BMC
Bioinformatics

METHODOLOGY ARTICLE

Open Access

Empirical Bayes estimation of posterior probabilities of enrichment: A comparative study of five estimators of the local false discovery rate

Zhenyu Yang[1], Zuojing Li[2] and David R Bickel[1*]

Abstract

Background: In investigating differentially expressed genes or other selected features, researchers conduct hypothesis tests to determine which biological categories, such as those of the Gene Ontology (GO), are enriched for the selected features. Multiple comparison procedures (MCPs) are commonly used to prevent excessive false positive rates. Traditional MCPs, e.g., the Bonferroni method, go to the opposite extreme: strictly controlling a family-wise error rate, resulting in excessive false negative rates. Researchers generally prefer the more balanced approach of instead controlling the false discovery rate (FDR). However, the q-values that methods of FDR control assign to biological categories tend to be too low to reliably estimate the probability that a biological category is not enriched for the preselected features. Thus, we study an application of the other estimators of that probability, which is called the local FDR (LFDR).

Results: We considered five LFDR estimators for detecting enriched GO terms: a binomial-based estimator (BBE), a maximum likelihood estimator (MLE), a normalized MLE (NMLE), a histogram-based estimator assuming a theoretical null hypothesis (HBE), and a histogram-based estimator assuming an empirical null hypothesis (HBE-EN). Since NMLE depends not only on the data but also on the specified value of π_0, the proportion of non-enriched GO terms, it is only advantageous when either π_0 is already known with sufficient accuracy or there are data for only 1 GO term. By contrast, the other estimators work without specifying π_0 but require data for at least 2 GO terms. Our simulation studies yielded the following summaries of the relative performance of each of those four estimators. HBE and HBE-EN produced larger biases for 2, 4, 8, 32, and 100 GO terms than BBE and MLE. BBE has the lowest bias if π_0 is 1 and if the number of GO terms is between 2 and 32. The bias of MLE is no worse than that of BBE for 100 GO terms even when the ideal number of components in its underlying mixture model is unknown, but has high bias when the number of GO terms is small compared to the number of estimated parameters. For unknown values of π_0, BBE has the lowest bias for a small number of GO terms (2-32 GO terms), and MLE has the lowest bias for a medium number of GO terms (100 GO terms).

Conclusions: For enrichment detection, we recommend estimating the LFDR by MLE given at least a medium number of GO terms, by BBE given a small number of GO terms, and by NMLE given either only 1 GO term or precise knowledge of π_0.

*Correspondence: dbickel@uottawa.ca
[1] Ottawa Institute of Systems Biology, Department of Biochemistry, Microbiology, and Immunology, Department of Mathematics and Statistics, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, K1H 8M5 Canada
Full list of author information is available at the end of the article

BioMed Central

Background

The development of microarray techniques and high-throughput genomic, proteomic, and bioinformatics scanning approaches (such as microarray gene expression profiling, mass spectrometry, and ChIP-on-chip) has enabled researchers to simultaneously study tens of thousands of biological features (e.g., genes, proteins, single-nucleotide polymorphisms [SNPs], etc.), and to identify a set of features for further investigation. However, there remains the challenge of interpreting these features biologically. For a given set of features, the determination of whether some biological information terms are enriched (i.e., differentially represented), compared to the reference feature set, is termed the *feature enrichment* problem. The biological information term may be, for instance, a Gene Ontology (GO) term [1,2] or a pathway in the Kyoto Encyclopedia of Genes and Genomes (KEGG) [3]. We call this problem the *feature enrichment problem.*

This problem has been addressed using a number of high-throughput enrichment tools, including DAVID, MAPPFinder, Onto-Express and GoMiner [4-7]. Huang et al. [8] reviewed 68 distinct feature enrichment analysis tools. These authors further classified feature enrichment analysis tools into 3 categories: singular enrichment analysis (SEA), gene set enrichment analysis (GSEA), and modular enrichment analysis (MEA). In this article, we propose empirical Bayes solutions to the SEA problem using genes as archetypal features. Without loss of generality, we consider whether some specific biological categories are enriched for differentially expressed genes with respect to the reference genes.

Indeed, like other enrichment-detection methods, our methods apply much more broadly. They can assess enrichment given any sub-list of features selected for future study, not just a list of genes considered differentially expressed. An anonymous referee pointed out these examples of such lists of candidate features that arise in the context of whole genome sequencing:

- genes with SNPs
- genes with copy number variations
- genes with loss of heterozygosity

These examples and those of our first paragraph do not exhaust contemporary applications, and the feature enrichment problem may occur in unforeseen domains of study. Thus, our illustrative use of differential gene expression as a running example should not be interpreted as a limitation.

Existing enrichment tools mainly address the feature enrichment problem using a p-value obtained from an exact or approximate statistical test (e.g., Fisher's exact test, the hypergeometric test, binomial test, or the χ^2 test). For each GO term or other biological category, the null hypothesis tested and its alternative hypothesis are as follows:

H_0: the GO term is not enriched for the preselected genes
H_1 : the GO term is enriched for the preselected genes

(1)

Here and in the remainder of the paper, we use GO terms as concrete examples of biological categories without excluding applications of the methods to categories from other relevant databases. The general process begins as follows:

- For each GO term, construct Table 1 based on the preselected genes (e.g., differentially expressed (DE) genes) and reference genes (e.g., all genes measured in a microarray experiment).
- Compute the p-value for each GO term using a statistical test that can detect enrichment for the preselected genes.

Multiple comparison procedures (MCPs) are then applied to the resulting p-values to prevent excessive false positive rates. The false discovery rate (FDR) [9] is frequently used to control the expected proportion of incorrectly rejected null hypotheses in gene enrichment studies [10-12] because it has lower false negative rates than Bonferroni correction and other methods of controlling the family-wise error rate. Methods of FDR control assign q-values [13] to biological categories, but q-values are too low to reliably estimate the probability that the biological category is not enriched for the preselected features. Thus, we study application of better estimators of that probability, which is technically known as the local FDR (LFDR). Hong et al. [14] used an LFDR estimator to solve a GSEA problem and pointed out that this was less biased than the q-value for estimating the LFDR, the posterior probability that the null hypothesis is true.

Efron [15,16] devised reliable LFDR estimators for a range of applications in microarray gene expression analysis and other problems of large-scale inference. However, whereas microarray gene expression analysis takes into account tens of thousands of genes, the feature enrichment problem typically concerns a much smaller

Table 1 The number of differentially expressed (DE) and equivalently expressed (EE) genes in a GO category

	DE genes	EE gene	Total
In GO category	x_1	x_2	$x_1 + x_2$
Not in GO category	$n - x_1$	$N - n - x_2$	$N - x_1 - x_2$
Total	n	$N - n$	N

Here, x_i ($i = 1, 2$) is the number of DE ($i = 1$) or EE genes ($i = 2$) in the GO category; n is the total number of DE genes; N is the total number of reference genes.

number of GO terms. While these methods are appropriate for microarray-scale inference, they are less reliable for enrichment-scale inference [17-19]. Thus, we will specifically adapt LFDR estimators that are appropriate for smaller-scale inference to address the SEA problem. Again, we will focus on genes and GO terms for the sake of concreteness. Nevertheless, the estimators used can be applied to other features and to other biological terms (e.g., metabolic pathways).

The sections of this paper are arranged as follows. We first introduce some preliminary concepts in the feature enrichment problem. Next, two previous LFDR estimators and three new LFDR estimators are described. Following this, we compare the LFDR estimators by means of a simulation study and an application to breast cancer data. Finally, we draw conclusions and make recommendations on the basis of our results.

