A Test for Partial Differential Expression

Wessel N. VAN WIERINGEN, Mark A. VAN DE WIEL, and Aad W. VAN DER VAART

Even in a single-tissue type cancer is often a collection of different diseases, each with its own genetic mechanism. Consequently, a gene may be expressed in some but not all of the tissues in a sample. Differentially expressed genes are commonly detected by methods that test for a shift in location that ignore the possibility of heterogeneous expression. This article proposes a two-sample test statistic designed to detect shifts that occur in only a part of the sample (partial shifts). The statistic is based on the mixing proportion in a nonparametric mixture and minimizes a weighted distance function. The test is shown to be asymptotically distribution free and consistent, and an efficient permutation-based algorithm for estimating the p value is discussed. A simulation study shows that the test is indeed more powerful than the two-sample t test and the Cramér–von Mises test for detecting partial shifts and is competitive for whole-sample shifts. The use of the test is illustrated on real-life cancer datasets, where the test is able to find genes with clear heterogeneous expression associated with reported subtypes of the cancer.

KEY WORDS: Asymptotics; Consistency; Differential gene expression; Nonparametric mixture; Permutation test; Weighted minimum distance.

1. INTRODUCTION

Knowledge of the human genome and its expression may greatly enhance our understanding of cancer (Brown and Botstein 1999). Microarrays are devices that can be used to measure the expression level of many genes simultaneously and have shown to be a promising means to acquire this knowledge. The reader is referred to Nguyen, Arpat, Wang, and Carroll (2002) for an excellent overview of the biological and technological aspects of microarrays.

In this article we focus on comparative microarray experiments carried out to identify genes that are differentially expressed between two conditions (e.g., normal tissue vs. cancerous tissue). Examples of such experiments are described, for example, in Schummer et al. (1999) and LaPointe et al. (2004). Knowing the differentially expressed genes may help us to understand the genetic mechanism underlying the disease.

The problem of finding differentially expressed genes is commonly restated as a problem of hypothesis testing (cf. Dudoit, Shaffer, and Boldrick 2003): A gene is said to be differentially expressed if the null hypothesis that the gene's expression level is unaffected by the condition is rejected. This null hypothesis is tested for each gene on the array. Multiple testing corrections are used to control the false discovery rate (FDR) or family-wise error rate (FWER).

Here we are interested in detecting genes that are possibly expressed in only a part of the cases or expressed at different levels among the cases. A well-known example is the ErbB2 gene, which is "over-expressed (with respect to normal tissue) in only 9–30% of ovarian carcinomas" (Pejovic 1995). Standard tests for differential expression (such as the t test) are designed to be sensitive to a general shift in location between two groups and are less suitable for detecting genes with partial differential expression. In this article we propose a new test statistic, motivated by a nonparametric mixture model. This statistic minimizes a weighted distance function.

Our mixture model is on a gene-by-gene basis and plays a very different role than the mixture models used in earlier articles on microarray studies (see, e.g., Broët, Richardson, and Radvanyi 2002; Efron, Tibshirani, Storey, and Tusher 2001). In previous studies the mixture components reflect the differentially expressed and nonexpressed genes, and the mixture model is used to calculate the (posterior) probability of a gene being differentially expressed. In our case the gene-specific mixture proportions are metrics for differential expression.

We derive the asymptotic distribution of the test statistic in Appendix A. This shows that for certain weight functions the test is approximately distribution free. We obtain an exact null distribution of the test statistic by permutation resampling, where we discuss a technique to speed up computation of the p values. The asymptotic results also show that the permutation test is consistent.

The power of our test is investigated in an extensive simulation study. We compare the proposed test with the two-sample permutation t test and the Cramér–von Mises test. We conclude that the proposed test is more powerful than those two tests when the alternative is a partial shift, and it is competitive for whole-sample shifts. Moreover, as opposed to the t test, it is robust against outliers and heavy tails.

Finally, we apply the technique to the datasets of Schummer et al. (1999) and LaPointe et al. (2004). The proposed test finds genes whose expression in the sample is clearly bimodal, indicating a partial shift, which are not found by the t test. It is also more powerful in detecting these genes than the Cramér–von Mises test.

2. MODEL

Consider a comparative microarray experiment involving a sample of n, say, normal tissues and m, say, cancerous tissues. Associated with each tissue is a gene expression profile $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})$, where X_{ij} is a random variable representing the expression level of gene j, $j = 1, \ldots, p$, of tissue i, $i = 1, \ldots, N = n + m$. Together the expression profiles of all

Wessel N. van Wieringen is Postdoctoral Scientist, Department of Mathematics, Vrije Universiteit, 1081 HV Amsterdam, The Netherlands (E-mail: *wvanwie@few.vu.nl*). Mark A. van de Wiel is Associate Professor, Department of Mathematics, Vrije Universiteit and Department of Pathology and Department of Biostatistics, VU University Medical Center, 1007 MB Amsterdam, The Netherlands (E-mail: *mark.vdwiel@vumc.nl*). Aad W. van der Vaart is Professor, Department of Mathematics, Vrije Universiteit, 1081 HV Amsterdam, The Netherlands (E-mail: *AW.van.der.Vaart@few.vu.nl*). This work was supported in part by the Center for Medical Systems Biology (CMSB) established by the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research (NGI/NWO). The authors thank Bauke Ylstra and Steven IJland for the discussions on the biological aspects of the article, and Michèl Schummer for providing the annotation of the ovarian cancer dataset. Supplementary material and software of the proposed method for the statistical computer program R (R Development Core Team 2005) is available on request from the authors.

Journal of the American Statistical Association
September 2008, Vol. 103, No. 483, Applications and Case Studies
DOI 10.1198/016214507000001319

tissues make up $\mathbf{X}$, the $(N \times p)$ expression matrix, with its realization $\mathbf{x}$, the outcome of the experiment.

We are interested in the situation that the expression in cancerous tissues may be heterogeneous, with only part of the sample being differentially expressed relative to expression in the normal tissues. This part may be different for different genes. The (population) proportion of cancerous tissues that is differentially expressed is modeled by a parameter τ_j. In Section 4 we extend our approach to the situation that normal tissues may also have heterogeneous expression.

Let the expression levels of gene j in normal tissue be distributed according to a density $f_j(x_{ij})$. We assume that the expression levels in the cancerous tissue follow a mixture density:

$$h_j(x_{ij}) = (1 - \tau_j) f_j(x_{ij}) + \tau_j g_j(x_{ij}). \tag{1}$$

Here $g_j(x_{ij})$ is an unspecified density modeling the expression level of gene j if it is differentially expressed. In the above we modeled the marginal densities; hence, we base the test statistic, related to τ_j, only on data for gene j. We discuss the effect of dependency between genes later on. The cumulative distribution functions of h_j, f_j, and g_j are denoted by capital letters: H_j, F_j, and G_j.

3. THE TEST STATISTIC

Model (1) will serve as the alternative in testing the differential expression of gene j. It is different from the null model $f_j(x_{ij})$ only if both $\tau_j > 0$ and $g_j \neq f_j$. To motivate our test statistic, we first assume that f_j and g_j are known and consider potential candidates for the value of τ_j. Initially, we assume positive shifts, that is, $G_j(x) \leq F_j(x)$. As a population parameter that reflects the degree of partial differential expression we use the value $\theta_j \in [0, 1]$ that minimizes a weighted distance $\delta((1-\theta_j)F_j(x), H_j(x); w_j(x))$ between $(1-\theta_j)F_j(x)$ and H_j, for suitably chosen nonnegative weights $w_j(x)$. This population parameter is inspired by the theory on the estimation of mixing proportions using minimum distance estimators (Titterington, Smith, and Makov 1985). It is based on shrinking to the left mixture component of $H_j(x)$ (see Fig. 1).

