Upper Bound of the Generalize p-Value for the Behrens-Fisher Problem with a Known Ratio of Variances

Rada Somkhuean, Suparat Niwitpong, and Sa-aat Niwitpong

Abstract—This paper presents the generalized p-values for testing the Behrens-Fisher problem when a ratio of variance is known. We also derive a closed form expression of the upper bound of the proposed generalized p-value.

Keywords—Generalized p-value, hypothesis testing, ratio of variances, upper bound.

I. INTRODUCTION

Schechtman and Sherman [1] described a situation with a known ratio of variances arises in practice when two instruments reports (averaged) response of the same object based on a number of replicates. If the two instruments have the same precision for a single measurement, then the ratio of the variance of the responses is known and it is simply the ratio of the number of replicates going into each response. They proposed a t-test statistic, which has each response. They proposed a t-test statistic, which has a ratio of variances.

II. GENERALIZED P-VALUES FOR THE BEHRENS-FISHER PROBLEM

Let \( X_1, X_2, \ldots, X_n \) and \( Y_1, Y_2, \ldots, Y_m \) be random samples from two independent normal distributions with means \( \mu_x, \mu_y \) and standard deviations \( \sigma_x \) and \( \sigma_y \), respectively.

Rada Somkhuean is with Department of Applied Statistic, Faculty of Applied Science, King Mongkuts University of Technology North Bangkok, Bangkok 10800, Thailand (e-mail: rada_m.1@hotmail.com).

Supatar Niwitpong is with Department of Applied Statistic, Faculty of Applied Science, King Mongkuts University of Technology North Bangkok, Bangkok 10800, Thailand (e-mail: snw@kmutnb.ac.th).

Sa-aat Niwitpong is with Department of Applied Statistic, Faculty of Applied Science, King Mongkuts University of Technology North Bangkok, Bangkok 10800, Thailand (e-mail: suparat8@gmail.com).

Let \( \theta = \mu_x - \mu_y \) be the parameter of interest. The problem is to test the hypothesis \( H_0 : \theta = 0 \) against the alternative hypothesis \( H_1 : \theta > 0 \) for some fixed \( \theta_0 \). The sufficient statistic of this problem is \( (\bar{X}, \bar{Y}, S^2_{X}, S^2_{Y}) \) (Tsui and Weerahandi [6]) where

\[
S^2_{X} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \quad \text{and} \quad S^2_{Y} = \frac{1}{m} \sum_{j=1}^{m} (Y_j - \bar{Y})^2
\]

The probability distributions of the statistics, \( \bar{X} \sim N(\mu_x, \frac{\sigma^2_x}{n}) \), \( \bar{Y} \sim N(\mu_y, \frac{\sigma^2_y}{m}) \), \( V = n \frac{\sigma^2_x}{\sigma^2_{X}} \sim \chi^2_{n-1} \) and \( U = m \frac{\sigma^2_y}{\sigma^2_{Y}} \sim \chi^2_{m-1} \) are independent of one another. Tsui and Weerahandi[6] proposed the generalized p-value for the above hypothesis as follow:

Suppose a random quantity \( T^*(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) \) can be expressed as

\[
T^*(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) = T(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) - \theta
\]

where

\[
T(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) = \frac{\bar{X} - \bar{Y} - \theta}{\sqrt{\frac{\sigma^2_x}{n} + \frac{\sigma^2_y}{m}}} \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{nS^2_{X}} + \frac{\sum_{j=1}^{m} (Y_j - \bar{Y})^2}{mS^2_{Y}}}
\]

and

\[
T(x, y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) = \bar{x} - \bar{y} - \theta
\]

It is straightforward to see that \( T(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) \) is free from nuisance parameters \( \sigma^2_x \) and \( \sigma^2_y \) and has the same distribution

\[
Z \sqrt{\frac{\bar{x}^2}{\bar{Y}^2} + \frac{\bar{y}^2}{\bar{X}^2}} \quad \text{where} \quad Z \sim N(0, 1)
\]

\( T^*(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) \) is defined to be a generalized test variable and \( T(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) \) is defined to be a generalized pivot statistic and \( T^*(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) \) is required to satisfy the following conditions:

C1. For a fixed \( x \) and \( y \), the probability distribution of \( T^*(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) \) is free of the unknown parameters.

C2. The observed value of \( T^*(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) \) namely \( T^*(x, y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) \) is simply \( \theta \).

C3. For fixed \( x, y \) and \( \delta = (\sigma^2_x, \sigma^2_y) \), \( T^*(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) \) is stochastically monotone in \( \theta \).

