Prediction of $s$th Ordered Observation for the Two-Parameter Exponential Distribution

J. Likeš
Praha, Czechoslovakia
Prague School of Economics

This paper deals with obtaining a prediction interval on a future observation $X_s$, in an ordered sample of size $n$ from a two-parameter exponential distribution for the situation where some or all the first $r$ observations $X_1 < X_2 < \cdots < X_r$, $1 \leq r < s \leq n$, have been observed. The intervals are based on the statistic $Z = (X_s - X_r) / S_r$, where $S_r$ is a function of the observations $X_1 = A < X_2 < \cdots < X_r$, such that $X_r - X_s$, and $S_r$ are independent variables and $2S_r/\sigma$ has the distribution $\chi^2(2\nu)$. The expressions for the quantiles $z_p$ are given and some problems of numerical determination of $z_p$'s are discussed. The results can be also applied to related distributions.

1. INTRODUCTION

Consider an ordered random sample $X_1 < X_2 < \cdots < X_n$ of size $n$ from the two-parameter exponential distribution

$$f(x; A, \sigma) = \frac{1}{\sigma} \exp \left[-\frac{(x - A)}{\sigma}\right], \quad x > A, \quad \sigma > 0. \quad (1)$$

We shall find a prediction interval on $s$th ordered observation $X_s$, for the situation where some or all first $r$ observations $X_1 < \cdots < X_r$, $1 < r < s < n$, in the same sample are available. This is a contrast to the frequently encountered situation of obtaining a prediction interval on a future and different sample.

For example, if in life testing $n$ items are put on test simultaneously and if the first $r$ failure times $X_1 < \cdots < X_r$ are observed, we wish to predict the $s$th failure time, $r < s \leq n$, assuming that the time to failure follows a two-parameter exponential distribution.

J. F. Lawless [2] solves the problem of predicting $X_s$ in the case of a one-parameter exponential distribution (i.e. distribution (1) with $A = 0$) and assumes that each of first $r$ values were observed. Hence the present paper can be thought of as extensions of paper [2]. In addition, the present paper indicates the calculations for a prediction interval on the $n$th (i.e. last ordered) observation and extends the results to some other distributions.

2. PREDICTION INTERVALS FOR $X_s$

Consider for given $r$ and $s$, $1 \leq r < s \leq n$, the statistic

$$Z = Z(r, s; n, \nu) = (X_s - X_r) / S_r, \quad (2)$$

where $S_r$ is a function of $X_1 = A, X_1, \cdots, X_r$, such that $X_s - X_r$, and $S_r$ are independent variables and $2S_r/\sigma$ has the distribution $\chi^2(2\nu)$.

Let $z_p = z_p(r, s; n, \nu)$ be the $P$th quantile of the statistic (2). Then

$$\Pr (Z < z_{1-\alpha}) = \Pr (X_s < X_r + z_{1-\alpha}S_r) = 1 - \alpha$$

and from this it follows that

$$X_s < X_r < X_r + z_{1-\alpha}S_r, \quad (3)$$

is a one-sided 100(1 - $\alpha$)% prediction interval for the $s$th ordered observation $X_s$ based on the first $r$ observations $X_1, X_2, \cdots, X_r$.

Note here that in the life testing context $z_{1-\alpha}S_r$ provides an upper bound on the elapsed time from the last observed failure to the $s$th observed failure.

We may also consider a second one-sided 100(1 - $\alpha$)% prediction interval

$$X_r > X_s + z_{1-\alpha}S_r, \quad (4)$$

or a two-sided 100(1 - $\alpha$)% prediction interval

$$X_r + z_{\alpha_2}S_r < X_s < X_r + z_{1-\alpha}S_r, \quad (5)$$

for $X_s$, where $0 < \alpha_1 < \alpha$ and $\alpha_2 = \alpha - \alpha_1$.

3. RELATIONS FOR $z_p$

Let

$$Y = Y(r, s; n, \nu) = (X_s - X_r) / (\nu S_r),$$

$$0 \leq r < s \leq n. \quad (6)$$

Received Aug. 1972; revised April 1973.
where \( X_s - X_r \) and \( S \) are defined as in Section 2. Then the probability \( \Pr(Y > y) \) equals (see relations (10) in [3])

\[
\Pr(Y > y) = \frac{1}{B(n - s + 1, s - r)} \cdot \sum_{i=1}^{s-1} \left( 
\begin{array}{c}
\frac{s - r - 1}{n - i + 1} \\
\end{array}
\right)
\cdot [1 + (n - i + 1)y]^{-s}, \quad y \geq 0. \tag{7}
\]

Consider now the following

**Lemma:** For the distribution (1), the variable \( X_s - X_r \) in a sample of size \( n \) has the same distribution as the variable \( X_{s+a} - X_{r+a} \) in a sample of size \( n + a \), where \( a \) is an integer such that \(-r \leq a \leq n - s\).