For the L_2 distance, a natural choice for a distance, the population parameter θ_j is given by

$$\arg \min_{\theta \in [0,1]} \int_{-\infty}^{\infty} \big((1-\theta)F_j(x) - H_j(x)\big)^2 w_j(x)\, dx$$

$$= 1 - \frac{\int_{-\infty}^{\infty} F_j(x)\, H_j(x)\, w_j(x)\, dx}{\int_{-\infty}^{\infty} F_j^2(x)\, w_j(x)\, dx}. \tag{2}$$

As a comparison we consider two robust distance measures, the L_1 distance and the Hellinger distance. The Hellinger distance has been shown to be quite robust and efficient when used in the estimation of finite mixture models (see, e.g., Cutler and Cordero-Braña 1996). For an overview of distances $\delta(\cdot, \cdot)$, including other choices, we refer to Titterington et al. (1985).

In the absence of (partial) differential expression, we have $H_j(x) = F_j(x)$; clearly, the population parameter equals 0. Given positive partial differential expression, that is, (1) with $\tau_j > 0$ and $G_j(x) \leq F_j(x)$, the population parameter, using the L_2 distance, equals

$$\tau_j - \tau_j \frac{\int_{-\infty}^{\infty} F_j(x) G_j(x) w_j(x)\, dx}{\int_{-\infty}^{\infty} F_j^2(x) w_j(x)\, dx}. \tag{3}$$

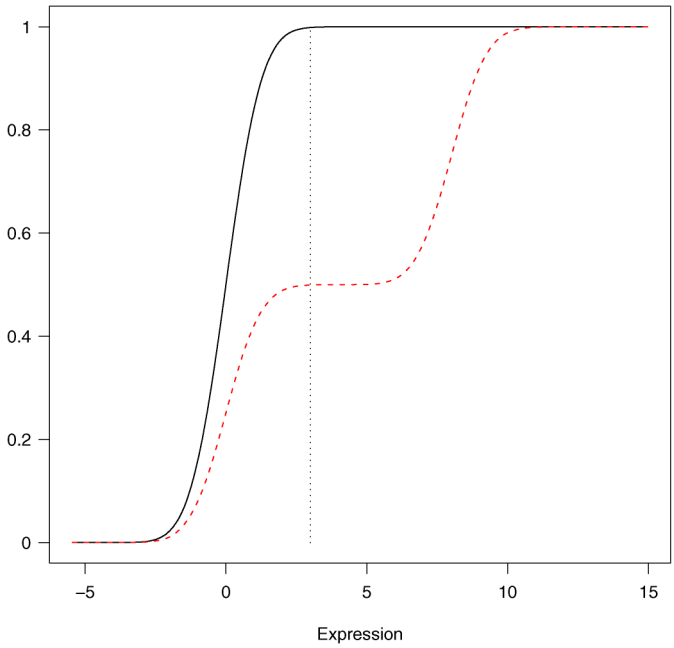

Figure 1. Cumulative distributions of F_j (—) and H_j (- - -). The dotted vertical line separates the supports of f_j and g_j.

We would like to choose the weights $w_j(x)$ such that this expression is large in this case. In the ideal situation where the supports of f_j and g_j are disjoint, we could choose the weight $w_j(x)$ equal to 0 if x is the support of g_j, and positive if x is the support of f_j, in which case the second term in (3) is 0, and the population parameter is maximized to τ_j. We refer to Figure 1 for an illustration of restricting the weights to the support of f_j.

Of course, in practice, we do not know F_j and G_j or their supports. We study several feasible weights that are nonzero on the support of f_j and almost zero on the support of g_j: the empirical distribution corresponding to $f_j(x)$, denoted $w_{1,j}(x)$; a kernel density estimator (normal kernel with a normal optimal smoothing parameter; see Bowman and Azzalini 1997) of $f_j(x)$, denoted $w_{2,j}$; $w_{3,j}(x) = F_j(x)(1 - F_j(x))$, whose support contains that of $f_j(x)$; and $w_{4,j}(x) = (1 - F_j(x))/(1 - H_j(x))$, motivated by the fact that $w_{4,j}(x_{ij}) \propto P(Z_{ij} = 0 | X_{ij} > x_{ij})$ if Z_{ij} is a binary variable indicating whether cancerous tissue i is differentially expressed on gene j. The supplementary material (*http://www.amstat.org/publications/jasa/supplemental_materials*) shows the behavior of the population parameter (with every combination of proposed distances and weights) under various degrees of partial and gradual differential expression.

The population parameter θ_j becomes a test statistic, denoted Θ_j, by replacing $F_j(x)$ and $H_j(x)$ by the empirical cumulative distribution functions $\mathbb{F}_m(x)$ and $\mathbb{H}_n(x)$ of expression on the jth gene in the normal and cancer group, respectively. We also use these estimates to obtain empirical versions of the weights $w_{3,j}(x)$ and $w_{4,j}(x)$. [We have chosen to set $w_{4,j}(x)$ equal to 0 if $1 - \mathbb{H}_n(x)$ equals 0 and $1 - \mathbb{F}_m(x)$ does not.] For the L_2 and Hellinger distances the minimum of (2) can be calculated explicitly (for L_2 see Sec. 4). For the test statistics involving the L_1 distance, the distance measure in (2) is evaluated on a fine grid of values for $\Theta_j \in [0, 1]$. The value of Θ_j that yields the minimum distance is the test statistic for gene j. If multiple Θ_j's

minimize the distance measure, the average over these Θ_j's is taken as the test statistic.

4. ONE– AND TWO–SIDED, SYMMETRIC AND ASYMMETRIC TEST STATISTICS

In this section we generalize the test statistic to deal with positive as well as negative shifts, two-sided alternatives, and the case that both samples are heterogeneous.

For the L_2 distance and weight w_1 the test statistic resulting from minimization problem (2) is

$$\Theta(\mathbb{F}_m, \mathbb{H}_n) = 1 - \min\left\{1, \frac{\sum_{x\in\mathcal{S}(d\mathbb{F}_m)} \mathbb{F}_m(x)\mathbb{H}_n(x)}{\sum_{x\in\mathcal{S}(d\mathbb{F}_m)} \mathbb{F}_m(x)\mathbb{F}_m(x)}\right\}, \quad (4)$$

where $\mathcal{S}(d\mathbb{F}_m)$ denotes the support of the empirical density $d\mathbb{F}_m$.

This is the one-sided statistic designed for detecting positive partial shifts $[G_j(x) \le F_j(x)]$ in model (1). Negative shifts can be detected by replacing $\mathbb{F}_m$ and $\mathbb{H}_n$ by $1 - \mathbb{F}_m$ and $1 - \mathbb{H}_n$, respectively. The corresponding two-sided statistic is

$$\Theta_2(\mathbb{F}_m, \mathbb{H}_n) = \max\{\Theta(\mathbb{F}_m, \mathbb{H}_n), \Theta(1 - \mathbb{F}_m, 1 - \mathbb{H}_n)\}. \quad (5)$$

This statistic is designed to detect both under- and overexpression.

So far we have concentrated on the asymmetric case, with one supposedly homogeneous "normal" group and one possibly heterogeneous "cancer" group. In the symmetric case, for example, with two distinct cancer groups, one wishes to detect partial expression in either of the groups. This can be accommodated by swapping the roles of f and h in model (1) and also in the statistic (4). This leads to the statistic:

$$\Theta_2^s(\mathbb{F}_m, \mathbb{H}_n) = \max\{\Theta_2(\mathbb{F}_m, \mathbb{H}_n), \Theta_2(\mathbb{H}_n, \mathbb{F}_m)\}. \quad (6)$$

5. A TEST FOR PARTIAL DIFFERENTIAL EXPRESSION

The test statistic Θ_j is used to decide on the partial differential expression of gene j. We implement the test as a permutation test, determining the null distribution of Θ_j (conditional on the data) by permutation resampling. To this end the columns of the gene expression matrix are permuted, thus preserving the dependency structure between the genes. For each permutation the test statistic Θ_j is calculated, resulting in $\Theta_j^1, \ldots, \Theta_j^P$, where P is the number of permutations. The (unadjusted) p value corresponding to gene j is given by

$$\{\#\ell | \Theta_j < \Theta_j^\ell \text{ for } \ell = 1, \ldots, P\}/P,$$

the proportion of the permutation null distribution that exceeds the observed value of the test statistic. The resulting test, which we will call the PDE test, is asymptotically distribution free and consistent, as is shown in Appendices A and B, respectively.