The generalized pivot statistic \( T(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) \) is also required to satisfy the following conditions:
C4. For a fixed \( x \) and \( y \), the probability distribution of 
\( T(X,Y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) \) is free of the unknown 
parameters \( \theta \) and \( \delta = (\sigma_x^2, \sigma_y^2) \).

C5. The observed value of \( T(X,Y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) \), 
namely \( T(x,y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) \) is simply equal to \( \theta \).

A 100(1-\( \alpha \)/2)% generalized lower confidence limit for \( \theta \) is 
then given by \( T(x,y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) \)\( _{1-\alpha} \), the 
100(1-\( \alpha \))th percentiles of \( T(x,y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) \).

Further, given the observed value \( x \), let \( t_1 \) and \( t_2 \) be such 
values that 
\[
P(t_1 < T(X,Y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) < t_2|\theta) = 1 - \alpha 
\]
for chosen significant level \( \alpha \in (0,1) \) is the confidence interval 
for parameter \( \theta \) defined by \( \{ \theta : t_1 < T(X,Y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) < t_2 \} \) is 
a 100(1-\( \alpha \))% 
generalized confidence interval for \( \theta \).

For the one-sided hypothesis given above they defined a 
data-based extreme region \( C_{x,y} \) of the form 
\[
C_{x,y}(\theta,\sigma_x^2,\sigma_y^2) = \{ (X,Y) : T(X,Y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) 
- T(x,y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) > 0 \}.
\]

For the one-sided Behrens-Fisher problem, the generalized 
p-value is 
\[
p^* = Pr(T(X,Y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) 
- T(x,y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2)|\theta = \theta_0).
\]

III. MAIN RESULTS FOR BEHRENS-FISHER PROBLEM WITH 
ONE VARIANCE UNKNOWN

Following Schechtman and Sherman [1], we suppose a ratio 
of variances is known i.e. \( \sigma_x^2 = \sigma_y^2 \), where \( \sigma \) is a constant.

According to Tsui and Weerahandi [6], one of the potential 
pivotal quantity can be defined as 
\[
Q(X,Y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) =
\]
\[
= \frac{X - Y - \theta}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m} + \theta}} 
= \frac{X - Y - \theta}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m} + \frac{\sigma_x^2}{m} \sigma_y^2}} 
= \frac{X - Y - \theta}{\sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{m} \sigma_x^2}} 
= \frac{X - Y - \theta}{\sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{m} \sigma_x^2}} \left( \frac{m + nc}{n m} \right) + \theta
= Z \sqrt{\frac{s_x^2}{m}} + \frac{m + nc}{m} + \theta 
\]

For the one-side Behrens-Fisher problem as stated, 
\( H_0 : \theta < \theta_0 \) against \( H_0 : \theta > \theta_0 \), we can assume \( \theta_0 = 0 \) 
without loss of generality, and the generalized p-value for the 
one-sided Behrens-Fisher problem is \( p(q) \) which is 
\[
Pr(Q(X,Y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) \geq q_{obs} = 0) = Pr \left( Z \sqrt{\frac{s_x^2}{m}} \left( \frac{m + nc}{m} \right) \geq \bar{x} - \bar{y} \right)
\]
\[
= Pr \left( Z \geq \frac{s_x^2}{\sqrt{m}} \left( \frac{m + nc}{m} \right) \right)^{-\frac{1}{2}}
\]
\[
= Pr \left( Z \leq \frac{s_x^2}{\sqrt{m}} \left( \frac{m + nc}{m} \right) \right)^{-\frac{1}{2}}
\]
\[
= Ev (\Phi (\frac{\bar{y} - \bar{x}}{s_x^2} \left( \frac{m + nc}{m} \right) )^{-\frac{1}{2}})
\]

where \( \Phi(\cdot) \) is a cdf of the standard normal distribution and 
\( Ev(\cdot) \) is an expectation operator with respect to \( V \).

To find the upper bound of \( p(q) \), we need Theorems 1-2 
based on Tang and Tsui [7] as follows:

Theorem 1. Define 
\[
h(v) = \Phi \left( z \sqrt{\frac{vm}{t}} \right) \quad for \quad v \in (0,1).
\]

Then for fixed \( z < 0, h(v) \) is a convex function of \( v \).

Proof: Letting 
\[
h(v) = z \sqrt{\frac{vm}{t}},
\]
we have \( f(v) = \Phi(h(v)) \). Let \( \phi \) be the probability density 
function of standard normal distribution.

