Exercise 2.

A simple Gaussian location model

A1. $p(w) \sim Gamma(\frac{d}{2}, \frac{\eta}{2})$, $p(\theta|w) \sim N(\mu, (wk)^{-1})$, prior $p(\theta)$?

$$p(\theta) = \int_w p(\theta, w)dw = \int_w p(\theta|w)\, p(w)\, dw$$

$$\propto \int_w w^{\frac{d+1}{2}-1} exp\{-w \cdot \frac{k(\theta-\mu)^2}{2}\} exp\{-w \cdot \frac{\eta}{2}\} dw$$

$$= \int_W w^{\frac{d+1}{2}-1} exp\{-\frac{w}{2}[k(\theta-\mu)^2+\eta]\} dw$$

Assume: $\alpha = \frac{d+1}{2}$, $\beta = k(\theta-\mu)^2+\eta$

$$= \int_W w^{\alpha-1} exp\{-\frac{w}{2}\beta\} dw$$

This is integration on a Gamma Function:

$$= \frac{\Gamma(\alpha)}{b^\alpha} \propto b^{-\alpha} = [k(\theta-\mu)^2+\eta]^{-\frac{d+1}{2}}$$

$$= \left[1 + \frac{1}{d} \frac{(\theta-\mu)^2}{(\frac{\eta}{dk})}\right]^{-\frac{d+1}{2}}$$

$\therefore$ center: $m = \mu$

degree of freedom: $\nu = d$ of $\left[1 + \frac{1}{\nu} \cdot \frac{(\cancel{\theta}-m)^2}{s^2}\right]^{-\frac{\nu+1}{2}}$

scale parameter: $s^2 = \frac{\eta}{dk}$

B). Bayesian theory:

$$p(\theta,w|y) \propto p(y|\theta,w)\, p(\theta,w)$$

$$p(\theta,w) \sim w^{\frac{d+1}{2}-1} exp\{-w\cdot\frac{k(\theta-\mu)^2}{2}\} exp\{-w\cdot\frac{\eta}{2}\}$$

$$p(y_i|\theta,w) \sim N(\theta, w^{-1}) \Rightarrow p(y|\theta,w) \propto \prod_{i=1}^{n} w^{\frac{1}{2}} exp\{-\frac{w}{2}(y_i-\theta)^2\}$$

$$= w^{n/2} exp\{-\frac{w}{2}\sum_{i=1}^{n}(y_i-\theta)^2\}$$

$$= w^{n/2} exp\{-\frac{w}{2}\sum_{i=1}^{n}(y_i-\bar{y}+\bar{y}-\theta)^2\}$$

$$= w^{n/2} exp\{-\frac{w}{2}[S_y + n(\bar{y}-\theta)^2]\},\quad S_y=\sum_{i=1}^{n}(y_i-\bar{y})^2$$

$$p(\theta,w|y) \propto w^{n/2} exp\{-\frac{w}{2}[S_y+n\cdot(\bar{y}-\theta)^2]\} w^{\frac{d+1}{2}-1} exp\{-w\cdot\frac{k(\theta-\mu)^2}{2}\} exp\{-w\cdot\frac{\eta}{2}\}$$

$$= w^{\frac{n+d+1}{2}-1} exp\{-\frac{w}{2}[S_y+n(\bar{y}-\theta)^2+k(\theta-\mu)^2]\} exp\{-w\cdot\frac{\eta}{2}\}$$

$$= w^{\frac{n+d+1}{2}-1} exp\{-\frac{w}{2}[(n+k)\theta^2-2\cdot(n\bar{y}+k\mu)\theta+(n\bar{y}^2+k\mu^2)]\} exp\{-\frac{w}{2}(S_y+\eta)\}$$

$$= w^{\frac{n+d+1}{2}-1} exp\{-\frac{w}{2}(n+k)\left[\theta-\left(\frac{n\bar{y}+k\mu}{n+k}\right)\right]^2\} exp\{-\frac{w}{2}[S_y+\eta-\frac{nk}{(n+k)}(\mu-\bar{y})^2]\} \quad \cdots (1)$$

$\therefore\ d^* = n+d,\quad \mu^* = \frac{k\mu+n\bar{y}}{n+k},\quad k^* = n+k,\quad \eta^* = \eta+S_y-\frac{nk}{n+k}(\mu-\bar{y})^2$

$$= w^{\frac{d^*+1}{2}-1} exp\{-\frac{wk^*}{2}(\theta-\mu^*)^2\} exp\{-\frac{w\eta^*}{2}\}$$

