A Novel Bayesian Approach for Differential Gene Expression Analysis with RNA-seq Data

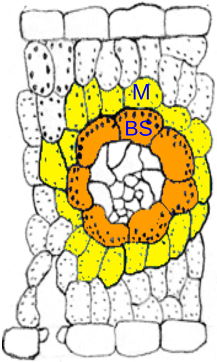
Peng Liu

Iowa State University
joint work with
Fangfang Liu, Chong Wang

Aug. 7, 2013

Example RNA-Seq Data

- Li et al (2010, *Nature Genetics*) studied the gene expression difference between BS and M cells in maize leaf.

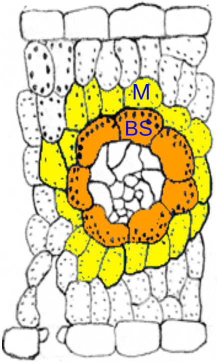


Gene ID	Length	BS		M	
1	1131	52	80	45	59
2	498	10	27	192	318
3	1896	2	3	0	0
4	990	177	382	7	6
5	363	1	12	12	38
...	...	...	...	...	...
G	609	102	135	88	127

- Question: Which genes are differentially expressed (DE) across treatment groups?

Example RNA-Seq Data

- Li et al (2010, *Nature Genetics*) studied the gene expression difference between BS and M cells in maize leaf.



Gene ID	Length	BS		M	
1	1131	52	80	45	59
2	498	10	27	192	318
3	1896	2	3	0	0
4	990	177	382	7	6
5	363	1	12	12	38
...	...	...	...	...	...
G	609	102	135	88	127

- Question: Which genes are differentially expressed (DE) across treatment groups?

Poisson Model

- Let X_{gij} denote the read count mapped to treatment i ($i = 1, 2$), replicate j and gene g.
- It has been assumed that $X_{gij} \sim \text{Poisson}(\lambda_{gij})$ with $\lambda_{gij} = S_{ij}\lambda_g \exp(\rho_i \delta_g)$.
  - S_{ij}: normalization factor, e.g., total number of mappable reads.
  - λ_g: the overall geometric mean expression level of gene g across both treatments
  - $\rho_1 = -1/2$ and $\rho_2 = 1/2$
  - δ_g: the log fold change (log-FC) between the two treatment means
- To identify genes differentially expressed between two treatments, test the hypothesis: $H_0^g : \delta_g = 0$ for each gene g.
- Poisson model has been shown to fit well for data with only technical replicates.

Poisson Model

- Let X_{gij} denote the read count mapped to treatment i ($i = 1, 2$), replicate j and gene g.
- It has been assumed that $X_{gij} \sim \text{Poisson}(\lambda_{gij})$ with $\lambda_{gij} = S_{ij}\lambda_g \exp(\rho_i\delta_g)$.
    - S_{ij}: normalization factor, e.g., total number of mappable reads.
    - λ_g: the overall geometric mean expression level of gene g across both treatments
    - $\rho_1 = -1/2$ and $\rho_2 = 1/2$
    - δ_g: the log fold change (log-FC) between the two treatment means
- To identify genes differentially expressed between two treatments, test the hypothesis: $H_0^g : \delta_g = 0$ for each gene g.
- Poisson model has been shown to fit well for data with only technical replicates.

Poisson Model

- Let X_{gij} denote the read count mapped to treatment i $(i = 1, 2)$, replicate j and gene g.
- It has been assumed that $X_{gij} \sim \text{Poisson}(\lambda_{gij})$ with $\lambda_{gij} = S_{ij}\lambda_g \exp(\rho_i\delta_g)$.
    - S_{ij}: normalization factor, e.g., total number of mappable reads.
    - λ_g: the overall geometric mean expression level of gene g across both treatments
    - $\rho_1 = -1/2$ and $\rho_2 = 1/2$
    - δ_g: the log fold change (log-FC) between the two treatment means
- To identify genes differentially expressed between two treatments, test the hypothesis: $H_0^g : \delta_g = 0$ for each gene g.
- Poisson model has been shown to fit well for data with only technical replicates.

