Simon Queric
simon.queric@telecom-paris.fr
Computational Statistics
Prof. Stéphanie Allassonnière

11/13/23
TP 1

TP2 : Expectation-Maximisation algorithm – Importance sampling

Exercice 1 : Discrete distributions

1. In order to generate a discrete random variable X with probability measure $\mu = \sum_{i=1}^{n} p_i \delta_{x_i}$, we can use the inversion method. indeed, the cumulative distribution function of X is $F_X(x) = \sum_{i=1}^{n} p_i \, \mathbb{1}_{(x \leqslant x_i)}$ dont l'inverse généralisée est $F_X^{-1}(u) = x_1 \, \mathbb{1}_{(0 \leqslant u \leqslant p_1)} + \sum_{i=2}^{n} x_i \, \mathbb{1}_{\left(\sum_{k=1}^{i-1} p_k \leqslant u \leqslant \sum_{k=1}^{i} p_k\right)}$. Hence, to generate under μ, we first generate $U \sim \mathcal{U}([0,1])$ and return $F_X^{-1}(U)$.

Exercise 2 : Gaussian mixture model and the EM algorithm

We note $\phi_j : x \mapsto \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)}{2}\right)$

1. (a) The parameters of the Gaussian Mixture Model are $\theta = (\theta_j)_{1 \leqslant j \leqslant p} = ((\alpha_j)_{1 \leqslant j \leqslant p}, (\mu_j)_{1 \leqslant j \leqslant p}, (\Sigma_j)_{1 \leqslant j \leqslant p})$.
   (b) The probability density function of X can be written, by using the law of total probability conditionnaly to latent variable Z, as follow $p_\theta(x) = \sum_{j=1}^{p} p(z=j) p_\theta(x|z=j) = \sum_{j=1}^{p} \alpha_j \phi_j(x_i)$.

   Then, the likelihood of the data is :

$$\begin{aligned} p_\theta(x_{1:n}) &= \prod_{i=1}^{n} p_\theta(x_i) \text{ by independance.} \\ &= \prod_{i=1}^{n} \sum_{j=1}^{p} \alpha_j \phi_j(x_i) \end{aligned}$$

2. In order to implement the EM algorithm, we compute the updates of the parameters. We note q the joint density of the couple (X, Z) and we define :

$$Q(\theta, \theta') = \mathbb{E}_{Z \sim p_{\theta'}(\cdot|X)}[\log q(X, Z, \theta)]$$

   We also introduce the conditionnal probabilities $\tau_{i,j}^t$ defined by the following relation :

$$\tau_{i,j}^t = p_{\theta^t}(z=j|x_i) = \frac{p_{\theta^t}(z=j, x_i)}{p_{\theta^t}(x_i)} = \frac{\alpha_j^t \phi_j^t(x_i)}{\sum_{k=1}^{p} \alpha_k^t \phi_k^t(x_i)}$$

   Then :

$$Q(\theta, \theta^t) = \sum_{i=1}^{n} \sum_{j=1}^{p} \tau_{i,j}^t \log(\alpha_j \phi_j(x_i)) = \sum_{i=1}^{n} \sum_{j=1}^{p} \tau_{i,j}^t (\log \alpha_j + \log \phi_j(x_i))$$

   The optimization problem with equality constraints is :

$$\theta^{t+1} = \mathrm{argmax}_\theta Q(\theta, \theta^t)$$

The lagrangian of the problem is :

$$L(\theta, \lambda) = Q(\theta, \theta^t) + \lambda \left(1 - \sum_{j=1}^{p} \alpha_j\right)$$

The cancellation of ∇L with respect to the θ_j gives us the parameters updates :

$$\mu_j^{t+1} \leftarrow \frac{\sum_{i=1}^N \tau_{i,j}^t x_i}{\sum_{i=1}^N \tau_{i,j}^t}, \Sigma_j^{t+1} \leftarrow \frac{\sum_{i=1}^N \tau_{i,j}^t (x_i - \mu_j)(x_i - \mu_j)^T}{\sum_{i=1}^N \tau_{i,j}^t}, \alpha_j^{t+1} \leftarrow \frac{1}{N} \sum_{i=1}^{N} \tau_{i,j}^t$$

We deduce the EM algorithm for the Gaussian Mixture Model.

