

**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

**Solution.**

1. It is

$$\begin{aligned} \sum_{i=1}^n (y_i - \theta)^2 &= \sum_{i=1}^n [(y_i - \bar{y}) - (\theta - \bar{y})]^2 \\ &= \sum_{i=1}^n \left[ (y_i - \bar{y})^2 + (\theta - \bar{y})^2 - 2(y_i - \bar{y})(\theta - \bar{y}) \right] \\ &= ns^2 + n(\bar{y} - \theta)^2, \text{ where } s^2 = \frac{1}{2} \sum_{i=1}^n (y_i - \bar{y})^2 \end{aligned}$$

2. I use the Bayes theorem

$$\begin{aligned} \pi(\mu, \sigma^2|y) &\propto f(y|\mu, \sigma^2) \pi(\mu, \sigma^2) = \prod_{i=1}^n N(y_i|\mu, \sigma^2) N(\mu|\mu_0, \sigma^2 \frac{1}{\tau_0}) \text{IG}(\sigma^2|a_0, k_0) \\ &\propto \left( \frac{1}{\sigma^2} \right)^{\frac{n}{2}} \exp \left( -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma^2} \right) \times \left( \frac{1}{\sigma^2} \right)^{\frac{1}{2}} \exp \left( -\frac{1}{2} \frac{(\mu - \mu_0)^2}{\sigma^2/\tau_0} \right) \times \left( \frac{1}{\sigma^2} \right)^{a_0+1} \exp \left( -\frac{1}{\sigma^2} k_0 \right) \\ &\propto \left( \frac{1}{\sigma^2} \right)^{\frac{n}{2} + \frac{1}{2} + a_0 + 1} \exp \left( \frac{1}{\sigma^2} \left[ -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{1} - \frac{1}{2} \frac{(\mu - \mu_0)^2}{1/\tau_0} \right] - \frac{1}{\sigma^2} k_0 \right) \end{aligned}$$

It is

$$-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{1} - \frac{1}{2} \frac{(\mu - \mu_0)^2}{1/\tau_0} = -\frac{1}{2} \underbrace{\frac{(\mu - \mu_n)^2}{v_n^2}}_{=1/\tau_n} + C_n$$

where

$$\begin{aligned}
 v_n &= \left( \sum_{i=1}^n \frac{1}{1} + \frac{1}{1/\tau_0} \right)^{-1} = \frac{1}{n + \tau_0} \implies \tau_n = n + \tau_0 \\
 \mu_n &= v_n \left( \sum_{i=1}^n \frac{y_i}{1} + \frac{\mu_0}{1/\tau_0} \right) \implies \mu_n = \frac{n\bar{y} + \tau_0\mu_0}{n + \tau_0} \\
 C_n &= \frac{1}{2} \frac{\mu_n^2}{v_n} - \frac{1}{2} \left( n \sum_{i=1}^n y_i^2 + \tau_0\mu_0^2 \right) = \frac{1}{2} \frac{(n\bar{y} + \tau_0\mu_0)^2}{n + \tau_0} - \frac{1}{2} \left( n \sum_{i=1}^n y_i^2 + \tau_0\mu_0^2 \right) \\
 &= \dots \text{Quest. 1} \dots = -\frac{1}{2} n s_n^2 - \frac{1}{2} \frac{\tau_0 n (\mu_0 - \bar{y})^2}{n + \tau_0}
 \end{aligned}$$

So

$$\begin{aligned}
 \pi(\mu, \sigma^2 | y) &\propto \left( \frac{1}{\sigma^2} \right)^{\frac{1}{2} + \frac{n}{2} + a_0 + 1} \exp \left( \frac{1}{\sigma^2} \left[ -\frac{1}{2} \frac{(\mu - \mu_n)^2}{1/\tau_n} + C_n \right] - \frac{1}{\sigma^2} k_0 \right) \\
 &\propto \underbrace{\left( \frac{1}{\sigma^2} \right)^{\frac{1}{2}} \exp \left( -\frac{1}{2} \frac{(\mu - \mu_n)^2}{\sigma^2/\tau_n} \right)}_{\propto N(\mu | \mu_n, \sigma^2/\tau_n)} \times \underbrace{\left( \frac{1}{\sigma^2} \right)^{\frac{n}{2} + a_0 + 1} \exp \left( -\frac{1}{\sigma^2} \overbrace{(k_0 - C_n)}^{=k_n} \right)}_{\propto \text{IG}(\sigma^2 | a_n, k_n)} \\
 &\propto N(\mu | \mu_n, \sigma^2/\tau_n) \text{IG}(\sigma^2 | a_n, k_n)
 \end{aligned}$$

