A Note on Combining Dependent Tests of Significance

Joachim Hartung

Department of Statistics
University of Dortmund
Vogelpothsweg 87
D–44221 Dortmund
Germany

Abstract

In combining several tests of significance the individual test statistics are allowed to be dependent. By choosing the weighted inverse normal method for the combination, the dependency of the original test statistics is then characterized by a correlation of the transformed statistics. For this correlation a confidence region, an unbiased estimator and an unbiased estimate of its variance are derived. The combined test statistic is extended to include the case of possibly dependent original test statistics. A simulation study shows the performance of the actual significance level.

Key Words: Combining dependent test statistics, Combining p–values, Non–parametric meta–analysis, Inverse normal method, Multiple endpoints.

1 Introduction

In many situations we are lead to combine several test statistics. For instance in a clinical trial there may be several aspects of the same underlying drug influence, so–called multiple endpoints, and one might be interested in the question, whether there is any drug effect at all. Or in the meta–analysis of a series of similar studies resp. experiments a common overall effect is of interest. Often the various test statistics are so different, that a direct combination is not possible, but the corresponding p–values of the test statistics under the individual null hypothesis have to be combined.

Now the cruical point in the assumptions of the known methods is the claim that the various test statistics have to be independent, cf. for instance Birnbaum (1954), Liptak (1958), van Zwet and Oosterhoff (1967), Hedges and Olkin (1985). Particularly in the first case of application mentioned above this condition is seldomly fulfilled.

Allowing for dependency in the original test statistics, in the class of quasi means of the p–values characterized by Liptak (1958) his proposal to work with the nowadays generally called inverse normal method is chosen, because then dependency becomes equivalent to correlation. Indeed a single parametric formulation of this correlation does have to be assumed, but of course it can also be regarded as a mean correlation approximating the case of possibly different correlations between the transformed statistics. Further, the parameter sets belonging to the individual hypotheses might not be different, cf. Hedges and Olkin (1985) for a general discussion of the inverse normal method, a slight modification of it is now given below to include the case of dependent test statistics, at large following the notations of van Zwet and Oosterhoff (1967).

2 Main Results

For $i = 1, \dots, n$, let T_i be one–sided test statistics for testing the null–hypotheses

$$H_{i,0} \ : \ \vartheta_i = \vartheta_{i0} \quad ,$$

for the real–valued parameters ϑ_i, against the one–sided alternatives

$$H_{i,1} \ : \ \vartheta_i > \vartheta_{i0} \quad ,$$

where large values of T_i may lead to a rejection of $H_{i,0}$.

It is desired to combine the results of these tests, i.e. to construct a function of $T_1, \dots, T_n$ that can be used to test the combined null–hypothesis

$$H_0 \ : \ \vartheta_i = \vartheta_{i0} \quad , \text{for all } i = 1, \dots, n \quad ,$$

against the alternative

$$H_1 \ : \ \vartheta_i \geq \vartheta_{i0} \quad , \ \text{for all } i = 1, \ldots, n$$
$$\text{with strict inequality in at least one case.}$$

If T_i has a continuous distribution function F_{i0} under the null–hypothesis $H_{i,0}$, then $F_{i0}(T_i)$, i.e. the $1-$ p–value(T_i), is uniformly distributed on $(0,1)$, and denoting Φ^{-1} the inverse of the standard normal distribution function, the 'probit'

$$(2.1) \qquad t_i \ = \ \Phi^{-1}(F_{i0}(T_i))$$

has a standard normal distribution under $H_{i,0}$, $i = 1, \ldots, n$.

Dependency in the original test statistics $T_1, \ldots, T_n$ leads to a dependency in the probits $t_1, \ldots, t_n$, which now is equivalent to some correlation of the t_i's, and we assume

$$\mathrm{Cov}(t_i, t_j) \ = \ \rho, \quad \text{for } i \neq j, \quad i, j = 1, \ldots, n \quad ,$$

with the real–valued parameter ρ in the natural parameter set, i.e. $-\frac{1}{n-1} \leq \rho \leq 1$.

