Computational Statistics
Homework 2

Salomé Do

November 10, 2018

Exercise 5.4. Simulated Annealing Algorithm

(a) Reproducing simulations from Example 5.5

With the code in the jupyter notebook, we reproduce Example 5.5., i.e. Simulated Annealing algorithm with :

$$\begin{aligned}
h(x) &= [\cos(50x) + \sin(20x)]^2 \\
a_t &= max(x^{(t)} - r, 0) \\
b_t &= min(x^{(t)} + r, 1) \\
u &\sim \mathcal{U}(a_t, b_t) \\
\rho^{(t)} &= \min\left\{\exp\left(\frac{h(u) - h(x^{(t)})}{T_t}\right), 1\right\} \\
T_t &= \frac{1}{\log(t)}
\end{aligned}$$

The algorithm is, at each time t:

1. Simulate $u \sim \mathcal{U}(a_t, b_t)$.
2. Accept $x^{(t+1)}$ with probability $\rho^{(t)}$, take$x^{(t+1)} = x^{(t)}$ otherwise.
3. Update T_t to T_{t+1}.

And we start a sequence of 2500 simulations with $x^{(0)} = 0$. As in Example 5.5, four sequences are simulated [1] and represented in Figure 1.

(b) Changing parameters in r and T_t

We now define $T_t = \frac{c}{\log(t)}$. We show in Figure 2 simulated sequences for a range of r and c values. We see that lower values of c seem to give a faster convergence. Regarding r, $r = 0.75$ seems to be the value that fastens convergence the most.

