Journal of Modern Applied Statistical Methods

Volume 5 | Issue 2 Article 5

11-1-2005

Inference for P(Y

Vee Ming Ng
Murdoch University, Australia, V.Ng@murdoch.edu.au

Recommended Citation
Ng, Vee Ming (2005) "Inference for P(Y," *Journal of Modern Applied Statistical Methods*: Vol. 5 : Iss. 2 , Article 5.
DOI: 10.22237/jmasm/1162353840

Journal of Modern Applied Statistical Methods
November, 2006, Vol. 5, No. 2, 323-327

1538 – 9472/06/$95.00

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.

Vee Ming Ng is lecturer at Murdoch University. His research interests include statistical inference and prediction and multivariate analysis. Email: V.Ng@murdoch.edu.au

Methodology and Results

A two-parameter exponential distribution with parameters (μ, σ) is defined by the probability density function:

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

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

$$R \equiv P(Y<X)$$
$$= \int \int_{\mu_y<y<x} f(x;\mu_x,\sigma_x) f(y;\mu_y,\sigma_y) dydx$$
$$= \begin{Bmatrix} \dfrac{\sigma_x}{\sigma_x+\sigma_y} e^{-\delta/\sigma_x}, \delta \ge 0 \\ 1-\dfrac{\sigma_y}{\sigma_x+\sigma_y} e^{\delta/\sigma_y}, \delta < 0 \end{Bmatrix}$$

where $\delta = \mu_y - \mu_x$. 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 (μ_x,σ_x) and (μ_y,σ_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

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

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

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 σ is unknown, then

$$R = \begin{Bmatrix} \dfrac{1}{2} e^{-\lambda}, \lambda \ge 0 \\ 1-\dfrac{1}{2} e^{-\lambda}, \lambda < 0 \end{Bmatrix}$$

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 $H_0: \frac{1}{2} e^{-\lambda} \ge p_0$, 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 $T = (Y_l - X_q)/S_p$. T is a maximal invariant and its distribution depends only on λ. A large value of T would be evidence against H_0. Hence, for an observed value t of T, P(T > t) for t ≥ 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 of density function of X_q, Y_l, S_p, the joint probability density function of $D = Y_l - X_q$ and S_p, $f(d,s;\delta,\sigma)$, which then yields the joint of density function $f(t,w;\lambda)$ of $T = D/S_p$ and $W = S_p/\sigma$:

$$f(d,s;\delta,\sigma) = \begin{cases} \sum_{i=0}^{q-1}\sum_{j=0}^{l-1} \dfrac{a(n,q,i)b(m,l,j)e^{m(l,j)\delta/\sigma}s^{v-1}e^{-[m(l,j)d+s]/\sigma}}{\Gamma(v)[n(q,i)+m(l,j)]\sigma^{v+1}} \\ ,d \geq \delta, s \geq 0 \\ \sum_{i=0}^{q-1}\sum_{j=0}^{l-1} \dfrac{a(n,q,i)b(m,l,j)e^{-n(q,i)\delta/\sigma}s^{v-1}e^{[n(q,i)d-s]/\sigma}}{\Gamma(v)[n(q,i)+m(l,j)]\sigma^{v+1}} \\ ,d < \delta, s \geq 0 \end{cases}$$

$$f(t,w;\lambda) = \begin{cases} \sum_{i=0}^{q-1}\sum_{j=0}^{l-1} \dfrac{a(n,q,i)b(m,l,j)e^{m(l,j)\lambda}w^{v}e^{-w[m(l,j)t+1]}}{\Gamma(v)[n(q,i)+m(l,j)]} \\ ,tw \geq \lambda, w \geq 0 \\ \sum_{i=0}^{q-1}\sum_{j=0}^{l-1} \dfrac{a(n,q,i)b(m,l,j)e^{-n(q,i)\lambda}w^{v}e^{w[n(q,i)t-1]}}{\Gamma(v)[n(q,i)+m(l,j)]} \\ ,tw < \lambda, w \geq 0 \end{cases}$$

