

1. Cautionary tale about importance sampling (10 points)

Suppose that we wish to estimate the normalizing constant Z_p for an un-normalized Gaussian $\tilde{p}(x) = \exp(-\frac{1}{2\sigma_p^2}x^2)$. In other words, we have $p(\cdot) \sim \mathcal{N}(\cdot|0, \sigma_p^2)$ with $p(x) = \tilde{p}(x)/Z_p$.

Given N i.i.d. samples $x^{(1)}, \dots, x^{(N)}$ from a standard normal $q(\cdot) \sim \mathcal{N}(0, 1)$, consider the importance sampling estimate:

$$\hat{Z} = \frac{1}{N} \sum_{i=1}^N \frac{\tilde{p}(x^{(i)})}{q(x^{(i)})}.$$

- Show that \hat{Z} is an unbiased estimator of Z_p .
- Letting $f(x) := \tilde{p}(x)/q(x)$, show that $\text{var}(\hat{Z}) = \frac{1}{N} \text{Var}(f(X))$ whenever $\text{Var}(f(X)) < \infty$.
- For what values of σ_p^2 is this variance finite?
Hint: Keep your variances non-negative!

2. Gibbs sampling and mean field variational inference (10 points)

Consider the Ising model with binary variables $X_s \in \{0, 1\}$ and a factorization of the form:

$$p(x; \eta) = \frac{1}{Z_p} \exp \left(\sum_{s \in V} \eta_s x_s + \sum_{\{s, t\} \in E} \eta_{st} x_s x_t \right).$$

- Derive the Gibbs sampling updates for this model.
- Consider fully factorized approximation q . Use the notation $q(X_s = 1) = \tau_s$. Derive an expression for $KL(q||p) - \log(Z_p)$.
Note: We can't directly compute the value of the KL divergence, as it would depend on the intractable quantity $\log(Z_p)$. When we subtract it, the remainder $KL(q||p) - \log(Z_p)$ can be computed efficiently.
- Derive the naive mean field updates, based on a fully factorized approximation. More specifically, we do cyclic coordinate descent on $KL(q||p)$, sequentially updating the parameter $\tau_s \in [0, 1]$. Note that Z_p is a constant with respect to the parameters τ_s .

3. Implementation of Gibbs sampling and mean field variational inference (20 points)

Follow the instructions in this Colab notebook:

<https://colab.research.google.com/drive/181XP5w5xEE2UCJU0bMEEDrpR77RYcz-U>