Advanced Simulation

Problem Sheet 2

Exercise 1 (Monte Carlo for Gaussians)

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 2 (Metropolis-Hastings)

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 $\phi : \mathbb{X} \to \mathbb{R}$ we have $\frac{1}{k}\sum_{i=1}^{k} \phi\left(Y^{(i)}\right) \to \sum_{x\in\mathbb{X}} \phi(x)\,\widetilde{\pi}(x)$ 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 $(\tau_k)_{k\geq 1}$.

Exercise 3 (Metropolis-Hastings)

Consider the following $\mathbb{X}$-valued Markov chain $(X_t)_{t\geq 1}$. It evolves over time as follows. At time t, with probability $\alpha(X_{t-1})$ sample

$$X_t \sim q(\cdot)$$

where $q(x)$ is a probability density function and otherwise set $X_t := X_{t-1}$. Hence its transition kernel is given by

$$K(x,y) = \alpha(x)\,q(y) + (1 - \alpha(x))\,\delta_x(y)$$

where $\delta_x(y)$ is the Dirac mass located at x.

1. Show that if
$$\int_{\mathbb{X}} \frac{q(x)}{\alpha(x)} dx < \infty$$
then K admits a stationary distribution of density
$$\pi(x) \propto \frac{q(x)}{\alpha(x)}.$$

2. Assume that $0 \leq \alpha(x) = \alpha < 1$ then it can be easily shown that a central limit theorem holds for $\frac{1}{t}\sum_{i=1}^{t} X_i$ as long as $\sigma^2 := \mathbb{V}_q[X_1] < \infty$. Compute the asymptotic variance $\sigma_X^2 = \mathbb{V}[X_1] + 2\sum_{k=2}^{\infty} \mathbb{C}ov[X_1, X_k]$ in the stationary regime as a function of α and σ^2.

   *(**Hint.** First prove that the marginal distribution of X_k is q for all k, then find a recursion formula for $\mathbb{C}ov(X_1, X_k)$.)*

Exercise 4 (Gibbs Sampler)

Suppose that we wish to use the Gibbs sampler on
$$\pi(x, y) \propto \exp\left(-\frac{1}{2}(x-1)^2 (y-2)^2\right).$$

1. Write down the two "full" conditional distributions associated to $\pi(x, y)$.
2. Does the resulting Gibbs sampler make any sense?

Exercise 5 (Gibbs Sampler)

For $i = 1, \ldots, T$ consider $Z_i = X_i + Y_i$ with independent X_i, Y_i such that
$$X_i \sim \mathcal{B}inomial(m_i, \theta_1), \; Y_i \sim \mathcal{B}inomial(n_i, \theta_2).$$

1. We assume $0 \leq z_i \leq m_i + n_i$ for $i = 1, \ldots, T$. We observe z_i for $i = 1, \ldots, T$ and the n_i, m_i, for $i = 1, \ldots, T$ are given. Give the expression of the likelihood function $p(z_1, \ldots, z_T | \theta_1, \theta_2)$.

2. Assume we set independent uniform priors $\vartheta_1 \sim \mathcal{U}_{[0,1]}$, $\vartheta_2 \sim \mathcal{U}_{[0,1]}$. Propose a Gibbs sampler to sample from $p(\theta_1, \theta_2 | z_1, \ldots, z_T)$. Recall that the Beta distribution of parameter $\alpha, \beta > 0$ admits a density $f(x) \propto x^{\alpha-1}(1-x)^{\beta-1}\mathbb{I}_{[0,1]}(x)$.

   (Hint: introduce auxiliary variables)

Simulation question (Normal mixture model — Gibbs sampling)

We aim at performing Gibbs sampling on the posterior distribution of a Normal mixture model. The observations $Y = \{Y_1, \ldots, Y_N\}$ are assumed independent and each of them is distributed according to the mixture with density
$$\forall x \in \mathbb{R} \quad f(x) = \sum_{k=1}^{K} p_k \varphi(x; \mu_k, \sigma_k^2)$$
where φ is the Gaussian density for $\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+$,
$$\forall x \in \mathbb{R} \quad \varphi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}.$$

Let $p = (p_1, \ldots, p_K)$ the vector of weights, $\mu = (\mu_1, \ldots, \mu_K)$ the vector of means, $\sigma^2 = (\sigma_1^2, \ldots, \sigma_K^2)$ the vector of variances.

1. Introduce a random vector $Z = \{Z_1, \dots Z_N\}$ taking values in $\{1, \dots, K\}^N$ and such that

$$\mathbb{P}(Z = z \mid p) = \prod_{i=1}^{N} \mathbb{P}(Z_i = z_i \mid p) = \prod_{i=1}^{N} p_{z_i},$$

i.e. $\mathbb{P}(Z_i = k \mid p) = p_k$ for all $k \in \{1, \dots, K\}$ and $i \in \{1, \dots, N\}$. Then introduce variables $X = \{X_1, \dots X_N\}$ such that

$$\forall i \in \{1, \dots, N\} \quad X_i \mid Z_i = z_i, p, \mu, \sigma^2 \quad \sim \quad \mathcal{N}(\mu_{z_i}, \sigma_{z_i}^2).$$

Check that X given p, μ, σ^2 has the same distribution as Y.

2. Assume the following prior distribution on p, μ, σ^2. First the weights follow a Dirichlet distribution with parameters $\gamma_1, \dots, \gamma_K$:

$$\pi(p) \propto \prod_{i=1}^{K} p_i^{\gamma_k - 1}.$$

Then the means all follow the same normal distribution with parameters (m, τ^2)

$$\forall k \in \{1, \dots, K\} \quad \pi(\mu_k) \propto \exp\left\{-\frac{1}{2\tau^2}(\mu_k - m)^2\right\}.$$

Finally the variances follow an inverse gamma distribution with parameters (α, β):

$$\forall k \in \{1, \dots, K\} \quad \pi(\sigma_k^2) \propto (\sigma_k^2)^{-\alpha-1} \exp\left\{-\beta\sigma_k^{-2}\right\}.$$

Derive the conditional distribution of Z given X, p, μ, σ^2 and describe how to sample from it.

3. Derive the conditional distribution of p given X, Z, μ, σ^2 and describe how to sample from it.

4. Derive the conditional distribution of μ given X, Z, p, σ^2 and describe how to sample from it.

5. Derive the conditional distribution of σ^2 given X, Z, p, μ and describe how to sample from it.

6. Explain how to sample a synthetic dataset $(y_1, \dots, y_N)$ from a mixture model, for a fixed value of K, p, μ, σ^2, and N of your choice. Implement it and plot the histogram of the dataset.

   (Hint: you can use the variables Z, X as introduced in the first question).

7. Choose hyperparameters $\gamma, m, \tau^2, \alpha, \beta$ and justify your choice, based on the generated dataset.

   (Remark: this corresponds to an empirical Bayes approach, where some features of the dataset are used to design the prior).

8. Implement a random scan Gibbs sampler for the full posterior distribution, to sample approximately from the posterior given the generated dataset and the hyperparameters. Represent the generated Markov chain in the way you want, as long as it allows to compare

   - the generated Markov chain with the fixed values of p, μ, σ^2 used to generate the dataset,
   - the approximation of the posterior distribution with the prior distribution.