Bayesian Multiple Regression

Dr. Daniel B. Rowe
Professor of Computational Statistics
Department of Mathematical and Statistical Sciences
Marquette University

Outline

Likelihood Distribution

Conjugate Estimation (Normal-Inverse Gamma)

Non-Conjugate Estimation 1 (Laplace-Inverse Gamma)

Non-Conjugate Estimation 2 (Normal/Laplace-Inverse Gamma)

Regularization

Discussion

Homework

Likelihood Distribution

We often observe a continuous random variable y that we believe has a linear (in the parameters) relationship to a (several) fixed known value(s) x (x's).

This can be motivated by a Taylor series and can be $y=\beta_0+\beta_1 x+\ldots+\beta_q x^q$.

So y is a linear function of x, $y=\beta_0+\beta_1 x$

that we observe with measurement error

$y=\beta_0+\beta_1 x+\varepsilon$.

Likelihood Distribution

We often observe n, data points, $(x_1,y_1),\ldots,(x_n,y_n)$ and aim to describe the relationship between x and y as $y_i=\beta_0+\beta_1 x_i+\varepsilon_i$, where $\varepsilon_i \sim N(0,\sigma^2)$ for $i=1,\ldots,n$.

This linear in the parameters model can be written as

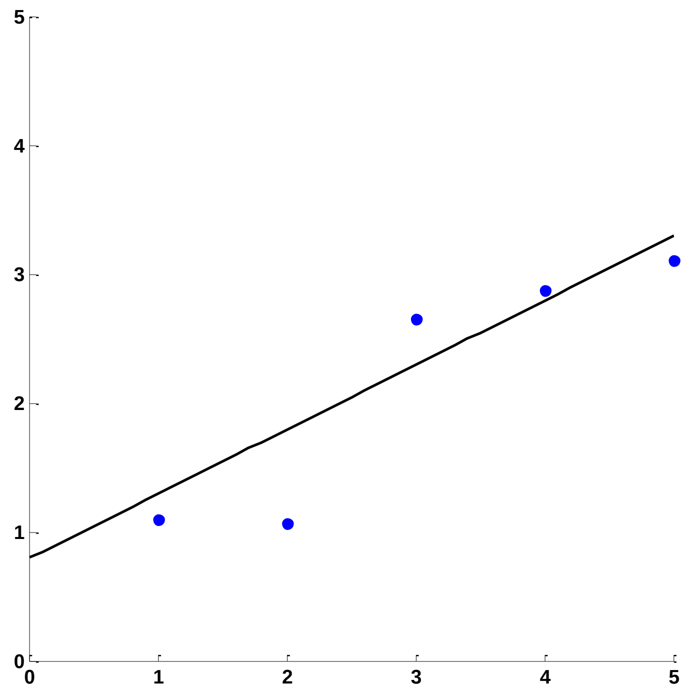
$$\underset{n\times 1}{y} = \underset{n\times(q+1)}{X}\ \underset{(q+1)\times 1}{\beta} + \underset{n\times 1}{\varepsilon}$$

$q=1$

where $\varepsilon \sim N(0,\sigma^2 I_n)$.

$$\underset{n\times 1}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \underset{n\times(q+1)}{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & \vdots \\ 1 & x_n \end{pmatrix} \quad \underset{(q+1)\times 1}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \quad \underset{n\times 1}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Likelihood Distribution



The linear regression model

$$\underset{n\times 1}{y} = \underset{n\times(q+1)}{X} \; \underset{(q+1)\times 1}{\beta} + \underset{n\times 1}{\varepsilon}, \quad \varepsilon \sim N(0,\sigma^2 I_n)$$

has likelihood

$$f(\underset{n\times 1}{y} \mid \underset{(q+1)\times 1}{\beta}, \sigma^2) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)} \qquad \beta \in \mathbb{R}^{q+1}, \sigma^2 \in \mathbb{R}^+$$

and MLEs $\underset{(q+1)\times 1}{\hat{\beta}_{MLE}} = (X'X)^{-1}X'y$ and $\hat{\sigma}^2_{MLE} = \frac{1}{n}(y - X\hat{\beta}_{MLE})'(y - X\hat{\beta}_{MLE})$.

Note MLE is not n-1.

Conjugate Estimation

With a normal distribution for the observations and likelihood,

$$f(\underset{n\times 1}{y}|\beta,\sigma^2)=(2\pi\sigma^2)^{-n/2}e^{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)}$$

similarities

Conjugate Prior For Normal RVs, UnKnown σ^2

$$f(x_1,\ldots,x_n|\mu)=(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

the conjugate prior distributions are

$$f(\underset{(q+1)\times 1}{\beta}|\beta_0,\sigma^2,n_0)=(2\pi\sigma^2/n_0)^{-(q+1)/2}e^{-\frac{n_0}{2\sigma^2}(\beta-\beta_0)'(\beta-\beta_0)}$$

$$f(\mu|\mu_0,n_0)=(2\pi\sigma^2/n_0)^{-1/2}\exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma^2/n_0}\right\}$$

a priori uncorrelated $\frac{\sigma^2}{n_0}I_{q+1}$

could have chosen $\frac{\sigma^2}{n_0}(X_0'X_0)^{-1}$

$$f(\sigma^2|h,\nu)=\frac{h^{\frac{\nu-2}{2}}}{2^{\frac{\nu-2}{2}}\Gamma(\frac{\nu-2}{2})}(\sigma^2)^{-\nu/2}e^{-\frac{h}{2\sigma^2}}$$

$$f(\sigma^2|h,\nu)=\frac{h^{(\nu-2)/2}(\sigma^2)^{-\nu/2}}{\Gamma(\frac{\nu-2}{2})2^{(\nu-2)/2}}\exp\left\{-\frac{h}{2\sigma^2}\right\}$$

and the posterior is

$$f(\beta,\sigma^2|y,\cdot)\propto(\sigma^2)^{-(n+q+1+\nu)/2}e^{-\frac{1}{2\sigma^2}[(y-X\beta)'(y-X\beta)+n_0(\beta-\beta_0)'(\beta-\beta)+h]}$$

.

