

# Exercise sheet 1

## Solutions

**Exercise 1.** Since the measurements are conditionally independent given  $\theta$ , we can use the result derived in the lecture notes (Section 1.3.1), that is, the posterior distribution of  $\theta$  is normal with mean and variance

$$\mu_n = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} \bar{y}_n + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2} \mu_0 \quad \text{and} \quad \sigma_n^2 = \frac{\sigma_0^2 \sigma^2}{n\sigma_0^2 + \sigma^2},$$

with  $\bar{y}_n = n^{-1} \sum_{i=1}^n y_i$ . In particular,  $n = 2$  since there are two measurements  $y_1 = 5$  and  $y_2 = 4$ , the corresponding variance is  $\sigma^2 = 9$  and the prior is defined by  $\mu_0 = 10$  and  $\sigma_0^2 = 4$ . It follows that

$$\mu_2 = \frac{2 \times 4}{2 \times 4 + 9} \times \frac{5 + 4}{2} + \frac{9}{2 \times 4 + 9} \times 10 = \frac{126}{17} \approx 7.41 \quad \text{and} \quad \sigma_2^2 = \frac{36}{17} \approx 2.12$$

The posterior mean  $\mu_2$  is a weighted average between the prior mean  $\mu_0$  and the observations  $y_1, y_2$ . Also, the posterior variance is smaller than the prior variance.

**Exercise 2.** 1. We first consider the likelihood of all the observations as follows

$$p_{\mathbf{y}_{1:n}|\theta}(y_1, \dots, y_n | \theta) = \prod_{k=1}^n p_{\mathbf{y}|\theta}(y_k | \theta) = \theta^n \exp(-\theta n \bar{y}_n)$$

with  $\bar{y}_n = n^{-1} \sum_{i=1}^n y_i$ . We proceed with the computation of the posterior distribution as follows

$$\begin{aligned} p_{\theta|\mathbf{y}_{1:n}}(\theta | y_1, \dots, y_n) &\propto p_{\mathbf{y}_{1:n}|\theta}(y_1, \dots, y_n | \theta) p_{\theta}(\theta) \\ &\propto \theta^n \exp(-\theta n \bar{y}_n) \theta^{\alpha-1} \exp(-\beta \theta) \\ &= \theta^{\alpha+n-1} \exp(-\theta(\beta + n \bar{y}_n)) \\ &\propto \text{Ga}(\theta; \alpha + n, \beta + n \bar{y}_n) \end{aligned}$$

which is a gamma distribution with parameters  $\alpha' = \alpha + n$  and  $\beta' = \beta + n \bar{y}_n$ . The posterior mean is

$$\mathbb{E}(\theta | \mathbf{y}_{1:n} = (y_1, \dots, y_n)) = \frac{\alpha'}{\beta'} = \frac{\alpha + n}{\beta + n \bar{y}_n} = \frac{\beta}{\beta + n \bar{y}_n} \frac{\alpha}{\beta} + \frac{n \bar{y}_n}{\beta + n \bar{y}_n} \frac{1}{\bar{y}_n},$$

where  $\alpha/\beta$  is the prior mean. It appears that, as  $n$  tends to infinity, the posterior mean tends to  $\bar{y}_n^{-1}$ , which can then be identified as the sample mean for  $\theta$ . The posterior distribution of  $\psi = \theta^{-1}$  can be found using the change of variable formula

$$p_{\psi|\mathbf{y}_{1:n}}(\psi | y_1, \dots, y_n) = p_{\theta|\mathbf{y}_{1:n}}(f^{-1}(\psi) | y_1, \dots, y_n) \left| \frac{df^{-1}}{d\psi}(\psi) \right|$$

with  $f : (0, \infty) \rightarrow (0, \infty)$  defined by  $f(\theta) = 1/\theta$ . It follows that

$$\begin{aligned} p_{\psi|\mathbf{y}_{1:n}}(\psi | y_1, \dots, y_n) &= \text{Ga}(1/\psi; \alpha + n, \beta + n \bar{y}_n) \frac{1}{\psi^2} \\ &= \frac{\beta'^{\alpha'}}{\Gamma(\alpha')} \psi^{-\alpha'-1} \exp(-\beta'/\psi). \end{aligned}$$

This distribution is referred to as the *inverse-gamma distribution*. We can compute its posterior expected value, or posterior mean, in the following way

$$\mathbb{E}(\psi | \mathbf{y}_{1:n} = (y_1, \dots, y_n)) = \int \psi p_{\psi | \mathbf{y}_{1:n}}(\psi | y_1, \dots, y_n) d\psi = \frac{\beta'^{\alpha'}}{\Gamma(\alpha')} \int \psi^{-(\alpha'-1)-1} \exp(-\beta'/\psi) d\psi$$

