Communications for Statistical Applications and Methods
2020, Vol. 27, No. 1, 141–148

https://doi.org/10.29220/CSAM.2020.27.1.141
Print ISSN 2287-7843 / Online ISSN 2383-4757

Comparison of methods for the proportion of true null hypotheses in microarray studies

Joonsung Kang[1,a]

[a]Department of Information Statistics, Gangneung-Wonju National University, Korea

Abstract

We consider estimating the proportion of true null hypotheses in multiple testing problems. A traditional multiple testing rate, family-wise error rate is too conservative and old to control type I error in multiple testing setups; however, false discovery rate (FDR) has received significant attention in many research areas such as GWAS data, FMRI data, and signal processing. Identify differentially expressed genes in microarray studies involves estimating the proportion of true null hypotheses in FDR procedures. However, we need to account for unknown dependence structures among genes in microarray data in order to estimate the proportion of true null hypothesis since the genuine dependence structure of microarray data is unknown. We compare various procedures in simulation data and real microarray data. We consider a hidden Markov model for simulated data with dependency. Cai procedure (2007) and a sliding linear model procedure (2011) have a relatively smaller bias and standard errors, being more proper for estimating the proportion of true null hypotheses in simulated data under various setups. Real data analysis shows that 5 estimation procedures among 9 procedures have almost similar values of the estimated proportion of true null hypotheses in microarray data.

Keywords: proportion of true null hypotheses, HMM, microarray

1. Introduction

We introduce a DNA microarray data that possesses gene expression levels in thousands of genes (Speed, 2003; Baldi and Hatfield, 2002). This study then simultaneously examines differentially expression genes among thousands of genes, which involves an appropriate simultaneous testing per each gene. The null hypothesis is to announce no association between gene expression levels and explanatory variables (Speed, 2003). For instance, microarray analysis could be conducted to examine differences in gene expression levels between cancer patients and healthy patient.

Applying the multiple testing framework to a microarray data, a true null hypothesis indicates no differentially expressed gene, whereas a non-true null hypothesis is a truly differentially expressed gene. Rejected hypothesis (gene) implies that this specific gene is declared as a differentially expressed gene (Table 1).

An older method in multiple testing framework is the family-wise error rate (FWER) defined as the probability of having any type I error among all hypotheses at assigned level α. However, it is too conservative to use when dealing with thousands of hypotheses (genes) as well as inadequate to test thousands of highly dependent genes (Benjamini and Yekutieli, 2001). The alternative to this FWER is the false discovery rate (FDR) which Benjamini and Hochberg (1995) introduced. It is

[1] Department of Information Statistics, Gangneung-Wonju National University, Jukheon-gil 7, Gangneung-si 25457, Republic of Korea. E-mail: mkang@gwnu.ac.kr

Published 31 January 2020 / journal homepage: http://csam.or.kr

Table 1: Multiple hypothesis testing

	Not rejected	Rejected	Total
True null	U	V	m_0
Non-true null	T	S	$m - m_0$
Total	$U + T = m - R$	$V + S = R$	m

defined as the expected proportion of type I errors among the rejected hypotheses (genes) such as $E(V/R|R > 0) \cdot P(R > 0)$. The proportion of true null hypotheses should be estimated in advance in order to control the FDR.

A dependent structure among the genes should be taken into account for the microarray data. Many researchers have developed various estimation procedures to assume a restricted dependent structure among genes and have estimated the proportion of true null hypotheses in a restricted or unrealistic manner. A dependent structure among genes in microarray studies is often unknown. The hidden Markov model (HMM) model exploits the local dependence structure and has been widely used in areas such as speech recognition, signal processing and DNA sequence analysis, see Rabiner (1989), Churchill (1992), Krogh *et al.* (1994), and Ephraim and Merhav (2002), among others (Sun and Cai, 2009). The HMM model is known to be more similar to the dependence structure among gene expression levels in microarray studies because it accounts for the observations in adjacent locations among genes (Sun and Cai, 2009). This study compares the performance of different estimation procedures for the proportion of true null hypotheses (genes) under realistic dependency structures. The HMM model known as a statistical Markov model is utilized when we believe that the system being modeled has a Markov process with unobserved (hidden) states. The HMM is shown to be an effective mechanism for modeling the dependence structure among genes. In a multiple testing framework, the HMM model shows that the sequence of the hidden states, $(\theta_i)_1^m = (\theta_1, \ldots, \theta_m)$ follows a Markov chain. $\theta_i = 1$ if the i^{th} hypothesis is non-null and $\theta_i = 0$ if the i^{th} hypothesis is true-null. Observed data $x = (x_1, \ldots, x_m)$ are independently created conditionally on the hidden states $(\theta_i)_1^m$. See Sun and Cai (2009) for more details.