Conjugate Estimation

With the normal likelihood and conjugate prior distribution

$$f(\beta,\sigma^2 \mid y,\cdot) \propto (\sigma^2)^{-(n+q+1+\nu)/2} e^{-\frac{1}{2\sigma^2}[(y-X\beta)'(y-X\beta)+n_0(\beta-\beta_0)'(\beta-\beta_0)+h]}$$

some algebra can be performed to arrive at

$$(y-X\beta)'(y-X\beta)+(\beta-\beta_0)'(\beta-\beta_0)+h$$
$$=(\beta-\hat{\beta})'(n_0 I_{q+1}+X'X)(\beta-\hat{\beta})+n_0\beta_0'\beta_0+h$$
$$+y'y-(n_0\beta_0+X'y)'(n_0 I_{q+1}+X'X)^{-1}(n_0\beta_0+X'y)$$

and posterior

$$f(\beta,\sigma^2 \mid y,\cdot) \propto (\sigma^2)^{-(n+q+1+\nu)/2} e^{-\frac{1}{2\sigma^2}\underbrace{[(\beta-\hat{\beta})'(n_0 I_{q+1}+X'X)(\beta-\hat{\beta})+\omega]}_{h_*}}$$

$$\hat{\beta}=(n_0 I_{q+1}+X'X)^{-1}(n_0\beta_0+X'X\hat{\beta}_{MLE})$$
$$\omega=n_0\beta_0'\beta_0+h+y'y-(n_0\beta_0+X'y)'(n_0 I_{q+1}+X'X)^{-1}(n_0\beta_0+X'y)$$

Conjugate Prior For Normal RVs, UnKnown σ^2

similarities

$$h_*=(n_0+n)(\mu-\hat{\mu})^2+\omega$$
$$\hat{\mu}=\frac{n_0}{n_0+n}\mu_0+\frac{n}{n_0+n}\bar{x}$$
$$\omega=n_0\mu_0^2+h+n\overline{x^2}-\frac{(n_0\mu_0+n\bar{x})^2}{(n_0+n)}$$

Conjugate Estimation

And upon differentiating $f(\beta,\sigma^2|y,\cdot)$ with respect to β

$$f(\beta,\sigma^2|y,\cdot) \propto (\sigma^2)^{-(n+q+1+\nu)/2} e^{-\frac{1}{2\sigma^2}[(\beta-\hat{\beta})'(n_0 I_{q+1}+X'X)(\beta-\hat{\beta})+\omega]}$$

$$\omega = n_0\beta_0'\beta_0 + h + y'y - (n_0\beta_0 + X'y)'(n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'y)$$

we obtain a MAP estimator for β similar to MLEs,

$$\underset{\beta}{\text{ArgMax}}\ f(\beta,\sigma^2|y,\cdot) = (n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'X\hat{\beta}_{MLE}) \ .$$

note a weighted average

Conjugate Prior For Normal RVs, UnKnown σ^2

$$\underset{\mu}{\text{ArgMax}}\ f(\mu,\sigma^2|x_1,\ldots,x_n,\cdot) = \frac{n_0\mu_0 + n\bar{x}}{n_0+n}$$

similarities

$$\hat{\beta} = A\beta_0 + (I-A)\hat{\beta}_{MLE}$$

$$A = (X'X + n_0 I_{q+1})^{-1} n_0 I_{q+1}$$

$$I - A = (X'X + n_0 I_{q+1})^{-1} X'X$$

This is the mode of the PDF.

(Take the second derivative to confirm.)

Conjugate Estimation

$$\hat{\beta} = (n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'X\hat{\beta}_{MLE})$$

And upon differentiating $f(\beta,\sigma^2|y,\cdot)$ with respect to σ^2

$$f(\beta,\sigma^2 | y,\cdot) \propto (\sigma^2)^{-(n+q+1+\nu)/2} e^{-\frac{1}{2\sigma^2}[(\beta-\hat{\beta})'(n_0 I_{q+1}+X'X)(\beta-\hat{\beta})+\omega]}$$

$$\omega = n_0\beta_0'\beta_0 + h + y'y - (n_0\beta_0 + X'y)'(n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'y)$$

we obtain a MAP estimator for σ^2 similar to MLEs,

$$\underset{\sigma^2}{\text{ArgMax}}\; f(\sigma^2 | \beta = \hat{\beta}, y,\cdot) = \frac{\omega}{\nu + n + q + 1} \;.$$

This is the mode of the PDF.

(Take the second derivative to confirm.)

Conjugate Prior For Normal RVs, UnKnown σ^2

$$\underset{\sigma^2}{\text{ArgMax}}\; f(\sigma^2 | \mu = \hat{\mu}, x_1,\ldots,x_n,\cdot) = \frac{\omega}{\nu + n + 1}$$

$$\omega = n_0\mu_0^2 + h + n\overline{x^2} - \frac{(n_0\mu_0 + n\bar{x})^2}{(n_0 + n)}$$

$$\hat{\mu} = \frac{n_0\mu_0 + n\bar{x}}{n_0 + n}$$

Conjugate Estimation

$$\hat{\beta} = (n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'X\hat{\beta}_{MLE})$$

This posterior distribution

$$f(\beta,\sigma^2 | y,\cdot) \propto (\sigma^2)^{-(n+q+1+\nu)/2} e^{-\frac{h_*}{2\sigma^2}}$$

$$h_* = (\beta - \hat{\beta})'(n_0 I_{q+1} + X'X)(\beta - \hat{\beta}) + \omega$$

$$\omega = n_0\beta_0'\beta_0 + h + y'y - (n_0\beta_0 + X'y)'(n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'y)$$

can also be integrated with respect to σ^2 to obtain

$$f(\beta | y,\cdot) \propto \frac{\Gamma(\frac{\nu_*-2}{2})2^{(\nu_*-2)/2}}{h_*^{(\nu_*-2)/2}} \underbrace{\int_{\sigma^2} \frac{h_*^{(\nu_*-2)/2}}{\Gamma(\frac{\nu_*-2}{2})2^{(\nu_*-2)/2}}(\sigma^2)^{-\nu_*/2} e^{-\frac{h_*}{2\sigma^2}} d\sigma^2}_{=1}$$

$$\nu_* = \nu + n + q + 1$$

$$f(\beta | y,\cdot) \propto \frac{1}{[(\beta - \hat{\beta})'(n_0 I_{q+1} + X'X)(\beta - \hat{\beta}) + \omega]^{(\nu_*-2)/2}}$$

$$f(x|\alpha,\beta) = \boxed{\frac{\beta^\alpha}{\Gamma(\alpha)}} x^{-\alpha-1} \mathrm{e}^{-\beta/x}$$

Conjugate Estimation

$$\hat{\beta} = (n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'X\hat{\beta}_{MLE})$$
$$\nu_* = \nu + n + q + 1$$

$$f(\beta \mid y,\cdot) \propto \frac{1}{[\omega + (\beta-\hat{\beta})'(n_0 I_{q+1} + X'X)(\beta-\hat{\beta})]^{(\nu_*-2)/2}}$$

$$\omega = n_0\beta_0'\beta_0 + h + y'y - (n_0\beta_0 + X'y)'(n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'y)$$

which becomes with some algebra

$$f(\beta \mid y) = \underbrace{\boxed{\frac{\Gamma(\frac{\nu_\# + q+1}{2})\left|T\right|^{-1/2}}{(\nu_\#\pi)^{(q+1)2}\Gamma(\frac{\nu_\#}{2})}}}_{c} \frac{1}{[1 + \frac{1}{\nu_\#}(\beta-\hat{\beta})'T^{-1}(\beta-\hat{\beta})]^{(\nu_\#+q+1)/2}} \qquad \nu_\# = n + \nu - 2$$

where $T = \dfrac{\omega}{\nu_\#}(n_0 I_{q+1} + X'X)^{-1}$ *a posteriori correlated*

$$E(\beta \mid y,\cdot) = \hat{\beta} \qquad \mathrm{cov}(\beta \mid y,\cdot) = \frac{\nu_\#}{\nu_\# - 2} T$$

$$f(s \mid \nu,\delta,T) = \boxed{\frac{\Gamma(\frac{\nu+p}{2})\left|T\right|^{-1/2}}{(\nu\pi)^{p/2}\Gamma(\frac{\nu}{2})}} \frac{1}{[1 + \frac{1}{\nu}(s-\delta)'T^{-1}(s-\delta)]^{(\nu+p)/2}}$$

