Master 2 MVA - Computational Statistics

TP3 : Hasting-Metropolis (and Gibbs) samplers

Mathis LE BAIL

November 2023

Exercise 1 : Hasting-Metropolis within Gibbs – Stochastic Approximation EM

1.A - A population model for longitudinal data

1. The information is as follows : there are the parameters $\theta = \{\bar{t_0}, \bar{v_0}, \sigma_\xi, \sigma_\tau, \sigma\}$, the latent variables $z = (z_{pop}, \{z_i\}) = (t_0, v_0, \{\alpha_i, \tau_i\}_i)$ and the fixed data $\sigma_{t_0}, \sigma_{v_0}, p_0 = 0, \bar{\bar{t_0}}, \bar{\bar{v_0}}, s_{t_0}, s_{v_0}, m = m_\xi = m_\tau, v = v_\xi = v_\tau, N$ and $(k_i)_i$ where $k_i = K$ for all $i \in [\![1, N]\!]$. $\{y_{i,j}\}_{i,j}$ are the assumed independent observations obtained at times $\{t_{i,j}\}_{j \in [\![1,K]\!]}$

The likelihood can be written

$$
\begin{aligned}
p(y, z, \theta) &= p(y|z, \theta)p(z, \theta) \\
&= p(y|z, \theta)p(z|\theta)p(\theta) \\
&= p(y|z, \theta)p(z_{pop}|\theta)p(\{z_i\}|\theta)p(\theta) \quad \text{by independence of } (z_{pop}, \{z_i\})
\end{aligned} \tag{1}
$$

Thus,

$$
\log(p(y, z, \theta)) = \log(p(y|z, \theta)) + \log(p(z_{pop}|\theta)) + \log(p(\{z_i\}|\theta)) + \log(p(\theta)) \tag{2}
$$

We detail the calculation of the four terms in the sum :

$$
\begin{aligned}
p(y|z, \theta) &= \prod_{i,j} p(y_{ij}|z, \theta) \qquad \text{by independence of the measurements } y_{ij} \\
y_{ij}|z, \theta &\sim \mathcal{N}(d_i(t_{ij})|\sigma^2) \\
p(y_{ij}|z, \theta) &\propto \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(y_{ij} - d_i(t_{ij}))^2}{\sigma^2}\right) \\
\log(p(y_{ij}|z, \theta)) &= -\frac{1}{2}\frac{(y_{ij} - d_i(t_{ij}))^2}{\sigma^2} - \log(\sigma) + cste \\
\log(p(y|z, \theta)) &= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{K}\frac{(y_{ij} - d_i(t_{ij}))^2}{\sigma^2} - KN\log(\sigma) + cste
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
p(z_{pop}|\theta) &= p(t_0, v_0|\theta) = p(t_0|\theta)p(v_0|\theta) \\
\log(p(z_{pop}|\theta)) &= -\frac{1}{2}\left(\frac{t_0 - \bar{t_0}}{\sigma_{t_0}}\right)^2 - \log(\sigma_{t_0}) - \frac{1}{2}\left(\frac{v_0 - \bar{v_0}}{\sigma_{v_0}}\right)^2 - \log(\sigma_{v_0}) + cste
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
p(\{z_i\}|\theta) &= \prod_i p(z_i|\theta) \qquad \text{by independence of the } \{z_i\} \\
p(z_i|\theta) &= p(\xi, \tau_i|\theta) = p(\xi|\theta)p(\tau_i|\theta) \\
\log(p(z_i|\theta)) &= -\frac{1}{2}\left(\frac{\xi_i}{\sigma_\xi}\right)^2 - \log(\sigma_\xi) - \frac{1}{2}\left(\frac{\tau_i}{\sigma_\tau}\right)^2 - \log(\sigma_\tau) + cste
\end{aligned} \tag{5}
$$

$$
\begin{aligned}
p(\theta) &= p(\bar{t_0}, \bar{v_0}, \sigma_\xi, \sigma_\tau, \sigma) \\
&= p(\bar{t_0})p(\bar{v_0})p(\sigma_\xi)p(\sigma_\tau)p(\sigma) \qquad \text{all parameters are independent in the prior} \\
\log(p(\theta)) &= -\frac{1}{2}\left(\frac{\bar{t_0} - \bar{\bar{t_0}}}{s_{t_0}}\right)^2 - \log(s_{t_0}) - \frac{1}{2}\left(\frac{\bar{v_0} - \bar{\bar{v_0}}}{s_{v_0}}\right)^2 - \log(s_{v_0}) - (m+2)\log(\sigma_\xi) - \frac{v^2}{2\sigma_\xi^2} \\
&\quad - (m+2)\log(\sigma_\tau) - \frac{v^2}{2\sigma_\tau^2} - (m+2)\log(\sigma) - \frac{v^2}{2\sigma^2} + cste
\end{aligned} \tag{6}
$$

The log-likelihood can be written under the explicit form $\log(p(y, z, \theta)) = -\Phi(\theta) + \langle\ S(y, z), \Psi(\theta)\ \rangle$ up to some constant independent of θ with :

$$
S(y,z) = \begin{bmatrix} t_0 \\ v_0 \\ \frac{1}{N}\sum_{i=1}^{N} \xi_i^2 \\ \frac{1}{N}\sum_{i=1}^{N} \tau_i^2 \\ \frac{1}{KN}\sum_{i=1}^{N}\sum_{j=1}^{K}(y_{ij} - d_i(t_{ij}))^2 \\ -\frac{1}{2}\frac{t_0^2}{\sigma_{t_0}^2} - \frac{1}{2}\frac{v_0^2}{\sigma_{v_0}^2} \end{bmatrix} = \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{bmatrix} \tag{7}
$$

$$
\Psi(\theta) = \begin{bmatrix} \frac{\bar{t_0}}{\sigma_{t_0}^2} \\ \frac{\bar{v_0}}{\sigma_{v_0}^2} \\ -\frac{N}{2\sigma_\xi^2} \\ -\frac{N}{2\sigma_\tau^2} \\ -\frac{NK}{2\sigma^2} \\ 1 \end{bmatrix} = \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \\ \Psi_4 \\ \Psi_5 \\ \Psi_6 \end{bmatrix} \tag{8}
$$

