6. Exercise sheet *Bayesian inference and data assimilation*

Exercise 1. Consider the two sequences of measures $\mu_n = \delta_{1-2^{-n}}$ and $\nu_n = \frac{1}{n}\delta_{-1} + (1-\frac{1}{n})\delta_1$. From the mere look of it one would guess that both of these sequences approach the measure $\pi = \delta_1$ in some way. The KL divergence $KL(\mu|\nu)$ from μ to ν is defined as ∞ if there is an x with $\mu(x) > 0$ and $\nu(x) = 0$ and

$$KL(\mu|\nu) = \sum_{x \in \mathrm{supp}(\mu)} \mu(x) \log\left(\frac{\mu(x)}{\nu(x)}\right)$$

otherwise.

(i) Calculate $\mathrm{KL}(\pi|\mu_n)$ and $\mathrm{KL}(\pi|\nu_n)$. Do the sequences μ_n and ν_n converge to π w.r.t. the KL-divergence? I.e. do the sequences $\mathrm{KL}(\pi|\mu_n)$ and $\mathrm{KL}(\pi|\nu_n)$ approach 0?

(ii) For each μ_n and ν, find the optimal coupling T between μ_n (resp. ν_n) and π with the procedure from Example 2.29 (by hand). The Wasserstein-2 distance is then given as $\mathcal{W}(T) = \sqrt{\sum_{i,j} t_{ij}|a_i - a_j|^2}$. Do the sequences μ_n and ν_n converge to π in the Wasserstein-2 distance?

Exercise 2. Let $X \sim \mathcal{N}(1, 3)$ and $f(x) = 1 + 2x + x^2$.

(i) Calculate $\mathbb{E}[f(X)]$ and $\mathrm{Var}[f(X)]$ by hand.

We now approximate the expectation value of f in a Monte-Carlo fashion with M samples, i.e.

$$\mathbb{E}[f(X)] \approx f_M := \frac{1}{M}\sum_{i=1}^{M} f(x_i), \qquad x_i \sim X.$$

(ii) Calculate the expectation $\mathbb{E}[f_M]$ and variance $\mathrm{Var}[f_M]$ of the estimator f_M by hand. The result will depend on M.

(iii) Let $M = 1, 2, 4, \cdots, 256$. For each M, do $N = 10\,000$ simulations to approximate the expectation value using f_M. For each M, calculate the mean and the f_M. Which value should it take? For each M, calculate the variance of the f_M. Which value should it take. Make a plot with M on the x-axis and the variance of the estimates on the y-axis. Overlay the plot with the $\mathrm{Var}[f_M]$ that you calculated in (ii).

Exercise 3. You are given 4 samples $a_1 = 1, a_2 = 2, a_3 = 3, a_4 = 4$ with uniform weights $w_i = \frac{1}{4}$. This represents the measure $\mu_1 = \sum_{i=1}^{4} \frac{1}{4}\delta_{a_i}$. After doing a Bayesian update step and multiplying each weight by the likelihood you end up with new weights, $w_1' = \frac{5}{8}, w_2' = w_3' = w_4' = \frac{1}{8}$. This represents the measure $\mu_2 = \sum_{i=1}^{4} w_i'\delta_{a_i}$.

(i) What is the mean of μ_2?

In task (ii) and (iii) you get ways to generate samples from μ_2. Repeat each of these sampling procedures $N = 2000$ times. For each experiment, calculate the mean of the resulting 4 samples. In the end, you should now have N mean estimates, both for (ii) and (iii) respectively. Calculate the average mean estimate and the variance of the mean estimates for (ii) and (iii).

(ii) Use Algorithm 3.27 from the book to generate $L = 4$ new samples.

(iii) Use the procedure described in Example 2.29 to find a coupling between the measures/random variables described by the above samples and weights. I.e. find a coupling between X_1 and X_2 with $\mathbb{P}[X_1 = a_i] = w_i = \frac{1}{4}$ and $\mathbb{P}[X_2 = a_i] = w_i'$. Recall that $T_{ij} = \mathbb{P}[X_1 = a_i, X_2 = a_j]$. Use this joint distribution and the formulas for the conditional expectation to find the values of $\mathbb{P}[X_2 = a_j | X_1 = a_i]$. Now for each row of T, generate one new sample from X_2, i.e. sample from $\mathbb{P}[X_2 | X_1 = a_i]$ for each i.

Which estimation procedure is better in this case?