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 difference 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 an exact t-distribution with the ratio of the variance of the responses is known and the ratio of the variance of the responses is not known.

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

Let \( x_i \) and \( y_j \) be random samples from two independent normal distributions with means \( \mu_x, \mu_y \) and standard deviations \( \sigma_x, \sigma_y \), respectively.

Let \( \theta = \mu_x - \mu_y \) be the parameter of interest. The problem is to test the hypothesis \( H_0 : \theta \leq 0 \) against the alternative hypothesis \( H_1 : \theta > 0 \) for some fixed \( \theta_0 \). The sufficient statistic of this problem is \( (X, Y, S_{S_x}^2, S_{S_y}^2) \) (Tsui and Weerahandi [6]) where \( X = n^{-1} \sum x_i, Y = m^{-1} \sum y_j, S_{S_x}^2 = \frac{\sum_i (x_i - X)^2}{n-1} \) and \( S_{S_y}^2 = \frac{\sum_j (y_j - Y)^2}{m-1} \).

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

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

\[
T^*(X,Y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) = T(X,Y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) - \theta
\]

where

\[
T(X,Y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) = \bar{X} - \bar{Y} - \theta \sqrt{\frac{nS_{S_x}^2}{n(S_{S_x}^2 + S_{S_y}^2)}} + \frac{\sigma_x^2 + \sigma_y^2}{mS_{S_y}^2} + \sqrt{\frac{nS_{S_x}^2}{mS_{S_y}^2}}
\]

and \( T(x,y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) = \bar{x} - \bar{y} - \theta_0 \). It is straightforward to see that \( T(X,Y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) \) is free from nuisance parameters \( \sigma_x^2 \) and \( \sigma_y^2 \) and has the same distribution \( Z \sqrt{\frac{nS_{S_x}^2}{n(S_{S_x}^2 + S_{S_y}^2)}} + \frac{\sigma_x^2 + \sigma_y^2}{mS_{S_y}^2} \) where \( Z \sim N(0,1) \).

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

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

C2. 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 \( \theta \).

C3. For fixed \( x, y \) and \( \delta = (\sigma_x^2, \sigma_y^2) \), \( T^*(X,Y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) \) is stochastically monotone in \( \theta \). The generalized pivot statistic \( T^*(X,Y,x,y,\mu_x,\mu_y,\sigma_x^2,\sigma_y^2) \) 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 $T(X, Y, x, y, \mu_x, \mu_y, \sigma_x^2, \sigma_y^2)$ is simply equal to $\theta$. A $(1 - \alpha/2)\%$ 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 $(1 - \alpha)th$ 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 $\alpha \in (0, 1)$ than 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 $(1 - \alpha)\%$ 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

\[ C_{x,y}(\theta, \sigma_x^2, \sigma_y^2) = \{ (X, Y) : T(X, Y, x, y, \mu_x, \mu_y, \sigma_x^2, \sigma_y^2) < T(x, y, x, y, \mu_x, \mu_y, \sigma_x^2, \sigma_y^2) \} \]

For the one-sided Behrens-Fisher problem, the generalised $p$-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. $\frac{\sigma_x^2}{\sigma_y^2} = c$, where $c$ is a constant. According to Tsui and Weerahandi [6], one of the potential pivotal quantity can be defined as

\[ Q(X, Y, x, y, \mu_x, \mu_y, \sigma_x^2, \sigma_y^2) = \frac{X - Y - \theta}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m} + \theta} \sqrt{\frac{n}{m}}} \]
\[ \quad - \frac{X - Y - \theta}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m} + \theta} \sqrt{\frac{n}{m}}} \frac{1}{\sqrt{\frac{m + nc}{m}}} + \theta \]
\[ \quad - \frac{X - Y - \theta}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m} + \theta} \sqrt{\frac{n}{m}}} \frac{(ns_x^2)(m + nc)}{nm} + \theta \]
\[ \quad - Z \sqrt{\frac{s_x^2}{V} \left( \frac{m + nc}{m} \right) + \theta} \]

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 generalised $p$-value for the one-sided Behrens-Fisher problem is $p(g)$ which is

\[ Pr(Q(X, Y, x, y, \mu_x, \mu_y, \sigma_x^2, \sigma_y^2) \geq q_{obs} = 0) \]

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

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

**Theorem 1. Define**

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

*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

\[ f''(v) = (f'(v))' = \left( \phi(h(v))h'(v) \right)'' = \phi'(h(v))(h'(v))^2 + \phi(h(v))h''(v) \]

For $Z < 0$, $h(v) < 0$. Hence $\phi'(h(v)) \geq 0$. Obviously, $\phi(h(v)) \geq 0$. Moreover,

\[ h''(v) = \left[ z \left( \frac{1}{2} \right) \left( \frac{vm}{t} \right) \right]' = -z \left( \frac{vm}{t} \right)^{-\frac{1}{2}} \left( \frac{m}{t} \right)^2 \]
\[ = -z \left( \frac{m}{t} \right)^2 > 0 \]

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 $z \sim N(0, 1)$, $C_{n-1} \sim X^2_{n-1}$. Then $g(a)$ is a convex function in $a$. 


For the one-sided Behrens Fisher problem,

\textbf{Theorem 3.}

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

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

By Jensen Inequality,

\[ p(q) = E_V \left[ f(V) \right] \geq f(E(V)) = f(n - 1) \]

\[ p(q) = \Phi \left( \sqrt{\frac{(n - 1)m}{C_{n-1} A(m + nc)}} \right) \equiv p_1(q) \]

Now observe that under \( \mu_1 - \mu_2 = 0 \), \( z \sim \mathcal{N}(0,1) \), \( C_{m-1} \sim \chi^2_{n-1} \) and \( z, C_{n-1} \) are independent of one another. For \( 0 < r < 0.5 \),

\[ P_q \{ q : p(q) \leq r \} \leq P_q \{ p_1(q) \leq r \} = g(A) \]

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

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

\[ = \Phi \left( r \sqrt{\frac{(n - 1)m}{C_{n-1} A(m + nc)}} \right) \leq r \]

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

where \( k = \sqrt{\frac{m + nc}{m}} \)

\section*{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 known 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.

\textbf{REFERENCES}


