Bayesian models for the multi-sample time-course microarray experiments

Claudia Angelini[1⋆], Daniela De Canditiis[1], Marianna Pensky[2], and Naomi Brownstein[3]

(1) Istituto per le Applicazioni del Calcolo CNR, Italy
(2) Department of Mathematics, University of Central Florida, US
(3) Department of Biostatistics, University of North Carolina at Chapel Hill, US

Abstract. *In this paper we present a functional Bayesian method for detecting genes which are temporally differentially expressed between several conditions. We identify the nature of differential expression (e.g., gene is differentially expressed between the first and the second sample but is not differentially expressed between the second and the third) and subsequently we estimate gene expression temporal profiles. The proposed procedure deals successfully with various technical difficulties which arise in microarray time-course experiments such as a small number of observations, non-uniform sampling intervals and presence of missing data or repeated measurements. The procedure allows to account for various types of errors, thus, offering a good compromise between nonparametric and normality assumption based techniques. In addition, all evaluations are carried out using analytic expressions, hence, the entire procedure requires very small computational effort. The performance of the procedure is studied using simulated data.*

Key words: Bayesian analysis, Classification, Hypothesis testing, Multi-sample problems, Time-course microarray

1 Introduction

Gene expression levels in a given cell can be influenced by various factors, namely, a pharmacological or a medical treatment, or a specific pathological or environmental state, or a specific experimental set-up. For the sake of brevity, we will simply use the term condition to describe any of such circumstances. One of the goals of modern molecular biology is the high-throughput identification

⋆ Corresponding Author

of genes associated with a particular condition of interest. The widely used technology of microarrays allows one to simultaneously monitor the expression levels of thousands of genes and time-course microarray experiments [2–4, 6, 9, 12, 14, 17] are an increasingly popular approach for understanding the dynamical behavior of a wide range of biological systems.

In this paper we shall consider microarray experiments made over the course of time which involve comparison between several biological conditions. In particular, we consider data consisting of measurements of the expression levels of N genes collected over time [0,T] under $H \geq 3$ different conditions. The objective is first to identify the genes that are differentially expressed between some of the H conditions and then to estimate the response.

In general, the problem can be formulated as follows. For condition $\aleph$, $\aleph = 1, \cdots, H$, data consists of the records on N genes which are taken at time points $t_{\aleph}^{(j)} \in [0, T]$, $j = 1, .., n_{\aleph}$, and, for gene i at a time point $t_{\aleph}^{(j)}$, there are $k_{\aleph i}^{(j)}$ records available, making the total number of records for gene i in condition $\aleph$ to be $M_{\aleph i} = \sum_{j=1}^{n_{\aleph}} k_{\aleph i}^{(j)}$. Note that the number of time points is relatively small ($n_{\aleph} \approx 10$) and very few replications are available at each time point ($k_{\aleph i}^{(j)} = 0, 1, \ldots K_{\aleph i}$, where $K_{\aleph i} = 1$, 2, 3 or 4) while the number of genes is very large ($N \approx 50,000$). Note that this is a much more general set up than the one which is usually considered since we require neither that the observations for the H conditions are made at the same time points nor that the number of observations for different samples is the same. The only requirement is that the samples are observed over the same period of time.

Each record can be modeled as a noisy measurement of a function $s_{\aleph i}(t)$ evaluated at a time point $t_{\aleph}^{(j)}$, where $s_{\aleph i}(t_{\aleph}^{(j)})$ represents the expression level of gene i measured on condition $\aleph$ at a time point $t_{\aleph}^{(j)}$, $\aleph = 1, 2, \cdots H$. The objective of the analysis is to identify differentially expressed genes (i.e. genes such that $s_{1i}(t) = \ldots = s_{Hi}(t)$ is not verified) and to estimate the expression profiles of the genes. Note that in the case of $H = 2$, this problem translates into selecting the curves $s_{1i}(t)$ and $s_{2i}(t)$, $i = 1, \cdots, N$, such that the difference $s_i(t) = s_{2i}(t) - s_{1i}(t)$ is not identical to zero (see [3]). However, in the case when $H \geq 3$ there are many more possibilities. For example, a particular gene i can be differentially expressed between all H conditions: $s_{1i}(t) \neq s_{2i}(t) \neq \cdots \neq s_{Hi}(t)$, or it can be not differentially expressed for the first $H - 1$ conditions and differentially expressed for the H-th condition: $s_{1\,i}(t) = s_{2\,i}(t) = s_{H-1\;i}(t) \neq s_{Hi}(t)$, and so on. In general, in the case of H conditions, one has B_H different possibilities, where B_H is the Bell exponential number, i.e the number of nonempty subsets of the set with H elements. The problem is to identify for each gene which of these B_H situations actually takes place.

The problem of identification of time-course differentially expressed genes under several biological conditions was considered in the recent papers by [17, 13, 3]. The last two papers do not examine every possible situation reducing the problem of finding differentially expressed genes between only two conditions.

In comparison, the paper by [17] studies all possible situations. However, the weakness of the approach is that it tests whether a gene is differentially expressed at a particular time point not overall, thus, ignoring the temporal dependence between the expression levels at different time points. Therefore, the objective of the present paper is to overcome this shortcoming by generalizing the functional approach of [2, 3] to the case of the H-conditions ($H \geq 3$) microarray experiments.

The paper is organized as follows. Section 2 introduces the hierarchical Bayesian model. Sections 2.2 and 2.3, respectively, describe modeling the gene expression profiles and the errors. Section 3 explains how to estimate the gene-dependent parameters. Section 4 describes the inference, while Section 5 outlines the procedures for estimating the gene expression profiles. Section 6 provides the techniques for estimating global parameters. Section 7 summarizes the algorithm. Finally, Section 8 provides an extensive simulation study. Section 8, Appendix, contains the derivation of the formulae in the previous sections.

2 Statistical modeling, estimation and classification of gene expression profiles

2.1 The data structure

The data are assumed to be already pre-processed to remove systematic sources of variation. Normalization procedures depend on the type of platforms used for performing the experiments. For a detailed discussion of the normalization procedures for microarray data we refer the reader to e.g. [7, 10, 15, 16].

The measurements are taken at $n_{\aleph}$, $\aleph = 1, \cdots, H$, different time points in $[0, T]$ where the sampling grid $t_{\aleph}^{(1)}, t_{\aleph}^{(2)}, \ldots, t_{\aleph}^{(n_{\aleph})}$ is not necessarily uniformly spaced and may be different for $\aleph = 1, \cdots, H$. For each array, the data consist of N measurements $z_{\aleph i}^{j,k}$, where $\aleph$ is the sample number, $i = 1, \ldots, N$, is the gene number, index j corresponds to the time point $t_{\aleph}^{(j)}$ and $k = 1, \ldots, k_{\aleph i}^{(j)}$, $k_{\aleph i}^{(j)} \geq 0$, accommodates for possible technical replicates at time $t_{\aleph}^{(j)}$. By the structure of the experimental design, $k_{\aleph i}^{(j)}$ are the same for each gene i; however, since some observations may be missing, we let $k_{\aleph i}^{(j)}$ to depend on i.