//

C). According to (B), then:

$$p(\theta|y,w) \propto exp\{-\frac{wk^*}{2}(\theta-\mu^*)^2\} \Rightarrow p(\theta|y,w) \sim N(\mu^*, (wk^*)^{-1})$$

D). Following (B), we know:

$$p(\theta,w|y) \propto w^{\frac{d^*+1}{2}-1} \exp\{-\frac{wk^*}{2}(\theta-\mu^*)^2\} \exp\{-\frac{w\eta^*}{2}\}$$

$$p(w|y) = \int_\theta p(w,\theta|y)\,d\theta$$

$$\propto w^{\frac{d^*+1}{2}-1} \exp\{-\frac{w\eta^*}{2}\} \underbrace{\int_\theta \exp\{-\frac{wk^*}{2}(\theta-\mu^*)^2\}}_{\to \text{ integration on normal function}}$$

$$\propto w^{\frac{d^*+1}{2}-1} \exp\{-\frac{w\eta^*}{2}\} \cdot \frac{\sqrt{2\pi}}{(wk^*)^{\frac{1}{2}}}$$

$$\propto w^{\frac{d^*}{2}-1} \exp\{-\frac{w\eta^*}{2}\} \qquad \text{(Gamma kernel)}$$

$$\Rightarrow p(w|y) \sim Gamma(\frac{d^*}{2}, \frac{\eta^*}{2})$$

E). $p(w|y) \sim Gammal(\frac{d^*}{2}, \frac{\eta^*}{2})$ → from (D)

$p(\theta|y,w) \sim N(\mu^*, (wk^*)^{-1})$ → from (C), what is $p(\theta|y)$? $\left[1+\frac{1}{v}\cdot\frac{(\theta-m)^2}{s^2}\right]^{-\frac{v+1}{2}}$

According to (A), the update of posterior distribution parameters are:

$$\begin{cases} \text{center: } m = \mu^* \\ \text{degree of freedom: } v = d^* \\ \text{scale parameter: } s^2 = \frac{\eta^*}{(d^* k^*)} \end{cases}$$

μ^*, d^*, k^*, η^* are on page ② from (B)

F) ① $p(w) \sim Gamma(\frac{d}{2}, \frac{\eta}{2}) \propto w^{\frac{d}{2}-1} e^{-\frac{\eta}{2}w}$

$\lim_{d,\eta\to 0} p(w) \propto w^{-1}$ $\qquad \int_{w^0}^{+\infty} p(w) \to +\infty$ $\qquad$ not pdf

② $p(\theta|w) \sim N(\mu,(wk)^{-1}) \propto (wk) \exp\{-\frac{wk}{2}(\theta-\mu)^2\}$

$\lim_{k\to 0} p(\theta|w) \propto 0$ $\qquad$ not pdf

$$\boxed{\begin{aligned} d^* &= n+d \\ K^* &= n+K \\ \mu^* &= (K\mu + n\bar{y})/(n+K) \\ \eta^* &= \eta + S_y - \frac{nK}{n+K}(\mu - \bar{y})^2 \end{aligned}}$$

(G) According to (E)

$$p(\theta|y) \propto \left[1 + \frac{1}{d^*} \cdot \frac{(\theta - \mu^*)^2}{\left(\frac{\eta^*}{d^* K^*}\right)} \right]^{-\frac{d^*+1}{2}}$$

$$\lim_{d,K,\mu,\eta \to 0} p(\theta|y) \propto \left[1 + \frac{n}{S_y}(\theta - \bar{y})^2 \right]^{-\frac{n+1}{2}}$$

(student dist) ✓ (t distribution)

According to (D)

$$p(w|y) \sim \text{Gamma}\left(\frac{d^*}{2}, \frac{\eta^*}{2}\right) \propto w^{\frac{d^*}{2}-1} \exp\left\{-w\frac{\eta^*}{2}\right\}$$

useful in hyperparameters

$p(\theta|w) \propto 1$ flat

$p(w) \propto \frac{1}{w}$

(improper prior ⇕ think limit)

"objective" priors

Jeffery's prior

reference prior

MDL prior

maximum Entropy prior

$$\lim_{d,\eta \to 0} p(w|y) \propto w^{\frac{n}{2}-1} \exp\left\{-w \cdot \frac{S_y}{2}\right\}$$

✓ (Gamma dist)

(H) $p(\theta|y) \propto \left[1 + \frac{1}{v} \cdot \frac{(\theta - m)^2}{s^2} \right]^{-\frac{v+1}{2}}$, $m = \mu^*$, $v = d^*$, $s^2 = \frac{\eta^*}{d^* K^*}$

$$\eta, d, K \to 0 \implies \begin{aligned} m &= \mu^* = (K\mu + n\bar{y})/(n+K) \to \bar{y} \\ v &= d^* = n+K \to n \\ s^2 &= \frac{\eta^*}{d^* K^*} \to \frac{S_y}{n^2} = \frac{Var(y)}{n} \end{aligned}$$

Credible interval of θ:

$$\theta \in \bar{y} \pm t^* s = \bar{y} \pm t^* \cdot \sqrt{\frac{Var(y)}{n}}$$

Same as classical freq. CI.

