Projet MCMC

Maria Bastiani, Vincent Farcinade, Amanda Gobeau

Question 1 :

Our initial model is defined by, for some $p \in [0,1]$, $\theta \in [0,1]$:

$$\mathbb{P}(X = AA) = p(1-\theta) + (1-p)(1-\theta)^2$$
$$\mathbb{P}(X = aa) = p\theta + (1-p)\theta^2$$
$$\mathbb{P}(X = aA) = 2(1-p)\theta(1-\theta)$$

By considering $Z \sim \mathcal{B}(p)$, we can write :

$$\mathbb{P}(X = AA, Z = 1) = p(1-\theta)$$
$$\mathbb{P}(X = AA, Z = 0) = (1-p)(1-\theta)^2$$
$$\mathbb{P}(X = aa, Z = 1) = p\theta$$
$$\mathbb{P}(X = aa, Z = 0) = (1-p)\theta^2$$
$$\mathbb{P}(X = aA, Z = 1) = 0$$
$$\mathbb{P}(X = aA, Z = 0) = 2(1-p)\theta(1-\theta)$$

Here, we only dispose of the allocation of the three alleles within the population, indeed the variable Z is not observed.

Thus, this representation is a latent representation of our model, using a Bernouilli random variable of parameter p.

Question 2 :

From now on, we denote $\eta = (p, \theta)$, and we consider $x = (x_1, \cdots, x_n)$ a sample of realizations of the variables $X_1, \cdots, X_n$ following the model defined in question 1 and $Z_1, \cdots, Z_n$ the associated latent variables, for $n \in \mathbb{N}$.

The Expectation - Maximization algorithm is defined as follows :

Algorithm 1: Expectation - Maximization Algorithm

Input: Initialization value $\eta^{(0)} \in \Theta$ and T iterations.

---
1 for $t \in [\![1, T]\!]$ **do**

2 $\quad$ EXPECTATION STEP : $Q\left(\eta; \eta^{(t)}\right) = \mathbb{E}_{\eta^{(t)}}\left[\ell\left(\eta \mid X, Z\right) \mid X = x\right]$

3 $\quad$ MAXIMIZATION STEP : $\eta^{(t+1)} = \arg\max_{\eta} Q\left(\eta; \eta^{(t)}\right)$;

4 end

Thus, for the first step, at a given iteration t, we want to compute :

$$Q\left(\eta; \eta^{(t)}\right) = \mathbb{E}_{\eta^{(t)}}\left[\ell\left(\eta \mid X, Z\right) \mid X = x\right]$$

$$= \mathbb{E}_{\eta^{(t)}}\left[\sum_{i=1}^{n} \mathbb{1}_{Z_i=1} \log\left\{\mathbb{P}(X_i = x_i, Z_i = 1)\right\} + \mathbb{1}_{Z_i=0} \log\left\{\mathbb{P}(X_i = x_i, Z_i = 0)\right\} \mid X_i = x_i\right]$$

$$= \sum_{i=1}^{n} \log\left\{\mathbb{P}(X_i = x_i, Z_i = 1)\right\} \mathbb{P}_{\eta^{(t)}}(Z_i = 1 \mid X_i = x_i) + \log\left\{\mathbb{P}(X_i = x_i, Z_i = 0)\right\} \mathbb{P}_{\eta^{(t)}}(Z_i = 0 \mid X_i = x_i)$$

In our case,

$$\mathbb{P}(X_i = x_i, Z_i = 1) = p((1-\theta)\mathbb{1}_{x_i=AA} + \theta\mathbb{1}_{x_i=aa})$$
$$\mathbb{P}(X_i = x_i, Z_i = 0) = (1-p)\left[(1-\theta)^2\right]^{\mathbb{1}_{x_i=AA}}\left[\theta^2\right]^{\mathbb{1}_{x_i=aa}}\left[2\theta(1-\theta)\right]^{\mathbb{1}_{x_i=aA}}$$

