Outline

- Bayesian Analysis
  - Proof 1
  - Proof 2
  - Proof 2-1
  - Proof 3
  - Truncated normal
- Gibbs sampler
- EM Algorithm

Bayesian Analysis ----(1)

$$Y = x\beta + \varepsilon \qquad \varepsilon \sim N(O, \sigma^2 I_T)$$

- frequentist : assume structure about prob. of ε

$$\Rightarrow L(Y \mid x, \beta, \Sigma) = \frac{1}{\sqrt{2\pi} \mid \Sigma \mid^{\frac{1}{2}}} e^{-(Y-x\beta)' \sum^{-1} (Y-x\beta)}$$

Bayesian Analysis ----(2)

- Bayesian : find Prob. $(\beta \mid X, Y)$ $P(\Sigma \mid X, Y)$

$$P(Y \mid X, \beta, \sigma) \propto \frac{1}{\sigma^n} \exp[-\frac{1}{2\sigma^2}(Y - x\beta)'(Y - x\beta)]$$

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

V= n-k , $\hat{\beta} = (x'x)^{-1} x'y, S^2 = \dfrac{(Y - x\hat{\beta})^{'}(Y - x\hat{\beta})}{V}$

Guess $\beta \sim Normal$ Guess $\Sigma \sim$ *inverse Gamma*

Bayesian Analysis ----(3)

$$f(x \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha x^{\alpha+1}} e^{-\frac{1}{x\beta}} \qquad x = (0, \infty), \alpha > 0, \beta > 0$$

$$E(x) = \frac{1}{\beta(\alpha - 1)} \quad \text{if} \quad \alpha > 1$$

$$\operatorname{var}(x) = \frac{1}{\beta^2(\alpha-1)^2(\alpha-2)} \quad \text{if} \quad \alpha > 2$$

$$\frac{1}{x} \sim Gamma(\alpha, \beta)$$

Bayesian Analysis ----(4)

$$Gamma : f(x \mid \ ,\ ) \ \frac{1}{(\ )} x^{\ 1} e^{\ \frac{x}{\ }}$$

$$E(x) = \alpha\beta$$

$$\operatorname{var}(x) = \alpha\beta^2$$

$$x^2(n) = G(\frac{n}{2}, 2)$$

× ∝ ×

is not function of parameters

If we ignore prior, focus on likelihood : L=L(data|θ)

As Bayesian , biggest trick: same likelihood but from perspective of θ being random & data fixed

Bayesian Analysis ----(6)

Ex:

$$Y = x\beta + \varepsilon$$

$$L = \frac{1}{\sqrt{2\pi}\,|\Sigma|^{\frac{1}{2}}} \exp[-(Y - x\beta)'\Sigma^{-1}(Y - x\beta)]$$

Bayesian $\quad L = L(\beta, \Sigma \mid data)$

Marginal likelihood for β: $\quad \int L(\beta, \Sigma \mid data) d\Sigma$

Bayesian Analysis ----(7)

For univariate case:

$$L(\beta, \sigma \mid data) = P(Y \mid x, \beta, \sigma) \propto \frac{1}{\sigma^n} \exp[\frac{-1}{2\sigma^2}(Y - x\beta)'(Y - x\beta)]$$

show as function of β (ignore the prior for this time being)

$$\propto \frac{1}{\sigma^n} \exp[-\frac{1}{2\sigma^2}(Y - x\beta)'(Y - x\beta)]$$

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

Bayesian Analysis ----(8)

$$f(\beta \mid y, x) = \text{ marginal likelihood of } \beta: \int_0^{\infty} L(\beta, \sigma \mid data) d\sigma$$

$$f(\sigma \mid y, x) = \text{ marginal likelihood of } \sigma: \int_{-\infty}^{\infty} L(\beta, \sigma \mid data) d\beta$$

$$f(\sigma \mid y, x) \propto \frac{1}{\sigma^V} \exp(\frac{-VS^2}{2\sigma^2}) \sim \text{ inverse gamma}$$

Bayesian Analysis ----(9)

- Diffuse or uninformative prior $\quad f(\beta) \propto \frac{1}{\sigma}$ i.e. some constant

  $\left(\text{but} \quad \int_{-\infty}^{\infty} f(\beta) d\beta \neq 1 \Rightarrow \quad \text{improper prior} \right)$

  frequentist ≈ Bayesian

- Conjugate Prior:

  give the posteriors same distribution as the likelihood function.

Bayesian Analysis ----(10)

Data Set #2:

(1) $Y = x\beta + \varepsilon$ OLS

(2) Bayes: posterior on $\theta = \{\beta, \sigma\}$

a) Diffuse Prior $\overset{(or\ P(\theta))}{f(\theta)} \propto \frac{1}{\sigma}$

b) Conjugate Prior: $\beta \sim N(b, Q)$

$\sigma \sim IG(V, S^2)$

inverse gamma

Bayesian Analysis ----(11)

