CHAPTER 19

Exponential Distributions

1 DEFINITION

The random variable X has an ***exponential*** (or ***negative exponential) distribution*** if it has a probability density function of form

$$p_X(x) = \sigma^{-1} \exp\left[-\frac{(x-\theta)}{\sigma} \right], \qquad x > \theta; \sigma > 0. \tag{19.1}$$

Figure 19.1 gives a graphical representation of this function, with $\theta > 0$. This is a special case of the gamma distribution, the subject of Chapter 17. The exponential distribution has a separate chapter because of its considerable importance and widespread use in statistical procedures.

Very often it is reasonable to take $8 = 0$. The special case of (19.1) so obtained is called the ***one-parameter*** exponential distribution. If $\theta = 0$ and $\sigma = 1$, the distribution is called the ***standard*** exponential distribution. The pdf is

$$p_X(x) = \exp(-x), \qquad x > 0. \tag{19.2}$$

The mathematics associated with the exponential distribution is often of a simple nature, and so it is possible to obtain explicit formulas in terms of elementary functions, without troublesome quadratures. For this reason models constructed from exponential variables are sometimes used as an approximate representation of other models that are more appropriate for a particular application.

2 GENESIS

There are many situations in which one would expect an exponential distribution to give a useful description of observed variation. One of the most widely quoted is that of events recurring "at random in time." In particular, suppose that the future lifetime of an individual has the same distribution, no matter

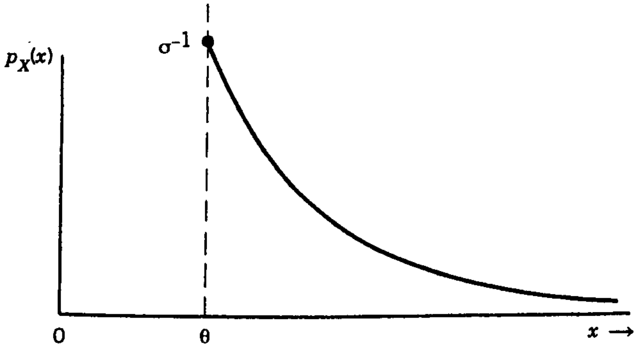

Figure 19.1 Exponential Density Function

how old it is at present. This can be written formally (X representing lifetime)

$$\Pr[X \le x_0 + x | X > x_0] = \Pr[X \le x] \qquad \text{for all } x_0 > 0,\ x > 0.$$

X must be a continuous positive random variable. If it has a probability density function $p_X(x)$, then the conditional probability density function, given that X is greater than x_0, is

$$\frac{p_X(x)}{1 - F_X(x_0)}, \qquad x > x_0 > 0.$$

Since the conditional distribution of the future lifetime $(X - x_0)$ is the same as the (unconditional) distribution of X, we have, say,

$$\frac{p_X(x_0)}{1 - F_X(x_0)} = p_X(0) = p_0.$$

It follows that if $F_X(x_0) \neq 1$, $p_0 > 0$ and $F_X(x)$ satisfy the differential equation

$$\frac{dF_X(x)}{dx} = p_0[1 - F_X(x)],$$

whence $1 - F_X(x)$ a $e^{-p_0 x}$. Introducing the condition $\lim_{x \to 0} F_X(x) = 0$, we find that

$$1 - F_X(x) = e^{-p_0 x}, \tag{19.3}$$

that is,

$$F_X(x) = 1 - e^{-p_0 x} = p_0 \int_0^x e^{-p_0 t}\,dt.$$

This shows that the probability density function of X is of form (19.1) with $\theta = 0$, $\alpha = p_0^{-1}$.

There are other situations in which exponential distributions appear to be the most natural. Many of these do, however, have as an essential feature the random recurrence (often in time) of an event.

In applying the Monte Carlo method it is often required to transform random variables from a standard rectangular distribution to exponential random variables. An ingenious method was suggested at an early date by von Neumann (1951). Let $\{X_i;\ i = 0, 1, \dots\}$ be a sequence of independent random variables from the standard rectangular distribution, and define a random variable N taking positive integer values through $\{X_i\}$ by the inequalities

$$X_1 < X_0, \quad \sum_{j=1}^{2} X_j < X_0, \dots, \quad \sum_{j=1}^{N-1} X_j < X_0, \quad \sum_{j=1}^{N} X_j > X_0.$$

We "accept" the sequence $\{X_i\}$ if N is odd, otherwise we "reject" it and repeat the process until N turns out odd. Let T be the number of sequences rejected before an odd N appears ($T = 0, 1, \dots$) and X_0 be the value of the first variable in the accepted sequence. Then $Y = T + X_0$ is an exponential random variable with the standard density e^{-x}.

A rather more convenient method was suggested by Marsaglia (1961). Let N be a nonnegative random integer with the geometric distribution [Chapter 5, Section 2, equation (5.8)]

$$\Pr[N = n] = (1 - e^{-\lambda})e^{-n\lambda}, \qquad n = 0, 1, \dots,$$

and let M be a positive random integer with the zero-truncated Poisson distribution [Chapter 4, Section 10, equation (4.73)]

$$\Pr[M = m] = (1 - e^{-\lambda})^{-1}\frac{e^{-\lambda}\lambda^m}{m!}, \qquad m = 1, 2, \dots .$$

Finally let $\{X_i;\ i = 1, 2, \dots\}$ be a sequence of independent random variables each having a standard rectangular distribution (Chapter 26). Then

$$Y = \lambda\{N + \min(X_1, \dots, X_M)\}$$

has the standard exponential distribution.

Sibuya (1962) gave a statistical interpretation of the procedure, recommended that the value of the parameter A be taken as 0.5 or log 2, and extended the technique to the chi-square distribution. Bánkövi (1964) investigated a similar technique. A table of 10,000 exponential random numbers is given by Barnett (1965).

3 SOME REMARKS ON HISTORY

Over the last 40 years the study of estimators based on samples from an exponential population has been closely associated with the study of order statistics. Lloyd (1952) described a method for obtaining the best linear unbiased estimators (BLUEs) of the parameters of a distribution, using order statistics. Epstein and Sobel (1953) presented the maximum likelihood estimator of the scale parameter a, of the one-parameter exponential distribution in the case of censoring from the right. Epstein and Sobel (1954) extended the foregoing analysis to the two-parameter exponential distribution. Sarhan (1954) employed the method derived by Lloyd to obtain the BLUEs of a and 6 for the two-parameter exponential distribution in the case of no censoring. Sarhan (1955) extended his results to censoring. Sarhan noted that in the special case of the one-parameter exponential distribution, his results agreed with those of Epstein and Sobel, and therefore his estimator of a was not only the best linear unbiased estimator but also the maximum likelihood estimator of a. Epstein (1960) extended his own results to estimators of a and θ for the one- and two-parameter exponential distributions in the cases of censoring from the right and/or left. For the two-parameter exponential distribution his maximum likelihood estimators coincided with the BLUEs of Sarhan (1960), but for the one-parameter exponential distribution there was agreement only in the case of censoring from the right. Many other contributions by Epstein and Sobel are included in the references.

In the light of the applicability of order statistics to the exponential distribution it became quite natural to attempt estimation of the parameter by use of the sample quasi-ranges. Rider (1959) derived the probability density function and the cumulants of the quasi-range of the standardized exponential distribution and Fukuta (1960) derived "best" linear estimators of a and 8 by two sample quasi-ranges. The next step would quite reasonably be that of determining the two order statistics that would supply the best linear unbiased estimator of σ and 8 for the two-parameter distribution; this was in fact done numerically by Sarhan, Greenberg, and Ogawa (1963). They employed the method of Lloyd to obtain the best linear estimators of a and 6 based on the two-order statistics $X'_{l:n}$ and $X'_{m:n}$, and then compared numerically the relative efficiencies of the estimators for various pairs of values $(1, m)$. Harter (1961), using a similar approach to that of Sarhan and his coworkers, presented the best linear estimators of σ for the one-parameter distribution based on one- and two-order statistics. Harter mentioned in this paper that he was not aware of any analytical process by which the optimum pair of order statistics $X'_{l:n}$ and $X'_{m:n}$ can be determined. Siddiqui (1963) presented an analytical method based on the Euler-Maclaurin formula for obtaining the optimum pair of BLUE order statistics. Since 1963 a considerable number of additional, more refined, results have been obtained. Some of these results are presented in Section 7.

4 MOMENTS AND GENERATING FUNCTIONS

The moment generating function of a random variable X with probability density function (19.1) is

$$E[e^{tX}] = (1-\sigma t)^{-1}e^{t\theta} \qquad (= (1-\sigma t)^{-1} \text{ if } \theta = 0). \tag{19.4}$$

The characteristic function is $(1 - i\sigma t)^{-1}e^{it\theta}$.

The central moment generating function is

$$E[e^{t(X-\theta-\sigma)}] = (1-\sigma t)^{-1}e^{-t\sigma}.$$

The cumulant generating function is $\log E[e^{tX}] = t\theta - \log(1-\sigma t)$. Hence the cumulants are

$$\kappa_1 = \theta + \sigma \qquad (= E[X])$$
$$\kappa_r = (r-1)!\sigma^r, \qquad r > 1. \tag{19.5}$$

Setting $r = 2, 3, 4$, we find that

$$\text{Var}(X) = \mu_2 = \sigma^2$$
$$\mu_3 = 2\sigma^3$$
$$\mu_4 = 9\sigma^4.$$

Note that if $\theta = 0$ and $\sigma = 1$, then $E[X] = 1 = \text{Var}(X)$.

The first two moment ratios are

$$\sqrt{\beta_1} = 2, \quad \beta_2 = 9.$$

The mean deviation is

$$2\sigma\int_1^\infty (x-1)e^{-x}\,dx = 2e^{-1}\sigma. \tag{19.6}$$

Note that

$$\frac{\text{Mean deviation}}{\text{Standard deviation}} = \frac{2}{e} = 0.736. \tag{19.7}$$

The median of the distribution is $\theta + \sigma \log_e 2$. The mode of this distribution is at the lowest value θ of the range of variation.

The information generating function [$(u-1)$-th frequency moment] is $\sigma^{1-u}u^{-1}$. The entropy is $1 + \log \sigma$. The hazard rate (σ^{-1}) is constant. This is an important property of exponential distributions.

5 APPLICATIONS

As has already been mentioned in Section 1, the exponential distribution is applied in a very wide variety of statistical procedures. Currently among the most prominent applications are in the field of life-testing. The lifetime (or life characteristic, as it is often called) can be usefully represented by an exponential random variable, with (usually) a relatively simple associated theory. Sometimes the representation is not adequate; in such cases a modification of the exponential distribution [often a Weibull distribution (Chapter 21) is used].

Another application is producing usable approximate solutions to difficult distributional problems. An ingenious application of the exponential distribution to approximate a sequential procedure is due to Ray (1957). He wished to calculate the distribution of the smallest n for which $\sum_{i=1}^{n} U_i^2 < K_n$, where $U_1, U_2, \ldots$ are independent unit normal variables and $K_1, K_2, \ldots$ are specified positive constants. By replacing this by the distribution of the smallest even n, he obtained a problem in which the sums $\sum_{i=1}^{n} U_i^2$ are replaced by sums of independent exponential variables (actually χ^2's with two degrees of freedom each).

Vardeman and Ray (1985) investigated the average sum lengths for CUSUM schemes when observations are exponentially distributed. They show that in this case the Page (1954) integral equation whose solution gives average sum lengths for one-sided CUSUM schemes can be solved without resorting to approximation. They provide tables of average run lengths for the exponential case.