$$
\begin{aligned}
-\Phi(\theta) = &-\frac{v^2}{2}\left(\frac{1}{\sigma_\xi^2} + \frac{1}{\sigma_\tau^2} + \frac{1}{\sigma^2}\right) - (m+2+N)\log(\sigma_\tau) - (m+2+N)\log(\sigma_\xi) \\
&- (m+2+NK)\log(\sigma) + \bar{t_0}\frac{\bar{\bar{t_0}}}{s_{t_0}^2} - \frac{1}{2}\frac{\bar{t_0}^2}{s_{t_0}^2} + \bar{v_0}\frac{\bar{\bar{v_0}}}{s_{v_0}^2} - \frac{1}{2}\frac{\bar{v_0}^2}{s_{v_0}^2} - \frac{1}{2}\frac{\bar{t_0}^2}{\sigma_{t_0}^2} - \frac{1}{2}\frac{\bar{v_0}^2}{\sigma_{v_0}^2}
\end{aligned} \tag{9}
$$

and the constant independent of θ is :

$$
-\log(\sigma_{t_0}) - \log(\sigma_{v_0}) - \log(s_{t_0}) - \log(s_{v_0}) - \frac{1}{2}\frac{\bar{\bar{t_0}}^2}{s_{t_0}^2} - \frac{1}{2}\frac{\bar{\bar{v_0}}^2}{s_{v_0}^2}
$$

2. We generate synthetic data from the model by taking some reasonable values for the parameters : $\sigma_{t_0} = \sigma_{v_0} = 0.1, s_{t_0} = s_{v_0} = 0.1, \bar{t_0} = \bar{v_0} = 1, m = m_\xi = m_\tau \in [5, 10], v = v_\xi = v_\tau \in [1, 5], N = 100, k_i = 20$
We obtain the synthetic ground truth θ_{real} :

```
real_theta :
array([0.96226364, 1.06828024, 1.09811226, 1.29905008, 1.26838235])
```

Figure 1 shows the associated trajectories for the measurements artificially generated :

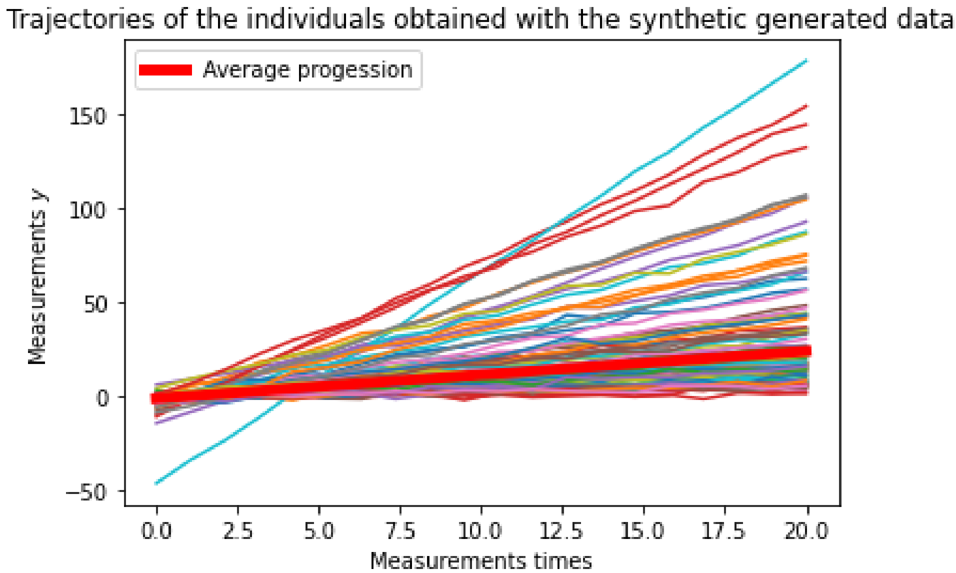

FIGURE 1 – Average and individual-specific progressions of some disease for the artificial observations $(y_{i,j})_{i,j}$ generated from our parameters and latent variables

1.B -HM-SAEM – Hasting-Metropolis sampler

3. In order to estimate – by a maximum a posteriori for instance – the parameter θ of this statistical model, we use the SAEM – Stochastic Approximation EM – algorithm. Since it requires to sample from the a posteriori distribution $p(z|y, \theta)$ of the latent variable, we use the Symmetric Random Walk Metropolis-Hastings sampler to try to sample from this distribution.

For the acceptance rate, the target distribution $\pi = p(z|y, \theta) = \frac{p(z,y,\theta)}{p(y,\theta)}$ so if we denote z^* the proposal state and z^k the current state, we have $\log(\pi(z^*)) - \log(\pi(z^k)) = \log(p(z^*, y\theta) - \log(p(y, \theta) - \log(p(z^*, y\theta) + \log(p(y, \theta) = \langle\ S(y, z^*) - S(y, z^k), \Psi(\theta)\ \rangle$. Thus, the logarithm of the acceptance ratio becomes

$$\min(0, \langle\ S(y, z^*) - S(y, z^k), \Psi(\theta)\ \rangle)$$

The results of a run of 50000 iterations of this sampler given the real synthetic θ_{real} is given in Figure 2 with a Q-Q plot representation. The acceptance rate of this run was between 0.2 and 0.4. We can see that the approximation seems reasonable, considering that the initial points are quite far from the true distribution. However, it is noteworthy that we struggle to reach the extreme values for τ_1 and ξ_1 in particular.