6. EFFICIENT p VALUE CALCULATION

A multiplicity correction is necessary to take account of the fact that we perform the test on all genes. Proper adjustment of the p values requires that the null distributions, which may be different for different genes, are sufficiently precise, particularly in the tails. For our permutation procedure this means that the number of permutations needs to be huge, so that sufficiently small p values can be reached (say, .001 or smaller).

Table 1. Noise distributions used in the simulation

	Noise distribution
1	N(0, 1)
2	Double exponential
3	t_5
4	$(1 - \tau)$ N(0, 1) + τ N(0, 3)
5	$(1 - \tau)$ N(0, 1) + τ N(0, 10)
8	$(1 - \tau)$ double exponential + τ N(0, 10)
7	$(1 - \tau)$ t_5 + τ N(0, 10)
8	Cauchy

NOTE: The mixing proportion is set at $\tau = 1/40$.

This has serious practical consequences for the computing time, which runs into hours because of the large numbers of genes (∼10,000) and permutations.

To reduce the computing time, we found the following procedure to be useful. First, we assume that genes with a marginal unadjusted p value larger than .01 will not be called significant after the multiple testing correction, although it is not determined a priori what effect the BH (Benjamini–Hochberg) multiple testing correction will have on the marginal p values. Given that the majority of the genes are probably not differentially expressed, we observe after a relatively small number of permutations (say 100) that most estimated marginal p values will highly exceed .01. The idea is to exclude those noninteresting genes from further calculations.

To make this more rigorous, we decide upon a cutoff for the estimated marginal p value after P permutations, denoted $\hat{p}_{j,P}$ for gene j. By the formula for proportions in binomial models, the approximate left 99.9% confidence bound can be written as

$$L_{j,P} = \hat{p}_{j,P} - z_{.001} \times \sqrt{\frac{\hat{p}_{j,P}(1 - \hat{p}_{j,P})}{P}},$$

where $z_{.001} = 3.09$ is the .001 upper quantile of the standard normal distribution. As it is very unlikely that the true p value will be smaller than .01 when $L_{j,P} > .01$, we exclude such genes from further permutations.

The procedure may be repeated for various values of P, which may lower the confidence level (of the entire procedure) somewhat. We found that, depending on the number of truly significant genes, use of this bounding rule speeds up computations between 5 and 50 times. Note that if the assumption were violated, one would notice this after the BH correction: If the largest adjusted p value is still smaller than the specified FDR cutoff (say .05), it is likely that more genes need to be included. In that (unlikely) case, one may simply increase the marginal unadjusted p value cutoff to, for example, .05, and repeat the procedure.

7. SIMULATION

In this section we compare the tests proposed in Section 5 to the Cramér–von Mises and the permutation t tests, two other commonly used methods for the detection of differentially expressed genes. All methods are applied to artificial datasets, which cover a wide range of situations: various degrees of differential expression (datasets 1 and 2 in Table 2), dependence between genes with a resultant correlation coefficient of

Table 2. Description of the datasets

Dataset	Noise distribution	Differentially expressed genes	Number of signal samples	Signal size	Dataset	Noise distribution	Differentially expressed genes	Number of signal samples	Signal size
1	1	1–50	20	3	17	5	1–50	20	3
2	1	1–50	20	.06–3	18	3	1–50	20	3
3*	1	1–50	20	3	19	3	1–50	10	.06–3
4*	1	1–50	20	.06–3	20	7	1–50	20	3
5	4	1–50	20	3			51–100	10	.12–6
6	5	1–50	20	3	21	2	1–50	10	3
7	4	1–50	20	.06–3	22	8	1–5	±20	3
8	5	1–50	20	.06–3			6–10	±19	3
9	2	1–50	20	3			11–15	±18	3
10	8	1–50	20	3			⋯–⋯		
11	1	1–50	10	3			96–100	±1	3
12	1	1–50	10	.12–6	23	4	1–50	20	3
13	1	1–50	20	3			51–100	10	3
		51–100	10	3	24	5	1–5	±20	3
14	1	1–50	20	.06–3			6–10	±19	3
		51–100	10	.12–6			11–15	±18	3
15	1	1–5	±20	3			⋯–⋯		
		6–10	±19	3			96–100	±1	3
		11–15	±18	3			101–150	20	3
		⋯–⋯			25	6	1–5	±20	3
		96–100	±1	3			6–10	±19	3
16	1	1–5	±20	3			11–15	±20	3
		6–10	±19	3			⋯–⋯		
		11–15	±18	3			96–100	±1	3
		⋯–⋯		3			101–150	20	3
		96–100	±1	3					

NOTE: Each dataset consists of 20 normal and 20 cancerous tissue samples and 1,000 genes. The coding for the noise distribution is given in Table 1. Datasets 3 and 4 have a random chip effect added: A vector of 40 N(0, .25) random variables is added to every gene. The next column specifies which genes are differentially expressed. In the "Number of signal samples" column the number of samples with differential expression is given. In the "Signal size" column the size of the signal is given. If a signal range is specified the signal linearly increases with the number of the differentially expressed gene.

.1 (datasets 3 and 4 in Table 2), presence of outliers (datasets 5–8 in Table 2), heavy-tailed distribution for expression (datasets 9 and 10 in Table 2), partial differential expression (datasets 11 and 12 in Table 2), and combinations of the aforementioned (datasets 13–25 in Table 2).

The design of the simulation study is as follows. Expression data (in accordance with the descriptions in Table 2) are simulated. Each dataset consists of 40 samples (20 for each group) and 1,000 genes. The group labels are permuted 5,000 times (cf. Klebanov, Gordon, Xiao, Land, and Yakovlev 2006, for a motivation of the number of permutations). All proposed tests (all distances with all combinations of weights), the two-sample Cramér-von Mises test (as given in Xiao, Gordon, and Yakovlev 2006), and the t test are applied to each gene in all 5,000 permutations. The 5,000 permutation values of the test statistic form the marginal null distribution for the gene. The observed statistic for each gene is compared to its marginal null distribution, and the corresponding p value is estimated by the proportion of the null distribution that exceeds the test statistic.

For ease of comparison the rejection level α for the BH-adjusted p values is set such that the FDR is controlled at $1/6$. Here the FDR is defined as the number of rejected true null hypotheses divided by the number of rejected null hypotheses. A gene is considered "truly" differentially expressed if it has a nonzero mean difference between the two groups; for example, in dataset 9 there are 50 "truly" expressed genes. This approach facilitates the comparison of the power of each test: One only has to compare the false negative rate (FNR). For each aforementioned dataset this is repeated 10 times. The average α, FDR, and FNR are calculated for each test statistic and dataset.

We first summarize the simulation results of the proposed PDE tests (given in the supplementary material). The test based on the L_2 distance in general outperforms (when combined with the same weight) tests based on the L_1 and Hellinger distances. The performances of the latter are comparable, except that the Hellinger distance is preferred for heavy-tailed distributions. Tests with any distance and weight w_1 outperform tests with the same distance but another weight. Weight w_4 is the worst choice of all weights (in particular, when combined with L_1). This is probably due to the fact that it is unstable if $H(x)$ is close to 1. Finally, tests with weight w_2 perform better than tests with weight w_3, especially in the presence of heavy tails. In short, the test with the L_2 distance and weight w_1 performs best. This is convenient, because of the simplicity of w_1, the analytical form of the test statistic (Sec. 4), and the tractability of the asymptotic distribution (App. A).