6 ORDER STATISTICS

Let $X_1' \le X_2' \le \cdots \le X_n'$ be the order statistics obtained from a sample of size n from the standard exponential distribution in (19.2). Then, the joint density of all n order statistics is

$$p_{X'_1,\dots,X'_n}(x_1,\dots,x_n) = n!e^{-\Sigma_{i=1}^n x_i}, \qquad 0 \le x_1 \le \cdots \le x_n < \infty. \quad (19.8)$$

By making the transformation

$$Y_i = (n-i+1)(X'_i - X'_{i-1}), \qquad i = 1,2,\dots,n \text{ (with } X'_0 \equiv 0), \quad (19.9)$$

we obtain from (19.8) the joint density function of $Y_1, Y_2, \dots, Y_n$ to be

$$p_{Y_1,\dots,Y_n}(y_1,\dots,y_n) = e^{-\Sigma_{i=1}^n y_i}, \qquad 0 \le y_1,\dots,y_n < \infty. \quad (19.10)$$

That is, the Y_i**'s** (termed normalized spacings) are independent and identically distributed standard exponential random variables. This result is originally due to Sukhatme (1937). Also from (19.9),

$$X'_i \stackrel{d}{=} \sum_{j=1}^{i} \left\{ \frac{Y_j}{n-j+1} \right\}, \qquad i = 1,2,\dots,n. \quad (19.11)$$

From (19.11) it is clear that the exponential order statistics form an additive Markov chain as shown originally by **Rényi** (1953).

The additive Markov chain representation in (19.11) makes it possible to write down explicit expressions for the single and product moments of X'_i. For example, we have

$$E[X'_i] = \sum_{j=1}^{i} \frac{E[Y_j]}{n-j+1} = \sum_{j=1}^{i} \frac{1}{n-j+1},$$

$$i = 1,2,\dots,n, \quad (19.12)$$

$$\text{Var}(X'_i) = \sum_{j=1}^{i} \frac{\text{Var}(Y_j)}{(n-j+1)^2} = \sum_{j=1}^{i} \frac{1}{(n-j+1)^2},$$

$$i = 1,2,\dots,n, \quad (19.13)$$

$$\text{Cov}(X'_i, X'_k) = \sum_{j=1}^{i} \frac{\text{Var}(Y_j)}{(n-j+1)^2} = \sum_{j=1}^{i} \frac{1}{(n-j+1)^2}$$

$= \text{Var}(X'_i)$ for 1 I i < k I n . (19.14)

This special structure of the variance-covariance matrix of exponential order statistics makes it possible to derive the best linear unbiased estimators of the parameters in an explicit form (see Section 7).

Interestingly the result that the normalized spacings Y_i**'s** defined in (19.9) are i.i.d. standard exponential random variables has been generalized by

Viveros and Balakrishnan (1994) to the case of a Type II progressively censored sample. To be precise, let m be the number of failures observed before termination of a life test of n individuals and $X'_1 \le X'_2 \le \cdots \le X'_m$ be the observed ordered life lengths. Let R_i designate the number of units removed at the time of the ith failure (Type II censoring),

$$0 \le R_i \le n - \sum_{j=1}^{i-1} (R_j + 1) - 1, \qquad i = 2, 3, \dots, m - 1,$$

with

$$0 \le R_1 \le n - 1,$$

$$R_n = n - \sum_{j=1}^{m-1} (R_j + 1) - 1.$$

The resulting data are referred to as a Type *II* progressively censored sample [e.g., see Nelson (1982); Lawless (1982); Cohen and **Whitten** (1988); Balakrishnan and Cohen (1991); Cohen (1991)]. Defining the ith normalized spacing between $X'_1, X'_2, \dots, X'_m$ as

$$Y_i = \left\{ n - \sum_{j=1}^{i-1} (R_j + 1) \right\} (X'_i - X'_{i-1}), \qquad i = 1, 2, \dots, m, \text{ (with } X'_0 \equiv 0), \tag{19.15}$$

Viveros and Balakrishnan (1994) have proved that Y_i's are i.i.d. exponential random variables. Sukhatme's result, presented earlier, is a particular case of this general result corresponding to R, $= R_2 = \dots = R_m = 0$. In this general case an additive Markov chain representation for X'_i similar to the one in (19.11) is possible; it can then be used in writing down explicit expressions for the means, variances, and covariances of X'_i that are similar to those in (19.12)–(19.14). The special structure of the variance-covariance matrix of exponential order statistics observed earlier in (19.13)–(19.14) occurs in this general case of Type II progressively censored sample and enables the best linear unbiased estimators of the parameters to be derived in an explicit form (as described in Section 7c).

Due to the close relationship between the geometric and the exponential distributions, there also exists a close relationship between the dependence structure of order statistics from the geometric distribution and those from the exponential distribution. Steutel and Thiemann (1989) showed some of these similarities. For example, by introducing a second subscript to denote sample size, (19.11) can be written as

$$X'_{i:n} \stackrel{d}{=} \sum_{j=1}^{i} \frac{Y_j}{(n-j+1)} \stackrel{d}{=} \sum_{j=1}^{i} X'_{1:n-j+1}, \tag{19.16}$$

where the $X'_{1:n-j+1}$'s are independent. Steutel and Thiemann (1989) established the following parallel relationship for the geometric order statistics:

$$Z'_{i:n} \stackrel{d}{=} \sum_{j=1}^{i} Z'_{1:n-j+1} + \left[\sum_{j=1}^{i} \left\langle \frac{Y_j}{n-j+1} \right\rangle \right], \tag{19.17}$$

where $Z'_{i:n}$ denotes the ith order statistic from a random sample of size n from a geometric (p) distribution, the Y_j's are independent exponential (σ) random variables with $\sigma = 1/[-\log(1-p)]$, $[Y]$ denotes the integer part of Y, and $\langle Y \rangle$ denotes the fractional part of Y. Further, all the random variables on the right-hand side of (19.17) are independent.

Arnold and Villaseñor (1989) and Arnold and Nagaraja (1991) discussed the Lorenz order relationships among order statistics from exponential samples. The Lorenz curve associated with X is

$$L_X(u) = \frac{\int_0^u F_X^{-1}(t)\, dt}{\int_0^1 F_X^{-1}(t)\, dt}, \qquad 0 \le u \le 1 \tag{19.18}$$

(see Chapter 12, Section 1). Given two nonnegative random variables X and Y (with finite positive mean), we say X exhibits less inequality (or variability) than Y in the Lorenz sense, and write X $<_L$ Y, if $L_X(u) \ge L_Y(u)$ for all $u \in [0,1]$; if the inequality is an equality, we write X $=_L$ Y. If $L_X(u)$ and $L_Y(u)$ cross, X and Y are not comparable in the Lorenz sense. Arnold and Nagaraja (1991) proved that for i $\le$ j,

$$X'_{j:m} <_L X'_{i:n} \quad \text{iff } (n-i+1)E[X'_{i:n}] \;\mathbf{I}\; (m-j+1)E[X'_{j:m}]. \tag{19.19}$$

As direct consequences of this result, they established for the exponential order statistics that

1. $X'_{i:n+1} <_L X'_{i:n}$,
2. $X'_{i+1:n+1} <_L X'_{i:n}$,
3. $X'_{i+1:n} <_L X'_{i:n}$ iff $E[X'_{i:n}] \;\mathbf{I}\; 1$; otherwise, $X'_{i:n}$ and $X'_{i+1:n}$ are not Lorenz ordered.

These authors have also discussed the Lorenz ordering of linear functions of exponential order statistics.

By making use of the underlying differential equation of a standard exponential distribution, given by $p_X(x) = 1 - F_X(x)$, Joshi (1978, 1982) derived the following recurrence relations:

$$E\left[X_{1:n}'^{m}\right] = \frac{m}{n}E\left[X_{1:n}'^{m-1}\right], \qquad n \ge 1, \quad m = 1,2,\dots, \tag{19.20a}$$

$$E\left[X_{i:n}'^{m}\right] = E\left[X_{i-1:n-1}'^{m}\right] + \frac{m}{n}E\left[X_{i:n}'^{m-1}\right],$$

and

$$2 \le i \le n, \quad m = 1,2,\dots, \tag{19.20b}$$

$$E\left[X_{i:n}'X_{i+1:n}'\right] = E\left[X_{i:n}'^{2}\right] + \frac{1}{n-i}E\left[X_{i:n}'\right], \qquad 1 \le i \le n-1, \tag{19.20c}$$

and

$$E\left[X_{i:n}'X_{j:n}'\right] = E\left[X_{i:n}'X_{j-1:n}'\right] + \frac{1}{n-j+1}E\left[X_{i:n}'\right],$$

$$1 \le i < j \le n, \quad j - i \ge 2. \tag{19.20d}$$

These recurrence relations can be used in a simple recursive manner in order to compute all the single and product moments (in particular, the means, variances, and covariances) of all order statistics. Balakrishnan and Gupta (1992) extended these results and derived relations that will enable one to find the moments and cross-moments (of order up to 4) of order statistics. They then used these results to determine the mean, variance, and the coefficients of skewness and kurtosis of a general linear function of exponential order statistics and approximate its distribution. Through this approach Balakrishnan and Gupta (1992) justify a chi-square approximation for the distribution of the best linear unbiased estimator of the mean lifetime based on doubly Type II censored samples (see Section 7). Interestingly the relations in (19.20a)–(19.20d), under certain conditions, can also be shown to be characterizations of the exponential distribution [Lin (1988, 198911 (see Section 8). Balakrishnan and Malik (1986) derived similar recurrence relations for the single and product moments of order statistics from a linear-exponential distribution with increasing hazard rate. Sen and Bhattacharyya (1994) have discussed inferential issues relating to this distribution.

By making use of the facts that $X = -\log U$ has a standard exponential distribution, when U has a uniform (0, 1) distribution (see Chapter 26), and that $-\log U$ is a monotonic decreasing function of U, we have

$$X_{i:n}' \stackrel{d}{=} -\log U_{n-i+1:n}', \qquad 1 \le i \le n,$$

and hence

$$V_i = \left(\frac{U'_{i:n}}{U'_{i+1,n}}\right)^i \stackrel{d}{=} \exp\{-i(X'_{n-i+1:n} - X'_{n-i:n})\} \stackrel{d}{=} e^{-Y'_{n-i+1}}, \quad (19.21)$$

where Y_i's are the normalized exponential spacings defined in (19.9). Since Y_i's are i.i.d. exponential, as seen earlier, it immediately follows from (19.21) that

$$V_1 = \frac{U'_{1:n}}{U'_{2:n}}, \quad V_2 = \left(\frac{U'_{2:n}}{U'_{3:n}}\right)^2, \dots, V_{n-1} = \left(\frac{U'_{n-1:n}}{U'_{n:n}}\right)^{n-1}, \quad V_n = U'^{n}_{n:n} \quad (19.22)$$

are i.i.d. uniform (0, 1) random variables. This result was derived originally by Malmquist (1950) and is now used effectively to simulate order statistics from uniform distribution without resorting to any ordering.

