

Problem class 3: Hypothesis tests ; Inference under model uncertainty ; Hierarchical Bayes

Lecturer: Georgios Karagiannis

georgios.karagiannis@durham.ac.uk

1 Hypothesis test

Exercise 1. (★★) Consider a Bayesian model

$$\begin{cases} x_i | \lambda & \stackrel{\text{iid}}{\sim} \text{Pn}(\lambda), \forall i = 1, \dots, n \\ \lambda & \sim \Pi(\lambda) \end{cases}$$

Hint-1 Poisson distribution has PMF: $\text{Pn}(x|\lambda) = \frac{1}{x!} \lambda^x \exp(-\lambda) 1_{\mathbb{N}}(x)$

Hint-2 Gamma distribution has PDF: $\text{Ga}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) 1_{(0, \infty)}(x)$, with $E(x) = a/b$, $\text{Var}(x) = a/b^2$.

Hint-3 Negative Binomial distribution has PMF: $\text{Nb}(x|r, \theta) = \binom{r+x-1}{r-1} \theta^r (1-\theta)^x 1_{\mathbb{N}}(x)$. with $\theta \in (0, 1)$, $r \in \mathbb{N}$.

Consider that we are interested in testing the hypothesis whether $\lambda = \lambda_0$, (where λ_0 is a fixed known number), or not.

Let $\pi_j = P(H_j)$ be the marginal prior probability of hypothesis H_j .

1. Design the test of hypotheses in Bayesian framework: Namely, set pair of hypotheses, specify priors, and compute the associated Bayes Factor.
2. Compute the posterior probability that $\lambda = \lambda_0$.
3. Perform the hypothesis test to test if $\lambda = 2$ or not based on the Jeffrey's scaling rule, by considering that
 - we have collected two observations $x_1 = 2, x_2 = 3$,
 - a priori the probability that $\{\lambda = 2\}$ is 0.5,
 - given $\{\lambda \neq 2\}$, the prior distr. of λ is a conjugate one with $E(\lambda) = 2$, and $\text{Var}(\lambda) = 1$.

1.

- The pair of hypotheses for this test is

$$\begin{cases} H_0 : & x_i \stackrel{\text{iid}}{\sim} \text{Pn}(\lambda_0 = 2), \text{ for all } i = 1, \dots, n \\ H_1 : & x_i \stackrel{\text{iid}}{\sim} \text{Pn}(\lambda), \lambda > 0 \text{ for all } i = 1, \dots, n \end{cases} \quad (1)$$

where H_0 is a single hypothesis, and H_1 is the general alternative.

- The overall prior can be specified as

$$\pi(\lambda) = \pi_0 1_{\{\lambda_0\}}(\lambda) + (1 - \pi_0) \text{Ga}(\lambda|a, b)$$

for $\pi_0 > 0$, which in this case is $\pi_0 = 1/2$, and $\lambda_0 = 2$.

- Do not get confused that the above notation in H_1 in (1) states $\lambda > 0$. Given H_1 , λ is a continuous random variable. Because λ is a continuous random variable and $\lambda \sim \text{Ga}(a, b)$ given H_1 , the probability that $\lambda = 2$ given on H_1 .

- The Bayes factor is

$$B_{01}(x_{1:n}) = \frac{p_0(x_{1:n})}{p_1(x_{1:n})} = \frac{\prod_{i=1}^n \text{Pn}(x_i|\lambda_0)}{\int \prod_{i=1}^n \text{Pn}(x_i|\lambda) \text{Ga}(\lambda|a, b) d\lambda}$$

where

$$p_0(x_{1:n}) = \prod_{i=1}^n \text{Pn}(x_i|\lambda_0) = \frac{1}{\prod_{i=1}^n x_i!} \lambda_0^{n\bar{x}} \exp(-n\lambda_0)$$

and

$$\begin{aligned} p_1(x_{1:n}) &= \int \prod_{i=1}^n \text{Pn}(x_i|\lambda) \text{Ga}(\lambda|a, b) d\lambda = \frac{1}{\prod_{i=1}^n x_i!} \frac{b^a}{\Gamma(a)} \int_0^\infty \lambda^{n\bar{x}+a-1} \exp(-(n+b)\lambda) d\lambda \\ &= \frac{1}{\prod_{i=1}^n x_i!} \frac{\Gamma(n\bar{x}+a)}{\Gamma(a)} \frac{b^a}{(n+b)^{n\bar{x}+a}} \end{aligned}$$