We consider the 9 most popular estimation procedures of the proportion of true null hypotheses as below. The least slope method (Benjamini and Hochberg, 2000), the smoother method described in Storey and Tibshirani (2003), the bootstrap method (Storey *et al.*, 2004), the Langaas method using a convex decreasing density estimate for p-values (Langaas *et al.*, 2005), the histogram method (Nettleton *et al.*, 2006), the average estimate method (Jiang and Doerge, 2008), the sliding linear model (SLIM) method of Wang *et al.* (2011) using a sliding linear model, the estimation method described in Jin and Cai (2007), and the robust method proposed by Pounds and Cheng (2006) are compared under the independent simulated data, the HMM model simulated data, and real microarray data. We compute estimates and standard errors in simulated data to evaluate the performance of each estimation procedure. We also conduct a real data analysis with the 9 procedures.

Section 2 introduces different estimation procedures. In Section 3, simulation studies are conducted under independence and the HMM dependence structure by comparing the procedures. In Section 4, real data analysis is tested with the value of each estimation procedure. The summary of this paper is devoted to the last Section 5.

2. 9 estimation procedures

The hypotheses, $H_1, \ldots, H_m$ correspond to p-values, $P_1, \ldots, P_m$. The p-values are assorted in ascending order denoted as $P_{(1)}, \ldots, P_{(m)}$. The least slope method (Benjamini and Hochberg, 2000) utilizes

the Lowest SLope (LSL) estimator $(1 - p_{(i)})/(m + 1 - i)$ as the slope of the line passing through the points $(m + 1, 1)$ and $(i, p_{(i)})$. We follow the steps:

- Calculate $S_i = (1 - p_{(i)})/(m + 1 - i)$, the i^{th} slope estimate.
- Starting with $i = 1$, proceed towards larger i as long as $S_i \geq S_{i-1}$, stop when the first time $S_j < S_{j-1}$, and use the proportion of true null hypotheses $\hat{\pi}_0 = \min[(1/S_j + 1), m]$.

The smoother method (Storey and Tibshirani, 2003) estimates the proportion of true null hypotheses as $\hat{\pi}_0 = \#\{p_i > \lambda\}/m(1 - \lambda)$ with the turning parameter λ.

The bootstrap method (Storey *et al.*, 2004) has the step as follows. Define $R(\lambda) = \#\{p_i : p_i \leq \lambda\}$ and $W(\lambda) = m - R(\lambda)$. We define the proportion of true null hypotheses as $\hat{\pi}_0 = W(\lambda)/\{(1 - \lambda)m\}$.

The rationale for this estimate is that p-values for true null hypotheses are uniformly distributed on the interval $(0, 1)$. They proposed a bootstrap method to automatically choose λ when estimating $\hat{\pi}_0(\lambda)$.

The Langaas method utilizes a convex decreasing density estimate for p-values (Langaas *et al.*, 2005). The p-values are independent and identically distributed random variables with mixture density

$$f(p) = \pi_0 + (1 - \pi_0)h(p), \quad 0 \leq p \leq 1.$$

We assume that h is decreasing on $[0, 1]$ with $h(1) = 0$, which implies that $\pi_0 = f(1)$. A way of estimating π_0 is through the estimation of f. Nonparametric estimators for f can be used with the special structure that we impose on f, namely decreasing property and convexity (and decreasing property) by requiring $h(p)$ to be convex in addition to decreasing with $h(1) = 0$.

