Advanced Simulation

Problem Sheet 2

Exercise 1

Consider the following "squeeze" rejection algorithm for sampling from a distribution with density $\pi(x) = \widetilde{\pi}(x)/Z_\pi$ on a state space $\mathbb{X}$ such that

$$h(x) \leq \widetilde{\pi}(x) \leq M\widetilde{q}(x)$$

where h is a non-negative function, $M > 0$ and $q(x) = \widetilde{q}(x)/Z_q$ is the density of a distribution that we can easily sample from. The algorithm proceeds as follows.

- **(a)** Draw independently $X \sim q$, $U \sim \mathcal{U}_{[0,1]}$.
- **(b)** Accept X if $U \leq h(X)/(M\widetilde{q}(X))$.
- **(c)** If X was not accepted in step (b), draw an independent $V \sim \mathcal{U}_{[0,1]}$ and accept X if

$$V \leq \frac{\widetilde{\pi}(X) - h(X)}{M\widetilde{q}(X) - h(X)}.$$

1. Show that the probability of accepting a proposed $X = x$ in either step (b) or (c) is

$$\frac{\widetilde{\pi}(x)}{M\widetilde{q}(x)}.$$

2. Deduce from the previous question that the distribution of the samples accepted by the above algorithm is π.
3. Show that the probability that step (c) has to be carried out is

$$1 - \frac{\int_{\mathbb{X}} h(x)\,dx}{MZ_q}.$$

4. Let $\widetilde{\pi}(x) = \exp\left(-x^2/2\right)$ and $\widetilde{q}(x) = \exp(-|x|)$. Using the fact that

$$\widetilde{\pi}(x) \geq 1 - \frac{x^2}{2}$$

for any $x \in \mathbb{R}$, how could you use the squeeze rejection sampling algorithm to sample from $\pi(x)$. What is the probability of not having to evaluate $\widetilde{\pi}(x)$? Why could it be beneficial to use this algorithm instead of the standard rejection sampling procedure?

Exercise 2

Optional. Do not hand in. Solutions will be released.

Consider the following algorithm known as Marsaglia's polar method.

- **Step a**: Generate independent U_1, U_2 according to $\mathcal{U}_{[-1,1]}$ until $Y = U_1^2 + U_2^2 \leq 1$.

- **Step b**: Define
$$Z = \sqrt{-2\log(Y)}$$
and return
$$X_1 = Z\frac{U_1}{\sqrt{Y}},\ X_2 = Z\frac{U_2}{\sqrt{Y}}.$$

1. Define $\vartheta = \text{arctan2}(U_1, U_2)$, where $\text{atan2} : \mathbb{R}^2 \to (-\pi, \pi]$, or $(0, 2\pi]$, sends $(x, y) \in \mathbb{R}^2$ to the angle θ the vector (x, y) forms with the positive x-axis. The difference with the standard $\arctan(y/x)$ function whose range is $(-\pi/2, \pi/2]$ is that it retains the information about the signs of x and y, which is lost when computing y/x.

   Show that the joint distribution of Y and ϑ has density
$$f_{Y,\vartheta}(y, \theta) = \mathbb{1}_{[0,1]}(y)\frac{\mathbb{1}_{[0,2\pi]}(\theta)}{2\pi}.$$

2. Show that X_1 and X_2 are independent standard normal random variables.

3. What are the potential benefits of this approach over the Box-Muller algorithm?

Exercise 3

Optional. Do not hand in. Solutions will be released.

Consider two probability densities π, q on $\mathbb{X}$ such that $\pi(x) > 0 \Rightarrow q(x) > 0$ and assume that you can easily draw samples from q. Whenever $\pi(x)/q(x) \leq M < \infty$ for any $x \in \mathbb{X}$, it is possible to use rejection sampling to sample from π. When M is unknown or when this condition is not satisfied, we can use importance sampling techniques to approximate expectations with respect to π. However it might be the case that most samples from q have very small importance weights.

