Exercise 1

Bayesian inference in simple conjugate families

(A). $\left.\begin{matrix} x_i \in \{0, 1\} \\ w = p(x_i = 1) \end{matrix}\right\} \Rightarrow p(x_1 \cdots x_N | w) = \prod_{i=1}^{N} w^{x_i}(1-w)^{(1-x_i)} = w^{\sum_{i=1}^{N} x_i}(1-w)^{\sum_{i=1}^{N}(1-x_i)}$

prior: $p(w) \propto w^{a-1}(1-w)^{b-1}$

$\Rightarrow$ posterior: $p(w|x_1 \cdots x_N) \propto p(x_1 \cdots x_N | w)\, p(w)$

$$= w^{\sum_{i=1}^{N} x_i + a - 1}(1-w)^{\sum_{i=1}^{N}(1-x_i) + b - 1}$$

$\therefore p(w|x_1 \cdots x_N) \sim Beta\left(\sum_{i=1}^{N} x_i + a,\ \sum_{i=1}^{N}(1-x_i) + b\right)$.

$\sum_{i=1}^{N} x_i$: times of $x_i = 1$

$\sum_{i=1}^{N} (1-x_i)$: times of $x_i = 0$

(B). change of variable: $p_y(y = h(x)) = p_x(x = h^{-1}(y)) \left| \frac{\partial h^{-1}(y)}{\partial (y)} \right|$

$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = h(X) = \begin{bmatrix} x_1/(x_1+x_2) \\ x_1 + x_2 \end{bmatrix}$, $\begin{cases} x_1/(x_1+x_2) = y_1 \\ x_1 + x_2 = y_2 \end{cases} \Rightarrow \begin{cases} x_1 = y_1 y_2 \\ x_2 = (1-y_1) y_2 \end{cases}$

$\Rightarrow h^{-1}(Y) = X = \begin{bmatrix} y_1 y_2 \\ (1-y_1) y_2 \end{bmatrix} \Rightarrow \left| \frac{\partial h^{-1}(Y)}{\partial (Y)} \right| = \begin{vmatrix} y_2 & y_1 \\ -y_2 & (1-y_1) \end{vmatrix} = |y_2| \xrightarrow[x_1>0,\ x_2>0]{y_1 = x_1 + x_2} y_2 \cdots ①$

$p(X) = p(x_1) p(x_2) \propto x_1^{a_1-1} x_2^{a_2-1} e^{-b_1 x_1} e^{-b_2 x_2}$

$p(y_1, y_2) = p(Y) \propto (y_1 y_2)^{a_1-1} [(1-y_1) y_2]^{a_2-1} e^{-b y_1 y_2} e^{-b(1-y_1) y_2} \cdot y_2 = (y_1 y_2)^{a_1-1} [(1-y_1) y_2]^{a_2-1} e^{-b y_2} \cdot y_2$

$\Rightarrow p(y_1) = \int_{y_2} p(y_1, y_2) \propto y_1^{(a_1-1)} (1-y_1)^{(a_2-1)} \underbrace{\int_{y_2} y_2^{(a_1+a_2)-1} e^{-b y_2}}_{}$ $\to$ Gamma(a_1+a_2, b) integral

$\Rightarrow p(y_1) \sim Beta(a_1, a_2)$

$p(y_2) = \int_{y_1} p(y_1, y_2) \propto y_2^{(a_1+a_2)-1} e^{-b y_2} \underbrace{\int_{y_1} y_1^{(a_1-1)} (1-y_1)^{(a_2-1)}}_{}$ $\to$ Beta(a_1, a_2) integral

$\Rightarrow p(y_2) \sim Gamma(a_1+a_2, b)$.

