This is a title

Henry Linder
mhlinder@gmail.com

October 31, 2016

We consider the R dataset Loblolly, a data frame with measurements of height (feet) and age (year), for 84 Loblolly pines.

A citation: Gelman, Carlin, Stern, & Rubin (2014)

# 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 $\epsilon_i$ are independent $N(0, \sigma^2)$ errors. Note that we can equivalently write

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

and the likelihood function for the data $\mathbf{y} = (y_1, \ldots, y_n)'$ is $f(\mathbf{y}|\theta, \mathbf{x})$, with $f(\cdot|\theta, \mathbf{x})$ the density of a $N(\beta_0 + \beta_1 x_i, \sigma^2)$ distribution; the data $\mathbf{x} = (x_1, \ldots, x_n)'$ is assumed known; and $\theta = (\beta_0, \beta_1)'$ a parameter vector.

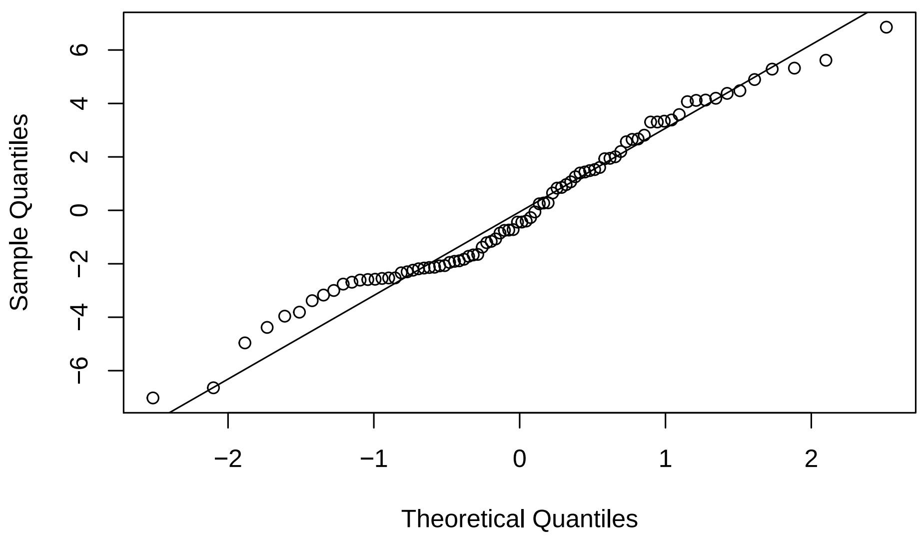
The least squares regression fit is

```
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 $\beta_1$ 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 $\beta_0$ and $\beta_1$. For simplicity, let $\beta_0 \sim N(0,1)$ and $\beta_1 \sim N(0,1)$, and we'll suppose the variance is known to be exactly the sample standard deviation, i.e., $\sigma^2 = 2.92894$.

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 joint prior for $\theta$ is

$$p(\theta) = p(\beta_0)p(\beta_1) \tag{5}$$

where $p(\cdot)$ is a standard normal density.

Then, the posterior distribution can be found by factoring the basic Bayes identity

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

It is possible, and only somewhat messy, to find an analytic solution for the posterior $p(\theta|\mathbf{y})$: a normal prior for normal linear regression coefficients is conventional because the posterior is still normal. In fact, as the number of observations increases asymptotically, the posterior mean values for $\theta$ will converge to $(\beta_0, \beta_1)'$.

Note that calculating the analytic form of the normal posterior can be performed without the normalizing constant, and hence written as a proportionality, because the form of the density uniquely determines the distribution, and only a single normalizing constant is valid. *This fact depends squarely on the prior being a proper probability density,* and will not hold for arbitrary improper prior. Because the parameters $\beta_1, \beta_2$ are valued on the entire real line, we might, for instance, choose to give an improper uniform prior, i.e., $\beta_0 \propto 1$. Then, it is necessary to make sure the posterior is valid by using the equality

$$p(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\beta_0, \beta_1, \mathbf{x})p(\theta)}{\int_{\Theta} f(\mathbf{y}|\beta_0, \beta_1, \mathbf{x})p(\theta)\, d\theta} \tag{7}$$

For all but the simplest models, this type of direct calculation can be tedious, or possibly analytically intractable. For that reason, much of the work can be practically side-stepped, by using the Gibbs sampler.

### Gibbs sampling

## References

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Chapman & Hall/CRC Boca Raton, FL, USA.