UQ5

Bayesian techniques for parameter estimation

$$Y_i = f_i(Q) + \varepsilon_i, \quad i = 1 \cdots n, \quad \varepsilon_i \text{ unbiased iid}$$
Y_i, ε_i, Q random variables

Parameters random variables, realizations $q = Q(\omega)$
Q has density, updated by incoming data
$\Rightarrow$ posterior density for Q based on sample(d) observations

prior density, likelihood $\Rightarrow$ posterior density
$\pi_0(q)$ $\quad \pi(y|q)$ $\quad \pi(q|y)$

$\pi_0(q)$ knowledge obtained before observations y
noninformative prior $\pi_0(q) = \chi_{(0,\infty)}(q)$

$$\chi_{(a,b)}(q) = \begin{cases} 1, & q \in (a,b) \\ 0, & \text{otherwise} \end{cases} \quad \text{indicator function}$$

$\pi(y|q) = L(q|y)$ based on information from samples

$\pi(q,y)$ joint density of Q and Y

$$\pi(y|q) = \frac{\pi(q,y)}{\pi_0(q)}$$

$$\Rightarrow \quad \pi(q|y_{obs}) = \frac{\pi(q, y_{obs})}{\pi(y_{obs})} \quad \text{posterior density}$$

$$\pi(y_{obs}) = \int_{\mathbb{R}^p} \pi(q, y_{obs})\,dq = \int_{\mathbb{R}^p} \pi(y_{obs}|q)\,\pi_0(q)\,dq$$

Q has prior density $\pi_0(q)$

$$\pi(q|y_{obs}) = \frac{\pi(y_{obs}|q)\pi_0(q)}{\pi(y_{obs})} = \frac{\pi(y_{obs}|q)\pi_0(q)}{\int_{\mathbb{R}^p} \pi(y_{obs}|q)\pi_0(q)\,dq}$$

$$\Rightarrow \int_{\mathbb{R}^p} \pi(q|y_{obs})\,dq = 1$$

like lect 4

Likelihood function ($y = y_{obs}$ from now on)

$\varepsilon_i \sim N(0, \sigma^2)$, iid, σ^2 fixed

$$\pi(y|q) = L(q, \sigma^2|y) = \frac{1}{(2\pi\sigma^2)^{n/2}} \underbrace{e^{-SS_q/2\sigma^2}}_{\prod_{i=1}^n e^{-(\underbrace{y_i - f_i(q)}_{\varepsilon_i})/2\sigma^2}}$$

$$SS_q = \sum_{i=1}^n (y_i - f_i(q))^2$$

Maximum a posteriori (MAP) estimate

$$q_{MAP} = \arg\max_q \pi(q|y)$$

normalization $\pi(y)$ does not change the max

$$\Rightarrow q_{MAP} = \arg\max \pi(y|q)\pi_0(q)$$

$\pi_0(q)$ uniform on $\mathbb{R}$ $\Rightarrow$

$$q_{MAP} = q_{MLE} = \arg\max_q \pi(y|q)$$

Take $\pi_0(q) = \chi_{[0,\infty)}(q)$

$$\Rightarrow \pi(q|y) = \frac{e^{-SS_q/2\sigma_0^2}}{\int_0^\infty e^{-SS_\zeta/2\sigma_0^2}\,d\zeta} = \frac{1}{\int_0^\infty e^{-(SS_\zeta - SS_q)/2\sigma_0^2}\,d\zeta}$$

Markov Chain Monte Carlo (MCMC)

Evaluation of integral over $\mathbb{R}^p$ for posterior definition is expensive when p is large

Quadrature or Monte Carlo techniques

Instead: construct Markov chains whose stationary distribution is the posterior density

Strategy $\quad q^{k-1} \in \mathbb{R}^p$

1. $X_{k-1} = q^{k-1}$ ← Markov property
2. $q^* \sim J(q^* | q^{k-1})$, J proposal or jumping distribution
3. With probability $\alpha(q^* | q^{k-1})$ accept q^* and let $X_k = q^*$, otherwise $X_k = q^{k-1}$

Metropolis algorithm (1953, Teller, Rosenbluth ...)

