Behrens-Fisher Problem with One Variance Unknown
Sa-aat Niwitpong, Rada Somkhuean, and Suparat Niwitpong

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

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

I. INTRODUCTION

MAITY and Sherman [1] mentioned that the situation of the hypothesis testing for the difference of two normal population means with one variance unknown, arises in practice. For example, when one is interested in comparing a standard treatment with a new treatment. A known variance comes from the standard treatment while an unknown variance comes from the new treatment. Maity and Sherman found that their proposed t-test has more power than the existing Satterthwaite’s test [2], [3]. However, they did not investigate the coverage probability and the expected length of the confidence interval for the difference of two normal population means when one variance is unknown. Niwitpong [4] also derived analytic expressions to find coverage probabilities and expected lengths of the confidence interval using the pivotal statistic t-statistic proposed by Maity and Sherman compared to Welch-Satterthwaite (WS) [5] confidence interval. In this paper, following Weerahandi [6], we propose the generalized p-value to test the hypothesis $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$, where $\theta$ is the parameter of interest, and, $\theta = \mu_x - \mu_y$ and $\theta_0$ is fixed and when one of variance is unknown.

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

Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be random samples from two independent normal distributions with means $\mu_x, \mu_y$ and standard deviations $\sigma_x$ and $\sigma_y$, respectively. Let $\theta = \mu_x - \mu_y$ be the parameter of interest. The problem is to test the hypothesis $H_0 : \theta \leq \theta_0$ against the alternative hypothesis $H_1 : \theta > \theta_0$ for some fixed $\theta_0$. The sufficient statistic of this problem is $(\bar{X}, \bar{Y}, S^2_x, S^2_y)$ (Tsu and Weerahandi [7]) where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \bar{Y} = \frac{1}{m} \sum_{j=1}^{m} Y_j$.

Sa-at 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).
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_i@hotmail.com).
Suparat 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).

$S^2_x = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}$ and $S^2_y = \frac{\sum_{j=1}^{m} (Y_j - \bar{Y})^2}{m}.$

The probability distributions of the statistics, $\bar{X} \sim N(\mu_x, \sigma_x^2/n)$, $\bar{Y} \sim N(\mu_y, \sigma_y^2/m)$, $V = \frac{S^2_x}{\sigma_x^2} / \frac{S^2_y}{\sigma_y^2} \sim \chi^2_{n-1}$ and $U = \frac{mS^2_x}{\sigma_y^2} \sim \chi^2_{m-1}$ are independent of one another. Tsu and Weerahandi [7] proposed the generalized p-value for the above hypothesis as follow:

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

$T(X, Y, x, y, \mu_x, \mu_y, \sigma_x^2, \sigma_y^2) = \bar{X} - \bar{Y} - \theta \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \sqrt{\frac{\sigma^2_x}{nS^2_x} + \frac{\sigma^2_y}{mS^2_y}}$

and $T(x, y, x, y, \mu_x, \mu_y, \sigma_x^2, \sigma_y^2) = x - y - \theta_0$. 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 $Z \sqrt{\frac{\sigma^2_x}{n} + \frac{\sigma^2_y}{m}}$ 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 statistic and $T (X, Y, x, y, \mu_x, \mu_y, \sigma_x^2, \sigma_y^2)$ is defined to be a generalized pivot statistic. $T^* (X, Y, x, y, \mu_x, \mu_y, \sigma_x^2, \sigma_y^2)$ 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 100(1 - $\alpha/2$)% generalized lower confidence limit for $\theta$
is then given by $T(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y)_{1 - \alpha}$, the 100(1 - $\alpha$)th percentiles of $T(x, y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y)$.