Moreover, we have

$$\mathbb{P}_{\eta^{(t)}}(Z_i = 1 \mid X_i = x_i) = \frac{\mathbb{P}(X_i = x_i, Z_i = 1)}{\mathbb{P}(X_i = x_i, Z_i = 1) + \mathbb{P}(X_i = x_i, Z_i = 0)}$$

$$\mathbb{P}_{\eta^{(t)}}(Z_i = 0 \mid X_i = x_i) = \frac{\mathbb{P}(X_i = x_i, Z_i = 0)}{\mathbb{P}(X_i = x_i, Z_i = 1) + \mathbb{P}(X_i = x_i, Z_i = 0)}$$

So :

$$\mathbb{P}_{\eta^{(t)}}(Z_i = 1 \mid X_i = AA) = \frac{p^{(t)}(1-\theta^{(t)})}{p^{(t)}(1-\theta^{(t)}) + (1-p^{(t)})(1-\theta^{(t)})^2} := p_{AA,1}^{(t)}$$

$$\mathbb{P}_{\eta^{(t)}}(Z_i = 1 \mid X_i = aa) = \frac{p^{(t)}\theta^{(t)}}{p^{(t)}\theta^{(t)} + (1-p^{(t)})\theta^{(t)^2}} := p_{aa,1}^{(t)}$$

$$\mathbb{P}_{\eta^{(t)}}(Z_i = 1 \mid X_i = aA) = 0$$

$$\mathbb{P}_{\eta^{(t)}}(Z_i = 0 \mid X_i = AA) = \frac{(1-p^{(t)})(1-\theta^{(t)})^2}{p^{(t)}(1-\theta^{(t)}) + (1-p^{(t)})(1-\theta^{(t)})^2} := p_{AA,0}^{(t)}$$

$$\mathbb{P}_{\eta^{(t)}}(Z_i = 0 \mid X_i = aa) = \frac{(1-p^{(t)})\theta^{(t)^2}}{p^{(t)}\theta^{(t)} + (1-p^{(t)})\theta^{(t)^2}} := p_{aa,0}^{(t)}$$

$$\mathbb{P}_{\eta^{(t)}}(Z_i = 0 \mid X_i = aA) = 1$$

By defining,

$$n_{AA} := \sum_{i=1}^{n} \mathbb{1}_{x_i=AA}$$

$$n_{aa} := \sum_{i=1}^{n} \mathbb{1}_{x_i=aa}$$

$$n_{aA} := \sum_{i=1}^{n} \mathbb{1}_{x_i=aA}$$

We finally have that :

$$Q\left(\eta;\eta^{(t)}\right) = \sum_{i=1}^{n} \log\left\{p((1-\theta)\mathbb{1}_{x_i=AA} + \theta\mathbb{1}_{x_i=aa})\right\}(p_{AA,1}^{(t)}\mathbb{1}_{x_i=AA} + p_{aa,1}^{(t)}\mathbb{1}_{x_i=aa})$$

$$+ \log\left\{(1-p)\left[(1-\theta)^2\right]^{\mathbb{1}_{x_i=AA}}\left[\theta^2\right]^{\mathbb{1}_{x_i=aa}}\left[2\theta(1-\theta)\right]^{\mathbb{1}_{x_i=aA}}\right\}(p_{AA,0}^{(t)}\mathbb{1}_{x_i=AA} + p_{aa,0}^{(t)}\mathbb{1}_{x_i=aa} + \mathbb{1}_{x_i=aA})$$

$$= \sum_{i=1}^{n}(n_{AA}p_{AA,0}^{(t)} + n_{aa}p_{aa,0}^{(t)} + n_{aA})\log\{1-p\}$$

$$+ 2n_{AA}p_{AA,0}^{(t)}\log\{1-\theta\} + 2n_{aa}p_{aa,0}^{(t)}\log\{\theta\} + n_{AA}\log\{2\theta(1-\theta)\} + n_{AA}p_{AA,1}^{(t)}\log\{p(1-\theta)\}$$

$$+ n_{aa}p_{aa,1}^{(t)}\log\{p\theta\}$$

$$= (n_{AA}pAA,0^{(t)} + n_{aa}p_{aa,0}^{(t)} + n_{aA})\log\{1-p\} + (n_{AA}p_{AA,1}^{(t)} + n_{aa}p_{aa,1}^{(t)})\log\{p\}$$

$$+ (2n_{aa}p_{aa,0}^{(t)} + n_{aa}p_{aa,1}^{(t)} + n_{aA})\log\{\theta\} + (2n_{AA}p_{AA,0}^{(t)} + n_{AA}p_{AA,1}^{(t)} + n_{aA})\log\{1-\theta\} + n_{aA}\log\{2\}$$

