Supplementary material for Lesson 5

1 Normalizing Constants and Proportionality

The full expression for a posterior distribution of some parameter θ is given by

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int_{-\infty}^{\infty} f(x|\theta)f(\theta)d\theta} \,.$$

As we will see in coming lessons, it is often more convenient to work with the numerator only: $f(\theta|x) \propto f(x|\theta)f(\theta)$, which is the likelihood times the prior. The symbol $\propto$ stands for "is proportional to." We can multiply a function of θ by any constant and maintain proportionality. For example, if $f(\theta) = 5\theta$, then $f(\theta) \propto \theta$. However, $f(\theta)$ is not proportional to $\theta + 1$. We maintain proportionality only by modifying constants which are *multiplied by the entire function* $f(\theta)$. Hence $5(\theta + 1) \propto \theta + 1$.

The reason we can write $f(\theta|x) \propto f(x|\theta)f(\theta)$ is because the denominator $\int_{-\infty}^{\infty} f(x|\theta)f(\theta)d\theta$ is free of θ. It is just a normalizing constant.[1] Therefore, we can ignore any multiplicative terms not involving θ. For example, if $\theta \sim \mathrm{N}(m, s^2)$, then

$$\begin{aligned} f(\theta) &= \frac{1}{\sqrt{2\pi s^2}} \exp\left[-\frac{1}{2s^2}(\theta - m)^2\right] \\ &\propto \exp\left[-\frac{1}{2s^2}(\theta - m)^2\right] \,. \end{aligned} \tag{1}$$

Clearly, the expression in (1) does not integrate to 1 (it integrates to $\sqrt{2\pi s^2}$). Although it is not a PDF, it is proportional to the $\mathrm{N}(m, s^2)$ PDF and can be normalized to represent the $\mathrm{N}(m, s^2)$ distribution *only*. Likewise, the posterior $f(\theta|x)$ maintains its uniqueness as long as we specify it up to a proportionality constant.

To evaluate posterior quantities such as posterior probabilities, we will eventually need to find the normalizing constant. If the integral required is not tractable, we can often still simulate draws from the posterior and approximate posterior quantities. In some cases, we can identify $f(x|\theta)f(\theta)$ as being proportional to the PDF of some known distribution. This will be a major topic of Lesson 6.

[1]Remember also that in the posterior distribution of θ, we are treating x as a known constant