Bayesian linear regression

Dr. Jarad Niemi

STAT 544 - Iowa State University

April 12, 2016

Outline

- Linear regression
  - Classical regression
  - Default Bayesian regression
  - Conjugate subjective Bayesian regression
- Simulating from the posterior
  - Inference on functions of parameters
  - Posterior for optimum of a quadratic

Linear Regression

Basic idea

- understand the relationship between response y and explanatory variables $x = (x_1, \ldots, x_k)$
- based on data from experimental units index by i.

If we assume

- linearity, independence, normality, and constant variance,

then we have

$$y_i \stackrel{ind}{\sim} N(\beta_1 x_{i1} + \cdots + \beta_k x_{ik}, \sigma^2)$$

where $x_{i1} = 1$ if we want to include an intercept. In matrix notation, we have

$$y \sim N(X\beta, \sigma^2 \mathrm{I})$$

where $y = (y_1, \ldots, y_n)'$, $\beta = (\beta_1, \ldots, \beta_k)'$, and X is $n \times k$ with each row being $x_i = (x_{i1}, \ldots, x_{ik})$.

Classical regression

How do you find confidence intervals for β?

What is the MLE for β?

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

What is the sampling distribution for $\hat{\beta}$?

$$\hat{\beta} \sim t_{n-k}(\beta, s^2(X'X)^{-1})$$

where $s^2 = SSE/[n-k]$ and $SSE = (Y - X\hat{\beta})'(Y - X\hat{\beta})$.

What is the sampling distribution for s^2?

$$\frac{[n-k]s^2}{\sigma^2} \sim \chi^2_{n-k} \implies \frac{1}{s^2} \sim \text{Inv-}\chi^2(n-k, \sigma^2)$$

Default Bayesian regression

Assume the standard noninformative prior

$$p(\beta, \sigma^2) \propto 1/\sigma^2$$

then the posterior is

$$\begin{array}{rl}
p(\beta, \sigma^2|y) & = p(\beta|\sigma^2, y)p(\sigma^2|y) \\
\beta|\sigma^2, y & \sim N(\hat{\beta}, \sigma^2 V_\beta) \\
\sigma^2|y & \sim \text{Inv-}\chi^2(n-k, s^2) \\
\beta|y & \sim t_{n-k}(\hat{\beta}, s^2 V_\beta)
\end{array}$$

$$\begin{array}{rl}
\hat{\beta} & = (X'X)^{-1}X'y \\
V_\beta & = (X'X)^{-1} \\
s^2 & = \frac{1}{n-k}(y - X\hat{\beta})'(y - X\hat{\beta})
\end{array}$$

The posterior is proper if $n > k$ and $\text{rank}(X) = k$.

Comparison to classical regression

For the regression coefficients, in a default Bayesian regression, we have

$$\beta|y \sim t_{n-k}(\hat{\beta}, s^2(X'X)^{-1}),$$

and in classical regression analysis, we have

$$\hat{\beta} \sim t_{n-k}(\beta, s^2(X'X)^{-1}).$$

For the error variance, in a default Bayesian regression, we have

$$\sigma^2|y \sim \text{Inv-}\chi^2(n-k, s^2)$$

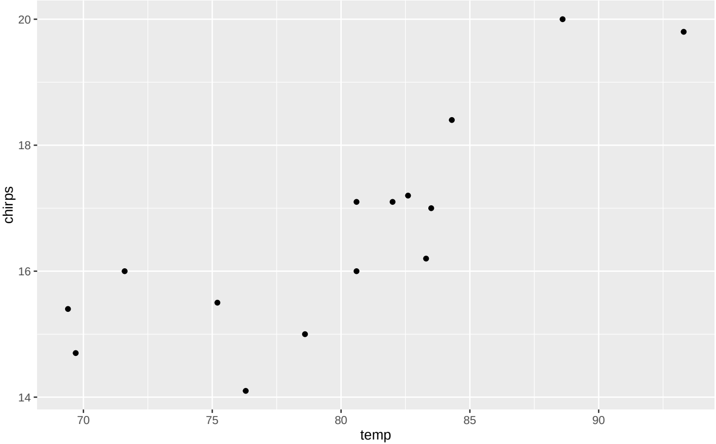
and in classical regression analysis, we have

$$1/s^2 \sim \text{Inv-}\chi^2(n-k, \sigma^2).$$

In the Bayesian statements, β is random and Y is fixed while in the classical statements, Y is random and β is fixed.

