Classification of Trends in Dose-response Microarray Experiments Using Information Theory Selection Methods

Dose-response microarray experiments consist of monitoring expression levels of thousands of genes with respect to increasing dose of the treatment under investigation. The primary goal of such an experiment is to establish a dose-response relationship, while the secondary goals are to determine the minimum effective dose level and to identify the shape of the dose-response curve. Recently, Lin *et al.* [1] discussed several testing procedures to test for monotone trend based on isotonic regression of the observed means [2,3]. Once a monotone relationship between the gene expression and dose is established, there is a set of R possible monotone models that can be fitted to the data. A selection of the best model from this set allows us to identify both the shape of dose-response curve and the minimum effective dose level. In this paper we focus on classification of dose-response curve shapes using the information theory model selection. In particular, the Order Restricted Information Criterion ($ORIC$) is discussed for the inference under order restriction. The posterior probability of the model is calculated using information criteria that take into account both the goodness-of-fit and the complexity of the models. The method is applied to a dose-response microarray experiment with 12 arrays (for three samples at each of the four dose levels) with 16,998 genes.

1   Introduction

A common experiment in the early drug development is a dose-response study that is set up to assess the biological activity of a chemical compound. In such a study, the response of primary interest is measured at several increasing dose levels. Typically, the first dose level is a control group with zero dose. In recent years, dose-response studies were extended to the microarray setting, in which the arrays are administered to measure expression of thousands of genes. The goal of the experiment is to identify genes whose expressions are affected by dose.

Recently, Lin *et al.* [1] discussed several testing procedures, namely Williams' [4,5], Marcus' [6], the global likelihood ratio test [2], M [7], and the modified M [1] that can be used to identify genes with a monotonic relationship between gene expression and doses. In this paper, we follow up the investigation on evaluating specific monotonic trend of dose response relationship based on the genes selected by one of the procedures mentioned above. The question of primary interest is the nature (or the curve shape) of the dose-response relationship. This question is closely related to the problem of determination of the minimum effective dose

(MED) - that is the smallest dose, at which the mean response is shifted from the mean of dose zero [8,9]. Several testing procedures were proposed for finding the MED. For example, Williams [4,5] proposed a step-down procedure, in which tests are performed sequentially from the highest to the lowest dose level. The procedure stops at the first dose level, for which the null hypothesis (of no dose effect) is not rejected. As a result, the MED is the first dose above that dose level. Other test procedures, proposed by Tamhane *et al.* [9], are based on contrasts among the sample means of gene expressions at different dose levels. Note that Williams' procedure assumes monotonicity of the dose-response relationship, while the tests based on contrasts of the sample means do not require this assumption.

In the microarray setting, the testing procedures mentioned above are additionally subject to the multiplicity problem. To avoid the multiple testing issue in determining the MED and in identifying the shape of the dose-response curve, we propose to classify possible dose-response trends using model selection based on information theory. Assuming a monotone relationship, the dose-response curve could be either linear, nonlinear, concave or convex. Furthermore, for an experiment with $K+1$ dose levels, there is a fixed number of monotonic models that can be fitted.

Table 1: The set of seven possible monotonic dose-response models (with different number of parameters) for an experiment with four dose levels. μ_i is the mean response at dose level d_i, MED: minimum effective dose.

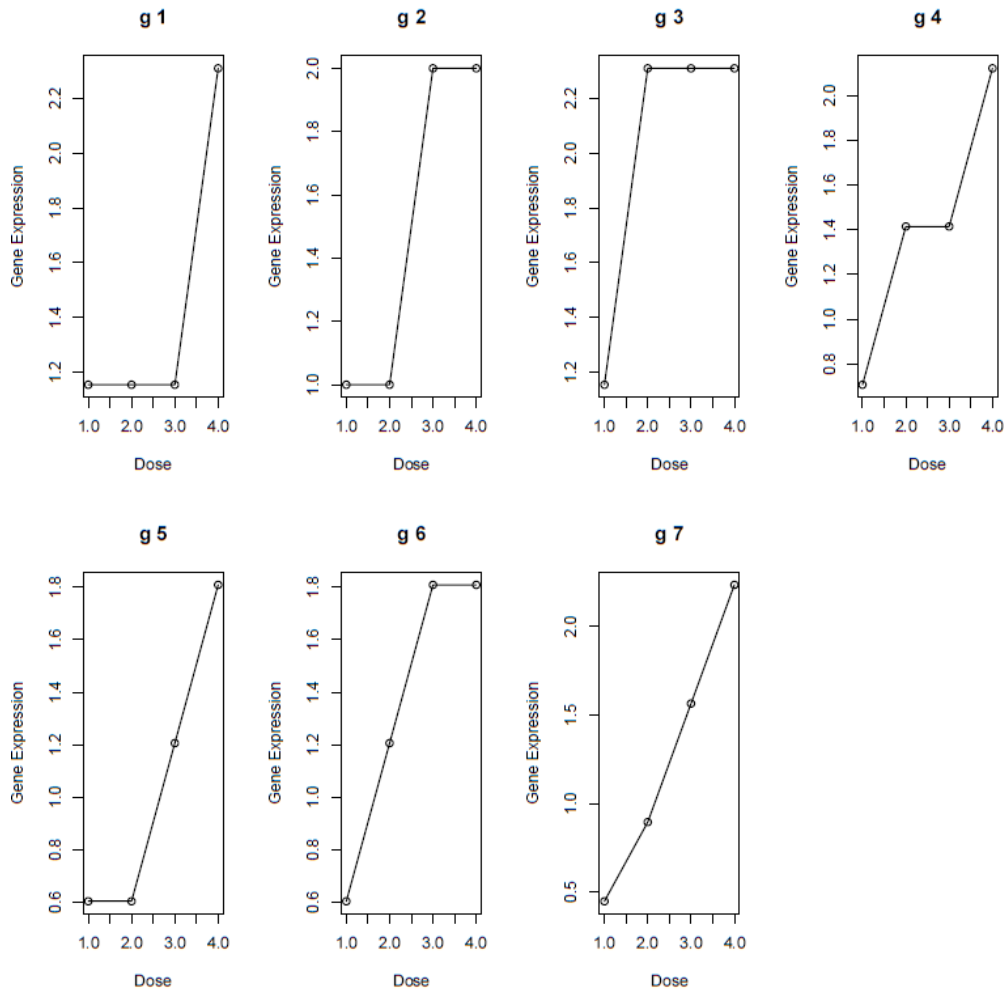| Model	Mean Structure	MED	Parameters
g_1	$\mu_1 = \mu_2 = \mu_3 < \mu_4$	4	2
g_2	$\mu_1 = \mu_2 < \mu_3 = \mu_4$	3	2
g_3	$\mu_1 < \mu_2 = \mu_3 = \mu_4$	2	2
g_4	$\mu_1 < \mu_2 = \mu_3 < \mu_4$	2	3
g_5	$\mu_1 = \mu_2 < \mu_3 < \mu_4$	3	3
g_6	$\mu_1 < \mu_2 < \mu_3 = \mu_4$	2	3
g_7	$\mu_1 < \mu_2 < \mu_3 < \mu_4$	2	4

