Computational statistics 2021.2
School of Applied Mathematics, Fundação Getulio Vargas
Professor Luiz Max de Carvalho

Problem sheet 2

Lucas Machado Moschen

Exercício 1. *(Monte Carlo for Gaussian)*
Let us consider the normal multivariate density on $\mathbb{R}^d$ with identity covariance, that is

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

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

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

Let ϕ be any measurable function and $\theta \in \mathbb{R}^d$. Denote I_2 the quantity in the right of the equation. Then,

$$\begin{aligned}
I_2 &= \int_{\mathbb{R}^d} \phi(x+\theta)\exp\left(-\frac{1}{2}\theta^T\theta - \theta^T x\right)\pi(x)\,dx \\
&= \int_{\mathbb{R}^d} \phi(x+\theta)\exp\left(-\frac{1}{2}\theta^T\theta - \theta^T x\right)\pi(x)\,dx \\
&= \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}}\phi(x+\theta)\exp\left(-\frac{1}{2}(x+\theta)^T(x+\theta)\right)\,dx \\
&= \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}}\phi(y)\exp\left(-\frac{1}{2}y^T y\right)\,dx \\
&= \mathbb{E}[\phi(X)].
\end{aligned}$$

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

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

of $\mathbb{E}[\phi(X)]$ is strongly consistent for any $\theta \in \mathbb{R}^d$ such that

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

The case $\theta = 0$ corresponds to the usual Monte Carlo estimate. The variance of $\hat{I}_n(\theta)$ is given by $\sigma^2(\theta)/n$ where

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

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

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

Let $\sigma^2(\theta) = \mathrm{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$ to simplify the writing. We already know that $\mathbb{E}[Y] = \mathbb{E}[\phi(X)]$ by the last exercise. Therefore, it remains to prove that

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

For that,

$$\begin{aligned}
\mathbb{E}[Y^2] &= \int_{\mathbb{R}^d} \phi^2(x+\theta)\exp\left(-\theta^T\theta - 2\theta^Tx\right)\pi(x)\,dx \\
&= \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}}\phi^2(x+\theta)\exp\left(-\theta^T\theta - 2\theta^Tx - \frac{1}{2}x^Tx\right)\,dx \\
&= \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}}\phi^2(x+\theta)\exp\left(-(x+\theta)^T(x+\theta) + \frac{1}{2}x^Tx\right)\,dx \\
&= \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}}\phi^2(y)\exp\left(-y^Ty + \frac{1}{2}(y-\theta)^T(y-\theta)\right)\,dx \\
&= \mathbb{E}\left[\phi^2(X)\exp\left(-X^TX + \frac{1}{2}(X-\theta)^T(X-\theta)\right)\right],
\end{aligned}$$

as we wanted to prove.

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.*

   For that, we will use the derived expression in the last exercise and we differentiate under the expected value using the Leibniz Rule. Then,

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

   and

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

   which is clearly positive definite since $(\theta - X)^T(\theta - X)$ is semi definite positive.

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

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

   Since $\sigma^2(\theta)$ is differentiable, its critical points are the solution of $\nabla_\theta\sigma^2(\theta) = 0$,

$$\begin{aligned}
&\mathbb{E}\left[\phi^2(X)(\theta - X)\exp\left(-X^TX + \frac{1}{2}(X-\theta)^T(X-\theta)\right)\right] = 0 \\
\implies\ &\mathbb{E}\left[\phi^2(X)(\theta - X)\exp\left(-\frac{1}{2}X^TX - \theta^TX\right)\right] = 0,
\end{aligned}$$

since $e^{\theta^T\theta/2}$ is a positive constant. Since the function is strictly convex, we already know that there is only one minimal and it occurs when the above expression is zero.

Exercício 2.

Exercício 3.

Exercício 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 $\pi(x, y)$.*

$$\begin{aligned}
\pi(x) &= \int_{y\in\mathbb{R}} \pi(x, y)\, dy \\
&= \int_{y\in\mathbb{R}} c\exp\left(-\frac{1}{2}(x-1)^2(y-2)^2\right) dy \\
&= c\int_{y\in\mathbb{R}} \exp\left(-\frac{1}{2}(x-1)^2(y-2)^2\right) dy \\
&= c\int_{y\in\mathbb{R}} \exp\left(-\frac{1}{2/(x-1)^2}(y-2)^2\right) dy \\
&= c\sqrt{2\pi}\frac{1}{|x-1|}
\end{aligned}$$

since setting $\sigma^2 = 1/(x-1)^2$, the integrand is the kernel of a normal distribution with mean 2 and variance σ^2. Therefore, its integral is the normalization constant. With the same reasoning, we have that

$$\pi(y) = c\sqrt{2\pi}\frac{1}{|y-2|}.$$

Then, we have that

$$\pi(x|y) = \frac{\pi(x, y)}{\pi(y)} = \frac{1}{\sqrt{2\pi}}|y-2|\exp\left(-\frac{1}{2}(x-1)^2(y-2)^2\right)$$

and

$$\pi(y|x) = \frac{\pi(x, y)}{\pi(x)} = \frac{1}{\sqrt{2\pi}}|x-1|\exp\left(-\frac{1}{2}(x-1)^2(y-2)^2\right),$$

which implies that

$$X \mid Y = y \sim \text{Normal}(1, |y-2|^2)$$

and

$$Y \mid X = x \sim \text{Normal}(2, |x-1|^2).$$

2. *Does the resulting Gibbs sampler make any sense?*

The problem with that Gibbs sampler is that the samples (x^n, y^n) converges to $(1, 2)$ when n is sufficiently high. This happens because the variances are very low in the region of greater mass. Therefore, even if the initial points are far from the mode, the sampling will explore it and wheen it happens, it will get stuck.