By transforming the density f confined to $[p_0, 1]$ to a density f^* on $[0, 1]$ given by $f^*(p) = f = ((1 - p_0)/p + p_0) \cdot (1 - p_0)/r_0$, $0 \leq p \leq 1$ with $r_0 = \#\{p_i > p_0\}/m$, we transform the p-values to $p_i^* = (p_i - p_0)/(1 - p_0)$ with $p_i > p_0$, which gives us an estimated function $\hat{f}^*$ with the transformed p-values and a corresponding estimate $\hat{\pi}_0^{*c} = \hat{f}^*(1)$. The proposed estimate of π_0 is $\hat{\pi}_0^{*c} r_0/(1 - p_0)$.

As for the histogram method (Nettleton *et al.*, 2006), an iterative algorithm that depends on a histogram of observed p-values shown in order to obtain the estimator. The limit of that iterative algorithm is characterized and shows that the estimator could be computed directly without iteration.

The proposed method (Pounds and Cheng, 2006) does not depend on assumptions that the tests are two-sided or produce continuously distributed p-values. The proposed method is proven to be conservative and has desirable large-sample properties. We estimate π_0 as follows. We consider four cases: (a) p-values are two-sided and continuous, (b) p-values are two-sided and discrete, (c) p-values are one-sided and continuous, and (d) p-values are one-sided and discrete. $\bar{p} = (1/m) \sum_{i=1}^{m} p_i$ and $\bar{a} = (1/m) \sum_{i=1}^{m} a_i$, where $a_i = 2 \min(p_i, 1 - p_i)$. $\hat{\pi}_0$ is computed as $\min(1, 2\bar{p})$ for cases (a) and (b). For case (c), it is defined as $\min(1, 2\bar{a})$. For case (d), $\hat{\pi}_0$ is $\min(1, 8\bar{a})$.

For the average estimate method (Jiang and Doerge, 2008), we estimate the proportion of true null hypotheses $\pi(B)$ as follows. We define $0 = t_1 < t_2 < \cdots < t_B < t_{B+1} = 1$ as equally spaced points in the interval $[0, 1]$ such that the interval $[0, 1]$ is divided into B small intervals with equal length $1/B$. Let NB_i denote the number of p-values greater than or equal to t_i and NS_i represent the number of p-values in the interval of $[t_i, t_{i+1})$.

$$\hat{\pi}(B) = \frac{1}{B - i + 2} \sum_{j=i-1}^{j=B} NB_j(1 - t_j)m, \quad i = \min\left\{i : NS_i \leq \frac{NB_i}{B - i + 1}\right\}.$$

The value of π_0 is estimated by the average of the $\hat{\pi}_0(B)$ over the average of B for each $B \in I$ for some interval I.

From the basic non-linear model of the q-value method described in Storey (2002), a simple linear algorithm is developed to probe local dependence blocks, resulting in uncovering a non-static relationship between p-values and corresponding q-values influenced by the data structure and π_0. By an optimization framework, these findings were used to propose a SLIM as a more reliable estimate π_0 under dependent data (Wang *et al.*, 2011). Let $\gamma = \pi_0\lambda + \pi_1$. The (λ, γ) plot presents the cumulative probability distribution of p-values.

We calculate the slope of that fitting line as the estimated π_0. Thus, $\hat{\pi}_0 = (\gamma_e - \gamma_s)/(\lambda_e - \lambda_s)$ for a given range of $\lambda, \Delta = [\lambda_s, \lambda_e]$, $0.05 \leq \lambda_s < \lambda_e \leq 1$, where γ_s and γ_e represent the cumulative probabilities at λ_s and λ_e. For a uniform p-value distribution, the above equation may be applied directly for π_0 estimation. However, we use the following strategy with non-uniform p-value distributions in order to keep as much information as possible about the null hypotheses p-value distribution. We first divide the (λ, γ) plot into a series of λ-segments (S) where $S = \{s_i : s_i = [\lambda_i, \lambda_{i+1}]\}_1^n$, $0.05 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n+1} = 1$. We then linearly regress by λ with the slope equation for each segment and obtain n local estimates of $\hat{p}_0^i = \hat{\pi}_0(\lambda_i, \lambda_{i+1})$, $i = 1, \ldots, n$ in accordance with S.

Then the estimate $\pi_0 = D^1(\alpha)$ represents the cumulative distribution function of $\hat{p}_0^i$. D^1 represents its inverse (quantile) function, and $0 \leq \alpha \leq 1$ represents a given quantile point.

