ASYMPTOTIC EFFICIENCY OF MULTIVARIATE NORMAL SCORE TEST[1]

By G. K. Bhattacharyya[2]

University of California, Berkeley

1. Introduction and summary. Let $X_\alpha = (X_{1\alpha}, X_{2\alpha}, \cdots, X_{p\alpha})$, $\alpha = 1, 2, \cdots, m$, and $\alpha = m + 1, \cdots, N$, be independent random samples of sizes m and $n = N - m$ from continuous p-variate cdf $\Psi^{(1)}(\mathbf{x})$ and $\Psi^{(2)}(\mathbf{x})$ respectively. For testing $H_0: \Psi^{(1)} = \Psi^{(2)}$ against shift alternatives $\Psi^{(1)}(\mathbf{x} - \mathbf{\Delta}) = \Psi^{(2)}(\mathbf{x})$ various tests are available in literature of which the important ones are the following:

(a) the classical Hotelling's T^2 test;

(b) the multivariate version of the Wilcoxon test proposed by Chatterjee and Sen ([4], [5]). The test statistic W is a quadratic form in the vector of coordinatewise Wilcoxon statistics.

(c) The multivariate version of the normal score test proposed by Bhattacharyya [1] and for the general multisample problem by Tamura [9] and Sen and Puri [8]. The test statistic M is a quadratic form in coordinatewise normal score statistics. Both M and W are members of the class of tests (4.7) of Tamura [9].

For consideration of asymptotic relative efficiency (ARE) let $\Psi(\mathbf{x})$ denote the common cdf under H_0 and $\mathbf{\Delta}_N = \boldsymbol{\delta}/N^{\frac{1}{2}}$, $(\boldsymbol{\delta} \neq \mathbf{0})$ a sequence of shift alternatives tending to H_0 along the direction $\boldsymbol{\delta}$. Using $e_{A:B}(\boldsymbol{\delta}, \Psi)$ as a general notation for the Pitman efficiency of a test A relative to a test B, which typically depends on Ψ and $\boldsymbol{\delta}$, we have under suitable regularity conditions (c.f. Theorem 3, Tamura [9])

$$e_{M:W}(\boldsymbol{\delta}, \Psi) = \Delta_M \Delta_W^{-1}, \qquad e_{M:T}(\boldsymbol{\delta}, \Psi) = \Delta_M \Delta_T^{-1} \tag{1.1}$$

where

$$\Delta_M = \boldsymbol{\delta}\Lambda^{-1}\boldsymbol{\delta}', \qquad \Delta_W = \boldsymbol{\delta}\Gamma^{-1}\boldsymbol{\delta}', \qquad \Delta_T = \boldsymbol{\delta}\Sigma^{-1}\boldsymbol{\delta}' \tag{1.2}$$

and $\Lambda = (\lambda_{ij})$, $\Gamma = (\gamma_{ij})$, $\Sigma = (\rho_{ij}\sigma_i\sigma_j)$ are nonsingular $p \times p$ matrices. Σ is the covariance matrix of Ψ. Denoting by ψ_i and Ψ_i the ith marginal density and cdf of Ψ, by Ψ_{ij} the joint (ij)th marginal cdf and by ϕ and Φ the density and cdf of standard normal distribution, the typical elements of Λ and Γ are given by:

$$\lambda_{ij} = \rho^*_{ij}\theta_i^{-1}\theta_j^{-1} \quad \text{and} \quad \gamma_{ij} = \rho'_{ij}\gamma_j\gamma_i \tag{1.3}$$

where

$$\theta_i = \int_{-\infty}^{\infty} [\psi_i^2(x)\,dx/\phi\{\Phi^{-1}[\Psi_i(x)]\}], \qquad \gamma_i = [(12)^{\frac{1}{2}} \int_{-\infty}^{\infty} \psi_i^2(x)\,dx]^{-1}; \tag{1.4}$$

Received 6 July 1966; revised 15 February 1967.

