Bayesian

marginal likelihood M-H

① Collection of models $\{M_1, \dots, M_L\}$: M_L charac. by model specific param vector: $\theta_L \in S_L \subseteq \mathbb{R}^{k_L}$ (Dimension)

② data $y = (y_1, \dots, y_n)$

③ two model M_i, M_j comparison $\frac{Pr(M_i|y)}{Pr(M_j|y)} = \frac{Pr(M_i)}{Pr(M_j)} \times \frac{m(y|M_i)}{m(y|M_j)}$

④ $m(y|M_L) = \int f(y|M_L, \theta_L) \cdot \pi_L(\theta_L|M_L) d\theta_L$

marginal likelihood — prior odds — Bayes factor

↳ normalising constant of posterior density $m(y|M_L) = \frac{f(y|M_L, \theta)\pi(\theta_L|M_L)}{\pi(\theta_L|y, M_L)}$

$\log m(y|M_L) = \log f(y|M_L, \theta_L^*) + \log \pi(\theta_L^*|M_L) - \underbrace{\log \pi(\theta_L^*|y, M_L)}_{\text{posterior ordinate}}$

$\theta^* = (\theta_1^*, \dots, \theta_B^*) \to$ parameter space split to Blocks

$\pi(\theta^*|y) = \pi(\theta_1^*|y)\pi(\theta_2^*|y, \theta_1^*) \cdots \underbrace{\pi(\theta_B^*|y, \theta_1^*, \dots, \theta_{B-1}^*)}_{\text{Conditional density (Ordinate)}}$

$\hat{\pi}(\theta_2^*|y, \theta_1^*) = M^{-1}\sum_{j=1}^{M} \pi(\theta_2^*|y, \theta_1^*, \theta_3^{(j)}, \dots, \theta_B^{(j)})$

$\{\theta_3^{(j)}, \dots, \theta_B^{(j)}\} \sim \pi(\theta_3, \dots, \theta_B|y, \theta_1^*)$

$\pi(\theta|y) \propto \pi(\theta) \cdot f(y|\theta)$

posterior density

goal: estimate posterior ordinate $\pi(\theta^*|y)$

$q(\theta, \theta'|y)$: Candidate generating density

(transition θ to θ')

$\alpha(\theta, \theta'|y) = \min\left\{1, \frac{f(y|\theta')\pi(\theta')}{f(y|\theta)\pi(\theta)} \cdot \frac{q(\theta', \theta|y)}{q(\theta, \theta'|y)}\right\}$

Probability of accepting (probability of move)

$p(\theta, \theta'|y) = \alpha(\theta, \theta'|y) \cdot q(\theta, \theta'|y)$

Sub kernel of M-H algorithm

reversability: $p(\theta, \theta^*|y)\pi(\theta|y) = \pi(\theta^*|y) \cdot p(\theta^*, \theta|y)$

$\pi(\theta^*|y) = \frac{\int \alpha(\theta, \theta^*|y) \cdot q(\theta, \theta^*)\pi(\theta|y)dy}{\int \alpha(\theta^*, \theta|y) q(\theta^*, \theta|y) d\theta}$

$\pi(\theta^*|y) = \frac{E_1\{\alpha(\theta, \theta^*|y) q(\theta, \theta^*|y)\}}{E_2\{\alpha(\theta^*, \theta|y)\}}$

$\hat{\pi}(\theta^*|y) = \frac{M^{-1}\sum_{g=1}^{M} \alpha(\theta^{(g)}, \theta^*|y) \cdot q(\theta^{(g)}, \theta^*|y)}{J^{-1}\sum_{j=1}^{J} \alpha(\theta^*, \theta^{(j)}|y)}$

$\log \hat{m}(y) = \log f(y|\theta^*) + \log \pi(\theta^*) - \log \hat{\pi}(\theta^*|y)$

Tobit

- Censored Reg.
  1. Coded data (eg. inc > 2000 = 2000)
  2. truncated data (nor x_i, nor y_i > 2000 observed)
  3. incidentally truncated data (joint s_i, at least x_i)
  4. Corner solution data (Constraint: non neg) or Capacity

tobit:

