STATISTICA, anno LXIII, n. 2, 2003

ON THE COMBINATION OF THE SIGN AND MAESONO TESTS FOR SYMMETRY AND ITS EFFICIENCY (*)

G. Burgio, Ya. Yu. Nikitin

1. INTRODUCTION

In our previous paper (Burgio and Nikitin, 2001) we considered the construction of a new test of symmetry with respect to zero for the univariate sample X_1, …, X_n with continuous distribution, and proposed the statistic

$$G = aS + bW \tag{1}$$

where S and W, correspondingly, are the classical sign and Wilcoxon statistics and a and $b \geq 0$ are some real constants. We proved the asymptotic normality of G under the null-hypothesis and found its Pitman efficacy against the shift alternative. It was shown how to choose the constants a and b in order to get the maximal efficiency.

The idea of taking the linear combination of S and W to get a more flexible test, with improved efficiency properties, is not new. See Burgio and Nikitin (2001) for some references to early papers on the subject, including the paper by Doksum and Thomson (1971) where the statistic

$$DT = W - 0.5\, S$$

was proposed and studied. Similar, though not identical, approach was initiated by Mehra and Madhava Rao (1990) in the two sample problem.

The aim of the present paper is to investigate the properties of the generalisation of the statistic G which arises if, in (1), we replace W with the so-called Maesono statistic $W_r, r \geq 2$. The latter was proposed by Maesono (1987) and is a direct generalisation of the Wilcoxon statistic W coinciding with it for $r = 2$. Mae-

---

(*) Research supported by the grants of RFBR No. 01-01-0245 and 02-01-01099 and a short term research award of the University of Rome "La Sapienza" Grants Commission.

Major part of this work was done while the second author was visiting the "Dipartimento di Studi Geoeconomici, Statistici e Storici per l'analisi regionale" of the University of Rome "La Sapienza".

sono showed that the statistics $W_r, r \geq 2$ are often more efficient than the classical Wilcoxon statistic W_2 and hence deserve their use when testing symmetry. Some further properties of Maesono statistics were recently studied by Nikitin and Ponikarov (2002). Hence, we propose the statistic

$$G_r = aS + W_r \tag{2}$$

where S is again the sign statistic, W_r is the Maesono statistic of order $r \geq 2$ (see below for its definition) and a is some real constant to be determined in an optimal way. Note that, instead of (1) with two constants a and b, we exclude now the possibility $b = 0$ and hence we can consider, without loss of generality, the statistic (2).

2. PROPERTIES OF THE COMBINED TEST

Now return to the Maesono statistic W_r. For any natural $r \geq 2$, it is a U-statistic

$$W_r = ({}_nC_r)^{-1} \sum_{1 \leq i_1 < .... < i_r \leq n} K_r(X_{i_1}, ..., X_{i_r})$$

with the kernel

$$K_r(s_1, ..., s_r) = r^{-1} \left(\sum_{i=1}^{r} \prod_{j \neq i}^{r} 1_{\{s_i + s_j > 0\}} \right).$$

For a clear idea of the structure of such kernels, e.g. for $r = 3$, we write out

$$K_3(s_1, s_2, s_3) = \frac{1}{3}(1_{\{s_1+s_2>0, s_1+s_3>0\}} + 1_{\{s_1+s_2>0, s_2+s_3>0\}} + 1_{\{s_2+s_3>0, s_1+s_3>0\}}) .$$

As for the sign statistic we clearly have the representation

$$S = n^{-1} \sum_{j=1}^{n} 1_{\{X_j > 0\}} = ({}_nC_r)^{-1} \sum_{1 \leq i_1 < .... < i_r \leq n} L_r(X_{i_1}, ..., X_{i_r}) ,$$

with

$$L_r(X_1, ..., X_r) = r^{-1}(1_{\{X_1 > 0\}} + ... + 1_{\{X_r > 0\}}) ,$$

the combination of these statistics $G_r = aS + W_r$ has the kernel

$$\Phi_r(s_1,....,s_r) = r^{-1}\left(a\sum_{i=1}^{r} 1_{\{s_i>0\}} + \sum_{i=1}^{r}\prod_{j\neq i}^{r} 1_{\{s_i+s_j>0\}} \right). \quad (3)$$

