Bayesian Simple Linear Regression

September 29, 2008

Reading HH 8, GIlI 4
Conjugate Priors for Regression

Model: \( Y_i \overset{ind}{\sim} N(\alpha + \beta x_i, \sigma^2) \)

Normal-Gamma distribution is conjugate for \( \alpha \) and \( \beta \) and precision \( \phi \equiv \frac{1}{\sigma^2} \)

- \((\alpha, \beta) | \phi \sim N((\alpha_0, \beta_0), \phi^{-1}\Sigma) \)
  \( \Sigma \) is a \( 2 \times 2 \) matrix of variances and covariance

- \( \phi \sim \text{Gamma}(\nu_0/2, \nu_0\sigma_0^2/2) \)

A Reference Prior for Regression:
Limiting case of conjugate prior as the prior variances goes to infinity (information goes to zero)

\[ p(\alpha, \beta, \phi) \propto \frac{1}{\phi} \]
Theory for Inference

\[ \frac{\beta - \hat{\beta}}{\sqrt{s^2_{Y|X} \frac{1}{s_{xx}}}} \sim t_{n-2}(0, 1) \]

- \( \hat{\beta} \) is OLS (MLE) estimate of \( \beta \), \( s^2_{Y|X} = \hat{\sigma}^2 \) is the MSE
- (marginal) posterior for \( \beta \) is a Student \( t \) distribution with \( n - 2 \) df.
- Sampling distribution of \( \hat{\beta} \) given \( \beta \) is Student \( t \) distribution with \( n - 2 \) df

Used for classical and Bayesian (Reference) analysis
(marginal) posterior for $\alpha$ is a Student $t$ distribution with $n - 2$ df.

Sampling distribution of $\hat{\alpha}$ is Student $t$ distribution with $n - 2$ df

\[
\frac{\alpha - \hat{\alpha}}{\sqrt{s^2_{Y|X} \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}} \sim t_{n-2}(0, 1)
\]
Significance of Regression

Measuring the “explanatory power” of predictor $X$

- Credible Intervals (HPD or equal-tailed):

$$\hat{\beta} \pm t_{\alpha/2} s_{\beta}$$

where $t_{\alpha/2}$ is 100 $\alpha/2\%$ quantile of a standard $t_{n-2}$

and $s_{\beta} = \sqrt{s_{Y|X}^2 / S_{xx}}$

- Confidence Intervals:

$$\hat{\beta} \pm t_{\alpha/2} \text{SE}(\hat{\beta})$$

and $\text{SE}(\hat{\beta}) = \sqrt{s_{Y|X}^2 / S_{xx}}$
### Body Fat Example

|                | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | -39.2802 | 2.6603     | -14.77  | 2.210^{-16} |
| Abdomen        | 0.6313   | 0.0286     | 22.11   | 2.210^{-16} |

95% HPD interval

\[
0.6313 \pm q_{t(.025, 250)} \times 0.0286 = (0.57, 0.69)
\]

For every additional cm of abdominal circumference, percent bodyfat increases by 0.57 to 0.69 percent with probability 0.95.
Significance of the Regression

Question: How probable is $\beta = 0$ under the posterior?

Informal Answer: Compute posterior probability on $\beta$ values with lower posterior density than $\beta = 0$

- “Measures” probability of $\beta$ “less likely” than $\beta = 0$
- Informal “test”: Probability in tails = significance level = (Bayesian) p-value

$$p-value = P(|t| > |\hat{\beta}/s_\beta|) = P(|t| > |\hat{\beta}/SE(\beta)|)$$

- Classical testing terminology:
  “The regression on $x$ is significant at the 5% level (or 1%, etc) if the p-value is smaller than 0.05 (or 0.01, etc)”
Lindley’s Method

Lindley suggested rejecting the hypothesis that $\beta = 0$ at the $\alpha 100\%$ level of significance if the $(1 - \alpha)100\%$ HPD region does not include 0.

$$0 \notin (\hat{\beta} - t_{1-\alpha/2} s_{\beta}, \hat{\beta} + t_{1-\alpha/2} s_{\beta})$$

Equivalent to comparing the p-value to $\alpha$ and concluding that the regression is significant if the p-value is less than $\alpha$.

