HOMEWORK 3 — APMA 1690 (FALL 2025)

Logistics. Collaboration is allowed (though you are encouraged to first attempt the problems on your own), but you must write the solution by yourself. It is okay to look at hints from sources and/or students but do not just copy. Submit your solution as pdf (written in latex) via Gradescope.

You must show your work (if you're unsure, go to office hours).

Due date: October 14, before beginning of class.

The precise definition of $a \bmod m$ is one of the following two equivalent definitions (you can use either one).

- $a \bmod m = a - \lfloor \frac{a}{m} \rfloor m$ where $\lfloor x \rfloor$ is the floor function, that is, the largest integer $\leq x$.
- $a \bmod m$ is the unique number $b \in [0, m)$ such that $a = b + km$ for some integer k.

1. Exercises

(1) (**1pt**) Compute by hand:

(a) $3 \bmod 7$
(b) $7 \bmod 7$
(c) $10 \bmod 7$
(d) $10^9 \bmod 7$
(e) $0 \bmod 7$
(f) $-(3 \bmod 7)$
(g) $(-3) \bmod 7$

(2) (**1pt**)

(a) Suppose a, b, m are all integers. Prove that $(ab) \bmod m = ((a \bmod m)(b \bmod m)) \bmod m$.

(b) Choose $r_0 \in \{1, \dots, m-1\}$ and set $r_{n+1} := (ar_n) \bmod m$. Show that

$$ar_n \pmod{m} = a^{n+1} r_0 \pmod{m}.$$

Hint. You need to use induction.

(3) (**2pt**) Let $E \subseteq \mathbb{R}^d$ be a bounded set and let f, g be probability densities over E. Let $h : E \to \mathbb{R}$. In class we computed the variance of the importance sampling estimator,

$$\operatorname{Var}\left[\frac{1}{N}\sum_{n=1}^{N}\frac{f(X^{(n)})}{g(X^{(n)})}h(X^{(n)})\right] = \frac{1}{N}\operatorname{Var}_{X\sim g}\left[\frac{f(X)}{g(X)}h(X)\right], \tag{1.1}$$

where $\{X^{(n)}\}_{n=1}^{N} \overset{\text{i. i. d.}}{\sim} g$. In this exercise we will find the function g which minimizes the variance (1.1), which is equivalent to minimizing $\operatorname{Var}_{X\sim g}[\frac{f(X)}{g(X)}h(X)]$ (because $\frac{1}{N}$ doesn't depend on g). Define $g^* : E \to \mathbb{R}$ by

$$g^*(x) = \frac{|h(x)|f(x)}{\int_E |h(y)|f(y)dy}. \tag{1.2}$$

(a) Show that

$$\operatorname{argmin}_g \operatorname{Var}_{X\sim g}\left[\frac{f(X)}{g(X)}h(X)\right] = \operatorname{argmin}_g E_{X\sim g}\left[\left(\frac{f(X)}{g(X)}h(X)\right)^2\right].$$

(b) Compute $E_{X\sim g^*}\left[\left(\frac{f(X)}{g^*(X)}h(X)\right)^2\right]$.

(c) Show that for all g,

$$E_{X\sim g}\left[\left(\frac{f(X)}{g(X)}h(X)\right)^2\right] \geq E_{X\sim g^*}\left[\left(\frac{f(X)}{g^*(X)}h(X)\right)^2\right].$$

Explain how the combination of (a),(b),(c) shows that g^* is optimal. Also explain why despite that, g^* will not be used in practice.

(4) (**2pt**)

(a) Consider the PDF

$$p(x) = \frac{1}{2\sqrt{x}}1_{x\in(0,1)}.$$

Write a Python code which samples from p by using the transportation of measure approach from class, that is, using a uniform random variable on $(0,1)$ and the inverse CDF of p. Include the code, as well as a plot comparing the histogram, using $N = 10^4$ samples, with the analytic formula for p.

