COMPARING TWO QUANTILES

THE BURR-TYPE-X AND WEIBULL CASES

MOHAMMED A. SHAYIB

mashayib@pvamu.edu

ALIAKBAR MONTAZER HAGHIGHI

amhaghighi@pvamu.edu

Department of Mathematics PVAMU, Prairie View, Texas 77446 USA

ABSTRACT

Interval estimation and hypothesis testing for the difference between two quantiles are investigated in this article. The underlying distributions that will be considered are the Weibull and Burr type X. The estimation procedures will be based on the generalized confidence interval procedure, while the hypothesis testing will be based on the generalized p-values procedure. Simulation will be carried to check on the accuracy of both procedures. In addition, a check will be made on the reliability of the system based on these quantiles.

1. INTRODUCTION

A statistical comparison between two populations based on their mean, variances or proportions is a common practice in the literature. This is carried out to check on the superiority of one population over the other. In this article we will make such a comparison between two populations based on their quantiles. In probability and statistics, the quantile function of the probability distribution of a random variable *specifies*, for a given probability, the value which the random variable will be at, or below, with that probability. A comparison between two quantiles for the Normal and Exponential distributions had been done; see Guo and Krishnamurthy, (2005). The quantile function is one way of prescribing a probability distribution. It is an alternative to the probability density or mass function, to the cumulative distribution function, and to the characteristic function. The quantile function of a probability distribution is the inverse F^{-1} of its cumulative distribution function (cdf) F. Assuming a continuous and strictly monotonic distribution function, $F: R \rightarrow (0,1)$, the quantile function returns the value below which the random variable drawn from the given distribution would fall, $p \times 100$ percent of the time. That is, it returns the value of x such that

$$F(x) = \Pr(X \leq x) = p. \tag{1.1}$$

If the probability distribution is discrete rather than continuous then there may be gaps between values in the domain of its cdf, while if the cdf is only weakly monotonic there may be "flat spots" in its range. In either case, the quantile function is

$$Q(p) = F^{-1}(p) = \inf \{x \in R : p \leq F(x)\}, \tag{1.2}$$

for a probability $0 < p < 1$, and the quantile function returns the minimum value of x for which the previous probability statement holds.

Consider two independent random variables X and Y. Let x_p and y_p denote the pth quantile of X and Y respectively. That is,

$$x_p = \inf\{x: P(X \leq x) \geq p\} \quad \text{and} \quad y_p = \inf\{y: P(Y \leq y) \geq p\}. \tag{1.3}$$

The problem of interest here is to make a statistical inference about x_p - y_p based on samples of sizes m and n observations on X and Y, respectively.

In this article, we propose methods for comparing the quantiles of two Burr – Type-X, and comparing the quantiles of two Weibull distributions. The generalized p-value has been introduced by Tsui and Weerahandi (1989), and the generalized confidence interval by Weerahandi (1993). Using this approach, we will give an inferential procedure for the difference between two Burr-Type-X quantiles in the following section. The performance of the procedure will be evaluated numerically through simulation. In Section 3 we present inferential procedure for x_p - y_p, when the underlying distributions are taken to be Weibull with different parameters. Section 4 will contain the conclusions and recommendations.

2. BURR-TYPE-*X* DISTRIBUTION

Burr (1942) introduced 12 different forms of cumulative distribution functions for modeling lifetime data or survival data. Out of those 12 distributions, *Burr-Type-X* and *Burr-Type-XII* have received the maximum attention. Several authors have considered different aspects of these two distributions. In this article we will present the generalized inferential procedures for the difference between the quantiles of two Burr-Type-X distributions. It is to be noted that the probability density function (pdf) of the Burr-Type-X distribution is given as follows:

$$f(x) = 2\theta x e^{-x^2}(1-e^{-x^2})^{\theta-1}, \qquad x > 0, \theta > 0 \qquad (2.1)$$

Moreover, for the one-parameter *Burr –Type-X* distribution, the cumulative distribution function F is given by

$$F(x|\theta)=\left\{1-\exp(-x^2)\right\}^{\theta} \qquad x>0,\ \theta>0\ , \tag{2.2}$$

For any given 0 < p < 1, the p^th quartile is the positive root of F(x) = p,
i.e. $p=(1-\exp(-x^2))^{\theta}$, or

$$x=\sqrt{\ln[1/(1-p^{1/\theta})]} \tag{2.3}$$

Thus if X_i ~ f(x|θ_i), i =1, 2, then the p_i th quantile of X_i can be expressed as

$$\eta_i=\sqrt{\ln[1/(1-p_i^{1/\theta_i})]}, \qquad i = 1, 2. \tag{2.4}$$

Theorem 2.1 Let X~ *Burr-Type-X* (θ), with pdf given by (2.1), then the

random variable $U=-\ln\left(1-e^{-x^2}\right)$

will have the one-parameter exponential distribution with mean $\lambda=1/\theta$, i.e., U will

have the following pdf

$$h(u)=\begin{cases}\dfrac{e^{-u/\lambda}}{\lambda}, & u>0,\\ 0, & otherwise.\end{cases} \tag{2.5}$$

Proof: Using the technique of transformations of a random variable of the continuous type will lead to the result.