Further, given the observed value $x$, let $t_1$ and $t_2$ be such values that $P(t_1 < T(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) < t_2|\theta) = 1 - \alpha$ for chosen significant level $\alpha \in (0, 1)$ than the confidence interval for parameter $\theta$ defined by $\{\theta : t_1 < T(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) < t_2\}$ 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

$$C_{x,y}(\theta, \sigma^2_x, \sigma^2_y) = \{(X, Y) : T(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) - T(x, y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) \geq 0\}.$$

For the one-sided Behrens-Fisher problem, the generalized $p$-value is

$$p^* = Pr(T(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y)$$

$$= Pr\left(Z \sqrt{\frac{\sigma^2_X}{n} + \frac{\sigma^2_Y}{U}} \geq \bar{x} - \bar{y}\right).$$

$$= Pr\left(Z \geq \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma^2_X}{n} + \frac{\sigma^2_Y}{U}}}\right).$$

$$= Pr\left(Z \geq \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{\sigma^2_X}{n} + \frac{\sigma^2_Y}{U}}}\right).$$

$$= E_U\left(\Phi\left(\frac{(\bar{y} - \bar{x})}{\sqrt{\frac{\sigma^2_X}{n} + \frac{\sigma^2_Y}{U}}}\right)\right).$$

(2)

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

Now to find the upper bound of $p(w)$ using the method described by Tang and Tsui [8], we need Theorems 1 and 2 as following.

**Theorem 1.** Define

$$f(u) = \Phi\left(z \sqrt{\frac{1}{t_1 + \frac{t_2}{u}}}\right) \text{ for } u \in (0, 1).$$

Then for fixed $z < 0$, $f(u)$ is a convex function of $u$.

**Proof:** Letting

$$h(u) = z \sqrt{\frac{1}{t_1 + \frac{t_2}{u}}} ,$$

we have $f(u) = \Phi(h(u))$. Let $F(\cdot)$ be the probability density function of the standard normal distribution.

Then

$$f''(u) = (f'(u))' = (\Phi(h(u))h''(u))'$$

$$= \Phi'(h(u))(h''(u))^2 + \Phi(h(u))h''(u)$$

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

$$h''(u) = z\left(\left(-\frac{1}{2}\right)\left(t_1 + \frac{t_2}{u}\right)^{-3/2}\left(-\frac{t_2}{u^2}\right)\right)'$$

$$= z\left[\frac{2}{u^3} t_1(t_1 + \frac{t_2}{u})^{-\frac{5}{2}} + \left(t_1 + \frac{t_2}{u}\right)^{-\frac{3}{2}} - \frac{2t_2}{u^2}\left(t_1 + \frac{t_2}{u}\right)^{-\frac{3}{2}}\right].$$

For the one-sided Behrens-Fisher problem as stated, $H_0 : \theta < \theta_0$ against $H_a : \theta > \theta_0$, we can assume $\theta_0 = 0$ without loss of generality, and the generalized $p$-value for the one-sided Behrens-Fisher problem is $p(w)$ which is

$$Pr(W(X, Y, x, y, \mu_x, \mu_y, \sigma^2_x, \sigma^2_y) \geq w_{obs} | \theta > 0)$$
Hence \( f(u) \geq 0 \), and \( f(u) \) is convex in \( u \).

**Theorem 2.** Let

\[
g(a) = P \left[ \Phi \left( \frac{1}{a + \frac{1}{m-1}} \leq r \right) \right],
\]

where \( z, C_{m-1} \) are independent random variables such that \( z \sim N(0,1) \), \( C_{m-1} \sim \chi^2_{m-1} \). Then \( g(a) \) is a convex function in \( a \).

**Proof:**

\[
g(a) = P \left[ \Phi \left( \frac{1}{a + \frac{1}{m-1}} \leq r \right) \right]
= P \left[ \frac{m-1}{a(m-1) + (1-a)C_{m-1}} \leq \Phi^{-1}(r) \right]
= P \left[ \frac{m-1}{a(m-1) + (1-a)C_{m-1}} \frac{1}{m-1} \right]
= E_{C_{m-1}} \left[ \Phi \left( \frac{a(m-1) + (1-a)C_{m-1}}{m-1} \Phi^{-1}(r) \right) \right]
\]

