AN ASYMPTOTIC EXPANSION FOR A ONE-SIDED RANK TEST IN A TWO-WAY LAYOUT

By

Taka-aki SHIRAISHI

Abstract. In the randomized block design with I blocks and two treatments, the within-block rank test is considered. It is found that the test is asymptotically efficient for a large number of observations per each cell and the asymptotic expansion of that under the null hypothesis is obtained.

§ 1. Introduction

The model is as follows,

\[ X_{ijk} = \mu + \beta_i + \tau_j + e_{ijk} \]

(1)

for \( i = 1, \ldots, I, \ j = 1, 2, \text{ and } k = 1, \ldots, s_j \), where

\[ \sum_{i=1}^{I} \beta_i = 0, \quad \tau_1 + \tau_2 = 0 \text{ and } \{ e_{ijk} : i = 1, \ldots, I, \ j = 1, 2, \ k = 1, \ldots, s_j \} \]

are independent and identically distributed random variables from a distribution function \( F(t) \) with density \( f(t) \).

The null hypothesis is \( H: \tau_1 = \tau_2 = 0 \) and the alternative is \( K: \tau_1 < \tau_2 \). If \( I = 1 \), (1) is the two-sample problem, the locally most powerful rank test exists and it is asymptotically optimum. Then the exact table of significance levels for small samples and special scores is given, the asymptotic normality is followed, and recently Bickel and Zwet [1] and Robinson [6] derived the asymptotic expansion of the test. In the present paper, we will propose the asymptotically optimum rank test and extend Robinson's result to the case of the model (1).

§ 2. Test statistic

In order to simplify the notations, we set \( s = s_1, \ t = s_2, \ s + t = n, \ p = s/n \text{ and } q = 1 - p \). Here define the scores function \( a_n(\cdot) \) by a mapping from \( [1, 2, \ldots, n] \) to \( R^t \) satisfying \( a_n(k) = -a_n(n-k+1) \) for \( k = 1, 2, \ldots, n \) and \( \sum_{k=1}^{n} |a_n(k)|^2 = 1 \), and define within-block rank \( R_{ijk} \) by the rank of \( X_{ijk} \) among the \( i \)-th block \( \{ X_{ijk} : j = 1, 2, k = 1, 2, \ldots, s_j \} \).
\[ \cdots, s_j \]. Then we reject \( H \), if \( S = \sqrt{(n-1)/(lpqn)} \sum_{i=1}^{l} \sum_{k=1}^{s_i} a_n(R_{i,k}) \) is too large. Hence if

\[
a_n(k) = E[-f'(X_n^{(k)})/f(X_n^{(k)})] / \sqrt{\sum_{k=1}^{l} [E[-f'(X_n^{(k)})/f(X_n^{(k)})]]^2}, \tag{2}
\]

it is a locally most powerful within-block rank test and we find the following proposition.

**Proposition.** Let

\[
P(x) = \prod_{i=1}^{l} \prod_{j=1}^{s_i} f(x_{ijk} - \mu - \beta_i) \quad \text{and} \quad Q_n(x) = \prod_{i=1}^{l} \prod_{j=1}^{s_i} f(x_{ijk} - \mu - \beta_i - \lambda_j / \sqrt{n})
\]

where \( \lambda_1 + \lambda_2 = 0 \). Also suppose that \( a_n(k) \) is defined by (2). If the sequence of the joint density functions of \( \{X_{ijk} : i = 1, \cdots, l, j = 1, 2, k = 1, 2, \cdots, s_i \} \) is \( \{Q_n(x)\} \) and

\[
\lim_{n \to \infty} s_n/n = p_j \quad \text{where} \quad p_j > 0,
\]

then

\[
\lim_{n \to \infty} \Pr[S \leq s_n] = \lim_{n \to \infty} \Pr[\log Q_n(x)p(X) \leq t_a], \tag{3}
\]

where \( s_n \) and \( t_a \) are upper \( \alpha \)-percentage points.