Negative Binomial Model

- Let X_{gij} denote the read count mapped to treatment i ($i = 1, 2$), replicate j and gene g.
- For experiment with biological replicates, we assume that

$$X_{gij} \sim \text{NegativeBinomial}(\lambda_{gij}, \phi_g)$$

with

$$mean(X_{gij}) = \lambda_{gij} = S_{ij}\lambda_g \exp(\rho_i \delta_g)$$

and

$$var(X_{gij}) = \lambda_{gij} + \phi_g \lambda_{gij}^2$$

- To identify genes differentially expressed between two treatments, test the hypothesis: $H_0^g : \delta_g = 0$ for each gene g.

Negative Binomial Model

- Let X_{gij} denote the read count mapped to treatment i ($i = 1, 2$), replicate j and gene g.
- For experiment with biological replicates, we assume that

$$X_{gij} \sim \text{NegativeBinomial}(\lambda_{gij}, \phi_g)$$

with

$$mean(X_{gij}) = \lambda_{gij} = S_{ij}\lambda_g \exp(\rho_i\delta_g)$$

and

$$var(X_{gij}) = \lambda_{gij} + \phi_g\lambda_{gij}^2$$

- To identify genes differentially expressed between two treatments, test the hypothesis: $H_0^g : \delta_g = 0$ for each gene g.

Proposed tests for RNA-seq data

- *Fisher's exact test*: based on a 2 x 2 table for a given gene and all other genes across 2 treatments.
- χ^2 *test*: likelihood ratio test or goodness-of-fit test.
- *edgeR*: a test for Negative Binomial model with shrinkage estimator of dispersion parameter, (Robinson and Smyth, 2007, 2008).
- *DESeq*: a test for Negative Binomial model with a shrinkage estimator of variance, (Anders and Huber, 2010).
- *baySeq*: an empirical Bayes test for Negative Binomial model, (Hardcastle and Kelly, 2010).
- More...

Remaining Challenges

- There is no theoretical justification for the optimality of these methods or discussions on how to search for optimal tests for RNA-seq data.
- There is little information on how well the false discovery rate (FDR) is controlled.
- Most existing tests are designed for testing H_0^g: $\delta_g = 0$. Sometimes, it is interesting to test for big fold-changes, i.e.: test H_0^g: $\delta_g \in \Delta_0$ where $\Delta_0 = \{\delta : |\delta| \leq c\}$ with $c = \log 1.5$ for example.

Remaining Challenges

- There is no theoretical justification for the optimality of these methods or discussions on how to search for optimal tests for RNA-seq data.
- There is little information on how well the false discovery rate (FDR) is controlled.
- Most existing tests are designed for testing H_0^g: $\delta_g = 0$. Sometimes, it is interesting to test for big fold-changes, i.e.: test H_0^g: $\delta_g \in \Delta_0$ where $\Delta_0 = \{\delta : |\delta| \leq c\}$ with $c = \log 1.5$ for example.

Remaining Challenges

- There is no theoretical justification for the optimality of these methods or discussions on how to search for optimal tests for RNA-seq data.

- There is little information on how well the false discovery rate (FDR) is controlled.

- Most existing tests are designed for testing H_0^g: $\delta_g = 0$. Sometimes, it is interesting to test for big fold-changes, i.e.: test H_0^g: $\delta_g \in \Delta_0$ where $\Delta_0 = \{\delta : |\delta| \leq c\}$ with $c = \log 1.5$ for example.

Question of Interest

- Considering the huge dimension of tests, we aim to search for tests with large average power.

- More specifically, our goal is to derive a test with *maximum average power* while controlling FDR.

- In addition, we want to allow the null hypothesis to be intervals, such as small range of fold-changes. So we test for H_0^g: $\delta_g \in \Delta_0$ where,
  $\Delta_0 = \{0\}$ when testing for differential expression;
  $\Delta_0 = (-\infty, 0]$ when testing for higher expression in the second treatment;
  $\Delta_0 = \{\delta : |\delta| \leq c\}$ with $c = \log 1.5$ when testing whether fold-changes of expressions is greater than 1.5 or not.

Question of Interest

- Considering the huge dimension of tests, we aim to search for tests with large average power.
- More specifically, our goal is to derive a test with *maximum average power* while controlling FDR.
- In addition, we want to allow the null hypothesis to be intervals, such as small range of fold-changes. So we test for H_0^g: $\delta_g \in \Delta_0$ where,
  $\Delta_0 = \{0\}$ when testing for differential expression;
  $\Delta_0 = (-\infty, 0]$ when testing for higher expression in the second treatment;
  $\Delta_0 = \{\delta : |\delta| \leq c\}$ with $c = \log 1.5$ when testing whether fold-changes of expressions is greater than 1.5 or not.

