Advanced Bayesian Methods - Assignment 4

통계데이터사이언스학과
박주연 (2022311137)

Q1

(a) $y_i = \mu(x_i) + \varepsilon_i \;,\; \varepsilon_i \sim N(0, \sigma^2)$ for $i = 1, \dots, n$, $\mu \sim GP(0, K)$

As vector notation,

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n, \quad \bar{\mu} = \begin{pmatrix} \mu(x_1) \\ \vdots \\ \mu(x_n) \end{pmatrix} \in \mathbb{R}^n, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \in \mathbb{R}^n, \quad y \sim N(\bar{\mu}, \sigma^2 I_n)$$

Since joint distribution of $\bar{\mu}$ and $\tilde{\mu}$ is given as

$$\bar{\mu}, \tilde{\mu} \sim N_{n+m}\left(\begin{pmatrix} 0_n \\ 0_m \end{pmatrix}, \begin{pmatrix} K(x,x) & K(x,\tilde{x}) \\ K(\tilde{x},x) & K(x,\tilde{x}) \end{pmatrix} \right)$$

the conditional distribution of $\tilde{\mu}$ given $\bar{\mu}$ is

$$\tilde{\mu} \mid \bar{\mu} \sim N_m\left(K(\tilde{x},x) K(x,x)^{-1} \bar{\mu} \,,\; K(\tilde{x},\tilde{x}) - K(\tilde{x},x) K(x,x)^{-1} K(x,\tilde{x}) \right)$$

① $$E\left(\begin{matrix} y \\ \bar{\mu} \\ \tilde{\mu} \end{matrix} \,\middle|\, \sigma^2 \right) = E\left(E\left[(y, \bar{\mu}, \hat{\mu}) \mid \bar{\mu}, \sigma^2 \right] \mid \sigma^2 \right)$$

$$= E\left(\begin{matrix} E(y \mid \bar{\mu}, \sigma^2) \\ E(\bar{\mu} \mid \bar{\mu}, \sigma^2) \\ E(\tilde{\mu} \mid \bar{\mu}, \cancel{\sigma^2}) \end{matrix} \,\middle|\, \sigma^2 \right) = E\left(\begin{matrix} \bar{\mu} \\ \bar{\mu} \\ K(\tilde{x},x) K(x,x)^{-1} \bar{\mu} \end{matrix} \,\middle|\, \cancel{\sigma^2} \right) = 0_{2n+m}$$

independent

② $$E\left(\begin{matrix} y \\ \bar{\mu} \\ \tilde{\mu} \end{matrix} \,\middle|\, \bar{\mu}, \sigma^2 \right) = \begin{pmatrix} E(y \mid \bar{\mu}, \sigma^2) \\ E(\bar{\mu} \mid \bar{\mu}, \cancel{\sigma^2}) \\ E(\tilde{\mu} \mid \bar{\mu}, \cancel{\sigma^2}) \end{pmatrix} = \begin{pmatrix} \bar{\mu} \\ \bar{\mu} \\ K(\tilde{x},x) K(x,x)^{-1} \bar{\mu} \end{pmatrix}$$

$\because \bar{\mu} \perp\!\!\!\perp \sigma^2 ,\; \tilde{\mu} \perp\!\!\!\perp \sigma^2$

$\because E(\bar{\mu} \mid \bar{\mu}, \sigma^2) = E(\bar{\mu} \mid \bar{\mu}) = \bar{\mu}$ ($\because$ constant)

③ $$\operatorname{cov}\left(\begin{matrix} y \\ \bar{\mu} \\ \tilde{\mu} \end{matrix} \,\middle|\, \bar{\mu}, \sigma^2 \right) = \begin{pmatrix} \operatorname{cov}(y \mid \bar{\mu}, \sigma^2) & \operatorname{cov}(y, \bar{\mu} \mid \bar{\mu}, \sigma^2) & \operatorname{cov}(y, \tilde{\mu} \mid \bar{\mu}, \sigma^2) \\ \operatorname{cov}(\bar{\mu}, y \mid \bar{\mu}, \sigma^2) & \operatorname{cov}(\bar{\mu} \mid \bar{\mu}, \sigma^2) & \operatorname{cov}(\bar{\mu}, \hat{\mu} \mid \bar{\mu}, \sigma^2) \\ \operatorname{cov}(\tilde{\mu}, y \mid \bar{\mu}, \sigma^2) & \operatorname{cov}(\tilde{\mu}, \bar{\mu} \mid \bar{\mu}, \sigma^2) & \operatorname{cov}(\tilde{\mu} \mid \bar{\mu}, \sigma^2) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2 I_n & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & K(\tilde{x},\tilde{x}) - K(\tilde{x},x) K(x,x)^{-1} K(x,\tilde{x}) \end{pmatrix}$$