Preliminary concepts

The feature enrichment problem described in the Background section is stated here more formally for the application of LFDR methods in the next section.

Likelihood functions

In Table 1, x_1 and x_2 are the observed numbers DE genes and EE genes in a given GO category, respectively. Whereas n is the total number of DE genes, N is the total number of reference genes. Thus, $N - n$ is the total number of EE genes. The columns gives the numbers of DE genes and EE genes, and the rows give the numbers of genes in the GO category and outside the GO category.

Let X_1 and X_2, respectively, denote the random numbers of DE and EE genes in a GO category. The observed values x_1 and x_2 are modeled as realizations of X_1 and X_2. X_1 and X_2 follow binomial distributions, namely, $X_1 \sim$ Binomial(n, Π_1) and $X_2 \sim$ Binomial($N - n, \Pi_2$), where Π_1 and Π_2 are the probabilities that a gene is DE and EE, respectively, given that it is in the GO category. Under the assumption that X_1 and X_2 are independent, the *unconditional likelihood* is

$$L(\Pi_1, \Pi_2; x_1, x_2, n, N) \qquad (2)$$
$$= \Pr(X_1 = x_1, X_2 = x_2; \Pi_1, \Pi_2, n, N)$$
$$= \binom{n}{x_1}\binom{N-n}{x_2}\Pi_1^{x_1}(1-\Pi_1)^{n-x_1}\Pi_2^{x_2}(1-\Pi_2)^{N-n-x_2},$$

where $0 \le x_1 \le n$, $0 \le x_2 \le N - n$, and $0 \le \Pi_i \le 1$, $i = 1, 2$.

If we define

$$\lambda = \ln[\,\Pi_2/(1 - \Pi_2)\,], \qquad (3)$$

and

$$\theta = \ln[\,\Pi_1/(1 - \Pi_1)\,] - \lambda \qquad (4)$$

then θ is the parameter of interest, representing the *log odds ratio* of the GO term, and λ is a nuisance parameter. Under the new parametrization, the unconditional likelihood function (2) is

$$L(\theta, \lambda; x_1, x_2, n, N) = \frac{\binom{n}{x_1}\binom{N-n}{x_2} \times e^{x_1(\theta+\lambda)}e^{x_2\lambda}}{(1 + e^{\theta+\lambda})^n(1 + e^{\lambda})^{N-n}}, \qquad (5)$$

where $0 \le x_1 \le n$ and $0 \le x_2 \le N - n$.

In equation (5), we take the interest parameter θ and also the nuisance parameter λ into consideration. Consider statistics T and S, functions of X_1 and X_2, such that $T(X_1, X_2) = X_1$ and $S(X_1, X_2) = X_1 + X_2$. Thus, T represents the number of DE genes in a GO category, and S represents the number of total genes in a GO category. Let t and s be the observed values of T and S. The probability mass function of $T(x_1, x_2) = t$ evaluated at $S(x_1, x_2) = x_1 + x_2 = s$, say $\Pr(T = t|S = s; \theta, \lambda, N, n)$, does not depend on the nuisance parameter λ [19]. See also Example 8.47 of Severini[20]. Thus, we derive the conditional probability mass function

$$f_\theta(t|s) = \Pr(T = t|S = s; \theta, n, N)$$
$$= \frac{\binom{n}{t}\binom{N-n}{s-t}e^{t\theta}}{\sum_{j=\max(0,s+n-N)}^{\min(s,n)} \binom{n}{j}\binom{N-n}{s-j}e^{j\theta}} \qquad (6)$$

understood as a function of t.

By eliminating the nuisance parameter λ, we can reduce the original data x_1 and x_2 by considering the statistic $T = t$. However, the use of the conditional probability mass function requires some justification because of concerns about losing information during the conditioning process. Unfortunately, in the presence of the nuisance parameter, the statistic $S(X_1, X_2) = X_1 + X_2$ is not an ancillary statistic for the parameter of interest. In other words, the probability mass function of the conditional variable $S(X_1, X_2)$ may contain some information about parameter θ [20]. However, following the explanation of Barndor-Nielsen and Cox ([21], §2.5), the expectation value of statistic $S(X_1, X_2)$ equals the nuisance parameter. Hence, from the observation of $S(X_1, X_2)$ alone, the distribution of $S(X_1, X_2)$ contains little information about θ [21]. $S(X_1, X_2)$ satisfies the other 3 conditions of an ancillary statistic defined by Barndor-Nielsen and Cox [21]: parameters θ and λ are variation independent; $(T(X_1, X_2), S(X_1, X_2))$ is the minimal sufficient statistic; and the distribution of $T(X_1, X_2)$, given $S(X_1, X_2) = s$, is independent of the parameter of interest, θ, given the nuisance parameter λ. Therefore, the probability mass function of $S(X_1, X_2)$ contains little information about the value of θ.

Hypotheses and LFDRs

Considering GO term i, we denote the T, S, t, s, and θ used in equation (6) as T_i, S_i, t_i, s_i, and θ_i. From Table 1, hypothesis comparison (1) of GO term i is equivalent to

$$H_0 : \theta_i = 0 \text{ versus } H_1 : \theta \neq 0. \tag{7}$$

Let $\mathbf{S} = \langle S_1, S_2, \cdots, S_m \rangle$ and $\mathbf{s} = \langle s_1, s_2, \cdots, s_m \rangle$. Let BF_i denote the *Bayes factor* of GO term i:

$$\mathrm{BF}_i = \frac{\Pr(T_i = t_i | \mathbf{S} = \mathbf{s}, \theta_i \neq 0)}{\Pr(T_i = t_i | \mathbf{S} = \mathbf{s}, \theta_i = 0)}. \tag{8}$$

It is called the Bayes factor because it yields posterior odds when multiplied by prior odds. More precisely, the *posterior odds* of the alternative hypothesis corresponding to GO term i is

$$\omega_i = \frac{\Pr(\theta_i \neq 0 | t_i)}{\Pr(\theta_i = 0 | t_i)} = \mathrm{BF}_i \times \frac{(1 - \pi_0)}{\pi_0}, \tag{9}$$

where π_0 is the *prior conditional probability* that a GO term is not enriched for the preselected genes given $\mathbf{s}$, i.e., $\pi_0 = \Pr(\theta_i = 0 | \mathbf{S} = \mathbf{s})$. Thus, $(1 - \pi_0)/\pi_0$ is the *prior odds* of the alternative hypothesis of enrichment. According to Bayes' theorem, the LFDR of GO term i is

$$\mathrm{LFDR}_i = \Pr(\theta_i = 0 | t_i) = \frac{1}{1 + \omega_i}, \tag{10}$$

where ω_i is defined in equation (9).

Methods

This section is divided into two parts:

1. **Previous LFDR estimators.** While not unique to this paper, these methods are included for comparison.
2. **New LFDR estimators.** Our main methodological innovations are the uses of a conditional probability mass function and of normalized maximum likelihood for LFDR estimation.

The other original contributions of this paper are the estimator comparisons of the next section. The comparisons are made by simulation and by a case study.

Previous LFDR estimators

Binomial-based LFDR estimator

The version of the FDR that generalizes the LFDR is the *nonlocal FDR*, which is defined as the ratio of the expected number of false discoveries to the expected total number of discoveries [17]. In our running example, a *discovery of enrichment* is a rejection of the null hypothesis of non-enrichment at some significance level α, and a *false discovery of enrichment* is a discovery of enrichment corresponding to a case of no actual enrichment. (This FDR has been called the "Bayesian FDR" [22]

to distinguish it from the FDR of Benjamini and Hochberg [9]).