$y_i = \begin{cases} x_i'\beta + u_i & \text{if } x_i'\beta + u_i > 0 \\ 0 & \text{otherwise} \end{cases}$

$f(y|\beta, \sigma^2) = \prod_{i \in C} \Phi(-x_i'\beta/\sigma) \prod_{i \in C^c} \phi([y_i - x_i'\beta]/\sigma)$

Since y_i has been lower than this, but since we didn't observe we just know its position is lower, but don't know where

when student t:

$B_1 = \left(\frac{X'\Lambda X}{\sigma^2} + B_0^{-1}\right)^{-1}$

$\beta_1 = B_1\left(B_0^{-1}\beta_0 + \frac{X'\Lambda Y}{\sigma^2}\right)$

$\alpha_1 = \alpha_0 + n$

$\delta_1 = \underbrace{(Y - X\beta)'\Lambda(Y - X\beta)}_{\hat{e}\hat{e}'} + \delta_0$

$\nu_1 = \nu + 1$

$\nu_{2i} = \nu + \frac{(y_i - x_i\beta)^2}{\sigma^2}$

$p(y_i|x_i\beta\sigma^2) = t_\nu(y_i|x_i\beta, \sigma^2)$

$f(y_i|\beta, \sigma^2, \lambda_i) = N(x_i\beta, \lambda_i^{-1}\sigma^2)$

$\lambda_i \sim G(\nu/2, \nu/2)$

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

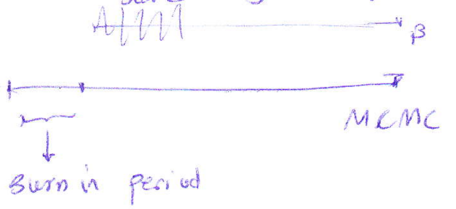
$\sigma^2 \sim IG(\alpha_0, \frac{\delta_0}{2})$

Inverse chisq → inverse Gamma

March 04 | Bayesian

1

- efficiently factor tells you about Convergence
- start two chain from different point, and make sure they end up to the same point



- Beta draw depend on Sigma and not Beta in Gibbs, but in metrapolis hasting depends on previous one
- logit has Gambel one errors, and since we used normal distribution, it did not Converge This change helped the Convergence

- Calculate Bays factor and Compare

$y \Rightarrow$ response variable $\quad x_i$ Covariates

$$p(y) = \frac{p(y|\theta)p(\theta)}{p(\theta|y)}$$

(*) we do not think about them as random var

- probit is way θ is created
- model assumption always remains

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

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

$$\hat{p}(\theta_1^*|y) = G^{-1}\sum_{g=1}^{G} p(\theta_1^*|y, \theta_1^{(g)})$$

$$p(y|\beta, \sigma^2) = \prod_{t=1}^{T} N_1(y_t | X_t\beta, \sigma^2)$$ model assumption

$$P(\beta) = N_K(\beta|\beta_0, B_0)$$ k dimensional normal density

$$P(\sigma^{-2}) = \Gamma(\sigma^{-2} | \tfrac{\alpha_0}{2}, \tfrac{\delta_0}{2})$$

$p(\beta, \sigma^2|y) = ?$

$$P(\beta|y, \sigma^2) = N_k\left(\beta \,\Big|\, \left(\frac{X'X}{\sigma^2} + B_0^{-1}\right)^{-1}\left(\frac{X'y}{\sigma^2} + B_0^{-1}\beta_0\right), \left(\frac{X'X}{\sigma^2} + B_0^{-1}\right)^{-1}\right)$$

$$P(\sigma^{-2}|y, \beta) = \Gamma\left(\frac{\alpha+n}{2}, \frac{\delta_0 + (Y - X\beta)'(Y - X\beta)}{2}\right)$$

$$P(y) = \frac{p(y|\beta, \sigma^2)P(\beta, \sigma^2)}{P(\beta, \sigma^2|y)}$$ $\Big|$ β^*, σ^{2*} Could be means, medians

