A Nonparametric Approach of Estimating the Number of True Null Hypotheses in Multiple Testing

Ma, Mi-Chia
Department of Statistics, National Cheng Kung University,
No. 1, University Road
Tainan City 70101, Taiwan R..O.C.
E-mail: mcma@stat.ncku.edu.tw

Chao, Weng-Chang
Department of Statistics, National Cheng Kung University,
No. 1, University Road
Tainan City 70101, Taiwan R..O.C.
E-mail: wcchao@stat.ncku.edu.tw

1 Introduction

For a single hypotheses problem, the maximal tolerable type I error rate (or called significance level) is chosen by experimenters in early planning stage. It is often suggested using the multiple comparison procedures (MCP) to control the type I error rate- so called familywise error rate (FWER) in literature, where FWER is the probability of rejecting at least one true null hypothesis in the given family of the hypothesis tests. When many statistical tests are simultaneously conducted, the chance of any false positive finding or FWER inflates as the number of hypotheses. Hochberg & Tamhane (1987) proposed that the problem of using MCP is too conservative, such that it often substantially reduces the power to detect a difference when the number of testing hypotheses is large. Benjamini & Hochberg (1995) introduced a multiple hypothesis testing error measure, called the false discovery rate (FDR). This quantity is the expected proportion of false positive findings among all the rejected hypotheses. Hereafter, Benjamini & Liu (1999) proposed sequential procedures to control the FDR. Benjamini & Hochberg (2000) proposed that lack of multiplicity control is too permissive and the protection resulting from controlling the FWER is too restrictive. Many papers have focused on FDR under independence or dependence assumption of test statistics (Storey, 2007; Friguet et al., 2009).

No matter to improve the statistical power of a MCP or have more precise results in FDR estimation, the estimation of the number of true null hypotheses, m_0, is important. There are some literatures about estimating m_0, e.g., Schweder and Spjψtvoll (1982), Storey (2002), Benjamini and Hochberg (2000), Hsueh, Chen and Kodell (2003). In section 2, five estimation methods are reviewed. In section 3, a nonparametric approach based on the McNemar test is presented. Furthermore, the statistical properties are also explored. In section 4, simulations studies are conducted. The means, standard deviations, and mean square errors of the estimates are used to express the behavior of different methods in simulation. All estimation methods are compared. Discussion and final remarks are provided in section 5.

2 Literature Review

The problem of simultaneously testing m null hypotheses H_{0i} $(i = 1,2,...,m)$ is considered. Let m_0 be the number of true null hypotheses and R be the number of significances declared. The mentioned symbols are summarized in Table 1. In Table1 except that m and R have already known, U, V, S and T are all unknown, and random variable R will increase as the significance level α increases. Hereafter, the capitalization of every symbol will be used to represent the random variable and small letter represent observed value.

Table 1 The result and probability of occurrence in testing m hypotheses.

	Declared Non-Significant	Declared Significant	Total
Null True	S $(1-\alpha)$	V (α)	m_0
Alternative True	T (β)	U $(1-\beta)$	$m-m_0$
Total	$m-$R	R	m

Suppose that H_{0i} is rejected when a test statistic Z_i is large. Let F_i be the cumulative distribution function of Z_i under H_{0i}. The p-value, i.e. significance probability, for the hypothesis H_{0i} is $p_i = 1 - F_i(z_i)$, with possible correction if the distribution is discrete (Cox, 1977). The distribution of Z_i is assumed completely known when H_{0i} is true, so that p_i does not depend upon unknown parameters. The following procedure will be based upon the observed significance probabilities $p_1,..., p_m$. If H_{0i} is true, the significance probability p_i is uniformly distributed on the interval (0,1). If H_{0i} is not true, p_i will tend to have small values.

Suppose that the corresponding p-value of the true null hypotheses H_{0i} is p_i, let $p_{(1)} \le p_{(2)} \le \cdots \le p_{(m)}$ be the ordered statistic of p_1, p_2,…, p_m, and $H_{(0i)}$ is the corresponding null hypotheses of $p_{(i)}$, hence FWER and FDR can be expressed as follows:

FWER=P(reject at least one null hypothesis $|H_{0i}$ is true, $i = 1,...,m$)=P(V ≥ 1),

FDR=$\sum_{i=1}^{m} P_i$ I(H_{0i} is true)/R = E(V)/R ,

The properties of FWER, FDR had stated by Benjamini and Hochberg (2000). About the estimation of m_0, the methods in literature are stated as follows.

