Journal of the Korean Statistical Society (1998), **27**: 4, pp 435 449

Confidence Intervals for the Stress-strength Models with Explanatory Variables

Sangyeol Lee[1] and Eunsik Park[2]

ABSTRACT

In this paper, we consider the problem of constructing the lower cofidence intervals for the reliability $P(X < Y|\mathbf{z}, \mathbf{w})$, where the stress X and the strength Y are the random variables with explanatory variables $\mathbf{z}$ and $\mathbf{w}$, respectively. As an estimator of the reliability, a Mann-Whitney type statistic is considered. It is shown that under regularity conditions, the proposed estimator is asymptotically normal. Based on the result, the distribution free lower confidence intervals are constructed.

Keywords: Stress-strength models with explanatory variables, reliabilty; Mann-Whitney type statistic, distribution free confidence intervals

1. INTRODUCTION

In the classical stress-strength reliability formulation, both the environmental stress X and the strength of the unit Y are treated as random, and there is concern for estimating $P(X < Y)$, called the reliability. Since the seminal work of Birnbaum (1956), many statistical methods have been developed for estimating the reliability, and a vast amount of literature is devoted to the study on stress-strength models and their applications. See, for example, Sen (1967), Govindarajulu (1968), Halperin et al. (1987), Johnson (1988), Mee (1990), and the papers cited in these articles.

In most articles, the stress and strength observations are assumed i.i.d. However, Guttman et. al. (1988) considered the stress-strength model where both the stress X and the strength Y have explanatory variables, say, $\mathbf{z}$ and $\mathbf{w}$, respectively. Using the normality assumption, they derived lower confidence bounds for the reliability $P(X < Y|\mathbf{z}, \mathbf{w})$. In practice, however, the normal assumption can

[1]Department of Statistics, Seoul National University, Seoul, 151-742, Korea
[2]Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27514, U.S.A.

be violated. Motivated by this viewpoint, we consider the problem of estimating $P(X < Y|\mathbf{z}, \mathbf{w})$, when X and Y are not necessarily normal.

Suppose that X and Y are related to nonstochastic p and q dimensional explanatory variables $\mathbf{z}$ and $\mathbf{w}$ through the linear relations:

$$X = \mu + \boldsymbol{\beta}'(\mathbf{z} - \overline{\mathbf{z}}) + \delta \quad \text{and} \quad Y = \nu + \boldsymbol{\gamma}'(\mathbf{w} - \overline{\mathbf{w}}) + \epsilon, \tag{1.1}$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)'$ are unknown regression coefficients, and the errors δ and ϵ are mutually independent random variables with the positive densities F and G, respectively, such that $E\delta = E\epsilon = 0$, $Var(\delta) = \sigma^2$ and $Var(\epsilon) = \tau^2$. The objective of this paper is to construct the lower confidence intervals for the reliability $P(X < Y|\,\mathbf{z}, \mathbf{w})$, where $\mathbf{z}, \mathbf{w}$ are given vectors. Here, to estimate it we employ the Mann-Whitney type statistic.

Assume that $(X_i, \mathbf{z}_i)$ and $(Y_j, \mathbf{w}_j), i = 1, \dots, m, j = 1, \dots, n$ are from the models in (1.1), and let $\overline{\mathbf{z}} = m^{-1}\sum_{i=1}^m \mathbf{z}_i$, $\overline{\mathbf{w}} = n^{-1}\sum_{j=1}^n \mathbf{w}_j$, and $P(\boldsymbol{\theta}) = P(X < Y|\,\mathbf{z}, \mathbf{w})$, where $\boldsymbol{\theta} = (\mu, \nu, \boldsymbol{\beta}, \boldsymbol{\gamma})'$. From (1.1), we can write that

$$P(\boldsymbol{\theta}) = P\left(\delta - \epsilon < -\mu + \nu - \boldsymbol{\beta}'(\mathbf{z} - \overline{\mathbf{z}}) + \boldsymbol{\gamma}'(\mathbf{w} - \overline{\mathbf{w}})\right),$$

so that, provided $\mu, \nu, \boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are known, $P(\boldsymbol{\theta})$ can be estimated by

$$\begin{aligned} U(\boldsymbol{\theta}) &= (mn)^{-1}\sum_{i=1}^m\sum_{j=1}^n I\left(\delta_i - \epsilon_j < -\mu + \nu - \boldsymbol{\beta}'(\mathbf{z} - \overline{\mathbf{z}}) + \boldsymbol{\gamma}'(\mathbf{w} - \overline{\mathbf{w}})\right) \\ &= (mn)^{-1}\sum_{i=1}^m\sum_{j=1}^n I\left(\delta_i - \epsilon_j < \kappa\right) = (mn)^{-1}\sum_{i=1}^m\sum_{j=1}^n U_{ij}, \end{aligned} \tag{1.2}$$

where I denotes the indicator function, $\kappa = -\mu + \nu - \boldsymbol{\beta}'(\mathbf{z} - \overline{\mathbf{z}}) + \boldsymbol{\gamma}'(\mathbf{w} - \overline{\mathbf{w}})$ and $U_{ij} = I(\delta_i - \epsilon_j < \kappa)$. Here, since the true errors δ_i and ϵ_j as well as the parameters $\mu, \nu, \boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are unknown, one should estimate these.

