Reverse marginal likelihood estimation

Vincent Fortuin

August 21, 2020

Bayes' theorem states that we can update our prior belief $p(w)$ given new data X and a likelihood $p(X|w)$ as

$$p(w|X) = \frac{p(X|w)p(w)}{p(X)} = \frac{p(X|w)p(w)}{\int_W p(X|w)p(w)}\,. \tag{1}$$

The term in the denominator is called *marginal likelihood* (or *evidence*) and is often rather a nuisance to the practitioner, since it is usually intractable to compute. In practice, one thus regularly resorts to techniques of performing approximate inference on the posterior that do not require evaluating this term.

However, when comparing different models, the marginal likelihood can come in extremely handy. Let's assume that we have two different priors, $p_1(w)$ and $p_2(w)$, and their associated marginal likelihoods $Z_1 = \int_W p(X|w)p_1(w)$ and $Z_2 = \int_W p(X|w)p_2(w)$. We can then compute the *Bayes factor* $\frac{Z_1}{Z_2}$, which is generally the most powerful statistical test to select one of the two priors over the other (c.f. Neyman-Pearson lemma).

In order to estimate these Bayes factors, we could just perform standard Monte Carlo estimation of the integrals involved, thas is,

$$Z_i \approx \frac{1}{N}\sum_{k=1}^{N} p(X|w_k) \quad \text{with} \quad w_k \sim p_i(w)\,. \tag{2}$$

As can be easily seen, this would require drawing N samples from each of the priors involved. In this note, we propose an alternative approach, which we call *reverse* marginal likelihood estimation. It will become clear that in this approach, we only have to draw N samples in total, regardless of the number of different priors.

In order for this approach to work, we define a very vague prior $\tilde{p}(w) = \mathcal{U}(-C, C)$, where C is some large constant, for instance $C = 10^{20}$. We

can then define a proposal distribution $q(w) = \frac{1}{\tilde{Z}} p(X|w)\tilde{p}(w)$, where $\tilde{Z}$ is a normalizing constant, such that $\tilde{Z} = \int_W p(X|w)\tilde{p}(w)$.

We can now use importance sampling to rewrite the marginal likelihoods of the other priors:

$$\begin{aligned} Z_i &= \mathbb{E}_{p_i}\left[p(X|w)\right] & &= \int_W p(X|w)p_i(w) \\ &= \int_W p(X|w)p_i(w)\frac{q(w)}{q(w)} & &= \mathbb{E}_q\left[\frac{p_i(w)}{q(w)}p(X|w)\right] \\ &= \mathbb{E}_q\left[\frac{p_i(w)\tilde{Z}}{p(X|w)\tilde{p}(w)}p(X|w)\right] & &= \mathbb{E}_q\left[\frac{\tilde{Z}}{\tilde{p}(w)}p_i(w)\right] \end{aligned}$$

Notice that by design, $\tilde{p}(w) = c$ for all $w \sim q(w)$ and some constant c. We can thus use linearity of expectation to get $Z_i = \frac{\tilde{Z}}{c}\mathbb{E}_q\left[p_i(w)\right]$ and therefore finally

$$\frac{Z_i}{Z_j} = \frac{\frac{\tilde{Z}}{c}\mathbb{E}_q\left[p_i(w)\right]}{\frac{\tilde{Z}}{c}\mathbb{E}_q\left[p_j(w)\right]} = \frac{\mathbb{E}_q\left[p_i(w)\right]}{\mathbb{E}_q\left[p_j(w)\right]}. \tag{3}$$

We can now draw N samples $w_k \sim q(w)$ and then use Monte Carlo estimation to approximate the Bayes factor as

$$\frac{Z_i}{Z_j} \approx \frac{\frac{1}{N}\sum_{k=1}^N p_i(w_k)}{\frac{1}{N}\sum_{k=1}^N p_j(w_k)} = \frac{\sum_{k=1}^N p_i(w_k)}{\sum_{k=1}^N p_j(w_k)}. \tag{4}$$