2nd tutorial

(1)(i). We derive the joint posterior of the normal likelihood with both mean and variance parameters unknown and based on three Gaussian priors. Suppose that $y = \{y_1, y_2, \dots, y_n\}$ is a sample of i.i.d Observations from $N(\mu, \sigma^2)$. The joint likelihood of (μ, σ^2) is given by,

$$L(\mu, \sigma^2 | y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\Sigma (y_i - \mu)^2 / \sigma^2\right)}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} \Sigma (y_i - \mu)^2}$$

We want the joint posterior $P(\mu, \sigma^2 | y)$ when both μ and σ^2 are unknown, we use non-informative prior.

$$P(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$$

Posterior (up to constant)

Posterior $\propto$ likelihood $\times$ prior.

$$P(\mu, \sigma^2 | y) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \Sigma (y_i - \mu)^2\right) \times \frac{1}{\sigma^2}$$

$$\propto (\sigma^2)^{-\frac{n}{2}-1} \exp\left(-\frac{1}{2\sigma^2} \Sigma (y_i - \mu)^2\right)$$

CS CamScanner

$$\therefore P(\mu, \sigma^2|y) \propto (\sigma^2)^{-(\frac{n}{2}+1)} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right)$$

Decompose the quadratic term,

$$\sum_{i=1}^{n}(y_i-\mu)^2 = \sum_{i=1}^{n}\left[(y_i-\bar{y})+(\bar{y}-\mu)\right]^2$$

while, $\bar{y} = \frac{1}{n}\sum y_i$, Expanding,

$$\sum(y_i-\mu)^2 = \sum_{i=1}^{n}(y_i-\bar{y})^2 + n(\bar{y}-\mu)^2$$

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

$$\therefore \sum(y_i-\mu)^2 = (n-1)s^2 + n(\bar{y}-\mu)^2$$

Now, we subtitute backe

$$P(\mu, \sigma^2|y) \propto \frac{1}{(\sigma^2)^{\frac{n}{2}+1}} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right)$$

$$\propto \frac{1}{\sigma^{n+2}} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right\}$$

So, the joint posterion distribution due to the non-informative prion,

$$P(\mu, \sigma^2|y) = \frac{1}{\sigma^{n+2}} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right\}$$

(ii) We assume the non informative prior, ~~$P(\mu,\sigma^2)$~~

$$P(\mu,\sigma^2) = 1/\sigma^2$$

Joint posterior (from likelihood x prior)

Likelihood of $y = (y_1, y_2, \dots, y_n)$:

$$L(\mu,\sigma^2 / y) \propto (\sigma^2)^{-\left(\frac{n}{2}\right)} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right]$$

with the prior $1/\sigma^2$ is,

$$P(\mu,\sigma^2/y) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right]\frac{1}{\sigma^2}$$

$$\propto (\sigma^2)^{-\left(\frac{n}{2}+1\right)} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right]$$

Use the identity:

$$\sum_{i=1}^{n}(y_i-\mu)^2 = \sum_{i=1}^{n}\left[(y_i-\bar{y})+(\bar{y}-\mu)\right]^2$$

where $\bar{y} = \frac{1}{n}\sum y_i$, expanding,

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

$$= (n-1)s^2 + n(\bar{y}-\mu)^2$$

with $\bar{y}$ sample mean and s^2 sample variance

So,

$$P(\mu,\sigma^2/y) \propto (\sigma)^{-(n+2)} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right\}$$

Conditional posterior of μ given σ^2

(a) For fixed σ^2, the term in μ is Gaussian:

$$P(\mu, \sigma^2 | y) \propto \exp\left\{-\frac{n}{2\sigma^2}(\mu - \bar{y})^2\right\}$$

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

So, we can say that when σ^2 is known then μ is normal. $P(\mu | \sigma^2, y)$ is $N\left(\bar{y}, \frac{\sigma^2}{n}\right)$