The integrand is proportional to yet another inverse-gamma distribution so we know the value of the integral to be the inverse of the normalising constant in that distribution, that is

$$\mathbb{E}(\psi | \mathbf{y}_{1:n} = (y_1, \dots, y_n)) = \frac{\beta'^{\alpha'}}{\Gamma(\alpha')} \frac{\Gamma(\alpha' - 1)}{\beta'^{\alpha'-1}} = \frac{\beta'}{\alpha' - 1} = \frac{\beta + n\bar{y}_n}{\alpha + n - 1} = \frac{\alpha - 1}{\alpha + n - 1} \frac{\beta}{\alpha - 1} + \frac{n}{\alpha + n - 1} \bar{y}_n$$

where  $\beta/(\alpha - 1)$  is the prior mean of  $\psi$  and  $\bar{y}_n$  is its sample mean.

2. The predictive distribution of  $\mathbf{y}_{n+1}$  given  $\mathbf{y}_{1:n} = (y_1, \dots, y_n)$  can be expressed as follows

$$\begin{aligned} p_{\mathbf{y}_{n+1} | \mathbf{y}_{1:n}}(y | y_1, \dots, y_n) &= \int p_{\mathbf{y}_{n+1} | \theta}(y | \theta) p_{\theta | \mathbf{y}_{1:n}}(\theta | y_1, \dots, y_n) d\theta \\ &= \int \text{Ex}(y; \theta) \text{Ga}(\theta; \alpha', \beta') d\theta \\ &= \frac{\beta'^{\alpha'}}{\Gamma(\alpha')} \int \theta^{(\alpha'+1)-1} \exp(-\theta(\beta' + y)) d\theta \end{aligned}$$

This integrand can be recognised to be a gamma distribution up to a normalising constant, so that

$$p_{\mathbf{y}_{n+1} | \mathbf{y}_{1:n}}(y | y_1, \dots, y_n) = \frac{\beta'^{\alpha'}}{\Gamma(\alpha')} \frac{\Gamma(\alpha' + 1)}{(\beta' + y)^{\alpha'+1}} = \frac{\alpha' \beta'^{\alpha'}}{(\beta' + y)^{\alpha'+1}}.$$

In order to verify the obtained expression is indeed a p.d.f., one can verify that its integral is equal to 1 (this is made easier by making the change of variable  $x = \beta' + y$ ). As a remark, this predictive distribution is a *Pareto distribution*.

**Exercise 3.** 1. The posterior distribution of  $\theta$  given the first  $n$  observations can be expressed up to a constant of proportionality as the product between the likelihood of all observations and the prior, that is

$$\begin{aligned} p_{\theta | \mathbf{y}_{1:n}}(\theta | y_1, \dots, y_n) &\propto \left[ \prod_{i=1}^n \theta^{y_i} \exp(-\theta) \right] \theta^{\alpha-1} \exp(-\beta\theta) \\ &= \theta^{n\bar{y}_n} \exp(-\theta n) \theta^{\alpha-1} \exp(-\beta\theta) \\ &= \theta^{\alpha+n\bar{y}_n-1} \exp(-\theta(\beta + n)) \end{aligned}$$

where, as usual,  $\bar{y}_n = n^{-1} \sum_{i=1}^n y_i$  is the sample mean of the  $n$  first observations. It follows that the posterior distribution is a gamma distribution with parameters  $\alpha' = \alpha + n\bar{y}_n$  and  $\beta' = \beta + n$  so that the posterior mean is

$$\mathbb{E}(\theta | \mathbf{y}_{1:n} = (y_1, \dots, y_n)) = \frac{\alpha'}{\beta'} = \frac{\alpha + n\bar{y}_n}{\beta + n} = \frac{\beta}{\beta + n} \frac{\alpha}{\beta} + \frac{n}{\beta + n} \bar{y}_n.$$