Let α denote any significance level chosen to be between 0 and 1. For all GO terms of interest, the nonlocal FDR may be estimated by

$$\widehat{\mathrm{FDR}}(\alpha) = \min \left(\frac{m\alpha}{\sum_{j=1}^{m} \mathbf{1}_{\{p_j \leq \alpha\}}}, 1 \right), \tag{11}$$

where m is the number of GO terms, p_j is the p-value of GO term j, and $\mathbf{1}_{\{p_j \leq \alpha\}}$ is the indicator such that $\mathbf{1}_{\{p_j \leq \alpha\}} = 1$ if $p_j \leq \alpha$ is true and $\mathbf{1}_{\{p_j \leq \alpha\}} = 0$ otherwise. Thus, $\sum_{j=1}^{m} \mathbf{1}_{\{p_j \leq \alpha\}}$ represents the number of discoveries of enriched GO terms, and $m\alpha$ estimates the number of such discoveries that are false.

Let r_i be the rank of the p-value of GO term i, e.g., $r_i = 1$ if the p-value of GO term i is the smallest among all p-values of m GO terms. Based on a modification of equation (11), the *binomial-based estimator* (BBE) of LFDR of the GO term i is

$$\widehat{\mathrm{LFDR}}_i = \begin{cases} \min \left(\frac{m p_{2r_i}}{2 r_i}, 1 \right), & r_i \leq \frac{m}{2}, \\ 1, & r_i > \frac{m}{2}. \end{cases} \tag{12}$$

It is conservative in the sense that it tends to overestimate LFDR [17].

Histogram-based LFDR estimator

Efron [15,16] devised reliable histogram-based LFDR estimators for a range of applications in microarray gene expression analysis and other problems of large-scale inference. Let $z_i = \Phi^{-1}(p_i)$ be the z-transformed statistic of GO term i, where Φ is the standard normal cumulative distribution function (cdf) and p_i is the 2-sided p-value of GO term i. For each GO term, the density is a mixture of the form

$$f(z_i) = \pi_0 f_0(z_i) + (1 - \pi_0) f_1(z_i), \tag{13}$$

where f_0 is the density function of z for the non-enriched GO terms, f_1 is that for the enriched GO terms, and π_0 is the probability that a GO term is non-enriched. The histogram-based LFDR of GO term i is estimated by equation (14):

$$\widehat{\mathrm{LFDR}}_i = \frac{\hat{f}_0(z_i)}{\hat{f}(z_i)}, \tag{14}$$

where $\hat{f}$ is the estimator of f that is estimated by a non-parametric Poisson regression method [15,16]. We call $\widehat{\mathrm{LFDR}}_i$ the *histogram-based estimator* (HBE) if the density function f_0 is assumed to be standard normal, $N(0, 1)$, and *the histogram-based estimator with empirical null* (HBE-EN) if the density function f_0 is estimated based on the truncated maximum likelihood technique of [16]. Dalmasso et al. [23] compared the precursor of HBE-EN [15] to other LFDR estimators.

New LFDR estimators

Type II maximum likelihood estimator

Bickel [17] follows Good [24] in calling the maximization of likelihood over a hyperparameter *Type II maximum likelihood* to distinguish it from the usual *Type I maximum likelihood*, which pertains only to models that lack random parameters. Type II maximum likelihood has been applied to parametric mixture models (PMMs) for the analysis of microarray data [25,26], proteomics data [18], and genetic association data [27]. In this section, we adapt the approach to the feature enrichment problem by using the conditional probability mass function defined above. The particular models we use in this framework correspond to new methods of enrichment analysis.

Let $\mathcal{G}(\mathbf{s}) = \{g_\theta(\bullet|\mathbf{s}); \theta \geq 0\}$ be a parametric family of probability mass functions with

$$g_\theta(\bullet|\mathbf{s}) = \frac{1}{2} \times \left[f_\theta(\bullet|\mathbf{s}) + f_{-\theta}(\bullet|\mathbf{s}) \right], \quad (15)$$

where $f_\theta(\bullet|\mathbf{s})$ is defined in equation (6). We define the *k-component PMM* as

$$g(\bullet|\mathbf{s}; \theta_0, \ldots, \theta_{k-1}, \pi_0, \ldots, \pi_{k-1}) = \sum_{j=0}^{k-1} \pi_j g_{\theta_j}(\bullet|\mathbf{s}), \quad (16)$$

where $\theta_0 = 0$ and $\theta_j \neq \theta_J$ for all $j, J \in \{0, \ldots, k-1\}$ such that $j \neq J$.

Let $\mathbf{T} = \langle T_1, T_2, \cdots, T_m \rangle$ and $\mathbf{t} = \langle t_1, t_2, \cdots, t_m \rangle$ be vectors of the T_is and t_is used in equation (8). Assuming T_i is independent of T_j and S_j for all $i, j \in \{1, \ldots, m\}$ such that $j \neq J$. $i \neq j$, the joint probability mass function is

$$\begin{aligned} &g(\mathbf{t}|\mathbf{s}; \theta_0, \ldots, \theta_{k-1}, \pi_0, \ldots, \pi_{k-1}) \\ &= \prod_{i=1}^{m} g(t_i|\mathbf{s}; \theta_0, \ldots, \theta_{k-1}, \pi_0, \ldots, \pi_{k-1}) \\ &= \prod_{i=1}^{m} g(t_i|s_i; \theta_0, \ldots, \theta_{k-1}, \pi_0, \ldots, \pi_{k-1}), \end{aligned} \quad (17)$$

where s_i is the observed value of S_i for GO term i, and $\mathbf{s} = \langle s_1, s_2, \cdots, s_m \rangle$.

Moreover, we assume that for a given number of genes in GO term i, T_i ($i = 1, \ldots, m$) satisfies the k-component PMM shown in equation (16). In other words, we assume that the possible log odds ratios of GO term i are the $\theta_0, \theta_1, \theta_2, \ldots, \theta_{k-1}$ of equation (16) if the alternative hypothesis H_1 in hypothesis comparison (7) is true.

Therefore, the log-likelihood function under the k-component PMM for all GO terms is

$$\begin{aligned} &\log L(\theta_0, \ldots, \theta_{k-1}, \pi_0, \ldots, \pi_{k-1}) \\ &= \log g(\mathbf{t}|\mathbf{s}; \theta_0, \ldots, \theta_{k-1}, \pi_0, \ldots, \pi_{k-1}) \\ &= \sum_{i=1}^{m} \left[\log \sum_{j=0}^{k-1} \pi_j g_{\theta_j}(t_i|s_i) \right]. \end{aligned} \quad (18)$$

The LFDR of GO term i is estimated by

$$\widehat{\text{LFDR}}_i^{(k)} = \frac{\widehat{\pi}_0 g_{\theta_0}(t_i|s_i)}{g(t_i|s_i; \theta_0, \widehat{\theta}_1, \ldots, \widehat{\theta}_{k-1}, \widehat{\pi}_0, \ldots, \widehat{\pi}_{k-1})}, \quad (19)$$

where $\widehat{\theta}_1, \ldots, \widehat{\theta}_{k-1}$ and $\widehat{\pi}_0, \ldots, \widehat{\pi}_{k-1}$ are maximum likelihood estimates of $\theta_1, \ldots, \theta_{k-1}$ and $\pi_0, \ldots, \pi_{k-1}$ in equation (18). We call $\widehat{\text{LFDR}}_i^{(k)}$ the *k-component maximum likelihood estimator* (MLE*k*). Our LFDRenrich and LFDRhat software suites of R functions that implement MLE2 and MLE3 are now available at www.statomics.com. Moreover, $\widehat{\theta}_i$ ($i = 1, \ldots, k-1$; $k = 2, 3$) and $\widehat{\pi}_j$ ($j = 0, \ldots, k-1$; $k = 2, 3$), also in LFDRenrich and LFDRhat, are maximum likelihood estimators of θ_i ($i = 1, \ldots, k-1$; $k = 2, 3$) and π_j ($j = 0, \ldots, k-1$; $k = 2, 3$) under given constraints .