Figure 1: The set of the seven possible monotonic dose response curves for an experiment with four dose levels.

For instance, in a dose-response experiment with four dose levels, upon the establishment of a monotonic relationship between gene expression and doses, there is a set of seven models, shown in Table 1 and Figure 0, that can be fitted to the data. Each model can be associated with a MED. For example, g_1 is a model with two parameters, and the MED is the last dose level. g_2 is also a model with two parameters, but the MED is the third dose level.

The aim is to choose the best model among these possible monotonic trends. To this purpose, we propose to use the information model selection theory proposed by Burnham and Anderson [10]. For a given set of candidate models the Order Restricted Information Criterion ($ORIC$), the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC) are used to calculate the posterior probability of each model in the set. The model with the highest posterior probability is selected. This allows us to identify both the shape of the dose-response curve and the MED level. In particular, in this paper, we address the importance of using the $ORIC$ to select the dose-response models under order restricted constraints.

For detailed description of data for analysis, we refer to Lin *et al.* [1]. The resulting data after log2 transformation and present/absent filtering consist of 12 samples (for three arrays at four dose levels) with 16,998 probe sets. For simplicity, we refer to probe sets as genes through our paper [11].

The contents of the paper are organized as follows. In Section 2 we briefly discuss the global likelihood test for monotonic trend proposed by Barlow *et al.* [2] and used in Lin *et al.* [1] to identify genes, for which there is a monotone relationship between gene expression and dose. After the initial inference step, in Section 3 we address the problem of trend classification within the framework of model selection. In Section 4, the $ORIC$ is discussed. We apply the proposed methods to our case study in Section 5. Section 6 describes the results of a simulation study conducted to compare the influence of the use of the four information criteria (likelihood, AIC, BIC, and $ORIC$). Section 7 concludes the paper with a discussion.

2   Testing No Dose Effect Against Ordered Alternatives by Using Isotonic Regression

We consider a dose-response microarray experiment described in Lin *et al.* [1]. The first dose level is a control (zero dose).

For each gene, the following ANOVA model is considered

$$Y_{ij} = \mu(d_i) + \varepsilon_{ij}, \ \varepsilon_{ij} \sim N(0, \sigma^2), \ i = 0,1,2,3, \quad j = 1,2,3, \tag{1}$$

where Y_{ij} is the gene expression at the ith dose level for array j and $\mu(d_i)$ is the mean gene expression at dose level d_i. We further assume that gene expression increases or decreases with doses, although not necessarily in a linear fashion. The null hypothesis of no dose effect is

$$H_0 : \mu(d_0) = \mu(d_1) = \mu(d_2) = \mu(d_3), \tag{2}$$

and the alternatives, under the assumption of a monotonic increasing/decreasing trend, is

$$H_1^{Up} : \mu(d_0) \le \mu(d_1) \le \mu(d_2) \le \mu(d_3),$$
$$or \quad H_1^{Down} : \mu(d_0) \ge \mu(d_1) \ge \mu(d_2) \ge \mu(d_3), \tag{3}$$

with at least one strict inequality. Let $\mu^* = (\hat{\mu}_0^*, \hat{\mu}_1^*, \hat{\mu}_2^*, \hat{\mu}_3^*)$ be the maximum likelihood estimates for the means under the ordered alternatives obtained by isotonic regression [2,3]. The $\overline{E}_{01}^2$ test statistic is given by:

$$\overline{E}_{01}^2 = 1 - \frac{\hat{\sigma}_{H_1}^2}{\hat{\sigma}_{H_0}^2} = \frac{\sum\limits_{ij}(y_{ij} - \hat{\mu})^2 - \sum\limits_{ij}(y_{ij} - \hat{\mu}_j^*)^2}{\sum\limits_{ij}(y_{ij} - \hat{\mu})^2}. \tag{4}$$

Lin *et al.* applied the likelihood ratio test statistic $\bar{E}_{01}^2$ to the case study. The distribution of the test statistic $\bar{E}_{01}^2$ was approximated by permutations. In total, the null hypothesis is rejected for 3499 out of the 16,998 genes that were tested (due to a random seed generation for permutations, the number of genes found significant can vary). For a more detailed discussion of the analysis of the data we refer to Lin *et al.* [1].