II

$$P(\beta, \sigma^2|y) = p(\beta|y)P(\sigma^2|\beta, y) \longleftarrow \text{I}$$

I

$$P(\sigma^{2*}|\beta^*, y) = P\left(\sigma^{-2*} \,\Big|\, \frac{\alpha_0+n}{2}, \frac{\delta_0 + (Y - X\beta^*)'(Y - X\beta^*)}{2}\right)$$

$\overset{?}{\Rightarrow} P(\beta^*|y)$

2

$$P(\beta^*|y) = \frac{1}{G}\sum_{g=1}^{G} p(\beta^*|y, \sigma^{2*})$$ II

draws

$$\int p(\beta|y, \sigma^2)p(\sigma^2|y)\, d\sigma^2$$ Gibbs Sampling

semi-Conjugate

- here we can Condition on beta and Calculate

- save('LRdata', x, y);
- load ('LRD', x, y);
- burn in should be excluded
- Sum over all Converged draws
- Jeffry scale tells you what that means diff. should be less than 3 ($e^{\ln \text{mlty}}$)
- numerical error should be a lot less than this
- HW → Tobit → linear Reg with binary tobit

three block

2 update β update σ update

- hetrocedasticity we Could have multiple blocks Panel model Could also be translated to this
- Random walk → cancel out
- g x g Covar but still two Block Variance is wishart (matrix form of Gamma)
- Surr seemingly unrelated regression
- G and J are same thing but shows they are not the same

- after Spring Break:
  1. test ; from text book & solutions — 1 Can use last page, 2 One sheet with any thing
  2. presentation for feedback (Just say whether you don't want to prent)
     - Explain dataset & feedback & what plan to do
     - doesn't matter
3. Do all examples for 25th March submit with homework

→ afternoon friday office out this week

- to create logit error term use:

$\underbrace{\log\left(\frac{u}{1-u}\right)}$ → Normal

inverse Gambel ←

- for metrapolis Hasting: assume it converges ⟹ Markov chain should have reversability (Backward)

- Standard error as an indicator of precision

- Correlation of real y and estimated $\hat{y}$ to see precision of model

- independence ⟹ t-student

- Random walk ⟹ densities cancel out but don't in independent

Bayesian
Feb 11

$y = X\beta + u$

$u \sim N(0, \sigma^2)$

$y_i \sim N(x_i\beta, \sigma^2)$

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

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

$p(\beta, \sigma^2) = p(\beta \mid \sigma^2)\, p(\sigma^2)$

$N(\beta_0, \sigma^2 B_0)$ — $IG\left(\frac{\alpha_0}{2}, \frac{\delta_0}{2}\right)$

out for semi Conjugate

$p(\beta, \sigma^2 \mid y) \propto p(y \mid \beta, \sigma^2)\, p(\beta, \sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)}] \times \left(\frac{1}{\sigma^2 |B_0|}\right)^{k/2} e^{-\frac{1}{2\sigma^2}(\beta-\beta_0)' B_0^{-1} (\beta-\beta_0)}$

out if semiconjugate

$\times \left(\frac{1}{\sigma^2}\right)^{\frac{\alpha_0}{2}+1} e^{-\left(\frac{\delta_0}{2\sigma^2}\right)}$

out semiconjugate

for semiconjugate out

$p(\sigma^2 \mid \beta, y) = \frac{p(\beta, \sigma^2 \mid y)}{p(\beta \mid y)} \Rightarrow p(\sigma^2 \mid \beta, y) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\frac{k}{2}+\frac{\alpha_0}{2}+1} \times e^{-\frac{1}{2\sigma^2}\left[(y-X\beta)'(y-X\beta) + (\beta-\beta_0)' B_0^{-1}(\beta-\beta_0) + \delta_0\right]}$,

