Variable selection

- Often a model includes covariates/predictors (fixed effects) as well as random effects
- In addition if a large number of effects are present then it is useful to evaluate which are most important
- Significance of individual effects can be assessed in a full model (i. e. including all predictors)
- However masking can occur due to co-linearities
- More often the choice is made to search amongst models to find the 'best' subset of predictors.

Some Notation

y_i : outcome $i = 1,....,m$

$x_{1i},......,x_{pi}$: p predictors

$\boldsymbol{\beta}$: $\{\beta_0, \beta_1,........,\beta_p\}^T$: regression parameters

$\phi_1,.......,\phi_p$:parameters

A linear model

- Assume first a Gaussian linear model of the form

$$y_i = \mu_i + e_i$$

$$\mu_i = \beta_0 + \beta_1 x_{1i} \dots\dots\dots + \beta_p x_{pi}$$

$$e_i \sim N(0, \tau^{-1})$$

Selection

- We could first of all fit different combinations of predictors and assess DIC or MSPE or some other GOF measure.
- This requires fitting a variety of models
- Sequential methods can also be used (backward elimination, forward selection, best subsets)
- Automatic methods may be preferred

Bayesian Variable Selection

- Automatic approach that allows variable suitability to be assessed while fitting a complete (full) model
- Recent reference :

O'Hara, R. and Sillanpää (2009) A Review of Bayesian Variable Selection Methods: what, how, which *Bayesian Analysis*, 4, 85-118

Also BBLL chapter 11

Four Main Methods

- Gibbs Variable Selection, Kuo & Mallick
- Stochastic search Variable Selection (SSVS)
- Reversible Jump McMC
- Adaptive shrinkage (Laplace priors)

Here we will only examine one of the simplest methods: Kuo & Mallick entry parameter (KMEP) method

Basic KMEP

- Define our Gaussian model for p predictors as

$$y_i = \beta_0 + \beta_1 x_{1i} \ldots\ldots\ldots\ldots \beta_p x_{pi} + e_i$$

$$= \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + e_i$$

$$e_i \sim N(0, \tau_y^{-1})$$

KMEP

- This is extended by introducing new parameters

$$y_i = \beta_0 + \phi_1\beta_1 x_{1i} \ldots\ldots\ldots\ldots \phi_p\beta_p x_{pi} + e_i$$

$$e_i \sim N(0, \tau_y^{-1})$$

$$\phi_* \sim Bern(p_*):$$ 0/1 entry parameter

$$p_* \sim Beta(1,1)$$: uniform prior

Sampling

- During posterior sampling the entry parameters will have values 1 or 0 denoting whether the variable is in the model or not.
- The posterior average of the entry parameter gives the probability of the variable being in the model:

$$P_{\phi_j} = \hat{\Pr}(x_j) = \sum_{g=1}^{G} \phi_j^g / G$$

Inclusion rules

- Basically the higher the value of P_{ϕ_j}
  - the more likely the j th variable is included
- Often 0.5 is assumed as cut off i.e.

