

On ratios of normalizing constants*

Erwann Rogard

2022-04-04

Abstract

This is a slightly revised version of a note on RC's dated 2007-04-24 that came after that on the stochastic mesh¹, hence some gaps (where I spootd them, they are indicated in the source file)

1 Ratios of normalizing constants

In the first subsection, we assume that the pseudo-data points have been determined and explain how to estimate the expectations of interest for a particular (m_1, m_2) .

Let $\mathbf{y}_1 = \mathbf{y}_i^{m_1}$, $\mathbf{y}_2 = \mathbf{y}_{i+1}^{m_2}$, $\mathbf{y}_3 = (\mathbf{y}_i^{m_1}, \mathbf{y}_{i+1}^{m_2})$. We denote the likelihood densities $L_l(\xi) = p(\mathbf{y}_l|\xi)$, $l = 1, 2, 3$ and the posterior densities $p_l(\xi) = p(\xi|\mathbf{y}_l)$, $l = 0, \dots, 3$. Let ζ_0 and $q_0(\cdot)$ denote the normalizing constant and unnormalized density, respectively, for the prior i.e. $p_0(\xi) = q_0(\xi)/\zeta_0$. Let $\zeta_l = p(\mathbf{y}_l) = E[L_l(\xi)]$, $l = 1, 2, 3$. We denote the un-normalized posterior densities $q_l(\xi)$, $l = 1, 2, 3$ i.e. $p_l(\xi) = p_0(\xi)L_l(\xi)/\zeta_l = q_l(\xi)/(\zeta_0\zeta_l)$, $l = 1, 2, 3$.

By assumption, each of \mathbf{y}_l , $l = 1, 2, 3$ is associated with a known pseudo data point \mathbf{y}_l^* , $l = 1, 2, 3$. The same conventions apply, with the addition of superscript *. In addition, let $\pi_l(\cdot)$ denote a completely known approximation to $p_l^*(\cdot)$ for $l = 0, \dots, 3$ [1, Chapter 9].

Our problem is to estimate

$$r = \frac{\zeta_3}{\zeta_1\zeta_2} \tag{1}$$

under the constraint that, we can only use knowledge of $p_l^*(\cdot)$, $l = 1, 2, 3$ for constructing importance distributions and generate samples from them.

For the standard ratio of normalizing constant problem say ζ_1/ζ_2 , a number of related methods are available: importance sampling (IS), bridge sampling (BS), path sampling (PS), ratio importance sampling (RIS) and linked importance sampling (LIS). See Gelman and Meng, 1997 for the first three, Chen and Shao, 1997 for the fourth, and Neal, 2005 for the last one. A summary of the

*Revisions: https://github.com/erwannr/statistics/commits/main/ratio_nc

¹<https://github.com/erwannr/statistics/tree/main/stochmesh>

first four is provided by Chen et al., 2000. IS, BS and RIS are both conceptually and practically simpler than the other two which were developed to address challenging cases.

We define the relative error as

$$\text{RE}^2(\hat{r}) = \mathbb{E}[(\hat{r} - r)^2]/r^2 \quad (2)$$

1.1 Importance sampling

Observe that $\mathbb{E}_{\pi_l}[q_l(\xi)/\pi_l(\xi)] = \zeta_0 \zeta_l$ and therefore

$$r_{\text{IS}} = \frac{\mathbb{E}_{\pi_0}[q_0(\xi)/\pi_0(\xi)]\mathbb{E}_{\pi_3}[q_3(\xi)/\pi_3(\xi)]}{\mathbb{E}_{\pi_1}[q_1(\xi)/\pi_1(\xi)]\mathbb{E}_{\pi_2}[q_2(\xi)/\pi_2(\xi)]} \quad (3)$$

By estimating the expectations above by iid draws from the π_l 's the relative error is, by the delta method,

$$\text{RE}^2 = \sum_{l=0}^3 \frac{1}{n_l^*} \mathbb{E}_{\pi_l} \left[\left(\frac{p_l(\xi) - \pi_l(\xi)}{\pi_l(\xi)} \right)^2 \right] + o(1/n_*) \quad (4)$$

We now turn to another importance sampling formulation. Define $r^* = \frac{\zeta_3^*}{\zeta_1^* \zeta_2^*}$ and $\tilde{r} = r/r^*$. From the decomposition $p_l(\xi) = q_0(\xi)L(\mathbf{y}_l|\xi)/(\zeta_0 \zeta_l)$ we see that $\zeta_l/\zeta_l^* = \mathbb{E}_{p_l^*}[L_l(\xi)/L_l^*(\xi)]$ and

$$\tilde{r}_{\text{IS}} = \frac{\mathbb{E}_{p_3^*}[L_3(\xi)/L_3^*(\xi)]}{\mathbb{E}_{p_1^*}[L_1(\xi)/L_1^*(\xi)]\mathbb{E}_{p_2^*}[L_2(\xi)/L_2^*(\xi)]} \quad (5)$$

$$r_{\text{IS}}^* = \frac{\mathbb{E}_{p_1^*}[\pi_1(\xi)/q_1^*(\xi)]\mathbb{E}_{p_2^*}[\pi_2(\xi)/q_2^*(\xi)]}{\mathbb{E}_{p_0}[\pi_0(\xi)/q_0(\xi)]\mathbb{E}_{p_3^*}[\pi_3(\xi)/q_3^*(\xi)]} \quad (6)$$

Each estimate of the components of r^* will be reused across various combinations (j, m) therefore we can afford to spend large amounts of simulation draws to estimate them. The formulation (5) is more appealing than (3) because it involves only likelihood calculations which are sometimes amenable to approximations that are faster to compute.

1.2 Ratio importance sampling

We now consider a ratio importance sampling formulation and $p_{1,2}^*(\cdot)$, an arbitrary distribution of the form $p_{1,2}^*(\xi) = p_0(\xi)L_{1,2}(\xi)/\zeta_{1,2}^*$. Define

$$\tilde{r}_{\text{R}} = \frac{\mathbb{E}_{p_{1,2}^*}[L_1(\xi)L_2(\xi)/L_1^*(\xi)]}{\mathbb{E}_{p_{1,2}^*}[L_1(\xi)/L_{1,2}^*(\xi)]\mathbb{E}_{p_{1,2}^*}[L_2(\xi)/L_{1,2}^*(\xi)]} \quad (7)$$

and note that $r = \tilde{r}_{\text{R}} \times \zeta_{1,2}^*$. The advantage of this formulation is that if the same sample $\{\xi^k\}_{k=1}^{k_*} \stackrel{\text{iid}}{\sim} p_{1,2}(\xi)$ is used for estimating each expectation, we can reuse the integrands in the denominator to compute the numerator. The price, however, is that we have to determine the optimal $p_{1,2}^*$ for each combination of clusters (c_i, c_{i+1}) . Again, this cost can be amortized over all combination of data points $(\mathbf{y}_i^j, \mathbf{y}_{i+1}^m)$ which map to (c_i, c_{i+1}) .