all items failed. The observed failure times are listed on Table 2.

Table 2. The observed failure times of 49 items placed on life test

1.2	13.7	38.9	72.4	102.8	151.6	203.0
2.2	15.1	47.9	73.6	108.5	152.6	204.3
4.9	15.2	48.4	76.8	128.7	164.2	229.5
5.0	23.9	49.3	83.8	133.6	166.8	253.1
6.8	24.3	53.2	95.1	144.1	178.6	304.1
7.0	25.1	55.6	97.9	147.6	185.2	341.7
12.1	35.8	62.7	99.6	150.6	187.1	354.4

Using procedure 3.1, the graph of y plotted against $z_i = \log\left(\frac{n+1}{n+1-i}\right)$ on semi-log paper appears approximately as a straight line passing through the origin.

3.2 The x^2 test for goodness of fit

This is also for the case when large samples are available. Given a large sample, it is first divided into k intervals of the form

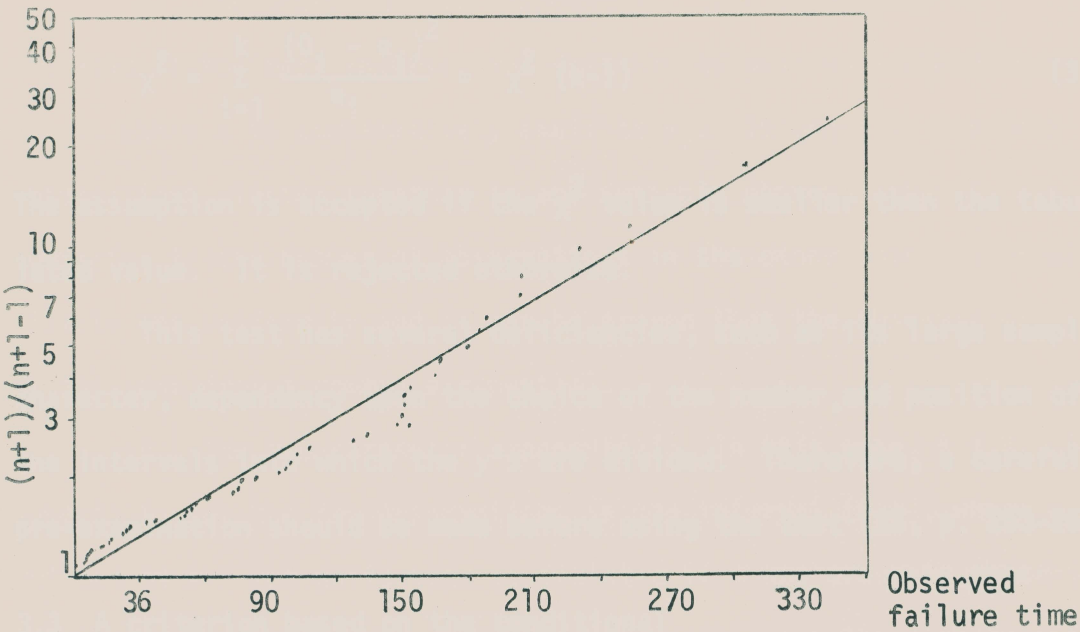

Figure 3. Graph of $z_i = \log \frac{(n+1)}{(n+1-i)}$ against y_i.

$y_{t_1} \leq y_{t_2} \leq \ldots, \leq y_{t_{k-1}}$, then the expected number of observations in each interval is found by $\hat{\gamma}$, the best estimate of γ based on the entire sample. If O_i is the observed number of observations in the ith interval and e_i the expected number of observations in the ith interval, where $e_i = np_i$ and p_i is given by

$$p_1 = \int_0^{t_1} \hat{\lambda} e^{-\hat{\lambda} y}\, dy$$

$$p_i = \int_{t_{i-1}}^{t_i} \hat{\lambda}\, e^{-\hat{\lambda} y}\, dy \qquad i = 1,2,\ldots,k-1 \tag{3.2.1}$$

$$p_k = \int_{t_{k-1}}^{\infty} \hat{\lambda} e^{-\hat{\lambda} y}\, dy$$

Then the usual chi-square goodness of fit test with (k-1) d.f. is given by

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} \simeq \chi^2\,(k-1) \tag{3.2.2}$$

The assumption is accepted if the χ^2 value is smaller than the tabulated value. It is rejected otherwise.

This test has several deficiencies, such as its large sample character, dependency upon the choice of the number, and position of the intervals into which the y's are divided. Therefore, a careful pre-examination should be made before using the test (23, p. 253-263).

3.3 A criterion based on the conditional distribution of total observations

This test utilizes the basic properties of Poisson processes (20, p. 1). B. Epstein has shown (14, p. 96-97) that if there are r observations from a Poisson process such that $y_1 \leq y_2 \leq \cdots \leq y_r$ and y_r is a preassigned termination point, then these (r-1) observations are independent of each other and are distributed uniformly over $(0, y_r)$. For large r, $\sum_{i=1}^{r} y_i$ is approximately normally distributed with mean $((r-1)y_r)/2$ and variance $((r-1)y_r^2)/12$. This is used to test whether the r observations are drawn from a Poisson process. Similarly, he has shown that the cumulative sum $n(y_1)$, $n(y_2)$, ... of $y_1 \leq y_2 \leq \cdots \leq y_r$, where $n(y_1) = ny_1$ and $n(y_r) = \sum_{i=1}^{r-1} y_i + (n-r+1)\,y_r$, $r=2\ldots,r \leq n$, from an exponential distribution is distributed uniformly over $(0,y_r)$.

The uniformity of the (r-1) observations under the assumption is then used to detect any deviation of the sample parameter γ from a constant. Two cases are considered.

A. The one-parameter exponential: The uniformity is violated if too many observations distributed near each other in $(0,\varepsilon)$, where ε is comparatively small to y_r. In general, too many observations cluster together in any suitable chosen interval, as comparing to the number of observations in the other intervals would violate the uniformity and it can be proven that the observations do not come from a common exponential.

B. The two-parameter exponential: In this case, $\theta > 0$. One expects to get too few observations in the interval $(0,\varepsilon)$, where ε is defined in A. The uniformity is violated if there are many observations in that interval. Then the assumption that the r observations are from a one-parameter exponential would not be true.

If r is small, only fairly large changes can be detected. With r large enough, a chi-square test can be used to detect whether the conditional distribution of cumulative sum, given some pre-assigned termination point, deviates badly from being uniform.

Example 3.3.1: A test is discontinued without replacement after 11 out of 20 items fail. The observed failure times are listed in Table 3.

Table 3. The observed failure times of the first 11 items

7	12	15	24	25	48	53	56	72	95	100

The theoretical mean = $\frac{(r\text{-}1\ n\ (y_r)}{2}$ = 7035 and the theoretical standard deviation = $\frac{(r\text{-}1\ n\ (y_r)^2}{12}$ = 1285. Under the assumption, $\sum_{i=1}^{10} n\ (y_i)$ is approximately normally distributed with mean 7035 and standard deviation 1285. A 95 percent acceptance interval is given by 7-35 ± 1.96 (1285) = (4517,9553). The observed sum is $\sum_{i=1}^{10} n(y_i)$ = 6613. So the assumption that the underlying distribution is exponential is accepted at .05 significance level.

3.4 A test for abnormally small observation

When a situation exists in which the underlying distribution is really exponential but with first or first two observations abnormally small, the deviation can be detected by the following test.

Suppose it is given r observations of the form $y_1 \leq y_2 \leq \cdot \cdot \cdot \leq y_r$ and one wants to test whether y_1 is abnormally small. B. Epstein and M. Sobel (17, p. 486-502; 18, p. 373-381) have shown that given $n(y_1)$ and $n(y_1 - y_1)$, the cumulative sum in $(0,y_1)$ and (y_1, y_r), it should distribute independently of each other if all y_i are drawn from a common exponential. Then,

$\frac{2n\ (y_1)}{\gamma}$ is distributed as chi-square with 2 d.f., and $\frac{2n\ (y_r - y_1)}{\gamma}$ is distributed as chi-square with (2r-2) d.f., and $\frac{(r\text{-}1)\ n\ (y_1)}{n\ (y_r - y_1)}$ is distributed as F-distribution with (2,2r-2) d.f. To test whether y_1 is abnormally small, one then uses the inequality (3.4.1), as a criterion. Given α the significance level of the test, one accepts the hypothesis that y_1 is abnormally small if

$$n(y_1) < F_\alpha \frac{n(y_r - y_1)}{r-1} \tag{3.4.1}$$

where F_α is a lower α point of the $F(2,2r-2)$ distribution such that

$$P_r\{ F(2,2r-2) \leq F_\alpha\} = \alpha \tag{3.4.2}$$

For testing both y_1 and y_2 abnormally small, B. Epstein and M. Sobel have shown, as in y_1 case that $n(y_2)$ and $n(y_r - y_2)$, the cumulative sum in $(0,y_2)$ and (y_2,y_r), are distributed independently of each other. Furthermore, $(2n(y_2))/\gamma$ is distributed as chi-square with 4 d.f. and $\frac{2n(y_r - y_2)}{\gamma}$ is distributed as chi-square with $(2r-4)$ d.f. Therefore, $\frac{(r-2)n(y_2)}{2n(y_r - y_2)}$ is distributed as F-distribution with $(4,2r-4)$ d.f. Given α is the significance level of the test, one accepts the hypothesis that y_1 and y_2 are abnormally small if

$$n(y_2) < F_\alpha \frac{2n(y_r - y_2)}{r-2} \tag{3.4.3}$$

where F_α is a lower α point of the F-distribution with $(4,2r-4)$ d.f. such that

$$P_r\ (F(4,2r-4) \leq F_\alpha) = \alpha \tag{3.4.4}$$

Example 3.4.1: Given 10 observations with $n(y_2) = 24$ and

$n(y_{10}) = 600$. Using Equation (3.4.1), it can be calculated that

$$n(y_2) = 24 < \frac{2(600 - 24)}{(5.84)\cdot 8} = 24.7$$

This proved that the first two observations are abnormal.

3.5 A test for an abnormally large first observation

This test is based on the fact that if the first observation is comparatively large, then the underlying distribution is really two-parameter exponential with (θ,λ) instead of one-parameter exponential with λ.

Given $\dfrac{(r-1)\, n\, (y_1)}{n\, (y_r - y_1)}$, which compares the cumulative sum in $(0,y_1)$ with the cumulative sum in (y_1,y_r), distributed as $F(2,2r-2)$ if $\theta = 0$. One accepts the hypothesis that y_1 is abnormally large if

$$n(y_1) > F_\alpha \, \frac{n(y_r - y_1)}{r-1} \qquad (3.5.1)$$

where F_α is an upper α point of the $F(2,2r-2)$ distribution such that

$$P_r \, \{F(2,2r-2) \geq F_\alpha\} = \alpha \qquad (3.5.2)$$

One rejects the hypothesis that y_1 is abnormally large if

$$n(y_1) < F_\alpha \, \frac{n(y_r - y_1)}{r-1} \qquad (3.5.3)$$

where F_α is given by (3.5.2).

3.6 A test for whether the parameter γ_1 in the first half sample differs significantly from the parameter γ_2 in the second half sample

Sometimes situations arise that one wants to investigate a gross change over an entire sample. This may occur in a situation that a certain portion of the sample does not follow from a common exponential. The actual choice of which two intervals of the sample for the test depends on some *a priori* knowledge.

As before, given $y_1 \leq y_2 \leq \cdots \leq y_r \leq \cdots \leq y_{2r}$ the first 2r ordered observations, $n(y_r)$ and $n(y_{2r} - y_r)$ are distributed independently of each other. Furthermore, $(2n(y_r))/\gamma$ is distributed as chi-square with 2r d.f. and $\frac{2n(y_{2r} - y_r)}{\gamma}$ is distributed as chi-square with 2r d.f. Therefore, $\frac{n(y_r)}{n(y_{2r} - y_r)}$ is distributed as F-distribution with (2r,2r) d.f. One accepts the hypothesis that there is no significant difference between the two intervals if

$$n(y_r) < F_\alpha \; n(y_{2r} - y_r) \tag{3.6.1}$$

where F_α is a lower α point of the F(2r,2r) distribution such that

$$P_r \; \{F(2r,2r) \leq F_\alpha\} = \alpha \tag{3.6.2}$$

In general, one can consider any two intervals of the form $(0,y_k)$ and (y_k, y_s) where $k < s \leq r$. Followed the same argument given above, one obtained

$$\frac{(s-k)\ n(y_k)}{k\ n\ (y_s - y_k)} \simeq F\ (2k,\ 2s - 2k) \qquad (3.6.3)$$

which can be used to test the hypothesis that there is no significant difference between intervals $(0, y_k)$ and (y_k, y_s) (17, p. 458-466).

Example 3.6.1: Twenty observations were obtained from a life test with $n(y_{10}) = 10000$ associated with the time interval $(0, y_{10})$ and $n(y_{20} - y_{10}) = 25000$ associated with the time interval (y_{10}, y_{20}). Using Equation (3.6.1)

$$\frac{n(y_{10})}{n(y_{20} - y_{10})} = 2.5 \nless \frac{1}{F_\alpha\ (20,20)} = 2.46$$

This proved that γ_1, the mean time of occurrence of an event in interval $(0, y_{10})$, and γ_2, the mean time of occurrence of an event in interval (y_{10}, y_{20}) are not the same.