Next, we compare simulation results of the proposed PDE test (using L_2 and w_1) with the Cramér–von Mises and two-sample t tests. Results are displayed in Table 3. In datasets 1–

Table 3. Simulation results

		PDE				CvM with L_2				t			
Dataset	# DEG	α	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	s.e. $\widehat{\text{FNR}}$	α	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	s.e. $\widehat{\text{FNR}}$	α	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	s.e. $\widehat{\text{FNR}}$
1	50	.17	.15	.00	.000	.19	.15	.00	.000	.19	.15	.00	.000
2	50	.22	.15	.29	.042	.23	.15	.29	.044	.20	.15	.30	.043
3	50	.20	.15	.00	.000	.19	.15	.00	.000	.19	.15	.00	.000
4	50	.25	.15	.30	.049	.24	.15	.31	.059	.21	.15	.28	.050
5	50	.19	.15	.00	.000	.16	.15	.00	.000	.15	.15	.00	.000
6	50	.20	.15	.00	.000	.18	.15	.00	.000	.20	.15	.06	.030
7	50	.27	.15	.30	.059	.26	.15	.29	.053	.22	.15	.31	.031
8	50	.22	.15	.29	.033	.24	.14	.29	.027	.27	.15	.42	.065
9	50	.18	.15	.00	.063	.15	.15	.00	.061	.17	.15	.00	.040
10	50	.19	.15	.13	.044	.21	.15	.06	.032	.40	.15	.73	.056
11	50	.20	.15	.23	.084	.27	.14	.45	.098	.21	.15	.27	.092
12	50	.27	.15	.43	.067	.30	.14	.54	.082	.27	.15	.39	.041
13	100	.21	.16	.05	.019	.20	.16	.12	.036	.20	.16	.06	.032
14	100	.23	.16	.29	.024	.29	.16	.31	.033	.23	.16	.28	.037
15	100	.27	.16	.35	.026	.26	.16	.40	.023	.24	.16	.34	.034
16	150	.25	.16	.21	.016	.26	.16	.24	.020	.21	.16	.21	.027
17	100	.21	.15	.13	.033	.24	.16	.16	.031	.22	.16	.22	.038
18	100	.22	.14	.12	.064	.22	.15	.17	.050	.23	.14	.13	.054
19	150	.23	.16	.25	.029	.23	.16	.26	.025	.26	.16	.31	.025
20	50	.22	.16	.32	.038	.24	.16	.46	.023	.21	.16	.36	.018
21	50	.27	.15	.46	.105	.31	.14	.61	.118	.28	.15	.49	.117
22	100	.33	.16	.59	.032	.33	.16	.58	.035	.61	.14	.85	.043
23	100	.20	.16	.07	.029	.20	.16	.14	.040	.23	.16	.09	.033
24	150	.22	.16	.21	.015	.22	.16	.24	.018	.25	.16	.28	.038
25	150	.23	.16	.26	.019	.25	.16	.27	.016	.25	.16	.33	.027

NOTE: The "Dataset" column is the number of the corresponding dataset, the "# DEG" column is the number of "truly" differentially expressed genes, the "α" column is the actual rejection criterion used to control the FDR, the "$\widehat{\text{FDR}}$" column is the estimated FDR corresponding to rejection criterion α, the "$\widehat{\text{FNR}}$" column is the estimated false negative rate corresponding to rejection criterion α, and the "s.e. $\widehat{\text{FNR}}$" column is the standard error of the estimated false negative rate.

4 with differential expression (with and without dependence) there is hardly any difference between the three methods. The datasets containing outliers (datasets 5–8) show that the PDE test is rather robust, as opposed to the t test, which performs somewhat worse than the other two. The t test also suffers most from the presence of heavy tails (datasets 9 and 10), whereas the PDE test is slightly outperformed by the Cramér–von Mises test. The proposed PDE test performs best for partial differential expression (datasets 11 and 12). Moreover, this test performs at least equally well in the remaining datasets (datasets 13–25), which are combinations of special cases.

For datasets 11, 12, 19, and 22, all containing some form of partial differential expression, we have also plotted the unadjusted p values of the differentially expressed genes (i.e., those with nonzero mean), against their order (Fig. 2). The performance of the Cramér–von Mises test falls behind in datasets 11 and 12; the t test performs less well in datasets 19 and 22. The PDE test performs best in all four cases, on a par with the Cramér–von Mises test in datasets 19 and 22 and with the t test in datasets 11 and 12.

We conclude that the proposed test with the L_2 distance and the empirical density weight w_1 performs particularly well in the situations that it was designed for: partial differential expression. Moreover, it is equally (if not more) powerful than traditional methods like the Cramér–von Mises and t tests for many other standard situations.

8. REAL–LIFE DATASETS

8.1 Ovarian Data

The ovarian cancer data (Schummer et al. 1999) consist of 53 tissue samples, of which 23 are normal and 30 are ovarian cancer tissues. The expression of 1,536 genes has been measured in each tissue using microarrays. Schummer et al. (1999) investigated differential expression between normal and cancerous ovarian tissue. Expression in the normal ovarian tissue serves as a reference. Because cancer of a particular tissue type is often a collection of different diseases, each with its own genetic mechanism, it is likely that the group of ovarian cancers is heterogeneous. This legitimates our interest in partial differential expression in the ovarian cancers: A part of the group may be shifted with respect to the normal while the rest remains unchanged. Although the group of normal ovarian tissues may be heterogeneous as well, this heterogeneity sheds no light on the genetic mechanism under study (that of ovarian cancer). Rather, one could even consider this heterogeneity as unwanted sampling variation. Therefore, we are not interested in partial differential expression in the normal group and use the two-sided asymmetric version of our statistic, that is, formula (5).

The number of permutations is set at 5,000 Table 4 gives the number of differentially expressed genes found by each method at several rejection levels. We observe that the PDE test detects more genes than its classical counterparts, in particular the t

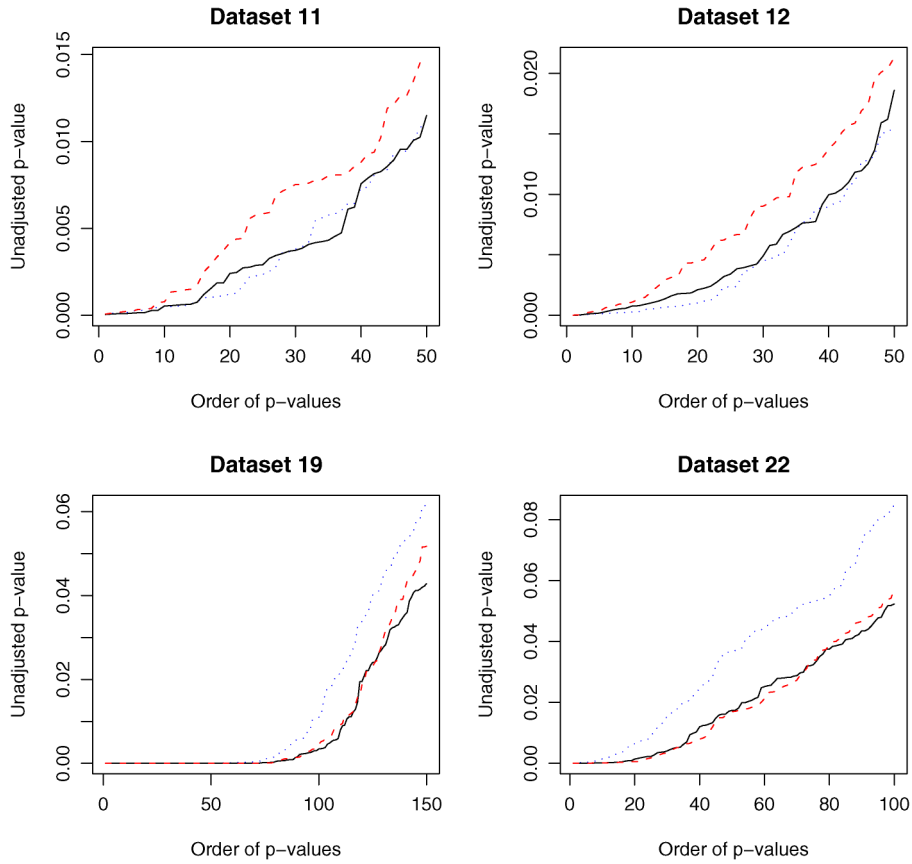

Figure 2. Ordered raw p values of the differentially expressed genes for the PDE test with weights w_1 (——), the L_2 Cramér–von Mises test (- - -), and the t test (· · ·), for four datasets with partial differential expression.

test. Those genes for which the PDE test conflicts with either the Cramér–von Mises test or the t test are of special interest. Here we focus on a comparison with the Cramér–von Mises test at FDR $\leq .01$. A cross-tabulation of differentially and nondifferentially expressed genes according to the PDE and Cramér–von Mises tests is given in Table 5.