where

$$\begin{aligned}
 \mu_n &= \frac{n\bar{y} + \tau_0\mu_0}{n + \tau_0}; & a_n &= \frac{n}{2} + a_0; \\
 \tau_n &= n + \tau_0; & k_n &= k_0 + \frac{1}{2} n s_n^2 + \frac{1}{2} \frac{\tau_0 n (\mu_0 - \bar{y})^2}{n + \tau_0}.
 \end{aligned}$$

3. It is

$$\pi(\mu | y) = \int \pi(\mu, \sigma^2 | y) d\sigma^2 = \int N(\mu | \mu_n, \sigma^2/\tau_n) \text{IG}(\sigma^2 | a_n, k_n) d\sigma^2$$

by change of variable  $\xi = \sigma^2 \frac{1}{2k_n}$ , it is

$$\begin{aligned}
 \pi(\mu | y) &= \int N(\mu | \mu_n, \xi 2k_n \frac{1}{\tau_n} \frac{2a_n}{2a_n}) \text{IG}(\xi | \frac{2a_n}{2}, \frac{1}{2}) d\xi = \int N(\mu | \mu_n, \xi \frac{1}{\tau_n} \frac{k_n}{a_n} 2a_n) \text{IG}(\xi | \frac{2a_n}{2}, \frac{1}{2}) d\xi \\
 &= T_1(\mu | \mu_n, \frac{k_n}{a_n} \frac{1}{\tau_n}, 2a_n)
 \end{aligned}$$

4. It is

$$\begin{aligned}
 g(z | y) &= \int f(z | \mu, \sigma^2) \pi(\mu, \sigma^2 | y) d\mu d\sigma^2 = \int N(z | \mu, \sigma^2) N(\mu | \mu_n, \sigma^2/\tau_n) \text{IG}(\sigma^2 | a_n, k_n) d\mu d\sigma^2 \\
 &= \int \left[ \int N(z | \mu, \sigma^2) N(\mu | \mu_n, \sigma^2/\tau_n) d\mu \right] \text{IG}(\sigma^2 | a_n, k_n) d\sigma^2
 \end{aligned}$$

Normal density is symmetric  $N(z | \mu, \sigma^2) N(\mu | \mu_n, \sigma^2/\tau_n) = N(\mu | z, \sigma^2) N(\mu | \mu_n, \sigma^2/\tau_n)$ , and by using the Hint

$$\int N(\mu | z, \sigma^2) N(\mu | \mu_n, \sigma^2/\tau_n) d\mu = \int N(\mu | \text{const.}, \text{const.}) N \left( z | \mu_n, \sigma^2 \left[ \frac{1}{\tau_n} + 1 \right] \right) d\mu = N \left( z | \mu_n, \sigma^2 \left[ \frac{1}{\tau_n} + 1 \right] \right)$$

So

$$g(z|y) = \int \mathbf{N}\left(z|\mu_n, \sigma^2 \left[\frac{1}{\tau_n} + 1\right]\right) \mathbf{IG}(\sigma^2|a_n, k_n) d\sigma^2$$

by change the variable  $\xi = \sigma^2 \frac{1}{2k_n}$ , it is

$$g(z|y) = \int \mathbf{N}\left(z|\mu_n, \xi \left[\frac{1}{\tau_n} + 1\right] \frac{k_n}{a_n} 2a_n\right) \mathbf{IG}(\xi|\frac{2a_n}{2}, \frac{1}{2}) d\xi = T_1\left(z|\mu_n, \left[\frac{1}{\tau_n} + 1\right] \frac{k_n}{a_n}, 2a_n\right)$$