Jin and Cai (2007) develops an approach based on the empirical characteristic function and Fourier analysis. The estimators are shown to be uniformly consistent over a wide class of parameters. They extend their approach to dependent data structures. Please see Jin and Cai (2007) for more details.

2.1. Dependence

In order to model the HMM, we need to utilize the notion of transition matrix with two states (0 and 1). Transition matrix is varied as:

$$T1 = \begin{bmatrix} 0.90 & 0.10 \\ 0.20 & 0.80 \end{bmatrix}. \tag{2.1}$$

$$T2 = \begin{bmatrix} 0.90 & 0.10 \\ 0.50 & 0.50 \end{bmatrix}. \tag{2.2}$$

$$T3 = \begin{bmatrix} 0.90 & 0.10 \\ 0.80 & 0.20 \end{bmatrix}. \tag{2.3}$$

$$T4 = \begin{bmatrix} 0.50 & 0.50 \\ 0.20 & 0.80 \end{bmatrix}. \tag{2.4}$$

$$T5 = \begin{bmatrix} 0.50 & 0.50 \\ 0.50 & 0.50 \end{bmatrix}. \tag{2.5}$$

$$T6 = \begin{bmatrix} 0.50 & 0.50 \\ 0.80 & 0.20 \end{bmatrix}. \tag{2.6}$$

$$T7 = \begin{bmatrix} 0.10 & 0.90 \\ 0.20 & 0.80 \end{bmatrix}. \tag{2.7}$$

$$T8 = \begin{bmatrix} 0.10 & 0.90 \\ 0.50 & 0.50 \end{bmatrix}. \tag{2.8}$$

$$T9 = \begin{bmatrix} 0.10 & 0.90 \\ 0.80 & 0.20 \end{bmatrix}. \tag{2.9}$$

The transition probability that a system goes a movement from state 0 (no differentially expressed gene) to state 1 (differentially expressed gene) or from state 1 to state 0 is constant over time. In the matrix, diagonal terms are generally not transitions of states such as from state 0 to state 0 or from state 1 to state 1. $T1, T2, \ldots, T9$ possess different transition probabilities (instantaneous rates). For example, the probability moving from 0 to 1 in $T1$ is 0.90, that in $T4$ 0.50, and that in $T7$ 0.10.

The conditional probability of future states relies only upon the present state in the Markov chain. We could consider m hypotheses as a Markov chain with m states. The dependencies moving from state 0 in the present state to state 1 in the future state are sorted in descending order. If the present state (the i^{th} hypothesis) is 0, the probability (extent) moving from the current state to state 1 in future state (the $(i + 1)^{th}$ hypothesis) is 0.10 for $T1$. $T1, T2, \ldots, T9$ have different dependency patterns between 0 and 1. Please refer to Sun and Cai (2009) for more details.

3. Numerical analysis

3.1. Simulation study

The least slope method (ABH), the smoother method (Spline), the bootstrap method (Boot), the langaas method with a convex decreasing density estimate (Langaas), the histogram method (Histo), the average method (Jiang), the SLIM method (SLIM), the robust method (Pounds) and the Jin and Cai method (Cai) (Jin and Cai, 2007) are assessed under independent simulated data, the HMM model simulated data under various setups, and real microarray data. Estimates and standard errors are calculated for each estimation procedure.

We present 9 estimation procedures in an independent data with a different true proportion of true null π_0 (= 0.25, 0.50, 0.75) and μ_1 (the mean under the alternative).

For the independence case, we simulate 1,000 independent normal random variables T_i, $i = 1, \ldots, 1000$ with the variance 1 and common correlation $\rho = 0$. 1,000 two-sided hypothesis tests are conducted with $\mu_0 = 0$ against $\mu_1 = 1, 2, 3$ (the mean under the alternative) using each of 9 procedures. Each individual hypothesis is tested by a z-test.

In a dependent simulation case, we model the two hidden state HMM (0 and 1) with a varying transition probability matrix in the previous section. We utilize Welsch t-test statistics for the p-value for each hypothesis.

