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

Lecturer: Georgios Karagiannis georgios.karagiannis@durham.ac.uk

1 Hypothesis test

Exercise 1. ($\star\star$)Consider a Bayesian model

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

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 $\text{E}(x)=a/b$, $\text{Var}(x)=a/b^2$.

Hint-3 Negative Binomial distribution has PMF: $\text{Nb}(x|r,\theta)=\binom{r+x-1}{r-1}\theta^r(1-\theta)^x1_{\mathbb{N}}(x)$. 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.

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 $\text{E}(\lambda)=2$, and $\text{Var}(\lambda)=1$.

---

2 Inference under model uncertainty

Exercise 2. ($\star\star$)Let $B_{k,j}(y)$ be the Bayes factor of model $\mathscr{M}_k$ against model $\mathscr{M}_j$, 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 $\forall k, i, j \in \mathcal{K}$.

---

3 Hierarchical Bayes

Exercise 3. ($\star\star\star$)[Relevance Vector Machine]

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

Consider that we are interested in recovering the mapping

$$x \overset{\eta}{\longmapsto} \eta(x)$$

in the sense that $y \in \mathbb{R}$ is the response (output quantity) that depends on $x = (x_1, ..., x_d) \in \mathcal{X} \subseteq \mathbb{R}^d$ 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 $\epsilon_i \overset{\text{IID}}{\sim} \text{N}\left(0, \sigma^2\right)$ 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

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

By assuming that $\mathcal{H}$ is a Reproducing Kernel Hilbert Space (RKHS), the solution to (1) is such that

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

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

Consider the following Bayesian model

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

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.

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 $P = K\left(K^\top K\right)^{-1} K^\top$. 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 $y^\top \left(I - P\right) y + 2d > 0$ and $c > -\frac{n}{2}$

3. Does the the improper Uniform prior on the joint $\log\left(\lambda_i\right)$ and $\log\left(\sigma^2\right)$, i.e. $\pi\left(\log\left(\lambda_i\right), \log\left(\sigma^2\right)\right) \propto 1$, lead to a valid inference?

4. Does the Jeffreys' prior $\pi\left(\lambda_i\right) \propto 1/\lambda_i$ lead to a valid inference?

Hint-1:

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

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

Hint-2: Sherman-Morrison-Woodbury formula:

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

Hint-3:

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

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

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