Question of Interest

- Considering the huge dimension of tests, we aim to search for tests with large average power.
- More specifically, our goal is to derive a test with *maximum average power* while controlling FDR.
- In addition, we want to allow the null hypothesis to be intervals, such as small range of fold-changes. So we test for H_0^g: $\delta_g \in \Delta_0$ where,
  $\Delta_0 = \{0\}$ when testing for differential expression;
  $\Delta_0 = (-\infty, 0]$ when testing for higher expression in the second treatment;
  $\Delta_0 = \{\delta : |\delta| \leq c\}$ with $c = \log 1.5$ when testing whether fold-changes of expressions is greater than 1.5 or not.

Question of Interest

- Considering the huge dimension of tests, we aim to search for tests with large average power.
- More specifically, our goal is to derive a test with *maximum average power* while controlling FDR.
- In addition, we want to allow the null hypothesis to be intervals, such as small range of fold-changes. So we test for H_0^g: $\delta_g \in \Delta_0$ where, $\Delta_0 = \{0\}$ when testing for differential expression; $\Delta_0 = (-\infty, 0]$ when testing for higher expression in the second treatment; $\Delta_0 = \{\delta : |\delta| \leq c\}$ with $c = \log 1.5$ when testing whether fold-changes of expressions is greater than 1.5 or not.

Question of Interest

- Considering the huge dimension of tests, we aim to search for tests with large average power.
- More specifically, our goal is to derive a test with *maximum average power* while controlling FDR.
- In addition, we want to allow the null hypothesis to be intervals, such as small range of fold-changes. So we test for H_0^g: $\delta_g \in \Delta_0$ where,
  $\Delta_0 = \{0\}$ when testing for differential expression;
  $\Delta_0 = (-\infty, 0]$ when testing for higher expression in the second treatment;
  $\Delta_0 = \{\delta : |\delta| \leq c\}$ with $c = \log 1.5$ when testing whether fold-changes of expressions is greater than 1.5 or not.

Question of Interest

- Considering the huge dimension of tests, we aim to search for tests with large average power.
- More specifically, our goal is to derive a test with *maximum average power* while controlling FDR.
- In addition, we want to allow the null hypothesis to be intervals, such as small range of fold-changes. So we test for H_0^g: $\delta_g \in \Delta_0$ where,
  $\Delta_0 = \{0\}$ when testing for differential expression;
  $\Delta_0 = (-\infty, 0]$ when testing for higher expression in the second treatment;
  $\Delta_0 = \{\delta : |\delta| \leq c\}$ with $c = \log 1.5$ when testing whether fold-changes of expressions is greater than 1.5 or not.

Notations

- Data: $\mathbf{X}_g = \{X_{gij} : i = 1, 2; j = 1, \cdots, n_i\} \in \mathcal{X}$.

- Likelihood: $f(\mathbf{X}_g|\lambda_g, \delta_g)$ for the Poisson model.

- Hypotheses: $H_0^g : \delta_g \in \Delta_0$ v.s. $H_1^g : \delta_g \in \Delta_1$.

- Critical Function: $\varphi(\mathbf{X}_g)$ so that the hypothesis H_0^g is rejected if and only if $\varphi(\mathbf{X}_g) = 1$.

Maximum Average Powerful (MAP) Test

- Assuming $(\lambda_g, \delta_g) \sim \pi_1(\lambda, \delta)$ for $\delta_g \in \Delta_1$, $(\lambda_g, \delta_g) \sim \pi_0(\lambda, \delta)$ for genes with $\delta_g \in \Delta_0$, and defining $\pi(\lambda, \delta) = p_0 \pi_0(\lambda, \delta) + (1 - p_0)\pi_1(\lambda, \delta)$, where p_0 is the proportion of genes with $\delta_g \in \Delta_0$, Si and Liu (2013) prove the following theorem.