LFDR estimator based on the normalized maximum likelihood

Combining equations (9)-(10), we obtain

$$\text{LFDR}_i = \left(1 + \text{BF}_i \times \frac{(1 - \pi_0)}{\pi_0} \right)^{-1}. \quad (20)$$

Therefore, given a guessed value of π_0, we may use an estimator of the Bayes factor to estimate the LFDR of a GO term.

We now develop such an estimator of the Bayes factor. For GO category i, let $\mathcal{E}_i$ stand for the set of all probability mass functions defined on $\{0, 1, \ldots, s_i\}$, the set of all possible values of t_i. Based on hypothesis comparison (7), the set of log odds ratios, denoted as $\mathcal{H}$, is $\{0\}$ under the null hypothesis and is $\mathbb{R} \setminus \{0\} = \{\theta \in \mathbb{R} : \theta \neq 0\}$, the set of all real values except 0, under the alternative hypothesis. With the assumption that random variable T_i is independent of random variable S_j for any $i \neq j$, the *regret* of a predictive mass function $\bar{f} \in \mathcal{E}_i$ is a measure of how well it predicts the observed value $t_i \in \{0, 1, \ldots, s_i\}$. The regret is defined as

$$\text{reg}(\bar{f}, t_i|s_i; \mathcal{H}) = \log \frac{f_{\widehat{\theta}_i(t_i|s_i)}(t_i|s_i)}{\bar{f}(t_i|s_i)}, \quad (21)$$

where $\widehat{\theta}_i(t_i|s_i)$ is a Type I MLE with respect to $\mathcal{H}$ under observed values t_i given s_i [28,29].

For all members of $\mathcal{E}_i$, the *optimal predictive conditional probability mass function* of GO category i and the hypothesis that $\theta_i \in \mathcal{H}$ is denoted by $f_i^\dagger(\bullet|s_i; \mathcal{H})$. It minimizes the maximal regret in sample space $\{0, 1, \ldots, s_i\}$ in the sense that it satisfies

$$f_i^\dagger(\bullet|s_i; \mathcal{H}) = \arg \min_{\bar{f} \in \mathcal{E}_i} \max_{t \in \{0, 1, \ldots, s_i\}} \text{reg}(\bar{f}, t|s_i; \mathcal{H}). \quad (22)$$

It is well known [28] that the predictive probability mass function that satisfies equation (22) is

$$f_i^\dagger(t_i|s_i; \mathcal{H}) = \frac{\max_{\theta \in \mathcal{H}} f_\theta(t_i|s_i)}{\mathcal{K}_i^\dagger(\mathcal{H})}, \quad (23)$$

where $f_\theta(t_i|s_i)$ is the conditional probability mass function defined in equation (6), and $\mathcal{K}_i^\dagger(\mathcal{H})$ is a constant defined as

$$
\begin{aligned}
\mathcal{K}_i^\dagger(\mathcal{H}) &= \max_{\theta \in \mathcal{H}} f_\theta(y|s_i) = \sum_{y=\max(0,s_i+n-N)}^{\min(s_i,n)} \max_{\theta \in \mathcal{H}} f_\theta(y|s_i) \\
&= \sum_{y=\max(0,s_i+n-N)}^{\min(s_i,n)} \frac{\binom{n}{y}\binom{N-n}{s_i-y}e^{y\hat{\theta}_i(y)}}{\sum_{j=\max(0,s_i+n-N)}^{\min(s_i,n)} \binom{n}{j}\binom{N-n}{s_i-j}e^{j\hat{\theta}_i(y)}},
\end{aligned}
\tag{24}
$$

where

$$
\hat{\theta}_i(y) = \arg\max_{\theta \in \mathcal{H}} f_\theta(y|s_i).
\tag{25}
$$

We call $f_i^\dagger(t_i|s_i; \mathcal{H})$ the *normalized maximum likelihood* (NML) associated with the hypothesis that $\theta_i \in \mathcal{H}$.

Thus, BF_i is approximated by

$$
\widehat{BF}_i^\dagger = \frac{f_i^\dagger(t_i|s_i; \mathbb{R}\setminus\{0\})}{f_i^\dagger(t_i|s_i; 0)},
\tag{26}
$$

which we call the *NML ratio*. (More generally, the logarithm of an NML ratio is interpreted as a measure of the evidential support for the alternative hypothesis over the null hypothesis [29,30]). Therefore, by combining equations (8) and (9), if we guess the prior probability π_0, the LFDR estimate of GO category i in the hypothesis comparison (7) is

$$
\widehat{LFDR}_i^\dagger = \left[1 + \frac{1-\pi_0}{\pi_0} \times \widehat{BF}_i^\dagger\right]^{-1},
\tag{27}
$$

where $\widehat{BF}_i^\dagger$ is defined in equation (26). We call this LFDR estimator the *NML estimator* (NMLE).

To assess the reliability of NML ratio $\widehat{BF}_i^\dagger$ for a particular data set, it will be compared to an empirical Bayes estimate of the Bayes factor that unlike NML, simultaneously takes all GO terms into account. Equations (19) and (20) suggest

$$
\widehat{BF}_i = \frac{1 - \widehat{LFDR}_i^{(k)}}{\widehat{LFDR}_i^{(k)}} \times \frac{1-\hat{\pi}_0}{\hat{\pi}_0}
\tag{28}
$$

as the empirical Bayes estimator of BF_i.

Results and discussion

In this section, we compared the LFDR estimators using simulation data and breast cancer data.

For each GO category, the p-value used in BBE to estimate LFDR is computed based on the 2-sided Fisher's exact test. In computing MLEk ($k = 2, 3$), θ_i ($i = 1, \ldots, k-1$) in equation (18) was constrained to lie between 0 and 10, whereas π_i ($i = 0, \ldots, k-1$) in equation (18) was allowed to take any value between 0 and 1 such that $\sum_{i=0}^{k-1} \pi_i = 1$.

Simulation studies

The aim of the following simulation studies is to compare the LFDR estimation biases of BBE, MLE2, MLE3, HBE, and HBE-EN. NMLE is not taken into account because its performance depends not only on the data, but also on the specified prior probability π_0.

The simulation setting involves $10,000$ genes in a microarray with 200 genes identified as DE and 100 GO terms. We conducted a separate simulation study using each of these values of π_0: 50%, 60%, 70%, 80%, 90%, and 94%.