(1) Schweder and Spjψtvoll's Method

For a relatively small β, the expected number of nonsignificant hypotheses can be approximated as

$$m - r(\alpha) \approx \mathrm{E}((m-R)|\alpha) \approx m_0(1-\alpha), \quad (1)$$

where $r(\alpha)$ is the observed number of rejections at level α. The number of rejections at the level $p_{(i)}$ is exactly i, that is, $r(p_{(i)}) = i$. Schweder and Spjψtvoll (1982) considered the cumulative plot of observed $1-p_{(i)}$ against $m-i$, $i = 1,...,m$. The procedure starts from $i = m$ and decreases one in each successive calculation. Because Hsueh, Chen and Kodell (2003) proposed that the method of Schweder and Spjψtvoll (1982) has the worst performance, so this method is not considered in the simulation.

(2) Storey's Method (ST)

Storey (2002) proposed $\hat{m}_0^{ST} = \{m - r(\lambda)\}/(1-\lambda)$ to estimate the slope directly based on Equation (1) where λ ideally is the change point of the p-values between true null and true alternative hypotheses. A bootstrapping procedure was suggested for the optimal λ in his paper. $\lambda = 0.5$ suggested by Hsueh et al. (2003) will be used in empirical evaluation of the Storey (ST) method.

(3) Benjamini and Hochberg's Lowest Slope Method (LSL)

Benjamini and Hochberg (2000) proposed the ordinary least squares estimator of the slope of the line restricted to pass through the point $(m+1,1)$, and the Lowest Slope (LSL) estimator $(1-p_{(i)})/(m+1-i)$ which is the slope of the line passing through the points $(m+1,1)$ and $(i, p_{(i)})$. The LSL method is given as: (a) Calculate $S_i = (1-p_{(i)})/(m+1-i)$, the i-th slope estimate. (b) Starting with $i = 1$, proceed towards larger i as long as $S_i \ge S_{i-1}$. (c) Stop when the first time $S_j < S_{j-1}$, and obtain the estimate.

$$\hat{m}_0 = min[(1/S_j + 1), m], \quad (2)$$

(4) Mean of Differences Method (MD)

As mentioned in Benjamini and Hochberg (2000), Hsueh, Chen and Kodell (2003) proposed the LSL

estimator can be derived in view of the difference $d_{(i)} = p_{(i)} - p_{(i-1)}$, $i = m - m_0 + 2, ..., m+1$, $p_{(0)} = 0$, $p_{(m+1)} = 1$. The differences $d(i)'s$ are identically $\text{Beta}(1, m_0)$ distributed and have the common mean $E(D) = 1/(m_0+1)$. Thus, m_0 can be estimated as

$$\hat{m}_0^{MD} = 1/\overline{d}_{m_0} - 1 \approx 1/E(D) - 1, \quad (3)$$

where $\overline{d}_{m_0} = \sum_{i=m+2-m_0}^{m+1} d_i / m_0 = \{1 - p_{(m-m_0+1)}\}/m_0$.

To have a conservative estimate, replacing m_0 with m in $\overline{d}_{m_0}$ is required and the search starts from $j = m$ with $\overline{d}_m$ and j proceeds downward. The search stops and $\hat{m}_0^{MD} = \hat{m}_0^{MD(j)}$ when the first time $\hat{m}_0^{MD(j-1)} \geq \hat{m}_0^{MD(j)}$.

(5) Least Squares Method (LS)

An alternative least squares (LS) estimation is described by Hsueh, Chen and Kodell (2003). At a given significance level α, the probability of rejection of the ith null hypothesis is $\{1 - \beta(\gamma_i, \alpha)\}$, where γ_i is the true mean under the ith alternative hypothesis. The expected number of declared significances is

$$\text{E}\{R(\alpha)\} = \sum_{i=1}^{m_0} \{1 - \beta(\gamma_i, \alpha)\} + \sum_{i=m_0+1}^{m} \{1 - \beta(\gamma_i, \alpha)\} = m_0\alpha + (m - m_0)\overline{\beta}(\alpha),$$