Theorem

The test that maximizes the average power with FDR controlled at level α is the test that rejects H_0^g when

$$T(\mathbf{X}_g) = \frac{\int_{\mathcal{R}^+} \int_{\Delta_0} f(\mathbf{X}_g | \lambda, \delta) \pi(\lambda, \delta) \mathrm{d}\delta \mathrm{d}\lambda}{\int_{\mathcal{R}^+} \int_{\mathcal{R}} f(\mathbf{X}_g | \lambda, \delta) \pi(\lambda, \delta) \mathrm{d}\delta \mathrm{d}\lambda} \leq c$$

for $g = 1, 2, ..., G$, and the constant c is the critical value so that the multiple testing procedure has FDR controlled at level α.

- This test is called the maximum average power (MAP) test.

Maximum Average Powerful (MAP) Test

- Assuming $(\lambda_g, \delta_g) \sim \pi_1(\lambda, \delta)$ for $\delta_g \in \Delta_1$, $(\lambda_g, \delta_g) \sim \pi_0(\lambda, \delta)$ for genes with $\delta_g \in \Delta_0$, and defining $\pi(\lambda, \delta) = p_0\pi_0(\lambda, \delta) + (1 - p_0)\pi_1(\lambda, \delta)$, where p_0 is the proportion of genes with $\delta_g \in \Delta_0$, Si and Liu (2013) prove the following theorem.

Theorem

The test that maximizes the average power with FDR controlled at level α is the test that rejects H_0^g when

$$T(\mathbf{X}_g) = \frac{\int_{\mathcal{R}^+} \int_{\Delta_0} f(\mathbf{X}_g|\lambda, \delta)\pi(\lambda, \delta)\mathrm{d}\delta\mathrm{d}\lambda}{\int_{\mathcal{R}^+} \int_{\mathcal{R}} f(\mathbf{X}_g|\lambda, \delta)\pi(\lambda, \delta)\mathrm{d}\delta\mathrm{d}\lambda} \leq c$$

for $g = 1, 2, ..., G$, and the constant c is the critical value so that the multiple testing procedure has FDR controlled at level α.

- This test is called the maximum average power (MAP) test.

Maximum Average Powerful (MAP) Test

- Assuming $(\lambda_g, \delta_g) \sim \pi_1(\lambda, \delta)$ for $\delta_g \in \Delta_1$, $(\lambda_g, \delta_g) \sim \pi_0(\lambda, \delta)$ for genes with $\delta_g \in \Delta_0$, and defining $\pi(\lambda, \delta) = p_0\pi_0(\lambda, \delta) + (1 - p_0)\pi_1(\lambda, \delta)$, where p_0 is the proportion of genes with $\delta_g \in \Delta_0$, Si and Liu (2013) prove the following theorem.

Theorem

The test that maximizes the average power with FDR controlled at level α is the test that rejects H_0^g when

$$T(\mathbf{X}_g) = \frac{\int_{\mathcal{R}^+} \int_{\Delta_0} f(\mathbf{X}_g|\lambda, \delta)\pi(\lambda, \delta)\mathrm{d}\delta\mathrm{d}\lambda}{\int_{\mathcal{R}^+} \int_{\mathcal{R}} f(\mathbf{X}_g|\lambda, \delta)\pi(\lambda, \delta)\mathrm{d}\delta\mathrm{d}\lambda} \leq c$$

for $g = 1, 2, ..., G$, and the constant c is the critical value so that the multiple testing procedure has FDR controlled at level α.

- This test is called the maximum average power (MAP) test.

Estimation of FDR Level

- $\pi(\lambda, \delta)$ can be considered as the prior distribution for (λ_g, δ_g), then:

$$T(\mathbf{X}_g) = \mathrm{P}(\delta_g \in \Delta_0 | \mathbf{X}_g)$$

- Estimated FDR for a test with critical function $\varphi(\mathbf{X}_g)$ is:

$$\widehat{\mathrm{FDR}} = \frac{\sum_g T(\mathbf{X}_g)\varphi(\mathbf{X}_g)}{\sum_g \varphi(\mathbf{X}_g)}, \quad (1)$$

Estimation of FDR Level

- $\pi(\lambda, \delta)$ can be considered as the prior distribution for (λ_g, δ_g), then:

$$T(\mathbf{X}_g) = \mathrm{P}(\delta_g \in \Delta_0 | \mathbf{X}_g)$$

- Estimated FDR for a test with critical function $\varphi(\mathbf{X}_g)$ is:

$$\widehat{\mathrm{FDR}} = \frac{\sum_g T(\mathbf{X}_g)\varphi(\mathbf{X}_g)}{\sum_g \varphi(\mathbf{X}_g)}, \quad (1)$$

