TSUKUBA J. MATH.
Vol. 8 No. 1 (1984). 119–123

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 asympotic expansion of that under the null hypothesis is obtained.

§1. Introduction

The model is as follows,

$$X_{ijk}=\mu+\beta_i+\tau_j+e_{ijk} \tag{1}$$

for $i=1, \cdots, I$, $j=1,2$, and $k=1, \cdots, s_j$, where

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

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 $s=s_1$, $t=s_2$, $s+t=n$, $p=s/n$ and $q=1-p$. Here define the scores function $a_n(\cdot)$ by a mapping from $\{1,2, \cdots, n\}$ to R^1 satisfying $a_n(k)=-a_n(n-k+1)$ for $k=1,2, \cdots, n$ 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 $\{X_{ijk}: j=1,2,\ k=1,2,$

Received April 8, 1983. Received July 25, 1983.

$\cdots, s_j\}$. Then we reject H, if $S=\sqrt{(n-1)/(Ipqn)}\sum_{i=1}^{I}\sum_{k=1}^{s}a_n(R_{i1k})$ is too large. Hence if

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

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

PROPOSITION. *Let*

$$P(\underline{x})=\prod_{i=1}^{I}\prod_{j=1}^{2}\prod_{k=1}^{s_j}f(x_{ijk}-\mu-\beta_i) \quad \textit{and} \quad Q_{n\Delta}(\underline{x})=\prod_{i=1}^{I}\prod_{j=1}^{2}\prod_{k=1}^{s_j}f(x_{ijk}-\mu-\beta_i-\Delta_j/\sqrt{n})$$

where $\Delta_1+\Delta_2=0$. *Also suppose that* $a_n(k)$ *is defined by* (2). *If the sequence of the joint density functions of* $\{X_{ijk}: i=1,\cdots,I,\ j=1,2,\ k=1,2,\cdots,s_j\}$ *is* $\{Q_{n\Delta}(\underline{x})\}$ *and* $\lim_{n\to\infty} s_j/n=p_j$ *where* $p_j>0$, *then*

$$\lim_{n\to\infty} Pr\{S\geqq s_\alpha\}=\lim_{n\to\infty} Pr\{\log\{Q_{n\Delta}(\underline{X})/P(\underline{X})\}\geqq t_\alpha\}\,, \tag{3}$$

where s_α *and* t_α *are upper* α*-percentage points.*

PROOF. As our proposed tests are similar to μ and β_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

$$\begin{aligned}\log\{Q_{n\Delta}(\underline{X})/P(\underline{X})\}&=\log\left[\prod_{i=1}^{I}\prod_{j=1}^{2}\prod_{k=1}^{s_j}f(X_{ijk}-\Delta_j/\sqrt{n})/f(X_{ijk})\right]\\ &\xrightarrow[L]{} N(-I(f)b^2/2,\, I(f)b^2) \qquad \text{under } H\\ &\xrightarrow[L]{} N(I(f)b^2/2,\, I(f)b^2) \qquad \text{under } \{Q_{n\Delta}(\underline{x})\} \text{ probability,}\end{aligned}$$

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

$$b^2=\sum_{j=1}^{2}Ip_j\Big(\Delta_j-\sum_{k=1}^{2}p_k\Delta_k\Big)^2 .$$

On the other hand, from [5],

$$\begin{aligned}X_i=\sum_{k=1}^{s}a_n(R_{i1k})/\sqrt{pq} &\xrightarrow[L]{} N(0,1) \qquad \text{under } H\\ &\xrightarrow[L]{} N(\sqrt{I(f)/I}\,b,1) \qquad \text{under } \{Q_{n\Delta}\} \text{ probability.}\end{aligned}$$

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

$$\begin{aligned}S &\xrightarrow[L]{} N(0,1) \qquad \text{under } H\\ &\xrightarrow[L]{} N(\sqrt{I(f)}b,1) \qquad \text{under } \{Q_{n\Delta}\} \text{ probability.}\end{aligned}$$

Therefore the left and right hands of the equation (3) of the proposition are equal to $1-\Phi(k_\alpha-\sqrt{I(f)}b)$ where $\Phi(\cdot)$ is the standard normal distribution function and k_α is the upper α-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 $K_n: \tau_j=\Delta_j/\sqrt{n}$ as $n\to\infty$. Hence when τ_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 $A_{rn}=\sum_{k=n}^{n}|a_n(k)|^r$, $b_n=\max_{1\leqq k\leqq n}|a_n(k)|$, $p=s/n$, $q=t/n$ 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 δ not depending on n and x which satisfy the condition:

for any $x\in R^1$,

$$\#\{k:|a_n(k)t-x-2r\pi|>\varepsilon \text{ for } r=0, \pm1, \pm2, \cdots \text{ and } t\in(cb_n^{-1}, c'A_{5n}^{-1})\}\geqq\delta n\,.$$

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

THEOREM

If we set

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

and suppose that Assumption (I) is satisfied, where $\Phi(x)$ is the standard normal distribution function and D^4 is the fourth differential, then $|Pr\{S\leqq x\}-G_{ns}(x)|<BA_{5n}$ for all x, where B is a function of p only.