Algorithm 1 EM Algorithm for Gaussian Mixture Model

Input : data $(x_i)_{1 \leq i \leq N}$, number of components $p \geqslant 2$

while not stopping criterion **do**

Expectation Step

$\tau_{i,j} \leftarrow \dfrac{\alpha_j \phi_j(x_i)}{\sum_{k=1}^p \alpha_k \phi_k(x_i)}$ with Log-Sum-Exp trick.

Maximization Step

$\mu_j \leftarrow \dfrac{\sum_{i=1}^N \tau_{i,j} x_i}{\sum_{i=1}^N \tau_{i,j}}$

$\Sigma_j \leftarrow \dfrac{\sum_{i=1}^N \tau_{i,j} (x_i - \mu_j)(x_i - \mu_j)^T}{\sum_{i=1}^N \tau_{i,j}}$

$\alpha_j \leftarrow \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} \tau_{i,j}$

end while

Output : $(\alpha_j)_{1 \leq j \leq p}, (\mu_j)_{1 \leq j \leq p}, (\Sigma_j)_{1 \leq j \leq p}$

Exercise 3 : Importance sampling

Adaptative Importance Sampling

At the step (iii) of *Population Monte Carlo Algorithm*, we want to minimize Kullback-Leibler divergence $K(\nu||q) = \displaystyle\int \log \frac{\nu(x)}{q(x)} \nu(x) \mathrm{d}x = \int \log \nu(x) \nu(x) \mathrm{d}x - \int \log q(x) \nu(x) \mathrm{d}x$ on the set of p.d.f q of Gaussian Mixture Model with M mixtures. it's equivalent to maximize $I(q) = \displaystyle\int \log q(x) \nu(x) \mathrm{d}x$, which can be estimated by Importance Sampling : $\widehat{I(q)} = \displaystyle\sum_{i=1}^{n} \widetilde{\omega}_i \log \sum_{j=1}^{p} \alpha_j \phi_j(X_i^{(0)})$ où $\widetilde{\omega}_i = \dfrac{\frac{\nu(X_i^{(0)})}{q^{(0)}(X_i^{(0)})}}{\sum_{k=1}^n \frac{\nu(X_k^{(0)})}{q^{(0)}(X_k^{(0)})}}$ with ν the target density, and $q^{(0)}$ the importance density.
By using Jensen's inequality, we get :

$$\begin{aligned}
\sum_{i=1}^{n} \widetilde{\omega}_i \log \sum_{j=1}^{p} \alpha_j \phi_j(x_i) &\geqslant \sum_{i=1}^{n} \widetilde{\omega}_i \sum_{j=1}^{p} \tau_{i,j} (\log \alpha_j + \log \phi_j(x_i)) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{p} \widetilde{\omega}_i \tau_{i,j} (\log \alpha_j + \log \phi_j(x_i)) \\
&= \widetilde{Q}(\theta, \theta^{(0)})
\end{aligned}$$

The function $\widetilde{Q}$ has the same form as Q of exercise 2, by replacing $\tau_{i,j}$ by $\widetilde{\omega}_i \tau_{i,j}$. The parameters updates are :

$$\mu_j \leftarrow \sum_{i=1}^{N} \widetilde{\omega}_i \tau_{i,j} x_i, \Sigma_j \leftarrow \sum_{i=1}^{N} \widetilde{\omega}_i \tau_{i,j} (x_i - \mu_j)(x_i - \mu_j)^T, \alpha_j \leftarrow \sum_{i=1}^{N} \widetilde{\omega}_i \tau_{i,j}$$