The heterosis problem: a comparison of Eric's method with `edgeR`, `baySeq`, and `ShrinkBayes`

Will Landau, Eric Mittman

Iowa State University

November 15, 2014

Outline

Simulate heterosis data with known heterosis genes

		Parent (1)				Parent (2)				Hybrid (3)				Truth
HPH	Feature 1	3	4	2	1	0	0	1	0	700	900	825	860	1
HPH	Feature 2	0	1	1	0	2	7	5	18	50	501	400	90	1
	Feature 3	100	225	0	15	300	106	200	400	70	279	100	123	0
LPH	Feature 4	893	400	760	901	1000	513	760	580	5	5	6	7	1
	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	Feature 25000	10	13	6	4	902	912	999	825	819	761	800	465	0

Apply edgeR to real data to get simulation parameters

Normalization factors

c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}

Main effects and dispersions

Parent (1)	Parent (2)	Hybrid (3)	Dispersion
$\mu_{1,1}$	$\mu_{1,2}$	$\mu_{1,3}$	ϕ_1
$\mu_{2,1}$	$\mu_{2,2}$	$\mu_{2,3}$	ϕ_2
$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\mu_{27888,1}$	$\mu_{27888,2}$	$\mu_{27888,3}$	ϕ_{27888}

Truth: which genes have heterosis?

Feature 1	$I(\mu_{1,3} > \max(\mu_{1,1}, \mu_{1,2}) \text{ or } < \min(\mu_{1,1}, \mu_{1,2}))$
Feature 2	$I(\mu_{2,3} > \max(\mu_{2,1}, \mu_{2,2}) \text{ or } < \min(\mu_{2,1}, \mu_{2,2}))$
...	...
Feature 27888	$I(\mu_{27888,3} > \max(\mu_{27888,1}, \mu_{27888,2}) \text{ or } < \min(\mu_{2,1}, \mu_{27888,2}))$

iid negative binomial counts (parent 1)

$\text{NB}(e^{c_1+\mu_{1,1}}, \phi_1)$	$\text{NB}(e^{c_2+\mu_{1,1}}, \phi_1)$	$\text{NB}(e^{c_3+\mu_{1,1}}, \phi_1)$	$\text{NB}(e^{c_4+\mu_{1,1}}, \phi_1)$
$\text{NB}(e^{c_1+\mu_{2,1}}, \phi_2)$	$\text{NB}(e^{c_2+\mu_{2,1}}, \phi_2)$	$\text{NB}(e^{c_3+\mu_{2,1}}, \phi_2)$	$\text{NB}(e^{c_4+\mu_{2,1}}, \phi_2)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\text{NB}(e^{c_1+\mu_{27888,1}}, \phi_{27888})$	$\text{NB}(e^{c_2+\mu_{27888,1}}, \phi_{27888})$	$\text{NB}(e^{c_3+\mu_{27888,1}}, \phi_{27888})$	$\text{NB}(e^{c_4+\mu_{27888,1}}, \phi_{27888})$

Remove low-count rows to get 25000 features

		Parent (1)				Parent (2)				Hybrid (3)				Truth
HPH	Feature 1	3	4	2	1	0	0	1	0	700	900	825	860	1
HPH	Feature 2	0	1	1	0	2	7	5	18	50	501	400	90	1
	Feature 3	100	225	0	15	300	106	200	400	70	279	100	123	0
LPH	Feature 4	893	400	760	901	1000	513	760	580	5	5	6	7	1
	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Feature 25000	10	13	6	4	902	912	999	825	819	761	800	465	0

Simulation workflow

- Simulate 30 datasets as above:
  - 10 datasets with 4 samples (libraries, columns, etc.) per group
  - 10 with 8 per group
  - 10 with 16 per group
- For each simulated dataset, test for heterosis with
  - empirical Bayes with STAN (Eric's method)
  - `edgeR`
  - `baySeq`
  - `ShrinkBayes`
- Compare methods with ROC curves

edgeR

- Fit a loglinear model to estimate main effects $\mu_{f,t}$
  - Feature $f = 1, \ldots, 25000$
  - Treatment group 1 (parent), 2 (parent), 3 (hybrid)
- Likelihood ratio tests to get p-values $p_{f,1},\ p_{f,2}$

$$H_{0,1} : \mu_{f,3} = \mu_{f,1} \qquad H_{a,1} : \mu_{f,3} \neq \mu_{f,1}$$
$$H_{0,2} : \mu_{f,3} = \mu_{f,2} \qquad H_{a,2} : \mu_{f,3} \neq \mu_{f,2}$$

-

Final p-value	if...
$p_{f,1}/2$	$\widehat{\mu}_{f,3} < \widehat{\mu}_{f,1} \leq \widehat{\mu}_{f,2}$ or $\widehat{\mu}_{f,3} > \widehat{\mu}_{f,1} \geq \widehat{\mu}_{f,2}$
$p_{f,2}/2$	$\widehat{\mu}_{f,3} < \widehat{\mu}_{f,2} \leq \widehat{\mu}_{f,1}$ or $\widehat{\mu}_{f,3} > \widehat{\mu}_{f,2} \geq \widehat{\mu}_{f,1}$
1	$\widehat{\mu}_{f,1} \leq \widehat{\mu}_{f,3} \leq \widehat{\mu}_{f,2}$ or $\widehat{\mu}_{f,2} \leq \widehat{\mu}_{f,3} \leq \widehat{\mu}_{f,1}$

baySeq

- Estimate main effects $\mu_{f,t}$ using edgeR.
- Calculate the posterior probability that each feature satisfies:

Model	Constraint
M_1	All $\mu_{f,t}$'s equal
M_2	$\mu_{f,1} = \mu_{f,2}$
M_3	$\mu_{f,1} = \mu_{f,3}$
M_4	$\mu_{f,2} = \mu_{f,3}$
M_5	All $\mu_{f,t}$'s distinct

-

Final posterior probability	if...
1	$\widehat{\mu}_{f,1} \le \widehat{\mu}_{f,3} \le \widehat{\mu}_{f,2}$ or $\widehat{\mu}_{f,2} \le \widehat{\mu}_{f,3} \le \widehat{\mu}_{f,1}$
$P(M_3 \mid \text{data}) + P(M_5 \mid \text{data})$	otherwise

ShrinkBayes

- Built on `inla` (integrated nested Laplace approximation).
- empirical Bayes with a zero-inflated NB likelihood and normal priors.
- I reparameterize

$$\phi_f = \frac{\mu_{f,1} + \mu_{f,2}}{2} \qquad \text{(parental mean)}$$

$$\alpha_f = \frac{\mu_{f,2} - \mu_{f,1}}{2} \qquad \text{(half parental difference)}$$

$$\delta_f = \mu_{f,3} - \frac{\mu_{f,1} + \mu_{f,2}}{2} \qquad \text{(hybrid effect)}$$

ShrinkBayes

ϕ_f	α_f	δ_f
parental mean	half parental difference	hybrid effect

-