Bayesian Statistics III/IV (MATH3361/4071) Michaelmas term 2021

Problem class 1 [a]

Nuisance parameters, the Normal model, and the Normal linear regression with unknown variance

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

[a]Author: Georgios P. Karagiannis.

Nuisance parameters

Exercise 1. ($\star\star$)Assume observable quantities $y=(y_1,...,y_n)$ forming the available data set of size n. Assume that the observations are drawn i.i.d. from a sampling distribution which is judged to be in the Normal parametric family of distributions $\mathrm{N}(\mu,\sigma^2)$ with unknown mean μ and variance σ^2. We are interested in learning μ and the next outcome $z=y_{n+1}$. We do not care about σ^2. <-story

Assume You specify a Bayesian model <-set-up

$$\begin{cases} y_i|\mu,\sigma^2 \sim \mathrm{N}(\mu,\sigma^2), \text{ for all } i=1,...,n & \text{, Statistical model} \\ \mu|\sigma^2 \sim \mathrm{N}(\mu_0,\sigma^2\frac{1}{\tau_0}) & \text{, prior} \\ \sigma^2 \sim \mathrm{IG}(a_0,k_0) & \text{, prior} \end{cases}$$

1. Show that
$$\sum_{i=1}^{n}(y_i-\theta)^2 = n(\bar{y}-\theta)^2+ns^2,$$
where $s^2=\frac{1}{2}\sum_{i=1}^{n}(y_i-\bar{y})^2$.

2. Show that the joint posterior distribution $\Pi(\mu,\sigma^2|y)$ is such as
$$\mu|y,\sigma^2 \sim \mathrm{N}(\mu_n,\sigma^2\frac{1}{\tau_n})$$
$$\sigma^2|y \sim \mathrm{IG}(a_n,k_n)$$
with
$$\mu_n=\frac{n\bar{y}+\tau_0\mu_0}{n+\tau_0}; \qquad \tau_n = n+\tau_0; \qquad a_n=a_0+n$$
$$k_n=k_0+\frac{1}{2}ns_n^2+\frac{1}{2}\frac{\tau_0 n(\mu_0-\bar{y})^2}{n+\tau_0}$$
Hint: It is
$$-\frac{1}{2}\frac{(\mu-\mu_1)^2}{v_1}-\frac{1}{2}\frac{(\mu-\mu_2)^2}{v_2}...-\frac{1}{2}\frac{(\mu-\mu_n)^2}{v_n}=-\frac{1}{2}\frac{(\mu-\hat{\mu})^2}{\hat{v}}+C$$
where
$$\hat{v}=\left(\sum_{i=1}^{n}\frac{1}{v_i}\right)^{-1}; \quad \hat{\mu}=\hat{v}\left(\sum_{i=1}^{n}\frac{\mu_i}{v_i}\right); \quad C=\frac{1}{2}\frac{\hat{\mu}^2}{\hat{v}}-\frac{1}{2}\sum_{i=1}^{n}\frac{\mu_i^2}{v_i}$$

3. Show that the marginal posterior distribution $\Pi(\mu|y)$ is such as
$$\mu|y \sim \mathrm{T}_1\left(\mu_n,\frac{k_n}{a_n}\frac{1}{\tau_n},2a_n\right)$$

Hint-1: If $x \sim \text{IG}(a, b)$, $y = cx$, then $y \sim \text{IG}(a, cb)$.

Hint-2: The definition of Student T is considered as known

4. Show that the predictive distribution $\Pi(z|y)$ is Student T such as

$$z|y \sim \text{T}_1\left(\mu_n, \frac{k_n}{a_n}(\frac{1}{\tau_n}+1), 2a_n\right)$$

Hint-1: Consider that

$$\text{N}(x|\mu_1, \sigma_1^2)\,\text{N}(x|\mu_2, \sigma_2^2) \;=\; \text{N}(x|m, v^2)\,\text{N}(\mu_1|\mu_2, \sigma_1^2+\sigma_2^2)$$

where

$$v^2 = \left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)^{-1}; \quad m = v^2\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right)$$

Hint-2: The definition of Student T is considered as known

Proper/improper priors

Exercise 2. ($\star\star$)Consider the Bayesian model

$$\begin{cases} x|\sigma & \sim \text{N}(0, \sigma^2) \\ \sigma & \sim \text{Ex}(\lambda) \end{cases}$$

where $\text{Ex}(\lambda)$ is the exponential distribution with mean $1/\lambda$. Show that the posterior distribution is not defined always.