For the second step, at a given iteration t, we want to compute :

$$\eta^{(t+1)} = \arg\max_{\eta} Q\left(\eta;\eta^{(t)}\right)$$

To do so, we are going to derivate Q with respect to p and θ.

$$\frac{\delta Q}{\delta p}(\eta,\eta^{(t)}) = \frac{n_{AA}p_{AA,1}^{(t)} + n_{aa}p_{aa,1}^{(t)}}{p} - \frac{n_{AA}p_{AA,O}^{(t)} + n_{aa}p_{aa,0}^{(t)} + n_{aA}}{1-p}$$

$$\frac{\delta Q}{\delta\theta}(\eta,\eta^{(t)}) = \frac{2n_{aa}p_{aa,0}^{(t)} + n_{aa}p_{aa,1}^{(t)} + n_{aA}}{\theta} - \frac{2n_{AA}p_{AA,0}^{(t)} + n_{AA}p_{AA,1}^{(t)} + n_{aA}}{1-\theta}$$

The derivate of Q with respect to p does not depend on θ and its derivate with respect to θ dos not depend on p. Consequently, we have :

$$\eta^{(t+1)} = \arg\max_{\eta} Q\left(\eta;\eta^{(t)}\right) = (p^{(t+1)},\theta^{(t+1)})$$

with

$$p^{(t+1)} = \arg\max_{p} Q\left(\eta;\eta^{(t)}\right)$$

$$\theta^{(t+1)} = \arg\max_{\theta} Q\left(\eta;\eta^{(t)}\right)$$

Moreover,

$$\frac{\delta^2 Q}{\delta p^2}(\eta,\eta^{(t)}) = -\frac{n_{AA}p_{AA,1}^{(t)} + n_{aa}p_{aa,1}^{(t)}}{p^2} - \frac{n_{AA}p_{AA,O}^{(t)} + n_{aa}p_{aa,0}^{(t)} + n_{aA}}{(1-p)^2} \leq 0$$

$$\frac{\delta^2 Q}{\delta\theta^2}(\eta,\eta^{(t)}) = -\frac{2n_{aa}p_{aa,0}^{(t)} + n_{aa}p_{aa,1}^{(t)} + n_{aA}}{\theta^2} - \frac{2n_{AA}p_{AA,0}^{(t)} + n_{AA}p_{AA,1}^{(t)} + n_{aA}}{(1-\theta)^2} \leq 0$$

3

The function Q is concave with respect to p and θ so $p^{(t+1)}$ and $\theta^{(t+1)}$ are such that :

$$\frac{\delta Q}{\delta p}(p^{(t+1)}, \theta, \eta^{(t)}) = 0$$

$$\frac{\delta Q}{\delta \theta}(p, \theta^{(t+1)}, \eta^{(t)}) = 0$$

On one hand we have :

$$\frac{\delta Q}{\delta p}(p^{(t+1)}, \theta, \eta^{(t)}) = 0 \iff n_{AA}p_{AA,1}^{(t)} + n_{aa}p_{aa,1}^{(t)} = p^{(t+1)}(n_{AA}p_{AA,0}^{(t)} + n_{aa}p_{aa,0}^{(t)} + n_{aA} + n_{AA}p_{AA,1}^{(t)} + n_{aa}p_{aa,1}^{(t)})$$

$$\iff p^{(t+1)} = \frac{n_{AA}p_{AA,1}^{(t)} + n_{aa}p_{aa,1}^{(t)}}{n_{AA}p_{AA,0}^{(t)} + n_{aa}p_{aa,0}^{(t)} + n_{aA} + n_{AA}p_{AA,1}^{(t)} + n_{aa}p_{aa,1}^{(t)}}$$