Theorem 2.2 Because η_i i=1, 2, are positive, testing
$H_0:\eta_1\le\eta_2$ *vs.* $H_1:\eta_1>\eta_2$ is equivalent to

$$H_0:\lambda_2/\lambda_1 \le c \quad vs. \quad H_1:\lambda_2/\lambda_1 > c, \tag{2.6}$$

where c = $\ln p_1/\ln p_2$, and λ_i I = 1, 2 as defined in Theorem 2.1. In the remaining of the article we will be referring to the means of the exponential distributions as cited in Theorem 2.1 above.

Proof: From (2.4), we see that $\eta_1 \le \eta_2$ if and only if

$\eta_1^2 \le \eta_2^2$ if and only if

$-\eta_2^2 \le -\eta_1^2$ iff

$1-e^{-\eta_1^2} \le 1-e^{-\eta_2^2}$

iff $\ln(1-e^{-\eta_1^2}) \le \ln(1-e^{-\eta_2^2})$

iff $-(1/\theta_2)\ln p_2 \le -(1/\theta_1)\ln p_1$

iff $-(\lambda_2)\ln p_2 \le -(\lambda_1)\ln p_1$

iff $\lambda_2/\lambda_1 \le \ln p_1/\ln p_2$ = c .

Moreover, we can easily see that the above test is equivalent to

$$H_0:\theta_1/\theta_2 \le c \quad vs \quad H_1:\theta_1/\theta_2 > c.$$

.

Utilizing Theorems 2.1 and 2.2, we find that all we need is to generate samples from exponential distributions with parameters λ_1 and λ_2 respectively. Let $X_{i1},...,X_{in_i}$ be a sample from F(u| λ_i), i = 1,2. Define $Y_i = \sum_{j=1}^{n_i} X_{ij}, i = 1,2$. Notice that Y_1 and Y_2 are independent with $2Y_i/\lambda_i \sim \chi^2_{2n_i}$, i = 1, 2, and hence Y_1 / Y_2 is distributed as a constant times an F random variable, (see Guo, and Krishnamoorthy (2005)). Thus it can be seen that the p-value for testing (2.6) is given by

$$P(F_{2n_1,2n_2} < cn_2 y_1 / (n_1 y_2)), \tag{2.7}$$

where $F_{a,b}$ denotes the F distribution with a degrees of freedom for the numerator, and b degrees of freedom for the denominator. Thus the null hypothesis in (2.6) will be rejected whenever this p-value is less than α ,(see Lawless, 1982)

3. THE WEIBULL DISTRIBUTION

A common application of the Weibull distribution (introduced by the Swedish physicist Waloddi Weibull in 1939), is to model the lifetimes of components such as bearings, ceramics, capacitors, and dielectrics. The Weibull distribution, in model fitting, is a strong competitor to the gamma distribution. Both the gamma and Weibull distributions are skewed, but they are valuable distributions for model fitting. Weibull distribution is commonly used as a model for life length because of the properties of its failure rate function $h(x) = (\alpha / \beta) x^{\alpha - 1}$, when the pdf is given by:

$$f(x) = \begin{cases} \dfrac{\alpha}{\beta} x^{\alpha-1} e^{-\frac{x^{\alpha}}{\beta}}, & x, \alpha, \beta > 0, \\ 0, & otherwise. \end{cases} \tag{3.1 A}$$

The failure rate function, for $\alpha > 1$, is a monotonically increasing function with no upper bound. This property gives the edge for the Weibull distribution over the gamma distribution, where the failure rate function is always bounded by $1/\beta$, when the probability density function for the gamma distribution is

$$g(y)=\frac{1}{\beta}\left(\frac{y}{\beta}\right)^{\alpha-1}\frac{e^{-y/\beta}}{\Gamma(\alpha)},\ y>0,\ \ \Gamma(\alpha)=\int_0^\infty e^{-u}u^{\alpha-1}du \tag{3.2}$$

Another property that gives the edge for the Weibull distribution over the gamma distribution is by varying the values of α and β, a wide variety of curves can be generated. Because of this, the Weibull distribution can be made to fit a wide variety of data sets.

Another for the Weibull probability density function is given by

$$f(x)=\begin{cases}\frac{\alpha}{\beta^{\alpha}}x^{\alpha-1}e^{-(\frac{x}{\beta})^{\alpha}}, & x,\alpha,\beta>0,\\ 0, & otherwise.\end{cases} \tag{3.1.B}$$

(See Hogg, R.V. and Ledolter, J.; 1992 p.126)

Yet another form is given by

$$f(x)=\begin{cases}\alpha\beta x^{\beta-1}e^{-\alpha x\beta}, & x,\alpha,\beta>0,\\ 0, & otherwise.\end{cases} \tag{3.1.C}$$

(See Walpole, R.E. and Myers, R.H., 4^{th} Ed.; 1989, p.171) (Misprinting)

$$f(x)=\begin{cases}\alpha\beta x^{\beta-1}e^{-\alpha x^{\beta}}, & x,\alpha,\beta>0,\\ 0, & otherwise.\end{cases} \tag{3.1.C'}$$

In this paper we will consider the Weibull distribution with two parameters, namely α and β, where α is the shape parameter while $\beta^{1/\alpha}$ is the scale parameter of the distribution, as given in the function form of (3.1.A).

