

Shrinkage estimation of expression fold change as an alternative to testing hypotheses of equivalent expression

Zahra Montazeri Corey M. Yanofsky David R. Bickel

Ottawa Institute of System Biology

December 2009



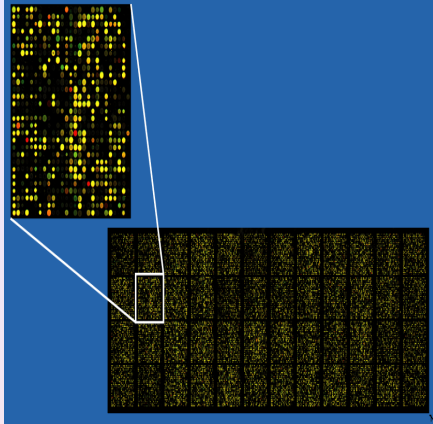


Figure: Example of an approximately 37,500 probe spotted oligo microarray with enlarged inset to show detail, en.wikipedia

Book1													
	A	B	C	D	E	F	G	H	I	J	K	L	
1	Name	Affy#175_2.0_Signal	Affy#2172_2.0_Signal	Affy#174_2.0_Signal									
2	AFFX-BioB-5_at	10627.7	6968.8	10925.6									
3	AFFX-BioB-M_at	11878.7	14586.9	17402.9									
4	AFFX-BioB-3_at	9633.9	10455.8	11152.1									
5	AFFX-BioC-5_at	15124.4	15018.5	16590.1									
6	AFFX-BioC-3_at	12678.2	14157	15198									
7	AFFX-BioDn-5_at	27626.6	26617.3	30816.9									
8	AFFX-BioDn-3_at	71775	69112.7	83522.4									
9	AFFX-CreX-5_at	130172.3	124006.3	150566.4									
10	AFFX-CreX-3_at	178951.4	173266.3	224628.8									
11	AFFX-DapX-5_at	199349.9	202840.5	280395.9									
12	AFFX-DapX-M_at	225677.2	226668.8	310669									
13	AFFX-DapX-3_at	233000	240969.3	321402									
14	AFFX-LysX-5_at	120485	136917.5	239509.8									
15	AFFX-LysX-M_at	218677.7	216654.2	298563.8									
16	AFFX-LysX-3_at	241432.3	241684	344666.8									
17	AFFX-PheX-5_at	36736.5	36207.6	89825.7									
18	AFFX-PheX-M_at	105890.8	119860.9	229873.5									
19	AFFX-PheX-3_at	195624.3	205276.4	278327.4									
20	AFFX-Thx-5_at	3182.5	4709.6	7202.9									
21	AFFX-Thx-M_at	17485.8	19010.2	53599.5									
22	AFFX-Thx-3_at	170854.9	180899.5	262971.5									
23	AFFX-TrpnX-5_at	7648.5	12248.9	22569.3									
24	AFFX-TrpnX-M_at	61673.5	134521.7	116461.9									
25	AFFX-TrpnX-3_at	115746.9	125389.4	171737.5									
26	AFFX-r2-Ec-bioB-5_ε	14077.8	13675.4	16634.8									
27	AFFX-r2-Ec-bioB-M_ε	14657.5	16088.3	16839.7									
28	AFFX-r2-Ec-bioB-3_ε	9931.8	8925.9	10594.2									
29	AFFX-r2-Ec-bioC-5_ε	20223	19282.3	24138									
30	AFFX-r2-Ec-bioC-3_ε	27999.2	22317.9	26847.6									
31	AFFX-r2-Ec-bioD-5_ε	65936.7	66141.4	86641.8									
32	AFFX-r2-Ec-bioD-3_ε	82285.6	78938.1	99192.7									
33	AFFX-r2-P1-cre-5_at	218382	213221.2	263908.9									
34	AFFX-r2-P1-cre-3_at	239042	239590	301087.1									
35	AFFX-r2-Bs-dap-5_at	219498.5	228909	331025.3									
36	AFFX-r2-Bs-dap-M_ε	259317.8	251629.9	380863.4									
37	AFFX-r2-Bs-dap-3_at	230784.4	231869.6	30079.8									
38	AFFX-r2-Bs-dap-5_ε	110477.7	127088.6	214517									

Figure: A piece of microarray data



- Applying microarray data has focus on the problem of identifying differentially expressed genes
- Researchers currently prioritize genes for further study
 - on the basis of volcano plots
 - simple estimates of the fold change after filtering the genes with an arbitrary statistical significance threshold
 - ▶ hard-threshold estimator of the expression ratio is not known to perform well in terms of mean-squared error, the sum of estimator variance and squared estimator bias.

- Applying microarray data has focus on the problem of identifying differentially expressed genes
- **Researchers currently prioritize genes for further study**
 - on the basis of volcano plots
 - subjective and informal nature of volcano method precludes quantification of its reliability.
 - simple estimates of the fold change after filtering the genes with an arbitrary statistical significance threshold
 - hard-threshold estimator of the expression ratio is not known to perform well in terms of mean-squared error, the sum of estimator variance and squared estimator bias.