[1] This paper was prepared with the partial support of the U. S. Army Research Office (Durham), Grant DA-ARO(D)-31-124-54.8 and the U. S. Navy through the Office of Naval Research under contract Nonr 1202 (17), Project No. NR 042 222.

[2] Now at the University of Wisconsin, Madison.

(1.5) $\rho_{ij}^* = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi^{-1}[\Psi_i(x)]\Phi^{-1}[\Psi_j(y)]\, d\Psi_{ij}(x, y)$;

(1.6) $\rho_{ij}' = 12\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Psi_i(x)\Psi_j(y)\, d\Psi_{ij}(x, y) - 3.$

(ρ_{ij}), (ρ_{ij}') and (ρ_{ij}^*) are respectively the product moment, the grade, and the normal score correlation matrices of Ψ. In this paper the ARE properties of the M test relative to the W and T^2 tests are studied by investigating the bounds of the efficiencies (1.1) for various important classes of multivariate distributions. Since the ARE expressions (1.1) for two sample tests are essentially of the same structure as those for their multisample analogues, no generality is lost, as far as efficiency bounds are concerned, by restricting discussions to the two sample situation. It is shown that in the class of nonsingular multivariate normal distributions the M test has efficiency 1 relative to T^2 and its efficiency relative to W exceeds 1 irrespective of the direction $\boldsymbol{\delta}$. The M test behaves very well when the parent distribution has marginal densities dropping down to zero discontinuously at either tail and also in gross error models when heavy-tails are present in the contaminating distribution.

2. Multivariate normal and other special cases. Taking $\Psi = \Phi_p(\mathbf{0}, \boldsymbol{\Sigma})$, the p-variate nonsingular normal cdf it can be easily verified from (1.4) and (1.5) that $\theta_i = \sigma_i^{-1}$, $\rho_{ij}^* = \rho_{ij}$ and consequently $e_{M:T}(\boldsymbol{\delta}, \Psi) = 1$. In fact, it is sufficient to assume that Ψ has normal marginals. It shows that when the parent distribution is nonsingular and has normal marginals the ARE property of the univariate normal score test vs. the t-test is preserved in the multivariate extensions. This property is not shared by W as can be seen from the result (5.6) of Bickel [3]. Regarding the relative performance between M and W tests the following theorem shows that for multivariate normal Ψ the M test behaves better than W in all directions $\boldsymbol{\delta}$.

THEOREM 2.1. *If ϕ_p denote the family of all nonsingular p-variate normal distributions*

(2.1) $$\inf_{\Psi\varepsilon\phi_p} \inf_{\boldsymbol{\delta}} e_{M:W}(\boldsymbol{\delta}, \Psi) \geqq 1 \qquad \textit{for all } p,$$

(2.2) $$\sup_{\Psi\varepsilon\phi_p} \sup_{\boldsymbol{\delta}} e_{M:W}(\boldsymbol{\delta}, \Psi) = \infty \qquad \textit{for } p \geqq 3$$
$$= 1.15 \qquad \textit{for } p = 2.$$

PROOF. As noted before $\Psi \varepsilon \phi_p$ implies $\Lambda = \Sigma$. Also noting that $\int\psi_i^2(x)\,dx = (2\pi^{\frac{1}{2}}\sigma_i)^{-1}$ we have $\gamma_{ij} = (\frac{1}{3}\pi)\sigma_i\sigma_j\rho_{ij}'$. Using these and letting $\boldsymbol{\delta}^* = (\delta_1/\sigma_1, \delta_2/\sigma_2, \cdots, \delta_p/\sigma_p)$ in (1.1)

(2.3) $$e_{M:W}(\boldsymbol{\delta}, \Psi) = [\boldsymbol{\delta}^*(\rho_{ij})^{-1}\boldsymbol{\delta}^{*\prime}][\boldsymbol{\delta}^*(\tfrac{1}{3}\pi\rho_{ij}')^{-1}\boldsymbol{\delta}^{*\prime}]^{-1}.$$

