STAT 431 — Applied Bayesian Analysis — Course Notes

Regression Models

Spring 2019

Consider regression of a response Y (random) on a predictor X (fixed).

Data come in pairs

$$(x_1, Y_1), \quad (x_2, Y_2), \quad \ldots \quad (x_n, Y_n)$$

Let $\boldsymbol{y}$ be the vector of the y_i values (observed Y_i values).

Predictor Centering

It is customary to **center** the predictor, i.e. to use

$$x_i^{\text{cent}} \;=\; x_i - \bar{x}$$

where $\bar{x}$ is its sample mean.

Advantages:

- Computationally: may improve Gibbs sampler mixing (because regression parameters are less correlated)
- Analytically: makes it easier to define and implement (semi-)conjugate priors (simpler expressions)

Linear Regression

$$Y_i \;=\; \beta_0 \;+\; \beta_1(x_i - \bar{x}) \;+\; \varepsilon_i \qquad i = 1, \ldots, n$$

$$\varepsilon_i \;\sim\; \text{i.i.d. } \mathrm{N}(0, \sigma^2)$$

Alternatively (but equivalently),

$$Y_i \mid \beta_0, \beta_1, \sigma^2 \;\sim\; \text{indep. } \mathrm{N}\big(\beta_0 + \beta_1(x_i - \bar{x}),\ \sigma^2\big)$$

Standard noninformative prior:

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

that is,

$$\beta_0, \beta_1, \sigma^2 \quad \sim \quad \frac{1}{\sigma^2}\, d\beta_0\, d\beta_1\, d\sigma^2$$

If $\hat{\beta}_0$ and $\hat{\beta}_1$ are the usual ordinary least squares estimates of β_0 and β_1, and

$$SSR = \sum_{i=1}^{n} \big(y_i - \hat{\beta}_0 - \hat{\beta}_1(x_i - \bar{x})\big)^2$$

then the posterior (under the noninformative prior) is

$$\sigma^2 \mid \boldsymbol{y} \sim \mathrm{IG}\big((n-2)/2,\, SSR/2\big)$$

$$\left.\begin{aligned} \beta_0 \mid \sigma^2, \boldsymbol{y} &\sim \mathrm{N}\big(\hat{\beta}_0,\, \sigma^2/n\big) \\ \beta_1 \mid \sigma^2, \boldsymbol{y} &\sim \mathrm{N}\big(\hat{\beta}_1,\, \sigma^2/\textstyle\sum(x_i - \bar{x})^2\big) \end{aligned}\right\} \text{conditionally independent}$$

(See Cowles for derivation.)

Since the pairs (β_0, σ^2) and (β_1, σ^2) have normal-inverse gamma posteriors, the posterior marginals for β_0 and β_1 are

$$\begin{aligned} \beta_0 \mid \boldsymbol{y} &\sim t\big(\hat{\beta}_0,\, s^2/n,\, n-2\big) \\ \beta_1 \mid \boldsymbol{y} &\sim t\big(\hat{\beta}_1,\, s^2/\textstyle\sum(x_i - \bar{x})^2,\, n-2\big) \end{aligned}$$

where

$$s^2 \;=\; SSR \,/\, (n-2)$$

Accordingly, the posterior credible intervals turn out to be equivalent to confidence intervals (just as in the case of a normal sample).

(See Cowles for a data example.)

Remarks:

- There is also a (fully) conjugate prior, based on normal-inverse gamma distributions.
- Normal priors are semi-conjugate for β_0 and β_1, and an inverse gamma prior is semi-conjugate for σ^2.
- As in the normal sample model, a (multivariate) Jeffreys prior exists, but is rarely used.

GLM Regression

Idea: Express a mean-related parameter of the model distribution of Y as a (transformed) linear regression on X.