- Applying microarray data has focus on the problem of identifying differentially expressed genes
- Researchers currently prioritize genes for further study
 - on the basis of volcano plots
 - subjective and informal nature of volcano method precludes quantification of its reliability.
 - simple estimates of the fold change after filtering the genes with an arbitrary statistical significance threshold
 - hard-threshold estimator of the expression ratio is not known to perform well in terms of mean-squared error, the sum of estimator variance and squared estimator bias.

- Applying microarray data has focus on the problem of identifying differentially expressed genes
- Researchers currently prioritize genes for further study
 - on the basis of volcano plots
 - subjective and informal nature of volcano method precludes quantification of its reliability.
 - simple estimates of the fold change after filtering the genes with an arbitrary statistical significance threshold
 - hard-threshold estimator of the expression ratio is not known to perform well in terms of mean-squared error, the sum of estimator variance and squared estimator bias.

- Applying microarray data has focus on the problem of identifying differentially expressed genes
- Researchers currently prioritize genes for further study
 - on the basis of volcano plots
 - subjective and informal nature of volcano method precludes quantification of its reliability.
 - simple estimates of the fold change after filtering the genes with an arbitrary statistical significance threshold
 - hard-threshold estimator of the expression ratio is not known to perform well in terms of mean-squared error, the sum of estimator variance and squared estimator bias.

- Applying microarray data has focus on the problem of identifying differentially expressed genes
- Researchers currently prioritize genes for further study
 - on the basis of volcano plots
 - subjective and informal nature of volcano method precludes quantification of its reliability.
 - simple estimates of the fold change after filtering the genes with an arbitrary statistical significance threshold
 - hard-threshold estimator of the expression ratio in not known to perform well in terms of mean-squared error, the sum of estimator variance and squared estimator bias.

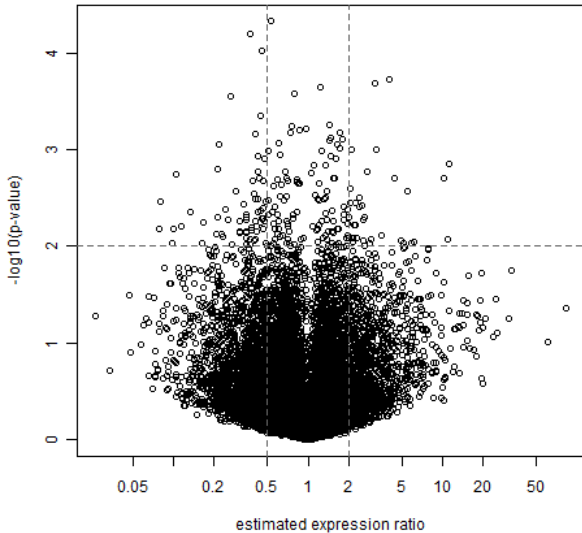


Figure: Volcano plot

Shrinkage estimation



- The main goal is to develop a method to identify differentially expressed genes
 - without using subjective and informal quantification
 - without using any hard-threshold estimators
 - that performs well in terms of mean-squared error

- The main goal is to develop a method to identify differentially expressed genes
 - without using subjective and informal quantification
 - without using any hard-threshold estimators
 - that performs well in terms of mean-squared error

- The main goal is to develop a method to identify differentially expressed genes
 - without using subjective and informal quantification
 - **without using any hard-threshold estimators**
 - that performs well in terms of mean-squared error

- The main goal is to develop a method to identify differentially expressed genes
 - without using subjective and informal quantification
 - without using any hard-threshold estimators
 - **that performs well in terms of mean-squared error**

- **Hard-threshold:** Comparing the corrected p-value, estimated global or local false discovery rate, or approximate posterior probability to a some arbitrary or subjective value α that sharply separate low-priority genes from high-priority genes.
- **Shrinkage:** Shrinks the estimate of the fold change toward 1 to the extent that the variability is too high for precise estimation.

- **Hard-threshold:** Comparing the corrected p-value, estimated global or local false discovery rate, or approximate posterior probability to a some arbitrary or subjective value α that sharply separate low-priority genes from high-priority genes.
- **Shrinkage:** Shrinks the estimate of the fold change toward 1 to the extent that the variability is too high for precise estimation.

Hard-threshold estimates

- General form:

$$\hat{\mu}_{\text{HT}} = \begin{cases} \hat{\mu}_{\text{MLE}} & \text{if p-value} < \alpha \\ 0 & \text{if p-value} \geq \alpha \end{cases} \quad (1)$$

Different methods of generating p-values lead to different hard-threshold estimators and we consider

- Standard p-value
- Adjusted p-value
- Empirical Bayes

Hard-threshold estimates

- General form:

$$\hat{\mu}_{\text{HT}} = \begin{cases} \hat{\mu}_{\text{MLE}} & \text{if p-value} < \alpha \\ 0 & \text{if p-value} \geq \alpha \end{cases} \quad (1)$$

Different methods of generating p-values lead to different hard-threshold estimators and we consider

