2. PRIOR INFORMATION

May be data-based or non-data-based.

Construction of Prior:

2 Analysts – prior for θ.

$$p_A(\theta) = \frac{1}{20\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\theta-900}{20}\right)^2\right]$$

$$p_B(\theta) = \frac{1}{80\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\theta-800}{80}\right)^2\right]$$

A. $Pr.[860 \le \theta \le 940] = Pr.[-2 \le z \le +2] \approx 0.95$

Only if offered odds of at least 20:1 would (s)he bet that θ differs from 900 by more than 40.

B. $Pr.[700 \le \theta \le 900] = Pr.[-1.25 \le z \le +1.25] \approx 0.8$

Only if offered odds of at least 5:1 would (s)he bet that θ differs from 800 by more than 100.

Check the odds – $x : 1$

Bet \$1 – lose it if wrong
– collect \$$x$ (including stak if right

Prior expected net payoff =

$$\$[(x-1)\,Pr(\text{right}) - (1)\,Pr(\text{wrong})]$$

Least acceptable payoff = \$0.

For "A":

$$0 = \left[(x-1)\left(\frac{1}{20}\right) - (1)\left(\frac{19}{20}\right)\right]$$

$$\Rightarrow x = 20$$

(Need odds of at least 20:1.)

Representing Prior Ignorance:

Harold Jeffreys' Rules:

(a) $-\infty < \theta < \infty$

$p(\theta) \propto \text{constant}$

i.e. $p(\theta)d\theta \propto d\theta$

Uniform & "improper" – Diffuse

$$\int_{-\infty}^{\infty} p(\theta)\,d\theta \propto \int_{-\infty}^{\infty} d\theta = \infty \quad (\text{not "1"})$$

(b) $0 < \theta < \infty$

$p(\theta) \propto 1/\theta$

or, $p(\theta)d\theta \propto d\theta/\theta$.

i.e. Take $\phi = \log(\theta)$ to be uniform:

$d\phi = d\theta/\theta \quad ; \quad -\infty < \phi < \infty$

So: $p(\phi) \propto \text{constant}$

$\Rightarrow p(\theta)\,|J_+| \propto \text{const.}$

$\Rightarrow p(\theta).\,\theta \propto \text{const} \Rightarrow p(\theta) \propto 1/\theta.$

In what sense do these improper priors represent "prior ignorance"?

(a) $-\infty < \theta < \infty$

$Pr.[a < \theta < b] / Pr.[c < \theta < d] = ?$

<u>Indeterminate.</u>

(b) $0 < \theta < \infty$

$$\int_0^a p(\theta)d\theta \propto \int_0^a (d\theta/\theta) \propto [\log|\theta|]_0^a = \infty$$

$$\int_a^\infty p(\theta)d\theta \propto \int_a^\infty (d\theta/\theta) \propto [\log|\theta|]_a^\infty = \infty$$

So: $Pr.[0 < \theta < a] / Pr.[a < \theta < \infty] =$

<u>Indeterminate.</u>

<u>Note</u>:

(i) Invariant to power transformations – $\phi = \theta^n$.

$(d\phi/d\theta) = n\theta^{n-1}$

$(d\phi/\phi) = n(d\theta/\theta) \propto (d\theta/\theta)$

So, in regression, work with σ or σ^2.

(ii) $$\int_0^\infty p(\theta)\,d\theta \propto \int_0^\infty (d\theta/\theta) \propto [\log|\theta|]_0^\infty$$
$$= \infty \qquad \text{(certainty)}$$

(Posterior will generally still be "proper".)

