Reversible Jump Details

Brian S. Yandell
www.stat.wisc.edu/~yandell/statgen
University of Wisconsin-Madison
April 2008

reversible jump idea

- expand idea of MCMC to compare models
- adjust for parameters in different models
  - augment smaller model with innovations
  - constraints on larger model
- calculus "change of variables" is key
  - add or drop parameter(s)
  - carefully compute the Jacobian
- consider stepwise regression
  - Mallick (1995) & Green (1995)
  - efficient calculation with Hausholder decomposition

model selection in regression

- known regressors (e.g. markers)
  - models with 1 or 2 regressors
- jump between models
  - centering regressors simplifies calculations

$$m = 1 : Y_i = \mu + a(Q_{i1} - \overline{Q}_1) + e_i$$

$$m = 2 : Y_i = \mu + a_1(Q_{i1} - \overline{Q}_1) + a_2(Q_{i2} - \overline{Q}_2) + e_i$$

slope estimate for 1 regressor

recall least squares estimate of slope
note relation of slope to correlation

$$\hat{a} = \frac{r_{1y}\, s_y}{s_1}, \quad r_{1y} = \frac{\sum_{i=1}^{n}(Q_{i1} - \overline{Q}_1)(Y_i - \overline{Y})/n}{s_1\, s_y}$$

$$s_1^2 = \sum_{i=1}^{n}(Q_{i1} - \overline{Q}_1)^2/n, s_y^2 = \sum_{i=1}^{n}(Y_i - \overline{Y})^2/n$$

2 correlated regressors

slopes adjusted for other regressors

$$\hat{a}_1 = \frac{(r_{1y} - r_{12} r_{2y}) s_y}{s_1} = \hat{a} - \frac{r_{2y} s_y}{s_2} c_{21}, \quad c_{21} = \frac{r_{12} s_2}{s_1}$$

$$\hat{a}_2 = \frac{(r_{2y} - r_{12} r_{1y}) s_y}{s_2}, s_{2\cdot 1}^2 = \frac{\sum_{i=1}^{n} (Q_{i2} - \overline{Q}_2 - c_{21}(Q_{i1} - \overline{Q}_1))^2}{n}$$

Gibbs Sampler for Model 1

- mean $\mu \sim \phi\left(\eta + B_n(\overline{Y}_. - \eta), B_n \frac{\sigma^2}{n}\right), B_n = \frac{n}{n+\kappa}$
- slope $a \sim \phi\left(B_n \frac{\sum_{i=1}^{n} (Q_{i1} - \overline{Q}_1)(Y_i - \overline{Y}_.)}{n s_1^2}, B_n \frac{\sigma^2}{n s_1^2}\right)$
- variance $\sigma^2 \sim \text{inv - }\chi^2\left(\nu + n, \frac{\nu \tau^2 + \sum_{i=1}^{n} \left(Y_i - \overline{Y}_. - a(Q_{i1} - \overline{Q}_1)\right)^2}{\nu + n}\right)$

Gibbs Sampler for Model 2

- mean $$\mu \sim \phi\left(\eta + B_n(\overline{Y}_\cdot - \eta), B_n \frac{\sigma^2}{n}\right)$$
- slopes $$a_2 \sim \phi\left(B_n \frac{\sum_{i=1}^{n}(Q_{i2} - \overline{Q}_2)(Y_i - \overline{Y}_\cdot - a_1(Q_{i1} - \overline{Q}_1))}{n s_{2\cdot 1}^2}, B_n \frac{\sigma^2}{n s_{2\cdot 1}^2}\right)$$
- variance $$\sigma^2 \sim \text{inv-}\chi^2\left(v + n, \frac{v\tau^2 + \sum_{i=1}^{n}\left(Y_i - \overline{Y}_\cdot - \sum_{k=1}^{2} a_k(Q_{ik} - \overline{Q}_k)\right)^2}{v + n}\right)$$

