EXERCISES 5

Exercise 5.1. Implement the importance sampling method for estimating[1]

$$\mathbb{P}(X > 4),$$

where $X \sim \mathcal{N}(0,1)$. Try two methods: (i) $\hat{\varphi}^N_{\text{MC}}$ and (ii) $\hat{\varphi}^N_{\text{IS}}$. What kind of proposals you can choose for this besides the one given in Example 4.6? What is a good criterion for this example? Choose different proposals and test their efficiency in terms of getting a low relative error vs. samples.

Exercise 5.2. Minimum variance is a good design principle when choosing proposals. However, as can be seen from Example 4.9, this can be tedious. In this example, we will artificially simplify this process in order to not solve quadratic equations to find optimum. Let $p(x)$ be the exponential distribution with $\lambda = 1$, i.e.,

$$p(x) = e^{-x}, \qquad x \geq 0.$$

Assume that we would like to estimate the expectation of $\varphi(x)$ w.r.t. this distribution where

$$\varphi(x) = e^{-\frac{Kx}{2}}$$

where $K > 0$. While this test function is artificial, it is good for testing your skills. Consider the exponential proposal

$$q_\mu(x) = \mu e^{-\mu x}.$$

Find the optimal $\mu_\star$ in terms of K that minimises the variance of the IS estimate (see Example 4.9 in the lecture notes – but the computation here is much simpler!).

Once you obtain $\mu_\star$ in terms of K, then evaluate $\mathsf{var}_{q_{\mu_\star}}(\hat{\varphi}^N_{\text{IS}})$ and $\mathsf{var}_p(\bar{\varphi}^N_{\text{MC}})$ (MC is the case of the usual Monte Carlo estimator). Note that both cases do not require any sampling (find them analytically). Verify the variance reduction.

Exercise 5.3. Let $p(x) = \mathcal{N}(x; 0, 1)$ and $\varphi(x) = x$ (estimating the mean!).

- Verify that that the standard MC estimator, $X_1, \dots, X_N \sim p(x)$:

  $$\hat{\varphi}^N_{\text{MC}} = \frac{1}{N}\sum_{i=1}^N \varphi(X_i) = \frac{1}{N}\sum_{i=1}^N X_i,$$

  has variance $1/N$, i.e.,

  $$\mathsf{var}_p(\hat{\varphi}^N_{\text{MC}}) = \frac{1}{N}.$$

  What would be the variance if we had $p(x) = \mathcal{N}(x; \mu, \sigma^2)$?

- Let

  $$q_\lambda(x) = \mathcal{N}(x; 0, 1/\lambda).$$

  Find the minimum variance IS proposal, i.e., find $\lambda_\star$ that minimises the IS estimator variance (again follow Example 4.9 and no quadratic equations are needed for this case either). Show that the variance is less than $1/N$ at this value by computing the exact number.

[1] See Example 4.6 in Lecture notes.

- Implement the MC sampler and IS sampler using this optimal $\lambda_\star$. Compute the estimate of the mean ($= 0$) and verify that your estimators work. Then compute the variance of your estimator empirically (in other words, choose N and run the same experiment M times (Monte Carlo runs)). Verify that the IS estimator variance is, indeed, less than naive MC estimator variance!

Exercise 5.4. Let us consider the standard Gaussian model

$$\begin{aligned} p(x) &= \mathcal{N}(x; 0, 1) \\ p(y|x) &= \mathcal{N}(y; x, 1), \end{aligned}$$

and, again, we would like to estimate the marginal likelihood $p(y) = \int p(y|x)p(x)\mathrm{d}x$ using i.i.d Monte Carlo (MC) and importance sampling (IS). This is a quantity that computes the model's likelihood under a particular observation.

1. Write the analytical expression of $p(y)$ (this is provided in lecture notes – you don't have to prove it). Now assume that we observe $y = 9$. Compute $p(y = 9)$ using your exact $p(y)$ expression. (Hint: It is a small number)

2. What happened? We observed an unlikely observation that could come from the model. It is of practical interest to estimate such probabilities (unlikely events under the model). Next, implement the standard MC estimator (i.i.d) to estimate the integral $p(y = 9) = \int p(y = 9|x)p(x)\mathrm{d}x$ by drawing i.i.d samples from $p(x)$. Describe your test function and estimator clearly. Provide results for $N = 10, 100, 1000, 10000, 100000$ by computing the relative absolute error (RAE) (note that you know the true value from Part 1, therefore use that to compute your RAE) for these N and plot it w.r.t. N.

3. Next choose a proposal of the form
$$q(x) = \mathcal{N}\left(x; \mu, \frac{1}{2}\right).$$
For any observation y, find the minimum variance proposal mean, i.e., $\mu_\star$ that minimises the variance in terms of y. Then set $y = 9$ and compute $\mu_\star$ for this case. Implement the IS estimator with $\mu_\star$ as the proposal mean and compute the RAE for $N = 10, 100, 1000, 10000, 100000$ by computing the relative absolute error (RAE) (note that you know the true value from Part 1, therefore use that to compute your RAE) for these N and plot it w.r.t. N.

4. Compare the RAE for MC and IS to each other (provide a plot w.r.t. N by plotting them together). What is the difference? Comment about the difference about the accuracy of these two estimators.

   Use `plt.loglog` to plot throughout this exercise to have clear figures.

Exercise 5.5. Consider a computed log-weight vector

```
logW = [1000, 1001, 999, 1002, 950]
```

These are computed log-weights which are, for various reasons, often the only available quantities in practice (people implement quadratics, rather than Gaussians, for example). Implement the naive normalisation procedure

$$\mathsf{w}_i = \frac{\exp(\log \mathsf{W}_\mathsf{i})}{\sum_{i=1}^N \exp(\log \mathsf{W}_\mathsf{i})}.$$

Implement the trick introduced in Sec. 4.4.1 in lecture notes and verify that the latter computation is stable.