For (2.1) it is therefore sufficient to show that for every $\Psi \varepsilon \phi_p$ the matrix difference $D = (\rho_{ij})^{-1} - (\pi\rho_{ij}'/3)^{-1}$ is positive semidefinite (psd). For a random vector $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_p)$ having cdf $\Phi_p(\mathbf{0}, (\rho_{ij}))$ the following are true:

(2.4) $$\operatorname{cov}[\Phi(Y_i), \Phi(Y_j)] = \rho_{ij}'/12, \qquad \operatorname{cov}[\Phi(Y_i), Y_j] = \rho_{ij}/2\pi^{\frac{1}{2}}.$$

The second equality follows by showing first that $\operatorname{cov}[\Phi(Y_i), Y_i] = (2\pi^{\frac{1}{2}})^{-1}$ and then for $i \neq j$ using the representation $Y_j = \rho_{ij}Y_i + (1 - \rho_{ij}^2)^{\frac{1}{2}}U$ where U is standard normal independent of Y_i. (2.4) shows that the $p \times p$ matrix $(\frac{1}{3}\pi\rho'_{ij} - \rho_{ij})$ is the covariance matrix of $\mathbf{Z} = (Z_1, Z_2, \cdots, Z_p)$ defined by

$$Z_i = 2\pi^{\frac{1}{2}}\Phi(Y_i) - Y_i, \qquad i = 1, 2, \cdots, p, \tag{2.5}$$

and hence is psd. Multiplication by the pd matrices $(\pi\rho'_{ij}/3)^{-1}$ to the left and $(\rho_{ij})^{-1}$ to the right yields D and the positive semi-definiteness is preserved. Incidentally, note that the lower bound 1 cannot be attained by any $\Psi \,\varepsilon\, \phi_p$ because nonsingularity of Ψ implies that there does not exist a linear relation among $Z_1, Z_2, \cdots, Z_p$ with probability 1 and hence that D is pd. However, the lower bound is sharp as is proved in Theorem 2.2. (2.2) readily follows from Bickel [3] by using the fact that for $\Psi \,\varepsilon\, \phi_p$, $e_{M:W} = e_{W:T}^{-1}$ and the proof is terminated.

From the univariate efficiency bounds of Chernoff-Savage [6] and Hodges-Lehman [7] it follows that if $\mathfrak{F}$ be the family of all continuous p-variate cdf with pairwise independent components

$$\inf_{\Psi\varepsilon\mathfrak{F}} \inf_{\boldsymbol{\delta}} e_{M:T}(\boldsymbol{\delta}, \Psi) = 1, \qquad \inf_{\Psi\varepsilon\mathfrak{F}} \inf_{\boldsymbol{\delta}} e_{M:W}(\boldsymbol{\delta}, \Psi) = \tfrac{1}{6}\pi, \tag{2.6}$$

and in each case the lower bound is attained by some $\Psi \,\varepsilon\, \mathfrak{F}$ and some $\boldsymbol{\delta} \neq \mathbf{0}$.The supremum in each case is infinity. Consider, next, the class of cdf Ψ which are homogeneous in the sense that the joint marginals Ψ_{ij} are the same for all $1 \leqq i \neq j \leqq p$. Exchangable random variables are a special case of this type. Denoting by σ^2, θ^{-2}, γ^2 the common diagonal elements and by $\rho\sigma^2$, $\rho^*\theta^{-2}$, $\rho'\gamma^2$ the common off diagonal elements of $\mathbf{\Sigma}$, $\mathbf{\Lambda}$, and $\mathbf{\Gamma}$ respectively, the bounds of ARE in (1.1) come out as

$$\inf_{\boldsymbol{\delta}} e_{M:T}(\boldsymbol{\delta}, \Psi) = \sigma^2\theta^2 q(\rho, \rho^*), \qquad \sup_{\boldsymbol{\delta}} e_{M:T}(\boldsymbol{\delta}, \Psi) = \sigma^2\theta^2 Q(\rho, \rho^*), \tag{2.7}$$