Conjugate Estimation

$$\hat{\beta} = (n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'X\hat{\beta}_{MLE})$$

$$h_* = (\beta - \hat{\beta})'(n_0 I_{q+1} + X'X)(\beta - \hat{\beta}) + \omega$$

$$\omega = n_0\beta_0'\beta_0 + h + y'y - (n_0\beta_0 + X'y)'(n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'y)$$

This posterior distribution is

$$f(\beta,\sigma^2 | y) \propto (\sigma^2)^{-(n+q+1+\nu)/2} e^{-\frac{h_\#}{2\sigma^2}} e^{-\frac{(\beta-\hat{\beta})'(n_0 I_{q+1}+X'X)(\beta-\hat{\beta})}{2\sigma^2}}$$

$$\hat{\beta} = (n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'X\hat{\beta}_{MLE})$$

$$h_\# = \omega$$

Integrating with respect to β leads to

$$f(\sigma^2 | y) \propto (\sigma^2)^{-(\nu+n)/2} e^{-\frac{h_\#}{2\sigma^2}} \underbrace{\int_\beta \left|\sigma^2 (n_0 I_{q+1} + X'X)^{-1}\right|^{-1/2} e^{-\frac{1}{2}(\beta-\hat{\beta})'\left[\frac{(n_0 I_{q+1}+X'X)}{\sigma^2}\right](\beta-\hat{\beta})} d\beta}_{=\text{constant}}$$

$$f(\sigma^2 | y) \propto (\sigma^2)^{-(\nu+n)/2} e^{-\frac{1}{2\sigma^2}[n_0\beta_0'\beta_0+h+y'y-(n_0\beta_0+X'y)'(n_0 I_{q+1}+X'X)(n_0\beta_0+X'y)]}$$

$$f(x|\mu,\Sigma) = (2\pi)^{-2/2} |\Sigma|^{-1/2} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}$$

Conjugate Estimation

The normalized posterior distribution is

C

$$f(\sigma^2|y) = \boxed{\frac{\omega^{\frac{n+\nu-2}{2}}}{\Gamma(\frac{n+\nu-2}{2})2^{\frac{n+\nu-2}{2}}}} (\sigma^2)^{-\frac{(n+\nu-2)+2}{2}} e^{-\frac{\omega}{2\sigma^2}}$$

$$\omega = n_0\beta_0'\beta_0 + h + y'y - (n_0\beta_0 + X'y)'(n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'y)$$

with expectation

$$E(\sigma^2|y) = \frac{\omega}{(n+\nu-3)}$$

and variance

$$\mathrm{var}(\sigma^2|y) = \frac{2\omega^2}{(n+\nu-3)^2(n+\nu-5)}$$

$$f(x|\alpha,\beta) = \boxed{\frac{\beta^\alpha}{\Gamma(\alpha)}} x^{-\alpha-1} \mathrm{e}^{-\beta/x}$$

Conjugate Estimation

$$f(y|\beta,\sigma^2)=(2\pi\sigma^2)^{-n/2}e^{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)}$$

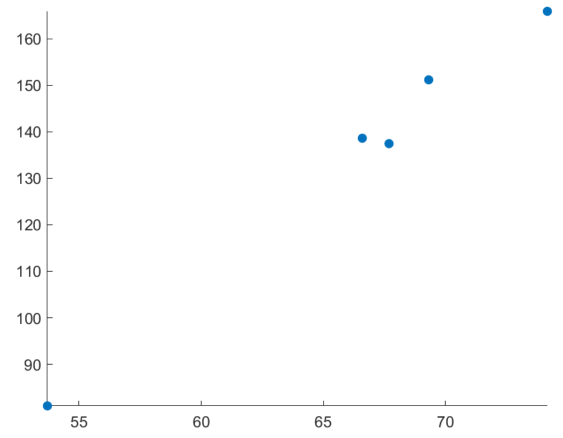
$$f(\beta|\beta_0,\sigma^2,n_0)=(2\pi\sigma^2/n_0)^{-(q+1)/2}e^{-\frac{n_0}{2\sigma^2}(\beta-\beta_0)'(\beta-\beta_0)}$$

$$f(\sigma^2|h,\nu)=\frac{h^{\frac{\nu-2}{2}}}{2^{\frac{\nu-2}{2}}\Gamma(\frac{\nu-2}{2})}(\sigma^2)^{-\nu/2}e^{-\frac{h}{2\sigma^2}}$$

Example: Heights and Weights

You are planning a study to predict weights from heights for Marquette undergraduate students. You expert assesses hyperparameters to be

$$n_0=25 \quad \beta_0=\begin{bmatrix}-150\\5\end{bmatrix} \quad \nu=n_0-1 \quad \sigma_0^2=5 \quad h=(n_0-1)\sigma_0^2=120$$