Since $p_{AA,0}^{(t)} + p_{AA,1}^{(t)} = p_{aa,0}^{(t)} + p_{aa,1}^{(t)} = 1$ and $n_{AA} + n_{aa} + n_{aA} = n$, we finally obtain :

$$p^{(t+1)} = \frac{n_{AA}p_{AA,1}^{(t)} + n_{aa}p_{aa,1}^{(t)}}{n}$$

On the other hand :

$$\frac{\delta Q}{\delta \theta}(p, \theta^{(t+1)}, \eta^{(t)}) = 0 \iff 2n_{aa}p_{aa,0}^{(t)} + n_{aa}p_{aa,1}^{(t)} + n_{aA} = \theta^{(t+1)}(2n_{AA}p_{AA,0}^{(t)} + n_{AA}p_{AA,1}^{(t)} + n_{aA}2n_{aa}p_{aa,0}^{(t)} + n_{aa}p_{aa,1}^{(t)} + n_{aA})$$

$$\iff \theta^{(t+1)} = \frac{2n_{aa}p_{aa,0}^{(t)} + n_{aa}p_{aa,1}^{(t)} + n_{aA}}{2n_{AA}p_{AA,0}^{(t)} + n_{AA}p_{AA,1}^{(t)} + n_{aA}2n_{aa}p_{aa,0}^{(t)} + n_{aa}p_{aa,1}^{(t)} + n_{aA}}$$

Since $p_{AA,0}^{(t)} + p_{AA,1}^{(t)} = p_{aa,0}^{(t)} + p_{aa,1}^{(t)} = 1$ and $n_{AA} + n_{aa} + n_{aA} = n$, we finally obtain :

$$\theta^{(t+1)} = \frac{(1 + p_{aa,0}^{(t)})n_{aa} + n_{aA}}{n + n_{aa}p_{aa,0}^{(t)} + n_{AA}p_{AA,0}^{(t)} + n_{aA}}$$

To conclude, the updates for estimating p and θ are given by :

$$p^{(t+1)} = \frac{n_{AA}p_{AA,1}^{(t)} + n_{aa}p_{aa,1}^{(t)}}{n}$$

$$\theta^{(t+1)} = \frac{(1 + p_{aa,0}^{(t)})n_{aa} + n_{aA}}{n + n_{aa}p_{aa,0}^{(t)} + n_{AA}p_{AA,0}^{(t)} + n_{aA}}$$

Question 3 :

We are now going to implement the corresponding EM algorithm, according to the updates for estimating p and θ, as found in the previous question.

We first start by computing the log-likelihood of our initial model. It will be useful for our EM algorithm since we want our algorithm to stop when it reaches an approximate of the maximum log-likelihood estimator. Indeed, our algorithm will stop when the difference between the value of the log-likelihood evaluated with the parameter computed at the $(t+1)$-th iteration and the value of the log-likelihood evaluated with the parameter computed at the (t)-th iteration is less than ϵ, for an ϵ chosen in advance.

The log-likelihood of our model is defined as follow :

$$\ell(p, \theta \mid x_1, \cdots, x_n) = n_{AA} \log \left\{ \mathbb{P}(X = AA) \right\} + n_{aa} \log \left\{ \mathbb{P}(X = aa) \right\} + n_{aA} \log \left\{ \mathbb{P}(X = aA) \right\}$$

We are now going to implement the EM algorithm and check graphically the convergence of our algorithm and the fit between the estimated model and the empirical distribution of the data.

To do so, we compute the vector of observations x_{obs}, according to the table given in the instructions and initialize our parameters.

For the initialization of our parameters, we choose $p^{(0)} = \frac{1}{2}, \theta^{(0)} = \frac{1}{2}$, $\epsilon = 10^{-16}$ and 10000 iterations.