Ex:

$$y = x\beta + \varepsilon \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2) \qquad \text{Let } \theta = \{\beta, \sigma_\varepsilon^2\}$$

$$f(\theta \mid y, x) \propto L(y \mid x, \theta) \cdot p(\theta)$$

prior

look at from the perspective of θ random

and y, x fixed

Bayesian Analysis ----(12)

a) Diffuse prior: $p(\theta) \propto \dfrac{1}{\sigma_\varepsilon}$

$$L(y \mid x, \theta) \propto \frac{1}{\sigma_\varepsilon^T} \exp\{\frac{-1}{2\sigma_\varepsilon^2}(Y - x\beta)'(Y - x\beta)\}$$

see proof 1 $\propto \dfrac{1}{\sigma_\varepsilon^T} \exp\{\dfrac{-1}{2\sigma_\varepsilon^2}[VS^2 + (\beta - \hat{\beta})' x' x (\beta - \hat{\beta})]\}$

$$V = T - k \;,\; \hat{\beta} = (x'x)^{-1} x' y \;,\; S^2 = \frac{(Y - x\hat{\beta})'(Y - x\hat{\beta})}{V}$$

Bayesian Analysis ----(13)

$$f(\beta \mid y, x) = \int_0^\infty f(\beta, \sigma \mid y, x) d\sigma$$

see proof 2 $= \{VS^2 + (\beta - \hat{\beta})' x' x (\beta - \hat{\beta})\}^{-\frac{T}{2}}$

$$f(\sigma \mid y, x) = \int_{-\infty}^{\infty} \ldots\ldots \int_{-\infty}^{\infty} f(\beta, \sigma \mid y, x) d\beta_1 d\beta_2 \ldots\ldots d\beta_k$$

see proof 2 $\propto \dfrac{1}{\sigma^{V+1}} \exp(\dfrac{-VS^2}{2\sigma^2})$ $\quad \therefore \sigma^2 \sim IG(V, S^2)$

inverse gamma

(V=T-k)

Bayesian Analysis ----(14)

b) Proper conjugate prior :

Choose a parametric from for θ that affords tractability to Bayes' rule.

If prior:

$$f(\beta \mid \sigma^{-2}, x) \sim N(b, \sigma^2 Q)$$

$$f(\sigma^{-2} \mid x) \sim G(\hat{S}^{-2}, \hat{V})$$

Bayesian Analysis ----(15)

Then posterior:

$$f(\beta \mid \sigma^{-2}, y, x) \propto N(\bar{\Sigma}(Q^{-1}b + x'x\hat{\beta}), \bar{\Sigma})$$

see proof 3 $\bar{\Sigma} = (Q^{-1} + x'x)^{-1}$

$$f(\sigma^{-2} \mid y, x) \propto G(\hat{\sigma}^{-2}, \hat{V}) \qquad \hat{V} = T + V$$

$$\hat{\sigma}^{-2}\hat{V} = VS^2 + (Y - x\hat{\beta})'(I_T + xQx')(Y - x\hat{\beta})$$

see proof 3

$f(\beta \mid y, x) \propto t$ distribution

Bayesian Analysis ----(16)

c) Proper conjugate prior:

Prior :

$$f(\beta) \sim N(b, Q) \equiv (2\pi)^{-\frac{k}{2}} |Q|^{-\frac{1}{2}} \exp[(\beta - b)' Q^{-1} (\beta - b)]$$

$$f(\sigma^{-2}) \sim G(\hat{S}^{-2}, \hat{V}) \equiv \frac{1}{\Gamma(\alpha)\beta^\alpha} (\sigma^{-2})^{\alpha - 1} \exp(-\frac{\hat{V}\hat{S}^2}{2\sigma^2})$$

$$= \frac{1}{\Gamma(\frac{\hat{V}}{2})(\frac{2}{\hat{V}\hat{S}^2})^{\frac{\hat{V}}{2}}} (\sigma^{-2})^{\frac{\hat{V}}{2} - 1} \exp\left(-\frac{\hat{V}\hat{S}^2}{2\sigma^2}\right)$$

Assume (prior β) $\perp$ (prior σ^{-2}) Then $f(\beta, \sigma^{-2} \mid x) = f(\beta) f(\sigma^2)$

Bayesian Analysis ----(17)

Posterior joint : $f(\beta, \sigma^{-2} \mid y, x) \propto f(y \mid \beta, \sigma^{-2}, x) \cdot f(\beta, \sigma^{-2} \mid x)$

$$= (2\pi\sigma^2)^{-\frac{T}{2}} \exp[\frac{-1}{2\sigma^2}(VS^2 + (\beta - \hat{\beta})' x' x (\beta - \hat{\beta}))] \bullet$$

$$(2\pi)^{-\frac{k}{2}} |Q|^{-\frac{1}{2}} \exp[-\frac{1}{2}(\beta - b)' Q^{-1} (\beta - b)] \bullet$$

$$\frac{1}{\Gamma(\frac{\hat{V}}{2})(\frac{2}{\hat{V}\hat{S}^2})^{\frac{\hat{V}}{2}}} (\sigma^{-2})^{\frac{\hat{V}}{2} - 1} \exp(-\frac{\hat{V}\hat{S}^2}{2\sigma^2})$$