PROOF. The characteristic function of $\sqrt{(n-1)/(npq)}\sum_{k=1}^{s}a_n(R_{i_1k})$ is from [3],

$$f_{ns}(t)=\binom{n}{s}^{-1}\sum\exp\left[i\sqrt{(n-1)/(npq)}t\{a_n(k_1)+\cdots+a_n(k_s)\}\right]$$
$$=[2\pi B_{ns}(p)]^{-1}\int_{-\pi}^{\pi}\prod_{k=1}^{n}[q+pe^{i(\sqrt{(n-1)/(npq)}ta_n(k)+\theta)}]e^{-i\theta s}d\theta$$

where $\sum$ is the summation over all vectors $(k_1,\cdots,k_s)$ with integer elements and

$$1\leqq k_1<\cdots k_s\leqq n \quad\text{and}\quad B_{ns}(p)=\binom{n}{s}p^sq^{n-s}\,.$$

Here we transform

$$f_{ns}(t)=[(npq)^{1/2}2\pi B_{ns}(p)]^{-1}\int_{-\pi(npq)^{1/2}}^{\pi(npq)^{1/2}}\prod_{k=1}^{n}\rho_k(\psi,t)d\psi\,,$$

where

$$\rho_k(\psi,t)=qe^{-ip\xi_{nk}(pq)^{-1/2}}+pe^{iq\xi_{nk}(pq)^{-1/2}}\quad\text{and}\quad \xi_{nk}=n^{-1/2}\psi+\sqrt{(n-1)/n}\,ta_n(k)\,.$$

Hence the characteristic function of S is $\{f_{ns}(t/\sqrt{I})\}^I$. Setting

$$g_{ns}(t)=e^{-t^2/2}[1+t^4\{(1-6pq)A_{4n}/(24pq)-(1-4pq)/(8pqn)\}]\,,$$

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

$$H_{ns}(x)=\Phi(x)+\sum_{k=1}^{I}D^{4k}\Phi(x)\binom{I}{k}\{(1-6pq)A_{4n}/(24I^2pq)-(1-4pq)/(8I^2pqn)\}^k$$

where D^{4k} is the $4k$-th differential.

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

$$\begin{aligned}|F_{ns}(x)-H_{ns}(x)|&\leqq 1/\pi\int_{-T}^{T}|\{f_{ns}(t/\sqrt{I})\}^I-\{g_{ns}(t/\sqrt{I})\}^I/t|dt+24d/(\pi T)\\&\leqq I/\pi\int_{-T}^{T}|f_{ns}(t/\sqrt{I})-g_{ns}(t/\sqrt{I})|/|t|dt+24d/(\pi T)\,,\end{aligned}$$

where $d=\sup_x\{H'_{ns}(x)\}$.

Hence if we take $T=c'A_{5n}^{-1}$, from the similar way of getting the equations (17) and (18) of [6], we find that

$$\int_{-T}^{-T^{-1}}+\int_{T^{-1}}^{T}|f_{ns}(t/\sqrt{I})-g_{ns}(t/\sqrt{I})|/|t|dt\leqq B_1A_{5n}$$

and

$$\int_{-T^{-1}}^{T^{-1}}|f_{ns}(t/\sqrt{I})-g_{ns}(t/\sqrt{I})|/|t|dt\leqq B_2A_{5n}\,.$$

Here we get $|F_{ns}(x)-H_{ns}(x)|<B_3A_{5n}$, where B_1, B_2 and B_3 are functions of p only. Now from the Hölder inequality, since $A_{4n}\geqq n^{-1}$ and

$$A_{4n}=\sum_{k=1}^{n}|a_n(k)|^4\leqq\left\{\sum_{k=1}^{n}|a_n(k)|^2\right\}^{1/3}\left\{\sum_{k=1}^{n}|a_n(k)|^5\right\}^{2/3}=A_{5n}^{2/3}\,,$$

we have $A_{4n}/n\leqq A_{4n}^2\leqq A_{5n}$.

Therefore we get the theorem.

References

[1] Bickel, P. J. and van Zwet, W. R., Asymptotic expansions for the power of distribution-free tests in the two sample problem. Ann. Statist. **6** (1978), 937–1004.

[2] Billingsley, P., Convergence of Probability Measures. New York: Wiley, (1968).

[3] Erdös, P. and Rényi, A., On the central limit theorem for samples from a finite population. Publi. Math. Inst. Hungar. Acad. Sci. **4** (1959), 49–61.
[4] Feller, W., An Introduction to Probability Theory and its Applications, 2. New York: Wiley, (1966).
[5] Hájek, J. and Šidák, Z., Theory of Rank Tests. New York: Academic Press, (1967).
[6] Robinson, J., An asymptotic expansion for samples from a finite population. Ann. Statist. **6** (1978), 1005–1011.

Institute of Mathematics
University of Tsukuba
Ibaraki, 305 Japan