3   Classification of Trends by Using Information Criteria

3.1   Classification of Trends Based on Posterior Probabilities

The initial testing procedure discussed in the previous section leads to identification of 3499 genes, for which the null hypothesis was rejected. In this section we address the problem of trend classification (or the identification of the dose-response curve shape). For each gene declared significant in the first step, the set $\{g_1, g_2, g_3, g_4, g_5, g_6, g_7\}$, given in Table 1 and shown in Figure 1, is the set of seven possible models with increasing trends for an experiment with four dose levels. Analogously, a set of seven models with decreasing trends is considered as well, with isotonic means given as in Table 1, but with negative values. Note that a model selection procedure that leads to a selection of the best model from the set of all candidate models will allow to identify both dose-response curve shape and the MED. On the other hand, using hypothesis testing for the determination of the MED will not always allow us to identify the shape of the dose-response curve.

The model selection criterion that we propose to use to select the best model is the posterior probability of the model, given the data (Burnham and Anderson 2002), defined by

$$P(g_i \mid D) = \frac{P(D \mid g_i)P(g_i)}{\sum_{r=1}^{R}P(D \mid g_r)P(g_r)} \quad i = 1,\ldots,R. \tag{5}$$

Here, $P(D \mid g_i)$ and $P(g_i)$ are the likelihood and the prior probabilities of the ith model, respectively, and R is the number of all the possible models. Note that if the data are monotonic and we use a non-informative prior, i.e., $P(g_i) = 1/R$, the isotonic regression model obtained will have the highest posterior probability. However, the posterior model probabilities (5) do not take the complexity of the model into account. In what follows we focus on the model selection procedures based on information criteria, that take into account both the goodness-of-fit and the model complexity. In other words, we focus on the question whether the monotonic model can be further simplified.

3.2   Akaike Weights and Bayesian Posterior Model Probabilities

3.2.1 Akaike Weights

The model selection theory discussed by Burnham and Anderson [10,12] allows to incorporate the need to balance between goodness-of-fit and model complexity within the model selection procedure. The starting point for Burnham and Anderson's model selection theory is the Kullback-Leibler information (K-L) given by

$$I(f,g) = \int f(x) log \frac{f(x)}{g(x|\theta)} dx. \tag{6}$$

Here, f represents the density function of the true and unknown model, g represents the density function of the model that is used to approximate f, and θ is the unknown parameter to be estimated. The K-L information is interpreted as the loss of information when the true model f is approximated by the model $g(x|\hat{\theta})$, where $\hat{\theta}$ is the parameter estimate for the unknown parameter θ. For a given set of candidate models $\{g_1, g_2, ..., g_R\}$, one can compare the K-L information for each model and select the model that minimizes the information loss across the considered set of models [10,12,13]. However, in practice $I(f,g)$ can not be computed since the true model f is unknown.

Akaike [14,15] made the link between the K-L information and the likelihood theory and showed that the expected Kullback-Leibler information can be expressed as

$$\hat{E}(K-L) = log\,\ell(\theta|D) - K, \tag{7}$$

where $\ell(\theta|D)$ is the likelihood of model $g(X|\theta)$ and K is the number of parameters in the model. The well-known Akaike's Information Criterion (AIC) is given by

$$AIC = -2log\,\ell(\theta|D) + 2K. \tag{8}$$

Akaike's approach allows for model selection that takes into account both goodness-of-fit and model complexity. Because the individual AIC values are not interpretable, as they contain arbitrary constants and are much affected by sample size, for a given set of R models, Burnham and Anderson (2004) proposed to rescale the AIC to

$$\Delta AIC_i = AIC_r - AIC_{min}, \qquad r = 1, ..., R, \tag{9}$$

with AIC_{min} being the smallest AIC value across the set of R models. The AIC differences, ΔAIC_r, are interpreted as the information loss when model g_r, rather than the best model g_{min}, is used to approximate f. Some simple rules of thumb are used in assessing the relative merits of the models in the set [10]: models with $\Delta AIC_r \leq 2$ have substantial support (evidence); those with $4 \leq \Delta AIC_r \leq 7$ have considerably less support; and models with $\Delta AIC_r > 10$ have essentially no support.

Akaike [16] advocated the use of $\exp(-\frac{1}{2}\Delta AIC_r)$ as the relative likelihood of the model given the data by

$$\ell(g_r \mid D) \propto e^{-\frac{1}{2}\Delta AIC_r}. \tag{10}$$

Note that the model likelihood $\ell(g_r \mid D)$ takes into account both goodness-of-fit and model complexity, while $P(g_r \mid D)$ takes into account only goodness-of-fit. Similar to the posterior probabilities in (5), Akaike [10] defined Akaike's weights by

$$P_A(g_r \mid D) = \frac{exp(-\frac{1}{2}\Delta AIC_r)p(g_r)}{\sum\limits_{r=1}^{R} exp(-\frac{1}{2}\Delta AIC_r)p(g_r)}. \tag{11}$$

Akaike's weight $P_A(g_r \mid D)$ can be interpreted as the weight of evidence that model g_r is the best K-L model given a set of R models and given that one of the models in the set must be the best K-L model. Note that for non-informative prior probabilities $P(g_r) = 1/R$, Akaike's weights can be interpreted as the posterior probabilities of the models.

