Fully Bayesian analysis of RNAseq data for gene expression heterosis detection

Dr. Jarad Niemi

Iowa State University

November 20, 2017

Dr. Dan Nettleton, Dr. Will Landau, Eric Mittman, and Dr. Ignacio Alvarez-Castro

This research was supported by National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health and the joint National Science Foundation / NIGMS Mathematical Biology Program under award number R01GM109458. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

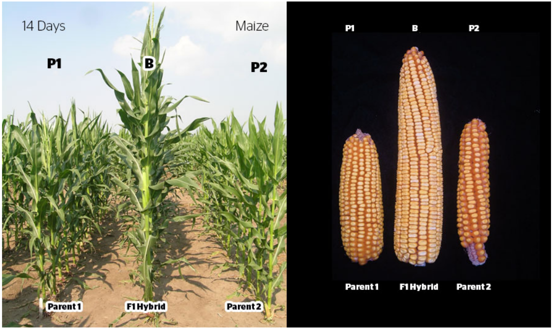
Outline

- Background
  - Heterosis
  - RNAseq data
- Modeling
  - Composite null hypothesis
  - Hierarchical overdispersed count regression model
- Inference
  - Empirical Bayes
  - Fully Bayes on graphics processing units
    - Memory transfers
    - Random number generation
- Simulation studies
  - Credible interval coverage
  - Heterosis detection via ROC curves
- Real data analysis

Heterosis

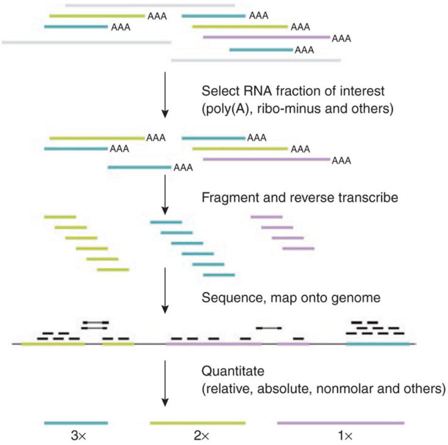
Definition

Heterosis, or hybrid vigor, is the enhancement of the phenotype of hybrid progeny relative to their inbred parents.



(http://www2.iastate.edu/~nscentral/news/06/may/vigor.shtml modified by Will Landau)

RNA fragmentation, sequencing, and alignment



(Pepke, Wold, and Mortazavi (2009) `http://www.nature.com/nmeth/journal/v6/n11s/fig_tab/nmeth.1371_F5.html`)

RNAseq data

Gene ID	B73				Mo17				B73 x Mo17				Mo17 x B73			
GRMZM2G107839	26	17	32	35	30	32	41	43	63	44	116	101	30	31	69	47
GRMZM5G899787	62	57	38	33	91	78	66	69	58	84	42	43	74	70	53	51
GRMZM5G899800	150	238	12	6	198	392	11	15	187	433	8	10	414	291	11	13
GRMZM2G301485	24	12	29	32	20	14	32	46	5	3	6	6	2	3	3	7
GRMZM5G899836	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

(Will Landau)

- Low parent heterosis (LPH): expression in hybrid is lower than both parents
- High parent heterosis (HPH): expression in hybrid is higher than both parents

Hypotheses

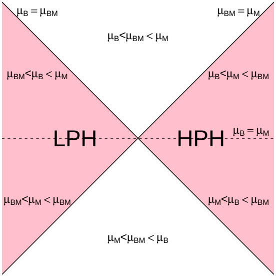
For a given gene, let

- μ_B : mean expression for B73 inbred parent,
- μ_M : mean expression for Mo17 inbred parent,
- μ_{BM}: mean expression for B73×Mo17 F1 hybrid, and
- μ_{MB}: mean expression for Mo17×B73 F1 hybrid.

We are interested in comparing hypotheses of the form

$$H_0 : \mu_{min} < \mu_{BM} < \mu_{max}, \qquad H_{LPH} : \mu_{BM} < \mu_{min}, \qquad H_{HPH} : \mu_{max} < \mu_{BM}$$

where $\mu_{min} = \min(\mu_B, \mu_M)$ and $\mu_{max} = \max(\mu_B, \mu_M)$.



Posterior hypothesis probabilities

If we had a posterior distribution for the mean expression levels, i.e.

$$p(\mu_B, \mu_M, \mu_{BM}, \mu_{MB}|y)$$

where y represents our data, then we could calculate the relevant hypothesis probabilities, i.e.

$$\begin{aligned} P(H_0 \quad |y) &= P(\mu_{min} < \mu_{BM} < \mu_{max}|y) \\ P(H_{LPH}|y) &= P(\mu_{BM} < \mu_{min}|y) \\ P(H_{HPH}|y) &= P(\mu_{max} < \mu_{BM}|y). \end{aligned}$$

Hierarchical modeling *partially pools* the information across genes and thereby provides a data-based multiple comparison adjustment by

- shrinking estimates toward a grand mean (or zero) based on the variability inherent in the data and
- reducing posterior uncertainty by borrowing information across genes.