For each gene i in the sample $\aleph$, we assume that evolution in time of its relative expression is governed by a function $s_{\aleph i}(t)$ and each of the measurements involves some measurement error, i.e.

$$z_{\aleph i}^{j,k} = s_{\aleph i}(t_{\aleph}^{(j)}) + \zeta_{\aleph i}^{j,k}, \qquad i = 1, \ldots, N, \quad j = 1, \ldots, n_{\aleph}, \quad k = 1, \ldots, k_{\aleph i}^{(j)}. \quad (1)$$

The measurement errors $\zeta_{\aleph i}^{j,k}$ are assumed to be i.i.d. with zero mean and finite variance. The function $s_{\aleph i}(t)$, $\aleph = 1, \cdots, H$, represents the temporal expression level of gene i in the sample $\aleph$ over the interval $[0, T]$.

2.2 Modeling the gene expression profiles

Each function $s_{\aleph i}(t)$ is globally estimated, since the measurements are available only at a few time points. Specifically, we expand each function over some standard orthonormal basis on the interval $[0, T]$

$$s_{\aleph i}(t) = \sum_{l=0}^{L_i} c_{\aleph i}^{(l)} \phi_l(t) \tag{2}$$

and characterize each of them by the vector of its coefficients $\mathbf{c}_{\aleph i}$. In the present paper we use Legendre polynomials suitably rescaled and normalized in $[0, T]$, but other choices are possible. We emphasize that the degree of the polynomial varies from gene to gene but is common for all H conditions. The values of the coefficients $c_{\aleph i}^{(l)}$ and the degrees of the polynomials L_i are estimated from the observations via a Bayesian approach.

We assume that the genes are conditionally independent, so that combination of (1) and (2) yields

$$\mathbf{z}_{\aleph i} = \mathbf{D}_{\aleph i} \mathbf{c}_{\aleph i} + \boldsymbol{\zeta}_{\aleph i} \tag{3}$$

where $\mathbf{z}_{\aleph i} = (z_{\aleph i}^{1,1} \dots z_{\aleph i}^{1,k_{\aleph i}^{(1)}}, \cdots, z_{\aleph i}^{n_\aleph,1}, \dots z_{\aleph i}^{n_\aleph,k_{\aleph i}^{(n_\aleph)}})^T \in R^{M_{\aleph i}}$ is the column vector of all measurements for gene i in condition $\aleph$, $\mathbf{c}_{\aleph i} = (c_{\aleph i}^0, \dots, c_{\aleph i}^{L_i})^T \in R^{L_i+1}$ is the column vector of the coefficients of $s_{\aleph i}(t)$ in the chosen basis, $\boldsymbol{\zeta}_{\aleph i} = (\zeta_{\aleph i}^{1,1}, \dots, \zeta_{\aleph i}^{1,k_{\aleph i}^{(1)}}, \cdots, \zeta_{\aleph i}^{n_\aleph,1}, \dots, \zeta_{\aleph i}^{n_\aleph,k_{\aleph i}^{(n_\aleph)}})^T \in R^{M_{\aleph i}}$ is the column vector of random errors and $\mathbf{D}_{\aleph i}$ is the $M_{\aleph i} \times (L_i + 1)$ block design matrix, the j-row of which is the block vector $[\phi_0(t_\aleph^{(j)})\ \phi_1(t_\aleph^{(j)})\ \dots\ \phi_{L_i}(t_\aleph^{(j)})]$ replicated $k_{\aleph i}^{(j)}$ times.

The proposed model is fully Bayesian, since we treat all parameters either as random variables or as nuisance parameters, thus recovered from data. We assume that given σ^2, the vectors of errors $\boldsymbol{\zeta}_{\aleph i}$ are normally distributed $\boldsymbol{\zeta}_{\aleph i} \mid \sigma^2 \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{M_{\aleph i}})$, hence

$$\mathbf{z}_{\aleph i} \mid L_i, \mathbf{c}_{\aleph i}, \sigma^2 \sim \mathcal{N}(\mathbf{D}_{\aleph i} \mathbf{c}_{\aleph i}, \sigma^2 \mathbf{I}_{M_{\aleph i}}). \tag{4}$$

We also assume that L_i a-priori has the truncated Poisson distribution $Pois^*(\lambda, L_{\max})$ with parameter λ truncated at $L_{\max}$, and we denote its pdf by $g_\lambda(L_i)$:

$$g_\lambda(L_i) = \left[\sum_{l=0}^{L_{\max}} (l!)^{-1} \lambda^l e^{-\lambda} \right]^{-1} (L_i!)^{-1} \lambda^{L_i} e^{-\lambda}, \quad L_i = 0, \dots, L_{\max}. \tag{5}$$

Parameter λ is proportional to the average degree of the polynomial and $L_{\max}$ refers to the maximal possible degree. The values of both parameters are treated as known constants. In general, λ and $L_{\max}$ should be chosen by considering the number of available time points and the nature of the problem.

Denote

$$\mathbf{C}_i = \{\mathbf{c}_{1i}, \cdots, \mathbf{c}_{Hi}\}, \quad \mathbf{Z}_i = \{\mathbf{z}_{1i}, \cdots, \mathbf{z}_{Hi}\} \tag{6}$$

the set of the vector coefficients and the set of the vector of data. The question of interest now is for every gene $i = 1, \cdots, N$, to identify whether it is differentially expressed and, if yes, then between which conditions. For this purpose, we introduce B_H classes $\omega_0, \omega_1, \cdots, \omega_{B_H-1}$ that represents all possible combination of the H conditions. We observe that B_H is the Bell exponential number and it can be evaluated recursively as $B_n = \sum_{k=0}^{n-1} \binom{n-1}{k} B_k$, with $B_0 = B_1 = 1$. The formula gives $B_2 = 2$, $B_3 = 5$, $B_4 = 15$, $B_5 = 52$, $B_6 = 203$. Clearly, it will be hard to analyse more than 4 or 5 samples simultaneously. For example, for $H = 3$ we have $B_3 = 5$ possible classes which can be described as follows:

$$\begin{array}{lll} \omega_0 : c_{1i} = c_{2i} = c_{3i} & \omega_1 : c_{2i} = c_{3i} \neq c_{1i} & \omega_2 : c_{1i} = c_{3i} \neq c_{2i} \\ \omega_3 : c_{1i} = c_{2i} \neq c_{3i} & \omega_4 : c_{1i} \neq c_{2i} \neq c_{3i} & \end{array} \tag{7}$$

