# Handout 5a: Nuisance parameters [a]

**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 <concept

**Definition 1.** 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 $\phi,\lambda$ 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* 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 $\mathrm{d}\Pi(\phi,\lambda)$ 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 $\phi$ given the data $y=(1_1,...,y_n)$ is given through the marginal posterior distribution $\mathrm{d}\Pi(\phi|y)$.

*Note* 3. To summarize; Specify the Bayesian model as: <sum-up

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

The joint posterior of $\theta$ 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 $\phi$ 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 $\phi$ 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

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

## Practice in challenging problems

**Exercise 4.** ($\star\star$)(Nuisance parameters are involved) <-story

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 $\mu$ 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

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

1. Show that 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

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

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

**Hint:** It is

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

where

$$\hat{v}^2 = \left(\sum_{i=1}^{n}\frac{1}{v_i^2}\right)^{-1}; \quad \hat{\mu} = \hat{v}^2\left(\sum_{i=1}^{n}\frac{\mu_i}{v_i^2}\right); \quad C = \frac{1}{2}\frac{\hat{\mu}^2}{\hat{v}^2} - \frac{1}{2}\sum_{i=1}^{n}\frac{\mu_i^2}{v_i^2}$$

2. 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 $x \sim \mathrm{IG}(a, b)$, $y = cx$, then $y \sim \mathrm{IG}(a, cb)$.

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

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

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

**Hint-1:** Consider that

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

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.** ($\star\star$)Consider the Bayesian model

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

**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 $\theta$ given $x$ is Gamma and compute its parameters.
3. Show that the predictive distribution $G(z|x)$ of a future $z$ given $x=(x_1,...,x_n)$, has PDF

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

---

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

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

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:** $\mathrm{Mu}_k$ denotes the Multinomial probability distribution with PMF

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

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

---