Let $\widehat{\boldsymbol{\theta}} = (\widehat{\mu}, \widehat{\nu}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}})'$ be the least squares estimate of $\boldsymbol{\theta}$, and $\widehat{\delta}_i$ and $\widehat{\epsilon}_j$ be the residuals computed by $\widehat{\delta}_i = X_i - \widehat{\mu} - \widehat{\boldsymbol{\beta}}'(\mathbf{z}_i - \overline{\mathbf{z}})$ and $\widehat{\epsilon}_j = Y_j - \widehat{\nu} - \widehat{\boldsymbol{\gamma}}'(\mathbf{w}_j - \overline{\mathbf{w}})$. In view of (1.2), we consider the following as an estimate of $P(\boldsymbol{\theta})$:

$$\begin{aligned} \widehat{U}(\widehat{\boldsymbol{\theta}}) &= (mn)^{-1}\textstyle\sum_{i=1}^m\sum_{j=1}^n I\left(\widehat{\delta}_i - \widehat{\epsilon}_j < -\widehat{\mu} + \widehat{\nu} - \widehat{\boldsymbol{\beta}}'(\mathbf{z} - \overline{\mathbf{z}}) + \widehat{\boldsymbol{\gamma}}'(\mathbf{w} - \overline{\mathbf{w}})\right) \\ &= (mn)^{-1}\textstyle\sum_{i=1}^m\sum_{j=1}^n I\left(\widehat{\delta}_i - \widehat{\epsilon}_j < \widehat{\kappa}\right) = (mn)^{-1}\sum_{i=1}^m\sum_{j=1}^n \widehat{U}_{ij}, \end{aligned} \tag{1.3}$$

where $\widehat{\kappa} = -\widehat{\mu} + \widehat{\nu} - \widehat{\boldsymbol{\beta}}'(\mathbf{z} - \overline{\mathbf{z}}) + \widehat{\boldsymbol{\gamma}}'(\mathbf{w} - \overline{\mathbf{w}})$ and $\widehat{U}_{ij} = I(\widehat{\delta}_i - \widehat{\epsilon}_j < \widehat{\kappa})$.

For constructing the confidence intervals, we first show that $N^{1/2}[\hat{U}(\hat{\theta})-P(\theta)]$, $N = m + n$, is asymptotically normal as N tends to ∞. For doing this, noticing that

$$\widehat{U}_{ij} = I\left(\delta_i - \epsilon_j < \kappa + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{z}_i - \mathbf{z}) - (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'(\mathbf{w}_j - \mathbf{w})\right),$$

we introduce the random variable

$$\widehat{P}(\widehat{\boldsymbol{\theta}}) = (mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n} H\left(\kappa + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{z}_i - \mathbf{z}) - (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'(\mathbf{w}_j - \mathbf{w})\right), \quad (1.4)$$

where $H(x) = P(\delta - \epsilon < x)$. Then we show that under regularity conditions,

$$N^{1/2}\left[\widehat{U}(\widehat{\boldsymbol{\theta}}) - \widehat{P}(\widehat{\boldsymbol{\theta}}) - U(\boldsymbol{\theta}) + P(\boldsymbol{\theta})\right] \longrightarrow_P 0 \text{ as } N \longrightarrow \infty, \quad (1.5)$$

(cf. Theorem 2.1). Next, we show that $N^{1/2}[\hat{U}(\hat{\boldsymbol{\theta}}) - P(\boldsymbol{\theta})]$ is asymptotically normal via establishing the asymptotic normality of $N^{1/2}\left[U(\boldsymbol{\theta})\text{-}P(\boldsymbol{\theta}) + \widehat{P}(\widehat{\boldsymbol{\theta}})\text{-}P(\boldsymbol{\theta})\right]$ (cf. Theorem 2.2). Finally, we propose a consistent estimator for the asymptotic variance of $N^{1/2}\left[\hat{U}(\hat{\boldsymbol{\theta}}) - P(\boldsymbol{\theta})\right]$ (cf. Theorem 3.1). Based on this result, we propose lower confidence intervals for $P(X < Y|\mathbf{z}, \mathbf{w})$ (cf. Theorem 3.2).

2. ASYMPTOTIC NORMALITY

Suppose that as $m, n \to \infty$,

$$\overline{\mathbf{z}} \longrightarrow \xi_1, \quad m^{-1}\sum_{i=1}^{m}(\mathbf{z}_i - \overline{\mathbf{z}})(\mathbf{z}_i - \overline{\mathbf{z}})' \longrightarrow \Gamma_{\mathbf{z}},$$

and (2.1)

$$\overline{\mathbf{w}} \longrightarrow \xi_2, \quad n^{-1}\sum_{j=1}^{n}(\mathbf{w}_j - \overline{\mathbf{w}})(\mathbf{w}_j - \overline{\mathbf{w}})' \longrightarrow \Gamma_{\mathbf{w}},$$

where $\xi_1 \in R^p$, $\xi_2 \in R^q$ and $\Gamma_{\mathbf{z}}$ and $\Gamma_{\mathbf{w}}$ are $p \times p$ and $q \times q$ positive definite matrices, respectively, and

$$\max_{1\le i\le m} ||m^{-1/2}(\mathbf{z}_i - \overline{\mathbf{z}})|| \longrightarrow 0 \text{ and } \max_{1\le j\le n} ||n^{-1/2}(\mathbf{w}_j - \overline{\mathbf{w}})|| \longrightarrow 0. \quad (2.2)$$

Further, assume that as $N \to \infty$,

$$m/N \longrightarrow \rho \quad \text{and} \quad n/N \longrightarrow 1-\rho, \quad 0<\rho<1, \tag{2.3}$$

and that

$$\sup_x \{|F'(x)| + |F''(x)| + |G'(x)| + |G''(x)|\} < \infty. \tag{2.4}$$

Note that (2.4) implies $\sup_x\{|H'(x)|+|H''(x)|\} < \infty$, where H is the distribution function in (1.4). The Condition (2.4) is a fairly mild condition; a broad class of distributions, such as the normal distribution, satisfy (2.4).