Denote by F the distribution function (d.f.) and by f the density of the initial observations X_1, $X_2,\dots$. Let us find some important characteristics of the kernel (3). First of all we need

$$\mu_a(F) = E_F\Phi_r(X_1,...,X_r) = r^{-1}[aP(X_1>0) + P(X_1+X_2>0,....,X_1+X_r>0)] =$$
$$= r^{-1}a(1-F(0)) + r^{-1}\int_{-\infty}^{+\infty}(1-F(-x))^{r-1}dF(x).$$

We are also interested in the projection of the kernel Φ_r, namely in the function

$$\psi_r(t) = E_F(\Phi_r(X_1,....,X_r) \mid X_1 = t) =$$
$$= r^{-1}\left[a1_{\{t>0\}} + a(r-1)(1-F(0)) + (1-F(-t))^{r-1} + (r-1)\int_{-t}^{+\infty}(1-F(-y))^{r-2}dF(y) \right].$$

We need also the value of the expectation

$$\mathrm{E}_F\psi_r(X_1) =$$
$$= r^{-1}\left[ar(1-F(0) + \int_{-\infty}^{+\infty}(1-F(-t))^{r-1}dF(t) + (r-1)\int_{-\infty}^{+\infty}\int_{-t}^{+\infty}(1-F(-y))^{r-2}dF(y)dF(t) \right].$$

Now we are ready to calculate the important variance

$$\sigma_r^2(a,F) = \mathrm{Var}_F\psi_r(X_1) =$$
$$= r^{-2}\,\mathrm{E}_F\Bigg[a1_{\{X_1>0\}} - a(1-F(0)) + (1-F(-X_1))^{r-1} - \int_{-t}^{+\infty}(1-F(-t))^{r-1}dF(t) +$$
$$+(r-1)\int_{-X_1}^{+\infty}(1-F(-t))^{r-2}dF(t) - (r-1)\int_{-\infty}^{+\infty}\int_{t}^{+\infty}(1-F(-y))^{r-2}dF(y)dF(t)\Bigg]^2.$$

Suppose that $\sigma_r^2(a,F) > 0$ for our F. Then we can apply the central limit theorem for non degenerate U-statistics (see, e.g., Serfling (1980) or Korolyuk and Borovskikh (1994)). Consequently we have the convergence in distribution

$$\sqrt{n}(G_r - \mu_a(F)) \to N(0, r^2\sigma_r^2(a,F)). \quad (4)$$

This asymptotic result enables us to construct the critical domain for our test using the normal approximation. In the next section we give the simplified expressions for $\mu_a(F)$ and $\sigma_r^2(a,F)$ in the case of null-hypothesis of symmetry.

3. PITMAN EFFICACY OF THE NEW TEST

In order to make the Pitman efficiency calculation we should precise the formulation of the statistical problem. Suppose that, under the null-hypothesis of simmetry H_0, the initial d.f. function F is absolutely continuous and symmetric with respect to zero. Hence, for every x,

$$1-F(x)-F(-x)=0 .$$

We suppose also that, under the alternative H_1, the observations have d.f. $F(x)=F_0(x-\theta),\theta\geq 0$ for some symmetric d.f. F_0 with a.e. differentiable density f_0. We consider the case of location alternative only for simplicity, the case of general parametric families may be treated in a similar way as shown in Ch. 6 of Nikitin (1995).

