New Results on the Cucconi Test

Nuovi risultati sul test di Cucconi

Marco Marozzi
Dipartimento di Economia e Statistica, Università della Calabria
e-mail: marozzi@unical.it

Keywords: the location-scale problem, rank testing, the Cucconi test, the Lepage test

1. The Cucconi test for the location-scale problem

Let X_1 and X_2 be the random variables underlying two populations with cdfs F_1 and F_2. The general system of hypotheses of interest in comparing them is H_0:$\{F_1(s)=F_2(s)\ \forall s\in]-\infty,+\infty[\}$ vs H_1:$\{\exists A\subset]-\infty,+\infty[$: $F_1(s)\neq F_2(s)$, $s\in A$ with $\Pr(A)>0\}$. We address the location-scale problem which corresponds to take $F_1(s)=G[(s-\mu_1)/\sigma_1]$ and $F_2(s)=G[(s-\mu_2)/\sigma_2]$, where $G(.)$ is the cdf of a continuous variable with location 0 and scale 1, μ_1 and μ_2 (σ_1 and σ_2) are the locations (scales) of population 1 and 2, respectively. Let observations $X_{11},...,X_{1n_1}$ and $X_{21},...,X_{2n_2}$ be random samples from population 1 and 2 respectively.

For the location-scale problem, Odoardo Cucconi (1968) proposed a rank test which is not very well-known in the literature. The test is based on

$$C=\frac{U^2+V^2-2\rho UV}{2(1-\rho^2)},$$

where $U=\left(6\sum_{i=1}^{n_1}W_{1i}^2-b\right)/\sqrt{c}$, $V=\left(6\sum_{i=1}^{n_1}(n+1-W_{1i})^2-b\right)/\sqrt{c}$, $b=n_1(n+1)(2n+1)$, $c=n_1n_2(n+1)(2n+1)(8n+11)/5$, $n=n_1+n_2$, W_{ji} denotes the rank of X_{ji} in the combined sample and $\rho=\frac{2(n^2-4)}{(2n+1)(8n+11)}-1$. Note that U is based on the squared ranks while V on the squared counter-ranks of the first sample. Under H_0, $E(U)=E(V)=0$ and $VAR(U)=VAR(V)=1$. Of course U and V are negative dependent. More precisely, $CORR(U,V)=COVAR(U,V)=\rho$ and takes values in $[-1,-7/8[$. It may be shown that under H_0 (U,V) is centered on $(0,0)$ but not under H_1. Standard asymptotics show that if $n_1,n_2\to\infty$ and $n_1/n\to d\in]0,1[$ then $\Pr(U\leq u)\to\Phi(u)$, where Φ is the standard normal cdf (the same applies for V), moreover

$$\Pr(U\leq u,V\leq v)\to\int_{-\infty}^{v}\int_{-\infty}^{u}\frac{1}{2\pi\sqrt{1-\rho_0^2}}\exp\left(-\frac{u^2+v^2-2\rho_0uv}{2(1-\rho_0^2)}\right)dudv=\int_{-\infty}^{v}\int_{-\infty}^{u}f(u,v)dudv,$$

where $\rho_0=-7/8$. The points (u,v) that speak in favor of H_0 are close to $(0,0)$ and satisfy $f(u,v)\geq h$, where the constant h is chosen so that the type-one error rate is α. Let $h=\alpha\left(2\pi\sqrt{1-\rho_0^2}\right)^{-1}$ then H_0 should be accepted if the point (u,v) is such that

$\left(u^2 + v^2 - 2\rho_0 uv\right)/\left(2\left(1-\rho_0^2\right)\right) < -\ln\alpha$. The test is unbiased and consistent (Cucconi 1968).

The test is of interest since, contrary to the other location-scale tests, is not a quadratic form of a test on location and one on scale, but it is directly devised for jointly detecting location and scale shifts. Note that the Lepage (1971) *L* test is the sum of the squares of the standardized Wilcoxon and Ansari-Bradley statistics.

2. Power of the Cucconi test and other results

The power of the Cucconi *C* test has been evaluated in detail for the first time and compared to that of the much more known and used *L* test by simulations. We consider the following distributions: standard normal, double exponential, Student's with 2 df; Cauchy, a 10% and a 30% outlier, a bimodal distribution with light tails. We considered (n_1,n_2)=(10,10), (10,30) and (30,30). The power has been estimated through 10,000 simulations. Only results under normal and Cauchy with n_1=n_2=30 are reported.

Table 1: *% power estimates with* α*=5%, under normal and Cauchy,* (n_1,n_2)*=(30,30)*

	Normal								
μ_1-μ_2	0	0.53	0.53	0.53	0.53	0	0.25	0.50	0.75
σ_1/σ_2	1	1	1.5	2	3	1.63	1.63	1.63	1.63
L	4.96	38.5	71.6	91.9	99.7	40.5	50.2	75.6	94.3
C	4.88	38.3	74.9	95.1	99.9	48.3	56.7	79.5	95.3
	Cauchy								
μ_1-μ_2	0	1	1	1	1	0	0.6	1.2	1.8
σ_1/σ_2	1	1	1.5	2	3	2.15	2.15	2.15	2.15
L	4.68	43.0	62.9	80.6	94.7	39.1	62.6	90.1	98.3
C	4.79	42.5	62.0	79.8	94.2	37.4	61.4	89.7	98.2

The tests are conservative but the *C* test with a higher degree and it performs better than the *L* test under light-tailed distributions, while the *L* test performs slightly better under heavy-tailed ones. Therefore we suggest the *C* test as a good alternative to the *L* test when addressing location-scale problems within the rank framework.

Actual applications and the multivariate *C* test, developed via the nonparametric combination framework (Pesarin, 2001), are not included for lack of space.

References

Cucconi O. (1968) Un nuovo test non parametrico per il confronto tra due gruppi campionari, *Giornale degli Economisti*, XXVII, 225-248.

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

Pesarin F. (2001) *Multivariate Permutation Tests with Applications in Biostatistics*, Wiley, Chichester.