(b) Conditional posterior of $\sigma^2 | \mu, y$

$$P(\sigma^2 | \mu, y) \propto (\sigma^2)^{-\frac{n}{2}} \left\{\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right] \times (\sigma^2)^{-1}\right.$$

$$\Rightarrow P(\sigma^2 | \mu, y) \propto (\sigma^2)^{-\left(\frac{n}{2}+1\right)} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right]$$

$$\therefore \sigma^2 | \mu, y \sim \text{Inverse-Gamma}\left(\frac{n}{2}, \frac{1}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right)$$

CamScanner

(iii) Derive the marginal posterior of μ:

$$L(\mu, \sigma^2 | y) \propto (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right]$$

$$\propto (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\bar{y}-\mu)^2\right\}\right]$$

$$P(\mu, \sigma^2 | y) \propto (\sigma^2)^{-(\frac{n}{2}+1)} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\bar{y}-\mu)^2\right\}\right]$$

Marginal posterior,

$$P(\mu | y) \propto \left[(n-1)s^2 + n \cancel{\bar{y}} (\mu - \bar{y})^2\right]^{-n/2}$$

$$\mu | y \sim t\left(\bar{y}, \frac{s}{\sqrt{n}}\right)$$

$$\frac{\mu - \bar{y}}{s/\sqrt{n}} \sim t_{n-1}$$

2. Symmetry and posterior summaries:

⇒ Student t density is perfectly symmetric around its location parameter.

therefore,

$$E(\mu | y) = \bar{y}, \quad \text{Median}(\mu | y) = \bar{y}, \quad \text{mode}(\mu | y) = \bar{y}$$

3. Posterior variance:

⇒ Student t - variance formula ($df = n-1$)

$$\text{Var}(\mu | y) = \frac{(n-1)}{n(n-1)} s^2$$

CamScanner

For σ^2

$Y_i \sim N(\mu, \sigma^2)$

Non-informative prior, $P(\mu, \sigma^2) \propto \sigma^{-2}$

Marginal posterior of σ^2,

$$P(\sigma^2 | y) \propto (\sigma^2)^{-\left(\frac{n}{2}+1\right)} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \times$$

$$\int_{-\infty}^{\infty} \exp\left(-\frac{n}{2\sigma^2}(\mu - \bar{y})^2\right) d\mu$$

Now,

$$\int_{-\infty}^{\infty} \exp\left(-\frac{n}{2\sigma^2}(\mu - \bar{y})^2\right) d\mu = \sqrt{\frac{2\pi\sigma^2}{n}}$$

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

That is kernel of an inverse gamma with,

$$\alpha = \frac{n-1}{2}, \quad \beta = \frac{(n-1)s^2}{2}$$

Equivalently,

$$\sigma^2 | y \propto \text{Inv-Gamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right) \equiv \text{Inv-}\chi^2\left[(n-1), s^2\right]$$

$$\therefore P(\mu, \sigma^2 | y) \propto N\text{-Inv-}\chi^2(\bar{y}, n, n-1, s^2)$$

CamScanner

(iv) Suppose we want to know the $dist^n$ of a future observation $\tilde{y}$ (normal $dist^n$ as y). In other words, we wish to know the posterior predictive distribution (PPD). We don't know about μ and σ^2

$$y_i \sim N(\mu, \sigma^2)$$

Non-informative prior, $P(\mu, \sigma^2) \propto (\sigma^2)^{-1}$

(1) The posterior predictive distribution for $\tilde{y}$ is,

$$P(\tilde{y}|y) = \iint N(\bar{y} | \mu, \sigma^2)\, P(\mu, \sigma^2 | y)\, d\mu\, d\sigma^2$$

After integrating over μ and σ^2, the result is student t-distribution.

$$\tilde{y} | y \sim t_{n-1}\left(\bar{y}, s^2(1 + 1/n)\right)$$

where,

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{(sample mean)}$$

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

$(n-1)$ d.f

location parameter $\bar{y}$
and scale parameter $s\sqrt{(1+\frac{1}{n})}$

CS CamScanner

The joint posterior (μ, σ^2) is normal inverse Gamma - marginal.

