Reflection - Monte carlo
Nicholas Hoernle   **February 26, 2018**

Mackay presents two problems that we may wish to solve with Monte carlo methods. Specifically, for **problem 2**, we may wish to estimate expectations of functions $\phi(\mathbf{x})$ under an unknown distribution $P(\mathbf{x})$. I.e. $\mathbf{\Phi} = \langle \phi(\mathbf{x}) \rangle$. Importance sampling assumes the ease of evaluation of both potentially unnormalized Q^* and P^* (with respective normalizing constants Z_Q and Z_P) such that *importance weights* $w_r = \frac{P^*(\mathbf{x}^{(r)})}{Q^*(\mathbf{x}^{(r)})} = \frac{Z_P}{Z_Q} \frac{P(\mathbf{x}^{(r)})}{Q(\mathbf{x}^{(r)})}$ can be calculated for samples $\mathbf{x}^{(r)}$ drawn from Q. The Monte carlo approximator for $\mathbf{\Phi}$ is $\hat{\mathbf{\Phi}} = \frac{1}{R} \frac{Z_P}{Z_Q} \sum_{r=1}^{R} w_r \phi(\mathbf{x}^{(r)})$. Now note that $\frac{Z_P}{Z_Q} = \int \frac{P^*}{Q^*} Q dx$ (as $\int P^* dx = Z_P$ and $Z_Q = \frac{Q^*}{Q}$) which in turn has the Monte carlo estimator $\int \frac{P^*}{Q^*} Q dx \approx \frac{1}{R} \sum_{r=1}^{R} w_r$. Hence, the importance weights automatically account for the unnormalized target distribution in $\hat{\mathbf{\Phi}} = \frac{\sum_r w_r \phi(\mathbf{x}^{(r)})}{\sum_r w_r}$. For rejection sampling, the points (x, u) are accepted uniformly on P^*. This means that the *density* of the x samples must follow P.

LDA presents a high-dimensional inference problem. The posterior distribution of the hidden variables given a document $p(\theta, z | w, \alpha, \beta)$ implies that we require samples from a high dimension distribution (due to the number of latent variables in this problem - e.g. z is indexed over words in a document and θ is indexed over documents thus for even moderately sized corpora this is a very high dimension problem). Mackay discusses how, for rejection sampling, the number of rejected samples scales exponentially with the number of dimensions. Therefore, rejection sampling cannot be expected to make useful samples from the high dimensional posterior present in the LDA problem.