<u>LARGE-SAMPLE RESULTS</u>:

$$p(\theta|y) \propto p(\theta).L(\theta|y)$$
$$\propto p(\theta)\exp\{\log L(\theta|y)\}$$

Let $\tilde{\theta}$ = MLE of θ.

$$p(\theta) = p(\tilde{\theta}) + (\theta-\tilde{\theta})p'(\tilde{\theta}) + \cdots$$
$$= p(\tilde{\theta})\left[1 + (\theta-\tilde{\theta})\left(\frac{p'(\tilde{\theta})}{p(\tilde{\theta})}\right) + \tfrac{1}{2}(\theta-\tilde{\theta})^2\left(\frac{p''(\tilde{\theta})}{p(\tilde{\theta})}\right) + \cdots\right]$$

Let $g(\theta) = \log L(\theta|y)$.

So, $$\exp[g(\theta)] = \exp\left[g(\tilde{\theta}) + (\theta-\tilde{\theta})g'(\tilde{\theta}) + \tfrac{1}{2}(\theta-\tilde{\theta})^2 g''(\tilde{\theta}) + \cdots\right]$$

But $g'(\tilde{\theta}) = 0$ (max.)

So, $$\exp[g(\theta)] = \exp[g(\tilde{\theta})].\exp[\tfrac{1}{2}(\theta-\tilde{\theta})^2 g''(\tilde{\theta})]$$
$$.\exp[\tfrac{1}{6}(\theta-\tilde{\theta})^3 g'''(\tilde{\theta})]\cdots$$
$$\propto \exp[\tfrac{1}{2}(\theta-\tilde{\theta})^2 g''(\tilde{\theta})]$$
$$.\exp[\tfrac{1}{6}(\theta-\tilde{\theta})^3 g'''(\tilde{\theta})]\cdots$$
$$\propto \exp[\tfrac{1}{2}(\theta-\tilde{\theta})^2 g''(\tilde{\theta})]$$

So, $$p(\theta|y) \propto e^{\frac{1}{2}(\theta-\tilde{\theta})^2 g''(\tilde{\theta})}\left\{1 + (\theta-\tilde{\theta})\frac{p'}{p} + \tfrac{1}{2}(\theta-\tilde{\theta})^2\frac{p''(\tilde{\theta})}{p(\tilde{\theta})} + \cdots\right\}$$

because $p(\tilde{\theta})$ = const.

So, $$p(\theta|y) \propto e^{\frac{1}{2}(\theta-\tilde{\theta})^2 g''(\tilde{\theta})}$$
$$= N[\tilde{\theta}, -g''(\tilde{\theta})^{-1}]$$

So, the posterior is Normal, centred at MLE, & I^{-1} as variance.

(This approximation improves as $n \to \infty$

3. BAYESIAN ANALYSIS OF THE MULTIPLE REGRESSION MODEL

$$y = X\beta + u \quad ; \quad u \sim N(0, \sigma^2 I)$$

Rank $(X) = k$; X non-stochastic.

$$L(\beta, \sigma | y) = p(y | \beta, \sigma)$$

$$\propto \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right\}$$

Now, $(y - X\beta)'(y - X\beta)$

$$= [(y - X\hat\beta) + (X\hat\beta - X\beta)]'[(y - X\hat\beta) + (X\hat\beta - X\beta)]$$

where : $\hat\beta = (X'X)^{-1}X'y$.

Now, $(y - X\hat\beta)'(X\hat\beta - X\beta) = 0$

So, $(y - X\beta)'(y - X\beta)$

$$= (y - X\hat\beta)'(y - X\hat\beta) + (\beta - \hat\beta)'X'X(\beta - \hat\beta)$$

$$= vs^2 + (\beta - \hat\beta)'X'X(\beta - \hat\beta) \quad ; \quad \begin{cases} v = (n-k) \\ s^2 = \frac{\hat u'\hat u}{v} \end{cases}$$

$$\therefore L(\beta, \sigma | y) \propto \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2}\left\{vs^2 + (\beta - \hat\beta)'X'X(\beta - \hat\beta)\right\}\right]$$

(i) Diffuse Prior Information:

$$p(\beta, \sigma) = p(\beta) \cdot p(\sigma)$$

$$p(\beta_i) \propto \text{const.}; \quad \begin{cases} i = 1, 2, \ldots, k \\ -\infty < \beta_i < \infty \end{cases}$$

$$p(\sigma) \propto \sigma^{-1} \quad ; \quad 0 < \sigma < \infty$$

So, $p(\beta, \sigma) \propto \sigma^{-1}$.

