9.07 Introduction to Statistics for Brain and Cognitive Sciences
Emery N. Brown

Lecture 10 Bayesian Methods (Corrigenda)

Derivation of the Importance Sampling Algorithm

Here is the correct derivation of the importance sampling algorithm stated in Eqs. 10.59 to 10.60 and the correct statement of **Algorithm 10.2**.

Assume we wish to compute

$$E(g(\theta)) = \int g(\theta) f(\theta \mid x) d\theta. \tag{10.A1}$$

Given an importance density $h(\theta)$ that has the same support as $f(\theta \mid x),$ rewrite Eq. 10.A1 as

$$\begin{aligned} E(g(\theta)) &= \int g(\theta) \frac{f(\theta \mid x)}{h(\theta)} h(\theta) d\theta \\ &= f(x)^{-1} \int g(\theta) \frac{f(\theta) f(x \mid \theta)}{h(\theta)} h(\theta) d\theta \\ &= f(x)^{-1} \int g(\theta) w(\theta \mid x) h(\theta) d\theta \end{aligned} \tag{10.A2}$$

where $w(\theta \mid x) = h(\theta)^{-1} f(\theta) f(x \mid \theta).$ Similarly, we can write

$$\begin{aligned} 1 = \int f(\theta \mid x) d\theta &= f(x)^{-1} \int \frac{f(\theta) f(x \mid \theta)}{h(\theta)} h(\theta) d\theta \\ &= f(x)^{-1} \int w(\theta \mid x) h(\theta) d\theta. \end{aligned} \tag{10.A3}$$

We can use Eqs. 10.A2 and 10.A3 to write Eq. 10.A1 as

$$\begin{aligned} E(g(\theta)) = \frac{E(g(\theta))}{1} &= \frac{f(x)^{-1} \int g(\theta) w(\theta \mid x) h(\theta) d\theta}{f(x)^{-1} \int w(\theta \mid x) h(\theta) d\theta} \\ &= \frac{\int g(\theta) w(\theta \mid x) h(\theta) d\theta}{\int w(\theta \mid x) h(\theta) d\theta}. \end{aligned} \tag{10.A4}$$

We can approximate the numerator and denominator in Eq. 10.A4 respectively by n draws from $h(\theta)$ as

$$\int g(\theta)w(\theta \mid x)h(\theta)d\theta \doteq \frac{\sum_{i=1}^{n} g(\theta_i)w(\theta_i \mid x)}{n}$$

$$\int w(\theta \mid x)h(\theta)d\theta \doteq \frac{\sum_{i=1}^{n} w(\theta_i \mid x)}{n}. \qquad (10.A5)$$

Therefore, we have the importance sampling approximation of $E(g(\theta))$ as

$$E(g(\theta)) \doteq \frac{\sum_{i=1}^{n} g(\theta_i)w(\theta_i \mid x)}{\sum_{i=1}^{n} w(\theta_i \mid x)}. \qquad (10.A6)$$

We can simulate Eq. 10.59 (10.A1) with the following algorithm

Algorithm 10.2 (Importance Sampling)

$\text{Sum} = 0$
$W = 0$
$\text{For } j = 1,...,10,000$

1. Draw θ_j from $h(\theta)$
2. Compute $w(\theta_j) \mid x) = \frac{f(\theta_j)f(x \mid \theta_j)}{h(\theta_j)}$ and $g(\theta_j)$
3. $\text{Sum} \leftarrow \text{Sum} + w(\theta_j \mid x)g(\theta_j)$
4. $W \leftarrow W + w(\theta_j \mid x)$

Compute $E[g(\theta)] \doteq W^{-1}\text{Sum}$

MIT OpenCourseWare
https://ocw.mit.edu

9.07 Statistics for Brain and Cognitive Science
Fall 2016

For information about citing these materials or our Terms of Use, visit: https://ocw.mit.edu/terms.