

**stichting
mathematisch
centrum**



AFDELING MATHEMATISCHE STATISTIEK

SW 27/74

MARCH

Y. LEPAGE
TESTING FOR k-SAMPLE LOCATION AND
SCALE ALTERNATIVES, I

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.

TESTING FOR k -SAMPLE LOCATION AND SCALE ALTERNATIVES, I ^{*)}

by YVES LEPAGE ^{**)}

ABSTRACT

In a k -sample case ($k \geq 2$), the problem of testing identity of distribution versus alternatives containing both location and scale parameters is studied. A contiguous sequence of alternatives is constructed and for those alternatives, an asymptotically most powerful rank test is found.

1. INTRODUCTION

The purpose of this work is to derive an asymptotically most powerful linear rank test for the k -sample ($k \geq 2$) problem where the distributions are differing both in their location and scale parameters.

A contiguous sequence of alternatives is constructed and the asymptotic distribution of linear rank statistics under such contiguous alternatives is found by specializing the results of Beran (1970). A rank test asymptotically most powerful among all tests is also deduced in a similar way as Hájek and Šidák (1967).

2. ASYMPTOTIC DISTRIBUTION

Let N_ν ($\nu=1,2,\dots$) be a sequence of positive integers such that $N_\nu \rightarrow \infty$ when $\nu \rightarrow \infty$. For $\nu=1,2,\dots$, let $(A_{\nu 1}, \dots, A_{\nu k})$, $k \geq 2$, be a partition of $\{1, \dots, N_\nu\}$ and put $n_{\nu j} = \text{card } A_{\nu j}$, $j=1, \dots, k$. Moreover, for each ν consider

^{*)} This work was done while the author was holding a postdoctoral fellowship from the National Research Council of Canada and visiting the Mathematisch Centrum, Amsterdam. It was also partially supported by the National Research Council of Canada under Grant No. A-8555. This paper is not for review, it is meant for publication in a journal.

^{**)} Université de Montréal; temporarily: Mathematisch Centrum, Amsterdam.

a sequence of random variables X_{v1}, \dots, X_{vN_v} and denote by R_{vi} the rank of X_{vi} among X_{v1}, \dots, X_{vN_v} ; $i=1, \dots, N_v$.

Suppose that under the hypothesis H_v , the random variables X_{v1}, \dots, X_{vN_v} are independently and identically distributed according to a continuous distribution function and suppose that under the alternatives K_v , the joint density of X_{v1}, \dots, X_{vN_v} is given by

$$(2.1) \quad q_v = \prod_{j=1}^k \prod_{i \in A_{vj}} e^{-c_j/\sqrt{N_v}} f(e^{-c_j/\sqrt{N_v}} x_i - d_j/\sqrt{N_v})$$

with $\underline{c} = (c_1, c_2, \dots, c_k)' \in \mathbb{R}^k$, $\underline{d} = (d_1, d_2, \dots, d_k)' \in \mathbb{R}^k$, $c_1 = d_1 = 0$ and at least one of the vectors \underline{c} or \underline{d} non null, and a density function f which satisfies the following condition:

Condition A.

Let $\Theta \subseteq \mathbb{R}^2$ be an open subset containing $(0,0)$ and for $\underline{\theta} = (\theta_1, \theta_2)' \in \Theta$, put

$$(2.2) \quad f(x, \underline{\theta}) = e^{-\theta_1} f(e^{-\theta_1} x - \theta_2) .$$

- (i) For almost all x , $f(x, \underline{\theta})$ is continuously differentiable with respect to $\underline{\theta}$ whenever $\underline{\theta} \in \Theta$.
- (ii) If $\|\cdot\|$ represents the usual Euclidean norm,

$$(2.3) \quad \lim_{\|\underline{\theta}\| \rightarrow 0} \int_{-\infty}^{\infty} \left[\left(\frac{\partial f(x, \underline{\theta})}{\partial \theta_1} \right)^2 / f(x, \underline{\theta}) \right] dx = I_1(f) < \infty$$

and

$$(2.4) \quad \lim_{\|\underline{\theta}\| \rightarrow 0} \int_{-\infty}^{\infty} \left[\left(\frac{\partial f(x, \underline{\theta})}{\partial \theta_2} \right)^2 / f(x, \underline{\theta}) \right] dx = I(f) < \infty$$