It is reasonable to assume that a-priori vectors $\mathbf{c}_{\aleph i}$, $\aleph = 1, 2, \cdots, H$, are either equal to each other or are independent and have identical distributions: the total of B_H different combinations. For example, in the case of $H = 3$ we elicit the following priors on the vectors $\mathbf{C}_i = \{\mathbf{c}_{1i}, \mathbf{c}_{2i}, \mathbf{c}_{3i}\}$:

$$\begin{array}{l} \mathbf{C}_i \mid L_i, \sigma^2, \omega_0 \sim \mathcal{N}(\mathbf{c}_{1i} \mid \mathbf{0}, \sigma^2\tau_{0i}^2\mathbf{Q}_i^{-1})\delta(\mathbf{c}_{1i} = \mathbf{c}_{2i} = \mathbf{c}_{3i}) \\ \mathbf{C}_i \mid L_i, \sigma^2, \omega_1 \sim \mathcal{N}(\mathbf{c}_{1i} \mid \mathbf{0}, \sigma^2\tau_{1i}^2\mathbf{Q}_i^{-1})\mathcal{N}(\mathbf{c}_{2i} \mid \mathbf{0}, \sigma^2\tau_{1i}^2\mathbf{Q}_i^{-1})\delta(\mathbf{c}_{2i} = \mathbf{c}_{3i}) \\ \mathbf{C}_i \mid L_i, \sigma^2, \omega_2 \sim \mathcal{N}(\mathbf{c}_{2i} \mid \mathbf{0}, \sigma^2\tau_{2i}^2\mathbf{Q}_i^{-1})\mathcal{N}(\mathbf{c}_{1i} \mid \mathbf{0}, \sigma^2\tau_{2i}^2\mathbf{Q}_i^{-1})\delta(\mathbf{c}_{1i} = \mathbf{c}_{3i}) \\ \mathbf{C}_i \mid L_i, \sigma^2, \omega_3 \sim \mathcal{N}(\mathbf{c}_{3i} \mid \mathbf{0}, \sigma^2\tau_{3i}^2\mathbf{Q}_i^{-1})\mathcal{N}(\mathbf{c}_{1i} \mid \mathbf{0}, \sigma^2\tau_{3i}^2\mathbf{Q}_i^{-1})\delta(\mathbf{c}_{1i} = \mathbf{c}_{2i}) \\ \mathbf{C}_i \mid L_i, \sigma^2, \omega_4 \sim \mathcal{N}(\mathbf{c}_{1i} \mid \mathbf{0}, \sigma^2\tau_{4i}^2\mathbf{Q}_i^{-1})\mathcal{N}(\mathbf{c}_{2i} \mid \mathbf{0}, \sigma^2\tau_{4i}^2\mathbf{Q}_i^{-1})\mathcal{N}(\mathbf{c}_{3i} \mid \mathbf{0}, \sigma^2\tau_{4i}^2\mathbf{Q}_i^{-1}). \end{array} \tag{8}$$

where $\mathbf{0} = (0, \cdots, 0)^T$. We assume that a-priori $P(\omega_l) = \pi_l$, $l = 0, \cdots, B_H - 1$, with $\sum_{l=0}^{B_H-1} \pi_l = 1$, so that

$$p(\mathbf{C}_i \mid L_i, \sigma^2) = \sum_{l=0}^{B_H-1} \pi_l \; p(\mathbf{C}_i \mid L_i, \sigma^2, \omega_l). \tag{9}$$

In formula (8), matrix $\mathbf{Q}_i$ is a diagonal matrix that can account for the decay of the coefficients in the chosen basis. Note that if no assumptions about smoothness of the gene expression profiles are made, we can assume $\mathbf{Q}_i = \mathbf{I}$. Parameters τ_{li}^2, $l = 0, \cdots, B_H - 1$, are gene and class specific and represent the strength of the signal with respect to the noise for a gene i in class l. We treat τ_{li}^2 as unknown nuisance parameters and estimate them from the data by maximizing the marginal likelihood for each gene independently (see Section 3).

2.3 Modeling the errors

We assume that parameter σ^2 is a random variable

$$\sigma^2 \quad \sim \quad \rho(\sigma^2). \tag{10}$$

The latter choice allows one to account for possibly non-Gaussian errors (quite common in microarray experiments), without sacrificing closed form expressions for estimators and test statistics. In particular, among the possible choices, we consider three types of priors $\rho(\cdot)$:

case 1: $\rho(\sigma^2) = \delta(\sigma^2 - \sigma_0^2)$, the point mass at σ_0^2. The marginal distribution of the error is normal.

case 2: $\rho(\sigma^2) = IG(\gamma, b)$, the Inverse Gamma distribution. The marginal distribution of the error is Student T.

case 3: $\rho(\sigma^2) = c_\mu \sigma^{(M-1)} e^{-\sigma^2 \mu/2}$, where M is the total number of arrays available in the experimental design set-up. If the gene has no missing data, i.e. all replications at each time point are available, then the marginal distribution of the error is double exponential.

The global hyperparameters, $\pi_0, \pi_1, \ldots, \pi_{B_H-1}$ and the $\rho(\sigma^2)$-specific parameters (σ_0^2 for case 1, γ and b for case 2 and μ for case 3), are estimated from the data. Possible strategies for doing this are discussed in Section 6. Once the hyperparameters are estimated, Bayesian analysis is carried out by combining the prior information and the data into the posterior distribution.

3 Estimation of gene-dependent parameters

If the global parameters of the model were known, one could proceed to a gene-by-gene analysis of vectors of coefficients $\mathbf{c}_{\aleph i}$, $\aleph = 1, \cdots, H$, $i = 1, \cdots, N$. In this section, we only provide the final formulae, referring the reader to the Appendix for the details of the calculations. To deal with different choices of $\rho(\sigma^2)$, we introduce a function

$$F(A, B) = \int_0^\infty \sigma^{-2A} e^{-B/2\sigma^2} \rho(\sigma^2) d\sigma^2 \tag{11}$$

that can be explicitly calculated in the three cases discussed above as:

$$F(A, B) = \begin{cases} \sigma_0^{-2A} e^{-B/2\sigma_0^2} & \text{in case 1,} \\ \frac{\Gamma(A+\gamma)}{\Gamma(\gamma)}\ b^{-A}(1 + \frac{B}{2b})^{-(A+\gamma)} & \text{in case 2,} \\ \frac{B^{(M+1-2A)/4}\mu^{(M+1+2A)/4}}{2^{(M-1)/2}\,\Gamma((M+1)/2)}\ K_{((M+1-2A)/2)}(\sqrt{B\mu}) & \text{in case 3.} \end{cases} \tag{12}$$

Here M denotes the number of available observations and $K_h(\cdot)$ is the Bessel function of degree h (see [8], Sections 8.4–8.5 for the definition). Note that in the sequel function F will appear with the argument $A = M_i/2$, where $M_i = \sum_{\aleph=1}^{H} M_{\aleph i}$, denotes the total number of records for gene i.