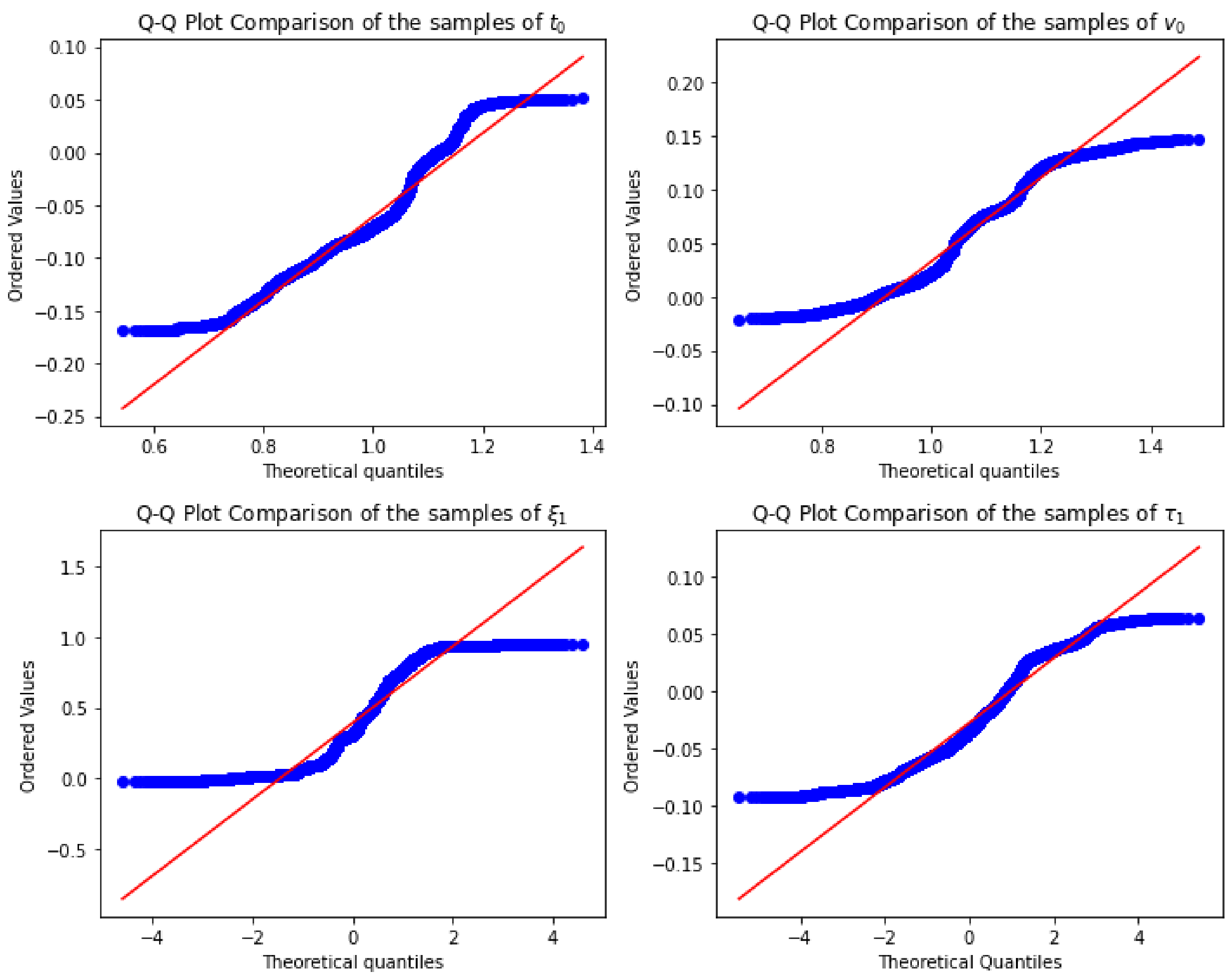

FIGURE 2 – Q-Q plot graph of the samples approximating the distributions of the latent variables t_0, v_0, ξ_1, τ_1 against their theoretical quantiles.

4. With our MH sampler, we can achieve the simulation requirement in the expectation step of the SAEM. As our model belongs to the curved exponential family, the Stochastic Approximation step can be written : $S^{(k+1)} = S^{(k)} + \epsilon_k(S(y, z^{(k+1)}) - S^{(k)})$ and the maximization step consists of finding : $\theta^{(k+1)} = \underset{\theta\in\Theta}{\operatorname{argmax}} \left\{-\Phi(\theta) + \left\langle S^{(k+1)} \mid \Psi(\theta)\right\rangle\right\}$. We denote the function $f^k = -\Phi + \left\langle S^{(k)} \mid \Psi\right\rangle$, it is log-concave. We compute the derivative of f^k with regard to θ to obtain $\theta^{(k)}$.

$$\begin{aligned} \frac{\partial f^k}{\partial \bar{t_0}}(\theta) &= -\frac{\partial \Phi}{\partial \bar{t_0}}(\theta) + S^{(k)} \frac{\partial \Psi}{\partial \bar{t_0}}(\theta) \\ &= \frac{\bar{\bar{t_0}}}{s_{t_0}^2} - \frac{\bar{t_0}}{s_{t_0}^2} - \frac{\bar{t_0}}{\sigma_{t_0}^2} + \frac{S_1^{(k)}}{\sigma_{t_0}^2} \end{aligned} \tag{10}$$

$$\begin{aligned} &\frac{\partial f^k}{\partial \bar{t_0}}(\theta) = 0 \\ \Leftrightarrow \; &\boxed{\bar{t_0} = \frac{\frac{S_1^{(k)}}{\sigma_{t_0}^2} + \frac{\bar{\bar{t_0}}}{s_{t_0}^2}}{\frac{1}{s_{t_0}^2} + \frac{1}{\sigma_{t_0}^2}}} \end{aligned} \tag{11}$$

and with the same calculations, we obtain :

$$\boxed{\bar{v_0} = \frac{\frac{S_2^{(k)}}{\sigma_{v_0}^2} + \frac{\bar{\bar{v_0}}}{s_{v_0}^2}}{\frac{1}{s_{v_0}^2} + \frac{1}{\sigma_{v_0}^2}}} \tag{12}$$

$$\begin{aligned}\frac{\partial f^k}{\partial \sigma_\xi^2}(\theta) &= -\frac{\partial \Phi}{\partial \sigma_\xi^2}(\theta) + S^{(k)} \frac{\partial \Psi}{\partial \sigma_\xi^2}(\theta) \\ &= \frac{v^2}{2} \frac{1}{(\sigma_\xi^2)^2} - \frac{m+2+N}{2} \frac{1}{\sigma_\xi^2} + S_3^{(k)} \frac{N}{2(\sigma_\xi^2)^2}\end{aligned} \tag{13}$$

$$\begin{aligned}&\frac{\partial f^k}{\partial \sigma_\xi^2}(\theta) = 0 \\ \Leftrightarrow \quad &\boxed{\sigma_\xi^2 = \frac{v^2 + S_3^{(k)} N}{m+2+N}}\end{aligned} \tag{14}$$

and with the same calculations, we obtain :

$$\boxed{\sigma_\tau^2 = \frac{v^2 + S_4^{(k)} N}{m+2+N}} \tag{15}$$

$$\boxed{\sigma^2 = \frac{v^2 + S_5^{(k)} KN}{m+2+KN}} \tag{16}$$

It enables us to proceed the maximization step.