So

$$\begin{aligned} B_{01}(x_{1:n}) &= \frac{\lambda_0^{n\bar{x}} \exp(-n\lambda_0)}{\frac{b^a \Gamma(n\bar{x}+a)}{\Gamma(a)(n+b)^{n\bar{x}+a}}} = \lambda_0^{n\bar{x}} (n+b)^{n\bar{x}+a} \exp(-n\lambda_0) \frac{1}{b^a} \frac{\Gamma(a)}{\Gamma(n\bar{x}+a)} \\ &= \lambda_0^{n\bar{x}} \exp(-n\lambda_0) \frac{(n+b)^{n\bar{x}+a}}{b^a} \frac{\Gamma(a)}{(n\bar{x}+a-1) \cdots a \Gamma(a)} \\ &= \frac{\lambda_0^{n\bar{x}} \exp(-n\lambda_0)}{(n\bar{x}+a-1) \cdots a} \frac{(n+b)^{n\bar{x}+a}}{b^a} \end{aligned}$$

2. Obviously, for the posterior probability that $\pi(\lambda = \lambda_0|x_{1:n})$, it is

$$\begin{aligned} \pi(\lambda = \lambda_0|x_{1:n}) &= \pi(\mathbf{H}_0|x_{1:n}) = \left(1 + \frac{1 - \pi_0}{\pi_0} \frac{p_1(x_{1:n})}{p_0(x_{1:n})}\right)^{-1} \\ &= \left(1 + \frac{1 - \pi_0}{\pi_0} \frac{b^a (n\bar{x}+a-1) \cdots a}{\lambda_0^{n\bar{x}} (n+b)^{n\bar{x}+a} \exp(-n\lambda_0)}\right)^{-1} \\ &= \frac{\pi_0 \lambda_0^{n\bar{x}} (n+b)^{n\bar{x}+a} \exp(-n\lambda_0)}{\pi_0 \lambda_0^{n\bar{x}} (n+b)^{n\bar{x}+a} \exp(-n\lambda_0) + (1 - \pi_0) b^a (n\bar{x}+a-1) \cdots a} \end{aligned}$$

3. This is actually the aforesaid hypothesis test with $\lambda_0 = 2$.

- Based on the prior information, it is $a = 4$, and $b = 2$ because

$$\left\{ \begin{array}{l} E^{\text{Ga}(a,b)}(\lambda) = 2 \\ \text{Var}^{\text{Ga}(a,b)}(\lambda) = 1 \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} a/b = 2 \\ a/b^2 = 1 \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} a/b = 2 \\ 2/b = 1 \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} a = 4 \\ b = 2 \end{array} \right\}$$

- Based on the sample I have $n\bar{x} = 2 + 3 = 5$, $n = 2$
- Hence,

$$\begin{aligned} B_{01}(x_{1:n}) &= \frac{\lambda_0^{n\bar{x}} (n+b)^{n\bar{x}+a} \exp(-n\lambda_0)}{b^a (n\bar{x}+a-1) \cdots a} \\ &= \frac{2^5 (2+2)^{5+4} \exp(-2 \times 2)}{2^4 (5+4-1) \cdots 4} = \frac{2^5 \times 4^9 \times \exp(-4)}{16 \times 8 \times 7 \times 6 \times 5 \times 4} \\ &\approx 1.42 \end{aligned}$$

- Then $B_{01}(x_{1:n}) \approx 1.42$, and $\log_{10}(B_{01}(x_{1:n})) \approx 0.15$. According to Jeffrey's scaling rule, H_0 is supported

2 Inference under model uncertainty

Exercise 2. (★★) Let $B_{k,j}(y)$ be the Bayes factor of model \mathcal{M}_k against model \mathcal{M}_j , for all $\forall k, i, j \in \mathcal{K}$. . Show that $B_{k,j}(y) = B_{k,i}(y)B_{i,j}(y)$, for all $\forall k, i, j \in \mathcal{K}$.