with

$$(2.5) \quad I_1(f) = \int_0^1 \phi_1^2(u, f) du \quad \text{and} \quad I(f) = \int_0^1 \phi^2(u, f) du$$

where, if F is the distribution function corresponding to f ,

$$(2.6) \quad \phi_1(u, f) = -1 - F^{-1}(u) \frac{f'(F^{-1}(u))}{f(F^{-1}(u))} \quad \text{and} \quad \phi(u, f) = - \frac{f'(F^{-1}(u))}{f(F^{-1}(u))},$$

$$0 < u < 1.$$

This regularity condition on the densities is the adaptation for a location and a scale parameter alternative of Condition A of Beran (1970). One can easily verify that the normal, the logistic and the Cauchy densities satisfy Condition A but the exponential, the double exponential and the double quadratic ($f(x) = \frac{1}{2}(1+|x|)^{-2}$) densities don't since from Nickerson, Spencer and Steenrod (1959), p.146, the continuous differentiability of $f(x, \underline{\theta})$ is equivalent to the existence and continuity of the column vector of first partial derivatives with respect to $\underline{\theta}$, $(\partial f(x, \underline{\theta})/\partial \theta_1, \partial f(x, \underline{\theta})/\partial \theta_2)'$. Also, if f satisfies Condition A, we conclude from Lemma 3.3 of Beran (1970), that

$$(2.7) \quad \int_0^1 \phi_1(u, f) du = \int_0^1 \phi(u, f) du = 0.$$

For simplicity of notation, let for $i \in A_{vj}$, $j=1, \dots, k$,

$$(2.8) \quad \underline{\theta}_{vi} = (c_j/\sqrt{N_v}, d_j/\sqrt{N_v})', \quad v=1, 2, \dots,$$

and

$$(2.9) \quad \bar{\underline{\theta}}_v = \frac{1}{N_v} \sum_{i=1}^{N_v} \underline{\theta}_{vi}.$$

Consider now the linear rank statistics

$$(2.10) \quad S_v = \sum_{i=1}^{N_v} \underline{\gamma}_{vi}' \underline{a}_v(R_{vi})$$

where $\underline{\gamma}_{v1}, \dots, \underline{\gamma}_{vN_v}$ are vectors and $\underline{a}_v(1), \dots, \underline{a}_v(N_v)$ are the values of a vector score function $\underline{a}_v(\cdot)$.

We will say that a sequence of vector score functions $\underline{a}_v(\cdot)$, $v=1,2,\dots$, is generated by a vector valued function $\underline{\phi}(u)$, $0 < u < 1$, if

$$(i) \quad \int_0^1 \underline{\phi}'(u) \underline{\phi}(u) du < \infty \quad \text{and} \quad \int_0^1 (\underline{\phi}(u) - \bar{\underline{\phi}})' (\underline{\phi}(u) - \bar{\underline{\phi}}) du > 0 \quad \text{where} \quad \bar{\underline{\phi}} = \int_0^1 \underline{\phi}(u) du .$$

$$(ii) \quad \lim_{v \rightarrow \infty} \int_0^1 \|\underline{a}_v(1 + [uN_v]) - \underline{\phi}(u)\|^2 du = 0 \quad \text{with } [uN_v] \text{ denoting the largest integer not exceeding } uN_v .$$

In Beran (1970), one can find methods for constructing vector score functions that are generated by a given vector function $\underline{\phi}(u)$, $0 < u < 1$.

Further, for an ordered sample $U_v^{(1)} < \dots < U_v^{(N_v)}$ from the uniform distribution on $[0,1]$, we will let

$$(2.11) \quad \underline{a}_v(i, f) = \begin{bmatrix} E \phi_1(U_v^{(i)}, f) \\ E \phi(U_v^{(i)}, f) \end{bmatrix} = \begin{bmatrix} a_{1v}(i, f) \\ a_v(i, f) \end{bmatrix}, \quad i=1, \dots, N_v .$$

One can easily show that if f satisfies Condition A then, the sequence of vector score functions $\underline{a}_v(\cdot, f)$, $v=1,2,\dots$, is generated by

$$(2.12) \quad \underline{\phi}(u, f) = \begin{bmatrix} \phi_1(u, f) \\ \phi(u, f) \end{bmatrix}, \quad 0 < u < 1 .$$

More generally, if for $j=1,\dots,k$ the sequence of score functions $a_v^{(j)}(\cdot)$, $v=1,2,\dots$, is generated by $\phi^{(j)}(u)$, $0 < u < 1$, then the sequence of vector score functions $\underline{a}_v(\cdot) = (a_v^{(1)}(\cdot), \dots, a_v^{(k)}(\cdot))'$, $v=1,2,\dots$, is generated by the vector valued function $\underline{\phi}(u) = (\phi^{(1)}(u), \dots, \phi^{(k)}(u))'$, $0 < u < 1$.