Estimation of Prior Distribution $\pi(\lambda, \delta)$

- We assume a K-component mixture Gamma-Normal (MGN) distribution for $\pi(\lambda, \delta)$:

$$\sum_{k=1}^{K} q_k G(\lambda|\alpha_k, \beta_k) N(\delta|\mu_k, \sigma_k).$$

- Changing the parameters $\{(q_k, \alpha_k, \beta_k, \mu_k, \sigma_k) : k = 1, 2, \cdots, K\}$ allows ample model flexibility.
- Parameters for the mixture distribution can be estimated by EM algorithm.
- Using the estimated $\pi(\lambda, \delta)$ results in the approximated MAP (AMAP) test.

MAP test for Negative Binomial model

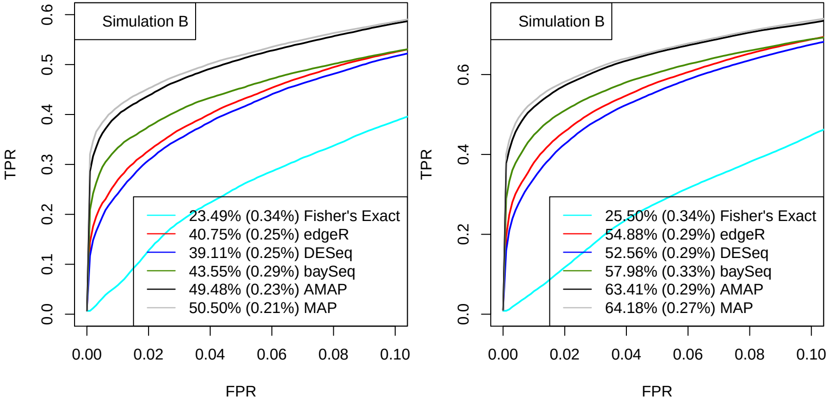
- To avoid expensive computation, we first estimate the dispersion parameter ϕ_g for each gene and then estimate $\pi(\lambda, \delta)$ in the same way as for Poisson model.
- With estimated $\hat{\pi}(\lambda, \delta)$ and $\hat{\phi}_g$, the AMAP statistic under Negative Binomial model is

$$T(\mathbf{X}_g) = \frac{\int_{\mathcal{R}^+} \int_{\Delta_0} f(\mathbf{X}_g | \lambda, \delta, \hat{\phi}_g) \hat{\pi}(\lambda, \delta) \mathrm{d}\delta \mathrm{d}\lambda}{\int_{\mathcal{R}^+} \int_{\mathcal{R}} f(\mathbf{X}_g | \lambda, \delta, \hat{\phi}_g) \hat{\pi}(\lambda, \delta) \mathrm{d}\delta \mathrm{d}\lambda}. \tag{2}$$

- The FDR for the AMAP test based on the NB model can also be estimated by equation (1).