$\boxed{p(\sigma^2 \mid \beta, y) = IG\left(\frac{\alpha_1}{2}, \frac{\delta_1}{2}\right)}$

out for semiconjugate

$\alpha_1 = \alpha_0 + n + k$

$\alpha_1 = \alpha_0 + n + k$

$\delta_1 = (y - X\beta)'(y - X\beta) + (\beta-\beta_0)' B_0^{-1} (\beta - \beta_0) + \delta_0$

for semi Conjugate

You just need to Recognize the part that has β inside

$p(\beta \mid \sigma^2, y) \propto e^{\frac{1}{2\sigma^2}}[y'y - 2y'X\beta + \beta'X'X\beta + \beta' B_0^{-1}\beta - 2\beta_0' B_0^{-1}\beta + \beta_0' B_0 \beta_0)$ $\beta' X' Y$]

take the following pages for exam $\boxed{474 - 477}$

You need to simplify as you did things mechanically

$= e^{-\frac{1}{2\sigma^2}\left(\beta'(X'X + B_0^{-1})\beta - 2(y'X + \beta_0' B_0^{-1})\beta\right)}$

σ^2 — For semi Conjugate

σ^2 — for semi Conjugate

$\boxed{\beta \sim N(\mu, V)}$

$(\beta - \mu)' V^{-1} (\beta - \mu)$

$\beta' V^{-1} \beta - 2\mu' V^{-1}\beta + \beta' V^{-1} \mu$

Variance of Normal is Quantity, so all needs to be idempotent

$*$ shows variance $\Rightarrow V^{-1} = \frac{X'X + B_0^{-1}}{\sigma^2}$

$\mu' V^{-1} = (y'X + \beta_0' B_0^{-1}) \Rightarrow$

Only remaining term after $*$ is $(Y'X + \beta_0' B_0^{-1})\beta$

$\Rightarrow \mu' = (y'X + \beta_0' B_0^{-1}) \cdot V$

transpose

$$\boxed{\begin{aligned} \mu &= V'(Y'X + \beta_0 B_0^{-1})' = (X'X + B_0^{-1})^{-1}(X'Y + B_0^{-1}\beta_0) \\ V &= \sigma^2 (X'X + B_0^{-1})^{-1} \end{aligned}}$$

σ^2 $\quad$ σ^2 $\quad$ σ^2 $\quad$ σ^2

for semi Conjugate

(*) The only way to go forward is to pinpoint the variables and find them and match them with distribution sheet

(*) You find is different distribution for each variable in Hasting and then calculate based on

→ You know the shape and you just you find the distribution

↳ proportion

* t-Student Robust

Gamma & normal would be t-student, you need to adjust

* logit → more complicated

→ if you don't know distribution you will not know Conjugate

(1) Semi Conjugate → ~~these~~ are not dependent (independent)

↳ it seems it does not have information

(2) parameter space exponential Family → the same for posterior family

Split parameter and have the same parametric form

(*) for hyper parameter Specification → You analyze data or assume something it is say based on the scale you kind of Guess

→ Specify Priors for your self and full around to see result you → check different parameters and see the result

$H_1 \Rightarrow \theta = 1$

$H_2 \Rightarrow$ Full model

$p(y) = p(y \mid \theta=1)\,p(\theta=1) + p(y \mid \theta=0)\,p(\theta=0)$

$\frac{p(y \mid H_1)}{p(y \mid H_2)} = \frac{1/4}{5/8}$

1/4, 1/8, 5/8, 1/16

2-son $p(y)$

3-son $p(y)$

$\frac{p(y \mid \theta=1)}{p(y \mid \theta=0)} = \frac{1/4}{1} = 1/4$

$\boxed{\theta = 1}$

how adding new data changes fact on

$p(\beta \mid \sigma^2, y)$

$y_t = x_t \beta + \varepsilon_t$ $\quad p(y) = ?$

$\varepsilon_t \sim N(0, \sigma^2)$

$\theta^* = \begin{pmatrix} \beta^* \\ \sigma^{2*} \end{pmatrix}$

$p(y \mid \theta^*) = \prod_t N(y_t \mid x_t \beta, \sigma^2)$

$p(\theta^*) = N_k(\beta \mid \beta_0, B_0)\, IG(\sigma^{-2} \mid \alpha_0, \delta_0)$