- **Standard p-value**
- Adjusted p-value
- Empirical Bayes

Hard-threshold estimates

- General form:

$$\hat{\mu}_{\text{HT}} = \begin{cases} \hat{\mu}_{\text{MLE}} & \text{if p-value} < \alpha \\ 0 & \text{if p-value} \geq \alpha \end{cases} \quad (1)$$

Different methods of generating p-values lead to different hard-threshold estimators and we consider

- Standard p-value
- **Adjusted p-value**
- Empirical Bayes

Hard-threshold estimates

- General form:

$$\hat{\mu}_{\text{HT}} = \begin{cases} \hat{\mu}_{\text{MLE}} & \text{if p-value} < \alpha \\ 0 & \text{if p-value} \geq \alpha \end{cases} \quad (1)$$

Different methods of generating p-values lead to different hard-threshold estimators and we consider

- Standard p-value
- Adjusted p-value
- **Empirical Bayes**

- Following set up is used:
 - $x_{i,j}$: logarithm of the measured expression intensity for the i^{th} gene and the j^{th} biological replicate of the control group.
 - $x'_{i,j}$: logarithm of the measured expression intensity for the i^{th} gene and the j^{th} biological replicate of the treatment group.
- The goal is to estimate:

$$\mu_i = E(X'_i - X_i)$$

- Expression ratio: $\exp(\mu_i)$
- Fold change: $\exp(|\mu_i|)$

- Following set up is used:
 $x_{i,j}$: logarithm of the measured expression intensity for the i^{th} gene and the j^{th} biological replicate of the control group.
 $x'_{i,j}$: logarithm of the measured expression intensity for the i^{th} gene and the j^{th} biological replicate of the treatment group.
- The goal is to estimate:

$$\mu_i = E(X'_i - X_i)$$

- Expression ratio: $\exp(\mu_i)$
- Fold change: $\exp(|\mu_i|)$

- Following set up is used:
 $x_{i,j}$: logarithm of the measured expression intensity for the i^{th} gene and the j^{th} biological replicate of the control group.
 $x'_{i,j}$: logarithm of the measured expression intensity for the i^{th} gene and the j^{th} biological replicate of the treatment group.
- The goal is to estimate:

$$\mu_i = E(X'_i - X_i)$$

- **Expression ratio: $\exp(\mu_i)$**
- Fold change: $\exp(|\mu_i|)$

- Following set up is used:
 $x_{i,j}$: logarithm of the measured expression intensity for the i^{th} gene and the j^{th} biological replicate of the control group.
 $x'_{i,j}$: logarithm of the measured expression intensity for the i^{th} gene and the j^{th} biological replicate of the treatment group.
- The goal is to estimate:

$$\mu_i = E(X'_i - X_i)$$

- Expression ratio: $\exp(\mu_i)$
- **Fold change: $\exp(|\mu_i|)$**

- Estimate μ_j based on general MLE method and normality assumption

- For paired data: Define $y_{i,j} = x'_{i,j} - x_{i,j}$, and

$$\hat{\mu}_{j,\text{MLE}} = \bar{y}_j$$

- For non-paired data: X'_j and X_j are independent, and

$$\hat{\mu}_{j,\text{MLE}} = \bar{x}'_j - \bar{x}_j$$

- Estimate μ_i based on general MLE method and normality assumption
 - For paired data: Define $y_{i,j} = x'_{i,j} - x_{i,j}$, and

$$\hat{\mu}_{i.MLE} = \bar{y}_i$$

- For non-paired data: X'_i and X_i are independent, and

$$\hat{\mu}_{i.MLE} = \bar{x}'_i - \bar{x}_i$$

- Estimate μ_j based on general MLE method and normality assumption
 - For paired data: Define $y_{i,j} = x'_{i,j} - x_{i,j}$, and

$$\hat{\mu}_{i,\text{MLE}} = \bar{y}_i$$

- For non-paired data: X'_j and X_j are independent, and

$$\hat{\mu}_{i,\text{MLE}} = \bar{x}'_j - \bar{x}_j$$

Statistical significance by error rate

- The standard measure of statistical significance is a p-value (control type I error at level α). Using this p-value in (1) leads to “raw p-value hard-threshold estimator”.
 - This method fails to control the probability of at least one type I error in multiple comparison.
- In detection of differential gene expression, the adjusted p-value is used to control the family-wise error rate. Using adjusted p-value in (1) leads to “adjusted p-value hard-threshold estimator”.

Statistical significance by error rate

- The standard measure of statistical significance is a p-value (control type I error at level α). Using this p-value in (1) leads to “raw p-value hard-threshold estimator”.
 - This method fails to control the probability of at least one type I error in multiple comparison.
- In detection of differential gene expression, the adjusted p-value is used to control the family-wise error rate. Using adjusted p-value in (1) leads to “adjusted p-value hard-threshold estimator”.

Statistical significance by error rate

- The standard measure of statistical significance is a p-value (control type I error at level α). Using this p-value in (1) leads to “raw p-value hard-threshold estimator”.
 - This method fails to control the probability of at least one type I error in multiple comparison.
- In detection of differential gene expression, the adjusted p-value is used to control the family-wise error rate. Using adjusted p-value in (1) leads to “adjusted p-value hard-threshold estimator”.