Then, if the i-th gene has no missing data (i.e. $M_i = M$), expression for case 3 in formula (12) simplifies to

$$F(M/2, B) = \sqrt{\pi}\,[\Gamma((M+1)/2)]^{-1}(\mu/2)^{M/2} \exp(-\sqrt{B\mu}).$$

Combining (4), (5), (9) and (10) in a joint pdf and integrating out $\mathbf{C}_i$ and σ^2, obtain

$$p(\mathbf{Z}_i|L_i) = (2\pi)^{-M_i/2}|\mathbf{Q}_i|^{1/2} \sum_{l=0}^{B_H-1} \pi_l A_{li}(\mathbf{Z}_i|L_i, \tau_{li}) \tag{13}$$

where $\mathbf{Z}_i$ is defined in (6) and

$$A_{li}(\mathbf{Z}_i|L_i,\tau_{li}) = (2\pi)^{M_i/2}\,|\mathbf{Q}_i|^{-1/2}\;p(\mathbf{Z}_i|\omega_l,L_i). \tag{14}$$

In the case of $H=3$, the expressions for $p(\mathbf{Z}_i|\omega_l,L_i)$ are given by formulae (22) and (23) below. Hence, the joint pdf of $\mathbf{Z}_i$ is

$$p(\mathbf{Z}_i) = \sum_{L_i=0}^{L_{\max}} p(\mathbf{Z}_i|L_i)\;g_\lambda(L_i), \tag{15}$$

and gene-dependent parameters τ_{li}^2, for $l=0,\ldots,B_H-1$, can be estimated as

$$(\hat\tau_{0i}^2,\ldots,\hat\tau_{B_H-1\;i}^2) = \arg\max_{\tau_{0i}^2,\ldots,\tau_{B_H-1\;i}^2} p(\mathbf{Z}_i). \tag{16}$$

However, since the joint pdf of $\mathbf{Z}_i$ is a sum of B_H positive terms (see formula (13)), each depending on a single τ_{li}^2, instead of one B_H-dimensional optimization, one can carry out B_H independent one-dimensional maximization procedures with respect to τ_{li}^2 for $l=0,\ldots,B_H-1$:

$$\hat\tau_{li}^2 = \arg\max_{\tau_{li}^2} \sum_{L_i=0}^{L_{\max}} A_{li}(\mathbf{Z}_i,L_i|\tau_{li})\;g_\lambda(L_i),\quad l=0,\cdots,B_H-1. \tag{17}$$

Maximization (17) represents the most computationally demanding step of the overall algorithm. However, since it is carried out independently for each gene, computations can be accelerated by using parallel computing.

The posterior pdf of the degree L_i given data $\mathbf{Z}$ is calculated as

$$p(L_i|\mathbf{Z}_i) = p(\mathbf{Z}_i|L_i)g_\lambda(L_i)\Big/p(\mathbf{Z}_i). \tag{18}$$

For each gene i, we estimate L_i by maximizing the posterior pdf (18) (MAP principle). After τ_{li}^2, $l=0,\ldots,B_H-1$, and L_i are estimated, we replace them with $\hat\tau_{li}^2$ and $\hat L_i$ in all the subsequent calculations.

4 Identification and classification of genes

4.1 Evaluation of class probabilities

Our main goal now is to carry out classification of genes to classes $\omega_l, l = 0,\cdots,B_H-1$. We evaluate probability that gene i belongs to class ω_l as

$$p(\omega_l|\mathbf{Z}_i) = \frac{\pi_l\;p(\mathbf{Z}_i|\omega_l)}{p(\mathbf{Z}_i)} = \pi_l\;\frac{\sum_{L_i=0}^{L_{\max}} g_\lambda(L_i)\;p(\mathbf{Z}_i|\omega_l,L_i)}{p(\mathbf{Z}_i)} \tag{19}$$

with

$$p(\mathbf{Z}_i|\omega_l,L_i) = \int\int p(\mathbf{Z}_i|\mathbf{C}_i,\sigma^2)\,p(\mathbf{C}_i|\omega_l,\sigma^2,L_i)\,d\mathbf{C}_i d\sigma^2 \tag{20}$$

and

$$p(\mathbf{Z}_i) = \sum_{L_i=0}^{L_{\max}} \sum_{k=0}^{B_H-1} \pi_l \; p(\mathbf{Z}_i|\omega_l, L_i) \; g_\lambda(L_i) \tag{21}$$

For example, in the case of $H = 3$, evaluation of $p(\mathbf{Z}_i|\omega_l, L_i)$, $l = 0, \cdots, B_H - 1$, yield

$$\begin{aligned}
p(\mathbf{Z}_i|\omega_0, L_i) &= \frac{|\mathbf{Q}_i|^{1/2} \, F(M_i/2, \mathbf{S}_i - \mathbf{v}_i^T \mathbf{V}_i^{-1} \mathbf{v}_i)}{(2\pi)^{M_i/2} \, |\mathbf{V}_i|^{1/2} \, \tau_{0i}^{(L_i+1)}}, \\
p(\mathbf{Z}_i|\omega_l, L_i) &= \frac{|\mathbf{Q}_i|^{1/2} \, F(M_i/2, \mathbf{S}_i - \mathbf{v}_{-li}^T \mathbf{V}_{-li}^{-1} \mathbf{v}_{-li} - \mathbf{v}_{li}^T \mathbf{V}_{li}^{-1} \mathbf{v}_{li})}{(2\pi)^{M_i/2} \, |\mathbf{V}_{-li}|^{1/2} \, |\mathbf{V}_{li}|^{1/2} \, \tau_{li}^{2(L_i+1)}}, \quad l = 1, \cdots, H, \\
p(\mathbf{Z}_i|\omega_4, L_i) &= \frac{|\mathbf{Q}_i|^{1/2} \, F(M_i/2, \mathbf{S}_i - \sum_{l=1}^H \mathbf{v}_{li}^T \mathbf{W}_{li}^{-1} \mathbf{v}_{li})}{(2\pi)^{M_i/2} \, \prod_{l=1}^H [|\mathbf{W}_{li}|^{1/2}] \, \tau_{4i}^{H(L_i+1)}},
\end{aligned} \tag{22}$$

where

$$\begin{array}{llll}
\mathbf{S}_i & = \sum_{k=1}^H \mathbf{z}_{ki}^T \mathbf{z}_{ki}, & \mathbf{D}_i & = \sum_{k=1}^H \mathbf{D}_{ki}^T \mathbf{D}_{ki}, \\
\mathbf{V}_i & = \mathbf{D}_i + \tau_{0i}^{-2} \mathbf{Q}_i, & \mathbf{v}_i & = \sum_{k=1}^H \mathbf{D}_{ki}^T \mathbf{z}_{ki}, \\
\mathbf{V}_{-li} & = \sum_{\substack{k=1 \\ k \neq l}}^H \mathbf{D}_{ki}^T \mathbf{D}_{ki} + \tau_{li}^{-2} \mathbf{Q}_i, & \mathbf{v}_{-li} & = \sum_{\substack{k=1 \\ k \neq l}}^H \mathbf{D}_{ki}^T \mathbf{z}_{ki}, \\
\mathbf{V}_{li} & = \mathbf{D}_{li}^T \mathbf{D}_{li} + \tau_{li}^{-2} \mathbf{Q}_i, & \mathbf{v}_{li} & = \mathbf{D}_{li}^T \mathbf{z}_{li}, \\
\mathbf{W}_{li} & = \mathbf{D}_{li}^T \mathbf{D}_{li} + \tau_{4i}^{-2} \mathbf{Q}_i, & l & = 1, \cdots, H.
\end{array} \tag{23}$$

Remark 1. Formulae (22) for the class $l = 1, 2, 3$ are analytically interchangeable, i.e. the H class are equivalent under label permutations.

4.2 Identification and classification of differentially expressed genes

In a standard Bayesian classification framework (**BC**), each gene is classified according to the highest posterior probability (20). However, with this approach, one has absolutely no control over the number of genes which are identified as differentially expressed. For this reason, we also propose a two-stage bayesian approach (**TSB**) as follows. At the first stage, we identify the genes which are differentially expressed among at least two of H conditions. At the second stage, we determine between which of the B_H cases occurs.

