Bayesian 22 April (1)

- sqrt(s2 t1)
- elearning slides of last session
- Check Priors → Gamma difference
- Check position wing

$\begin{cases} s2\,t1 = Cov(y - x * bols); \\ bols = regress(y, x); \end{cases}$ → For prior you must Calculate it rather than putting number that is unrelated

- Consistent with data starting value to avoid problem of Computer processing

☒ Lesson learned: pick priors more Carefully

- increasing Error term would make it diffuse
- exploding Error term - informative prior
- less variance will hinder the exploding
- Exploding = variance is not Constant
- This makes Error term is assymetric
- Coup douglas - log transformation Could be a solution for positive error term
- You know that estimate could not be very big so solution is to put boundary → {inside loop, outside loop}
- α and β → solve equation and put boundary based on result of solving systems of Equations

---

- For next week try to solve this: not yet Converged

AR(1)

$y_t = X_t\beta + \varepsilon_t \quad t = 1, \ldots, T$

$\Phi(L)\varepsilon_t = u_t \sim N(0, \sigma^2)$

$\varepsilon_t = \phi_1 \varepsilon_{t-1} + u_t \quad , t = 2, \ldots, T$

$\Phi(L) = 1 - \Phi_1 L$

$\begin{cases} \beta \mid \sigma^2, \Phi_1 \quad \text{I} \\ \sigma^2 \mid \beta, \Phi_1 \quad \text{⊛ II} \\ \Phi_1 \mid \beta, \sigma^2 \quad \text{III} \end{cases}$

$\Rightarrow y_t - \phi y_{t-1} = (X_t - \phi X_{t-1})\beta + \underline{u_t}$ → normal form

→ use like simple OLS ⊛ I II

For III we need to use $\varepsilon_t^* = Y_t - X_t\beta$

(2)

$\Rightarrow p(\Phi_1 \mid \beta, \sigma^2) \propto$

Φ would be normal

- since we previously we had $u_t \sim$ Normal ⇒ beta normal

$\underbrace{\varepsilon}_{\text{"}y\text{"}} = \phi_1 \underbrace{\varepsilon}_{\text{"}x\text{"}}$ → lag of ε_{t-1} on the second side

$\boxed{-1 < \Phi_1 < 1}$ ⊛ truncated normal

→ where each is $\varepsilon_t^* = Y_t - X_t\beta$

Mixture model

- two different states → Markov process → $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ transition: once start does not leave state

$p(s_t \mid s_{t+1}) = Mp(p, \pi_1)$ → Communicat Starting point

→ different universes

→ $= \{0, 1\}$

Probabilities of belonging to each of segments ← iid

- unobserved markov process

$p(y_t \mid s_t, \beta, \sigma^2)$ → You could have come either from this mixture normal or other

$y_t = X_t \beta_{s_t} + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma^2)$

- where I am today only depends on where I was today

- General case $p = \begin{bmatrix} p_1 & 1-p_1 \\ 1-q_1 & q_1 \end{bmatrix}$

- This is discrete hetrogeneity case

this is as opposed to Contineous hetrogeneity → discrete hetrogeneity

price sensitivity does not change over time

⊛ Could have Come from each of the states

$$\prod_{t=1}^{T} \frac{\pi_1}{\sqrt{2\pi}\sigma} e^{\frac{-(y_t - X_t\beta_1)^2}{2\sigma^2}} \times I[s_t = 0]$$

$$+ \frac{\pi_2}{\sqrt{2\pi}\sigma} e^{\frac{-(y_t - X_t\beta_2)^2}{2\sigma^2}} \times I[s_t = 1]$$

integrate out the parameter

$L = \prod (\quad)^{s_t} (\quad)^{1-s_t}$ → Switch on and off likelihood

π → percentage of people in each of the states

- like expected value of likelihood
- usually you can not pinpoint the value before so you use expected

parameters

β_1

β_2

σ^2

π_1

S_t

β
- if you observe state and estimate beta things would be sorted out

→ σ^2 you put related element

$\left\{ \begin{array}{l} \sigma_1 \\ \sigma_2 \end{array} \right.$ → easier

π_1 → gamma update π_1, π_2 → number of

for each indep # obsrv in each state

go and calculate ratio

S_t → you draw b/w 0/1 uniform
and calculate π and based on threshod
you decide whether it is 0 or 1

→ you need to not only have π_1
but the whole probability for each
observation

- inside markov chain you separate vector of
x_1, x_2 from x, and y_1, y_2 from y
based on prior vector S_t

- try to solve and debug program

- may 10th - read 1996 paper again

Chib 96

mixture $\rightarrow$ unobserved finite state random var $s_t \in \{1, \dots, m\}$

$S_t = (s_1, \dots, s_t)$ history

$S^{t+1} = (s_{t+1}, \dots, s_n)$ future

$$p(s_t \mid Y_n, S^{t+1}, \theta) \propto p(s_t \mid Y_t, \theta) \times f(Y^{t+1}, S^{t+1} \mid Y_t, s_t, \theta)$$
$$\propto p(s_t \mid Y_t, \theta) \propto p(s_{t+1} \mid s_t, \theta) \times f(Y^{t+1}, S^{t+2} \mid Y_t, s_t, s_{t+1}, \theta)$$
$$\propto p(s_t \mid Y_t, \theta) \times \underbrace{p(s_{t+1} \mid s_t, \theta)}_{\text{transition prob}} \quad \text{indep. of } s^t$$

according to normalize with sum

- Population select from an obs to next unobserved markov proces
- Simplification: (1) Data augmentation (2) population index vc into the list of unknown parameters
- Simulate latent $\rightarrow$ from joint dist given data, remaining param
- Stochastic EM Algorithm: SEM

markove mixture model (MMM)
Good for modeling persistence $\rightarrow$ serial correlation

- poisson
mixture of multivar normal
autoreg time series

chain: time homogeneous irreducible aperiodic

$s_t \mid s_{t-1} \sim \text{Markov}(P, \pi_1)$ one state transition probability

$y_t \mid Y_{t-1}, \theta_k \sim f(y_t \mid Y_{t-1}, \theta_k) \quad k = 1, \dots, m$

$$f(y_t \mid Y_{t-1}, s_{t-1}, \theta) = \begin{cases} \sum_{k=1}^{m} f(y_t \mid Y_{t-1}, \theta_k)\, \pi_1(s_t = k) & t = 1 \\ \sum_{k=1}^{m} f(y_t \mid Y_{t-1}, \theta_k)\, P(s_t = k \mid s_{t-1}) \end{cases}$$

e.g. pop of searcher vs not searcher

$\theta = \{ \bigcup_k \theta_k \} \cup \{ p_{ij} \mid \le i, j \le m-1 \}$ — Probability of state

Hidden markov = ~~mixture~~ markov mixture model

$s_1, s_2, \dots, s_n \mid Y_n, \theta \quad S_n \in \mathcal{S} = \{1, 2, \dots, m\}^n$
Conditional on full conditional distrib data and params

- at each step starting with terminal state s_n, only a single state has to be drawn

$$p(S_n \mid Y_n, \theta) = p(s_n \mid Y_n, \theta) \times \dots \times p(s_t \mid Y_n, S^{t+1}, \theta) \times \dots \times p(s_1 \mid Y_n, S^2, \theta)$$

$*$ Finding (1) $\rightarrow$ recursive $P(s_t \mid Y_{t-1}, \theta)$?

(1) prediction step
(2) update step

$P(s_{t-1} \mid Y_{t-1}, \theta)$ available

(1) $$P(s_t \mid Y_{t-1}, \theta) = \sum_{k=1}^{m} p(s_t \mid s_{t-1} = k, \theta) \times p(s_{t-1} = k \mid Y_{t-1}, \theta)$$

utilize $p(s_t \mid Y_{t-1}, s_{t-1}, \theta)$

(2) $$p(s_t \mid Y_t, \theta) \propto p(s_t \mid Y_{t-1}, \theta) \times f(y_t \mid Y_{t-1}, \theta)$$

$t = 1 \quad p(s_1 \mid Y_0, \theta)$ stationary dist. of chain
left eigen vector corres. eigenvalue of 1.