Bayesian Analysis ----(18)

$$\propto (\sigma^{-2})^{\frac{1}{2}(T+\hat{V}-2)} \exp\{\frac{-1}{2\sigma^2}[\underbrace{VS^2+\hat{V}\hat{S}^2+(\beta-\hat{\beta})'x'x(\beta-\hat{\beta})}_{A}+\sigma^2\underbrace{(\beta-b)'Q^{-1}(\beta-b)}_{B}]\}$$

$$= y^{\frac{1}{2}a}\exp[-\frac{1}{2}y(A+\frac{B}{y})]$$

where $\left(\begin{array}{l} y=\sigma^{-2} \\ a=T+\hat{V}-2 \end{array} \right)$

$$= y^{\frac{1}{2}a}\exp[-\frac{1}{2}(Ay+B)]$$

Bayesian Analysis ----(19)

Posterior

$$f(\beta \mid y,x)=\int_0^\infty f(\beta,\sigma^{-2}\mid y,x)d\sigma^{-2}$$

$$\propto e^{-\frac{B}{2}}\int_0^\infty y^{\alpha-1}e^{-\frac{A}{2}y}dy$$

$\left(\text{where} \quad \alpha=\frac{T+\hat{V}}{2} \right)$

Bayesian Analysis ----(20)

posterior

$$f(\beta \mid y,x)\propto e^{-\frac{B}{2}}\int_0^\infty y^{\alpha-1}e^{-\frac{A}{2}y}dy$$

$$= e^{-\frac{B}{2}}\int(\frac{2x}{A})^{\alpha-1}e^{-x}(\frac{2}{A})dx$$

$$= e^{-\frac{B}{2}}(\frac{2}{A})^{\alpha}\Gamma(\alpha)$$

$\left(\begin{array}{l} \text{where} \quad x=\frac{A}{2}y \\ \qquad \alpha=\frac{T+\hat{V}}{2} \end{array} \right)$

$$\propto \{\exp[\frac{-1}{2}(\beta-b)'Q^{-1}(\beta-b)]\}\cdot\{\underbrace{VS^2+\hat{V}\hat{S}^2+(\beta-\hat{\beta})'xx'(\beta-\hat{\beta})}_{M}\}^{-\alpha}$$

Bayesian Analysis ----(21)

< d.f. >

Let $M = VS^2+\hat{V}\hat{S}^2+(\beta-\hat{\beta})'xx'(\beta-\hat{\beta})$

$\beta \sim$ multivariate t dist. with $(T+\hat{V}-1)$

$$E(\beta)=\hat{\beta} \qquad \text{var}(\beta)=\frac{VS^2+\hat{V}\hat{S}^2}{T+\hat{V}-3}(x'x)^{-1}$$

see proof 2-1

Bayesian Analysis ----(22)

If $T\to\infty$, $\beta \sim N(\hat{\beta},S^2(x'x)^1)$

Hence

$$f(\beta \mid y,x) \overset{T\to\infty}{\propto} N(\bar{\Sigma}(Q^{-1}b+x'x\hat{\beta}),\bar{\Sigma})$$

where $\bar{\Sigma}=(Q^{-1}+\frac{x'x}{S^2})^{-1}$

Q.E.D.

Bayesian Analysis ----(23)

posterior

$$f(\sigma^{-2}\mid y,x)=\int f(\beta,\sigma^{-2}\mid y,x)d\beta$$

$$\propto (\sigma^{-2})^{\frac{1}{2}(T+\hat{V}-2)}\exp[\frac{-1}{2\sigma^2}(VS^2+\hat{V}\hat{S}^2)]\bullet$$

$$\int\exp[\frac{-1}{2\sigma^2}(\beta-\hat{\beta})'x'x(\beta-\hat{\beta}) + \frac{-1}{2}(\beta-b)'Q^{-1}(\beta-b)]d\beta$$

$\longrightarrow$ It's not *IG*!!

<Back>

Proof 1: ----(1)

$$\theta=\{\beta,\sigma^2\} \qquad Y=x\beta+\varepsilon \qquad \hat{\varepsilon}=Y-x\hat{\beta}$$

$$L(y\,|\,x,\theta)=(2\pi)^{-\frac{T}{2}}(\sigma^2)^{-\frac{T}{2}}\exp\{\frac{-1}{2\sigma^2}(Y-x\beta)'(Y-x\beta)\}$$

$$\begin{aligned}(Y-x\beta)'(Y-x\beta) &= (Y-x\hat{\beta}+x\hat{\beta}-x\beta)'(Y-x\hat{\beta}+x\hat{\beta}-x\beta)\\ &= (Y-x\hat{\beta})'(Y-x\hat{\beta})+(\hat{\beta}-\beta)'x'x(\hat{\beta}-\beta)\\ &\quad +(\underbrace{Y-x\hat{\beta}}_{\hat{\varepsilon}})'x(\hat{\beta}-\beta)+(\hat{\beta}-\beta)'x'(\underbrace{Y-x\hat{\beta}}_{\hat{\varepsilon}})\\ &= (Y-x\hat{\beta})'(Y-x\hat{\beta})+(\hat{\beta}-\beta)'x'x(\hat{\beta}-\beta)\end{aligned}$$