$$\begin{aligned} \because \operatorname{cov}(\bar{\mu}, y \mid \bar{\mu}, \sigma^2) &= \operatorname{cov}(\bar{\mu}, \bar{\mu} + \varepsilon \mid \bar{\mu}, \sigma^2) \\ &= \operatorname{cov}(\bar{\mu}, \bar{\mu} \mid \bar{\mu}, \sigma^2) + \underbrace{\operatorname{cov}(\bar{\mu}, \varepsilon \mid \bar{\mu}, \sigma^2)}_{=0 \text{ (indep)}} \\ &= 0 \quad (\because \bar{\mu} \text{ is constant given } \bar{\mu}) \end{aligned}$$

$$\begin{aligned} \operatorname{cov}(\tilde{\mu}, y \mid \bar{\mu}, \sigma^2) &= \operatorname{cov}(\tilde{\mu}, \bar{\mu} + \varepsilon \mid \bar{\mu}, \sigma^2) \\ &= \operatorname{cov}(\tilde{\mu}, \bar{\mu} \mid \bar{\mu}, \sigma^2) + \underbrace{\operatorname{cov}(\tilde{\mu}, \varepsilon \mid \bar{\mu}, \sigma^2)}_{=0 \text{ (indep)}} \\ &= 0 \end{aligned}$$

$$\operatorname{cov}(\tilde{\mu}, \bar{\mu} \mid \bar{\mu}, \cancel{\sigma^2}) = 0$$

④ $$\operatorname{cov}\left(E\left[\begin{pmatrix} y \\ \bar{\mu} \\ \tilde{\mu} \end{pmatrix} \middle| \bar{\mu}, \sigma^2\right] \middle| \sigma^2\right) = \operatorname{cov}\left[\begin{pmatrix} \bar{\mu} \\ \bar{\mu} \\ K(\tilde{x},x)K(x,x)^{-1}\bar{\mu} \end{pmatrix} \middle| \sigma^2\right] \quad (\text{by } ②)$$

$$= \begin{pmatrix} K(x,x) & K(x,x) & K(x,\tilde{x}) \\ K(x,x) & K(x,x) & K(x,\tilde{x}) \\ K(\tilde{x},x) & K(\tilde{x},x) & K(\tilde{x},x)K(x,x)^{-1}K(x,\tilde{x}) \end{pmatrix}$$

$$* \operatorname{cov}(\bar{\mu}, K(\tilde{x},x)K(x,x)^{-1}\bar{\mu}) = \operatorname{cov}(\bar{\mu})K(x,x)^{-1}K(x,\tilde{x})$$
$$= \cancel{K(x,x)K(x,x)^{-1}}K(x,\tilde{x})$$
$$= K(x,\tilde{x})$$

$$* \operatorname{cov}(K(\tilde{x},x)K(x,x)^{-1}\bar{\mu}) = K(\tilde{x},x)K(x,x)^{-1}\cancel{\operatorname{cov}(\bar{\mu})K(x,x)^{-1}}K(x,\tilde{x})$$
$$= K(\tilde{x},x)K(x,x)^{-1}K(x,\tilde{x})$$

⑤ $$E\left(\operatorname{cov}\left[\begin{pmatrix} y \\ \bar{\mu} \\ \tilde{\mu} \end{pmatrix} \middle| \bar{\mu}, \sigma^2\right] \middle| \sigma^2\right) = E\left[\begin{pmatrix} \sigma^2 I_n & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & K(\tilde{x},\tilde{x}) - K(\tilde{x},x)K(x,x)^{-1}K(x,\tilde{x}) \end{pmatrix} \middle| \sigma^2\right]$$

$$= \begin{pmatrix} \sigma^2 I_n & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & K(\tilde{x},\tilde{x}) - K(\tilde{x},x)K(x,x)^{-1}K(x,\tilde{x}) \end{pmatrix}$$