Since the PMM behind MLE is optimal when the number of enriched GO terms is equal to the number non-enriched GO terms, we assessed the sensitivity of MLE to that symmetry assumption by using strongly asymmetric log odds ratios and by using symmetric ones. For each GO term, two configurations were used in this simulation to choose log odds ratios: the *asymmetric configuration*
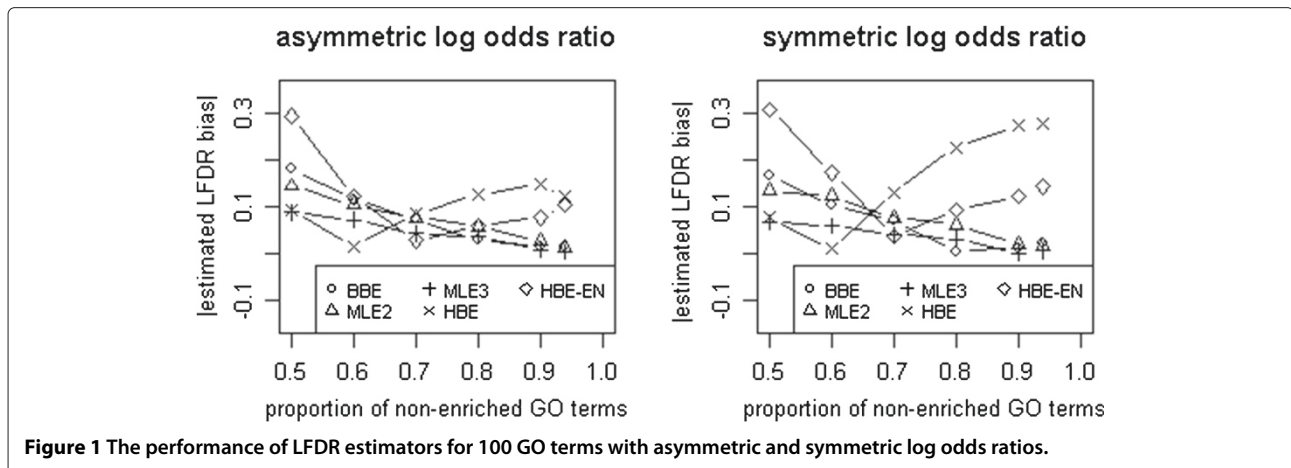

Figure 1 The performance of LFDR estimators for 100 GO terms with asymmetric and symmetric log odds ratios.

shown in equation (29) and the *symmetric configuration* shown in equation (30). We used these values of odds ratio of the ith GO term:

$$\phi_i^{\text{asymmetric}} = \begin{cases} \frac{5i}{100(1-\pi_0)}, & 1 \le i \le 100(1-\pi_0), \\ 0, & 100(1-\pi_0) < i \le 100; \end{cases}$$

(29)

$$\phi_i^{\text{symmetric}} = \begin{cases} \frac{5 \times 2i}{100(1-\pi_0)}, & 1 \le i \le 50(1-\pi_0), \\ 5 - \frac{5 \times 2i}{100(1-\pi_0)}, & 50(1-\pi_0) < i \le 100(1-\pi_0), \\ 0, & 100(1-\pi_0) < i \le 100. \end{cases}$$

(30)

Considering the log odds ratios of all GO terms in each simulation study, we generated Table 1 for GO term i and for each of the 20 simulated data sets as follows:

- x_1 is generated from a binomial distribution with parameter Π_1 used in equation (2); Π_1 is a real value randomly picked from 0 to 1.
- x_2 is obtained from a binomial distribution with parameter $\Pi_2 = \left[\frac{(1-\Pi_1) \times 2^{\phi_i}}{\Pi_1} + 1\right]^{-1}$, obtained by solving

$$\phi_i = \log_2[\Pi_1/(1 - \Pi_1)] - \log_2[\Pi_2/(1 - \Pi_2)]. \quad (31)$$

Thus, according to equation (4), we obtain $\phi_i = \theta_i \log_2 e$ for GO term i. Each of those artificial data sets represents what might have been a real data set such as that of the next subsection.

The p-value of each GO term used in BBE, HBE, and HBE-EN is obtained from the 2-sided Fisher's exact test. The k-component PMM ($k = 2$ or $k = 3$) used in MLE is shown in equation (16) with $\pi_j = (1 - \pi_0)/k$ $[j = 1, \ldots, k]$ and $g_{\theta_i}(t_i|s_i)$ defined in equation (15). For every log odds ratio sequence, we estimated the LFDR for each GO term and each data set using BBE, MLE2, MLE3, HBE, and HBE-EN. We compared the performances of the 5 estimators by means of computing the absolute value of the estimated LFDR bias. The true LFDR is computed by equation (10), where

$$f_0(t_i) = \frac{\binom{n}{t_i}\binom{N-n}{s_i-t_i}}{\sum_{j=\max(0,s_i+n-N)}^{\min(s_i,n)} \binom{n}{j}\binom{N-n}{s_i-j}}$$

and $f_1(t_i)$ is computed by

$$\frac{1}{J}\sum_{j=1}^{J} f_{\theta_j}(t_i|s_i),$$

where $f_\theta(t|s)$ is defined in equation (6).

Figure 1 shows the performance comparisons of the 5 LFDR estimators (i.e., BBE, MLE2, MLE3, HBE, and HBE-EN) for simulation data obtained from asymmetric

Table 2 The proportion of non-enriched GO terms and the log$_2$ odds ratios of GO terms used in the simulation studies

Number of GO terms	π_0	log$_2$ odds ratio (ϕ_i)	
		0.5,	$i = 1$
2	50.0%	0,	$i = 2$
	100.0%	0,	$i = 1, 2$
		3,	$i = 1$
	50.0%	-3,	$i = 2$
		0,	$i = 3, 4$
		0.5,	$i = 1$
4	75.0%	0,	$i = 2, \ldots, 4$
	100.0%	0,	$i = 1, \ldots, 4$
		$(0.5+1.5 \times (i-1))$,	$i = 1, 2$
	50.0%	$-(0.5+1.5 \times (i-3))$,	$i = 3, 4$
		0,	$i = 5, \ldots, 8$
		$(0.5+1.5 \times (i-1))$,	$i = 1, 2$
	62.5%	$-(0.5+1.5 \times (i-2))$,	$i = 3$
		0,	$i = 4, \ldots, 8$
		2,	$i = 1$
8	75.0%	-2,	$i = 2$
		0,	$i = 3, \ldots, 8$
		2,	$i = 1$
	87.5%	0,	$i = 2, \ldots, 8$
	100.0%	0,	$i = 1, \ldots, 8$
		$(0.32+0.64 \times (i-1))$,	$i = 1, \ldots, 8$
	50.0%	$-(0.32+0.64 \times (i-9))$,	$i = 9, \ldots, 16$
		0,	$i = 17, \ldots, 32$
		$(0.8+0.8 \times (i-1))$,	$i = 1, \ldots, 3$
	62.5%	$-(0.8+0.8 \times (i-4))$,	$i = 4, \ldots, 6$
		0,	$i = 7, \ldots, 32$
		$(0.2+1.6 \times (i-1))$,	$i = 1, 2$
32	75.0%	$-(0.2+1.6 \times (i-1))$,	$i = 3, 4$
		0,	$i = 5, \ldots, 32$
		1.8,	$i = 1$
	87.5%	-1.8,	$i = 2$
		0,	$i = 3, \ldots, 32$
	100.0%	0,	$i = 1, \ldots, 32$

Here, index i labels the GO term.

and symmetric log odds ratios. The absolute LFDR biases estimated by BBE, MLE2, MLE3, and HBE-EN are similar. The absolute bias of LFDR estimated by HBE on symmetric log odds ratios is a little higher than that on asymmetric log odds ratios when the proportion of non-enriched GO terms is greater than 80%. Therefore, the estimated LFDR biases of the estimators are not strongly affected by whether the log odds ratios are symmetric or asymmetric.

To assess the performance of the 5 estimators for smaller GO terms, we added simulation studies using 2, 4, 8, and 32 as the total number of GO terms. The proportion of non-enriched GO terms (π_0) and log odds ratios of simulation studies are shown in Table 2. The simulation studies were otherwise the same as those for 100 GO terms. Figure 2 shows the performance of LFDR estimators by means of computing the absolute estimated LFDR bias for 2, 4, 8, and 32 GO terms with log odds ratios based on formulas shown in Table 2.

Considering every (m, π_0) pair of the simulation studies with symmetric log odds ratios for the case of 100 GO terms, we recorded the LFDR estimator with the lowest absolute estimated LFDR bias among the 5 LFDR estimators (BBE, MLE2, MLE3, HBE, and HBE-EN). Moreover, we determined the maximum absolute LFDR bias over the proportion of non-enriched GO terms (π_0) in order to evaluate the worst-case bias of each estimator at each value of *m*. Figure 3 shows the results.