Eg: Logistic Regression

$$Y_i \mid \pi_i \quad \sim \quad \text{binomial}(n_i, \pi_i)$$

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) \quad = \quad \text{logit}(\pi_i) \quad = \quad \beta_0 \; + \; \beta_1(x_i - \bar{x})$$

Eg: Poisson Loglinear Regression

$$\begin{aligned} Y_i \mid \lambda_i &\sim \text{Poisson}(\lambda_i) \\ \ln(\lambda_i) &= \beta_0 + \beta_1(x_i - \bar{x}) \end{aligned}$$

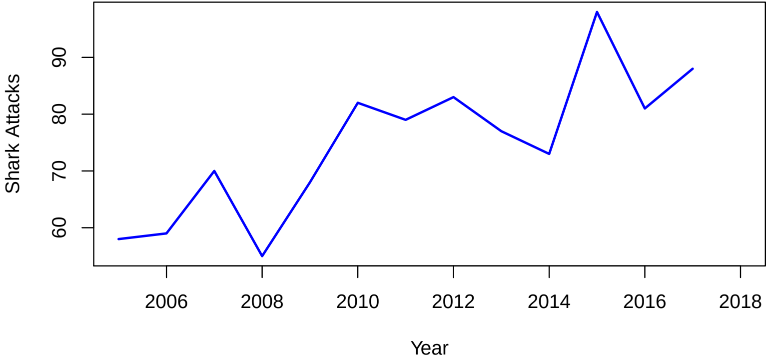
Priors on β_0 and β_1 can be specified similarly to linear regression, e.g.

$$\beta_0, \beta_1 \sim 1\, d\beta_0\, d\beta_1$$

Example: Shark Attacks

$$Y_i = \text{number of shark attacks (worldwide)}$$

$$x_i = \text{year (2005–2017)}$$



- Are shark attacks becoming more frequent?
- How many were predicted for 2018? (Actual: 66)

Since attacks are "rare" and usually unrelated, suppose

$$Y_i \mid \lambda_i \quad \sim \quad \text{indep. Poisson}(\lambda_i)$$

$$\ln(\lambda_i) \quad = \quad \beta_0 + \beta_1(x_i - \bar{x})$$

We will choose “vague” but proper priors:

$$\beta_0, \beta_1 \quad \sim \quad \text{indep. N}(0, 100^2)$$

[Draw preliminary model graph ...]

The data:

```
x     y
2005  58
2006  59
2007  70
2008  55
2009  68
2010  82
2011  79
2012  83
2013  77
2014  73
2015  98
2016  81
2017  88
2018  NA
```

Note the response of `NA` for the year 2018.

The "missing" Y value for 2018 will be sampled as an unobserved random node, to give its posterior predictive distribution.

The JAGS code:

```
data {
  xmean <- mean(x[1:(length(x)-1)])
  for(i in 1:length(x)) {
    xcent[i] <- x[i] - xmean
  }
}
model {
  for(i in 1:length(y)) {
    y[i] ~ dpois(lambda[i])
    log(lambda[i]) <- beta0 + beta1 * xcent[i]
  }

  beta0 ~ dnorm(0, 0.0001)
  beta1 ~ dnorm(0, 0.0001)

  beta1.gt.0 <- beta1 > 0
}
```

Notes:

- To get the centered version of x (`xcent`), we subtract `xmean`.

  Only the observed cases (`1:(length(x)-1)`) are used in the mean.

- We define `beta1.gt.0` so we can calculate the posterior probability that $\beta_1 > 0$.

R/JAGS Example 10.1:

Poisson Regression

Hierarchical Normal Regression

Now suppose that, in addition to X and Y, there is a grouping variable.

Let

$$
\begin{aligned}
Y_{ij} &= \text{response of } j\text{th observation in group } i \\
x_{ij} &= \text{its predictor value}
\end{aligned}
$$

Let

$$\bar{x} = \text{average of } \textit{all } x_{ij} \text{ values}$$

(We will use the same covariate centering for all groups.)

Each group can have its own regression line:

$$Y_{ij} \;=\; \alpha_{0i} + \alpha_{1i}(x_{ij} - \bar{x}) + \varepsilon_{ij}$$

$$\varepsilon_{ij} \;\sim\; \text{i.i.d. } \mathrm{N}(0, \sigma_y^2)$$

The model becomes

$$Y_{ij} \mid \alpha_{0i}, \alpha_{1i}, \sigma_y^2 \quad \sim \quad \text{indep. } \mathrm{N}\big(\alpha_{0i} + \alpha_{1i}(x_{ij} - \bar{x}),\ \sigma_y^2\big)$$

A semi-conjugate prior for the variance:

$$\sigma_y^2 \sim \mathrm{IG}(a_y, b_y)$$

We will assume it is independent of the other parameters.

Two potential prior formulations for α_{0i} and α_{1i}:

- Univariate: assumes α_{0i} and α_{1i} are (a priori) independent
- Bivariate: allows (conditional) prior correlations between α_{0i} and α_{1i}

Correlations between α_{0i} and α_{1i} are frequently encountered ...

