ON THE ROBUST-EFFICIENCY OF THE COMBINATION OF INDEPENDENT NONPARAMETRIC TESTS*

PRANAB KUMAR SEN

(Received July 20, 1968)

1. Summary

Optimum combination of independent one-sample and two-sample nonparametric tests of a general class has been studied by Puri [3], [4], and this unifies the basic work by Elteren [2] and others. The object of the present investigation is to show that the combined tests considered by Puri and Elteren are asymptotically more robust-efficient when the different sets have heteroscedastic distributions.

2. The main results

Let $X_{i1},\cdots,X_{im_i}$ be m_i independent and identically distributed random variables (i.i.d.r.v.) having a continuous cumulative distribution function (c.d.f.) $F_i(x)$, and let $Y_{i1},\cdots,Y_{in_i}$ be a second set of i.i.d.r.v.'s having a continuous c.d.f. $G_i(x)$, for $i=1,\cdots,c(\geqq 2)$; all the $2c$ samples are assumed to be mutually independent. It is desired to have a combined nonparametric test for the null hypothesis H_0: $G_i(x)=F_i(x)$ for all $i=1,\cdots,c$, against $G_i(x)=F_i(x-\theta)$, $\theta\neq 0$. For the ith set, let $N_i=m_i+n_i$ and define the usual Chernoff-Savage [1] type of statistic as

$$T_{N,i}=(1/m_i)\sum_{r=1}^{N_i} E_{N_i,r}Z_{N_i,r}\ , \tag{2.1}$$

where $E_{N,r}=J_N(r/N)$, $1\leqq r\leqq N$ statisfies the conditions of Theorem 1 of Chernoff and Savage [1] and $Z_{N_i,r}$ is one or zero according as the rth smallest observation in the combined sample is from an X-observation or not, $r=1,\cdots,N_i$, $i=1,\cdots,c$. The combined tests are based on the statistics of the type

$$T_N=\sum_{i=1}^{c} h_i T_{N,i}\ , \qquad (N=N_1+\cdots+N_c)\ , \tag{2.2}$$

* Work supported by the Army Research Office, Durham, Grant DA-ARO-D-3 1-124-G746.

where $h_1, \cdots, h_c$ are suitable compounding coefficients. Puri [3] has obtained the locally best test statistic when $F_1 = \cdots = F_c = F$, and has shown it to be

$$T_N^* = \sum_{i=1}^{c} m_i T_{N,i} . \tag{2.3}$$

We shall study the robust-efficiency of T_N^* when $F_1, \cdots, F_c$ are not necessarily all identical. For this purpose, let $J(u) = \lim_{N\to\infty} J_N(u)$, $0<u<1$, and assume that it satisfies condition 3 of Theorem 1 of Chernoff and Savage [1]. Define

$$A^2 = \int_0^1 J^2(u)du - \left(\int_0^1 J(u)du\right)^2 , \qquad B(F) = \int_{-\infty}^{\infty} (d/dx) J(F(x)) dF(x) . \tag{2.4}$$

Also, let

$$n_i^0 = m_i n_i / N_i , \quad i=1,\cdots,c , \qquad N^0 = \sum_{i=1}^{c} n_i^0 \quad \text{and} \qquad \rho_i = n_i^0 / N^0 , \quad i=1,\cdots,c . \tag{2.5}$$

Assume that $\rho_1, \cdots, \rho_c$ are all bounded away from zero and one as N^0 (or N)$\to\infty$.

Then, proceeding precisely on the same line as in Sections 5, 6 and 7 of Puri [3] we arrive at the following theorem.

THEOREM 2.1. *Under the sequence of alternatives* $\{H_N: \theta=\theta_N=\lambda/(N^0)^{1/2}$, λ *real and finite*$\}$, $(N^0)^{-1/2}\left(T_N^* - M\int_0^1 (Ju)du\right)$ (*where* $M=m_1+\cdots+m_c$) *has asymptotically a normal distribution with mean* $\sum_{i=1}^{c} \rho_i B(F_i)$ *and variance* A^2, *provided the conditions of Lemma* 7.1 *of Puri* [3] *hold for all* $F_1, \cdots, F_c$.