Proof 1: ----(2)

$$\because \hat{\varepsilon}'x=x'\hat{\varepsilon}=0$$

$$\therefore L(y\,|\,x,\theta)=(2\pi)^{-\frac{T}{2}}(\sigma^2)^{-\frac{T}{2}}\exp\{\frac{-1}{2\sigma^2}[VS^2+(\beta-\hat{\beta})'x'x(\beta-\hat{\beta})]\}$$

where $\begin{cases} V=T-k\ ,\ S^2=\dfrac{(Y-x\hat{\beta})'(Y-x\hat{\beta})}{V} = \dfrac{\hat{\varepsilon}'\hat{\varepsilon}}{T-k}=\hat{\sigma}^2 \\ \hat{\beta}=(x'x)^{-1}xy\ ,\ \operatorname{var}(\hat{\beta})=\sigma^2(x'x)^{-1}\end{cases}$

Q.E.D.

or

$$f(y\,|\,x,\theta)\propto\frac{1}{\sigma^T}\exp\{\cancel{\frac{-1}{2\sigma^2}}[\cancel{VS^2}-\frac{1}{2}[\frac{\hat{\sigma}^2}{\sigma^2}+\frac{(\beta-\hat{\beta})'(\beta-\hat{\beta})}{\operatorname{var}(\hat{\beta})}]]\}$$

<Back>

Proof 2: ----(1)

$$f(\theta\,|\,y,x)=\frac{f(y,\theta\,|\,x)}{f(y\,|\,x)}=\frac{f(y\,|\,x,\theta)\overbrace{f(\theta)}^{\text{prior}}}{f(y\,|\,x)}$$

$$\theta=\{\beta,\sigma\}$$

$$\therefore f(\beta,\sigma\,|\,y,x)\propto\frac{1}{\sigma^{T+1}}\exp\{\frac{-1}{2\sigma^2}[\underbrace{VS^2+(\beta-\hat{\beta})'x'x(\beta-\hat{\beta})}_{A}]\}$$

$$(\text{if}\quad f(\theta)\propto\frac{1}{\sigma}\ )$$

Proof 2: ----(2)

1) $f(\beta\,|\,y,x)=\displaystyle\int_0^\infty f(\beta,\sigma\,|\,y,x)\,d\sigma$

$$\propto +\frac{1}{2}\int_0^\infty y^{\frac{T}{2}-1}\exp(-\frac{A}{2}y)\,dy$$

$$=\frac{1}{2}\Gamma(\frac{T}{2})(\frac{2}{A})^{\frac{T}{2}}\int_0^\infty\frac{1}{\Gamma(\frac{T}{2})(\frac{2}{A})^{\frac{T}{2}}}y^{\frac{T}{2}-1}e^{-\frac{y}{2/A}}\,dy$$

see Proof 2-1.

$$\propto [A]^{-\frac{T}{2}} \sim \text{multivariate } t \text{ dist. with } \underset{T-k}{\underset{\shortparallel}{\upsilon}} \text{ d.f.}$$

Proof 2: ----(3)

Note:

Let $\quad y=\sigma^{-2}$

$$\Rightarrow y^{-\frac{1}{2}}=\sigma$$

$$-\frac{1}{2}y^{-\frac{3}{2}}dy=d\sigma$$

Let $\quad A=VS^2+(\beta-\hat{\beta})'x'x(\beta-\hat{\beta})$

Proof 2: ----(4)

2) $f(\sigma\,|\,y,x)=\displaystyle\int_{-\infty}^{\infty}f(\beta,\sigma\,|\,y,x)\,d\beta \qquad\qquad \beta=k\times 1$

$$\propto\frac{1}{\sigma^{T+1}}\exp(\frac{-VS^2}{2\sigma^2})\int_{-\infty}^{\infty}\exp(\frac{-1}{2\sigma^2}A)\,d\beta$$

$$=\frac{1}{\sigma^{T+1}}\exp(\frac{-VS^2}{2\sigma^2})(2\pi\sigma^2)^{\frac{k}{2}}\int_{-\infty}^{\infty}(2\pi\sigma^2)^{-\frac{k}{2}}\exp(\frac{-1}{2\sigma^2}A)\,d\beta$$

$$\propto\frac{1}{\sigma^{T-k+1}}\exp(\frac{-VS^2}{2\sigma^2}) \qquad [\ \text{where}\ \ \alpha=\frac{V-1}{2}\ ,\ \beta=\frac{2}{VS^2}\ ]$$

$$\xrightarrow{called}\sigma^2\sim IG(V,S^2)\ \text{ or }\ \sigma^2\sim IG(\alpha,\beta)$$

Proof 2: ----(5)

if $\quad x \sim G(\alpha, \beta): \; f(x) = \dfrac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}}$

if $\quad x \sim IG(\alpha, \beta): \; f(x) = \dfrac{1}{\Gamma(\alpha)\beta^{\alpha} x^{\alpha+1}} e^{-\frac{1}{x\beta}}$

$$E(\sigma^2) = \frac{1}{\beta(\alpha-1)} = \frac{VS^2}{V-3}$$

$$\operatorname{var}(\sigma^2) = \frac{1}{\beta^2(\alpha-1)^2(\alpha-2)} = \frac{2V^2S^4}{(V-3)^2(V-5)}$$

