SA

stichting mathematisch centrum

MC

AFDELING MATHEMATISCHE STATISTIEK   SW 31/74   JUNE

Y. LEPAGE

A CLASS OF NONPARAMETRIC TESTS FOR LOCATION AND SCALE PARAMETERS

Prepublication

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

AMS (MOS) subject classification scheme (1970): 62G10, 62G30

A CLASS OF NONPARAMETRIC TESTS FOR LOCATION AND SCALE PARAMETERS*)

by

YVES LEPAGE**)

ABSTRACT

A class of nonparametric two-sample tests for testing identity of distributions versus alternatives containing both location and scale parameters is proposed. The asymptotic distribution of the statistics of the class is given under both the hypothesis and a contiguous sequence of alternatives. Some asymptotic optimality properties are deduced for particular tests of the class and finally, the asymptotic efficiency is found.

---

*) This work is part of the author's Ph.D. dissertation written at the Université de Montréal under the direction of Professor Constance van Eeden and it was partially supported by the National Research Council of Canada under Grant No.A-8555. The manuscript was completed while the author was holding a postdoctoral fellowship from the National Research Council of Canada and visiting the Mathematisch Centrum, Amsterdam. The paper is not for review; it is meant for publication in a journal.

**) Université de Montréal; temporarily: Mathematisch Centrum, Amsterdam, The Netherlands.

1. INTRODUCTION

The purpose of this paper is to propose and study a class of nonparametric two-sample tests for testing the null hypothesis of randomness versus alternatives containing both location and scale parameters. In section 2, the statistics of the class are defined by a quadratic combination of two linear rank statistics.

Section 3 contains the asymptotic distribution of the statistics of the class under both the null hypothesis and a contiguous sequence of alternatives. In section 4, it is proved that the asymptotically maximin most powerful test belongs to the proposed class. Finally, the asymptotic efficiency of the tests of the class relative to the asymptotically maximin most powerful test is given in section 5.

2. THE CLASS OF TESTS

Let $(m_\nu, m_\nu), \nu = 1,2,\ldots,$ be a sequence of pairs of positive integers such that $N_\nu = m_\nu + m_\nu \to \infty$ when $\nu \to \infty$. For each ν, consider a sequence of random variables $X_{\nu 1},\ldots,X_{\nu N_\nu}$ and denote by $R_{\nu i}$ the rank of $X_{\nu i}$ among $X_{\nu 1},\ldots,X_{\nu N_\nu}$.

Suppose that under H_ν, the random variables $X_{\nu 1},\ldots,X_{\nu N_\nu}$ are independently and identically distributed according to a continuous distribution and that under the alternatives, the joint density of $X_{\nu 1},\ldots,X_{\nu N_\nu}$ is given by

$$(2.1)\qquad q_\nu = \prod_{i=1}^{m_\nu} e^{-c_{\nu i}} f(e^{-c_{\nu i}} x_i - d_{\nu i}) \prod_{i=m_\nu+1}^{N_\nu} f(x_i)$$

with $c_{\nu i} = \Delta_1 (m_\nu n_\nu / N_\nu)^{-\frac{1}{2}}$ and $d_{\nu i} = \Delta_2 (m_\nu n_\nu / N_\nu)^{-\frac{1}{2}}$, $i=1,\ldots,m_\nu$, where $(\Delta_1, \Delta_2) \in \mathbb{R}^2$ and a known density f in the class C of absolutely continuous density functions on $\mathbb{R}$ such that

$$(2.2)\qquad I(f) = \int_0^1 \phi^2(u,f)du < \infty \quad \text{and} \quad I_1(f) = \int_0^1 \phi_1^2(u,f)du < \infty,$$

where if F(x) is the distribution function corresponding to f(x), (2.3) $$\phi(u,f) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))} \text{ and } \phi_1(u,f) = -1 - F^{-1}(u)\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}, \quad 0 < u < 1.$$