The asymptotically locally optimum parametric test (when $F_1 = \cdots = F_c$) is based on the statistic $t^* = \sum_{i=1}^{c} n_i^0 (\bar{X}_i - \bar{Y}_i)$, where $\bar{X}_i$ and $\bar{Y}_i$ are respectively the sample averages of the X's and Y's in the ith set, $i=1,\cdots,c$ (cf. Puri [3]). Under $\{H_N\}$, it is also seen (by some standard computations) that $(N^0)^{-1/2} t^*$ has asymptotically a normal distribution with mean λ and variance $\sum_{i=1}^{c} \rho_i \sigma_i^2$, where σ_i^2 is the variance of the cdf F_i, $i=1,\cdots,c$. Hence, for $\{H_N\}$, the asymptotic relative efficiency (A.R.E.) of T_N^* with respect to t^* is equal to

$$e(T_N^*/t^*) = \left(\sum_{i=1}^{c} \rho_i B(F_i)\right)^2 \left(\sum_{i=1}^{c} \rho_i \sigma_i^2\right) \Big/ A^2 . \tag{2.6}$$

Recalling that for the cdf F_i, the A.R.E. of $T_{N,i}$ with respect to $(\bar{X}_i - \bar{Y}_i)$ is equal to $e_i = B^2(F_i)\sigma_i^2/A^2$, $i=1,\cdots,c$, (2.6) may be rewritten as

$$(2.7)\qquad e(T_N^*/t^*)=\left(\sum_{i=1}^{c}\rho_i\sqrt{e_i}/\sigma_i\right)^2\left(\sum_{i=1}^{c}\rho_i\sigma_i^2\right),$$

which is a general expression that may be simplified under certain conditions. Before considering these, we present side by side the one sample case.

Let $X_{i1},\cdots,X_{im_i}$ be m_i be i.i.d.r.v.'s having a continuous cdf $F_i(x)$, $i=1,\cdots,c(\geqq 2)$ all the c samples are assumed to be independent. The null hypothesis states that $F_1,\cdots,F_c$ are all symmetric about zero, against their symmetry about some non-zero θ. For the ith source, we define the statistic

$$(2.8)\qquad T_{m_i}=(1/m_i)\sum_{r=1}^{m_i}E_{m_i,r}Z_{m_i,r},$$

where $E_{m_i,r}$ are also Chernoff-Savage [1] type of rank-scores and $Z_{m_i,r}$ is one or 0 according as the rth smallest observation among $|X_{i1}|,\cdots,|X_{im_i}|$ is from a positive X or not, $r=1,\cdots,m_i$, $i=1,\cdots,c$. For $F_1=\cdots=F_c=F$, the locally asymptotically best statistic is deduced by Puri [4] as

$$(2.9)\qquad T_M^*=\sum_{i=1}^{c}m_iT_{m_i},\qquad M=m_1+\cdots+m_c.$$

Let $\Psi^*(x)$ be a continuous cdf symmetric about 0, and let $\Psi(x)=2\Psi^*(x)-1$, $x\geqq 0$. Then, if $E_{m,r}$ is the expected value of the rth smallest observation of a sample of size m from the cdf $\Psi(x)$, the corresponding $J(u)$ is $\Psi^{-1}(u)=\Psi^{*-1}((1+u)/2)=J^*((1+u)/2)$, $0<u<1$. We then define

$$(2.10)\qquad \mu=\int_0^1 J(u)du,\qquad A^2=\int_0^1 J^2(u)du\quad\text{and}$$

$$B(F)=\int_{-\infty}^{\infty}(d/dx)J^*(F(x))dF(x).$$

Then, proceeding precisely on the same line as in Puri [4], we arrive at the following theorem by straight-forward generalizations.