For the step-sizes (ϵ_k), the burn-in parameter N_b is chosen to be equal to the maximum number of iterations divided by 4 and to ensure convergence $\alpha = 0.75$. For an initial $\theta_{\text{init}} = [0, 0, 1, 1, 1]$ and $z_{\text{init}} = 0_{\mathbb{R}^{2N+2}}$, one obtains after 10000 iterations and with a MH sampler of number of iterations equals to 500 :

```
estimated_theta :
array([1.02639014, 0.95906801, 1.11179078, 0.88143494, 1.1212836 ])
np.linalg.norm(real_theta-estimated_theta)
0.46072449975045926
```

Graphically, Figure 3 depicts the evolution of the parameters θ estimated over iterations. It seems to reasonably approach the true values of θ_{real}, except for σ_τ.

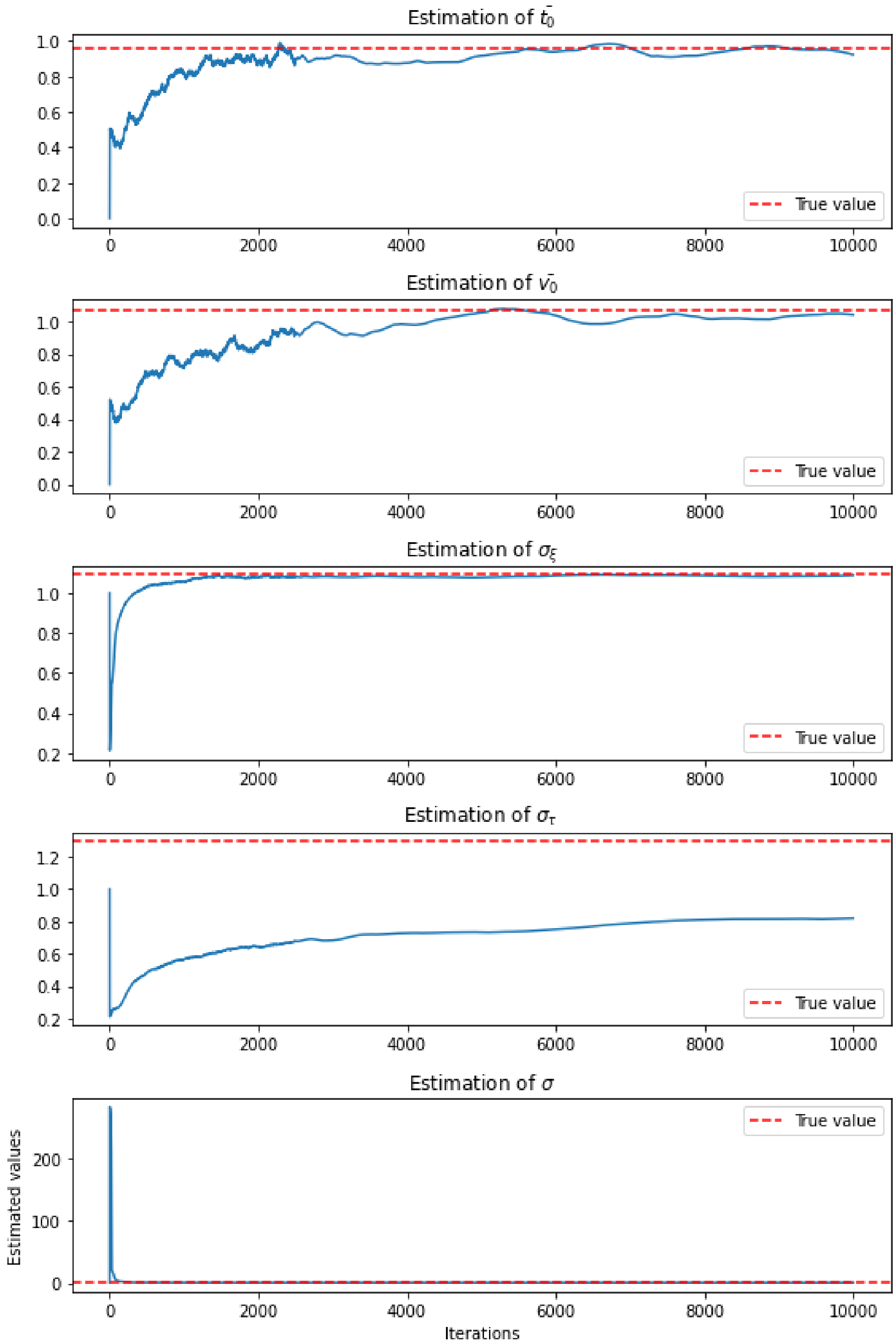

FIGURE 3 – Evolution of the estimated parameters θ over iterations of the SAEM

1.C – HMwG-SAEM – Hasting-Metropolis within Gibbs sampler

5. We now want to sample (sub)-variable of the latent space conditionally on the current values of the other (sub)-variables to reduce the dimension of the space we sample within. We implement the Metropolis-Hastings within Gibbs sampler. First, we sample each coordinates of z separately and conditionally on the others.

If $z_i = (\xi_i, \tau_i)$ denotes a particular individual we have :

$$\begin{aligned} p(z_i|z_{pop}, y, \theta) &= \frac{p(z_i, z_{pop}, y, \theta)}{p(z_{pop}, y, \theta)} \\ &= \frac{p(z_i, z_{pop}, y, \theta) \prod_{j \neq i} p(z_j, y, \theta)}{p(z_{pop}, y, \theta) \prod_{j \neq i} p(z_j, y, \theta)} \\ &= \frac{p(\{z_i\}, z_{pop}, y, \theta)}{p(z_{pop}, y, \theta) \prod_{j \neq i} p(z_j, y, \theta)} \end{aligned} \tag{17}$$

Thus, again if we look at the logarithm of the acceptance ratiofor the MH part, the expression undergoes simplification as the denominator does not depend on z :

$$\log(p(z^*|z_{pop}, y, \theta)) - \log(p(z^k|z_{pop}, y, \theta)) = \left\langle S(y, z^*) - S(y, z^k) \mid \Psi(\theta) \right\rangle$$

and it it the same calculations for the *a posteriori* distribution $p(z_{pop}|\{z_i\}, y, \theta)$.