You have a random sample of n=5 undergraduates

67.69	137.44
74.17	165.90
53.71	81.05
69.31	151.17
66.59	138.61

$$\hat{\beta}=(n_0I_{q+1}+X'X)^{-1}(n_0\beta_0+X'X\hat{\beta}_{MLE})$$

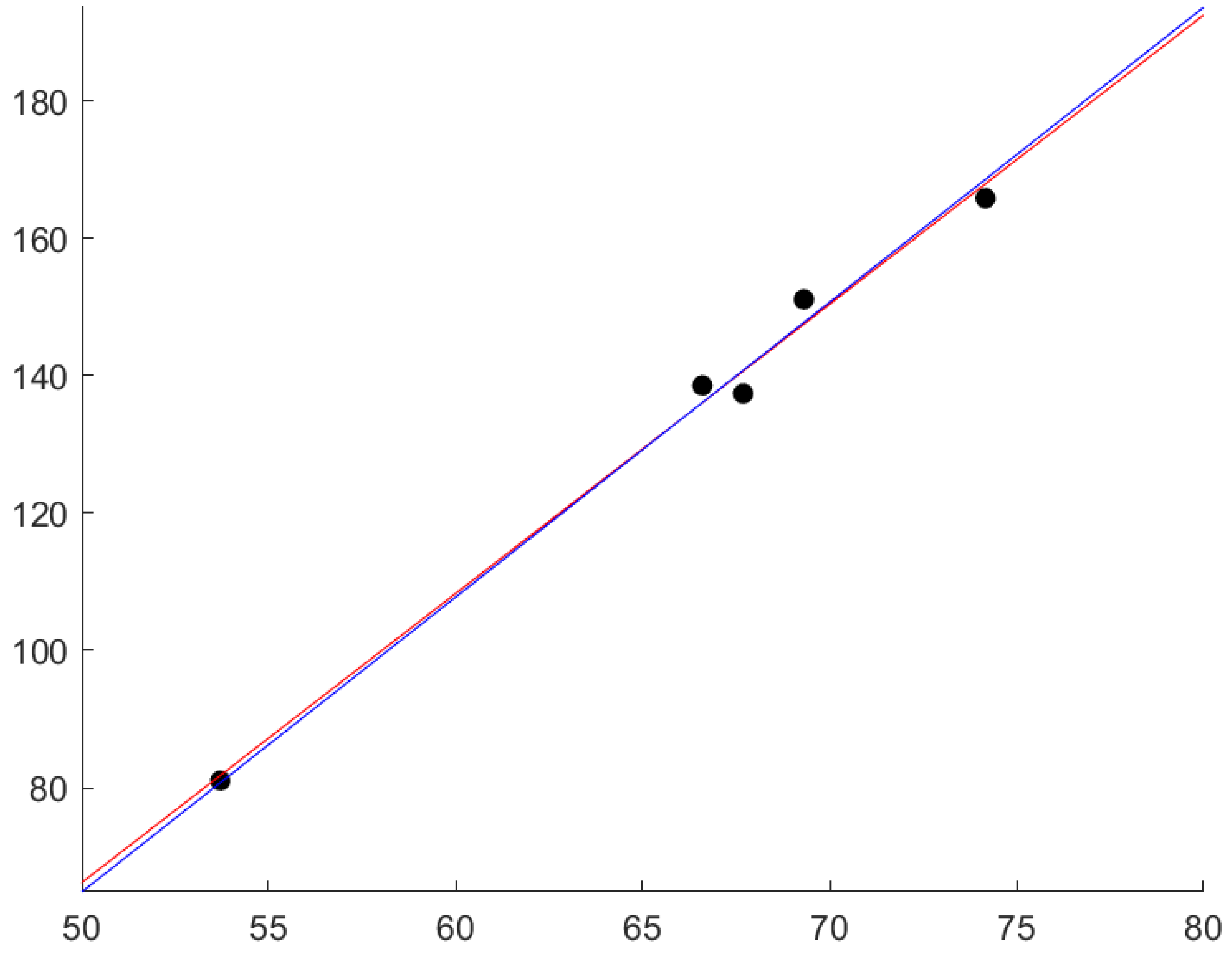
$$\hat{\beta}=\begin{bmatrix}0.9977 & -0.0149\\-0.0153 & 0.0013\end{bmatrix}\begin{bmatrix}-150\\5\end{bmatrix}+\begin{bmatrix}0.0023 & 0.0149\\0.0149 & 0.9987\end{bmatrix}\begin{bmatrix}-144.4407\\4.2127\end{bmatrix}$$

$$\hat{\beta}_{MLE}=\begin{bmatrix}-144.4407\\4.2127\end{bmatrix} \qquad \hat{\beta}=\begin{bmatrix}-149.9989\\4.2964\end{bmatrix}$$

Conjugate Estimation

$$\hat{\beta} = (n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'X\hat{\beta}_{MLE})$$

```
% DB Rowe Univariate
rng('default')
% Priors Hyperparameters
n0=25;
nu=n0-1;
b0=[-150;5];
sigma02=5;
h=nu*sigma02;
% Likelihood data
n=5; mu=[65;150]; sigma2=2;
ht=5*randn(n,1)+mu(1,1);
X=[ones(n,1),ht];
betaT=[-145;4.2]
y=X*betaT+sqrt(sigma2)*randn(n,1);
figure;
scatter(ht,y,'filled')
axis tight
% Bayesian Conjugate Estimation
bMLE=inv(X'*X)*X'*y;
A=inv(n0*eye(2)+X'*X)*(n0*eye(2))
ImA=inv(n0*eye(2)+X'*X)*(X'*X)
bhat=A*b0+(eye(2)-A)*bMLE
```



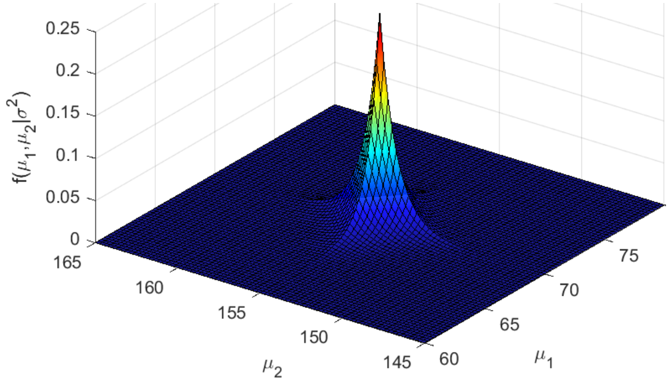
Non-Conjugate Estimation 1

With a normal distribution for the observations and likelihood,

$$f(y|\beta,\sigma^2)=(2\pi\sigma^2)^{-n/2}e^{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)}$$

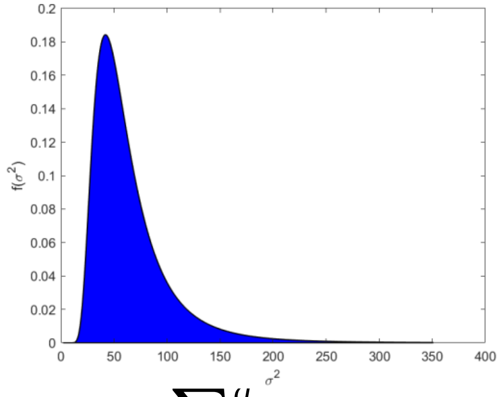
a non-conjugate prior distribution is Laplace-inverse Gamma

$$f(\beta|\sigma^2,\beta_0,n_j)=(4\sigma^2)^{-(q+1)}\left(\prod_{j=0}^{q}n_j\right)e^{-\frac{1}{2\sigma^2}\sum_{j=0}^{q}n_j|\beta_j-\beta_{0j}|}$$



a priori uncorrelated

$$f(\sigma^2|q,\nu)=\frac{h^{\frac{\nu-2}{2}}}{2^{\frac{\nu-2}{2}}\Gamma(\frac{\nu-2}{2})}(\sigma^2)^{-\nu/2}e^{-\frac{h}{2\sigma^2}}$$



and the posterior is

$$f(\beta,\sigma^2|y)=C(\sigma^2)^{-(n+2q+2+\nu)/2}e^{-\frac{1}{2\sigma^2}[(y-X\beta)'(y-X\beta)+\sum_{j=0}^{q}n_j|\beta_j-\beta_{0j}|+h]}$$