where $\overline{\beta}(\alpha) = \sum_{i=m_0+1}^{m} \{1 - \beta(\gamma_i, \alpha)\}/(m - m_0)$ is the average power among the $(m - m_0)$ nontrue null hypotheses tested at the α level. Given $p_{(1)}, ..., p_{(m)}$, and $\overline{\beta}(p_{(i)})$, the expected number of rejected hypotheses at significance level $p_{(i)}$ is $\text{E}(i) = m_0 p_{(i)} + (m - m_0)\overline{\beta}(p_{(i)})$. Thus, an estimate of m_0 can be obtained by minimizing the sum of squares $\sum_{i=1}^{m} (i - m_0 p_{(i)} - (m - m_0)\overline{\beta}(p_{(i)}))^2$, and is given as

$$\hat{m}_0^{LS} = \sum_{i=1}^{m} x_i y_i \Big/ \sum_{i=1}^{m} x_i^2, \quad (4)$$

where $y_i = i - m\overline{\beta}(p_{(i)})$ and $x_i = p_{(i)} - \overline{\beta}(p_{(i)})$.

3 Proposed Statistical Procedure

3.1 Proposed Method

The McNemar test is used to obtain the estimate of m_0 by replacing V and T in Table 1 by αm_0 and $(m - m_0 - \text{R} + \alpha m_0)$ respective. Therefore, the McNemar test can be shown as

$$Z^2 = \frac{(\text{V} - \text{T})^2}{\text{V} + \text{T}} = \frac{[\alpha m_0 - (m - m_0 - \text{R} + \alpha m_0)]^2}{\alpha m_0 + (m - m_0 - \text{R} + \alpha m_0)} = \frac{(m - m_0 - \text{R})^2}{m - m_0 - \text{R} + 2\alpha m_0}$$

where $\text{R} = \sum_{i=1}^{m} I(p_i < \alpha)$. Let $Z^2 \leq \chi^2_{1,\alpha}$, where $\chi^2_{1,\alpha}$ is the 100α th upper percentile of chi-square distribution with 1 degree of freedom, the equation will be obtained below. When the equality holds, we have

$$m_0^2 - [2(m - \text{R}) + 2\chi^2_{1,\alpha}\alpha - \chi^2_{1,\alpha}]m_0 + (m - \text{R})^2 - \chi^2_{1,\alpha}(m - \text{R}) = 0.$$

The m_0 can be estimated by solving the equation, hence $\hat{m}_0^{MC} = \text{A} \pm B$. Where

$A = (m - R) + (\alpha - 0.5)\chi^2_{1,\alpha}$ and $B = \sqrt{(\chi^2_{1,\alpha})^2(\alpha - 0.5)^2 + 2\chi^2_{1,\alpha}\alpha(m - R)}$.

Here the part of plus sign and minus sign will be denoted by $\hat{m}_0^{MC(+)}$ and $\hat{m}_0^{MC(-)}$. To avoid $\hat{m}_0^{MC} > m$ or $\hat{m}_0^{MC} < 0$. Hence, the $\hat{m}_0^{MC(+)}$ and $\hat{m}_0^{MC(-)}$ can be adjusted as

$$\hat{m}_0^{MC(+)} = \min\{\text{A} + B, m\} \quad (5)$$

and

$$\hat{m}_0^{MC(-)} = \max\{0, \text{A} - B\} \quad (6)$$

The method of MCC is that the estimator $\hat{m}_0^{MCA}$=A is used.

3.2 Applying $\hat{m}_0$ to Multiple Testing Procedure

The power of Bonferroni-type multiple testing procedures can be improved by the estimates of m_0. If all hypotheses are independent, then testing each individual hypothesis at level α will have the FWER $Pr(\mathrm{V} \geq 1) = 1 - Pr(\mathrm{V} = 0) = 1 - (1-\alpha)^{m_0} = \alpha_0$. Hence, under the Bonferroni-type multiple testing procedures each individual test will be required to be rejected at the level $\alpha = 1 - (1-\alpha_0)^{1/m_0}$ for a FWER-controlling test set at the level α_0. On the contrary, if all test statistics are dependent, then testing each individual hypothesis at the level α_0 / m_0 and that will have FWER< α_0, proposed by Hsueh et al. (2003).

4 Simulation Studies

In this section, a simulation study was conducted to compare the performance of the number of true null hypotheses among seven different methods. Fortran 90 and IMSL's STAT/LIBRARY Fortran subroutines were used in the simulation study. In order to compare the behaviors of different method, the simulated data is generated using the simulated method proposed Hsueh et al. (2003).