We produced violin plots for the $28 + 21 = 49$ "conflicting genes" (see the supplementary material). A specific feature of those plots for genes detected by the PDE test and not detected by the Cramér–von Mises test is the concentration of expression values of the normal class within either the left or right half of the expression values of the cancer class. In quite a number of cases bimodality in the cancer class is clearly observed. Such genes may further distinguish groups of individuals within the class of ovarian cancer tissues. For genes detected by the Cramér–von Mises test and not detected by the PDE test, dispersion within each of the two classes is often comparable, and the mean shift between the two classes is quite small. Moreover, the Cramér–von Mises test detects some genes displaying (possibly spurious) heterogeneity in the normal group when compared to the cancer group. Such genes may be of minor interest to biologists. A similar story holds for the comparison of the "conflicting genes" of the PDE and t tests.

We have studied the well-known oncogene ErbB2, known to be upregulated in 9–30% of the ovarian carcinomas (Pejovic 1995) and the new, promising biomarker for ovarian cancers mesothelin (e.g., Bast 2003; Hassan et al. 2006; Hellstrom et al. 2006) in more detail, investigating which test is most powerful in detecting the (partial) differential expression of these genes. To this end smaller versions of the original dataset were created by resampling and reanalyzed. We chose a set of percentages (40%, 45%, . . . , 95%, and 100%) and randomly selected (without replacement) for each percentage a corresponding number of samples from the normal and cancer groups, leaving the ratio between the group sample sizes unchanged. For the newly created dataset we calculated the marginal p value for the three

Table 4. Number of differentially expressed genes in the ovarian cancer dataset for the Θ, Cramér–von Mises, and t tests

Method	FDR	# DEG	FDR	# DEG
Θ	.05	673	.01	501
CvM.L2	.05	667	.01	494
t	.05	638	.01	480

Table 5. Number of differentially and nondifferentially expressed genes in the ovarian cancer dataset for the Θ and Cramér–von Mises tests, with FDR controlled at .01

	CvM.L2	
Θ	# NDEG	# DEG
# NDEG	1,014	21
# DEG	28	473

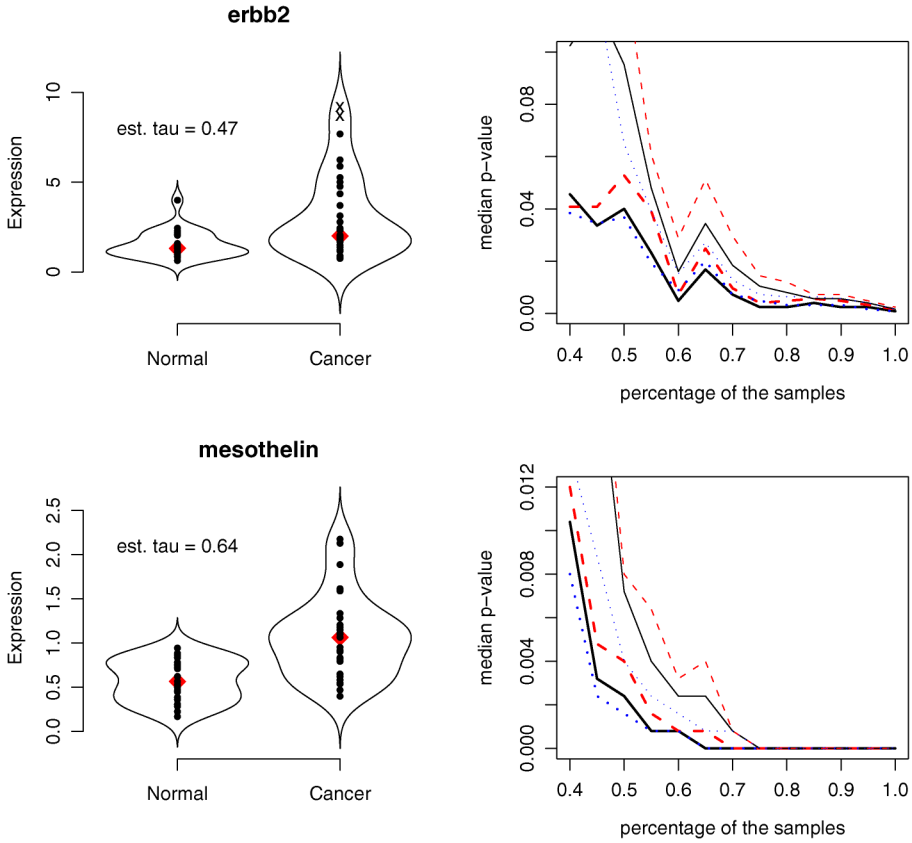

Figure 3. Violin plots of the well-known oncogene ErbB2 and the new, promising ovarian cancer biomarker mesothelin, with their median (and 75% quantile: thinner line) p values for reduced sample sizes. The expression data of ErbB2 contain two outliers in the cancer group. For the violin plot they were shrunk and plotted as "X." (——, PDE; – –, Cramér–von Mises; · · ·, t.)

tests of the selected genes using 5,000 permutations. We repeated this procedure 25 times. The resulting median marginal p values are plotted against the percentage (next to their violin plots) in Figure 3. The PDE and t tests perform equally well, with the Cramér–von Mises test a little behind in power.

8.2 Prostate Data

The prostate cancer data (LaPointe et al. 2004) consist of 103 tissue samples, 41 of which are normal and 62 of which are cancerous. The expression of 5,153 genes was measured for each tissue.

We summarize the results of our analysis. The PDE test detects 2,775 and 2,690 genes at FDR $\leq .05$ and FDR $\leq .01$, respectively. Again, this is more than the Cramér–von Mises test (2,659 and 2,557) and the test t (2,578 and 2,478). Genes detected by the PDE test and not by one of the other two tests often exhibit the same characteristics as the conflicting genes in the ovarian cancer dataset (see the supplementary material).

We use the prostate cancer dataset of Dhanasekaran et al. (2001) as a validation dataset for genes found by the PDE test and not by the Cramér–von Mises test and the t test in the LaPointe et al. (2004) prostate cancer dataset. Violin plots of two genes, IGF-1R and RPN2, present on both arrays, are plotted in Figure 4. In the LaPointe et al. (2004) data set a small group of the cancer samples exhibits up-regulated expression for IGF-1R, which is confirmed in the cancer samples of the Dhanasekaran et al. (2001) dataset. Overexpression of the IGF-1R gene is associated with tumor growth in prostate cancer (Hellawell et al. 2002), and silencing IGF-1R can make prostate cancer cells more sensitive to radiotherapy and certain kinds of chemotherapy (Rochester, Riedemann, Hellawell, Brewster, and Macaulay 2005). For the other gene, RPN2, there is some form of bimodality present in the cancer group of the LaPointe et al. (2004) dataset, a phenomenon that is more pronounced in the Dhanasekaran et al. (2001) dataset. RPN2 has already been found to be differentially expressed in prostate cancer (Covell, Wallqvist, Rabow, and Thanki 2003). RPN2 has also been found to be overexpressed in other cancers (e.g., colorectal cancer; Hufton et al. 1999) and reported to be involved in apoptosis, programmed cell death (Sun et al. 2004). This makes a promising candidate oncogene for prostate cancer found by the PDE test.

One of the objectives of LaPointe et al. (2004) was to find biologically and clinically relevant gene expression tumor subtypes. Three subtypes are found by means of hierarchical clustering and principal component analysis. Having chosen the subtypes, the expression heat map is used to visually select genes associated with the three subtypes, which may function as "surrogate markers" for the prostate cancer subtypes.

