SOLUTIONS: Example Sheet 1

Christopher J. Moore

Jan 2024

1.

$$\mathcal{L}(\{x_i\}|X) = \frac{\exp\left(\frac{-\sum_{i=1}^{N}(X-x_i)^2}{2\sigma^2}\right)}{(2\pi\sigma^2)^{N/2}}$$

2. Using the notation for the indicator function introduced in lectures,

$$\pi(X) = \frac{1}{2\Delta}\mathbb{1}_{(-\Delta,+\Delta)}(X),$$
$$P(X|\{x_i\}) = \frac{\mathcal{L}(\{x_i\}|X)\pi(X)}{Z}.$$

3.

$$Z = \int_{-\Delta}^{+\Delta} \frac{\mathcal{L}(\{x_i\}|X)}{2\Delta}$$

4. Most likely (maximum *a posteriori*) value $\hat{X}$ is location of the of posterior, i.e.

$$\frac{\partial P(X|\{x_i\})}{\partial X}\bigg|_{X=\hat{X}} = 0, \tag{1}$$

If the posterior is smooth and differentiable. The mean and standard deviation are given by

$$\langle X\rangle = \frac{1}{Z}\int_{-\Delta}^{+\Delta} \mathrm{d}X\; X\mathcal{L}(\{x_i\}|X)$$
$$\sigma_X = \frac{1}{Z}\int_{-\Delta}^{+\Delta} \mathrm{d}X\; (X-\langle X\rangle)^2\mathcal{L}(\{x_i\}|X)$$

Provided the peak of the likelihood lies in the interval $(-\Delta, \Delta)$, the value of $\hat{X}$ will not depend on the choice of Δ. However, both $\langle X \rangle$ and σ_X do depend on Δ. Because the support of the likelihood (and hence posterior) is localised around $\langle X \rangle$, in the limit $\Delta \to \infty$ the posterior doesn't depend on Δ and approaches a Gaussian, $X|\{x_i\} \sim \mathcal{N}(\bar{x}, \sigma^2/N)$.

$$\hat{X} = \langle X \rangle \to \bar{x} \quad \text{where } \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \text{ is the sample mean}$$

$$\sigma_X \to \frac{\sigma^2}{N} \quad \text{(i.e. the width scales as } \propto N^{-1/2})$$

5. The conjugate prior is another Gaussian distribution, e.g. with mean μ and variance Σ^2. (Assume σ^2 is known.)

$$\begin{aligned} X|\mu,\Sigma &\sim \mathcal{N}(\mu, \Sigma^2) && \text{(Gaussian prior)} \\ x_i|X &\overset{\text{iid}}{\sim} \mathcal{N}(X, \sigma^2) \ \text{ for } i = 1, 2, \ldots, N && \text{(the likelihood } \mathcal{L}(\{x_i\}|X)) \\ X|\{x_i\},\mu,\Sigma &\sim \mathcal{N}(\mu', \Sigma') && \text{(Gaussian posterior)} \end{aligned}$$

Updating rule $\mu \to \mu' = \frac{\sigma^2\mu + N\Sigma^2\bar{x}}{\sigma^2 + N\Sigma^2}$ and $\Sigma^{-2} \to \Sigma'^{-2} = \Sigma^{-2} + N\sigma^{-2}$, where $\bar{x} = \frac{1}{N}\sum_i x_i$ is the sample mean.

6. No integration required. Comparing the prior and likelihood with the known form (already normalised) of Gaussian distribution allows us to find the evidence. Directly. After a bit of algebra

$$Z = \frac{\sigma}{\sqrt{\sigma^2 + N\Sigma^2}} (2\pi\sigma^2)^{-N/2} \exp\left(\frac{-1}{2}\left[\frac{\mu'^2}{\Sigma'^2} - \frac{\mu^2}{\Sigma^2} - \frac{\sum_i x_i^2}{\sigma^2}\right]\right), \quad (2)$$

where μ' and Σ' are defined in terms of μ, Σ, σ and N as above.

7. The inverse-gamma distribution is the conjugate prior for σ^2. (Assume X is known.)

$$\begin{aligned} \sigma^2|\alpha,\beta &\sim \text{InvGamma}(\alpha, \beta) && \text{(inv-gamma prior)} \\ x_i|\sigma^2 &\overset{\text{iid}}{\sim} \mathcal{N}(X, \sigma^2) \ \text{ for } i = 1, 2, \ldots, N && \text{(the likelihood)} \\ \sigma^2|\{x_i\},\alpha,\beta &\sim \text{InvGamma}(\alpha', \beta') && \text{(inv-gamma posterior)} \end{aligned}$$

Updating $\alpha \to \alpha' = \alpha + N/2$ and $\beta \to \beta' = \beta + \sum_{i=1}^{N}(x_i - X)^2/2$.

8. The product of the Gaussian prior for X used in part 5 and the inverse-gamma prior for σ^2 used in part 7 gives (in an alternative parametrisation to that on Wikipedia) the normal-inverse-gamma distribution. Despite being formed from the product of two distributions, this prior distribution is *not* separable; this is because the Gaussian prior depends on σ^2 which also appears in the inverse-gamma prior.

   This is an example of a 2D conjugate prior. It also shows why conjugate priors are rarely used in practice. Even in this simple 2D example, the conjugate prior has become restrictive and difficult to work with. Although it allows us to solve the Bayesian inference problem analytically, in practice it is usually easier more flexible and more interpretable to define a separable prior $\pi(X)\pi(\sigma^2)$ on the individual parameters and to solve the problem numerically using the techniques that will be described later in this course.