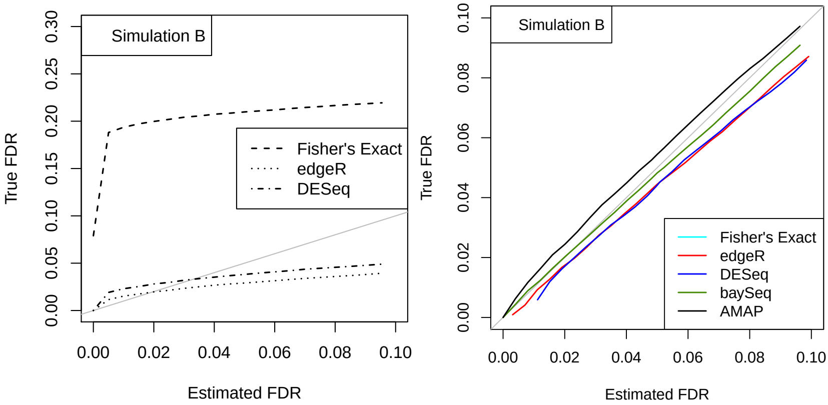
Simulation Study Based on NB

- Left: $H_0^g : \delta_g = 0$, Right: $H_0^g : |\delta_g| \leq \log(1.5)$

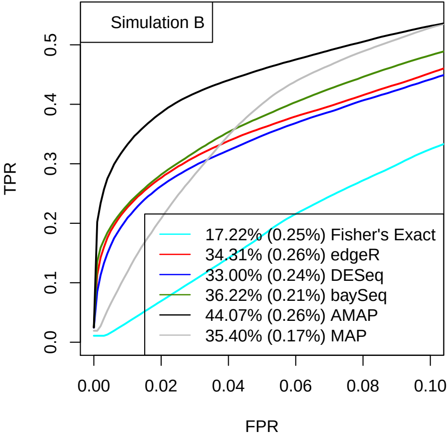


Simulation Study Based on NB

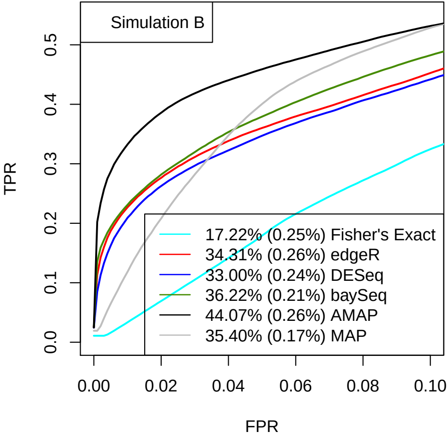
- FDR estimation: Left: BH method, Right: AMAP method



Simulation Study Based on data outliers



Simulation Study Based on data with discrete fold-changes



Motivation of the Model-based Nonparametric Bayesian Approach

- Negative binomial distribution is very popular in modeling gene expression, however, this distribution has no conjugate prior, thus we use Poisson-Gamma Mixture model instead.
- Our main interest is checking hypothesis testing corresponding to fold change between treatment means, so it is very important to decide what prior distribution is used for the fold change for each gene. To provide maximal flexibility, DP modelling framework is applied to the fold change parameters.
- Mixture of a point-mass at 1 and continuous distributions is used as the base distribution for the DP prior for the fold change parameters.

Bayesian Nonparametrics

Basic Bayesian Framework

- y_i's are i.i.d. random samples from an unknown distribution F
- prior $F \sim p(F)$

- if $F = F(\theta)$, $p(F)$ parametric Bayes
- if F is a random distribution, $p(F)$ is nonparametric Bayes. Nonparametric Bayesian models are used to avoid critical dependence on parametric assumptions.
- Most common random distributions are Dirichlet Process, Polya Trees, Bernstein Polynomials, etc.

Our Hierarchical Model

$$\begin{aligned} X_{gij} | \lambda_{gij} &\sim \text{Poisson}(S_{ij}\lambda_{gij}), \\ \lambda_{g1j} | \alpha_g, \beta_g &\sim \text{Gamma}(\alpha_g,\ \beta_g), \\ \lambda_{g2j} | \alpha_g, \beta_g, \rho_g &\sim \text{Gamma}(\alpha_g,\ \beta_g \rho_g) \end{aligned}$$

- ρ_g is the fold change between the two treatment means
- S_{ij} is a normalization factor
- The dispersion parameter for gene g is $\phi_g = 1/\alpha_g$

Our Hierarchical Model Cont.

- Priors for α_g and β_g,

$$\alpha_g \sim \text{Exp}(r), \beta_g \sim \text{Gamma}(a_0, b_0),$$

- the prior distribution for ρ_g is Dirichilet Process

$$\rho_g | G \overset{iid}{\sim} G \text{ and } G \sim DP(M,\ G_0),$$

$$G_0 \sim p_0 \mathbf{1}_{\{1\}} + (1 - p_0)\text{Gamma}(\alpha_0, \beta_0).$$

- r, a_0, b_0 and α_0, β_0 are hyperparameters, set $M = 1$.
- The priors for α_g, β_g and ρ_g are independent.
- At present hyperparameters are assumed to be known and are not assigned hyperpriors.