2. The predictive distribution of  $\mathbf{y}_{n+1}$  given  $\mathbf{y}_{1:n} = (y_1, \dots, y_n)$  can be expressed as follows

$$\begin{aligned} p_{\mathbf{y}_{n+1} | \mathbf{y}_{1:n}}(y | y_1, \dots, y_n) &= \int p_{\mathbf{y}_{n+1} | \theta}(y | \theta) p_{\theta | \mathbf{y}_{1:n}}(\theta | y_1, \dots, y_n) d\theta \\ &= \frac{\beta'^{\alpha'}}{\Gamma(\alpha') y!} \int \theta^y \exp(-\theta) \theta^{\alpha'-1} \exp(-\beta'\theta) d\theta \\ &= \frac{\beta'^{\alpha'}}{\Gamma(\alpha') y!} \int \theta^{\alpha'+y-1} \exp(-\theta(\beta' + 1)) d\theta \\ &= \frac{\Gamma(\alpha' + y)}{y! \Gamma(\alpha')} \beta'^{\alpha'} \frac{1}{(\beta' + 1)^{\alpha'+y}}, \end{aligned}$$

where the last line follows from identifying the integrand as being proportional to a gamma distribution with parameters  $\alpha' + y$  and  $\beta' + 1$ . As a remark, this last distribution can be identified as a negative-binomial distribution  $\text{NB}(\cdot; r, p)$  with parameters  $r > 0$  and  $p \in (0, 1)$ , defined as

$$\text{NB}(k; r, p) = \frac{\Gamma(k+r)}{k! \Gamma(r)} p^r (1-p)^k,$$

with the following identification of the variable and parameters:

$$k = y, \quad r = \alpha', \quad p = \frac{1}{\beta' + 1}.$$

The negative-binomial distribution as for mean  $pr/(1-p)$  and variance  $pr/(1-p)^2$  so that

$$\begin{aligned} \mathbb{E}(\mathbf{y}_{n+1} | \mathbf{y}_{1:n} = (y_1, \dots, y_n)) &= \frac{\alpha'}{\beta'} \\ \text{var}(\mathbf{y}_{n+1} | \mathbf{y}_{1:n} = (y_1, \dots, y_n)) &= \frac{\alpha'(\beta' + 1)}{\beta'^2}. \end{aligned}$$

**Exercise 4.** In Section 1.1.3 of the lecture notes, it was demonstrated that the normal-gamma prior is conjugate for a normal likelihood with unknown mean and variance and that the corresponding posterior distribution is of the form

$$p_{\mu, \tau | \mathbf{y}_{1:n}}(\mu, \tau | y_1, \dots, y_n) = \text{N}(\mu; \mu_n, (k_n \tau)^{-1}) \text{Ga}(\tau; \alpha_n, \beta_n)$$

with

$$\mu_n = \frac{k\mu_0 + n\bar{y}_n}{k+n}, \quad k_n = k+n, \quad \alpha_n = \alpha + \frac{n}{2}, \quad \beta_n = \beta + \frac{n\hat{v}}{2} + \frac{nk}{2(n+k)}(\mu_0 - \bar{y}_n)^2,$$

where the usual notations for the prior parameters have been used and where  $\hat{v} = n^{-1} \sum_{i=1}^n (y_i - \bar{y}_n)^2$ . It follows that the predictive distribution of  $\mathbf{y}_{n+1}$  given  $\mathbf{y}_{1:n} = (y_1, \dots, y_n)$  can be expressed as

$$\begin{aligned} p_{\mathbf{y}_{n+1} | \mathbf{y}_{1:n}}(y | y_1, \dots, y_n) &= \int \int p_{\mathbf{y}_{n+1} | \mu, \tau}(y | \mu, \tau) p_{\mu, \tau | \mathbf{y}_{1:n}}(\mu, \tau | y_1, \dots, y_n) d\mu d\tau \\ &= \int \int \text{N}(y; \mu, \tau^{-1}) \text{N}(\mu; \mu_n, (k_n \tau)^{-1}) \text{Ga}(\tau; \alpha_n, \beta_n) d\mu d\tau \\ &\propto \int \int \sqrt{\tau} \exp\left(-\frac{\tau}{2}(y - \mu)^2\right) \sqrt{\tau} \exp\left(-\frac{k_n \tau}{2}(\mu - \mu_n)^2\right) \tau^{\alpha_n - 1} \exp(-\beta_n \tau) d\mu d\tau \\ &\propto \int \int \tau^{\alpha_n + \frac{1}{2} - 1} \sqrt{\tau} \exp\left(\underbrace{-\frac{\tau}{2}(y - \mu)^2 - \frac{k_n \tau}{2}(\mu - \mu_n)^2 - \beta_n \tau}_{\doteq A(y, \mu, \tau)}\right) d\mu d\tau. \end{aligned}$$