\( E_{C_{m-1}}(\cdot) \) is an expectation operator with respect to \( C_{m-1} \) with \( (n-1) \) degree of freedom and \( \Phi(.) \) is a cdf of the standard normal distribution, denote

\[
h_1(a) = \sqrt{\frac{a(m-1) + (1-a)C_{m-1}}{m-1}} \Phi^{-1}(r)
\]
and \( g_1(a) = \Phi(h_1(a)) \),

\[
g''_1(a) = (g'_1(a))' = (\Phi(h_1(a)))' h'_1(a) + \Phi(h_1(a)) h''_1(a)
\]

For \( r \leq 0.5 \), \( h_1(a) \leq 0 \), and consequently, \( \phi'(h_1(a)) \geq 0 \).

Moreover,

\[
h''_1(a) = \frac{1}{2} \left( \frac{a(m-1) + (1-a)C_{m-1}}{m-1} \right)^{-\frac{1}{2}} \Phi^{-1}(r) \left( \frac{(m-1) - C_{m-1}}{m-1} \right)^2
\]

\[
\geq 0.
\]

Hence \( g''_1(a) \geq 0 \). That is \( g_1(a) \) is convex in \( a \). As a result, \( g(a) = E_{C_{m-1}}(g_1(a)) \) is convex in \( a \).

**Theorem 3.** For the one-sided Behrens Fisher problem, when one of variation is unknown with \( H_0 : \mu_1 - \mu_2 \leq \theta_0 \) and any \( 0 < r < 0.5 \), the generalized \( p \)-value, \( p(w) \) in (2), has the following property under \( H_0 \):

\[
P_w(p(w) \leq r) < \Phi(\Phi^{-1}(r))
\]

Where \( \Phi(.) \) is a cdf of the standard normal distribution and \( \Phi^{-1}(\cdot) \) is the inverse function of \( \Phi(.) \).

**Proof:** Denote

\[
A = \frac{\mu_1 \nu_1 + \bar{x} \mu_2}{\sqrt{\frac{\mu_1}{n_1} + \frac{\bar{x}^2}{n}}}, \quad C_{m-1} = \frac{m \sigma^2}{\sigma_2^2}
\]

From (2)

\[
p(w) = E_U \left[ \Phi \left( \frac{\tilde{g} - \tilde{x}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n}}} \right) \right]
= E_U \left[ \Phi \left( \frac{\tilde{g} - \tilde{x}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n}}} \right) \right],
\]

\[
= E_U \left[ Z \frac{1}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n} \frac{C_1}{m}}} \right] = E_U \left[ Z \frac{1}{\sqrt{A + \frac{1}{m} (1-A) C_{m-1}}} \right]
\]

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

\[
f(U) = \Phi \left( Z \frac{1}{\sqrt{A + \frac{1}{m} (1-A) C_{m-1}}} \right)
\]

By Jensens Inequality,

\[
p(w) = E_U (f(U)) = f(m-1)
\]

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

\[
P_u(w : p(w) \leq r) \leq P_u (p_1(w) \leq r) = g(A)
\]

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

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

\[
= \max \left\{ P(\Phi (Z \leq r), P \left( Z \frac{1}{\sqrt{C_{m-1}}} \leq \Phi^{-1}(r) \right) \right\}
\]

\[
= \max \left\{ P \left( Z \leq \Phi^{-1}(r) \right), P \left( Z \frac{1}{\sqrt{C_{m-1}}} \leq \Phi^{-1}(r) \right) \right\}
\]

\[
= \max (\Phi(\Phi^{-1}(r)), \Psi_{m-1}(\Phi^{-1}(r))) = \Phi(\Phi^{-1}(r))
\]

where \( \Phi(.) \) is cdf of standard normal distribution.
IV. CONCLUSION

In this paper, we derive an expression of the upper bound of the generalized p-value for the Behrens-Fisher problem with one unknown variance used the method described by Tang and Tsui [8]. This upper bound can be easily computed by R program with command: pnorm(qnorm(r)), when r is a fixed real value between 0 to 0.5.

REFERENCES