Simulate Beta: ① Simulate pair (x_1, x_2) from Gamma
② $y = \frac{x_1}{x_1+x_2}$
③ y follows Beta distribution

(C) $p(x_i|\theta) \sim N(\theta, \sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}(x_i-\theta)^2\right\}$ ①

$p(\theta) \sim N(m, v) \propto \exp\left\{-\frac{1}{2v}(\theta-m)^2\right\}$ ②

① $\Rightarrow p(X=x_1\cdots x_N|\theta) = \prod_{i=1}^{N} p(x_i) \propto \prod_{i=1}^{N} \exp\left\{-\frac{1}{2\sigma^2}(x_i-\theta)^2\right\} = \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\theta)^2\right\}$

posterior:

$$p(\theta|X) \propto p(X|\theta)p(\theta) = \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\theta)^2 - \frac{1}{2v}(\theta-m)^2\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[\left(\frac{N}{\sigma^2}+\frac{1}{v}\right)\theta^2 - 2\left(\frac{\sum_{i=1}^{N}x_i}{\sigma^2}+\frac{m}{v}\right)\theta + const\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\frac{N}{\sigma^2}+\frac{1}{v}\right)\left[\theta - \left(\frac{\sum_{i=1}^{N}x_i}{\sigma^2}+\frac{m}{v}\right)\Big/\left(\frac{N}{\sigma^2}+\frac{1}{v}\right)\right]^2\right\}$$

$$Var(\theta|X) = \left(\frac{N}{\sigma^2}+\frac{1}{v}\right)^{-1} = \left[\left(\frac{\sigma^2}{N}\right)^{-1} + v^{-1}\right]^{-1}$$

(Sum of precision)$^{-1}$ via $\left[\bar{X} \sim N\left(\theta, \frac{\sigma^2}{N}\right)\right]$

$$E(\theta|X) = \left(\frac{\sum_{i=1}^{N}x_i}{\sigma^2}+\frac{m}{v}\right)\Big/\left(\frac{N}{\sigma^2}+\frac{1}{v}\right) \overset{\sum x_i = N\bar{X}}{=\!=} \frac{v}{v+\frac{\sigma^2}{N}}\bar{X} + \frac{\frac{\sigma^2}{N}}{v+\frac{\sigma^2}{N}}m$$

Sum of cross-weighted mean.

(D). known mean, unkown Var for Normal

$p(x_i|w) \sim N(\theta, w^{-1}) = \left(\frac{w}{2\pi}\right)^{1/2}\exp\left\{-\frac{w}{2}(x_i-\theta)^2\right\}$ ---- ①

$p(w) \sim Gamma(a,b) \propto w^{a-1}\exp\{-bw\}$ ②

① $\Rightarrow p(X=x_1\cdots x_N|w) = \prod_{i=1}^{N}\left(\frac{w}{2\pi}\right)^{1/2}\exp\left\{-\frac{w}{2}(x_i-\theta)^2\right\} \propto w^{N/2}\exp\left\{-\frac{w}{2}\sum_{i=1}^{N}(x_i-\theta)^2\right\}$

posterior:

$$p(w|X) \propto p(X|w)p(w) \propto w^{N/2}\exp\left\{-\frac{w}{2}\sum_{i=1}^{N}(x_i-\theta)^2\right\} w^{a-1}\exp\{-bw\}$$

$$= w^{(a+\frac{N}{2})-1}\exp\left\{-\frac{\sum_{i=1}^{N}(x_i-\theta)^2+b}{2}w\right\}$$

$$\begin{cases} a^* = a + \frac{N}{2} \\ b^* = \frac{1}{2}\left[\sum_{i=1}^{N}(x_i-\theta)^2 + b\right] \end{cases}$$

$p(w|X) \sim Gamma(a^*, b^*)$

(E) unkown θ, known but idiosyncratic variance σ_i^2 (Exercise 1)

$x_i \sim N(\theta, \sigma_i^2)$, $\theta \sim N(m, v)$, $p(\theta | X) = ?$

$$p(\theta|X) \propto p(x_1 \cdots x_N|\theta)\, p(\theta) \propto \prod_{i=1}^{N} \exp\left\{-\frac{1}{2\sigma_i^2}(x_i-\theta)^2\right\} \exp\left\{-\frac{1}{2v}(\theta-m)^2\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{N} \frac{1}{\sigma_i^2}(x_i-\theta)^2 + \frac{1}{v}(\theta-m)^2\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^{N}\frac{1}{\sigma_i^2}+\frac{1}{v}\right)\theta^2 - 2\cdot\left(\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2}+\frac{m}{v}\right)\theta + \text{const}\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{N}\frac{1}{\sigma_i^2}+\frac{1}{v}\right)\left[\theta - \frac{\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2}+\frac{m}{v}}{\sum_{i=1}^{N}\frac{1}{\sigma_i^2}+\frac{1}{v}}\right]^2\right\}$$