[1] Code available here

Figure 1: Reproduction of Example 5.5 : Simulated Annealing algorithm

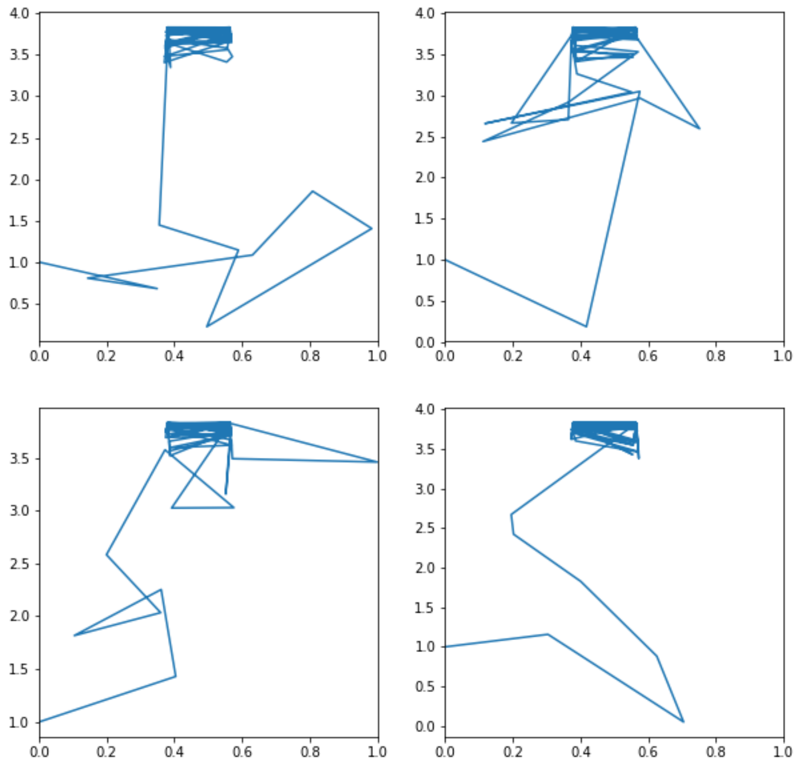
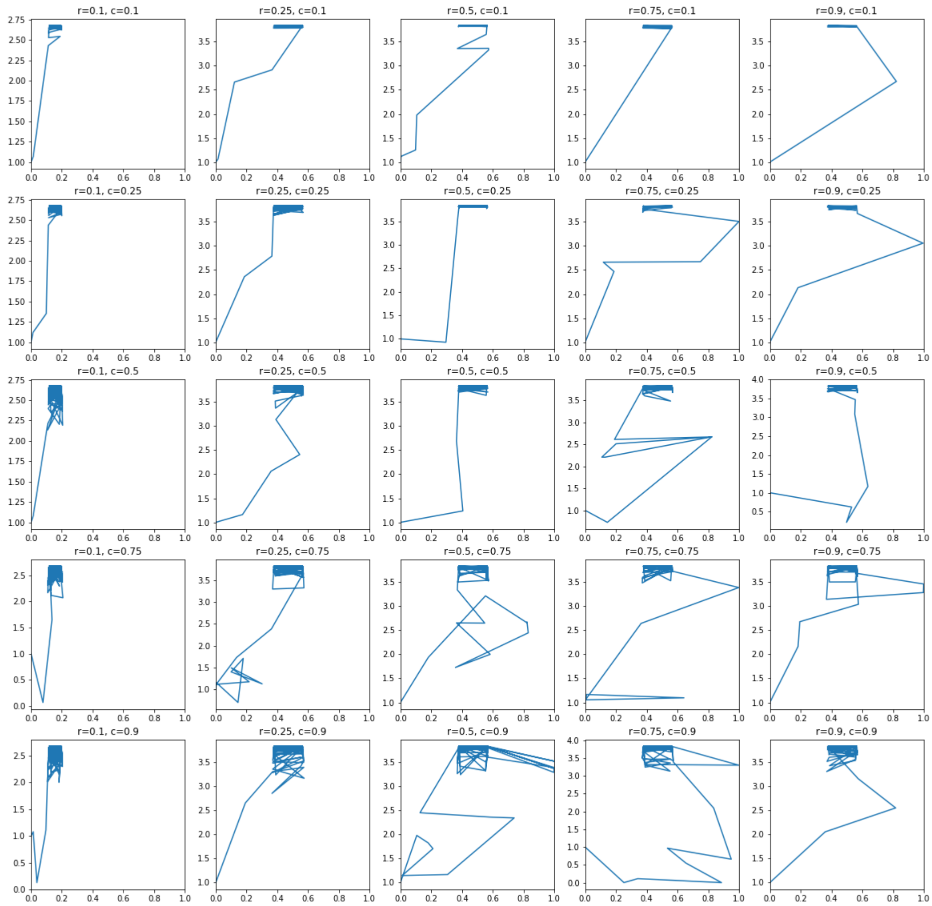


Figure 2: Simulated Annealing for various values of r and c



Exercise 5.22. EM Algorithm on Bernouilli Variables

Here, we observe $X_1, ..., X_n$ i.i.d. depending on $Z_1, ..., Z_n$ independently distributed as $\mathcal{N}(\zeta, \sigma^2)$. We have, for a known threshold u :

$$X_i = \begin{cases} 0 & \text{if} \quad Z_i \leq u \\ 1 & \text{if} \quad Z_i > u \end{cases}$$

Our aim is to obtain MLE estimates for ζ, σ^2.

(a) Likelihood function

We want to compute $\mathcal{L}(x; (\zeta, \sigma))$. We have :

$$\begin{aligned} \mathcal{L}(x; (\zeta, \sigma)) &= \prod_{i=1}^{n} \mathbb{P}(X_i = x_i | \zeta, \sigma) \\ &= p^S (1-p)^{n-S} \end{aligned}$$

As the X_i are drawn independently from a Bernouilli. p is the probability for X_i to be equal to 1, thus:

$$\begin{aligned} p &= \mathbb{P}(Z_i > u) \\ &= \mathbb{P}\left(\frac{Z_i - \zeta}{\sigma} > \frac{u - \zeta}{\sigma}\right) \\ &= 1 - \mathbb{P}\left(\frac{Z_i - \zeta}{\sigma} \leq \frac{u - \zeta}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{u - \zeta}{\sigma}\right) \\ &= \Phi\left(\frac{\zeta - u}{\sigma}\right) \end{aligned}$$