The conjugate Gaussian Linear Model

A). Derive $p(\beta|y,\omega)$

$\omega \sim \Gamma(\frac{d}{2},\frac{\eta}{2})$, $(\beta|\omega) \sim N(m,(\omega K)^{-1})$, $(y|\beta,\omega) \sim N(X\beta,(\omega\Lambda)^{-1})$

$$p(\beta,\omega|y) \propto p(y|\beta,\omega)\,p(\beta,\omega) = p(y|\beta,\omega)\,p(\beta|\omega)\,p(\omega)$$

$$\propto \omega^{\frac{n}{2}} \exp\{-\tfrac{1}{2}(y-X\beta)^T(\omega\Lambda)(y-X\beta)\}\, \omega^{\frac{p}{2}} \exp\{-\tfrac{1}{2}(\beta-m)^T(\omega K)(\beta-m)\}\, \omega^{\frac{d}{2}-1}\exp\{-\tfrac{\eta}{2}\omega\}$$

$$= \omega^{\frac{n+p+d}{2}-1} \exp\{-\tfrac{\omega}{2}[(y-X\beta)^T\Lambda(y-X\beta) + (\beta-m)^T K(\beta-m) + \eta)]\}$$

$$= \omega^{\frac{n+p+d}{2}-1} \exp\{-\tfrac{1}{2}(\beta-m^*)^T(\omega K^*)(\beta-m^*)\}\cdot \exp\{-\tfrac{\omega}{2}[y^T\Lambda y + m^T K m + \eta - m^{*T} K^* m^*]\}$$

$K^* = X^T\Lambda X + K$, $m^* = (K^*)^{-1}(X^T\Lambda y + Km)$,

Normal - Gamma distribution

$$= \underbrace{\omega^{\frac{p}{2}} \exp\{-\tfrac{1}{2}(\beta-m^*)^T(\omega K^*)(\beta-m^*)\}}_{\text{Normal } p(\beta|y,\omega)}\ \underbrace{\omega^{\frac{d^*}{2}-1}\exp\{-\tfrac{\eta^*}{2}\omega\}}_{\text{Gamma } p(\omega|y)}$$

$d^* = n+d$, $\eta^* = \eta + y^T\Lambda y + m^T K m - m^{*T} K^* m^*$

$\Rightarrow p(\beta|y,\omega) \sim N(m^*,(\omega K^*)^{-1})$ (A) ✓

$p(\omega|y) = \int_\beta p(\omega,\beta|y) \sim \Gamma(\frac{d^*}{2},\frac{\eta^*}{2})$ (B) ✓

(C). Derive $p(\beta|y)$?

$$p(\beta|y) = \int_\omega p(\omega,\beta|y) \propto \int_\omega \omega^{\frac{p}{2}} \exp\{-\tfrac{1}{2}(\beta-m^*)^T(\omega K^*)(\beta-m^*)\}\,\omega^{\frac{d^*}{2}-1}\exp\{-\tfrac{\eta^*}{2}\omega\}$$

$$= \int_\omega \omega^{\frac{p+d^*}{2}-1}\exp\{-\omega[\tfrac{1}{2}(\beta-m^*)^T K^*(\beta-m^*) + \tfrac{\eta^*}{2}]\}$$ (integrate on Γ distr)

$$\propto [\eta^* + (\beta-m^*)^T K^*(\beta-m^*)]^{-\frac{p+d^*}{2}}$$

$$\propto \left[1 + \frac{1}{d^*}(\beta-m^*)^T \frac{K^*}{(\frac{\eta^*}{d^*})}(\beta-m^*)\right]^{-\frac{d^*+p}{2}}$$

$$\boxed{\begin{aligned} &p(x) \sim N(\mu,\Sigma) \\ &\Leftrightarrow p(x) = \left(\frac{1}{2\pi|\Sigma|}\right)^{\frac{1}{2}} \exp\{-\tfrac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\} \end{aligned}}$$