STAT 530: Regression

Mar. 10, 2010

Basic regression model

$$Y = \beta_1 X_1 + \ldots + \beta_p X_p + \epsilon$$

Or in matrix notation:

$$Y|X \sim N_n(X\beta, \sigma^2 I_n)$$

Prior on (β, σ^2)? Embarrassment of riches!

Considerations

- Computational convenience
  - closed-form expressions?
  - Gibbs sampling?
- Expression of belief about sparsity
  - Some components of β are zero?
  - Shrinkage of β toward zero?

Starting point: semi-conjugate prior

$\beta \sim N_p(\beta_0, \Sigma_0)$,
$\sigma^2 \sim IG(\nu_0/2, \nu_0\sigma_0^2/2)$

Parallels to multivariate normal case (think $\Sigma_0 \to \infty$, etc.)

Simplifications with $(\beta|\sigma^2) \sim N_p(\beta_0, \sigma^2 V_0)$

Zellner's 'g-prior' is particularly convenient and intuitive

$$\begin{aligned} \beta|\sigma^2 &\sim N_p\{0, g\sigma^2(X^TX)^{-1}\} \\ \sigma^2 &\sim IG(\nu_0/2\nu_0\sigma_0^2/2) \end{aligned}$$

- prior depends on data!
  - but only via X matrix, not Y vector
- choice $g = n$ gives 'unit information' interpretation
- **fully conjugate**
  - $(\sigma^2|y, X) \sim IG()$
  - $(\beta|\sigma^2, y, X) \sim N_p\{\frac{g}{g+1}\hat{\beta}_{OLS}, \frac{g}{g+1}\sigma^2(X^TX)^{-1}\}$

Comments on text example using g-prior

$n = 12$, Y=OXG, X=(1, PROG, AGE, PROG*AGE)

Hyperparameters $g = n$, $\nu_0 = 1$, $\sigma_0^2 =$ OLS estimate.

Direct MC sampling of $(\beta, \sigma^2|X, y)$

Posterior correlation between (β_2, β_4)

Nice display of inference for age-specific program effect, $\beta_2 + \beta_4$age

Aside: shape of prior matters

Consider

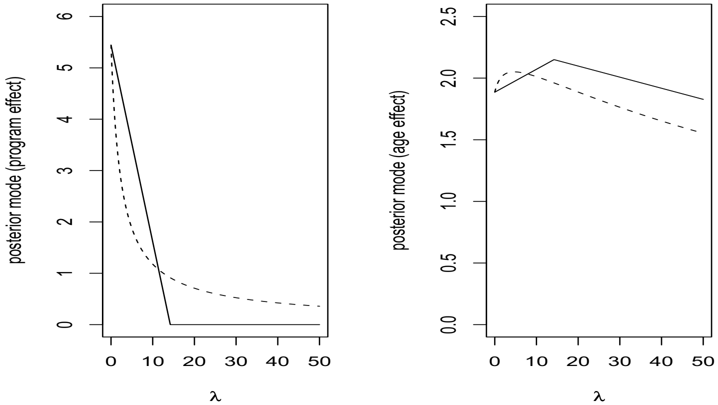
$$p(\beta|\sigma) \quad \propto \quad \prod_{i=1}^{p} exp\left\{-\frac{\lambda}{2\sigma^2}\beta_i^2\right\}$$

versus

$$p(\beta|\sigma) \quad \propto \quad \prod_{i=1}^{p} exp\left\{-\frac{\lambda}{2\sigma^2}|\beta_i|\right\}$$

How does posterior **mode** of β vary with λ under each prior?

Ex. (text), n=12, Y=OXG, X=(1, PROG, AGE)



Prior as regularizer

What if $(X^T X)$ is not invertible?

Possible reasons/contexts?

If $\beta | \sigma^2 \sim N_p(0, \sigma^2 V_0)$ then

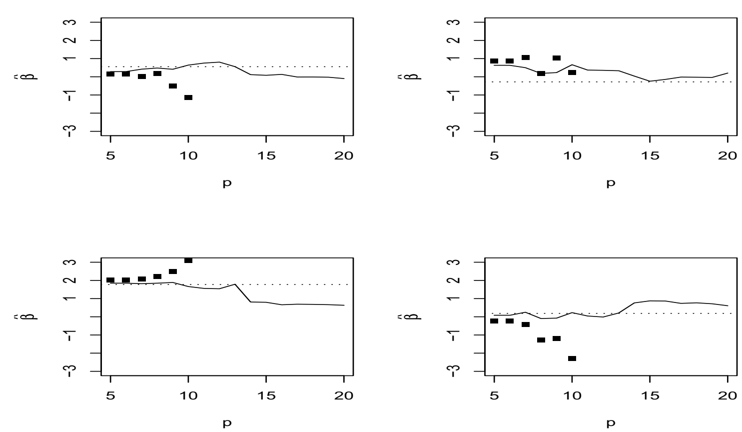
$$E(\beta | y, X) = (X^T X + V_0^{-1})^{-1} X^T y$$

Little experiment

```
beta <- rnorm(20)
xmat <- matrix(rnorm(10*20),10,20) +
        matrix(rnorm(10),10,20)
err <- rnorm(10)

for (p in 5:20) {
  y <- xmat[,1:p]%*%beta[1:p] + err
  est.ols <- lm(y~xmat[,1:p]-1, singular.OK=F)$coef
  est.bys <- solve(t(xmat[,1:p])%*%xmat[,1:p]+diag(p))%*%
             t(xmat[,1:p])%*%y
}
```

Estimates of β_j $(j = 1, \ldots, 4)$ as p increases



How much shrinkage?

Hierarchical model idea again.

A more complex semi-conjugate prior would be $p(\beta, \gamma^2, \sigma^2)$, with:

$$\begin{aligned} \beta|\gamma^2, \sigma^2 &\sim N_p(\beta_0, \gamma^2 V_0) \\ \gamma^2 &\sim IG(\omega_0/2, \omega_0\gamma_0^2/2) \\ \sigma^2 &\sim IG(\nu_0/2, \nu_0\sigma_0^2/2) \end{aligned}$$

Full conditionals OK.

Up next?

What about a prior expressing a belief of the form: I think some/most of the regression coefficients are zero, but I don't know which ones.