In the following, we first prove (1.5) and show the asymptotic normality of $N^{1/2}[\hat{U}(\boldsymbol{\theta}) - P(\boldsymbol{\theta})]$. The following lemmas are useful for proving (1.5).

Lemma 2.1. *Let κ be the number in (1.2). Then, under Conditions (2.1)-(2.4),*

$$C_N = N^{1/2}(mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}[\,F(\epsilon_j + \kappa + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{z}_i - \mathbf{z}) - (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'(\mathbf{w}_j - \mathbf{w}))$$
$$- H(\kappa + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{z}_i - \mathbf{z}) - (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'(\mathbf{w}_j - \mathbf{w})) + H(\kappa) - F(\epsilon_j + \kappa)\,] \longrightarrow_P 0.$$

Proof: We only provide the proof for the case $p = q = 1$ since the proofs of the other cases will follow essentially the same arguments. Rewrite the models in (1.1) as follows:

$$X = \mu + \beta(z - \overline{z}) + \delta \quad \text{and} \quad Y = \nu + \gamma(w - \overline{w}) + \varepsilon.$$

Since $\hat{\beta}$ and $\hat{\gamma}$ are $m^{1/2}$ and $n^{1/2}$-consistent, respectively, by our assumptions, for any $\eta \in (0,1)$, there exists $K>0$, such that for all m, n,

$$P(|m^{1/2}(\widehat{\beta} - \beta)| \le K) > 1-\eta, \quad \text{and} \quad P(|n^{1/2}(\widehat{\gamma} - \gamma)| \le K) > 1-\eta.$$

Hence, putting $S_m(K) = (|m^{1/2}(\widehat{\beta} - \beta)| \le K)$, it suffices to show that $\sup_{|t|\le K}$ $|C_N(t)|\ I(S_m(K)) = o_P(1)$, where

$$\begin{aligned} C_N(t) &= N^{1/2}(mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}[F\left(\epsilon_j + \kappa + (\widehat{\beta} - \beta)(z_i - z) - n^{-1/2}t(w_j - w)\right) \\ &\quad -H\left(\kappa + (\widehat{\beta} - \beta)(z_i - z) - n^{-1/2}t(w_j - w)\right) + H(\kappa) - F(\epsilon_j + \kappa)]. \end{aligned}$$

Let $t_l = -K + 2Kl/N^2, l = 0, \ldots, N^2$, $t_{jl}{}^+ = \sup_{t_l \le t < t_{l+1}} n^{-1/2} t(w - w_j)$ and $t_{jl}{}^- = \inf_{t_l \le t < t_{l+1}} n^{-1/2} t(w - w_j)$. Putting $b_i = (\hat{\beta} - \beta)(z_i - z)$, we have

$$\begin{aligned}
& \sup_{|t| \le K} |C_N(t)|\, I(S_m(K)) \\
\le \;& \max_{0 \le l \le N^2} N^{1/2}(mn)^{-1} \Bigg| \sum_{i=1}^{m} \sum_{j=1}^{n} \Big[F\left(\epsilon_j + \kappa + b_i + t_{jl}{}^+\right) - H\left(\kappa + b_i + t_{jl}{}^+\right) \\
& \qquad + H(\kappa) - F(\epsilon_j + \kappa) \Big] \Bigg| I(S_m(K)) \qquad (2.5) \\
+ \;& \max_{0 \le l \le N^2} N^{1/2}(mn)^{-1} \Bigg| \sum_{i=1}^{m} \sum_{j=1}^{n} \Big[F\left(\epsilon_j + \kappa + b_i + t_{jl}{}^-\right) - H\left(\kappa + b_i + t_{jl}{}^-\right) \\
& \qquad + H(\kappa) - F(\epsilon_j + \kappa) \Big] \Bigg| I(S_m(K)) \qquad (2.6) \\
+ \;& K_1 N^{-3/2},
\end{aligned}$$

where K_1 is a positive constant. We only deal with the argument in (2.5) since (2.6) can be handled in a similar fashion. Let

$$\begin{aligned}
\Pi_{jl} \;=\; & \sum_{i=1}^{m} \left[F\left(\epsilon_j + \kappa + b_i + t_{jl}{}^+\right) - H\left(\kappa + b_i + t_{jl}{}^+\right) + H(\kappa) - F(\epsilon_j + \kappa) \right] \\
& \cdot I(S_m(K)).
\end{aligned}$$

Then, $|\Pi_{jl}| \le m, E(\Pi_{jl}|\delta_1, \ldots, \delta_m) = 0$, and due to (2.1), (2.3) and (2.4),

$$\begin{aligned}
\sum_{j=1}^{n} E(\Pi_{jl}^2|\delta_1, \ldots, \delta_m) \;&\le\; m \sum_{i=1}^{m} \sum_{j=1}^{n} \left| H\left(\kappa + b_i + t_{jl}{}^+\right) - H(\kappa) \right| \, I(S_m(K)) \\
&\le\; m \sum_{i=1}^{m} \sum_{j=1}^{n} \left(|b_i| + \left|t_{jl}{}^+\right| \right) \sup_x |H'(x)|\, I(S_m(K)) \\
&\le\; K_2 N^{5/2} \;\; (K_2 > 0),
\end{aligned}$$

regardless of $\delta_1, \ldots, \delta_m$. Here, applying Bernstein's inequality, we have for all $a > 0$,

$$
\begin{aligned}
P&\left(\max_{0\le l\le N^2} N^{1/2}(mn)^{-1}\left|\sum_{j=1}^{n}\Pi_{jl}\right|\ge a\right)\\
&\le E\sum_{l=0}^{N^2} P\left(N^{1/2}(mn)^{-1}|\sum_{j=1}^{n}\Pi_{jl}|\ge a\middle|\delta_1,\dots,\delta_m\right)\\
&\le 2(N^2+1)\exp\left(-\frac{2^{-1}a^2m^2n^2N^{-1}}{K_2N^{5/2}+3^{-1}am^2nN^{-1/2}}\right)\\
&\le 2(N^2+1)\exp\left(-K_3N^{1/2}\right)\longrightarrow 0,
\end{aligned}
$$

where K_3 is a positive number. This establishes the lemma. □

The following Lemma 2.2 is a direct result of Welsh (1987, Lemma 1). In fact, F and G satisfy the condition (C3)$'$ in Welsh (1987, P. 25).