In order to proceed, we note that $p(\omega_0|\mathbf{Z}_i)$ can be presented as

$$p(\omega_0|\mathbf{Z}_i) = \pi_0 \; \left(\pi_0 + \frac{1 - \pi_0}{BF_i(\mathbf{Z}_i)} \right)^{-1}$$

where $BF_i(\mathbf{Z}_i)$ is the Bayes factors, the quotient between the posterior odds ratio and the prior odds ratio, for testing hypotheses $H_{0i} : i \in \omega_0$ versus $H_{1i} : i \notin \omega_0$ (see e.g. [5]):

$$BF_i(\mathbf{Z}_i) = (1-\pi_0)\frac{\sum_{L_i=0}^{L_{\max}} g_\lambda(L_i) p(\mathbf{Z}_i|\omega_0, L_i)}{\sum_{l=1}^{B_H} \pi_l \sum_{L_i=0}^{L_{\max}} g_\lambda(L_i) p(\mathbf{Z}_i|\omega_l, L_i)}. \tag{24}$$

Note that, although Bayes factors BF_i can be used for independent testing of the null hypotheses H_{0i}, $i = 1, \ldots, N$, the classical Bayesian approach [5] does not account for the multiplicity of comparisons. However, since in microarray experiments N is large, the problem of multiplicity cannot be ignored, therefore we apply the Bayesian multiple testing procedure of [1].

5 Estimation of gene expression profiles

There are at last two approaches for estimating the gene expression profiles: the model-selection based estimator or the model-average based estimator. The model-selection based estimator is constructed under the assumption that gene i belongs to class ω_l, i.e. $\mathbf{c}_{\aleph i}$ is estimated by

$$\hat{\mathbf{c}}_{\aleph i}(\omega_l) = \mathrm{E}\left(\mathbf{c}_{\aleph i}|\omega_l, \mathbf{Z}_i, \hat{L}_i\right).$$

Here, $\hat{L}_i$ is estimated degree of the polynomial for expression profiles for gene i. Hence, $\hat{\mathbf{c}}_{\aleph i}(\omega_l)$ can be evaluated as

$$\hat{\mathbf{c}}_{\aleph i}(\omega_l) = \frac{\int \mathbf{c}_{\aleph i}\ p(\mathbf{C}_i, \mathbf{Z}_i|\omega_l, \sigma^2, \hat{L}_i)\, \rho(\sigma^2)\ d\mathbf{C}_i d\sigma^2}{p(\mathbf{Z}_i|\omega_l, \hat{L}_i)}. \tag{25}$$

In the case of $H = 3$ we derive the following expressions for $\hat{\mathbf{c}}_{\aleph i}(\omega_l)$, $\aleph = 1, 2, 3$, $l = 0, \cdots, 5$:

$$\begin{array}{lll}
\hat{\mathbf{c}}_{\aleph i}(\omega_0) = \mathbf{V}_i^{-1}\mathbf{v}_i, & & \aleph = 1,2,3, \\
\hat{\mathbf{c}}_{1i}(\omega_1) = \mathbf{V}_{1i}^{-1}\mathbf{v}_{1i}, & \hat{\mathbf{c}}_{2i}(\omega_1) = \hat{\mathbf{c}}_{3i}(\omega_1) = \mathbf{V}_{-1i}^{-1}\mathbf{v}_{-1i}, & \\
\hat{\mathbf{c}}_{2i}(\omega_2) = \mathbf{V}_{2i}^{-1}\mathbf{v}_{2i}, & \hat{\mathbf{c}}_{1i}(\omega_1) = \hat{\mathbf{c}}_{3i}(\omega_1) = \mathbf{V}_{-2i}^{-1}\mathbf{v}_{-2i}, & \\
\hat{\mathbf{c}}_{3i}(\omega_3) = \mathbf{V}_{3i}^{-1}\mathbf{v}_{3i}, & \hat{\mathbf{c}}_{1i}(\omega_1) = \hat{\mathbf{c}}_{2i}(\omega_1) = \mathbf{V}_{-3i}^{-1}\mathbf{v}_{-3i}, & \\
\hat{\mathbf{c}}_{\aleph i}(\omega_4) = \mathbf{W}_{\aleph i}^{-1}\mathbf{v}_{\aleph i}, & & \aleph = 1,2,3,
\end{array} \tag{26}$$

In alternative one can use the model average estimator:

$$\hat{\mathbf{c}}_{\aleph i} = \mathrm{E}\left(\mathbf{c}_{\aleph i}|\mathbf{Z}_i, \hat{L}_i\right)$$

which can be evaluated as

$$\hat{\mathbf{c}}_{\aleph i} = \frac{\sum_{l=0}^{B_H-1} \pi_l \int \mathbf{c}_{\aleph i}\ p(\mathbf{C}_i, \mathbf{Z}_i|\omega_l, \sigma^2, \hat{L}_i)\, \rho(\sigma^2)\ d\mathbf{C}_i d\sigma^2}{\sum_{l=0}^{B_H-1} \pi_l p(\mathbf{Z}_i|\omega_l, \hat{L}_i)}. \tag{27}$$

Since it follows from (26) that the integrals in the numerator of formula (27) are equal to $\hat{\mathbf{c}}_{\aleph i}(\omega_l) p(\mathbf{Z}_i|\omega_l, \hat{L}_i)$, one can easily evaluate the estimators $\hat{\mathbf{c}}_{\aleph i}$ as

$$\hat{\mathbf{c}}_{\aleph i} = \frac{\sum_{l=0}^{B_H-1} \pi_l\ \hat{\mathbf{c}}_{\aleph i}(\omega_l)\ p(\mathbf{Z}_i|\omega_l, \hat{L}_i)}{\sum_{l=0}^{B_H-1} \pi_l\ p(\mathbf{Z}_i|\omega_l, \hat{L}_i)} \tag{28}$$

where $\hat{\mathbf{c}}_{\aleph i}(\omega_l)$ and $p(\mathbf{Z}_i|\omega_l, \hat{L}_i)$ are defined in (26) and (20), respectively.

6 Estimation of global parameters and prior hyperparameters

In this section we consider some possible strategies for estimation of the global parameters σ^2, π_l, $l = 0, \cdots, H$, and the $\rho(\sigma^2)$-specific parameters σ_0^2 (case 1), γ and b (case 2), and μ (case 3).