Joshi (1978, 1982) also considered order statistics from a right-truncated exponential population with density

$$p_X(x) = \begin{cases} \dfrac{1}{P}e^{-x}, & 0 \le x \le P_1, \\ 0, & \text{otherwise}, \end{cases} \quad (19.23)$$

where $P_1 = -\log(1-P)$ and $1 - P$ $(0 < P < 1)$ is the proportion of truncation on the right of the standard exponential distribution. By making use of the underlying differential equation given by

$$p_X(x) = [1 - F_X(x)] + \left(\frac{1-P}{P}\right)$$

and proceeding along the lines used to prove relations (19.20a)–(19.20d), Joshi (1978, 1982) established the following recurrence relations:

$$E[X'^{m}_{1:n}] = \frac{m}{n}E[X'^{m-1}_{1:n}] - \left(\frac{1-P}{P}\right)E[X'^{m}_{1:n-1}], \qquad n \ge 2, m = 1, 2, \dots, \quad (19.24a)$$

$$E[X'^{m}_{i:n}] = \frac{1}{P}E[X'^{m}_{i-1:n-1}] + \frac{m}{n}E[X'^{m-1}_{i:n}] - \left(\frac{1-P}{P}\right)E[X'^{m}_{i:n-1}], \qquad 2 \le i \le n-1, m = 1, 2, \dots, \quad (19.24b)$$

$$E\left[X'^{m}_{n:n}\right] = \frac{1}{P}E\left[X'^{m}_{n-1:n-1}\right] + \frac{m}{n}E\left[X'^{m-1}_{n:n}\right] - \left(\frac{1-P}{P}\right)P_1^m, \qquad n \geq 2, m = 1,2,\ldots, \tag{19.24c}$$

$$E\left[X'_{n-1:n}X'_{n:n}\right] = E\left[X'^{2}_{n-1:n}\right] + E\left[X'_{n-1:n}\right] - n\left(\frac{1-P}{P}\right)\left\{P_1 E\left[X'_{n-1:n-1}\right] - E\left[X'^{2}_{n-1:n-1}\right]\right\}, \qquad n \geq 2, \tag{19.24d}$$

$$E\left[X'_{i:n}X'_{i+1:n}\right] = E\left[X'^{2}_{i:n}\right] + \frac{1}{n-i} \times \left\{E\left[X'_{i:n}\right] - n\left(\frac{1-P}{P}\right)\left\{E\left[X'_{i:n-1}X'_{i+1:n-1}\right] - E\left[X'^{2}_{i:n-1}\right]\right\}\right\}, \qquad 1 \leq i \leq n-2, \tag{19.24e}$$

$$E\left[X'_{i:n}X'_{j:n}\right] = E\left[X'_{i:n}X'_{j-1:n}\right] + \frac{1}{n-j+1} \times \left\{E\left[X'_{i:n}\right] - n\left(\frac{1-P}{P}\right)\left\{E\left[X'_{i:n-1}X'_{j:n-1}\right] - E\left[X'_{i:n-1}X'_{j-1:n-1}\right]\right\}\right\},$$
$$1 \leq i < j \leq n-1;\ j-i \geq 2, \tag{19.24f}$$

and

$$E\left[X'_{i:n}X'_{n:n}\right] = E\left[X'_{i:n}X'_{n-1:n}\right] + E\left[X'_{i:n}\right] - n\left(\frac{1-P}{P}\right)\left\{P_1 E\left[X'_{i:n-1}\right] - E\left[X'_{i:n-1}X'_{n-1:n-1}\right]\right\}, \qquad 1 \leq i \leq n-2. \tag{19.24g}$$

Saleh, Scott, and Junkins (1975) derived exact (but somewhat cumbersome) explicit expressions for the first two single moments and the product moments of order statistics in this right-truncated exponential case. Balakrishnan and Gupta (1992) extended the results of Joshi and established recurrence relations that will enable one to find the moments and cross-moments (of order up to four) of order statistics.

By considering the doubly truncated exponential distribution with density

$$p_X(x) = \begin{cases} \dfrac{1}{P-Q} e^{-x}, & Q_1 \le x \le P_1, \\ 0, & \text{otherwise,} \end{cases} \tag{19.25}$$

where Q and 1 – P (0 < Q < P < 1) are the proportions of truncation on the left and right of the standard exponential distribution, respectively, and $Q_1 = -\log(1-Q)$ and $P_1 = -\log(1-P)$, Joshi (1979) and Balakrishnan and Joshi (1984) derived several recurrence relations satisfied by the single and the product moments of order statistics. Khan, Yaqub, and Parvez (1983) tabulated these quantities for some value of P, Q, and n. Distributions of some systematic statistics like the sample range and quasi-range were derived in this case by Joshi and Balakrishnan (1984).

7 ESTIMATION

Before 1959 a considerable amount of work had been done on inference procedures for the exponential distribution with both censored and uncensored data. (See numerous references at the end of this chapter.) It was realized, in the 1960s and 1970s, that although the exponential distribution can be handled rather easily, the consequent analysis is often poorly robust [e.g., see Zelen and Dannemiller (1961)]. Nevertheless, the study of properties of this distribution, and especially construction of estimation and testing procedures has continued steadily, during the last 30 years, with some emphasis on Bayesian analysis and order statistics methodology, and an explosion of results on characterizations. To keep this chapter within reasonable bounds, it was necessary to be very selective in our citations and descriptions of results. We first discuss classical estimators.

7.1 Classical Estimation

If $X_1, X_2, \dots, X_n$ are independent random variables each having the probability density function (19.1), then the maximum likelihood estimators of 8 and a are

$$\hat{\theta} = \min(X_1, X_2, \dots, X_n),$$
$$\hat{\sigma} = n^{-1} \sum_{i=1}^{n} (X_i - 0) = \bar{X} - \hat{\theta}. \tag{19.26}$$

If 8 is known, the maximum likelihood estimator of a is $(\bar{X} - \theta)$. Even with a known, 8" above is still the maximum likelihood estimator of 8.

The probability density function of $\hat{\theta}$ is

$$p_{\hat{\theta}}(t) = \left(\frac{n}{\sigma}\right)\exp\left[-\frac{n(t-\theta)}{\sigma}\right], \qquad t > \theta, \tag{19.27}$$

which is of the same form as (19.1) but with σ replaced by σ/n. The variance of $\hat{\theta}$ is therefore σ^2/n^2, and its expected value is 8 $+ \sigma/n$. It is interesting to note that the variance is proportional to n^{-2} and not to n^{-1}.

The expected value of $\hat{\sigma}$ $[= \bar{X} - \hat{\theta}]$ is $\sigma(1 - n^{-1})$, and its variance is

$$\sigma^2[n^{-1} + n^{-2} - 2n^{-3}].$$

The expected value of $(\bar{X} - \theta)$ is σ and its variance is $\sigma^2 n^{-1}$.

A function of special interest in some applications is the value of the probability that X exceeds a value x_0; the reliability function $R(x_0)$. If $\theta = 0$ so that (19.1) becomes

$$p_X(x) = \sigma^{-1}\exp\left(-\frac{x}{\sigma}\right), \qquad x > 0, \sigma > 0, \tag{19.28}$$

then

$$R(x_0) = e^{-x_0/\sigma}. \tag{19.29}$$

Inserting the maximum likelihood estimator, $\hat{\sigma} = n^{-1}\sum_{i=1}^n X_i$ in place of σ, would give the estimator $\exp(-x_0 n/\sum_{i=1}^n X_i)$. This is the maximum likelihood estimator of the reliability $R(x_0)$. It is biased, but a minimum variance unbiased estimator can be obtained by using the Blackwell-Rao theorem.

The statistic

$$T = \begin{cases} 1 & \text{if } X_1 > x_0, \\ 0 & \text{if } X_1 \le x_0, \end{cases}$$

is an unbiased estimator of $\exp(-x_0/\sigma)$. Since $\sum_{i=1}^n X_i$ is a complete sufficient statistic for σ [and so also for $\exp(-x_0/\sigma)$], the required minimum variance unbiased estimator is

$$E\left[T \,\middle|\, \sum_{i=1}^n X_i\right] = \Pr\left[X_1 > x_0 \,\middle|\, \sum_{i=1}^n X_i\right].$$

The ratio $X_1/\sum_{i=1}^n X_i$ has a beta distribution with parameters 1, $n - 1$, and is *independent of* $\sum_{i=1}^{n} X_i$ (Chapter 17, Section 6). Hence

$$E\left[T \middle| \sum_{i=1}^{n} X_i\right] = (n-1)\int_{x_0/\sum_{i=1}^{n} X_i}^{1} (1-z)^{n-2}\, dz$$

$$= \begin{cases} \left(1 - \dfrac{x_0}{\sum_{i=1}^{n} X_i}\right)^{n-1}, & x_0 < \sum_{i=1}^{n} X_i \\ 0, & x_0 \geq \sum_{i=1}^{n} X_i \end{cases} \qquad (19.30)$$

which is the required minimum variance unbiased estimator. This formula was obtained by Pugh (1963). (Pugh claims this is the "best" estimator but does not compare its mean square error with that of competing estimators.) The sampling distribution of the maximum likelihood estimator of parameter σ in **(19.28)**, based on a "time-censored" sample, was derived by Bartholomew (1963).

Moment estimators $(\tilde{\theta}, \tilde{\sigma})$ of (θ, σ) can be obtained by equating sample and population values of the mean and variance. They are

$$\tilde{\theta} = \bar{X} - \tilde{\sigma} \qquad (19.31a)$$

$$\tilde{\sigma} = \text{sample standard deviation.} \qquad (19.31b)$$

Cohen and Helm (1973) discuss modified moment estimators obtained by replacing **(19.31b)** by an equation that puts the first-order statistic X_1' equal to its expected value. This gives

$$\tilde{\theta}^* + n^{-1}\tilde{\sigma}^* = X_1', \qquad (19.31c)$$

which leads to

$$\tilde{\theta}^* = \frac{nX_1' - \bar{X}}{n-1}, \qquad (19.31d)$$

and

$$\tilde{\sigma}^* = \frac{n(\bar{X} - X_1')}{n-1} \qquad (19.31e)$$

They show that these are minimum variance unbiased estimators (and a fortiori **BLUEs**). Also

$$\mathrm{Var}(\tilde{\theta}^*) = \frac{\sigma^2}{n(n-1)}, \qquad (19.32a)$$

$$\mathrm{Var}(\tilde{\sigma}^*) = \frac{\sigma^2}{n-1}, \qquad (19.32b)$$

$$\mathrm{Cov}(\tilde{\theta}^*, \tilde{\sigma}^*) = \frac{\sigma^2}{n(n-1)}, \qquad (19.32c)$$

so that $\mathrm{Corr}(\tilde{\theta}^*, \tilde{\sigma}^*) = 1/\sqrt{n}$. Further, since $\tilde{\sigma}^*$ is distributed as $\frac{1}{2}(n-1)^{-1}\chi^2_{2(n-1)}$, a 100(1 – *a*)% confidence interval for σ is (in the obvious notation)

$$\left(\frac{2(n-1)}{\chi^2_{2(n-1),1-\frac{\alpha}{2}}}\tilde{\sigma}^*, \frac{2(n-1)}{\chi^2_{2(n-1),\frac{\alpha}{2}}}\tilde{\sigma}^* \right). \qquad (19.33)$$

7.2 Grouped Data

In a monograph **Kulldorff** (1961) discussed a general theory of estimation based on grouped or partially grouped samples. By ***grouped*** we mean that in disjoint intervals of the distribution range, only the numbers of observed values that have fallen in the intervals are available, and not the individual sample values. The distribution of the observed numbers is a **multinomial** distribution with probabilities that are functions of the parameter. If individual observations are available in ***some*** intervals, the sample is ***partially grouped.***

Kulldorff devoted a large part of his book to the estimation of the exponential distribution because of its simplicity. The cases studied included completely or partially grouped data, 8 unknown, a unknown, both 8 and σ unknown, a finite number of intervals, an infinite number of intervals, intervals of equal length, and intervals of unequal length. Here we describe only the maximum likelihood estimator of a when 8 is known, the intervals are not of equal length, and the number of intervals is finite.