- transition matrix P given states

$P_i = (p_{i1}, \dots, p_{im})$ $\rightarrow$ indep $(Y_n, \bigcup_{k=1}^{m} \theta_k)$

Prior P_i: Dirichlet m-dim. simplex $P_i \sim \mathcal{D}(\alpha_{i1}, \alpha_{im})$

Post: Dirich

$$P \mid S_n \sim \mathcal{D}(\alpha_{i1} + n_{i1}, \dots, \alpha_{i1} + n_{im}) \quad i = 1, \dots, m$$

$\rightarrow$ total # one step transition

$$p_{i1} = \frac{x_1}{\sum_{j=1}^{m} x_j}, \dots, p_{im} = \frac{x_m}{\sum_{j=1}^{m} x_j} \quad x_j \sim \text{Gamma}(\alpha_{ij} + n_{ij}, 1)$$

vector from simplex

Monte Carlo EM

$$Q(\theta, \theta^i) = \int_{S_n} \log(\pi(\theta \mid Y_n, S_n))\, d(S_n \mid Y_n, \theta^i)$$

$$\hat{Q}(\theta, \theta^i) = \frac{\sum_{j=1}^{N} \log(\pi(\theta \mid Y_n, S_{n,j}))}{N}$$

$$\sum \log(\pi(p_k \mid Y_n, S_{n,j})) \propto \sum_{j=1}^{N} \sum_{l=1}^{m-1} (n_{kl,j} + \alpha_{kl} - 1) \log(p_{kl}) + (n_{km,j} + \alpha_{km} - 1) \log(1 - p_{k1} - \dots - p_{km-1})$$

transition state k to state l in simulation $S_{n,j}$

$$\hat{p}_{kl} = \frac{\sum_{j=1}^{N} (\alpha_{kl} + n_{kl,j})}{\sum_{j=1}^{N} \left(\sum_{l=1}^{m} (\alpha_{kl} + n_{kl,j}) \right)}$$

Poisson $f(y_t \mid \lambda_k) = \dfrac{\lambda_k^{y_t} e^{-\lambda_k}}{y_t!} \quad t = 1, 2, \dots, 240 \quad k = 1, \dots, m$

conjugate Gama

$\lambda_k \mid Y_n, S_n, P \sim \mathcal{G}(a_k, b_k)$ Prior

$\mathcal{G}(a_k + \sum_{t=1}^{n} y_t I\{s_t = k\}, b_k + N_k)$ full condi $k = 1, \dots, m$

$$\sum_{j=1}^{N} \log(\pi(\lambda_k \mid Y_n, S_{n,j})) \propto \sum_{j=1}^{N} (U_{k,j} + a_k - 1) \log(\lambda_k) - \sum_{j=1}^{N} (b_k + N_{k,j}) \lambda_k$$

$U_{k,j} = \sum_{t=1}^{n} y_t I[s_{t,j} = k]$

$$\hat{\lambda}_k = \sum_{j=1}^{N} (U_{k,j} + a_k - 1) \Big/ \sum_{j=1}^{N} (b_k + N_{k,j})$$

change in number $\rightarrow$ = new state

Autoreg GNP %change in GNP

$$f(y_t \mid Y_{t-1}, s_{t-1}, \alpha, \gamma) = \sum_{k=1}^{4} P(s_t = k, s_{t-1})\, f(y_t \mid Y_{t-1}, \alpha_k, \gamma, \sigma^2)$$

$$\alpha_k \mid Y_n, S_n, \gamma, \sigma^2 \sim N_1\left(V_k (A_0 \alpha_{0k} + \sigma^{-2} \sum_{t=1}^{n} z_t I(s_t = k)), N_k \right)$$

$$\gamma \mid Y_n, S_n, \alpha, \sigma^2 \propto N_p\left(V(T_0 \gamma_0 + \sigma^{-2} \sum_{t=p+1}^{n} x_t (y_t - \alpha_t)), N \right)$$

$$V_k = (A_{0k} + N_k \sigma^{-2})^{-1}$$

Econometrics | 15-April | Bayesian | 1

- Stochastic frontier model (today) SFM exponential two sided error
- mixture model - Hidden markov model
- time series

DEA → no assumptions
Data Envelop analysis
↳ Error term does not Exist

$y_i = f(x) = X_i\beta$

- most efficient use of time
- output based on input
- different outcome based on effort
- different output even if same input

$y_i = X_i\beta + u_i$, $i = 1, \dots, N$

optimal response $X_i\beta = \bar{y}_i$

everybody has not put u_i enough effort

$u_i \to$ negative (based on definition)

- not average response function, but best response function $y_i < \bar{y}_i$

analyzing efficiency

$\min_{\beta} \sum (y_i - X_i\beta)^2$

- Could be something else happen - unobserved events shift up or down maybe whether, and not enough effort of sales people

- which location well and which bad
- how relation input & output → ideal, negligance and inability

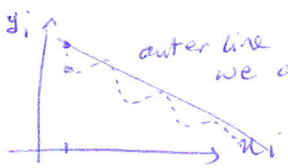
$y_i = X_i\beta + u_i + \varepsilon_i$

ε_i → two sided error (unobservable)

$u_i \geq 0$ one sided error (personal inability)

$X_i \to$ different outputs of different stores some large parking lot, some small parking lot

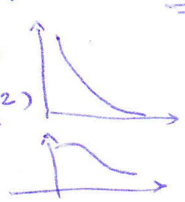
- all the stores assume to have same frontier



- need to solve → assume error structure

positive $u_i \sim \Gamma(1, \phi)$

$u_i \sim TN(0, \sigma_u^2)$ $(0, \infty)$



⇒ Drive MLE, likelihood & then gibbs sampler

$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ -- Completing - squares

2

$$p(y \mid \mu, \sigma_\varepsilon^2, \phi) = \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} e^{-\frac{(y_i - X_i\beta + u_i)^2}{2\sigma_\varepsilon^2}} e^{-\phi u_i} \phi \, du_i \propto \phi^N \prod_{i=1}^N \Phi(X_i\beta - y_i - \phi\sigma_\varepsilon^2) \cdot \underbrace{e^{2\phi \sum (y_i - X_i\beta) + \phi^2 \sigma_\varepsilon^2 \cdot N}}$$

Likelihood

↓

Natural parameter

---

for Bayesian

$$p(y \mid \mu, \sigma_\varepsilon^2, \phi, u_i) = \frac{\mathbb{I}(u_i > 0)}{\sqrt{2\pi}\sigma_\varepsilon} e^{-\frac{(y_i - X_i\beta + u_i)^2}{2\sigma_\varepsilon^2}} \prod_{i=1}^N e^{-\phi u_i} \phi \, du_i$$

- augmented variable here is $(u_i > 0)$
- The form should be Gamma for ϕ

$p(\beta) \sim N(\beta_0, B_0)$

$p(\sigma_\varepsilon^{-2}) \sim \Gamma(\frac{\alpha_0}{2}, \frac{\delta_0}{2})$

$p(\phi) \sim \Gamma(\alpha_0, \mu_0)$

$p(\beta \mid y, \sigma_\varepsilon^2, \phi, \{u_i\})$

difference : $\to y + u$

$p(\sigma_\varepsilon^2 \mid y, \beta, \phi, \{u_i\})$

$p(\phi \mid y, \beta, \sigma_\varepsilon^2, \{u_i\})$

$p(u_i \mid \beta, y, \sigma_\varepsilon^2, \phi)$

- any Constraint Can be added as augmented variable

$\underbrace{\underbrace{(\frac{X'X}{\sigma^2} + B_0^{-1})^{-1}}_{V} (\frac{X'Y}{\sigma^2} + B_0^{-1}\beta_0)}_{\text{mean}}$

