STAT 530: Regression, Continued

Mar. 15, 2010

Model comparison in general

z indicates which model, θ_z are the parameters within this model.

Must specify both 'across' and 'within' priors, $p(z)$ and $p(\theta_z|z)$

Both prior and posterior are of a mixture form:

$$p(z, \theta_z) = \sum_z p(z)p(\theta_z|z)$$

$$p(z, \theta_z|D) = \sum_z p(z|D)p(\theta_z|D, z)$$

Second term: posterior formed within model, as usual

First term?

Model comparison in general, continued

Need 'marginal probability' for each model:
$p(D|z) = \int p(D|\theta_z, z)p(\theta_z|z)d\theta_z$

Then 'Bayes as usual'

$$p(z|D) = \frac{p(D|z)p(z)}{\sum_{\tilde{z}} p(D|\tilde{z})p(\tilde{z})}$$

Note (recall?) Bayes factor interpretation:

$$\frac{p(z_i|D)}{p(z_j|D)} = \frac{p(D|z_i)}{p(D|z_j)} \frac{p(z_i)}{p(z_j)}$$

Back to regression

2^p possible models indexed by z

e.g. if $p = 4$ and regressors are (1,PRG,AGE,PRG*AGE), what model does $z = (1, 0, 1, 0)$ represent?

g-prior (with mean zero) convenient for 'within' model, so hyperparameters: g, ν_0, σ^2_{0z}

Much algebra... (but also much intuition)

$$p(D|z) = c(1+g)^{-p_z/2} \frac{(\nu_0 \sigma^2_{0z})^{\nu_0/2}}{(\nu_0\sigma^2_{0z} + SSR^z_g)^{(\nu_0+n)/2}}$$

where $SSR^z_g = y^T(I - \frac{g}{g+1}H_z)y$,
with H_z being the 'hat matrix' for model z.

Summarize so far...

$\beta | z, D \sim N_{p_z}\left(\frac{g}{g+1}\hat{\beta}_{z,OLS}, \ldots\right)$

$p(D|z)$ readily computed

Hence $p(z|D)$ readily computed

well, as long as the number of possible models isn't too large...

Smaller p

- exhaustive evaluation of $p(z|D)$ for all z
- conjugate representation for $(\theta_z | z, D)$

What to report as estimated coefficients???

$E(\beta | z = z^*, D)$, where $z^* = \text{argmax}_z p(z|D)$

$E(\beta|D)$

$E(\beta | z = z^{**}, D)$

where

$$z_j^{**} = \begin{cases} 1 & \text{if } p(z_j = 1|D) > 0.5 \\ 0 & \text{otherwise} \end{cases}$$

Oxygen Uptake Ex.: Posterior distribution over models

1	PRG	AGE	PRG*AGE	p(z|D)
1	1	1	0	0.395
1	0	1	1	0.335
1	1	1	1	0.121
1	0	1	0	0.114
0	1	0	1	0.009
1	1	0	1	0.007
0	1	1	1	0.005
0	0	0	1	0.004
1	0	0	1	0.003
0	1	0	0	0.002
0	0	1	1	0.002

...

inclusion probs

0.98 0.54 0.97 0.49

Ex., continued: model-specific and averaged estimates

1	PRG	AGE	P*A	p(z|D)
-46.457	5.443	1.886	0.000	0.395
-43.374	0.000	1.756	0.218	0.335
-51.294	13.107	2.095	-0.318	0.121
-53.346	0.000	2.278	0.000	0.114
0.000	-38.187	0.000	1.776	0.009
-2.767	-35.420	0.000	1.776	0.007
0.000	-38.187	-0.098	1.875	0.005
0.000	0.000	0.000	0.320	0.004
-3.272	0.000	0.000	0.444	0.003
0.000	7.705	0.000	0.000	0.002
0.000	0.000	-0.098	0.418	0.002

...

model averaged

-45.191 2.983 1.845 0.076

p larger

$p(z, \beta_z, \sigma_z^2 | D)$ not readily Gibbs sampled. Why?

$p(z|D)$ readily Gibbs sampled. Why?

Aside: General Monte Carlo strategy. Always looking for simplifications via either higher-D or lower-D!

Gibbs sampling output readily augmented by $(\beta_z, \sigma_z$ draws.

What to do with Gibbs sampling output

Have $z^{(1)}, \ldots z^{(S)}$.

Say M is the set of all possible models, i.e., 2^p elements.
M^* is the subset that are visited at least once by the sampler.

For $z \in M^*$, two possible estimates of $p(z|D)$.