Estimation of σ^2 . If replications are available (i.e. , $k_{\aleph i}^{(j)} > 1$ for some $j, \aleph$ and i), one can form statistics

$$\beta_{\aleph i}^{(j)} = \sum_{k=1}^{k_{\aleph i}^{(j)}} (z_{\aleph i}^{j,k} - \bar{z}_{\aleph i}^{j})^2 \quad \text{where} \quad \bar{z}_{\aleph i}^{j} = (k_{\aleph i}^{(j)})^{-1} \sum_{k=1}^{k_{\aleph i}^{(j)}} z_{\aleph i}^{j,k}.$$

It is easy to notice that $\beta_{\aleph i}^{(j)}/\sigma^2$ has chi-squared distribution with $k_{\aleph i}^{(j)} - 1$ degrees of freedom $\chi(k_{\aleph i}^{(j)} - 1)$ if $k_{\aleph i}^{(j)} > 1$ and is identical zero otherwise. Hence,

$$\beta = \sum_{\aleph=1}^{H} \sum_{i=1}^{N} \sum_{j=1}^{n_\aleph} \beta_{\aleph i}^{j} \sim \sigma^2 \, \chi(\upsilon) \quad \text{with} \quad \upsilon = \sum_{\aleph=1}^{H} \sum_{i=1}^{N} \sum_{j=1}^{n_\aleph} (k_{\aleph i}^{(j)} - 1) I(k_{\aleph i}^{(j)} > 1), \quad (29)$$

so that $\hat{\sigma}^2 = \beta/\upsilon$ is an unbiased estimator of σ^2.

In a general situation, when replications are not available or they are available only at few time points, one can apply the U-statistics version of the Rice estimator derived in [11] with the kernel $K(x) = 3(1 - x^2)_+/4$ gene by gene, After that, the global estimator of the variance σ^2 is obtained by pooling the estimators of the variance for single genes, see also [3] for details.

Estimation of $\boldsymbol{\pi} = (\boldsymbol{\pi_0}, \cdots, \boldsymbol{\pi_{B_H - 1}})$. Recall that parameters τ_{li}, $l = 0, \cdots, B_H - 1$, can be estimated without any knowledge of the values of π_l, $l = 1, \cdots, B_H - 1$. Hence, in principle, π_l, $l = 1, \cdots, B_H - 1$, can be obtained as a solution of a $(B_H - 1)$-dimensional optimization problem

$$\hat{\boldsymbol{\pi}} = \arg\max_{\boldsymbol{\pi}} \; \log p(\mathbf{Z}) = \arg\max_{\boldsymbol{\pi}} \; \sum_{i=1}^{N} \log p(\mathbf{Z}_i) \quad (30)$$

where $p(\mathbf{Z}_i)$ is given by formulae (13) and (15). However, solution of this optimization problem is highly unstable and computationally demanding. Therefore we used equally likely prior probability.

Estimation of case-specific parameters . In *case 1*, the natural estimator of σ_0^2 is $\hat{\sigma}^2$. In *case 2*, one can either fix one of the two parameters, γ or b, and then estimate another one by matching the mean of the distribution $IG(\gamma, b)$ with $\hat{\sigma}^2$, or use the MLE estimate of both parameters as proposed in [2]. Similarly, in *case 3*, μ is estimated by $\hat{\mu} = (M - 1)/\hat{\sigma}^2$, so that the mean of the prior distribution $\rho(\sigma^2)$ is centered at $\hat{\sigma}^2$. Other alternative strategies can also be used for estimating parameters without changing the general algorithm.

7 Algorithm

The algorithm can be carried out as follows:

1. Fix prior parameters λ, $L_{\max}$ and ν.
2. Estimate global parameters: σ^2 and case-specific hyper-parameters σ_0^2 (for case 1), γ and b (for case 2) or μ (for case 3), see Section 6.
3. For each gene i, estimate the gene specific parameters $\tau_{\aleph i}^2$ for $\aleph = 0, \ldots, B_H - 1$, by maximizing the marginal pdf of the data $p(\mathbf{Z}_i)$.
4. Apply **BC** classification procedure by placing gene i into a class with the highest posterior probability (20). Alternatively, apply **TSB** classification procedure, i.e., for each gene i, compute Bayes Factor BF_i using formula (24). See Section 4.2
5. For each differentially expressed gene i, estimate the most appropriate degree L_i as the mean or the mode of the posterior pdf (18).
6. Estimate the gene expression profiles $s_{\aleph i}(t)$ for differentially expressed genes using formulae (2) and (26) or (27).

Since all evaluations are based on explicit expressions, the algorithm is very computationally efficient.

8 Simulations results and discussion

In order to evaluate the performance of the proposed method under different possible scenarios, we carried out a simulation study over various kinds of synthetic data-sets. For simplicity, we dealt only with the case $H = 3$. We considered $N = 10000$ genes and analyze four different structure of data-sets: DATASET1, DATSET2, DATASET3 and DATSET4.

DATASET1 has the same time grid, $t_{\aleph}^{(\cdot)} = [1, 2, 3, 4, 6, 8, 10, 11, 12]$ with 2 replicates at each time point except $k_{\aleph i}^{(3,5,7)} = 1$ for all three conditions $\aleph = 1, 2, 3$, and equal cardinalities of all classes $|\omega_l| = 2000$, for $l = 0, ..., 4$. In DATASET2, the time grids are $t_1^{(\cdot)} = [1, 2, 4, 6, 8, 10, 11, 12]$, $t_2^{(\cdot)} = [1, 2, 3, 4, 6, 8, 10, 12]$ and $t_3^{(\cdot)} = [1, 2, 3, 4, 6, 8, 11, 12]$, respectively, with 2 replicates at each time point except $k_{1i}^{(4,6)} = 1$, $k_{2i}^{(5,7)} = 1$ and $k_{3i}^{(3,5)} = 1$, and equal classes' cardinalities. DATASET3 has time grids as in DATSET1, while classes' cardinalities are $|\omega_0| = 5000$, $|\omega_1| = 2500$, $|\omega_2| = 1000$, $|\omega_3| = 500$ and $|\omega_4| = 1000$. DATASET4 has time grids as in DATASET2 while classes' cardinalities as DATASET3.

The data were generated according to model (1) with the noise $\zeta_{\aleph i}^{j,k}$ following the Student distribution with 5 degrees of freedom and scaled so that its standard deviation is $\sigma = 0.2$. The gene expression profiles $s_{\aleph i}(t_{\aleph}^{(j)})$ were generated according to model (2). In particular, for each gene, we first sampled the degree of the polynomial L_i from the discrete uniform distribution $U[0, 6]$. Then, we sampled one, two or three different vectors of coefficients $\mathbf{c}_{\aleph i}$ from a normal distribution $\mathcal{N}(0, \sigma^2 \tau_i^2 \mathbf{Q}_i^{-1})$ according to the class participation of the current

gene. For example, if the gene belongs to the first class (ω_0), then only one vector of coefficient is generated since all the three samples $\aleph = 1, 2, 3$ are the same. If the gene belongs to class 1,2, or 3, then two different vectors of coefficients were generated with one of the vectors used for the sample which differentiates from the other two. Finally, if the gene belongs to class ω_4, then three different vectors of coefficients were generated, each used for one sample. Matrix $\mathbf{Q}_i$ is set equal to $\mathrm{diag}(1^{2\nu_i}, 2^{2\nu_i}, ..., L^{i^{2\nu_i}})$ with ν_i sampled from the uniform distribution $U([0, 1])$. For each gene, the values of τ_i^2 were independently and uniformly sampled in order to produce the signal-to-noise ratio (SNR) in the interval $[2, 6]$.