We will use the notation X ~ Weibull (α , β) for the random variable X having a Weibull probability density function with parameters α and β, as shown above in (3.1). We will also take the shape parameter as in Hogg and Tanis (1988) to be 1< α < 5. In addition, the cdf for the Weibull distribution given in (3.1.A) is $F(x) = 1 - \exp\left(-x^{\alpha} / \beta\right)$, and thus the pth quartile is the solution of the equation F(x) = p, namely,

$$\eta = [\beta \ln\{1/(1-p)\}]^{1/\alpha} \qquad (3.3)$$

If the form for Weibull probability density function, as given in (3.1.B), is used instead of the form in (3.1.A), then the new quartile in (3.3) will be presented as

$$\eta = \beta[\ln\{1/(1-p)\}]^{1/\alpha} \qquad (3.3.A)$$

Clearly, it is seen the quartile in (3.3.A) will be greater than that in (3.3), for the same values of p, and β when α > 1,

Now, as it was the case in the Burr –Type-X, we have the following Theorem.

Theorem 3.1 **If** X ~ Weibull(α, β_1) as given in (3.1.A), then

i) $U = X^{\alpha}$ will have an exponential distribution with parameter β_1, i.e.,

$$h(u) = \begin{cases} \dfrac{e^{-u/\beta_1}}{\beta_1}, & u > 0, \\ 0, & otherwise, \end{cases}$$

and

ii) $\dfrac{2m\bar{U}}{\beta_1} \sim \chi^2_{2m}$ (see Baklizi and Abu Dayyeh (2003)).

iii) Similarly, if $Y \sim$ Weibull(α, β_2), then $V = Y^{\alpha}$ will have an exponential distribution with parameter β_2, i.e.

$$g(v) = \begin{cases} \dfrac{1}{\beta_2} e^{-v/\beta_2}, & v > 0 \\ 0, & otherwise, \end{cases}$$

iv) $\dfrac{2n\bar{V}}{\beta_2} \sim \chi^2_{2n}$ (See ibid),

and

v) $\dfrac{\beta_1 \bar{V}}{\beta_2 \bar{U}} \sim F_{2n,2m}$ (See ibid).

Theorem 3.2 From (3.3) and because η_i i=1, 2, are positive, testing

$$H_0: \eta_1 \le \eta_2 \quad vs. \quad H_1: \eta_1 > \eta_2$$

is equivalent to

$$H_0: \beta_1 / \beta_2 \le c \quad vs \quad H_1: \beta_1 / \beta_2 > c,$$

where c = $\ln(1-p_2)/\ln(1-p_1)$.

Under H_0, $W = \dfrac{R_0}{1-R_0}\dfrac{\bar{V}}{\bar{U}} \sim F_{2n,2m}$

The *p*-value for this test is

$$p = 2min\,[P_{H_0}(W > w),\ P_{H_0}(W < w)] = 2min\,[1- F(w), F(w)],$$

where *w* is the observed value of the test statistic *W* and *F* is the distribution function of *W* under H_0. The *p*-value of this test indicates how strongly H_0 is supported by the data. A large *p*-value indicates that *R* is close to the prior estimate R_0 (Tse and Tso, 1996). Thus, we can use this *p*-value to form the shrinkage estimator.

REFERENCES

Baklizi, A. and Abu Dayyeh, W. (2003) Shrinkage Estimation of *P*(*Y* < *X*) in the Exponential Case. *Communications in Statistics, Simulation and Computation* 32(1): 31-42

Guo, H. and Krishnamoorthy, K. (2005) Comparison between two quantiles: the Normal and Exponential cases, *Communications in Statistics-Simulation and Computation*, 34, 2, 243 – 252

Haghighi, A. M. and Shayib, M. A. (2008) Shrinkage Estimators for Calculating Reliability, Weibull case, *Journal of Applied Statistical Science* (JASS), Vol. 17, No.2, 1-14

Hogg, R. V. and Tanis, E. A. (1988) *Probability and Statistical Inference*, 3rd Edition, Macmillan

Hogg, R.V. and Ledolter, J.; (1992) *Applied Statistics for Engineers and Physical Scientists*, 2nd Edition, Macmillan

Tsui, K. W. and Weerahandi, S. (1989) Generalized p-value in significance testing of hypotheses in the presence of nuisance parameters, *J. Amer. Statist. Assoc*. 84, 602 - 607

Walpole, R.E. and Myers, R.H., (1989) *Probability and Statistics for Engineers and Scientists*, 4th edition, Macmillan

Weerahandi, S. (1993) Generalized confidence interval. *J. Amer. Statist. Assoc*. 88, 899-905