We obtain the following graphs :
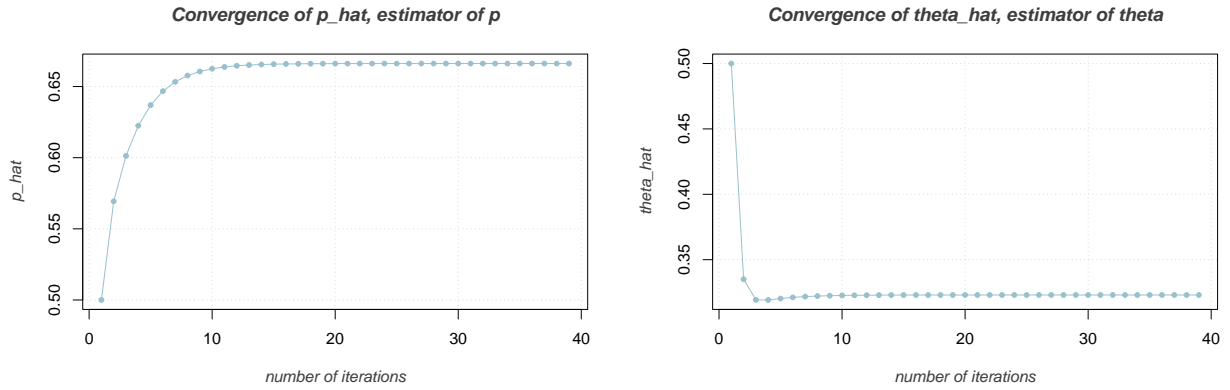
We can see that our estimators of the parameter p and θ converges respectively to 0.66 and 0.32. Our algorithm starts converging after approximately 10 iterations and stops after 40 iterations.
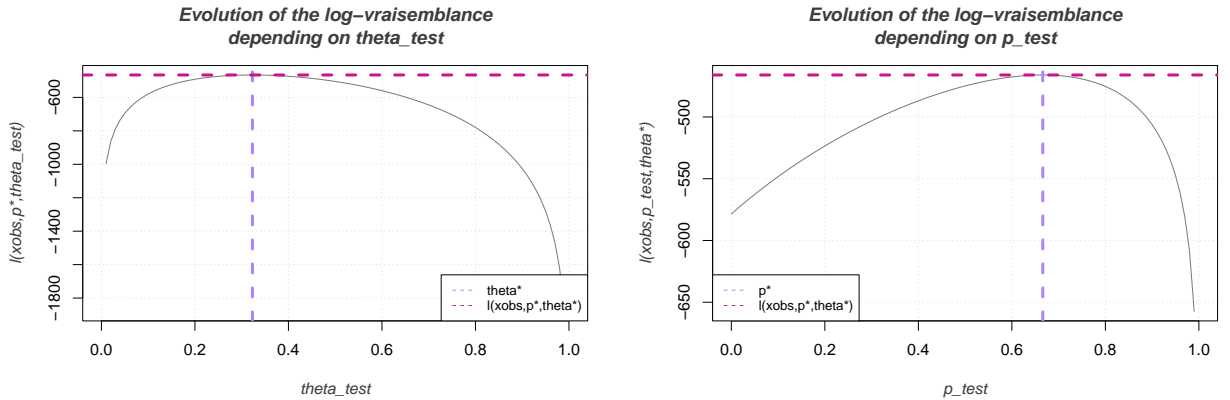
The estimators of p and θ found at the last iteration correspond to the maximum of the log-likelihood of our initial model.

Question 4 :

We would like to sample from the posterior distribution of (p, θ) using the Metropolis-Hastings algorithm with a gaussian proposal density. Therefore, we will implement the Random Walk Metropolis-Hastings algorithm.

At each iteration t, the p's and θ's candidates are sampled from the following gaussian distributions (σ_θ et σ_p are fixed) :

$$\theta_{new}^{(t)} \sim \mathcal{N}(\theta^{(t-1)}, \sigma_\theta)$$

$$p_{new}^{(t)} \sim \mathcal{N}(p^{(t-1)}, \sigma_p)$$

Accept-reject function : We have the following formula : $\pi(p, \theta|x_{obs}) = \pi(p, \theta)l(\theta, p|x_{obs})$, where $\pi(p, \theta)$ is the prior distribution of (p, θ) and $l(\theta, p|x_{obs})$ is the likelihood function of (p, θ).

$$l(\theta, p|x_{obs}) = (p(1-\theta) + (1-p)(1-\theta)^2))^{302}(2(1-p)\theta(1-\theta))^{73}(p\theta + (1-p)\theta^2)^{125}$$

At each iteration t, we want to compute the report :

$$\rho(p_{new}, \theta_{new}, p^{(t-1)}, \theta^{(t-1)}) := \frac{\pi(p_{new}, \theta_{new}|x_{obs})}{\pi(p^{(t-1)}, \theta^{(t-1)}|x_{obs})}$$