**Proof.** As our proposed tests are similar to \( \mu \) and \( \beta_i \), we may assume \( \mu = \beta_i = 0 \) in the model (1). Then from LeCam's third lemma stated in VI 1.4 of [5] and from theorem V 2.1 of [5], it follows that

\[
\log Q_n(x)p(X) = \log \left[ \prod_{i=1}^{l} \prod_{j=1}^{s_i} f(x_{ijk} - \lambda_j / \sqrt{n})/f(x_{ijk}) \right] \xrightarrow{L} N(-l(f)b^2/2, l(f)b^2) \quad \text{under} \quad H
\]

\[
\xrightarrow{L} N(I(f)b^2/2, I(f)b^2) \quad \text{under} \quad \{Q_n(x)\} \text{ probability},
\]

where \( I(f) \) is the Fisher information number and

\[
b^2 = \sum_{j=1}^{s_i} Ip_j(D_j - \sum_{k=1}^{s_i} p_k D_k)^2.
\]

On the other hand, from [5],

\[
X_i = \sum_{k=1}^{s_i} a_n(R_{i,k}) / \sqrt{pq} \xrightarrow{L} N(0,1) \quad \text{under} \quad H
\]

\[
\xrightarrow{L} N(\sqrt{I(f)}b, 1) \quad \text{under} \quad \{Q_n(x)\} \text{ probability}.
\]

Since \( \{X_i : 1 \leq i \leq n\} \) are independent random variables, from theorem 3.2 of [2],

\[
S \xrightarrow{L} N(0,1) \quad \text{under} \quad H
\]

\[
\xrightarrow{L} N(\sqrt{I(f)}b, 1) \quad \text{under} \quad \{Q_n(x)\} \text{ probability}.
\]

Therefore the left and right hands of the equation (3) of the proposition are equal to \( 1 - \Phi(k_a - \sqrt{I(f)}b) \) where \( \Phi(\cdot) \) is the standard normal distribution function.
and $k_{z}$ is the upper $\alpha$-percentage point of $\Phi(\cdot)$. So the result follows.

The right hand of the equation (3) is the asymptotic power of the most powerful test for $H$ versus $K_{n}: \tau_{j} = d_{j}/\sqrt{n}$ as $n \to \infty$. Hence when $\tau_{j}$ is small and $a_{n}(k)$ is defined by (2), the test based on $S$ is efficient. Hájek and Šidák [5] refer to the rank test satisfying this proposition as asymptotically optimum.

§ 3. Asymptotic expansion

In order to investigate the asymptotic expansion of the test based on $S$ under the null hypothesis, we need the following definitions. Let $A_{n} = \sum_{k=1}^{n} |a_{n}(k)|^{r}$, $b_{n} = \max_{1 \leq k \leq n} |a_{n}(k)|$, $p = s/n$, $q = t/n$ and let the notation $\#$ denote the number of elements. Also we set Assumption (I).

**Assumption (I)**

For any $c>0$, there exist $\varepsilon>0$, $c'>0$ and $\delta$ not depending on $n$ and $x$ which satisfy the condition:

for any $x \in \mathbb{R}$,

$$\#(k: |a_{n}(k)|t-x-2r\pi| > \varepsilon \text{ for } r=0, \pm 1, \pm 2, \cdots \text{ and } t \in (cb_{n}, c'A_{n}^{*})] \geq \delta n.$$  

Robinson [6] showed that Wilcoxon and normal scores satisfy Assumption (I). Here we get the theorem.

**Theorem**

If we set

$$G_{n}(x) = \Phi(x) + D^{*} \Phi(x)((1-6pq)A_{n}/(24pqI) - (1-4pq)/(8pqIn))$$

and suppose that Assumption (I) is satisfied, where $\Phi(x)$ is the standard normal distribution function and $D^{*}$ is the fourth differential, then $|Pr\{S \leq x\} - G_{n}(x)| < BA_{n}$ for all $x$, where $B$ is a function of $p$ only.