Alternative approach is to compute a Bayes Factor.
Bayes Factors

Testing $H_0 : \beta = 0$ versus $H_a : \beta \neq 0$

- Assign prior probabilities to $H_0$ and $H_a$
- Find $P(H_i \mid Y)$ via Bayes Theorem (Ch 7)

Bayes Factor for comparing evidence in favor of $H_0$

$$BF[H_0 : H_a] = \frac{p(H_0 \mid Y)/p(H_o)}{p(H_a \mid Y)/p(H_a)}$$

Often difficult to calculate, instead use lower bound based on p-values (Berger, Selke and Bayarri)

$$BF[H_0 : H_a] = -ep \log(p)$$
Bodyfat Example

- $P(|t| > 22.11) = 2.2 \times 10^{-16}$
- The regression of bodyfat on abdominal circumference is highly significant (p-value = $2.2 \times 10^{-16}$).
- Lower bound on Bayes Factor
  \[ BF[H_0 : H_a] = 2.15 \times 10^{-14} \]
  \[-2.2 \times 10^{-16} \exp(1) \log(2.2 \times 10^{-16}) = 2.156043 \times 10^{-14} \]
- Approximate posterior probability of $H_0 = 2.15 \times 10^{-14}$
  \[ P(H_0 \mid Y) = \frac{BF[H_0 : H_a]O[H_0 : H_a]}{1 + BF[H_0 : H_a]O[H_0 : H_a]} \]
# Jeffreys Scale of Evidence

<table>
<thead>
<tr>
<th>Bayes Factor</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B \geq 1$</td>
<td>$H_0$ supported</td>
</tr>
<tr>
<td>$1 &gt; B \geq 10^{-\frac{1}{2}}$</td>
<td>minimal evidence against $H_0$</td>
</tr>
<tr>
<td>$10^{-\frac{1}{2}} &gt; B \geq 10^{-1}$</td>
<td>substantial evidence against $H_0$</td>
</tr>
<tr>
<td>$10^{-1} &gt; B \geq 10^{-2}$</td>
<td>strong evidence against $H_0$</td>
</tr>
<tr>
<td>$10^{-2} &gt; B$</td>
<td>decisive evidence against $H_0$</td>
</tr>
</tbody>
</table>

$B = \text{BF} [H_o : B_a]$

Decisive evidence against hypothesis that bodyfat is not associated with abdominal circumference.
F Tests, ANOVA and Deviances

Deviances = Sums of squares:
Deviance decomposition: TSS = SSE + SSR

\[ S_{yy} = Q(\hat{\alpha}, \hat{\beta}) + \hat{\beta}^2 / s^2_\beta \]

- Total deviance (TSS) \( S_{yy} = Q(\hat{\alpha}, 0) \sum_{i=1}^{n} (y_i - \bar{y})^2 \)
- Residual deviance (SSE) \( Q(\hat{\alpha}, \hat{\beta}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \)
- Explained deviance or Sum of Squares Regression: (SSR) = TSS - SSE = \( \hat{\beta}^2 S_{xx} \)
- Large deviance explained \( \equiv \) large \( F \) \( \equiv \) significant regression

ANOVA: analysis of variance (deviance)
Extra Sum of Squares F test of regression fit: SSR = TSS - SSE, the “extra” SS due to adding $X$ to the model. $F_{\text{obs}} = \frac{MSR}{MSE}$ where MSR = SSR/1

$$F_{\text{obs}} = \frac{\hat{\beta}^2}{s^2_{Y|X} \frac{1}{s_{xx}}} = t^2$$

Theory: If $t \sim t_k(0,1)$ then $F = t^2 \sim F_{1,n-2}$

- $P(F \geq F_{\text{obs}}) = P(|t| > |t_{\text{obs}}|)$
- $t$ and $F$ tests are equivalent: same $p-$value
**Example**

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdomen</td>
<td>1</td>
<td>11631.53</td>
<td>11631.53</td>
<td>488.93</td>
<td>2.2 × 10^{-16}</td>
</tr>
<tr>
<td>Residuals</td>
<td>250</td>
<td>5947.46</td>
<td>23.79</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: $F = t^2$ and same p-value
Predictions

The (posterior) predictive distribution for a new case, $y_{n+1} = \alpha + \beta x_{n+1} + \varepsilon_{n+1}$ is also a Student t distribution with $n - 2$ df.

$$y_{n+1} | y_1, \ldots, y_n \sim t_{n-2}(\hat{y}, s^2_{y_{n+1}})$$

$$\hat{y} = \hat{\alpha} + \hat{\beta} x_{n+1}$$

$$s^2_{y_{n+1}} = s^2_{Y|X} \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{S_{xx}}\right)$$

- posterior uncertainty about $\alpha + \beta x_{n+1}$
- depends on $x_{n+1}$ spread is higher for $x_{n+1}$ far from $\bar{x}$
- additional variability $+ s^2_{Y|X}$ due to $\varepsilon_{n+1}$
Intervals: `ci.plot(bodyfat.lm)`

95% confidence and prediction intervals for bodyfat.lm
Intervals: without case 39

95% confidence and prediction intervals for bodyfat.lm2