Exercise Session: Sampling

Exercise 1: Classical Sampling Methods

1. Apply the inversion of the cdf method to simulate a random variable Y with a Weibull distribution of density $3x^2e^{-x^3}$ over $\mathbb{R}^+$.

   The function $f : x \mapsto 3x^2e^{-x^3}$ is a density over $\mathbb{R}_+$, since it is the derivative of $x \mapsto -e^{-x^3}$ and we have
$$\int_0^\infty 3x^2e^{-x^3}\,dx = \left[-e^{-x^3}\right]_0^\infty = 1.$$

   The CFD associated with this probability distribution is the function F defined by
$$F(x) = \int_{-\infty}^{x} 3y^2e^{-y^3}\mathbf{1}_{y\geq 0}\,dy = \left[-e^{-y^3}\right]_0^x = 1 - e^{-x^3}.$$

   The inverse of F can be computed as follows. For any $u \in (0,1)$ and $x \in \mathbb{R}_+$, we have
$$\begin{aligned} F(x) = u &\iff 1 - e^{-x^3} = u \iff 1 - u = e^{-x^3} \\ &\iff x = \big[-\log(1-u)\big]^{1/3}. \end{aligned}$$

   To sample from the density f, we can therefore apply the inversion of the CDF method by simulating $U \sim \mathrm{Unif}([0,1])$ and outputting $F^{-1}(U) = \big[-\log(1-U)\big]^{1/3}$. Noting that U has the same law as $1-U$, one can also output $\big[-\log(U)\big]^{1/3}$ instead.

2. Compute the distribution of $C\sqrt{-\ln(U)}$, if $U \sim \mathrm{Unif}([0,1])$ and for any constant $C > 0$. Deduce a way to simulate a Rayleigh distribution, with density $\frac{x}{\sigma^2}e^{-x^2/2\sigma^2}$ over $\mathbb{R}^+$.

   We use the change of variable technique to identify the law of $\sqrt{-\ln(U)}$, if $U \sim \mathrm{Unif}(0,1)$ and let $X = C\sqrt{-\ln(U)}$. Let g be a bounded measurable function. We use the change of variable $x = C\sqrt{-\log(u)}$, i.e. $u = \exp(-(x/C)^2)$, and associated with
$$dx = C \cdot \frac{-1/u}{2\sqrt{-\log(u)}}du = -\frac{C}{2(x/C)}\exp\big((x/C)^2\big)du = -\frac{C^2}{2x}\exp\big((x/C)^2\big)du$$

   or equivalently $du = -2\dfrac{x}{C^2}\exp\big(-(x/C)^2\big)dx$. We obtain
$$\begin{aligned} \mathbb{E}\big[g(X)\big] = \mathbb{E}\Big[g\Big(\sqrt{-\log(U)}\Big)\Big] &= \int_0^1 g\Big(\sqrt{-\log(u)}\Big)du \\ &= \int_\infty^0 g(x)\Big[-2\frac{x}{C^2}\exp\big(-(x/C)^2\big)\Big]dx \\ &= \int_0^\infty g(x)\cdot 2\frac{x}{C^2}\exp\big(-(x/C)^2\big)dx. \end{aligned}$$

   To simulate a Rayleigh random variable with density $x \mapsto \frac{2x}{\sigma^2}\exp\big(-\frac{x^2}{2\sigma^2}\big)$, it suffices to output $\sigma\sqrt{2}\sqrt{-\log(U)}$ where $U \sim \mathrm{Unif}([0,1])$.

Exercise 2: Other Integrals

Write the following integrals as an expectation, and provide Monte-Carlo methods to estimate them:

1. $\int_0^1 \sin(\sqrt{x})dx$.

   Output $\frac{1}{N}\sum_{i=1}^N \sin(\sqrt{U_i})$ where $U_1,\dots,U_N \sim \text{Unif}([0,1])$ are iid.

2. $\int_0^{+\infty} \sin(x^4)\exp(-2x)\exp(-x^2/2)dx$ (propose two methods for this integral).

   We have

$$\int_0^{+\infty} \sin(x^4)\exp(-2x)\exp(-x^2/2)dx = \int_{-\infty}^{+\infty} \mathbf{1}_{x\geq 0}\sqrt{2\pi}\sin(x^4)\exp(-2x)\cdot \underbrace{\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)}_{\text{Density of } N(0,1)}\,dx$$

   One can output $\frac{\sqrt{2\pi}}{N}\sum_{i=1}^N \mathbf{1}_{X_i\geq 0}\sin(X_i^4)\exp(-2X_i)$ where $X_1,\dots,X_N \sim N(0,1)$ are iid.