Therefore

$$E[(y, \bar{\mu}, \tilde{\mu}) \mid \sigma^2] = 0$$

$$\operatorname{cov}[(y,\bar{\mu},\tilde{\mu}) \mid \sigma^2] = \operatorname{cov}(E[(y,\bar{\mu},\tilde{\mu}) \mid \bar{\mu},\sigma^2] \mid \sigma^2) + E(\operatorname{cov}[(y,\bar{\mu},\tilde{\mu}) \mid \bar{\mu},\sigma^2] \mid \sigma^2)$$

$$= \begin{pmatrix} K(x,x) + \sigma^2 I_n & K(x,x) & K(x,\tilde{x}) \\ K(x,x) & K(x,x) & K(x,\tilde{x}) \\ K(\tilde{x},x) & K(\tilde{x},x) & K(\tilde{x},\tilde{x}) \end{pmatrix}$$

(b)

$$\begin{pmatrix} y \\ \bar{\mu} \\ \tilde{\mu} \end{pmatrix} \middle| \sigma^2 \sim N\left(\begin{pmatrix} 0_n \\ 0_n \\ 0_m \end{pmatrix}, \left(\begin{array}{c|cc} K(x,x)+\sigma^2 I_n & K(x,x) & K(x,\tilde{x}) \\ \hline K(x,x) & K(x,x) & K(x,\tilde{x}) \\ K(\tilde{x},x) & K(\tilde{x},x) & K(\tilde{x},\tilde{x}) \end{array}\right)\right)$$

Marginal distributions of MVN are MVN.

Therefore

$$m_* = E(\mu \mid \sigma^2) = 0_{n+m}$$

$$K_* = \operatorname{var}(\mu \mid \sigma^2) = \begin{pmatrix} K(x,x) & K(x,\tilde{x}) \\ K(\tilde{x},x) & K(\tilde{x},\tilde{x}) \end{pmatrix}$$

(c)

1. Generating simulation data

- Draw $x = \{x_1, \cdots, x_{200}\}$ where $x_i \sim Unif(0,1)$ $i = 1, \cdots, 200$
- Draw $y = \{y_1, \cdots, y_{200}\}$ where $y_i \sim N(\mu(x_i), 0.1^2)$ with $\mu(x_i) = \sin^3(2\pi x^3)$
- Draw $\tilde{x} = \{\tilde{x}_1, \cdots, \tilde{x}_{999}\} = \{1/1000, \cdots, 999/1000\}$

2. Calculate posterior distribution $p(\sigma^2 | y)$ using grid sampling and choose suitable range

① Grid Sampling for some range

② Calculate unnormalized log posterior $\log p(\sigma^2 | y)$

- Likelihood (marginal MVN)

$$p(y | \sigma^2) \sim N_n(0, K(x,x) + \sigma^2 I_n)$$

- Prior (Jeffrey's)

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

- log posterior

$$p(\sigma^2 | y) \propto p(\sigma^2)\, p(y | \sigma^2)$$

$$\log p(\sigma^2 | y) \propto \log p(\sigma^2) + \log p(y | \sigma^2)$$

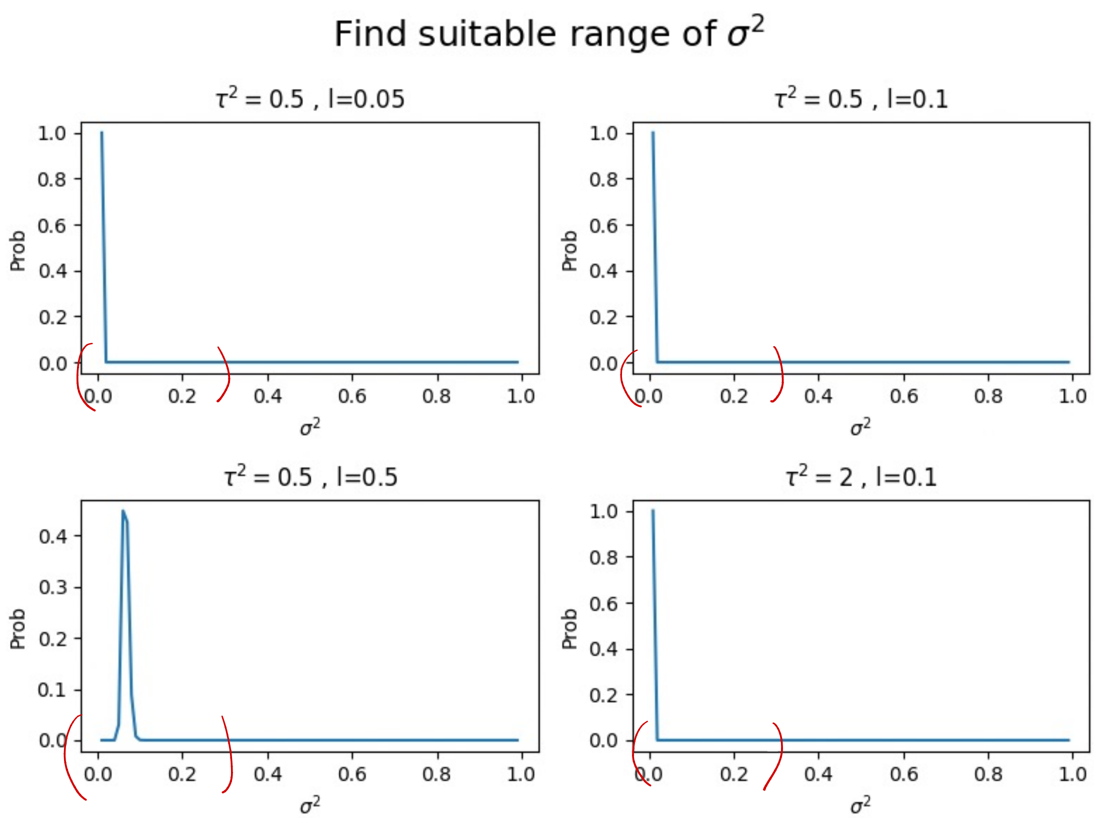
$$\propto \log N(y | 0, K(x,x) + \sigma^2 I_n) - \log \sigma^2$$