3.2.2   Bayesian Posterior Model Probabilities

The Bayesian Information Criterion (BIC), proposed by Schwartz [17], is given by

$$BIC = -2log\,\ell(\theta \mid D) + Klog(N), \tag{12}$$

where N is the number of observations (in our case, arrays). The BIC uses a higher penalty on the number of observations than the AIC that penalizes on the number of parameters in the model. Therefore BIC leads to the selection of less complicated models. The posterior model probabilities are given by

$$P_B(g_r \mid D) = \frac{exp(-\frac{1}{2}\Delta BIC_r)p(g_r)}{\sum\limits_{r=1}^{R} exp(-\frac{1}{2}\Delta BIC_r)p(g_r)}. \tag{13}$$

Burnham and Anderson (2002) show that, for prior probabilities defined by

$$P(g_r) = B\exp(\frac{1}{2}\Delta BIC_r)\exp(-\frac{1}{2}\Delta AIC_r) \tag{14}$$

it follows that $P_B(g_r \mid D) = P_A(g_r \mid D)$, where B is a constant.

4 Order Restricted Information Criterion

Anraku [18] proposed an information criterion for parameters under an order restriction ($ORIC$). In contrast to the AIC, the $ORIC$ method is particularly suitable for detecting the configuration of the isotonic means under order restriction.

In contrast to the AIC and BIC, which penalize on the number of parameters in the model or on the sample size, the $ORIC$ takes into account the level probability for the number of parameters under each order restricted models in set R. The $ORIC$, proposed by Anraku [18] is given by,

$$ORIC = -2log\,\ell(\theta\,|\,D) + \sum_{i=1}^{K} iP(i,K,w_i),\tag{15}$$

where $P(i,K,w_i)$ is the level probability [3] that for given $K+1$ doses under H_0 the isotonic regression will result in i unique isotonic means, $w_i = n_i/\tau_i$, n_i is the number of arrays at dose i ($i = 0,\ldots,K$), and τ_i is the variance at dose i. For the case that $w_i = \ldots = w_K$, or $K = 2$, it follows that $ORIC = -2log\,\ell(\theta\,|\,D) + \sum 1/i$ [3].

The posterior model probabilities (with non-informative priors) are given by

$$P_{OR}(g_r\,|\,D) = \frac{exp(-\frac{1}{2}\Delta ORIC_r)}{\sum_{r=1}^{R} exp(-\frac{1}{2}\Delta ORIC_r)}.\tag{16}$$

Under simple order alternatives, i.e., H_1^{Up} or H_1^{Down}, in the setting of four doses (control and three higher doses) and equal number of arrays per dose, the level probabilities $P(i,K,w_i)$, given by Robertson *et al.* [3], are equal to 0.25, 0.45833, 0.25, and 0.04167 for $i = 0$, 1, 2, and 3, respectively.

Table 2 lists the value of penalty given by the AIC, BIC, and $ORIC$. Note that the relative magnitude of penalty used by the $ORIC$ for models g_1, g_2, g_3 (two parameters) and models g_4, g_5 g_6 (three parameters) is large as compared to that used for model g_7 (four parameters). Thus, due to the small penalty difference between models g_4, g_5, g_6, and model g_7, the latter is more likely classified as the best model in set R.

Table 2: Comparison of penalty values by the AIC, BIC and $ORIC$ for models g_1 to g_7. N is the total number of arrays.

Models	Parameters	*AIC*	*BIC*	*ORIC*

g_1, g_2, g_3	2	2×2	$2 \times log(N)$	2×1.16666
g_4, g_5, g_6	3	2×3	$3 \times log(N)$	2×1.91666
g_7	4	2×4	$4 \times log(N)$	2×2.08334

So far we have discussed the model selection by using the information theory in order to classify the dose-response curve shape and identify the MED level. An alternative approach to determine the MED are the hypothesis testing procedures, such as Williams' [4,5] and Marcus' [6] procedures, which identify the MED based on hypothesis test conducted in a step down fashion. However, identification of the MED by using these procedures does not imply the shape of the dose-response curve. On the other hand, model selection procedures based on the information theory can simultaneously address two objectives - determination of the MED and classification of trends. Once the best model from the candidates is selected, the MED can be identified from the selected model. Note that the methods discussed above are conditional on the result of testing for monotonic trend that is essential. The model selection procedure is applied to those genes, for which there is an evidence of a monotone relationship between gene expression and dose. Hence, the best monotone model must be in the model set as required. Moreover, the initial testing step reduces the computational time for the model selection, as the direction of the trend (upward or downward trend) is known from the initial step.

5 Application to the Data

Based on the 3499 genes found to be significant in the initial inference step (see Section 3.1), the classification of trends for these genes is to be obtained. Using the four information criteria (likelihood, AIC, BIC, and $ORIC$) the genes are classified into the seven curve shapes shown in Figure 1 (see Table 3). Using the likelihood-based posterior probabilities, defined in (5), 1710 genes (48.85%) are classified as g_7 - the isotonic regression model with four parameters. As shown in Table 3, when the AIC and BIC criteria are used to calculate the posterior probabilities, the number of genes that are classified as g_1 (isotonic regression with two parameters) increases from 344 to 1528 and 1648, respectively. The same pattern is observed for models g_2 and g_3 (both are isotonic regression models with two parameters). For the $ORIC$, the number of genes classified to models from g_1 to g_6 decreases as compared to the AIC and BIC. On the other hand, 816 genes are classified as g_7.

Table 3: Classification of genes into dose-response trends. Numbers of genes classified as each model (from g_1 to g_7) are shown in the table.