Empirical Bayes methods

- An alternative to the control of type I error is a hard-threshold estimator based on an empirical Bayes estimate of the local false discovery rate (FDR). Using local FDR in (1) leads to “local false discovery rate hard-threshold estimator”.

Shrinkage estimators

- The shrinkage approach estimates levels of differentially expression by shrinkage the raw estimate toward equivalent expression.
- When p-values are large, estimates are shrunk strongly towards equivalent expression; when p-values are small, estimates are close to the MLE.
- Unlike the hard-threshold, intermediate p-values give intermediate estimates.

Shrinkage estimators

- The shrinkage approach estimates levels of differentially expression by shrinkage the raw estimate toward equivalent expression.
- When p-values are large, estimates are shrunk strongly towards equivalent expression; when p-values are small, estimates are close to the MLE.
- Unlike the hard-threshold, intermediate p-values give intermediate estimates.

Shrinkage estimators

- The shrinkage approach estimates levels of differentially expression by shrinkage the raw estimate toward equivalent expression.
- When p-values are large, estimates are shrunk strongly towards equivalent expression; when p-values are small, estimates are close to the MLE.
- Unlike the hard-threshold, intermediate p-values give intermediate estimates.

Shrinkage methods

We developed three methods of shrinkage:

- **Frequentist shrinkage estimators**
- Bayesian shrinkage model
- Local-FDR-based shrinkage estimator

Shrinkage methods

We developed three methods of shrinkage:

- Frequentist shrinkage estimators
- **Bayesian shrinkage model**
- Local-FDR-based shrinkage estimator

Shrinkage methods

We developed three methods of shrinkage:

- Frequentist shrinkage estimators
- Bayesian shrinkage model
- **Local-FDR-based shrinkage estimator**

Frequentist shrinkage estimators

- When a reasonable initial μ_0 is available, it is possible to reduce mean squared error by shrinking MLE toward μ_0 .
- Willink (2008) proposed two different shrinkage estimators for the mean

$$Q_1(h, c) = \mu_0 + \frac{\hat{\mu}_{\text{MLE}} - \mu_0}{1 + hR} + cV$$

$$Q_2(a, b, c) = \hat{\mu}_{\text{MLE}} - a(\hat{\mu}_{\text{MLE}} - \mu_0) \exp\left(-\frac{b}{R}\right) + cV$$

where

$$R = \frac{s^2/n}{(\hat{\mu}_{\text{MLE}} - \mu_0)^2}$$

$$V = \frac{s/\sqrt{n}}{1 + R} \times \text{sgn}(\hat{\mu}_{\text{MLE}} - \mu_0)$$



Frequentist shrinkage estimators

- When a reasonable initial μ_0 is available, it is possible to reduce mean squared error by shrinking MLE toward μ_0 .
- Willink (2008) proposed two different shrinkage estimators for the mean

$$Q_1(h, c) = \mu_0 + \frac{\hat{\mu}_{\text{MLE}} - \mu_0}{1 + hR} + cV$$

$$Q_2(a, b, c) = \hat{\mu}_{\text{MLE}} - a(\hat{\mu}_{\text{MLE}} - \mu_0) \exp\left(-\frac{b}{R}\right) + cV$$

where

$$R = \frac{s^2/n}{(\hat{\mu}_{\text{MLE}} - \mu_0)^2}$$

$$V = \frac{s/\sqrt{n}}{1 + R} \times \text{sgn}(\hat{\mu}_{\text{MLE}} - \mu_0)$$



Bayes shrinkage model

- The data are treated as a mixture of equivalently expressed and differentially expressed genes.
- The model achieves shrinkage by two mechanism
 - The data distribution contains a component explicitly modeling equivalently expressed genes as having a mean of zero.
 - Given that a gene is differentially expressed, the prior distribution of its expression is centered at a mean of zero.

Bayes shrinkage model

- The data are treated as a mixture of equivalently expressed and differentially expressed genes.
- **The model achieves shrinkage by two mechanism**
 - The data distribution contains a component explicitly modeling equivalently expressed genes as having a mean of zero.
 - Given that a gene is differentially expressed, the prior distribution of its expression is centered at a mean of zero.



Bayes shrinkage model

- The data are treated as a mixture of equivalently expressed and differentially expressed genes.
- The model achieves shrinkage by two mechanism
 - The data distribution contains a component explicitly modeling equivalently expressed genes as having a mean of zero.
 - Given that a gene is differentially expressed, the prior distribution of its expression is centered at a mean of zero.



Bayes shrinkage model

- The data are treated as a mixture of equivalently expressed and differentially expressed genes.
- The model achieves shrinkage by two mechanism
 - The data distribution contains a component explicitly modeling equivalently expressed genes as having a mean of zero.
 - Given that a gene is differentially expressed, the prior distribution of its expression is centered at a mean of zero.