Cricket chirps

As an example, consider the relationship between the number of cricket chirps (in 15 seconds) and temperature (in Fahrenheit). From example in `LearnBayes::blinreg`.



Default Bayesian regression

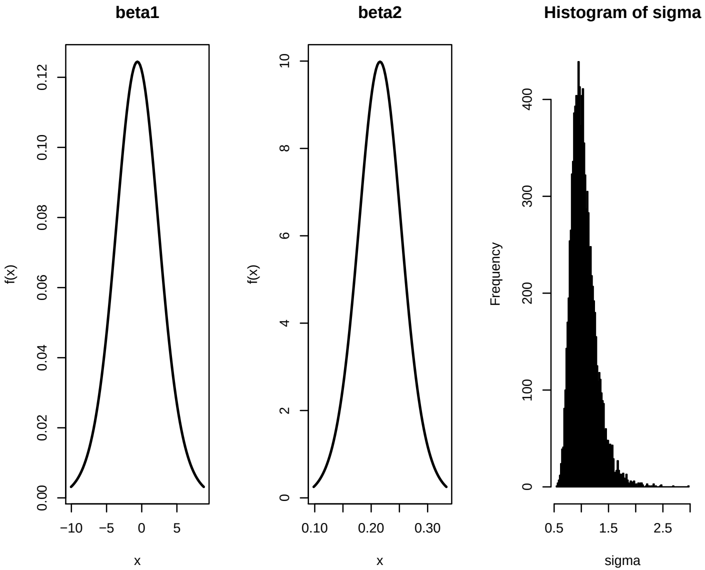
```
summary(m <- lm(chirps~temp))
```

```

Call:
lm(formula = chirps ~ temp)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.74107 -0.58123  0.02956  0.58250  1.50608 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.61521    3.14434  -0.196 0.847903    
temp         0.21568    0.03919   5.504 0.000102 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9849 on 13 degrees of freedom
Multiple R-squared:  0.6997, Adjusted R-squared:  0.6766 
F-statistic: 30.29 on 1 and 13 DF,  p-value: 0.0001015
```

```
confint(m) # Credible intervals
```

```
                 2.5 %    97.5 %
(Intercept) -7.4081577 6.1777286
temp         0.1310169 0.3003406
```

Default Bayesian regression - Full posteriors



Fully conjugate subjective Bayesian inference

If we assume the following normal-gamma prior,

$$\beta|\sigma^2 \sim N(m_0, \sigma^2 C_0) \qquad \sigma^2 \sim \text{Inv-}\chi^2(v_0, s_0^2)$$

then the posterior is

$$\beta|\sigma^2, y \sim N(m_n, \sigma^2 C_n) \qquad \sigma^2|y \sim \text{Inv-}\chi^2(v_n, s_n^2)$$

with

$$\begin{array}{rl} m_n & = m_0 + C_0 X'(XC_0X' + \mathrm{I})^{-1}(y - Xm_0) \\ C_n & = C_0 - C_0X'(XC_0X' + \mathrm{I})^{-1}XC_0 \\ v_n & = v_0 + n \\ v_n s_n^2 & = v_0 s_0^2 + (y - Xm_0)'(XC_0X' + \mathrm{I})^{-1}(y - Xm_0) \end{array}$$

Information about chirps per 15 seconds

Let

- Y_i is the average number of chirps per 15 seconds and
- X_i is the temperature in Fahrenheit.

And we assume

$$Y_i \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_i, \sigma^2)$$

then

- β_0 is the expected number of chirps at 0 degrees Fahrenheit
- β_1 is the expected increase in number of chirps (per 15 seconds) for each degree increase in Fahrenheit.

Based on prior experience the prior $\beta_1 \sim N(0,1)$ might be reasonable.