Let $0 = x_0 < x_1 < \cdots < x_{k-1} < \infty$ be the dividing points and $N_1, \dots, N_k$, $(\sum_{i=1}^k N_i = n)$ the numbers of observed values in the respective intervals. Then the maximum likelihood estimator $\hat{\sigma}$ is the unique solution of

$$\sum_{i=1}^{k-1} \frac{N_i(x_i - x_{i-1})}{e^{(x_i - x_{i-1})/\hat{\sigma}} - 1} - \sum_{i=2}^{k} N_i x_{i-1} = 0 \qquad (19.34)$$

which exists if and only if $N_1 < n$ and $N_k <$ n. For large n,

$$n\,\mathrm{Var}(\hat{\sigma}^{-1}) \doteqdot \left(\sum_{i=1}^{k-1} \frac{(x_i - x_{i-1})^2}{e^{x_i/\sigma} - e^{x_{i-1}/\sigma}} \right)^{-1}. \tag{19.35}$$

For a given k, the dividing points that minimize the asymptotic variance are

$$\frac{x_i}{\sigma} = \sum_{j=k-i}^{k-1} \delta_j, \tag{19.36}$$

where

$$\delta_1 = g^{-1}(2), \quad \delta_i = g^{-1}\{2 + \delta_{i-1} - g(\delta_{i-1})\},$$
$$g(x) = x(1 - e^{-x})^{-1}.$$

For example,

$$k = 2, \quad \frac{x_1}{\sigma} = 1.5936;$$
$$k = 3, \quad \frac{x_1}{\sigma} = 1.0176, \qquad \frac{x_2}{\sigma} = 2.6112.$$

The simplicity of mathematical analysis for the exponential distribution permits us to construct convenient Bayesian estimators of parameters of (19.1). Some initial results in this area (for censored samples) are presented in Varde **(1969),** who also compared their performance with more natural (at least in this case) and efficient maximum likelihood and minimum variance unbiased estimators.

7.3 Estimators Using Selected Quantiles

In life-test analysis it is often supported that lifetime can be represented by a random variable with probability density function

$$p_X(x) = \sigma^{-1} \exp\left(-\frac{x}{\sigma} \right), \qquad x > 0, \sigma > 0. \tag{19.28$'$}$$

If a number n of items are observed with lifetimes commencing simultaneously, then, as each life concludes, observations of lifetime become available sequentially, starting with the shortest lifetime X_1' of the n items, followed by the **2nd,** 3rd.. . shortest lifetimes $X_2', X_3', \dots,$ respectively. Clearly it will be advantageous if useful inferences can be made at a relatively early stage, without waiting for completion of the longer lifetimes. This means that inference must be based on observed values of the k, say, shortest lifetimes, or in more general terms, the first k-order statistics (see Section 6). On account of the practical importance of these analyses, statistical techniques have been worked out in considerable detail. Here we will describe only methods of estimation, but a considerable range of test procedures is also available.

From (19.8) we find that if

$$V_2 = X_2' - X_1', \quad V_3 = X_3' - X_2', \dots, V_n = X_n' - X_{n-1}',$$

then

1. $X_1', V_2, V_3, \dots, V_n$ are mutually independent,
2. the distribution of X_1' is exponential (19.1) with $\theta = 0$, $\sigma = n^{-1}$,
3. the distribution of V_j is exponential **(19.1)** with $\theta = 0$, $\sigma = (n - j + 1)^{-1}$, j = 2, . . . , n.

Since $X_j' = X_1' + V_2 + \cdots + V_j$ $(j \geq 2)$, it follows that all linear functions of the order statistics can be expressed as linear functions of the independent random variables $X_1', V_2, \dots, V_n$. This form of representation [suggested by **Sukhatme (1937)** and **Rényi** (1953); see also Epstein and **Sobel (1954)** for a similar result] is very helpful in solving distribution problems associated with methods described in the remainder of this section. A similar kind of representation can be applied to gamma distributions, though the results are not so simple.

It is necessary to distinguish between censoring (often refered as Type *II* censoring), in which the order statistics that will be used **(e.g.,** the r smallest values) are decided in advance, and truncation (or Type *I* censoring) in which the range of values that will be used **(e.g.,** all observations less than T_0) is decided in advance (regardless of how many observations fall within the specific limits). Truncation (or Type I censoring) by omission of all observations less than a **fixed** value T_0 (> 0) has the effect that observed values may be represented by a random variable with probability density function

$$\sigma^{-1} \exp\left[-\frac{x - T_0}{\sigma}\right], \qquad x > T_0, \sigma > 0, \qquad (19.37)$$

which is again of form **(19.1)** [with T_0 (known) replacing θ**]**, and so presents no special difficulties. However, if (as is more commonly the case) truncation is by omission of all values greater than T_0 (> 0), then the corresponding probability density function is

$$p_X(x) = \left[1 - \exp\left(-\frac{T_0}{\sigma}\right)\right]^{-1} \sigma^{-1} \exp\left(-\frac{x}{\sigma}\right), \qquad 0 < x < T_0, \sigma > 0.$$

(19.38)

If m observations are obtained, they can be represented by independent random variables $X_1, X_2, \ldots, X_m$, each with distribution (19.38). The maximum likelihood equation for an estimator $\hat{\sigma}_{T_0}$ of a is

$$\hat{\sigma}_{T_0} = m^{-1} \sum_{j=1}^{m} x_j + T_0 \left[\exp\left(\frac{T_0}{\hat{\sigma}_{T_0}} \right) - 1 \right]^{-1}. \tag{19.39}$$

This equation may be solved by an iterative process. In this work the table of Barton, David, and Merrington (1963) is useful.

Wright, Engelhardt, and Bain (1978) and Piegorsch (1987) studied inference on θ and a for distribution (19.1) under Type I censoring. Wright, Engelhardt, and Bain (1978) presented procedures based on the conditional distribution of "failure" times, given the number D of failures occurring before the censoring time τ. They distinguished between testing with and without replacement. In the first case X_1' and D are sufficient statistics, and in the second D and $S = \sum_{j=1}^{D} X_j'$ are sufficient.

Assuming testing without replacement, they utilized the facts that given D = d, the conditional distribution of $[(X - X_1')/(X - \theta)]^d$ is uniform on (0, 1), and that for fixed 8, D is sufficient and complete for a , while for fixed σ, X_1' is sufficient and complete for 8.

Assuming testing without replacement leads to a rather complicated conditional density for Xi, given D and S. However, Wright, Engelhardt, and Bain (1978) provide a table of exact percentage points of X_1' for small and moderate D and an approximation for large D. [In the case where 8 = 0, Bartholomew (1963) developed confidence intervals based on the maximum likelihood estimator

$$\hat{\sigma} = \frac{\sum_{i=1}^{D} X_i' + (n - D)\tau}{D} = \frac{T_{D:n}}{D} \tag{19.40}$$

where τ is the termination time, (provided that D > 0)].

Piegorsch (1987) uses somewhat simpler methods, based on the approximate distribution of the likelihood ratio (LR) test statistic and some other approximations reviewed in Lawless (1982, Sec. 3.5.2). He discusses small sample performance of these procedures, based on Monte Carlo evaluation. He introduces the sets

$$A = (\text{i}: \min(t_i, \tau) = t_i, \text{i} = 1, \ldots, n\}$$

and

$$\Gamma = \{i: \min(t_i, \tau) = \tau, i = 1, \ldots, n\}$$

and proposes estimators

$$\tilde{\theta} = X_1' \tag{19.41a}$$

and

$$\tilde{\sigma} = \left[\Sigma_\Delta\left(X_i - \tilde{\theta}\right) + \Sigma_\Gamma \max(\tau - \tilde{\theta}, 0)\right], \qquad D > 0, \tag{19.41b}$$

and observed lifetimes t_i and **fixed** censoring time at τ.

The LR statistic for testing $\sigma = \sigma_0$ is

$$2D\left\{\tilde{\sigma}\sigma_0^{-1} - \log\left(\tilde{\sigma}\sigma_0^{-1}\right) - 1\right\}. \tag{19.42}$$

The asymptotic distribution of this statistic is χ^2 with one degree of freedom if $\theta = \theta_0$. To construct approximate $100(1 - \alpha)\%$ confidence limits for 8, Piegorsch (1987) suggests solving the equations

$$\tilde{\sigma}\sigma^{-1} - \log(\tilde{\sigma}\sigma^{-1}) = 1 + \chi^2_{1,1-\alpha}(2D)^{-1}.$$

This equation always has two solutions, one for $\tilde{\sigma}/\sigma < 1$ and the other for $\tilde{\sigma}/\sigma > 1$. These solutions L_σ and U_σ are then used to construct an asymptotic $100(1 - \alpha)\%$ confidence interval on a of the form

$$\frac{\tilde{\sigma}}{U_\sigma} \le \sigma \le \frac{\tilde{\sigma}}{L_\sigma}. \tag{19.43}$$

A similar approach yields asymptotic 100(1 − a)% confidence limits for 8. For smaller sample sizes ($n \le 10$), the conditional inference on σ given by Wright, Engelhardt, and Bain (1978) may be preferred. However, Piegorsch's method for 8 performs well even for n = 5 with an F-based approximation providing errors closer to nominal than the LR-based values particularly at $\alpha = 0.05$.

Joint regions for (σ, θ) can be constructed using Bonferroni's approach, with a rectangular region corresponding to the cross-product of univariate $1 - \frac{\alpha}{2}$ confidence intervals on 8 and σ (the elliptic procedure using the asymptotic normality of the MLEs breaks down, yielding a hyperbolic rather than an elliptic region). Ranking and subset selection procedures for exponential population with Type I (and Type II) censored data are discussed in Berger and Kim (1985).

We will now restrict ourselves to Type II censored samples only. We will give details only for the case where $\theta = 0$ [so that the probability density function is as in (19.28)] and where censoring results in omission of the largest n − k values (i.e., observation of the k smallest values, where k is specified prior to obtaining the observations). The joint probability density function of the k (> 1) smallest observations in a random sample of size n is

$$p_{X'_1,\ldots,X'_k}(x_1,\ldots,x_k) = \frac{n!}{(n-k)!\sigma^k}\exp\left(-\frac{\sum_{j=1}^{k-1}x_j + (n-k+1)x_k}{\sigma}\right),$$

$$0 \le x_1 \le x_2 \le \cdots \le x_k. \quad (19.44)$$

The maximum likelihood estimator of σ is

$$Y'_k = k^{-1}\left\{\sum_{j=1}^{k-1} X'_j + (n-k+1)X'_k\right\} = k^{-1}T_{k:n}. \quad (19.45)$$

This statistic is distributed as $(\frac{1}{2}\sigma/k) \times (\chi^2$ with 2k degrees of freedom). The expected value of Y'_k is therefore σ, and its variance is σ^2/k. The limits

$$\frac{2kY'_k}{\chi^2_{2k,1-\frac{1}{2}\alpha}} \quad \text{and} \quad \frac{2kY'_k}{\chi^2_{2k,\frac{1}{2}\alpha}}$$

define a confidence interval for σ with confidence coefficient $100(1-\alpha)\%$.

A wide variety of estimators of σ and 8 based on order statistics is available. Many references at the end of this chapter contain discussions of such estimators. Among the problems discussed are estimation

1. by linear functions of a limited number (usually not more than about five) of order statistics—this includes both the choice of coefficients in the linear function, and of the order statistics to be used;
2. when only the k ***largest*** values are observed;
3. when predetermined numbers of both the smallest and largest values are omitted;
4. conversely, when only these values are known—that is, the middle group of observations is omitted.