By Bayes' Theorem:

$$p(\beta, \sigma | y) \propto p(\beta, \sigma) . L(\beta, \sigma | y)$$

$$\propto \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[vs^2 + (\beta - \hat\beta)'X'X(\beta - \hat\beta)\right]\right\}$$

Now, suppose σ^2 is known:

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

$$\propto \exp\left\{-\frac{1}{2\sigma^2} . vs^2\right\} . \exp\left\{-\frac{1}{2\sigma^2}(\beta - \hat\beta)'X'X(\beta - \hat\beta)\right\}$$

$p(\beta|\sigma,y) \propto \exp\left[-\frac{1}{2}(\beta-\hat\beta)'(\sigma^2(X'X)^{-1})^{-1}(\beta-\hat\beta)\right\}$

& so $p(\beta|\sigma,y) = MVN[\hat\beta, \sigma^2(X'X)^{-1}]$

& the MEL estimator of β is $\hat\beta$, under various loss fctns.

(Bayes $\Longleftrightarrow$ MLE)

Of course, generally σ^2 is unknown.

Then, $p(\beta|y) = \int_0^\infty p(\beta,\sigma|y)\,d\sigma$

$$\propto \int_0^\infty \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[vs^2 + (\beta-\hat\beta)'X'X(\beta-\hat\beta)\right]\right\} . d\sigma$$

Let : $c = [vs^2 + (\beta-\hat\beta)'X'X(\beta-\hat\beta)]$

$z = c/\sigma^2$

So : $dz = -(2c/\sigma^3)\,d\sigma$

or, $d\sigma = -(\sigma^3/2c)\,dz$

Also, $\sigma = (c/z)^{1/2}$

So, $\sigma^{-(n+1)} = (c/z)^{-(n+1)/2}$

Then :

$$p(\beta|y) \propto \int_\infty^0 (c/z)^{-(n+1)/2} e^{-z/2} \left(\frac{-\sigma^3}{2c}\right) dz$$

$$\propto \int_0^\infty (c/z)^{-(n+1)/2} e^{-z/2} \left(\tfrac{1}{2}\right)\left(\tfrac{1}{z}\right)^{3/2} c^{3/2}$$

$$\propto c^{-n/2} \int_0^\infty z^{n/2-1} e^{-z/2}\,dz$$

But $\int_0^\infty k' z^{n/2-1} e^{-z/2}\,dz = 1$

So,

$$p(\beta|y) \propto c^{-n/2}\left(\frac{1}{k'}\right) \propto c^{-n/2}$$

$$\propto [vs^2 + (\beta-\hat\beta)'X'X(\beta-\hat\beta)]^{-n/2}$$

$$= MVST\left(\hat\beta,\ s^2(X'X)^{-1}\left(\frac{v}{v-2}\right)\right)$$

Now, the marginals of a MVST are also Student-t. So:

$$p(\beta_i|y) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\beta|y)\,d\beta_1 \cdots d\beta_{i-1}\,d\beta_{i+1} \cdots d$$

$= $ Student-t (mean $= \hat\beta_i$)

Similarly:

$$p(\sigma|y) = \int_{-\infty}^{\infty} p(\beta,\sigma|y)\, d\beta$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\beta_1, \ldots, \beta_k, \sigma|y)\, d\beta_1 \cdots d\beta_k$$

& it turns out that

$$p(\sigma|y) \propto \left(\frac{1}{\sigma^{\nu+1}}\right) \exp\left[-\frac{\nu s^2}{2\sigma^2}\right].$$

This is the density for an Inverted Gamma distribution.