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

$P(\mu / y) \sim t_{n-1} (\bar{y}, s^2/n)$

$P(\sigma^2 | y) \sim Inv\text{-}Gamma \left(\frac{n-1}{2}, \frac{(n-1)s^2}{2} \right)$

$\tilde{y} | y \sim t_{n-1} (\bar{y}, s^2 (1 + \frac{1}{n}))$

$\frac{\tilde{y} - y}{s\sqrt{1 + 1/n}} \sim t_{n-1}$

$E(\tilde{y} | y) = \bar{y}$

$var(\tilde{y} | y) = \frac{(n+1)s^2}{n-3} \quad ; \; n > 3$

(v) ~~Show~~ ~~that~~ ~~joint~~

(v) Joint posterior. $P(\mu, \sigma^2 / y) \propto P(\bar{y} | \mu, \sigma^2) \times P(\mu, \sigma^2)$ ~~using~~

Using the chain rule of probability.

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

Then,

$$P(\mu, \sigma^2 | y) = N(\bar{y}, \sigma^2/n) \times \text{Inverse-Gamma}\left(\frac{n-1}{2}, \frac{\Sigma(x_i-\bar{x})^2}{2}\right)$$

This explicitly shows that the joint posterior is the product of the conditional posterior of μ given σ^2 and marginal posterior of σ^2, which is general property in Bayesian inference.

<u>Solution (2):</u> Given, $n = 5$

$$L = P(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

$$= \begin{cases} \binom{5}{x} \left(\frac{1}{2}\right)^x \left(1-\frac{1}{2}\right)^{5-x} & ; \text{fair} \\ \binom{5}{x} (0.7)^x (0.3)^{5-x} & ; \text{loaded} \end{cases}$$

Prior, $P(\theta) = 0.7$ (head) ; $P(\text{fair}) = 0.4$

$P(\theta) = 0.3$ (tail) ; $P(\text{loaded}) = 0.6$

$$P(\text{fair} | x=2) = \frac{P(x|\text{fair}) \cdot P(\text{fair})}{P(x|\text{fair}) \cdot P(\text{fair}) + P(x|\text{loaded}) \cdot P(\text{loaded})}$$

$$= \frac{\left[\binom{5}{2}\left(\frac{1}{2}\right)^5\right] \times 0.4}{\left[\binom{5}{2}\left(\frac{1}{2}\right)^5\right] 0.4 + \left[\binom{5}{2}(0.7)^2 (0.3)^3\right] 0.6}$$

CamScanner

$$= \frac{0.125}{0.20438}$$

$\therefore$ ~~P(fair)~~ $= 0.612$

$\therefore P(\text{fair} \mid x=2) = 61\%$

~~Now~~ And,

$$P(\text{loaded} \mid x=2) = \frac{P(x \mid \text{loaded}) \cdot P(\text{loaded})}{P(x \mid \text{loaded}) \cdot P(\text{loaded}) + P(x \mid \text{fair}) \cdot P(\text{fair})}$$

$$= \frac{\left[\binom{5}{2} (0.7)^2 (0.3)^3\right] 0.6}{\left[\binom{5}{2} (0.7)^2 (0.3)^3\right] 0.6 + \left[\binom{5}{2} \left(\frac{1}{2}\right)^5\right] 0.4}$$

$$= \frac{0.07938}{0.20938}$$

$$= 0.388$$

$\therefore P(\text{loaded} \mid x=2) = 39\%$

(i) Metropolis - Hasting steps to estimate posterior probabilities

Posterior on the two state space $\{L, F\}$: | L = loaded, F = fair

$\pi(L) = P(L|y)$, $\pi(F) = P(F|y)$

Metropolis - Hastings algorithm

1. Choose an initial state $X_0 \in \{L, F\}$

2. For $t = 0, 1, 2, \ldots$

· Propose $Y \sim q(\cdot | X_t)$

· Accept the proposal with probability

$$\alpha(X_t \to Y) = \min\left\{1, \frac{\pi(Y) q(X_t | Y)}{\pi(X_t) q(Y | X_t)}\right\}$$

· Generate $U \sim \text{Uniform}(0, 1)$ set

$$X_{t+1} = \begin{cases} Y, & U \le \alpha(X_t \to Y) \\ X_t, & U > \alpha(X_t \to Y) \end{cases}$$