Local-FDR-based shrinkage estimator

- Testing the local false discovery rate as an approximate posterior probability of equivalent expression for the purpose of estimating the degree of differential expression by the approximate posterior mean. This estimator shrinks the MLE by the posterior probability that the gene is actually differentially expressed.

$$\mu_{\text{LFDR shrinkage}} = [1 - \text{LFDR}] \hat{\mu}_{\text{MLE}}.$$



Comparing reliability of estimators

To quantify the performance of each estimators, the simulations and parametric bootstrap sampling procedures are employed.

- the mean-squared error (MSE) for each gene was estimated by the empirical MSE for each estimator, i.e.,

$$\widehat{MSE}_i = \frac{1}{B} \sum_{b=1}^B (\hat{\mu}_{b,i} - \mu_i)^2$$

where

- μ_i = true mean used to generate the simulated data
- $\hat{\mu}_{b,i}$ = estimate of μ_i for the b^{th} simulated example sample for each estimator
- B = number of simulated sample



Comparing reliability of estimators

To quantify the performance of each estimators, the simulations and parametric bootstrap sampling procedures are employed.

- the mean-squared error (MSE) for each gene was estimated by the empirical MSE for each estimator, i.e.,

$$\widehat{MSE}_i = \frac{1}{B} \sum_{b=1}^B (\hat{\mu}_{b,i} - \mu_i)^2$$

where

- μ_i = true mean used to generate the simulated data
- $\hat{\mu}_{b,i}$ = estimate of μ_i for the b^{th} simulated example sample for each estimator
- B = number of simulated sample



Comparing reliability of estimators

To quantify the performance of each estimators, the simulations and parametric bootstrap sampling procedures are employed.

- the mean-squared error (MSE) for each gene was estimated by the empirical MSE for each estimator, i.e.,

$$\widehat{MSE}_i = \frac{1}{B} \sum_{b=1}^B (\hat{\mu}_{b,i} - \mu_i)^2$$

where

- μ_i = true mean used to generate the simulated data
- $\hat{\mu}_{b,i}$ = estimate of μ_i for the b^{th} simulated example sample for each estimator
- B = number of simulated sample



Comparing reliability of estimators

To quantify the performance of each estimators, the simulations and parametric bootstrap sampling procedures are employed.

- the mean-squared error (MSE) for each gene was estimated by the empirical MSE for each estimator, i.e.,

$$\widehat{MSE}_i = \frac{1}{B} \sum_{b=1}^B (\hat{\mu}_{b,i} - \mu_i)^2$$

where

- μ_i = true mean used to generate the simulated data
- $\hat{\mu}_{b,i}$ = estimate of μ_i for the b^{th} simulated example sample for each estimator
- B = number of simulated sample



- Estimated “risk” associated with each estimator is

$$\widehat{\text{risk}} = \sum_{i=1}^M \widehat{\text{MSE}}_i$$

where M is number of genes.

Simulation study

- Two classes of simulations were carried out. For both cases, data sets with 2 and 8 biological replicates were simulated.
 - **Case I.** Half of the simulated genes were equivalently expressed across conditions, true means set to 0 (M.Langaas and Ferkingstad (2005)).
 - True means for the other half were drawn from a symmetric bi-triangular density and genes were dependent.

Simulation study

- Two classes of simulations were carried out. For both cases, data sets with 2 and 8 biological replicates were simulated.
 - **Case I.** Half of the simulated genes were equivalently expressed across conditions, true means set to 0 (M.Langaas and Ferkingstad (2005)).
 - True means for the other half were drawn from a symmetric bi-triangular density and genes were dependent.

Simulation study

- Two classes of simulations were carried out. For both cases, data sets with 2 and 8 biological replicates were simulated.
 - **Case I.** Half of the simulated genes were equivalently expressed across conditions, true means set to 0 (M.Langaas and Ferkingstad (2005)).
 - True means for the other half were drawn from a symmetric bi-triangular density and genes were dependent.

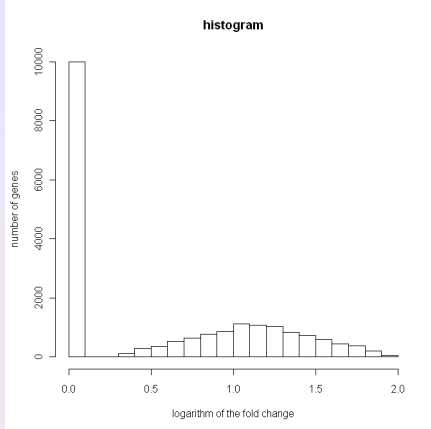


Figure: Logarithm of the true fold change for the simulation of 20,000 genes where half of genes are equivalently expressed.