The proof is very simple: The variable \( X_s - X_r \) can be expressed as

\[
X_s - X_r = \frac{1}{2} \sum_{i=1}^{n} \frac{1}{n - i + 1} X_i^2
\]

where \( X_i^2 = 2(n - i + 1)(X_i - X_{i-1})/\sigma \), \( i = 1, 2, \ldots, n \) and \( X_i^{(2)} = X_i^{(2)} \).

Since the variables \( X_i^2, \ldots, X_n^2 \) are [1] independent all having the distribution \( \chi^2(2) \), then the variable \( X_s - X_r \) in a sample of size \( n \) has the same distribution as the last expression in (8); but this last expression represents the variable \( X_{s+a} - X_{r+a} \) in a sample of size \( n + a \).

Further, since the variables \( X_s - X_r \) and \( S \), in (6) and (2) are independent, then the statistic \( Y(r; s; n, \nu) \) has the same distribution as the statistic \( Y(\nu; s + a; n + a, \nu) \) and the statistic \( Z(r; s; n, \nu) \) has the same distribution as the statistic \( Z(\nu; s + a; n + a, \nu) \), \(-r \leq a \leq n - s\). From this follow the relations for the quantiles \( z_p \):

\[
z_p(r, s; n, \nu) = z_p(r + a, s + a; n + a, \nu) \tag{9}
\]

for \(-r \leq a \leq n - s\) and for every \( P, 0 < P < 1 \) and \( \nu \).

Specifically, for \( a = -r \) we obtain

\[
z_p(r, s; n, \nu) = z_p(0, s - r; n - r, \nu) \tag{10}
\]

whence only the determination of the quantiles \( z_{p}(0, k; N, \nu) \) of the variables \( Z(0, k; N, \nu) = (X_k - A)/S \) is required for given \( k, N \) \((k = s - r, N = n - r)\), \( 1 \leq k \leq N \) and \( \nu \) is sufficient.

Putting \( j = k - i \) or \( l = j + 1 \), we obtain from (7) the relation

\[
\frac{1}{B(N - k + 1, k)} \sum_{i=0}^{k-1} (-1)^{k-1-j} \left( \begin{array}{c} k - 1 \\ j \end{array} \right) \frac{1}{N - k + j + 1} [1 + (N - k + j + 1)z_{p}]^{-s} = 1 - P \tag{11}
\]

for \( z_{p} = z_{p}(0, k; N, \nu) \).

Specifically, for the next observation, i.e. \( k = 1 \), one obtains the simple result

\[
z_{p}(0, 1; N, \nu) = \frac{\nu}{N} [1 - (1 - P)^{-1/c} - 1] = \frac{1}{N} F_{p}(2, 2\nu) \tag{12}
\]

where \( F_{p}(v_1, v_2) \) denotes the \( P \)th quantile of the \( F \)-distribution \( F(v_1, v_2) \).

Consider now the case \( s = n \), which is the one of most frequent interest. In [4] \( P \)th quantiles \( q_{p}(n, \nu) \) of the standardized range \( Q = Q(n, \nu) = (X_s - X_1)/S \) are tabulated for \( P = 0.90, 0.95, 0.99, n = 1(1)20, \nu = 1(1)20, 24, 30, 40, 60, 120, \infty \). Here \( X_s \) and \( X_s \) is the smallest and the largest observation respectively in a sample of size \( n \) from the distribution (1) and \( S \) is a statistic independent of \( X_s - X_1 \) such that \( 2\nu S/\sigma \) has the distribution \( \chi^2(2\nu) \).

