Automatic variable selection in hierarchical models: application to durations of untreated psychoses

Dave Lunn  1    James Kirkbride  2

1MRC Biostatistics Unit, Cambridge, UK
2Dept. Psychiatry, University of Cambridge, UK

July 1, 2008
Outline

▶ Background
▶ DUP data
▶ Random effects models
▶ Mixed effects models
▶ Variable selection
▶ Interpretation of results
Outline

- Background
- DUP data
- Random effects models
  - Mixed effects models
  - Variable selection
- Interpretation of results
Outline

- Background
- DUP data
- Random effects models
- Mixed effects models
- Variable selection
- Interpretation of results
Outline

- Background
- DUP data
- Random effects models
- Mixed effects models
- Variable selection
- Interpretation of results
Family involvement in help-seeking associated with shorter duration of untreated psychoses (DUP)

Are factors at neighbourhood level relevant to DUP?

Recent study in Netherlands found +ve relationship between hospital admissions and neighbourhood level of informal social control

Perhaps neighbourhoods with greater social cohesion are associated with shorter DUPs?
Background

- Family involvement in help-seeking associated with shorter duration of untreated psychoses (DUP)

- Are factors at neighbourhood level relevant to DUP?
  - Recent study in Netherlands found +ve relationship between hospital admissions and neighbourhood level of informal social control
  - Perhaps neighbourhoods with greater social cohesion are associated with shorter DUPs?
Background

- Family involvement in help-seeking associated with shorter duration of untreated psychoses (DUP)

- Are factors at neighbourhood level relevant to DUP?

- Recent study in Netherlands found positive relationship between hospital admissions and neighbourhood level of informal social control

- Perhaps neighbourhoods with greater social cohesion are associated with shorter DUPs?
Family involvement in help-seeking associated with shorter duration of untreated psychoses (DUP)

Are factors at neighbourhood level relevant to DUP?

Recent study in Netherlands found +ve relationship between hospital admissions and neighbourhood level of informal social control

Perhaps neighbourhoods with greater social cohesion are associated with shorter DUPs?
DUP data

- Data from SE London centre of ÆSOP first episode psychoses study
  - 314 durations from individuals in 32 electoral wards
  - Mode number of durations = 7; min = 1, max = 31
  - Durations range from 1 day to 9118 days (25 years)
    median = 70 days, mean = 450 days!!
  - Distribution a little skewed perhaps...
DUP data

- Data from SE London centre of ÆSOP first episode psychoses study
- 314 durations from individuals in 32 electoral wards
- Mode number of durations = 7; min = 1, max = 31
- Durations range from 1 day to 9118 days (25 years) median = 70 days, mean = 450 days!!
- Distribution a little skewed perhaps...
DUP data

- Data from SE London centre of ÆSOP first episode psychoses study
- 314 durations from individuals in 32 electoral wards
- Mode number of durations = 7; min = 1, max = 31
- Durations range from 1 day to 9118 days (25 years) median = 70 days, mean = 450 days!!
- Distribution a little skewed perhaps...
Distribution of DUPs

Histogram of DUP

Histogram of log.DUP

▶ Measurement error model?
## Covariates

<table>
<thead>
<tr>
<th>Neighbourhood level</th>
<th>Individual level</th>
</tr>
</thead>
<tbody>
<tr>
<td>population density</td>
<td>gender</td>
</tr>
<tr>
<td>proportion non-white British</td>
<td>type of disorder</td>
</tr>
<tr>
<td>proportion black Caribbean</td>
<td>(affective or broad schizophrenia)</td>
</tr>
<tr>
<td>proportion black African</td>
<td>ethnicity, either binary or 7-level</td>
</tr>
<tr>
<td>voter turnout</td>
<td>relationship status (3-level)</td>
</tr>
<tr>
<td>segregation between white and non-white British</td>
<td>highest ever occupation (6-level)</td>
</tr>
<tr>
<td>proportion single</td>
<td>age</td>
</tr>
<tr>
<td>proportion of Asian groups</td>
<td>family involvement</td>
</tr>
<tr>
<td>multiple deprivation</td>
<td>mode of onset, binary or 6-level</td>
</tr>
<tr>
<td>social capital</td>
<td></td>
</tr>
<tr>
<td>social disorganisation</td>
<td></td>
</tr>
<tr>
<td>social cohesion</td>
<td></td>
</tr>
</tbody>
</table>
Basic random effects model