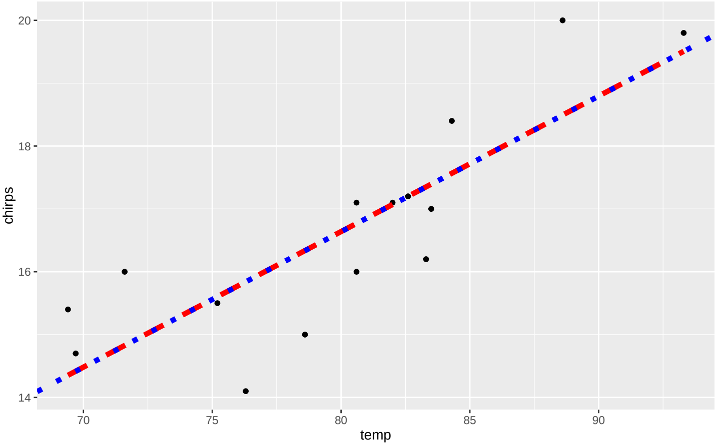
Subjective Bayesian regression

```
m = arm::bayesglm(chirps~temp,   # Default prior for \beta_0 is N(0,Inf)
                  prior.mean=0,  # E[\beta_1]
                  prior.scale=1, # V[\beta_1]
                  prior.df=Inf)  # normal prior
summary(m)



Call:
arm::bayesglm(formula = chirps ~ temp, prior.mean = 0, prior.scale = 1,
    prior.df = Inf)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.73940  -0.57939   0.03139   0.58435   1.50809

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.61478    3.14415  -0.196 0.847999
temp         0.21565    0.03919   5.503 0.000102 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.9700575)

    Null deviance: 41.993  on 14  degrees of freedom
Residual deviance: 12.611  on 13  degrees of freedom
AIC: 45.966

Number of Fisher Scoring iterations: 11
```

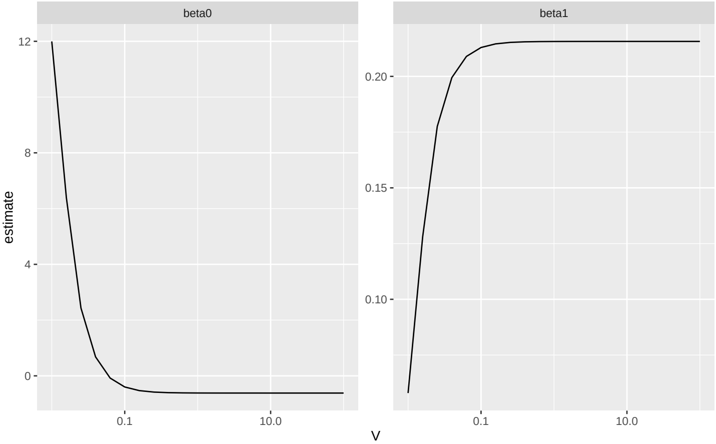
Subjective vs Default

```
# Subjective analysis
m$coefficients

(Intercept)        temp
 -0.6147847   0.2156511

confint(m)

                2.5 %    97.5 %
(Intercept) -6.7780731 5.5476365
temp         0.1388701 0.2924879

# compared to default analysis
tmp = lm(chirps~temp)
tmp$coefficients

(Intercept)        temp
 -0.6152146   0.2156787

confint(tmp)

                2.5 %    97.5 %
(Intercept) -7.4081577 6.1777286
temp         0.1310169 0.3003406
```

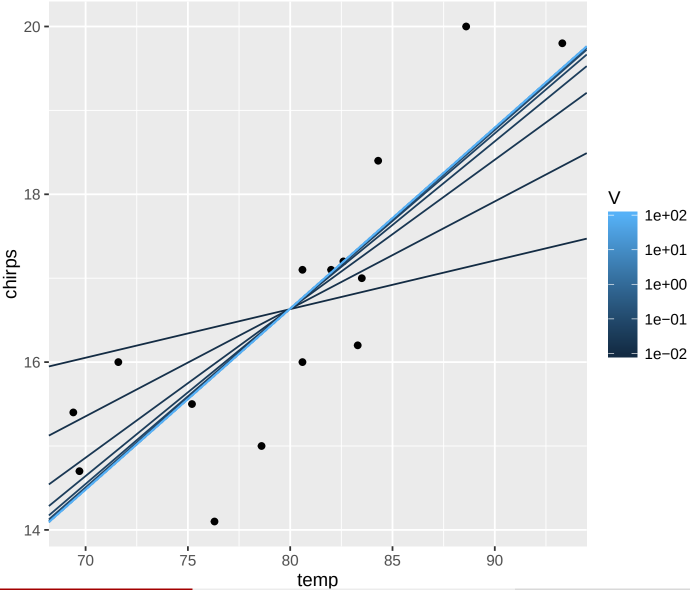
Subjective vs Default



Shrinkage (as $V[\beta_1]$ gets smaller)