Solution. It is

$$B_{k,j}(y) = \frac{\pi(\mathcal{M}_k|y) / \pi(\mathcal{M}_k)}{\pi(\mathcal{M}_j|y) / \pi(\mathcal{M}_j)} = \frac{\pi(\mathcal{M}_k|y) / \pi(\mathcal{M}_k)}{\pi(\mathcal{M}_i|y) / \pi(\mathcal{M}_i)} \frac{\pi(\mathcal{M}_i|y) / \pi(\mathcal{M}_i)}{\pi(\mathcal{M}_j|y) / \pi(\mathcal{M}_j)} = B_{k,i}(y)B_{i,j}(y)$$

3 Hierarchical Bayes

Exercise 3. (★★)[Relevance Vector Machine]

Regarding the statistical model: Long story short (supplementary material)

Consider that we are interested in recovering the mapping

$$x \mapsto^\eta \eta(x)$$

in the sense that $y \in \mathbb{R}$ is the response (output quantity) that depends on $x = (x_1, \dots, x_d) \in \mathcal{X} \subseteq \mathbb{R}^d$ which is the independent variable (input quantity) in a procedure; E.g.:

- y : precipitation in log scale
- $x = (\text{longitude}, \text{latitude})$: geographical coordinates.

Consider a set of observed data $\{(y_i, x_i)\}_{i=1}^n$, which may be contaminated by additive noise of unknown variance; i.e.

$$y_i = \eta(x_i) + \epsilon_i,$$

where $\epsilon_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ and $\sigma^2 > 0$ is unknown. We wish to recover $\eta(x)$ by using the Tikhonov regularization on the functional space \mathcal{H} such that

$$\eta = \arg \min_{\tilde{\eta} \in \mathcal{H}} \left\{ \sum_{i=1}^n L(y_i - \tilde{\eta}(x_i)) + \lambda \|\tilde{\eta}\|_{\mathcal{H}}^2 \right\} \quad (2)$$

By assuming that \mathcal{H} is a Reproducing Kernel Hilbert Space (RKHS), the solution to (2) is such that

$$\eta(x) = \beta_0 + \sum_{j=1}^n k(x, x_j) \beta_j = k(x)^\top \beta$$

where $k(x) = (1, k(x, x_1), \dots, k(x, x_n))^\top$, $k(x, x_j)$ is the reproducing kernel (such as $k_\phi(x, x_j) = \exp(-\phi \|x - x_j\|^2)$ for some known parameter $\phi > 0$), and $\beta \in \mathbb{R}^{n+1}$ is an unknown vector.

Consider the following Bayesian model¹

$$\begin{cases} y|\beta, \sigma^2 & \sim \mathcal{N}(K\beta, I\sigma^2) \\ \beta|\lambda & \sim \mathcal{N}(0, D^{-1}), \quad D = (\lambda_0, \lambda_1, \dots, \lambda_n) \\ \lambda_i & \stackrel{\text{iid}}{\sim} d\Pi(\lambda_i) \propto \lambda_i^{a-1} \exp(-b\lambda_i) d\lambda_i, \quad \forall i = 1, \dots, n \\ \sigma^2 & \sim d\Pi(\sigma^2) \propto (\sigma^2)^{c-1} \exp(-\frac{1}{\sigma^2}d) d\sigma^2 \\ \beta, \sigma^2 & \text{a priori independent} \end{cases}$$

where K is a known matrix with size $n \times (n+1)$ such that

$$K = \begin{bmatrix} 1 & k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & k(x_n, x_1) & \cdots & k(x_n, x_n) \end{bmatrix}.$$

The quantities $a > 0$, $b > 0$, $c > 0$, $d > 0$, and $\phi > 0$ are considered as fixed.