As usually when calculating the Pitman efficiency, we take the parameter θ in the form $\theta=\theta_n=\delta/\sqrt{n}$ for some $\delta\geq 0$. It is assumed that the condition

$$\int_{-\infty}^{+\infty} f_0^2(y)dy<\infty \tag{5}$$

is valid all along this paper. We impose also the natural condition

$$f_0(\pm\infty)=0 . \tag{6}$$

As noted above, in the case of location alternative, the expressions for $\mu_a(F)$ and $\sigma_r^2(a,F)$ become simpler if we use the symmetry of F_0. Then we can simplify the expression for $\mu_a(F_0,\theta)$ as follows:

$$\mu_a(F_0,\theta)=r^{-1}\left[aF_0(\theta)+\int_{-\infty}^{+\infty}(F_0(x+2\theta))^{r-1}f_0(x)dx\right]. \tag{7}$$

The expression for $\sigma_r^2(a,F_0,\theta)$ is more complicated. Denote for simplicity

$$Z(F_0,x,r,\theta)=(F_0(x-\theta))^{r-1}-(1-F_0(x-\theta))^{r-1} .$$

We have

$$\sigma_r^2(a,F_0,\theta)=r^{-2}\left[\frac{a^2}{4}+\int_{-\infty}^{+\infty}(Z(F_0,x,r,\theta))^2\,dF_0(x-\theta)+\right.$$

$$\left.-a\int_{-\infty}^{+\infty}\big(Z(F_0,x,r,\theta)\big)dF_0(x-\theta)+2a\int_{0}^{+\infty}(Z(F_0,x,r,\theta))dF_0(x-\theta)\right].$$

In the case of the null-hypothesis, when $\theta=0$, after some easy calculations, we obtain

$$\sigma_r^2(a,F_0,0)=r^{-2}\left[\frac{a^2}{4}+2\left(\frac{1}{2r-1}-\frac{(r-1)!^2}{(2r\text{-}1)!}\right)+\frac{2a}{r}(1-2^{-r+1})\right]. \tag{8}$$

From (8) it follows that, for any $r\geq 2$, any continuous F_0 and any *a*, $\sigma_r^2(a,F_0,0)>0$ so that our kernel and our U-statistic, under the hypothesis of symmetry, are non-degenerate. Moreover, the analysis of the above formula for $\sigma_r^2(a,F_0,\theta)$ shows that this expression is continuous in θ and, for sufficiently small θ, is close to the value of $\sigma_r^2(a,F_0,0)$ and hence positive.

Using the classical rate of convergence result for nondegenerate U-statistics (see, e.g., Korolyuk and Borovskikh, 1994), we may state that the convergence in (4) is uniform with respect to $0\leq\theta<\theta_0$ for sufficiently small θ_0. Due to the regularity conditions (5) and (6), from the expression (7) we obtain that

$$\frac{d}{d\theta}\mu_a(F_0,\theta)\Big|_{\theta=\theta_0}=r^{-1}\left[af_0(0)+2(r-1)\int_{-\infty}^{+\infty}F_0^{r-2}(y)f_0^2(y)dy\right]. \tag{9}$$

We suppose that *a* is such that the expression in (9) is positive (this condition, ensuring the consistency of our test, is always required when calculating Pitman efficiency). According to the well-known formulae and conditions for the calculation of Pitman efficacy (Rao, 1965; Hettmansperger, 1984) we get the following expression for Pitman efficacy of the test based on G_r

$$\mathit{eff}^{\,P}(G_r,f_0)=\frac{\left[af_0(0)+2(r-1)\int_{-\infty}^{+\infty}F_0^{r-2}(x)f_0^2(x)dx\right]^2}{\frac{a^2}{4}+2\left[\frac{1}{2r-1}-\frac{(r-1)!^2}{(2r-1)!}\right]+\frac{2a}{r}(1-2^{-r+1})}.$$

Now we want to maximise this expression with respect to *a.* To make the expression more compact, let us denote

$$f_0(0) = m, \qquad 2(r-1)\int_{-\infty}^{+\infty} F_0^{r-2}(x) f_0^2(x)dx = i\,,$$

$$\frac{2}{r}(1-2^{-r+1}) = d\,, \qquad 2\left[\frac{1}{2r-1} - \frac{(r-1)!^2}{(2r-1)!}\right] = w\,.$$