Breast cancer data analysis

The single-channel microarray data set used here to illustrate our new methods is from an experiment applying an estrogen treatment to cells of a human breast cancer cell line [31]. The Affymetrix human genome U-95Av2 genechip data are from four samples from an estrogen receptor positive breast cancer cell line. Two of the samples were exposed to estrogen and then harvested after 10 hours. The remaining two samples were left untreated and then harvested after 10 hours. For simplicity of terminology, we call probes in the microarray experiment "genes." The relevant data consist of measurements of gene expression across the reference class of 12, 625 genes. The purpose of the study was to determine which genes are affected by the estrogen treatment. (For further information concerning the data, see Gentleman et al. [32].)
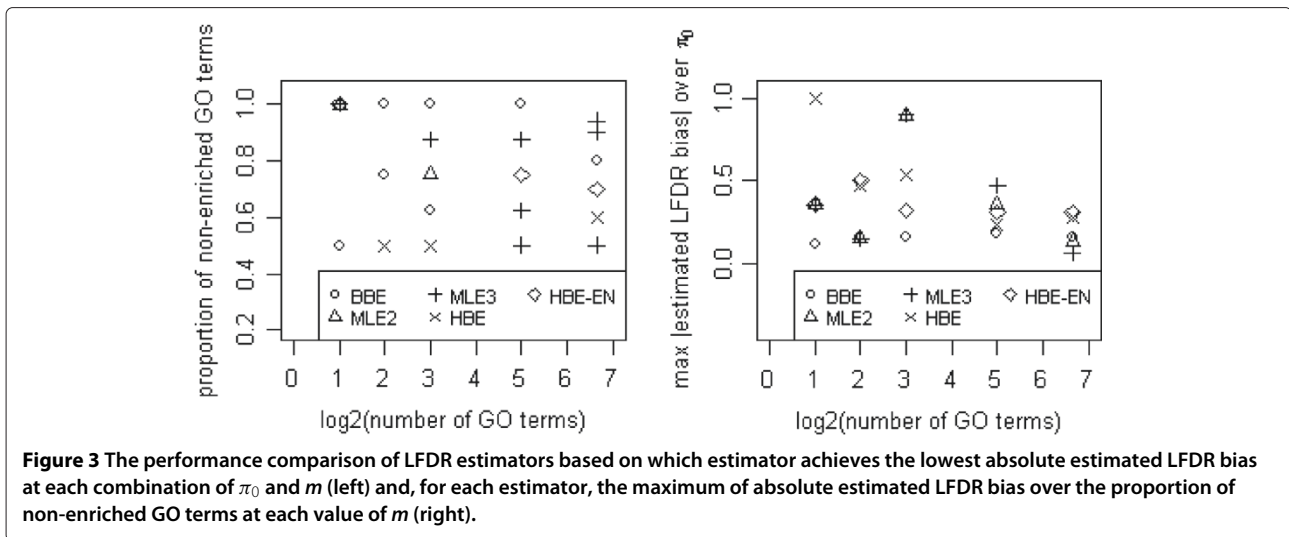
We applied the R function `expresso` in the `affy` package [33] of Bioconductor [34] to convert the raw probe intensities from the the CEL data files to logarithms of gene expression levels without background correction. In doing so, we applied the "quantiles," "pmonly," and "medianpolish" [35] preprocessing settings.

We selected as genes of interest those that were differentially expressed between the treatment group and the control group according to the following criterion.
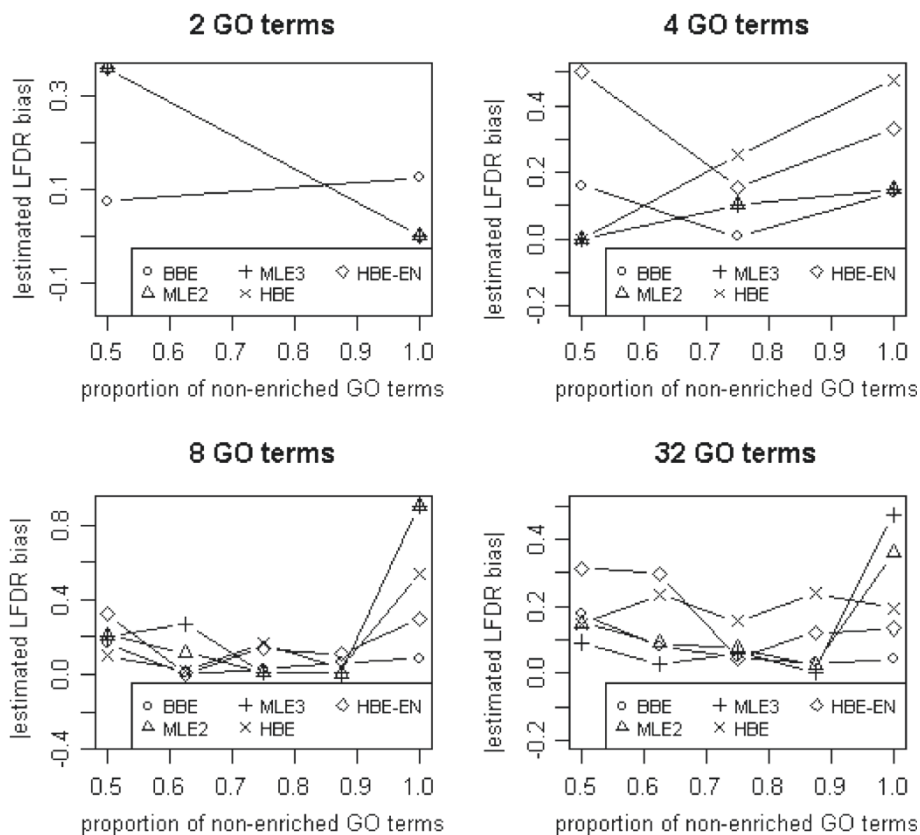
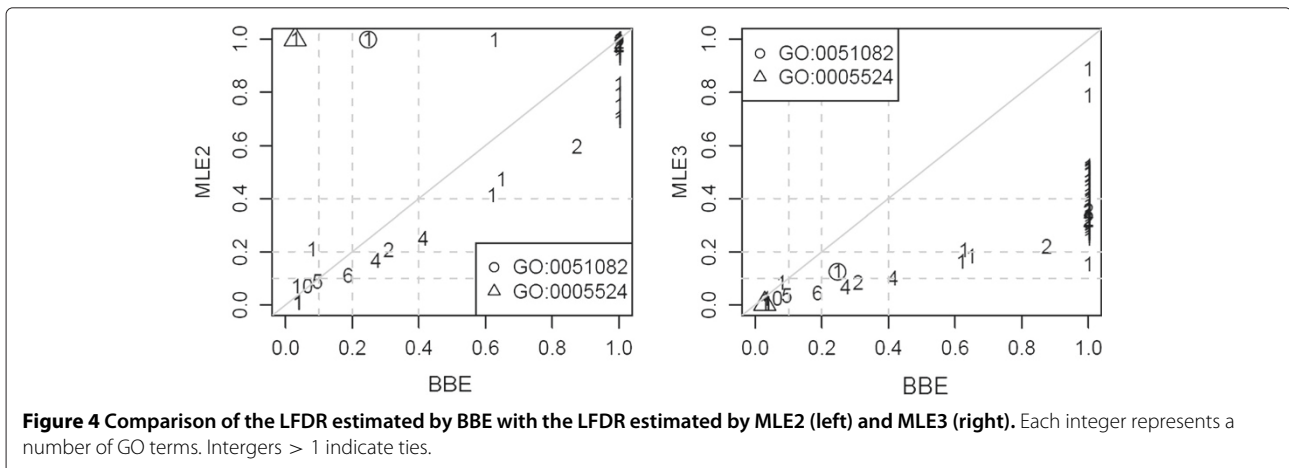**Figure 2 The performance of LFDR estimators for 2, 4, 8, and 32 GO terms with log odds ratios.**

Figure 3 The performance comparison of LFDR estimators based on which estimator achieves the lowest absolute estimated LFDR bias at each combination of π_0 and m (left) and, for each estimator, the maximum of absolute estimated LFDR bias over the proportion of non-enriched GO terms at each value of m (right).