3.7 A test for whether or not the parameter γ fluctuates when the sample is divided into k intervals

This test is a natural extension of 3.6. Instead of two, one wants to test whether the parameter γ is constant in k groups, where the k,4 observations are arranged as $y_{11} \leq y_{12} \leq \cdots \leq y_{1r} \leq y_{21} \leq \cdots \leq y_{2r} \leq \cdots \leq y_{k1} \leq \cdots \leq y_{kr}$. M. S. Bartlett has derived a test for homogeneity of variance and is used here by B. Epstein.

First, $n(y_{1r})$, $n(y_{2r} - y_{1r})$, ..., $n(y_{kr} - y_{(k-1)r})$, respectively, are mutually independent and

$\dfrac{2n\ (y_{ir} - y_{(i-1)r})}{\gamma}$ is for each $i = 1,2,\ldots,k$ distributed as chi-square with 2r d.f. Furthermore, $\dfrac{2n\ (y_{kr})}{\gamma}$ is distributed as chi-square with (2rk) d.f., where $n(y_{kr}) = n(y_{1r}) + n(y_{2r} - y_{1r}) + \ldots + n(y_{kr} - y_{(k-1)r})$, B. Epstein verified that

$$\frac{2rk\{\log\frac{n(y_{kr})}{k} - \frac{1}{k}[\log n(y_{1r}) + \log n(y_{2r}-y_{1r}) + \ldots + \log n(y_{kr}-y_{(1-1)r})]\}}{1 + \frac{k+1}{6rk}}$$

$$\doteq \chi^2(k-1) \qquad (3.7.1)$$

One rejects the hypothesis that the parameter γ in the k intervals is a constant if

$$(3.7.1) > \chi^2_{\alpha}(k-1) \qquad (3.7.2)$$

where $\chi^2_{\alpha}(k-1)$ is the upper α point of $\chi^2(k-1)$ such that

$$P_r\ \{\chi^2(k-1) > \chi^2_{\alpha}\ (k-1)\} = \alpha \qquad (3.7.3)$$

Example 3.7.1: Five homogeneous samples of 10 observations each were obtained with $n(y_{1r}) = 12.74$, $n(y_{2r}) = 10.02$, $n(y_{3r}) = 7.25$, $n(y_{4r}) = 9.18$, $n(y_{5r}) = 12.99$. Using Equation (3.7.1)

$$\frac{-2rk\ \log L_1}{1 + \frac{k+1}{6rk}} = 2.26 < 9.49$$

(the upper 5 percent point of the $\chi^2(4)$ distribution). This proved that γ is constant for the five samples.

3.8 A special case of Procedure 3.7 when r = 1

Two situations are discussed by B. Epstein which are based on the alternatives of the hypothesis.

A. H_o: the distribution of cumulative sum between successive observations is exponential with constant γ.

H_A: the p.d.f. is Type III or a member of the Weibull family.

Given $y_1 \leq y_2 \leq \ldots \leq y_r$, (3.7.1) reduced to

$$\frac{2k\{\log\frac{n(y_k)}{k} - \frac{1}{k}[\log n(y_1)+\log n(y_2-y_1)+\ldots+\log n(y_k-y_{(k-1)})]\}}{1+\frac{k+1}{6k}}$$

$$\simeq \chi^2(k-1) \qquad (3.8.1)$$

when r =1 . This can be used to test whether γ is constant from one observation to the next.

The likelihood-ratio test, derived by P. A. P. Moran, is used here by B. Epstein for testing

H_o: the underlying p.d.f. is given by $f(x) = \lambda e^{-\lambda(x-\theta)}$

H_A: the p.d.f. is given by $f(x;\gamma,\beta) = \frac{1}{\tau(\alpha)\gamma^\beta} \chi^{\beta-1} e^{-\chi/\gamma}$ $\beta>0$

(noted that if $\beta = 1$, $f(x;\theta,1) = \lambda e^{-\lambda x}$).

D. J. Bartholomew (4, p. 64-78) has shown that if one tests

B. H_o: a stochastic process is Poisson with constant λ

H_A: a time dependent Poisson with rate $\lambda(x) = \lambda(\lambda x)^a$

Equation (3.8.1) can be considered as an alternative to 3.2 or 3.3. From 3.3, if an observation χ is taken at random on a variable distributed uniformly in the interval (0,1), then $-2\log \chi$ is distributed as chi-square with 2 d.f. It follows that

$$-2 \sum_{1}^{r} \log \left[\frac{n(y_1)}{n(y_r)} \right] \tag{3.8.2}$$

where $n(y_r)$ is a preassigned cumulative sum and r is a random variable. This can be used to test the hypothesis defined in B.

3.9 A test based on the maximum F distribution

H. O. Hartley (28, p. 308-312) has derived a quick test for homogeneity for the situation described in 3.7. This is given by

$$Z = \frac{\mathrm{MAX}[n(y_{1r}), n(y_{2r} - y_{1r}), \ldots , n(y_{kr} - y_{(1-1)r}]}{\mathrm{MIN}[n(y_{1r}), n(y_{2r} - y_{1r}), \ldots , n(y_{kr} - y_{(1-1)r}]} \tag{3.9.1}$$

One rejects the hypothesis of homogeneity if Z is too large. A table derived by H. O. Hartley (19, p. 308-312) gives the 5 percent points for Z for various values of k and r. When r - 1

$$Z = \frac{\mathrm{MAX}[n(y_1), n(y_2 - y_1), \ldots , n(y_k - y_{k-1})]}{\mathrm{MIN}[n(y_1), n(y_2 - y_1), \ldots , n(y_k - y_{k-1})]} \tag{3.9.2}$$

which is an alternative test to the one given in 3.8. A table derived by Hartley gives 5 percent and 1 percent points values.

3.10 Tests for abnormally long intervals in which there are no observation

This is a general case of 3.5. The purpose is to test whether any of the intervals between successive observations are too long.

Given $y_1 \leq y_2 \leq \ldots \leq y_n$ are the n independent observations from $f(y) = \lambda e^{-\lambda y}$, $y \S \gamma \S 0$. The statistic

$$Z = \frac{\underset{1 \leq i \leq n}{\mathrm{MAX}} (y_1)}{\sum_1^n y_i}$$

has the distribution of the form

$$P_r (Z > Z_o) = \sum_{k=1}^{r} \binom{n}{k} (-1)^{k-1} (1-k Z_o)^{n-1} \qquad (3.10.1)$$

where r is the largest integer less than 1/Z. One rejects the hypothesis of homogeneity if Z is too large. A table derived by R. A. Fisher (21, p. 54-59) gives values of Z_o such that $P_r(Z > Z_1) \leq \alpha$ for $\alpha = .05$ for values of n up to 50.

In addition, Cochran (9, p. 47-52) dealt with a more general case of similar nature. Given a sample from a normal distribution with variance σ^2, k independent variance estimates of σ^2, $s_1^2, \ldots, s_k^2$, each based on v d.f. gives the statistic

$$g_i = \frac{s_i^2}{\sum_{i=1}^{n} s_i^2} \qquad i = 1,2,\ldots,k \qquad (3.10.2)$$

Since $\frac{v\, s_i^2}{\sigma^2}$ is distributed as $\chi^2(v)$, (3.10.1) becomes a special case of (3.10.2) when k = n and v = 2.

Example 3.10.1: Given $n(y_1) = 68$, $n(y_2) = 1516$, $n(y_3) = 49$, $n(y_4) = 22$, $n(y_5) = 358$. Using Equation (3.10.1), $Z_1 = \frac{1516}{2013} = .753$. According to a table given by R. A. Fisher, $P_r(Z \geq .684) = .05$. Since $Z_1 > Z$, this proved that $n(y_2 - y_1)$ is abnormally long.

3.11 A graphical procedure based on the Kolmogorov-Smirnov test

If a test is terminated after a preassigned cumulative sum (y_r), then the observations can be arranged in ascending order of magnitude $y_1 \leq y_2 \leq \ldots \leq y_r$. The associated cumulative sums $n(y_1)$, $\ldots$, $n(y_r)$ are uniformly distributed over $(0, n(y_r))$, providing the underlying p.d.f. is exponential with constant parameter γ. From G. A. Barnard (2, p. 212-213), one can plot a random walk diagram in which the proportion of observations (relative to r) that occur at or before $n(y_r)$ is plotted against $n(y_i)$. The random walk diagram should fluctuate around the straight line joining the points (0,0) and $(n(y_r),1)$, providing the hypothesis is accepted. The results of Kolmogotov and Smirnov (47, p. 279-281) are used here by B. Epstein which gives

$$
{}_rD_r = r \max_{0 < n(y_i) < n(y_r)} \left| F_r(n(y)) - F(n(y)) \right|
$$

$$
= \max_{0 < n(y_i) < n(y_r)} \left| (\text{number of observations up to } n(y_r) - \frac{rn(y)}{n(y_r)} \right|
$$

where

$$
{}_rF_r\,(n(y)) = \text{number of observations} \leq n(y_r)
$$

$$
F\,(n(y)) = \frac{n(y)}{n(y_r)} \text{ for } 0 \leq n(y_i) \leq n(y_r)
$$

Epstein (14, p. 95), Birnbaum (7, p. 425-441), and Massey (35, p. 68-78) give values for accepting or rejecting the hypothesis.

3.12 A test based on conditional probability

The characteristic property of the exponential distribution is that the conditional probability of occurrence of an event in the interval $(\chi, \chi + \Delta\chi)$, given that it has occurred up to time χ, is independent of χ. The conditional probability is given by

$$\frac{f(x)\Delta x}{1 - F(x)} = \frac{\lambda e^{-\lambda x}\ \Delta x}{e^{-\lambda x}} = \lambda\Delta x \qquad (3.12.1)$$

When n is large and is divided into intervals $(0,\chi_i)$, (χ_i, χ_{2i}) and if n_1, n_2, ... are the numbers of observations in each interval, then

$$\frac{n_1}{N}, \frac{n_2}{N-n_1}, \frac{n_3}{N-n_1-n_2}, \ldots, \frac{n_k}{N-n_1-n_2- \ldots n_{k-1}}$$

should fluctuate within reasonable limits about a constant value, i.e., the parameter γ.

ESTIMATION OF THE DISTRIBUTION PARAMETERS

In this section, theorems relevant to the optimum properties of the estimators of the parameters of exponential distribution, the distribution of these estimators, and the sufficiency and completeness of a sample used for estimation are introduced together with methods of estimation.

4.1 Theorems relevant to the estimators and their distributions

B. Epstein and M. Sobel (18, p. 373-381) developed several lemmas concerning the first r out of n observations when the common underlying p.d.f. for each observation is given by

$$f(x;\theta,\lambda) = \lambda e^{-\lambda(x-\theta)} \qquad \theta \geq 0,\ 0 \leq x < \infty,\ \gamma > 0 \tag{4.1.1}$$
$$= 0$$

N observations are divided into k sets s_i (each containing $n_i > 0$, $\sum_{i=1}^{k} n = N$) and for each set s_i, only the first r_i observations are taken ($0 < r_i \leq n_k$). Three cases are considered.

1. The n_i items in each set s_i have a common known $\theta_i (i = 1,2,\ldots k)$

2. All n observations have a common unknown θ

3. The n_i observations in each set s_i have a common unknown θ_i $(i = 1,2,\ldots k)$

The reason for classifying n observations into the three cases is that they include as many as possible of types of situations when the underlying distribution is exponential. This can be seen from one-parameter exponential distribution of Equation (2.1.1) which is the case when k = 1 and θ = 0.

Let $y_1 \leq y_2 \leq \ldots \leq y_r$ denote the r smallest ordered observations from (4.1.1). The joint p.d.f. of the first r out of n observations is given by

$$P\ (y_1,y_2)\ldots y_r;\theta,\gamma) = \frac{n!}{(n-r)\gamma^r}\ e^{\frac{1}{\gamma}[\sum_{i=1}^{r}(y_i-\theta) + (n-r)(y_r-\theta)]}$$

$$\theta \leq y_1 \leq \ldots \leq y_r < \infty \qquad (4.1.2)$$

Lemma 1. For $1 \leq s \leq r \leq n$, the conditional joint density of

$$z_i = y_{i+1} - y_s \qquad i = s,\ s+1,\ \ldots\ ,\ r-1$$

given $Y_s = y_s$ (as well as the unconditional joint density), is (4.1.2) with (n,r,θ) replaced by $(n-s,r-s,0)$ respectively.

Lemma 2. For $1 \leq r \leq n$ and any preassigned constant $c \geq \theta$, the conditional joint density of the set $Y_i' = Y_i - c$ $(i = 1, 2,\ldots,r)$, given that $Y_1 \geq c$, is (4.1.3) with θ replaced by 0.