Shrinkage (as $V[\beta_1]$ gets smaller)



Simulating from the posterior

Although the full posterior for β and σ^2 is available, the decomposition

$$p(\beta, \sigma^2|y) = p(\beta|\sigma^2, y)p(\sigma^2|y)$$

suggests an approach to simulating from the posterior via

1. $(\sigma^2)^{(j)} \sim \text{Inv-}\chi^2(n-k, s^2)$ and
2. $\beta^{(j)} \sim N(\hat{\beta}, (\sigma^2)^{(j)} V_\beta)$.

This also provides an approach to obtaining posteriors for any function $\gamma = f(\beta, \sigma^2)$ of the parameters via

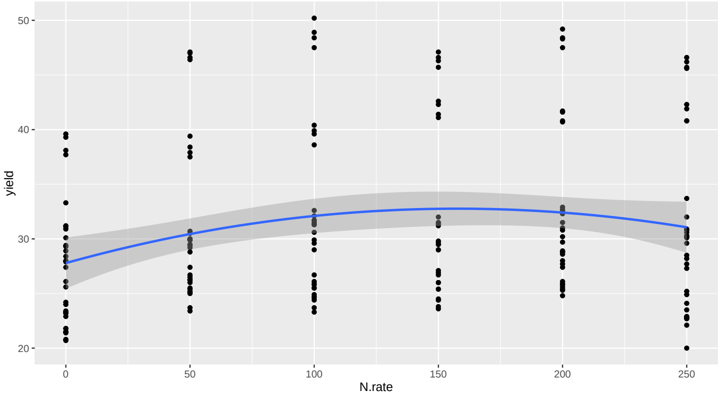
$$\begin{array}{rl} p(\gamma|y) & = \int\int p(\gamma|\beta,\sigma^2,y)p(\beta|\sigma^2,y)p(\sigma^2|y)d\beta d\sigma^2 \\ & = \int\int p(\gamma|\beta,\sigma^2)p(\beta|\sigma^2,y)p(\sigma^2|y)d\beta d\sigma^2 \\ & = \int\int \mathrm{I}(\gamma = f(\beta,\sigma^2))p(\beta|\sigma^2,y)p(\sigma^2|y)d\beta d\sigma^2 \end{array}$$

by adding the step

3. $\gamma^{(j)} = f(\beta^{(j)}, (\sigma^2)^{(j)})$.

Posterior for global maximum

Consider this potato yield data set



with a goal of estimating the optimal nitrogen rate.

Posterior for global maximum

Let

- Y_i be the potato yield and
- X_i be the nitrogen rate.

We assume the model

$$Y_i \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_i + \beta_2 X_i^2, \sigma^2)$$

Assuming this quadratic curve is correct, the maximum occurs at $\gamma = -\beta_1/[2\beta_2]$.

```
m = LearnBayes::blinreg(d$yield, cbind(1,d$N.rate, d$N.rate^2), 1e4)
beta1 = m$beta[,2]; beta2 = m$beta[,3]; gamma = -beta1/(2*beta2)
round(quantile(gamma, c(.025,.5,.975)))

 2.5%   50% 97.5%
  124   157   282
```

This does not require any data asymptotics or approximations, e.g. delta method.

Summary

- Model: $y \sim N(X\beta, \sigma^2 \mathrm{I})$
- Default Bayesian analysis corresponds exactly to classical regression analysis

$$p(\beta, \sigma^2) \propto 1/\sigma^2 \implies$$

$$\beta|\sigma^2, y \sim N(\hat{\beta}, \sigma^2[X'X]^{-1}), \sigma^2|y \sim \text{Inv-}\chi^2(n-k, s^2)$$

- Conjugate subjective Bayesian analysis:

$$\beta|\sigma^2 \sim N(m_0, \sigma^2 C_0), \sigma^2 \sim \text{Inv-}\chi^2(v_0, s_0^2) \implies$$

$$\beta|\sigma^2, y \sim N(m_n, \sigma^2 C_n), \sigma^2|y \sim \text{Inv-}\chi^2(v_n, s_n^2)$$

- Obtain functions of parameters and their uncertainty by simulating the parameters from their joint posterior, calculating the function, and taking posterior quantiles.