(Gelman, Hill, and Yajima (2012))

Overdispersed count regression model

Let

- g $(g = 1, \ldots, G)$ identify the gene,
- n $(n = 1, \ldots, N)$ identify the sample,
- y be the $G \times N$ matrix of RNAseq counts and
- X be the $N \times L$ model matrix that connects the N samples to the varieties, blocking factors, etc.

We assume

$$y_{gn} \stackrel{ind}{\sim} \text{Po}\left(e^{h_n + \varepsilon_{gn} + x_n' \beta_g}\right)$$

where

- h_n are *normalization factors*,
- $\varepsilon_{gn} \stackrel{ind}{\sim} N(0, \gamma_g)$ allow for gene-specific overdispersion,
- x_n is the n^{th} row of X, and
- β_g is a vector of length L that account for effects on gene expression of variables of interest.

Hierarchical model

Recall

$$y_{gn} \stackrel{ind}{\sim} \text{Po}\left(e^{h_n+\varepsilon_{gn}+x_n'\beta_g}\right) \qquad \text{and} \qquad \varepsilon_{gn} \stackrel{ind}{\sim} N(0,\gamma_g).$$

We construct a hierarchical model for both γ_g and β_g to borrow information across genes. Specifically, we assume

$$1/\gamma_g \stackrel{ind}{\sim} \text{Ga}\left(\nu/2, \nu\tau/2\right)$$

such that $E[1/\gamma_g] = 1/\tau$ and $CoV[1/\gamma_g] = \sqrt{2/\nu}$ and

$$\beta_{g\ell} \stackrel{ind}{\sim} N(\theta_\ell, \sigma_\ell^2)$$

for $\ell = 1, \ldots, L$.

Model matrix for our heterosis experiment

Experimental design: 4 varieties, 2 flow cells, 2 replicates per variety per flow cell

$$X = \left(\begin{bmatrix} 1 & 1 & -1 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix} \otimes J_{(N/4)\times 1} \qquad J_{(N/4)\times 1} \otimes \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \right)$$

where $\otimes$ denotes the Kronecker product and $J_{m\times n}$ is the m by n matrix with all entries equal to 1.

Interpretations of the gene-specific parameters (dropping the g subscript) are

- β_{g1} is the parental mean
- β_{g2} is the half difference of hybrid mean vs M
- β_{g3} is the half difference of hybrid mean vs B
- β_{g4} is the half difference between hybrids
- β_{g5} is the flow cell block effect

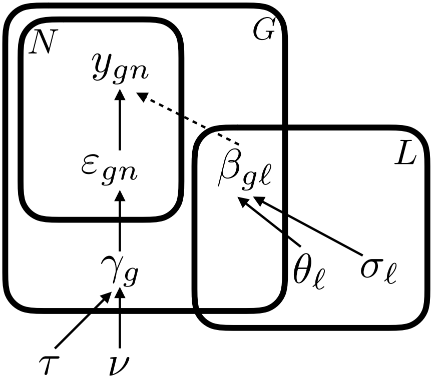
Heterosis hypotheses

Heterosis	With log-scale group means	With $\beta_{g\ell}$ parameters
high-parent BM	$\mu_{g,\text{BM}} > \max\left(\mu_{g,\text{B}}, \mu_{g,\text{M}}\right)$	$2\beta_{g2} + \beta_{g4},\quad 2\beta_{g3} + \beta_{g4} > 0$
low-parent BM	$\mu_{g,\text{BM}} < \min\left(\mu_{g,\text{B}}, \mu_{g,\text{M}}\right)$	$-2\beta_{g2} - \beta_{g4}, -2\beta_{g3} - \beta_{g4} > 0$
high-parent MB	$\mu_{g,\text{MB}} > \max\left(\mu_{g,\text{B}}, \mu_{g,\text{M}}\right)$	$2\beta_{g2} - \beta_{g4},\quad 2\beta_{g3} - \beta_{g4} > 0$
low-parent MB	$\mu_{g,\text{MB}} < \min\left(\mu_{g,\text{B}}, \mu_{g,\text{M}}\right)$	$-2\beta_{g2} + \beta_{g4},\quad 2\beta_{g3} + \beta_{g4} > 0$
high-parent mean	$\mu_{g,\text{BM}} + \mu_{g,\text{MB}} > 2\max\left(\mu_{g,\text{B}}, \mu_{g,\text{M}}\right)$	$\beta_{g2},\quad \beta_{g3} > 0$
low-parent mean	$\mu_{g,\text{BM}} + \mu_{g,\text{MB}} < 2\min\left(\mu_{g,\text{B}}, \mu_{g,\text{M}}\right)$	$-\beta_{g2}, -\beta_{g3} > 0$