$\hat{p}(\theta^* \mid y) = \hat{p}(\beta^* \mid y) \cdot \hat{p}(\sigma^{2*} \mid y, \beta^*)$

$\tilde{\beta^*, \sigma^2}$

① First

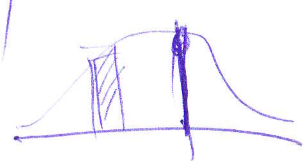
② How Bayes factor on the Binomial

①

Beta

$\binom{200}{115} (\cancel{0.5})^{115} (\cancel{0.5})^{85}$

$\theta \qquad 1-\theta$



H_0 $\boxed{\theta = 0.5}$

H_1 $\theta \neq 0.5$

$\int \longrightarrow$ Beta

Marginal likelihood

using Beta

integration of Beta

emily

① Classical likelihood 0.006

0.057

Emily

$\max_\theta p(y \mid \theta)$

$\boxed{0.575} = \frac{115}{200}$

MLE Estimate

$p = 0.5 \Rightarrow 0.006$

Marginal 0.005

$\frac{p(y \mid \theta=1)}{p(y \mid \theta=0)} \Rightarrow$

Long Term Bayesian $\Rightarrow$ Next term series or other book ..., use

$\Rightarrow$ = keep thing simple, but not simpler blogit bayesian diagram

```
n = 1000
k = 3
x = [ones(n,1) rand(n,k-1)];
beta = [1 -1 -2];
y = zeros(n,1);
Zt = zeros(n,1);
```

$\begin{bmatrix} 1 & x_1 & x_2 \cdots \\ \vdots & \cdots & \cdots \\ 1 & \cdots & \cdots \end{bmatrix}$ $\begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}$ $\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$ $\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$

for i = 1:n $1 < i < n$

$u \sim N(0,1)$

```
ei = (S2)^2 * u ;
Zti = xi * β + ei ;
if Zt > 0
   y(i) = 1
end;
```

probit

```
m = 2000;
betam = zeros(m,k);
Z0 = 2y - 1
```

β_m

t = 1:m   …. m

$\sigma_t = (x'x)^{-1}$ sigt = inv(x'x)

$M_t = \sigma_t (x'Z)$ mut = sigt * (x'Z0)

$\beta_t = M_t + \sqrt{\sigma_t}\,\varepsilon$ betat = mut + chol(sigt) x randn(k,1)

σ_t

$Z_1 = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}_n$ Z1 = zeros(n,1);

$M = X \times \beta$ Mt = x × betat;

$\begin{bmatrix} \vdots \end{bmatrix}$ N( )

i = 1:n   …. n

```
if y(i) == 0
   Zi = Mi + ur;        ~N(0,1)
   while (Zi > 0);      ~N(0,1)
      Zi = Mi + ur;
   end;
else;                   ~N(0,1)
   Zi = Mi + ur;
   while Zi <= 0;       ~N(0,1)
      Zi = Mi + ur;
   end;
end;
Z1(i) = zi
```

$\beta_m = \beta_t$ betam(t:) = betat;

β_t = ?

$Crr(Z_t Z_i)$ = ?

$\sigma_t \to \beta \to \hat\beta \xrightarrow{X\hat\beta} Z \to$ $X'Z$ A-R

logit

β_1 beta1 = zeros(k,1) $\begin{array}{c} m \\ k \begin{bmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{bmatrix} \end{array}$

β_m m = 2000;

betam = zeros(m,k);

Nothing

σ Sig = .1 × eye(k); $\begin{bmatrix} 1 & \\ 0 & \end{bmatrix}$

acr = 0

t = 1:m   …. m $N($ $)$

```
βp = β1 + col(σ) × randn(k,1);
% calc alpha:                         βp
num = lnlik(y, x, betap);           loglik
den = lnlik(y, x, beta1);           β1
lnalpha = num - den;                ln(α)   probab of accepting
u = log(u);                         ~Unif(0,1)
if (u < lnalpha);
   beta1 = betap;                   β1 = βp
   acr = acr + 1;
end;
betam(t,:) = beta1;
```

