Inference for $P(Y$

Vee Ming Ng

*Murdoch University, Australia, VNg@murdoch.edu.au*

Follow this and additional works at: [http://digitalcommons.wayne.edu/jmasm](http://digitalcommons.wayne.edu/jmasm)

Part of the [Applied Statistics Commons](https://digitalcommons.wayne.edu/jmasm), [Social and Behavioral Sciences Commons](https://digitalcommons.wayne.edu/jmasm), and the [Statistical Theory Commons](https://digitalcommons.wayne.edu/jmasm)

Recommended Citation


DOI: 10.22237/jmasm/1162353840

Available at: [http://digitalcommons.wayne.edu/jmasm/vol5/iss2/5](http://digitalcommons.wayne.edu/jmasm/vol5/iss2/5)

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in *Journal of Modern Applied Statistical Methods* by an authorized editor of DigitalCommons@WayneState.
Inference for $P(Y < X)$ for Exponential and Related Distributions

Vee Ming Ng
School of Engineering Sciences,
Murdoch University, Australia

Some tests and confidence bounds for the reliability parameter $R = P(Y < X)$ are proposed, where $X$ and $Y$ are independent random variables from a two-parameter exponential distribution. The results are based on missing or incomplete data and are applicable to some related distributions.

Key words - Confidence bounds, exponential distribution, missing data, P-value, reliability parameter

Introduction

The problem of estimating and testing the reliability parameter $R = P(Y < X)$ has been widely researched in the literature. The problem originated in the context of reliability of a component of strength $X$ subjected to a stress $Y$, the component failing if and only if at any time the applied stress is greater than its strength. Other applications for the reliability parameter exists when $X$ and $Y$ have different interpretation, such as when $Y$ is the response for a control group and $X$ is the response for the treatment group. Inference on $R$ shall be considered when $X$ and $Y$ are random variables from a two-parameter exponential distribution. Inference on $R$ for the one-parameter exponential distribution can be found in Enis and Geisser (1971), Tong (1977), and Chao (1982) among others.

Gupta and Gupta (1988) derived and compared some point estimators for $R$ in the case of two independent exponential variables having a common scale parameter. For the case in which the location parameter is common, Bai and Hong (1992) discussed point and interval estimation of $R$ and Baklizi (2003) compared the performance of several types of asymptotic, approximate, and bootstraps confidence intervals. Ali, Woo, and Pal (2004) considered test and estimation of $R$ when the scale parameters are equal and known and also inference procedures for $R$ which are based on likelihood ratio tests for equality of scale and equality of location parameters.

This article considers some tests and confidence bounds for $P(Y < X)$ for the two-parameter exponential distribution with a common but unknown scale parameter and also with a common but unknown location parameter. Exact tests and confidence bounds are derived in situations where data may be missing or incomplete, the situation with complete data being a special case. These results are extended to some related distributions.

Methodology and Results

A two-parameter exponential distribution with parameters $(\mu, \sigma)$ is defined by the probability density function:

$$f(x; \mu, \sigma) = \frac{1}{\sigma} e^{-(x-\mu)/\sigma}, \quad x > \mu, \quad \sigma > 0$$

Suppose $X$ and $Y$ are independent exponential random variables with parameters $(\mu_x, \sigma_x)$ and $(\mu_y, \sigma_y)$ and probability density...
functions \( f(x; \mu_x, \sigma_x) \) and \( f(y; \mu_y, \sigma_y) \) respectively. Then

\[
R = P(Y < X) = \int_{\mu_y}^{\mu_x} \int_{\sigma_y}^{\sigma_x} f(x; \mu_x, \sigma_x) f(y; \mu_y, \sigma_y) \, dy \, dx
\]

\[
= \left\{ \frac{\sigma_x}{\sigma_x + \sigma_y} e^{-\delta/\sigma_x}, \delta \geq 0 \right\}
\]

\[
= \left\{ 1 - \frac{\sigma_y}{\sigma_x + \sigma_y} e^{-\delta/\sigma_y}, \delta < 0 \right\}
\]

where \( \delta = \mu_x - \mu_y \). Inference on \( R \) is considered for two cases: (a) scale parameters are equal and unknown and (b) location parameters are equal and unknown.