$$\inf_{\boldsymbol{\delta}} e_{M:W}(\boldsymbol{\delta}, \Psi) = \theta^2\gamma^2 q(\rho'\rho^*), \qquad \sup_{\boldsymbol{\delta}} e_{M:W}(\boldsymbol{\delta}, \Psi) = \theta^2\gamma^2 Q(\rho', \rho^*), \tag{2.8}$$

where $q(a, b)$ and $Q(a, b)$ are respectively the minimum and maximum of the quantities $(1 - a)/(1 - b)$ and $[1 + (p - 1)a]/[1 + (p - 1)b]$. The exact values of (2.7) and (2.8) depend, among other things, on the common correlations of the three types and hence cannot be given in general. Specializing (2.8) to $\Psi \varepsilon\, \phi_p$ once again we now show that the bound in (2.1) is sharp.

THEOREM 2.2. *If $\Psi \varepsilon\, \phi_p$ has identical correlations ρ, then*

$$\begin{aligned} \text{(a)} \quad \sup_{\boldsymbol{\delta}} [\inf] e_{M:W}(\boldsymbol{\delta},\Psi) &= A(\rho)[B_p(\rho)] && \textit{if} \quad 0 < \rho < 1 \\ &= B_p(\rho)[A(\rho)] && \textit{if} \quad -(p-1)^{-1} < \rho \leqq 0 \end{aligned} \tag{2.9}$$

where

$$A(\rho) = \tfrac{1}{3}\pi[1 - (6/\pi)\sin^{-1}(\rho/2)]/(1 - \rho) \tag{2.10}$$

$$\textit{and} \quad B_p(\rho) = \tfrac{1}{3}\pi[1 + (6/\pi)(p - 1)\sin^{-1}(\rho/2)]/[1 + (p - 1)\rho].$$

(b) *For every $\epsilon > 0$ there exists a nonsingular multivariate normal cdf Ψ_ϵ such that*

$$1 < \inf_{\boldsymbol{\delta}} e_{M:W}(\boldsymbol{\delta},\Psi_\epsilon) \leqq 1 + \epsilon. \tag{2.11}$$

PROOF. Using the representation (2.3) and the elementary inequality: $\sin(\pi x/3) \geqq x$ for $0 \leqq x \leqq \frac{1}{2}$, the part (a) follows from (2.8) by noting that $\rho' = (6/\pi)\sin^{-1}(\rho/2)$. For (b) note that the function

$$g(\rho) = (2/\rho)\sin^{-1}(\rho/2) \downarrow 1 \quad \text{and} \quad \rho \downarrow 0.$$

Select $\rho_\epsilon > 0$ such that $g(\rho_\epsilon) < 1 + \epsilon/2$. The inequality $(6/\pi)\sin^{-1}(\rho/2) < \rho$ for $0 < \rho < 1$ implies that as a function of positive integer p $B_p(\rho_\epsilon) \downarrow g(\rho_\epsilon)$ as $p \to \infty$, so that for p^* sufficiently big $B_{p^*}(\rho_\epsilon) \leqq 1 + \epsilon$. Thus, taking Ψ_ϵ to be a p^*-variate normal cdf with identical correlation ρ_ϵ (2.11) is satisfied.

The following extends to the multivariate case the result of Hodges and Lehmann [7] on the infinite ARE of the univariate normal score test vs. the Wilcoxon for distributions having discontinuous density.

THEOREM 2.3. *If the p-variate* cdf Ψ *be nondegenerate and have marginal densities satisfying the conditions* (3.8) *and* (3.10) *of Hodges-Lehmann* [7], *then for every* $\boldsymbol{\delta} \neq \mathbf{0}$, $e_{M:W}(\boldsymbol{\delta}, \Psi) = \infty$.