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

$$P(T>t) = \sum_{i=0}^{q-1}\sum_{j=0}^{l-1} \int_t^{\infty} \left\{ \int_{\lambda[m(l,j)+1/u]}^{\infty} \frac{\dfrac{a(n,q,i)b(m,l,j)e^{m(l,j)\lambda}z^{v}e^{-z}}{\Gamma(v)[n(q,i)+m(l,j)]}}{[1+m(l,j)u]^{v+1}} dz \right\} du$$

Integration by parts yields

$$P(T>t) = \sum_{i=0}^{q-1}\sum_{j=0}^{l-1} \frac{a(n,q,i)b(m,l,j)}{m(l,j)[n(q,i)+m(l,j)]} \left\{ \begin{array}{l} \dfrac{e^{m(l,j)\lambda}}{(1+m(l,j)t)^{v}} P(G > \lambda[m(l,j)+1/t]) \\ +P(G < \lambda/t) \end{array} \right\}$$

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

$$P(T>t) = \frac{n}{n+m} \left\{ \begin{array}{l} \dfrac{e^{m\lambda}}{(1+mt)^{v}} P(G > \lambda[m+1/t]) \\ +P(G < \lambda/t) \end{array} \right\}$$

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 λ in the equation for R.

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

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

$$\theta \equiv \frac{\sigma_x}{\sigma_x + \sigma_y}$$

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 $F = \frac{S_x}{S_y} \frac{v_y (1-q_0)}{v_x q_0}$ as the test statistic.

An estimate of θ is $\hat{\theta} = \frac{S_x / v_x}{S_x / v_x + S_y / v_y}$. A $(1-\alpha)$ confidence interval for θ is obtainable from the F distribution with v_x and v_y degrees of freedom via $P\{F_l < \frac{S_x \sigma_y v_y}{S_y \sigma_x v_x} < F_u\}$ where F_l and F_u satisfies $1-\alpha = P\{F_l < F < F_u\}$. 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, $S_x = \sum_{i=2}^{n} (X_i - X_1)$, $S_y = \sum_{j=2}^{m} (Y_j - Y_1)$ one of which is slightly different from those used in Bai and Hong (1992). They used $\sum_{i=1}^{n} (X_i - \min(X_1, Y_1))$ $\sum_{j=1}^{m} (Y_j - \min(X_1, Y_1))$ 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 φ is a monotonic function with inverse φ^{-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, φ, 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 $l = q = 1$, $c_i, d_j = 1$ for $i = j = 1,2,...,15$, $t = 0.0193$, $s_y = 609$, $s_x = 789$, and $v_x = v_y = 14$. 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

Ali, M. M., Woo, J. & Pal, M. (2004). Inference on reliability P(Y<X) in two-parameter exponential distribution. *International Journal of Statistical Sciences, 3*, 119-125.

Baklizi, A. (2003). Confidence intervals for P(X>Y) in the exponential case with common location parameter. *Journal of Modern Applied Statistical Methods, 2*, 341-349.

Bai, D. S. & Hong (1992). Estimation of P(X<Y) in the exponential case with common location parameter. *Communications in Statistics-Theory & Method, 21*, 269-282.

Chao, A. (1982). On comparing estimators of Pr{Y<X} in the exponential case. *IEEE Transaction on Reliability, R31*, 389-392.

Enis, P. & Geisser, S. (1971). Estimation of the probability that Y < X. *Journal American Statistical Association, 66*, 162-168

Gupta, R. D. & Gupta, R. C. (1988). Estimation of $P(Y_p > \max(Y_1, Y_2, ..., Y_{p-1}))$ in the exponential case. *Communication in Statistics-Theory & Methods, 17*, 911-924.

Likes, J. (1974). Prediction of sth ordered observation for the two-parameter exponential distribution. *Technometrics, 16*, 241-244.

Ranganathan, J. & Kale, B. K. (1979). Tests of the hypotheses for reliability functions in two-parameter exponential models. *Canadian Journal of Statistics, 7,* 117-184.

Tanis, E. A. (1964). Linear forms in the order statistics from an exponential distribution. *Annals Mathematical Statistics, 35*, 270-276.

Tong, H. (1977). On the estimation of P(Y<X) for the exponential families. *IEEE Transactions on Reliability, R26*, 54-56