## 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  $x \sim \text{G}(a, b)$  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

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

**Solution.**

1. The parametric model is

$$\text{Ex}(x|\theta) = \theta \exp(-\theta x) 1(x \geq 0)$$

It is member of the exponential family

$$\text{Ef}_1(x|u, g, h, c, \phi, \theta, c) = u(x)g(\theta) \exp\left(\sum_{j=1}^k c_j \phi_j(\theta) h_j(x)\right)$$

with  $u(x_{1:n}) = 1$ ,  $g(\theta) = \theta$ ,  $c_1 = -1$ ,  $\phi_1(\theta) = \theta$ ,  $h_1(x) = x$ . The sufficient statistic is  $t_n = (n, \sum_{i=1}^n x_i)$ .

2. I can get the posterior by using the Bayes theorem

$$\begin{aligned} \pi(\theta|x) &\propto f(x|\theta) \pi(\theta|a, b) \propto \prod_{i=1}^n \text{Ex}(x_i|\theta) \text{Ga}(\theta|a, b) \\ &\propto \theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right) \theta^{a-1} \exp(-\theta b) \propto \theta^{a+n-1} \exp\left(-\theta \left(\sum_{i=1}^n x_i + b\right)\right) \\ &\propto \text{Ga}\left(\theta \left| \underbrace{a+n}_{=a^*}, \underbrace{b + \sum_{i=1}^n x_i}_{=b^*} \right.\right) \end{aligned}$$

3. By using the definition of the predictive distribution, it is ...

$$\begin{aligned}
 g(z|x) &= \int_{\mathbb{R}_+} f(z|\theta)\pi(\theta|x)d\theta \stackrel{z \geq 0}{=} \int_{\mathbb{R}_+} \theta \exp(-\theta z) \frac{(b^*)^{a^*}}{\Gamma(a^*)} \theta^{a^*-1} \exp(-\theta b^*) d\theta \\
 &= \frac{(b^*)^{a^*}}{\Gamma(a^*)} \int_{\mathbb{R}_+} \theta^{a^*+1-1} \exp(-\theta(z+b^*)) d\theta = \frac{(b^*)^{a^*}}{\Gamma(a^*)} \frac{\Gamma(a^*+1)}{(z+b^*)^{a^*+1}} = \frac{a^*(b^*)^{a^*}}{(z+b^*)^{a^*+1}} \\
 &= \frac{a^*(b^*)^{a^*}}{(z+b^*)^{a^*+1}}
 \end{aligned}$$

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

**Solution.**

1. There are  $k-1$  independent parameters in  $\text{Mu}_k(\theta)$  because  $\sum_{j=1}^k \theta_j = 1$ . I consider as parameters  $(\theta_1, \dots, \theta_{k-1})$  and the last one is a function of them as  $\theta_k = 1 - \sum_{j=1}^{k-1} \theta_j$ .

It is

$$\text{Mu}_k(x|\theta) = \prod_{j=1}^k \theta_j^{x_j} = \prod_{j=1}^{k-1} \theta_j^{x_j} (1 - \sum_{j=1}^{k-1} \theta_j)^{1 - \sum_{j=1}^{k-1} x_j} = (1 - \sum_{j=1}^{k-1} \theta_j) \exp\left(\sum_{j=1}^{k-1} x_j \log\left(\frac{\theta_j}{1 - \sum_{j=1}^{k-1} \theta_j}\right)\right)$$

This is the  $k - 1$  exponential family PDF with

$$\begin{aligned} u(x) &= 1; & g(\theta) &= (1 - \sum_{j=1}^{k-1} \theta_j); & c &= (1, \dots, 1) \\ h(x) &= (x_1, \dots, x_{k-1}); & \phi(\theta) &= (\log(\frac{\theta_1}{1 - \sum_{j=1}^{k-1} \theta_j}), \dots, \log(\frac{\theta_{k-1}}{1 - \sum_{j=1}^{k-1} \theta_j})), \end{aligned}$$