Assume J symmetric: $J(q^* | q^{k-1}) = J(q^{k-1} | q^*)$

Take $\quad J(q^* | q^{k-1}) = N(q^{k-1}, V)$

$\quad\quad\; J(q^* | q^{k-1}) = N(q^{k-1}, D)$ ← diagonal

V is covariance matrix for Q

$$J(q^* | q^{k-1}) = \frac{1}{\sqrt{(2\pi)^p |V|}} e^{-\frac{1}{2}[(q^* - q^{k-1}) V^{-1} (q^* - q^{k-1})^T]}$$

$|V|$: det V $\quad\quad$ V^{-1}: symmetric

1. Initialization : q^0 such that $\pi(q^0|y) > 0$

2. For $k = 1 \cdots M$

a For $z \sim N(0,1)$: $q^* = q^{k-1} + Rz$
R Choleski Decomposition of V or D, $V = RR^T$
$\Rightarrow q^* \sim N(q^{k-1}, V)$ or $N(q^{k-1}, D)$

b Compute

$$r(q^*|q^{k-1}) = \frac{\pi(q^*|y)}{\pi(q^{k-1}|y)} = \frac{\pi(y|q^*)\pi_0(q^*)}{\pi(y|q^{k-1})\pi_0(q^{k-1})}$$

c

$$q^k = \begin{cases} \cancel{q^*} & \cancel{\text{with probability } \alpha = \min(1, r)} \\ q^{k-1} & \text{else} \end{cases}$$

$\Rightarrow r \geq 1 \Rightarrow q^k = q^*$ $(\pi(q^*|y) \geq \pi(q^{k-1}|y))$
otherwise q^* is accepted with probab. r
rejected $1-r$

Likelihood (normally Distributed $\overbrace{\text{errors}}^{iid}$, uniform prior)

$$\pi(y|q) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$

$$SS_q = \sum_{i=1}^{n} (y_i - f_i(q))^2$$

$$r(q^*|q^{k-1}) = \frac{\pi(y|q^*)}{\pi(y|q^{k-1})} = \frac{e^{-SS_{q^*}/2\sigma^2}}{e^{-SS_{q^{k-1}}/2\sigma^2}}$$

$$= e^{-(SS_{q^*} - SS_{q^{k-1}})/2\sigma^2}$$

$SS_{q^{k-1}} \geq SS_{q^*} \Rightarrow r \geq 1$ and $q^k = q^*$

$SS_{q^{k-1}} < SS_{q^*} \Rightarrow r < 1$ and $q^k = q^*$ with prob. r

large variance $SS_{q^*} \Rightarrow$ often rejected, chain stagnates

small variance $SS_{q^*} \Rightarrow$ often accepted but small part of parameter space is explored

Choice of V: $V = \sigma_{OLS}^2 \left(\mathcal{X}^T(q_{OLS}) \mathcal{X}(q_{OLS}) \right)^{-1}$

$$\sigma_{OLS}^2 = \frac{1}{n-p} \sum_{i=1}^{n} \left(y_i - f_i(q_{OLS}) \right)^2$$

<u>Random Walk Metropolis algorithm</u> (RWMA)

1. M, n_s, σ_s
2. $q^0 = \arg\min_q \sum_{i=1}^{N} (y_i - f_i(q))^2$
3. $SS_{q^0} = \sum_{i=1}^{N} (y_i - f_i(q^0))^2$
4. $s_0^2 = \frac{SS_{q^0}}{n-p}$
5. $V = s_0^2 \left(\mathcal{X}^T(q^0) \mathcal{X} \right)^{-1} = R R^T$
6. For $k = 1, \dots M$
   - a Sample $z_k \sim N(0, I_p)$
   - b. $q^* = q^{k-1} + R z_k$
   - c. Sample $u \sim U(0,1)$
   - d. $SS_{q^*} = \sum_{i=1}^{N} (y_i - f_i(q^*))^2$
   - e. $\alpha(q^* | q^{k-1}) = \min\left(1, e^{-(SS_{q^*} - SS_{q^{k-1}})/2 s_{k-1}^2}\right)$
   - f. If $u < \alpha$: $q^k = q^*$, $SS_{q^k} = SS_{q^*}$
     else $q^k = q^{k-1}$ $SS_{q^k} = SS_{q^{k-1}}$
   - g. Update $s_k^2 \sim \text{Inv-gamma}(aval, bval)$
     $aval = \frac{1}{2}(n + n_s)$, $bval = \frac{1}{2}(n_s \sigma_s^2 + SS_{q^k})$