$$P_{\phi_j} > 0.5$$

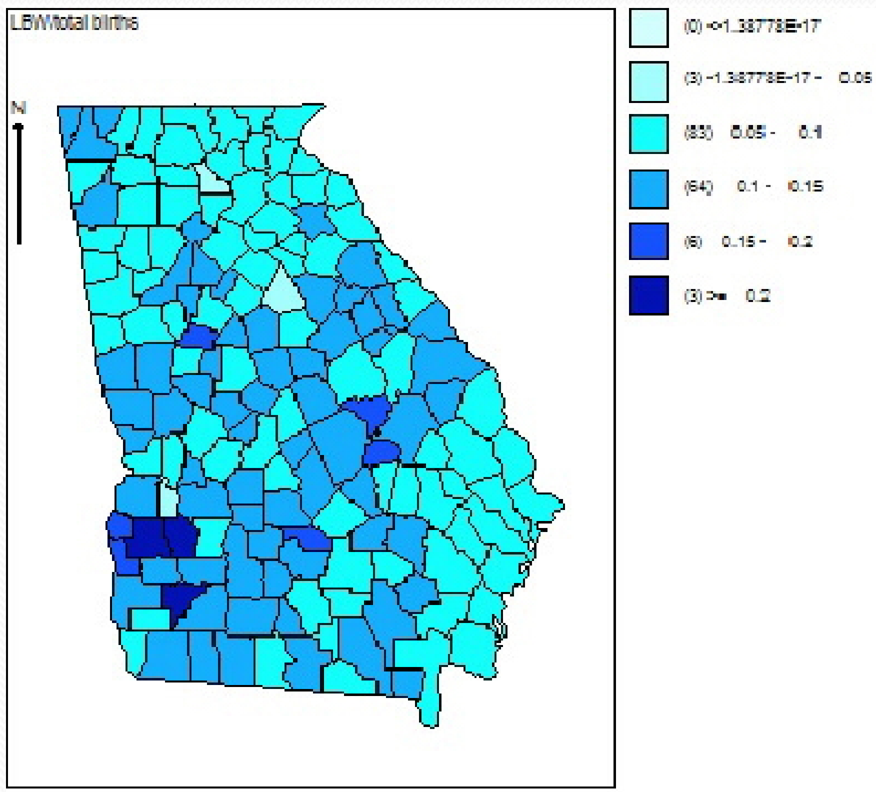
More Generally

- This method can be applied to a wide class of models: logistic or Poisson regression, generalized linear mixed models or additive mixed models

A Spatial example

- Georgia: counts of low birth weight (LBW) in 159 counties for the year 2007
- Model: outcome (LBW) with predictors thought to affect LBW

Map of ratio of LBW to total births



Predictors

- County level: ARF sourced (http://arf.hrsa.gov)
- Black (%), income (median income), poverty(%), unemployment (%), population density

Model

$$y_i \sim bin(p_i, n_i)$$

$$\text{logit}(p_i) = \beta_0 + \phi_1\beta_1 x_{1i} \ldots + \phi_5\beta_5 x_{5i}$$

where

x_{1i} : population density

x_{2i} : Black %

x_{3i} : median income

x_{4i} : poverty %

x_{5i} : unemployment %

Full Model

- DIC for full model:
- Model 1 : fixed predictors
- Model 2: plus random intercept

Model	DIC	pD
Model 1	1139.9	5.9
Model 2	1071.6	59.8

KMEP variable selection

- We have a program that you can run:

VAR_SELECTexample.odc

Results

node	mean	sd	2.50%	median	97.50%
psi[1]	0.04211	0.2008	0	0	1
psi[2]	0.9835	0.1274	1	1	1
psi[3]	0.1001	0.3001	0	0	1
psi[4]	1	0	1	1	1
psi[5]	0.06715	0.2503	0	0	1

Notes

- Only those estimates >0.5 are to be considered
- In this case, the tow variables with high entry probability are Black% and poverty%
- These variables also have the most significant parameter estimates
- Model with only Black% and poverty% should be favored over lowest DIC model

Results

node	mean	sd	2.50%	median	97.50%
b[1]	0.006272	1.124	-2.453	0.00294	2.496
b[2]	0.1066	0.1706	0.05103	0.1119	0.1548
b[3]	0.001495	1.091	-2.442	0.001884	2.362
b[4]	0.1193	0.0363	0.0727	0.1152	0.1969
b[5]	0.009985	1.108	-2.411	-0.00137	2.49

Discussion

- Results are sometimes not as clear as this example
- You can get predictors 'hovering' around 0.3 or 0.4 in some models with no clear 'winners'
- There is also an issue of prior sensitivity
- We assumed:

$$\phi_j \sim Bern(p_j)$$

$$p_j \sim Beta(0.5, 0.5)$$

- This is the Jeffrey's prior and is mostly vague
- Other priors?