$\alpha_1 = \frac{\alpha_0 + N}{2}$

$\delta_1 = (\frac{\delta_0 + e'e}{2})$

- inverse cdf sometimes better than accept rejection

- Homework to resolve this

Bayesian | April 8 | 1.502 | 1

→ during finals week submit paper, so no presentat

- today talk about estimation
- normaly estimatiation & derivative
- markov mixure model

- today estimating panel model & some version

$i = 1,\dots,N$
$t = 1,\dots,T$

$y_{it} = x_{it}\beta + w_{it}\beta_i + \varepsilon_{it}$

β_i → Contineous variable

- Prior bayesian discrete hetrogeneity

Assumptions

$\beta_i \sim N(0, D)$

$\sigma^{-2} \sim \Gamma(\frac{a_0}{2}, \frac{\delta_0}{2})$

$\beta \sim N(\beta_0, B_0)$ → if seperate w_{it}, x_{it} mean should also be here

Blocks

$\beta \mid \{\beta_i\}, \sigma^2, y$

$\sigma^2 \mid \{\beta_i\}, \beta, Y$

$\beta_i \mid D, \beta, \sigma^2, y$

$D \mid \{\beta_i\}$

(1) beta = [-1 1 2];

q → w

k → x

alternative way index [Po: Lo:];

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{2T} \\ \vdots \\ y_{5T} \end{bmatrix} \quad \begin{bmatrix} x_{11} \\ \vdots \\ x_{1T} \\ \vdots \\ \vdots \end{bmatrix} \quad \begin{bmatrix} e_{11} \\ e_{1T} \\ \vdots \end{bmatrix}$$

does not have to be norma

column in x — column w

(1) Simulate data
N = 100; T = 5; nn = N*T; k = 3; q = 1;
x = [Ones(nn,1) randn(nn, k-1)]; W = X(:,1);
et = sqr(Sigt)* randn(nn,1);
betaim = zeros(N, q); for i = 1:N; {betai = Col(D)* rand(q,1);
betaim(i,:) = betai;
eti = et(ii:ii+T-1,:);
xi = x(ii:ii+T-1,:); y(ii:ii+T-1,:) = yi; ii+=T;}

(2) set priors hyper paramete
beta0 = zeros(k,1); B0 = eye(k)*10; B0in = inv(B0);
alpha0 = 10; delta0 = 1; qq = q*(q+1)/2
Starting value s21 = 4.7; wbeta = zeros

(3) mcmc procedure ; m = bi + ni ; bi = 100; ni = 2000;
→ burn in

for m = 1:m;

(3-1) simulate beta

(3-2) simulate s2

(3-3) simulate betai

(3-4) simulate D

loop

(4) analysis
(Mean(betam)) sqrt[Cov(betam))]
other parameters → check the order to make sure it has Converged

/95

2

- hetrogenity
- Random effect
- Check cov & mean of draws of β_i to make sure thata is generated Correctly

- way to check data (Sanity)
: no post would be zero
wrong index skip

- look at y and measure in 1000,000 or ...
post variance with same order

Sol - Normaligzing could help in this case, to make it small

- rather than saving β_i, we can capture hetrogeneity by looking at D

- Change prior hyperparameter from 10 to 3 helps since it is saying we have 10 segments (degree of freedom of wishart)

(1) Bounded Rationality - stochastic mind

(2) Saying if I know I will predict ← dont observe everything
- decision science or game theoric
another paradigm → implications

- if you have all data you can not know what I do even if you have all the informatio
- mixure equilibrium
- inherent stochasticity

- metrapolis Hasting
- multinomial probit
- identification is normally pobrem
{ - multivariate probit ⟼ Covariate & basket levels
{ - multinomial probit ⟼ IIA
biased Coefficients

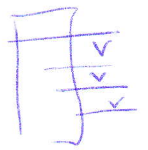
Covariate matrix ⇐ { look at them and say Complement or substitute

- assumptions are made to make identification possible so first variance equal to one

- as base more frequently purchase rather than spersely purchased

- Category choice → choice at high level

- Cut off points → utility simulating - larger than zero



① - truncated from above for J-2 select
② - select from current select minus infinity

- The difference of draws for beta is that this time we have SUR updates

- Same set up → we have J equations

- metrapolis hasting → wishart but in the form of restrictions for identification

  - not like MLE all at the same time isolate parameters and conditioning helps it would not be random walk anymore

⟹ makes Convergence faster

Conditioning: ① Simplify: Just parent (one stage) & not grand parent (today see, yesterday what?)
② subset => set Removed (in Condition)
independent condition removed
Predict
③ set theory: set & Complement (or all subset) Divide & Conquer

Prob = weight
① Set outcome (Complete)
② Freq
③ Area Uan
④ multiplicity

info = Variance

Rand: ① all permut ② naturalizer

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

$f(y_1, y_2, \dots y_n | \beta, \sigma^2) = \prod f(y_i|\beta,\sigma^2) = \left(\frac{1}{\sigma^2}\right)^{-n/2} e^{\frac{-1}{2\sigma^2}\sum_{i=1}^n (y_i - x_i\beta)^2}$

$= \left(\frac{1}{\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)}$

Conjugate, for not Remove = Semi Conj

$p(\beta,\sigma^2) = p(\beta|\sigma^2)p(\sigma^2) \to IG\left(\frac{\alpha_0}{2}, \frac{\delta_0}{2}\right) \to N(\beta_0, \sigma^2 B_0)$

$p(\beta,\sigma^2|y) \propto p(y|\beta,\sigma^2)p(\beta,\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)}$

$\times \left(\frac{1}{\sigma^2 |B_0|}\right)^{k/2} e^{\frac{-1}{2\sigma^2}(\beta-\beta_0)'B_0^{-1}(\beta-\beta_0)} \times \left(\frac{1}{\sigma^2}\right)^{\frac{\alpha_0}{2}+1} e^{-\left(\frac{\delta_0}{2\sigma^2}\right)}$

$\alpha_1 = \alpha_0 + n + k, \quad \delta_1 = (y-X\beta)'(y-X\beta) + (\beta-\beta_0)'B_0^{-1}(\beta-\beta_0) + \delta_0$

$p(\sigma^2|\beta,y) = IG\left(\frac{\alpha_1}{2}, \frac{\delta_1}{2}\right) \qquad B_1^{-1} = \frac{X'X + B_0^{-1}}{\sigma^2}$

$\beta \propto e^{-\frac{1}{2\sigma^2}}[\beta'(X'X+B_0^{-1})\beta - 2(Y'X+\beta_0'B_0^{-1})\beta] \qquad \mu = v'(Y'X + \beta_0 B_0^{-1})'$
$= (X'X+B_0^{-1})^{-1}(X'Y + B_0^{-1}\beta_0)$

$BF = \frac{P(y|H_2)}{P(y|H_1)} = \frac{\int p(y|H_2,\theta_2)p(\theta_2|H_2)d\theta_2}{\int p(y|\theta_1,H_1)p(\theta_1|H_1)d\theta_1} \Rightarrow \frac{P(H_2|y)}{P(H_1|y)} = \frac{P(H_2)}{P(H_1)} \cdot BF(H_2,H_1)$
$\hookrightarrow$ MLE
$\hookrightarrow$ dist, mean, var, mmnt

Conjugate Prior: $p(y_i|\theta) = f(y_i)g(\theta)e^{\phi(\theta)'u(y_i)}$
natural parameter

Jeffry: $p(\theta) \propto [J(\theta)]^{1/2} \to$ Fisher info Jacobian

$J(\theta) = E\left[\left(\frac{\partial \log p(y|\theta)}{\partial\theta}\right)^2 \Big|\theta\right] = -E\left[\frac{\partial^2 \log p(y|\theta)}{\partial\theta^2}\Big|\theta\right]$

$p(y_t|\beta) = \left(\frac{e^{x_t\beta}}{1+e^{x_t\beta}}\right)^{y_t}\left(\frac{1}{1+e^{x_t\beta}}\right)^{1-y_t}, \quad p(\beta|y) \propto p(y|\beta) \qquad y_t = \begin{cases}0\\1\end{cases}$