Assuming two independent samples of size \( n \) and \( m \) from the exponential distributions with parameters \((\mu_x, \sigma_x)\) and \((\mu_y, \sigma_y)\) respectively, let \( X_q < X_{q+1} < \ldots < X_r \) and \( Y_l < Y_{l+1} < \ldots < Y_p \) denote the ordered observations; some of these could be missing where \( q = 1, r = n \), and \( l = 1, p = m \) would correspond to all observations being available.

Let \( S_x = \sum_{i=q+1}^{r} c_i (n-i+1)(X_i - X_{i-1}) \) \( c_i = 1 \)

or \( 0; \)

\( S_y = \sum_{j=l+1}^{p} d_j (m-j+1)(Y_j - Y_{j-1}) \)

\( d_j = 1 \) or \( 0 \), and \( S_p = S_x + S_y \), \( v_x = \sum_{i=q+1}^{r} c_i \), \( v_y = \sum_{j=l+1}^{p} d_j \), \( v = v_x + v_y \). It is well known that:

- \( X_q, Y_l, S_x, S_y, S_p \) are statistically independent (see Tanis (1964), Likes (1974)).
- \( 2S_x / \sigma_x, 2S_y / \sigma_y, 2S_p / \sigma \) when \( \sigma_x = \sigma_y = \sigma \), have chi-square distributions with \( 2v_x, 2v_y, 2v \) degrees of freedom respectively.

- The probability density functions of the ordered statistics \( X_q \) and \( Y_l \) can be written, respectively, as

\[
f(x; \mu_x, \sigma_x, n, q) = \sum_{i=0}^{q-1} a(n, q, i) \frac{1}{\sigma_x} e^{-n(q + x - \mu)/\sigma_x} \]

\[
f(y; \mu_y, \sigma_y, m, l) = \sum_{j=0}^{l-1} b(m, l, j) \frac{1}{\sigma_y} e^{-m(l + y - \mu)/\sigma_y} \]

where

\[
a(n, q, i) = q \binom{n}{q} \binom{q-1}{i} (-1)^i \]

\[
b(m, l, j) = l \binom{m}{l} \binom{l-1}{j} (-1)^j \]

\[
n(q, i) = n - q + i + 1 \]

\[
m(l, j) = m - l + j + 1 \]

Test of hypothesis when \( \sigma_x = \sigma_y = \sigma \)

Suppose that \( \sigma_x = \sigma_y = \sigma \) but \( \sigma \) is unknown, then

\[
R = \begin{cases} \frac{1}{2} e^{-\lambda}, \lambda \geq 0 \\ 1 - \frac{1}{2} e^{-\lambda}, \lambda < 0 \end{cases}
\]

where

\[
\lambda = (\mu_y - \mu_x) / \sigma .
\]

A test procedure is now derived for testing hypotheses about the reliability parameter \( R \); a similar procedure is considered in Ranganathan and Kale (1979) for a 1-sample reliability problem. Because \( P(X < Y) = 1 - R \), it suffices to consider the problem of testing the null hypothesis \( H_0: \frac{1}{2} e^{-\lambda} \geq p_0 \), against the
alternative $H_1: \frac{1}{2} e^{-\lambda} < p_0$, $p_0$ being a specified value less than 0.5. As these hypotheses are equivalent to $H_0: \lambda \geq -\ln(2p_0)$ against $H_1: \lambda < -\ln(2p_0)$, consider the test statistic $T = (Y_i - X_q) / S_p$. $T$ is a maximal invariant and its distribution depends only on $\lambda$. A large value of $T$ would be evidence against $H_0$. Hence, for an observed value $t$ of $T$, $P(T > t)$ for $t \geq 0$, $\lambda \geq 0$ is the P-value of the test, a small value of which would indicate sufficient evidence against $H_0$. In order to get an expression for the P-value, one must first obtain, from the joint density function of $p_{lq}, Y, X$, the joint probability density function of $q_{XYD} - = p_{S}, \lambda \geq 0$, which then yields the joint density function $p_{SDT} / \lambda = p_{SW} / \lambda$:

$$f(d, s; \delta, \sigma) = \begin{cases} \sum_{i=0}^{q-1} \sum_{j=0}^{l-1} \frac{a(n, q, i)b(m, l, j)e^{m(l, j)\delta}e^{-n(m, j)e^{l(i+1)}\delta}}{\Gamma(n(m, i)+m(l, j))\sigma^{m+1}} & \text{if } d \geq s, s \geq 0 \\ \sum_{i=0}^{q-1} \sum_{j=0}^{l-1} \frac{a(n, q, i)b(m, l, j)e^{-n(m, j)e^{l(i+1)}\delta}}{\Gamma(n(m, i)+m(l, j))\sigma^{m+1}} & \text{if } d < s, s \geq 0 \end{cases}$$

$$f(t, w; \lambda) = \begin{cases} \sum_{i=0}^{q-1} \sum_{j=0}^{l-1} \frac{a(n, q, i)b(m, l, j)e^{m(l, j)\lambda}w^{e^{-n(m, j)e^{l(i+1)}\delta}}}{\Gamma(n(m, i)+m(l, j))} & \text{if } tw \geq \lambda, w \geq 0 \\ \sum_{i=0}^{q-1} \sum_{j=0}^{l-1} \frac{a(n, q, i)b(m, l, j)e^{-n(m, j)e^{l(i+1)}\delta}}{\Gamma(n(m, i)+m(l, j))} & \text{if } tw < \lambda, w \geq 0 \end{cases}$$

The P-value, $P(T > t)$, is obtained from $f(t, w; \lambda)$, $tw \geq \lambda$, $w \geq 0$ as

$$P(T > t) = \sum_{i=0}^{q-1} \sum_{j=0}^{l-1} \left[ \frac{a(n, q, i)b(m, l, j)e^{m(l, j)\lambda}w^{e^{-n(m, j)e^{l(i+1)}\delta}}}{\Gamma(n(m, i)+m(l, j))\sigma^{m+1}} \right]$$

Integration by parts yields

$$P(T > t) = \left[ \frac{e^{m(l, j)\lambda}}{(1+m(l, j)t)^{n+m}} P(G > \lambda[m(l, j)+1/t]) \right] + P(G < \lambda/t)$$

where $G$ the Gamma random variable with shape parameter $v + 1$.

In many situations the first ordered statistics are available i.e. $q = 1$, $l = 1$ and the above simplifies to

$$P(T > t) = \left[ \frac{e^{m\lambda}}{(1+m)^n} P(G > \lambda[1+m]) \right] + P(G < \lambda/t)$$

Point estimators of $R$ for the case $q = 1$, $l = 1$ are considered in Gupta and Gupta (1988) where the maximum likelihood estimator of $R$ is obtained with $T / (m + n)$ as an estimator of $\lambda$ in the equation for $R$.

Inference when $\mu_x = \mu_y = \mu$

When $\mu_x = \mu_y = \mu$ but $\mu$ is unknown then $R$ reduces to

$$\theta = \frac{\sigma_x}{\sigma_x + \sigma_y}$$
Consider the null hypothesis

\[ H_0: \theta \geq q_0 \]

or equivalently

\[ H_0: \frac{\sigma_x}{\sigma_y} \geq \frac{q_0}{1-q_0} \]

where \( q_0 \) is a specified probability. \( 2S_x/\sigma_x \) and \( 2S_y/\sigma_y \) are independently distributed as chi-square with \( 2v_x \) and \( 2v_y \) degrees of freedom and \( \frac{S_x/(v_x\sigma_x)}{S_y/(v_y\sigma_y)} \) has a F distribution with \( v_x \) and \( v_y \) degrees of freedom. Hence, one can use \( F = \frac{S_x/v_x (1-q_0)}{S_y/v_y q_0} \) as the test statistic.