Sampling from Posterior Distribution

Update λ_{gij}'s , α_g's, β_g's

- Full Conditionals

$$\begin{aligned} p(\lambda_{g1j}|.) &\sim \text{Gamma}(X_{g1j}+\alpha_g,\ \beta_g+S_{1j}) \\ p(\lambda_{g2j}|.) &\sim \text{Gamma}(X_{g2j}+\alpha_g,\ \beta_g\rho_g+S_{2j}) \\ p(\beta_g|.) &\sim \text{Gamma}(\alpha_g(n_1+n_2)+a_0, \sum\lambda_{g1j}+\rho_g\sum\lambda_{g2j}+b_0) \end{aligned}$$

- $p(\alpha_g|.)$ is a logConcave function so we use adaptive rejection sampling method to obtain the posterior samples for α_g.
- We use collapsed Gibbs sampling scheme for $p(\rho_g|.)$.

Test for the Hierarchical Model and FDR Control

- Hypothesis test

$$H_0^g : \rho_g = 1 \text{ v.s. } H_1^g : \rho_g \neq 1$$

- For each g, posterior probability $P(\rho_g = 1|\mathbf{X}_g)$ is estimated by the proportion of the posterior samples for ρ_g that equals to 1.
- Estimated FDR

$$\widehat{\text{FDR}} = \frac{\sum_g P(\rho_g = 1|\mathbf{X}_g)\psi(\mathbf{X}_g)}{\sum_g \psi(\mathbf{X}_g)},$$

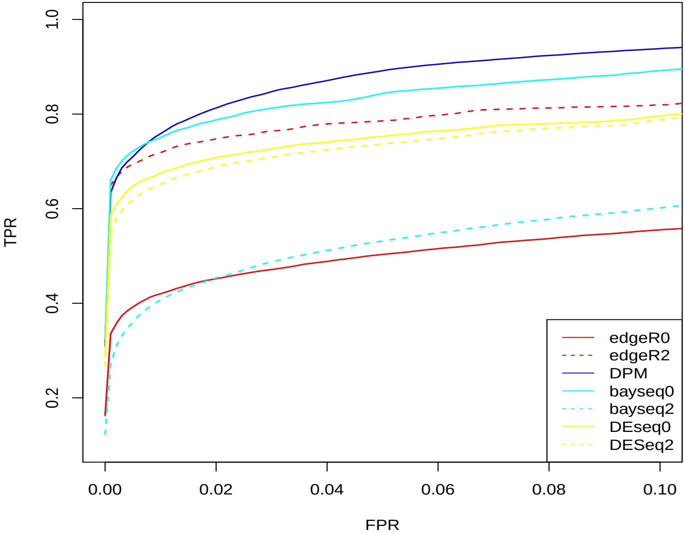
where $\psi(\mathbf{X}_g)$ is an indicator function and $\psi(\mathbf{X}_g) = 1$ iff the null hypothesis H_0^g is rejected.

Simulation Study

10 independent data sets were generated and each data set included 10000 genes with 5 replicates for each of the two treatments.

- α_g and β_g empirically estimated from the maize data set using edgeR where $\alpha_g = \frac{1}{\phi_g}$ and $\beta_g = \frac{\mu_{1g}}{\alpha_g}$.
- 50% genes $\rho_g = 1$, the other 50% genes randomly set 1250 genes of them to be 4, 8, 0.25, 0.125.
- The normalizing factor S_{ij} for all i and j were set to be 1.
- X_{gij} were simulated from Negative binomial distribution with mean parameter $\frac{\alpha_g}{\beta_g \rho_g}$ and the size parameter α_g.

Simulation Result



Concluding Remarks

- Under empirical Bayesian framework, we develop an optimal test that achieves MAP while controlling the FDR.
- The AMAP test performs better than other tests in simulation studies in terms of both the average power and the FDR control.
- For the test based on negative binomial model, there is room to improve the performance of AMAP test by better estimating the dispersion parameter and the prior distribution for the means.
- We have built an R package, `AMAP.seq`, that is available from CRAN, to implement the proposed test.
- We have applied the method to derive optimal test in other cases, such as testing for alternative splicing events using RNA-seq data.