- **Case II.** No simulated genes had exactly equivalent expression across groups although most genes had very small levels of differential expression (Bickel, 2008). The distributions were

$$\mu_i = \begin{cases} -\left(\frac{10001-i}{10000}\right)^8 & i \in \{1, \dots, 10^4\} \\ \left(\frac{i-10000}{10000}\right)^8 & i \in \{10001, \dots, 20^4\} \end{cases}$$

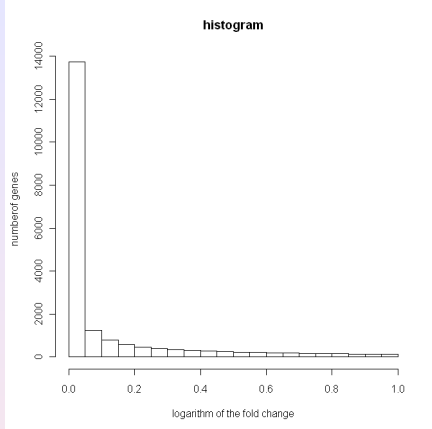


Figure: Logarithm of the true fold change for the simulation of 20,000 genes where all genes have at least some differential expression.

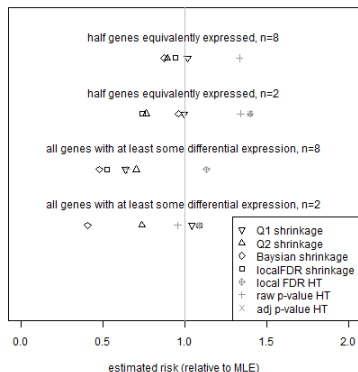


Figure: Estimated risk for all estimators relative to that of the MLE for the simulated data sets. The local FDR HT estimator for the first row and the adjusted p-value estimator for the first and third rows have relative risks greater than 2.0 and are not plotted.

Comparison of shrinkage and hard-threshold estimation

- We compare the best shrinkage estimators, Q2 shrinkage, with the best hard-threshold estimator, the raw p-value hard threshold estimator.
- The first and the third quartiles are plotted against the true expression ratio.

Comparison of shrinkage and hard-threshold estimation

- We compare the best shrinkage estimators, Q2 shrinkage, with the best hard-threshold estimator, the raw p-value hard threshold estimator.
- The first and the third quartiles are plotted against the true expression ratio.

The case where half of genes are equivalently expressed, $n=2$

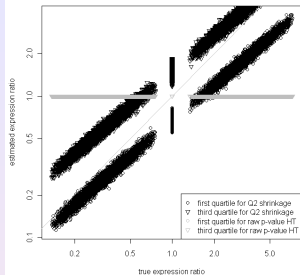


Figure: True expression ratio versus the first and third quartiles of both estimators over the 200 simulations for the case where half of genes are equivalently expressed and sample size is 2. When the sample size is small, the hard-threshold estimator cannot detect any differential expression.

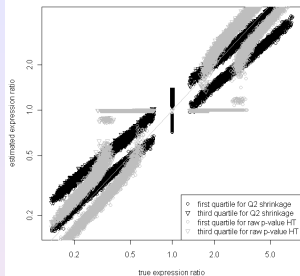


Figure: True expression ratio versus the first and third quartiles of both estimators over the 200 simulations for the case where half of genes are equivalently expressed and sample size is 8. The hard-threshold estimator tends to collapse to zero at fold changes between about 1.5 and 3.0. For fold changes greater than 3.0 it tracks the true expression ratio.



All genes have at least some differential expression, $n=2$

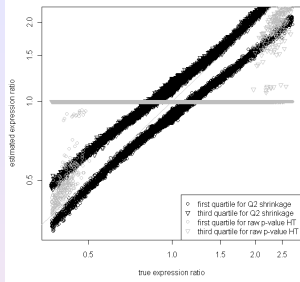


Figure: True expression ratio versus the first and third quartiles of both estimators over the 200 simulations for the case where all genes have at least some differential expression and sample size is 2.

Empirical validation

The estimators were evaluated by a frequentist method (parametric bootstrap) and by a Bayesian method (posterior expected loss) for two experimental data sets:

- **Data set I:** From an experiment applying an estrogen treatment to cells of a human breast cancer cell line (Scholtens *et al.* 2004). Two non-paired biological replicates were collected for each group in 2 steps.
- **Data set II:** A subset of data from the MAQC study (Guo *et al.* 2006) in which rat liver cells were treated with comfrey six non-paired biological replicates were collected for both treatment and control conditions.



Empirical validation

The estimators were evaluated by a frequentist method (parametric bootstrap) and by a Bayesian method (posterior expected loss) for two experimental data sets:

- **Data set I:** From an experiment applying an estrogen treatment to cells of a human breast cancer cell line (Scholtens *et al.* 2004). Two non-paired biological replicates were collected for each group in 2 steps.
- **Data set II:** A subset of data from the MAQC study (Guo *et al.* 2006) in which rat liver cells were treated with comfrey six non-paired biological replicates were collected for both treatment and control conditions.



Parametric bootstrap

- It is a strategy for estimating expectations over the sampling distribution when the parameters of that distribution are unknown. The parameters of the sampling distribution are estimated and replicate data sets are sampled from the sampling distribution with the parameter values fixed to the estimate from the original data.