6.7. The HMwG-SAEM is implemented using the previous functions that leverage the HMwG algorithm to sample from the *a posteriori* distribution $p(z_{pop}|\{z_i\}, y, \theta)$ and $p(z_i|z_{pop}, y, \theta)$. It is stoped after 500 iterations with $N_b = 50$ and $\alpha = 0.75$ and a number of iterations equals to 20 for the MH part of the HMwG sampler. We obtain the following estimated θ for the same θ_{init} and z_{init} as the previous method :

```
estimated_theta :
array([0.99294535, 1.01362528, 1.09676481, 1.11137909, 1.11539255])
np.linalg.norm(real_theta-estimated_theta)
0.2501132241230222
```

The norm of the difference between $\theta_{\text{estimated}}$ and θ_{real} is divided by 2 compare to the same metric obtained with the previous approach. It is also possible to visualize the improvement of our estimator for θ_{real} graphically in Figure 4. The parameters converge towards values closer to the true ones.

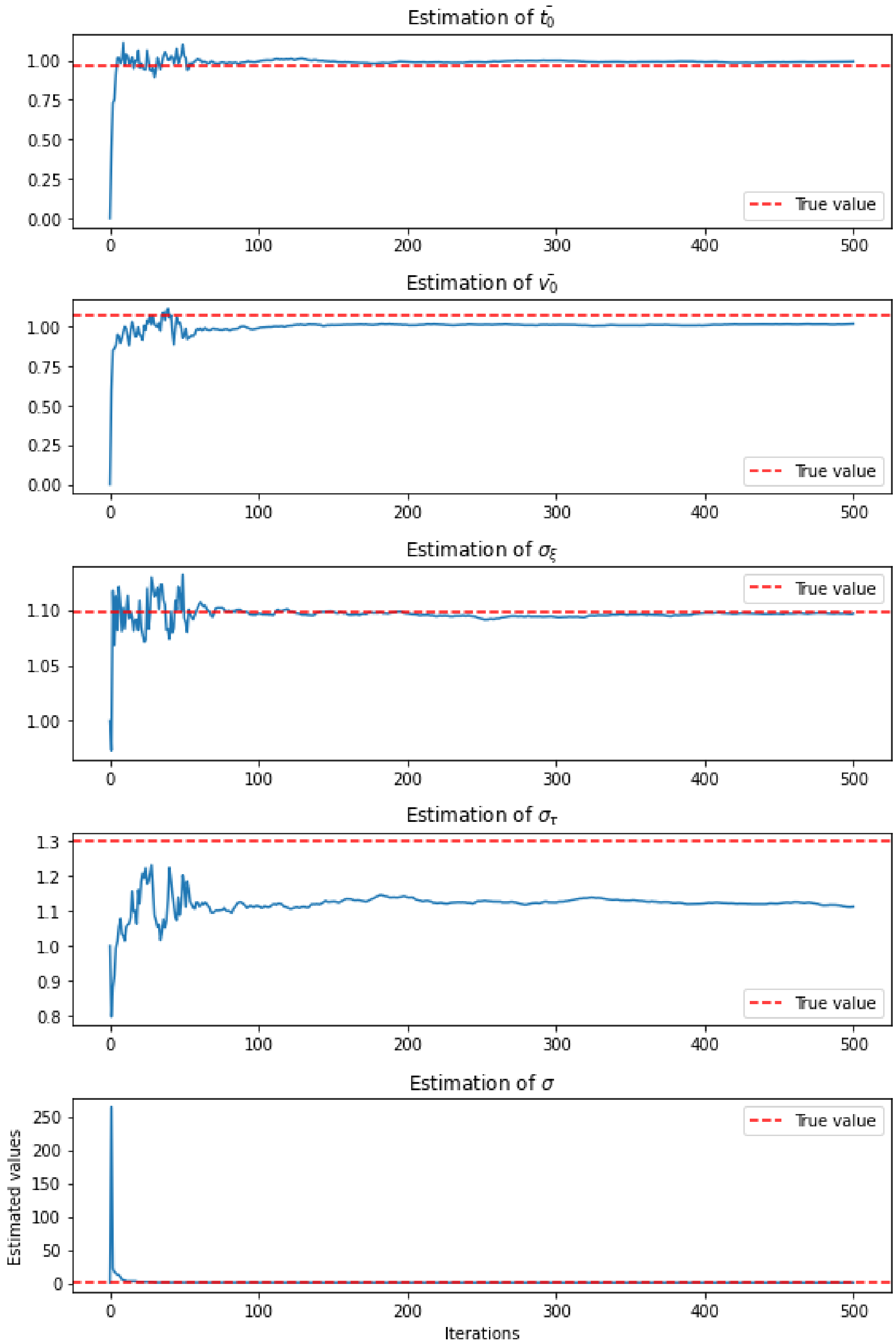

FIGURE 4 – Evolution of the estimated parameters θ over iterations of the HMwG-SAEM

We can improve the sampling step for big dataset by considering a Block HMwG sampler instead of a "one-at-a-time" as described above HMwG sampler. In the Block version, each Metropolis-Hastings step of the algorithm consists in a multivariate symmetric random walk. Then, the Block MHwG sampler updates simultaneously block (or sets) of latent variables given the others.

8. The block Gibbs sampler strikes a balance between the traditional Hastings-Metropolis sampler and the Hastings-Metropolis within Gibbs sampler. By sampling a block at a time, it reduces the iterations required compared to the HMwG which samples each coordinate, so it can speed up the process. Additionally, it minimizes rejected proposals by focusing on one block at a time as it reduces the dimension of the space in which we seek a new state to move to unlike the classic HM that may face challenges in too high-dimension. This algorithm excels when well-chosen blocks are employed, optimizing its performance.

9. We implement a Block HMwG sampler by choosing a block for the fixed effects and a block by individuals, in the SAEM framework. The acceptance ratio is the same, the only difference is in the update of the proposal. Given the same configuration and initial conditions as for the HMwG-SAEM algorithm, we obtain the following estimated θ :

```
estimated_theta :
array([0.99330513, 1.00566061, 1.0929226 , 1.12790498, 1.11913776])
np.linalg.norm(real_theta-estimated_theta)
0.23764746243382856
```

The norm of the difference between $\theta_{\text{estimated}}$ and θ_{real} seems slightly better for the Block HMwG sampler. Figure 5 illustrates the graphical results of the algorithm's evolution for the θ parameters. The convergence of $\bar{t_0}$ and $\bar{v_0}$ seems more stable. More globally the estimated parameters values seem to converge closer to the real values. From an implementation point of view, with the same starting configuration, the Block HMwG-SAEM took more than twice less time ($\sim 40s$) to execute than the classic HMwG-SAEM ($\sim$ 1min30s). This is what we could have expected as we divide the number of iterations by 2.