2. The likelihood is

$$f(x_{1:n}|\theta) = \prod_{i=1}^n \text{Mu}_k(x_i|\theta) = \prod_{j=1}^k \theta_j^{\sum_{i=1}^n x_{i,j}} = \prod_{j=1}^k \theta_j^{x_{*,j}} = (1 - \sum_{j=1}^{k-1} \theta_j)^n \exp \left( \sum_{j=1}^{k-1} x_{*,j} \log(\frac{\theta_j}{1 - \sum_{j=1}^{k-1} \theta_j}) \right)$$

and the sufficient statistic is

$$t_n = (n, x_{*,1}, \dots, x_{*,k-1})$$

3. It is

$$\pi(\theta|x_{1:n}) = \prod_{i=1}^n \text{Mu}_k(x_i|\theta) \text{Di}_k(\theta|a) \propto \prod_{j=1}^k \theta_j^{x_{*,j}} \prod_{j=1}^k \theta_j^{a_j-1} = \prod_{j=1}^k \theta_j^{x_{*,j}+a_{*,j}-1} \propto \text{Di}_k(\theta|\tilde{a})$$

where  $\tilde{a} = (\tilde{a}_1, \dots, \tilde{a}_k)$ , with  $\tilde{a}_j = a_j + x_{*,j}$  for  $j = 1, \dots, k$ . So the posterior is  $\theta|x_{1:n} \sim \text{Di}_k(\tilde{a})$ .

4. It is

$$\begin{aligned} p(y|x_{1:n}) &= \int \text{Mu}_k(y|\theta) \text{Di}_k(\theta|\tilde{a}) d\theta = \int \prod_{j=1}^k \theta_j^{y_j} \frac{\Gamma(\sum_{j=1}^k \tilde{a}_j)}{\prod_{j=1}^k \Gamma(\tilde{a}_j)} \prod_{j=1}^k \theta_j^{\tilde{a}_j-1} d\theta \\ &= \frac{\Gamma(\sum_{j=1}^k \tilde{a}_j)}{\prod_{j=1}^k \Gamma(\tilde{a}_j)} \int \prod_{j=1}^k \theta_j^{y_j+\tilde{a}_j-1} d\theta = \frac{\Gamma(\sum_{j=1}^k \tilde{a}_j)}{\prod_{j=1}^k \Gamma(\tilde{a}_j)} \frac{\prod_{j=1}^k \Gamma(y_j + \tilde{a}_j)}{\Gamma(\sum_{j=1}^k (y_j + \tilde{a}_j))} \\ &= \frac{\Gamma(\sum_{j=1}^k (a_j + x_{*,j}))}{\prod_{j=1}^k \Gamma(a_j + x_{*,j})} \frac{\prod_{j=1}^k \Gamma(y_j + a_j + x_{*,j})}{\Gamma(\sum_{j=1}^k (y_j + a_j + x_{*,j}))} \\ &= \frac{\Gamma(a_* + x_{*,*})}{\prod_{j=1}^k \Gamma(a_j + x_{*,j})} \frac{\prod_{j=1}^k \Gamma(y_j + a_j + x_{*,j})}{\Gamma(\sum_{j=1}^k y_j + a_* + x_{*,*})} \\ &= \frac{\Gamma(a_* + n)}{\prod_{j=1}^k \Gamma(a_j + x_{*,j})} \frac{\prod_{j=1}^k \Gamma(y_j + a_j + x_{*,j})}{\Gamma(1 + a_* + n)} \\ &= \frac{\Gamma(n + a_*)}{\Gamma(1 + a_* + n)} \prod_{j=1}^k \frac{\Gamma(y_j + a_j + x_{*,j})}{\Gamma(a_j + x_{*,j})} \end{aligned}$$

so  $p(y|x_{1:n}) = \frac{1}{n+a_*} (a_{j'} + x_{*,j'})$ , where  $j'$  such that  $y_{j'} = 1$ .