Lemma 2.2. *Suppose that $\{\mathbf{x}_i\}$ are $p\times 1$ nonstochastic vectors such that as $n\to\infty$,*

$$
n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\to 0, n^{-1/2}\max_{1\le i\le n}||\mathbf{x}_i||\to 0 \quad\text{and}\quad n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\to\Gamma,
$$

where Γ is a $p\times p$ positive definite matrix. Further, assume that $\{e_i\}$ is a sequence of i.i.d. random variables whose distribution function F_e satisfies the condition (C3)$'$ in Welsh. Then, for all $K>0$,

$$
\begin{aligned}
\sup_{-\infty<t<\infty}\ \sup_{||s||\le n^{-1/2}K} n^{-1/2}\Big|\sum_{i=1}^{n}&\big[I(e_i\le t+s'\mathbf{x}_i)-\\
&F_e(t+s'\mathbf{x}_i)+F_e(t)-I(e_i\le t)\big]\Big|\longrightarrow_P 0.
\end{aligned}
$$

The following lemma is due to the arguments in Billingsley (1968, P. 106-108).

Lemma 2.3. *Suppose that $\{e_i\}$ is a sequence of i.i.d. random variables with distribution F_e, $Ee_1=0$ and $Var(e_1)<\infty$, and let $\{a_n\}$ be a sequence of positive random variables decaying to 0 in probability. Then,*

$$
\sup_{|F_e(u)-F_e(v)|\le a_n} n^{-1/2}\left|\sum_{i=1}^{n}[I(e_i\le u)-F_e(u)+F_e(v)-I(e_i\le v)]\right|\longrightarrow_P 0.
$$

Theorem 2.1. *Under Conditions (2.1)-(2.4),*

$$D_N = N^{1/2}\left[\widehat{U}(\widehat{\boldsymbol{\theta}}) - \widehat{P}(\widehat{\boldsymbol{\theta}}) + P(\boldsymbol{\theta}) - U(\boldsymbol{\theta})\right] \longrightarrow_P 0 \quad \text{as } N \longrightarrow \infty.$$

Proof: Let C_N denote the random variable in Lemma 2.1. Write $D_N = A_N + B_N + C_N$, where

$$\begin{aligned} A_N &= N^{1/2}(mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\Big[I\big(\delta_i < \epsilon_j + \kappa + (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\mathbf{z}_i - \mathbf{z}) - (\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})' \\ &\quad (\mathbf{w}_j - \mathbf{w})\big) - F\left(\epsilon_j + \kappa + (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\mathbf{z}_i - \mathbf{z}) - (\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'(\mathbf{w}_j - \mathbf{w})\right) \\ &\quad + F\left(\epsilon_j + \kappa - (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\mathbf{z} - \overline{\mathbf{z}}) - (\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'(\mathbf{w}_j - \mathbf{w})\right) \\ &\quad - I\left(\delta_i < \epsilon_j + \kappa - (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\mathbf{z} - \overline{\mathbf{z}}) - (\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'(\mathbf{w}_j - \mathbf{w})\right) \Big]. \end{aligned}$$

and

$$\begin{aligned} B_N &= N^{1/2}(mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\Big[I\big(\delta_i < \epsilon_j + \kappa - (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\mathbf{z} - \overline{\mathbf{z}}) - (\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})' \\ &\quad (\mathbf{w}_j - \mathbf{w})\big) - F\big(\epsilon_j + \kappa - (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\mathbf{z} - \overline{\mathbf{z}}) - (\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'(\mathbf{w}_j - \mathbf{w})\big) \\ &\quad + F(\epsilon_j + \kappa) - I(\delta_i < \epsilon_j + \kappa)\Big]. \end{aligned}$$

In view of Lemmas 2.2 and 2.3, we have $A_N = o_P(1)$ and $B_N = o_P(1)$. Since C_N goes to zero in probability by Lemma 2.1, the theorem is established. □

The remainder of this section is concerned with the asymptotic normality of $N^{1/2}\left[\hat{U}(\boldsymbol{\theta}) - P(\boldsymbol{\theta})\right]$. Put $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m)'$, $\overline{\mathbf{Z}} = (\overline{\mathbf{z}}', \overline{\mathbf{z}}', \ldots, \overline{\mathbf{z}}')'$, $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n)'$ and $\overline{\mathbf{W}} = (\overline{\mathbf{w}}', \overline{\mathbf{w}}', \ldots, \overline{\mathbf{w}}')'$.