Rejection control is a method combining rejection and importance weighting. It relies on an arbitrary threshold value $c > 0$. We introduce the notation $w(x) = \pi(x)/q(x)$ and
$$Z_c = \int_{\mathbb{X}} \min\{1, w(x)/c\}\, q(x)\, dx.$$

Rejection control proceeds as follows.

- **Step a.** Generate independent $X \sim q$, $U \sim \mathcal{U}_{[0,1]}$ until $U \leq \min\{1, w(X)/c\}$.
- **Step b**. Return X.

1. Give the expression of the probability density $q^*(x)$ of the accepted samples.

2. Prove that
$$\mathbb{E}_{q^*}\left([w^*(X)]^2\right) = Z_c \mathbb{E}_q(\max\{w(X), c\}\, w(X))$$
   where $w^*(x) = \pi(x)/q^*(x)$.

3. Establish that
$$\mathbb{E}_q(\min\{w(X), c\})\, \mathbb{E}_q(\max\{w(X), c\}\, w(X)) \leq \mathbb{E}_q(\min\{w(X), c\} \max\{w(X), c\}\, w(X))$$
   (Hint. Show first that for any $c > 0, w_1 > 0, w_2 > 0$
$$h(w_1, w_2) = [\min\{w_1, c\} - \min\{w_2, c\}]\,[w_1 \max\{w_1, c\} - w_2 \max\{w_2, c\}] \geq 0.$$
   This is related to the Harris inequality).

4. Deduce from the results established in (2) and (3) that
$$\mathbb{V}_{q^*}(w^*(X)) \leq \mathbb{V}_q(w(X)).$$

Exercise 4

We want to use Monte Carlo methods to approximate the integral

$$I = \int_{\mathbb{X}} \phi(x) \pi(x) \, dx$$

where $\phi : \mathbb{X} \to \mathbb{R}$ and π is a probability density on $\mathbb{X}$. Assume we have access to a proposal probability density q such that $w(x) = \pi(x) / q(x) \leq M < \infty$ for any $x \in \mathbb{X}$.

1. Consider the extended probability density $\overline{\pi}_{X,U}$ on $\mathbb{X} \times [0,1]$ defined as

$$\overline{\pi}_{X,U}(x,u) = \begin{cases} Mq(x) & \text{for } x \in \mathbb{X}, u \in \left[0, \frac{w(x)}{M}\right] \\ 0 & \text{otherwise.} \end{cases}$$

   Verify that $\overline{\pi}_X(x) = \pi(x)$.

2. Using the identity

$$I = \int_0^1 \int_{\mathbb{X}} \phi(x) \overline{\pi}_{X,U}(x,u) \, dx du,$$

   give the expression of the normalised importance sampling estimate $\widehat{I}_n$ of I when using n independent samples (X_i, U_i) such that $(X_i, U_i) \sim \overline{q}_{X,U}$ where $\overline{q}_{X,U}(x,u) = q(x) \times \mathbb{I}_{[0,1]}(u)$ (that is under $\overline{q}_{X,U}$ we have $X \sim q$, $U \sim \mathcal{U}_{[0,1]}$ and X and U are independent). Express this estimate as a function the importance weight function

$$\overline{w}(x,u) = \frac{\overline{\pi}_{X,U}(x,u)}{\overline{q}_{X,U}(x,u)}.$$

   Show that this estimate is equivalent to the estimate one would obtain by sampling from π using rejection sampling using n proposals from q.

3. Show that

$$\mathbb{V}_q(w(X)) \leq \mathbb{V}_{\overline{q}_{X,U}}(\overline{w}(X,U)).$$

Exercise 5

Let us consider the normal multivariate density on $\mathbb{R}^d$ with identity covariance, that is

$$\pi(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2} x^{\mathrm{T}} x\right).$$

We write $\mathbb{E}$ and $\mathbb{V}$ for the expectation and variance under π.