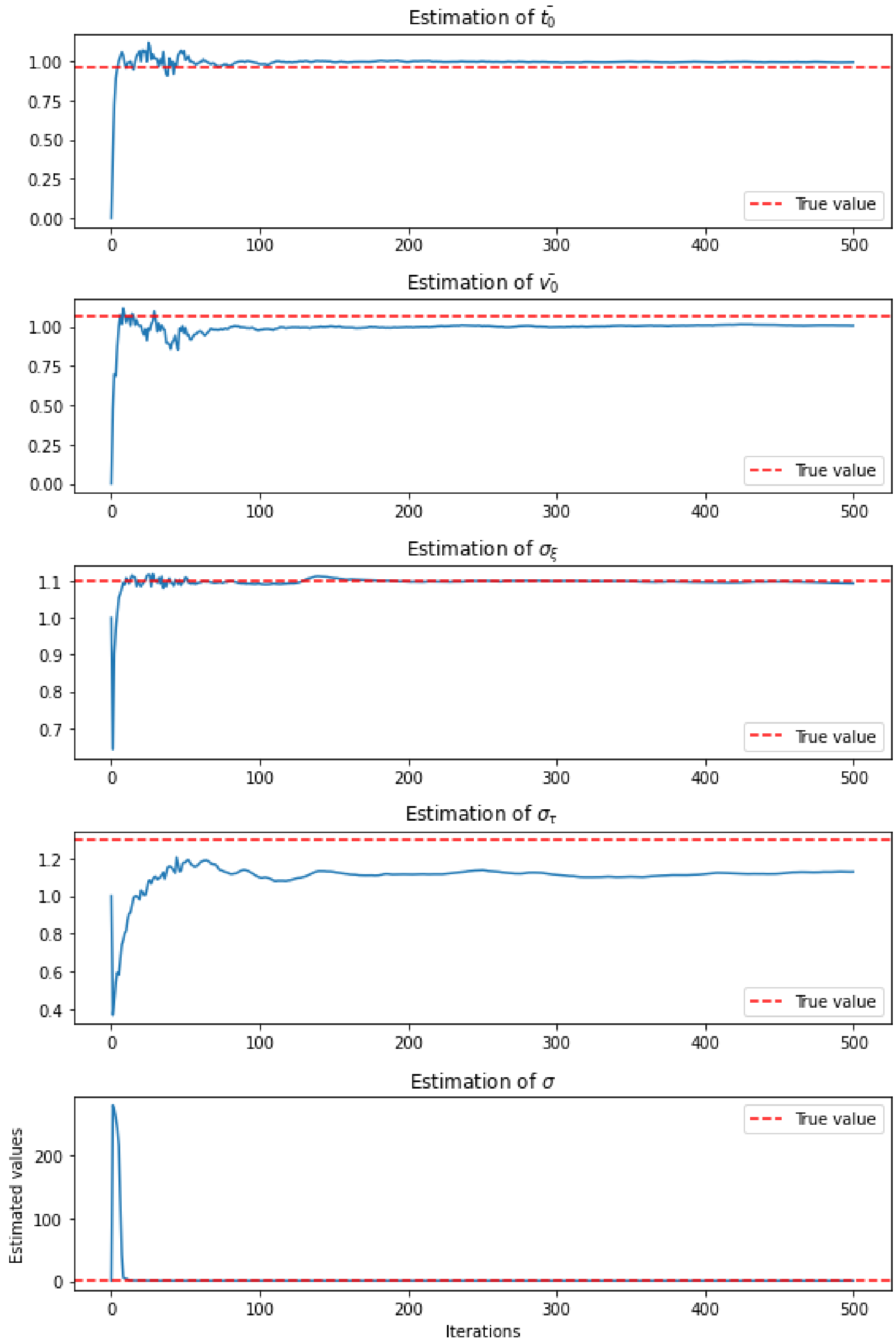

FIGURE 5 – Evolution of the estimated parameters θ over iterations of the Block HMwG-SAEM

Exercise 2 : Multiplicative Hasting-Metropolis

Let f be the density of some distribution π_f supported on $]-1,1[$. We consider the multiplicative Hasting-Metropolis algorithm defined as follows. Let X be the current state of the Markov chain.

(i) We sample ε from π_f and from a random variable $\mathcal{B}$ that has the Bernoulli distribution with parameter $\frac{1}{2}$.
(ii) If $\mathcal{B}=1$, we set $Y=\varepsilon X$. Otherwise, we set $Y=\frac{X}{\varepsilon}$. Then, we accept the candidate Y with a probability given by $\alpha(X,Y)$, the usual Hasting-Metropolis acceptance ratio.

1. Let denote $A \in \mathcal{X}$ and a current state x, the kernel is of the form

$$P(x,A)=\int_A q(x,y)\alpha(x,y)\,dy+\delta_x(A)\int_{\mathcal{X}}(1-\alpha(x,y))q(x,y)\,dy \tag{18}$$

where q is the proposal kernel of the MCMC step and $\alpha(x,y)$ is the acceptance rate.

First, we show that for a fixed $x\neq 0$, $\{\varepsilon x|\varepsilon\in \mathrm{Supp}(\pi_f)\}$ is disjoint with $\{\frac{x}{\varepsilon}|\varepsilon\in\mathrm{Supp}(\pi_f)\}$. Indeed, $\mathrm{Supp}(\pi_f)\subset]-1,1[$ so ε takes values in the open interval $]-1,1[$ and there is no value a in this interval such that $ax=\frac{x}{a}$. So based on the algorithm description, P can be written like :

$$\begin{aligned}P(x,A)=&\frac{1}{2}\int_{-1}^{1}\mathbb{1}_{\varepsilon x}(A)\alpha(x,\varepsilon x)f(\varepsilon)\,d\varepsilon+\frac{1}{2}\int_{-1}^{1}\mathbb{1}_{\frac{x}{\varepsilon}}(A)\alpha(x,\frac{x}{\varepsilon})f(\varepsilon)\,d\varepsilon\\&+\delta_x(A)\left[\frac{1}{2}\int_{-1}^{1}(1-\alpha(x,\varepsilon x))f(\varepsilon)\,d\varepsilon+\frac{1}{2}\int_{-1}^{1}(1-\alpha(x,\frac{x}{\varepsilon}))f(\varepsilon)\,d\varepsilon\right]\end{aligned} \tag{19}$$