- Let $T_{ij} = \log(\text{DUP}_{ij})$, where DUP$_{ij}$ is the DUP for individual $j$ in neighbourhood $i$ ($i = 1, \ldots, N = 32, j = 1, \ldots, n_i$)

  \[ T_{ij} = \alpha + u_i + e_{ij} \]
  \[ u_i \sim N(0, \sigma_u^2), \quad e_{ij} \sim N(0, \sigma_e^2) \]

- Or, equivalently: $T_{ij} \sim N(\mu_i, \sigma_e^2)$, $\mu_i = \alpha + u_i$

- Could also allow spatially correlated random effects, e.g.

  \[ u_i = r_i + s_i, \quad r_i \sim N(0, \sigma_r^2), \quad s \sim \text{CAR}(\sigma_s^2) \]
Basic random effects model

- Let $T_{ij} = \log(\text{DUP}_{ij})$, where DUP$_{ij}$ is the DUP for individual $j$ in neighbourhood $i$ ($i = 1, \ldots, N = 32, j = 1, \ldots, n_i$)

$$T_{ij} = \alpha + u_i + e_{ij}$$

$$u_i \sim N(0, \sigma_u^2), \quad e_{ij} \sim N(0, \sigma_e^2)$$

- Or, equivalently: $T_{ij} \sim N(\mu_i, \sigma_e^2)$, $\mu_i = \alpha + u_i$

- Could also allow spatially correlated random effects, e.g.

$$u_i = r_i + s_i, \quad r_i \sim N(0, \sigma_r^2), \quad s \sim \text{CAR}(\sigma_s^2)$$
Basic random effects model

- Let $T_{ij} = \log(\text{DUP}_{ij})$, where DUP$_{ij}$ is the DUP for individual $j$ in neighbourhood $i$ ($i = 1, \ldots, N = 32$, $j = 1, \ldots, n_i$)

$$T_{ij} = \alpha + u_i + e_{ij}$$

$$u_i \sim N(0, \sigma^2_u), \ e_{ij} \sim N(0, \sigma^2_e)$$

- Or, equivalently: $T_{ij} \sim N(\mu_i, \sigma^2_e)$, $\mu_i = \alpha + u_i$

- Could also allow spatially correlated random effects, e.g.

$$u_i = r_i + s_i, \ r_i \sim N(0, \sigma^2_r), \ s \sim \text{CAR}(\sigma^2_s)$$
Basic CAR model

\[ s_i \mid s_{-i} \sim \text{Normal} \left( \frac{1}{\nu_i} \sum_{j \in \delta_i} s_j, \frac{\sigma_s^2}{\nu_i} \right) \]

\[ \delta_i = \text{set of neighbours, area } i \]

\[ \nu_i = \text{no. of neighbours, area } i \]
Basic CAR model

\[ s_i | s_{-i} \sim \text{Normal} \left( \frac{1}{\nu_i} \sum_{j \in \delta_i} s_j, \frac{\sigma_s^2}{\nu_i} \right) \]

\( \delta_i = \text{set of neighbours, area } i \)
\( \nu_i = \text{no. of neighbours, area } i \)
Basic CAR model

\[ s_i | s_{-i} \sim \text{Normal} \left( \frac{1}{\nu_i} \sum_{j \in \delta_i} s_j, \frac{\sigma_s^2}{\nu_i} \right) \]

\( \delta_i = \text{set of neighbours, area } i \)

\( \nu_i = \text{no. of neighbours, area } i \)
BUGS code

model {
  for (i in 1:N) {
    for (j in start[i]:(start[i + 1] - 1)) {
      T[j] ~ dnorm(mu[i], tau.e)
    }
    mu[i] <- alpha + u[i]
    u[i] ~ dnorm(0, tau.u)
  }
...}
BUGS code

model {
  for (i in 1:N) {
    for (j in start[i]:(start[i + 1] - 1)) {
      T[j] ~ dnorm(mu[i], tau.e)
    }
    mu[i] <- alpha + r[i] + s[i]
    r[i] ~ dnorm(0, tau.r)
  }
  s[1:N] ~ car.normal(adj[], weights[], nu[], tau.s)
  ...
}

Dave Lunn
Uncorrelated and spatially correlated random effects

\[ \hat{\alpha} = 4.39 \ (4.15, 4.62), \quad \hat{\sigma}_e = 1.97 \ (1.82, 2.14) \]
\[ \hat{\sigma}_r = 0.129 \ (0.00715, 0.440), \quad \hat{\sigma}_s = 0.185 \ (0.00592, 0.696) \]
Uncorrelated random effects only

\[ \hat{\alpha} = 4.39 \ (4.16, 4.62) \]
\[ \hat{\sigma}_e = 1.97 \ (1.83, 2.14) \]
\[ \hat{\sigma}_u = 0.120 \ (0.0065, 0.419) \]
Which model is better?