$p(y_t|\beta) = \left(\frac{e^{X_t\beta}}{1+e^{X_t\beta}}\right)^{y_t} \left(\frac{1}{1+e^{X_t\beta}}\right)^{1-y_t}$

$p(y|\beta) = \prod_t p(y_t|\beta)$ $p(\beta) \propto 1$

$p(\beta|y) \propto p(y|\beta)$

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

lnlik

$y_t \begin{cases} 0 \\ 1 \end{cases}$ Complete model & not partial

Normal multivariate Non informative prior dis

Jeffry density: $p(\mu,\Sigma) \propto |\Sigma|^{-\frac{(d+1)}{2}}$ $\begin{cases} \kappa_0 \to 0 \\ \nu_0 \to -1 \\ |\Lambda_0| \to 0 \end{cases}$

Posterior:
$\Sigma | y \sim \text{Inv-wishart}_{n-1}(S)$
$\mu | \Sigma, y \sim N(\bar{y}, \Sigma/n)$

d: # Variance Matrix params & $\frac{d(d-1)}{2}$: Covariance params

$p(y_i|\alpha,\beta,n_i,x_i) \propto [\text{logit}^{-1}(\alpha+\beta x_i)]^{y_i}[1-\text{logit}^{-1}(\alpha+\beta x_i)]^{n_i-y_i}$

x_i: ith of k dose level
n_i: animals given
y_i: Response outcome

$p(\alpha,\beta|y,n,x) \propto p(\alpha,\beta|n,x)\,p(y|\alpha,\beta,n,x) \propto p(\alpha,\beta)\prod_{i=1}^{k} p(y_i|\alpha,\beta,n_i,x_i)$

$\alpha+\beta x_i = \theta_i$: Prob. of Death

$0<\theta_i<1$
$\text{logit}(\theta_i)=\alpha+\beta x_i$

problem of too low high dose

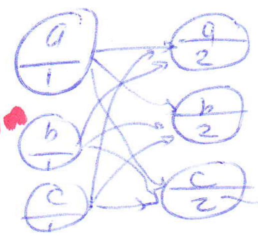
Sample size: n_i
dose level: x_i

Counter plot

Scatter plot

Sample draw

markov chain



t: time

Cholesky decomposition

Hermitian

$A=LL^*$

Conjugate transpose

Positive semidefinit

$z^T M z \leftrightarrow$ non neg

$[a\ b]\,I\begin{bmatrix}a\\b\end{bmatrix} = a^2+b^2$

Tailor expansion

$\log(\theta|y) = \log(\hat{\theta}|y) + \frac{1}{2}(\theta-\hat{\theta})^T \left[\frac{d^2}{d\theta^2}\log p(\theta|y)\right]_{\theta=\hat{\theta}}(\theta-\hat{\theta}) + \dots$

Grid of points

other def

non positive ← Neg semidefinite

positive ← Positive def

Rejection sampling

Simulation

Simulate in part

① draw from marginal posterior dist of hyperparameters

② simulate other parameters conditional on data and the hyperparameters simulated

- sequence of random iteration
- θ: Common stationary distribution

alternating conditional sampling

Gibbs Sampler

higher probability when higher area

① Sample θ at random from prob. density proportional to $g(\theta)$

② with probability $p(\theta|y)/(Mg(\theta))$ accept θ as draw from p

Markov chain

transition distribution
drawing θ^t $T_t(\theta^t|\theta^{t-1})$ → previous draw

used when direct draw not possible

Converge to unique stationary distribution

multidimension Sampling

draws $p(\theta_j|\theta_{-j}^{t-1})$ ← subvector of $\theta = (\theta_1,\dots,\theta_d)$