S is the number of x_i equal to one, thus we can write :

$$S = \sum_{i=1}^{n} x_i$$

(b) Complete Likelihood

We are now interested in the complete likelihood function $\mathcal{L}((z); (\zeta, \sigma))$. We have :

$$\begin{aligned} \mathcal{L}(z; (\zeta, \sigma)) &= \prod_{i=1}^{n} \mathbb{P}(Z_i = z_i \mid \zeta, \sigma) \\ &= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(z_i - \zeta)^2}{2\sigma^2}\right) \end{aligned}$$

Thus,

$$\begin{aligned}\log \mathcal{L}((z);(\zeta,\sigma)) &= \sum_{i=1}^{n} -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(z_i-\zeta)^2}{2\sigma^2} \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n}\frac{(z_i-\zeta)^2}{2\sigma^2}\end{aligned}$$

Then, we take the expectation of this function, on the observed data, i.e. the (x_i), regarding the random variables Z_i :

$$\begin{aligned}\mathbb{E}\left[\log \mathcal{L}((Z_i)_i;(\zeta,\sigma))|x_i\right] &= -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n}\mathbb{E}\left[\left.\frac{(Z_i-\zeta)^2}{2\sigma^2}\right| x_i\right] \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\mathbb{E}[Z_i^2 - 2\zeta Z_i + \zeta^2|x_i] \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathbb{E}[Z_i^2|x_i] - 2\zeta\mathbb{E}[Z_i|x_i] + \zeta^2)\end{aligned}$$

(c) EM Sequence

We now want to give the EM sequence. We first treat $\zeta^{(t)}$ sequence :

$$\begin{aligned}\frac{\partial \mathbb{E}\left[\log \mathcal{L}((Z_i)_i;\zeta=\zeta^{(t)},\sigma=\sigma^{(t)})|x_i\right]}{\partial \zeta^{(t)}} &= -\frac{1}{2(\sigma^{(t)})^2}\sum_{i=1}^{n}(-2\mathbb{E}[Z_i|x_i,\zeta^{(t)},\sigma^{(t)}] + 2\zeta^{(t)}) \\ &= -\frac{n}{(\sigma^{(t)})^2}\zeta^{(t)} + \frac{1}{(\sigma^{(t)})^2}\sum_{i=1}^{n}\mathbb{E}[Z_i|x_i,\zeta^{(t)},\sigma^{(t)}]\end{aligned}$$

We chose $\zeta^{(t+1)}$ to minimize the expected log-likelihood of the complete data, thus:

$$\zeta^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Z_i|x_i,\zeta^{(t)},\sigma^{(t)}]$$

We are now interested in $\sigma^{(t)}$. We compute :

$$\begin{aligned}\frac{\partial \mathbb{E}\left[\log \mathcal{L}((Z_i)_i;\zeta=\zeta^{(t)},\sigma=\sigma^{(t)})|x_i\right]}{\partial (\sigma^{(t)})^2} &= -\frac{n}{2}\frac{2\pi}{2\pi(\sigma^{(t)})^2} \\ &+ \frac{1}{2((\sigma^{(t)})^2)^2}\left[\sum_{i=1}^{n}\mathbb{E}[Z_i^2|x_i,\zeta^{(t)},\sigma^{(t)}] - 2\zeta^{(t)}\mathbb{E}[Z_i|x_i,\zeta^{(t)},\sigma^{(t)}] + (\zeta^{(t)})^2\right]\end{aligned}$$