- Spiegelhalter et al (2002, RSSB) propose Deviance Information Criterion (DIC) for comparing models
- Model fit measured by posterior mean deviance $\overline{D}$ (deviance $= -2 \times \text{log-likelihood}$)
- Penalized by ‘effective number of parameters’ $p_D = \overline{D} - D(\bar{\theta})$ (posterior mean deviance – deviance evaluated at posterior mean parameters)
- DIC $= \overline{D} + p_D = D(\bar{\theta}) + 2p_D$
- Models with smaller DIC are better supported by data...
## Compare DICs

<table>
<thead>
<tr>
<th>Model</th>
<th>$\bar{D}$</th>
<th>$D(\bar{\theta})$</th>
<th>$p_D$</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncorr. + spat. corr.</td>
<td>1308.14</td>
<td>1308.14</td>
<td>6.071</td>
<td>1314.21</td>
</tr>
<tr>
<td>Uncorrelated only</td>
<td>1308.17</td>
<td>1304.01</td>
<td>4.158</td>
<td>1312.33</td>
</tr>
<tr>
<td>No area-level effects</td>
<td>1308.63</td>
<td>1306.66</td>
<td>1.97</td>
<td>1310.60</td>
</tr>
</tbody>
</table>

$\hat{\alpha} = 4.39 \ (4.17, 4.61)$

$\hat{\sigma}_e = 1.97 \ (1.83, 2.14)$
## Compare DICs

<table>
<thead>
<tr>
<th>Model</th>
<th>$\bar{D}$</th>
<th>$D(\bar{\theta})$</th>
<th>$p_D$</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncorr. + spat. corr.</td>
<td>1308.14</td>
<td>1308.14</td>
<td>6.071</td>
<td>1314.21</td>
</tr>
<tr>
<td>Uncorrelated only</td>
<td>1308.17</td>
<td>1304.01</td>
<td>4.158</td>
<td>1312.33</td>
</tr>
<tr>
<td>No area-level effects</td>
<td>1308.63</td>
<td>1306.66</td>
<td>1.97</td>
<td>1310.60</td>
</tr>
</tbody>
</table>

$$\hat{\alpha} = 4.39 \ (4.17, 4.61)$$
$$\hat{\sigma}_e = 1.97 \ (1.83, 2.14)$$

model {
  for (i in 1:N) {
    for (j in start[i]:(start[i + 1] - 1)) {
      T[j] ~ dnorm(mu[i], tau.e)
    }
    mu[i] <- alpha
  }
  ...
Variable selection

- Model:

\[ T_j \sim N(\mu_j, \sigma^2_e), \quad \mu_j = \alpha + X_j \beta, \quad j = 1, \ldots, N_{\text{tot}} = 314 \]

where \( X_j \) is a row-vector of area-level and/or individual-level covariates for observation \( j \) (\( X_j \) is \( j^{\text{th}} \) row of matrix \( X \))

- Which covariates should form the columns of \( X \)?
- Forward/backward selection can be unreliable
- Include all of them?
  - interpretability issues
  - spurious effects
Variable selection

- Model:

\[ T_j \sim N(\mu_j, \sigma_e^2), \quad \mu_j = \alpha + X_j \beta, \quad j = 1, \ldots, N_{\text{tot}} = 314 \]

where \( X_j \) is a row-vector of area-level and/or individual-level covariates for observation \( j \) (\( X_j \) is \( j^{\text{th}} \) row of matrix \( X \))

- Which covariates should form the columns of \( X \)?
- Forward/backward selection can be unreliable
- Include all of them?
  - interpretability issues
  - spurious effects
Variable selection continued

- Let $\mathbf{Z}$ denote the matrix of all area- and individual-level covariates
- Let $\phi$ denote the set of column indices that form $\mathbf{X}$:

$$X_{jk} = Z_{j\phi_k}, \quad k = 1, \ldots, d = \text{dim}(\phi)$$

- Perhaps we can ‘estimate’ $\phi$?
- Note, however, that it has variable dimension, as has $\beta$ ($X_j$ expands and contracts as covariates are included in and discarded from the model)
- Can handle in Bayesian setting using ‘reversible jump’
Variable selection continued

- Let \( Z \) denote the matrix of all area- and individual-level covariates.
- Let \( \phi \) denote the set of column indices that form \( X \):

\[
X_{jk} = Z_{j\phi_k}, \quad k = 1, \ldots, d = \text{dim}(\phi)
\]