Proof 2: ----(6)

or $\quad \begin{cases} \sigma^{-2} \sim G(V, S^2) \\ \sigma^{-2} \sim G(\alpha, \beta) \end{cases}$ $\qquad$ where $\quad \begin{cases} \alpha = \dfrac{V+3}{2} \\ \beta = \dfrac{2}{VS^2} \end{cases}$

$$E(\sigma^{-2}) = \alpha\beta = \frac{V+3}{VS^2}$$

$$\operatorname{var}(\sigma^{-2}) = \alpha\beta^2 = \frac{2(V+3)}{V^2S^4}$$

<Back>

Proof 2-1: ----(1)

$$\boxed{(VS^2 + \Delta\beta' x'x \Delta\beta)^{-\frac{T}{2}} \sim t \text{ dist. with } \upsilon = T-k \text{ d.f. } (\Delta\beta = \beta - \hat{\beta})}$$

Note: $\quad x \sim t-dist. : f(x) = k[1 + \dfrac{x^2}{n}]^{-\frac{1}{2}(n+k)}$ $\quad$ with n d.f.

Let $\quad A = (VS^2 + \Delta\beta' x'x \Delta\beta)^{-\frac{T}{2}} = k[1 + \dfrac{\Delta\beta'\Sigma\Delta\beta}{T-k}]^{-\frac{1}{2}(T-k+k)}$

Where $\quad k = (VS^2)^{-\frac{T}{2}} \;, \qquad \Sigma = \dfrac{\cancel{(T-k)}x'x}{\cancel{V}S^2} = \dfrac{xx'}{S^2}$

Proof 2-1: ----(2)

Let $\quad \Delta\beta = Ay \quad$ and $\quad A'\Sigma A = I \;, \; n = T-k = V$

Then $\quad A = k'[1 + \dfrac{y'y}{n}]^{-\frac{1}{2}(n+k)}$

$\therefore y \sim$ multivariate t dist. $\qquad$ Q.E.D

$$E(y) = 0, \qquad \operatorname{var}(y) = \frac{n}{n-2} I_k = \frac{\upsilon}{\upsilon-2} I_k$$

Proof 2-1: ----(3)

$$\operatorname{var}(\Delta\beta) = \operatorname{var}(\beta) = A(\operatorname{var} y)A'$$

note: $\; k' = k \cdot |A| \;$, where $\; |A| = |\Sigma|^{-\frac{1}{2}}$

$\because \; |A'\Sigma A| \; = \; |A'||\Sigma||A| \; = 1$

since $\quad A'\Sigma A = I \;, \; \Sigma = A'^{-1}A^{-1} = (AA')^{-1} \Rightarrow AA' = \Sigma'$

$\therefore \begin{cases} \operatorname{var}(\beta) = A(\dfrac{\upsilon}{\upsilon-2} I_k)A' = \dfrac{\upsilon}{\upsilon-2} AA' = \dfrac{\upsilon}{\upsilon-2}\Sigma^{-1} = \dfrac{\upsilon}{\upsilon-2} S^2(x'x)^{-1} \\ E(\Delta\beta) = E(Ay) = 0 \qquad \therefore \; E(\beta) = \hat{\beta} \end{cases}$

Proof 2-1: ----(4)

$x \sim$ multivariate t dist. : $\; [t_k, (\mu, \Sigma, \overset{\text{d.f.}}{\upsilon})]$

$$f(x) = \underset{\text{const.}}{\underset{\downarrow}{k}}[\upsilon + (x-\mu)'\Sigma^{-1}(x-\mu)]^{-\frac{1}{2}(\upsilon+k)}$$

$$E(x) = \mu$$

$$\operatorname{var}(x) = \frac{\upsilon}{\upsilon-2}\Sigma$$

<Back>

Proof 3: ----(1)

Prior

$$\begin{cases} f(\beta \mid \sigma^{-2}, x) \sim N(b, \sigma^2 Q) \equiv (2\pi\sigma^2)^{-\frac{k}{2}} \mid Q \mid^{-\frac{1}{2}} \exp[-\frac{1}{2\sigma^2}(\beta - b)' Q^{-1}(\beta - b)] \\ f(\sigma^{-2} \mid x) \sim G(\hat{S}^{-2}, \hat{V}) \equiv \frac{1}{\Gamma(\alpha)\beta^{\alpha}} (\sigma^{-2})^{\alpha - 1} e^{-\frac{\sigma^{-2}}{\beta}} \end{cases}$$

$$= \frac{1}{\Gamma(\frac{\hat{V}}{2})(\frac{2}{\hat{V}\hat{S}^2})^{\frac{\hat{V}}{2}}} (\sigma^{-2})^{\frac{\hat{V}}{2} - 1} \exp(-\frac{\hat{V}\hat{S}^2}{2\sigma^2})$$

(where $\alpha = \frac{\hat{V}}{2}, \beta = \frac{2}{\hat{V}\hat{S}^2}$)

Proof 3: ----(2)

$$E(\sigma^{-2}) = \alpha\beta = \hat{S}^{-2}$$

$$\operatorname{var}(\sigma^{-2}) = \alpha\beta^2 = \frac{2}{\hat{V}\hat{S}^4}$$

$$\xrightarrow{\text{prior :}} f(\beta, \sigma^{-2} \mid x) = f(\beta \mid \sigma^{-2}, x) \cdot f(\sigma^{-2} \mid x)$$

Assume: (prior indep. btw β, σ^{-2})

Proof 3: ----(3)

