

**Problem class 1<sup>a</sup>****Nuisance parameters, the Normal model, and the Normal linear regression with unknown variance**

Lecturer: Georgios Karagiannis

georgios.karagiannis@durham.ac.uk

<sup>a</sup>Author: Georgios P. Karagiannis.**Nuisance parameters**

&lt;concept

**Definition 1.** Assume observable quantities  $y = (y_1, \dots, y_n)$ . Assume that the sampling distribution is  $dF(y|\theta)$  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  $\phi, \lambda$  are parts of the statistical model parameterisation. The unknown quantity  $\lambda \in \Lambda$  is called nuisance parameter. We can call  $\phi \in \Phi$  parameter of interest.

*Note 2.* In Bayesian Stats, learning (or quantifying uncertainty about) parameter of interest  $\phi$  under the presence of a nuisance parameter  $\lambda \in \Lambda$  is performed according to the Bayesian paradigm as usual: You specify a prior  $d\Pi(\phi, \lambda)$  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  $d\Pi(\theta|y)$  of  $\theta = (\phi, \lambda)^\top$  via the Bayesian theorem. Reasonably, Your posterior degree of belief about the parameter of interest  $\phi$  given the data  $y = (y_1, \dots, y_n)$  is given through the marginal posterior distribution  $d\Pi(\phi|y)$ .

*Note 3.* To summarize; Specify the Bayesian model as:

&lt;sum-up

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

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

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

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

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

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

$$\pi(\phi|y) = \underbrace{\int_{\Lambda} \pi(\underbrace{\phi, \lambda}_{=\theta}|y)d\lambda}_{=E_{\Pi(\lambda|y)}(\pi(\phi|y, \lambda))} \quad \text{or equivalently} \quad \pi(\phi|y) = \frac{f(y|\phi)\pi(\phi)}{f(y)}$$

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

$$g(z|y) = \int f(y|\underbrace{\phi, \lambda}_{=\theta}) d\Pi(\underbrace{\phi, \lambda}_{=\theta}|y)$$

and the marginal likelihood  $f(y)$  is

$$f(y) = \int f(y|\underbrace{\phi, \lambda}_{=\theta}) \pi(\underbrace{\phi, \lambda}_{=\theta}) d\phi d\lambda$$

## Practice in challenging problems

**Exercise 4.** (★★)(Nuisance parameters are involved)

<-story

Assume observable quantities  $y = (y_1, \dots, y_n)$  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  $N(\mu, \sigma^2)$  with unknown mean  $\mu$  and variance  $\sigma^2$ . We are interested in learning  $\mu$  and the next outcome  $z = y_{n+1}$ . We do not care about  $\sigma^2$ .

Assume You specify a Bayesian model

<-set-up

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

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

$$\begin{aligned} \mu|y, \sigma^2 &\sim N(\mu_n, \sigma^2 \frac{1}{\tau_n}) \\ \sigma^2|y &\sim \text{IG}(a_n, k_n) \end{aligned}$$

with

$$\mu_n = \frac{n\bar{y} + \tau_0\mu_0}{n + \tau_0}; \quad \tau_n = n + \tau_0; \quad a_n = a_0 + n$$

$$k_n = k_0 + \frac{1}{2}ns_n^2 + \frac{1}{2}\frac{\tau_0n(\mu_0 - \bar{y})^2}{n + \tau_0}$$

**Hint:** It is

$$-\frac{1}{2} \frac{(\mu - \mu_1)^2}{v_1} - \frac{1}{2} \frac{(\mu - \mu_2)^2}{v_2} \dots - \frac{1}{2} \frac{(\mu - \mu_n)^2}{v_n} = -\frac{1}{2} \frac{(\mu - \hat{\mu})^2}{\hat{v}} + C$$

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

$$\mu|y \sim T_1 \left( \mu_n, \frac{k_n}{a_n} \frac{1}{\tau_n}, 2a_n \right)$$

**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

$$z|y \sim T_1 \left( \mu_n, \frac{k_n}{a_n} \left( \frac{1}{\tau_n} + 1 \right), 2a_n \right)$$

**Hint-1:** Consider that

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

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

## General practice

**Exercise 5.** (★★) Consider the Bayesian model

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

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

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

1. Show that the parametric model is member of the Exponential family, and the sufficient statistic for a sample of observables  $x = (x_1, \dots, x_n)$ .
2. Show that the posterior distribution  $\theta$  given  $x$  is Gamma and compute its parameters.
3. Show that the predictive distribution  $G(z|x)$  of a future  $z$  given  $x = (x_1, \dots, x_n)$ , has PDF

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

**Exercise 6.** (★★) Consider the Bayesian model

$$\begin{cases} x_i|\theta & \stackrel{\text{iid}}{\sim} \text{Mu}_k(\theta) \\ \theta & \sim \text{Di}_k(a) \end{cases}$$

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

**Hint-1:**  $\text{Mu}_k$  denotes the Multinomial probability distribution with PMF

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

**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. Show that the parametric model (1) is a member of the  $k - 1$  exponential family.
2. Compute the likelihood  $f(x_{1:n}|\theta)$ , and find the sufficient statistic  $t_n := t_n(x_{1:n})$ .
3. Compute the posterior distribution. State the name of the distribution, and express its parameters with respect to the observations and the hyper-parameters of the prior. Justify your answer.
4. Compute the probability mass function of the predictive distribution for a future observation  $y = x_{n+1}$  in closed form.

**Hint**  $\Gamma(x) = (x - 1)\Gamma(x - 1)$ .

---