THEOREM 2.2. *Under the sequence of alternatives* $\{H_M:\ \theta=\theta_M=\lambda/M^{1/2}$, λ *real and finite*$\}$, $2M^{-1/2}(T_M^*-\mu/2)/A$ *has asymptotically a normal distribution with mean* $\lambda\sum_{i=1}^{c}\rho_iB(F_i)/A$ *and unit variance (where* $\rho_i=m_i/M$, $i=1,\cdots,c$*), provided* $F_1,\cdots,F_c$ *satisfy the conditions of Lemma* 7.1 *of Puri* [3].

Under $\{H_M\}$, the parametrically optimum test-statistic $t^*=\sum_{i=1}^{c}m_i\bar{X}_i$ leads to the following result by virtue of the well-known central limit theorem: $M^{-1/2}t^*$ has asymptotically a normal distribution with mean λ and variance $\sum_{i=1}^{c}\rho_i\sigma_i^2$. Thus, the A.R.E. of T_M^* with respect to t^* will have the same expression as in (2.6) and (2.7), with the only difference that A^2 and $B(F)$ in (2.4) and (2.10) are different. Hence, bounds for (2.7) appear to be equally applicable for both the one sample and two sample situations.

We shall specifically simplify (2.7) when $F_1, \cdots, F_c$ have the same form but they may differ possibly in scale factors, viz.,

$$F_i(x)=F((x-a_i)/\sigma_i), \qquad i=1,\cdots,c; \tag{2.11}$$

where $a_1, \cdots, a_c$ are arbitrary locations and $\sigma_1, \cdots, \sigma_c$ are scale parameters. Under

$$e_i=B^2(F_i)\sigma_i^2/A^2=B^2(F)/A^2 \qquad \text{for all } i=1,\cdots,c, \tag{2.11}$$

and hence, (2.7) reduces to

$$e(T_N^*/t^*)=(B^2(F)/A^2)\left(\sum_{i=1}^{c}\rho_i/\sigma_i\right)^{-2}\left(\sum_{i=1}^{c}\rho_i\sigma_i^2\right). \tag{2.12}$$

Now by elementary moment-inequalities, we obtain that

$$\left(\sum_{i=1}^{c}\rho_i/\sigma_i\right)^2 \geqq \left(\sum_{i=1}^{c}\rho_i\sigma_i\right)^{-2} \geqq \left(\sum_{i=1}^{c}\rho_i\sigma_i^2\right)^{-1}, \tag{2.13}$$

where the equality signs hold iff $\sigma_1=\cdots=\sigma_c$. Thus, from (2.12) and (2.13)

$$e(T_N^*/t^*)\geqq e=B^2(F)/A^2, \tag{2.14}$$

where the equality signs holds iff $\sigma_1=\cdots=\sigma_c$. This clearly exhibits the robust-efficiency of the combined test by Puri and Elteren when the parent cdf's are heteroscedastic. Moreover, if $B^2(F)/A^2$ is $\geqq e^0$ for all F (viz., the normal scores or Wilcoxon scores statistics for which e^0 are respectively 1 and 0.864), it readily follows from (2.7) and (2.13) that $e(T_N^*/t^*)$ will also be $\geqq e^0$, uniformly in $F_1, \cdots, F_c$. Thus, the lower bounds for $e(T_N^*/t^*)$, deduced for $F_1=\cdots=F_c=F$, also remain valid when $F_1, \cdots, F_c$ are arbitrarily different from each other.

UNIVERSITY OF NORTH CAROLINA, CHAPELL HILL

REFERENCES

[1] Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics, *Ann. Math. Statist.*, **29**, 972-994.

[2] Elteren, Ph. Van. (1960). On the combination of independent two sample tests of Wicoxon, *Bull. Inst. Internat. Statist.*, **37**, 351-361.

[3] Puri, M. L. (1965). On the combination of independent two sample tests of a general class, *Rev. Inst. Internat. Statist.*, **33**, 229-241.

[4] Puri, M. L. (1967). Combining independent one sample tests of significance. *Ann. Inst. Statist. Math.*, **19**, 285-300.