All hypothesis regions are intersections of linear combination events, but we can also accommodate unions of contrast events via

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Directed acyclic graphical model



$$p(\varepsilon,\beta,\gamma,\theta,\sigma,\tau,\nu|y) =$$

$$p(\varepsilon,\beta,\gamma|\tau,\nu,\theta,\sigma,y) \times p(\tau,\nu,\theta,\sigma|y) \propto$$

$$\prod_{g=1}^{G}\left\{\left[\prod_{n=1}^{N} p(y_{gn}|\beta_g,\varepsilon_{gn})p(\varepsilon_{gn}|\gamma_g)\right]\left[\prod_{\ell=1}^{L} p(\beta_{g\ell}|\theta_\ell,\sigma_\ell)p(\theta_\ell)p(\sigma_\ell)\right] p(\gamma_g|\tau,\nu)\right\} p(\tau)p(\nu)$$

(Will Landau)

where $G \approx 40\,000$, $N = 16$, and $L = 5$ in our application and thus we have

- $G \times N + G + 2 + G \times L + 2 \times L \approx 800\,000$ parameters
- and $G \times N \approx 640\,000$ observations.

Empirical Bayes

First we tried to fit this model using black-box Bayesian software, i.e. JAGS and Stan, but it appears computationally intractable using these platforms.

Niemi, Mittman, Landau, and Nettleton (2015) used an empirical Bayes approach by

1. using moment matching techniques on independent gene-specific parameter estimates to obtain hyperparameter estimates
2. and then parallelizing the MCMC across genes, i.e.

$$p(\varepsilon,\beta,\gamma|\hat\tau,\hat\nu,\hat\theta,\hat\sigma,y)\propto\prod_{g=1}^{G}\left\{\left[\prod_{n=1}^{N}p(y_{gn}|\beta_g,\varepsilon_{gn})p(\varepsilon_{gn}|\gamma_g)\right]\left[\prod_{\ell=1}^{L}p(\beta_{g\ell}|\hat\theta_\ell,\hat\sigma_\ell)\right]p(\gamma_g|\hat\tau,\hat\nu)\right\}$$

```
if (require(doMC)) {
  registerDoMC()
} else {
  parallel=FALSE
}

analysis = adply(d,
                 1,
                 function(x) single_gene_analysis(x),
                 .id = 'gene',
                 .parallel = parallel,
                 .paropts = list(.export=c('single_gene_analysis','model','hyperparameters'), .packages='rstan'))
```

Took about 10 hours to run.

Priors

All priors are constructed to be vague, proper, and (if possible) conditionally conjugate. There are $2(L+1)$ hyperparameters and we assign the following independent priors

$$\begin{array}{llll} \tau & \sim Ga(a,b) & & \text{conditionally conjugate} \\ \nu & \sim Unif(0,d) & & \\ \theta_\ell & \stackrel{ind}{\sim} N(0,c_\ell^2) & & \text{conditionally conjugate} \\ \sigma_\ell & \stackrel{ind}{\sim} Unif(0,s_\ell) & & \text{(Gelman (2006))} \end{array}$$

As we'll see, posterior distributions for these parameters are extremely tight relative to their priors.

Constructing a Gibbs sampler

Conditional independence within a step:

$$
\begin{array}{rl}
p(\varepsilon|\ldots) & \propto \prod_{g=1}^{G}\prod_{n=1}^{N} Po\left(y_{gn}|e^{h_n+\varepsilon_{gn}+x_n'\beta_g}\right) N(\varepsilon_{gn}|0,\gamma_g) \\
p(\gamma|\ldots) & \propto \prod_{g=1}^{G}\prod_{n=1}^{N} N(\varepsilon_{gn}|0,\gamma_g) IG(\gamma_g|\nu/2,\nu\tau/2) \\
p(\beta_\ell|\ldots) & \propto \prod_{g=1}^{G}\prod_{n=1}^{N} Po\left(y_{gn}|e^{h_n+\varepsilon_{gn}+x_n'\beta_g}\right) N(\beta_{g\ell}|\theta_\ell,\sigma_\ell^2)
\end{array}
$$

Sufficient “statistics”:

$$
\begin{array}{rlrl}
p(\tau|\ldots) & \sim Ga(\tau|a',b') & (a',b') & = f_\tau(\gamma,\nu,a,b) \\
p(\nu|\ldots) & \sim p(\nu|d')\mathrm{I}(0<\nu<d) & d' & = f_\nu(\gamma,\tau,d) \\
p(\theta_\ell|\ldots) & \sim N(\theta_\ell|m_\ell',C_\ell') & (m_\ell',C_\ell') & = f_{\theta_\ell}(\beta_\ell,\sigma_\ell,c_\ell^2) \\
p(\sigma_\ell^2|\ldots) & \sim IG(e',f')\mathrm{I}(0<\sigma_\ell^2<s_\ell^2) & (e',f') & = f_{\sigma_\ell}(\beta_\ell,\theta_\ell)
\end{array}
$$

where the functions calculate means, variances, products, etc. over G terms.