Features:

(i) Single mode at $\sigma_m = s\left(\frac{\nu}{\nu+1}\right)^{1/2}$

(ii) Mean $= \frac{\Gamma[(\nu-1)/2]}{\Gamma(\nu/2)} \cdot \nu^{1/2} s$; $\nu > 1$

(iii) variance $= \left(\frac{\nu s^2}{\nu-2}\right) - \text{mean}^2$; $\nu > 2$

(iv) skew $= \left[\frac{[\Gamma(\nu-2)/2]}{\Gamma(\nu/2)} \left(\frac{\nu}{2}\right)^{1/2} - \left(\frac{\nu}{\nu+1}\right)^{1/2}\right] \cdot \frac{s}{\sqrt{\text{var}}}$; $\nu > 2$

(Generally +ve skewness)

So, with diffuse prior, Bayes estimator of

(ii) Natural-Conjugate Prior:

For a Normal L.F. the NCP for (β, σ) is

$$p(\beta,\sigma) = p(\beta|\sigma)\, p(\sigma)$$

where: $\begin{cases} p(\beta|\sigma) \text{ is conditionally MVN} \\ p(\sigma) \text{ is inverted gamma.} \end{cases}$

So, $p(\beta|\sigma) \propto |A|^{1/2} \sigma^{-k} \exp\left\{-\frac{1}{2\sigma^2}(\beta-\bar{\beta})' A (\beta-\bar{\beta})\right\}$

$= MVN(\bar{\beta} ; \sigma^2 A^{-1})$; A pd

$p(\sigma) \propto \sigma^{-(w+1)} \exp\left[-\frac{wc^2}{2\sigma^2}\right]$; $c, w >$

Choose $\bar{\beta}$ & A to reflect prior beliefs abo[ut] β ; choose c and w in the case of σ.

The likelihood function is

$$L(\beta,\sigma|y) \propto \frac{1}{\sigma^n} \exp\left[-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right.$$

Applying Bayes' Theorem:

$$p(\beta,\sigma|y) \propto p(\beta,\sigma) \cdot L(\beta,\sigma|y)$$

$$\propto \left(\frac{1}{\sigma^{n+k+w+1}}\right) \exp\left[-\frac{1}{2\sigma^2}\left\{wc^2 + (\beta-\bar{\beta})' A(\beta-\bar{\beta}) + (y-X\beta)'(y-X\beta)\right\}\right]$$

Now, let

$$\begin{aligned} \gamma &= (\beta-\bar{\beta})'A(\beta-\bar{\beta}) + (y-X\beta)'(y-X\beta) \\ &= [\beta'A\beta - \beta'A\bar{\beta} - \bar{\beta}'A\beta + \bar{\beta}'A\bar{\beta} + y'y - y'X\beta \\ &\quad - \beta'X'y + \beta'X'X\beta] \\ &= \beta'(A+X'X)\beta - 2\beta'(A\bar{\beta}+X'y) + (y'y + \bar{\beta}'A\bar{\beta}) \end{aligned}$$

Complete the square:

$$q = ax^2 + bx + c$$

$$\Rightarrow q = a(x + b/2a)^2 + (c - b^2/4a)$$

So, $q = x'A^*x + x'b + c$

$$\Rightarrow q = (x + \tfrac{1}{2}A^{*-1}b)'A^*(x + \tfrac{1}{2}A^{*-1}b) + (c - \tfrac{1}{4}b'A^{*-1}b)$$

So,

$$\begin{aligned} \gamma = {} & [\beta - (A+X'X)^{-1}(A\bar{\beta}+X'y)]'(A+X'X) \\ & [\beta - (A+X'X)^{-1}(A\bar{\beta}+X'y)] + y'y + \bar{\beta}'A\bar{\beta} \\ & - (A\bar{\beta}+X'y)'(A+X'X)^{-1}(A\bar{\beta}+X'y) \end{aligned}$$

Let $\tilde{\beta} = (A+X'X)^{-1}(A\bar{\beta}+X'y)$.