Lemma 3. For $1 \leq r \leq n$, the set of random variables

$$W_i - (n-i\_1)(Y_i - Y_{i-1}) \qquad i = 1,2,\ldots r$$

are mutually independent with common p.d.f. (4.1.1) except that $\theta = 0$.

These lemmas together with several corollaries (18, p. 373-375) lead to the following theorems.

Theorem 1. The distribution of the maximum likelihood $\hat{\gamma}$ estimator of (4.1.1) depends only on R, γ (and also on k in Case 3), where $R = \sum_{i=1}^{k} r_i$. The random variable $\frac{2R\hat{\gamma}}{\gamma}$ is distributed as a chi-square with (2R-2) d.f., and chi-square with (2R-2k) d.f., and chi-square with (2R-2k) d.f. in Case 1, 2, and 3 respectively.

The unbiased estimate $E(\gamma_i)$ for Cases 1, 2, and 3, respectively, are

$$E(\hat{\gamma}_1) = \hat{\gamma}_1, \quad E(\hat{\gamma}_2) = \frac{R\hat{\gamma}_2}{(R-1)}, \quad E(\hat{\gamma}_3) = \frac{R\hat{\gamma}_3}{(R-K)} \tag{4.1.3}$$

which depends on the observations z_i $(i = 1,2,3)$ only where z_i are given by

$$Z_1 = \sum_{i=1}^{k} \Big[\sum_{j=1}^{r_i} Y_{ij} + (n_i + r_i)\, Y_i r_i \Big] = R\hat{\gamma}_1 + \sum_{i=1}^{k} n_{1\theta 1}$$

$$Z_2 = (Z_{20}, Z_{21}), \; Z_{20} = Z_1, \; Z_{21} = \min_i Y_{i1}$$

$$Z_3 = (Z_{30}, Z_{31}, \ldots, Z_{3k}), \; Z_{30} = Z_1, \; Z_{3i} = Y_{i1}, \; i = 1, 2, \ldots, k$$

The author proved that $E(\hat{\gamma}_i)$ are uniformly minimum variance unbiased estimates by showing that z_i is complete and sufficient for estimating γ_i in three cases.

Theorem 2. z_1 is sufficient and complete for estimating γ.

Proof: Sufficiency is proved by showing that the joint density in

Case 1 can be written as $g(y, \ldots, y_n;\theta,\gamma) = h(\hat{\theta},\hat{\gamma};\theta,\gamma)\ k\ (y_1, y_2, \ldots, y_n)$ where $k(y_1, y_2, \ldots, y_n)$ do not involve the parameter θ, and γ (38, p. 170-171). Completeness is proved by using one sided Lapace transforms (18, p. 377-378).

Theorem 3: $z_z = (Z_{20},Z_{21})$, is sufficient and complete for estimating the pair (θ,γ).

Proof: Sufficiency is proved in the usual way as in the previous case. Completeness is also proved similarly except a two-dimensional uniqueness theorem for Lapace transforms is used (18, p. 378-379).

So far, discussion has been on the varied properties of the m.l.e. $\hat{\theta}$ and $\hat{\gamma}$. But quite a few empirical phenomena which follow the exponential law are, to some extent, involved in what can be called as a sequential process of time (or space) such that the observations can be arranged by order. The knowledge of order statistics, then, becomes important. A. G. Laurent, A. E. Sarhan, etc., have contributed their studies in the form Lemmas and Theorems. These Lemmas and Theorems are used somewhat through the report (32, p. 652-657; 45, p. 844-906; 44).

4.2 Method of maximum likelihood

W. L. Deemer and D. F. Votaw (12, p. 498-504) have derived the m.l.e. $\hat{\theta}$ and $\hat{\gamma}$ $(1/\hat{\lambda})$ of θ and γ of a single exponential.

A. Based on a complete sample. Given the p.d.f. of a one-parameter exponential by

$$f(x;\lambda) = \lambda e^{-\lambda x} \qquad \lambda > 0,\ 0 \leq x < \infty$$

The estimator $\hat{\lambda}$ of λ based on a complete sample of n observations arranged in ascending order according to their magnitude $y_1 \leq y_2 \leq \cdots \leq y_n$ is given by

$$\hat{\lambda} = \frac{1}{\bar{y}} \tag{4.2.1}$$

Proof: The likelihood function $L(\lambda)$ of λ is given by

$$L(\lambda) = (\lambda e^{-\lambda\bar{y}})^N \tag{4.2.2}$$

where $\bar{y}$ is the sample mean. It is easily shown by $\frac{\partial \log L(\lambda)}{\partial\lambda} = 0$ that (4.2.1) holds.

The asymptotic variance of $\sqrt{n}(\hat{\lambda}-\lambda)$ is

$$\frac{1}{-E[\frac{\partial^2 \log f(y)}{\partial\lambda\partial\lambda}]}$$

(38, p. 174-499; 48, p. 136-139; 37, p. 208-212). From (4.2.1), this equals to λ^2. Therefore, for large n

$$\text{VARIANCE } [\sqrt{n}(\hat{\lambda}-\lambda)] = \lambda^2 \tag{4.2.3}$$

From section 4.1, $\frac{2n\hat{\lambda}}{\lambda}$ has a chi-square distribution with 2n d.f. This can be used to construct the confidence interval, and derive a test for testing hypothesis.

<u>B. In a censored sample</u>. If a sample from (4.2.1) is censored to the right, only the first r observations $y_1 \leq y_2 \leq \cdots \leq y_r$ are used in the estimation.

The m.l.e. of λ is given by

$$\hat{\lambda}_c = \frac{r}{[(n-r)\,y_r + \sum_1^r y_i]} \tag{4.2.4}$$

Proof: The p.d.f. under the condition is given by

$$f(x;\lambda) = \lambda e^{-\lambda x} \qquad 0 < x \leq x_r \tag{4.2.5}$$

$$P_r(x > x_r) = e^{-\lambda x_r} = \eta \tag{4.2.6}$$

The likelihood function $L(\lambda_c)$ of λ_c is consequently

$$L(\lambda_c) = \begin{cases} \dfrac{r!}{(n-r)!\; r!}\, \lambda^r e^{-\lambda \sum_1^r y_i - (n-r)\lambda y_r} & r > 0 \\ e^{-r\lambda y_r} & r = 0 \end{cases} \tag{4.2.7}$$

the partial derivative of $\log L(\lambda_c)$ with respect to λ and set the result to zero gives (4.2.5).

Halperin (25, p. 226-238) proved that the $\hat{\lambda}_c$ in (4.2.5) has the properties of consistency, asymptotic normality, and minimum asymptotic variance.

The asymptotic variance of $\sqrt{n}(\hat{\lambda}_c-\lambda)$, from the results of Halperin, is reciprocal of

$$\int_0^{y_r} \left[\frac{\partial \log f(y)}{\partial\lambda} \right]^2 f(y)\,dy + \eta \left[\frac{\partial \log \eta}{\partial\lambda} \right]^2 \tag{4.2.8}$$

where η is defined in (4.2.6). The expression in (4.2.8) then becomes

$$\gamma^2 (1 - e^{-\lambda y_r}) \tag{4.2.9}$$

For large n,

$$\text{VARIANCE } [\sqrt{n}\,(\hat{\lambda}_c - \lambda)] = \frac{\lambda^2}{1 - e^{-\lambda y_r}}$$

which is naturally always greater than the asymptotic variance in 4.1 which is based on a complete sample.

Example 4.2.1: The first 10 failure times were observed from 20 items placed on a life test (16, p. 405-406).

Assuming that the underlying distribution is exponential, it is found that

$$\hat{\gamma} = \frac{[(n-r)y_r + \sum_1^r y_1]}{r-1} = \frac{12000}{9} = 1333 \tag{4.2.10}$$

and $\hat{\theta}$ is found by

$$\hat{\theta} = y_1 - \frac{\hat{\gamma}}{n} = 452.3 \tag{4.2.11}$$

<u>C. In a truncated sample</u>. If the population is truncated to the right, the number of observations greater than y_r, the truncated point, is usually unknown. Therefore, the p.d.f. of f(y), given $y \leq y_r$ is a conditional density and is given by

$$f(y) = \lambda e^{-\lambda y} (1 - e^{-\lambda y_r})^{-1} \qquad 0 < y \leq y_r \tag{4.2.12}$$

The m.l.e. $\hat{\lambda}_t$ of (4.2.12) is given by

$$\hat{\lambda}_t = \begin{cases} \frac{1}{\gamma'} & 0 < \bar{y} < \frac{1}{2} y_r \\ 0 & \bar{y} \geq \frac{1}{2} y_r \end{cases} \tag{4.2.13}$$

Proof: Let the r observations in the sample be arranged in ascending order such that $y_1 \leq y_2 \leq \cdots \leq y_r$, the likelihood function $L(\lambda_t)$ is given by

$$L(\lambda_t) = \frac{c^n}{(1 - e^{-\lambda y_r})^n} e^{-\lambda \sum_1^n y_i} = [\lambda e^{-\lambda \bar{y}} (1 - e^{-\lambda y_r})^{-1}]^n \tag{4.2.14}$$

the partial derivative of $\log L(\lambda_t)$ with respect to λ gives

$$n[\gamma - y_r e^{-\lambda y_r} (1 - e^{-\lambda y_r})^{-1} - \bar{y}] \tag{4.2.15}$$

The function in the bracket is monotonic decreasing in λ. It tends to $\frac{1}{2} y_r$ as λ tends to zero, and tends to zero as λ tends to infinity. When $0 \leq \bar{y} < \frac{1}{2} y_r$, there exists a solution $\frac{1}{\gamma'}$ by setting (4.2.15) to zero. When $\bar{y} \geq \frac{1}{2} y_r$, (4.2.15) assumes its maximum value for $\lambda = 0$.

The asymptotic variance of $\sqrt{n}\,(\hat{\lambda}_t - \lambda)$ equals to

$$\frac{1}{-E\left[\frac{\partial^2 \log f(y)}{\partial\lambda\partial\lambda}\right]} \tag{4.2.16}$$

where f(y) is given by (4.2.12).

For large n,

$$\text{VARIANCE } [\sqrt{n}\,(\hat{\lambda}_t - \lambda) = \frac{1}{[\gamma^2 - y_r^2 e^{-\lambda y_r} (1-e^{-\lambda y_r})^{-2}]} \qquad (4.2.17)$$

D. Interval estimation of λ_c. When sample size is large, the following approximation gives an approximate 100 (1-q) percent confidence limits for λ in the censored case (4.2.7).

$$P_r\,(-z_q < z < z_q) = P \qquad (4.2.18)$$

where z_q is the 100 (1-q) percent point of the standard normal distribution and

$$z = \frac{\sqrt{n}\,(\hat{\lambda}_c - \lambda)}{[\lambda(1-e^{-\lambda y_r})^{-\frac{1}{2}}]} \qquad (4.2.19)$$

A similar 100 (1-q) percent confidence limits for λ_t in truncated case can be obtained by

$$P_r\,(-z_q < z < z_q) = P \qquad (4.2.20)$$

where

$$y = \sqrt{n'}\,(\hat{\lambda}_t - \lambda)\,[\gamma^2 - y_r^2 e^{-\lambda y_r} (1-e^{-\lambda y_r})^{-2}]^{\frac{1}{2}} \qquad (4.2.21)$$

Examples of finding confidence interval for $\hat{\theta}$ and $\hat{\gamma}$ has been given by B. Epstein (16, p. 406; 15, p. 448-451). He used equations derived under slightly different assumptions from those ones given above. But, nevertheless, they are very much the same in the sense that they are m.l.e. of optimum characteristics.

4.3 Method of least-squares

Lloyd (33, p. 88-95) has discussed the possibility of applying general least-squares theory to obtain the estimates for the parameters of the distributions which depend on location and scale parameter only. Using an ordered sample, the resulting estimates by method of least-squares are unbiased, linear in the ordered observations, and of minimum variance. Formulae are obtained for the estimates and for their variances and covariance.

Let $(x_1, x_2, \ldots, x_n)$ be a sample of n independent observations from a distribution which depends only on location and scale parameters θ and λ. The observations in the sample can be arranged in ascending order of magnitude $(y_1, y_2, \ldots, y_n)$ such that $y_1 \leq y_2 \leq \cdots \leq y_n$ the standardized variate

$$z_i = \lambda(y_i - \theta) \tag{4.3.1}$$

transforms the ordered set $(y_1, y_2, \ldots, y_n)$ into $(z_1, z_2, \ldots, z_n)$ such that

$$z_1 \leq z_2 \leq \cdots \leq z_n \qquad 1 \leq i \leq n$$

Let $\varepsilon(z_i) = u_i$

$$\text{VARIANCE } (z_i) = V_{ii} \qquad \text{COVARIANCE } (z_i, z_j) = V_{ij} \tag{4.3.2}$$

which, given the form of the parent distribution, will have known values irrelevant of the parameters θ and λ.

A linear transformation of (4.3.2) by the relation of (4.3.1) gives

$$E(y_i) = \Theta + \gamma U_i, \text{ var } (y_i) = \gamma^2 V_{ii}, \text{ voc}(y_i, y_j) = \gamma^2 V_{ij}$$

have expectation which are linear functions of the parameters θ and λ with known coefficients. The variance and covariances of these expectations are known up to a scale factor γ^2. From Aitken's paper "On least-squares and linear combination of observation," it is known that the least-squares theorem of Gauss and Markoff can be applied to the situation. The parameters thus estimated by the linear functions of y_i are unbiased, linear, and have minimum variance.