Define for k = 1,2, the linear rank statistics

(2.4) $$S_{\nu k} = \sum_{i=1}^{m_\nu} a_{\nu k}(R_{\nu i}),$$

where $a_{\nu k}(1),\ldots,a_{\nu k}(N_\nu)$ are the values of a score function $a_{\nu k}(\cdot)$. Further, let

(2.5) $$\mu_{\nu k} = E(S_{\nu k}|H_\nu) = \frac{m_\nu}{N_\nu}\sum_{i=1}^{N_\nu} a_{\nu k}(i) = m_\nu \bar{a}_{\nu k}$$

and

(2.6) $$\sigma^2_{\nu k} = \mathrm{Var}(S_{\nu k}|H_\nu) = \frac{m_\nu n_\nu}{N_\nu(N_\nu-1)}\sum_{i=1}^{N_\nu}(a_{\nu k}(i)-\bar{a}_{\nu k})^2.$$

Finally, the class of statistics that we will consider is of the form

(2.7) $$U_\nu = \sum_{k=1}^{2}\left(\frac{S_{\nu k}-\mu_{\nu k}}{\sigma_{\nu k}}\right)^2.$$

It should be noted that if $a_{\nu 1}(i) = i$ and $a_{\nu 2}(i) = (N_\nu+1)/2-|i-(N+1)/2|$, $i=1,\ldots,N_\nu$, the statistics U_ν reduce to the combined Wilcoxon and Ansari-Bradley statistics studied by LEPAGE (1971).

3. ASYMPTOTIC DISTRIBUTION

A sequence of score functions $a_\nu(\cdot), \nu = 1,2,\ldots,$ is said to be generated by a real valued function $\phi(u)$, $0 < u < 1$, if

(i) $\int_0^1 \phi^2(u)\,du < \infty$ and $\int_0^1 (\phi(u)-\bar{\phi})^2\,du > 0$ where $\bar{\phi} = \int_0^1 \phi(u)\,du$

and,

(ii) $\lim_{\nu\to\infty}\int_0^1 (a_\nu(1+[uN_\nu]) - \phi(u))^2 du = 0$ with $[uN_\nu]$ denoting the largest integer not exceeding uN_ν.

This condition will be needed to obtain the asymptotic distribution under H_ν and q_ν of the statistics U_ν given by (2.7).

Theorem 3.1. *Suppose that for* k=1,2, *the sequence of score functions* $a_{\nu k}(\cdot), \nu = 1,2,\ldots,$ *is generated by* $\phi_k(u)$, $0 < u < 1$, *with*

$$\int_0^1 (\phi_2(u)-\bar{\phi}_2)(\phi_2(u)-\bar{\phi}_2)du = 0. \tag{3.1}$$

Then, under H_ν, *the statistics* U_ν *are, for* $\min(m_\nu,n_\nu) \to \infty$ *when* $\nu \to \infty$, *asymptotically* χ^2 *- distributed with* 2 *degrees of freedom.*

Proof. Let $(\lambda_1,\lambda_2) \in \mathbb{R}^2 - \{(0,0)\}$ and define $T_\nu = \sum_{i=1}^{m_\nu} a(R_{\nu i})$ where for $i = 1,\ldots,N_\nu$, $a_\nu(i) = \lambda_{\nu 1}\, a_{\nu 1}(i) + \lambda_{\nu 2}\, a_{\nu 2}(i)$ with

$$\lambda_{\nu k} = \lambda_k\Big(\frac{1}{N_\nu}\sum_{i=1}^{N_\nu}(a_{\nu k}(i)-\bar{a}_{\nu k})^2\Big)^{-\frac{1}{2}}, \tag{3.2}$$

k=1,2. From HÁJEK & ŠIDÁK (1967), p.163, it follows that under H_ν, the statistics T_ν are asymptotically normal $(0,\lambda_1^2+\lambda_2^2)$ and consequently, the proof is complete. □