4.1 Simulation procedure

(A) Independent model

The data is generated from the problem in terms of testing m univariate means of an m-variate normal random vector, $H_{oi} : \mu_i = 0$ *vs.* $H_{1i} : \mu_i \neq 0$, $i = 1, ..., m$. The detail of the simulation process will be described below.

1. The univariate normal random variables are independent and each has variance 1. The m_0 true null variables were generated from a m_0-variate normal random vector with zero mean vector.
2. The nontrue variables, a nonzero effect size, γ, is added to each random variate. Two alternative models were considered for the effect sizes: (a) a simple alternative model, in which the effect size is constant $\gamma = \gamma_0$ (b) a multiplicity alternative model, in which the effective size γ is generated from a truncated normal distribution $C \cdot N(\gamma_0, 1) I\{\gamma > 0\}$ with some normalizing constant $C > 0$.
3. The parameter γ_0 has two cases $\gamma_0 = 2$ and γ_0 have 80% power, $1 - \beta(\gamma_0, \alpha) = 1 - \Phi(\Phi^{-1}(1-\alpha/2) - \gamma_0) + \Phi(\Phi^{-1}(-\alpha/2) - \gamma_0) \approx 1 - \Phi(\Phi^{-1}(1-\alpha/2) - \gamma_0)$. The individual α was set to ensure FWER=0.25 under independent model. That is, $\alpha = 1 - (1-0.25)^{1/m_0}$.
4. A constraint $\hat{m}_0 \leq m$ is given in solving each estimate. The simulation trials are repeated 10,000 times and the sample means, standard deviations (SD) and the root mean square error (RMSE) of the number of true hypotheses under independent models at the nominal significance level of α for these methods were calculated.

(B) Equicorrelated model

Similar steps are conducted for equicorrelated model, only the generated data in step 1 is different from independent model in simulation study. The pairwise correlations of the normal random variables for the true hypotheses are $\sqrt{0.2}$, the same as the pairwise correlations for the nontrue hypotheses. The correlations between the true variables and the nontrue variables are set to be zero in equicorrelated model.

4.2 Simulation Results

$m = 1000, 500$; $m_0 = 0.9m$ and m are chosen for comparisons of seven methods. The parameter γ_0

is set to be $\gamma_0 = 2$ and have 80% power. There are 4 combinations for different γ_0 and model. For saving space, only the simulated results of $m = 1000$ and $\gamma_0 = 2$ are showed. Table 2 is based on the simple alternative model and multiplicity alternative model. The method of MCC(+) is the estimator $\hat{m}_0^{MC(+)}$ and the method of MCC(-) is the estimator $\hat{m}_0^{MC(-)}$. In LS method, in order to avoid the mistake of m_0 unknown, so $\hat{m}_0{}^{MD}$ is suggested to instead of m_0. Generally speaking, the MCC(+), MCC(-) and the MCC estimators have the most desired performance, least bias and variation and RMSE as $m_0 \neq m$. The LSL, MD and LS have smaller RMSE than the proposed method and ST has larger RMSE than LSL, MD and LS as $m_0 = m$. For saving space, the simulated results of empirical FWERs are summarized below. The empirical FWERs of LS method often exceed the nominal level as $m_0 \neq m$. On the other hand, the proposed methods always well control the FWER. Generally speaking, the proposed methods perform well, and the difference between simple and multiplicity alternative is small.

5 Conclusion

The FWER approaches have been proposed to control an error rate in multiple hypothesis testing. Here seven methods are used to evaluate the number of true null hypotheses in two-sided test, and consider if they properly control the FWER. These different methods have the similar results for simple alternative and multiple alternatives. The MCC(+), MCC(-) and MCC have least bias, variation and root mean square error as $m_0 \neq m$, and perform well to control the FWER among the seven methods considered. The proposed estimation has the advantage of directly computing. It is not like the other estimation methods that need to compute iteratively. Besides, MCC(-) and MCC(+) can be regarded as the confidence lower limit and upper limit of the number of true null hypotheses. Hence, the large sample properties may further be researched in the future.