Non-Conjugate Estimation 1

The normalizing constant for the posterior

$$f(\beta,\sigma^2 \mid y) = C(\sigma^2)^{-(n+2q+2+\nu)/2} e^{-\frac{1}{2\sigma^2}[(y-X\beta)'(y-X\beta)+\sum_{j=0}^{q} n_j|\beta_j-\beta_{0j}|+h]}$$

is

$$C = \frac{(2\pi)^{-n/2} h^{\frac{\nu-2}{2}} \prod_{j=0}^{q} n_j}{4^{(q+1)} 2^{\frac{\nu-2}{2}} \Gamma(\frac{\nu-2}{2}) f(y)}$$

$$f(y) = \int f(y \mid \beta,\sigma^2) f(\beta_0 \mid \sigma^2) \ldots f(\beta_q \mid \sigma^2) f(\sigma^2) d\beta_1 \ldots d\beta_2 d\sigma^2$$

$$f(y) = \int \frac{(2\pi)^{-n/2} h^{\frac{\nu-2}{2}} \prod_{j=0}^{q} n_j}{4^{(q+1)} 2^{\frac{\nu-2}{2}} \Gamma(\frac{\nu-2}{2}) f(y)} (\sigma^2)^{-(n+2q+2+\nu)/2} e^{-\frac{1}{2\sigma^2}[(y-X\beta)'(y-X\beta)+\sum_{j=0}^{q} n_j|\beta_j-\beta_{0j}|+h]} d\beta_1 \ldots d\beta_2 d\sigma^2$$

Non-Conjugate Estimation 1

We can do a little algebra to turn

$$f(\beta,\sigma^2 \mid y) = C(\sigma^2)^{-(n+2q+2+\nu)/2} e^{-\frac{1}{2\sigma^2}[(y-X\beta)'(y-X\beta)+\sum_{j=0}^{q} n_j|\beta_j-\beta_{0j}|+h]}$$

into

$$C = \frac{(2\pi)^{-n/2} h^{\frac{\nu-2}{2}} \prod_{j=0}^{q} n_j}{4^{(q+1)} 2^{\frac{\nu-2}{2}} \Gamma(\frac{\nu-2}{2}) f(y)}$$

$$f(\beta,\sigma^2 \mid y) = C(\sigma^2)^{-(n+2q+2+\nu)/2} e^{-\frac{h_*}{2\sigma^2}}$$

$$h_* = (\beta-\hat{\beta}_{MLE})'(X'X)(\beta-\hat{\beta}_{MLE}) + \sum_{j=0}^{q} n_j \mid \beta_j - \beta_{0j} \mid + \underbrace{(y-X\hat{\beta}_{MLE})'(y-X\hat{\beta}_{MLE}) + h}_{\text{Does not depend on } \beta.}$$

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

Non-Conjugate Estimation 1

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

The posterior distribution can then be maximized wrt β

$$\underset{\beta}{\text{ArgMax}} \; f(\beta,\sigma^2 \mid y) = C(\sigma^2)^{-(n+2q+2+\nu)/2} e^{-\frac{h_*}{2\sigma^2}}$$

same as

$$\underset{\beta}{\text{ArgMin}} \; (\beta-\hat{\beta}_{MLE})'(X'X)(\beta-\hat{\beta}_{MLE}) + \sum_{j=0}^{q} n_j \mid \beta_j - \beta_{0j} \mid$$

to obtain $\hat{\beta}_{MAP}$. However no probabilities such as CIs or percentiles.

But one can computationally calculate expectations!

$$E(\beta_j \mid x_1,\ldots,x_n)$$

Non-Conjugate Estimation 1

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

The posterior can be maximized with respect to σ^2

$$\underset{\sigma^2}{\text{ArgMax}} \; f(\sigma^2 \mid \beta = \hat{\beta}_{MAP}, y, \cdot) = \frac{h_*}{n+2q+2+\nu}$$

$$h_* = (\beta_{MAP} - \hat{\beta}_{MLE})'(X'X)(\beta_{MAP} - \hat{\beta}_{MLE}) + \sum_{j=0}^{q} n_j \mid \beta_{MAP\,j} - \beta_{0\,j} \mid$$
$$+ (y - X\hat{\beta}_{MLE})'(y - X\hat{\beta}_{MLE}) + h$$

to obtain $\hat{\sigma}^2_{MAP}$. However no probabilities such as CIs or percentiles.

But one can computationally calculate expectations!

$$E(\beta_j \mid x_1,\ldots,x_n) \qquad E(\sigma^2 \mid x_1,\ldots,x_n)$$

Non-Conjugate Estimation 1

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

The $\hat{\beta}_{MAP}$ estimator from

$$\underset{\beta}{\text{ArgMin}} \; (\beta - \hat{\beta}_{MLE})'(X'X)(\beta - \hat{\beta}_{MLE}) + \sum_{j=0}^{q} n_j \, | \beta_j - \beta_{0j} |$$

is called LASSO regression when the β_{0j}=0!

When we assess β_{0j}=0, we are saying my prior mean is zero!

A prior mean of zero biases toward zero!

Non-Conjugate Estimation 1

$$f(y\,|\,\beta,\sigma^2)=(2\pi\sigma^2)^{-n/2}e^{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)}$$

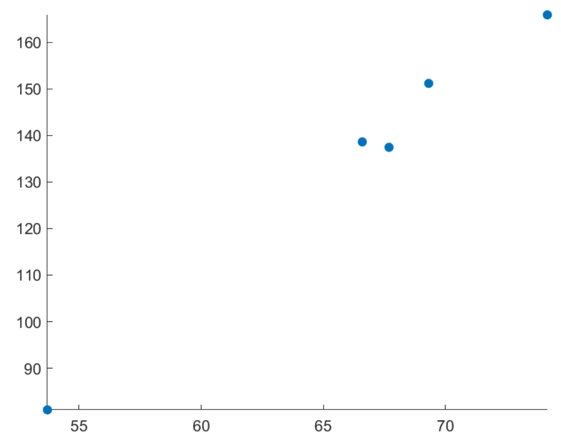
$$f(\beta\,|\,\sigma^2,\beta_0,n_j)=(4\sigma^2)^{-(q+1)}\left(\prod_{j=0}^{q}n_j\right)e^{-\frac{1}{2\sigma^2}\sum_{j=0}^{q}n_j|\beta_j-\beta_{0j}|}$$

$$f(\sigma^2\,|\,h,\nu)=\frac{h^{\frac{\nu-2}{2}}}{2^{\frac{\nu-2}{2}}\Gamma(\frac{\nu-2}{2})}(\sigma^2)^{-\nu/2}e^{-\frac{h}{2\sigma^2}}$$