3.2. Independence

Table 2 summarizes the result for independent p-values for Gaussian random variables. Cai procedure and SLIM procedures have better performance in that they have smaller biases and standard errors under all configurations of independent structures; however, Langaas, Spline, Boot, and ABH have larger biases and standard errors compared to other procedures.

3.3. Dependence

As transition matrix varies, we compute each $\hat{\pi}_0$ and a standard error of each estimator in Table 3. Cai procedure and SLIM procedure have relatively better performance in that they have smaller biases and standard errors under all configurations of dependent structures; however, Langaas, Spline, Boot, and ABH have larger biases and standard errors compared to other procedures.

Table 2: Independent p-values: each $\hat{\pi}_0$: an estimated proportion of true null hypotheses for Spline, Boot, Jiang, Histo, Langaas, Pounds, ABH, SLIM, and Cai

μ_1	π_0	Spline	Boot	Jiang	Histo	Langaas	Pounds	ABH	SLIM	Cai
1	0.25	0.5500	0.5430	0.4770	0.4900	0.5415	0.4877	0.6104	0.3215	0.2356
		(0.3572)	(0.3461)	(0.2673)	(0.3152)	(0.3162)	(0.2861)	(0.4157)	(0.1963)	(0.1952)
	0.50	0.7299	0.6171	0.5335	0.5364	0.5951	0.5846	0.6104	0.4286	0.4985
		(0.3532)	(0.2186)	(0.1565)	(0.1623)	(0.1603)	(0.1612)	(0.3642)	(0.1055)	(0.0864)
	0.75	0.8324	0.8313	0.7846	0.8165	0.6951	0.7134	0.6104	0.7219	0.7386
		(0.3805)	(0.1972)	(0.1329)	(0.1294)	(0.1603)	(0.1419)	(0.3914)	(0.1198)	(0.0785)
2	0.25	0.5610	0.5391	0.4691	0.4896	0.5512	0.4912	0.6084	0.3175	0.2413
		(0.3467)	(0.3378)	(0.2631)	(0.3254)	(0.3098)	(0.2918)	(0.3981)	(0.1895)	(0.1823)
	0.50	0.7177	0.6319	0.5413	0.5429	0.6019	0.5912	0.6409	0.4809	0.4998
		(0.3742)	(0.2218)	(0.1618)	(0.1719)	(0.1579)	(0.1701)	(0.3591)	(0.1193)	(0.0711)
	0.75	0.8264	0.8516	0.7984	0.8093	0.6991	0.7231	0.6104	0.7410	0.7485
		(0.3519)	(0.3609)	(0.1410)	(0.1310)	(0.1578)	(0.1092)	(0.4109)	(0.0909)	(0.0682)
3	0.25	0.4811	0.4491	0.4519	0.4789	0.5029	0.4892	0.6139	0.3091	0.2331
		(0.3561)	(0.3451)	(0.2718)	(0.3348)	(0.2966)	(0.2798)	(0.4001)	(0.1886)	(0.1765)
	0.50	0.6156	0.6291	0.5542	0.5324	0.5967	0.6589	0.6340	0.4798	0.4999
		(0.3697)	(0.2315)	(0.2234)	(0.1698)	(0.1498)	(0.1688)	(0.3610)	(0.1210)	(0.0691)
	0.75	0.8109	0.8402	0.7975	0.7654	0.6580	0.7368	0.6291	0.7443	0.7495
		(0.3509)	(0.3598)	(0.1409)	(0.1309)	(0.1569)	(0.1066)	(0.4091)	(0.0889)	(0.0661)

π_0: a true proportion of true null hypotheses, ( ): a standard error of an estimator and μ_1: the mean under the alternative. Spline = smoother method; Boot = bootstrap method; Jiang = average method; Histo = histogram method; Langaas = Langaas method with a convex decreasing density estimate; Pounds = robust method; ABH = least slope method; SLIM = sliding linear model method; Cai = Jin and Cai method.