Lemma 2.4. *Let*

$$h^{(1,0)}(x) = 1 - G(x-\kappa) - P(\boldsymbol{\theta}),\ H^{(1,0)} = m^{-1}\sum_{i=1}^{m} h^{(1,0)}(\delta_i),$$

$$h^{(0,1)}(x) = F(x+\kappa) - P(\boldsymbol{\theta}) \text{ and } H^{(0,1)} = n^{-1}\sum_{j=1}^{n} h^{(0,1)}(\epsilon_j).$$

Suppose that (2.1)-(2.4) hold. Then, for $\mathbf{a} \in R^p$ *and* $\mathbf{b} \in R^q$, *as* $N \longrightarrow \infty$,

$$N^{1/2}\left[\mathbf{a}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + H^{(1,0)}\right] \longrightarrow_D N(0, \Sigma_1) \tag{2.7}$$

and

$$N^{1/2}\left[\mathbf{b}'(\widehat{\gamma}-\gamma)+H^{(0,1)}\right]\longrightarrow_D N(0,\Sigma_2), \tag{2.8}$$

where

$$\Sigma_1=\rho^{-1}\left[\mathbf{a}'\Gamma_{\mathbf{z}}^{-1}\mathbf{a}\,\sigma^2+\int(1-G(t-\kappa))^2\,dF(t)-P^2(\boldsymbol{\theta})\right]$$

and

$$\Sigma_2=(1-\rho)^{-1}\left[\mathbf{b}'\Gamma_{\mathbf{w}}^{-1}\mathbf{b}\,\tau^2+\int F^2(t+\kappa)dG(t)-P^2(\boldsymbol{\theta})\right].$$

Proof: We only prove (2.7) since the proof of (2.8) is similar to that of (2.7). Set

$$(c_1,\dots,c_m)=(Nm)^{1/2}\,\mathbf{a}'\left[(\mathbf{Z}-\overline{\mathbf{Z}})'(\mathbf{Z}-\overline{\mathbf{Z}})\right]^{-1}(\mathbf{Z}-\overline{\mathbf{Z}})',$$

$$Y_{mi}=c_i\delta_i,\quad Z_{mi}=N^{1/2}m^{-1/2}h^{(1,0)}(\delta_i)\text{ and }\ X_{mi}=Y_{mi}+Z_{mi}.$$

Putting $Y_m=\sum_{i=1}^m Y_{mi}$, $Z_m=\sum_{i=1}^m Z_{mi}$ and $X_m=\sum_{i=1}^m X_{mi}$, we can write that $Y_m=(Nm)^{1/2}\mathbf{a}'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$, $Z_m=(Nm)^{1/2}\,H^{(1,0)}$ and $X_m=Y_m+Z_m$. Since $N^{1/2}\left[\mathbf{a}'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})+H^{(1,0)}\right]$ equals to $m^{-1/2}\sum_{i=1}^m X_{mi}$, we only have to check that $\{X_{mi}\}$ satisfies Lindeberg's condition.

Note that for all m,i, $|Z_{mi}|\le d$ for some $d>0$, and $EX_{mi}=0$. Denoting $s^2(W)=Var(W)$ for any random variable W, we can write

$$s^2(Y_m)=Nm\,\mathbf{a}'\left[(\mathbf{Z}-\overline{\mathbf{Z}})'(\mathbf{Z}-\overline{\mathbf{Z}})\right]^{-1}\mathbf{a}\,\sigma^2$$

and

$$s^2(Z_m)=N\left[\int(1-G(t-\kappa))^2\,dF(t)-P^2(\boldsymbol{\theta})\right].$$

Since $\sum_{i=1}^m c_i=0$, it follows that $Cov(Y_m,Z_m)=0$ and thus $s^2(X_m)=s^2(Y_m)+s^2(Z_m)$. Now, for every $\theta\ge 2d/s(Y_m)$,

$$\begin{aligned}
&\sum_{i=1}^m\int_{|X_{mi}|\ge\theta s(X_m)}X_{mi}^2dP\,/s^2(X_m)\\
\le\ &\sum_{i=1}^m\int_{|Y_{mi}|\ge\theta s(Y_m)-d}2(Y_{mi}^2+d^2)dP\,/s^2(Y_m)\\
\le\ &2\sum_{i=1}^m\int_{|Y_{mi}|\ge\theta s(Y_m)/2}Y_{mi}^2dP\,/s^2(Y_m)\,+2d^2\sum_{i=1}^m P\left[|Y_{mi}|\ge\theta s(Y_m)/2\right]/s^2(Y_m)
\end{aligned}$$

Note that

$$2\sum_{i=1}^{m}\int_{|Y_{mi}|\geq\theta s(Y_m)/2} Y_{mi}^2 dP\Big/s^2(Y_m) \leq$$
$$2\int_{|x|\geq\theta\sigma(\sum_{i=1}^{m}c_i^2)^{1/2}/(2\max_{1\leq i\leq m}|c_i|)} x^2 dF(x)\Big/\sigma^2,$$

which goes to zero since $\max_{1\leq i\leq m}|c_i|\,/(\sum_{i=1}^{m}c_i^2)^{1/2}$ tends to zero by (2.1) and (2.2). Meanwhile, applying Chebyshev's inequality, we have

$$2d^2\sum_{i=1}^{m}P\left[|Y_{mi}|\geq\theta s(Y_m)/2\right]/s^2(Y_m)\leq 8d^2\,/\theta^2 s^2(Y_m)\longrightarrow 0,$$

which completes the proof. □

Theorem 2.2. *Under Conditions (2.1)-(2.4),*

$$N^{1/2}\left[\widehat{U}(\widehat{\boldsymbol{\theta}})-P(\boldsymbol{\theta})\right]\longrightarrow_D N(0,\Sigma),$$

where

$$\begin{aligned}\Sigma &= \rho^{-1}\left[\int(1-G(t-\kappa))^2\,dF(t)-P^2(\boldsymbol{\theta})+(\mathbf{z}-\xi_1)'\Gamma_{\mathbf{z}}^{-1}(\mathbf{z}-\xi_1)(H'(\kappa))^2\sigma^2\right]\\ &\quad+(1-\rho)^{-1}\left[\int F^2(t+\kappa)dG(t)-P^2(\boldsymbol{\theta})+(\mathbf{w}-\xi_2)'\Gamma_{\mathbf{w}}^{-1}(\mathbf{w}-\xi_2)(H'(\kappa))^2\tau^2\right].\end{aligned}$$