③ Calculate normalized posterior probability

$$Pr(\sigma^2 = \sigma_i^2 | y) = \frac{p(y | \sigma^2 = \sigma_i^2)\, p(\sigma_i^2)}{\sum_k p(y | \sigma^2 = \sigma_k^2)\, p(\sigma_k^2)}$$

④ Choose suitable range for σ^2

As a result of drawing posterior distribution for various l and τ^2, the range of σ^2 is appropriate from 0 to 0.25



3. Draw $\tilde{\mu}$ from $p(\tilde{\mu} \mid y, \sigma^2)$

① Define covariance matrix $k(x, x)$, $k(x, \tilde{x})$, $k(\tilde{x}, x)$ and $k(\tilde{x}, \tilde{x})$

② Calculate posterior distribution $p(\sigma^2 \mid y)$ (use result of step 2)

③ Draw $\tilde{\mu}$ and σ^2 from posterior (jointly)

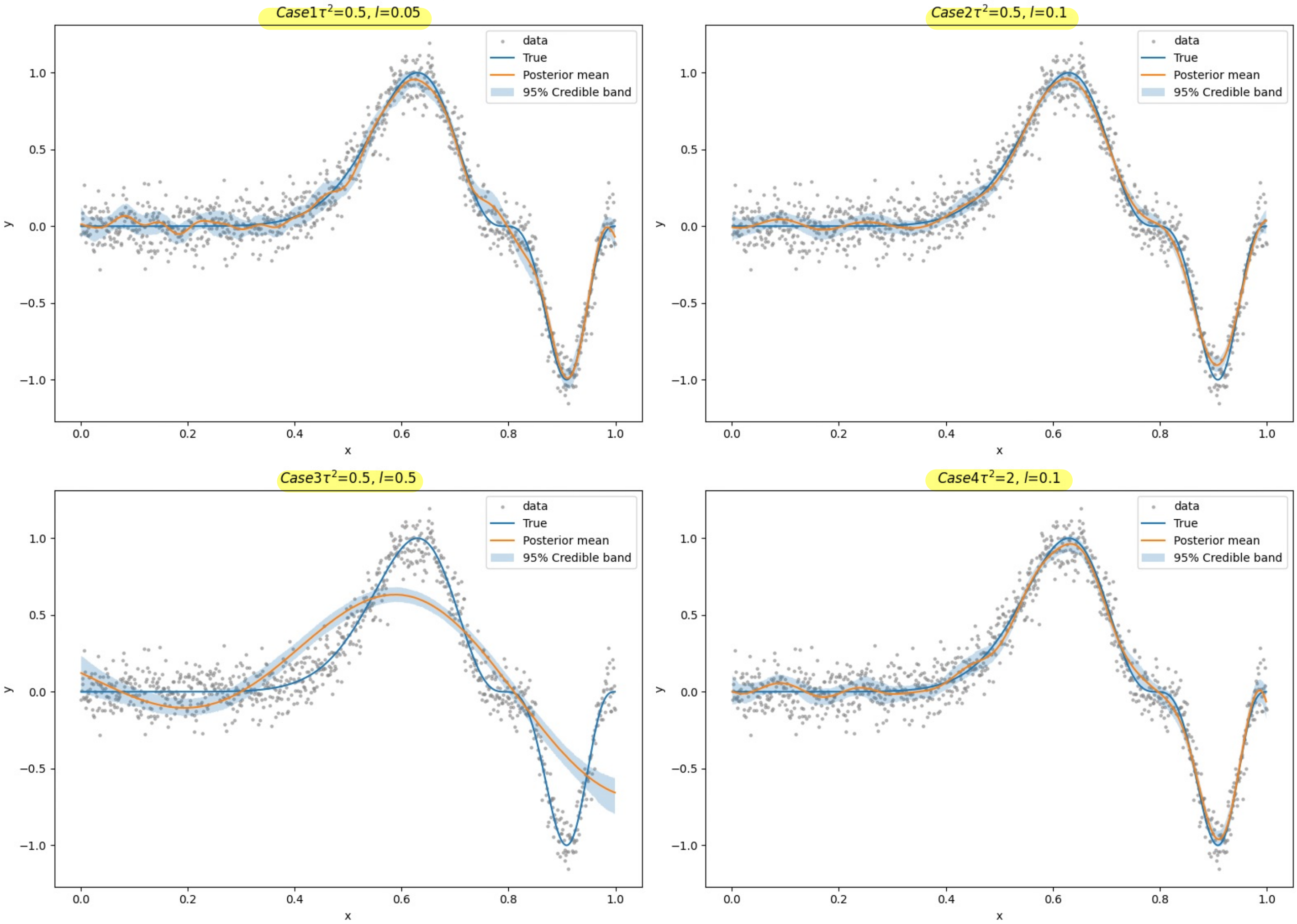
- Draw $\sigma^2 \sim p(\sigma^2 \mid y)$
- plug-in σ^2 and draw $\tilde{\mu} \sim p(\tilde{\mu} \mid y, \sigma^2)$