Table 3: Dependent p-values: each $\hat{\pi}_0$: an estimated proportion of true null hypotheses for Spline, Boot, Jiang, Histo, Langaas, Pounds, ABH, SLIM, and Cai

Transition matrix	π_0	Spline	Boot	Jiang	Histo	Langaas	Pounds	ABH	SLIM	Cai
T1	0.7330	0.9990	0.9717	0.9290	0.9303	0.9477	0.9390	1.0000	0.9200	0.9042
		(0.2608)	(0.2506)	(0.1698)	(0.1699)	(0.2018)	(0.1729)	(0.3608)	(0.1589)	(0.1546)
T2	0.8456	0.9897	0.9799	0.8955	0.9380	0.9620	0.9400	1.0000	0.8654	0.8569
		(0.2658)	(0.2109)	(0.1462)	(0.1589)	(0.1603)	(0.1609)	(0.3756)	(0.1357)	(0.1164)
T3	0.8950	0.9871	0.9698	0.9665	0.9457	0.9620	0.9461	0.9963	0.8837	0.8900
		(0.1998)	(0.1888)	(0.1112)	(0.1236)	(0.1603)	(0.1320)	(0.3756)	(0.1087)	(0.0975)
T4	0.8826	0.9823	0.9717	0.9125	0.9148	0.9414	0.9308	0.9985	0.9044	0.8977
		(0.2965)	(0.1865)	(0.1131)	(0.1248)	(0.1686)	(0.1319)	(0.2995)	(0.1067)	(0.0825)
T5	0.8963	0.9716	0.9324	0.9133	0.9156	0.9414	0.9310	0.9966	0.9133	0.9126
		(0.2876)	(0.1897)	(0.1234)	(0.1267)	(0.1698)	(0.1345)	(0.2898)	(0.0976)	(0.0784)
T6	0.8136	0.9810	0.9708	0.8857	0.9282	0.9707	0.8927	0.9987	0.8823	0.8589
		(0.2619)	(0.2245)	(0.1198)	(0.1478)	(0.2198)	(0.1456)	(0.3101)	(0.1089)	(0.0884)
T7	0.8800	0.9822	0.9681	0.9278	0.9480	0.9502	0.9378	0.9999	0.9160	0.9155
		(0.3109)	(0.2365)	(0.1287)	(0.1645)	(0.1780)	(0.1503)	(0.3589)	(0.1123)	(0.0967)
T8	0.9160	0.9678	0.9598	0.9227	0.9333	0.9456	0.9394	0.9897	0.9136	0.9150
		(0.3069)	(0.2438)	(0.1277)	(0.1310)	(0.2101)	(0.1659)	(0.3090)	(0.1209)	(0.1095)
T9	0.8686	0.9780	0.9666	0.9016	0.8019	0.7889	0.8109	0.9888	0.8938	0.8922
		(0.3109)	(0.2670)	(0.1310)	(0.1896)	(0.2546)	(0.1783)	(0.3245)	(0.1298)	(0.1156)

π_0: a true proportion of true null hypotheses, ( ): a standard error of an estimator and a transition matrix. Spline = smoother method; Boot = bootstrap method; Jiang = average method; Histo = histogram method; Langaas = Langaas method with a convex decreasing density estimate; Pounds = robust method; ABH = least slope method; SLIM = sliding linear model method; Cai = Jin and Cai method.

Table 4: Data analysis in Van't Wout *et al.* (2003): each $\hat{\pi}_0$: an estimated proportion of true null hypotheses for Spline, Boot, Jiang, Histo, Langaas, Pounds, ABH, SLIM, and Cai

Spline	Boot	Jiang	Histo	Langaas	Pounds	ABH	SLIM	Cai
0.8970344	0.9045681	0.8672398	0.9021678	0.9468924	0.8510235	0.9820108	0.8702348	0.8847652

Spline = smoother method; Boot = bootstrap method; Jiang = average method; Histo = histogram method; Langaas = Langaas method with a convex decreasing density estimate; Pounds = robust method; ABH = least slope method; SLIM = sliding linear model method; Cai = Jin and Cai method.

4. Application to real data

The microarray data in an HIV study (Van't Wout *et al.*, 2003) is analyzed. This study identifies differentially expressed genes between HIV positive samples and HIV negative controls. They measured gene expression levels for 4 HIV-positive patients and 4 HIV-negative controls with 7,680 genes. 7680 two-sample t-tests are conducted and two-sided p-values are computed. Preprocessing this raw data was already done.

Table 4 describes different estimation procedures in real data analysis. Spline, Pounds, SLIM, Cai and Jiang procedures have relatively similar values of the proportion of true null hypotheses in the data.