So posterior join dist.

$$f(\beta, \sigma^{-2} \mid y, x) = \frac{f(y \mid \beta, \sigma^{-2}, x) \cdot f(\beta, \sigma^{-2} \mid x)}{f(y \mid x)}$$

$$\propto f(y \mid \beta, \sigma^{-2}, x) \cdot f(\beta, \sigma^{-2} \mid x)$$

Proof 3: ----(4)

$$f(\beta, \sigma^{-2} \mid y, x) = (2\pi\sigma^2)^{-\frac{T}{2}} \exp[\frac{-1}{2\sigma^2}(VS^2 + (\beta - \hat{\beta})' x' x (\beta - \hat{\beta}))] \bullet$$

$$(2\pi\sigma^2)^{-\frac{k}{2}} \mid Q \mid^{-\frac{1}{2}} \exp[\frac{-1}{2\sigma^2}(\beta - b)' Q^{-1}(\beta - b)] \bullet$$

$$\frac{1}{\Gamma(\frac{\hat{V}}{2})(\frac{2}{\hat{V}\hat{S}^2})^{\frac{\hat{V}}{2}}} (\sigma^{-2})^{\frac{\hat{V}}{2} - 1} \exp(-\frac{\hat{V}\hat{S}^2}{2\sigma^2})$$

$$\propto (\sigma^{-2})^{\frac{1}{2}(T + k + \hat{V} + 2)} \exp\{\frac{-1}{2\sigma^2}[VS^2 + \hat{V}\hat{S}^2 + \underbrace{(\beta - b)' Q^{-1}(\beta - b) + (\beta - \hat{\beta})' x' x (\beta - \hat{\beta})}_{M}]\}$$

Proof 3: ----(5)

Let $M = (\beta - b)' Q^{-1}(\beta - b) + (\beta - \hat{\beta})' x' x (\beta - \hat{\beta})$

$$= (\hat{\beta} - b)' x' x \overline{Q} Q^{-1}(\hat{\beta} - b) + (\beta - \overline{b})' \overline{Q}^{-1} x (\beta - \overline{b})$$

Where $\begin{cases} \overline{Q} = (Q^{-1} + x'x)^{-1} \\ \overline{b} = \overline{Q}(Q^{-1}b + x'x\hat{\beta}) \end{cases}$ $\begin{cases} E(\overline{b}) = \beta \\ \operatorname{var}(\overline{b}) = \sigma^2 \overline{Q} \end{cases}$

Proof 3: ----(6)

posterior join

$$\therefore\ f(\beta, \sigma^{-2} \mid y, x) \propto (\sigma^{-2})^{\frac{1}{2}(T + k + \hat{V} + 2)} \exp\{\frac{-1}{2\sigma^2}[VS^2 + \hat{V}\hat{S}^2 + M]\}$$

$$\Rightarrow f(\sigma^{-2} \mid y, x) = \int_{-\infty}^{\infty} f(\beta, \sigma^{-2} \mid y, x) d\beta = \ldots\ldots\ldots\ldots\ldots$$

$$= \left[\frac{(\sigma^{-2})^{\frac{N}{2} - 1}}{\Gamma(\frac{N}{2})(\frac{2}{\lambda})^{\frac{N}{2}}} \exp[-\frac{\lambda^{\sigma^{-2}}}{2}] \right] \sim Gamma(\alpha, \beta), \alpha = \frac{N}{2}, \beta = \frac{2}{\lambda}$$

Proof 3: ----(7)

where

$$N = \hat{V} + T$$

$$\lambda = \hat{V}\hat{S}^2 + \underbrace{(y - x\hat{\beta})'(y - x\hat{\beta})}_{VS^2} + (\hat{\beta} - b)'Q^{-1}\,\overline{Q}x'x(\hat{\beta} - b)$$

$$E(\sigma^{-2}) = \alpha\beta = \frac{N}{\lambda}$$

$$\operatorname{var}(\sigma^{-2}) = \alpha\beta^2 = \frac{2N}{\lambda^2}$$

Proof 3: ----(8)

or $\quad f(\sigma^2 \mid y, x) \sim IG(\alpha', \beta') \quad \begin{cases} \alpha' = \dfrac{N}{2} - 2 \\ \beta' = \dfrac{2}{\lambda} \end{cases}$

$$E(\sigma^2) = \frac{1}{\beta(\alpha - 1)} = \frac{\lambda}{N - 6}$$

$$\operatorname{var}(\sigma^2) = \frac{1}{\beta^2(\alpha - 1)^2(\alpha - 2)} = \frac{2\lambda^2}{(N - 6)^2(N - 8)}$$

Proof 3: ----(9)

posterior

$$f(\beta \mid \sigma^{-2}, x, y) = (2\pi\sigma^2)^{-\frac{k}{2}} \mid \overline{Q} \mid^{-\frac{1}{2}} \exp\{-\frac{1}{2\sigma^2}(\beta - \overline{b})'\overline{Q}^{-1}(\beta - \overline{b})\}$$

$$\begin{cases} E(\beta \mid \sigma^{-2}) = \overline{b} \\ \operatorname{var}(\beta \mid \sigma^{-2}) = \sigma^2\overline{Q} \end{cases}$$