We investigated which test is most powerful in detecting the claimed partial differential expression in these genes. In the "regular" analysis, comparing the normal and cancer group and not exploiting the subtype information, most of these genes were found by all three tests at FDR $= .01$, with the PDE test finding more of the genes reported by LaPointe et al. (2004)

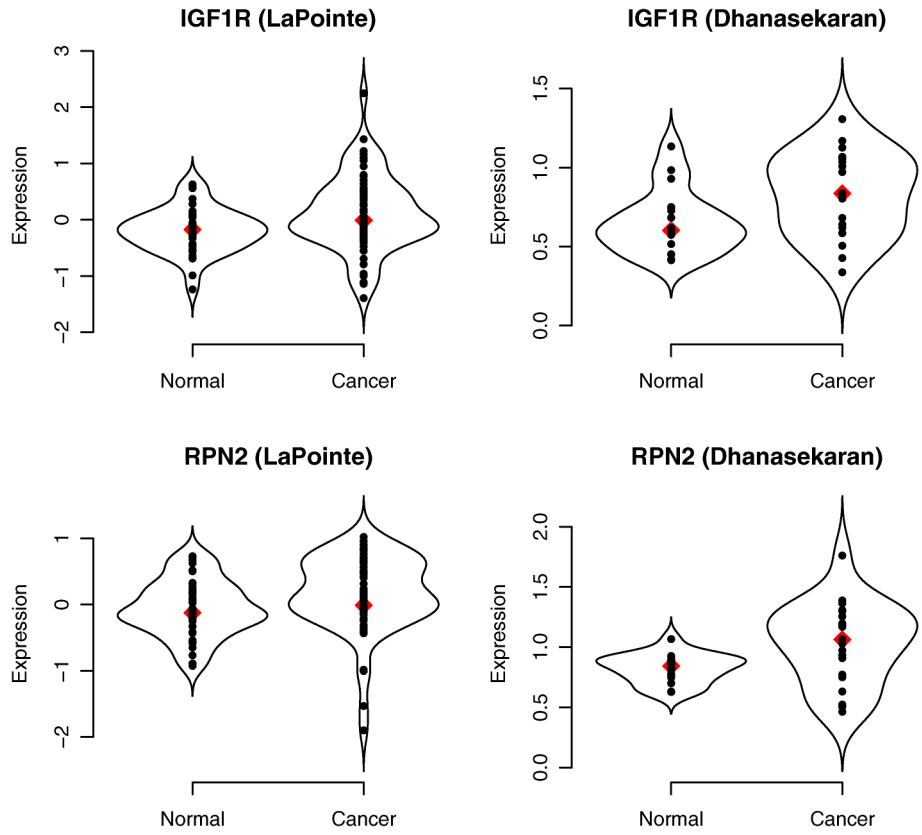

Figure 4. Violin plots of the IGF-1R and RPN2 genes in the datasets of LaPointe et al. (2004) and Dhanasekaran et al. (2001).

than the Cramér–von Mises and t tests. Heterogeneity within the cancer group is thus detected without knowledge of the subtypes. Notably, some genes were not found by any of the three tests. This is due to the fact that the genes (e.g., MUC1) have been selected primarily for their discriminative power between the tumor subtypes and not for their discriminative power with the normal samples, whose expression distribution is sometimes rather diffuse, making it difficult to detect the partial differential expression.

In addition, we studied the RPL13 and NRP1 genes, associated with tumor subtype III and subtypes II and III (as defined by LaPointe et al. 2004), respectively, in a similar fashion as ErbB2 and mesothelin for the ovarian cancer dataset. The results are given in Figure 5. The PDE test performs best for both genes, outperforming both the t test and the Cramér–von Mises test.

We also assessed the association of the sets of (partial) differentially expressed genes found by each method and the clinically relevant subgroups as reported by LaPointe et al. (2004) by means of the global test of Goeman, Van de Geer, De Kort, and Van Houwelingen (2004). The set of genes found by the PDE test has a stronger association with these subgroups than the set of genes found by the other two methods. This remains true if we limit ourselves to the "conflicting" genes. This is visually confirmed by the heat maps of the PDE and Cramér–von Mises "conflicting" genes as given in the supplementary material. They show that the genes found by the PDE test and not by the Cramér–von Mises test relate very well with the LaPointe et al. (2004) subgroups, even though LaPointe et al. (2004) used all genes to construct these subgroups.

More details of the analyses can be found in the supplementary material.

9. CONCLUSION AND DISCUSSION

We developed a test that is specifically designed to detect partial shifts of one population versus another. The test is applicable to a wide range of data, because it is based on a nonparametric mixture model, thereby avoiding parametric assumptions. We show that the statistic is asymptotically distribution free and the corresponding permutation procedure consistent. We supply efficient procedures for the calculation of the marginal p values.

We discuss a possible extension of our test, which incorporates ideas to use the multiplicity of genes more effectively in testing procedures. Several authors have proposed using shrinkage to borrow information across genes. This may be beneficial, especially when group sizes are small (Allison, Cui, Page, and Sabripour 2006). In general, we do not recommend using the PDE test for very small group sizes (say ≤ 7). The test still has power comparable to the permutation t test and Cramér–von Mises test for whole-group shifts, but its ability to reliably detect heterogeneity in the cancer group is lost. Shrinkage may be useful but cannot be applied to the variance parameters, as is commonly done (e.g., Tusher, Tibshirani, and Chu 2001), because this would require unwanted parameterization of F_j and H_j. In the case with few normal tissues and relatively many cancerous tissues (which is not uncommon in practice), the following approach deserves further study. We may assume without loss of generality that F_j is located at 0, because the Θ statistic is invariant to common shifts in F_j and H_j. Then, shifting

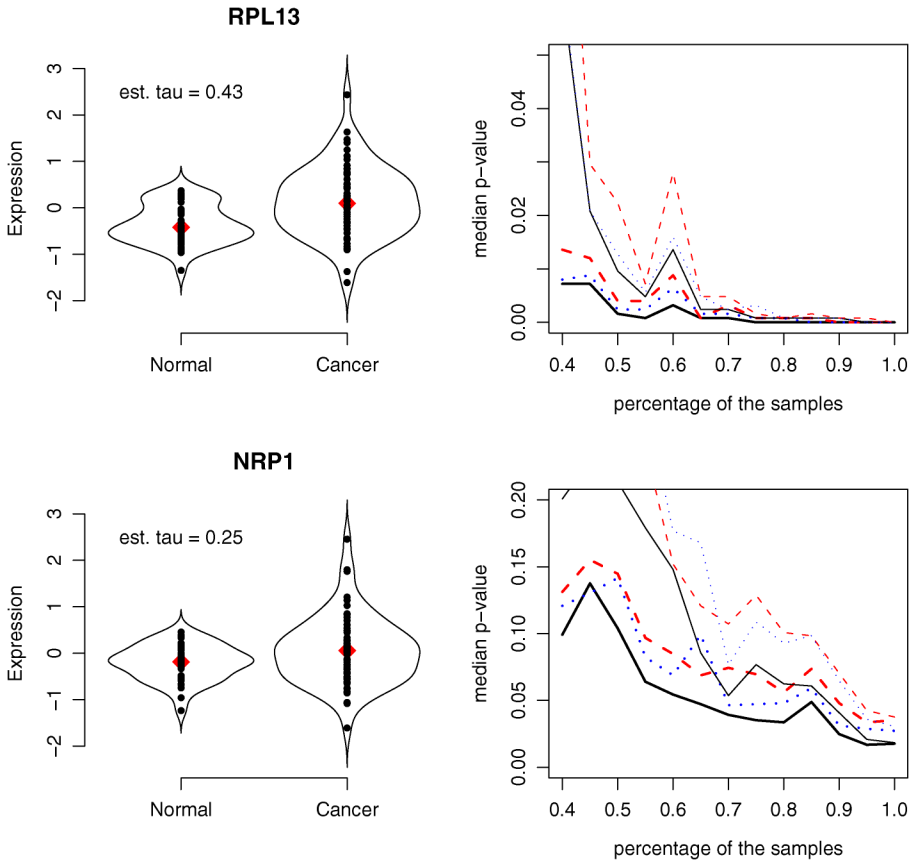

Figure 5. Violin plots of two "surrogate markers" for subgroups found by LaPointe et al. (2004), with their median (and 75% quantile: ——) p values for reduced sample sizes. (——, PDE; – –, Cramér–von Mises; · · ·, t.)

all values for each gene such that the median equals 0, we use instead of the edf $\mathbb{F}_j$ in (2) the estimate:

$$\tilde{F}_j = \lambda \mathbb{F}_j + (1-\lambda)\hat{F},$$

where $\hat{F}$ is the overall empirical distribution function (also with location 0), constructed from all genes. Among other techniques, cross-validation could be used to find an optimal λ.