1. (Cameron-Martin formula). Show that for any $\theta \in \mathbb{R}^d$ and function $\phi : \mathbb{R}^d \to \mathbb{R}$

$$\mathbb{E}[\phi(X)] = \mathbb{E}\left[\phi(X+\theta) \exp\left(-\frac{1}{2}\theta^{\mathrm{T}}\theta - \theta^{\mathrm{T}} X\right)\right].$$

2. It follows directly from the Cameron-Martin formula and the strong law of large numbers that, for independent $X_1, ..., X_n \sim \pi$, the estimator

$$\widehat{I}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \phi(X_i + \theta) \exp\left(-\frac{1}{2}\theta^{\mathrm{T}}\theta - \theta^{\mathrm{T}} X_i\right)$$

   of $\mathbb{E}[\phi(X)]$ is strongly consistent for any $\theta \in \mathbb{R}^d$ such that $\mathbb{E}\left[\left|\phi(X+\theta)\exp\left(-\frac{1}{2}\theta^{\mathrm{T}}\theta - \theta^{\mathrm{T}}X\right)\right|\right] < \infty$. The case $\theta = (0, ..., 0)^{\mathrm{T}}$ corresponds to the usual Monte Carlo estimate. The variance of $\widehat{I}_n(\theta)$ is given by $\sigma^2(\theta)/n$ where

$$\sigma^2(\theta) = \mathbb{V}\left[\phi(X+\theta) \exp\left(-\frac{1}{2}\theta^{\mathrm{T}}\theta - \theta^{\mathrm{T}} X\right)\right].$$

We assume in the sequel that $\sigma^2(\theta) < \infty$ for any θ.

Show that

$$\sigma^2(\theta) = \mathbb{E}\left[\phi^2(X)\exp\left(-\frac{1}{2}X^{\mathrm{T}}X + \frac{1}{2}(X-\theta)^{\mathrm{T}}(X-\theta)\right)\right] - (\mathbb{E}[\phi(X)])^2$$

3. A twice differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ is strictly convex if $\nabla^2 f(\theta)$ (called the Hessian of f) is a positive definite matrix for any $\theta \in \mathbb{R}^d$. Deduce from the expression of $\sigma^2(\theta)$ given in (2) that the function $\theta \to \sigma^2(\theta)$ is strictly convex.

4. Show that the minimum of $\theta \to \sigma^2(\theta)$ is reached at θ^* such that

$$\mathbb{E}\left[\phi^2(X)(\theta^* - X)\exp\left(-\theta^{*\mathrm{T}}X\right)\right] = 0.$$

5. We apply the previous results to a simple model of European options in a Black-Scholes model. We want to compute

$$I = \exp(-rT)\,\mathbb{E}[\max\{0, \lambda\exp(\sigma X) - K\}]$$

where $X \sim \mathcal{N}(0,1)$ and r, λ, K, σ, T are positive real numbers such that $\lambda < K$. Show that $\theta \to \sigma^2(\theta)$ is decreasing on $D = \left(-\infty, \sigma^{-1}\log(K/\lambda)\right)$. Deduce from this result that there exists a range of values of θ such that the variance of $\widehat{I}_n(\theta)$ is strictly lower than the variance of the usual Monte Carlo estimate.

Exercise 6

Let $\mathbb{X}$ be a finite state-space. Consider the following Markov transition kernel

$$T(x,y) = \alpha(x,y)\,q(x,y) + \left(1 - \sum_{z\in\mathbb{X}}\alpha(x,z)\,q(x,z)\right)\,\delta_x(y)$$

where $q(x,y) \geq 0$, $\sum_{y\in\mathbb{X}} q(x,y) = 1$ and $0 \leq \alpha(x,y) \leq 1$ for any $x, y \in \mathbb{X}$. $\delta_x(y)$ is the Kronecker symbol; i.e. $\delta_x(y) = 1$ if $y = x$ and zero otherwise.

1. Explain how you would simulate a Markov chain with transition kernel T.

2. Let π be a probability mass function on $\mathbb{X}$. Show that if

$$\alpha(x,y) = \frac{\gamma(x,y)}{\pi(x)\,q(x,y)}$$

where $\gamma(x,y) = \gamma(y,x)$ and $\gamma(x,y)$ is chosen such that $0 \leq \alpha(x,y) \leq 1$ for any $x, y \in \mathbb{X}$ then T is $\pi-$reversible.