PROOF. Observe first that under the condition (3.8) of [7] the expression for $e_{M:W}$ given in (1.1) is not valid because under the sequence Δ_N of local shift alternatives M_N does not have an asymptotic noncentral χ_p^2 distribution whereas $\lim_{N\to\infty} \mathcal{L}_{\Delta_N}(W_N) = \chi_p^2(\Delta_W)$. The proof is accomplished, however, by showing that $\Delta_W < \infty$ and that $M_N \to \infty$ in P_{Δ_N}-probability. By a well known inequality of pd quadratic form

$$\Delta_W = \boldsymbol{\delta}\Gamma^{-1}\boldsymbol{\delta}' \leqq \boldsymbol{\delta}\boldsymbol{\delta}'/C_{\min}(\Gamma), \tag{2.12}$$

where $C_{\min}(\Gamma)$ denotes the minimum eigen value of Γ. (3.10) of [7] and nondegeneracy of Ψ imply that each $\gamma_i < \infty$ and (ρ_{ij}') is nonsingular and hence that $C_{\min}(\Gamma) > 0$. Thus, for fixed $\boldsymbol{\delta}$ $\Delta_W < \infty$. Let $\mathbf{T}_N$ denote the vector of coordinatewise normal scores test statistics and $(\hat{\rho}_{ij(N)}^*)$ the $p \times p$ matrix with elements

$$\hat{\rho}_{ij(N)}^* = N^{-1} \textstyle\sum_{\alpha=1}^N h_N(R_{i\alpha})h_N(R_{j\alpha}),$$

where $h_N(\alpha)$ is the expected value of the αth normal order statistics in a sample of size N and $R_{i\alpha}$, $\alpha = 1, 2, \cdots, N$, are the combined sample ranks in the ith coordinate. The test statistic M_N is then given by the quadratic form $M_N = \mathbf{T}_N(\hat{\rho}_{ij(N)}^*)^{-1}\mathbf{T}_N'$. If $\mathbf{u} = (u_1, u_2, \cdots, u_p)$, we have

$$\begin{aligned}\mathbf{u}(\hat{\rho}_{ij(N)}^*)\mathbf{u}' &= N^{-1} \textstyle\sum_{\alpha=1}^N [\sum_{i=1}^p u_i h_N(R_{i\alpha})h_N(R_{j\alpha})]^2 \\ &\leqq (\textstyle\sum_{i=1}^p u_i^2)N^{-1} \sum_{\alpha=1}^N \sum_{i=1}^p [h_N(R_{i\alpha})]^2 \\ &= pN^{-1}(\mathbf{uu}') \textstyle\sum_{\alpha=1}^N [h_N(\alpha)]^2.\end{aligned}$$

This implies that $C_{\max}(\hat{\rho}_{ij(N)}^*) \leqq pN^{-1} \sum_{\alpha=1}^N [h_N(\alpha)]^2$ and hence we have

$$\begin{aligned}M_N &\geqq \mathbf{T}_N\mathbf{T}_N'/C_{\max}(\hat{\rho}_{ij(N)}^*) \\ &\geqq \{pN^{-1} \textstyle\sum_{\alpha=1}^N [h_N(\alpha)]^2\}^{-1}\mathbf{T}_N\mathbf{T}_N'.\end{aligned}$$

As $N \to \infty$, $\sum_{\alpha=1}^{N} [h_N(\alpha)]^2/N \to 1$ and under the condition (3.8) of [7] $\mathbf{T}_N\mathbf{T}_N' \to \infty$ in probability. This completes the proof of the theorem.

3. Contamination model. The following theorem shows that with an increasing amount of heavy tails in the contaminating multivariate distribution the M test tends to behave much better than T^2 just as the W test does (c.f. Bickel [3]).