Equation (4.3.3) in matrix form can be written as

$$\varepsilon(Y) = \theta I + \gamma U \tag{4.3.4}$$

Y is the vector of the y_i, I a vector of identity, and U the vector of the U_i. The equation can be written more compactly as

$$\varepsilon(Y) = P\theta' \tag{4.3.5}$$

where P is the (nx2) matrix (1,U), and $\theta' = (\theta,\gamma)$.

The variance matrix of the y_i is

$$V(Y) = \gamma^2 V$$

where V is the (nxn) symmetric positive-definite matrix of the V_{ij}.

The estimator of the vector θ' of the parameters is easily shown to be

$$\hat{\gamma} = (P'WP)^{-1} P'WY \tag{4.3.6}$$

where

$$W = V^{-1}$$

The variance matrix of the estimates is $(P'WP)^{-1}\gamma^2$, where

$$P'WP = \begin{pmatrix} 1'W1 & 1'WU \\ 1'WU & U'WU \end{pmatrix} \tag{4.3.7}$$

The inverse of P'WP is given by

$$(P'WP)^{-1} = \frac{1}{D} \begin{pmatrix} U'WU & -1'WU \\ -1'WU & 1'W1 \end{pmatrix} \tag{4.3.8}$$

Where D is the determinant of the matrix P'WP.

By (4.3.6), the estimates are given by

$$\hat{\theta} = U'GY, \; \hat{\gamma} = 1'GY \tag{4.3.9}$$

where G is the skew-symmetric matrix given by

$$G = [W(1U' - U1')\,W]\,D^{-1} \tag{4.3.10}$$

The variance and the covariance of $\hat{\theta}, \hat{\gamma}$ are given by

$$\text{VARIANCE } (\hat{\theta}) = \frac{U'Wu\gamma^2}{D}, \quad \text{var } (\hat{\gamma}) = \frac{1'W1\gamma^2}{D},$$

$$\text{cov } (\hat{\theta},\hat{\gamma}) = \frac{-1'WU\gamma^2}{D} \qquad (4.3.11)$$

F. Downton (13, p. 457-458), using the results obtained by Lloyd, generated the least-squares estimator for one-parameter exponential distribution. The equations in the two-parameter case are reduced to one-parameter case which measures the dispersion of the distribution. The ordered least-squares estimate $\hat{\gamma}$ of γ is given by

$$\hat{\gamma} = \frac{U'WX}{U'WU}, \quad \varepsilon\left(\frac{X}{\gamma}\right) = \varepsilon(Y) = U, \text{ VAR } (Y) = V, \ W = V^{-1}$$

He later extended the conditions described by Lloyd to unsymmetrical case. He derived the least-squares estimate $\hat{\gamma}$ of γ by using Cauchy-Schwarz inequality (39, p. 363-364). Again, he obtained $\hat{\gamma} = \frac{1'Y}{n}$. The estimator is minimum variance unbiased estimate of γ in the Newman-Pearson sense (38, p. 292).

4.4 Method of moments

The estimators derived by the method of moments are usually the same as the maximum-likelihood estimators. Under quite general conditions, they are

1. (sinple) consistent estimators and squared-error consistent estimators

2. asymptotically normal but not, in general (asymptotically), efficient nor BAN (38, p. 186).

Rider (42, p. 143-147) derived estimators by the method of moments for a mixed exponential distribution. It can be reduced to a single exponential when $p = 1$ and $q = 1 - p = 0$ with just a slight adjustment.

Suppose that two populations of the form

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & \gamma > 0, \quad 0 \leq x < \infty \\ 0 & \text{elsewhere} \end{cases}$$

with parameters λ_1 and λ_2 respectively have been mixed in the unknown proportions p and q. The resulting p.d.f. is given by

$$f(x;\ \lambda_1,\lambda_2) = p\lambda_1 e^{-\lambda_1 x} + q\lambda_2 e^{-\lambda_2 x} \tag{4.4.2}$$

A simple method of estimating the three parameters p, λ_1, λ_2, is derived from the first three moments of a sample. Let m_1', m_2', m_3' denote the first, second, and the third moment about zero of a random sample of size $n(x_1, x_2, \ldots, x_n)$ from (4.4.2). The estimators $\hat{P}$, $\hat{\gamma}_1$, $\hat{\gamma}_2$ of P, γ_1, γ_2 are obtained by equating $m_i' = U_i'$ for $i = 1,2,3$, and U_i' is the population moment about zero from (4.4.2) (38, p. 186).

$$\hat{P}\hat{\gamma}_1 + \hat{q}\ \hat{\gamma}_2 = m_1' \tag{4.4.3}$$

$$\hat{P}\hat{\gamma}_1^2 + \hat{q}\hat{\gamma}_2^2 = \frac{1}{2}\ m_2' \tag{4.4.4}$$

$$\hat{P}\hat{\gamma}_1^3 + \hat{q}\hat{\gamma}_2^3 = \frac{1}{6} m_3' \qquad (4.4.5)$$

After rearrangement, (4.4.3) becomes

$$\hat{P} = \frac{m_1' - \hat{\gamma}_2}{\hat{\gamma}_1 - \hat{\gamma}_2} \qquad (4.4.6)$$

Substituting (4.4.6) for $\hat{P}$ in (4.4.4) and (4.4.5) gives

$$(m_1' - \hat{\gamma}_2)(\hat{\gamma}_1 + \hat{\gamma}_2) = \frac{1}{2} m_2' - \hat{\gamma}_2^2 \qquad (4.4.7)$$

$$(m_1' - \hat{\gamma}_2)(\hat{\gamma}_1^2 + \hat{\gamma}_1 \hat{\gamma}_2 + \hat{\gamma}_2^2) = \frac{1}{6} m_3' - \hat{\gamma}_2^2 \qquad (4.4.8)$$

From the preceding two equations we can obtain an equation for $\hat{\gamma}_i$ $(i = 1,2)$ in terms of m_1', m_2', and $\hat{\gamma}_j$ $(j = 2$, or 1 according to $i = 1$, or $2)$. Substituting the $\hat{\gamma}_i$ in (4.4.8) will, after some simplifications, give an equation for $\hat{\gamma}_j$

$$6(2m_1'^2 - m_2')\hat{\gamma}_j^2 + 2(m_3' - 3m_1'm_2')\hat{\gamma}_j + 3m_2'^2 - 2m_1'm_3' = 0 \qquad (4.4.9)$$

The quadratic equation (4.4.9) yields two roots, $\hat{\gamma}_1$ and $\hat{\gamma}_2$. P is then found by (4.4.6).

The variances of these estimators, although not impossible, are not easy to obtain due to difficulty in calculation. However, in order to have some idea of the reliability of the estimators $\hat{\gamma}_1$ and $\hat{\gamma}_2$, let us assume that P is known and the variances of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ will be derived under this assumption.

According to Cramer (10), the asymptotic variance of $\hat{\gamma}_1$ is found by

$$\text{VARIANCE}(\hat{\gamma}_1) = V(m_1')\left(\frac{\partial\hat{\gamma}1}{\partial m_1'}\right)^2 + 2\ \text{COV}\ (m_1', m_2')\ \frac{\partial\hat{\gamma}1}{\partial m_1'}\ \frac{\partial\hat{\gamma}2}{\partial m_2'}\ +$$

$$+\ V(m_2')\left(\frac{\partial\hat{\gamma}1}{\partial m_2'}\right)^2 \tag{4.4.10}$$

where $V(m_1')$ and $V(m_2')$ are the variances of m_1' and m_2' respectively. $\text{COV}(m_1', m_2')$ is the covariance of these two moments. The partial derivatives are evaluated at the point

$$m_1' = P\gamma_1 + q\gamma_2$$

$$m_2' = 2(P\gamma_1^2 + q\gamma_2^2) \tag{4.4.11}$$

The formulas given by Kendell (30) are used to find the value of the coefficient of the partial derivative in (4.4.11).

$$V(m_1') = n^{-1}\ [(2P-P^2)\gamma_1^4 - 2Pq\gamma_1^2\gamma_2^2 + (2q - q^2)\gamma_2^4] \tag{4.4.12}$$

$$V(m_2') = 4n^{-1}\ [(6P-P^2)\ \gamma_1^4 - 2Pq\gamma_1^2\gamma_2^2 + (6q - q^2)\ \gamma_2^4\ ] \tag{4.4.13}$$

$$\text{COV}(m_1', m_2') = 2n^{-1}\ [(3P-P^2)\gamma_1^3 - Pq\gamma_1^2\gamma_2 - Pq\gamma_1\gamma_2^2 +$$

$$+\ (3q - q^2)\ \gamma_2^3\ ] \tag{4.4.14}$$

The partial derivatives at the point (4.4.11) have the values

$$\frac{\partial \hat{\gamma}_1}{\partial m_1} = \frac{-\gamma_2}{P(\gamma_1-\gamma_2)} \qquad \frac{\partial \hat{\gamma}_1}{\partial m_2'} = \frac{1}{4P(\gamma_1-\gamma_2)} \tag{4.4.15}$$

Substituting (4.4.12), (4.4.13), (4.4.14), and (4.4.15) in (4.4.10), after some simplification, gives the variance of the asymptotic distribution of $\hat{\gamma}_1$

$$\frac{1}{4nP^2(\gamma_1-\gamma_2)}[P(6-P)\gamma_1^4 - 4P(3-P)\gamma_1^3\gamma_2 + 2P(5-3P)\gamma_1^2\gamma_2^2 -$$

$$- 4P(1-P)\gamma_1\gamma_2^3 + (1-P^2)\gamma_2^4] \tag{4.4.16}$$

The variance of the asymptotic distribution of $\hat{\gamma}_2$ is obtained by substituting q for P and interchanging γ_1 and γ_2 in (4.4.16).

4.5 Simplified estimates

When the parameters can be estimated from a linear function of specific subsets of the order statistics, these estimators are optimal in that they provide the most efficient linear combinations of a given number of order statistics.

A. Based on one ordered statistic. H. L. Harter (27, p. 1078-1084) derived an estimator $\hat{\gamma}$ of γ for the one-parameter exponential population based on one ordered statistic from a sample of any size up through n = 100. The minimum variance unbiased estimate of γ is, of course, the sample mean $\bar{y}$ with variance = $\frac{\gamma^2}{n}$.

The expected value and the variance of the kth order statistic of a sample of size n is given by

$$E(y_k) = \gamma \sum_{1}^{k} a_i \tag{4.5.1}$$

$$\text{VARIANCE } (y_k) = \gamma^2 \sum_{1}^{k} a_i^2,\ a_i = \frac{1}{(n-i+1)} \tag{4.5.2}$$

An unbiased estimator of γ, based on the ordered statistic y_k, is

$$\hat{\gamma}_k = b_k\, y_k,\quad b_k = \frac{1}{\sum_{1}^{k} a_i} \tag{4.5.3}$$

$$\text{VARIANCE}(\hat{\gamma}_k) = \frac{\gamma^2 \sum_{1}^{k} a_i^2}{\left(\sum_{1}^{k} a_i\right)^2} \tag{4.5.4}$$

And its efficiency (relative to the minimum variance unbiased estimator $\bar{y}$) is

$$\text{EFF}(\hat{\gamma}_k) = \frac{\text{VARIANCE}(\bar{y})}{\text{VARIANCE}(\hat{\gamma}_k)} \tag{4.5.5}$$

It then follows that the best estimator of γ, based on one ordered statistic y_k, is the one for that value of k which minimizes variance $\hat{\gamma}_k$ (maximizes $\text{EFF}(\hat{\gamma}_k)$). A table given by H. L. Harter (27, p. 1085) gives the value of k, b_k, $(\text{VAR}(\hat{\gamma}_k))/\gamma^2$, and $\text{EFF}(\hat{\gamma}_k)$. To determine the value of k which yields the best estimator of γ (i.e., after setting up the equation for the relative efficiency of a linear combination of one or two ordered statistics), an analytical method for determining the best combinations is given by M. M. Siddiqui (46, p. 117-121).

The method is based on the Euler-Maclaurin formula (29, p. 281).

$$\sum_{r=0}^{k-1} f(r) = \int_0^k f(x)dx - \frac{1}{2}[f(k) - f(0)] + (\frac{1}{12})[f^{(1)}(k) - f^{(1)}(0)] - (\frac{1}{720})$$

$$[f^{(3)}(\frac{1}{720})[f^{(3)}(k) - f^{(3)}(0)] + \ldots \qquad (4.5.6)$$

Substituting $z = \frac{n-k+1}{n-1}$ in (4.5.6), and after a lengthy derivation, it is found that the optimum k is the nearest integer to

$$(n+1)(1-z_o) \simeq 0.79681\ (n+1) - 0.39841 + 1.16312\ (n+1)^{-1}$$

where

$$z_o = 0.20319 + 0.39841\ (n+1)^{-1} - 1.16312\ (n+1)^{-2} \qquad (4.5.7)$$

A quick check verifies the correctness of this method to the table given by H. L. Hartley (27, p. 1078-1090).

