

22S:138 Bayesian Statistics

Bayesian Linear Regression

Lecture 15
Oct. 18, 2006

Kate Cowles
374 SH, 335-0727
kcowles@stat.uiowa.edu

Intro to Bayesian simple linear regression

- likelihood
 $y_i | x_i, \beta_0, \beta_1, \sigma^2 \sim N(\beta_0 + \beta_1(x_i - \bar{x}), \sigma^2)$
- “reference prior”: independently uniform on $\beta_0, \beta_1, \log \sigma^2$

$$p(\beta_0, \beta_1, \sigma^2) \propto \frac{1}{\sigma^2}$$

- IG(0, 0) on σ^2
- We approximate this prior in WinBUGS with
 - * vague normals (or “dflat()”) priors on β_0 and β_1
 - * vague gamma on precision

Centering the covariate in (frequentist) simple linear regression

- particularly useful when all values of the covariate are far away from zero and of the same sign
- in this case, without centering, intercept is estimated very imprecisely
- example: heart rate and body temperature data
 - response variable: heart rate in beats/minute
 - covariate: body temperature in degrees F.
 - subjects: 130 healthy adults

Analytically computing joint posterior

- notation

$$\hat{\beta}_0 = \bar{y}$$

$$\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$$

$$SSE = \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1(x_i - \bar{x}))^2$$

- these are all statistics – functions of the data alone

- *joint* posterior using reference prior

$$p(\beta_0, \beta_1, \sigma^2 | \mathbf{y}) \propto \frac{1}{\sigma^2} \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp \left[\frac{-\sum_i (y_i - \beta_0 - \beta_1(x_i - \bar{x}))^2}{2\sigma^2} \right]$$

$$= \frac{1}{(\sigma^2)^{\frac{n+2}{2}}} \exp \left[\frac{SSE + n(\beta_0 - \hat{\beta}_0)^2 + \sum_i (x_i - \bar{x})^2 (\beta_1 - \hat{\beta}_1)^2}{2\sigma^2} \right]$$

Steps to find marginal posterior distributions

- So conditional on σ^2

$$p(\beta_0|\mathbf{y}, \sigma^2) = N(\hat{\beta}_0, \frac{\sigma^2}{n})$$

$$p(\beta_1|\mathbf{y}, \sigma^2) = N(\hat{\beta}_1, \frac{\sigma^2}{\sum(x_i - \bar{x})^2})$$

- if we integrate β_1 out of the joint posterior we get

$$p(\beta_0, \sigma^2|\mathbf{y}) \propto \frac{1}{(\sigma^2)^{\frac{n+1}{2}}} \exp\left[-\frac{SSE + n(\beta_0 - \hat{\beta}_0)^2}{2\sigma^2}\right]$$

- now to get marginal of β_0 , integrate σ^2 out of the preceding expression

$$p(\beta_0|\mathbf{y}) = t(\hat{\beta}_0, \frac{s^2}{n}, n - 2)$$

a t distribution with mean $\hat{\beta}_0$, scale $\frac{s^2}{n}$, and degrees of freedom $n - 2$

When will posterior be proper with the improper reference prior?

- $n > 2$
- x_i s not all the same

– recall that $s^2 = \frac{SSE}{n-2}$

- similarly

$$p(\beta_1|\mathbf{y}) = t(\hat{\beta}_1, \frac{s^2}{\sum(x_i - \bar{x})^2}, n - 2)$$

- finally

$$p(\sigma^2|\mathbf{y}) = IG\left(\frac{n-2}{2}, \frac{SSE}{2}\right)$$

what GCSR calls a scaled Inverse $\chi^2(n - 2, s^2)$

Informative priors in simple linear regression

- If a previous dataset is available:
 - See handout from P.M. Lee book for exact method if you have only summary statistics from previous dataset
 - Or: just combine old and new datasets and use reference prior (if you have all the data from old dataset)
 - Or derive the following simplified, independent priors (formulas apply if covariate was centered in previous analysis and you will center it in your analysis):

$$p(\beta_0) = N\left(\hat{\beta}_{0,old}, \frac{s_{old}^2}{n_{old}}\right)$$

$$p(\beta_1) = N\left(\hat{\beta}_{1,old}, \frac{s_{old}^2}{\sum_i(x_{i,old} - \bar{x}_{old})^2}\right)$$

$$p(\sigma^2) = IG\left(\frac{n_{old} - 2}{2}, \frac{SSE_{old}}{2}\right)$$

Example

- You wish to use an article in the literature regarding a previous study to construct a prior for a simple linear regression model.
- The investigators centered their covariate and report the following:

$$\begin{aligned} n &= 100 \\ \hat{\beta}_0 &= 5 \quad (s.e.2) \\ \hat{\beta}_1 &= -2 \quad (s.e.1) \end{aligned}$$

or

$$\begin{aligned} \hat{\beta}_0 &= 5 \quad 95\% \text{ c.i. } (1.04, 8.96) \\ \hat{\beta}_1 &= -2 \quad 95\% \text{ c.i. } (-3.98, -0.02) \end{aligned}$$

- Recall that:

$$\begin{aligned} s.e.(\hat{\beta}_0) &= \sqrt{\frac{s^2}{n}} \\ s.e.(\hat{\beta}_1) &= \sqrt{\frac{s^2}{\sum_i (x_i - \bar{x})^2}} \end{aligned}$$

Multiple regression

- likelihood

$$\begin{aligned} y_i | \mathbf{x}_i, \boldsymbol{\beta}, \sigma^2 &\sim N(\beta_0 + \beta_1(x_{1i} - \bar{x}_1) \\ &+ \beta_2(x_{2i} - \bar{x}_2) + \cdots + \beta_{k-1}(x_{k-1,i} - \bar{x}_{k-1} \end{aligned}$$

- Reference prior

$$p(\boldsymbol{\beta}, \sigma^2) \propto \frac{1}{\sigma^2}$$

- conditional on σ^2 , joint posterior of β s

$$p(\boldsymbol{\beta} | \sigma^2, \mathbf{x}, \mathbf{y}) = N(\hat{\boldsymbol{\beta}}, \sigma^2 (X^T X)^{-1})$$

- marginally

- $p(\boldsymbol{\beta} | \mathbf{x}, \mathbf{y})$ is multivariate t with n-k degrees of freedom
- $p(\sigma^2 | \mathbf{x}, \mathbf{y})$ is $IG\left(\frac{n-k}{2}, \frac{SSE}{2}\right)$

$$s^2 = \frac{SSE}{n-2}$$

- also:

- width of 95% c.i. for $\beta_0 = 2 t_{n-2} s.e.(\hat{\beta}_0)$
- width of 95% c.i. for $\beta_1 = 2 t_{n-2} s.e.(\hat{\beta}_1)$
- get t coefficients from t table

When is posterior proper with improper reference prior

- $n > k$
- columns of \mathbf{X} matrix are linearly independent

What if observations y_i are not conditionally independent, given β , σ^2 , \mathbf{x} ?

- hierarchical linear models
- time series models