REFERENCES

1. Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery reat: a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* B 57:289-300.
2. Benjamini, Y., Liu, W. (1999). A step down multiple hypotheses testing procedure that controls the false discovery rate under independence. *J. Statist. Plann. Inference* 82:163-170.
3. Benjamini, Y., Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Statist.* 25:60-83.
4. Cox, D. R. (1977). The role of significance tests. *Scand. J. Statist.* 4, 49-70.
5. Hochberg, Y., Tamhane, A. C. (1987). *Multiple Comparison Procedures*. New York. John Wiley & Sons.
6. Hsueh, H. M., Chen, J. J. and Kodel, R. L. (2003). Comparison of methods for estimating the number of true null hypotheses in mulitplicity testing. *Journal of Biopharmaceutical Statistics*, 13, 675-689.
7. Schweder, T., Spjψtvoll, E. (1982). Plots of p-values to evaluate many tests simultaneously. *Biometrika* 69:493-502.
8. Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Statist. Soc. B* 64:479-498.
9. Storey, J. D. (2007). The optimal discovery procedure: a new approach to simultaneous significance testing. *J. R. Statist. Soc. B* 69:347-368.
10. Friguet, C., Kloareg, M. and Causeur, D. (2009) A factor model approach to multiple testing under dependence. *JASA,* 104:1406-1415.

Table 2 Comparisons of various estimations for m_0, the number of true null hypotheses in two-sided test. The parameter $\gamma_0 = 2$. FWER=0.25

			Simple alternative model							
			Independent				Equicorrelated			
m	m_0	Method	Mean	SD	RMSE	FWER	Mean	SD	RMSE	FWER
1000	1000	ST	987.46	18.05	21.98	0.247	952.28	129.02	140.52	0.220
		LSL	999.84	0.63	0.65	0.250	994.56	48.47	48.97	0.246
		MD	998.79	1.34	1.80	0.249	994.42	37.89	38.39	0.246
		LS	997.07	5.19	6.08	0.249	987.50	42.59	44.89	0.237
		MCC(+)	967.39	6.95	33.33	0.242	961.66	39.61	55.25	0.237
		MCC(-)	929.11	6.81	71.14	0.234	928.95	42.00	82.45	0.232
		MCC	948.34	6.88	52.16	0.238	945.38	40.35	67.92	0.234
	900	ST	917.68	29.87	34.71	0.249	891.88	129.57	130.65	0.235
		LSL	978.46	8.01	78.87	0.251	949.30	55.04	76.91	0.248
		MD	976.05	8.07	76.48	0.251	952.63	44.42	70.38	0.247
		LS	971.99	7.86	72.48	0.262	954.62	45.83	72.54	0.266
		MCC(+)	920.39	8.32	21.98	0.243	903.82	44.59	47.30	0.242
		MCC(-)	883.05	8.15	18.77	0.234	867.24	44.15	56.98	0.232
		MCC	901.79	8.24	8.41	0.238	885.61	44.33	49.03	0.237
			Multiplicity alternative model							
			Independent				Equicorrelated			
m	m_0	Method	Mean	SD	RMSE	FWER	Mean	SD	RMSE	FWER
1000	1000	ST	945.26	140.84	153.62	0.247	942.24	145.76	158.90	0.220
		LSL	993.51	53.09	53.64	0.250	993.06	54.96	55.52	0.246
		MD	984.59	48.82	51.68	0.249	987.35	47.08	49.15	0.246
		LS	985.67	46.47	49.10	0.249	984.84	48.06	50.79	0.237
		MCC(+)	960.58	43.26	58.66	0.242	960.11	44.75	60.06	0.237
		MCC(-)	928.99	45.91	84.53	0.234	929.01	47.48	85.41	0.232
		MCC	944.86	44.09	70.65	0.238	944.64	45.60	71.79	0.234
	900	ST	886.37	143.27	144.68	0.250	883.58	148.08	149.59	0.235
		LSL	934.29	58.26	70.42	0.250	927.28	58.92	67.20	0.248
		MD	944.70	52.61	70.33	0.250	940.35	51.24	66.26	0.246
		LS	945.98	50.07	69.19	0.261	941.68	51.59	67.29	0.264
		MCC(+)	894.08	47.01	49.26	0.243	889.97	47.68	50.13	0.242
		MCC(-)	850.85	7.00	49.65	0.234	845.72	38.51	66.65	0.232
		MCC	876.00	46.66	54.19	0.238	871.92	47.29	56.29	0.237