Example: Heights and Weights

You are planning a study to predict weights from heights for Marquette undergraduate students. You expert assesses hyperparameters to be

$$n_0=25\quad \beta_0=\begin{bmatrix}-150\\5\end{bmatrix}\quad \nu=n_0-1\quad \sigma_0^2=5\quad h=(n_0-1)\sigma_0^2=120$$



You have a random sample of n=5 undergraduates

67.69	137.44
74.17	165.90
53.71	81.05
69.31	151.17
66.59	138.61

$$\underset{\beta}{\text{ArgMin}}\;(\beta-\hat{\beta}_{MLE})'(X'X)(\beta-\hat{\beta}_{MLE})+\sum_{j=0}^{q}n_j\,|\beta_j-\beta_{0j}|$$

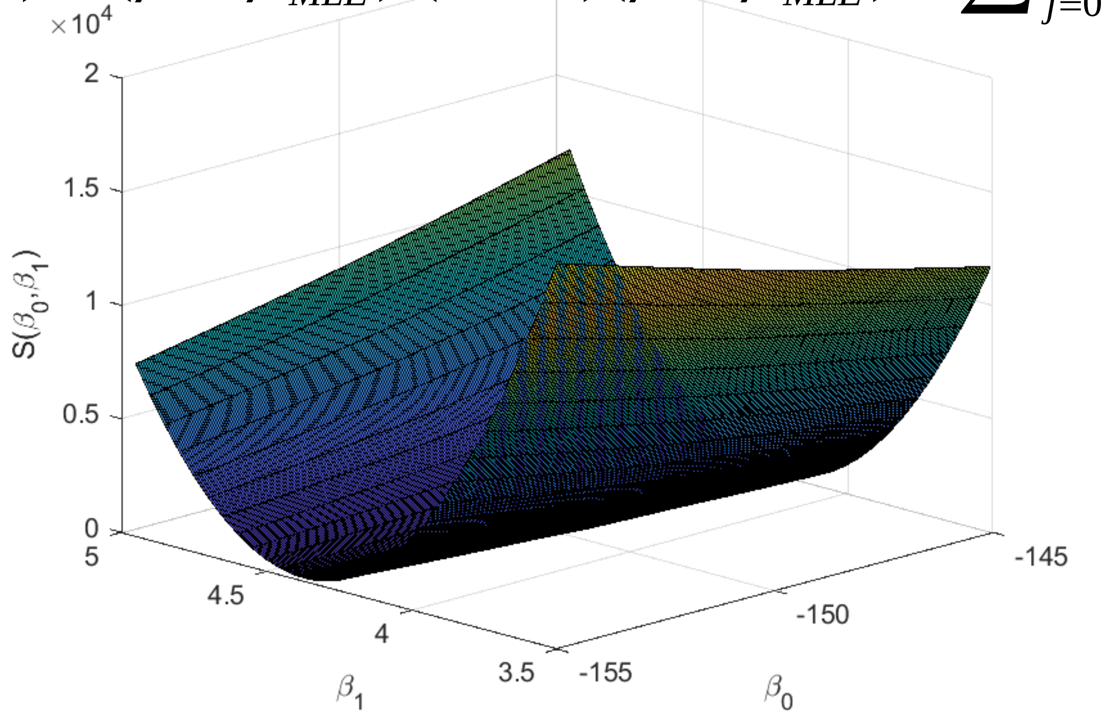
Non-Conjugate Prior   *Conjugate Prior*

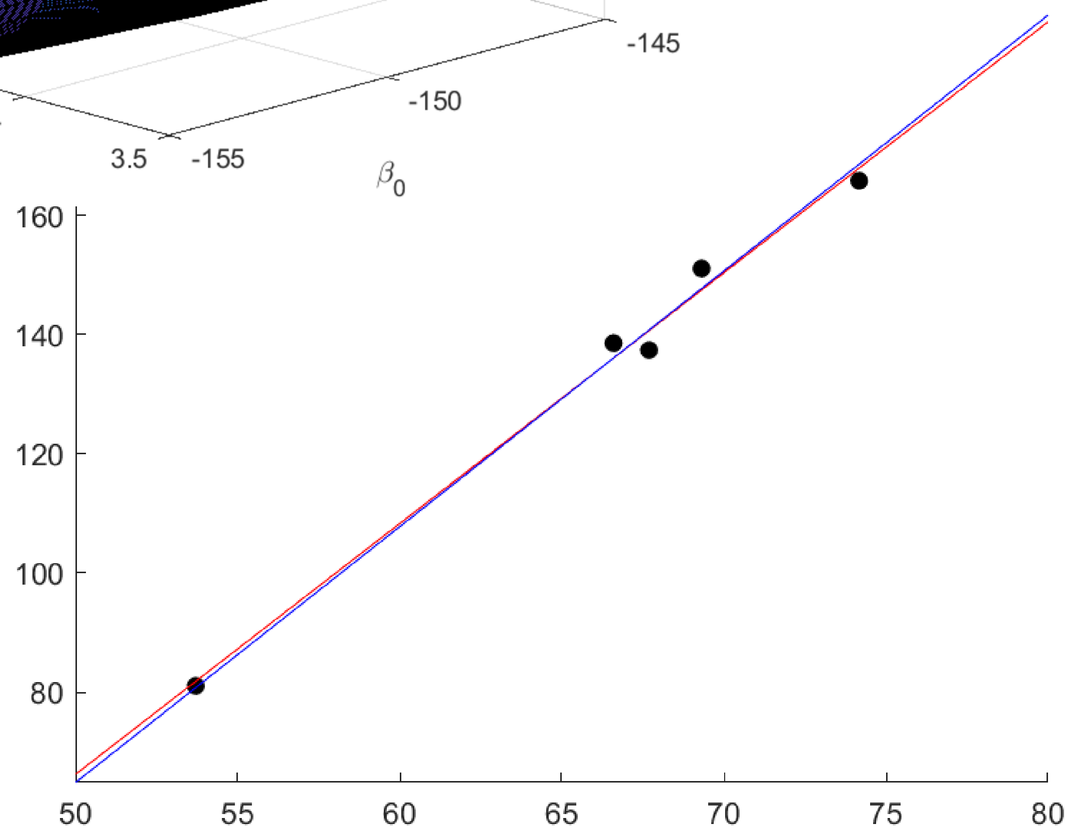
$$\hat{\beta}_{MLE}=\begin{bmatrix}-144.4407\\4.2127\end{bmatrix}\qquad \hat{\beta}=\begin{bmatrix}-150.0000\\4.2960\end{bmatrix}\qquad \hat{\beta}=\begin{bmatrix}-149.9989\\4.2964\end{bmatrix}$$