¹Dixit, A., & Roy, V. (2021). Posterior impropriety of some sparse Bayesian learning models. Statistics & Probability Letters, 171, 109039.

1. When $b = 0$, show that a necessary condition for a valid posterior inference is $a \in (-1/2, 0)$ for any choice of prior for τ (i.e. any choice of (c, d)).
2. Let $P = K (K^\top K)^{-1} K^\top$. Show that (2a) and (2b) are sufficient conditions for the Bayesian model to lead to a valid posterior inference
 - (a) if $a > 0$ and $b > 0$, or
 - (b) if $y^\top (I - P) y + 2d > 0$ and $c > -\frac{n}{2}$
3. Does the the improper Uniform prior on the joint $\log(\lambda_i)$ and $\log(\sigma^2)$, i.e. $\pi(\log(\lambda_i), \log(\sigma^2)) \propto 1$, lead to a valid inference?
4. Does the Jeffreys' prior $\pi(\lambda_i) \propto 1/\lambda_i$ lead to a valid inference?

Hint-1:

$$(y - K\beta)^\top (y - K\beta) + (\beta - \mu)^\top V^{-1}(\beta - \mu) = (\beta - \mu^*)^\top (V^*)^{-1}(\beta - \mu^*) + S^*;$$

$$S^* = \mu^\top V^{-1}\mu - (\mu^*)^\top (V^*)^{-1}(\mu^*) + y^\top y; \quad V^* = (V^{-1} + K^\top K)^{-1}; \quad \mu^* = V^* (V^{-1}\mu + K^\top y)$$

Hint-2: Sherman-Morrison-Woodbury formula:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U (C^{-1} + VA^{-1}U)^{-1} VA^{-1}$$

Hint-3:

$$-\frac{y^\top y}{2\sigma^2} \leq -\frac{y^\top (I\sigma^2 + KD^{-1}K^\top)^{-1} y}{2} \leq -\frac{1}{2\sigma^2} y^\top (I - P) y$$

where $P = K (K^\top K)^{-1} K$.

Hint-4: It is given that $\int_{(0, \infty)} \frac{t^{-(a+1)}}{(\xi+t)^{1/2}} dt < \infty$ if and only if $a \in (-1/2, 0)$.

Solution. The posterior pdf is given by

$$\pi(\beta, \sigma^2, \lambda | y) = \frac{f(y|\beta, \sigma^2) \pi(\beta, \sigma^2, \lambda)}{f(y)}$$

and is proper iff $f(y) < \infty$ where

$$f(y) = \int \left(\underbrace{\int \left(\underbrace{\int f(y|\beta, \sigma^2) \pi(\beta|\lambda) d\beta}_{=f(y|\lambda, \sigma^2)} \right) \pi(\lambda) d\lambda}_{=f(y|\sigma^2)} \right) \pi(\sigma^2) d\sigma^2$$

It is

$$\begin{aligned} f(y|\lambda, \sigma^2) &= \int f(y|\beta, \sigma^2) \pi(\beta, \sigma^2) d\beta \\ &= (2\pi)^{-\frac{n+n+1}{2}} \left(\frac{1}{\sigma^2} \right)^{\frac{n}{2}} \det(D)^{\frac{1}{2}} \int \exp \left(-\frac{1}{2\sigma^2} \left((y - K\beta)^\top (y - K\beta) + \beta^\top (D\sigma^2) \beta \right) \right) d\beta \\ &= (2\pi)^{-\frac{n+n+1}{2}} \left(\frac{1}{\sigma^2} \right)^{\frac{1}{2}} \det(D)^{\frac{1}{2}} \left[\int \exp \left(-\frac{1}{2\sigma^2} (\beta - \mu^*)^\top V^* (\beta - \mu^*) \right) d\beta \right] \left[\exp \left(-\frac{1}{2\sigma^2} S^* \right) \right] \end{aligned}$$