5. Concluding remarks

Analyzing microarray data involves an appropriate multiple testing procedure to control type I error. Significant attention has been paid to the FDR procedure as a less conservative alternative to the FWER. So as to control the FDR, we need to compute the proportion of true null hypotheses in a multiple testing framework. We need to account for the structure since microarray data have high dependency structure among genes.

We assess various estimation procedures with independent data and dependent data with the HMM model and conduct real data analysis. Simulation result indicate that Cai and SLIM procedures have relatively smaller biases and standard errors, being more appropriate for estimating the proportion of true null hypotheses. Spline, Pounds, SLIM, Cai and Jiang procedures have almost similar values of the proportion of true null hypotheses in real data analysis.

Acknowledgements

This work was supported by the Research Institute of Natural Science of Gangneung-Wonju National University.

References

Baldi P and Hatfield W (2002). *DNA Microarrays and Gene Expression*, Cambridge University Press, Cambridge.

Benjamini Y and Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society: Series B*, **57**, 289–300.

Benjamini Y and Hochberg Y (2000). On the adaptive control of the false discovery rate in multiple testing with independent Statistics, *Journal of Educational and Behavioral Statistics*, **25**, 60–83.

Benjamini Y and Yekutieli D (2001). The control of the false discovery rate in multiple testing under dependency, *Annals of Statistics*, **29**, 1165–1188.

Churchill G (1992). Hidden Markov chains and the analysis of genome structure, *Computers and Chemistry*, **16**, 107–115.

Ephraim Y and Merhav N (2002). Hidden Markov processes, *IEEE Transactions on Information Theory*, **48**, 1518–1569.

Jiang H and Doerge RW (2008). Estimating the proportion of true null hypotheses for multiple comparisons, *Cancer Informatics*, **6**, 25–32.

Jin J and Cai TT (2007). Estimating the null and the proportion of non-null effects in large-scale multiple comparisons, *Journal of the American Statistical Association*, **102**, 495–506.

Krogh A, Brown M, Mian I, Sjölander K, and Haussler D (1994). Hidden Markov models in computational biology. Applications to protein modeling, *Journal of Molecular Biology*, **235**, 1501–1531.

Langaas M, Lindqvist BH, and Ferkingstad E (2005). Estimating the proportion of true null hypotheses, with application to DNA microarray data, *Journal of the Royal Statistical Society: Series B*, **67**, 555–572.

Nettleton D, Hwang JTG, Caldo RA, and Wise RP (2006). Estimating the number of true null hypotheses from a histogram of p values, *Journal of Agricultural, Biological, and Environmental Statistics*, **11**, 337–356.

Pounds S and Cheng C (2006). Robust estimation of the false discovery rate, *Bioinformatics*, **22**, 1979–1987.

Rabiner L (1989). A tutorial on hidden Markov models and selected applications in speech recognition, *IEEE*, **77**, 257–286.

Speed T (2003). *Statistical Analysis of Gene Expression Microarray Data*, Chapman and Hall/CRC, New York.

Storey JD (2002). A direct approach to false discovery rates, *Journal of the Royal Statistical Society: Series B*, **64**, 479–498.

Storey JD, Taylor JE, and Siegmund D (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach, *Journal of the Royal Statistical Society: Series B*, **66**, 187–205.

Storey JD and Tibshirani R (2003). Statistical significance for genomewide studies. In *Proceedings of the National Academy of Sciences*, **100**, 9440–9445.

Sun W and Cai TT (2009). Large-scale multiple testing under dependence, *Journal of the Royal Statistical Society: Series B*, **71**, 393–424.

Van't Wout AB, Lehrman GK, Mikheeva SA, O'Keeffe GC, Katze MG, Bumgarner RE, Geiss GK, and Mullins JI (2003). Cellular gene expression upon human immunodeficiency virus type 1 infection of CD4(+)-T-cell lines, *Journal of Virology*, **77**, 1392–1402.

Wang HQ, Tuominen LK, and Tsai CJ (2011). SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures, *Bioinformatics*, **27**, 225–231.

Received October 18, 2019; Revised November 27, 2019; Accepted November 28, 2019