Non-Conjugate Estimation 1

$$S(\beta)=(\beta-\hat{\beta}_{MLE})'(X'X)(\beta-\hat{\beta}_{MLE})+\sum_{j=0}^{q} n_j\,|\beta_j-\beta_{0j}|$$

```
%Bayesian Non-Conjugate Estimation
delta=.0005; minS=10^6;
b0MAP=( -155:delta:-145-delta)';
b1MAP=(3.5:delta:5-delta)';
S=zeros(length(b0MAP),length(b1MAP));
for i=1:length(b0MAP)
    for j=1:length(b1MAP)
        bMAP=[b0MAP(i,1);b1MAP(j,1)];
        S(i,j)=(bMAP-bMLE)'*(X'*X)*(bMAP-bMLE)+...
            n0*abs(b0MAP(i,1)-b0(1,1))+...
            n0*abs(b1MAP(j,1)-b0(2,1));
        if (S(i,j)<minS)
            minS=S(i,j);
            b0formin=b0MAP(i,1);
            b1formin=b1MAP(j,1);
        end
    end
end
[minS,b0formin,b1formin]
figure;% repeat above with delta=.05
[X,Y]=meshgrid(b0MAP,b1MAP);
surf(X',Y',S,'FaceAlpha',.9,'FaceColor','interp')
az=-45; el=20; view(az,el)
xlabel('\beta_0'),ylabel('\beta_1'),zlabel('S(\beta_0,\beta_1)')
```





Non-Conjugate Estimation 1

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

This posterior distribution

$$f(\beta,\sigma^2 | y) = C(\sigma^2)^{-\nu_\#/2} e^{-\frac{h_\#}{2\sigma^2}}$$

$$\nu_\# = n + 2q + 2 + \nu$$

$$h_\# = (\beta - \hat{\beta}_{MLE})'(X'X)(\beta - \hat{\beta}_{MLE}) + \sum_{j=0}^{q} n_j |\beta_j - \beta_{0j}|$$

$$+ (y - X\hat{\beta}_{MLE})'(y - X\hat{\beta}_{MLE}) + h$$

$$\hat{\beta} = (n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'X\hat{\beta}_{MLE})$$

can also be integrated with respect to σ^2 to obtain

$$f(\beta | y) = C\frac{\Gamma(\frac{\nu_\#-2}{2})2^{(\nu_\#-2)/2}}{h_\#^{(\nu_\#-2)/2}} \underbrace{\int_{\sigma^2} \frac{h_\#^{(\nu_\#-2)/2}}{\Gamma(\frac{\nu_\#-2}{2})2^{(\nu_\#-2)/2}} (\sigma^2)^{-\nu_\#/2} e^{-\frac{h_\#}{2\sigma^2}} d\sigma^2}_{=1}$$

$$f(\beta | y) = C_\# \frac{1}{[\omega + (\beta - \hat{\beta}_{MLE})'(X'X)(\beta - \hat{\beta}_{MLE}) + \sum_{j=0}^{q} n_j |\beta_j - \beta_{0j}|]^{(\nu_\#-2)/2}}$$

Non-Conjugate Estimation 1

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

This marginal posterior distribution for $\beta|y$

$$f(\beta|y) = C_{\#}\frac{1}{[\omega + (\beta-\hat{\beta}_{MLE})'(X'X)(\beta-\hat{\beta}_{MLE}) + \sum_{j=0}^{q} n_j\,|\beta_j - \beta_{0j}|]^{(\nu_{\#}-2)/2}}$$

is not a Student-t and does not have a closed form expression for the expectation (mean).

$$C_{\#} = \frac{\Gamma(\frac{\nu_{\#}-2}{2})2^{(\nu_{\#}-2)/2}h^{\frac{\nu-2}{2}}\prod_{j=0}^{q} n_j}{(2\pi)^{n/2}4^{(q+1)}2^{\frac{\nu-2}{2}}\Gamma(\frac{\nu-2}{2})f(y)}$$

There is also no closed form for the marginal posterior distribution for $\sigma^2|y$.

Deterministic numerical/stochastic MCMC integration.

Non-Conjugate Estimation 2

With a normal distribution for the observations and likelihood,

$$f(y|\beta,\sigma^2)=(2\pi\sigma^2)^{-n/2}e^{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)}$$

another non-conjugate prior is Normal/Laplace-IG

$$f(\beta|\sigma^2,\beta_0,n_j)=C(2\pi\sigma^2/n_*)^{-(q+1)/2}e^{-\frac{n_*}{2\sigma^2}(\beta-\beta_0)'(\beta-\beta_0)}(4\sigma^2)^{-(q+1)}\left(\prod_{j=0}^{q}n_j\right)e^{-\frac{1}{2\sigma^2}\sum_{j=0}^{q}n_j|\beta_j-\beta_{0j}|}$$

a priori uncorrelated

$$f(\sigma^2|h,\nu)=\frac{h^{\frac{\nu-2}{2}}}{2^{\frac{\nu-2}{2}}\Gamma(\frac{\nu-2}{2})}(\sigma^2)^{-\nu/2}e^{-\frac{h}{2\sigma^2}}$$

and the posterior is

$$f(\beta,\sigma^2|y)\propto(\sigma^2)^{-(n+3q+3+\nu)/2}e^{-\frac{1}{2\sigma^2}[(y-X\beta)'(y-X\beta)+n_*(\beta-\beta_0)'(\beta-\beta_0)+\sum_{j=0}^{q}n_j|\beta_j-\beta_{0j}|+h]}$$

Non-Conjugate Estimation 2

We can do a little algebra to turn

$$f(\beta,\sigma^2 \mid y) \propto (\sigma^2)^{-(n+3q+3+\nu)/2} e^{-\frac{1}{2\sigma^2}[(y-X\beta)'(y-X\beta)+n_*(\beta-\beta_0)'(\beta-\beta_0)+\sum_{j=0}^{q} n_j|\beta_j-\beta_{0j}|+h]}$$

into

$$f(\beta,\sigma^2 \mid y) \propto (\sigma^2)^{-(n+3q+3+\nu)/2} e^{-\frac{h_*}{2\sigma^2}}$$

$$h_* = (\beta-\hat{\beta}_{MLE})'(X'X)(\beta-\hat{\beta}_{MLE}) + n_*(\beta-\beta_0)'(\beta-\beta_0)$$
$$+\sum_{j=0}^{q} n_j \mid \beta_j - \beta_{0j} \mid + (y-X\hat{\beta}_{MLE})'(y-X\hat{\beta}_{MLE}) + h$$

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

Non-Conjugate Estimation 2

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

The posterior distribution can then be maximized wrt β

$$\underset{\beta}{\text{ArgMax}}\ (\sigma^2)^{-(n+3q+3+\nu)/2} e^{-\frac{h_*}{2\sigma^2}}$$

same as

$$\underset{\beta}{\text{ArgMin}}\ (\beta-\hat{\beta}_{MLE})'(X'X)(\beta-\hat{\beta}_{MLE}) + n_*(\beta-\beta_0)'(\beta-\beta_0) + \sum_{j=0}^{q} n_j\,|\beta_j-\beta_{0j}|$$

to obtain $\hat{\beta}_{MAP}$.