Exercício 5. *(Gibbs Sampler) For $i = 1, \ldots, T$ consider $Z_i = X_i + Y_i$ with independent X_i, Y_i such that*

$$X_i \sim \text{Bin}(m_i, \theta_1), Y_i \sim \text{Bin}(n_i, \theta_2).$$

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 $p(z_1, \ldots, z_T | \theta_1, \theta_2)$.*

   Supposing conditionally independent samples, we have that

   $$p(z_1, \ldots, z_T \mid \theta_1, \theta_2) = \prod_{i=1}^{T} p(z_i \mid \theta_1, \theta_2)$$

   Note that

   $$\begin{aligned} \mathbb{P}(Z_i = X_i + Y_i = k) &= \sum_{j=0}^{k} \mathbb{P}(X_i = j)\mathbb{P}(Y_i = k - j) \\ &= \sum_{j=0}^{k} \binom{m_i}{j}\binom{n_i}{j}(\theta_1 - \theta_1\theta_2)^j(\theta_2 - \theta_1\theta_2)^{k-j}. \end{aligned}$$

   Therefore,

   $$p(z_1, \ldots, z_T \mid \theta_1, \theta_2) = \prod_{i=1}^{T}\sum_{j=0}^{z_i} \binom{m_i}{j}\binom{n_i}{j}(\theta_1 - \theta_1\theta_2)^j(\theta_2 - \theta_1\theta_2)^{z_i-j}.$$

2. *Assume we set independent uniform priors $\theta_1 \sim \text{Unif}[0, 1], \theta_2 \sim \text{Unif}[0, 1]$. Propose a Gibbs sampler to sample from $p(\theta_1, \theta_2 \mid z_1, \ldots, z_T)$.*

   We know that

   $$p(\theta_1, \theta_2 \mid z_1, \ldots, z_T) \propto p(z_1, \ldots, z_T \mid \theta_1, \theta_2).$$

   and we want to specify

   $$p(\theta_1^t \mid \theta_2^{t-1}, z_1, \ldots, z_T)$$

   and

   $$p(\theta_2^t \mid \theta_1^t, z_1, \ldots, z_T).$$

   Since we do not know how to sample from these distributions, we have to introduce auxiliary variables to help out. Let $X_1, \ldots, X_T$ and $Y_1, \ldots, Y_T$ be the variables introduced in the exercise. We denote $X_{1:T} = X_1, \ldots, X_T$. Then

   $$\mathbb{P}(\theta_1 \mid \theta_2, X_{1:T}, Y_{1:T}, Z_{1:T}) = \mathbb{P}(\theta_1 \mid X_{1:T}),$$

   since given $X_{1:T}$, we have that θ_1 is independent of the other random variables. Analogously,

   $$\mathbb{P}(\theta_2 \mid \theta_1, X_{1:T}, Y_{1:T}, Z_{1:T}) = \mathbb{P}(\theta_2 \mid Y_{1:T}).$$

   Given that $\{X_i \mid \theta_1\}_{i=1}^T$ are independent and have binomial distribution, and θ_1 has Beta distribution with parameters $\alpha = 1$ and $\beta = 1$, we know that,

   $$\theta_1 \mid X_{1:T} = x_{1:T} \sim \text{Beta}\left(1 + \sum_{i=1}^{T} x_i, 1 + \sum_{i=1}^{T} m_i - x_i\right)$$

and

$$\theta_2 \mid Y_{1:T} = y_{1:T} \sim \text{Beta}\left(1 + \sum_{i=1}^{T} y_i, 1 + \sum_{i=1}^{T} n_i - y_i\right).$$

It remains to sample from

$$X_{1:T}, Y_{1:T} \mid \theta_1, \theta_2, Z_{1:T}.$$

Note that

$$\begin{aligned}\mathbb{P}(X_{1:T} = x_{1:T}, Y_{1:T} = y_{1:T} \mid \theta_1, \theta_2, Z_{1:T}) = \mathbb{P}(X_{1:T} = x_{1:T} \mid \theta_1, \theta_2, Z_{1:T}, Y_{1:T}) \\ \times \mathbb{P}(Y_{1:T} = y_{1:T} \mid \theta_1, \theta_2, Z_{1:T}).\end{aligned}$$

The first term is $1_{\{x_{1:T} = Z_{1:T} - Y_{1:T}\}}$, while the second can be writing as follows:

$$\begin{aligned}\mathbb{P}(Y_{1:T} = y_{1:T} \mid \theta_1, \theta_2, Z_{1:T}) &= \prod_{i=1}^{T} \mathbb{P}(Y_i = y_i \mid \theta_1, \theta_2, Z_i) \\ &= \prod_{i=1}^{T} \text{Bin}(y_i; n_i, \theta_2) \cdot \text{Bin}(z_i - y_i; m_i, \theta_1).\end{aligned}$$

Since this distribution has finite support, one can calculate its constant. Therefore, we have defined our Gibbs sampler.