Proof: Write

$$\begin{aligned}N^{1/2}\left[\widehat{U}(\widehat{\boldsymbol{\theta}})-P(\boldsymbol{\theta})\right] &= N^{1/2}\left[\widehat{U}(\widehat{\boldsymbol{\theta}})-\widehat{P}(\widehat{\boldsymbol{\theta}})+P(\boldsymbol{\theta})-U(\boldsymbol{\theta})\right]\\ &+ N^{1/2}\left[\widehat{P}(\widehat{\boldsymbol{\theta}})-P(\boldsymbol{\theta})+U(\boldsymbol{\theta})-P(\boldsymbol{\theta})\right].\end{aligned}$$

The first term of the RHS of the above equality goes to 0 in probability by Theorem 2.1. Since by Taylor's series expansion, (2.1) and (2.2),

$$\Bigg|N^{1/2}(mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\Big[H\left(\kappa+(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\mathbf{z}_i-\mathbf{z})-(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'(\mathbf{w}_j-\mathbf{w})\right)$$
$$-H\left(\kappa-(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\mathbf{z}-\overline{\mathbf{z}})+(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'(\mathbf{w}-\overline{\mathbf{w}})\right)\Big]\Bigg|$$

$$
\begin{aligned}
&\le N^{1/2}(mn)^{-1}\left|\sum_{i=1}^{m}\sum_{j=1}^{n}\left[\left((\widehat{\boldsymbol\beta}-\boldsymbol\beta)'(\mathbf z_i-\overline{\mathbf z})-(\widehat{\boldsymbol\gamma}-\boldsymbol\gamma)'(\mathbf w_j-\overline{\mathbf w})\right)\times\right.\right.\\
&\qquad\left.\left. H'\left(\kappa-(\widehat{\boldsymbol\beta}-\boldsymbol\beta)'(\mathbf z-\overline{\mathbf z})+(\widehat{\boldsymbol\gamma}-\boldsymbol\gamma)'(\mathbf w-\overline{\mathbf w})\right)\right]\right|\\
&+\ 2^{-1}N^{1/2}(mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\left((\widehat{\boldsymbol\beta}-\boldsymbol\beta)'(\mathbf z_i-\overline{\mathbf z})-(\widehat{\boldsymbol\gamma}-\boldsymbol\gamma)'(\mathbf w_j-\overline{\mathbf w})\right)^2\sup_x|H''(x)|\\
&=\ O_P(N^{-1/2}),
\end{aligned}
$$

$$
\begin{aligned}
N^{1/2}\left[\widehat P(\widehat{\boldsymbol\theta})-P(\boldsymbol\theta)\right] &= N^{1/2}\Big[H\left(\kappa-(\widehat{\boldsymbol\beta}-\boldsymbol\beta)'(\mathbf z-\overline{\mathbf z})+(\widehat{\boldsymbol\gamma}-\boldsymbol\gamma)'(\mathbf w-\overline{\mathbf w})\right)\\
&\qquad -H(\kappa)\Big]+O_P(N^{-1/2}).
\end{aligned}
$$

Thus, by Taylor's series expansion,

$$
N^{1/2}\left[\widehat P(\widehat{\boldsymbol\theta})-P(\boldsymbol\theta)\right]=N^{1/2}\left(-(\widehat{\boldsymbol\beta}-\boldsymbol\beta)'(\mathbf z-\overline{\mathbf z})+(\widehat{\boldsymbol\gamma}-\boldsymbol\gamma)'(\mathbf w-\overline{\mathbf w})\right)H'(\kappa)+o_P(1). \tag{2.9}
$$

Meanwhile, using the H-decomposition of U-statistics, we can write that

$$
N^{1/2}\left(U(\boldsymbol\theta)-P(\boldsymbol\theta)\right)=N^{1/2}\left[H^{(1,0)}+H^{(0,1)}\right]+o_P(1), \tag{2.10}
$$

where $H^{(1,0)}$ and $H^{(0,1)}$ are the random variables in Lemma 2.4. Combining (2.9), (2.10) and Lemma 2.4, we establish the theorem. □

Theorem 2.2 implies that to obtain the estimator producing small asymptotic variance, one should choose the design vectors whose arithmetic means are close to given points $\mathbf z$ and $\mathbf w$. The following corollary is a direct result of Theorem 2.2.

Corollary 2.1. *Under the conditions of Theorem 2.2, if* $\mathbf z=\overline{\mathbf z},\mathbf w=\overline{\mathbf w}$, *then*

$$
N^{1/2}\left[\widehat U(\widehat{\boldsymbol\theta})-P(\boldsymbol\theta)\right]\longrightarrow_D N(0,\Sigma'),
$$

where

$$
\begin{aligned}
\Sigma' = \rho^{-1}&\left[\int(1-G(t-\kappa))^2\,dF(t)-P^2(\boldsymbol\theta)\right]\\
&\qquad+(1-\rho)^{-1}\left[\int F^2(t+\kappa)dG(t)-P^2(\boldsymbol\theta)\right].
\end{aligned}
$$

3. DISTRIBUTION FREE CONFIDENCE INTERVALS FOR $P(\boldsymbol{\theta})$

In this section, we propose a weakly consistent estimator for the asymptotic variance in Theorem 2.2, and construct the lower confidence intervals for $P(\boldsymbol{\theta})$ by using the estimator.

Let $p_1 = P(U_{ij}U_{kj} = 1)$ and $p_2 = P(U_{ij}U_{ik} = 1)$. Note that $p_1 = \int F^2(t+\kappa)dG(t)$ and $p_2 = \int (1 - G(t-\kappa))^2 \, dF(t)$.