④ Repeat ③ until converge

4. Result

When comparing case 1 and 3, higher τ^2 led to oversmootness to density estimation. Although parameter l does not affect as much as τ^2, high value of l also led to high correlation between observations.

Nonparametric regression GP prior with different values of $\tau^2, \tilde{\mu}$ GP prior



Q2

(a)

log prior

$$\log P(\mu_h, \tau_h^2) = \log \underbrace{P(\mu_h \mid \tau_h^2)}_{= N(\mu_h \mid \mu_0,\, k\tau_h^2)} + \log P(\tau_h^2) \longrightarrow = IG(\tau_h^2 \mid a_\tau, b_\tau) \propto (\tau_h^2)^{-a_\tau - 1} \cdot \exp(-b_\tau (\tau_h^2)^{-1})$$

$$\propto (k\tau_h^2)^{-1/2} \exp\left(-\frac{1}{2k\tau_h^2}(\mu_h - \mu_0)^2\right)$$

$$= -\frac{1}{2}\log(k\tau_h^2) - \frac{1}{2k\tau_h^2}(\mu_h - \mu_0)^2 - (a_\tau + 1)\log \tau_h^2 - b_\tau(\tau_h^2)^{-1}$$

loglikelihood

$$\log P(y_i \mid z_i, \mu_{z_i}, \tau_{z_i}) = \log N(y_i \mid \mu_{z_i}, \tau_{z_i}^2)$$

$$\propto -\frac{1}{2}\log \tau_{z_i}^2 - \frac{1}{2\tau_{z_i}^2}(y_i - \mu_{z_i})^2$$

conditional posterior

$$p(\mu_h, \tau_h^2 \mid -) = p(\mu_h, \tau_h^2 \mid y, \pi, z, \mu_{-h}, \tau_{-h}^2)$$

$$\propto p(\mu_h, \tau_h^2)\, p(y \mid z, \mu_h, \tau_h^2, \mu_{-h}, \tau_{-h}^2)$$

$$= p(\mu_h, \tau_h^2) \prod_{i=1}^{n} p(y_i \mid z_i, \mu_{z_i}, \tau_{z_i}^2)$$

$$\propto p(\mu_h, \tau_h^2) \prod_{i: z_i = h} p(y_i \mid z_i, \mu_h, \tau_h^2)$$ (leave terms only related to component h)

log conditional posterior distribution

$$\log p(\mu_h, \tau_h^2 \mid -) \propto \log p(\mu_h, \tau_h^2) + \sum_{i: z_i = h} \log p(y_i \mid z_i, \mu_h, \tau_h^2)$$

$$\propto -\frac{1}{2}\log(k\tau_h^2) - \frac{1}{2k\tau_h^2}(\mu_h - \mu_0)^2 - (a_\tau + 1)\log\tau_h^2 - b_\tau(\tau_h^2)^{-1} + \sum_{i: z_i = h}\left\{-\frac{1}{2}\log \tau_h^2 - \frac{1}{2\tau_h^2}(y_i - \mu_h)^2\right\}$$