3. After burn - in estimate

$$\hat{P}(L|y) = \frac{1}{N} \sum_{t=1}^{N} 1\{X_t = L\}$$

which converges to $P(L|y) = 0.388$

CamScanner

(~~ii~~) ~~The MH algorithm~~

(ii) Two-state Markov chain diagram and show $\pi P = \pi$

$\Rightarrow$ Let the state space be

$S = \{F, L\}$ ~~where~~, ~~F = fair~~ ~~L = loaded~~

where,

F: fair coin $(p = 0.5)$

L: loaded coin $(p = 0.7)$

The stationary distribution is

$$\pi(F) = P(F|y), \quad \pi(L) = P(L|y)$$

From Bayes' result in (i) we have

$\pi(F) = 0.612$

$\pi(L) = 0.388$

Now, we use a symmetric proposal:

$q_{FL} = q_{LF} = 0.5, \quad q_{FF} = q_{LL} = 0.5$

General Metropolis hasting acceptance probability is:

$$\alpha_{ij} = \min\left(1, \frac{\pi(j)\, q_{ji}}{\pi(i)\, q_{ij}}\right)$$

Since q_{ij} is symmetric $(q_{ij} = q_{ji})$

$$\alpha_{ij} = \min\left(1, \frac{\pi(j)}{\pi(i)}\right)$$

Now,

$$\alpha_{FL} = \min\left(1, \frac{\pi(L)}{\pi(F)}\right) = \min\left(1, \frac{0.388}{0.612}\right)$$

$$= \min(1, 0.634)$$

$$= 0.634$$

and

$$\alpha_{LF} = \min\left(1, \frac{\pi(F)}{\pi(L)}\right) = \min\left(1, \frac{0.612}{0.388}\right)$$

$$= \min(1, 1.577)$$

$$= 1$$

Also,

$$\alpha_{FF} = \alpha_{LL} = 1$$

Now,

Transition probabilities P_{ij}

for $i \neq j$, $P_{ij} = q_{ij}\,\alpha_{ij}$

So, From F:

$$P_{FL} = q_{FL}\,\alpha_{FL} = 0.5 \times (0.634) = 0.317$$

$$P_{FF} = 1 - P_{FL} = 1 - 0.317 = 0.683$$

From L:

$$P_{LF} = q_{LF}\,\alpha_{LF} = 0.5 \times 1 = 0.5$$

$$P_{LL} = 1 - P_{LF} = 1 - 0.5 = 0.5$$

Transition matrix P

Ordering states as (F, L)

$$P = \begin{pmatrix} P_{FF} & P_{FL} \\ P_{LF} & P_{LL} \end{pmatrix} = \begin{pmatrix} 0.683 & 0.317 \\ 0.5 & 0.5 \end{pmatrix}$$

Now,

$$\pi = (\pi(F), \pi(L)) = (0.612, 0.388)$$

$$(\pi P)_F = \pi(F) P_{FF} + \pi(L) P_{LF}$$

$$= 0.612 \times 0.683 + \cancel{0.612 \times} \; 0.388 \times 0.5$$

$$= 0.418 + 0.194$$

$$= 0.612$$

$$= \pi(F)$$

and, $$(\pi P)_L = \pi(F) P_{FL} + \pi(L) P_{LL}$$

$$= 0.612 \times (0.317) + 0.388 \times 0.5$$

$$= 0.194 + 0.198$$

$$= 0.388$$

$$= \pi(L)$$

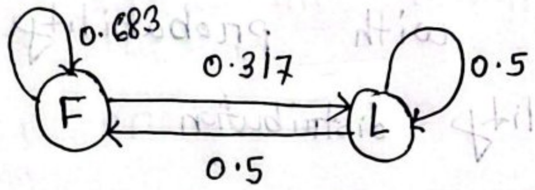
$\therefore$ Hence

$$\pi P = \pi$$

So, the Markov chain generated by Metropolis-Hastings has stationary distribution equal to the posterior distribution π.