Parallel hardware platforms

(All values are orders of magnitude)

	Platform		
	Multicore	Accelerator (GPU)	Cluster
# of nodes	10	1000	100+
node speed (GHz)	1	1	1
memory (GB)	10	10	10/node
node comm. speed (GB/s)	10	10	1
data transfer speed (GB/s)	N/A	1	1

The combination of

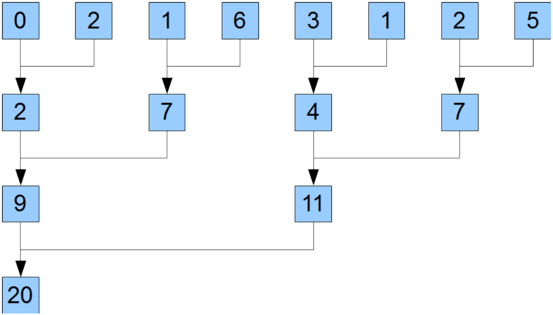
- number of nodes and
- node communication speed

make accelerators a good choice for parallelizable MCMC. (Lee et. al. (2010))

Parallelization translated to a GPU

If there are G nodes, then

- Conditional independence $\to$ embarrassingly parallel - possible speedup is G
- Calculate sufficient "statistics" $\to$ reduction - possible speedup is $[G-1]/\log_2(G) \approx 2,500$



(https://scs.senecac.on.ca/~gpu610/pages/images/parallel_reduction.png)

GPU computing algorithm constraints

Constraint	Solution
memory coalescence	set up proper data structures
only uniform and normal RNGs	step out slice sampler
data transfer speed	thinning
	return draws for a small subset of parameters hyperparameters random set of gene-specific parameters
	calculate running sums for convergence diagnostics normal-based credible intervals hypothesis probabilities

Convergence diagnostics

For each parameter θ over the M iterations of chain c, calculate

$$M\overline{\theta}_c = \sum_{m=1}^{M} \theta_c^{(m)} \qquad \text{and} \qquad M\overline{\theta^2}_c = \sum_{m=1}^{M} \left(\theta_c^{(m)}\right)^2.$$

using a numerically stable one-pass (or online) algorithm.

Compute the Gelman-Rubin convergence diagnostic amongst C chains using

$$\widehat{R} = \sqrt{1 + \frac{1}{M}\left(\frac{B}{W} - 1\right)}$$

where

$$B = \frac{M}{C-1}\sum_{c=1}^{C}\left(\overline{\theta}_c - \overline{\theta}\right)^2, W = \frac{1}{C}\sum_{c=1}^{C} S_c^2,$$

$$\overline{\theta} = \frac{1}{C}\sum_{c=1}^{C}\overline{\theta}_c, \text{ and } S_c^2 = \frac{M}{M-1}\left[\overline{\theta^2}_c - \overline{\theta}_c^2\right] \approx \overline{\theta^2}_c - \overline{\theta}_c^2.$$

Normal-based credible intervals

For the collection of parameters ψ and under regularity conditions, we have

$$p_N(\psi|y_N) \xrightarrow{d} N\left(\psi_0, [\mathrm{I}_N(\psi_0)]^{-1}\right)$$

as $N \to \infty$ where ψ_0 is the true value and $\mathrm{I}_N(\psi_0)$ is the Fisher information.

For any scalar parameter θ, we have

$$\theta|y \dot{\sim} N\left(\overline{\theta}, \overline{\theta^2} - \overline{\theta}^2\right)$$

and can construct normal-based credible intervals with

$$\overline{\theta} \pm z_{\alpha/2}\sqrt{\overline{\theta^2} - \overline{\theta}^2}$$

where $P(Z > z_\alpha) = \alpha$ and Z is a standard normal distribution.

Estimating hypothesis probabilities

Recall we are interested in estimating probabilities similar to

$$P(\text{high parent heterosis for the B73}\times\text{Mo17 hybrid}|y)$$

$$= P\left(2\beta_{g2} + \beta_{g4} > 0 \text{ and } 2\beta_{g3} + \beta_{g4} > 0 |\, y\right)$$

$$\approx \frac{1}{M}\sum_{m=1}^{M} \mathrm{I}\left(2\beta_{g2}^{(m)} + \beta_{g4}^{(m)} > 0\right) \mathrm{I}\left(2\beta_{g3}^{(m)} + \beta_{g4}^{(m)} > 0\right)$$

We can use the running sums to keep track of this sum.