     dimension (n), small (n_s)

Metropolis-Hastings algorithm for unsymmetric J: $\underbrace{J(q^*|q^{k-1})}_{\text{jump distribution}} \neq J(q^{k-1}|q^*)$

Markov chain has stationary distribution coinciding with posterior density

Sufficient condition for π to be stationary (detailed balance condition)

$$\pi_{k-1} p_{k-1,k} = \pi_k p_{k,k-1} \qquad (*)$$

constructed by Metropolis alg.

Let $\pi_k = \pi(q^k|y)$, $p_{k-1,k} = P(X_k = q^k | X_{k-1} = q^{k-1})$

$(*) \Rightarrow \pi(q^{k-1}|y) p_{k-1,k} = \pi(q^k|y) p_{k,k-1}$

$$\begin{aligned} p_{k-1,k} &= P(\text{proposing } q^k) P(\text{accepting } q^k) \\ &= J(q^k|q^{k-1}) \alpha(q^k|q^{k-1}) \\ &= J(q^k|q^{k-1}) \min\left(1, \frac{\pi(q^k|y) J(q^{k-1}|q^k)}{\pi(q^{k-1}|y) J(q^k|q^{k-1})}\right) \end{aligned}$$

$$z \min(1, x/z) = \min(x, z) = x \min(1, z/x)$$

$$\Rightarrow \pi(q^{k-1}|y) p_{k-1,k} = \pi(q^{k-1}|y) J(q^k|q^{k-1}) \min\left(1, \frac{\overbrace{\pi(q^k|y) J(q^{k-1}|q^k)}^{x}}{\underbrace{\pi(q^{k-1}|y) J(q^k|q^{k-1})}_{z}}\right)$$

$$= \pi(q^k|y) J(q^{k-1}|q^k) \underbrace{\min\left(1, \frac{\pi(q^{k-1}|y) J(q^k|q^{k-1})}{\pi(q^k|y) J(q^{k-1}|q^k)}\right)}_{\alpha(q^{k-1}|q^k)}$$

$$= \pi(q^k|y) p_{k,k-1}$$

Detailed balance $\Rightarrow$ stationary solution π

chain iterated sufficiently long ⇒
samples from posterior density

Convergence criteria for termination: after
burn-in period
acceptance ratio: percentage of accepted points
$\sim 0.1 - 0.5$
quantifies whether sampling is adequate
☆ →

Delayed Rejection Adaptive Metropolis (DRAM)
Adaptive algorithms use part of chain history
to update proposal function J (no longer
Markovian)

* Update V_k using $q^0, q^1 \cdots q^{k-1}$ (Adaptive Metropolis)
permanent changes of q^*
* If q^* is rejected construct q^{*j} (Delayed Rejection)
temporary changes in J, α

Still: DRAM inefficient for various parameter
regimes: multimodal, highly complex, heavy tails

⇒ parallel chains of AM algorithms

Exercise 8.2, use Random Walk Metropolis

☆ $q^k, q^{k+x}, q^{k+2x} \cdots$ approximates stationary π
x such that $q^{k+nx}, q^{k+(n+1)x} \simeq$ independent
test independence with autocorrelation

$$R(k) = \frac{\sum_{i=1}^{M-k} (q_i - \bar{q})(q_{i+k} - \bar{q})}{\sum_{i=1}^{M} (q_i - \bar{q})^2} = \frac{cov(q_i, q_{i+k})}{var(q_i)}$$