Let now $\lambda_1, \ldots, \lambda_n$ be a set of real valued weights, with $\sum_{i=1}^n \lambda_i \neq 0$, so according to the 'weighted inverse normal method' the combined test statistic, which under H_0 is standard normally distributed, is given as follows:

$$(2.2) \qquad t(\rho) \ = \ \frac{\sum_{i=1}^n \lambda_i t_i}{\sqrt{(1-\rho)\sum_{i=1}^n {\lambda_i}^2 + \rho\left(\sum_{i=1}^n \lambda_i\right)^2}} \ \underset{H_0}{\sim} \ N(0,1) \quad ,$$

and we see, for $\rho = 0$ we get the usual 'weighted inverse normal method' test statistic $t(0)$ for independent test statistics T_i, c.f. for instance Hedges and Olkin (1985), and for $\rho = 1$ we have just the weighted mean of the probits $t_1, \ldots, t_n$, with weights $\lambda_i / \sum_{j=1}^n \lambda_j$.

If there is no further information available about ρ, the t_i's have to be used themselves for drawing some inference about ρ. So let be defined the quadratic form

$$(2.3) \qquad q \;=\; \frac{1}{n-1}\sum_{i=1}^{n}\left(t_i - \frac{1}{n}\sum_{i=1}^{n} t_i\right)^2 \quad ,$$

then, under H_0, a confidence interval for ρ at level $1-\gamma$, $0<\gamma<1$, is given by

$$(2.4) \qquad 1-\frac{(n-1)\cdot q}{\chi^2_{n-1;\gamma_1}} \;\le\; \rho \;\le\; 1-\frac{(n-1)\cdot q}{\chi^2_{n-1;1-\gamma_2}}, \qquad \begin{array}{l}\gamma_1+\gamma_2=\gamma\ ,\\ \gamma_1\ge 0 \quad , \quad \gamma_2 \ge 0\ ,\end{array}$$

where $\chi^2_{n-1;\gamma}$ denotes the γ–quantile of the central χ^2 distribution with $(n-1)$ degrees of freedom.

Further, under H_0, an unbiased estimator of ρ is given by

$$(2.5) \qquad \hat{\rho} \;=\; 1-q$$

with the variance

$$(2.6) \qquad \mathrm{Var}(\hat{\rho}) \;=\; \frac{2}{n-1}(1-\rho)^2 \quad ,$$

which can be estimated unbiasedly by

$$(2.7) \qquad \widehat{\mathrm{Var}(\hat{\rho})} \;=\; \frac{2}{n+1}\cdot q^2 \quad .$$

Extending the usual 'weighted inverse normal method' in order to include the case of possibly dependent test statistics, an estimated form of (2.2), that under H_0 is approximately standard normally distributed, is now given by

$$(2.8) \qquad t(\hat{\rho}^*,\kappa) \;=\; \frac{\sum\limits_{i=1}^{n}\lambda_i t_i}{\sqrt{\sum\limits_{i=1}^{n}{\lambda_i}^2 + \left[\left(\sum\limits_{i=1}^{n}\lambda_i\right)^2 - \sum\limits_{i=1}^{n}{\lambda_i}^2\right]\left\{\hat{\rho}^* + \kappa\cdot\sqrt{\frac{2}{n+1}}(1-\hat{\rho}^*)\right\}}}$$

$$\underset{H_0}{\overset{\text{approx.}}{\sim}} N(0,1) \quad ,$$

where

$$(2.9) \qquad \hat{\rho}^* = \max\left\{-\frac{1}{n-1}, \hat{\rho}\right\}$$

and $\kappa \geq 0$ is a parameter regulating the actual significance level, chosen for instance as $\kappa = \kappa_1 = 0.2$ or as $\kappa = \kappa_2 = (1 + \frac{1}{n-1} - \hat{\rho}^*) \cdot 0.1$, as in Table 1.

For demonstrating the influence of correlation on the significance level given by the usual test statistic $t(0)$, cf. (2.2), and its corrections by the statistic $t(\hat{\rho}^*, \kappa)$, cf. (2.8), with $\kappa = \kappa_1$ and $\kappa = \kappa_2$, in table 1 for some constellations of $n, \lambda_1, \ldots, \lambda_n$ and ρ the realizations of significance levels $\hat{\alpha}$ are simulated (10.000 runs each) for a prescribed nominal significance level of $\alpha = 0.05$.

For the most practical case of a nonnegative correlation the correction statistics (2.8) have a good performance, whereas for negative correlations — being more of a theoretical interest — all considered statistics are too conservative. For improvements in that case the factor κ in (2.8) could be weakened, allowing then, of course, the significance level to increase for nonnegative correlations.