The usual regularity condition on the vectors of constants $\gamma_{v1}, \dots, \gamma_{vN_v}$ is represented by

Condition E.

$$\text{If } \bar{\gamma}_v = \frac{1}{N_v} \sum_{i=1}^{N_v} \gamma_{vi} ,$$

$$(i) \quad \text{for } v=1,2,\dots, \sum_{i=1}^{N_v} \|\gamma_{vi} - \bar{\gamma}_v\|^2 > 0 .$$

$$(ii) \quad \lim_{v \rightarrow \infty} \sum_{i=1}^{N_v} \|\gamma_{vi} - \bar{\gamma}_v\|^2 / \max_{1 \leq i \leq N_v} \|\gamma_{vi} - \bar{\gamma}_v\|^2 = \infty .$$

The following theorem gives the asymptotic distribution of linear rank statistics under the hypothesis H_v . The proof is omitted since it is a direct consequence of Theorem 2.3 of Beran (1970).

Theorem 2.1. *Let the sequence of vector score functions $\underline{a}_v(\cdot)$, $v=1,2,\dots$, be generated by a vector function $\underline{\phi}(u)$, $0 < u < 1$, and assume that Condition E is satisfied. Then, under H_v , the statistics S_v , given by (2.10), are asymptotically normal (μ_v, σ_v^2) with*

$$(2.13) \quad \mu_v = \sum_{i=1}^{N_v} \gamma_{vi}' \bar{\underline{\phi}}$$

and

$$(2.14) \quad \sigma_v^2 = \sum_{i=1}^{N_v} (\gamma_{vi} - \bar{\gamma}_v)' D (\gamma_{vi} - \bar{\gamma}_v)$$

where

$$(2.15) \quad D = \int_0^1 (\underline{\phi}(u) - \bar{\underline{\phi}}) (\underline{\phi}(u) - \bar{\underline{\phi}})' du .$$

In the next theorem, the contiguity of the alternatives K_v with respect to the hypothesis H_v is established.

Theorem 2.2. *Suppose that $\lim_{v \rightarrow \infty} n_{vj}/N_v = \lambda_j$ for $j=1,\dots,k$. Then, if f satisfies Condition A, K_v are contiguous to H_v .*

Proof. Let $p_v = \prod_{i=1}^{N_v} f(x_i)$. From Hájek and Šidák (1967), p.202, it is sufficient to show that the densities $\{q_v\}$ are contiguous to the densities $\{p_v\}$.

We have that

$$(2.16) \quad \max_{1 \leq i \leq N_v} \|\theta_{vi}\|^2 = \max_{2 \leq j \leq k} \left(\frac{c_j^2 + d_j^2}{N_v} \right) \rightarrow 0 \quad \text{when } v \rightarrow \infty,$$

$$(2.17) \quad \sum_{i=1}^{N_v} \|\theta_{vi}\|^2 = \sum_{j=2}^k \frac{n_{vj}}{N_v} (c_j^2 + d_j^2) \leq \sum_{j=2}^k (c_j^2 + d_j^2) < \infty \quad (v=1,2,\dots)$$

and,

$$(2.18) \quad \begin{aligned} \sum_{i=1}^{N_v} \theta'_{vi} \left[\int_0^1 \phi(u, f) \phi(u, f)' du \right] \theta_{vi} &= \\ &= \sum_{j=2}^k \frac{n_{vj}}{N_v} (c_j^2 I_1(f) + 2c_j d_j \int_0^1 \phi_1(u, f) \phi(u, f) du + d_j^2 I(f)) \\ &\rightarrow \sum_{j=2}^k \lambda_j \int_0^1 (c_j \phi_1(u, f) + d_j \phi(u, f))^2 du < \infty \quad \text{when } v \rightarrow \infty. \end{aligned}$$

Thus, since by hypothesis f satisfies Condition A, we conclude from Theorem 3.1 of Beran (1970) that $\{q_v\}$ are contiguous to $\{p_v\}$ and the proof is complete. \square

The last theorem of this section gives the asymptotic distribution of linear rank statistics under the contiguous sequence of alternatives K_v .