Model	Likelihood	AIC	BIC	$ORIC$
g_1	344	1528	1648	1348

g_2	25	307	369	221
g_3	14	106	126	86
g_4	343	370	337	253
g_5	885	823	715	655
g_6	178	170	149	120
g_7	1710	195	155	816
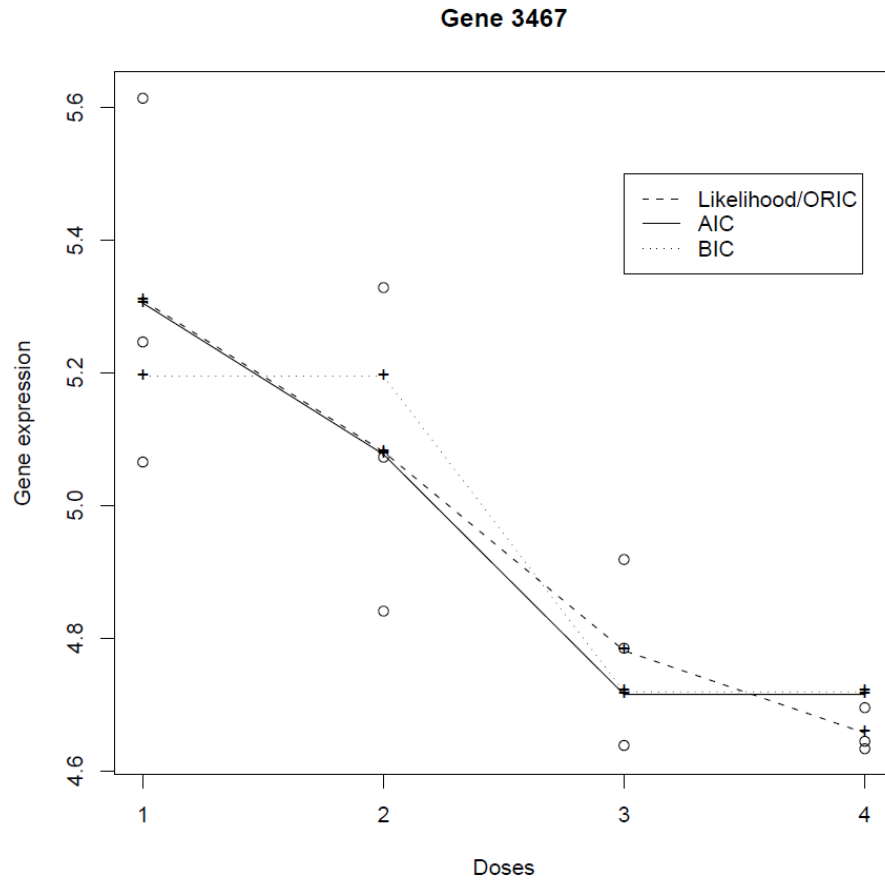
Figure 2: Classification of trend for an example gene. The best model according to the likelihood, AIC, BIC, and $ORIC$. Solid line is the isotonic regression model g_7 (four parameters) selected by the likelihood and $ORIC$, dotted line is g_6 (three parameters) selected by the AIC, and dashed line is g_2 (two parameters) selected by the BIC.

Figure 2 shows the data for gene 3467. Based on the likelihood and $ORIC$, the best model for the gene is g_7 (solid line). Using the AIC, the gene is classified as g_6 (dotted line). For both models, the second dose level is selected to be the MED level. Using the BIC, the model is further reduced and the gene is classified as g_2 (dashed line), and the MED level

is identified as the third dose level.

Note that, although models g_1, g_2, and g_3 have the same number of parameters, the MED are not the same. The effective dose level for model g_1 is the highest dose level, while for g_2 and g_3 the MED are the third and the second dose levels, respectively. In general, the AIC and BIC criteria favor the two-parameter models, as compared to the three-parameter models g_4, g_5, g_6, and to the four-parameter model g_7, which is favored by the likelihood. Table 3 shows that the AIC and BIC classify most of genes as g_1 and g_5, the $ORIC$ classifies most of genes as g_1 and g_7, while the likelihood criterion classifies genes mostly as g_7.

6 Simulation Study

Two simulation studies are conducted to investigate the performance of the model selection procedure. In particular, two questions are of interest: (1) is the initial inference step of filtering out genes with monotonic trends necessary, and (2) when the AIC, BIC, and $ORIC$ are used to reduce the complexity of the models, does the selection favor some models over others? In the first study, we compare the proposed model selection methods for classification of dose response curve shapes with and without the initial step of testing monotonic trends. In the second study, we evaluate whether the model selection criteria are sensitive to the shape of the seven possible dose-response relationships.

6.1 The Influence of the Initial Inference Step

In this study, we investigate two different approaches to classification of genes into the seven models under ordered alternative (see Figure 1), under the null model and under non-monotonic models. The first approach is to classify all the genes into different models from the experiment by using the four information criteria. Hence, the classification is done without an initial inference step. The second approach is divided into two steps. In the first step null hypothesis (2) is tested against order restricted alternatives (3) or (4). After the initial step of filtering out non significant genes, in the second step, monotonic genes are further classified into one of the seven monotonic models. The first approach avoids the problem of hypothesis testing and multiplicity. The second approach zooms in on the genes of interest with monotonic trends, and controls the FDR.