Table 1: (*) Simulated actual significance levels $\hat{\alpha}$ to the nominal significance level $\alpha = 0.05$ in several constellations of $n, \lambda_1, \ldots, \lambda_n$ and ρ for the combined test statistics $t(0)$, cf. (2.2), and $t(\hat{\rho}^*, \kappa)$, c.f. (2.8), with $\kappa_1 = 0.2$ and $\kappa_2 = [1 + 1/(n-1) - \hat{\rho}^*] \cdot 0.1$.

$\alpha = 0.05$ nominal significance level			realized significance level $\hat{\alpha}$							
			ρ							
			< 0		≥ 0					
n	λ_i	test stat.	$-\frac{1}{n}$	$-\frac{1}{2(n-1)}$	0.0	0.05	0.1	0.2	0.5	1.0
3	1	$t(0)$	0.002	0.009	0.05	0.06	0.07	0.08	0.13	0.17
		$t(\hat{\rho}^*, \kappa_1)$	0.011	0.019	0.05	0.05	0.05	0.05	0.06	0.05
		$t(\hat{\rho}^*, \kappa_2)$	0.013	0.023	0.05	0.06	0.06	0.07	0.06	0.05
	$\sqrt{i}$	$t(0)$	0.003	0.013	0.05	0.06	0.07	0.08	0.12	0.16
		$t(\hat{\rho}^*, \kappa_1)$	0.014	0.022	0.05	0.05	0.05	0.05	0.05	0.05
		$t(\hat{\rho}^*, \kappa_2)$	0.016	0.025	0.05	0.06	0.06	0.06	0.06	0.05
	i^2	$t(0)$	0.024	0.026	0.05	0.06	0.06	0.07	0.09	0.12
		$t(\hat{\rho}^*, \kappa_1)$	0.040	0.040	0.05	0.05	0.05	0.05	0.05	0.05
		$t(\hat{\rho}^*, \kappa_2)$	0.040	0.040	0.05	0.05	0.05	0.06	0.05	0.05
5	1	$t(0)$	0.001	0.009	0.05	0.07	0.08	0.11	0.18	0.24
		$t(\hat{\rho}^*, \kappa_1)$	0.001	0.014	0.04	0.05	0.05	0.06	0.06	0.05
		$t(\hat{\rho}^*, \kappa_2)$	0.003	0.026	0.05	0.06	0.07	0.07	0.07	0.05
	$\sqrt{i}$	$t(0)$	0.001	0.012	0.05	0.06	0.08	0.11	0.16	0.22
		$t(\hat{\rho}^*, \kappa_1)$	0.003	0.018	0.04	0.04	0.05	0.06	0.05	0.05
		$t(\hat{\rho}^*, \kappa_2)$	0.007	0.030	0.05	0.06	0.07	0.07	0.06	0.05
	i^2	$t(0)$	0.015	0.027	0.05	0.06	0.07	0.08	0.13	0.17
		$t(\hat{\rho}^*, \kappa_1)$	0.023	0.030	0.04	0.04	0.05	0.05	0.05	0.05
		$t(\hat{\rho}^*, \kappa_2)$	0.028	0.037	0.05	0.05	0.06	0.06	0.06	0.05

(*) The notation 0.001 for $\hat{\alpha}$ should be understood as $\hat{\alpha} \leq 0.001$.

Table1: continued.