$\theta_{-j}^{t-1} = (\theta_1^t,\dots,\theta_{j-1}^t,\theta_{j+1}^{t-1},\dots,\theta_d^{t-1})$

zeros (n,1) m = zeros (m,1) m = zeros (m,1); y(i) = yi ~~...~~ y(i) = y(i)

y(i) y(i+1) y(i-1) y(i^2) Gamma $\frac{1}{s}$ $\left(\frac{1}{s\,nt}^{n}\right) \times m$ $\frac{1}{s}$

man Duality pdf hist duality pd

σ

μ → Gen Norm → IGen Gam ← many μ/σ Inverse Marginal → mean sigma

assum

y ← many yi

$\bar{y}$ $S_2 = \frac{\sigma}{n-1}$

$S_2 t$

$\mu = \bar{y} + \sqrt{\frac{S_2 t}{n}}$ → Normal

$p(\mu, \sigma^2 | y) = p(\mu | \sigma^2, y)\, p(\sigma^2 | y)$

$p(\mu | \sigma^2, y) \sim N\left(\bar{y}, \frac{\sigma^2}{n}\right)$

Likelihood $p(y | \mu, \sigma^2) \propto \sigma^{-n} e^{-\frac{1}{2}\sum_{i=1}^{n} \frac{(y_i - \mu)^2}{\sigma^2}}$

Prior $p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$

Full joint posterior $\propto \sigma^{-n-2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} \frac{(y_i - \mu)^2}{\sigma^2}}$

$p(\mu, \sigma^2 | y) \propto \sigma^{-n-2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} \left((n-1)s^2 + n(\bar{y} - \mu)^2\right)}$

$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$

$p(\sigma^2 | y) \propto \int \sigma^{-n-2} e^{-\frac{1}{2\sigma^2}\left((n-1)s^2 + n(\bar{y} - \mu)^2\right)} d\mu$

$p(\sigma^2 | y) \propto \sigma^{-n-2} e^{-\frac{1}{2\sigma^2}(n-1)s^2} \sqrt{\frac{2\pi\sigma^2}{n}}$

$\propto \sigma^{-\frac{(n+1)}{2}} e^{-\frac{1}{2\sigma^2}(n-1)s^2}$

$\sim IG\left(\frac{n-1}{2}, \frac{n-1}{2} s^2\right)$

P288

Large Sample infer

102-113

hierarch model → 125-150

Gibbs Sampler

Metropolis Markov chain

310

Effective Gibbs

$02/03/05$

Q1

Solution manual to the boo

pdf

① Ram souf

population

X eye

xx → blue

X, Xx → heterozygot

XX → brown

$\alpha p()$

① $\frac{\text{blue}}{\text{pop}} = p^2$

$\frac{\text{hetro}}{\text{pop}} = 2p(1-p)$

② parent hetro → P(fam\|heterozyg) X = 0.5

③ Rand mating

Q1 $\boxed{E(\text{heto}) = 2p/(1+2p)}$ show

Q2 Jude

brown eyes Jude

brown eyed parents

hetrozygot

n

all brown eyes

Q2 → P(Judy = hetrozygote)

P(first child) = bule eyes

Q2

$\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

$Z \sim Beta(\alpha,\beta)$

Q? $E[Z^m(1-Z)^n] = ?$ $\forall$ $m, n \in Z$

Q-2.2? $E(Z) = ?$
$V(Z) = ?$

Q1 Likelihood?

Q2? Prior?
Gibbs Sampler?

robust altrative Gaussian Linear Reg

H: sample y from $t_\nu(\mu,\sigma^2)$: ① Simulate α from $G(\nu/2, \nu/2)$

② simulate y from $N(\mu, \alpha^{-1}\sigma^2)$

Q3 Gibbs Sampler when deg freedom ν ~ Rand ν

takes either $\{5, 10\}$

Q3 Student Linear Reg

$y_t = X_t\beta + \epsilon_t$ $t = 1,\dots,n$

$\epsilon_t \sim iid \sim t_5(0,\sigma^2)$

deg 5

Q4 how simulate draws

truncated normal (inverse CDF)

$TN_{(a,\infty)}(\mu,\sigma^2)$ & $TN_{(a,b)}(\mu,\sigma^2)$

explicitely show pdfs & cdfs

HW: Gibbs sampler Gaussian linear Reg

Semi Conjugate prior (algorithm?)