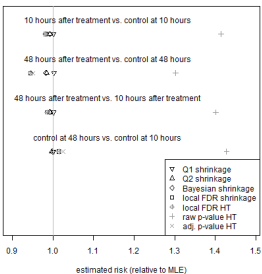


Figure: Estimated risk for all estimators relative to that of the MLE for the breast cancer data sets. The relative risk for the adjusted p-value hard-threshold estimator does not appear in the plot because it was greater than 1.5 in all cases.

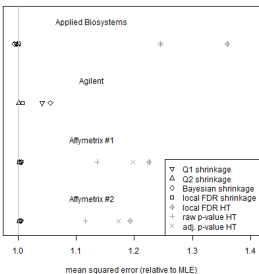


Figure: Estimated risk for all estimators relative to that of the MLE for the MAQC data sets. The relative risks for the hard-threshold estimators set were not plotted because they were greater than 1.9 in all cases.

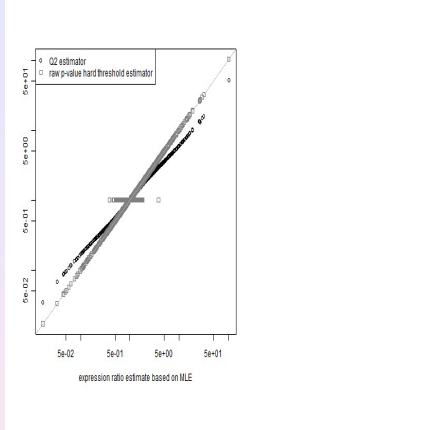


Figure: Maximum likelihood estimate versus both estimators for the first Affymetrix data set of the Microarray Quality Control experiment.

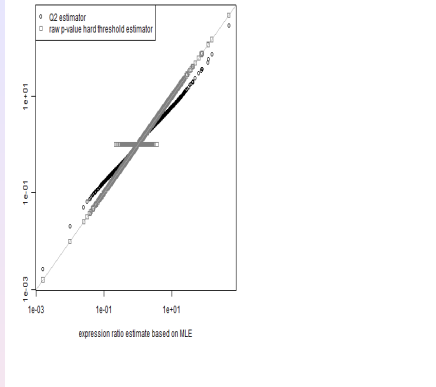


Figure: Maximum likelihood estimate versus both estimators for the Applied Biosystems data set of the Microarray Quality Control experiment.

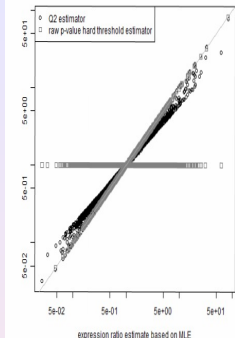


Figure: Maximum likelihood estimate versus both estimators for the 48 hours after treatment versus control at 48 hours data set of the breast cancer cell line experiment.

Discussion and conclusion

- In the simulation study:
 - The best hard-threshold estimator performed about as well as the best shrinkage estimator in one set of simulations, but in the other three sets of simulations, each of the shrinkage estimators proved to be substantially more reliable than each of the hard-threshold estimators.
- In the bootstrap study:
 - Cancer data: Each estimator except for the raw p-value hard-threshold estimator performed essentially as well as the MLE.
 - MAQC data: All shrinkage estimators performed essentially identically to the MLE, and all hard threshold estimators performed notably worse than the MLE.

Discussion and conclusion

- In the simulation study:
 - The best hard-threshold estimator performed about as well as the best shrinkage estimator in one set of simulations, but in the other three sets of simulations, each of the shrinkage estimators proved to be substantially more reliable than each of the hard-threshold estimators.
- In the bootstrap study:
 - Cancer data: Each estimator except for the raw p-value hard-threshold estimator performed essentially as well as the MLE.
 - MAGC data: All shrinkage estimators performed essentially identically to the MLE, and all hard threshold estimators performed notably worse than the MLE.

Discussion and conclusion

- In the simulation study:
 - The best hard-threshold estimator performed about as well as the best shrinkage estimator in one set of simulations, but in the other three sets of simulations, each of the shrinkage estimators proved to be substantially more reliable than each of the hard-threshold estimators.
- In the bootstrap study:
 - Cancer data: Each estimator except for the raw p-value hard-threshold estimator performed essentially as well as the MLE .
 - MAQC data: All shrinkage estimators performed essentially identically to the MLE, and all hard threshold estimators performed notably worse than the MLE.

Discussion and conclusion

- In the simulation study:
 - The best hard-threshold estimator performed about as well as the best shrinkage estimator in one set of simulations, but in the other three sets of simulations, each of the shrinkage estimators proved to be substantially more reliable than each of the hard-threshold estimators.
- In the bootstrap study:
 - **Cancer data: Each estimator except for the raw p-value hard-threshold estimator performed essentially as well as the MLE .**
 - MAQC data: All shrinkage estimators performed essentially identically to the MLE, and all hard threshold estimators performed notably worse than the MLE.