From the above Lemma it follows that the statistic \( Z(r, s; n, \nu) = (X_s - X_r)/S \) has the same distribution as the statistic \( Z(\nu; r + a; n + a, \nu) \), \(-r \leq a \leq n - s\), i.e. the same as the statistic \( Q(\nu; n - r + 1; n - r + 1, \nu) \), \(-r \leq a \leq n - s\). Hence

\[
z_{p}(r, s; n, \nu) = q_{p}(n - r + 1, \nu) \tag{13}
\]

When the parameter \( \sigma \) is known, we consider the variable

\[
Z = Z(r, s; n, \infty) = (X_s - X_r)/\sigma,
\]

\[
0 \leq r < s \leq n, \tag{14}
\]

which has the same distribution as the variable \( Z(\nu; r + a; n + a, \infty) \), \(-r \leq a \leq n - s\). The \( P \)th quantile \( z_{p}(0, k; N, \nu) \) is given by the relation (following from (11) for \( \nu \rightarrow \infty \))

\[
\frac{1}{B(N - k + 1, k)} \sum_{i=0}^{k-1} (-1)^{k-1-j} \left( \begin{array}{c} k - 1 \\ j \end{array} \right) \frac{1}{N - k + j + 1} \exp[-(N - k + j + 1)z_{p}] = 1 - P. \tag{15}
\]

\( z_{p}(0, k; N, \nu) \) can also be obtained from the tables of the incomplete beta function by means of relations

\[
I_{c}(N - k + 1, k) = 1 - I_{1-c}(k, N - k + 1)
\]

\[
= 1 - P, \quad z_{p}(0, k; N, \nu) = -\ln c \tag{15'}
\]

or by the relation

\[
\text{TECHNOMETRICS®, VOL. 16, NO. 2, MAY 1974}
\]

J. LIKES
TWO-PARAMETER EXPONENTIAL DISTRIBUTION

\[ z_p(0, k; N, \infty) = \ln \left[ 1 + \frac{k}{N - k + 1} F_p(2k, 2(N - k + 1)) \right] \]

Note yet that for \( k = 1 \) we obtain from (15)

\[ z_p(0, 1; N, \infty) = -\frac{1}{N} \ln (1 - P) = \frac{1}{N} x_p^2(2). \]

For \( \sigma \) known, 100(1 - \( \alpha \))% prediction intervals for \( X \), based on \( X_r \), 0 \( \leq \) \( r \) \( \leq \) \( s \) \( \leq \) \( n \), are given by (3)-(5) with \( S_r \) and with \( z_p = z_p(0, s - r; n - r, \infty) \) for \( P = 1 - \alpha \), \( P = \alpha \), and \( P = \alpha_1, 1 - \alpha \) respectively.

4. CHOICE OF \( S_r \)

From the properties of the variables \( x_i^2 = 2(n - i + 1)(X_i - X_{i-1})/\sigma^2 \), \( i = 1, 2, \ldots, n \), it follows that the random variable \( 2vS_r/\sigma = \sum_{i=1}^{r-1} c_ix_i^2 \) with \( c_i = 0 \) or \( c_i = 1, i = 1, 2, \ldots, r \), \( \sum_{i=1}^{r-1} c_i \geq 1 \), has the distribution \( x^2(2\sigma) \) with \( r - 1 \) degrees of freedom and \( X_r - X_1 = (\sigma/2) \sum_{i=1}^{r-1} c_i x_i \) with \( c_i \)'s, fulfill the assumptions for the statistic \( S_r \) in (2) with \( \nu = \sum_{i=1}^{r-1} c_i \).

J. Lawless [2] considers the case \( A = 0 \) and \( Y(r, s; n, r) = (X_r - X_1)/r\sigma \), where \( rS_r = c_i \sum_{i=1}^{r-1} X_i + (n - r)X_1 \), i.e. \( rS_r \) is the statistic (17) with \( c_i = c_2 = \cdots = c_r = 1 \) (note that in [2] \( Y \) = \( \mu \), \( r = k \) and \( s = r \)).

For \( A \) unknown we may consider in (2) \( S_{r-1} = \sum_{i=1}^{r-1} (n - i + 1)(X_i - X_{i-1})/(r - 1) = \sum_{i=1}^{r-1} X_i + (n - r)X_1 - (n - 1)X_0/(r - 1) \). This assumes that all \( r \) observations \( X_1, X_2, \ldots, X_r \) are available. When some of these observations are missing, \( S_r \) can be easily found from the remaining observations by means of (17). For example, when \( X_0, X_1, \ldots, X_5 \), 0 \( \leq \) \( b \) \( \leq \) \( r \), are missing, we have \( S_{r-b} = \sum_{i=b+1}^{r-1} (n - i + 1)(X_i - X_{i-1})/(r - b) = \sum_{i=1}^{r-1} X_i + (n - r)X_1 - (n - b)X_0/(r - b) \).

