MA3227 Numerical Analysis II

Assignment 4

Simon Etter, 2019/2020
Deadline: 17 April 2020, 18.00
Total marks: 10

1 Uniformly distributed points on the disk [5 marks]

Lecture 23 introduced the transformation and rejection sampling techniques for sampling according to a given target distribution $\mathcal{F}$. The goal of this exercise is to use these techniques for sampling $\operatorname{Uniform} D$, where $D = \{x \in \mathbb{R}^2 \mid \|x\|_2 \leq 1\}$ denotes the unit disk and $\operatorname{Uniform} S$ for a two-dimensional set $S \subset \mathbb{R}^2$ is defined through

$$\big(\operatorname{Uniform} S\big)(A) = \frac{\operatorname{area}(A)}{\operatorname{area}(S)} \qquad \text{for all } A \subset S.$$

1. Complete `randdisk_rejection()` such that it samples $X \sim \operatorname{Uniform} D$ using rejection sampling with proposal distribution $\mathcal{G} = \operatorname{Uniform}[-1,1]^2$.
2. How many samples of $\operatorname{Uniform}[-1,1]^2$ does `randdisk_rejection()` require on average to produce a single sample of $\operatorname{Uniform} D$?
3. Let $\mathcal{R}$ be the distribution on $[0,1]$ given by $\mathcal{R}([a,b]) = \int_a^b 2r\,dr$ for all $[a,b] \subset [0,1]$. Show that if $R \sim \mathcal{R}$ and $\Phi \sim \operatorname{Uniform}[0,2\pi]$ independently, then

   $$X(R,\Phi) = \begin{pmatrix} R\,\cos(\Phi) \\ R\,\sin(\Phi) \end{pmatrix} \sim \operatorname{Uniform} D.$$

   Hint. You may use that for any[1] $f : D \to \mathbb{R}$, we have that

   $$\int_D f(x)\,dx = \int_0^1 \int_0^{2\pi} f\big(X(r,\phi)\big)\,r\,d\phi\,dr \qquad .$$

4. Complete `randdisk_transformation()` such that it samples $X \sim \operatorname{Uniform} D$ using the result from Task 3.

Hint. You can check your answers to Tasks 1 and 4 using `plot_samples()`.

2 Importance sampling for highly concentrated integrals [5 marks]

We have seen in Lecture 22 that $\int_0^1 f(x)\,dx = \mathbb{E}[f(X)]$ for $X \sim \operatorname{Uniform}[0,1]$ and that the integral on the right-hand side can be estimated using Monte Carlo sampling. In this task, we will first revise the Monte Carlo error estimate discussed at the end of Lecture 24 (Tasks 1 to 3), and then we will see how for a specific class of functions $f(x)$, the sampling efficiency can be improved using the importance sampling theorem introduced in Lecture 23 (Tasks 4 to 7).

[1] More precisely, the statement holds for any integrable $f : D \to \mathbb{R}$, but characterising integrable functions is beyond the scope of this module and all practically relevant functions are integrable.

1. Complete `uniform_sampling(f,N)` such that it uses Monte Carlo sampling to estimate the expectation and variance of $f(X)$ with $X \sim \text{Uniform}[0,1]$.

2. Complete `plot_histogram()` as described in the code.

3. (Unmarked) The provided function `sin_integral()` uses `uniform_sampling()` to estimate

$$\int_0^1 \tfrac{\pi}{2}\,\sin(\pi\,x)\,dx = 1, \tag{1}$$

and `plot_histogram()` to visualise the accuracy of the estimates. Use this function to check your answers to Tasks 1 and 2. If your answers are correct, you will observe that $\tilde{\mathbb{E}}_{100}[f(X)]$ achieves a relative accuracy of roughly 10%.

4. (Unmarked) The provided function `concentrated_integral()` uses `uniform_sampling()` to estimate

$$\int_0^1 \tfrac{1}{0.0906401}\,\exp\Big(-\big(20\,(x-0.5)\big)^4\Big)\,dx \approx 1. \tag{2}$$

Run this function and observe how the relative accuracy of $\tilde{\mathbb{E}}_{100}[f(X)]$ drops to about 50% when applied to (2) instead of (1).

5. (Unmarked) The importance sampling theorem from Lecture 23 can easily be generalised to the following statement.

Importance sampling theorem. Assume X and Y are random variables distributed according to probability densities $p_X(x)$ and $p_Y(x)$, respectively. Then,

$$\mathbb{E}[f(X)] = \mathbb{E}\big[f(Y)\,\tfrac{p_X(Y)}{p_Y(Y)}\big]. \tag{3}$$

Our aim for the remainder of this exercise is to use this result with a $p_Y(x)$ chosen such that Monte Carlo sampling applied to the expectation on the right-hand side of (3) becomes as accurate as possible. We have seen in Lecture 23 that this is equivalent to minimising the variance of $f(Y)\,\frac{p_X(Y)}{p_Y(Y)}$, and clearly this variance is minimised if we choose $p_Y(x) = C\,f(x)\,p_X(x)$ for some $C > 0$ since then $f(Y)\,\frac{p_X(Y)}{p_Y(Y)} = \frac{1}{C}$ has variance 0. Unfortunately, in the case of (2) we have

$$f(x)\,p_X(x) = \tfrac{1}{0.0906401}\,\exp\Big(-\big(20\,(x-0.5)\big)^4\Big),$$

and it is unclear how we could sample $p_Y(x) = C\,f(x)\,p_X(x)$ efficiently. Instead, we will resort to choosing Y normally distributed with mean $\mu = 0.5$ and standard deviation $\sigma = 0.05$, i.e. $Y \sim \mathcal{N}(0.5, 0.05^2)$. Run `plot_concentrated_integrand()` and observe how $p_Y(x)$ and $f(x)$ have roughly the same shape and hence $\frac{f(Y)}{p_Y(Y)}$ is reasonably close to being constant.

6. Complete `importance_sampling(f,N,m,s)` such that it uses Monte Carlo sampling to estimate the expectation and variance of $f(Y)\,\frac{p_X(Y)}{p_Y(Y)}$ where $Y \sim \mathcal{N}(m, s^2)$ and $p_X(x)$, $p_Y(x)$ denote the probability density functions of Uniform$[0,1]$ and $\mathcal{N}(m, s^2)$, respectively.

   Hints. You can sample $Y \sim \mathcal{N}(m, s^2)$ using `y = m + s*randn()`. $p_Y(x)$ is implemented in `normal_pdf()`.

7. (Unmarked) Uncomment the line `plot_histogram(... importance_sampling ...)` in `concentrated_integral()`. Observe how importance sampling achieves a much better accuracy than uniform sampling in the case of (3).