Discussion and conclusion

- In the simulation study:
 - The best hard-threshold estimator performed about as well as the best shrinkage estimator in one set of simulations, but in the other three sets of simulations, each of the shrinkage estimators proved to be substantially more reliable than each of the hard-threshold estimators.
- In the bootstrap study:
 - Cancer data: Each estimator except for the raw p-value hard-threshold estimator performed essentially as well as the MLE .
 - MAQC data: All shrinkage estimators performed essentially identically to the MLE, and all hard threshold estimators performed notably worse than the MLE.

- In general:

- No hard-threshold estimator ever had the least estimated risk and they were frequently far worse than the MLE and any shrinkage estimator.
- The shrinkage estimators often outperform the MLE and are never much worse.
- In particular, local FDR shrinkage had the best overall performance and is a good default choice, especially if only a few genes are expected to be differentially expressed at notable levels.
- We recommend the replacement of threshold-based prioritization of genes with shrinkage-based prioritization.

- In general:

- No hard-threshold estimator ever had the least estimated risk and they were frequently far worse than the MLE and any shrinkage estimator.
- The shrinkage estimators often outperform the MLE and are never much worse.
- In particular, local FDR shrinkage had the best overall performance and is a good default choice, especially if only a few genes are expected to be differentially expressed at notable levels.
- We recommend the replacement of threshold-based prioritization of genes with shrinkage-based prioritization.

Conclusion

- In general:
 - No hard-threshold estimator ever had the least estimated risk and they were frequently far worse than the MLE and any shrinkage estimator.
 - **The shrinkage estimators often outperform the MLE and are never much worse.**
 - In particular, local FDR shrinkage had the best overall performance and is a good default choice, especially if only a few genes are expected to be differentially expressed at notable levels.
 - We recommend the replacement of threshold-based prioritization of genes with shrinkage-based prioritization.

Conclusion

- In general:
 - No hard-threshold estimator ever had the least estimated risk and they were frequently far worse than the MLE and any shrinkage estimator.
 - The shrinkage estimators often outperform the MLE and are never much worse.
 - In particular, local FDR shrinkage had the best overall performance and is a good default choice, especially if only a few genes are expected to be differentially expressed at notable levels.
 - We recommend the replacement of threshold-based prioritization of genes with shrinkage-based prioritization.



- In general:
 - No hard-threshold estimator ever had the least estimated risk and they were frequently far worse than the MLE and any shrinkage estimator.
 - The shrinkage estimators often outperform the MLE and are never much worse.
 - In particular, local FDR shrinkage had the best overall performance and is a good default choice, especially if only a few genes are expected to be differentially expressed at notable levels.
 - **We recommend the replacement of threshold-based prioritization of genes with shrinkage-based prioritization.**

Acknowledgment

- This research was supported by
 - Agriculture and Agri-Foods Canada
 - Canada Foundation for Innovation
 - Ministry of Research and Innovation; Faculty of Medicine
- Special thanks to
 - Dr. David P. Bickel
 - Corey M. Yanofsky
 - Pei-Chun Hsieh

Acknowledgment

- This research was supported by
 - Agriculture and Agri-Foods Canada
 - Canada Foundation for Innovation
 - Ministry of Research and Innovation; Faculty of Medicine
- Special thanks to
 - Dr. David P. Bickel
 - Corey M. Yanofsky
 - Pei-Chun Hsieh



Acknowledgment

- This research was supported by
 - Agriculture and Agri-Foods Canada
 - **Canada Foundation for Innovation**
 - Ministry of Research and Innovation; Faculty of Medicine
- Special thanks to
 - Dr. David R. Bickel
 - Corey M. Yanofsky
 - Pei-Chun Hsieh

Acknowledgment

- This research was supported by
 - Agriculture and Agri-Foods Canada
 - Canada Foundation for Innovation
 - **Ministry of Research and Innovation; Faculty of Medicine**
- Special thanks to
 - Dr. David R. Bickel
 - Corey M. Yanofsky
 - Pei-Chun Hsieh

Acknowledgment

- This research was supported by
 - Agriculture and Agri-Foods Canada
 - Canada Foundation for Innovation
 - Ministry of Research and Innovation; Faculty of Medicine
- **Special thanks to**
 - Dr. David R. Bickel
 - Corey M. Yanofsky
 - Pei-Chun Hsieh



Acknowledgment

- This research was supported by
 - Agriculture and Agri-Foods Canada
 - Canada Foundation for Innovation
 - Ministry of Research and Innovation; Faculty of Medicine
- Special thanks to
 - **Dr. David R. Bickel**
 - Corey M. Yanofsky
 - Pei-Chun Hsieh



Acknowledgment

- This research was supported by
 - Agriculture and Agri-Foods Canada
 - Canada Foundation for Innovation
 - Ministry of Research and Innovation; Faculty of Medicine
- Special thanks to
 - Dr. David R. Bickel
 - **Corey M. Yanofsky**
 - Pei-Chun Hsieh



Acknowledgment

- This research was supported by
 - Agriculture and Agri-Foods Canada
 - Canada Foundation for Innovation
 - Ministry of Research and Innovation; Faculty of Medicine
- Special thanks to
 - Dr. David R. Bickel
 - Corey M. Yanofsky
 - Pei-Chun Hsieh



Thanks

Thank you