Proof 3: ----(10)

marginal posterior

$$f(\beta \mid x, y) = \{\frac{\Gamma(\frac{k+N}{2})}{\Gamma(\frac{N}{2})(\pi N)^{\frac{k}{2}}} \mid (\frac{\lambda}{N})\overline{Q} \mid^{-\frac{1}{2}} \cdot [1 + (\frac{1}{N})(\beta - \overline{b})'[\frac{\lambda}{N}\overline{Q}]^{-1}(\beta - \overline{b})]^{-\frac{k+N}{2}}\}$$

k-dimentional t dist. with N d.f. , mean $\overline{b}$ and scale matrix $(\frac{\lambda}{N}) \cdot \overline{Q}$

$$\begin{cases} E(\beta) = \overline{b} \\ \operatorname{var}(\beta) = (\dfrac{N}{N - 2})(\dfrac{\lambda}{N}\overline{Q}) = \dfrac{\lambda}{N - 2}\overline{Q} \end{cases}$$

<Back>

Truncated normal: ----(1)

(1) to the left (at $x = \tau$) $\quad x \sim N(\mu, \sigma^2)$

$$f(x) = \begin{cases} 0 & \text{if } x < \tau \\ \dfrac{k}{\sqrt{2\pi}\sigma}\exp[-\dfrac{1}{2}(\dfrac{x - \mu}{\sigma})^2] & \text{if } x \geq \tau \end{cases}$$

$$\Rightarrow f(x) = \frac{\frac{1}{\sigma}\phi(\frac{x - \mu}{\sigma})}{1 - \Phi(\alpha)}$$

where

$$k = \frac{1}{1 - \Phi(\frac{\tau - \mu}{\sigma})} \qquad \alpha = \frac{\tau - \mu}{\sigma}$$

Truncated normal: ----(2)

$$E(x) = \mu + \frac{\sigma}{1 - \Phi(\frac{\tau - \mu}{\sigma})}\frac{1}{\sqrt{2\pi}}\exp[-\frac{1}{2}(\frac{\tau - \mu}{\sigma})^2] = \mu + \sigma\lambda(\alpha)$$

where $\quad \lambda(\alpha) = \dfrac{\phi(\alpha)}{1 - \Phi(\alpha)} \qquad \alpha = \dfrac{\tau - \mu}{\sigma}$

if $\quad \tau = 0 \Rightarrow \lambda(\alpha) = \dfrac{\phi(\frac{\mu}{\sigma})}{\Phi(\frac{\mu}{\sigma})}$

Truncated normal: ----(3)

(2) To the right (at $x = \tau$)

$$f(x) = \begin{cases} 0 & \text{if} \quad x > \tau \\ \dfrac{k}{\sqrt{2\pi}\sigma}\exp[-\dfrac{1}{2}(\dfrac{x-\mu}{\sigma})^2] & \text{if} \quad x \le \tau \end{cases}$$

where

$$\Rightarrow f(x) = \frac{\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})}{\Phi(\alpha)}$$

$$k = \frac{1}{\Phi(\frac{\tau-\mu}{\sigma})}$$

Φ : cdf of std. normal

Truncated normal: ----(4)

$$E(x) = \mu - \frac{\sigma}{\Phi(\frac{\tau-\mu}{\sigma})}\frac{1}{\sqrt{2\pi}}\exp[-\frac{1}{2}(\frac{\tau-\mu}{\sigma})^2] = \mu + \sigma\lambda(\alpha)$$

where $\lambda(\alpha) = -\dfrac{\phi(\alpha)}{\Phi(\alpha)}$ if $\tau = 0 \Rightarrow \lambda(\alpha) = \dfrac{-\phi(\frac{\mu}{\sigma})}{1-\Phi(\frac{\mu}{\sigma})}$

$$\boxed{\begin{array}{l} \mathrm{var}(x) = \sigma^2[1-\sigma(\alpha)] \quad \text{where} \quad \sigma(\alpha) = \lambda(\alpha)[\lambda(\alpha)-\alpha] \\ \qquad\qquad 0 < \sigma(\alpha) < 1 \quad \text{for all } \alpha \end{array}}$$

Truncated normal: ----(5)

Tobit: (for censored data)

$$y_i^* = x_i'\beta + \varepsilon_i \qquad \varepsilon_i \sim iid \;\; N(0, \tau^{-2}) \qquad \tau = \frac{1}{\sigma}$$

$$y_i = \max\{y_i^*, 0\} \qquad i = 1, 2, \ldots\ldots\ldots n$$

ex:

y_i^* : observed stock price

[$y_i^* = y$ rounded to nearest 1/8 spread]

$$n = \underbrace{n_0}_{\#\text{of } \; y_i = 0} + \underbrace{n_1}_{\#\text{ of } \; y_i > 0}$$

Truncated normal: ----(6)

$$L(\beta, \tau^{-2}) = \{\prod_{i \in c}[1-\Phi(x_i'\beta\tau)]\}(2\pi)^{-\frac{n_1}{2}}(\tau^2)^{\frac{n_1}{2}}\exp\left\{\frac{-\tau^2\|y_u - x_u'\beta\|^2}{2}\right\}$$

$c = \{i \mid y_i^* = 0\}$

$u = \{i \mid y_i^* > 0\}$

y_u : set of uncensored

x_u : corresponding x_i

<Back>

Gibbs Sampler ----(1)

$$1 - \Phi(x_i'\beta\tau) = p(y_i = 0) = p(y_i^* < 0)$$

If we could observe n_o , y_i corresponding to c , then this would be a standard problem.