Theorem 2.3. Let the sequence of vector score functions $\underline{a}_v(\cdot)$, $v=1,2,\dots$, be generated by a vector function $\underline{\phi}(u)$, $0 < u < 1$, and assume that f satisfies Condition A, and Condition E is verified. Then, under K_v , the statistics S_v , given by (2.10), are asymptotically normal (η_v, σ_v^2) with

$$(2.19) \quad \eta_v = \sum_{i=1}^{N_v} (\underline{y}_{vi} - \bar{\underline{y}}_v)' B(\theta_{vi} - \bar{\theta}_v) + \sum_{i=1}^{N_v} \underline{y}'_{vi} \bar{\underline{\phi}}$$

where $B = \int_0^1 \phi(u) \phi(u, f)' du$ and σ_v^2 given by (2.14).

Proof. From the proof of Theorem 2.2, we have that $\max_{1 \leq i \leq N_v} \|\theta_{vi}\|^2 \rightarrow 0$ when $v \rightarrow \infty$, $\sum_{i=1}^{N_v} \|\theta_{vi}\|^2 < \infty$ ($v=1,2,\dots$) and, by hypothesis, the density f satisfies Condition A of Beran (1970). Thus, the result is obtained by applying Theorem 3.2 of Beran (1970). \square

3. ASYMPTOTIC OPTIMALITY

The following theorem establishes an asymptotically optimum rank test among the class of all possible tests.

Theorem 3.1. Consider testing H_v versus q_v given by (2.1) with a density f satisfying Condition A. Then, if $\lim_{v \rightarrow \infty} n_{vj}/N_v = \lambda_j$, $0 < \lambda_j < 1$, for $j=1,\dots,k$, the test based on

$$(3.1) \quad S_v^0 = \sum_{i=1}^{N_v} \theta_{vi}' a_v(R_{vi}, f)$$

with critical region

$$(3.2) \quad S_v^0 \geq k_{1-\alpha} \cdot b$$

where $k_{1-\alpha}$ is the $(1-\alpha)$ -quantile of the standardized normal distribution and

$$(3.3) \quad b^2 = \sum_{j=2}^k \lambda_j \int_0^1 (c_j \phi_1(u, f) + d_j \phi(u, f))^2 du + \\ - \int_0^1 \left(\sum_{j=2}^k \lambda_j (c_j \phi_1(u, f) + d_j \phi(u, f)) \right)^2 du ,$$

is an asymptotically most powerful test for H_v versus q_v at level α . Furthermore, the asymptotic power is given by $1 - \Phi(k_{1-\alpha} - b)$ where $\Phi(\cdot)$ is

the distribution function of the standardized normal distribution.

Proof. Denote by $\beta(\alpha, H_v, q_v)$ the power of the most powerful test for H_v versus q_v at level α , and let $p_v = \prod_{i=1}^{N_v} f(x_i)$. It is clear that

$$(3.4) \quad \beta(\alpha, H_v, q_v) \leq \beta(\alpha, p_v, q_v) .$$

Moreover, from Theorem 3.1 of Beran (1970) and since

$$(3.5) \quad \lim_{v \rightarrow \infty} \sum_{i=1}^{N_v} (\theta_{vi} - \bar{\theta}_v)' \left[\int_0^1 \phi(u, f) \phi(u, f)' du \right] (\theta_{vi} - \bar{\theta}_v) = b^2 > 0$$

because $\int_0^1 \phi(u, f) \phi(u, f)' du$ is a positive definite 2×2 matrix, we have that $\log(q_v/p_v)$ is asymptotically normal $(-\frac{1}{2}b^2, b^2)$ under p_v and, from relation (3.40) of Beran (1970), Le Cam's third lemma (see Hájek and Šidák (1967), p.208) and Theorem 2.2, $\log(q_v/p_v)$ is asymptotically normal $(\frac{1}{2}b^2, b^2)$ under q_v . Consequently, the most powerful test for p_v versus q_v at level α has the following asymptotic power:

$$(3.6) \quad \lim_{v \rightarrow \infty} \beta(\alpha, H_v, q_v) = 1 - \Phi(k_{1-\alpha} - b) .$$