$\log(p(\beta|\beta)) = \sum_{t=1}^N y_t x_t\beta - \log(1+e^{x_t\beta})$

$p(y|\sigma^{-2}) \Rightarrow$ Likelihood $\qquad v = F^{-1}(U) \qquad TN = p(x|a<x<b)$
$x \sim N(\mu,\sigma^2) \quad \Phi = N(0,1)$

$p(\tilde{y}|y) = \iint p(\tilde{y},\beta,\sigma^2|y)d\beta d\sigma^2 = \iint p(\tilde{y}|\beta,\sigma^2,y)p(\beta,\sigma^2|y)d\beta d\sigma^2$

$= \iint p(\tilde{y}|\beta,\sigma^2)p(\beta,\sigma^2|y)d\beta d\sigma^2$

$1 \to 3$; $3 \to 10$; $10 \to 30$; $30 \to 100$; 100
Barely worth; Substantial; strong; Very strong; Decisive

H_1: Could be just $p(y)$ assuming $\theta = v$ Freq $(p(\theta = v) = 1)$
H_2: Could be weighted Avg $p(y)$ on each θ Probability
$p(y)$ is likelihood MLE

$p(y) = \frac{p(y|\theta)\cdot p(\theta)}{p(\theta|y)} \qquad \ln \hat{p}(y) = \ln p(y|\theta^*) + \ln p(\theta^*) - \ln \hat{p}(\theta^*|y)$

① Reversability ② apperiodicity

$p(\theta_1^*,\theta_2^*|y) = p(\theta_1^*|y)\cdot p(\theta_2^*|y,\theta_1^*)$

$p(\theta_1^*|y) = \int p(\theta_1^*|y,\theta_2)\cdot p(\theta_2|y)d\theta_2 \Rightarrow \hat{p}(\theta_1^*|y) = G^{-1}\sum^G p(\theta_1^*|y,\theta_2^{(g)})$

$\hat{p}(\theta_r|y,\theta_1^*,\dots,\theta_{r-1}^*) = G^{-1}\sum^G p(\theta_r^*|y,\theta_1^*,\dots\theta_{r-1}^*,\theta_{r+1}^{(g)},\dots,\theta_k^{(g)})$

$p(Y_i = 1|\beta) = H(x_i\beta) \quad i = 1,\dots,N$

$p(\beta,Z|y) \propto p(\beta)\prod^N \{1(Z_i>0)1(y_i=1) + 1(Z_i\le 0)1(y_i=0)\} \times N(Z_i|x_i\beta,1)$

$p(\beta|Z,Y) \propto \prod^N N(Z_i|x_i\beta,1) \propto N_k(\beta|(X'X)^{-1}(X'Z)(X'X)^{-1})$

$\alpha(x,y) = \begin{cases}\min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right) & \text{if } \pi(x)q(x,y) > 0\\ 0 & \text{otherwise}\end{cases}$

$\pi^*_{i|-i}(dy_i|x_{-i}) = \int p_i(x_i, dy_i|x_{-i})\pi_{i|-i}(x_i|x_{-i})dx_i$ Meisam Hejazi

$\hat{p}(\theta_r^*|y,\theta_1^*,\dots,\theta_{r-1}^*) + G^{-1}\sum_{g=1}^G p(\theta_r^*|y,\theta_1^*,\dots,\theta_{r-1}^*,\theta_{r+1}^{(g)},\dots,\theta_k^{(g)})$

$p(\theta'|y) = \frac{\int p(\theta'|\theta,y)\cdot p(\theta|y)d\theta}{\int p(\theta|\theta',y)d\theta} = \frac{\int \alpha(\theta'|\theta,y)\cdot q(\theta'|\theta,y)\cdot p(\theta|y)d\theta}{\int \alpha(\theta'|\theta,y)q(\theta|\theta',y)d\theta}$

$\Rightarrow \hat{p}(\theta^*|y) = \frac{G^{-1}\sum_{g=1}^G \alpha(\theta^*|\theta^{(g)},y)q(\theta^*|\theta^{(g)},y)}{J^{-1}\sum_{j=1}^J \alpha(\theta^{(j)}|\theta^*,y)}$

$\alpha(\theta_1|y,\theta_1,\theta_2,Z) = \min\left\{\frac{p(\theta_1'|y,\theta_2,Z)q(\theta_1|y,\theta_1',\theta_2,Z)}{p(\theta_1|y,\theta_2,Z)\cdot q(\theta_1'|y,\theta_1,\theta_2,Z)}, 1\right\}$

① Find likelihood (Condition on param, mean, var)
② use prior (Connect by var)
③ Calc posterior (① Refactor, Change var ② Transfer Jacobian ③ eye)
(Guess pattern match, IG test)
Ⓘ pinpoint Ⓘ find Ⓘ match

student t: $\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\left(1+\frac{x^2}{\nu}\right)^{\frac{-(\nu+1)}{2}}$ parameters of dist config. on iterations eg. $\Omega \Rightarrow \Omega \Rightarrow L$

student t: $f(y_i|\beta,\sigma^2,\lambda_i) = N(x_i'\beta,\lambda_i^{-1}\sigma^2), \quad \lambda_i \sim G(\nu/2,\nu/2)$
Reg $\quad \beta \sim N_k(\beta_0,B_0), \quad \sigma^2 \sim IG(\alpha_0/2, \delta_0/2)$

tobit: $y_i = \begin{cases}x_i'\beta + u_i & \text{if } x_i'\beta + u_i \\ 0 & \text{otherwise}\end{cases} \qquad y_i = y_i^* 1(x_i'\beta + u_i > 0)$
observed latent

$f(y|\beta,\sigma^2) = \prod_{i\in C}\Phi(-x_i'\beta/\sigma)\prod_{i\in C^c}\phi([y_i - x_i'\beta]/\sigma) \qquad \beta \sim N(\beta_0,B_0)$
$\sigma^2 \sim IG(\alpha_0/2,\delta_0/2)$

$f(y_i|\beta,\sigma^2,y_{Ci}^*) = [1(y_i=0)1(y_{Ci}^*\le 0) + 1(y_i>0)]N(y_i|x_i'\beta,\sigma^2)$
$= x_i\beta + u_i$ latent

$\pi(\beta,\sigma^2,y_C^*) = \pi(y_C^*|\beta,\sigma^2)\pi(\beta)\pi(\sigma^2) \qquad y_i^* \begin{cases}y_{Ci} & \text{if } i\in C \text{ i: } y_i = 0\\ y_i & \text{if } i\in C^c\end{cases}$

$\pi(y_C^*|\beta,\sigma^2) = \prod_{i\in C} N(y_{Ci}^*|x_i'\beta,\sigma^2)$

$\pi(\beta,\sigma^2,y_C^*|y) \propto \prod_i^n [1(y_i=0)1(y_{Ci}^*\le 0) + 1(y_i>0)] \times (\sigma^2)^{-\frac{n}{2}} \times$
$e^{-\frac{1}{2\sigma^2}(y^*-X\beta)'(y^*-X\beta)} \times e^{-\frac{1}{2}(\beta-\beta_0)'B_0^{-1}(\beta-\beta_0)} \times \left(\frac{1}{\sigma^2}\right)^{\frac{\alpha_0}{2}-1} e^{\frac{-\delta_0}{2\sigma^2}}$

$\pi(y_{Ci}^*|\beta,\sigma^2) = N(y_i^*|x_i\beta,\sigma^2)$

⊛ Draw $y_i^{*(g)}$ for $i \in C$ from $TN_{(-\infty,0]}(x_i'\beta^{(g)},\sigma^{2(g)})$
$y_i^* = x_i'\beta + u_i \quad u_i \sim N(0,\sigma^2)$

student t binary probit: $y_i^* = x_i'\beta + u_i \quad u_i \sim N(0,\lambda_i^{-1}) \quad \lambda_i \sim G(\nu/2,\nu/2)$
$E(y_i) = G(x_i'\beta)$ link func $p(y_i = 1) = \Phi(x_i\beta)$
$y_i^* = x_i'\beta + u_i \quad u_i \sim N(0,1)$
$y_i \begin{cases}0 & \text{if } y_i^* < 0\\ 1 & \text{otherwise}\end{cases} \qquad y_i = 1(y_i^* > 0)$
$u_i \sim N(0,1)$ for $N(0,\sigma^2)$