$y_t = \beta x_t + \epsilon_t$ $t = 1,\dots,n$

$N(\beta,\sigma^2)$ $p(\sigma^2,\beta \mid y)$ $p(\sigma^2,\beta) = p(\sigma^2)p(\beta)$

$= (\sigma^2 \beta)^T$

Bayesian

→ Condition on main variable

① find likelihood (normal form of function) $P(\theta \mid \alpha, \beta)$ → don't forget to plug in the same variable that is not know we are willing to estimate

② use prior

③ calculate posterior { Refactor, change variable / Transform / Jacobian

* Single vs. multiple observation
* Conjugate vs. semi conjugate (independent & not conditional)
* informative vs. non informative

Importance of parameter and Conditioning on it

Just look for parameter that should appear in result

① Pinpoint var: e.g $\beta^\nu \beta$ ② find ③ match

- parametric space

Matrix multiplication

→ Reduction at dimension

$$(\square \ \square)\begin{pmatrix}\square & \square \\ \square & \square\end{pmatrix}\begin{bmatrix}\square \\ \square\end{bmatrix} = (\bigcirc)$$

Bayesian

Candidate generating density $q(x,y)$, $\int q(x,y)dy=1$

March 02

- Spread
- acceptance rate $\alpha(x,y)=\min\left[\frac{\pi(y)q(x,y)}{\pi(x)q(x,y)},1\right]$
- Region of sample covered by the chain
- auto correlation across sample values
- invariant density / dist: $\pi^*(dy)=\int_{R^d} P(x,dy)\pi(x)dx = \pi(x)dy$
- Reversability Condition: $p(x)p(x,y)=p(y)p(y,x)$
- Choleski: $P'P=\Sigma$
- Transition Kernel $P(x,A)$ $x\in R^d$ $A\in B$ (Borel set)
  $P(x,R^d)=1$ $P(x,\{x\})\geq 0$

MCMC

$P(x,dy)=p(x,y)dy+r(x)\delta_x(dy)$, $p(x,x)=0$

transition kernel $\delta_x(dy)\begin{cases}1 & x\notin dy\\ 0 & x\in dy\end{cases}$

$r(x)=1-\int_{R^d}p(x,y)dy$

M-H Algorithm

- for $j=1,2,\dots,N$
- Gen y from $q(x^{(j)},y)$ and u from $U(0,1)$
- Calc. prob. of move: if $u\leq\alpha(x^{(j)},y)$ set $x^{(j+1)}=y$
  otherwise, set $x^{(j+1)}=x^{(j)}$

$\pi^*_{i|-i}(dy_i|x_{-i})=\int P_i(x_i,dy_i|x_{-i})\pi_{i|-i}(x_i|x_{-i})dx_i$: Block at a time

Candidate generating density: ① Random walk chain $q(x,y)=q_1(y-x)$
② independence chain $q(x,y)=q_2(y)$

- marginal likelihood

- metropolice hasting
- accept reject
- Markov chain Monte Carlo
- moving probability $q(y)\alpha(x,y)$: satisfy reversability
  $\alpha(x,y)$
- Block at a time / variable at a time
- Kernel principle
- invarient density
- Random walk generating density $(y=x+\varepsilon)$

$\alpha(x,y)=\min\left\{\frac{e^{-\frac{1}{2}(y-\mu)'\Sigma(y-\mu)}}{e^{-\frac{1}{2}(x-\mu)'\Sigma(x-\mu)}},1\right\}$
$x,y\in R^2$

$\to$ not hold: $\pi(x).q(x,y)>\pi(y)q(y,x)$

$\Rightarrow$ transition kernel: $P_{MH}(x,y)=q(x,y)\alpha(x,y)$, $x\neq y$

$\alpha(x,y)=\frac{\pi(y).q(y,x)}{\pi(x).q(x,y)}$

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

$\Downarrow$

symmetric ratio: $\frac{\pi(y)}{\pi(x)}$

Convergence Req: Regularity ① irreducibility: possible finite iter. move
② Aperiodicity: not multiple of int

- Number of burn-in
- auto Correlation