Similarly, if \( r \geq 5 \) and \( X_2, X_3 \) are missing (i.e. \( A \) is unknown and \( X_0 \) is missing), we consider \( S_{r-2} = \sum_{i=b+2}^{r-1} (n - i + 1)(X_i - X_{i-1})/(r - 3) = \sum_{i=1}^{r-1} X_i + (n - r)X_1 + (n - 4)X_0/(r - 3) \) and so on.

Example. Suppose that \( n = 8 \) items are put on test simultaneously and that the first \( r = 4 \) items have the lifetimes 62, 84, 106 and 144 hours. Let the lifetimes of all \( n \) items be distributed according to the exponential distribution (1) with the same parameters \( A \) and \( \sigma \).

We wish to find a 95% prediction interval of the type (3) for \( X_8 \). In this case \( S_3 = (\sum_{i=1}^{3} X_i + 4X_4 - 7X_3)/3 \) and \( z_p(4, 8; 8, 3) = q_p(5, 3) \).

From Table 1b in [4] we find \( q_{0.95}(3, 3) = 8.879 \). Since \( S_3 = 476/3 \), we obtain the 95% prediction interval \( 144 < X_8 < 144 + (8.879) 476/3 \), i.e. \( 144 < X_8 < 1408.8 \). We can be 95% confident that the total elapsed time will not exceed 1409 hours.

5. APPLICATIONS TO SOME RELATED DISTRIBUTIONS

The above results can be also applied to the prediction of the \( s \)th ordered observation in samples from other distributions:

(i) Let \( T \) be a continuous random variable such that the variable \( X = \phi(T) \) has the distribution (1); where \( \phi(T) \) is a strictly increasing function of \( T \). Let \( T_1 < T_2 < \cdots < T_r \) be an ordered random sample of size \( n \) from such a distribution. Then

\[ Z = Z(r, s; n, v) = [\phi(T_r) - \phi(T_s)]/S, \]

with

\[ S = \sum_{i=1}^{r-1} c_i(n - i + 1)[(X_i - X_{i-1})]/v, \]

\[ v = \sum_{i=1}^{r-1} c_i, \]

\( c_i = 0 \) or \( c_i = 1, i = 1, 2, \ldots, r, r = \sum_{i=1}^{r-1} c_i \geq 1 \), has the same distribution as the variable (2). Hence

\[ \phi^{-1}[\phi(T_r) + z_{1-\alpha}S] < \phi(T_s) < \phi^{-1}[\phi(T_r) + z_{1-\alpha}S] \]

is a 100(1 - \( \alpha \))% prediction interval for \( \phi(T) \) (for \( v = \infty \), we set \( S = \sigma \) in (19) and (20)). From this we obtain a 100(1 - \( \alpha \))% prediction interval for \( T_r \)

\[ \phi^{-1}[\phi(T_r) + z_{1-\alpha}S] < T_r < \phi^{-1}[\phi(T_r) + z_{1-\alpha}S] \]

where \( \phi^{-1}(X) \) is the inverse function to the function \( \phi(T) \).

For example, for the Rayleigh distribution

\[ h(t) = (t/\theta) \exp (-t^2/2\theta), \theta > 0, \]

we have \( \phi(T) = T^2/2 \) and \( \phi^{-1}(X) = \sqrt{2X} \). Similarly, for the Pareto distribution

\[ h(t) = \theta^\beta t^{\beta-1} \]

we have \( \phi(T) = \ln T \) and \( \phi^{-1}(X) = \exp(X) \).
nential distribution, i.e. the distribution (1) with \( A = 0, \sigma = 1 \).

Let \( T_1 < T_2 < \cdots < T_n \) be an ordered random sample of size \( n \) from such a distribution. Then a 100(1 - \( \alpha \))% prediction interval for \( -\ln [1 - H(T_i)] \) is given by (20) with \( \varphi(T_i) = -\ln [1 - H(T_i)] \), \( i = 1, 2, \ldots, r, s \). From here it follows that

\[
1 - [1 - H(T_i)] \exp (-z_{1-\alpha} S_i) < H(T_i) < 1 - [1 - H(T_i)] \exp (-z_{1-\alpha} S_i)
\]

is a 100(1 - \( \alpha \))% prediction interval for \( H(T_i) \) (for \( \nu = \infty \), we set \( S_i = 1 \) in (22)).

6. Acknowledgment

I should like to thank the referees for suggesting improvements in the presentation of this paper.

References


TECHNOMETRICS©, VOL. 16, NO. 2, MAY 1974