$p(y_i|y_i^*) = 1(y_i=0)1(y_i^*\le 0) + 1(y_i=1)1(y_i^*>0)) \times N_k(y^*|X\beta, I)$

$\pi(\beta,y^*|y) = \prod[1(y_i=0)1(y_i^*<0) + 1(y_i=1)1(y_i^*>0)) \times N_k(\beta|\beta_0,B_0)$

Meisan

Table A.1 Continuous distributions

Distribution	Notation	Parameters
Uniform	$\theta \sim \mathrm{U}(\alpha, \beta)$ $p(\theta) = \mathrm{U}(\theta\|\alpha, \beta)$	boundaries α, β with $\beta > \alpha$
Normal	$\theta \sim \mathrm{N}(\mu, \sigma^2)$ $p(\theta) = \mathrm{N}(\theta\|\mu, \sigma^2)$	location μ scale $\sigma > 0$
Multivariate normal	$\theta \sim \mathrm{N}(\mu, \Sigma)$ $p(\theta) = \mathrm{N}(\theta\|\mu, \Sigma)$ (implicit dimension d)	symmetric, pos. definite, $d \times d$ variance matrix Σ
Gamma	$\theta \sim \mathrm{Gamma}(\alpha, \beta)$ $p(\theta) = \mathrm{Gamma}(\theta\|\alpha, \beta)$	shape $\alpha > 0$ inverse scale $\beta > 0$
Inverse-gamma	$\theta \sim \text{Inv-gamma}(\alpha, \beta)$ $p(\theta) = \text{Inv-gamma}(\theta\|\alpha, \beta)$	shape $\alpha > 0$ scale $\beta > 0$
Chi-square	$\theta \sim \chi^2_\nu$ $p(\theta) = \chi^2_\nu(\theta)$	degrees of freedom $\nu > 0$
Inverse-chi-square	$\theta \sim \text{Inv-}\chi^2_\nu$ $p(\theta) = \text{Inv-}\chi^2_\nu(\theta)$	degrees of freedom $\nu > 0$
Scaled inverse-chi-square	$\theta \sim \text{Inv-}\chi^2(\nu, s^2)$ $p(\theta) = \text{Inv-}\chi^2(\theta\|\nu, s^2)$	degrees of freedom $\nu > 0$ scale $s > 0$
Exponential	$\theta \sim \mathrm{Expon}(\beta)$ $p(\theta) = \mathrm{Expon}(\theta\|\beta)$	inverse scale $\beta > 0$
Wishart	$W \sim \mathrm{Wishart}_\nu(S)$ $p(W) = \mathrm{Wishart}_\nu(W\|S)$ (implicit dimension $k \times k$)	degrees of freedom ν symmetric, pos. definite $k \times k$ scale matrix S
Inverse-Wishart	$W \sim \text{Inv-Wishart}_\nu(S^{-1})$ $p(W) = \text{Inv-Wishart}_\nu(W\|S^{-1})$ (implicit dimension $k \times k$)	degrees of freedom ν symmetric, pos. definite $k \times k$ scale matrix S

Density function	Mean, variance, and mode
$p(\theta) = \frac{1}{\beta-\alpha},\ \ \theta \in [\alpha, \beta]$	$\mathrm{E}(\theta) = \frac{\alpha+\beta}{2}$, $\mathrm{var}(\theta) = \frac{(\beta-\alpha)^2}{12}$ no mode
$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(\theta-\mu)^2\right)$	$\mathrm{E}(\theta) = \mu$, $\mathrm{var}(\theta) = \sigma^2$ $\mathrm{mode}(\theta) = \mu$
$p(\theta) = (2\pi)^{-d/2}\lvert\Sigma\rvert^{-1/2}$ $\times \exp\left(-\frac{1}{2}(\theta-\mu)^T\Sigma^{-1}(\theta-\mu)\right)$	$\mathrm{E}(\theta) = \mu$, $\mathrm{var}(\theta) = \Sigma$ $\mathrm{mode}(\theta) = \mu$
$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta},\ \ \theta > 0$	$\mathrm{E}(\theta) = \frac{\alpha}{\beta}$ $\mathrm{var}(\theta) = \frac{\alpha}{\beta^2}$ $\mathrm{mode}(\theta) = \frac{\alpha-1}{\beta}$, for $\alpha \geq 1$
$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{-(\alpha+1)}e^{-\beta/\theta},\ \ \theta > 0$	$\mathrm{E}(\theta) = \frac{\beta}{\alpha-1}$, for $\alpha > 1$ $\mathrm{var}(\theta) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \alpha > 2$ $\mathrm{mode}(\theta) = \frac{\beta}{\alpha+1}$
$p(\theta) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)}\theta^{\nu/2-1}e^{-\theta/2},\ \ \theta > 0$ same as Gamma($\alpha = \frac{\nu}{2}, \beta = \frac{1}{2}$)	$\mathrm{E}(\theta) = \nu$, $\mathrm{var}(\theta) = 2\nu$ $\mathrm{mode}(\theta) = \nu - 2$, for $\nu \geq 2$
$p(\theta) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)}\theta^{-(\nu/2+1)}e^{-1/(2\theta)},\ \ \theta > 0$ same as Inv-gamma($\alpha = \frac{\nu}{2}, \beta = \frac{1}{2}$)	$\mathrm{E}(\theta) = \frac{1}{\nu-2}$, for $\nu > 2$ $\mathrm{var}(\theta) = \frac{2}{(\nu-2)^2(\nu-4)}, \nu > 4$ $\mathrm{mode}(\theta) = \frac{1}{\nu+2}$
$p(\theta) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} s^\nu \theta^{-(\nu/2+1)}e^{-\nu s^2/(2\theta)},\ \ \theta > 0$ same as Inv-gamma($\alpha = \frac{\nu}{2}, \beta = \frac{\nu}{2}s^2$)	$\mathrm{E}(\theta) = \frac{\nu}{\nu-2}s^2$ $\mathrm{var}(\theta) = \frac{2\nu^2}{(\nu-2)^2(\nu-4)}s^4$ $\mathrm{mode}(\theta) = \frac{\nu}{\nu+2}s^2$
$p(\theta) = \beta e^{-\beta\theta},\ \ \theta > 0$ same as Gamma($\alpha = 1, \beta$)	$\mathrm{E}(\theta) = \frac{1}{\beta}$, $\mathrm{var}(\theta) = \frac{1}{\beta^2}$ $\mathrm{mode}(\theta) = 0$
$p(W) = \left(2^{\nu k/2}\pi^{k(k-1)/4}\prod_{i=1}^k \Gamma\left(\frac{\nu+1-i}{2}\right)\right)^{-1}$ $\times \lvert S\rvert^{-\nu/2}\lvert W\rvert^{(\nu-k-1)/2}$ $\times \exp\left(-\frac{1}{2}\mathrm{tr}(S^{-1}W)\right)$, W pos. definite	$\mathrm{E}(W) = \nu S$
$p(W) = \left(2^{\nu k/2}\pi^{k(k-1)/4}\prod_{i=1}^k \Gamma\left(\frac{\nu+1-i}{2}\right)\right)^{-1}$ $\times \lvert S\rvert^{\nu/2}\lvert W\rvert^{-(\nu+k+1)/2}$ $\times \exp\left(-\frac{1}{2}\mathrm{tr}(SW^{-1})\right)$, W pos. definite	$\mathrm{E}(W) = (\nu-k-1)^{-1}S$