We make the change of variable $y=\varepsilon x$ and $y=\frac{x}{\varepsilon}$ so we have to look at two cases, either $x>0$ or $x<0$. If $x>0$:

$$\begin{aligned}P(x,A)=&\frac{1}{2}\int_{]-x,x[\setminus 0}\mathbb{1}_y(A)\alpha(x,y)f(\frac{y}{x})\frac{1}{x}\,dy+\frac{1}{2}\int_{+\infty}^{x}\mathbb{1}_y(A)\alpha(x,y)f(\frac{x}{y})\frac{-x}{y^2}\,dy+\\&\frac{1}{2}\int_{-x}^{-\infty}\mathbb{1}_y(A)\alpha(x,y)f(\frac{x}{y})\frac{-x}{y^2}\,dy+\\&\delta_x(A)\left[\frac{1}{2}\int_{]-x,x[\setminus 0}(1-\alpha(x,y))f(\frac{y}{x})\frac{1}{x}\,dy+\frac{1}{2}\int_{+\infty}^{x}(1-\alpha(x,y))f(\frac{x}{y})\frac{-x}{y^2}\,dy+\frac{1}{2}\int_{-x}^{-\infty}(1-\alpha(x,y))f(\frac{x}{y})\frac{-x}{y^2}\,dy\right]\\=&\frac{1}{2}\int_{\mathbb{R}\setminus 0}\mathbb{1}_y(A)\alpha(x,y)f(\frac{y}{x})\frac{1}{x}\,dy+\frac{1}{2}\int_{\mathbb{R}\setminus 0}\mathbb{1}_y(A)\alpha(x,y)f(\frac{x}{y})\frac{x}{y^2}\,dy+\\&\delta_x(A)\left[\frac{1}{2}\int_{\mathbb{R}\setminus 0}(1-\alpha(x,y))f(\frac{y}{x})\frac{1}{x}\,dy+\frac{1}{2}\int_{\mathbb{R}\setminus 0}(1-\alpha(x,y))f(\frac{x}{y})\frac{x}{y^2}\,dy\right]\\=&\int_{A\setminus 0}\frac{1}{2}\left[f\left(\frac{y}{x}\right)\frac{1}{x}+f\left(\frac{x}{y}\right)\frac{x}{y^2}\right]\alpha(x,y)\,dy+\delta_x(A)\left[\int_{\mathbb{R}\setminus 0}(1-\alpha(x,y))\left[f\left(\frac{y}{x}\right)\frac{1}{x}+f(\frac{x}{y})\frac{x}{y^2}\right]dy\right]\end{aligned} \tag{20}$$

And similarly, if $x<0$:

$$P(x,A)=\int_{A\setminus 0}\frac{1}{2}\left[f\left(\frac{y}{x}\right)\frac{-1}{x}+f\left(\frac{x}{y}\right)\frac{-x}{y^2}\right]\alpha(x,y)\,dy+\delta_x(A)\left[\int_{\mathbb{R}\setminus 0}(1-\alpha(x,y))\left[f\left(\frac{y}{x}\right)\frac{-1}{x}+f(\frac{-x}{y})\frac{x}{y^2}\right]dy\right] \tag{21}$$

Thus by identification of the formula (18), we deduce the following writing of q that unifies the case $x<0$ and $x>0$:

$$q(x,y)=\frac{1}{2}\left[f\left(\frac{y}{x}\right)\frac{1}{|x|}+f\left(\frac{x}{y}\right)\frac{|x|}{y^2}\right]\qquad\text{for } x\neq 0 \text{ and } y\neq 0 \tag{22}$$

and we take $q(x,0)=0$ for all x and we do not initialize with $x=0$.

2. For a given distribution π, we take the usual acceptance ratio $\alpha(x,y)=\min\left(1, \frac{\pi(y)q(y,x)}{q(x,y)\pi(x)}\right)$

$$\begin{aligned}
\frac{\pi(y)q(y,x)}{q(x,y)\pi(x)} &= \frac{\pi(y)}{\pi(x)}\frac{f\left(\frac{x}{y}\right)\frac{1}{|y|}+f\left(\frac{y}{x}\right)\frac{|y|}{x^2}}{f\left(\frac{y}{x}\right)\frac{1}{|x|}+f\left(\frac{x}{y}\right)\frac{|x|}{y^2}} \\
&= \frac{\pi(y)}{\pi(x)}\frac{|y|}{|x|}\frac{f\left(\frac{y}{x}\right)\frac{1}{|x|}+f\left(\frac{x}{y}\right)\frac{|x|}{y^2}}{f\left(\frac{y}{x}\right)\frac{1}{|x|}+f\left(\frac{x}{y}\right)\frac{|x|}{y^2}} \\
&= \frac{\pi(y)}{\pi(x)}\frac{|y|}{|x|}
\end{aligned} \tag{23}$$

So finally, we have the acceptance ratio $\alpha(x,y)=\min\left(1, \frac{\pi(y)}{\pi(x)}\frac{|y|}{|x|}\right)$ for $x\neq 0$ and $y\neq 0$.

3. 4. We implement this sampler where f is given by the uniform distribution on $(-1,1)$. First, we apply it on a Gaussian distribution of mean $\mu=-1$ and variance $\sigma^2=1$. Figure 6 shows the histogram of the sampled points.

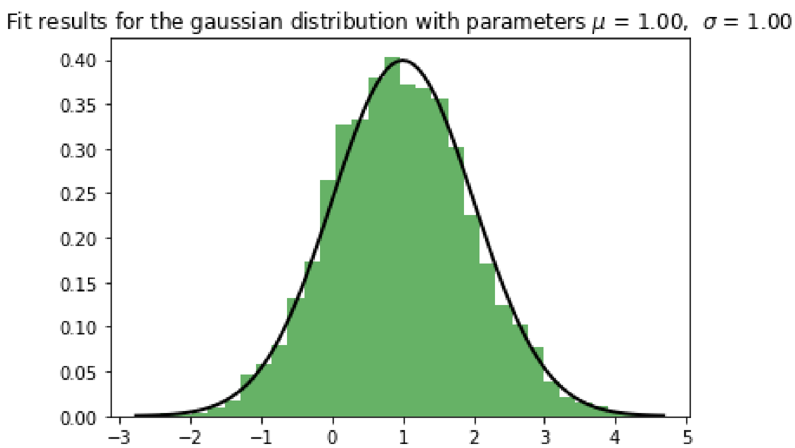

Figure 6 – Histogram of the points obtained from the Multiplicative Hasting-Metropolis algorithm with 5000 points obtained from the sampler with 30 iterations for each. The black curve represents the pdf of the target Gaussian distribution