Computation

For numerical stability and efficiency, the QR decomposition can be used to calculate posterior quantities.

Definition

For an $n \times k$ matrix X, a QR decomposition is $X = QR$ for an $n \times k$ matrix Q with orthonormal columns and a $k \times k$ upper triangular matrix R.

The quantities of interest are

$$\begin{aligned} V_\beta &= (X'X)^{-1} = ([QR]'QR)^{-1} = (R'Q'QR)^{-1} = (R'R)^{-1} \\ &= R^{-1}[R']^{-1} \end{aligned}$$

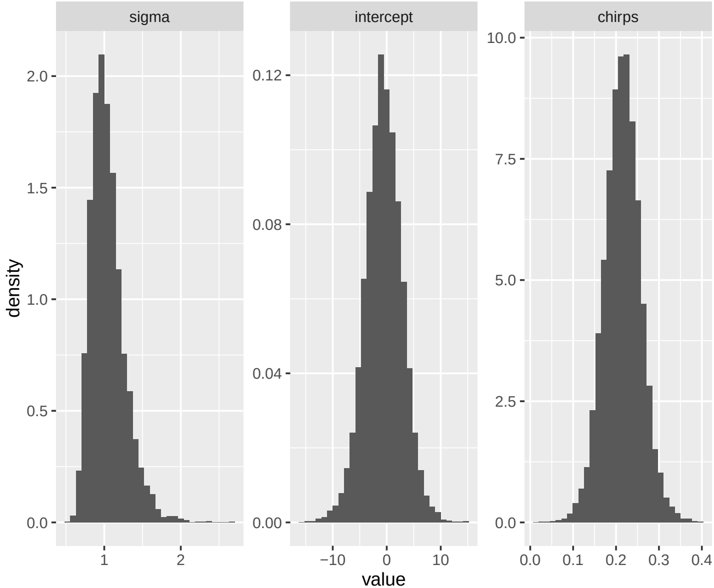
$$\begin{aligned} \hat\beta &= (X'X)^{-1}X'y = R^{-1}[R']^{-1}R'Q'y = R^{-1}Q'y \\ R\hat\beta &= Q'y \end{aligned}$$

The last equation is useful because R is upper triangular and therefore the system of linear equations can be solved without requiring the inverse of R.

Cricket chirps

```
library(MASS)
X = cbind(1,temp)
n = nrow(X)
k = ncol(X)
y = matrix(chirps,n,1)

qr = qr(X); Q = qr.Q(qr); R = qr.R(qr)
stopifnot(all.equal(X, Q%*%R),
          all.equal(rep(1,k), colSums(Q^2)),
          all.equal(diag(nrow=k), t(Q)%*%Q))

# Check for posterior propriety
stopifnot(n>k,qr$rank==k)

# Calculate posterior hyperparameters
Rinv = solve(qr.R(qr))
vbeta = Rinv%*%t(Rinv)
betahat = qr.solve(qr,y)
df = n-k
e = qr.resid(qr,y)
s2 = sum(e^2)/df

# Simulate from the posterior
n.sims = 10000
sigma = sqrt(1/rgamma(n.sims, df/2, df*s2/2))
beta = matrix(betahat, n.sims, k, byrow=T) + sigma * mvrnorm(n.sims, rep(0,k), vbeta)
```

Cricket chirps



Monte Carlo error

```
# sigma^2
sqrt(df*s2/qchisq(c(.975,.025),df)) # Exact

[1] 0.7140166 1.5867368

quantile(sigma,c(.025,.975)) # MC

     2.5%     97.5%
0.7147342 1.5801627

# beta
confint(lm(chirps~temp)) # Exact

                 2.5 %    97.5 %
(Intercept) -7.4081577 6.1777286
temp         0.1310169 0.3003406

t(apply(beta, 2, quantile, probs=c(.025,.975))) # MC

           2.5%     97.5%
     -7.4576088 6.1585177
temp  0.1309921 0.3010475
```