Implementation

The computation for this hierarchical overdispersed count regression model is provided in the following three R packages at `https://github.com/wlandau/`:

- `fbseq`: user interface
- `fbseqOpenMP`: multithreaded backend
- `fbseqCUDA`: NVIDIA GPU backend

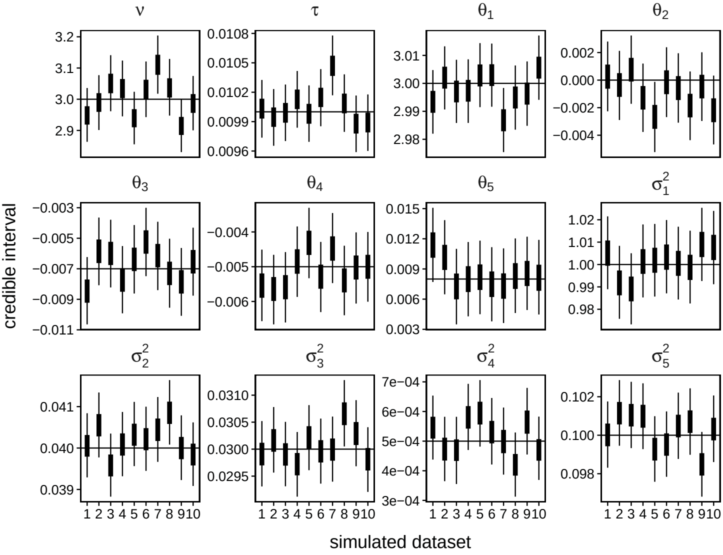
```
library(fbseq)
data(paschold) # Paschold et. al. (2012) data

paschold@contrasts[[5]]


## beta_1 beta_2 beta_3 beta_4 beta_5
##      0      2      0      1      0


paschold@contrasts[[6]]


## beta_1 beta_2 beta_3 beta_4 beta_5
##      0      0      2      1      0


paschold@propositions$`high-parent_B73xMo17`


## high-parent_B73xMo17_1 high-parent_B73xMo17_2
##                      5                      6
```

```
configs    = Configs(burnin = 10, iterations = 10, thin = 1)
chain      = Chain(paschold, configs)
chain_list = fbseq(chain, backend = "CUDA")
```

Simulation studies

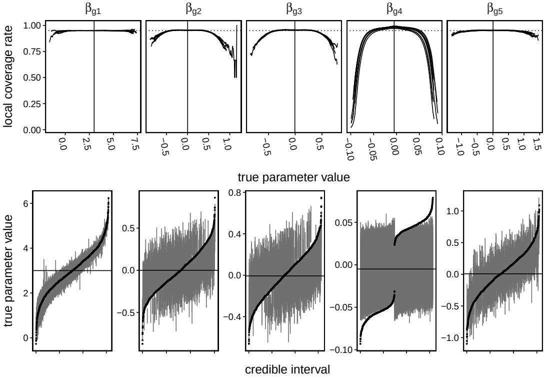
- Simulation model
  - Model: hierarchical count regression model with data-based hyperparameters
  - Simple: count regression model with sparsity in gene-specific parameters
  - edgeR: negative binomial model with data-based gene-specific parameters
- Inference
  - edgeR: non-hierarchical except for overdispersion, negative binomial model
  - fully Bayes: Bayesian analysis with hierarchical count regression model
  - eBayes (Means): empirical Bayesian analysis with hierarchical count regression model with hyperparameter estimated from posterior means of the fully Bayes approach
  - eBayes (Oracle): empirical Bayesian analysis with hierarchical count regression model with true values for the hyperparameters
- Data size
  - $G = 30\,000$
  - $N = 16$ and $N = 32$