Non-Conjugate Estimation 2

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

The $\hat{\beta}_{MAP}$ estimator from

$$\underset{\beta}{\text{ArgMin}}\, (\beta - \hat{\beta}_{MLE})'(X'X)(\beta - \hat{\beta}_{MLE}) + n_*(\beta - \beta_0)'(\beta - \beta_0) + \sum_{j=0}^{q} n_j \,|\, \beta_j - \beta_{0j} \,|$$

is called Elastic Net regression when the β_{0j}=0!

When we assess β_{0j}=0, we are saying my prior mean is zero.

A prior mean of zero biases toward zero!

Regularization

A general framework for the prior on β is:

$$f(\beta|\sigma^2,\beta_0,n_*,n_j)=C(2\pi\sigma^2/n_*)^{-(q+1)/2}e^{-\frac{n_*}{2\sigma^2}(\beta-\beta_0)'(\beta-\beta_0)}(4\sigma^2)^{-(q+1)}\left(\prod_{j=0}^{q}n_j\right)e^{-\frac{1}{2\sigma^2}\sum_{j=0}^{q}n_j|\beta_j-\beta_{0j}|}$$

$$S(\beta)=\underbrace{(\beta-\hat{\beta}_{MLE})'(X'X)(\beta-\hat{\beta}_{MLE})}_{\text{Likelihood}}+\underbrace{n_*(\beta-\beta_0)'(\beta-\beta_0)}_{\text{Normal}}+\underbrace{\sum_{j=0}^{q}n_j|\beta_j-\beta_{0j}|}_{\text{Laplace}}$$

Model	n_*	n_j	β_0
Ridge (Normal)	n_*	0	0
LASSO (Laplace)	0	n_j	0
Elastic Net (Normal/Laplace)	n_*	n_j	0

Homework 11

1. Show that

$$(y - X\beta)'(y - X\beta) + n_0(\beta - \beta_0)'(\beta - \beta_0) + h$$
$$= (\beta - \hat{\beta})'(n_0 I_{q+1} + X'X)(\beta - \hat{\beta}) + n_0\beta_0'\beta_0 + h$$
$$+ y'y - (n_0\beta_0 + X'y)'(n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'y)$$

2*. Integrate $f(\beta,\sigma^2|y)$ with conjugate priors to obtain the marginal posterior distributions $f(\beta|y)$ and $f(\sigma^2|y)$.

Hint: This is a Normal PDF integral.

Hint: This is an Inverse Gamma PDF integral.

* For students in MSSC 5790.

Homework 11

$$T = \frac{\omega}{\nu_{\#}}(n_0 I_{q+1} + X'X)^{-1}$$

3. You plan to see data points $(x_1,y_1),\ldots,(x_n,y_n)$ of heights (x's) and weights (y's) and aim to use your available info in a Bayesian fashion to estimate regression coefficients β in the model $y=X\beta+\varepsilon$, $\varepsilon \sim N(0,\sigma^2 I_n)$.

   Generate one $(x_1,y_1),\ldots,(x_n,y_n)$ dataset by selecting your own true values. Using conjugate priors, assess your own hyperparameters, β_0, n_0, h, ν, and estimate your regression coefficients and their covariance.

$$\hat{\beta} = (n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'X\hat{\beta}_{MLE})$$

$$\mathrm{cov}(\beta \mid y,\cdot) = \frac{\nu_{\#}}{\nu_{\#}-2}T$$

Homework 11

$$\omega = n_0\beta_0'\beta_0 + h + y'y - (n_0\beta_0 + X'y)'(n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'y)$$

4*. You plan to see data points $(x_1,y_1),\ldots,(x_n,y_n)$ of heights (x's) and weights (y's) and aim to use your available info in a Bayesian fashion to estimate regression coefficients β in the model $y=X\beta+\varepsilon$, $\varepsilon \sim N(0,\sigma^2 I_n)$.

Generate one $(x_1,y_1),\ldots,(x_n,y_n)$ dataset by selecting your own true values. Using conjugate priors, assess your own hyperparameters, β_0, n_0, h, v, and estimate your regression residual variance and its variance.

$$E(\sigma^2 \mid y) = \frac{\omega}{(n+q+\nu-3)} \qquad \operatorname{var}(\sigma^2 \mid y) = \frac{2\omega^2}{(n+q+\nu-3)^2(n+q+\nu-5)}$$

* For students in MSSC 5790.

Homework 11

$$T = \frac{\omega}{\nu_{\#}}(n_0 I_{q+1} + X'X)^{-1}$$

5**.Repeat problem 3 a very large number of times L.

Using the same conjugate priors and assessed hyperparameters estimate your regression coefficients for each data set.

$$\hat{\beta} = (X'X + n_0 I_{q+1})^{-1}(X'y + n_0\beta_0) \qquad \text{cov}(\beta \mid y, \cdot) = \frac{\nu_{\#}}{\nu_{\#} - 2} T$$

$$\omega = n_0 \beta_0' \beta_0 + h + y'y - (n_0\beta_0 + X'y)'(n_0 I_{q+1} + X'X)^{-1}(n_0\beta_0 + X'y)$$

Make histograms, calculate means, variances and covariances. Compare to what the true values should be.

** For students that have had 6010 and 6020.

Homework 11

6***.Using the same data from problem 3., estimate regression coefficients using Laplace prior and Normal/Laplace prior.

$$\underset{\beta}{\text{ArgMin}} (\beta-\hat{\beta}_{MLE})'(X'X)(\beta-\hat{\beta}_{MLE})+\sum_{j=0}^{q} n_j |\beta_j-\beta_{0j}|$$

$$\underset{\beta}{\text{ArgMin}} (\beta-\hat{\beta}_{MLE})'(X'X)(\beta-\hat{\beta}_{MLE})+n_*(\beta-\beta_0)'(\beta-\beta_0)+\sum_{j=0}^{q} n_j |\beta_j-\beta_{0j}|$$

Compare your answers to those in problem 3.

*** For enthusiastic students.

Homework 11

7***.Using the same data from problem 4., estimate regression coefficients using Laplace prior and Normal/Laplace prior.

$$\underset{\beta}{\text{ArgMin}} (\beta - \hat{\beta}_{MLE})'(X'X)(\beta - \hat{\beta}_{MLE}) + \sum_{j=0}^{q} n_j \,|\, \beta_j - \beta_{0j} \,|$$

$$\underset{\beta}{\text{ArgMin}} (\beta - \hat{\beta}_{MLE})'(X'X)(\beta - \hat{\beta}_{MLE}) + n_*(\beta - \beta_0)'(\beta - \beta_0) + \sum_{j=0}^{q} n_j \,|\, \beta_j - \beta_{0j} \,|$$

Make histograms, calculate means, variances and covariances. Compare to what the true values should be.

*** For enthusiastic students.