To conclude, this new asymptotically distribution-free test is shown to share robustness properties with nonparametric tests such as the Cramér–von Mises two-sample test while remaining competitively powerful with respect to the two-sample permutation t test for simple whole-sample shifts. Moreover, it deviates from such conventional tests in two ways. First, it has specific power for detecting partial shifts, an alternative that is very biologically relevant in gene expression studies. Second, the asymmetric version of the test is able to deal with asymmetric interest in two populations (e.g., normal and cancerous tissues). Therefore, we believe it to be a strong competitor for conventional tests in gene expression studies.

APPENDIX A: ASYMPTOTIC DISTRIBUTION

Here the asymptotic distribution of the Θ statistic under the null hypothesis $H_0 : \Theta = 0$ is derived. The asymptotic distribution is not to be used for the calculation of the p value in the test for partial differential expression, as the approximation of the tail probabilities by the asymptotic distribution is often crude. In the derivation of the asymptotic distribution, we use the theory developed in van der Vaart (1998). We limit ourselves to the L_2 distance with weights $w_1(x)$, the empirical density.

Let $X_1, \ldots, X_m$ be iid random variables with distribution function F and empirical distribution function $\mathbb{F}_m$. Similarly, $Y_1, \ldots, Y_n$ are iid random variables with $H_\tau = (1-\tau)F + \tau G$ and $\mathbb{H}_n$ the (empirical) distribution function. We assume that the $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are independent and write the test statistic Θ as

$$\Theta = 1 - \frac{\int \mathbb{F}_m(x)\mathbb{H}_n(x)\, d\mathbb{F}_m}{\int \mathbb{F}_m^2(x)\, d\mathbb{F}_m} = 1 - \psi_0(\mathbb{F}_m, \mathbb{H}_n). \tag{A.1}$$

We study the asymptotic behavior by means of the delta method. To this end, we need the derivative of ψ_0. We shall consider the following more general functional, which allows for a general weight function:

$$\psi(F, H, W) = \frac{\int FH\, dW}{\int F^2\, dW}.$$

Theorem 3.9.17 of van der Vaart and Wellner (1996) implies that ψ is Hadamard differentiable on the subset of $D[-\infty, \infty]^3$ consisting of (F, H, W) with $\int |dW| \le 2$ and $\int F^2\, dW > 0$.

The derivative of ψ is, in view of the chain rule (van der Vaart 1998, thm. 20.9), and the derivative of the Wilcoxon functional (van der Vaart 1998, lemma 20.10), then given by

$$\begin{aligned}\psi'_{F,H,W}(f,h,w) &= \frac{1}{(\int F^2\, dW)^2}\left(\int F^2\, dW\left(\int (fH + Fh)\, dW + \int FH\, dw\right)\right.\\ &\qquad \left. - \int FH\, dW\left(\int 2fF\, dW + \int F^2\, dw\right)\right).\end{aligned}$$

Here $\int FH\,dW$ and $\int F^2\,dW$ are defined by the partial integration formula. Note that, if $F = H$, we have

$$\psi'_{F,F,W}(f,g,h) = \frac{\int (h-f)F\,dW}{\int F^2\,dW}, \tag{A.2}$$

which is independent of w, the perturbation of W. It follows that the type of estimator of W has no influence on the asymptotic distribution of the test statistic, as long as it converges sufficiently fast to a limiting function W.

Because both $\sqrt{m}(\mathbb{F}_m - F)$ and $\sqrt{n}(\mathbb{H}_n - F)$ converge in distribution in $D[-\infty,\infty]$, by Donsker's theorem (van der Vaart 1998, thm. 19.3), for any estimators $\mathbb{W}_N$ such that $\sqrt{N}(\mathbb{W}_N - W)$ converges in distribution in the same space, the delta method (van der Vaart 1998, thm. 20.8) then gives, under the null hypothesis,

$$\begin{aligned}\sqrt{N}\Theta &= \sqrt{N}\big(\psi(\mathbb{F}_m,\mathbb{H}_n,\mathbb{W}_N) - \psi(F,F,W)\big)\\ &\approx \sqrt{N}\frac{\int(\mathbb{H}_n - F) - (\mathbb{F}_m - F)F\,dW}{\int F^2\,dW}\\ &= \frac{1}{\int F^2(u)\,dW(u)}\int\left(\frac{1}{\sqrt{1-\lambda}}\mathbb{G}_n^Y - \frac{1}{\sqrt{\lambda}}\mathbb{G}_m^X\right)F\,dW, \end{aligned}\tag{A.3}$$

where $m/N \to \lambda$, $n/N \to 1-\lambda$, $\lambda \in (0,1)$, $\mathbb{G}_m^X = \sqrt{n}(\mathbb{F}_m^X - F)$, and $\mathbb{G}_n^Y = \sqrt{n}(\mathbb{H}_n - F)$. As $\mathbb{G}_n^X$ and $\mathbb{G}_m^Y$ are independent, this converges in distribution to

$$\frac{1}{\int F^2\,dW}\int\frac{1}{\sqrt{\lambda(1-\lambda)}}\mathbb{G}_F F\,dW, \tag{A.4}$$

where $\mathbb{G}_F$ is a Brownian bridge corresponding to F. Finally, for $F = H = W$ a continuous distribution, the variable (A.4) is equal in distribution to

$$\mathrm{N}\left(0, \frac{\int\int (F(x\wedge y) - F(x)F(y))F(x)F(y)\,dF(x)\,dF(y)}{\lambda(1-\lambda)(\int F^2(x)\,dF(x))^2}\right) \tag{A.5}$$

if F is continuous. As this equals $\mathrm{N}(0, 1/(5\lambda(1-\lambda)))$, the asymptotic distribution of the test statistic with Euclidean distance and weights equal to the empirical density is (asymptotically) distribution free.

APPENDIX B: CONSISTENCY

Here we show the consistency of the proposed test that employs the marginal permutational null distribution. First, the asymptotic distribution of the permutated samples is derived. Then we show that asymptotically the marginal permutational null distribution equals the null distribution of Θ.

Let $Z_1,\ldots,Z_{m+n}$ be the pooled sample from $X_1,\ldots,X_m, Y_1,\ldots, Y_n$ with empirical distribution $\mathbb{K}_{m+n} = \frac{m}{m+n}\mathbb{F}_m + \frac{n}{m+n}\mathbb{H}_n$, and $R_1,\ldots,R_{m+n}$ a random permutation of $1,\ldots,m+n$. Define

$$\mathbb{F}_m^P = \frac{1}{m}\sum_{i=1}^{m} 1_{\{Z_{R_i}\le x\}} \qquad \text{and} \qquad \mathbb{H}_n^P = \frac{1}{n}\sum_{i=m+1}^{m+n} 1_{\{Z_{R_i}\le x\}}.$$

From theorem 3.7.1 of van der Vaart and Wellner (1996), it follows that, given $X_1,\ldots,X_m, Y_1,\ldots,Y_n$, $\sqrt{m}(\mathbb{F}_m^P - \mathbb{K}_{m+n})$ and $\sqrt{n}(\mathbb{H}_n^P - \mathbb{K}_{m+n})$ converge in distribution to $\sqrt{1-\lambda}\mathbb{G}_K$ and $\sqrt{\lambda}\mathbb{G}_K$, respectively. Here $\lambda = \lim_{m,n\to\infty}\frac{m}{m+n}$, and $\mathbb{G}_K$ is a tight Brownian bridge process corresponding to the measure $K = \lambda F + (1-\lambda)H_\tau$, which is F under the null hypothesis.