   Alternatively, we also have

$$\int_0^{+\infty} \sin(x^4)\exp(-2x)\exp(-x^2/2)dx = \int_{-\infty}^{+\infty} \frac{1}{2}\sin(x^4)\exp(-x^2/2)\cdot \underbrace{2\exp(-2x)\mathbf{1}_{x\geq 0}}_{\text{Density of Exp(2)}}\,dx.$$

   Therefore, one can output $\frac{1}{2N}\sum_{i=1}^N \sin(Y_i^4)\exp(-Y_i^2/2)$ where $Y_1,\dots,Y_N \sim \text{Exp}(2)$ are iid.

Exercise 3: Importance sampling

1. Suppose that p is the $\mathcal{N}(0,1)$ distribution, and that $f(x) = \exp\left(-\frac{(x-10)^2}{2}\right)$. Find the optimal importance sampling density q.

   Hint: a random variable X is distributed as $N(\mu,\sigma^2)$ iff its density is *proportional to* $x \mapsto \exp\big(-\frac{(x-\mu)^2}{2\sigma^2}\big)$.

   We recall that the optimal importance sampling density in this case is

$$g^*(x) = \frac{|f(x)|p(x)}{\int_{\mathbb{R}}|f(x)|p(x)}.$$

   The denominator is a constant. If we prove that the numerator is proportional to an expression of the form $\exp\big(-\frac{(x-\mu)^2}{2\sigma^2}\big)$, then we will have proved that the optimal importance sampling density is that of $N(\mu,\sigma^2)$. We have

$$\begin{aligned}|f(x)|p(x) &= \exp\left(-\frac{(x-10)^2}{2}\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\\ &= \frac{1}{\sqrt{2\pi}}\exp\left(-x^2+10x-50\right)\\ &= \frac{1}{\sqrt{2\pi}}\exp\left(-(x-5)^2-25\right)\\ &= \frac{1}{\sqrt{2\pi}}e^{-25}\exp\left(-\frac{(x-5)^2}{2\times 1/2}\right).\end{aligned}$$

   The last expression is a constant times the function $x\mapsto \exp\big(-\frac{(x-\mu)^2}{2\sigma^2}\big)$ where $\mu = 5$ and $\sigma^2 = 1/2$. We have therefore obtained that the optimal importance sampling density is the density of $N(5,\frac{1}{2})$.

2. Suppose that p is the $\mathcal{N}(0,1)$ distribution, and that f is $\exp(kx)$ for $k \neq 0$. Find the optimal importance sampling density q.

   We use the same method for the rest of the exercise: We compute $|f(x)|p(x)$ and identify two parameters μ, σ^2 such that $|f(x)|p(x)$ is proportional to $\exp\big(-\frac{(x-\mu)^2}{2\sigma^2}\big)$. We have

$$\begin{aligned}|f(x)|p(x) &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}+kx\right)\\ &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-k)^2}{2}+\frac{k^2}{2}\right)\\ &\propto \exp\left(-\frac{(x-k)^2}{2}\right).\end{aligned}$$

   This is proportional to the density of $N(k,1)$, hence g^* is the density of $N(k,1)$.

3. Let p be the $\mathcal{N}(0,I)$ distribution in dimension $d \geq 1$.

   (a) Generalize Question 1 to the case where f is the density of $\mathcal{N}(\theta, I)$.
   Hint: a random variable $X \in \mathbb{R}^d$ is distributed as $N(\mu, \sigma^2 I)$ iff its density is *proportional to* $x \mapsto \exp\big(-\frac{\|x-\mu\|^2}{2\sigma^2}\big)$.
   We have

$$\begin{aligned}|f(x)|p(x) &\propto \exp\left(-\frac{\|x-\theta\|^2}{2}\right)\exp\left(-\frac{\|x\|^2}{2}\right)\\ &= \exp\left(-\|x\|^2+\theta^\top x-\frac{\|\theta\|^2}{2}\right)\\ &= \exp\left(-\left\|x-\frac{\theta}{2}\right\|^2-\frac{\|\theta\|^2}{4}\right)\\ &= e^{-\|\theta\|^2/4}\exp\left(-\frac{\|x-\theta/2\|^2}{2\times 1/2}\right).\end{aligned}$$

   The last expression is a constant times the function $x \mapsto \exp\big(-\frac{\|x-\mu\|^2}{2\sigma^2}\big)$ where $\mu = \frac{\theta}{2}$ and $\sigma^2 = 1/2$. We have therefore obtained that the optimal importance sampling density is the density of $N(\theta/2, \frac{1}{2}I)$, hence g^* is the density of $N(\theta/2, \frac{1}{2}I)$.