Theorem 3.2. *Suppose that for* k=1,2, *the sequence of score functions* $a_{\nu k}(\cdot), \nu = 1,2,\ldots,$ *is generated by* $\phi_k(u)$, $0 < u < 1$, *with relation* (3.1) *satisfied. Then, under* q_ν, *the statistics* U_ν *are, for* $\Delta_1 \neq 0$ *and* $\min(m_\nu,n_\nu) \to \infty$ *when* $\nu \to \infty$ *asymptotically noncentral* χ^2 *- distributed with* 2 *degrees of freedom and non-centrality parameter*

$$\delta^2 = \sum_{k=1}^{2}\Big[\Big(\int_0^1 \phi_k(u)(\Delta_2\phi(u,f)+\Delta_1\phi_1(u,f))du\Big)^2 \Big/ \int_0^1 (\phi_k(u)-\bar{\phi}_k)^2 du\Big]. \tag{3.3}$$

Proof. Consider the statistics T_ν defined in the proof of theorem 3.1. From theorem 5.1 of LEPAGE (1973b), one can deduce that under q_ν, the statistics T_ν are asymptotically normal $(\mu_\nu,\lambda_1^2+\lambda_2^2)$ where

$$\mu_\nu = \sum_{k=1}^{2} [\lambda_k (\int_0^1 \phi_k(u)(\Delta_2\phi(u,f)+\Delta_1\phi_1(u,f))du)^2 \Big/ \int_0^1 (\phi_k(u)-\bar{\phi}_k)^2 du]. \tag{3.4}$$

Consequently, the proof follows. □

4. ASYMPTOTIC OPTIMALITY

Let C_1 be the class of density functions of C such that

$$\int_0^1 \phi(u,f)\phi_1(u,f)du = 0. \tag{4.1}$$

It should be observed that if $f \in C$ is symmetric with respect to the origin, then $f \in C_1$. Further, define the alternatives $K_\nu(b)$, $b > 0$, by the joint density q_ν given by (2.1) with $f \in C_1$, $\Delta_1 \neq 0$ and

$$\Delta_2^2\, I(f) + \Delta_1^2\, I_1(f) = b^2. \tag{4.2}$$

In the next theorem, an asymptotically optimal test for H_ν versus $K_\nu(b)$ will be found in the class U_ν given by (2.7). The $(1-\alpha)$-quantile of the χ^2-distribution with k degrees of freedom will be denoted $\chi^2_{k,\alpha}$ while $F_k(\cdot,\delta^2)$ will represent the distribution function of the non-central χ^2-distribution with non-centrality parameter δ^2 and k degrees of freedom.

Theorem 4.1. *Suppose that* $\min(m_\nu,n_\nu) \to \infty$ *and that for* $k=1,2$ *the sequence of score functions* $a_{\nu k}(\cdot)$, $\nu = 1,2,\ldots$, *are generated by* $\phi(u,f)$ *and* $\phi_1(u,f)$, $0 < u < 1$, *respectively. Then, the test based on* U_ν *with critical region*

$$U_\nu \geq \chi^2_{2,\alpha} \tag{4.3}$$

is an asymptotically maximin most powerful test for H_ν *versus* $K_\nu(b)$ *at level* α. *Furthermore, the asymptotic power is given by*

$$1 - F_2(\chi^2_{2,\alpha}, b^2). \tag{4.4}$$

Proof. The fact that the asymptotic power of the U_ν-test equals $1 - F_2(\chi^2_{2,\alpha}, b^2)$ follows immediately from theorem 3.2. Consequently, if $\beta(\alpha, H_\nu, K_\nu(b))$ denotes the power of the maximin most powerful test (among all tests), it remains to prove that

$$(4.5) \qquad \lim_{\nu\to\infty} \beta(\alpha, H_\nu, K_\nu(b)) = 1 - F_2(\chi^2_{2,\alpha}, b^2).$$

Suppose that (4.5) does not hold. Thus, passing to a subsequence if necessary, we may assume that

$$(4.6) \qquad \lim_{\nu\to\infty} \beta(\alpha, H_\nu, K_\nu(b)) > 1 - F_2(\chi^2_{2,\alpha}, b^2).$$