The performance of the two approaches is evaluated by the misclassification error and correct classification for each model. In the simulations 100 data sets with 12 microarrays (three arrays for each of four dose levels) and 17,000 genes each, are randomly generated. Among all the genes, 250 genes are generated under each of the seven models with an increasing and decreasing trend that accounts for 3500 genes. Another 3500 genes are generated under a non-monotonic trend. Finally, the remaining 10,000 genes are generated under the null model. The sample means used for simulation are given in Table 4, with scale factor $\lambda = 1$ and the variance $\sigma^2 = 0.01$.

Table 4: Means for simulated data. For each model (from g_1 to g_7), each μ_i is multiplied by the factor λ given in the last column, where $\lambda = 1$.

Model	μ_1	μ_2	μ_3	μ_4	$\lambda = 1$
g_1:	(1	1	1	2)	$\times 2\lambda/\sqrt{3}$
g_2:	(1	1	2	2)	$\times \lambda$
g_3:	(1	2	2	2)	$\times 2\lambda/\sqrt{3}$
g_4:	(1	2	2	3)	$\times \lambda/\sqrt{2}$
g_5:	(1	1	2	3)	$\times 2\lambda/\sqrt{11}$
g_6:	(1	2	3	3)	$\times 2\lambda/\sqrt{11}$
g_7:	(1	2	3.5	4)	$\times \lambda/\sqrt{5}$
null:	(0	0	0	0)	λ
non-mon:	(1	-1	2	-2)	$\times \lambda/\sqrt{5}$

The results of the simulations (with correct classification and misclassification error rates) are summarized in Table 5 (first approach: classification without initial inference step) and Table 6 (second approach: initial inference step follows by classification step).

Table 5: Misclassification and correct classification rates under each simulated model (from g_1 to g_7). Fist approach: classification without initial inference step. L=Likelihood, A=AIC, B=BIC, O=$ORIC$, NA=Not Available.

Trend	Model	Error				Correct			
		L	A	B	O	L	A	B	O
	g_1	NA	0.7113	0.6608	0.8593	0.0000	1	1	0.9995
	g_2	NA	0.6968	0.6501	0.8481	0	0.9960	0.9981	0.9885
U	g_3	NA	0.7125	0.6617	0.8593	0	0.9999	0.9999	0.9993
p	g_4	0.9421	0.2780	0.2499	0.3389	0.0011	0.7842	0.7903	0.6741
	g_5	0.8847	0.0266	0.0146	0.0512	0.0019	0.9834	0.9895	0.8513
	g_6	0.8674	0.2378	0.2659	0.0979	0.0018	0.9830	0.9890	0.8469
	g_7	0.7884	0.0710	0.0503	0.3540	0.9965	0.7176	0.6526	0.9545
	g_1	NA	0.7118	0.7144	0.9330	0	0.99988	1	0.9992
	g_2	NA	0.6962	0.6500	0.8480	0	0.9966	0.9984	0.9890
D	g_3	NA	0.7103	0.6588	0.8592	0	0.9999	0.9999	0.9996
o	g_4	0.8835	0.098	0.0608	0.1766	0.0024	0.9792	0.9861	0.8352
w	g_5	0.8799	0.0273	0.0153	0.0537	0.0018	0.9825	0.9886	0.8447
n	g_6	0.8495	0.2415	0.2672	0.0977	0.0023	0.985	0.9900	0.8492

		0.7882	0.0713	0.0523	0.3567	0.9970	0.7101	0.6483	0.9544
	g_7								
	null	NA	0	0	0	0	0.6163	0.7053	0.0954
	non.mon	0.7606	0.0398	0.0077	0.0091	1	0.9999	0.9604	0.4434

Table 6: Misclassification and correct classification rates under each simulated model (from g_1 to g_7). Approach two: initial inference step followed by classification step. L=Likelihood, A= AIC, B= BIC, O= $ORIC$, NA=Not Available.

Trend	Model	Error				Correct			
		L	A	B	O	L	A	B	O
	g_1	0.022	0.0291	0.0320	0.0263	0.3369	0.9999	1	0.9995
	g_2	0.0272	0.0310	0.03560	0.0261	0.2505	0.9960	0.99810	0.9885
U	g_3	0.0253	0.0300	0.0327	0.0265	0.333	0.9999	0.9999	0.9993
p	g_4	0.6112	0.2188	0.2160	0.2191	0.4033	0.7842	0.7903	0.6741
	g_5	0.5093	0.0149	0.0108	0.0182	0.5041	0.9834	0.9895	0.8513
	g_6	0.5106	0.2307	0.2638	0.0661	0.4995	0.9830	0.9890	0.8469
	g_7	0.6796	0.0710	0.0503	0.3350	0.9965	0.7176	0.6526	0.9545
	g_1	0.0255	0.0456	0.0480	0.0427	0.3336	0.9999	1	0.9992
	g_2	0.0216	0.0289	0.0341	0.0236	0.2507	0.9966	0.9984	0.9890
D	g_3	0.0270	0.0300	0.0322	0.0269	0.3399	0.9999	0.9999	0.9996
o	g_4	0.5220	0.0238	0.0210	0.0259	0.4972	0.9792	0.9861	0.8352
w	g_5	0.5141	0.0144	0.0113	0.0179	0.4987	0.9825	0.9886	0.8447
n	g_6	0.5097	0.2338	0.2649	0.0660	0.5025	0.985	0.9900	0.8492
	g_7	0.6798	0.0713	0.0523	0.3991	0.9970	0.7102	0.6483	0.9544
	null	NA	0	0	0	0	0.8796	0.9390	0.8796
	non.mon	0.7396	0.2449	0.1337	0.2449	0.9989	0.9988	0.9988	0.9988

The misclassification and correct classification rates are computed as follows. Suppose that X_i genes are classified as model g_i. Among these X_i genes, Y_i genes are truly generated under g_i. The proportion of correct classifications for model g_i is thus equal to $Y_i/250$; for non-monotonic model, it is equal to $Y_i/3500$; and for the null model, it is equal to $Y_i/10000$. The misclassification rate is defined as the relative proportion, mimicking the concept of FDR: the proportion of misclassifications (false positives) among the number of classifications (total discoveries). The proportion of misclassifications for each model is equal to $(X_i - Y_i)/X_i$. From Tables 5 and 6, it is easy to note that the correct classification rate obtained for the two approaches, is quite similar for the AIC and BIC, while the likelihood criterion under the second approach yields a higher correct classification rate than that under the first approach.