CamScanner

Diagram



▣ Show that, the MCMC outcome satisfies the stochastic process property (i.e. Markovian property holds) and total $\pi P = \pi$

Answer:

The MH algorithm creates a reversible Markov chain, i.e. a Markov chain that satisfies the detailed balance condition.

The proof goes as ~~foo~~ follows: Suppose that, π is a discrete distribution with $\pi_j = P(\theta = x_j)$ with possible values $S = \{x_1, x_2, \ldots, x_n\}$. Further suppose that $\theta = (q_{ij})_{ij}$ is a matrix that describes the moves, i.e. if the chain is in state x_i, then the state x_j is drawn with probability q_{ij}. Now, one can either move from x_i to a different x_j or stay at x_i. the discrete version of the MH algorithm states that the move should happen with probability.

π = target probability distribution

P = transition Prob^y matrix of the Markov chain.

$$\alpha_{ij} = \min\left(1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right)$$

The probability to move from x_i to x_j is therefore $p_{ij} = \alpha_{ij} q_{ij}$. The detailed balance condition which implies here that $\pi_i p_{ij} = \pi_j p_{ji}$ is statisfied because

$$\pi_i p_{ij} = \alpha_{ij} q_{ij} \pi_i$$

$$= \pi_i \min\left(1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right) q_{ij}$$

$$= \min\left(\pi_i q_{ij}, \pi_j q_{ji}\right)$$

$$= \min\left(1, \frac{\pi_i q_{ij}}{\pi_j q_{ji}}\right) \pi_j q_{ji}$$

$$= \pi_j p_{ji}$$

then,

$$(\pi P)_j = \sum_j \pi_j p_{ji}$$

$$= \pi_j \sum_i p_{ji} = \pi_j \qquad \left[\because \text{ Here } \sum_i p_{ji} = 1\right]$$

∴ In vector form,

$$\pi P = \pi$$

This proves π is a stationary distribution of the Markov chain defined by P.

Answer: 3

i) <u>Conditional conjugate prior:</u>

A prior is called conditional conjugate when the prior distribution of a parameter, given other parameters and the data, belongs to the same family as its posterior.

It allows efficient Bayesian updating and simplifies sampling methods such as Gibbs or Metropolis-Hastings.

(ii) <u>Jeffry's prior:</u>

A non-informative prior defined as:

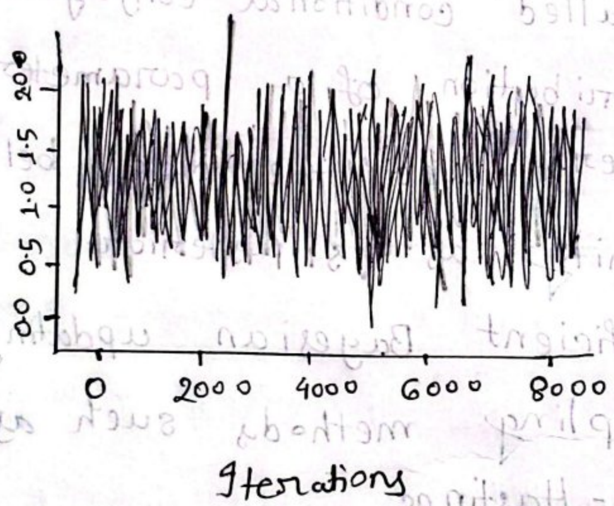
$$P(\theta) \propto \sqrt{I(\theta)}$$

where $I(\theta)$ is the Fisher information.
It is invariant under reparameterization, objective and used when no prior knowledge

is available.

Example: for a Bernoulli coin, Jeffrey's Prior is Beta$(1/2, 1/2)$

(iii) Traceplot: A traceplot shows the history of a parameter value across iterations of the chain. It shows precisely where the chain has been exploring.



(iv) Burn in Period:

Burn in period refers to the initial iterations of an MCMC chain that are strongly influnced by the starting values and do not represent the stationary (Posterior) distribution.

Therefore, these early samples are discarded before performing Monte Carlo estimation.

Removing burn in helps ensure that posterior summaries such as the mean, credible intervals and probabilities are reliable and unbiased.