Table A.1 **Continuous distributions** *continued*

Distribution	Notation	Parameters
Student-t	$\theta \sim t_\nu(\mu, \sigma^2)$ $p(\theta) = t_\nu(\theta\|\mu, \sigma^2)$ t_ν is short for $t_\nu(0,1)$	degrees of freedom $\nu > 0$ location μ scale $\sigma > 0$
Multivariate Student-t	$\theta \sim t_\nu(\mu, \Sigma)$ $p(\theta) = t_\nu(\theta\|\mu, \Sigma)$ (implicit dimension d)	degrees of freedom $\nu > 0$ location $\mu = (\mu_1, \ldots, \mu_d)$ symmetric, pos. definite $d \times d$ scale matrix Σ
Beta	$\theta \sim \text{Beta}(\alpha, \beta)$ $p(\theta) = \text{Beta}(\theta\|\alpha, \beta)$	'prior sample sizes' $\alpha > 0, \beta > 0$
Dirichlet	$\theta \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_k)$ $p(\theta) = \text{Dirichlet}(\theta\|\alpha_1, \ldots, \alpha_k)$	'prior sample sizes' $\alpha_j > 0$; $\alpha_0 \equiv \sum_{j=1}^k \alpha_j$

Table A.2 **Discrete distributions**

Distribution	Notation	Parameters
Poisson	$\theta \sim \text{Poisson}(\lambda)$ $p(\theta) = \text{Poisson}(\theta\|\lambda)$	'rate' $\lambda > 0$
Binomial	$\theta \sim \text{Bin}(n, p)$ $p(\theta) = \text{Bin}(\theta\|n, p)$	'sample size' n (positive integer) 'probability' $p \in [0, 1]$
Multinomial	$\theta \sim \text{Multin}(n; p_1, \ldots, p_k)$ $p(\theta) = \text{Multin}(\theta\|n; p_1, \ldots, p_k)$	'sample size' n (positive integer) 'probabilities' $p_j \in [0, 1]$; $\sum_{j=1}^k p_j = 1$
Negative binomial	$\theta \sim \text{Neg-bin}(\alpha, \beta)$ $p(\theta) = \text{Neg-bin}(\theta\|\alpha, \beta)$	shape $\alpha > 0$ inverse scale $\beta > 0$
Beta-binomial	$\theta \sim \text{Beta-bin}(n, \alpha, \beta)$ $p(\theta) = \text{Beta-bin}(\theta\|n, \alpha, \beta)$	'sample size' n (positive integer) 'prior sample sizes' $\alpha > 0, \beta > 0$

Density function	Mean, variance, and mode
$p(\theta) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi}\sigma}(1 + \frac{1}{\nu}(\frac{\theta-\mu}{\sigma})^2)^{-(\nu+1)/2}$	$\mathrm{E}(\theta) = \mu$, for $\nu > 1$ $\mathrm{var}(\theta) = \frac{\nu}{\nu-2}\sigma^2$, for $\nu > 2$ $\mathrm{mode}(\theta) = \mu$
$p(\theta) = \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}}\|\Sigma\|^{-1/2}$ $\times(1 + \frac{1}{\nu}(\theta-\mu)^T\Sigma^{-1}(\theta-\mu))^{-(\nu+d)/2}$	$\mathrm{E}(\theta) = \mu$, for $\nu > 1$ $\mathrm{var}(\theta) = \frac{\nu}{\nu-2}\Sigma$, for $\nu > 2$ $\mathrm{mode}(\theta) = \mu$
$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$ $\theta \in [0,1]$	$\mathrm{E}(\theta) = \frac{\alpha}{\alpha+\beta}$ $\mathrm{var}(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ $\mathrm{mode}(\theta) = \frac{\alpha-1}{\alpha+\beta-2}$
$p(\theta) = \frac{\Gamma(\alpha_1+\cdots+\alpha_k)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)}\theta_1^{\alpha_1-1}\cdots\theta_k^{\alpha_k-1}$ $\theta_1,\ldots,\theta_k \geq 0; \sum_{j=1}^k \theta_j = 1$	$\mathrm{E}(\theta_j) = \frac{\alpha_j}{\alpha_0}$ $\mathrm{var}(\theta_j) = \frac{\alpha_j(\alpha_0-\alpha_j)}{\alpha_0^2(\alpha_0+1)}$ $\mathrm{cov}(\theta_i,\theta_j) = -\frac{\alpha_i\alpha_j}{\alpha_0^2(\alpha_0+1)}$ $\mathrm{mode}(\theta_j) = \frac{\alpha_j-1}{\alpha_0-k}$

Density function	Mean, variance, and mode
$p(\theta) = \frac{1}{\theta!}\lambda^\theta \exp(-\lambda)$ $\theta = 0, 1, 2, \ldots$	$\mathrm{E}(\theta) = \lambda$, $\mathrm{var}(\theta) = \lambda$ $\mathrm{mode}(\theta) = \lfloor\lambda\rfloor$
$p(\theta) = \binom{n}{\theta}p^\theta(1-p)^{n-\theta}$ $\theta = 0, 1, 2, \ldots, n$	$\mathrm{E}(\theta) = np$ $\mathrm{var}(\theta) = np(1-p)$ $\mathrm{mode}(\theta) = \lfloor(n+1)p\rfloor$
$p(\theta) = \binom{n}{\theta_1\ \theta_2 \cdots \theta_k}p_1^{\theta_1}\cdots p_k^{\theta_k}$ $\theta_j = 0, 1, 2, \ldots, n; \sum_{j=1}^k \theta_j = n$	$\mathrm{E}(\theta_j) = np_j$ $\mathrm{var}(\theta_j) = np_j(1-p_j)$ $\mathrm{cov}(\theta_i,\theta_j) = -np_ip_j$
$p(\theta) = \binom{\theta+\alpha-1}{\alpha-1}\left(\frac{\beta}{\beta+1}\right)^\alpha\left(\frac{1}{\beta+1}\right)^\theta$ $\theta = 0, 1, 2, \ldots$	$\mathrm{E}(\theta) = \frac{\alpha}{\beta}$ $\mathrm{var}(\theta) = \frac{\alpha}{\beta^2}(\beta+1)$
$p(\theta) = \frac{\Gamma(n+1)}{\Gamma(\theta+1)\Gamma(n-\theta+1)}\frac{\Gamma(\alpha+\theta)\Gamma(n+\beta-\theta)}{\Gamma(\alpha+\beta+n)}$ $\times\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$, $\theta = 0, 1, 2, \ldots, n$	$\mathrm{E}(\theta) = n\frac{\alpha}{\alpha+\beta}$ $\mathrm{var}(\theta) = n\frac{\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Bayesian — 1-April — 1

- Slides are up
- Exam

$y_{it} = X_{it}\beta_i + \varepsilon_{it}$

$i = 1, \dots, N$

$t = 1, \dots, M$

discrete Case

$s = 1, \dots, S$

$\beta_s \rightarrow$ for segments

Choice model $\sum_{s=1}^{S} P_s P(y|\theta_s)$

Segment probability

BIC $\rightarrow$ Bad criteria, since has preference for smaller number of segments

- sensible result is important

$\beta_i \sim N(\beta, D)$

Correl b/w sensitivity

Bayesian $\longrightarrow$ mixture Normal $\longrightarrow$ two kind of beta

On Structural model it is used

~~Continuous heterogeneity~~

$N \gg M$ panel

$T = M \gg N$ SUR

$M \simeq N \gg 0$ click stream data

- if you have for each person thousand, hundreds per person, you just run that, and don't need heterogeneity

Continuous heterogeneity

$$\underbrace{y_{ij}}_{1\times1} = \underbrace{X_{ij}}_{1\times k}\underbrace{\beta}_{k\times1} + \underbrace{w_{ij}}_{1\times q}\underbrace{\beta_i}_{q\times1} + \underbrace{\varepsilon_{ij}}_{1\times1}$$

$\varepsilon_{ij} \sim N(0, \sigma^2)$

$P(y_{ij} | \beta, \{\beta_i\}, \sigma^2, D)$

w has everything X has $w \subset X$

$\beta_i \sim N_q(0, D)$

$\hookrightarrow$ centered on zero since non zero part is spited for X_{it}

Likelihood

$$P(y_{ij} | \beta, \{\beta_i\}, \sigma^2, D) = \prod_{j=1}^{M} N(y_{ij} | x_{ij}\beta + w_{ij}\beta_i, \sigma^2)$$