For each data-set the simulations were repeated using 100 randomly generated sets of profiles s_{1i}, s_{2i} and s_{3i} and noise realizations, with the choice $L_{max} = 6$, $\lambda = 9$ and $\nu = 0$. The data were processed using the BC and TBS approaches described in Section 4.2. Results are summarized in Tables 1 and 2, where means and standard deviations (in parenthesis) of, respectively, the sensitivity and the specificity for the 5 classes are reported. Sensitivity measures the proportion of actual members of the class which are correctly identified as such (i.e., for each class, the sensitivity is the percentage of genes correctly classified into the current class). Specificity measures the proportion of samples which do not belong to the class and are correctly identified (i.e., for each class, specificity is the percentage of genes which are correctly classified into a class different from the current). Tables 1 and 2 confirm that the two approaches, BC and TSB, are similar. However TSB is slightly more sensitive in the first class, being proposed for controlling the FDR on ω_0; while is less sensitive on classes ω_l, $l = 1, 2, 3$. On the contrary TSB is slightly less specific on ω_0 and more specific on classes ω_l $l = 1, 2, 3$. The two procedure have the same performance on class ω_4. Moreover, the techniques are invariant under permutation of class labels (if the class labels are permuted, then the results permute accordingly, result not showed here). The unbalanced scenarios, both in terms of time grids and of class cardinality, do not affect the classification precision since the procedures yield very similar results on all four data sets.

In order to be fair in the evaluation of the proposed procedures, we carried out an additional set of simulations which was not model-based. Specifically, in the same four scenarios for time grids and class cardinalities as in data sets DATSET1–4, we generated data according to smooth functions which are not directly represented as linear combinations of Legendre polynomials. For each class, we randomly drawn one or more functions from a set of predefined shapes: linear, quadratic, sine, cosine and exponential then we randomly pick its coefficients.

Results of these simulations are presented in Tables 3 and 4.

Acknowledgments. Claudia Angelini and Daniela De Canditiis were partially supported by the CNR-Bioinformatics projects. Marianna Pensky was partially supported by the National Science Foundation (NSF), grant DMS-1106564. Naomi Brownstein was partially supported by the NSF, grant ???.

Table 1. Sensitivities of BC and TSB procedures for model-based data sets.

		class ω_0	class ω_1	class ω_2	class ω_3	class ω_4
DATASET1	BC	.9318	.9716	.9720	.9713	.9942
		(00030)	(.0020)	(.0017)	(.0018)	(.0009)
	TSB	.9488	.9684	.9689	.9681	.9942
		(.0024)	(.0021)	(.0018)	(.0019)	(.0009)
DATASET2	BC	.9254	.9685	.9695	.9684	.9935
		(.0027)	(.0018)	(.0019)	(.0020)	(.0010)
	TSB	.9456	.9647	.9657	.9648	.9935
		(.0024)	(.0019)	(.0020)	(.0021)	(.0010)
DATASET3	BC	.9281	.9844	.9857	.9861	.9965
		(.0037)	(.0012)	(.0012)	(.0010)	(.0006)
	TSB	.9463	.9816	.9829	.9832	.9965
		(.0033)	(.0014)	(.0013)	(.0011)	(.0006)
DATASET4	BC	.9219	.9827	.9843	.9843	.9961
		(.0037)	(.0014)	(.0015)	(.0010)	(.0006)
	TSB	.9427	.9797	.9807	.9811	.9961
		(.0032)	(.0016)	(.0017)	(.0011)	(.0006)

Table 2. Specificities of BC and TSB procedures for model-based data sets.

		class ω_0	class ω_1	class ω_2	class ω_3	class ω_4
DATASET1	BC	.9935	.8945	.8939	.8940	.6872
		(.0018)	(.0071)	(.0074)	(.0069)	(.0118)
	TSB	.9864	.9066	.9065	.9067	.6873
		(.0026)	(.0066)	(.0069)	(.0066)	(.0118)
DATASET2	BC	.9922	.8841	.8819	.8834	.6593
		(.0018)	(.0066)	(.0074)	(.0074)	(.0107)
	TSB	.9840	.8991	.8970	.8979	.6594
		(.0026)	(.0062)	(.0073)	(.0069)	(.0107)
DATASET3	BC	.9935	.8942	.8930	.8948	.6883
		(.0010)	(.0064)	(.0093)	(.0136)	(.0143)
	TSB	.9864	.9074	.9058	.9082	.6883
		(.0015)	(.0062)	(.0091)	(.0126)	(.0143)
DATASET4	BC	.9923	.8833	.8830	.8849	.6590
		(.0013)	(.0059)	(.0102)	(.0148)	(.0144)
	TSB	.9843	.8981	.8979	.8996	.6590
		(.0017)	(.0057)	(.0096)	(.0142)	(.0144)

Table 3. Sensitivities of BC and TSB procedures for model-free data sets.

		class ω_0	class ω_1	class ω_2	class ω_3	class ω_4
DATASET1	BC	.9945	.9599	.9599	.9596	.9937
		(.0009)	(.0022)	(.0020)	(.0019)	(.0008)
	TSB	.9970	.9562	.9562	.9560	.9937
		(.0006)	(.0023)	(.0022)	(.0020)	(.0008)
DATASET2	BC	.9937	.9000	.9565	.9585	.8924
		(.0008)	(.0034)	(.0022)	(.0019)	(.0032)
	TSB	.9967	.8952	.9520	.9548	.8924
		(.0006)	(.0034)	(.0023)	(.0020)	(.0032)
DATASET3	BC	.9942	.9725	.9769	.9782	.9954
		(.0011)	(.0017)	(.0013)	(.0013)	(.0008)
	TSB	.9968	.9636	.9694	.9710	.9954
		(.0007)	(.0019)	(.0017)	(.0015)	(.0008)
DATASET4	BC	.9934	.8187	.9696	.9771	.9476
		(.0013)	(.0039)	(.0017)	(.0014)	(.0023)
	TSB	.9964	.8069	.9604	.9700	.9476
		(.0009)	(.0038)	(.0019)	(.0017)	(.0023)

Table 4. Specificities of BC and TSB procedures for model-free data sets.

		class ω_0	class ω_1	class ω_2	class ω_3	class ω_4
DATASET1	BC	.9635	.9822	.9821	.9822	.5604
		(.0047)	(.0029)	(.0029)	(.0028)	(.0114)
	TSB	.9239	.9840	.9838	.9839	.5604
		(.0064)	(.0027)	(.0028)	(.0028)	(.0114)
DATASET2	BC	.7012	.9493	.8092	.7891	.5554
		(.0105)	(.0047)	(.0082)	(.0092)	(.0109)
	TSB	.6549	.9510	.8116	.7912	.5554
		(.0109)	(.0047)	(.0081)	(.0090)	(.0109)
DATASET3	BC	.9638	.9824	.9822	.9828	.5605
		(.0027)	(.0026)	(.0046)	(.0055)	(.0149)
	TSB	.9243	.9841	.9838	.9844	.5605
		(.0039)	(.0025)	(.0044)	(.0053)	(.0149)
DATASET4	BC	.7016	.9499	.8088	.7900	.5581
		(.0058)	(.0041)	(.0120)	(.0189)	(.0150)
	TSB	.6552	.9518	.8109	.7922	.5581
		(.0059)	(.0039)	(.0119)	(.0188)	(.0150)