By introducing the random variables

$$(4.7) \qquad \xi_{\nu 1} = \left(\frac{N_\nu}{m_\nu n_\nu}\right)^{\frac{1}{2}} (I(f))^{-\frac{1}{2}} S_{\nu 1} \quad , \quad \xi_{\nu 2} = \left(\frac{N_\nu}{m_\nu n_\nu}\right)^{\frac{1}{2}} (I_1(f))^{-\frac{1}{2}} S_{\nu 2}$$

and the constants

$$(4.8) \qquad \theta_1 = \Delta_2 (I(f))^{\frac{1}{2}} \quad , \quad \theta_2 = \Delta_1 (I_1(f))^{\frac{1}{2}}$$

one can, in view of theorem 4.1 of LEPAGE (1973b), proceed in a similar way as HÁJEK & ŠIDÁK (1967), p.256-258, and contradict corollary 1 of LEPAGE (1973a). □

Corollary 4.1. *Under the hypothesis of theorem 4.1, the test based on* U_ν *with critical region* (4.3) *is an asymptotically uniformly maximin most powerful test for* H_ν *versus* $\bigcup_{b>0} K_\nu(b)$ *at level* α.

Proof. In view of relation (4.4) and the definition of uniformly maximin most powerful tests (see HÁJEK & ŠIDÁK (1967),p.29), the result is immediate. □

Corollary 4.2. *In theorem 4.1 and corollary 4.1, the densities* q_ν *given by* (2.1) *can be replaced by*

$$(4.9)\qquad q'_\nu = \prod_{i=1}^{m_\nu} \exp(-\Delta_1(m_\nu n_\nu/N_\nu)^{-\frac{1}{2}}) f(\exp(-\Delta_1(m_\nu n_\nu/N_\nu)^{-\frac{1}{2}})(x_i-\Delta_2(m_\nu n_\nu/N_\nu)^{-\frac{1}{2}})) \prod_{i=m_\nu+1}^{N_\nu} f(x_i).$$

Proof. In view of corollary 5.1 of LEPAGE (1973b), the proof of theorem 4.1 still holds and consequently, corollary 4.1 also. □

Corollary 4.3. *In theorem 4.1 and corollary 4.1, the densities* q_ν *given by* (2.1) *can be replaced by*

$$(4.10)\qquad q_{\nu,\omega} = \prod_{i=1}^{m_\nu} \exp(-\Delta_1(m_\nu n_\nu/N_\nu)^{-\frac{1}{2}}-\omega) f(\exp(-\Delta_1(m_\nu n_\nu/N_\nu)^{-\frac{1}{2}}-\omega)x_i-\Delta_2(m_\nu n_\nu/N_\nu)^{-\frac{1}{2}}) \prod_{i=m_\nu+1}^{N_\nu} e^{-\omega} f(e^{-\omega}x_i)$$

where $\omega \in \mathbb{R}$ *is unknown. Furthermore, the test based on* U_ν *with critical region given by* (4.3) *is then an asymptotically uniformly maximin most powerful* α *level test for* H_ν *versus* $\bigcup_{\omega\in\mathbb{R}} K_\nu(b)$.

Proof. In view of corollary 5.2 of LEPAGE (1973b), the proof of theorem 4.1 still holds and consequently, the result follows. □

5. ASYMPTOTIC EFFICIENCY

Let U_ν^0 represent the asymptotically maximin most powerful test of theorem 4.1 and consider the tests based on U_ν given by (2.7) with critical region $U_\nu \geq \chi^2_{2,\alpha}$.