Clearly, the maximum of the function

$$a \to \frac{4(am+i)^2}{a^2+4ad+4w}$$

is attained for

$$a_0 = \frac{2di-4mw}{2md-i} \tag{10}$$

and hence the maximal value of efficacy is

$$k^2(r, f_0) = \frac{i^2-4mdi+4m^2w}{w-d^2}\,. \tag{11}$$

We see that this expression, for $r = 2$, completely corresponds to the maximal efficacy obtained in Burgio and Nikitin (2001) for the combination of the sign and Wilcoxon tests.

In practice, the optimal value a_0 is certainly unknown and we can estimate it according to (10) and using, say, the kernel estimates for m and i. This recommendation should work well in large samples. It differs, however, from Mehra and Madhava Rao (1990) where, in similar situation, they recommend to use the value of a_0 calculated for the normal case.

4. COMPARISON WITH THE STUDENT TEST

Since the famous paper by Hodges and Lehmann (1956), it became usual to compare the new nonparametric tests of symmetry with the Student test in the case of the normal law. Denote by $e^P(U,V)$ the Pitman efficiency of the test U with respect to V and let t denote the Student test. The classical result by Hodges and Lehmann says that, for the normal law and the location alternative,

$$e^P(W,t) = 3/\pi \approx 0.9549\,.$$

In Burgio and Nikitin (2001) it was shown that the statistic G_2, with the optimal coefficients, gives a better result

$$e^P(G_2,t) \approx 0.9643 .$$

In this section, we will show that the statistic G_4, with the appropriate choice of *a*, gives an even better result, namely

$$e^P(G_4,t) \approx 0.9794 . \tag{12}$$

The case $r = 3$ does not present any interest because, as shown by Maesono (1987), the statistics W_2 and W_3 have the same Pitman efficiency. However, see Nikitin and Ponikarov (2002) on the comparison of W_2 and W_3 from the point of view of Bahadur efficiency.

To justify the result (12), note that the efficiency of the test based on G_4, with respect to the *t* test, is given by the expression (see, e.g., Hodges and Lehmann, 1956; Hettmansperger, 1984)

$$e^P(G_4,t) = \sigma^2 k^2(4,f_0), \tag{13}$$

where σ^2 is the variance of the underlying symmetric distribution. Clearly, when calculating the right-hand side of (13), we may assume that $\sigma^2 = 1$. First of all we need the value of the integral

$$I_2 = \int_{-\infty}^{+\infty} \Phi^2(x)\varphi^2(x)dx ,$$

where Φ and φ, as usually, denote the standard normal d.f. and density. This integral can be found numerically but, in order to get maximal precision, we will evaluate its exact value. It is sufficient to prove the following

Lemma. For any real α, $h(\alpha) := \int_{-\infty}^{+\infty} \Phi^2(\alpha x)d\Phi(x) = \frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{1}{\sqrt{1+2\alpha^2}}\right).$

Proof. The statement is proved using the differentiation with respect to α. Clearly

$$h'(\alpha) = \frac{1}{\pi}\int_{-\infty}^{+\infty} x\exp\left(-\frac{(1+\alpha^2)x^2}{2}\right)\Phi(\alpha x)dx = -\frac{1}{\pi(1+\alpha^2)}\int_{-\infty}^{+\infty}\Phi(\alpha x)d\left[\exp\left(-\frac{(1+\alpha^2)x^2}{2}\right)\right]$$

$$= \frac{\alpha}{\pi(1+\alpha^2)}\int_{-\infty}^{+\infty}\exp\left(-\frac{(1+\alpha^2)x^2}{2}\right)\varphi(\alpha x)dx =$$

$$= \frac{\alpha}{\pi\sqrt{2\pi}(1+\alpha^2)}\int_{-\infty}^{+\infty}\exp\left(-\frac{(2+\alpha^2)x^2}{2}\right)dx = \frac{\alpha}{\pi(1+\alpha^2)\sqrt{1+2\alpha^2}} .$$