$\alpha = 0.05$ nominal significance level			realized significance level $\hat{\alpha}$							
			ρ							
			< 0		≥ 0					
n	λ_i	test stat.	$-\frac{1}{n}$	$-\frac{1}{2(n-1)}$	0.0	0.05	0.1	0.2	0.5	1.0
10	1	$t(0)$	0.001	0.009	0.05	0.09	0.12	0.16	0.25	0.30
		$t(\hat{\rho}^*, \kappa_1)$	0.001	0.006	0.03	0.04	0.05	0.06	0.05	0.05
		$t(\hat{\rho}^*, \kappa_2)$	0.001	0.021	0.05	0.07	0.08	0.08	0.06	0.05
	$\sqrt{i}$	$t(0)$	0.001	0.013	0.05	0.08	0.12	0.15	0.23	0.29
		$t(\hat{\rho}^*, \kappa_1)$	0.001	0.009	0.03	0.04	0.05	0.06	0.05	0.05
		$t(\hat{\rho}^*, \kappa_2)$	0.001	0.025	0.05	0.06	0.08	0.08	0.06	0.05
	i^2	$t(0)$	0.010	0.027	0.05	0.07	0.08	0.12	0.18	0.25
		$t(\hat{\rho}^*, \kappa_1)$	0.012	0.022	0.04	0.04	0.05	0.05	0.05	0.05
		$t(\hat{\rho}^*, \kappa_2)$	0.019	0.034	0.05	0.05	0.06	0.06	0.05	0.05
25	1	$t(0)$	0.001	0.009	0.05	0.14	0.19	0.25	0.34	0.37
		$t(\hat{\rho}^*, \kappa_1)$	0.001	0.002	0.02	0.04	0.05	0.06	0.05	0.05
		$t(\hat{\rho}^*, \kappa_2)$	0.001	0.009	0.04	0.07	0.08	0.08	0.06	0.05
	$\sqrt{i}$	$t(0)$	0.001	0.012	0.05	0.13	0.18	0.23	0.32	0.36
		$t(\hat{\rho}^*, \kappa_1)$	0.001	0.003	0.02	0.04	0.05	0.06	0.05	0.05
		$t(\hat{\rho}^*, \kappa_2)$	0.001	0.014	0.04	0.07	0.08	0.08	0.05	0.05
	i^2	$t(0)$	0.007	0.027	0.05	0.10	0.14	0.20	0.28	0.33
		$t(\hat{\rho}^*, \kappa_1)$	0.004	0.012	0.02	0.04	0.05	0.06	0.05	0.05
		$t(\hat{\rho}^*, \kappa_2)$	0.010	0.024	0.04	0.05	0.06	0.07	0.05	0.05

3 Proof of the Results

Denote $\mathbf{A}^T$ the transpose of a real matrix $\mathbf{A}$, $\mathbf{I}$ the $(n \times n)$–identity matrix and $\mathbf{1}$ the $(n \times 1)$ vector of ones, i.e. $\mathbf{1}^T = (1, \ldots, 1)_{1 \times n}$. With respect to distributional properties of linear and quadratic forms let us in general refer for instance to Mathai and Provost (1992).

Putting now, c.f. (2.1),

$$(3.1) \qquad \mathbf{x} \;=\; (t_1, \ldots, t_n)^T$$

we get under H_0 the covariance matrix of $\mathbf{x}$

$$(3.2) \qquad \mathbf{Cov}(\mathbf{x}) \;=\; (1-\rho)\mathbf{I} + \rho \mathbf{1}\mathbf{1}^T \;,$$

and with $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)^T$ for the variance of $\boldsymbol{\lambda}^T\mathbf{x}$:

$$(3.3) \qquad \begin{aligned} \mathrm{Var}(\boldsymbol{\lambda}^T\mathbf{x}) &= \boldsymbol{\lambda}^T \mathbf{Cov}(\mathbf{x}) \boldsymbol{\lambda} \\ &= (1-\rho)\boldsymbol{\lambda}^T \mathbf{I} \boldsymbol{\lambda} + \rho \cdot \boldsymbol{\lambda}^T \mathbf{1}\mathbf{1}^T \boldsymbol{\lambda} \\ &= (1-\rho) \sum_{i=1}^n \lambda_i^2 + \rho \cdot \left(\sum_{i=1}^n \lambda_i \right)^2 \;, \end{aligned}$$

which yields (2.2). Let us define now the projection matrix

$$(3.4) \qquad \mathbf{K} \;=\; \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T \quad , \; \mathbf{K} = \mathbf{K}^T \; , \; \mathbf{K} = \mathbf{K}^2 \; ,$$

then for the quadratic form q of (2.3) we have the equivalent representation

$$(3.5) \qquad q \;=\; \frac{1}{n-1} \mathbf{x}^T \mathbf{K} \mathbf{x} \; ,$$

such that by the following property

$$\begin{aligned}\mathbf{K}\cdot\mathbf{Cov}(\mathbf{x}) &= (1-\rho)\mathbf{K}+\rho\cdot\mathbf{K}\cdot\mathbf{1}\mathbf{1}^T\\ &= (1-\rho)\mathbf{K}+\rho\cdot(\mathbf{I}-\frac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{1}\mathbf{1}^T\\ &= (1-\rho)\mathbf{K}+\rho\cdot(\mathbf{1}-\frac{1}{n}\mathbf{1}\cdot n)\mathbf{1}^T\\ &= (1-\rho)\mathbf{K}\ ,\end{aligned}$$