Minimization of the expected log-likehood of the complete data gives:

$$
\begin{aligned}
(\sigma^{(t+1)})^2 &= \frac{1}{n}\left[\sum_{i=1}^{n}\left(\mathbb{E}[Z_i^2|x_i,\zeta^{(t)},\sigma^{(t)}] - 2\zeta^{(t+1)}\mathbb{E}[Z_i|x_i,\zeta^{(t)},\sigma^{(t)}] + (\zeta^{(t+1)})^2\right)\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}\mathbb{E}[Z_i^2|x_i,\zeta^{(t)},\sigma^{(t)}] - 2\zeta^{(t+1)}\sum_{i=1}^{n}\mathbb{E}[Z_i|x_i,\zeta^{(t)},\sigma^{(t)}] + n(\zeta^{(t+1)})^2\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}\mathbb{E}[Z_i^2|x_i,\zeta^{(t)},\sigma^{(t)}] - 2\zeta^{(t+1)}\sum_{i=1}^{n}\mathbb{E}[Z_i|x_i,\zeta^{(t)},\sigma^{(t)}] + n(\zeta^{(t+1)})^2\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}\mathbb{E}[Z_i^2|x_i,\zeta^{(t)},\sigma^{(t)}] - 2\zeta^{(t+1)}n\zeta^{(t+1)} + n(\zeta^{(t+1)})^2\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}\mathbb{E}[Z_i^2|x_i,\zeta^{(t)},\sigma^{(t)}] - n(\zeta^{(t+1)})^2\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}\mathbb{E}[Z_i^2|x_i,\zeta^{(t)},\sigma^{(t)}] - n(\zeta^{(t+1)})^2\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}\mathbb{E}[Z_i^2|x_i,\zeta^{(t)},\sigma^{(t)}] - \frac{1}{n}\left(\sum_{i=1}^{n}\mathbb{E}[Z_i|x_i,\zeta^{(t)},\sigma^{(t)}]\right)^2\right]
\end{aligned}
$$

(d) Expectations Computation

We want to compute $\mathbb{E}[Z_i|x_i,\zeta,\sigma]$. Let's first suppose that $X_i = 1$. Then :

$$
\begin{aligned}
\mathbb{E}[Z_i|X_i = 1\zeta,\sigma]] &= \frac{1}{\mathbb{P}(X_i = 1)}\int_{u}^{\infty} z\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(z-\zeta)^2}{2\sigma^2}}dz \\
&= \frac{1}{\mathbb{P}(Z_i > u)}\int_{\frac{u-\zeta}{\sigma}}^{\infty}(\zeta+\sigma x)\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-x^2}{2}}\sigma dx \qquad \leftarrow x = \frac{z-\zeta}{\sigma} \\
&= \frac{1}{1-\Phi(\frac{u-\zeta}{\sigma})}\int_{\frac{u-\zeta}{\sigma}}^{\infty}(\zeta+\sigma x)\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}dx \\
&= \frac{1}{1-\Phi(\frac{u-\zeta}{\sigma})}\left[\zeta\int_{\frac{u-\zeta}{\sigma}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}dx - \sigma\frac{1}{\sqrt{2\pi}}\int_{\frac{u-\zeta}{\sigma}}^{\infty}(-x)e^{\frac{-x^2}{2}}dx\right] \\
&= \frac{1}{1-\Phi(\frac{u-\zeta}{\sigma})}\left[\zeta(1-\Phi\left(\frac{u-\zeta}{\sigma}\right)) - \sigma\frac{1}{\sqrt{2\pi}}\left[e^{\frac{-x^2}{2}}\right]_{\frac{u-\zeta}{\sigma}}^{\infty}\right] \\
&= \frac{1}{1-\Phi(\frac{u-\zeta}{\sigma})}\left[\zeta(1-\Phi\left(\frac{u-\zeta}{\sigma}\right)) + \sigma\frac{1}{\sqrt{2\pi}}e^{-\frac{(u-\zeta)^2}{2\sigma}}\right] \\
&= \zeta + \sigma\frac{\varphi\left(\frac{u-\zeta}{\sigma}\right)}{1-\Phi\left(\frac{u-\zeta}{\sigma}\right)}
\end{aligned}
$$