- HINT: Precisely, show that the posterior is not defined in the case that you collect only one observation $x = 0$.

Conjugate priors

Exercise 3. ($\star\star$)Consider the Bayesian model

$$\begin{cases} x_i|\theta & \overset{\text{IID}}{\sim} \text{Mu}_k(\theta) \\ \theta & \sim \Pi(\theta) \end{cases}$$

where $\theta \in \Theta$, with $\Theta = \{\theta \in (0,1)^k | \sum_{j=1}^k \theta_j = 1\}$ and $\mathcal{X}_k = \{x \in \{0, ..., n\}^k | \sum_{j=1}^k x_j = 1\}$.

Hint-1: Mu_k denotes the Multinomial probability distribution with PMF

$$\text{Mu}_k(x|\theta) = \begin{cases} \prod_{j=1}^k \theta_j^{x_j} & \text{, if } x \in \mathcal{X}_k \\ 0 & \text{, otherwise} \end{cases}$$

Hint-2: $\text{Di}_k(a)$ denotes the Dirichlet distribution with PDF

$$\text{Di}_k(\theta|a) = \begin{cases} \frac{\Gamma(\sum_{j=1}^k a_j)}{\prod_{j=1}^k \Gamma(a_j)} \prod_{j=1}^k \theta_j^{a_j-1} & \text{, if } \theta \in \Theta \\ 0 & \text{, otherwise} \end{cases}$$

1. Derive the conjugate prior distribution for θ, and recognize that it is a Dirichlet distribution family of distributions.
2. Verify that the prior distribution you derived above is indeed conjugate by using the definition.

Jeffreys priors

Exercise 4. ($\star\star$)Consider the trinomial distribution

$$\begin{aligned} p(x, y|\pi, \rho) &= \frac{n!}{x!\,y!\,z!}\pi^x\rho^y\sigma^z, \qquad (x+y+z=n) \\ &\propto \pi^x\rho^y(1-\pi-\rho)^{n-x-y}. \end{aligned}$$

Specify a Jeffreys' prior for (π, ρ).

HINT: It is $\mathrm{E}(x) = n\pi$, $\mathrm{E}(y) = n\rho$.

Exercise 5. ($\star\star$)Consider the Bayesian model

$$\begin{cases} x_i|\theta & \overset{\text{iid}}{\sim} \mathrm{Ga}(\alpha, \beta),\ \forall i = 1, ..., n \\ (\alpha, \beta) & \sim \Pi(\alpha, \beta) \end{cases}$$

where $\mathrm{Ga}(a, \beta)$ is the Gamma distribution with expected value α/β. Specify a Jeffrey's prior for $\theta = (\alpha, \beta)$.

Hint-1: Gamma distr.: $x \sim \mathrm{Ga}(a, b)$ has pdf $f(x) = \frac{b^a}{\Gamma(a)}x^{a-1}\exp(-bx)1_{(0,+\infty)}(x)$, and Expected value $\mathrm{E}_{\mathrm{Ga}}(x|a, b) = \frac{a}{b}$

Hint-2: You may also need that the second derivative of the logarithm of a Gamma function is the 'polygamma function of order 1'. Ie,

- $F^{(0)}(\alpha) = \frac{\mathrm{d}}{\mathrm{d}\alpha}\log(\Gamma(a))$
- $F^{(1)}(\alpha) = \frac{\mathrm{d}^2}{\mathrm{d}\alpha^2}\log(\Gamma(a))$

Hint-3: You may leave your answer in terms of function $F^{(1)}(\alpha)$.

..