If I knew τ & β , then I could characterize probability of $y_i \in c$ (Integrate over this space)

Gibbs Sampler ----(2)

Integration can be accomplished by taking successive draws from conditional posterior densities:

1) make up $\hat{y}_i$ for $y_i \in c(\hat{y}_i \in c)$ & τ i.e. all negative 1 , initially

2) draw β from $f(\beta \mid x, y, \hat{y}, \tau) \rightarrow \hat{\beta}$

3) draw τ from $f(\tau \mid x, \hat{\beta}, y, \hat{y},) \rightarrow \hat{\tau}$

4) draw $\hat{y}$ from $f(\hat{y} \mid x, \hat{y} < 0, \hat{\beta}, \hat{\tau}) \rightarrow \hat{\hat{y}}$

Throw if draw $>$ 0
Keep if draw $\le$ 0

continued step 2) ~ 4)

full conditional densities

Gibbs Sampler ----(3)

Use step 2) & 3) 1000 draws to construct marginal

(i.e. after throwing first 1000 draws)

posterior of $\beta,\ \tau$

If Prior : $\begin{cases} \beta \propto \tau \\ \tau \propto 1 \end{cases}$

Then the full conditional densities :

(like marginal posterior given augmented data)

Gibbs Sampler ----(4)

$$\begin{cases} f(y \in c \mid y, \beta, \tau) = \dfrac{N(x_i'\beta, \tau^{-2})}{1-\Phi(x_i'\beta\tau)} & < truncated \quad normal > \\ \tau^2 \sim G(\dfrac{V+3}{2}, \dfrac{2}{VS^2}) & y_z = \{y, \hat{y}\} \quad y_z \text{ :augmented data} \\ f(\beta \mid \tau, y, x, y \in c) = N(\hat{\beta}_z, \tau^{-2}(x'x)^{-1}) \end{cases}$$

$\hat{\beta}_z$: based on augmented data

The draws from step 2) , 3) , 4) consist of Markov chain process.

Gibbs Sampler ----(5)

Call DRNGAM (N,A,R)

N = 1 (# of rv's)

A = shape parameters

R = o/p (vector of length N)

Gibbs Sampler ----(6)

Call DRMVW (N, K, R, LR, S, LS)

N = # of vectors (1)

K = length of vector

R = upper-triangular Cholesky factor of var-cov matrix

LR = Ldg dim of R in calling program

S = o/p matrix

LS = Ldg dim of S

Gibbs Sampler ----(7)

Call DCHFAC (K, SIGMA, LDS, TOL, IRANK, R, LR)

Cholesky factorization

TOL : tolerance i.e. 1.0D-8

SIGMA : R'R $\longrightarrow$ var-cov matrix

K: order of sigma

Gibbs Sampler ----(8)

Iteration 1 : $\hat{\beta}$, $\hat{\tau}$, ... $\{\hat{y} \in c\}$

burning sample (throw out) $\vdots$

Iteration 1000: $\hat{\beta}$, $\hat{\tau}$, ... $\{\hat{y} \in c\}$

construct dist. of β & τ $\vdots$

Iteration 11,000: $\hat{\beta}$, $\hat{\tau}$, ... $\{\hat{y} \in c\}$

<Back>

EM Algorithm (Expectation Max) ----(1)

- Frequentists concentrate missing variables (m.v.) into likelihood.
- Instead of making draws for missing variables, we use the expectation of m.v.'s and put them in likelihood function. Then maximize the likelihood to get new MLE of parameters. Use new MLE to get new set of expectation of m.v.'s ……Iterate the procedure until $\hat{\theta}^{t} \approx \hat{\theta}^{t+1}$

  $p(y_j = 0 \mid x_j, \beta) = p(y_j < 0 \mid x_j, \beta)$

EM Algorithm ----(2)

1) $E(y_j^* \mid y_j = 0, x_j, \hat{\beta}, \tilde{\tau}) = x_j{}'\hat{\beta} + \hat{\tau}^{-1}\hat{\lambda}_i$

$$\hat{\lambda}_i = \frac{-\hat{\phi}_i}{1-\hat{\Phi}_i} \qquad \begin{cases} -\hat{\phi}_j = \phi(x_j{}'\hat{\beta}\hat{\tau}) \\ \hat{\Phi}_j = \Phi(x_j{}'\hat{\beta}\hat{\tau}) \end{cases}$$

2) Put $E(y_j^* \mid y_j = 0)$ along with $y_j^* \mid y_j > 0$ to get an augmented data which we use to get new MLE for β, τ

(same as OLS in this case)

EM Algorithm ----(3)

3) Then use new $\hat{\beta}$, $\hat{\tau}$ to get new

$E(y_j^* \mid y_j = 0, x_j, \hat{\beta}, \hat{\tau})$

4) Iterate the procedure until $\hat{\theta}^{t} \approx \hat{\theta}^{t+1}$

note: $\theta = (\beta, \tau)'$