**Proof.** The characteristic function of $\sqrt{(n-1)/(npq)} \sum_{k=1}^{n} a_{n}(R_{ik})$ is from [3],

$$f_{n}(\ell) = \left(\frac{n}{s}\right)^{r} \sum \exp \left[i\sqrt{(n-1)/(npq)}t(a_{n}(k_{1}) + \cdots + a_{n}(k_{s}))\right]$$

$$= [2\pi B_{n}(p)]^{-1} \int_{-\pi}^{\pi} \prod_{k=1}^{n} \left|q + p e^{i(x(n-1)/\sqrt{npq})a_{n}(k_{j})} \right| e^{i\theta q} d\theta$$

where $\Sigma$ is the summation over all vectors $(k_{1}, \cdots, k_{s})$ with integer elements and

$$1 \leq k_{1} < \cdots < k_{s} \leq n \text{ and } B_{n}(p) = \left(\frac{n}{s}\right)^{p} q^{n-1}.$$  

Here we transform
\[ f_{ns}(t) = (n^2pq)^{1/2} 2\pi B_{ns}(p) \left[ \prod_{k=1}^{n} T_{n}(\phi, t) d\phi \right] \]

where

\[ T_{n}(\phi, t) = \eta_{n}^{-1/2} e^{-t \phi} + \eta_{n}^{1/2} e^{t \phi} \]

and \( \xi_{nk} = n^{-1/2} \phi + \sqrt{n-1} n \tau_{ns}(k) \).

Hence, the characteristic function of \( S \) is \( \{ f_{ns}(t/\sqrt{T}) \} \). Setting

\[ g_{ns}(t) = e^{-T/2} \left[ 1 + t \right] (1 - 6pq) A_{ns} / (24pq) - (1 - 4pq) / (8pq) \]

the distribution function with characteristic function \( \{ g_{ns}(t/\sqrt{T}) \} \) is

\[ H_{ns}(x) = \phi(x) + \sum_{k=1}^{J} D^{4k} \phi(x) \left( \frac{1}{k} \right) (1 - 6pq) A_{ns} / (24pq) - (1 - 4pq) / (8pq) \]

where \( D^{4k} \) is the \( 4k \)-th differential.

Then from XVI. 3 lemma 1 of [4],

\[ |F_{ns}(x) - H_{ns}(x)| \leq 1 \int_{-T}^{T} |f_{ns}(t/\sqrt{T}) - g_{ns}(t/\sqrt{T})| / |t| dt + 24d / (\pi T) \]

where \( d = \sup \{ H_{ns}(x) \} \).

Hence, if we take \( T = c' A_{ns}^{-1} \), from the similar way of getting the equations (17) and (18) of [6], we find that

\[ \int_{-T}^{T} |f_{ns}(t/\sqrt{T}) - g_{ns}(t/\sqrt{T})| / |t| dt \leq B_{1} A_{ns} \]

and

\[ \int_{-T}^{T} |f_{ns}(t/\sqrt{T}) - g_{ns}(t/\sqrt{T})| / |t| dt \leq B_{2} A_{ns} \].

Here we get \( |F_{ns}(x) - H_{ns}(x)| < B_{3} A_{ns} \), where \( B_{1} \) and \( B_{2} \) are functions of \( p \) only.

Now from the Hölder inequality, since \( A_{ns} \approx n^{-1} \) and

\[ A_{4n} = \sum_{k=1}^{n} |a_{n}(k)|^{4} \leq \left( \sum_{k=1}^{n} |a_{n}(k)|^{2} \right)^{1/2} \left( \sum_{k=1}^{n} |a_{n}(k)|^{2} \right)^{1/4} = A_{2n}^{2/3} \]

we have \( A_{4n} / n \leq A_{2n}^{2} \leq A_{ns} \).

Therefore, we get the theorem.

**References**


An Asymptotic Expansion for A One-sided Rank Test


Institute of Mathematics
University of Tsukuba
Ibaraki, 305 Japan