An estimate of \( \theta \) is \( \hat{\theta} = \frac{S_x}{S_x/v_x + S_y/v_y} \). A \((1-\alpha)\) confidence interval for \( \theta \) is obtainable from the F distribution with \( v_x \) and \( v_y \) degrees of freedom via

\[ P(F_l < \frac{S_x/v_x}{S_y/v_y} < F_u) \]

where \( F_l \) and \( F_u \) satisfy \( 1-\alpha = P(F_l < F < F_u) \). The confidence interval can be written, after some algebraic manipulation, as

\[
\left( \frac{-\hat{\theta}}{\hat{\theta} + (1-\hat{\theta})F_u}, \frac{-\hat{\theta}}{\hat{\theta} + (1-\hat{\theta})F_l} \right)
\]

When complete samples are available,

\[ S_x = \sum_{i=2}^{n} (X_i - X_1), \quad S_y = \sum_{j=2}^{m} (Y_j - Y_1) \]

one of which is slightly different from those used in Bai and Hong (1992). They used

\[ \sum_{i=1}^{n} (X_i - \min(X_i,Y_i)) \quad \sum_{j=1}^{n} (Y_j - \min(X_i,Y_i)) \]

instead of \( S_x, \) \( S_y \) respectively and obtained approximate confidence interval based on a mixed beta distribution.

Applications to Related Distributions

Suppose \( X \) and \( Y \) are independent two-parameter exponential random variables and \( \varphi \) is a monotonic function with inverse \( \varphi^{-1} \). Because

\[ P(Y < X) = P(\varphi(Y) < \varphi(X)) \]

the tests and confidence bounds developed in the previous sections are also applicable to the variables \( \varphi(X) \) and \( \varphi(Y) \); the results are to be applied after making the transformation, \( \varphi \), to the observations. The results are applicable to the Rayleigh distribution with \( \varphi(X) = \sqrt{2X} \), \( \varphi^{-1}(X) = X^2/2 \) and the Pareto distribution with \( \varphi(X) = \exp(X) \), \( \varphi^{-1}(X) = \ln(X) \).

Numerical example

Suppose a system has two main parts, \( Y \) and \( X \), whose lifetimes are exponentially distributed. Suppose \( m=n=15 \) component parts are put on test simultaneously and the failure times are \( \{106, 108, 109, 113, 116, 126, 127, 132, 138, 141, 147, 164, 185, 202, 285\} \) and \( \{79, 82, 88, 89, 91, 107, 112, 118, 133, 149, 165, 167, 170, 202, 222\} \) for \( Y \) and \( X \) respectively. Then \( l = q = 1, \quad c_i, d_j = 1 \) for \( i = j = 1,2,...,15, \quad t = 0.0193, \quad s_y = 609, \quad s_x = 789, \quad v_x = v_y = 14 \). To test whether system failure may be equally likely due to either part, the test of \( H_0: \lambda \geq 0 \) \((R \geq 0.5)\) against \( H_1: \lambda < 0 \) yields a P-value of 0.0004 which is sufficient evidence that \( X \) is more likely to fail before \( Y \). If instead one is interested to test, say, \( H_0: R = \frac{1}{2}e^{-\lambda} \geq 0.4 \) against \( H_1: R < 0.4 \) then the P-value is 0.011. There is sufficient evidence to reject \( H_0 \); the probability that system failure will be due to \( Y \) is less than 0.4. If, for example, the values 108
and 109 for Y are missing, then one would set \(d_2 = d_3 = 0\) and the recalculated values for the test of \(H_0: R \geq 0.4\) are \(t = 0.0199, s_y = 568\), and \(v_y = 12\) with a P-value equal 0.016.

**Conclusion**

Tests of hypotheses and confidence bounds for R have been developed for the two-parameter exponential distribution in two cases, namely one involving a common scale parameter and the other a common location parameter. Exact tests for the two cases are derived for situations in which data may be missing or incomplete. Exact confidence bounds for R in the common location case are also proposed and they provide an alternative to the approximate bounds that have been considered in a complete sample situation. Furthermore, these results are applicable to a larger class of distributions which includes the Raleigh and the Pareto distributions.

**References**