Hyperparameter coverage



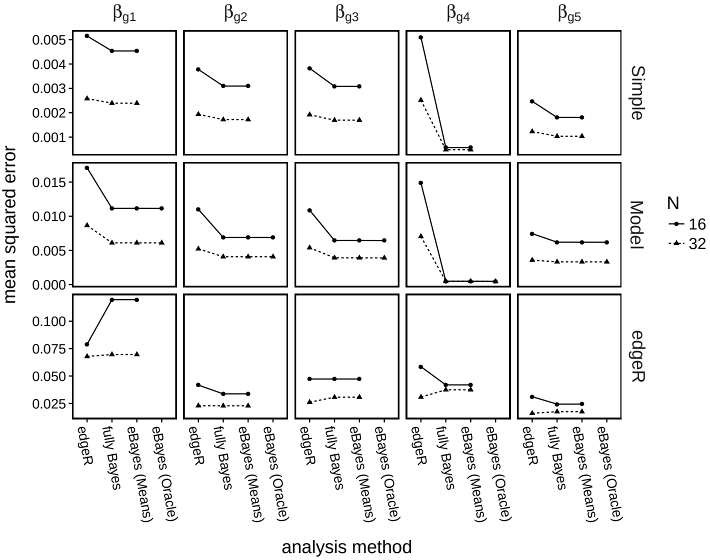
Gene-specific parameter coverage

Overall, we had approximately 95% coverage of 95% credible intervals, but

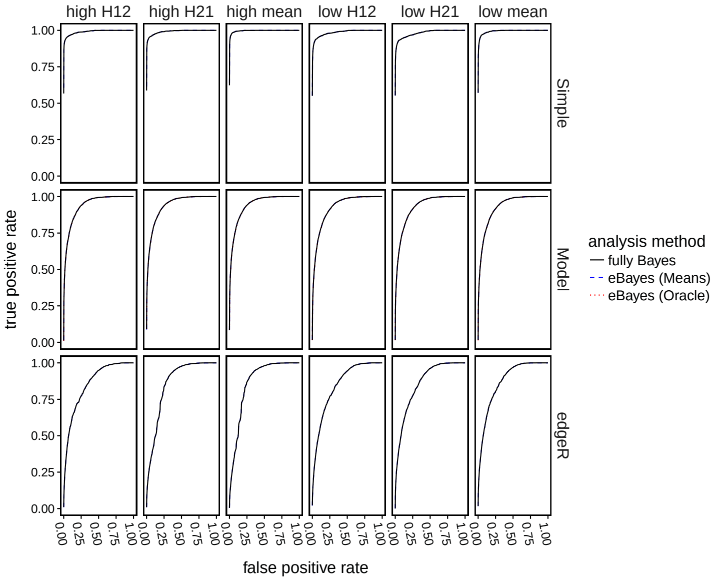


where the lower plot indicates that when the intervals miss, we are overshrinking.

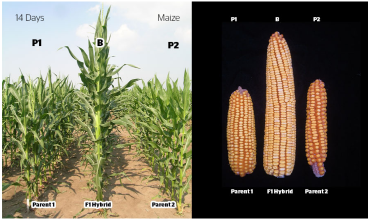
Mean squared error



ROC curves for heterosis detection

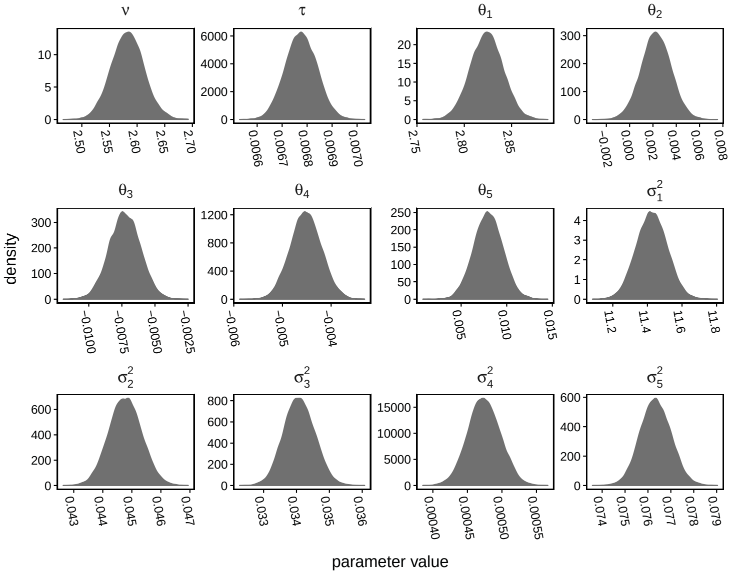


Analysis of Paschold et. al. (2012) data



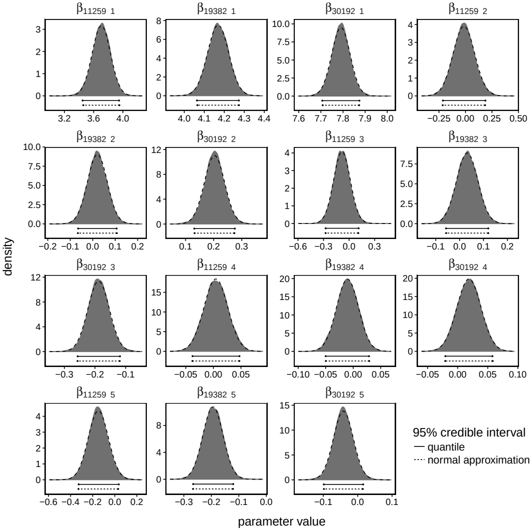
- $N = 16$ with 4 replicates/variety on 2 plates
  - varieties: B73, Mo17, B73×Mo17, Mo17× B73
- $G = 39\,656$
- 21% of genes have mean counts less than 1
- 39% have mean counts less than 10
- count: median 37, mean 260, and max 38 010

Hyperparameter posterior



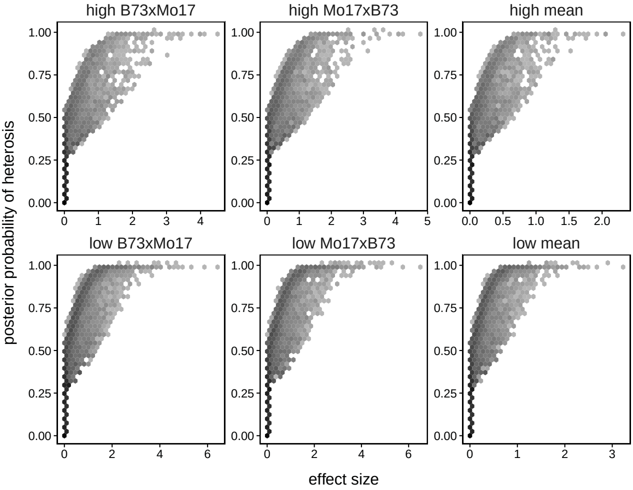
Gene-specific parameter posteriors

Compare posterior distribution (gray area) to normal-based approximation (black dashed line).