Then 
\[
f''(v) = (f'(v)')' = (\phi(h(v)))' = (\phi(h(v)))' = \phi(h(v))h''(v) + \phi(h(v))h'(v)
\]

For \( Z < 0, h(v) < 0. \) Hence \( \phi'(h(v)) \geq 0. \) Obviously, 
\( \phi(h(v)) \geq 0. \) Moreover, 
\[
h''(v) = \left[ z \left( \frac{1}{2} \right) \left( \frac{vm}{t} \right)^{-\frac{1}{2}} \left( \frac{m}{t} \right) \right]^{-1}
= -z \left( \frac{m}{t} \right)^{-\frac{1}{2}} \left( \frac{m}{t} \right)^{2}
= -z \left( \frac{m}{t} \right)^{2} > 0
\]

Hence \( h(v) \geq 0, \) and \( h(v) \) is convex in \( v \).

Theorem 2. Let 
\[
g(a) = P \left[ \Phi \left( z \sqrt{\frac{(n-1)m}{C_{n-1}a(m + nc)}} \leq r \right) \right],
\]

where \( z, C_{n-1} \) independent random variables such that 
\( z \sim N(0,1), C_{n-1} \sim \chi^2_{n-1}. \) Then \( g(a) \) is a convex function 
in \( a. \)
For the one-sided Behrens Fisher problem, 

\[ H \]

Hence 

\[ g \]

and any 

\[ g \]

\[ E \]

Moreover, 

\[ h_{1}(a) = \sqrt{\frac{a(m+nc)}{m}} \Phi^{-1}(r) \] and 

\[ g_{1}(a) = \Psi^{-1}_{n-1}(h_{1}(a)) \]

Let \( \psi_{n-1} \) be the probability density function of \( t \) distribution with \( n-1 \) degrees of freedom. we have

\[ g''_{1}(a) = g'_{1}(a)' = (\psi_{n-1}(h_{1}(a)))'h_{1}(a) + \psi_{n-1}(h_{1}(a))h_{1}(a)' \]

For \( r \leq 0.5, h_{1}(a) \leq 0 \), and consequently, \( \psi_{n-1}(h_{1}(a)) \geq 0 \).

Moreover, 

\[ A = \frac{\sigma_{X}^{2}}{n_{x} \sigma_{X}^{2} + \sigma_{Y}^{2} / m} \]

\[ z = \frac{\bar{y} - \bar{x}}{\sqrt{\frac{\sigma_{X}^{2}}{n_{x} \sigma_{X}^{2} + \sigma_{Y}^{2} / m}}} \]

\[ C_{n-1} = \frac{n_{x} \sigma_{X}^{2}}{\sigma_{X}^{2}} \]

From (2)

\[ p(q) = E_{V} \left[ \Phi \left( \sqrt{\frac{V_{n}}{C_{n-1}^{A}(m+nc)}} \right) \right] \]

For any \( r < 0.5 \) and \( p(q) < r \), we must have. Hence by theorem 1

\[ f(v) = E_{V} \left[ \Phi \left( \sqrt{\frac{V_{n}}{C_{n-1}^{A}(m+nc)}} \right) \right] \] is convex in \( V \).

By Jensen’s Inequality,

\[ p(q) = E_{V} \left[ f(V) \right] \geq f \left( E(V) \right) \]

\[ p(q) = \Phi \left( Z \left( \frac{\sqrt{n_{x} \sigma_{X}^{2}}}{C_{n-1}} \right) \right) \equiv p_{1}(q) \]

Now observe that under \( \mu_{1} - \mu_{2} = 0 \), \( Z \sim N(0,1) \) and \( Z, C_{n-1} \) are independent of one another. For \( 0 < r < 0.5 \).

\[ P_{q} \{ p(q) \leq r \} \leq P_{q} \{ p_{1}(q) \leq r \} = g(A) \]

where \( g(a) \) is a defined in theorem 2. Next by theorem 2 for \( 0 < r < 0.5, g(A) \) is convex in \( A \).

\[ g(A) \leq \max \{ g(0), g(1) \} \]

\[ = \Phi \left( Z \left( \frac{(n_{x} \sigma_{X}^{2})}{C_{n-1}^{A}(m+nc)} \right) \right) \leq r \]

\[ = \Phi \left( \sqrt{\frac{n_{x} \sigma_{X}^{2}}{C_{n-1}^{A}(m+nc)}} \right) \]

\[ = \Psi_{n-1} \leq k \Phi^{-1}(r) \]

where \( k = \sqrt{\frac{n_{x} \sigma_{X}^{2}}{C_{n-1}^{A}(m+nc)}} \).

IV. CONCLUSION

In this paper, we derive an expression of the upper bound of the generalized \( p \)-value for the Behrens-Fisher problem with a know ratio of variances used the method described by Tang and Tsui [7]. This upper bound can be easily computed by R program with command: pnorm(k*qnorm(r)), when \( r \) is a fixed real value between 0 to 0.5.

REFERENCES