Now, if $X_i = 0$:

$$\mathbb{E}[Z_i|X_i = 0, \zeta, \sigma] = \frac{1}{\mathbb{P}(X_i = 0)} \int_{-\infty}^{u} z \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(z-\zeta)^2}{2\sigma^2}} dz$$

Using the same substitution $x = \frac{z-\zeta}{\sigma}$ as in the first case, we have:

$$\begin{aligned}
\mathbb{E}[Z_i|X_i = 0, \zeta, \sigma] &= \frac{1}{\Phi(\frac{u-\zeta}{\sigma})} \left[\zeta\Phi\left(\frac{u-\zeta}{\sigma}\right) - \sigma \frac{1}{\sqrt{2\pi}} e^{-\frac{(u-\zeta)^2}{2\sigma}} \right] \\
&= \zeta - \sigma \frac{\varphi\left(\frac{u-\zeta}{\sigma}\right)}{\Phi\left(\frac{u-\zeta}{\sigma}\right)}
\end{aligned}$$

Thus, we generally have :

$$\mathbb{E}[Z_i|x_i, \zeta, \sigma] = \zeta + \sigma H_i\left(\frac{u-\zeta}{\sigma}\right)$$

With H_i as described in the exercice. Let's compute the expectation $\mathbb{E}[Z_i^2|x_i, \zeta, \sigma]$, in the same way. First :

$$\begin{aligned}
\mathbb{E}[Z_i^2|X_i = 1, \zeta, \sigma] &= \frac{1}{\mathbb{P}(X_i = 1)} \int_{u}^{\infty} z^2 \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(z-\zeta)^2}{2\sigma^2}} dz \\
&= \frac{1}{\mathbb{P}(Z_i > u)} \int_{\frac{u-\zeta}{\sigma}}^{\infty} (\zeta + \sigma x)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2}} \sigma dx \qquad \leftarrow x = \frac{z-\zeta}{\sigma} \\
\mathbb{E}[Z_i^2|X_i = 1, \zeta, \sigma](1 - \Phi\left(\frac{u-\zeta}{\sigma}\right)) &= \zeta^2 \int_{\frac{u-\zeta}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \\
&\quad - 2\zeta\sigma \int_{\frac{u-\zeta}{\sigma}}^{\infty} (-x) \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \\
&\quad + \sigma^2 \int_{\frac{u-\zeta}{\sigma}}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \\
&= \zeta^2(1 - \Phi\left(\frac{u-\zeta}{\sigma}\right)) + 2\zeta\sigma\varphi\left(\frac{u-\zeta}{\sigma}\right) + A
\end{aligned}$$

Where $A = \sigma^2 \int_{\frac{u-\zeta}{\sigma}}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx$. We use integration by parts to compute A.

$$\begin{aligned}
A &= -\frac{\sigma^2}{\sqrt{2\pi}} \left(\left[x e^{-\frac{x^2}{2}} \right]_{\frac{u-\zeta}{\sigma}}^{\infty} - \int_{\frac{u-\zeta}{\sigma}}^{\infty} e^{-\frac{x^2}{2}} dx \right) \\
&= \sigma(u-\zeta)\varphi\left(\frac{u-\zeta}{\sigma}\right) + \sigma^2(1 - \Phi\left(\frac{u-\zeta}{\sigma}\right))
\end{aligned}$$

Thus, we have:

$$\mathbb{E}[Z_i^2|X_i = 1, \zeta, \sigma] = \zeta^2 + \sigma^2 + \sigma(u+\zeta) \frac{\varphi\left(\frac{u-\zeta}{\sigma}\right)}{1 - \Phi\left(\frac{u-\zeta}{\sigma}\right)}$$

