Bayesian inference and data assimilation, SS2019
Dr. Ester Mariucci, Dr. Han Cheng Lie, Mr. Jens Fischer
Institut für Mathematik, Universität Potsdam

If you have questions about the exercise sheet, write to hanlie@uni-potsdam.de. E-mails will be replied to during working hours.

You may use results from the previous Exercise Sheets. Justify each step in your proofs.

Exercise Sheet 5 (Due date: Thursday, 06.06.2019)

Exercise 1. (Monte Carlo integration; **4 points**). Let $\phi : [0,1] \to [0,1]$ be defined by

$$\phi(x) = \begin{cases} 2x & 0 \leq x \leq \frac{1}{2} \\ 2-2x & \frac{1}{2} < x \leq 1, \end{cases}$$

and consider for $n \in \mathbb{N}$ the function $f_n : [0,1] \to [0,1]$ defined by the n-times composition of ϕ with itself, i.e.

$$g_n(x) := \underbrace{\phi(\phi(\cdots(\phi}_{n \text{ times}}(x)\cdots)).$$

For this exercise, you may use the fact that for any $m \in \mathbb{N}$, $\mathbb{E}[g_m(X_1)] = \frac{1}{2}$.

(a) Plot the function g_n on the unit interval $[0,1]$ for $n = 4$ using a grid with step size $\Delta x = 0.005$. (Working code with comments - **0.5 point**; plot - **0.5 point**)

(b) Let $X_1 \sim \text{Unif}([0, \frac{1}{2}])$ and let $n = 10$. Using Monte Carlo integration, compute the empirical mean (i.e. the Monte Carlo estimator) of $\mathbb{E}[g_n(X_1)]$ using $M = 125, 250, 500, 1000, 2000, 4000, 8000$ samples drawn from the distribution of X_1. Use $L = 20$ samples of the empirical mean for each value of M to compute the mean square error (MSE). Repeat this procedure, using $X_2 \sim \text{Unif}([0, 0.3])$ instead of X_1. On the same graph, display a log-log plot of the MSE curves as a function of $1/M$. (Working code with comments - **1 point**; plot - **0.5 point**).

(c) Repeat part (b) for $n = 4$. (Working code with comments, plus plot - **0.5 point**)

(d) Describe the convergence behaviour that you observe in the plots from parts (b) and (c). Is the convergence behaviour the same in both plots? Explain your observations, using results and terminology from the lectures. (**1 point**)

Exercise 2. (The transform method; **1 point**). Let $X_1 \sim \text{Unif}([0,1])$, and let $X_2 \sim \text{Exponential}(\lambda)$ for $\lambda = 0.5$. Let T be the transport map given by the transform method (see the lecture notes for week 4). Draw $M = 5000$ sample values $(x_i)_{i=1}^M$

of X_1, apply the transform method to each sample to obtain another sequence $(T(x_i))_{i=1}^M$, and plot the frequency histogram of the $(T(x_i))_{i=1}^M$ over the interval $[t_{\min}, t_{\max}]$, where $t_{\min} := \min\{T(x_i); 1 \leq i \leq M\}$ and $t_{\max} := \max\{T(x_i); 1 \leq i \leq M\}$, using $N = 20$ bins of equal width. On the same figure, plot the PDF of X_2 and explain what you observe in the plot using results and terminology from the lectures (working code with comments - **0.5 point**; plot with observation and explanation - **0,5 point**).

Exercise 3. (Computing empirical probabilities; **2.5 points**) Let $(X_i)_{i\in\mathbb{N}}$ be i.i.d. copies of $X \sim \text{Exponential}(\lambda)$ for $\lambda = 0.5$. Define $\widehat{\mu}_M := \frac{1}{M}\Sigma_{i=1}^M X_i$.

(a) For $M = 20, 40, 80, 160, 320, 640$, sample $L = 500$ values of $\widehat{\mu}_M$ using your code for Exercise 2. For each value of M, compute the empirical probability that $\widehat{\mu}_M$ lies in the 95% confidence interval that was described in the lectures. Plot the empirical probability (linear scale) as a function of M (log scale). Is the empirical probability that $\widehat{\mu}_M$ lies in the 95% confidence interval always equal to 95%? (working code with comments - **0.5 point**; plot and correct answer to the question - **0.5 point**)

(b) Using your data from part (a) above, plot on a single figure the curves of the non-normalised histogram of $\widehat{\mu}_M$ for each value of M using $N = 40$ bins of equal width, using a different colour for each value of M. How do the curves of the histograms change as M increases? (working code with comments - **0.5 point**; plot and correct answers - **0.5 point**).

(c) Explain your answers to parts (a) and (b) using what you learned from the lectures. (correct explanations - **0.5 point**)

Exercise 4. (Convergence of random variables; **2.5 points**)

(a) Let $(X_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of $\mathbb{R}$-valued random variables with $X_n \sim \text{Ber}(p)$, and let $Y := 1 - X$ for $X \sim \text{Ber}(p)$ independent of X_n for all $n \in \mathbb{N}$.

  (i) Show that when $p = \frac{1}{2}$, that $X_n = Y$ in distribution for every $n \in \mathbb{N}$. (**0.5 point**)

  (ii) Let $p = \frac{1}{2}$. Calculate the largest possible number $\epsilon > 0$ such that $\mathbb{P}(|X_n - Y| \geq 1) \geq \epsilon$ for all $n \in \mathbb{N}$. (**0.5 point**)

  (iii) Explain why your answers from parts (i) and (ii) provide a counterexample to the claim that convergence in distribution implies convergence in probability. Make sure to use the definitions of the two types of convergence in your answer. (**0.5 point**)

(b) Let $Y \sim \text{Unif}([0,1])$ and let $X_n \sim \text{Unif}(\{\frac{k}{2^n} \mid k = 1, 2, \ldots, 2^n - 1, 2^n\})$. Show that X_n converges to Y in distribution. You must justify your answer, using the definition of convergence in distribution, the definition of Riemann integrability, and Lebesgue's condition for Riemann integrability. (**0.5 point**) What do you conclude about the convergence of random variables from this example? (**0.5 point**)