B. Based on two ordered statistics. A. E. Sarhan, B. G. Greenburg, and J. Ogawa (44, p. 102-116) have discussed the use of only two observations from a sample up to size 20 to construct a best linear combination for estimating the parameters of an exponential distribution.

$$\text{Given } f(x) = \lambda e^{-\lambda(x-\theta)} \qquad 0 \le \theta \le x$$
$$= 0 \qquad \text{elsewhere}$$

The minimum variance unbiased linear estimates for θ and γ based on two ordered statistics y_ℓ, y_m, where $y_1 \le y_\ell \le y_m \le y_n$, from a sample of size n is given by

$$\hat{\theta} = \frac{1}{\sum_{\ell+1}^{m} a}\left[y_\ell \sum_{1}^{m} a - y_m \sum_{1}^{\ell} a\right] \tag{4.5.8}$$

$$\hat{\gamma} = \frac{1}{\sum_{\ell+1}^{m} a}(y_m - y_\ell) \tag{4.5.9}$$

Proof: If the matrix of coefficients of (θ,γ) for the expected value (B) of the n order statistics and the variance-covariance matrix V are given by

$$B = \begin{bmatrix} 1 & \sum_{1}^{\ell} a \\ & \\ 1 & \sum_{1}^{m} b \end{bmatrix} \qquad V = \begin{bmatrix} \sum_{1}^{\ell} b & \sum_{1}^{\ell} b \\ & \\ \sum_{1}^{\ell} b & \sum_{1}^{m} b \end{bmatrix} \tag{4.5.10}$$

where

$$\sum_{1}^{\ell} a = \sum_{1}^{\ell} \frac{1}{(n-i+1)}$$

and

$$\sum_{1}^{\ell} b = \sum_{1}^{\ell} \frac{1}{(n-i+1)^2}$$

then (4.5.8) and 4.5.9) can be obtained from

$$(B'V^{-1}B)^{-1}B'V^{-1} = \left(\frac{1}{\sum_{\ell+1}^{m} a}\right)\begin{bmatrix} \sum_{1}^{m} a & - \sum_{1}^{\ell} a \\ & \\ -1 & 1 \end{bmatrix} \tag{4.5.11}$$

The variance for $\hat{\theta}$, $\hat{\gamma}$, taken from $(B^{-1}V^{-1}B)^{-1}$ are

$$\mathrm{VAR}(\hat{\theta}) = \{\sum_{1}^{\ell} b + [((\sum_{1}^{\ell} a)^2 \sum_{\ell+1}^{m} b)/(\sum_{\ell+1}^{m} a)^2]\}\ \gamma^2 \tag{4.5.12}$$

$$\mathrm{VAR}(\hat{\gamma}) = \{\frac{\sum_{\ell+1}^{m} b}{(\sum_{\ell+1}^{m} a)^2}\}\ \gamma^2 \tag{4.5.13}$$

The estimate of the linear combination $\hat{U} = \hat{\theta} + \hat{\gamma}$ is given by

$$\hat{U} = (\frac{1}{\sum_{\ell+1}^{m} a})\ \{(\sum_{1}^{m} a-1)\ y_\ell + (1 - \sum_{1}^{\ell} a)\ y_m\} \tag{4.5.14}$$

and the variance of $\hat{U}$ is given by

$$\mathrm{VAR}(\hat{U}) = \{[\sum_{1}^{\ell} b] + \frac{(\sum_{1}^{\ell} a-1)^2 \sum_{\ell+1}^{m} b}{(\sum_{\ell+1}^{m} a)^2}\}\ \gamma^2 \tag{4.5.15}$$

The last expression in (4.5.13) attains a minimum for $\ell = 1$ and m according to the values of n shown in a table given by Sarhan, etc. (44, p. 104). The same results hold for estimating θ and U.

For one-parameter exponential distribution (θ=0), the comparable equations are given by

$$\hat{\gamma} = \frac{(\sum_{\ell+1}^{m} b \sum_{1}^{\ell} a - \sum_{1}^{\ell} b \sum_{\ell+1}^{m} a)\ y_\ell + (\sum_{1}^{\ell} b \sum_{\ell+1}^{m} a)\ y_m}{(\sum_{1}^{\ell} a)^2 \sum_{\ell+1}^{m} b + \sum_{1}^{\ell} b\ (\sum_{\ell+1}^{m} a)^2} \tag{4.5.16}$$

$$VAR(\hat{\gamma}) = \left\{ \frac{\sum_{1}^{\ell} b \sum_{\ell+1}^{m} b}{(\sum_{1}^{\ell} a)^2 \sum_{\ell+1}^{m} b + \sum_{1}^{\ell} b \, (\sum_{\ell+1}^{m} a)^2} \right\} \gamma^2 \qquad (4.5.17)$$

The variance for estimating γ with two order statistics attains a minimum at different points than previously. The results are summarized in a table given by Sarhan et al. (44, p. 106).

To simplify the selection, estimation using two symmetric observations in a sample up to size 20 is available. Tables and figures given by Sarhan et al. (44, p. 107-115) give the relative efficiency of $\hat{\theta}, \hat{\gamma}$ (in one- and two-parameter case).

Later, H. L. Harter obtained an unbiased linear estimate of the parameter γ based on two order statistics, y_ℓ and y_m, up to sample size of 100 instead of 20 given by Sarhan et al. The equations are given by

$$\hat{\gamma}_{\ell m} = C_\ell y_\ell + C_m y_m \qquad (4.5.18)$$

where

$$C_\ell E(y_\ell) + C_m E(y_m) = \gamma$$

and

$$VAR(\hat{\gamma}_{\ell m}) = C_\ell^2 \, VAR(y_\ell) + C_m^2 \, VAR(y_m) + 2C_\ell C_m \, COV(y_\ell, y_m)$$

It can be shown that for given values of ℓ and m, the values of C_ℓ and C_m which yield the unbiased estimator $\hat{\gamma}_{\ell m}$ with minimum variance are

$$C_\ell = \frac{1}{\sum_1^\ell a_i + \eta \sum_1^m a_i}\ , \quad C_m = \eta C_\ell \tag{4.5.19}$$

where

$$\eta = \left(\sum_{\ell+1}^m a_i \sum_1^\ell a_i^2\right)\cdot\left(\sum_1^\ell a_i \sum_1^m a_i^2 - \sum_1^m a_i \sum_1^\ell a_i^2\right)^{-1} \tag{4.5.20}$$

It then followed by substituting (4.5.19) and (4.5.20) in (4.5.18), the minimum variance of $\hat{\gamma}_{\ell m}$ for given ℓ and m, is

$$\mathrm{VAR}(\hat{\gamma}_{\ell m}) = \frac{\gamma^2\left[(1+2\eta)\sum_1^\ell a_i^2 + \eta^2 \sum_1^m a_i^2\right]}{\left(\sum_1^\ell a_i + \eta \sum_1^m a_i\right)^2} \tag{4.5.21}$$

The efficiency of $\hat{\gamma}_{\ell m}$ relative to the minimum variance unbiased estimator $\bar{y}$ is

$$\mathrm{EFF}(\hat{\gamma}_{\ell m}) = \frac{\mathrm{VAR}(\bar{y})}{\mathrm{VAR}(\hat{\gamma}_{\ell m})} \tag{4.5.22}$$

The best $\hat{\gamma}$ based on y_ℓ and y_m is the one for those values of ℓ and m which minimize the variance $\hat{\gamma}_{\ell m}$ (i.e., maximize $\mathrm{EFF}(\hat{\gamma}_{\ell m})$). Again, M. M. Siddiqui provided a method for finding the optimum k (46. p. 117-121).

For estimating θ,γ for two-parameter exponential population using two ordered statistics, it is a direct consequence of the one-parameter case. The unbiased linear estimates are of the form

$$\hat{\theta}_{\ell m} = C_{\theta\ell}\, y_\ell + C_{\theta m}\, y_m \tag{4.5.23}$$

$$\hat{\gamma}_{\ell m} = C_{\gamma\ell}\, y_\ell + C_{\gamma m}\, y_m \tag{4.5.24}$$

$$\hat{U}_{\ell m} = C_{U\ell}\, y_\ell + C_{Um}\, y_m \tag{4.5.25}$$

Those values of ℓ and m which minimize variance $\hat{\theta}_{\ell m}$, $\hat{\gamma}_{\ell m}$, $\hat{U}_{\ell m}$ (i.e., maximize EFF($\hat{\theta}_{\ell m}$), EFF($\hat{\gamma}_{\ell m}$), and EFF($\hat{U}_{\ell m}$)) are the best estimators. Again, M. M. Siddiqui provided a similar method for obtaining the optimum k which is the closest integer to

$$n - nz_o + 1 \simeq 0.79681n + 0.60159 + 1.16312n^{-1} \tag{4.5.26}$$

where

$$z_o \simeq 0.20319 + 0.39841n^{-1} - 1.16312n^{-2}$$

C. In the small sample case. So far, discussion has been around the small sample situation for k = 1 and k = 2. For k greater than two in the small sample situation, G. Kulldorff (31, p. 1419-1431) gave estimators for one- and two-parameter exponential on the basis of suitably chosen ordered statistics.

First, in deriving the BLUE (beat linear unbiased estimator), he used the results from Gumbel and Sarhan (11, p. 317-328) which says

$$E(y_j) = \theta + \gamma \sum_1^k (n - j + 1)^{-1}$$

$$VAR(y_j) = COV(y_j, y_j') = \gamma^2 \sum_1^n (n - j + 1)^{-2} \quad j < j'$$

and

$$\delta_{ri} = \varepsilon(n-j)^{-r} \quad r = 1,2, \quad i = 1,2, \ldots , k$$

Estimators are then derived under three situations:

1. the BLUE of γ when θ is known,
2. the BLUE of θ when γ is unknown,
3. the BLUE of θ and γ when both are unknown.

Given $y_1, y_2, \ldots, y_k$, where $1 \leq 2 \leq \ldots \leq k$, the k ordered statistics from

$$f(x) = \begin{cases} \lambda e^{-\lambda(x-\theta)} & x > \theta \geq 0,\ \gamma > 0 \\ 0 & \text{elsewhere} \end{cases}$$

For (1), the BLUE of γ is

$$\hat{\gamma} = a_0\theta + \sum_1^k a_i\, y_i \tag{4.5.27}$$

where

$$a_i = \left(\frac{\delta_{1i}}{\delta_{2i}} - \frac{\delta_{1,i+1}}{\delta_{2,i+1}}\right) C^{-1} \qquad i = 0,1,\ \ldots,\ k$$

$$C = \frac{\sum_1^k \delta_{1i}^{\ 2}}{\delta_{2i}}$$

$$\mathrm{VAR}(\hat{\gamma}) = \frac{\gamma^2}{C}$$

For (2), the BLUE of θ is

$$\hat{\theta} = y_1 - \theta\delta_{11}, \qquad \mathrm{VAR}(\hat{\theta}) = \gamma^2\delta_{21} \tag{4.5.28}$$

For (3), the BLUE of θ and γ are

$$\hat{\gamma} = \sum_{1}^{k} b_i \, y_i \tag{4.5.29}$$

$$\hat{\theta} = y_1 - \hat{\gamma}\, \delta_{11} \tag{4.5.30}$$

where

$$b_i = - \frac{\delta_{12}}{C'\delta_{22}} \qquad i = 1$$

$$= \left(\frac{\delta_{1i}}{\delta_{2i}} - \frac{\delta_{1,i+1}}{\delta_{2,i+1}}\right) C'^{-1} \qquad i = 2,3, \ldots ,k$$

$$C' = \sum_{i=2}^{k} \frac{\delta_{1i}^{2}}{\delta_{2i}}$$

and the BLUE of $U = \theta + \gamma$ is given by

$$\hat{U} = y_1 + \hat{\gamma}\,(1 - \delta_{11}) \tag{4.5.31}$$

$$\mathrm{VAR}(\hat{U}) = \gamma^2[\delta_{21} + (1 - \delta_{11})^2 \, C'^{-1}] \tag{4.5.32}$$

The efficiency of $\hat{\theta}$, $\hat{\gamma}$, $\hat{U}$ when compared with the BLUE of θ, γ, and U based on all n observations is given by

$$\mathrm{EFF}(\hat{\theta}) = \frac{1}{n(n-1)(\delta_{21} + \frac{\delta_{11}^{2}}{C'}} \tag{4.5.33}$$

$$\mathrm{EFF}(\hat{\gamma}) = \frac{C'}{(n-1)} \tag{4.5.34}$$

$$\text{EFF}(\hat{U}) = \frac{1}{n[\delta_{21} + \frac{(1-\delta_{11})^2}{C'}]} \tag{4.5.35}$$

D. <u>In the large sample case</u>. The general large sample theory for estimating one-parameter exponential distribution based upon sample quantiles is discussed below (44, p. 103-116).