① Find $p(\mu_h \mid -)$

From log joint posterior distribution, the quadratic terms in μ_h can be collected together :

$$-\frac{1}{2}\log(k\tau_h^2) - \frac{1}{2k\tau_h^2}(\mu_h - \mu_0)^2 - (a_\tau + 1)\log\tau_h^2 - b_\tau(\tau_h^2)^{-1} + \sum_{i: z_i = h}\left\{-\frac{1}{2}\log \tau_h^2 - \frac{1}{2\tau_h^2}(y_i - \mu_h)^2\right\}$$

$$\propto -\frac{1}{2k\tau_h^2}(\mu_h - \mu_0)^2 - \sum_{i: z_i = h}\frac{1}{2\tau_h^2}(y_i - \mu_h)^2$$

$$= -\frac{1}{2k\tau_h^2}(\mu_h - 2\mu_0\mu_h + \mu_0^2) - \frac{1}{2\tau_h^2}\sum_{i: z_i = h}\{\mu_h^2 - 2y_i\mu_h + y_i^2\}$$

$$\propto -\frac{1}{2k\tau_h^2}(\mu_h^2 - 2\mu_0\mu_h) - \frac{1}{2\tau_h^2}(n_h\mu_h^2 - 2\mu_h \cdot n_h\bar{y}_h) \quad \text{where } \bar{y}_h = n_h^{-1}\sum_{i: z_i = h} y_i$$

$$= -\frac{1}{2\tau_h^2}\left[(k^{-1} + n_h)\mu_h^2 - 2(k^{-1}\mu_0 + n_h\bar{y}_h)\mu_h\right]$$

$$= -\frac{1}{2}\left[\tau_h^2(k^{-1} + n_h)^{-1}\right]^{-1}\left[\mu_h^2 - 2(k^{-1} + n_h)^{-1}(k^{-1}\mu_0 + n_h\bar{y}_h)\mu_h\right] \quad : \text{Quadratic form in } \mu_h$$

$$\sim N(\mu_h \mid (k^{-1} + n_h)^{-1}(k^{-1}\mu_0 + n_h\bar{y}_h),\ (k^{-1} + n_h)^{-1}\tau_h^2)$$

$$\therefore p(\mu_h \mid -) = N(\mu_h \mid \hat{\mu}_h, \hat{k}_h\tau_h^2) \quad \text{where } \hat{k}_h = (k^{-1} + n_h)^{-1} \text{ and } \hat{\mu}_h = \hat{k}_h(k^{-1}\mu_0 + n_h\bar{y}_h)$$

② Find $p(\tau_h^2 \mid -\ )$ (except μ_h)

From the log joint posterior dist'n ,

$$-\frac{1}{2}\log(k\tau_h^2) - \frac{1}{2k\tau_h^2}(\mu_h - \mu_0)^2 - (a_\tau + 1)\log\tau_h^2 - b_\tau(\tau_h^2)^{-1} + \sum_{i: z_i = h}\left\{\underbrace{-\frac{1}{2}\log \tau_h^2}_{= -\frac{n_h}{2}\log\tau_h^2} - \underbrace{\frac{1}{2\tau_h^2}(y_i - \mu_h)^2}_{= -\frac{1}{2\tau_h^2}\sum_{i: z_i = h}(y_i - \mu_h)^2}\right\}$$

$$\propto \left(-\frac{1}{2} - (a_\tau + n_h/2 + 1)\right)\cdot\log\tau_h^2 \times \frac{-1}{\tau_h^2}\left[b_\tau + \frac{1}{2}\left\{k^{-1}(\mu_h - \mu_0)^2 + \sum_{i: z_i = h}(y_i - \mu_h)^2\right\}\right]$$

$$* \ k^{-1}(\mu_h - \mu_0)^2 + \sum_{i: z_i = h}(y_i - \mu_h)^2 = k^{-1}(\mu_h^2 - 2\mu_0\mu_h + \mu_0^2) + \sum_{i: z_i = h}(y_i^2 - 2\mu_h y_i^2 + \mu_h^2)$$

$$= (k^{-1} + n_h)\mu_h^2 - 2\left(k^{-1}\mu_0 + \sum_{i: z_i = h} y_i\right)\mu_h + k^{-1}\mu_0^2 + \sum_{i: z_i = h} y_i^2$$