Lemma 3.1. *Under Conditions (2.1)-(2.4)* ,

$$\widehat{p}_1 = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k\neq i}^{m} \widehat{U}_{ij}\widehat{U}_{kj} \,/mn(m-1) \longrightarrow_P p_1 \tag{3.1}$$

and

$$\widehat{p}_2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k\neq j}^{n} \widehat{U}_{ij}\widehat{U}_{ik} \,/mn(n-1) \longrightarrow_P p_2. \tag{3.2}$$

Proof: Here, we only provide the proof of (3.1). Note that

$$\begin{aligned} Q &= \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k\neq i}^{m} \widehat{U}_{ij}\widehat{U}_{kj} \,/mn(m-1) - p_1 \\ &= \sum_{j=1}^{n}\left[\left(\sum_{i=1}^{m}\widehat{U}_{ij}\right)^2 - \left(\sum_{i=1}^{m}\widehat{U}_{ij}\right)\right] \Big/ mn(m-1) - p_1, \end{aligned}$$

and that $Q = R - S + T$, where

$$\begin{aligned} R &= \sum_{j=1}^{n}\left[\left(\sum_{i=1}^{m}\widehat{U}_{ij}\right)^2 - \left(\sum_{i=1}^{m}U_{ij}\right)^2\right] \Big/ mn(m-1) \\ S &= \sum_{j=1}^{n}\sum_{i=1}^{m}\left(\widehat{U}_{ij} - U_{ij}\right) /mn(m-1) \\ T &= \sum_{j=1}^{n}\left[\left(\sum_{i=1}^{m}U_{ij}\right)^2 - \left(\sum_{i=1}^{m}U_{ij}\right)\right] \Big/ mn(m-1) - p_1. \end{aligned}$$

Now, to prove (3.1), we need to show that R, S and T converge to 0 in probability Let κ be the number in (1.2) and $\kappa_j^* = \kappa - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{z} - \overline{\mathbf{z}}) - (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'(\mathbf{w}_j - \mathbf{w}) + \epsilon_j$

Observe that

$$\begin{aligned}
R &\le (mn(m-1))^{-1}\sum_{j=1}^{n}\left|\sum_{i=1}^{m}\left(\widehat{U}_{ij}+U_{ij}\right)\right|\left|\sum_{i=1}^{m}\left(\widehat{U}_{ij}-U_{ij}\right)\right| \\
&\le 2\,(n(m-1))^{-1}\sum_{j=1}^{n}\left|\sum_{i=1}^{m}\left(\widehat{U}_{ij}-U_{ij}\right)\right| \\
&= 2\,(n(m-1))^{-1}\sum_{j=1}^{n}\left|\sum_{i=1}^{m}\left[I\left(\delta_i<(\widehat{\boldsymbol\beta}-\boldsymbol\beta)'(\mathbf{z}_i-\overline{\mathbf{z}})+\kappa_j^*\right)-I\left(\delta_i<\epsilon_j+\kappa\right)\right]\right| \\
&\le R_1+R_2+R_3,
\end{aligned}$$

where

$$\begin{aligned}
R_1 &= 2\,(n(m-1))^{-1}\sum_{j=1}^{n}\left|\sum_{i=1}^{m}\left[I\left(\delta_i<(\widehat{\boldsymbol\beta}-\boldsymbol\beta)'(\mathbf{z}_i-\overline{\mathbf{z}})+\kappa_j^*\right)\right.\right. \\
&\qquad \left.\left. -F\left((\widehat{\boldsymbol\beta}-\boldsymbol\beta)'(\mathbf{z}_i-\overline{\mathbf{z}})+\kappa_j^*\right)+F(\kappa_j^*)-I(\delta_i<\kappa_j^*)\right]\right| \\
R_2 &= 2(n(m-1))^{-1}\sum_{j=1}^{n}\left|\sum_{i=1}^{m}\left[I(\delta_i<\kappa_j^*)-F(\kappa_j^*)+F(\epsilon_j+\kappa)-I(\delta_i<\epsilon_j+\kappa)\right]\right| \\
R_3 &= 2\,(n(m-1))^{-1}\sum_{j=1}^{n}\left|\sum_{i=1}^{m}\left[F\left((\widehat{\boldsymbol\beta}-\boldsymbol\beta)'(\mathbf{z}_i-\overline{\mathbf{z}})+\kappa_j^*\right)-F(\epsilon_j+\kappa)\right]\right|.
\end{aligned}$$

In view of Lemmas 2.2 and 2.3, we have that $R_1=o_P(1)$ and $R_2=o_P(1)$. Since Taylor's series expansion yields $R_3=o_P(1)$, we have $R=o_P(1)$.

Meanwhile, it is obvious that $S=o_P(1)$. Since by Hoeffding (1948),

$$T=(mn(m-1))^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k\neq i}^{m}U_{ij}U_{kj}-p_1=o_P(1),$$

we have $Q=o_P(1)$. □

Lemma 3.2. *Let*

$$\widetilde{H}(x)=(mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}I(\widehat{\delta}_i-\widehat{\epsilon}_j<x).$$

Under Conditions (2.1)-(2.4), $\widehat{p}_3=(2h)^{-1}\left[\widetilde{H}(\widehat{\kappa}+h)-\widetilde{H}(\widehat{\kappa}-h)\right]\longrightarrow_p H'(\kappa)$, *where* $\widehat{\kappa}$ *is the random variable in (1.3) and* h *is a bandwidth of order* N^b, $-1/2<b<0$.