- Perhaps we can ‘estimate’ \( \phi \)?
- Note, however, that it has variable dimension, as has \( \beta \) (\( X_j \) expands and contracts as covariates are included in and discarded from the model).
- Can handle in Bayesian setting using ‘reversible jump’
Priors for variable selection

- Need to specify priors for $\beta$, $\phi$ and $d$
- Typically

\[
\beta | d \sim \text{MVN}_d(0_d, \sigma^2_{\beta} I_d) \quad \text{(e.g. } \sigma^2_{\beta} = 50)\]

\[
\phi | d \sim \text{Uniform} \Rightarrow p(\phi | d) = \frac{d!(Q - d)!}{Q!}
\]

\[
d \sim \text{Poisson}(.) , \quad d \sim \text{Binomial}(p, Q) \quad \text{(e.g. } p = 0.5)\]
BUGS code

model {
  for (i in 1:N) {
    for (j in start[i]: (start[i + 1] - 1)) {
      T[j] ~ dnorm(mu[j], tau.e)
    }
  }
  mu[1:N.tot] <- jump.lin.pred(X[,], d, prec)
  d ~ dbin(0.5, Q)
  ...
}
BUGS output

var #2 = affective/broad schizophrenia (binary)
var #6 = age (continuous)
var #8 = mode of onset (6-level)
var #2 = affective/broad schizophrenia (binary)
var #6 = age (continuous)
var #8 = mode of onset (6-level)
### BUGS output continued

<table>
<thead>
<tr>
<th>model structure</th>
<th>posterior prob.</th>
<th>cumulative prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>00000101000000000000</td>
<td>0.1943107039</td>
<td>0.1943107039</td>
</tr>
<tr>
<td>01000101000000000000</td>
<td>0.09628279883</td>
<td>0.2905935027</td>
</tr>
<tr>
<td>00000101000000100000</td>
<td>0.04008329863</td>
<td>0.3306768013</td>
</tr>
<tr>
<td>01000101000000100000</td>
<td>0.0231153686</td>
<td>0.3537921699</td>
</tr>
<tr>
<td>00000101000010000000</td>
<td>0.02144314869</td>
<td>0.3752353186</td>
</tr>
</tbody>
</table>

... 

<table>
<thead>
<tr>
<th>variable no.</th>
<th>marginal prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.02665139525</td>
</tr>
<tr>
<td>2</td>
<td>0.3389046231</td>
</tr>
<tr>
<td>3</td>
<td>0.02906497293</td>
</tr>
<tr>
<td>4</td>
<td>0.0168971262</td>
</tr>
<tr>
<td>5</td>
<td>0.000000000000</td>
</tr>
<tr>
<td>6</td>
<td>0.9702936277</td>
</tr>
<tr>
<td>7</td>
<td>0.03944814661</td>
</tr>
<tr>
<td>8</td>
<td>1.0</td>
</tr>
</tbody>
</table>

... 

histogram of d
Effect sizes

\( \hat{\sigma}_e : 1.97 \rightarrow 1.50 \)

affective?

\[ \mathbb{E}[\beta_2 | 2 \in \phi, T] = -0.488 \]

age

\[ \mathbb{E}[\beta_{11} | 6 \in \phi, T] = 1.62 \]

sudden onset precipitous

\[ \mathbb{E}[\beta_{13} | 8 \in \phi, T] = -3.00 \quad \mathbb{E}[\beta_{14} | 8 \in \phi, T] = -1.63 \]

acute, no symptoms acute, symptoms

\[ \mathbb{E}[\beta_{15} | 8 \in \phi, T] = -1.79 \quad \mathbb{E}[\beta_{16} | 8 \in \phi, T] = -1.68 \]

no clear demarcation

\[ \mathbb{E}[\beta_{17} | 8 \in \phi, T] = 1.24 \]
Comments

- Can use for selecting ‘best’ model or for averaging inferences over all (or some) models – *Bayesian Model Averaging*
- Advisable to transform all covariates onto same scale
- Care needed when choosing priors – sensitivity
- Can also use ‘reversible jump’ for other modelling challenges... e.g. controlling for complex covariate effects
Acknowledgements

- Joint work with James Kirkbride in Dept. Psychiatry, University of Cambridge
- Thanks to Craig Morgan, Julia Lappin, Paul Fearon and Paola Dazzan (Institute of Psychiatry, King’s College London) for the DUP data
References

- WinBUGS website: http://www.mrc-bsu.cam.ac.uk/bugs/
- OpenBUGS website: http://mathstat.helsinki.fi/openbugs/
- WinBUGS Development website (e.g. Jump): http://www.winbugs-development.org.uk/