Using the LFDR as the probability that a gene is EE, we considered genes with LFDR estimates below 0.2 as DE. In other words, we selected as DE genes those that were differentially expressed with estimated posterior probability of at least 80%. Considering four samples of each gene in the microarray, we used the unpaired t-test with equal variances to compute the p-value. The LFDR of every gene is estimated using the theoretical null hypothesis method of Efron [15,16]; the empirical null hypotheses method can lead to excessive bias due to deviations from normality [36]. When we compared gene expression data for the presence and absence of estrogen after 10 hours of exposure, we obtained 74 DE genes.
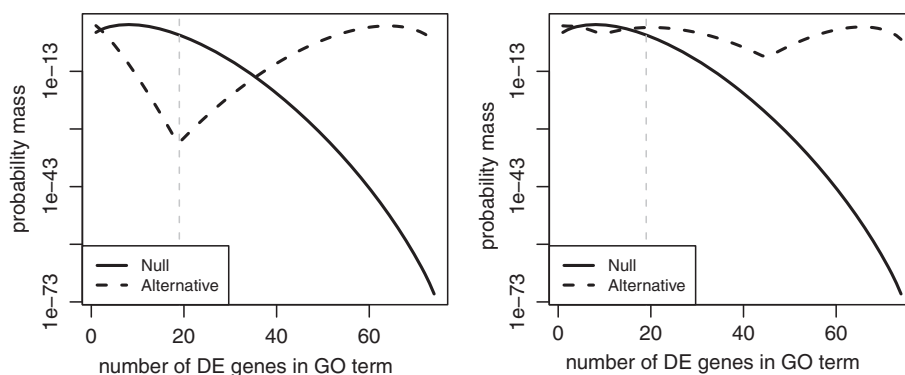
Defining *unrelated* pairs of GO terms as those that do not share any common ancestor, we selected for analysis all unrelated GO molecular function terms with at least 1 DE gene, thereby obtaining a total of 82 GO terms

of interest. Figure 4 compares the BBE to the MLEs based on the 2-component (MLE2) and 3-component (MLE3) PMM. Figure 5 displays the probability mass of GO:0005524 under the null and alternative hypotheses of hypothesis comparison (7). Figure 6 compares MLE-based estimates of the Bayes factor given by equation (28) to the NML ratios given by equation (28).

For two GO terms, opposite conclusions would be drawn about their enrichment, depending on which estimator is used. As seen in Figure 4, the estimated LFDRs of GO:0051082 and GO:0005524 using MLE2 were 100%. However, the LFDRs estimated by MLE3 were essentially 0.

Using the MLE formula shown in equation (19), and the k-component PMM shown in equation (16), we conclude that the sensitivity of the LFDRs of GO term i estimated by MLE2 and MLE3 depended mainly on the sensitivity of the Bayes factor, based on the number of

Figure 4 Comparison of the LFDR estimated by BBE with the LFDR estimated by MLE2 (left) and MLE3 (right). Each integer represents a number of GO terms. Intergers > 1 indicate ties.

Figure 5 The conditional probability mass functions given the number of genes in GO:0005524 under a null hypothesis, and alternative hypotheses based on 2-component PMM (left) and 3-component PMM (right). The grey dashed line is the number of DE genes in GO:0005524.

PMM components. Comparing the probability masses of GO:0005524, based on the 2- and 3-component PMMs shown in Figure 5, we found that the probability mass of GO:0005524 under the null hypothesis is larger than that under the alternative hypothesis based on the 2-component PMM (left plot in Figure 5). In contrast, the probability mass under the null hypothesis is smaller than that under the alternative hypothesis based on the 3-component PMM (right plot in Figure 5). Thus, the LFDR estimated by MLE is strongly dependent on the number of PMM components.

While a real data set can in that way indicate the impact of selecting an appropriate method, that impact does not in itself say which method has lowest bias. For that, we rely on the simulation study of the previous subsection.

Conclusions

As seen in Figure 1 and Figure 2, HBE and HBE-EN have relative high biases for a small number and a medium

number of GO terms, respectively. The performance comparison displayed in the left-hand side of Figure 3 indicates that BBE contains the lowest minimum estimated LFDR bias for a small number of GO terms (i.e., 2-32 GO terms) when the proportion of non-enriched GO terms is 1. Although the minimum bias of BEE is not the lowest for some π_0s under a small number of GO terms, it is very close to the lowest value of bias based on plots shown in Figure 2. The right-hand side of Figure 3 indicates that MLE3 has the lowest maximum absolute estimated LFDR bias in 100 GO terms. MLE exhibits bias similar to that of BBE when the number of GO terms is much larger than k except for when the proportion of non-enriched GO terms is high (close to 1). Moreover, MLE3 has lower bias than MLE2 as an LFDR estimator. Due to its conservatism and freedom from PMM, we recommend using BBE for a small number of GO terms of interest (2-32 GO terms) and MLE for a medium number of GO terms of interest (100 GO terms).
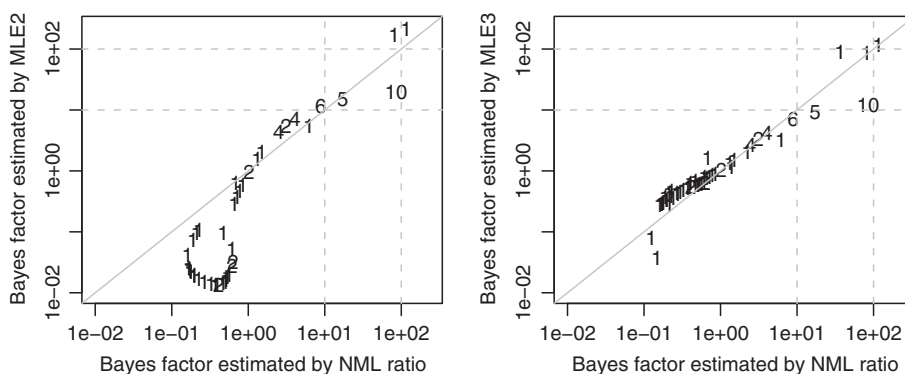
Figure 6 Comparison of the Bayes factor approximated by the NML ratio with that estimated by MLE2 (left) and MLE3 (right) on the basis of equations (26) and (28). The integers are defined in Figure 4. The grey dashed lines mark commonly used thresholds for strong and overwhelming evidence [37,38].

Finally, we recommend that NMLE be used when there is only 1 GO term of interest since none of the other estimators is able to estimate LFDR in such a case except by conservatively giving 1 as the estimate. Otherwise, unless π_0 is known with sufficient accuracy, NMLE is not recommended since it depends not only on the data but also on a guess of the value of π_0, which in the absence of strong prior information, is often set to the default value of 50%. A closely related approach is to use the logarithm of the NML ratio as a measure of statistical support for the enrichment hypothesis [30] without guessing π_0. By using 10 and 100 as thresholds of the approximate Bayes factors from equations (26) and (28) to determine whether a GO term is enriched, we reached similar conclusions with both NML and MLE (Figure 6). Thus, in our data set, the NML ratio tends to estimate the Bayes factor almost as accurately as methods that simultaneously use information across GO terms. While we do not expect the same for all data sets, we note that similar results have been found for an application of a modified NML [29] to a proteomics data set [30].