In all cases formulas have been obtained appropriate to estimation of σ, knowing 8; of θ, knowing σ; and of both θ and σ, neither being known. We now discuss some of the more useful of these formulas.

The variance-covariance matrix of the order statistics $X'_1 \le X'_2 \le \cdots < X'_n$ has elements

$$\text{Var}(X'_r) = \sigma^2\sum_{j=1}^{r}(n-j+1)^{-2} = \text{Cov}(X'_r, X'_s), \qquad r < s. \quad (19.46)$$

Also

$$E[X'_r] = \theta + \sigma \sum_{j=1}^{r} (n - j + 1)^{-1}. \tag{19.47}$$

From these relationships it is straightforward to construct best linear unbiased estimators based on k-selected order statistics $X'_{n_1}, X'_{n_2}, \ldots, X'_{n_k}$ with

$$n_1 < n_2 < \cdots < n_k.$$

It will be convenient to use the notation

$$w_{mi} = \sum_{j=n_{i-1}}^{n_i - 1} (n - j)^{-m}, \tag{19.48}$$

with $n_0 = 0$, and w_{10}/w_{20}, $w_{1,k+1}/w_{2,k+1}$ each defined to be zero.

If θ is known, the best linear unbiased estimator of σ is

$$\tilde{\sigma} = \left[\sum_{i=1}^{k} \left(\frac{w_{1i}}{w_{2i}} - \frac{w_{1,i+1}}{w_{2,i+1}} \right) X'_{n_i} - \frac{w_{11}}{w_{21}} \theta \right] \left[\sum_{i=1}^{k} \frac{w_{1i}^2}{w_{2i}} \right]^{-1}. \tag{19.49}$$

The variance of $\tilde{\sigma}$ is $\sigma^2(\sum_{i=1}^{k} w_{1i}^2 w_{2i}^{-1})^{-1}$. Some special cases are

1. $n_i = i$, $k = n$ (complete sample):

$$\tilde{\sigma} = n^{-1} \sum_{i=1}^{n} X_i,$$

$$\operatorname{Var}(\tilde{\sigma}) = \frac{\sigma^2}{n}.$$

2. $n_i = r_1 + i$, $k = n - r_1 - r_2$ (censoring r_1 smallest and r_2 largest values):

$$\tilde{\sigma} = K^{-1} \left[\frac{\sum_{i=1}^{r_1+1} (n - i + 1)^{-1}}{\sum_{i=1}^{r_1+1} (n - i + 1)^{-2}} (X'_{r_1+1} - \theta) - (n - r_1) X'_{r_1+1} \right.$$

$$\left. + (n - r_1 - k) X'_{r_1+k} + \sum_{i=r_1+1}^{r_1+k} X'_i \right] \tag{19.50}$$

(1963)]. Epstein (1962) has noted that the efficiency of the unbiased estimator based on $X'_r - 8$ is never less than 96% if $r/n \le \frac{1}{2}$, or 90% if $r/n \le \frac{2}{3}$.

5. $k = 2$ (estimation from two order statistics X'_{n_1}, X'_{n_2}). The variance of $\tilde{\sigma}$ is a minimum when n_1 is the nearest integer to $0.6386(n + \frac{1}{2})$ and n_2 is the nearest integer to $0.9266(n + \frac{1}{2})$ [Siddiqui (1963)].

For small samples Sarhan and Greenberg (1958) give the following optimal choices:

Sample Size (n)	2–4	5–7	8–11	12–15	16–18	19–21
n_1	$n-1$	$n-2$	$n-3$	$n-4$	$n-6$	$n-7$
n_2	n	n	n	n	$n-1$	$n-1$

If θ is not known, the optimal choices are different. Sarhan and Greenberg (1958) give the following optimal choices for n_1 and n_2 (with $k = 2$):

Sample Size (n)	2–6	7–10	11–15	16–20	21
n_1	1	1	1	1	1
n_2	n	$n-1$	$n-2$	$n-3$	$n-4$

The best (in fact only) linear unbiased estimator of a, using only X'_{n_1} and X'_{n_2} is

$$(X'_{n_2} - X'_{n_1})\left[\sum_{j=n_1}^{n_2-1}(n-j)^{-1}\right]^{-1} \tag{19.53}$$

and its variance is

$$\left(\sum_{j=n_1}^{n_2-1}(n-j)^{-2}\right)\left(\sum_{j=n_1}^{n_2-1}(n-j)^{-1}\right)^{-2}\sigma^2. \tag{19.54}$$

The optimal value of n_1 is always 1, whatever the values of n and n_2.

Kulldorff (1962) considered the problem of choosing $n_1, n_2, \cdots, n_k$ to minimize $\operatorname{Var}(\tilde{\sigma})$ or $\operatorname{Var}(\tilde{\theta})$ for k fixed. He tabulated the optimal n_i's and coefficients for small values of k and n. Earlier, Sarhan and Greenberg (1958) treated the asymptotic (large n) case (for θ known) giving the optimal percentiles $n_1/n, n_2/n, \ldots, n_k/n$, for k fixed. These tables are reproduced by Ogawa (1962). Saleh and Ali (1966) and Saleh (1966) proved the uniqueness of the optimal selection and extended the results to censored cases.

The table of Zabransky, Sibuya, and Saleh (1966) is most exhaustive and covers uncensored and censored samples, finite and asymptotic cases for a wide range. Sibuya (1969) gave the algorithms for computing their tables, and unified previous results in simpler form.

(1963)]. Epstein (1962) has noted that the efficiency of the unbiased estimator based on $X'_r - \theta$ is never less than 96% if $r/n \leq \frac{1}{2}$, or 90% if $r/n \leq \frac{2}{3}$.

5. $k = 2$ (estimation from two order statistics X'_{n_1}, X'_{n_2}). The variance of $\tilde{\sigma}$ is a minimum when n_1 is the nearest integer to $0.6386(n + \frac{1}{2})$ and n_2 is the nearest integer to $0.9266(n + \frac{1}{2})$ [Siddiqui (1963)].

For small samples Sarhan and Greenberg (1958) give the following optimal choices:

Sample Size (n)	2–4	5–7	8–11	12–15	16–18	19–21
n_1	$n-1$	$n-2$	$n-3$	$n-4$	$n-6$	$n-7$
n_2	n	n	n	n	$n-1$	$n-1$

If θ is not known, the optimal choices are different. Sarhan and Greenberg (1958) give the following optimal choices for n_1 and n_2 (with $k = 2$):

Sample Size (n)	2–6	7–10	11–15	16–20	21
n_1	1	1	1	1	1
n_2	n	$n-1$	$n-2$	$n-3$	$n-4$

The best (in fact only) linear unbiased estimator of σ, using only X'_{n_1} and X'_{n_2} is

$$(X'_{n_2} - X'_{n_1})\left[\sum_{j=n_1}^{n_2-1}(n-j)^{-1}\right]^{-1} \tag{19.53}$$

and its variance is

$$\left[\sum_{j=n_1}^{n_2-1}(n-j)^{-2}\right]\left[\sum_{j=n_1}^{n_2-1}(n-j)^{-1}\right]^{-2}\sigma^2. \tag{19.54}$$

The optimal value of n_1 is always 1, whatever the values of n and n_2.

Kulldorff (1962) considered the problem of choosing $n_1, n_2, \ldots, n_k$ to minimize $\operatorname{Var}(\tilde{\sigma})$ or $\operatorname{Var}(\tilde{\theta})$ for k fixed. He tabulated the optimal n_i's and coefficients for small values of k and n. Earlier, Sarhan and Greenberg (1958) treated the asymptotic (large n) case (for θ known) giving the optimal percentiles $n_1/n, n_2/n, \ldots, n_k/n$, for k fixed. These tables are reproduced by Ogawa (1962). Saleh and Ali (1966) and Saleh (1966) proved the uniqueness of the optimal selection and extended the results to censored cases.

The table of Zabransky, Sibuya, and Saleh (1966) is most exhaustive and covers uncensored and censored samples, finite and asymptotic cases for a wide range. Sibuya (1969) gave the algorithms for computing their tables, and unified previous results in simpler form.

Epstein (1962) gave a number of useful results. He pointed out that $X'_r - \theta$ is approximately distributed as $\frac{1}{2}(\sigma/r) \times (\chi^2$ with $2r$ degrees of freedom). Approximate confidence intervals for σ can be constructed on this basis. Epstein also gave formulas for the minimum variance unbiased estimators of θ and σ for the two-parameter distribution (19.1), based on $X'_1, X'_2, \dots, X'_k$. These are

$$\sigma^* = (k-1)^{-1} \sum_{j=1}^{k-1} (n-j)(X'_{j+1} - X'_j), \tag{19.55a}$$

$$\theta^* = X'_1 - n^{-1}\sigma^*; \tag{19.55b}$$

σ^* is distributed as $\frac{1}{2}\sigma(k-1)^{-1} \times [\chi^2$ with $2(k-1)$ degrees of freedoml. $100(1-$ a)% confidence limits for σ are

$$\frac{2(k-1)\sigma^*}{\chi^2_{2(k-1),1-\frac{\alpha}{2}}} \quad \text{and} \quad \frac{2(k-1)\sigma^*}{\chi^2_{2(k-1),\frac{\alpha}{2}}}.$$

$100(1-$ a)% confidence limits for 8 are

$$X'_1 - F_{2,2(k-1),1-\alpha}\sigma^* n^{-1} \quad \text{and} \quad X'_1 \tag{19.56}$$

(using the notation of Chapter 27, Section 1).

If only $X'_{k_1+1}, \dots, X'_{n-k_2}$ are to be used, the minimum variance unbiased estimators are

$$\sigma^* = (n-k_1-k_2-1)^{-1} \sum_{j=k_1+1}^{n-k_2-1} (n-j+1)(X'_{j+1} - X'_j), \tag{19.57a}$$

$$\theta^* = X'_{k_1+1} - \sigma^* \sum_{j=0}^{k_1} (n-j)^{-1}; \tag{19.57b}$$

σ^* is distributed as $\frac{1}{2}\sigma(n-k_1-k_2-1)^{-1} \times [\chi^2$ with $2(n-k_1-k_2-1)$ degrees of freedoml and

$$\operatorname{Var}(\theta^*) = \sigma^2 \left[\sum_{j=0}^{k_1} (n-j)^{-2} + (n-k_1-k_2-1)^{-1} \left\{ \sum_{j=0}^{k_1} (n-j)^{-1} \right\}^2 \right]. \tag{19.58}$$

The minimum variance unbiased estimator (MVUE) of the reliability **func-**

tion based on $X_1', \ldots, X_k'$ is

$$\hat{R}(t) = \begin{cases} \left(1 - \dfrac{t}{T_{k:n}}\right)^{k-1} & \text{if } t \le T_{k:n}, \\ 0 & \text{if } t \ge T_{k:n}. \end{cases} \tag{19.59}$$

[$T_{k:n}$ is defined in (19.45)].