Exercise 6. ($\star\star$)Consider the Bayesian model

$$\begin{cases} x_i|\theta & \overset{\text{iid}}{\sim} \mathrm{Ex}(\theta),\ \forall i = 1, ..., n \\ \theta & \sim \mathrm{Ga}(a, b) \end{cases}$$

Hint-1: The PDF of $x \sim \mathrm{G}(a, b)$ is $\mathrm{Ga}(x|a, b) = \frac{b^a}{\Gamma(a)}x^{a-1}\exp(-bx)1_{(0,+\infty)}(x)$

Hint-2: The PDF of $x \sim \mathrm{Ex}(\theta)$ is $\mathrm{Ex}(x|\theta) = \mathrm{Ga}(x|1, \theta)$

1. Show that the parametric model is member of the Exponential family, and the sufficient statistic for a sample of observables $x = (x_1, ..., x_n)$.

2. Show that the posterior distribution θ given x is Gamma and compute its parameters.

3. Show that the predictive distribution $G(z|x)$ of a future z given $x = (x_1, ..., x_n)$, has PDF

$$g(z|x) = \frac{a^*(b^*)^{a^*}}{(z + b^*)^{a^*+1}} 1(x \geq 0)$$

Further practice

From the exercise sheet, have a look at Exercises 26, 31, 47, 6, and 50.

A About Nuisance parameters

Assume observable quantities $y = (y_1, ..., y_n)$. Assume that the sampling distribution is $\mathrm{d}F(y|\theta)$ labeled by an unknown parameter $\theta \in \Theta$. Let $\theta = (\phi, \lambda)^\top$ with $\phi \in \Phi$ and $\lambda \in \Lambda$. Assume You are interested in learning parameter $\phi \in \Phi$, and You are not interested in learning the unknown parameter $\lambda \in \Lambda$; but both ϕ, λ are parts of the statistical model parameterisation. The unknown quantity $\lambda \in \Lambda$ is called nuisance parameter. We an call $\phi \in \Phi$ parameter of interest.

Note 7. In Bayesian Stats, learning (or quantifying uncertainty about) parameter of interest ϕ under the presence of a nuisance parameter $\lambda \in \Lambda$ is performed according to the Bayesian paradigm as usual: You specify a prior $\mathrm{d}\Pi(\phi, \lambda)$ with PDF/PMF $\pi(\phi, \lambda) = \pi(\phi|\lambda)\pi(\lambda)$ on the joint space of ALL Your unknown parameters $\theta = (\phi, \lambda)^\top$; you compute the joint posterior distribution $\mathrm{d}\Pi(\theta|y)$ of $\theta = (\phi, \lambda)^\top$ via the Bayesian theorem. Reasonably, Your posterior degree of believe about the parameter of interest ϕ given the data $y = (1_1, ..., y_n)$ is given through the marginal posterior distribution $\mathrm{d}\Pi(\phi|y)$.

Note 8. To summarize; Specify the Bayesian model as: <sum-up

$$\begin{cases} y|\overbrace{\phi, \lambda}^{=\theta} \sim F(y|\overbrace{\phi, \lambda}^{=\theta}) & \text{, the statistical model} \\ (\underbrace{\phi, \lambda}_{=\theta}) \sim \Pi(\underbrace{\phi, \lambda}_{=\theta}) & \text{, the prior model} \end{cases}$$

The joint posterior of θ given y is $\mathrm{d}\Pi(\theta|y) = \mathrm{d}\Pi(\lambda|y, \phi)\mathrm{d}\Pi(\phi|y)$ is with PDF/PMF

$$\pi(\overbrace{\phi, \lambda}^{=\theta}|y) = \frac{f(y|\overbrace{\phi, \lambda}^{=\theta})\pi(\overbrace{\phi, \lambda}^{=\theta})}{f(y)} = \underbrace{\frac{f(y|\phi, \lambda)\pi(\lambda|\phi)}{f(y|\phi)}}_{=\pi(\lambda|y,\phi)}\underbrace{\frac{f(y|\phi)\pi(\phi)}{f(y)}}_{=\pi(\phi|y)} = \pi(\lambda|y, \phi)\pi(\phi|y)$$