since $N(\mu_h \mid (k^{-1} + n_h)^{-1}(k^{-1}\mu_0 + n_h\bar{y}_h),\ (k^{-1} + n_h)^{-1}\tau_h)$

$$= (k^{-1} + n_h)\left[\mu_h^2 - 2(k^{-1}\mu_0 + \Sigma y_i)(k^{-1} + n_h)^{-1}\mu_h + (k^{-1} + n_h)^{-1}(k^{-1}\mu_0^2 + \Sigma y_i^2)\right]$$

$$+\left((k^{-1}\mu_0 + n_h\bar{y})(k^{-1} + n_h)^{-1}\right)^2 - \left((k^{-1}\mu_0 + n_h\bar{y})(k^{-1} + n_h)^{-1}\right)^2$$

$$= (k^{-1} + n_h)\underbrace{\left[\mu_h - (k^{-1}\mu_0 + \Sigma y_i)(k^{-1} + n_h)^{-1}\right]^2}_{\text{quadratic in } \mu_h} - (k^{-1}\mu_0 + \Sigma y_i)^2(k^{-1} + n_h)^{-1} + (k^{-1}\mu_0^2 + \Sigma y_i^2)$$

$\propto \log N(\mu_h \mid (\kappa^{-1}+n_h)^{-1}(\kappa^{-1}\mu_0 + n_h\bar{y}_h), (\kappa^{-1}+n_h)^{-1}\tau_h^2) \times (-a_\tau - n_h/2 - 1)\log\tau_h^2 \left(-\frac{1}{\tau_h^2}\right)[b_h + \frac{1}{2}\{-(\kappa^{-1}\mu_0 + n_h\bar{y}_h)^2(\kappa^{-1}+n_h)^{-1} + \kappa^{-1}\mu_0^2 + \Sigma y_i^2\}]$

$$
\begin{aligned}
* \; & -(\kappa^{-1}\mu_0 + n_h\bar{y}_h)^2(\kappa^{-1}+n_h)^{-1} + \kappa^{-1}\mu_0^2 + \Sigma_{i:z_i=h} y_i^2 \\
&= -\frac{1}{\kappa^2(\kappa^{-1}+n_h)}\mu_0^2 - \frac{2}{\kappa^{-1}+n_h}\kappa^{-1}\mu_0 n_h\bar{y}_h - \frac{1}{\kappa^{-1}+n_h}n_h^2\bar{y}_h^2 + \kappa^{-1}\mu_0^2 + \Sigma_{i:z_i=h} y_i^2 \\
&= -\left(\frac{1}{1+\kappa n_h} - 1\right)\kappa^{-1}\mu_0^2 - 2\frac{n_h}{1+\kappa n_h}\bar{y}_h\mu_0 - \frac{n_h^2}{\kappa^{-1}+n_h}\bar{y}_h^2 + \Sigma y_i^2 \\
&= \frac{n_h}{1+\kappa n_h}\mu_0^2 - 2\frac{n_h}{1+\kappa n_h}\bar{y}_h\mu_0 - \frac{n_h^2}{\kappa^{-1}+n_h}\bar{y}_h^2 + \Sigma y_i^2 \\
&= \frac{n_h}{1+\kappa n_h}[\mu_0^2 - 2\bar{y}_h\mu_0 + \bar{y}_h^2 - \bar{y}_h^2] - \frac{n_h^2}{\kappa^{-1}+n_h}\bar{y}_h^2 + \Sigma y_i^2 \\
&= \frac{n_h}{1+\kappa n_h}(\bar{y}_h - \mu_0)^2 - \frac{n_h}{1+\kappa n_h}\bar{y}_h^2 - \frac{\kappa \cdot n_h^2}{1+\kappa n_h}\bar{y}_h^2 + \Sigma y_i^2 \\
&= \frac{n_h}{1+\kappa n_h}(\bar{y}_h - \mu_0)^2 - \frac{n_h(1+\kappa n_h)}{1+\kappa n_h}\bar{y}_h^2 + \Sigma y_i^2 \\
&= \frac{n_h}{1+\kappa n_h}(\bar{y}_h - \mu_0)^2 - \underline{n_h\bar{y}_h^2 + \Sigma y_i^2} \\
&\qquad\qquad = \Sigma y_i^2 - n_h\bar{y}_h^2 - n_h\bar{y}_h^2 + n_h\bar{y}_h^2 \\
&\qquad\qquad = \Sigma y_i^2 - 2n_h\bar{y}_h^2 + n_h\bar{y}_h^2 \\
&\qquad\qquad = \Sigma y_i^2 - 2\Sigma y_i\bar{y}_h + \Sigma\bar{y}_h^2 \\
&\qquad\qquad = \Sigma_{i:z_i=h}(y_i - \bar{y}_h)^2 \\
&= \frac{n_h}{1+\kappa n_h}(\bar{y}_h - \mu_0)^2 + \Sigma_{i:z_i=h}(y_i - \bar{y}_h)^2
\end{aligned}
$$