By re-using the same substitution and integration by parts for $\mathbb{E}[Z_i^2|X_i = 0]$, we find that $\mathbb{E}[Z_i^2|X_i = 0, \zeta, \sigma] = \zeta^2 + \sigma^2 - \sigma(u+\zeta)\frac{\varphi\left(\frac{u-\zeta}{\sigma}\right)}{\Phi\left(\frac{u-\zeta}{\sigma}\right)}$, so we can conclude that :

$$\mathbb{E}[Z_i^2|x_i, \zeta, \sigma] = \zeta^2 + \sigma^2 + \sigma(u+\zeta)H_i\left(\frac{u-\zeta}{\sigma}\right)$$

(e) Convergence

Let's return to the expected complete-data log-likelihood, which is equal to :

$$\mathbb{E}\left[\log \mathcal{L}(Z|x_i, \zeta^*, \sigma^*)\right] =$$
$$-\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathbb{E}[Z_i^2|x_i, \zeta^*, \sigma^*] - 2\zeta\mathbb{E}[Z_i|x_i, \zeta^*, \sigma^*] + \zeta^2)$$
$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[\zeta^* + \sigma^* H_i\left(\frac{u-\zeta^*}{\sigma^*}\right) - 2\left[(\zeta^*)^2 + (\sigma^*)^2 + \sigma^*(u+\zeta^*)H_i\left(\frac{u-\zeta^*}{\sigma^*}\right)\right] + \zeta^2\right]$$

This expression is continuous in ζ, σ and ζ^*, σ^* We can thus use Theorem 5.16. Every limit point of the EM sequence is a stationnary point of the log-likehood. As the log-likelihood is log-concave, is only has one maximum. The EM sequence converges to the points of this maximum, and thus maximize the log-likelihood.

Exercise 5.33. EM on Bayesian Hierarchical Models

We have a hierarchical bayesian model, i.e. :

$$\begin{aligned} X|\theta &\sim f(x|\theta) \\ \theta|\lambda &\sim \pi(\theta|\ \lambda) \\ \lambda &\sim \gamma(\lambda) \end{aligned}$$

We want to estimate the posterior $\pi(\theta|x)$, and in order to do so, we use EM Algorithm.

(a) Log-Likelihood

Our aim is to compute $\log \pi(\theta|x)$. Let θ^* be any value defined in the same set as θ. We use the fact that :

$$\begin{aligned} &\pi(\theta|x) = \frac{\pi(\theta, \lambda|x)}{k(\lambda|\theta, x)} \\ &\Leftrightarrow \log \pi(\theta|x) = \log \pi(\theta, \lambda\ |x) - \log k(\lambda|\theta, x) \\ &\Leftrightarrow \log \pi(\theta|x) k(\lambda|\theta^*, x) = \log \pi(\theta, \lambda|x) k(\lambda|\theta^*, x) - \log k(\lambda|\theta, x) k(\lambda|\theta^*, x) \\ &\Leftrightarrow \int \log \pi(\theta|x) k(\lambda|\theta^*, x) d\lambda = \int \log \pi(\theta|\lambda, x) k(\lambda|\theta^*, x) d\lambda - \int \log k(\lambda|\theta, x) k(\lambda|\theta^*, x) d\lambda \\ &\Leftrightarrow \log \pi(\theta|x)) \underbrace{\int k(\lambda|\theta^*, x) d\lambda}_{=1} = \int \log \pi(\theta|\lambda, x) k(\lambda|\theta^*, x) d\lambda - \int \log k(\lambda|\theta, x) k(\lambda|\theta^*, x) d\lambda \end{aligned}$$

Thus, we have, for any θ^*:

$$\log \pi(\theta|x) = \int \log \pi(\theta|\lambda, x) k(\lambda|\theta^*, x) d\lambda - \int \log k(\lambda|\theta, x) k(\lambda|\theta^*, x) d\lambda \tag{1}$$

(b) EM Sequence

We want to show that the EM Sequence improves $\log \pi(\theta^{(j)}|x)$ at each step j, i.e. $\log \pi(\theta^{(j+1)}|x) \geq \log \pi(\theta^{(j)}|x), \forall j \in \mathbb{N}$. We re-write Equation 1 as:

$$\log \pi(\theta|x) := Q(\theta|\theta^*, x) - \mathbb{E}_{\theta^*}[\log k(\lambda|\theta, x)] \tag{2}$$