Because

$$\begin{aligned} \int \exp \left(-\frac{1}{2\sigma^2} (\beta - \mu^*)^\top V^* (\beta - \mu^*) \right) d\beta &= (2\pi)^{\frac{n+1}{2}} \det(V^*/\sigma^2)^{-\frac{1}{2}} \\ &= (2\pi)^{\frac{n+1}{2}} \det(K^\top K + D\sigma^2)^{-\frac{1}{2}} \end{aligned}$$

$$\begin{aligned} \exp \left(-\frac{1}{2\sigma^2} S^* \right) &= \exp \left(-\frac{1}{2\sigma^2} \mu^\top (D\sigma^2) \mu - (\mu^*)^\top (V^*)^{-1} (\mu^*) + y^\top y \right) \\ &= \exp \left(-\frac{1}{2\sigma^2} \left(y^\top y - y^\top K (K^\top K + D\sigma^2)^{-1} K^\top y \right) \right) \\ &= \exp \left(-\frac{1}{2\sigma^2} \left(y^\top \left(I - K (K^\top K + D\sigma^2)^{-1} K^\top \right) y \right) \right) \\ &= \exp \left(-\frac{1}{2\sigma^2} \left(y^\top (K^\top D^{-1} K + I\sigma^2)^{-1} y \right) \right) \end{aligned}$$

So

$$\begin{aligned} f(y|\lambda, \sigma^2) &= (2\pi)^{-\frac{n}{2}} \left(\frac{1}{\sigma^2} \right)^{\frac{1}{2}} \det(D)^{\frac{1}{2}} \det(K^\top K + \sigma^2 D)^{-\frac{1}{2}} \\ &\quad \times \exp \left(-\frac{1}{2\sigma^2} \left(y^\top (I\sigma^2 + K^\top D^{-1} K)^{-1} y \right) \right) \end{aligned}$$

1. I have

$$\begin{aligned} f(y|\sigma^2) &= \int f(y|\lambda, \sigma^2) \pi(\lambda) d\lambda \\ &= (2\pi)^{-\frac{n}{2}} (\sigma^2)^{\frac{1}{2}} \int \left[\det(D)^{\frac{1}{2}} \right] \left[\det(K^\top K + D\sigma^2)^{-\frac{1}{2}} \right] \\ &\quad \times \exp \left(-\frac{1}{2} \left(y^\top (I\sigma^2 + K^\top D^{-1} K)^{-1} y \right) \right) \left[\prod_{i=0}^n \lambda_i^{a-1} \right] d\lambda_0 \dots d\lambda_n \end{aligned}$$

because $b = 0$.

- It is $\exp \left(-\frac{y^\top y}{2\sigma^2} \right) \leq \exp \left(-\frac{y^\top (I\sigma^2 + K^\top D^{-1} K)^{-1} y}{2} \right)$
- It is $\det(D)^{\frac{1}{2}} = \prod_{i=0}^n \lambda_i^{\frac{1}{2}}$.
- If $\{e_j\}_{j=0}^{n-1}$ are eigenvalues of $K^\top K$ and $e_{\max} = \max(\{e_j\})$, then $K^\top K + D\sigma^2 \leq Ie_{\max} + D\sigma^2$, consequently $\det(K^\top K + D\sigma^2)^{-\frac{1}{2}} \geq \prod_{j=0}^n (\lambda_j \sigma^2 + e_{\max})^{-\frac{1}{2}}$.