This is because the testing procedure filters out non-significant genes and controls the FDR. On the other hand, the misclassification errors, obtained by using the AIC, BIC, and $ORIC$, for the seven models under the order alternatives in the second approach, are much lower than those in the first approach. The initial testing step in the second approach selects the genes rejected with a test that results in favor of the monotonic ordered alternatives and allows them to be correctly classified into one of the seven models. This greatly improves the misclassification error. On the other hand, in the first approach, a large number of non-monotonic genes are misclassified into one of the seven models under the order restriction after the penalization used by the AIC, BIC, and $ORIC$, that results in high misclassification errors. Thus, the second approach is recommended to ensure a correct classification of genes of interests.

We also observe that, for g_7, both the misclassification error and the correct classification proportion, obtained by using the $ORIC$, are higher than those obtained by using the AIC and BIC. However, the opposite results can be observed for models g_4, g_5, and g_6. This is because the penalty used by the $ORIC$ favors g_7 as noted in Section 4. As the number of genes classified as g_7 (X_7) increases, the number of correct classification (Y_7) increases as expected. Thus, the proportion of correct classification $Y_7/250$ also increases. On the other hand, the misclassification error $(X_7 - Y_7)/X_7$ increases at the same time, due to an increase in X_7.

6.2 Evaluation of the Four Criteria

In the simulation study, 100 data sets are generated, each with 12 microarrays (three arrays for each of four dose levels) containing 5000 genes. The proportion of genes generated under different models, follows the results obtained by the AIC shown in Table 3. The isotonic means at different dose levels are assumed to correspond to those shown in Table 4 with $\lambda =1$ and $\sigma^2 = 0.01$. The results obtained by using the four information criteria are presented in Table 7. As we observe, except for model g_7, classification based on the AIC and BIC achieves better results when compared to the classification based on the likelihood. The former reduces not only the misclassification error, but also its standard error. However, as compared to the AIC and BIC, the $ORIC$ achieves a better classification for model g_7, while it performs worse for models g_1-g_6.

For example, 2184 genes (i.e., 43% of 5000 genes in total in each data set) are generated under g_1. Using likelihood criterion for classification, 25% of these genes are misclassified as g_4, 24.87% are misclassified as g_5, and 16.5% as g_7. The total misclassification error rate (number of correct classification/2184) is 66.46% (=25%+24.87%+16.5%) with the standard error of 0.0105. When using the AIC and BIC, the error rates are reduced to 0.01% and 0.004%, respectively. Similar patterns are observed for all the models except g_7. For model g_7, the misclassification is due to the fact that g_6 that is similar to g_7 when the isotonic means are considered with only a slight difference at the two highest dose levels.   Table 7:

Simulation results for misclassification rates under each simulated model (from g_1 to g_7) based on likelihood (L), *AIC* (A), *BIC* (B) and *ORIC* (O). Misclassification error in % (standard error).

True Model	Misclassification Models	L	A	B	O
g_1	Total error	66.46	0.01	0.00	0.05
		(0.0105)	(0.0002)	(0.0001)	(0.0005)
	g_4	25.10	0.00	0.00	0.02
	g_5	24.87	0.01	0.00	0.03
	g_5	16.50	0.00	0.00	0.00
g_2	Total error	74.69	0.36	0.14	1.11
		(0.0212)	(0.0027)	(0.0018)	(0.005)
	g_5	25.10	0.20	0.06	0.37
	g_6	24.93	0.15	0.07	0.30
	g_7	24.66	0.01	0.00	0.44
g_3	Total error	66.72	0.01	0.00	0.03
		(0.0443)	(0.0009)	0.00	(0.0013)
	g_4	25.07	0.01	0.00	0.01
	g_6	25.19	0.01	0.00	0.01
	g_7	16.46	0.00	0.00	0.00
g_4	Total error	49.59	2.07	1.38	15.83
		(0.0202)	(0.0058)	(0.0048)	(0.0145)
	g_7	49.59	2.07	1.38	15.83
g_5	Total error	49.90	1.78	1.18	15.26
		(0.0154)	(0.0039)	(0.0033)	(0.0102)
	g_1	0.00	0.00	0.00	0.00
	g_2	0.00	0.00	0.00	0.00
	g_7	49.90	1.78	1.17	15.26
g_6	Total error	50.06	1.67	1.05	15.42
		(0.0307)	(0.0098)	(0.0075)	(0.0241)
	g_2	0.00	0.00	0.00	0.00
	g_7	50.06	1.67	1.05	15.42
g_7	Total error	0.29	28.08	34.61	4.41
		(0.0032)	(0.0261)	(0.025)	(0.0121)
	g_2	0.00	0.11	0.26	0.01

	g_5	0.00	0.05	0.06	0.00
	g_6	0.29	27.92	34.28	4.40
Mod g_7	Total error	0.00	1.23	2.42	0.00
		(0.0032)	(0.0261)	(0.0250)	(0.0121)
	g_4	0.00	0.61	1.22	0.00
	g_5	0.00	0.61	1.20	0.00

To enhance a distinction between g_6 and g_7, we modified the isotonic means for model g_7 to $(1\ 2\ 3\ 5)\times/\sqrt{5}$, that makes them substantially different from those for model g_6. After the modification, the results of this simulation for model g_1-g_6 are the same as shown in Table 7. The results for model g_7 are summarized in the last row of Table 7. It shows a substantially improved performance for model g_7 for the AIC and BIC.