The mean square error of $\hat{R}(t)$ is

$$MSE(\hat{R}(t)) = \int_c^{\infty} \left(1 - \frac{c}{u}\right)^{2r-2} \gamma(u; r)\, du$$

where

$$\gamma(u; r) = \frac{1}{\Gamma(r)} u^{r-1} e^{-u},$$

$$c = \frac{t}{\sigma}$$

[Basu (1964)]. Based on past data, one may have a prior (guessed) value of a , a,, say, that can be utilized for statistical inference. The so-called shrinkage estimators arising in this situation perform better than the MVUE if the value σ_0 is close to a. A preliminary test can be conducted to check the closeness of σ_0 to a. Chiou (1987) proposes incorporating a preliminary test on a , using an estimator of type

$$R_{PT}(t) = \begin{cases} \exp\left(-\dfrac{t}{\sigma_0}\right) & \text{if } H_0\colon \sigma = \sigma_0 \text{ is not rejected}, \\ \hat{R}(t) & \text{otherwise}. \end{cases} \tag{19.60}$$

Since $2T_{k:n}/\sigma$ has a χ^2 distribution with 2k degrees of freedom, H_0 is not rejected if

$$C_1 \le \frac{2T_{k:n}}{\sigma_0} \le C_2,$$

where C_1 (C_2) are the $\frac{\alpha}{2}$ lower (upper) percentage points of χ^2_{2k}. Thus

$$\hat{R}_{PT}(t) = \begin{cases} \exp\left(-\dfrac{t}{\sigma_0}\right) & \text{if } C_1 \le \dfrac{2T_{k:n}}{\sigma_0} \le C_2, \\ \hat{R}(t) & \text{otherwise}. \end{cases} \tag{19.61}$$

Chiou (1987) provides optimal critical values for the preliminary test and their corresponding level of significance, based on the minimax regret criterion.

Alternatively, Chiou (1992) proposes the "shrinkage estimator"

$$\hat{R}_2(t) = \begin{cases} \omega \exp\left(-\dfrac{t}{\sigma_0}\right) + (1-\omega)\hat{R}(t) & \text{if } C_1 \le \dfrac{2T_{k:n}}{\sigma_0} \le C_2 \\ \hat{R}(t) & \text{otherwise,} \end{cases} \tag{19.62}$$

where ω is a shrinkage coefficient suitably specified. Chiou provides a table of optimum values of shrinkage coefficient w for $t/\sigma_0 = 0.25(0.25)2.0$ and $k = 4(2)10$ as well as critical values C_1 and C_2 for the preliminary test.

Zacks and Even (1966) compared the performance of MVUE and the maximum likelihood estimator in terms of mean square error for small samples. The MLE is more efficient than MVUE over the interval $0.5 < t/\sigma < 3.5$. Over the "effective intervals where σ_0/σ is in the vicinity of 1 [(0.7, 1.4) for $r = 4$ and $t/\sigma_0 = 2$ as an example], Chiou's (1992) shrinkage estimators are more efficient, but none of the estimators for $R(t)$ investigated so far is uniformly better than others over the whole possible range of σ/σ_0."

Cohen and Whitten (1988) discuss estimation of parameters in the case of progressively censored samples. Censoring is progressively carried out in k stages at times $\tau_1 < \tau_2 < \cdots < \tau_j < \cdots < \tau_k$. At jth stage of censoring, C_j sample specimens, selected randomly from the survivors at time τ_j, are removed. In addition we have n full-term observations and our sample consists of $(X_1, X_2, \dots, X_n)$ plus k partial term observations $\{C_j\tau_j\}$ $(j \ge 1, \dots, k)$. Thus $N = n + r$, where $r = \Sigma_1^k C_j$ ($C_j \ge 1$ corresponds to censoring and $C_j = 0$ to noncensoring), and the sum total (ST) term of all N observations in the sample is

$$ST = \sum_{i=1}^{n} X_i + \sum_{j=1}^{k} C_j\tau_j.$$

The modified maximum likelihood estimators (MMLEs) of σ and 8 in the case of progressive censoring, obtained by solving

$$\frac{\partial \ln L}{\partial \sigma} = 0 \quad \text{and} \quad E[X_1'] = X_1'$$

are given by

$$\tilde{\sigma} = \frac{ST - N\tilde{\theta}}{n}$$

$$\tilde{\theta} = \frac{ST - NX_1'}{n-1}. \tag{19.63a}$$

Explicitly

$$\tilde{\sigma} = \frac{ST - NX'_1}{n-1},$$

$$\tilde{\theta} = \frac{nX'_1 - N^{-1}ST}{n-1}. \qquad (19.63b)$$

(For the uncensored case $C_j = 0$, n = N, and ST = $n\bar{X}$.)

Estimation of parameters from general location-scale families (of which the two-parameter exponential distribution is a member) under progressive censored sampling was studied by Viveros and Balakrishnan (1994). They follow the classical arguments for conditional inference for location and scale parameters as expounded, for example, in Lawless (1982).

A modified hybrid censoring model was investigated by Zacks (1986). Let τ_0 be the fixed time at which the (Type I) censoring occurs. Let $X'_{k:n}$ denote the kth-order statistic of a sample of n i.i.d. random variables from the one-parameter exponential distribution (19.28).

Let $X_k^* = \min(X'_{k:n}, \tau_0)$. The recorded random variable is the time-censored kth-order statistic. Let $X_{k1}^*, \dots, X_{km}^*$ denote m i.i.d. random variables distributed as X_k^*. Zacks (1986) shows the existence and uniqueness of a MLE of σ based on the sample $\mathbf{X}_{k,m}^* = (X_{k1}^*, \dots, X_{km}^*)$.

He also investigates the properties of the moment estimator (ME) that is the root of the equation in σ

$$\sigma \sum_{j=0}^{k-1} \frac{1}{n-j}\left[1 - B(j; n; e^{-\tau_0/\sigma})\right] = \bar{X}_k^* = m^{-1} \sum_{j=1}^{m} X_{kj}^*, \qquad (19.64)$$

where $B(j; n; p)$ is the cdf of the binomial distribution with parameters (n, p).

Numerical comparisons show that in the case of MLE "censoring has a dramatic effect on the possibility of estimating σ efficiently. The efficiency of the MLE drops to almost zero when σ is in the neighborhood of τ_0." The asymptotic relative efficiency of ME relative to MLE is an increasing function of $\tau_0/\sigma \equiv$ q. When $\eta = 1$, the ARE of the ME is about 62%, whereas at $\eta = 3$ it is 99%. On the other hand, under censoring, the MLE is considerably more efficient than the ME when η is close to 1.

7.4 Estimation of Quantiles

Robertson (1977), in an important paper, develops estimation procedures for quantiles of the exponential distribution (19.28) such that mean square errors in the predicted distribution function is minimized. For complete random samples of size n, a particular quantile, represented as $K\sigma$, has estimators of the form $K^*\bar{X}$. The optimal estimator of $K\sigma$ (with squared error loss) is in fact $K_0\bar{X}$, where

$$K_0 = n\left[\frac{\exp\{K/(n+1)\} - 1}{2 - \exp\{K/(n+1)\}}\right]. \tag{19.65}$$

Its mean square error is

$$e^{-2K} - e^{-K}\left[2 - \exp\left\{\frac{K}{n+1}\right\}\right]^{n+1} \tag{19.66}$$

For a linear estimator $\Sigma_{i=1}^n a_i X_i$, the mean square error is

$$\int_0^\infty \cdots \int_0^\infty \{e^{-K} - e^{\Sigma a_i x_i}\}^2 e^{-\Sigma x_i}\, dx_1\, dx_2 \cdots dx_n$$
$$= e^{-2K} - 2e^{-K}\prod_{i=1}^n (1 + a_i)^{-1} + \prod_{i=1}^n (1 + 2a_i)^{-1} \tag{19.67}$$

where $\sigma = 1$ for convenience, yielding minimizing values

$$a_i = \frac{[\exp\{K/(n+1)\} - 1]}{2 - \exp\{K(n+1)\}} \quad \text{for all } i \tag{19.68}$$

For $n < (K/\log 2) - 1$, this estimator will break down, and the optimum choice will be $K_0 = \infty$. [Indeed one cannot reasonably expect to estimate the 99% point ($K = 4.605$) using a sample of size 5 [$(< (K/\log 2) - 1)$.] Comparison of $K_0\bar{X}$ with $K\bar{X}$ (the natural K) and $K_2\bar{X}$ [chosen to make $\exp(-K_2\bar{X})$ an unbiased estimator of $\exp(-K\sigma)$] has been carried out by Robertson (1977). For small n or large K (extreme percentiles) there are considerable differences among the three estimators.

To estimate $K\sigma$ by means of a single order statistic the estimator is of the form K_3X_r', where the optimal choice for large n turns out to be $r \doteqdot an$ with $a = 0.79681213$ (the positive root of equation $e^{-2\alpha} = 1 - \alpha$) and

$$K_3 = \frac{K}{2\alpha} + \frac{3K^2 - 2(1-\alpha)K}{16\alpha^2(1-\alpha)n} + O(n^{-2}). \tag{19.69}$$

For large n, K_3X_r' has efficiency approximately 65% [$\doteqdot 4\alpha(1-\alpha)\%$] relative to $K_0\bar{X}$, the optimal estimator based on the complete sample.

7.5 Bayesian Estimation

One-Parameter Exponential Distributions

In Bayesian estimation prior distributions are often assigned to the hazard rate ($\lambda = \sigma^{-1}$) rather than to the expected value (σ). We therefore use the

pdf

$$p_X(x|\lambda) = \lambda e^{-\lambda x} \qquad (x > 0; \lambda > 0). \tag{19.28)''$$

If the prior on A is uniform over $(0, M)$, then the posterior density of λ, given a random sample $X_1, X_2, \dots, X_n$ is

$$\frac{\lambda^n \exp(-\lambda \Sigma_{i=1}^n X_i)}{\int_0^M \lambda^n \exp(-\lambda \Sigma_{i=1}^n X_i)\, d\lambda}, \qquad 0 < \lambda < M. \tag{19.70}$$

For large M this is approximately a gamma $(n + 1, \{\Sigma_{i=1}^n X_i\}^{-1})$ distribution with pdf

$$\frac{(\Sigma_{i=1}^n X_i)^{n+1} \lambda^n}{\Gamma(n+1)} \exp\left(-\lambda \sum_{i=1}^n X_i\right). \tag{19.71}$$

If this posterior pdf is used as a new prior, and a random sample $Y_1, Y_2, \dots, Y_m$ obtained, the new posterior is of the same form as (19.71), with n increased to $(n + m)$ and $\Sigma_{i=1}^n X_i$ increased by $\Sigma_{j=1}^m Y_j$. Thus the gamma prior is a "natural conjugate" prior for A [see Barlow and Proschan (1979)].

Waller et al. (1977) develop an interesting and potentially fruitful approach to determination of the parameters of a gamma (a, β) prior on A in the pdf (19.28)'', based on values of

$$\Pr[\lambda < \lambda_j] = p_j, \qquad j = 1, 2; \lambda_1 < \lambda_2,$$

decided upon by a researcher. Extensive tables and graphs are provided giving values of a and β for selected values A, and p_0 for which

$$\Pr[\lambda < A,] = p_0. \tag{19.72}$$

By overlapping transparencies of graphs for (A,, p_1) and (A,, p_2), appropriate values of a and β can be determined. If (19.72) is satisfied then λ_0/β is constant for given a. For small values of a, Waller et al. (1977) recommended the approximation A $\sim \beta p^{1/\alpha}$.

For situations where only the first k "failures" are observed, with k specified in advance (Type II censoring), the MLE of A is

$$\hat{\lambda} = kT_k^{-1}, \tag{19.73}$$

with $T_k = \Sigma_{i=1}^k X_i' + (n - k)X_k'$. With a gamma (a, b^{-1}) prior on A, we arrive at a posterior gamma $(a + k, (b + T_k)^{-1})$ distribution of A. Similar results are obtained for truncated sampling (Type I censoring), and also for inverse binomial sampling (when a failed or truncated unit is replaced).