$\propto \log N(\mu_h \mid (\kappa^{-1}+n_h)^{-1}(\kappa^{-1}\mu_0 + n_h\bar{y}_h), (\kappa^{-1}+n_h)^{-1}\tau_h^2) \times \underbrace{\left[\{-(a_\tau - n_h/2) - 1\}\log\tau_h^2 \left(-\frac{1}{\tau_h^2}\right)\left[b_\tau + \frac{1}{2}\left\{\frac{n_h}{1+\kappa n_h}(\bar{y}_h - \mu_0)^2 + \Sigma_{i:z_i=h}(y_i - \bar{y}_h)^2\right\}\right]\right]}_{\text{log pdf of Inv-Gamma}}$

$\log\left[(\tau_h^2)^{-(a_\tau + n_h/2)-1} \exp\left[\left\{b_\tau + \frac{1}{2}\left(\Sigma_{i:z_i=h}(y_i - \bar{y}_h)^2 + \frac{n_h}{1+\kappa n_h}(\bar{y}_h - \mu_0)^2\right)\right\}(\tau_h^2)^{-1}\right]\right]$

$\therefore$ $p(\tau_h^2 \mid -)$ except τ_h^2 is

$\text{Inv-Gamma}(\tau_h^2 \mid \hat{a}_{\tau_h}, \hat{b}_{\tau_h})$ where $\hat{a}_{\tau_h} = a_\tau + n_h/2$, $\hat{b}_{\tau_h} = b_\tau + \frac{1}{2}\left(\Sigma_{i:z_i=h}(y_i - \bar{y}_h)^2 + \frac{n_h}{1+\kappa n_h}(\bar{y}_h - \mu_0)^2\right)$

Therefore, $p(\mu_h, \tau_h^2 \mid -) = N(\mu_h \mid \hat{\mu}_h, \hat{\kappa}_h\tau_h^2)\,\text{Inv-Gamma}(\tau_h^2 \mid \hat{a}_{\tau_h}, \hat{b}_{\tau_h})$

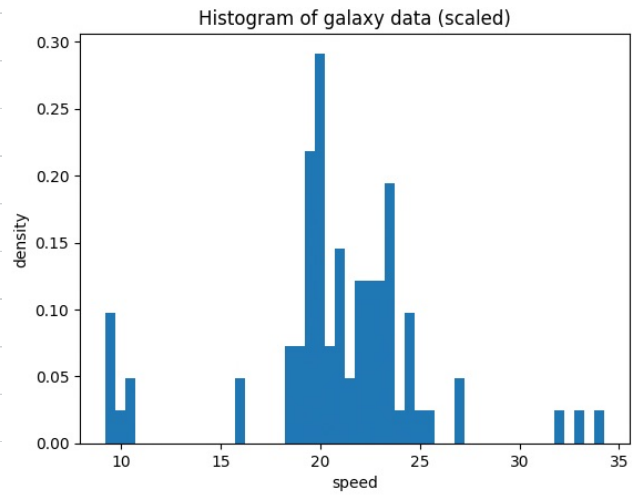
(b) The prior is given as

$$
\begin{aligned}
(\pi_1, \cdots, \pi_H) &\sim \text{Dir}(a, \cdots, a) \\
\mu_h \mid \tau_h^2 &\sim N(\mu_0, \kappa\tau_h^2) \\
\tau_h^2 &\sim \text{Inv-Gamma}(a_\tau, b_\tau) \qquad h = 1, \cdots, H
\end{aligned}
$$

The hyperparametes and H that I use are

$a = 1/H$, $\mu_0 = 0$, $\kappa = 100$, $a_\tau = 1$, $b_\tau = 1$, $K = 25$

It looks like $H = 3$ or 4 in the histogram below. But by letting large H as a upper bound and $a = 1/H$ for parameter of dirichlet prior, data choose the effective number $\hat{H}$.



Result

Since the shape of the distribution varies with H, it should be held large H as an upper bound, and let the data choose the effective number $\hat{H}$.