The (marginal) likelihood $f(y|\phi)$ of y given ϕ is

$$f(y|\phi) = \underbrace{\int_\Lambda f(y|\overbrace{\phi, \lambda}^{=\theta})\mathrm{d}\Pi(\lambda|\phi)}_{=\mathrm{E}_{\Pi(\lambda|\phi)}(f(y|\phi,\lambda)|\phi)} = \begin{cases} \int_\Lambda f(y|\phi, \lambda)\pi(\lambda|\phi)\mathrm{d}\lambda & \text{, if } \lambda \text{ cont} \\ \\ \sum_{\forall \lambda \in \Lambda} f(y|\phi, \lambda)\pi(\lambda|\phi) & \text{, if } \lambda \text{ discr} \end{cases}$$

The PDF/PMF $\pi(\phi|y)$ of marginal posterior $\mathrm{d}\Pi(\phi|y)$ of ϕ is

$$\pi(\phi|y) = \underbrace{\int_\Lambda \pi(\overbrace{\phi, \lambda}^{=\theta}|y)\mathrm{d}\lambda}_{=\mathrm{E}_{\Pi(\lambda|y)}(\pi(\phi|y,\lambda))} \qquad \text{or equivalently} \qquad \pi(\phi|y) = \frac{f(y|\phi)\pi(\phi)}{f(y)}$$

The predictive distribution $\mathrm{d}G(z|y)$ of the next outcome $z = (y_{n+1}, ...y_{n+m})$ given y has pdf/pmf

$$g(z|y) = \int f(y|\overbrace{\phi, \lambda}^{=\theta})\mathrm{d}\Pi(\overbrace{\phi, \lambda}^{=\theta}|y)$$

and the marginal likelihood $f(y)$ is

$$f(y) = \int f(y|\overbrace{\phi, \lambda}^{=\theta})\pi(\overbrace{\phi, \lambda}^{=\theta})\mathrm{d}\phi\mathrm{d}\lambda$$

B Criteria for integrals

General: Let integrable functions $f(x)$, and $g(x)$ for $x \geq a$.

Let

$$0 \leq f(x) \leq g(x), \quad \text{for } x \geq a$$

Then

$$\int_a^\infty g(x)\mathrm{d}x < \infty \implies \int_a^\infty f(x)\mathrm{d}x < \infty$$
$$\int_a^\infty f(x)\mathrm{d}x = \infty \implies \int_a^\infty g(x)\mathrm{d}x = \infty$$

Type I: Let integrable functions $f(x)$, and $g(x)$ for $x \geq a$, and let $g(x)$ be positive.

Let

$$\lim_{n\to\infty} \frac{f(x)}{g(x)} = c$$

Then

- If $c \in (0, \infty)$:
$$\int_a^\infty g(x)\mathrm{d}x < \infty \iff \int_a^\infty f(x)\mathrm{d}x < \infty$$
- If $c = 0$:
$$\int_a^\infty g(x)\mathrm{d}x < \infty \implies \int_a^\infty f(x)\mathrm{d}x < \infty$$
- If $c = \infty$:
$$\int_a^\infty f(x)\mathrm{d}x = \infty \implies \int_a^\infty g(x)\mathrm{d}x = \infty$$

Type II: Let integrable functions $f(x)$, and $g(x)$ for $a < x \leq b$, and let $g(x)$ be positive.

Let

$$\lim_{n\to a^+} \frac{f(x)}{g(x)} = c$$

Then

- If $c \in (0, \infty)$:
$$\int_a^\infty g(x)\mathrm{d}x < \infty \iff \int_a^\infty f(x)\mathrm{d}x < \infty$$
- If $c = 0$:
$$\int_a^\infty g(x)\mathrm{d}x < \infty \implies \int_a^\infty f(x)\mathrm{d}x < \infty$$
- If $c = \infty$:
$$\int_a^\infty f(x)\mathrm{d}x = \infty \implies \int_a^\infty g(x)\mathrm{d}x = \infty$$

Note: A useful test function is

$$\int_0^\infty \left(\frac{1}{x}\right)^p \mathrm{d}x \quad \begin{cases} < \infty & , \text{ when } p > 1 \\ = \infty & , \text{ when } p \leq 1 \end{cases}$$