   (b) Generalize Question 2 to the case where $f(x) = \exp(k^\top x)$ for $k \in \mathbb{R}^d$.
   We have

$$\begin{aligned}|f(x)|p(x) &\propto \exp\left(-\frac{\|x\|^2}{2}+k^\top x\right)\\ &= \exp\left(-\frac{\|x-k\|^2}{2}+\frac{\|k\|^2}{2}\right)\\ &\propto \exp\left(-\frac{\|x-k\|^2}{2}\right).\end{aligned}$$

   This is proportional to the density of $N(k,I)$, hence g^* is the density of $N(k,I)$.

Exercise 4: Rejection method

We define the following quantities.

- Let X be random variable with probability density $f(x) = \frac{1}{Z}\exp(-x^3)$ over $\{x \in \mathbb{R} : x \geq 1\}$. Here, the normalizing constant $Z = \int_1^{+\infty} e^{-x^3} dx$ is not given explicitly, but its value does not matter for this exercise.
- Let Y be a random variable with probability density $g(x) = \frac{1}{x^2}$ over $\{x \in \mathbb{R} : x \geq 1\}$.

1. Give a procedure to simulate Y using the inversion of the CDF method.

   We compute the CDF of Y. We denote by f_Y the density of the random variable Y. This density can be expressed as $f_Y(x) = g(x)1_{g\geq 1}$. For any $y \geq 1$, we have

   $$F_Y(y) = \int_1^y g(x)dx = \int_1^y \frac{1}{x^2}dx = \left[-\frac{1}{x}\right]_1^y = 1 - \frac{1}{y}.$$

   To compute the inverse of F_Y, we can write, for any $u \in (0,1)$ and $y \geq 1$:

   $$F(y) = u \Longleftrightarrow 1 - \frac{1}{y} = u \Longleftrightarrow y = \frac{1}{1-u}.$$

   Therefore, to simulate Y, it suffices to output $\frac{1}{1-U}$ where $U \sim \mathrm{Unif}([0,1])$. The variable $1-U$ has the same distribution as U, hence it also suffices to output $\frac{1}{U}$.

2. Show that the function $x \mapsto \frac{f(x)}{g(x)}$ is decreasing over $\{x \in \mathbb{R} : x \geq 1\}$.
   For any $x \geq 1$, we have

   $$\frac{f(x)}{g(x)} = \frac{1}{Z}\frac{\exp(-x^3)}{1/x^2} = \frac{1}{Z}x^2\exp(-x^3).$$

   The constant Z is independent of x. To show that this function is decreasing, we define the function $h(x) = x^2\exp(-x^3)$ and compute its derivative:

   $$h'(x) = 2x\exp(-x^3) + x^2 \cdot (-3x^2)\exp(-x^3) = (2 - 3x^3)x\exp(-x^3).$$

   Since $x \geq 1$, we have $3x^2 \geq 3 \cdot 1^2 \geq 2$, hence $h' < 0$ and h is decreasing over $[1, \infty)$.

3. Deduce that $f(x) \leq \dfrac{1}{eZ}g(x)$ for any $x \geq 1$.

   Since h is decreasing over $[1,\infty)$, we have $h(x) \leq h(1) = 1/(eZ)$, or equivalently, $f(x) \leq \frac{1}{eZ}g(x)$ for any $x \geq 1$.

4. Deduce a rejection algorithm to simulate $X \sim f$ using the auxiliary law g.

   We first check the assumptions needed to apply the rejection method

   - There exists a function g and a constant $c = \frac{1}{eZ}$ such that $f \leq cg$,
   - g is easy to sample from
   - It is easy to evaluate the ratio $f(x)/(cg(x)) = ex^2\exp(-x^3)$.

   To obtain one sample from f, it suffices to sample iid random variables $Y_1, \dots Y_n, \dots$ with density g and $U_1, \dots, U_n, \dots$ iid uniform random variables independent of $Y_1, \dots, Y_n, \dots$ and to output the first Y_N such that $U_N \leq \frac{f(Y_N)}{cg(Y_N)} = eY_N^2\exp(-Y_N^3)$.