A Bayesic regression analysis

Henry Linder
Test

A classical (frequentist) approach

Consider a simple linear regression of tree height y_i on age x_i

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{1}$$

where ϵ_i are independent $N(0, \sigma^2)$ errors. For how it will pertain to the Bayesian analysis, note two things: first, we can equivalently write

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2) \tag{2}$$

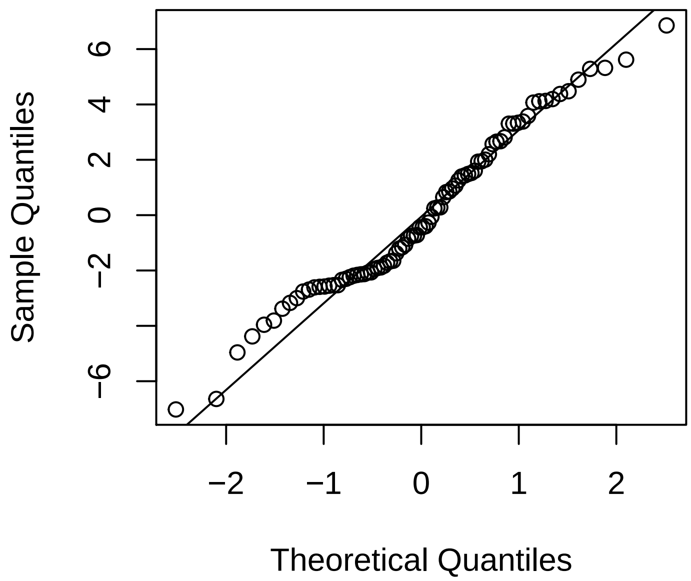
and second, this implies that the likelihood function for the data $\mathbf{y} = (y_1, \ldots, y_n)'$ is $f(\mathbf{y}|\theta, \mathbf{x})$ where the data $\mathbf{x} = (x_1, \ldots, x_n)'$ is assumed known and the parameter vector is $\theta = (\beta_0, \beta_1)'$.

A regression fit is given by

```r
m <- lm(height ~ age, data = Loblolly)
summary(m)
```

```
##
## Call:
## lm(formula = height ~ age, data = Loblolly)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0207 -2.1672 -0.4391  2.0539  6.8545
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.31240    0.62183  -2.111   0.0379 *
## age          2.59052    0.04094  63.272   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.947 on 82 degrees of freedom
## Multiple R-squared:  0.9799, Adjusted R-squared:  0.9797
## F-statistic:  4003 on 1 and 82 DF,  p-value: < 2.2e-16
```

Note, in particular, that the regression coefficient β_1 is highly significant, the R^2 is high, and a QQ plot indicates the residuals are approximately normal:



A Bayesian approach

A conventional, convenient, and conjugate choice in Bayesian regression gives independent normal priors to β_0 and β_1. For simplicity, let $\beta_0 \sim N(0, 1)$ and $\beta_1 \sim N(0, 1)$. A Bayesian analysis revolves around the fundamental relationship between the prior distribution $p(\theta)$ and the posterior distribution $p(\theta|\mathbf{y})$ of the parameter vector $\theta = (\beta_0, \beta_1)'$:

$$p(\theta|\mathbf{y}) \propto f(\mathbf{y}|\theta, \mathbf{x})p(\theta) \tag{3}$$

where $f(\mathbf{y}|\theta, \mathbf{x})$ is the likelihood of the observed data, and we take $\mathbf{x}$ as given. In the present case, we can write the likelihood as

$$f(\mathbf{y}|\theta, \mathbf{x}) = \prod_{i=1}^{n} f(y_i|\theta, x_i) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right] \tag{4}$$

where we omit the constant of proportionality for clarity.
Too, the prior for θ is

$$p(\theta) = p(\beta_0)p(\beta_1) \propto \exp\left[-\frac{1}{2}\left(\beta_0^2 + \beta_1^2\right)\right] \tag{5}$$

So, we can find the posterior distribution as

$$p(\theta|\mathbf{y}) \propto p(\beta_0|\mathbf{y})p(\beta_1|\mathbf{y}) \tag{6}$$