From the last row of Table 7 we can observe that the performance of classification by using the AIC and BIC for the modified model g_7 is satisfactory. The misclassification errors are 0.0122 and 0.0242 when using the AIC and BIC, respectively. When model g_7 is substantially different from the other models, the AIC and BIC perform as good as the likelihood.

7 Discussion

In this paper, we have investigated the issue of classifying dose response curve shapes for the genes that tested positively for a monotonic trend. After the initial inference step, the set of all possible models is considered and a non-informative prior is given to these models. The likelihood, AIC, BIC, and $ORIC$ are used to assess the goodness-of-fit and complexity of the models. The Akaike weights, representing the posterior probabilities of the models, are calculated from the AIC of the models; Bayesian posterior probabilities are obtained from the BIC; the posterior probability obtained from the $ORIC$ takes order restriction into account. We have shown that using the likelihood to calculate the posterior probabilities of the model leads to the selection of the isotonic regression model, which maximizes the likelihood under the ordered constraints. For the setting of four doses, the $ORIC$ tends to classify genes as model g_7 as opposed to models g_4, g_5, and g_6 that have one parameter less. And, as expected, using the AIC and BIC leads to a selection of simpler models with less parameters (compared with the model selection based on the likelihood). Moreover, we have shown that after a classification is made, the mean structure of the selected model can be used to estimate the minimum effective dose level.

From our simulation study, we note that the initial inference step of testing monotonic trends reduces the misclassification error by filtering out the genes with non-monotonic trends. Moreover, the classification based on the AIC and BIC leads to a smaller classification

error when compared to the use of likelihood. The $ORIC$ performs better for model g_7, but worse for the other models in set R.

The R is used to perform the analysis in our paper.

Acknowledgement

References

[1]   D. Lin, Z. Shkedy, D. Yekutieli, T. Burzykowki, H.W.H. Göhlmann, A. De Bondt, T. Perera, T. Geerts, L. Bijnens, "Testing for trend in dose-response microarray experiments: comparison of several testing procedures, multiplicity, and resampling-based inference". *Stat. Appl. Genet. Mol. Biol*. Vol. 6 : Iss. 1, Article 26, 2007.

[2]   R.E. Barlow, D.J. Bartholomew, M.J. Bremner, and H.D. Brunk,   *Statistical Inference Under OrderRestriction*, Wiley: New York, 1972.

[3]   T. Robertson, F.T. Wright, and R.L. Dykstra, *Order Restricted Statistical Inference*, Wiley, 1988.

[4]   D.A. Williams, "A test for differences between treatment means when several dose levels are compared with a zero dose control".   *Biometrics*, vol. 27, pp. 103-117, 1971.

[5]   D.A. Williams, "The comparison of several dose levels with a zero dose control". *Biometrics*, vol. 28, pp. 519--531, 1972.

[6]   R. Marcus, "The powers of some tests of the quality of normal means against an ordered alternative",   *Biometrika*, vol. 63, pp. 177-83, 1976.

[7]   J. Hu, M. Kapoor, W. Zhang, S.R. Hamilton, and K.R. Coombes, "Analysis of dose response effects on gene expression data with comparison of two microarray platforms", *Bioinformatics*, vol. 21(17), pp. 3524--3529, 2005.

[8]   S.J. Ruberg, "Contrasts for identifying the minimum effective dose".   *J. Am. Stat. Assoc.*, vol. 84, pp. 816-822, 1989.

[9]   A.C. Tamhane, Y. Hochberg, and C.W. Dummett, "Multiple test procedures for dose findings".   *Biometrics*, vol. 52, pp. 21-37, 1996.

[10]   K.P. Burnham, and D.R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* Springer:New-York,   2002.

[11]   E. Hubbell, W.M. Liu, and R. Mei. "Robust estimators for expression analysis", *Bioinformatics*,   vol.18(12), pp.1585-1592, 2002.

[12]   K.P. Burnham, and D.R. Anderson, "Multimodel inference: understanding AIC and BIC in model selection",   *Soc.   Methods Res.*, vol. 33, pp. 261-304, 2004.

[13]   E. Poeter and D. Anderson. "Multimodel ranking and inference in ground water modeling".   *Ground Water*, vol. 43(4), pp. 597-605, 2005.

[14]   H. Akaike, *"Information theory and an extension of the maximum likelihood principle",   B. Petrov and B. Csaki, Eds.* Second International Symposium on Information Theory,

Academiai Kiado, Budapest, pp.267–281, 1973.

[15]　H. Akaike, "A new look at the statistical model identification". *IEEE Transact. Autom. Control*, AC-19, pp.716-723, 1974.

[16]　H. Akaike, "Likelihood of a model and information criteria". *J.Econom.*, vol. 16, pp. 3–14, 1981.

[17]　M.J. Schwarz, "Estimating the dimension of a model". *Ann. Stat.*, vol. 6, pp. 461-464, 1978.

[18]　Anraku K., "An information criterion for parameters under a simple order restriction". *Biometrika*, vol. 86(1), pp.141-152(12), 1999.