updates from 2->1

- drop 2nd regressor
- adjust other regressor

$$a \rightarrow a_1 + a_2 c_{21}$$
$$a_2 \rightarrow 0$$

updates from 1->2

- add 2nd slope, adjusting for collinearity
- adjust other slope & variance

$$z \sim \phi(0,1), \qquad J = \frac{\sigma}{s_{2\cdot 1}\sqrt{n}}$$

$$a_2 \to \hat{a}_2 + z \times J, \quad \hat{a}_2 = \frac{\sum_{i=1}^{n}(Q_{i2} - \overline{Q}_2)\left(Y_i - \hat{\mu} - \hat{a}_1(Q_{i1} - \overline{Q}_1)\right)}{n s_{2\cdot 1}^2}$$

$$a_1 \to a - a_2 c_{21} = a - z \times c_{21} J - \hat{a}_2 c_{21}$$

model selection in regression

- known regressors (e.g. markers)
  - models with 1 or 2 regressors
- jump between models
  - augment with new innovation *z*

m	parameters	innovations	transformations
$1 \to 2$	$(\mu, a, \sigma^2; z)$	$z \sim \phi(0,1)$	$\left\{ \begin{array}{c} a_2 \to \hat{a}_2 + z \times J \\ a_1 \to a - a_2 c_{21} \end{array} \right\}$
$2 \to 1$	$(\mu, a_1, a_2, \sigma^2)$		$\left\{ \begin{array}{c} a \to a_1 + a_2 c_{21} \\ z \to 0 \end{array} \right\}$

change of variables

- change variables from model 1 to model 2
- calculus issues for integration
  - need to formally account for change of variables
  - infinitessimal steps in integration (*db*)
  - involves partial derivatives (next page)

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{bmatrix} 1 & -c_{21}J & -c_{21} \\ 0 & J & 1 \end{bmatrix} \times \begin{pmatrix} a \\ z \\ \hat{a}_2 \end{pmatrix} = g(a; z \mid Y, Q_1, Q_2)$$

$$\int \pi(a_1, a_2 \mid Y, Q_1, Q_2) da_1 da_2 = \int \pi(a; z \mid Y, Q_1, Q_2) J da dz$$

Jacobian & the calculus

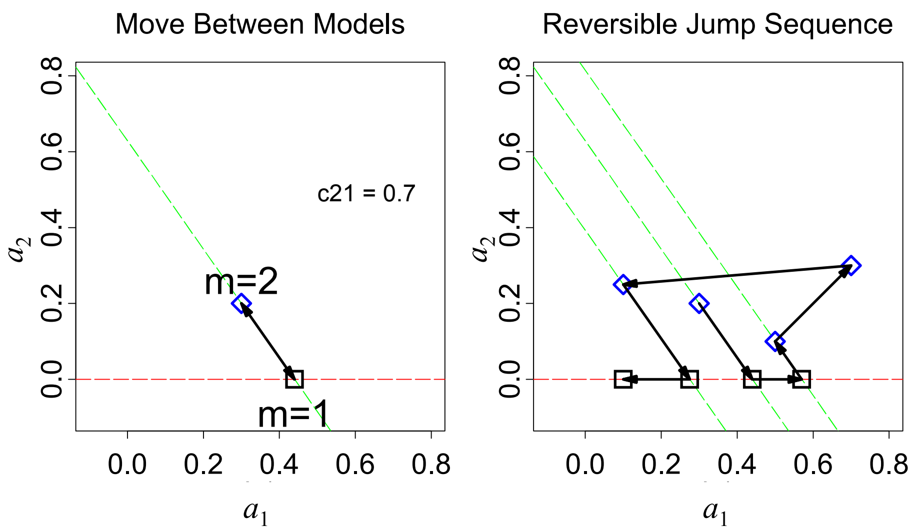
- Jacobian sorts out change of variables
  - <u>careful</u>: easy to mess up here!