If λ^{-1} $(=\sigma)$ is ascribed a "uniform" distribution, the posterior distribution for σ would have pdf

$$\frac{(\Sigma_{i=1}^{n}X_i)^{n-1}}{\Gamma(n-1)}\left(\frac{1}{\sigma}\right)^{n}\exp\left(-\frac{1}{\sigma}\sum_{i=1}^{n}X_i\right), \qquad \sigma>0, \tag{19.74}$$

which is called the inverted gamma distribution.

Since an inverted gamma prior on σ also results in an inverted gamma posterior distribution for σ, it is a natural conjugate in this situation. This is not surprising since, if λ has a gamma distribution, σ $(=\lambda^{-1})$ has an inverted gamma distribution.

Bayesian estimation of the reliability function

$$R(t)=\Pr[X>t]=\exp\left(-\frac{t}{\sigma}\right) \tag*{(19.29)$'$}$$

has been studied by Sinha and Guttman (1976). Assuming a so-called vague Jeffreys's prior density (proportional to σ^{-1}) on σ, and using only the first k order statistics, they show that the posterior density of $R(t)$ is

$$\frac{S_k^k}{\Gamma(k)}\{-\log R(t)\}\{R(t)\}^{S_k-1}, \qquad R(t)\geq 0, \tag{19.75}$$

where $S_k=T_k/t$, which can be regarded as the "cumulative life per unit time" up to time t. The posterior density for σ in this situation is

$$\frac{T_k^k}{\Gamma(k)}\sigma^{-(k+1)}\exp\left(-\frac{T_k}{\sigma}\right). \tag{19.76}$$

See also Shoukri (1983) for further discussions.

Villén-Altamizano (1990) used a gamma (p,a^{-1}) prior for λ $(=\sigma^{-1})$, obtaining the posterior density of $R(1)=\exp(-\sigma^{-1})$ given the results (W_j,I_j) $(j=1,\dots,n)$ from a randomly censored sample of size n with lifetimes $\{X_j\}$ and censoring times $\{Y_j\}$, where

$$W_j=\min(X_j,Y_j)$$

and

$$I_j = I(X_j \le Y_j) = \begin{cases} 1 & \text{if } X_j \le Y_j, \\ 0 & \text{if } X_j > Y_j, \end{cases}$$

in the form

$$p_R(r|W,I) = \frac{(a+W)^{p+I}}{\Gamma(p+1)} r^{a+W-1}(-\log r)^{p+I-1}, \qquad 0 < r < 1, \quad (19.77)$$

where W $= \Sigma_{j=1}^{n} W_j$ and $I = \Sigma_{j=1}^{n} I_j$.

The posterior expected value is

$$\tilde{R}_2 = E[R|W,I] = \left(\frac{a+W}{a+W+1}\right)^{p+I}. \qquad (19.78)$$

This is the Bayesian estimator (optimal for quadratic loss). The mode of the posterior distribution is

$$\tilde{R}_3 = \begin{cases} \exp\left(-\dfrac{p+I-1}{a+W-1}\right) & \text{if } W > 1-a, \\ 0 & \text{otherwise.} \end{cases} \qquad (19.79)$$

Two-Parameter Exponential Distributions

For the two-parameter distribution

$$p_X(x) = \sigma^{-1} \exp\left\{-\frac{x-\theta}{\sigma}\right\}, \qquad x \ge \theta > 0; \sigma > 0, \qquad (19.1)'$$

Sinha and Guttman (1976) ascribed a joint prior to 8 and σ that is proportional to σ^{-a} (a > 0), and obtained the following results for Type I censoring using the first k order statistics in a random sample of size n. The posterior expected values (Bayes estimators) of θ and σ are

$$E[\theta|\mathbf{X}] = \frac{C_k\{n(k+a-3)X_1' - C_{k-1}^{-1}S_k'\}}{n(k+a-3)}, \qquad (19.80)$$

$$E[\sigma|\mathbf{X}] = \frac{C_k S_k'}{C_{k-1}(k+a-3)}, \qquad (19.81)$$

where

$$S'_k = \sum_{i=1}^{k} (X'_i - X'_1) + (n-k)(X'_k - X'_1),$$

$$C_k = \left\{1 - \left(1 + \frac{nX'_1}{S'_k}\right)^{-(k+a-2)}\right\}^{-1}.$$

For the reliability function

$$R(t) = \exp\left(-\frac{t-\theta}{\sigma}\right),$$

we have

$$E[R(t)|\mathbf{X}] = \frac{n}{n+1}C_k\left\{\left(1 + \frac{t - X'_1}{S_k}\right)^{-(k+a-2)} - \left(1 + \frac{t + nX'_1}{S_k}\right)^{-(k+a-2)}\right\}, \qquad t > X'_1. \quad (19.82)$$

This estimator was found to be reasonably robust with respect to a. The posterior distribution of $R(t)$ is quite complicated [see also Pierce (1973)]. Trader (1985) used a truncated normal prior.

7.6 Miscelianeous

Maximum probability estimation **(MPE)** is a method of estimation proposed by Weiss and Wolfowitz (1967) [see also Weiss (1985)]. The essential idea is to seek an interval of predetermined length 1, say, that maximizes the integral of the likelihood function over the interval. In the case of the two-parameter exponential distribution (19.1), the analysis for estimation of *8* takes an especially simple form. Given a random sample of size n with values $X_1, \dots, X_n$, the likelihood function is

$$L(\mathbf{X}|\theta, a) = \begin{cases} \exp\left\{-\dfrac{\sum_{i=1}^{n}(X_i - \theta)}{\sigma}\right\} & \sigma > 0,\ \theta \le \min X_i = X'_1, \\ 0, & \theta > X'_1. \end{cases} \quad (19.83)$$

Since L = 0 for 8 > X_1' and L is an increasing function of θ for 8 $\leq X_1'$, the MPE interval of length l is

$$(X_1' - l, X_1') \tag{19.84}$$

(whatever the value of a). Blyth (1982) followed this analysis by considering choice of attractive values for 1, if $(X_1' - \frac{1}{2}l)$ is used as a point estimator of 8. He found that

1. for minimum mean square error, $\frac{1}{2}l = n^{-1}\sigma$,
2. for minimum absolute error, $\frac{1}{2}l =$ **n- 'a** log 2.

Introducing a quadratic loss function, Arnold (1970), and independently Zidek (1973) and Brewster (1974), derived optimal estimators of a for cases where θ is unknown. The MLE of a, $\hat{\sigma} = \bar{X} - X_1'$, is also the best affine equivariant estimator (BAEE) of a. However, the estimator is inadmissible for a wide class of loss functions satisfying conditions of differentiability and bowl-shapedness. Modified estimators, aimed at reducing this drawback, were suggested by the above authors.

Estimators of the ratio $(\delta = \sigma_2/\sigma_1)$ of scale parameters of two distributions π_1, π_2 of form (19.1) with values (σ_j, θ_j) $(j = 1, 2)$ for the parameters (σ, θ), respectively, were investigated by Madi and Tsui (1990). Suppose that there are two independent ordered random sample values

$$X_{11}' \text{ I } X_{12}' \leq \cdots \leq X_{1n_1}' \quad \text{from } \pi_1$$

and

$$X_{21}' \text{ I } X_{22}' \leq \cdots \leq X_{2n_2}' \quad \text{from } \pi_2.$$

With loss function $W(\delta, \delta^*)$ whence δ^* is an estimator of 6, the statistics

$$\left(\bar{X}_1 - X_{11}', \bar{X}_2 - X_{21}'\right) \left(\frac{X_{11}'}{\bar{X}_1 - X_{11}'}, \frac{X_{21}'}{\bar{X}_2 - X_{21}'}\right)$$

are sufficient for the four parameters a,, a,, θ_1, and θ_2. Assuming that $W(\delta, \delta^*) = W(\delta^*/\delta)$, the BAEE estimator is

$$\delta^* = \begin{cases} \dfrac{\bar{X}_1 - X_{11}'}{\bar{X}_2 - X_{21}'} \min\left[c_0, c_1 \dfrac{\bar{X}_1}{\bar{X}_1 - X_{11}'}\right] & \text{if } X_{11}' > 0, \\ \dfrac{\bar{X}_1' - X_{11}'}{\bar{X}_2 - X_{21}'} c_0 & \text{if } X_{11}' \leq 0, \end{cases} \tag{19.85}$$

where c_0 minimizes $\int_0^\infty t^{n_1-2}W(ct)(1 + n_1t/n_2)^{-(n_1+n_2-2)}\,dt$ and c_1 minimizes $\int_0^\infty t^{n_1-1}W(ct)(1 + n_1t/n_2)^{-(n_1+n_2-1)}\,dt$ provided that

1. W is differentiable,
2. $W(y)$ is bowl-shaped [i.e., $W(y)$ decreases with y for $y \le y_0$, increases for $y > y_0$], attaining its minimum value at $y = 1$,
3. if $\sigma_1 = \sigma_2 = 1$, then

$$E\left[\left|W'\left(c\frac{\overline{X}_1 - X'_{11}}{\overline{X}_2 - X'_{21}}\right)\right|\right] \text{ is finite for all } c > 1.$$

A "smooth" version of δ^* can be obtained by replacing the multiplier of $(\overline{X}_1 - X'_{11})/(\overline{X}_2 - X'_{21})$ by a more refined function. The resulting estimator has

$$c_0 = 1 - 3n_2^{-1},$$

$$c_1 = n_2^{-1}(n_1 + 1)^{-1}n_1(n_2 - 3) \qquad \text{for } n_2 > 3 \quad \text{and } W\left(\frac{\delta^*}{\delta}\right) = 1 - \left(\frac{\delta^*}{\delta}\right)^2; \tag{19.86a}$$

$$c_0 = n_2^{-1}(n_1 - 1)^{-1}n_1(n_2 - 2),$$

$$c_1 = 1 - 2n_2^{-1} \qquad \text{for } W\left(\frac{\delta^*}{\delta}\right) = \left(\frac{\delta^*}{\delta}\right) - \log\left(\frac{\delta^*}{\delta}\right) - 1. \tag{19.86b}$$

Effects of Outliers

The effects of outliers on the estimation of a in the one-parameter case (19.28) has received considerable attention in the literature since the early 1970s. Among the first studies were those of Kale and Sinha (1971) and Joshi (1972). They initially studied situations where $n - 1$ independent random variables each had distribution (19.28), while one further random variable (the "outlier") has a pdf of the same form, but with a replaced by σ/α.

For situations where the identity of the outlier is not known (and the probability that X_j is the outlier is n^{-1} for $j = 1, \ldots, n$), Kale and Sinha (1971) suggest consideration of the class of estimators (for σ):

$$\tilde{\sigma}_k = \frac{T_k}{k + 1}, \tag{19.87}$$

where T_k is defined as in (19.73). Joshi (1972) tabulates optimum values for k, minimizing the mean square error. If $k = n$ [with $\tilde{\sigma}_n = \sum_{i=1}^{n} X_i/(n + 1)$], we have

$$\text{MSE}(\tilde{\sigma}_n/\sigma) = (n + 1)^{-1} + 2(\alpha^{-1} - 1)^2(n + 1)^{-2}, \tag{19.88}$$

where σ/α is the mean of the outlier.

Generally, the optimum k has to be found numerically. Joshi (1972) found that for $0.55 \le a \le 1$ the optimal value of k is n. Later [Joshi (1988)] he found that the optimal value of k is fairly stable over wide ranges of values of a and suggested, as a rule of thumb,

Take $k = n$ for $0.5 \le a < 1.0$
Take $k = n - 1$ for $0.25 \le a < 0.5$
Take $k = n - 2$ for $0.05 \le \alpha \le 0.25$.