$[M\times1]$

$N(\beta_i, 0, D^2)$

$|D|^{-1/2} e^{-1/2 \beta_i' D^{-1} \beta_i}$

$$= c\prod_{i=1}^{N} |D|^{-1/2} e^{-1/2 \beta_i' D^{-1} \beta_i}$$

Gibbs Sampler

Prior

$D \sim \text{wishart}_{\nu_0}(\Omega_0)$

$\beta \sim N(\beta_0, B_0)$

$\sigma^2 \sim \Gamma(\frac{\alpha_0}{2}, \frac{S_0}{2})$

2

D^{-1}

$\underset{k\times k}{W} \sim \text{wish}_{\nu}(S) \rightarrow \beta_i'\beta_i$

$\propto |W|^{\frac{\nu - k - 1}{2}} e^{-1/2 \operatorname{tr}(S^{-1} W)}$

Posterior

$$D^{-1} | y, \beta, \{\beta_i\}, \sigma^2 \sim |D^{-1}|^{\frac{\nu_0 - q - 1}{2}} e^{-1/2 \operatorname{tr}(\Omega_0^{-1} D^{-1})} \times |D^{-1}|^{N/2} e^{-1/2 \operatorname{tr}[\sum^{N} \beta_i'\beta_i) D^{-1}]}$$

(*) indep of σ^2 and β, so just $N(\beta_i, D)$

$$\propto \text{wishart}_{N+\nu_0}[(\Omega^{-1} + \sum_{i=1}^{N} \beta_i'\beta_i)^{-1}]$$

(**)

$\beta | y, \{\beta_i\}, \sigma^2, D \sim N(VA, V)$

- You can take w_{ij} on the other side $\longrightarrow$ Convert to liner regression

$y_{ij} - w_{ij}\beta_i = X_{ij}\beta + \varepsilon_{ij} = y_{ij}^*$

then indep D (**)

$$\beta \sim N\left((X'X + B_0)^{-1}\left(\frac{X'Y_i^*}{\sigma^2} + B_0^{-1}\beta_0\right), (*)\right)$$

$$\begin{array}{c} x_{i1} \dots x_{ik} \\ \begin{bmatrix} x_{11} \\ \vdots \\ x_{12} \\ x_{13} \\ x_{23} \\ \vdots \end{bmatrix} \end{array}$$

$[NM \times K]$

$\downarrow$

# of Covariates

$V = \left(\frac{X'X}{\sigma^2} + B_0^{-1}\right)^{-1}$

$A = \left(\frac{\sum_i \sum_j x_{ij}'(y_{ij} - w_{ij}\beta_i)}{\sigma^2} + B_0^{-1}\beta_0\right)$

$\sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}' x_{ij}$

$$\sigma^{-2} | y, \beta, \{\beta_i\}, \cancel{D} \sim \Gamma\left(\frac{\alpha_0 + NM}{2}, \frac{S_0 + (y^* - X\beta)'(y^* - X\beta)}{2}\right)$$

falls out since once you have it it will fall out

$$y^* = \begin{bmatrix} y_{11} - w_{11}\beta_1 \\ y_{12} - w_{12}\beta_1 \\ \vdots \\ y_{21} - w_{21}\beta_2 \\ \vdots \end{bmatrix}$$

$[NM \times 1]$ matrix

another hierarchy could be

$\beta_i \sim f(z)$

$\beta_i = d_i + z_i \beta_i$ } $\longrightarrow$ Convert to form of interaction

update β_i:

appears somehow in all the equations

$W_i = w_{ij}\beta_i$

$X_i = x_{ij}\beta$

$y_{ij} = Y$

$\Rightarrow p(y_i \mid \beta, \beta_i, \sigma^2, D) = N_m(y_i \mid X_i\beta + w_i\beta_i, \underset{M\times M}{I} \otimes \underset{1\times 1}{\sigma^2}) \cdot N_q(\beta_i \mid 0, D)$

$\Rightarrow \cancel{|\sigma^2|^{-\frac{M}{2}}}\ e^{-\frac{1}{2}(y_i - X_i\beta - w_i\beta_i)(I\otimes\sigma^{-2})(Y - X_i\beta - w_i\beta_i) - \frac{1}{2}\beta_i' D^{-1}\beta_i}$

don't care about $\cancel{\beta_i}$, since just care about β_i

$e^{-\frac{1}{2}(\beta_i'(D^{-1} + w_i'(I\otimes\sigma^{-2})w_i)\beta_i - 2[(y_i - X_i\beta)'(I\otimes\sigma^{-2})w_i]\beta_i}$

$\Rightarrow \beta_i \mid y, \beta, \sigma^2, \rho^2 \sim N(VA, V)$

$V = (w_i'(I\otimes\sigma^{-2})w_i + D^{-1})^{-1})$

$A = (w_i'(I\otimes\sigma^{-2})(y_i - X_i\beta)$

- β would be bias if you have not put this
- numerical integration and data augmentation will work here

SUR

① Journal of marketing → Good problem, method not important

② JMS, JMR → have complex enough Fancy & Realistic model

- in marketing assume it is indogenous
- Economics competition is huge. Good outlet
- psychology → JCB, other method

SUR

multinomial multivariate problems will be discussed next time

$y_{ij} = X_{ij}\beta_i + \varepsilon_{ij}$ → correlation of error term

$\varepsilon_{ij} \sim N(0, \Sigma)$

$i = 1 \ldots N \quad j = 1, \ldots, M$

$M >> N$ a lot of observation from individual

2 years of suppermarket for example

- Categories correlated with each other. Could be firm financial metric
- long time series, not lot of observation on the individual (other dimension)

$\underset{c\times 1}{y_t} = \underset{c\times K}{x_t}\underset{K\times 1}{\beta} + \underset{c\times 1}{\varepsilon_t}$ $\quad c = 1, \ldots, C$, $\quad t = 1, \ldots, T$

$\beta = \begin{pmatrix} \beta_1 \to k_1\times 1 \\ \beta_2 \to k_2\times 2 \\ \vdots \\ \beta_c \to k_c\times 1 \end{pmatrix}$ $\quad y_t \underset{\text{staching}}{=} \begin{pmatrix} y_{t1} \\ y_{t2} \\ \vdots \\ y_{tc} \end{pmatrix}$ $\quad K = \sum_{c=1}^{C} k_c$

$\varepsilon_t \sim N_c(0, \Sigma)$

$y_{ct} = X_{ct}\beta_c + \varepsilon_{ct}$ $\quad T >> C$

$\varepsilon_t \sim N(0, \Sigma)$

$X_t = \begin{pmatrix} x_{t1} & 0 & 0 & 0 \\ \vdots & x_{t2} & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \ldots & x_{tc} \end{pmatrix}$ Sur form of matrix

$\beta \mid Y, \sigma^2 = ?$

$\sigma^2 \mid y, \beta = ?$

---

prior

$\beta = N_K(\bar{\beta}, B_0)$

$\Sigma^{-1} = \text{wish}_{\nu_0}(S_0)$

$p(y \mid \beta, \Sigma) = \prod_{t=1}^{T} N(y_t \mid X_t\beta, \Sigma)$

$\Sigma^{-1} \mid y, \beta \propto |\Sigma^{-1}|^{\frac{\nu_0 - c - 1 + T}{2}} (-\frac{1}{2}\operatorname{tr}(S_0^{-1} + \sum^{T}(y_t - X_t\beta)(y_t - X_t\beta)^{T})\Sigma^{-1}$

$\rightarrow \frac{1}{2}\operatorname{tr}(\Sigma ee' \Sigma^{-1})$

5

$$p(\beta \mid Y, \Sigma) = e^{-1/2 \{ \beta'(B_0^{-1} + \Sigma x_t' \Sigma^{-1} x_t)\beta - 2\beta'(B_0^{-1}\beta_0 + \Sigma x_t' \Sigma^{-1} y_t)\}}$$