Acknowledgement

- Yaqing Si
- National Science Foundation Grant Awards 0701736 and 1127017

References

- Muller, P. and Quintana F.A. (2007), Nonparametric Bayesian Data Analysis,Statistical Science 19(1), 95-110.
- Do, K.A., Muller, P. and Tang, F.(2005), A Bayesian mixture model for differential gene expression, Applied Statistics, 54, 627-644.
- Y. Si and P. Liu (2013), An Optimal Test with Maximum Average Power While Controlling FDR with Application to RNA-seq Data, accepted by Biometrics

Collapsed Gibbs Sampling Scheme

Update λ_{gij}'s , α_g's, β_g's

- Full Conditionals

$$
\begin{aligned}
p(\lambda_{g1j}|.) &\sim \text{Gamma}(Y_{g1j} + \alpha_g,\ \beta_g + S_{1j}) \\
p(\lambda_{g2j}|.) &\sim \text{Gamma}(Y_{g2j} + \alpha_g,\ \beta_g \rho_g + S_{2j}) \\
p(\beta_g|.) &\sim \text{Gamma}(\alpha_g(n_1 + n_2) + a_0, \sum \lambda_{g1j} + \rho_g \sum \lambda_{g2j} + b_0)
\end{aligned}
$$

- $p(\alpha_g|.)$ is a logConcave function so we use adaptive rejection sampling method to obtain the posterior samples for α_g.

Collapsed Gibbs Sampling Scheme

Reparameterization:

- K : number of distinct values (clusters) in the vector $(\rho_1, \dots, \rho_G)^T$ (assume the distinct values are $\rho_1^*, \dots, \rho_K^*$)
- $\xi = (\xi_1, \dots, \xi_G)^T$ denotes the configuration indicators, defined by $\xi_g = k$ if and only if $\rho_g = \rho_k^* = \rho_{\xi_g}^*$.

The model could be rewritten as

$$\begin{array}{rcl} Y_{gij}|\lambda_{gij} & \sim & Poisson(S_{ij}\lambda_{gij}), \\ \lambda_{g1j}|\alpha_g, \beta_g & \sim & Gamma(\alpha_g,\ \beta_g), \\ \lambda_{g2j}|\alpha_g, \beta_g, \{\rho_k^*\} & \sim & Gamma(\alpha_g,\ \beta_g \rho_{\xi_g}^*), \end{array}$$

with $\rho_k^* \overset{iid}{\sim} G_0$, and $(\xi_1, ..., \xi_G)|M \sim \mathrm{CRP}(M)$, where CRP is Chinese Restaurant Process.

Collapsed Gibbs Sampling Scheme

Update configuration vector $(\xi_1, \ldots, \xi_G)$

- If $\xi = \xi_l$ for some $l \neq g$:

$$\begin{aligned} P(\xi_g = \xi | \xi_{-g}, rest) &= c \quad n_{\xi}^{-g} \Pi_{j=1}^{n_2} p(\lambda_{g2j} | \alpha_g, \beta_g, \rho_{\xi}^{*}) \\ &= c \quad n_{\xi}^{-g} \Pi_{j=1}^{n_2} \frac{\beta_g^{\alpha_g} (\rho_{\xi}^{*})^{\alpha_g}}{\Gamma(\alpha_g)} \lambda_{g2j}^{\alpha_g - 1} \exp(-\beta_g \rho_{\xi}^{*} \lambda_{g2j}) \end{aligned}$$

- otherwise

$$P(\xi_g \neq \xi_l \text{ for all } l \neq g | \xi_{-g}, rest) = c \int \Pi_{j=1}^{n_2} p(\lambda_{g2j} | \alpha_g, \beta_g, \rho) G_0(\rho) d\rho$$

Collapsed Gibbs Sampling Scheme

Update $(\rho_1^*, \ldots, \rho_K^*)$

$$p(\rho_k^*|.) \propto C_1 1_{\{\rho_k^*=1\}} + C_2 \text{Gamma}(a_k, b_k)$$

where

$$a_k = n_2 \sum_{\{g:\xi_g=k\}} \alpha_g + \alpha_0, \;\; b_k = \beta_0 + \sum_{\{g:\xi_g=k\}} \sum_{j=1}^{n_2} \beta_g \lambda_{g2j}$$

,

$$C_1 = p_0 \exp\{-(\sum_{\{g:\xi_g=k\}} \sum_{j=1}^{n_2} \beta_g \lambda_{g2j})\},$$

$$C_2 = (1-p_0)\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \frac{\Gamma(n_2 \sum_{\{g:\xi_g=k\}} \alpha_g + \alpha_0)}{(\beta_0 + \sum_{\{g:\xi_g=k\}} \sum_{j=1}^{n_2} \beta_g \lambda_{g2j})^{n_2 \sum_{\{g:\xi_g=k\}} \alpha_g + \alpha_0}}$$