Where the expectation is taken with respect to $k(\lambda|\theta^*, x)$. Let's define the EM Sequence $(\theta^{(j)})_{j\in\mathbb{N}}$ for each step with :

$$\begin{aligned} \theta^{(j+1)} &= \arg\max_\theta\ Q(\theta|\theta^{(j)}, x) \\ &= \arg\max_\theta \int \log \pi(\theta|\lambda, x) k(\lambda|\theta^{(j)}, x) d\lambda \end{aligned}$$

By definition of $\theta^{(j+1)}$:

$$Q(\theta^{(j+1)}|\theta^{(j)}, x) \geq Q(\theta^{(j)}|\theta^{(j)}, x) \tag{3}$$

We finally want to show that :

$$\mathbb{E}_{\theta^{(j)}}[\log k(\lambda|\theta^{(j+1)},x)] \leq \mathbb{E}_{\theta^{(j)}}[\log k(\lambda|\theta^{(j)},x)]$$

Jensen's inequality for concave functions is the following: for any concave function f,

$$f(\mathbb{E}[X]) \geq \mathbb{E}[f(X)]$$

Taking $f = \log$ and $X = \frac{k(\lambda|\theta^{(j+1)},x)}{k(\lambda|\theta^{(j)},x)}$, we have :

$$\begin{aligned}
\mathbb{E}_{\theta^{(j)}}\left[\log\left(\frac{k(\lambda|\theta^{(j+1)},x)}{k(\lambda|\theta^{(j)},x)}\right)\right] &\leq \log \mathbb{E}_{\theta^{(j)}}\left[\frac{k(\lambda|\theta^{(j+1)},x)}{k(\lambda|\theta^{(j)},x)}\right] \\
&= \log \int \frac{k(\lambda|\theta^{(j+1)},x)}{k(\lambda|\theta^{(j)},x)} k(\lambda|\theta^{(j)},x)d\lambda \\
&= 0
\end{aligned}$$

Which directly implies that :

$$\mathbb{E}_{\theta^{(j)}}[\log k(\lambda|\theta^{(j+1)},x)] \leq \mathbb{E}_{\theta^{(j)}}[\log k(\lambda|\theta^{(j)},x)] \tag{4}$$

Thus, by taking $\theta^* = \theta^{(j)}$ in (2), we showed (3) and (4), directly giving :

$$\log \pi(\theta^{(j+1)}|x) \geq \log \pi(\theta^{(j)}|x)$$

According to Theorem 5.16., if $Q(\theta|\theta^*,x)$ is continuous in both θ and θ^*, every limit point of $(\theta^{(j)})_{j\in\mathbb{N}}$ is a stationnary point of $\log \pi(\theta|x)$, and $\log \pi(\theta^{(j)}|x)$ converges monotonically to $\log \pi(\hat{\theta}|x)$ for some stationnary point $\hat{\theta}$.

(c) Application

We apply this EM strategy to the hierarchical model :

$$\begin{aligned}
X|\theta &\sim \mathcal{N}(\theta, 1) \\
\theta|\lambda &\sim \mathcal{N}(\lambda, 1)
\end{aligned}$$

With $\pi(\lambda) = 1$. At a given step j, we want to compute :

$$\begin{aligned}
\theta^{(j+1)} &= \arg\max_{\theta} Q(\theta|\theta^{(j)},x) \\
&= \arg\max_{\theta} \int \log \pi(\theta|\lambda,x)k(\lambda|\theta^{(j)},x)d\lambda
\end{aligned}$$

Following Baye's rule, we have:

$$\pi(\theta,\lambda|x) = \frac{\pi(x|\theta)\pi(\theta\ |\lambda)\pi(\lambda)}{\pi(x)}$$