$$Var(\theta|X) = \left(\sum_{i=1}^{N}\frac{1}{\sigma_i^2}+\frac{1}{v}\right)^{-1} = v^*$$

$$E(\theta|X) = \frac{\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2}+\frac{m}{v}}{\sum_{i=1}^{N}\frac{1}{\sigma_i^2}+\frac{1}{v}} = v^*\left(\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2}+\frac{m}{v}\right) = \sum_{i=1}^{N}\left(\frac{v^*}{\sigma_i^2}\right)x_i + \frac{v^*}{v}m \quad \text{(weighted Sum)}$$

(F)

$$p(w=\tfrac{1}{\sigma^2}) \sim Gamma(a,b) \propto w^{a-1}\exp\{-bw\}$$

$$p(x|\sigma^2) = p(x|w) = N(0, w^{-1}) \propto w^{\frac{1}{2}}\exp\{-\tfrac{w}{2}x\}$$

$$p(x) = \int_w p(x,w) = \int_w p(x|w)\,p(w) \propto \int_w w^{a-1}\exp\{-bw\}\exp\{-\tfrac{w}{2}x\}\,w^{\frac{1}{2}}$$

$$= \int_w \underbrace{w^{(a+\frac{1}{2})-1}\exp\{-(b+\tfrac{x}{2})w\}}_{\text{Gamma kernel}}$$

$$= \frac{\Gamma(a+\frac{1}{2})}{(b+\frac{x}{2})^{(a+\frac{1}{2})}} \propto \left(b+\frac{x}{2}\right)^{-\frac{2a+1}{2}}$$

$$\propto \left(1+\frac{1}{(2a)}\cdot\frac{x}{(\frac{b}{a})}\right)^{-\frac{(2a+1)}{2}} \sim t_{2a}\left(\mu=0, \sigma^2=\frac{b}{a}\right)$$

$x \sim N(\mu, w)$, $w \sim Gamma\left(\frac{v}{2}, \frac{v\sigma_0^2}{2}\right) \Rightarrow \underset{\text{marginal}}{p(x)} \sim t_v(\mu; \sigma_0^2)$

useful when model <u>hyper</u>parameters

(Basics)

(A). 1. prove $cov(x) = E(xx^T) - \mu\mu^T$

$$cov(X) = E\left[(X-\mu)(X-\mu)^T\right] = E(XX^T - X\mu^T - \mu X^T + \mu\mu^T)$$

$$= E(XX^T) - E(X\mu^T) - E(\mu X^T) + \mu\mu^T$$

$$= E(XX^T) - E(X)\mu^T - \mu E(X^T) + \mu\mu^T$$

$$= E(XX^T) - \mu\mu^T \cancel{- \mu\mu^T + \mu\mu^T}$$

$$= E(XX^T) - \mu\mu^T \quad \text{-----} \quad 1$$

2. prove: $cov(Ax+b) = A\,cov(x)A^T$

According to 1,

$$cov(Ax+b) = E\left[(Ax+b)(Ax+b)^T\right] - (A\mu+b)(A\mu+b)^T$$

$$= E\left[Axx^TA^T + Axb^T + bx^TA^T + bb^T\right] - A\mu\mu^TA - A\mu b^T - b\mu^TA^T - bb^T$$

$$= A\,E(XX^T)A^T + \cancel{A E(x) b^T} + \cancel{b E(x^T) A^T} + \cancel{bb^T} - A\mu\mu^TA - \cancel{A\mu b^T} - \cancel{b\mu^T A^T} - \cancel{bb^T}$$

$$= A\,E(XX^T)A^T - A\mu\mu^TA^T$$

$$= A\left[\underbrace{E(XX^T) - \mu\mu^T}_{cov(X)}\right]A^T$$

$$= A\,cov(X)A^T$$