Chikkagoudar and Kunchur (1980) suggest the estimator

$$V = \sum_{j=1}^{n} \left(\frac{1}{n} - \frac{2j}{n^2(n+1)} \right) X'_j \tag{19.89}$$

Comparison between $\tilde{\sigma}_k$ and V shows that neither estimator dominates the other (in terms of mean square error loss). [T_k (for optimal k) is superior for values of a near 0 or 1.] Balakrishnan and Barnett (1994) have recently proposed some generalized estimators of this form and discussed their properties.

From (1991) has studied robust estimators that are general linear functions of order statistics, aiming to obtain optimal (or near-optimal) values for the coefficients c_j in $\sum_{j=1}^{n} c_j X'_j$, for various values of α and n. Optimal choice calls for rather complicated calculations, so From advocates use of simplified estimators with very nearly optimal mean square error, which is always less than those of $\tilde{\sigma}_k$ and V. This estimator uses

$$c_j = \begin{cases} d_1 & \text{for } 1 \le j \le m, \\ d_2 & \text{for } m+1 \le j \le n, \end{cases}$$

with $d_1 > d_2$ and appropriate integer m. Tables are provided giving optimal values of (d_1, d_2, m) for n = 2(1)15(5)30(10)50 and a = 0.05, 0.15.

If the value of a is known, Joshi (1972) suggests using the estimator

$$(n - 1 + \alpha^{-1})\tilde{\sigma}_{n-1}/n. \tag{19.90}$$

He also suggests estimating a (if it is unknown) by an iterative procedure, solving the equation

$$n\tilde{\sigma}_n = (n - 1 + \alpha^{-1})\tilde{\sigma}_{n-1} \tag{19.91}$$

to estimate a, and then using optimal k to get a new estimator $\tilde{\sigma}_n$ of a, and so on; see also Jevanand and Nair (1993).

Through a systematic study of order statistics from independent and non-identically distributed exponential random variables, Balakrishnan (1994) has established several recurrence relations for the single and the product moments of order statistics. These results will enable the computation of all the single and product moments of order statistics arising from a sample containing p outliers (recursively in p by starting with p = 0 or the *i.i.d.* case). Through this recursive computational process, Balakrishnan (1994) has extended the work of *Joshi* (1972) to the p-outlier case and determined the optimal trimmed and Winsorized estimators. The robustness properties of various linear estimators, including the Chikkagoudar–Kunchur estimator in **(19.89)**, have also been examined by Balakrishnan **(1994)**.

Veale (1981) has investigated cases where the identity of the outlier is known. This of course leads to much simpler analysis.

Returning to uncontaminated (no-outliers) data, the estimation of a in the standard (one-parameter) case **(19.28)**, subject to the condition that a is no less than a,, was studied by Gupta and Basu (1980). [Related cases $\sigma <$ a,, or a in (a,, a,) can be analyzed similarly by using appropriate transformations.]

Natural estimators of a (given $\sigma > \sigma_0$) are

$$6 = \max(\bar{X}, \sigma_0) \qquad \text{(the MLE)} \tag{19.92}$$

or

$$\sigma^* = n^{-1} \sum_{j=1}^{n} \max(X_j, \sigma_0). \tag{19.93}$$

Numerical studies show that the mean square error **(MSE)** of a * is less than that of 6 for small n, when σ_0/σ is small. Indeed, even for n = 30, a * has the smaller MSE if $\sigma_0/\sigma \leq 0.3$.

Estimation of the probability P = $\Pr[Y < X]$, where X and Y are independent exponential variables, has received prominent attention in the literature. The common interpretations of this probability is a measure of the reliability or performance of an item of strength Y subject to a stress X, or probability that one component fails prior to another component of some device.

Tong (1974, **1975)** obtained the uniformly minimum variance unbiased estimator (UMVUE) of P when X and Y are independent one-parameter exponential (19.28) variables. Kelley, Kelley, and Schucany (1976) derived the variance of the UMVUE of P. Beg (1980) obtained the UMVUE of P when X and Y have two-parameter exponential (19.1) distributions with unequal scale and location parameters. Gupta and Gupta (1988) obtained the MLE, the UMVUE, and a Bayesian estimator of P when the location parameters are unequal but there is a common scale parameter. Bartoszewicz (1977) tackled the problem in the exponential case for different types of censoring. Reiser, Faraggi, and Guttman (1993) discuss the choice of sample sizes for the experiments dealing with inference on $\Pr[Y < X\}$ in an acceptance sampling theory framework with exponential variables. Most recently Bai and Hong *(1992)* revisited the problem, obtaining the UMWE of P with unequal sample sizes when X and Y are independent two-parameter exponential random variables with an unknown common location parameter. Kocherlakota and Balakrishnan *(1986)* have discussed one-sided as well as two-sided acceptance sampling plans based on Type-I1 censored samples.

If X and Y are independent one-parameter exponential *(19.28)* random variables with hazard rates (σ^{-1}) A and μ, respectively, then

$$P = \Pr[Y < X] = \mu(\lambda + \mu)^{-1}. \tag{19.94}$$

Given two independent random samples $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$, the MVUE of P is

$$P^* = \begin{cases} \displaystyle\sum_{j=0}^{m-1} (-1)^j \frac{(m-1)^{(j)}}{n^{[j]}} \left(\frac{n\bar{Y}}{m\bar{X}} \right)^j & \text{if } n\bar{Y} \le m\bar{X}, \\ \displaystyle 1 - \sum_{j=0}^{n-1} (-1)^n \frac{(n-1)^{(j)}}{m^{[j]}} \left(\frac{m\bar{X}}{n\bar{Y}} \right)^j & \text{if } n\bar{Y} > m\bar{X}. \end{cases} \tag{19.95}$$

[Tong (1974)] where $a^{(b)} = a(a-1)\cdots(a-b+1)$, $a^{[b]} = a(a+1)\cdots(a+b-1)$.

If μ is known so that there is no need to sample for Y values, the MVUE is

$$P^{**} = \sum_{j=0}^{m-1} (-1)^j (m-1)^{(j)} (m\mu\bar{X})^{-j} + (-1)^m (m\mu\bar{X})^{-m+1} \exp(-m\mu\bar{X}). \tag{19.96}$$

If there is Type II censoring, with only the first g order statistics $X_1', \dots, X_g'$ of X and the first h order statistics $Y_1', \dots, Y_h'$ of Y being available, the UMVUE of P is

$$\hat{P}^* = \begin{cases} \displaystyle\sum_{j=0}^{g-1} (-1)^j \frac{(g-1)^{(j)}}{h^{[j]}} \left(\frac{H}{G} \right)^j & \text{if } H \le G, \\ \displaystyle 1 - \sum_{j=0}^{h-1} (-1)^j \frac{(h-1)^{(j)}}{g^{[j]}} \left(\frac{G}{H} \right)^j & \text{if } H > G, \end{cases} \tag{19.97}$$

where $G = \sum_{j=1}^{g} X_j' + (m-g)X_g'$, $H = \sum_{j=1}^{h} Y_j' + (n-h)Y_h'$. For $g = m$ and $h = n$, *(19.97)* reduces to *(19.95)*. If each failed item is replaced immediately by a new item having the same lifetime distribution, the UMVUE of P is obtained from (19.97) by replacing G by mX'_g and H by nY'_h.

The maximum likelihood estimator of P (given complete sample values) is

$$\hat{P} = \frac{\bar{X}}{\bar{X} + \bar{Y}}. \tag{19.98}$$

If $m = n$ the MLE is unbiased ($E[\hat{P}] = \mu(\lambda + \mu)^{-1}$), and hence it is also the UMVUE of P, as noted by Chious and Cohen (1984). Expressions for the UMVUE of P for the general case of two-parameter exponential random variables are complex, even in the case of common (unknown) location parameter.

However, the MLE of P is given by

$$\hat{P} = \frac{\bar{X} - T}{\bar{X} + \bar{Y} - 2T}, \tag{19.99}$$

where $T = \min(X, Y) = \min(X'_1, Y'_1)$ is the maximum likelihood estimator of the common location parameter. As $m, n \to \infty$, with $m/n \to \gamma$, asymptotically

$$\sqrt{n}\,(\hat{P} - P) \xrightarrow{d} N(0, \sigma^2), \tag{19.100}$$

where $\sigma^2 = P^2(1 - P)^2\{\gamma(1 - \gamma)\}$ and the $\hat{P}$ and the UMVUE are asymptotically equivalent in this case.

In view of (19.94), the problem of choice of sample sizes for estimating P with specified accuracy is equivalent to the solution to the sample size problem provided by Reiser, Faraggi, and Guttman (1993) for the ratio of two exponential scale parameters. Given P_1, α, P_2, β with $0 < P_2 < P_1 < 0$, $0 < \alpha < 1$, $0 < \beta < 1$, it is required to find an acceptance rule of the form $P > P_c$, with m and n such that (1) if $P = P_1$, the probability of acceptance is $1 - \alpha$, and (2) if $P = P_2$, the probability of acceptance is β. In this case m and n must satisfy

$$\frac{F_{2n, 2m; \alpha}}{F_{2n, 2m; (1-\beta)}} = \frac{P_2}{P_1}, \tag{19.101}$$

where $F_{2n, 2m; \alpha}$ is the α percentage point of the $F_{2n, 2m}$ distribution.

When a random sample of size n has been censored by omission of the greatest $n - k$ observed values, it may be of interest to estimate (predict) the value of a specific one of the omitted values X'_r, say $(r > k)$. Lawless (1971) used the pivotal variable

$$G_1 = \frac{X'_r - X'_k}{T_k} \qquad [T_k \text{ as defined in (19.73)}] \tag{19.102}$$

to obtain the upper $100(1 - \alpha)\%$ prediction limit (UPL),

$$\text{UPL}(X'_r) = X'_k + g_{1,1-\alpha}T_k \tag{19.103}$$

where $\Pr[G_1 \le g_{1,1-\alpha}] = 1 - \alpha$. The cdf of G_1 is

$$\Pr[G_1 \le g] = \frac{k}{B(r-k, n-r+1)} \times \sum_{i=1}^{r-k-1} \binom{r-k-1}{i}(-1)^i(n-r+i+1)^{-1} \times \{1 + (n-r+i+1)g\}^{-k}. \tag{19.104}$$

Lingappaiah (1973) suggested using

$$G_2 = \frac{X'_r - X'_k}{X'_k} \tag{19.105}$$

in place of G_1, noting that these would lead to simpler calculations and that X'_k contains "all the prior information" about X'_r. However, Kaminsky (1977) pointed out that this is only true if σ is known. He also noted that the probability that the UPL based on G_1 exceeds that based on G_2 is close to (but below) 0.5 in a number of representative cases. Furthermore he showed that the asymptotic probability (as $n \to \infty$) that the length of predictor interval based on G_2 exceeds that based on G_1 is zero.

For the two-parameter distribution (19.1) similar procedures can be followed, noting that the variables $X'_i - X'_1$ are distributed as order statistics from a random sample of size $n - 1$ from the distribution (19.28). Likeš (1974) extends Lawless's (1971) treatment to this case, using the statistic

$$G'_1 = \frac{X'_r - X'_k}{T'_k} \tag{19.106}$$

with $T'_k = \sum_{i=2}^{k} X'_i + (n-k)X'_k - (n-1)X'_1$. Tables of quantiles of G'_1 were constructed by Likeš and Nedělka (1973). Further results on these lines were obtained by Lawless (1977), and later incorporated in his important textbook,