Theorem 5.1. *Consider testing* H_ν *versus* q_ν *given by* (2.1) *with* $f \in C_1$ *and suppose that for* k=1,2, *the sequence of score functions* $a_{\nu k}(\cdot), \nu = 1,2,\ldots,$ *is generated by* $\phi_k(u)$, $0 < u < 1$, *with relation* (3.1) *satisfied. Then, if* $\Delta_1 \neq 0$ *and* $\min(m_\nu,n_\nu) \to \infty$ *when* $\nu \to \infty$, *the asymptotic efficiency of* U_ν *relative to* U_ν^0, *denoted* $e(U_\nu,U_\nu^0)$, *is given by*

$$(5.1)\qquad e(U_\nu,U_\nu^0) = \sum_{k=1}^{2} \frac{\left(\int_0^1 \phi_k(u)(\Delta_2\phi(u,f)+\Delta_1\phi_1(u,f))du\right)^2}{\left(\int_0^1 (\phi_k(u)-\bar{\phi}_\kappa)^2 du\right)\ (\Delta_2^2 I(f)+\Delta_1^2 I_1(f))} .$$

Proof. It follows from theorem 3.2 that the asymptotic power of the test based on U_ν is given by $1 - F_2(\chi^2_{2,\alpha}, \Delta_2^2 I(f)+\Delta_1^2(f))$. Thus, in view of relation (3.3) and ANDREWS (1954), the proof is complete. □

It should be noted that if $\phi_1(u) = \phi(u,f_1)$ and $\phi_2(u) = \phi_1(u,f_2)$, $0 < u < 1$, with $f_1,f_2 \in C_1$ and

$$(5.2)\qquad \int_0^1 \phi(u,f_1)\phi_1(u,f)du = \int_0^1 \phi(u,f)\phi_1(u,f_2)du = 0,$$

the asymptotic efficiency given by (5.1) can be written as

$$(5.3)\qquad e(U_\nu,U_\nu^0) = \frac{\Delta_2^2 I(f)e(f_1,f)+\Delta_1^2 I_1(f)e_1(f_2,f)}{\Delta_1^2 I(f)+\Delta_1^2 I_1(f)} ,$$

where

$$(5.4)\qquad e(f_1,f) = \frac{\left(\int_0^1 \phi(u,f_1)\phi(u,f)du\right)^2}{I(f_1)I(f)} \text{ and } e_1(f_2,f) = \frac{\left(\int_0^1 \phi_1(u,f_2)\phi_1(u,f)du\right)^2}{I_1(f_2)I_1(f)}$$

correspond, respectively to the asymptotic efficiency of the test based on $S_{\nu 1}$ where the sequence of score functions $a_{\nu 1}(\cdot), \nu = 1,2,\ldots,$ is generated by $\phi(u,f_1)$, $0 < u < 1$, relative to the asymptotically most powerful rank test for contiguous location alternatives for a density f and to the asymptotic efficiency of the test based on $S_{\nu 2}$ where the sequence of score functions $a_{\nu 2}(\cdot), \nu = 1,2,\ldots,$ is generated by $\phi_1(u,f_2)$, $0 < u < 1$, relative to the asymptotically most powerful rank test for contiguous scale alternatives for a density f (see HÁJEK & ŠIDÁK (1967),p.267-270).

Furthermore, from (5.3), it can be seen that for all $(\Delta_1,\Delta_2) \in \mathbb{R}^2$,

$$(5.5)\qquad \min(e(f_1,f),e_1(f_2,f)) \le (U_\nu,U_\nu^0) \le \max(e(f_1,f),e_1(f_2f)).$$

Acknowledgements. I would like to mention my sincere gratitude to Professor Constance van Eeden for suggesting me this problem and also for her constructive comments.

REFERENCES

ANDREWS, F.C. (1954). Asymptotic behavior of some rank tests for analysis of variance. *Ann.Math.Statist.* 25, 724-736.

HÁJEK, J. & ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.

LEPAGE, Y. (1971). A combination of Wilcoxon's and Ansari-Bradley's statistics. *Biometrika* 58, 213-217.

LEPAGE, Y. (1973a). A maximin test for means. *Statistica Neerlandica* 27, 55-57.

LEPAGE, Y. (1973b). *Asymptotically optimum rank tests for contiguous location and scale alternatives*. Report SW 20, Mathematisch Centrum, Amsterdam.