The argument  $A(y, \mu, \tau)$  of the exponential can be rewritten as

$$\begin{aligned} A(y, \mu, \tau) &= -\frac{\tau}{2}(k_n \mu^2 - 2k_n \mu_n \mu + k_n \mu_n^2 + y^2 - 2y\mu + \mu^2) - \beta_n \tau \\ &= -\frac{\tau}{2}\left((k_n + 1)\mu^2 - 2(k_n \mu_n + y)\mu + \frac{(k_n \mu_n + y)^2}{k_n + 1} - \frac{(k_n \mu_n + y)^2}{k_n + 1} + k_n \mu_n^2 + y^2\right) - \beta_n \tau \\ &= -\frac{\tau(k_n + 1)}{2}\left(\mu - \frac{k_n \mu_n + y}{k_n + 1}\right)^2 - \frac{\tau}{2(k_n + 1)}(-k_n^2 \mu_n^2 - 2y k_n \mu_n - y^2 + k_n(k_n + 1)\mu_n^2 + (k_n + 1)y^2) \\ &= -\frac{\tau(k_n + 1)}{2}\left(\mu - \frac{k_n \mu_n + y}{k_n + 1}\right)^2 - \frac{\tau k_n}{2(k_n + 1)}(-2y\mu_n + \mu_n^2 + y^2) \\ &= -\frac{\tau(k_n + 1)}{2}\left(\mu - \frac{k_n \mu_n + y}{k_n + 1}\right)^2 - \frac{\tau k_n}{2(k_n + 1)}(y - \mu_n)^2 \end{aligned}$$

so that a normal distribution can be identified in the variable  $\mu$  as follows

$$p_{\mathbf{y}_{n+1} | \mathbf{y}_{1:n}}(y | y_1, \dots, y_n) \propto \int \tau^{\alpha_n + \frac{1}{2} - 1} \exp\left(-\tau\left[\beta_n + \frac{k_n}{2(k_n + 1)}(y - \mu_n)^2\right]\right) \int \sqrt{\tau} \exp\left(-\frac{\tau(k_n + 1)}{2}\left[\mu - \frac{k_n \mu_n + y}{k_n + 1}\right]^2\right) d\mu d\tau$$

so that the inner integral does not depend on  $y$  which means that it is just a normalising constant. Focusing on the first integral, whose integrand is proportional to a gamma distribution, it follows that

$$p_{\mathbf{y}_{n+1}|\mathbf{y}_{1:n}}(y | y_1, \dots, y_n) \propto \left(1 + \frac{1}{2\alpha_n} \frac{\alpha_n k_n}{\beta_n(k_n + 1)} (y - \mu_n)^2\right)^{-\frac{2\alpha_n + 1}{2}}$$

which is a generalised Student's t distribution with  $2\alpha_n$  degrees of freedom, with location parameter  $\mu_n$  and scale parameter  $\beta_n(k_n + 1)/(\alpha_n k_n)$ . As a remark, whenever  $\alpha_n > 1$ , the variance of  $\mathbf{y}_{n+1}$  given  $\mathbf{y}_{1:n} = (y_1, \dots, y_n)$  can be expressed as

$$\text{var}(\mathbf{y}_{n+1} | \mathbf{y}_{1:n} = (y_1, \dots, y_n)) = \frac{\beta_n(k_n + 1)}{\alpha_n k_n} \frac{2\alpha_n}{2\alpha_n - 2} = \frac{\beta_n}{\alpha_n - 1} + \frac{\beta_n}{k_n(\alpha_n - 1)},$$

so that the predictive variance equals the sum of the posterior expectation of the sampling variance  $\tau^{-1}$  and the posterior variance of the unknown mean  $\boldsymbol{\mu}$ , indeed,

$$\begin{aligned} \text{var}(\mathbf{y}_{n+1} | \mathbf{y}_{1:n} = y_{1:n}) &= \mathbb{E}(\text{var}(\mathbf{y}_{n+1} | \boldsymbol{\mu}, \boldsymbol{\tau}) | \mathbf{y}_{1:n} = y_{1:n}) + \text{var}(\mathbb{E}(\mathbf{y}_{n+1} | \boldsymbol{\mu}, \boldsymbol{\tau}) | \mathbf{y}_{1:n} = y_{1:n}) \\ &= \mathbb{E}(\boldsymbol{\tau}^{-1} | \mathbf{y}_{1:n} = y_{1:n}) + \text{var}(\boldsymbol{\mu} | \mathbf{y}_{1:n} = y_{1:n}) \\ &= \frac{\beta_n}{\alpha_n - 1} + \frac{\beta_n}{k_n(\alpha_n - 1)} \end{aligned}$$

the last line following from results in the lecture notes.