$$g(a; z) = (a_1, a_2), \ \frac{\partial g(a; z)}{\partial a \partial z} = \begin{bmatrix} 1 & -c_{21}J \\ 0 & J \end{bmatrix}$$

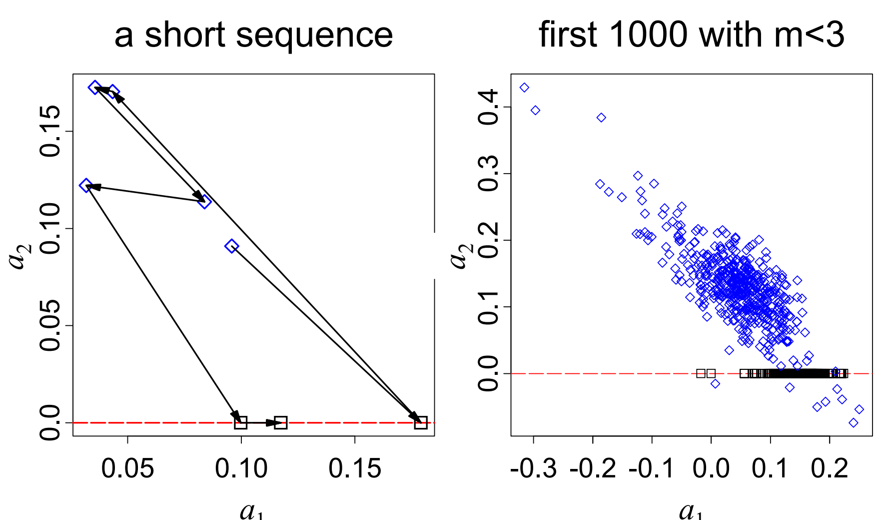
$$\left| \det\left(\begin{bmatrix} 1 & -c_{21}J \\ 0 & J \end{bmatrix} \right) \right| = \left| 1 \times J - 0 \times (-c_{21}J) \right| = J$$

$$da_1 da_2 = \left| \det\left(\frac{\partial g(\mu, a, \sigma^2; z)}{\partial a \partial z} \right) \right| da_1 da_2 = J da dz$$

geometry of reversible jump



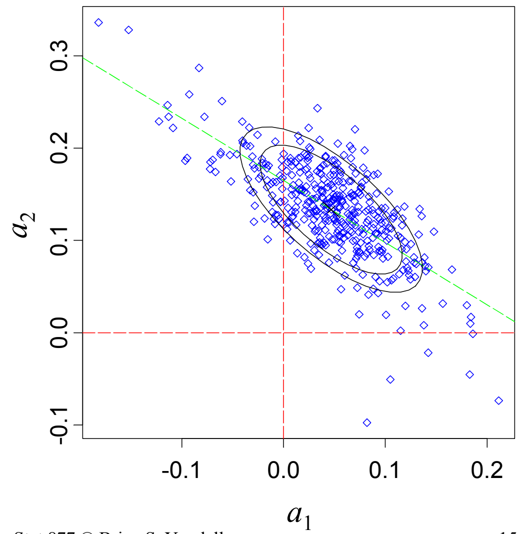
QT additive reversible jump



credible set for additive

90% & 95% sets based on normal

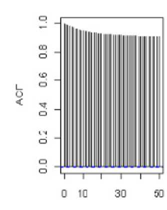
regression line corresponds to slope of updates



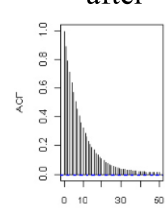
multivariate updating of effects

- more computations when $m > 2$
- avoid matrix inverse
  - Cholesky decomposition of matrix
- simultaneous updates
  - effects at all loci
- accept new locus based on
  - sampled new genos at locus
  - sampled new effects at all loci
- also long-range positions updates

before



after



References

- Satagopan, Yandell (1996); Heath (1997); Sillanpää, Arjas (1998); Stephens, Fisch (1998)
- Green (1995); Richardson, Green (1997); Green 2003, 2004