References

1. Abramovich, F., Angelini, C.: Bayesian maximum a posteriori multiple testing procedure. Sankhya, **68**, 436-460, (2006)
2. Angelini, C., De Canditiis, D., Mutarelli, M., Pensky, M.: A Bayesian Approach to Estimation and Testing in Time-course Microarray Experiments, Stat. Appl. Gen. Mol. Bio. **6,** Art. 24, (2007)
3. Angelini, C., D. De Canditiis, D., Pensky, M.: Bayesian Models for the Two-Sample Time-course Microarray Experiments, CSDA, **53,** 1547-1565, (2009)
4. Bar–Joseph, Z.:, Analyzing time series gene expression data. Bioinformatics, **20**, 2493–2503, (2004)
5. Berger, O.J.,: Statistical Decision Theory and Bayesian Analysis. Springer series in Statistics (1985)
6. Conesa, A., Nueda, M.J., Ferrer, A., Talon, M.: MaSigPro: a method to identify significantly differential expression profiles in time-course microarray-experiments. Bioinformatics, **22,** 1096–1102, (2006)
7. Cui, X., Kerr, M.K., Churchill G.A.: Transformation for cDNA Microarray Data. Stat. Appl. Gen. Mol. Bio., **2**, (2002)
8. Gradshteyn, I.S., Ryzhik, I.M.: Tables of Integrals, Series, and Products. New York: Academic Press, (1980).
9. Heard, N.A., Holmes, C.C., Stephens, D.A.: A quantitative study of gene regulation involved in the Immune response of Anopheline Mosquitoes: An application of Bayesian hierarchical clustering of curves. JASA. **101**, 18-29,(2006)
10. McLachlan, G., Do, K.A., Ambroise, C.: Analyzing microarray gene expression data. Wiley series in Probability and Statistics, (2004)

11. Müller, U., Schick, A., Wefelmeyer W.: Estimating the error variance in nonparametric regression by a covariate-matched U-statistic. Statistics, **37,** 179-188, (2003)
12. Storey, J.D., Xiao, W., Leek, J.T., Tompkins, R.G.,Davis, R.W.: Significance analysis of time course microarray experiments. PNAS, **12**, 12837-12842, (2005)
13. Tai, Y.C., Speed, T.P.: On gene ranking using replicated microarray time course data. *Biometrics*, **65**, 40-51, (2009)
14. Vinciotti, V., Yu. K.: M-quantile regression analysis of temporal gene expression data, Stat. Appl. Gen. Mol. Bio., vol. 8, art. 41, (2009)
15. Wit, E., McClure J.: Statistics for Microarrays: Design, Analysis and Inference, Wiley, (2004).
16. Yang, Y.H., Dudoit, S., Luu, P., Lin M.D., Peng, V., Ngai,J., Speed, T.P.: Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Research, **30,** (2002)
17. Yuan, M., Kendziorski, C.: Hidden Markov Models for microarray time course data in multiple biological conditions. JASA, **101,** 1323–1340, (2006)

Appendix

Without loss of generality, let us consider derivation for $H = 3$, $B_H = 5$, $\mathbf{C}_i = \{\mathbf{c}_{1i}, \mathbf{c}_{2i}, \mathbf{c}_{3i}\}$ and $\mathbf{Z}_i = \{\mathbf{z}_{1i}, \mathbf{z}_{2i}, \mathbf{z}_{3i}\}$. Combine (4), (8) and (10) in a joint pdf

$$
\begin{aligned}
p(\mathbf{C}_i, \mathbf{Z}_i, \sigma^2|L_i) = \frac{\rho(\sigma^2)}{(2\pi\sigma^2)^{M_i/2}} \exp\left\{-\sum_{\aleph=1}^{3} \frac{(\mathbf{z}_{\aleph i} - \mathbf{D}_{\aleph i}\mathbf{c}_{\aleph i})^T(\mathbf{z}_{\aleph i} - \mathbf{D}_{\aleph i}\mathbf{c}_{\aleph i})}{2\sigma^2}\right\} \\
\times \left[\frac{\pi_0|\mathbf{Q}_i|^{1/2}}{(2\pi\sigma^2\tau_{0i}^2)^{(L_i+1)/2}} \exp\left(-\frac{\mathbf{c}_{1i}^T\mathbf{Q}_i\mathbf{c}_{1i}}{2\sigma^2\tau_{0i}^2}\right) \delta(\mathbf{c}_{1i} = \mathbf{c}_{2i} = \mathbf{c}_{3i})\right. \\
+ \frac{\pi_1|\mathbf{Q}_i|}{(2\pi\sigma^2\tau_{1i}^2)^{(L_i+1)}} \exp\left(-\frac{\mathbf{c}_{1i}^T\mathbf{Q}_i\mathbf{c}_{1i}}{2\sigma^2\tau_{1i}^2}\right) \exp\left(-\frac{\mathbf{c}_{2i}^T\mathbf{Q}_i\mathbf{c}_{2i}}{2\sigma^2\tau_{1i}^2}\right) \delta(\mathbf{c}_{2i} = \mathbf{c}_{3i}) \\
+ \frac{\pi_2|\mathbf{Q}_i|}{(2\pi\sigma^2\tau_{2i}^2)^{(L_i+1)}} \exp\left(-\frac{\mathbf{c}_{1i}^T\mathbf{Q}_i\mathbf{c}_{1i}}{2\sigma^2\tau_{2i}^2}\right) \exp\left(-\frac{\mathbf{c}_{2i}^T\mathbf{Q}_i\mathbf{c}_{2i}}{2\sigma^2\tau_{2i}^2}\right) \delta(\mathbf{c}_{1i} = \mathbf{c}_{3i}) \\
+ \frac{\pi_3|\mathbf{Q}_i|}{(2\pi\sigma^2\tau_{3i}^2)^{(L_i+1)}} \exp\left(-\frac{\mathbf{c}_{1i}^T\mathbf{Q}_i\mathbf{c}_{1i}}{2\sigma^2\tau_{3i}^2}\right) \exp\left(-\frac{\mathbf{c}_{3i}^T\mathbf{Q}_i\mathbf{c}_{3i}}{2\sigma^2\tau_{3i}^2}\right) \delta(\mathbf{c}_{1i} = \mathbf{c}_{2i}) \\
\left. + \frac{\pi_4|\mathbf{Q}_i|^{3/2}}{(2\pi\sigma^2\tau_{4i}^2)^{3(L_i+1)/2}} \exp\left(-\sum_{\aleph=1}^{3} \frac{\mathbf{c}_{\aleph i}^T\mathbf{Q}_i\mathbf{c}_{\aleph i}}{2\sigma^2\tau_{4i}^2}\right)\right].
\end{aligned}
$$

Completing the squares with respect to $\mathbf{c}_{1i}, \mathbf{c}_{2i}, \mathbf{c}_{3i}$ for each of the terms separately and integrating out $\mathbf{c}_{1i}, \mathbf{c}_{2i}, \mathbf{c}_{3i}$ and then σ^2, we arrive at $p(\mathbf{Z}_i|\omega_l, L_i)$, $l = 0, \cdots, B_H - 1$, given by (22).