Competing interests

Authors' contributions

ZY implemented the NMLE function, co-designed and executed the simulation study, carried out the comparison among 3 LFDR estimators (BBE, MLE, and NMLE) using the breast cancer data, and drafted the manuscript. ZL implemented the 3-component MLE functions. DRB suggested and guided the project, contributed to writing the paper, co-designed the simulation study, and provided the BBE function. All authors have read and approved the final manuscript.

Acknowledgements

Author details

[1]Ottawa Institute of Systems Biology, Department of Biochemistry, Microbiology, and Immunology, Department of Mathematics and Statistics, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, K1H 8M5 Canada. [2]School of Foundation, Shenyang Pharmaceutical University, No. 103 Wenhua Road, Shenyang, Liaoning, 110016, China.

References

1. Altshuler D, Daly MJ, Lander ES: **Genetic mapping in human disease.** *Science* 2008, **322:**881–888.
2. Rhee SY, Wood V, Dolinski K, Draghici S: **Use and misuse of the gene ontology annotations.** *Nat Rev Genet* 2008, **9**(7):509–515.
3. Kanehisa M, Goto S: **KEGG: Kyoto Encyclopedia of Genes and Genome.** *Nucleic Acids Res* 2000, **28:**27–30.
4. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki R: **DAVID: Database for Annotation, Visualization, and Integrated Discovery.** *Genome Biol* 2003, **4:**P3.
5. Doniger SW, Salomonis N, Dahlquist KD, Vranizan K, Lawlor SC, Conklin BR: **MAPPFinder: using gene ontology and GenMAPP to create a global gene-expression profile from microarray data.** *Genome Biol* 2003, **4:**R7.
6. Khatri P, Draghici S, Ostermeier G, Krawetz S: **Profiling gene expression using onto-express.** *Genomics* 2002, **79:**266–270.
7. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, Riss J, Barrett JC, Weinstein JN: **GoMiner: a resource for biological interpretation of genomic and proteomic data.** *Genome Biol* 2003, **4:**R28.
8. Huang DW, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools paths toward the comprehensive functional analysis of large gene lists.** *Nucleic Acids Res* 2009, **37:**1–13.
9. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Stat Soc B* 1995, **57:**289–300.
10. Min JL, Barrett A, Watts T, Pettersson FH, Lockstone HE, Lindgren CM, Taylor JM, Allen M, Zondervan KT, McCarthy MI: **Variability of gene expression profiles in human blood and lymphoblastoid cell lines.** *BMC Genomics* 2010, **11:**96.
11. Reyal F, van Vliet MH, Armstrong NJ, Horlings HM, de Visser KE, Kok M, Teschendorff AE, Mook S, van't Veer L, Caldas C, Salmon RJ, Vijver MJVD, Wessels LFA: **A comprehensive analysis of prognostic signatures reveals the high predictive capacity of the proliferation, immune response and RNA splicing modules in breast cancer.** *Breast Cancer Res* 2008, **10:**R93.
12. Wang R, Bencic D, Lazorchak J, Villeneuve D, Ankley GT: **Transcriptional regulatory dynamics of the hypothalamic-pituitary-gonadal axis and its peripheral pathways as impacted by the 3-beta HSD inhibitor trilostane in zebrafish (*Danio rerio*).** *Ecotoxicol Environ Saf* 2011, **74:**1461–1470.
13. Storey JD: **The positive false discovery rate: a Bayesian interpretation and the q-value.** *Ann Stat* 2003, **31:**2013–2035.
14. Hong WJ, Tibshirani R, Chu G: **Local false discovery rate facilitates comparison of different microarray experiments.** *Nucleic Acids Res* 2009, **37:**7483–7497.
15. Efron B: **Large-scale simultaneous hypothesis testing: The choice of a null hypothesis.** *J Am Stat Assoc* 2004, **99:**96–104.
16. Efron B: *Large-Scale Inference: Empirical Bayes Methods for Estimation,Testing, and Prediction Cambridge*: Cambridge University Press; 2010.
17. Bickel DR: **Simple estimators of false discovery rates given as few as one or two p-values without strong parametric assumptions.** *Stat Appl Genet Mol Biol*. in press.
18. Bickel DR: **Small-scale inference: empirical Bayes and confidence methods for as few as a single comparison.**: Tech Rep, Ottawa Inst Syst Biol; 2011:arXiv:1104.0341.
19. Padilla M, Bickel DR: **Empirical Bayes methods corrected for small numbers of tests.** *Stat Appl Genet Mol Biol* 2012, **11**(5):art. 4.
20. Severini T: *Likelihood Methods in Statistics Oxford*: Oxford University Press; 2000.
21. Barndorff-Nielsen OE, Cox DR: *Inference and Asymptotics*. London: CRC Press; 1994.
22. Efron B, Tibshirani R: **Empirical Bayes methods and false discovery rates for microarrays.** *Genet Epidemiol* 2002, **23:**70–86.
23. Dalmasso C, Bar-Hen A, Broët P: **A constrained polynomial regression procedure for estimating the local false discovery rate.** *BMC Bioinformatics* 2007, **8:**229.
24. Good IJ: **How to estimate probabilities.** *IMA J Appl Math* 1966, **2:**364–383.
25. Pawitan Y, Murthy K, Michiels S, Ploner A: **Bias in the estimation of false discovery rate in microarray studies.** *Bioinformatics* 2005, **21:**3865–3872.
26. Muralidharan O: **An empirical Bayes mixture method for effect size and false discovery rate estimation.** *Ann Appl Stat* 2010, **4:**422–438.
27. Yang Y, Aghababazadeh FA, Bickel DR: **Parametric estimation of the local false discovery rate for identifying genetic associations.** *IEEE/ACM Trans Comput Biol Bioinformatics* 2012. online ahead of print at doi:10.1109/TCBB.2012.140.

28. Grünwald PD: *The Minimum Description Length Principle*. London: MIT Press; 2007.
29. Bickel DR: **A predictive approach to measuring the strength of statistical evidence for single and multiple comparisons.** *Can J Stat* 2011, **39:**610–631.
30. Bickel DR: **Minimax-optimal strength of statistical evidence for a composite alternative hypothesis.** *Int Stat Rev* 2013. in press. 2011 version available at arXiv:1101.0305.
31. Scholtens D, Miron A, Merchant FM, Miller A, Miron PL, Iglehart JD, Gentleman R: **Analyzing factorial designed microarray experiments.** *J Multivariate Anal* 2004, **90:**19–43.
32. Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S (Eds): *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. New York: Springer; 2005.
33. Gautier L, Cope L, Bolstad BM, Irizarry RA: **Affy—analysis of Affymetrix Gene Chip data at the probe level.** *Bioinformatics* 2004, **20**(3):307–315.
34. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J: **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biol* 2004, **5:**R80.
35. Tukey JW: *Exploratory Data Analysis*. Reading: Addison-Wesley; 1977 .
36. Bickel DR: **Estimating the null distribution to adjust observed confidence levels for genome-scale screening.** *Biometrics* 2011, **67:**363–370.
37. Jeffreys H: *Theory of Probability*. London: Oxford University Press; 1948.
38. Bickel DR: **The strength of statistical evidence for composite hypotheses: inference to the best explanation.** *Statistica Sinica* 2012, **22:**1147–1198.