Given a sample of size n from

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0,\ \gamma > 0 \\ 0 & \text{elsewhere} \end{cases}$$

there are k fixed real numbers $q_1, q_2, \ldots, q_k$ such that $0 < q_1 < q_2 < \ldots < q_{k-1} < 1$. One can select the k sample q_i quantiles, where i = 1,2,...,k, to estimate γ (10, p. 179-180). The k ordered statistics are $y_1, y_2, \ldots, y_k$, where $i = [kq_i] + 1$ and $[kq_i]$ stands for the greatest integer not exceeding kq_i. Then the standardized exponential distribution is given by

$$g(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & \text{elsewhere} \end{cases} \tag{4.5.36}$$

If the q_i-quantile of the standardized distribution is U_i, so that $q_i = 1 - e^{-U_i}$ and the original distribution is x_i, then $x_i = U_i\gamma$, $i = 1,2, \ldots, k$. From Ogawa (44, p. 103-116), the asymptotically best linear unbiased estimate $\hat{\gamma}$ of γ is given by

$$\hat{\gamma} = \frac{D}{B} \tag{4.5.37}$$

where

$$D = \frac{\sum_{1}^{k+1} [U_i e^{-U_i} - U_{i-1} e^{-U_{i-1}}][y_i e^{-U_i} - y_{i-1} e^{-U_{i-1}}]}{[e^{-U_{i-1}} - e^{-U_i}]} = \sum_{1}^{k} C_i \, y_i$$

and

$$B = \frac{\sum_{1}^{k+1} [U_i e^{-U_i} - U_{i-1} e^{-U_{i-1}}]^2}{[e^{-U_{i-1}} - e^{-U_i}]}$$

$$C_i = e^{-U_i} \left\{ \frac{[U_i e^{-U_i} - U_{i-1} e^{-U_{i-1}}]}{[e^{-U_{i-1}} - e^{-U_i}]} - \frac{[U_{i+1} e^{-U_{i+1}} - U_i e^{-U_i}]}{[e^{-U_i} - e^{-U_{i+1}}]} \right\}$$

Simplification of (4.5.37) brought

$$\hat{\gamma} = \sum_{1}^{k} d_i \, y_i, \qquad d_i = \frac{C}{B} \tag{4.5.38}$$

A table is obtained for values of $d_i (i = 1, \ldots, 15)$, γ_i, and θ_i such that B is a maximum; i.e., values for determining the asymptotically optimum spacings for estimates, asymptotic relative efficiencies, and the coefficients of best estimates for the one-parameter exponential (44, p. 112-113).

H. L. Harter also obtained an estimate using sample quasi-ranges from a one-parameter exponential.

The rth quasi-range, q_u, of a sample of size n is defined as the range of (n-2r) sample value, omitting the r largest and the r smallest.

This is basically a double censored situation. Symbolically, $q_u = y_{n-r} - y_{r+1}$, where $y_1 \leq y_2 \leq \ldots \leq y_n$ are the ordered sample values.

For $f(x) = e^{-x}$ $\quad 0 \leq x < \infty$

i.e., mean = standard deviation = 1

$$\hat{\gamma}_r = \frac{q_r}{E(q_r)}$$

$$\mathrm{VAR}(\hat{\gamma}_r) = \frac{\sum_{j=r+1}^{n-r-1} \frac{1}{j^2}}{\left(\sum_{j=r+1}^{n-r-1} \frac{1}{j^2}\right)^2} \tag{4.5.39}$$

where

$$E(q_r) = \sum_{j=r+1}^{n-r-1} \frac{1}{j}$$

For

$$f(x) = \frac{1}{\gamma} e^{-\frac{x}{\gamma}} \qquad 0 \leq x < \infty$$

$$\hat{\gamma} = \bar{y} \qquad \mathrm{VAR}(\hat{\gamma}) = \frac{\theta^2}{n} \tag{4.5.40}$$

The efficiency of the estimate $\hat{\gamma}_r$ based on the rth quasi-range is given by

$$\mathrm{EFF}(\hat{\gamma}_r) = \frac{\left(\sum_{j=r+1}^{n-r-1} \frac{1}{j}\right)^2}{\left(n \sum_{j=r+1}^{n-r-1} \frac{1}{j^2}\right)} \tag{4.5.41}$$

which varies from 50 percent to 61.73 percent and is not very satisfactory. But, when the lower limit is known, γ can be estimated more efficiently from a single order statistic (41, p. 252-254; 26, p. 980-999).

4.6 M.L.E. of the parameters of bivariate and mixed exponential

Up to now, estimation for the parameter of one- or two-parameter exponential has been discussed. Some methods for estimating the parameters of bivariate and mixed exponential will be given below.

A. Bivariate exponential. The m.l.e. of λ_1, λ_1', λ_2, λ_2' of λ_1, λ_1', λ_2, λ_2' the parameters of a bivariate exponential distribution given by J. E. Freund in Section 2.2 will be given below.

In a random sample of size n from a population having the bivariate p.d.f. (2.2.24), the four parameters are defined as before. Let the first r observations come from $f(x;\lambda_1)$ and the rest (n-r) observations from $f(y;\lambda_2)$ of a sample of n ordered statistics and denoted the respective sum by Σx and Σy. Let the first r observations come from $f(y;\lambda_2)$ and the rest (n-r) observations from $f(x;\lambda_1)$ and denoted the respective sum by $\Sigma' y$ and $\Sigma' x$. The likelihood function of the sample is

$$L = (\lambda_1\lambda_2')^r(\lambda_1'\lambda_2)^{n-r}\ e^{-(\lambda_1+\lambda_2-\lambda_2')\Sigma x-\lambda_2'\Sigma y-\lambda_1'\Sigma' x\ -\ (\lambda_1+\lambda_2-\lambda_1')\Sigma' y}$$

$$1 \leq r \leq n \qquad (4.6.1)$$

Then the partial derivative of (4.6.1) with respect to λ_1, λ_1', λ_2, and λ_2' respectively gives m.l.e. of

$$\hat{\lambda}_1 = \frac{r}{\Sigma x + \Sigma' y} \qquad (4.6.2)$$

$$\hat{\lambda}_1' = \frac{n-r}{\Sigma' x - \Sigma' y} \qquad (4.6.3)$$

$$\hat{\lambda}_2 = \frac{n-r}{\Sigma x + \Sigma' y} \tag{4.6.4}$$

$$\hat{\lambda}'_2 = \frac{r}{(\Sigma y - \Sigma x)} \tag{4.6.5}$$

The mean and variance of $\hat{\lambda}_1$, $\hat{\lambda}_2$ can be shown to be

$$E(\hat{\lambda}_1) = \frac{n}{(n-1)} \lambda_1$$

$$VAR(\hat{\lambda}_1) = \frac{n\lambda_1[\lambda_1 n + \lambda_2(n-1)]}{(n-1)^2(n-2)} \tag{4.6.6}$$

$$E(\hat{\lambda}_2) = \frac{n}{(n-1)} \lambda_2$$

$$VAR(\hat{\lambda}_2) = \frac{n\lambda_2[\lambda_2 n + \lambda_1(n-1)]}{(n-1)^2(n-2)} \tag{4.6.7}$$

The mean of $\frac{1}{\hat{\lambda}'_1}$ and $\frac{1}{\hat{\lambda}'_2}$ are given by

$$E(\frac{1}{\hat{\lambda}'_1}) = \frac{1}{\hat{\lambda}'_1}$$

$$E(\frac{1}{\hat{\lambda}'_2}) = \frac{1}{\hat{\lambda}'_2}$$

The asymptotic expressions for variance $\frac{1}{\hat{\lambda}'_1}$ and $\frac{1}{\hat{\lambda}'_2}$ can be obtained by using the methods developed by Mendenhall and Lehmann (36, p. 227-242).

$$VAR(\frac{1}{\hat{\lambda}_1'}) = \frac{1}{n\lambda_1'^2}(1 + \frac{\lambda_1}{\lambda_2}) \tag{4.6.8}$$

$$VAR(\frac{1}{\hat{\lambda}_2'}) = \frac{1}{n\lambda_2'^2}(1 + \frac{\lambda_2}{\lambda_1}) \tag{4.6.9}$$

B. Mixed exponential. The m.l.e. of the parameters P, γ_1, γ_2 is discussed for two cases, i.e., when the relative magnitude of V_1, V_2 are known and are not known (36, p. 504-520).

Given the p.d.f. is described in Equation (2.3.2).

1. Let y_r denote the termination point; no observations greater than y_r will be taken. This is obviously a censored case. Further, define two random variables, $Z = \frac{y}{y_r}$, $\gamma_i \frac{1}{\lambda_i y_r}$, for i = 1,2. Then the cumulative distribution function becomes

$$F_i(Z) = 1 - e^{\frac{-Z}{\gamma_i}} \qquad 0 \leq Z < \infty \tag{4.6.10}$$

Given a random sample of size n, the probability of r_1 events occur in $f_1(Z)$, r_2 events occur in $f_2(Z)$, and (n-r) events (where $r = r_1 + r_2$) not occur is the multinomial

$$P_r(r_1, r_2, n-r|n) = \frac{n!}{r_1!r_2!(n-r)!}[PF_1(1)]^{r_1} \cdot [qF_2(1)]^{r_2} \cdot [G(1)]^{n-r} \tag{4.6.11}$$

The conditional density of obtaining the ordered observations $Z_{i1}, Z_{i2}, \ldots, Z_{ir_2}$ given r_i and $Z_{ij} \leq 1$ is

$$P(Z_{i1}, Z_{i2}, \ldots, Z_{ir_i} \mid r_i; Z_{ij} \leq 1) = \frac{r_i! \prod_{j=1}^{r_i} f_i(Z_{ij})}{[F_i(1)]^{r_i}} \tag{4.6.12}$$

It then follows that the likelihood function for the sample is

$$L = \frac{n!}{(n-r)!} G(1)^{n-r} p^{r_1} q^{r_2} \prod_{j=1}^{r_1} f_1(Z_{1j}) \prod_{j=1}^{r_2} f_2(x_{2j}) \tag{4.6.13}$$

The partial derivative of lnL with respect to γ_1, γ_2, P gives the m.l.e. of γ_1, γ_2, P as

$$\hat{P} = \frac{r_1}{n} + \hat{k}\,\frac{(n-r)}{n} \tag{4.6.14}$$

$$\hat{\gamma}_1 = \overline{Z}_1 + \hat{k}\,\frac{(n-r)}{r_1} \tag{4.6.15}$$

$$\hat{\gamma}_2 = \overline{Z}_2 + \frac{(1-\hat{k})(n-r)}{r_2}$$

where

$$\overline{Z}_i = \frac{\prod_{j=1}^{r_i} Z_{ij}}{r_i}, \quad \hat{k} = \frac{1}{1 + (\frac{\hat{q}}{\hat{p}})\, e^{[\frac{1}{\hat{\gamma}_1} - \frac{1}{\hat{\gamma}_2}]}} \tag{4.6.16}$$

2. In some situations when the relative magnitude of r_1 and r_2 is known, then the m.l.e. of $r_1 = V_2 = V$ is given by

$$\hat{\gamma} = \frac{r_1\bar{Z}_1 + r_2\bar{Z}_2 + (n-r)}{r_i}, \ \hat{P} = \frac{r_1}{r} \tag{4.6.17}$$

The m.l.e. of the parameter γ of a single exponential distribution is

$$\hat{\gamma} = \frac{1}{r} n(y_r), \ n(y_r) = \sum_{i=1}^{r} y_i + (n-r)\, y_r \tag{4.6.18}$$

It can be seen that this reduced consequence from the mixed case is the familiar single exponential distribution estimator.

Excellent examples are given by P. R. Rider (42, p. 143-147) and W. Mendenhall and R. J. Hader (36, p. 504-520) about estimation of the parameters of a mixed exponential case.

TEST OF HYPOTHESIS

It is often useful to test whether the parameter(s) of an exponential distribution is equal to, greater than, or less than some fixed value. Several methods for deriving a test of hypothesis in complete, truncated, and censored case will be introduced in this section.

5.1 For a complete sample

A. Case I. Given the p.d.f. of an exponential distribution as

$$f(x) = \begin{cases} \lambda e^{-\lambda(x-\theta)} & x \geq \theta \geq 0 \\ \\ 0 & \text{elsewhere} \end{cases}$$

H_0: θ = some fixed value, assuming $\gamma = 1$. The likelihood-ratio test derived by E. Paulson (40, p. 317-328) is given below.

Given Ω the parameter space from which the sample might have been drawn to be $\{-\infty < \theta < +\infty, \gamma = 1\}$, while the sample space under the hypothesis H_0 is $\{\theta = 0, \gamma = 1\}$. The likelihood ratio is given by

$$\rho = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \frac{e^{-\sum_{1}^{h} x_i}}{e^{-\sum_{1}^{h} (x_i - x_1)}} = e^{-hx_1} \tag{5.1.1}$$

The region of acceptance consists of all points in the sample space for which $\rho_\varepsilon \leq \rho \leq 1$, where ρ_ε is chosen such that

$$\int_{\rho_\varepsilon}^{1} g(\rho)\, d\rho = 1 - \alpha.$$

α is the level of significance used and $g(\rho)$ is the distribution of ρ when θ really equals to zero. Also for any value of θ,

$$\phi(x_1)\, dx_1 = he^{-n(x_1-\theta)} dx \tag{5.1.2}$$

It can be proved that the power function $\rho(\theta)$ of θ is greater than α if $\theta \neq 0$. The test is therefore completely unbiased in the sense of Daly (11, p. 2). In addition, it is proved that the test is a uniformly most powerful test with respect to all alternatives.