On the other hand, since the vectors $\theta_{v1}, \dots, \theta_{vN_v}$ satisfy condition E, we get from Theorem 2.3 that the statistics S_v^0 are asymptotically normal (b^2, b^2) under q_v . Thus, the asymptotic power of a test based on S_v^0 with critical region (3.2) is given by $1 - \Phi(k_{1-\alpha} - b)$ and therefore

$$(3.7) \quad \liminf_{v \rightarrow \infty} \beta(\alpha, H_v, q_v) \geq 1 - \Phi(k_{1-\alpha} - b) .$$

The rest follows by combining (3.4), (3.6) and (3.7). \square

Corollary 3.1. In Theorem 3.1, the densities q_v can be replaced by

$$(3.8) \quad q'_v = \prod_{j=1}^k \prod_{i \in A_{vj}} e^{-c_j/\sqrt{N_v}} f\left(e^{-c_j/\sqrt{N_v}} (x_i - d_j/\sqrt{N_v})\right).$$

Proof. Define for $i \in A_{vj}$, $j=1, \dots, k$,

$$(3.9) \quad \Delta_{vi} = \left(c_j/\sqrt{N_v}, e^{-c_j/\sqrt{N_v}} d_j/\sqrt{N_v} \right)'$$

One can easily verify that $\max_{1 \leq i \leq N_v} \|\Delta_{vi}\|^2 \rightarrow 0$ when $v \rightarrow \infty$ and

$$(3.10) \quad \sum_{i=1}^{N_v} \|\Delta_{vi}\|^2 \leq \sum_{j=2}^k (c_j^2 + d_j^2 e^{2c})$$

with $c = \max_{2 \leq j \leq k} |c_j|$. Thus, from Theorem 3.2 of Beran (1970), the linear rank statistics S_v^0 given by (3.1) are, under q'_v , asymptotically normal (b^2, b^2) since

$$(3.11) \quad \lim_{v \rightarrow \infty} \sum_{i=1}^{N_v} (\Delta_{vi} - \bar{\Delta}_v)' \left[\int_0^1 \phi(u, f) \phi(u, f)' du \right] (\Delta_{vi} - \bar{\Delta}_v) = b^2.$$

The rest follows in the same way as in Theorem 3.1. \square

Corollary 3.2. In Theorem 3.1, if the densities q_v are replaced by

$$(3.12) \quad q_{v, \omega} = \prod_{j=1}^k \prod_{i \in A_{vj}} e^{-(c_j/\sqrt{N_v} + \omega_1)} f\left(e^{-(c_j/\sqrt{N_v} + \omega_1)} x_i - (d_j/\sqrt{N_v} + \omega_2)\right)$$

where $\omega = (\omega_1, \omega_2) \in \mathbb{R}^2$ is unknown, then, the test based on S_v^0 given by (3.1) with critical region (3.2) is an asymptotically uniformly most powerful α level test for H_v versus

$$(3.13) \quad \{q_{v, \omega} : \omega \in \mathbb{R}^2\}.$$

Proof. Define for $i \in A_{vj}$, $j=1, \dots, k$,

$$(3.14) \quad \Delta_{vi} = (c_j/\sqrt{N_v} + \omega_1, d_j/\sqrt{N_v} + \omega_2)' .$$

Since $\Delta_{vi} - \bar{\Delta}_v = \theta_{vi} - \bar{\theta}_v$, the result is deduced by an argument similar as for the Theorem 3.1. \square

Corollary 3.3. *The results of Theorem 3.1 and Corollaries 3.1, 3.2 still hold if the score vector functions $\underline{a}_v(\cdot, f)$ are replaced by score vector functions $\underline{a}_v(\cdot)$ generated by $\phi(u, f)$, $0 < u < 1$.*

Proof. In view of Theorem 2.3, the result is immediate. \square

ACKNOWLEDGEMENTS

I would like to thank Professor Constance van Eeden of the Université de Montréal for a careful reading of a first draft of this paper.

REFERENCES

- Beran, R.J. (1970). *Linear rank statistics under alternatives indexed by a vector parameter*. Ann. Math. Statist. 41, 1896-1905.
- Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.
- Nickerson, H.K., Spencer, D.C. and Steenrod, N.E. (1959). *Advanced Calculus*. Van Nostrand, Princeton.