Proof: Write

$$
\begin{aligned}
& (2h)^{-1}\left[\widetilde{H}(\widehat{\kappa}+h)-\widetilde{H}(\widehat{\kappa}-h)\right]-H^{'}(\kappa) \\
= \; & (2hmn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[I(\widehat{\delta}_i-\widehat{\epsilon}_j<\widehat{\kappa}+h)-I(\widehat{\delta}_i-\widehat{\epsilon}_j<\widehat{\kappa}-h)\right]-H^{'}(\kappa) \\
= \; & \sum_{i=1}^{5}H_i,
\end{aligned}
$$

where

$$
\begin{aligned}
H_1 &= (2hmn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[I(\widehat{\delta}_i-\widehat{\epsilon}_j<\widehat{\kappa}+h)-I(\delta_i-\epsilon_j<\kappa+h)\right] \\
H_2 &= -(2hmn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[I(\widehat{\delta}_i-\widehat{\epsilon}_j<\widehat{\kappa}-h)-I(\delta_i-\epsilon_j<\kappa-h)\right] \\
H_3 &= (2h)^{-1}\left[(mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}I(\delta_i-\epsilon_j<\kappa+h)-H(\kappa+h)\right] \\
H_4 &= -(2h)^{-1}\left[(mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}I(\delta_i-\epsilon_j<\kappa-h)-H(\kappa-h)\right] \\
H_5 &= (2h)^{-1}\left[H(\kappa+h)-H(\kappa-h)\right]-H^{'}(\kappa).
\end{aligned}
$$

It is obvious that $H_i = o_P(1)$, $i = 3, 4, 5$. Following the similar arguments in the proof of Lemma 2.1 and Theorem 2.1, one can readily check that $H_i = o_P(1), i = 1, 2$, which completes the proof. □

Theorem 3.1. *Let* $\widehat{p}_i$, $i = 1, 2, 3$ *be the random variables in Lemmas 3.1 and 3.2, and*

$$\widehat{\sigma}^2 = \sum_{i=1}^{m}\left(X_i-\widehat{\mu}-\widehat{\boldsymbol{\beta}}^{'}(\mathbf{z}_i-\overline{\mathbf{z}})\right)^2/(m-p-1)$$

and

$$\widehat{\tau}^2 = \sum_{j=1}^{n}\left(Y_j-\widehat{\nu}-\widehat{\boldsymbol{\gamma}}^{'}(\mathbf{w}_j-\overline{\mathbf{w}})\right)^2/(n-q-1)$$

Then, under the condition of Lemma 3.1,

$$
\begin{aligned}
\widehat{\Sigma} &= Nm^{-1}\left[\widehat{p}_2-\widehat{U}(\widehat{\boldsymbol{\theta}})^2\right]+N(\mathbf{z}-\overline{\mathbf{z}})'\left[(\mathbf{Z}-\overline{\mathbf{Z}})'(\mathbf{Z}-\overline{\mathbf{Z}})\right]^{-1}(\mathbf{z}-\overline{\mathbf{z}})\widehat{p}_3^2\,\widehat{\sigma}^2 \\
&+ Nn^{-1}\left[\widehat{p}_1-\widehat{U}(\widehat{\boldsymbol{\theta}})^2\right]+N(\mathbf{w}-\overline{\mathbf{w}})'\left[(\mathbf{W}-\overline{\mathbf{W}})'(\mathbf{W}-\overline{\mathbf{W}})\right]^{-1}(\mathbf{w}-\overline{\mathbf{w}})\,\widehat{p}_3^2\,\widehat{\tau}^2
\end{aligned}
$$

is a consistent estimator of Σ *in Theorem 2.2.*

Proof: It is well-known that $\widehat{\sigma}^2 \longrightarrow_P \sigma^2$ and $\widehat{\tau}^2 \longrightarrow_P \tau^2$. Then, the result follows from Lemmas 3.1 and 3.2. □

Here, we construct distribution free lower confidence intervals based on Theorems 2.2 and 3.1.

Theorem 3.2. *Let* $\hat{\Sigma}$ *be the random variable in Theorem 3.1. Under Conditions (2.1)-(2.4),*

$$N^{1/2}(\widehat{U}(\widehat{\boldsymbol{\theta}}) - P(\boldsymbol{\theta})) \leq z_\alpha \widehat{\Sigma}^{1/2}$$

is an approximate $100(1-\alpha)\%$ *distribution free lower confidence interval for* $P(\boldsymbol{\theta})$, *where* z_α *is the upper* $100\alpha\%$ *quantile of the standard normal distribution.*

References

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Birnbaum, Z. W. (1956). "On a use of the Mann-Whitney statistic," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **1**, 13-17.

Govindarajulu, Z. (1968). "Distribution-free confidence bounds for $P(X < Y)$," *Annals of Institute of Statistical Mathematics*, **20**, 229-238.

Guttman, I., Johnson, R. A., Bhattacharyya, G. K. and Reiser, B. (1988). "Confidence limits for stress-strength models with explanatory variables," *Technometrics*, **30**, 161-168.

Halperin, M., Gilbert, P. R. and Lachin, J. M. (1987). "Distribution-free confidence intervals for $Pr(X_1 < X_2)$," *Biometrics,* **43**, 71-80.

Hoeffding, W. (1948). "A class of statistics with asymptotically normal distribution," *The Annals of Mathematical Statististics*, **19**, 293-325.

Johnson, R. A. (1988). "Stress-strength models for reliability," In *Handbook of Statistics* (P. R. Krishnaiah and C. R. Rao eds.) Vol. 7, North-Holland, Amsterdam.

Mee, R. W. (1990). "Confidence intervals for probabilities and tolerance regions based on a generalization of a Mann-Whitney statistic," *Journal of American Statistical Association*, **85**, 793-800.

Sen, P. K. (1967). "A note on asymptotically distribution-free confidence bounds for $P(X < Y)$ based on two independent samples," *Sankhyā,* **A 29**, 95-102.

Welsh, A. H. (1987). "The trimmed mean in the linear model," *The Annals of Statistics*, **15**, 20-36.