To show that the asymptotic distribution of the estimate using the permutations equals that of Θ, we define

$$\Theta^P = 1 - \frac{\int \mathbb{F}_m^P(x)\mathbb{H}_n^P(x)\,d\mathbb{F}_m^P}{\int(\mathbb{F}_m^P)^2(x)\,d\mathbb{F}_m^P}.$$

Under the null hypothesis (at $\Theta = 0$ when $F = H = W$) we have, in analogy to (A.3),

$$\begin{aligned}&\sqrt{N}(\Theta^P - \Theta)\\ &\quad\approx \psi'_{F,F,F}\big(\sqrt{N}(\mathbb{F}_m^P - \mathbb{F}_m), \sqrt{N}(\mathbb{H}_n^P - \mathbb{H}_n), \sqrt{N}(\mathbb{F}_m^P - \mathbb{F}_m)\big)\\ &\quad= \frac{\int\sqrt{N}(\mathbb{H}_n^P - \mathbb{H}_n) - \sqrt{N}(\mathbb{F}_m^P - \mathbb{F}_m)F\,dF}{\int F^2\,dF}\\ &\quad= \frac{\int\sqrt{N}(\mathbb{H}_n^P - \mathbb{F}_m^P)F\,dF}{\int F^2\,dF} - \Theta.\end{aligned}$$

As $\frac{n}{m+n}(\mathbb{F}_m^P - \mathbb{H}_n^P) = \mathbb{F}_m^P - \mathbb{K}_{m+n}$ and the latter process converges conditionally in distribution as noted earlier in this section, it follows that $\sqrt{N}\Theta^P$ converges conditionally in distribution to

$$\frac{1}{\int F^2\,dF}\int\frac{\mathbb{G}_F F}{\sqrt{\lambda(1-\lambda)}}\,dF.$$

Hence, $\sqrt{n}\Theta^P$ converges in distribution to the asymptotic distribution of Θ as given in (A.5).

[Received February 2006. Revised August 2007.]

REFERENCES

Allison, D. B., Cui, X., Page, G. P., and Sabripour, M. (2006), "Microarray Data Analysis: From Disarray to Consolidation and Consensus," *Nature Reviews Genetics*, 7, 55–65.

Bast, R. C. (2003), "Status of Tumor Markers in Ovarian Cancer Screening," *Journal of Clinical Oncology*, 21, 200s–205s.

Bowman, A. W., and Azzalini, A. (1997), *Applied Smoothing Techniques for Data Analysis: The Kernel Approach With S–Plus Illustrations*, Oxford, U.K.: Oxford University Press.

Broët, P., Richardson, S., and Radvanyi, F. (2002), "Bayesian Hierarchical Model for Identifying Changes in Gene Expression From Microarray Experiments," *Journal of Computational Biology*, 9, 671–683.

Brown, P. O., and Botstein, D. (1999), "Exploring the New World of the Genome With DNA Microarrays," *Nature Genetics Supplement*, 21, 33–37.

Covell, D. G., Wallqvist, A., Rabow, A. A., and Thanki, N. (2003), "Molecular Classification of Cancer: Unsupervised Self-Organizing Map Analysis of Gene Expression Microarray Data," *Molecular Cancer Therapeutics*, 2, 317–332.

Cutler, A., and Cordero-Braña, O. I. (1996), "Minimum Hellinger Distance Estimation for Finite Mixture Models," *Journal of the American Statistical Association*, 91, 1716–1723.

Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. (2001), "Delineation of Prognostic Biomarkers in Prostate Cancer," *Nature*, 412, 822–826.

Dudoit, S., Shaffer, J. P., and Boldrick, J. C. (2003), "Multiple Hypothesis Testing in Microarray Experiments," *Statistical Science*, 18, 71–103.

Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. (2001), "Empirical Bayes Analysis of a Microarray Experiment," *Journal of the American Statistical Association*, 96, 1151–1160.

Goeman, J. J., Van de Geer, S. A., De Kort, F., and Van Houwelingen, H. C. (2004), "A Global Test for Groups of Genes: Testing Association With a Clinical Outcome," *Bioinformatics*, 20, 93–99.

Hassau, R., Remaley, A. T., Sampson, M. L., Zhang, J., Cox, D. D., Pingpank, J., Alexander, R., Willingham, M., Pastan, I., and Onda, M. (2006), "Detection and Quantitation of Serum Mesothelin, a Tumor Marker for Patients With Mesothelioma and Ovarian Cancer," *Clinical Cancer Research*, 12, 447–453.

Hellawell, G. O., Turner, G. D. H., Davies, D. R., Poulsom, R., Brewster, S. F., and Macaulay, V. M. (2002), "Expression of the Type 1 Insulin-Like Growth Factor Receptor Is Up-Regulated in Primary Prostate Cancer and Commonly Persists in Metastatic Disease," *Cancer Research*, 62, 2942–2950.

Hellstrom, I., Raycraft, J., Kanan, S., Sardesai, N. Y., Verch, T., Yang, Y., and Hellstrom, K. E. (2006), "Mesothelin Variant 1 Is Released From Tumor Cells as a Diagnostic Marker," *Cancer Epidemiology Biomarkers and Prevention*, 15, 1014–1020.

Hufton, S. E., Moerkerk, P. T., Brandwijk, R., De Bruin, A. P., Arends, J. W., and Hoogenboomb, H. R. (1999), "A Profile of Differentially Expressed Genes in Primary Colorectal Cancer Using Suppression Subtractive Hybridization," *FEBS Letters*, 463, 77–82.

Klebanov, L., Gordon, A., Xiao, Y., Land, H., and Yakovlev, A. (2006), "A Permutation Test Motivated by Microarray Data Analysis," *Computational Statistics & Data Analysis*, 50, 3619–3628.

LaPointe, J., Li, C., Higgins, J. P., Van de Rijn, M., Bair, E., Montgomery, K., Ferrari, M., Egevad, L., Rayford, W., Bergerheim, U., Ekman, P., DeMarzo, A. M., Tibshirani, R., Botstein, D., Brown, P. O., Brooks, J. D., and Pollack, J. R. (2004), "Gene Expression Profiling Identifies Clinically Relevant Subtypes of Prostate Cancer," *Proceedings of the National Academy of Sciences*, 101, 811–816.

Nguyen, D. V., Arpat, A. B., Wang, B., and Carroll, R. J. (2002), "Microarray Experiments: Biological and Technical Aspects," *Biometrics*, 58, 701–717.

Pejovic, T. (1995), "Genetic Changes in Ovarian Cancer," *Annals of Medicine*, 27, 73–78.

R Development Core Team (2005), *R: A Language and Environment for Statistical Computing*, Vienna: R Foundation for Statistical Computing. For more information, please see *http://www.r-project.org*.

Rochester, M. A., Riedemann, J., Hellawell, G. O., Brewster, S. F., and Macaulay, V. M. (2005), "Silencing of the IGF1R Gene Enhances Sensitivity to DNA–Damaging Agents in Both PTEN Wild-Type and Mutant Human Prostate Cancer," *Cancer Gene Therapy*, 12, 90–100.

Schummer, M., Ng, W. V., Bumgarner, R. E., Nelson, P. S., Schummer, B., Bednarski, D. W., Hassell, L., Baldwin, R. L., Karlan, B. Y., and Hood, L. (1999), "Comparative Hybridization of an Array of 21500 Ovarian cDNAs for the Discovery of Genes Overexpressed in Ovarian Carcinomas," *Gene*, 238, 375–385.

Sun, X. M., Butterworth, M., MacFarlane, M., Dubiel, W., Ciechanover, A., and Cohen, G. M. (2004), "Caspase Activation Inhibits Proteasome Function During Apoptosis," *Molecular Cell*, 14, 81–93.

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985), *Statistical Analysis of Finite Mixture Distributions*, New York: Wiley.

Tusher, V. G., Tibshirani, R., and Chu, G. (2001), "Significance Analysis of Microarrays Applied to Ionizing Radiation Response," *Proceedings of the National Academy of Sciences*, 98, 5116–5121.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge, U.K.: Cambridge University Press.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer.

Xiao, Y., Gordon, A., and Yakovlev, A. (2006), "The L_1-Version of the Cramér–von Mises Test for Two-Sample Comparisons in Microarray Data Analysis," *EURASIP Journal of Bioinformatics and Computational Biology*, dio:10.1155/BSB/2006/85769.