(b) Let the target distribution f be the Beta distribution on $(0,1)$ with parameters $\alpha = 2$ and $\beta = 5$,

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)},$$

where $B(\alpha,\beta)$ is the Beta function. You can get the distribution f in Python using scipy.stats. Let the proposal distribution be

$$g(x) = \frac{1}{2\sqrt{x}}1_{x\in(0,1)}.$$

Write a Python code that implements rejection sampling to sample from the target f using the proposal g. Include the code, as well as a plot comparing the histogram, using $N = 10^4$ samples, with the analytic formula for f.

Hint. Use Desmos plots to figure out which M to choose.

(5) (**2pt**) Let $h : [0,1]^{100} \to \mathbb{R}$ be given by

$$h(x_1,\ldots,x_{100}) = \left|\sin\left(2\pi x_1 \sum_{i=1}^{100} x_i\right)\right| \left(\cos\left(2\pi x_2 \sum_{j=1}^{100} x_j^2\right)\right)^2.$$

Write a Python code which estimates $\int_{[0,1]^{100}} h(x)dx$ using standard Monte Carlo where the target f is the uniform distribution on $[0,1]^{100}$. Include:

(i) Your code

(ii) Compute the estimate $\frac{1}{N}\sum_{n=1}^{N} h(X^{(n)})$ for $N = 10^6$ where $\{X^{(n)}\}_{n=1}^{N}$ are i.i.d. from f. Because N is very large we will think of the estimate $\frac{1}{N}\sum_{n=1}^{N} h(X^{(n)})$ as the "true" $\int_{[0,1]^{100}} h(x)f(x)$.

(iii) Compute the estimate

$$\frac{1}{N}\sum_{n=1}^{N}[h(X^{(n)})]^2 - \left(\frac{1}{N}\sum_{n=1}^{N} h(X^{(n)})\right)^2$$

for $N = 10^6$ where $\{X^{(n)}\}_{n=1}^{N}$ are i.i.d. from f. Because N is very large we will think of the estimate $\frac{1}{N}\sum_{n=1}^{N}[h(X^{(n)})]^2 - \left(\frac{1}{N}\sum_{n=1}^{N} h(X^{(n)})\right)^2$ as the "true" $\mathrm{Var}_{X\sim f}[h(X)]$.

(iv) Let $N = 100$ and estimate $E\left[\frac{1}{N}\sum_{n=1}^{N} h(X^{(n)})\right]$ by averaging over $M = 1000$ realizations of $\frac{1}{N}\sum_{n=1}^{N} h(X^{(n)})$. What is the relation with part (ii)?.

(v) Let $N = 100$ and estimate the variance $\mathrm{Var}\left[\frac{1}{N}\sum_{n=1}^{N} h(X^{(n)})\right]$ by averaging over $M = 1000$ realizations of

$$\frac{1}{N}\sum_{n=1}^{N}[h(X^{(n)})]^2 - \left(\frac{1}{N}\sum_{n=1}^{N} h(X^{(n)})\right)^2.$$

What is the relation with part (iii)?

(6) (**2pt**) Let $p = \mathbb{P}[X > 4]$ where $X \sim f := \mathcal{N}(0,1)$.

(i) Classical Monte Carlo would approximate p, using $h(x) := 1_{x>4}$, by

$$\hat{p} := \frac{1}{N}\sum_{n=1}^{N} h(X^{(n)}) = \frac{1}{N}\sum_{n=1}^{N} 1_{\{X^{(n)}>4\}}, \qquad \{X^{(n)}\}_{n=1}^{N} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1).$$

Explain why Classical Monte Carlo is likely to not be efficient in this case.

(ii) Instead of using Classical Monte Carlo, we use importance sampling using a proposal distribution g which an exponential distribution with parameter 1, shifted to the right by 4, so that its samples are around the points we are trying to estimate. The weight function in this case is

$$w(x) = \frac{f(x)}{g(x)} = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2 + (x-4)\right).$$

Write a python code that estimates $\hat{p}$ using classical Monte Carlo **and** importance sampling with g. Repeat both estimates for $N = 10^4$ to $N = 10^6$ with jumps of 2×10^4 (50 points overall) and plot both estimates, on the same figure, as a function of N. Comment on the plot.