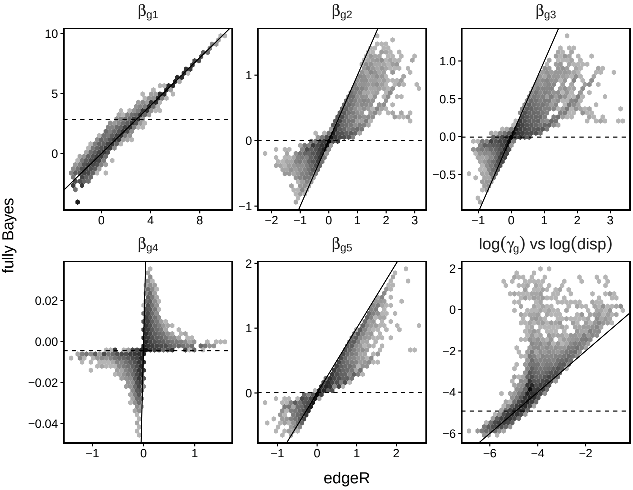


Smokestack plot

Effect size for HPH BM is the maximum of 0 and $\min(2\beta_{g2} + \beta_{g4}, 2\beta_{g3} + \beta_{g4})/\sqrt{\gamma_g}$.



Shrinkage



Summary

- Introduced a hierarchical overdispersed count regression model and
- a GPU-implementation of a fully Bayesian analysis.

This slides are available

- `https://github.com/jarad/UCI2017`
- `http://www.jarad.me/research/presentations.html`

Thank you!

Ongoing work

- Alternative hierarchical distributions via scale mixtures of normals:
  - t
  - Laplace
  - Horseshoe
- Semi-parametric Bayesian model
  - Gene-specific parameters arise from an unknown distribution $\mathcal{F}$
  - assume Dirichlet process prior on $\mathcal{F}$
- Allele specific expression
  - Some counts can be tracked by their origin: B vs M
  - How does allele specific expression relate to heterosis or differences between hybrids

Scale-mixture of normals

Recall the model

$$y_{gn} \stackrel{ind}{\sim} \text{Po}\left(e^{h_n+\varepsilon_{gn}+x_n'\beta_g}\right)$$

with hierarchical distribution

$$\beta_{g\ell}|\xi_{g\ell} \stackrel{ind}{\sim} N(\theta_\ell, \xi_{g\ell}\sigma_\ell^2)$$

where the distribution of $\xi_{g\ell}$ will determine the marginal distribution for $\beta_{g\ell}$ via

$$p(\beta_{g\ell}) = \int N(\beta_{g\ell}|\theta_\ell, \xi_{g\ell}\sigma_\ell^2)p(\xi_{g\ell})d\xi_{g\ell}.$$

For example,

- $\xi_{g\ell} = 1 \implies \beta_{g\ell} \stackrel{ind}{\sim} N(\theta_\ell, \sigma_\ell^2)$
- $\xi_{g\ell} \stackrel{ind}{\sim} IG(q_\ell, r_\ell) \implies \beta_{g\ell} \stackrel{ind}{\sim} t_{2q_\ell}(\theta_\ell, \sigma_\ell^2 r_\ell/q_\ell)$
- $\xi_{g\ell} \stackrel{ind}{\sim} Exp(k_\ell) \implies \beta_{g\ell} \stackrel{ind}{\sim} Laplace(\theta_\ell, \sigma_\ell^2/2k_\ell)$ (Park and Casella (2008), Hans (2009))
- $\xi_{g\ell} \stackrel{ind}{\sim} Ca^+(0,1) \implies \beta_{g\ell} \stackrel{ind}{\sim} Horseshoe(\theta_\ell, \sigma_\ell)$ (Carvalho, Polson, and Scott 2010)

These are all implemented in `fbseq`.