[Illustrate with regression lines ...]

Univariate Formulation

$$\left.\begin{array}{rcl} \alpha_{0i} \mid \beta_0, \sigma^2_{\alpha_0} & \sim & \mathrm{N}(\beta_0, \sigma^2_{\alpha_0}) \\ \alpha_{1i} \mid \beta_1, \sigma^2_{\alpha_1} & \sim & \mathrm{N}(\beta_1, \sigma^2_{\alpha_1}) \end{array}\right\} \begin{array}{c} \text{all} \\ \text{conditionally} \\ \text{independent} \end{array}$$

$$\left.\begin{array}{rcl} \beta_0 & \sim & \mathrm{N}(\mu_0, \sigma^2_0) \\ \beta_1 & \sim & \mathrm{N}(\mu_1, \sigma^2_1) \\ \sigma^2_{\alpha_0} & \sim & \mathrm{IG}(a_{\alpha_0}, b_{\alpha_0}) \\ \sigma^2_{\alpha_1} & \sim & \mathrm{IG}(a_{\alpha_1}, b_{\alpha_1}) \end{array}\right\} \text{independent}$$

[Draw model graph ...]

Example: Baby Rat Weights

$$
\begin{aligned}
Y_{ij} &= \text{mass of rat } i \text{ (g?) at } j\text{th measurement} \\
x_{ij} &= \text{age of rat } i \text{ (days) at } j\text{th measurement}
\end{aligned}
$$

The measurements were synchronous:

$$
x_{ij} = x_j \qquad (8,\ 15,\ 22,\ 29,\ \text{or } 36)
$$

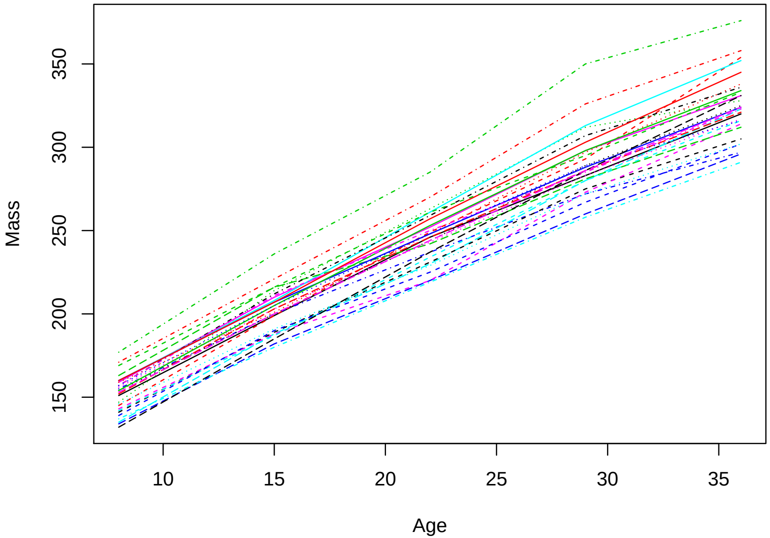
so $\bar{x} = 22$.

Rat mass values in file `ex10.2data.txt` (truncated):

```
151   199   246   283   320
145   199   249   293   354
147   214   263   312   328
155   200   237   272   297
135   188   230   280   323
159   210   252   298   331
...
```

Each row is a different rat, and each column is a different age.

Plot of the “growth curves”:



We can create the necessary data objects in R:

```
ages <- c(8, 15, 22, 29, 36)

d <- list(Y = read.table("ex10.2data.txt"),
          x = ages,
          xbar = mean(ages))
```

`Y` is a data frame, but it can be indexed like a matrix.

The JAGS code:

```
data {
  dim.Y <- dim(Y)
}
model {
  for(i in 1:dim.Y[1]) {

    for(j in 1:dim.Y[2]) {
      Y[i,j] ~ dnorm(mu[i,j], tausq.y)
      mu[i,j] <- alpha0[i] + alpha1[i] * (x[j] - xbar)
    }

    alpha0[i] ~ dnorm(beta0, tausq.alpha0)
    alpha1[i] ~ dnorm(beta1, tausq.alpha1)
  }

  tausq.y ~ dgamma(0.001, 0.001)
  sigma.y <- 1 / sqrt(tausq.y)

  beta0 ~ dnorm(0.0, 1.0E-6)
  beta1 ~ dnorm(0.0, 1.0E-6)
  tausq.alpha0 ~ dgamma(0.001, 0.001)
  tausq.alpha1 ~ dgamma(0.001, 0.001)
  sigma.alpha0 <- 1 / sqrt(tausq.alpha0)
  sigma.alpha1 <- 1 / sqrt(tausq.alpha1)
}
```