Then

$$\begin{aligned} f(y|\sigma^2) &\geq (2\pi)^{-\frac{n}{2}} (\sigma^2)^{\frac{1}{2}} \int \prod_{j=0}^n \lambda_j^{\frac{1}{2}} \prod_{j=0}^n (\lambda_j \sigma^2 + e_{\max})^{-\frac{1}{2}} \exp \left(-\frac{1}{2\sigma^2} y^\top y \right) \prod_{j=0}^n \lambda_j^{a-1} d\lambda_0 \dots d\lambda_n \\ &= (2\pi)^{-\frac{n}{2}} (\sigma^2)^{\frac{1}{2}} \exp \left(-\frac{1}{2\sigma^2} y^\top y \right) \int \dots \int \prod_{j=0}^n \left[\lambda_j^{\frac{1}{2}} \right] \left[\prod_{j=0}^n (\lambda_j \sigma^2 + e_{\max})^{-\frac{1}{2}} \right] \left[\prod_{j=0}^n \lambda_j^{a-1} \right] d\lambda_0 \dots d\lambda_n \\ &= (2\pi)^{-\frac{n}{2}} (\sigma^2)^{\frac{1}{2}} \exp \left(-\frac{1}{2\sigma^2} y^\top y \right) \prod_{j=0}^n \int \frac{\lambda_j^{a-\frac{1}{2}}}{(\lambda_j \sigma^2 + e_{\max})^{\frac{1}{2}}} d\lambda_j \end{aligned}$$

Let $t_i = 1/\lambda_i$, then

$$f(y|\sigma^2) \geq (2\pi)^{-\frac{n}{2}} (\sigma^2)^{\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} y^\top y\right) \prod_{j=0}^n \int \frac{t_{jj}^{-a-1}}{\left(t_j + \frac{\sigma^2}{e_{\max}}\right)^{\frac{1}{2}}} d\lambda_j$$

which is finite if and only if $a \in (-1/2, 0)$.

2.

(a) If $a > 0$, $b > 0$ then $\lambda_i \stackrel{\text{iid}}{\sim} \text{Ga}(a, b)$ for all $i = 1, \dots, n$, and if $c > 0$, $d > 0$ then $\tau \stackrel{\text{iid}}{\sim} \text{Ga}(c, d)$ which are proper. So $\Pi(\beta, \sigma^2, \lambda, \tau)$ is a proper prior, and hence it leads to proper posterior.

(b) I have

$$f(y|\sigma^2) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{\frac{1}{2}} \int \left[\det(D)^{\frac{1}{2}} \right] \left[\det(K^\top K + D\sigma^2)^{-\frac{1}{2}} \right] \\ \times \exp\left(-\frac{1}{2} \left(y^\top (I\sigma^2 + K^\top D^{-1}K)^{-1} y \right)\right) \pi(\lambda) d\lambda$$

It is $\det(D)^{\frac{1}{2}} = \prod_{i=0}^n \lambda_i^{\frac{1}{2}}$. Also, it is $K^\top K + D\sigma^2 \geq D\sigma^2$ then $\det(K^\top K + D\sigma^2)^{-\frac{1}{2}} \leq \prod_{j=0}^n (\lambda_j \sigma^2)^{-\frac{1}{2}}$. Hence

$$f(y|\sigma^2) \leq (2\pi)^{-\frac{n}{2}} (\sigma^2)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} y^\top (I - P) y\right) \int \pi(\lambda) d\lambda$$

which implies that $f(y|\sigma^2) < \infty$ if $\pi(\lambda)$ is proper. Yet,

$$f(y) = \int f(y|\sigma^2) \pi(\sigma^2) d\sigma^2 \\ \leq (2\pi)^{-\frac{n}{2}} \int (\sigma^2)^{-\frac{n}{2}+c+1} \exp\left(-\frac{1}{\sigma^2} \left(\frac{y^\top (I - P) y}{2} + d \right)\right) d\sigma^2$$

which is finite if $y^\top (I - P) y + 2d > 0$ and $c > -\frac{n}{2}$.

(c) No. This implies $\pi(\lambda, \sigma^2) \propto \sigma^2 \prod_{j=0}^n \lambda_j^{-1}$. It is improper prior as $\int \pi(\lambda, \sigma^2) d(\lambda, \sigma^2) = \infty$, and $(a, b, c, d) = (0, 0, 0, 0)$ which violates the necessary conditions.

(d) No, it violates the necessary conditions.