We know $\pi(x|\theta), \pi(\theta\ |\lambda)$ and $\pi(\lambda)$. $\pi(x)$ is a constant only depending on the observed data. Q can be re-wrote as:

$$Q(\theta|\theta^{(j)},x) = R(\theta) + \underbrace{\int \log \frac{\pi(\lambda)}{\pi(x)} k(\lambda|\theta^{(j)},x)d\lambda}_{=C}$$

Thus, $\arg\max_{\theta} Q(\theta|\theta^{(j)}, x) = \arg\max_{\theta} R(\theta)$. We have :

$$R(\theta) = \int \underbrace{\log[\pi(x|\theta)\pi(\theta\ |\lambda)]}_{A(\theta,\lambda)}\underbrace{k(\lambda|\theta^{(j)}, x)}_{B(\lambda,\theta^{(j)})}d\lambda$$

Compute :

$$\begin{aligned}
A(\theta, \lambda) &= \log[\pi(x|\theta)] + \log[\pi(\theta\ |\lambda)] \\
&= -\frac{1}{2}\log(2\pi) - \frac{(x-\theta)^2}{2} - \frac{1}{2}\log(2\pi) - \frac{(\theta-\lambda)^2}{2}, \\
\frac{\partial A(\theta, \lambda)}{\partial \theta} &= \theta(x-\theta) - \theta(\theta-\lambda) \\
&= \theta(x - 2\theta + \lambda)
\end{aligned}$$

Finally, as $\int B(\lambda, \theta^{(j)})d\lambda = 1$,

$$\begin{aligned}
\frac{\partial R(\theta)}{\partial \theta} &= \int B(\lambda, \theta^{(j)})\frac{\partial A(\theta, \lambda)}{\partial \theta}d\lambda \\
&= \theta(x - 2\theta) + \theta\int \lambda B(\lambda, \theta^{(j)})d\lambda \\
&= \theta(x - 2\theta + \mathbb{E}[\lambda|\theta^{(j)}, x])
\end{aligned}$$

This leads us to :

$$\theta^{(j+1)} = \frac{1}{2}(x + \mathbb{E}[\lambda|\theta^{(j)}, x])) \tag{5}$$

However, we haven't computed $\mathbb{E}[\lambda|\theta^{(j)}, x])$ yet. Following Bayes's rule :

$$k(\lambda|\theta^{(j)}, x) = \frac{\pi(x|\theta^{(j)})\pi(\theta^{(j)}|\lambda)\pi(\lambda)}{\pi(x)\pi(\theta^{(j)}|x)}$$

We know $\pi(\theta^{(j)}|x)$ as we have calculated it during the following step. $\pi(x)$ is a constant that we can estimate by other means. We also know $\pi(x|\theta^{(j)}) = \frac{1}{\sqrt{2\pi}}e^{\frac{(x-\theta^{(j)})^2}{2}}$ and $\pi(\theta^{(j)}\lambda) = \frac{1}{\sqrt{2\pi}}e^{\frac{(\theta^{(j)}-\lambda)^2}{2}}$. Using $\pi(\lambda) = 1$ seems strange (as $\pi(\lambda)$ should be a p.d.f), but we will stick to the problem's rules. Then,

$$\begin{aligned}
\int \lambda k(\lambda|\theta^{(j)}, x)d\lambda &= \frac{e^{\frac{(x-\theta^{(j)})^2}{2}}}{\sqrt{2\pi}\pi(x)\pi(\theta^{(j)}|x)}\int \lambda\frac{1}{\sqrt{2\pi}}e^{\frac{(\theta^{(j)}-\lambda)^2}{2}}\pi(\lambda)d\lambda & (6) \\
&= \frac{e^{\frac{(x-\theta^{(j)})^2}{2}}}{\sqrt{2\pi}\pi(x)\pi(\theta^{(j)}|x)}\theta^{(j)},\ \text{if } \pi(\lambda) = 1 & (7)
\end{aligned}$$

Using (5) and (7), we can apply EM Algorithm to this bayesian hierarchical model.