Integrating this expression by substituting $x = (\tan t)/\sqrt{2}$, for some constant C, we obtain

$$h(\alpha) + C = \pi^{-1}\int \frac{\alpha}{(1+\alpha^2)\sqrt{1+2\alpha^2}} d\alpha = \pi^{-1}\int \frac{\tan t \cos t}{\cos^2 t(2+\tan^2 t)} dt = \pi^{-1}\int \frac{\sin t}{1+\cos^2 t} dt =$$

$$= -\pi^{-1}\arctan(\cos t) = -\pi^{-1}\arctan(1+2\alpha^2)^{-1/2}.$$

For $\alpha = 0$, we have $h(0) = \frac{1}{4}$, hence $C = -\frac{1}{2}$. □

Now take in consideration that

$$I_2 = (2\pi)^{-1}\int_{-\infty}^{+\infty} \Phi^2(x)\exp(-x^2)dx = \frac{1}{2\sqrt{\pi}} h\left(\frac{1}{\sqrt{2}}\right).$$

Hence we get easily

$$I_2 = \frac{1}{4\sqrt{\pi}} - \frac{1}{2\pi^{3/2}}\arctan\frac{1}{\sqrt{2}} \approx 0.08578 .$$

Let continue the calculations for $r = 4$. We have easily

$$m \approx 0.39894,\ w \approx 0.27143,\ d = 0.4375$$

and finally $i = 6I_2 \approx 0.51469$. With these values, the formula (11) takes the value

$$\frac{i^2 - 4mdi + 4m^2 w}{w - d^2} \approx \frac{0.26490 - 0.35933 + 0.17280}{0.27143 - 0.19141} \approx \frac{0.07837}{0.08002} \approx 0.9794$$

and using (13), we get (12). Therefore, it is possible to conclude that, even for the normal law, the efficiency of G_4 with respect to the Student test (which is in our context the likelihood ratio test) is very high.

The calculations for the logistic distribution, with

$$F_0(x) = \frac{1}{1+\exp(-x)}, \qquad f_0(x) = \frac{\exp(-x)}{(1+\exp(-x))^2},$$

are easier and the new test becomes more efficient than the Student test. We may expect such result as the Maesono statistic is closely related to the Wilcoxon test and the latter is Pitman optimal in this context (Hettmansperger, 1984). The calculations are as follows

$$I_2 = \int_{-\infty}^{+\infty} F_0^2(y) f_0^2(y) dy = \int_{-\infty}^{+\infty} \frac{\exp(4x)}{(1+\exp(x))^6} dx = \int_0^{+\infty} \frac{x^3}{(1+x)^6} dx = \frac{\Gamma(2)\Gamma(4)}{\Gamma(6)} = \frac{1}{20}.$$

The arguments similar to the previous section give us k^2 (4, *logistic*) = 0.3325 whereas the variance of logistic distribution is equal to $\pi^2/3$. Hence the Pitman efficiency of G_4 with respect to the *t* test is

$$e^P(G_4, t) = (\pi^2/3) \cdot 0.3325 \approx 1.0939.$$

It is interesting to calculate such relative efficiency for other model distributions, say, for generalised normal distribution of order *p* (for related results see Burgio and Nikitin, 1998). We can compare our test with any other test with known efficacy, for instance with the locally most powerful linear rank test which is optimal in our problem (see, e.g., Hettmansperger, 1984).

It is natural to ask for what density f_0 our test has the same efficiency as such rank test. Unfortunately we cannot expect getting the explicit solution for $G_r, r \geq 4$ as it turned out that it is impossible even in the case of pure Maesono statistic (Nikitin and Ponikarov, 2002). The solution for G_2 was described in Burgio and Nikitin (2001).