For a second application, the algorithm is used to sample according to a Cauchy distribution which has heavy tails compare to a distribution belonging to the family of exponential distributions. Figure 7 shows the resulting histogram for a Cauchy distribution of parameters $x_0=-2$ and $\gamma=2$:

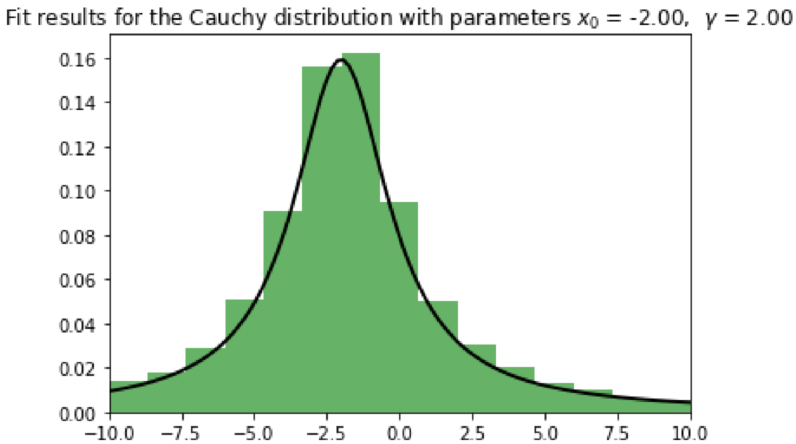

FIGURE 7 – Histogram of the points obtained from the Multiplicative Hasting-Metropolis algorithm with 5000 points obtained from the sampler with 30 iterations for each. The black curve represents the pdf of the target Cauchy distribution

In both cases, the empirical histograms align well graphically with the theoretical pdfs. Figure 8 shows the QQ-plots of the samples against their true distributions :

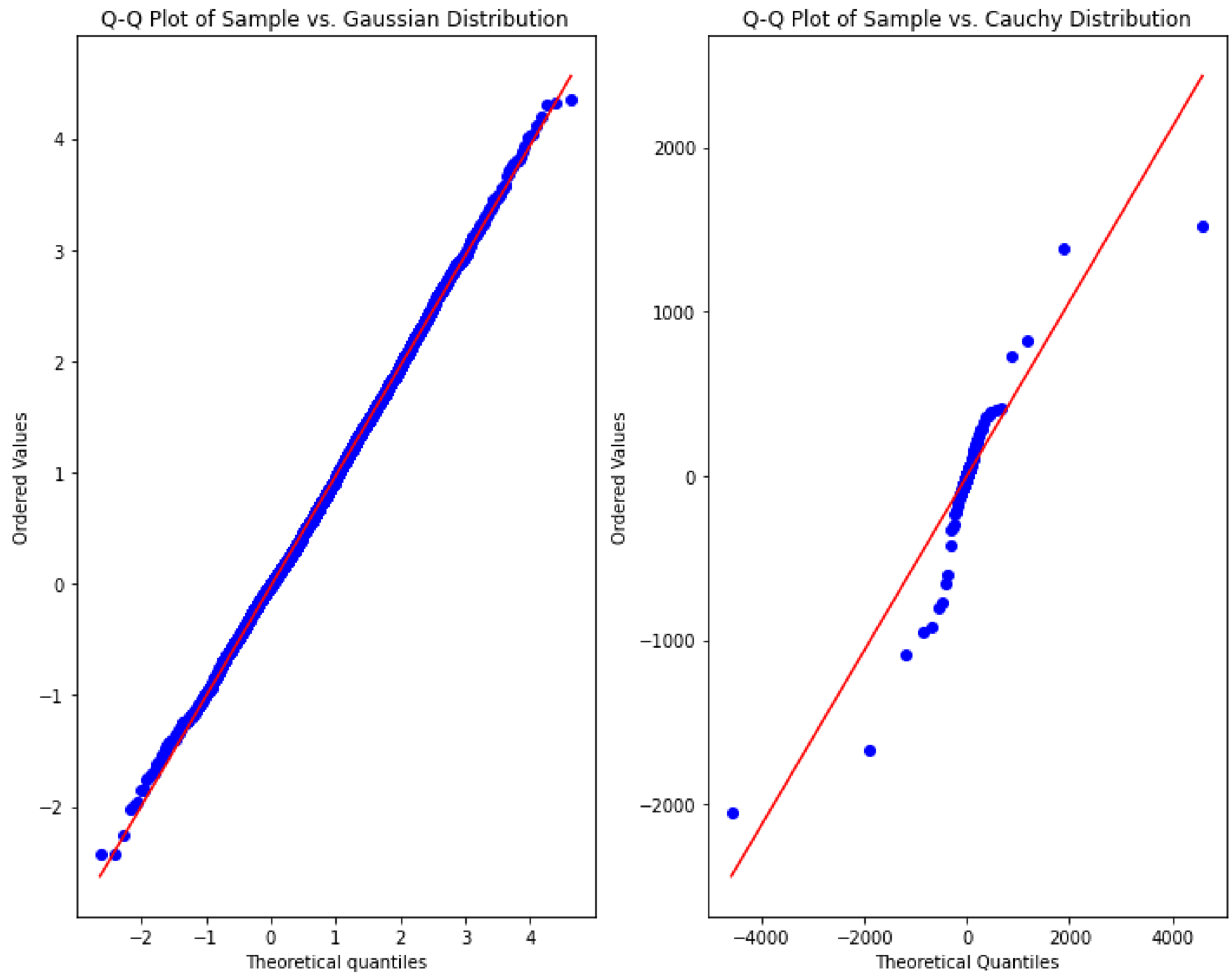

FIGURE 8 – QQ-plot for the sampled points against the theoretical quantiles of their corresponding true distributions (left : Gaussian, right : Cauchy)

We observe that the sampled points for the Gaussian case align very well on the line, indicating successful sampling from the true distribution. This is less pronounced for the Cauchy distribution, especially in the more extended extreme values. It is more challenging to sample from this distribution due to its heavy tails, requiring exploration of a larger state space to approach its true distribution