Then, $\gamma = y'y + \bar{\beta}'A\bar{\beta} + (\beta-\tilde{\beta})'(A+X'X)(\beta-\tilde{\beta}) - \tilde{\beta}'(A+X'X)\tilde{\beta}$.

So:

$$\begin{aligned} p(\beta,\sigma|y) &\propto \left(\frac{1}{\sigma^{n+k+w+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[wc^2 + y'y + \bar{\beta}'A\bar{\beta} + (\beta-\tilde{\beta})'(A+X'X)(\beta-\tilde{\beta}) - \tilde{\beta}'(A+X'X)\tilde{\beta}\right]\right\} \\ &\propto \left(\frac{1}{\sigma^{n'+k+1}}\right) \exp\left[-\frac{1}{2\sigma^2}\left\{n'c_1^2 + (\beta-\tilde{\beta})'(A+X'X)(\beta-\tilde{\beta})\right\}\right] \end{aligned}$$

where: $n' = n+w$; $n'c_1^2 = (wc^2 + y'y + \bar{\beta}'A\bar{\beta} - \tilde{\beta}'(A+X'X)\tilde{\beta})$

Now, if we condition on σ, note that:

$$p(\beta \mid \sigma, y) \propto \exp\left[-\frac{1}{2\sigma^2}(\beta-\tilde{\beta})'(A+X'X)(\beta-\tilde{\beta})\right]$$

$$= MVN\left[\tilde{\beta},\ \sigma^2(A+X'X)^{-1}\right].$$

Marginalizing:

$$p(\beta \mid y) = \int_0^\infty p(\beta,\sigma \mid y)\, d\sigma .$$

Now, in the diffuse prior case, we found that

$$\int_0^\infty \left(\frac{1}{\sigma^{n+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})\right]\right\} d\sigma$$

$$\propto \left[\nu s^2 + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})\right]^{-n/2}.$$

So:

$$p(\beta \mid y) \propto \left[n'c_1^2 + (\beta-\tilde{\beta})'(A+X'X)(\beta-\tilde{\beta})\right]^{-\left(\frac{n'+k}{2}\right)}$$

$$= MVST\left[\tilde{\beta},\ c_1^2(A+X'X)^{-1}\left(\frac{n'}{n'-2}\right)\right]$$

NOTE:

i) $\tilde{\beta}$ is defined even if rank $(X) < k$!

ii) $\tilde{\beta}$ is a matrix weighted average of $\bar{\beta}$ and $\hat{\beta}$:

$$\tilde{\beta} = (A+X'X)^{-1}(A\bar{\beta}+X'y)$$

$$= \left[\frac{1}{\sigma^2}A + \frac{1}{\sigma^2}X'X\right]^{-1}\left[\frac{1}{\sigma^2}A\bar{\beta} + \frac{1}{\sigma^2}X'X\hat{\beta}\right]$$

(iii) If $\bar{\beta} = \hat{\beta}$, then $\tilde{\beta} = \hat{\beta}$ $(=\bar{\beta})$

(iv) $\tilde{\beta}$ does not depend on ν or c.

Finally, to draw inferences about σ:

$$p(\sigma \mid y) = \int_{-\infty}^{\infty} p(\beta,\sigma \mid y)\, d\beta$$

$$\propto \left(\frac{1}{\sigma^{n'+1}}\right)\exp\left[-\left(\frac{n'c_1^2}{2\sigma^2}\right)\right]$$

which is Inverted Gamma, with

$$E(\sigma \mid y) = \left[\frac{\Gamma(n'/2)-1}{\Gamma(n'/2)}\right]\left(\frac{n'}{2}\right)^{1/2} . c_1$$

Once we combine proper prior information with the sample information, the Bayes estimator differs from the MLE of β.