3. Show that the Metropolis-Hastings algorithm corresponds to a particular choice of $\gamma(x,y)$.

4. Let π be a probability mass function on the finite space $\mathbb{X}$ such that $\pi(x) > 0$ for any $x \in \mathbb{X}$. To sample from π, we run a Metropolis-Hastings chain $\left(X^{(t)}\right)_{t\geq 1}$ with proposal $q(x,y) \geq 0$, such that $\sum_{y\in\mathbb{X}} q(x,y) = 1$ and $q(x,x) = 0$ for any $x \in \mathbb{X}$. Consider here the sequence $\left(Y^{(k)}\right)_{k\geq 1}$ of accepted proposals: $Y^{(1)} = X^{(\tau_1)}$ where $\tau_1 = 1$ and, for $k \geq 2$, $Y^{(k)} = X^{(\tau_k)}$ where $\tau_k := \min\left\{t : t > \tau_{k-1},\ X^{(t)} \neq Y^{(k-1)}\right\}$.

Let $\phi : \mathbb{X} \to \mathbb{R}$ be a test function. Show that the estimate $\frac{1}{\tau_k - 1}\sum_{t=1}^{\tau_k - 1}\phi\left(X^{(t)}\right)$ can be rewritten as a function of $\left(Y^{(k)}\right)_{k\geq 1}$ and $(\tau_k)_{k\geq 1}$ and prove that the sequence $\left(Y^{(k)}\right)_{k\geq 1}$ is a Markov chain with transition kernel

$$K(x,y) = \frac{\alpha(x,y)\,q(x,y)}{\sum_{z\in\mathbb{X}}\alpha(x,z)\,q(x,z)}.$$

5. Show that the transition kernel $K(x, y)$ of the Markov chain $\left(Y^{(k)}\right)_{k \geq 1}$ is $\widetilde{\pi}$-reversible where

$$\widetilde{\pi}(x) = \frac{\pi(x)\, m(x)}{\sum_{z \in \mathbb{X}} \pi(z)\, m(z)}$$

with

$$m(x) := \sum_{z \in \mathbb{X}} \alpha(x, z)\, q(x, z).$$

6. Assume that for some test function $\phi : \mathbb{X} \rightarrow \mathbb{R}$ we have $\frac{1}{k} \sum_{i=1}^{k} \phi\left(Y^{(i)}\right) \rightarrow \sum_{x \in \mathbb{X}} \phi(x)\, \widetilde{\pi}(x)$ almost surely and additionally assume that $m(x)$ can be computed exactly for any $x \in \mathbb{X}$.

   Propose a strongly consistent estimate of $\sum_{x \in \mathbb{X}} \phi(x)\, \pi(x)$ based on the Markov chain $\left(Y^{(k)}\right)_{k \geq 1}$ which does not rely on $(\tau_k)_{k \geq 1}$.

1 Programming questions

1. Consider the genetic linkage model as in the slides of Lecture 3. Sample some simulated data with a fixed value of θ of your choice. Implement rejection sampling and reproduce the histograms of the posterior of θ and the waiting time before acceptance. Experiment with different proposal distributions.

2. Implement a sampler to draw from a mixture of Gaussians

$$\pi(x) = \omega_1 \phi(x; \mu_1, \sigma_1^2) + \omega_2 \phi(x; \mu_2, \sigma_2^2)$$

   where ϕ is the Gaussian pdf. You are allowed to use R's Gaussian generator (but feel free to reimplement Box-Muller from Lecture 3 or Marsaglia's method from Question 1 of this sheet, just for fun).

3. Let

$$h(x) = [\cos(50x) + \sin(20x)].$$

   We consider estimating $\int_0^1 h(x)dx$ through Monte Carlo methods.

   - first of all, what is the exact answer, to accuracy within 10^{-4}?
   - Can you implement rejection sampling with a uniform proposal?
   - Find a way to assess how good you are doing.
   - Implement an importance sampling solution with a smart proposal (*hint: plot h and find a matching q*).