Dirichlet process prior

Recall the model

$$y_{gn} \stackrel{ind}{\sim} \text{Po}\left(e^{h_n+\varepsilon_{gn}+x_n'\beta_g}\right) \qquad \varepsilon_{gn} \stackrel{ind}{\sim} N(0,\gamma_g)$$

Since G is so large, we should be able to learn the distribution of the gene-specific parameters. One possibility is

$$\theta_g = \left(\begin{array}{c} \beta_g' \\ \log(\gamma_g) \end{array}\right) \stackrel{ind}{\sim} \mathcal{F}$$

for some unknown distribution $\mathcal{F}$. Then we can place a prior on $\mathcal{F}$, e.g.

$$\mathcal{F} \sim DP(aF_0),$$

i.e. a Dirichlet process prior with concentration parameter $a > 0$ and base measure F_0 on $\mathbb{R}^{L+1}$. This has been done for differential expression in RNAseq data. (Liu, Wang, and Liu (2015))

Stick-breaking representation

A stick-breaking representation of $\mathcal{F}$ is the probability mass function

$$p(\theta) = \sum_{k=1}^{\infty} \pi_k \delta_{\tilde{\theta}_k}(\theta)$$

where $\tilde{\theta}_k \sim F_0$, $\pi_k = \pi'_k \prod_{i=1}^{k-1}(1-\pi'_i)$, and $\pi'_k \stackrel{ind}{\sim} Be(1,\alpha)$.

We can approximate the infinite sum with a finite sum of K components with $\pi_K = 1 - \sum_{k=1}^{K-1} \pi_k$ and introduce labels $\zeta_g \in \{1,\ldots,K\}$ that indicating component membership for gene g.

The additional Gibbs sampling steps are

- $p(\pi|\ldots)$ which has a conjugate update,
- $p(\zeta|\ldots)$ which is conditionally independent across genes and therefore embarrassingly parallel and requires a parallel scan, i.e. cumulative sum, and
- $p(\tilde{\theta}|\ldots)$ which is
  - conditionally independent across the K components and
  - requires sufficient "statistics" of the genes in each component. (Suchard et. al. 2008)

Allele specific expression

In addition to the total count from a particular gene, sometimes we are able to identify the origin, i.e. allele B or M. Scientists would be interested in detecting alleles of genes that have a ratio of expression in the hybrid that differs from

- one and/or
- the ratio of parental expression.

From a modeling perspective, this brings up at least two questions:

- how do we construct a model matrix for arbitrary scenarios (Lithio and Nettleton 2015) and
- how do we incorporate the idea of random effects, i.e. create non-independence in the samples.

References

- Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* 97(2)" 465–480
- Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis*, 1(3), 515–534.
- Gelman, A., Hill, J., and Yajima, M. (2012) Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, 5(2), 189–211.
- Hans, C. (2009). Bayesian lasso regression. Biometrika, 96(4), 835-845.
- Lee, A., Yau, C., Giles, M. B., Doucet, A., and Holmes, C. C. (2010). On the utility of graphics cards to perform massively parallel simulation of advanced Monte Carlo methods. *Journal of computational and graphical statistics*, 19(4), 769-789.
- Lithio, A., and Nettleton, D. (2015). Hierarchical modeling and differential expression analysis for rna-seq experiments with inbred and hybrid genotypes. *Journal of Agricultural, Biological, and Environmental Statistics*, 20(4), 598-613.
- Liu, Fangfang, Chong Wang, and Peng Liu. (2015) A Semi-parametric Bayesian Approach for Differential Expression Analysis of RNA-seq Data. *Journal of Agricultural, Biological, and Environmental Statistics* 20, no. 4 : 555–576.
- Park, T., and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103(482), 681-686.
- Paschold, A., Jia, Y., Marcon, C., Lund, S., Larson, N.B., Yeh, C.T., Ossowski, S., Lanz, C., Nettleton, D., Schnable, P.S. and Hochholdinger, F., (2012) Complementation contributes to transcriptome complexity in maize (Zea mays L.) hybrids relative to their inbred parents. *Genome research*, 22(12), pp.2445–2454.
- Suchard, M. A., Wang, Q., Chan, C., Frelinger, J., Cron, A., and West, M. (2010). Understanding GPU programming for statistical computation: Studies in massively parallel massive mixtures. *Journal of Computational and Graphical Statistics*, 19(2), 419-438.

Markov chain Monte Carlo integration

Consider approximating an integral via it's Markov chain Monte Carlo (MCMC) estimate, i.e.

$$E_{\theta|y}[h(\theta)|y] = \int_\Theta h(\theta)p(\theta|y)d\theta \quad \text{and} \quad \hat{h}_T = \frac{1}{T}\sum_{t=1}^{(t)} h\left(\theta^{(t)}\right).$$

where $\theta^{(t)}$ is the t^{th} iteration from the MCMC. Under regularity conditions,

- SLLN: $\hat{h}_T \stackrel{a.s.}{\to} E[h(\theta)|y]$ as $T \to \infty$.
- CLT: under stronger regularity conditions,

$$\hat{h}_T \stackrel{d}{\to} N\left(E[h(\theta)|y], \sigma^2/T\right)$$

where

$$\sigma^2 = Var[h(\theta)|y]\left(1 + 2\sum_{k=1}^{\infty}\rho_k\right)$$

where ρ_k is the k^{th} autocorrelation of the $h(\theta)$ values.