<u>B. Case II.</u> To test the hypothesis H_0: = some fixed value, when γ unknown, the likelihood-ratio function is given by

$$\rho = \left[\frac{\sum_1^n (x_1 - x_1)}{\sum_1^n x_i}\right]^n = \left[\frac{1}{1 + \dfrac{n\, x_1}{\sum_1^n (x_i - x_1)}}\right]^n \tag{5.1.3}$$

The region of acceptance consists of all points in the sample space for which $\rho_\varepsilon \leq \rho \leq 1$ such that

$$\int_{\rho_\varepsilon}^{1} g(\rho) d\rho = 1 - \alpha, \text{ and is equivalent to}$$

$$0 \leq x_1 \leq \frac{k_2 s}{n} \tag{5.1.4}$$

where

$$s = \sum_1^n (x_i - x_1)\cdot(n-1)^{-1}$$

After defining the joint distribution of x_1 and s (40, p. 303), the power function $P(\theta)$ of θ for $\theta \leq 0$, $\theta \geq 0$ is found by performing double integrals of $\Psi(x_1, s)$ over region defined in (5.1.4). For $-\infty < \theta \leq 0$, $P(\theta)$ is greater than α, for $0 \leq \theta < \infty$, $P'(\theta)$ is positive and monotonically increasing in the interval, so $P(\theta)$ is $> \alpha$ when $\theta > 0$. Therefore, the test is also completely unbiased.

C. Case III. Given a sample of size n_1 from a distribution with p.d.f.

$$f(x) = \begin{cases} \lambda e^{-\lambda(x-\theta_1)} & x \geq \theta_1 \geq 0,\ \lambda > 0 \\ 0 & \text{elsewhere} \end{cases}$$

and a sample of size n_2 with p.d.f.

$$f(y) = \begin{cases} \lambda e^{-\lambda(y-\theta_2)} & x \geq \theta_2 \geq 0,\ \lambda > 0 \\ 0 & \text{elsewhere} \end{cases}$$

To test the hypothesis that $\theta_1 = \theta_2$, we let x_1 denote the smallest observation in $f(x;\theta_1,\lambda)$, y_1 the smallest observation in $f(y;\theta_2,\lambda)$, and L the smallest of $n_1 + n_2 = N$. The likelihood-ratio function is given by

$$\rho = \left[\frac{\sum_1^{n_1}(x_i - x_1) + \sum_1^{n_2}(y_i - y_1)}{\sum_1^{n_1}(x_i - L) + \sum_1^{n_2}(y_i - L)}\right]^2 = \left[\frac{1}{1 + \frac{Z}{n}}\right]^n$$

$$Z = n_2(y_1 - x_1) \text{ for } y_1 > x_1$$

$$= n_1(x_1 - y_1) \quad \text{for } x_1 > y_1$$

$$\eta = \sum_{1}^{n_1} (x_i - x_1) + \sum_{1}^{n_2} (y_i - y_1) \tag{5.1.5}$$

The region of acceptance, $\rho_\varepsilon \leq \rho \leq 1$, is equivalent to the region

$$0 \leq Z \leq k_3 U \tag{5.1.6}$$

where $\alpha = \dfrac{1}{(1 + k_3)^{N-Z}}$ after defining $f(Z)$ and $\phi(\eta)$ (40, p. 304-305).

The power function $P(D)$ of $D = \theta_2 - \theta_1$ is found by double integrals over $\phi(Z)$ and $\phi(\eta)$ over the region defined in (5.1.6). To prove that $P(D) > \alpha$ when $D \neq 0$, it is done by showing $P'(D)$ is always positive when $D \neq 0$, and always negative when $D < 0$ (40, p. 305). Therefore, the test is completely unbiased.

5.2 For a censored sample

B. Epstein and M. Sobel have derived a best test based on the first r out of n ordered observations from an exponential distribution for

$H_0: \gamma - \gamma_1$

$H_A: \gamma = \gamma_2$, where $\gamma_1 < \gamma_2$

Their derivation is based on Neyman-Pearson Lemma (38, p. 292).

A best test for which the region of rejection is found from

$$\frac{f(y_1,y_2,\ldots,y_r;\gamma_2)}{f(y_1,y_2,\ldots,y_r;\gamma_1)} \quad k \tag{5.2.1}$$

The region of rejection for $H_o:\gamma = \gamma_1$ is given by

$$\hat{\gamma}_{r,n} < C \text{ such that } P_r(\hat{\gamma}_{r,n} < C|\gamma = \gamma_1) = \alpha \tag{5.2.2}$$

Given $\hat{\gamma}_{r,n} = \frac{1}{\hat{\lambda}_c}$ from Equation (4.2.4) and its p.d.f. as

$$f_r(y) = \frac{1}{(r-1)!}\left(\frac{r}{\gamma}\right)^r y^{r-1} e^{-\frac{ry}{\gamma}} \tag{5.2.3}$$

It is easy to verify that $W = (2r\hat{\gamma}_{r,n})/\gamma$ is distributed as a chi-square with 2r d.f. Then (5.2.2) can be written as

$$P_r\left(W < \frac{2rc}{\gamma_1}\right) = \alpha \qquad \text{or} \qquad P_r\left(W > \frac{2rc}{\gamma_1}\right) = 1-\alpha \tag{5.2.4}$$

which, again, can be rewritten as

$$C = [\gamma_1 x^2_{1-\alpha}(2r)]\cdot(2r)^{-1}$$

$$\hat{\gamma}_{r,n} < \frac{\gamma_1 x^2_{1-\alpha}(2r)}{2r} \tag{5.2.5}$$

This region is best in the Neyman-Pearson sense for that it has a greater chance of rejecting $\gamma = \gamma_1$ when $\gamma = \gamma_2$ is true. It is a uniformly most powerful test for the hypothesis

$H_o:\gamma=\gamma_1$ vs $H_A:\gamma<\gamma_1$

If the acceptance region rather than reject region is used, the test of hypothesis with Type I error equals to α is given by

$$\hat{\gamma}_{r,n} > \frac{\gamma_1 x^2_{1-\alpha}(2r)}{2r} \qquad (5.2.6)$$

The operating characteristic curve $L(\gamma)$ = probability of accepting $\gamma = \gamma_1$ when γ is the true value =

$$P_r\left(\hat{\gamma}_{r,n} > \frac{\gamma_1 x^2_{1-\alpha}(2r)}{2r}\right) = P_r\left(x^2(2r) > \frac{\gamma_1 x^2_{1-\alpha}(2r)}{\gamma}\right)$$

is given in Figure 4.

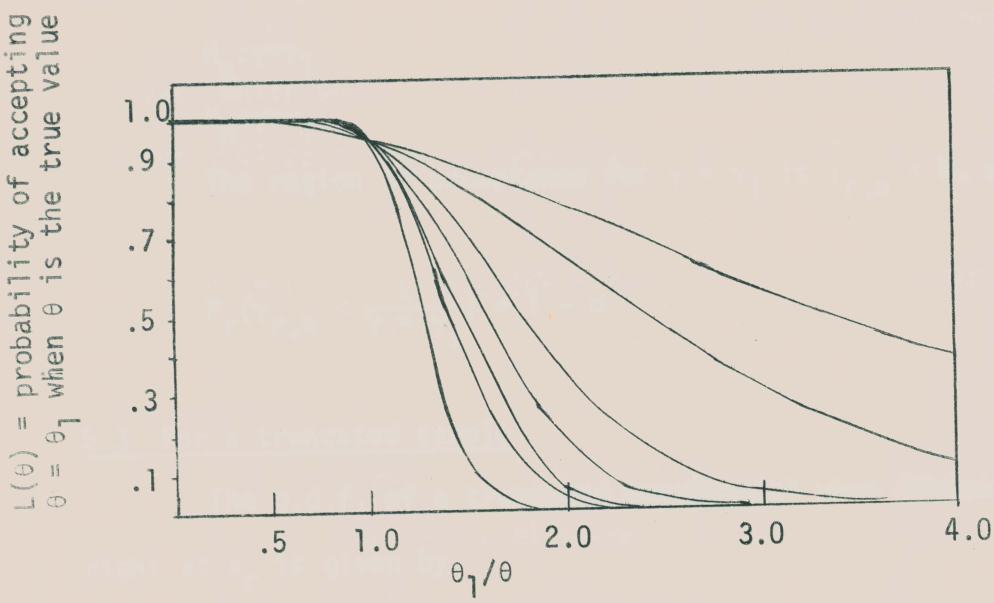

Figure 4. Operating characteristics of test of the form $\hat{\gamma}_{r,n} > C \cdot L(\gamma_1) = 1 - \alpha = .95$.

If r and C are initially unknown, the test is derived according to the following:

Choose r and C such that the o.c. curve will have the property $L(\gamma_1) = 1-\alpha$ and $L(\gamma_2) \leq \beta$, where $\gamma_2 < \gamma_1$, and α,β are prescribed Type 1 and Type 2 error. The condition is met by substituting γ_2 for γ in (5.2.7) and requiring that r satisfies the inequality $\frac{\gamma_1}{\gamma_2} x^2_{1-\alpha}(2r) \geq x^2_\beta(2r)$ or $\frac{\gamma_2}{\gamma_1} \leq x^2_{1-\alpha}(2r)\ [x^2_\beta(2r)]^{-1}$.

r, the integer is then found in such fashion that it leads the o.c. curve passing most nearly through the points $[\gamma_1, L(\gamma_1) = 1 - \alpha]$ and $[\gamma_2, L(\gamma_2) = \beta]$. Using this r, the acceptance region $\gamma = \gamma_1$ is given by $\hat{\gamma}_{r,n} > C$ where $C = \gamma_1 x^2_{1-\alpha}(2r)\cdot[2r]-1$.

In a completely similar way, for a uniformly most powerful test in the Neyman-Pearson sense for

$H_o: \gamma=\gamma_1$

$H_A: \gamma>\gamma_1$

The region of acceptance for $\gamma = \gamma_1$ is $\hat{\gamma}_{r,n} < k$, where

$$P_r(\hat{\gamma}_{r,n} < \frac{k}{\gamma = \gamma_1}) = 1 - \alpha$$

5.3 For a truncated sample

The p.d.f. of a truncated exponential when truncated to the right at x_r is given by

$$g(x) = \begin{cases} \lambda k e^{-\lambda x} & x_r \geq x \geq 0 \\ 0 & \text{elsewhere} \end{cases} \qquad (5.3.1)$$

The distribution of the sample mean $\bar{x} = \frac{\sum_1^n x_i}{n}$ is given by

$$f(\bar{x}|\gamma)d\bar{x} = \frac{1}{\Gamma(n)} e^{\frac{-n\bar{x}}{\gamma}} (\frac{n\bar{x}}{\gamma})^{n-1} \alpha (\frac{n\bar{x}}{\gamma}) \qquad (5.3.2)$$

The uniformly most powerful test for the one-sided hypothesis

$H_o: \gamma = \gamma_1$

$H_A: \gamma < \gamma_1$

for a given sample of size n from (5.3.1) and (5.3.2) is given by

$$\int_0^{c_1} f_n(\bar{x}|\gamma_1)d\bar{x} = \int_0^{c_2} g(\bar{x}|\gamma_1)d\bar{x} = \alpha \qquad (5.3.3)$$

Reject H_o if $\bar{x} < c_i (i = 1,2)$; accept otherwise. α is the type 1 error; c_2 is a function of x_r, the point of truncation.

To obtain the "error" incurred if the usual procedure is followed while sampling is actually from a truncated distribution, one can utilize the three power functions given below.

$$P_u(\gamma_1 x_r) = \int_0^{c_1} f_n(\bar{x}|\gamma)d\bar{x} = \text{power of usual test when x is}$$

untruncated (5.3.4)

$$P(\gamma_1 x_r) = \int_0^{c_2} g(\bar{x}|\gamma)d\bar{x} = \text{power of usual test when x is}$$

truncated (5.3.5)

$$P_c(\gamma_1 x_r) = \int_0^{c_2} g(\bar{x}|\gamma)d\bar{x} = \text{power of test when x is truncated}$$

(5.3.6)

The error can then be found by

$$E(\gamma_1 x_r) = P_u(\gamma_1 x_r) - P(\gamma_1 x_r) \quad \text{for } \gamma = \gamma_1$$

$$E(\gamma_1 x_r) = \alpha - \alpha' \tag{5.3.7}$$

A table has been tabulated for a few values of α' for different sample size and for $x_r = 1.0, 5.0$ when $\alpha = .05$ (6, p. 209-213).