THEOREM 3.1. *Let Ψ and Ψ' be two continuous nondegenerate p-variate* cdf *having means* $\mathbf{0}$ *and marginal densities ψ_i and ψ_i', $i = 1, 2, \cdots, p$. For $0 < \epsilon < 1$, $\mathbf{c} = (c_1, c_2, \cdots, c_p)$ and $\mathbf{x}/\mathbf{c} = (x_1/c_1, x_2/c_2, \cdots x_p/c_p)$ define the mixture*

$$F_{\mathbf{c}}(\mathbf{x}) = (1 - \epsilon)\Psi(\mathbf{x}) + \epsilon\Psi'(\mathbf{x}/\mathbf{c}). \tag{3.1}$$

If $e(\boldsymbol{\delta}, \mathbf{c})$ denote the ARE *of the M test vs. T^2 for the parent* cdf $F_{\mathbf{c}}(\mathbf{x})$, *then for any* $\boldsymbol{\delta} \neq \mathbf{0}$ $\lim_{\mathbf{c}\to\infty} e(\boldsymbol{\delta}, \mathbf{c}) = \infty$.

PROOF. Let

$$\theta_i(c_i) = \int_{-\infty}^{\infty} [f_i^2(c_i, x)\, dx/\phi\{\Phi^{-1}[F_i(c_i, x)]\}] \tag{3.2}$$

where $F_i(c_i, x)$ and $f_i(c_i, x)$ respectively denote the ith marginal cdf and density of $F_{\mathbf{c}}$. Applying the Hodges-Lehmann [7] bound to this marginal we have $\theta_i(c_i) \geqq (2\pi)^{\frac{1}{2}} \int_{-\infty}^{\infty} f_i^2(c_i, x)\, dx$. Use of (3.1) to decompose $f_i(c_i, x)$ and application of Schwarz's inequality yield

$$\lim_{c_i\to\infty} \int_{-\infty}^{\infty} f_i^2(c_i, x)\, dx = (1 - \epsilon)^2 \int_{-\infty}^{\infty} \psi_i^2(x)\, dx > 0$$

and hence

$$\liminf_{c_i\to\infty} \theta_i(c_i) > 0. \tag{3.3}$$

Let $\sigma_i^2(c_i)$ and $\rho_{ij}(c_k, c_j)$ denote the variances and the correlations of the distribution (3.1) and ρ_{ij0} denote the correlations of the distribution Ψ'. As $\mathbf{c} \to \boldsymbol{\infty}$ each $\sigma_i^2(c_i) \to \infty$ and $\rho_{ij}(c_i, c_j) \to \rho_{ij0}$. Using (3.3) the proof now follows in the same lines as Theorem 6.1 of Bickel [2].

4. Acknowledgment. The author is grateful to Professor E. L. Lehmann for his inspiration and guidance throughout this investigation. Thanks also go to the referee for making useful comments.

REFERENCES

[1] BHATTACHARYYA, G. K. (1965). Multivariate two sample normal scores test for shift (preliminary report). *Ann. Math. Statist.* **36** 1905–1906.

[2] BICKEL, P. J. (1964). On some alternative estimates for shift in the p-variate one sample problem. *Ann. Math. Statist.* **35** 1079–1090.

[3] BICKEL, P. J. (1965). On some asymptotically nonparametric competitors of Hotelling's T^2. *Ann. Math. Statist.* **36** 160–173.

[4] CHATTERJEE, S. K. and SEN, P. K. (1964). Nonparametric tests for the bivariate two sample location problem. *Calcutta Statist. Assoc. Bulletin* **13** Nos. 49–50.

[5] CHATTERJEE, S. J. and SEN, P. K. (1965). Nonparametric tests for the multivariate location problem. (Unpublished).

[6] CHERNOFF, H. and SAVAGE, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric tests. *Ann. Math. Statist.* **29** 972–994.
[7] HODGES, J. L., JR. and LEHMANN, E. L. (1961). Comparison of the normal scores and Wilcoxon tests. *Proc. Fourth Berk. Symp. on Math. Statist. and Prob.* **1** 307–317. Univ. of California Press.
[8] SEN, P. K. and PURI, M. L. (1967). Submitted to *Ann. Math. Statist.*
[9] TAMURA, R. (1966). Multivariate nonparametric several-sample tests. *Ann. Math. Statist.* **37** 611–618.