R/JAGS Example 10.2:

Hierarchical Normal Regression:
Univariate Formulation

Bivariate Formulation

$$\boldsymbol{\alpha}_i = \begin{bmatrix} \alpha_{0i} \\ \alpha_{1i} \end{bmatrix} \Bigg| \; \boldsymbol{\beta}, \boldsymbol{\Sigma}_\alpha \quad \sim \quad \text{i.i.d. } \mathrm{N}_2(\boldsymbol{\beta}, \boldsymbol{\Sigma}_\alpha)$$

where

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad\qquad \boldsymbol{\Sigma}_\alpha = \begin{bmatrix} \sigma^2_{\alpha_0} & \sigma_{\alpha_{01}} \\ \sigma_{\alpha_{01}} & \sigma^2_{\alpha_1} \end{bmatrix}$$

But we probably want to let $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_\alpha$ be chosen by the data, rather than arbitrarily specified, so we add another prior level ...

A semi-conjugate hyperprior specification:

$$\left.\begin{array}{rcl} \boldsymbol{\beta} & \sim & \mathrm{N}_2(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \\ \boldsymbol{\Sigma}_\alpha^{-1} & \sim & \mathrm{Wishart}_2(\boldsymbol{\Omega}, \nu) \end{array}\right\} \text{ independent}$$

where

$$\boldsymbol{\mu}_0 \quad \text{is a } 2 \times 1 \text{ vector}$$

$$\boldsymbol{\Sigma}_0 \quad \text{and} \quad \boldsymbol{\Omega} \quad \text{are } 2 \times 2 \text{ matrices (positive definite)}$$

$$\text{and} \quad \nu \quad \text{is a positive scalar.}$$

The $\mathrm{Wishart}_p$ distribution generalizes the gamma distribution to $p \times p$ non-negative definite matrices (see Cowles, Sec. 10.4.7).

Remarks:

- Need $\nu > p - 1$ for the $\mathrm{Wishart}_p$ distribution to be non-degenerate.

  This suggests $\nu = p$ might be a good choice — not quite "vague," but at least has relatively little information.

- JAGS and OpenBUGS use a special parameterization of the Wishart — see Cowles.

  In this parameterization, $\mathbf{\Omega}$ is like a prior location parameter for $\mathbf{\Sigma}_\alpha$, not for $\mathbf{\Sigma}_\alpha^{-1}$.

[Draw model graph ...]

Example: Baby Rat Weights (continued)

As before, we define the data in R.

We also add to the data some objects to help specify the prior:

```
ages <- c(8, 15, 22, 29, 36)

d <- list(Y = read.table("ex10.3data.txt"),
          x = ages,
          xbar = mean(ages),
          Omega = rbind(c(200, 0),
                        c(0, 0.2)),
          mu0 = c(0,0),
          Sigma0.inv = rbind(c(1.0E-6, 0),
                             c(0, 1.0E-6)))
```

(File `ex10.3data.txt` contains the mass data, as before.)

The JAGS code:

```
data {
  dim.Y <- dim(Y)
}
model {
  for(i in 1:dim.Y[1]) {

    for(j in 1:dim.Y[2]) {
      Y[i,j] ~ dnorm(mu[i,j], tausq.y)
      mu[i,j] <- alpha[i,1] + alpha[i,2] * (x[j] - xbar)
    }

    alpha[i,1:2] ~ dmnorm(beta, Sigma.alpha.inv)
  }

  tausq.y ~ dgamma(0.001, 0.001)
  sigma.y <- 1 / sqrt(tausq.y)

  beta ~ dmnorm(mu0, Sigma0.inv)
  Sigma.alpha.inv ~ dwish(Omega, 2)
  Sigma.alpha <- inverse(Sigma.alpha.inv)

  rho <- Sigma.alpha[1,2] / sqrt(Sigma.alpha[1,1] * Sigma.alpha[2,2])
}
```

R/JAGS Example 10.3:

Hierarchical Normal Regression: Bivariate Formulation