5.4 For a sample when some extreme observations are present

From B. Epstein and M. Sobel, the m.l.e. of γ for one-parameter exponential distribution is given by

$$\hat{\gamma} = \{\sum_1^r (y_i - \theta) + (n-r)(y_r - \theta)\} \cdot r^{-1} \tag{5.4.1}$$

$$\hat{\gamma}' = \frac{\sum_1^r (y_i - y_1) + (n-r)(y_r - y_1)}{r-1} = \frac{\sum_2^r Z_i}{r-1} \tag{5.4.2}$$

where $Z_i = y_i - y_{i-1}$.

To test the hypothesis $H_0{:}\theta=\theta_1$, given a complete sample of size n from (4.1.1) and assuming the extreme observations undetected, Carlson (5, p. 550-559) proposed the statistics

$$h_n = \frac{n}{\gamma}(y_1 - \theta_1) \quad \text{when } r \text{ is known} \tag{5.4.3}$$

$$h_n' = \frac{y_1 - \theta_1}{y_n - y_1} \tag{5.4.4}$$

which is not the minimum variance estimate. An alternative, proposed by A. P. Basu, is given as

$$\tau = \frac{n(n-1)(y_1 - \theta_1)}{\sum_{2}^{n} (n-i+1)(y_i - y_{i-1}} = \frac{n(n-1)(y_1-\theta)}{(y_2+ \ldots +y_n)(n-1)y_1} \simeq F(2,2n-2) \tag{5.4.5}$$

which is free from all the criticisms raised against Carlson. Examples are also given in the paper.

5.5 For a doubly censored sample

A doubly censored sample, ℓ observations and m observations missing at each end, i.e., $y_\ell \leq y_{\ell+1} \leq \ldots \leq y_{\ell+r}$ of size r from a complete sample of size n having the p.d.f.

$$f(x) = \begin{cases} \lambda e^{-\lambda(x-\theta)} & 0 \leq \theta \leq x < \infty, \quad \gamma > 0 \\ 0 & \text{elsewhere} \end{cases}$$

The joint density of $y_{\ell+1}, y_{\ell+2} \ldots, y_{\ell+r}$ is given by

$$h(y_{\ell+1};\ldots y_{\ell+r}) = \frac{n!}{\ell!(n-\ell-r)!} \{1-e^{-\frac{1}{\gamma}(y_{\ell+1}-\theta)}\}^{\ell}$$

$$e^{\{-\frac{1}{\gamma}[\sum_{1}^{r}(y_{\ell+i}-\theta) + (n-\ell-r)(y_{\ell+r}-\theta)]\}} \tag{5.5.1}$$

The m.l.e. $\hat{\theta},\hat{\gamma}$ of θ and γ are given by

$$\hat{\theta} = y_{\ell+1}$$

$$\hat{\gamma} = \frac{1}{r}\left\{\sum_{2}^{r} (n-\ell-i+1)\, z_{\ell+i}\right\} \tag{5.5.2}$$

and

$$\frac{2(r-1)\hat{\gamma}}{r} \simeq \chi^2(2r-2) \tag{5.5.3}$$

where

$$z_i = y_i - y_{i-1}$$

The minimum variance unbiased estimates of θ and γ are given by

$$\hat{\theta} = y_{\ell+1} - k\hat{\gamma} \tag{5.5.4}$$

$$\hat{\gamma} = \frac{1}{r-1}\sum_{2}^{r} (n-\ell-i+1) z_{\ell+i}$$

where

$$k = \sum_{1}^{\ell+1} \frac{1}{n-i+1} \tag{5.5.5}$$

If a sample is considered to be a censored one, one may want to test the hypothesis

$$H_0{:}\gamma=\gamma_0$$

by using the fact that $(2(r-1)\hat{\gamma})/\gamma$ is distributed as a chi-square with $(2r-2)$ d.f. Also, one may want to test the hypothesis

$$H_0{:}\theta=\theta_0$$

by using the statistic

$$\tau = \frac{y_{\ell+1} - \theta_o}{\sum_{2}^{r} (n-\ell-i+1)\ z_{\ell+i}} \tag{5.5.6}$$

When θ and γ are known, the following tests

$$\tau_1 = \frac{y_n - \theta}{\gamma} \tag{5.5.7}$$

$$\tau_2 = \frac{y_n - y_1}{\gamma} \tag{5.5.8}$$

$$\tau_3 = \frac{y_n - y_{n-1}}{\gamma} \tag{5.5.9}$$

can be used to test whether the largest observation in a sample of size n is an extreme observation. Similar test for the smallest observation y_1.

When θ and γ are not known, the standardized deviate

$$\tau_n = \frac{y_n - y_1}{\Sigma(y_i - y_1)} \tag{5.5.10}$$

can be used to test the hypothesis whether τ_n is an extreme observation.

Example 5.5.1: Given an ordered sample as follows:

1, 3, 3, 15, 25, 33, 39, 70, 680

Test whether the value y_9 = 680 is really an extreme observation.

Using Equation (5.5.10)

$$\tau_9 = \frac{680 - 1}{860} = .79$$

By consulting the table given by Laurent, we find that y_9 is an extreme observation to the sample. To test $y_8 = 70$

$$\tau_8 = \frac{69}{181} = .38$$

and this is found not an extreme observation.

CONCLUSION

It has been observed that, in numerous practical situations (not mentioning those fields cited in the Introduction), for example, quality control, life testing, fatigue testing, and other kinds of destructive testing situations, such as test of life of electric bulbs, radio tubes, ball bearings, etc., most of the data obtained are approximately exponentially distributed. Knowledge developed concerning every aspect of exponential distributions certainly is very useful. Jacobson (17, p. 502) has compared the operating characteristic curves of a test procedure based on the lowest three out of five observations with that based on the average of five out of five, and four out of four. He showed that the operating characteristic curves are almost identical for three cases. This means the true distribution parameter values, although unknown, can be obtained with the same Type I and Type II error using three instead of four or five items. If a process involves testing large amount of expensive items, considerable savings can thus be obtained.

Finally, this report is by no means meant to be a comprehensive summary of the topics discussed in Section I through Section V. There are possibly several hundred published papers of related interests. It is not possible to include all of them here. However, this report does provide the basic concepts, procedures, and equations. For example, the Introduction section illustrates the general area where problems could arise or could be applied. Section II illustrates, besides a score of distribution functions, how a special exponential function can be derived under different assumptions along with methods of finding basic characteristics of a particular exponential distribution. Section III provides a dozen of procedures for testing the validity of an assumption that the underlying distribution is really of a certain type of exponential. Once the assumption is established with certain associated characteristics, methods, equations, and procedures are provided in Sections IV and V for estimation and test of hypothesis. Optimum decision, rather than a decision based on experience, can be made based on the result of the estimation or test of hypothesis.

LITERATURE CITED

1. Bailey, N. T. The elements of stochastic process. John Wiley and Sons, New York. 1964.

2. Barnard, G. A. Time intervals between accidents--a note on Maguire, Pearson, and Wynn's paper. Biometrika 40:212-213. 1953.

3. Bartholomew, D. J. A sequential test of randomness for events occurring in time or space. Biometrika 43:64-78. 1956.

4. Bartholomew, D. J. Testing for departure from the exponential distribution. Biometrika 44:253-256. 1957.

5. Basu, A. P. Effect of truncation on a test for the scale parameter of the exponential distribution. A.M.S. 35:203-213. 1964.

6. Basu, A. P. On some tests of hypothesis relating to the exponential distribution when some outliers are present. J.A.S.A. 60:548-559. 1965.

7. Birnbaum, Z. W. Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size. J.A.S.A. 47:425-441. 1952.

8. Bulmer, M. G. Principles of statistics. MIT Press, Cambridge, Massachusetts. 1965.

9. Cochran, W. G. The distribution of the largest of a set of estimated variance as a fraction of their total. Ann. of Eugenics 11:47-52. 1941.

10. Cramer, H. Mathematical methods of statistics. Princeton University Press, Princeton, New Jersey. 1946.

11. Daly, J. F. On the unbiased character of likelihood-ratio test for independence in normal systems. A.M.S. 11:2. 1940.

12. Deemer, W. L., and D. F. Votaw. Estimation of parameters of truncated or censored exponential distributions. A.M.S. 26:498-504. 1955.

13. Downton, F. A note on ordered least-squares estimation. Biometrika 40:457-458. 1953.

14. Epstein, B. Estimation from life test data. Technometrics 2:447-451. 1960.

15. Epstein, B. Estimation of the parameters of two-parameter exponential distribution from censored sample. Technometrics 2:403-406. 1960.

16. Epstein, B. Tests for the validity of the assumption that the underlying distribution of life is exponential. Part I. Technometrics 2:83-101. 1960.

17. Epstein, B., and M. Sobel. Life testing. J.A.S.A. 48:486-502. 1953.

18. Epstein, B., and M. Sobel. Some two-sample tests based on ordered observations from the exponential distribution. A.M.S. 24:458-466. 1953.

19. Epstein, B., and M. Sobel. Some theorems relevant to life testing from an exponential distribution. A.M.S. 25:373-381. 1954.

20. Feller, W. An introduction to probability theory and its applications. Second edition. John Wiley and Sons, New York. 1957.

21. Fisher, R. A. Test of significance on harmonic analysis. Proc. Roy. Soc., Series A. 125:54-59. 1929.

22. Freund, J. E. A Bivariate extension of the exponential distribution. J.A.S.A. 56:971-977. 1961.

23. Gumbel, E. J. On the reliability of the classical chi-square test. A.M.S. 14:253-263. 1943.

24. Gumbel, E. J. Bivariate exponential distributions. J.A.S.A. 55:698-707. 1960.

25. Halperin, M. Maximum likelihood estimates in truncated samples. A.M.S. 23:226-238. 1952.

26. Harter, H. L. The use of sample quasi-ranges in estimating population standard deviation. A.M.S. 30:980-999. 1959.

27. Harter, H. L. Estimating the parameters of negative exponential populations from one or two order statistics. A.M.S. 32:1078-1090. 1961.

28. Hartley, H. O. The maximum F-ratio as a short-cut test for heterogeneity of variance. Biometrika 37:308-312. 1950.

29. Jeffreys, Sir Harold. Methods of mathematical physics. Third edition. Cambridge University Press, Cambridge, England. 1956.

30. Kendall, M. G. The advanced theory of statistics. Vol. II. Charles Gridfin and Company, Ltd., London. 1946.

31. Kulldorff, G. Estimation of one or two parameters of the exponential distribution on the basis of suitably chosen statistics. A.M.S. 34:1419-1431. 1963.

32. Laurent, A. G. Conditional distribution of order statistics and distribution of the reduced ith order statistics of the exponential mode. A.M.S. 34:652-657. 1963.

33. Lloyd, E. H. Least-squares estimation of location and scale parameters using order statistics. Biometrika 39:88-95. 1952.

34. Maguire, B. A., E. S. Pearson, and A. H. A. Wynn. The times intervals between industrial accidents. Biometrika 39:168-180. 1952.

35. Massey, F. J., Jr. The Kolmogorov-Smirnov Test for goodness of fit. J.A.S.A. 46:68-78. 1951.

36. Mendenhall, W., and R. J. Hader. Estimation of parameters of mixed exponentially distributed failure time distribution from censored life test data. Biometrika 45:504-520. 1958 .

37. Mood, A. M. Introduction to the theory of statistics. McGraw-Hill Book Company, Inc., New York. 1950.

38. Mood, A. M. and F. A. Graybill. Introduction to the theory of statistics. Second edition. McGraw-Hill Book Company, Inc., New York. 1963.

39. Parzen, E. Modern probability theory and its applications. John Wiley and Sons, Inc., New York. 1964.

40. Paulson, E. On certain likelihood-ratio tests associated with the exponential distribution. A.M.S. 12:301-306. 1941.

41. Rider, P. R. Quasi-range of samples from an exponential population. A.M.S. 30:252-254. 1959.

42. Rider, P. R. The method of moments applied to a mixture of two exponential distributions. A.M.S. 32:143-147. 1961.

43. Sarhan, A. E., and B. G. Greenburg. Contributions to order statistics. John Wiley and Sons, Inc., New York. 1962.

44. Sarhan, A. E., B. G. Greenburg, and J. Ogawa. Simplified estimates for the exponential distribution. A.M.S. 34:102-116. 1963.

45. Savage, R. Bibliography of nonparametric statistics and related topics. J.A.S.A. 48:844-906. 1953.

46. Siddiqui, M. M. Optimum estimators of the parameters of negative esponential distributions from one or two order statistics. A.M.S. 34:117-121. 1963.

47. Smirov, N. Tables for estimating the goodness of fit of empirical distributions. A.M.S. 19:279-281. 1948.

48. Wilks, S. S. Mathematical statistics. Princeton University Press, Princeton, New Jersey. 1943.

VITA

Michael Chang-Yu Wang

Candidate for the Degree of

Master of Science

Report: A Study of the Exponential Distributions and Their Applications

Major Field: Applied Statistics

Biographical Information:

Personal Data: Born at Kweiyang, China, September 23, 1938, son of Chong-Chieh and Ging-Hwa Wang.

Education: Received the Bachelor of Arts degree from Tunghai University, Taiwan; completed requirements for the Master of Science degree in Applied Statistics at Utah State University, in 1969.

Professional Experience: 1967 to present, statistician, Programmer, S. D. D., I. B. M., Poughkeepsie, New York.