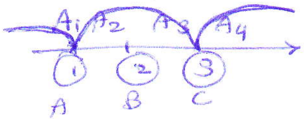
→ panel structure binary probit marrying two models. Condition on Z

6

Bayesian 25 March

$y_i = X_i\beta + \varepsilon_i \quad \varepsilon_i \sim N(0,1) \quad i=1,\dots,n$ (1)

$y_i = \{1,2,3\}$



- ~~orderde~~ cutoff points: difference may not be similar
- The problem is censored higher or lower value
- 3 categories creates four areas
- show result force to not precise
- Assumption: I know some latent variable

$\gamma_0 = -\infty \quad \gamma_1 = 0 \qquad \gamma_2 \quad \gamma_3 = \infty$

$$P(y_i \mid x_i\beta, 1) = \int TN_{[-\infty,0]}(y_i^* \mid X_i\beta, 1)\cdot I(y_i=1) + TN_{[0,\gamma_2]}(y_i^* \mid X_i\beta,1)\, I(y_i=2) + TN_{[\gamma_2,\infty)}(y_i \mid x_i\beta, 1)\, I(y_i=3)\, dy_i$$

pdf

$$P(y_i \mid x_i\beta, 1) = \int TN_{[-\infty,0)}(y_i^* \mid x_i\beta,1)\, I(y_i=1) + TN_{[0,\gamma_2]}(y_i^* \mid x_i\beta,1)\, I(y_i=2) + TN_{[\gamma_2,\infty)}(y_i^* \mid x_i\beta,1)\, I(y_i=3)\, dy_i$$

in tobit

$$p(y_i \mid \beta, \gamma_2) = TN_{[0,\infty)}(y_i^* \mid x_i\beta, 1)\cdot I(y_i>0) + \int TN_{[-\infty,0]}(y_i \mid x_i\beta,1)\cdot I(y_i=0)$$

- MLE should be one of the parameter that you get derivative of
- two equation and two unknown that you try to solve
  - one derivative with respect to β
  - one derivative with respect to γ_2

Bayesian

$$P(y_i \mid \beta, \gamma_2^2, z_i) \propto N_{(-\infty,0)}(z_i \mid X_i\beta, 1)\cdot I(y_i=1)\cdot I(z_i<0) + N_{(0,\gamma_2)}(z_i \mid x_i\beta, 1)\cdot I(y_i=2)\, I(0<z_i<\gamma_2) + N_{(\gamma_2,\infty)}(z_i \mid x_i\beta,1)\, I(y_i=3)\cdot I\cdot I(z_i>\gamma_2)$$

number w.r.t γ_2

respect γ_2 w.r.t.

pdf

$P(\beta, \gamma_2 \mid y) = ?$ $\{z_i\}$

$\boxed{\text{assum } P(\beta,\gamma_2) \propto 1}$

$$P(\beta, \gamma_2, z_i \mid y) = ? \begin{cases} TN_{(-\infty,0)} & \text{if } y_i=1 \\ TN_{(0,\gamma_2)} & \text{if } y_i=2 \\ TN_{(\gamma_2,\infty)} & \text{if } y_i=3 \end{cases}$$

$\beta \mid \{z_i\}, \cancel{\gamma_2}\ \cancel{y} \sim N(V(x'z), V)$, $V = (x'x)^{-1}$

irrelevant

$V = x'x$

$P(\beta \mid \{z_i\}) \sim N(V(x'z), V)$, $V \to (x'x)^{-1}$

$\gamma_2 \mid \{z_i\}, y \propto \text{Uniform}\left[\max_{y_i=2} z_i, \min_{y_i=3} z_i\right]$

You take minimum and maximum over all z_i calculated

→ Proportion to constant

- if 3 category you will have two more parameter,

- multinomial probit
- multiparameter

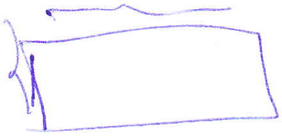
$\boxed{t\text{-student} = \dfrac{\text{Normal}}{\text{Gamma}}}$

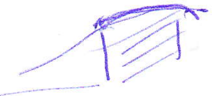
$$\begin{cases} x_{it} = y_{it} - a_i \\ x_{it-1} = y_{it-1} - a_i \end{cases} \Rightarrow$$

$L(\beta, \alpha) = \prod_{i=1} Pr(y_i = j \mid x_j)$

$L(\beta, \alpha) = \prod_{i=1}^{n} \Phi(\alpha_{y_i+1} - x_i\beta) - \underline{\Phi(\alpha_{y_i} - x_i\beta)}$

1000





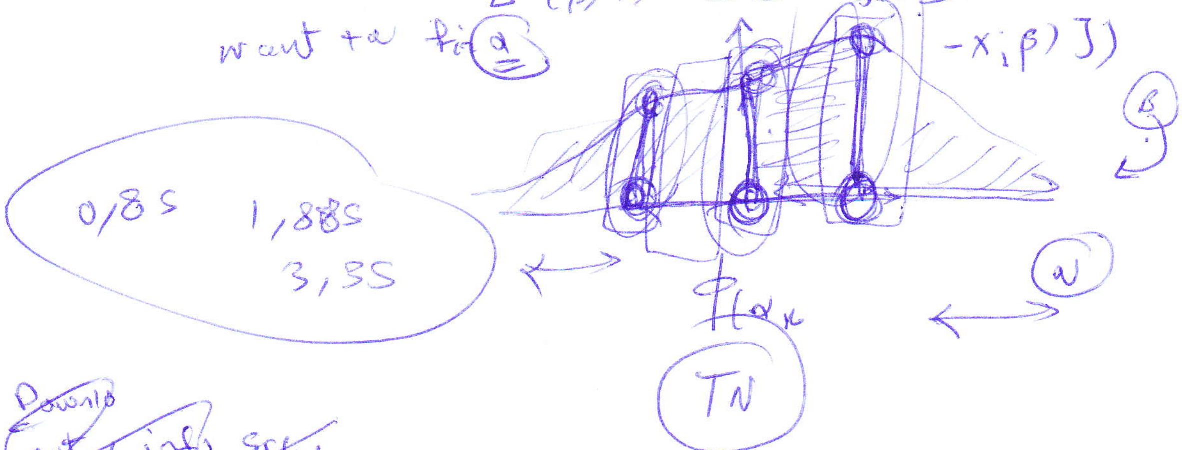
$\alpha = [\alpha_3 \quad \alpha_4 \quad \cdots \quad \alpha_J]$

$L_{\alpha_k}(\beta, \alpha) = \sum_{i: y_i = k} \frac{\Phi(\alpha_k - x_i\beta)}{p_i}$

$Z_{ij} = I(y_i = j)$

$L(\beta, \alpha) = \sum^{n} \sum^{J} z_{ij} (\log[\Phi(\alpha_{j+1} - x_i\beta) - \Phi(\alpha_j - x_i\beta)])$

want to fit α



~~Cost~~ Pownld, info src,

Response

Cont, str

Regressor

1 infor src, 2 Complmnt, 3 Alternat, 4 Experience 5 ~~A~~ lack expue

6 ~~Cost, time effort~~ usage and Experience 7 product attribute 8 Product customer interact 9 critic

10 Cost/time & Effort 11 terminology 12 Positive 13 negative

14 Comparison 15 Social network 16 Feeling intensity 17 uncertainty

18 patterns 19 Broad 20 narrow

4 average freq per day

5, 6 Stars ~ Doc Category (dc)

7 Pownld ~ viewed Category (365) (vc)

d

1212

Star1 → 23 ① ~ Doc category

Star2 → 144 ② ~ "

Star3 → 156 ③ ~ "

Star4 → 235 ④ ~ "

Star5 → 884 ⑤ ~ "

Data & definition → List categories
Model 3

all tables

① 1440 Ewes
② $
③ Cont

Terminology ← Poisson Reg ✓

// Not that much → to hit ✓

// Baysian Binary Pro ✓

First frame replace sea

Poisson Signific ✓