Regression Models: Maximum Likelihood Estimation

Assume a regression model with the following structure:

$$y_i = \beta_1 x_{i,1} + \ldots + \beta_k x_{i,k} + \epsilon_i,$$

for $i = 1, \ldots, n$ and ϵ_i independent random variables with $\epsilon_i \sim N(0, v)$ for all i. This model can be written in matrix form as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, v\boldsymbol{I}), \tag{1}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)'$ is an n-dimensional vector of responses, $\boldsymbol{X}$ is an $n \times k$ matrix containing the explanatory variables, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ is the k-dimensional vector of regression coefficients, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$ is the n-dimensional vecror of errors, $\boldsymbol{I}$ is an $n \times n$ identity matrix.

If $\boldsymbol{X}$ is a full rank matrix with rank k the maximum likelihood estimator for $\boldsymbol{\beta}$, denoted as $\hat{\boldsymbol{\beta}}_{MLE}$ is given by:

$$\hat{\boldsymbol{\beta}}_{MLE} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y},$$

and the MLE for v is given by

$$\hat{v}_{MLE} = \frac{(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{MLE})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{MLE})}{n}.$$

$\hat{v}_{MLE}$ is not an unbiased estimator of v, therefore, the following unbiased estimator of v is typically used:

$$s^2 = \frac{(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{MLE})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{MLE})}{n-k}.$$

Regression Models: Bayesian Inference

Assume once again we have a model with the structure in (1), which results in a likelihood of the form

$$p(\boldsymbol{y}|\boldsymbol{\beta}, v) = \frac{1}{(2\pi v)^{n/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right\}.$$

If a prior of the form

$$p(\boldsymbol{\beta}, v) \propto \frac{1}{v}$$

is used, we obtain that the posterior distribution is given by

$$p(\boldsymbol{\beta}, v|\boldsymbol{y}) \propto \frac{1}{v^{n/2+1}} \exp\left\{-\frac{1}{2v}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right\}.$$

In addition it can be shown that

- $(\boldsymbol{\beta}|v, \boldsymbol{y}) \sim N(\hat{\boldsymbol{\beta}}_{MLE}, v(\boldsymbol{X}'\boldsymbol{X})^{-1})$,
- $(v|\boldsymbol{y}) \sim IG((n-k)/2, d/2)$, with

$$d = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{MLE})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{MLE}).$$

with $k = dim(\boldsymbol{\beta})$.

Given that $p(\boldsymbol{\beta}, v|\boldsymbol{y}) = p(\boldsymbol{\beta}|v, \boldsymbol{y})p(v|\boldsymbol{y})$ the equations above provide a way to directly sample from the posterior distribution of $\boldsymbol{\beta}$ and v by first sampling v from the inverse-gamma distribution above and then conditioning on this sampled value of v, sampling $\boldsymbol{\beta}$ from the normal distribution above.