Dip. di studi geoeconomici, statistici e storici per l'analisi regionale
Università di Roma "La Sapienza"
GIUSEPPE BURGIO

Department of Mathematics and Mechanics
Saint Petersburg State University
YAKOV YU. NIKITIN

REFERENCES

G. BURGIO, YA.YU. NIKITIN, (1998), *Goodness-of-fit tests for normal distribution of order p and their asymptotic efficiency*, "Statistica", LVIII, 2, pp. 213-230.

G. BURGIO, YA.YU. NIKITIN, (2001), *The combination of the sign and Wilcoxon tests for symmetry and their Pitman efficiency*, in "Asymptotic Methods in Probability and Mathematical Statistics, with Applications", Birkhaeuser, Boston, pp. 395-408.

K. DOKSUM, R. THOMPSON, (1971), *Power bounds and asymptotic minimax results for one-sample rank tests,* "Annals of Mathematical Statistics", 42, pp. 12-34.

T. HETTMANSPERGER, (1984), *Statistical inference based on ranks*, Wiley, New York.

J. HODGES, E. LEHMANN, (1956), *The efficiency of some nonparametric distributions*, "Annals of Mathematical Statistics", 19, pp. 293-325.

V.S. KOROLYUK, YU.V. BOROVSKIKH, (1994), *Theory of U-statistics,* Dordrecht, Kluver.

Y. MAESONO, (1987), *Competitors of the Wilcoxon signed rank test*, "Annals of the Institute of Statistical Mathematics", 39, Pt.A, pp. 363-375.

K.L. MEHRA, K.S. MADHAVA RAO, (1990), *On a class of nonparametric tests for the treatment vs control problem*, "Communication in Statistics – Theory and Methods", 19, 6, pp. 2323-2336.

YA.YU. NIKITIN, (1995), *Asymptotic efficiency of nonparametric tests*, Cambridge University Press, New York.

YA.YU. NIKITIN, E.V. PONIKAROV, (2002), *Asymptotic efficiency of Maesono statistics for testing of simmetry*, "Annals of the Institute of Statistical Mathematics", 54, 2, pp. 382-390.
C.R. RAO, (1965), *Linear statistical inference and its applications*, Wiley, New York.
R. SERFLING, (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

RIASSUNTO

Un nuovo test di simmetria, combinazione dei test dei segni e di Maesono, e sua efficienza

In questo lavoro viene proposto un nuovo test statistico per l'ipotesi di simmetria, che è una combinazione del test dei segni e del test W_r di Maesono. Quest'ultimo è una generalizzazione del test di Wilcoxon e coincide con esso per $r = 2$. Il nuovo test combinato appartiene alla classe delle statistiche U non degeneri e pertanto ha distribuzione asintotica normale. L'efficienza di Pitman del nuovo test è calcolata e confrontata con quella del test t di Student. Si dimostra, tra l'altro, che nel caso normale e per $r = 4$ il nuovo test, al confronto con il test t, è più efficiente di quello di Wilcoxon. Difatti, l'efficienza di quest'ultimo rispetto a t è 0,9549, mentre quella di G_4 rispetto a t è 0,9794. Nel caso di popolazione logistica, il nuovo test G_4 risulta più efficiente di t, con un'efficienza pari a 1,0939.

SUMMARY

On the combination of the sign and Maesono tests for simmetry and its efficiency

We propose a new test for the simmetry hypothesis which is a combination of the sign statistic and the W_r Maesono statistic. The latter generalizes the Wilcoxon statistic and coincides with it for $r = 2$. The proposed statistic belongs to the class of non-degenerate U-statistics and hence it has asymptotically normal distribution. We calculate its Pitman efficacy and compare it with the t-test. For instance, in the normal case, for $r = 4$, the new test, with respect to the t-test, has a higher efficiency (0.9794) than the Wilcoxon test (0.9549). In the logistic case, G_4 has a higher efficiency (1,0939) than the t-test.