with $trace(\boldsymbol{K}) = n-1$, we get under H_0

$$(3.6)\qquad \frac{1}{1-\rho}\cdot(n-1)q \ \stackrel{H_0}{\sim}\ \chi^2_{n-1}\ ,$$

which, with $\gamma_1+\gamma_2=\gamma$, gives the $(1-\gamma)$–confidence intervall for $1/(1-\rho)$:

$$(3.7)\qquad \frac{1}{(n-1)\cdot q}\cdot\chi^2_{n-1;\gamma_1} \ \leq\ \frac{1}{1-\rho}\ \leq\ \frac{1}{(n-1)\cdot q}\cdot\chi^2_{n-1;1-\gamma_2}\ ,$$

respectively for $(1-\rho)$:

$$(3.8)\qquad \frac{(n-1)\cdot q}{\chi^2_{n-1;1-\gamma_2}} \ \leq\ 1-\rho\ \leq\ \frac{(n-1)\cdot q}{\chi^2_{n-1;\gamma_1}}\ ,$$

yielding now directly the $(1-\gamma)$–confidence intervall for ρ given in (2.4).

Now by (3.6), under H_0, we get for the expectation

$$(3.9)\qquad \begin{aligned}&\mathrm{E}\left(\frac{1}{1-\rho}(n-1)q\right) = n-1\ ,\quad \text{respectively}\\ &\mathrm{E}\,q = 1-\rho\ ,\end{aligned}$$

and for the variance

$$(3.10)\qquad \begin{aligned}2(n-1) &= \mathrm{Var}\left(\frac{1}{1-\rho}(n-1)q\right)\\ &= \frac{1}{(1-\rho)^2}\cdot(n-1)^2\cdot\mathrm{Var}(q)\ ,\quad \text{respectively}\\ \mathrm{Var}(q) &= 2\cdot\frac{1}{n-1}\cdot(1-\rho)^2\ ,\end{aligned}$$

an unbiased estimator which is given by, cf. Hartung and Voet (1986),

$$\text{(3.11)} \quad \widehat{\mathrm{Var}(q)} \;=\; 2 \cdot \frac{1}{n+1} q^2 \;.$$

Now $\mathrm{Var}(q) = \mathrm{Var}(1-q)$, so that (2.5), (2.6) and (2.7) are also shown.

Finally, although under H_0 in (2.2) we may use the unbiased estimator (2.5) for ρ, we have to recognize the square root function to be concave, so that in expectation the denominator would be underestimated. Therefore in (2.8) the estimate of ρ is corrected by adding a 'small amount' of its estimated standard deviation under H_0.

4 Final Remark

In this paper we have shown how to modify the 'inverse normal method' of non–parametric meta–analysis in order to include the case of possibly dependent test statistics resp. dependent p–values. In the worst case one can take at least a weighted mean of the probits.

So we recommend to take our procedure in consideration if one is not absolutely sure whether the statistics to be combined are really independent.

Acknoledgement: Thanks are due to Boris Weimann, University of Dortmund, for carrying out the simulations.

References

Birnbaum, A. (1954). Combining independent tests of significance. *Journal of the American Statistical Association*, **49**, 559–547.

Hartung, J. and Voet, B. (1986). Best invariant unbiased estimators for the mean squared error of variance component estimators. *Journal of the American Statistical Association*, **81**, 689–691.

Hedges, L. V. and Olkin, I. (1985). *Statistical methods for meta–analysis*. Academic Press, Orlando (Flor.).

Liptak, T. (1958). On the combination of independent tests. *Magyar Tudomanyos Akademia Matematikai Kutato Intezetenek Kozlemenyei*, **3**, 171–197.

Mathai, A. M. and Provost, S. B. (1992). *Quadratic forms in random variables*. Marcel Decker, New York.

van Zwet, W. R. and Oosterhoff, J. (1967). On the combination of independent test statistics. *Annals of Mathematical Statistics*, **38**, 659–680.