This can be simplified as follows :

$$\begin{aligned}
\rho(p_{new}, \theta_{new}, p^{(t-1)}, \theta^{(t-1)}) &= \frac{\pi(p_{new}, \theta_{new})l(p_{new}, \theta_{new}|x_{obs})}{\pi(p^{(t-1)}, \theta^{(t-1)})l(p^{(t-1)}, \theta^{(t-1)}|x_{obs})} \\
&= \frac{\pi(p_{new}, \theta_{new})}{\pi(p^{(t-1)}, \theta^{(t-1)})} \frac{l(p_{new}, \theta_{new}|x_{obs})}{l(p^{(t-1)}, \theta^{(t-1)}|x_{obs})}
\end{aligned}$$

The report between the prior is implemented as follows :

Beta prior : We assume that $\theta \sim Beta(\alpha_\theta, \beta_\theta)$ and $p \sim Beta(\alpha_p, \beta_p)$.

Then :

$$\begin{aligned}
\frac{\pi(p_{new}, \theta_{new})}{\pi(p_{(t-1)}, \theta_{(t-1)})} &= \frac{p_{new}^{\alpha_p-1}(1-p_{new})^{\beta_p-1}\theta_{new}^{\alpha_\theta-1}(1-\theta_{new})^{\beta_\theta-1}}{p_{(t-1)}^{\alpha_p-1}(1-p_{(t-1)})^{\beta_p-1}\theta_{(t-1)}^{\alpha_\theta-1}(1-\theta_{(t-1)})^{\beta_\theta-1}} \mathbb{1}_{[0,1]}(\theta_{new}, p_{new}, \theta_{(t-1)}, p_{(t-1)}) \\
&= (\frac{p_{new}}{p_{(t-1)}})^{\alpha_p-1}(\frac{1-p_{new}}{1-p_{(t-1)}})^{\beta_p-1}(\frac{\theta_{new}}{\theta_{(t-1)}})^{\alpha_\theta-1}(\frac{1-\theta_{new}}{1-\theta_{(t-1)}})^{\beta_\theta-1} \mathbb{1}_{[0,1]}(\theta_{new}, p_{new}, \theta_{(t-1)}, p_{(t-1)})
\end{aligned}$$

Uniform prior : We assume that $\theta \sim \mathcal{U}(0, 1)$ et $p \sim \mathcal{U}(0, 1)$.

Then :

$$\frac{\pi(p_{new}, \theta_{new})}{\pi(p_{(t-1)}, \theta_{(t-1)})} = \mathbb{1}_{[0,1]}(\theta_{new}, p_{new}, \theta_{(t-1)}, p_{(t-1)})$$

Following this, we can implement the accept-reject function.

Random Walk Metropolis-Hastings : Finally, we can implement the Random Walk Metropolis-Hastings algorithm and initialize our parameters.
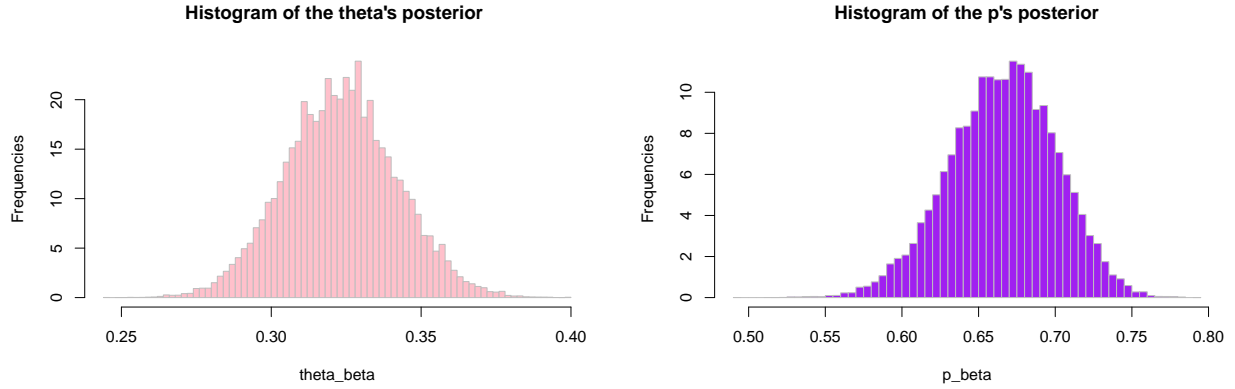
We will test our algorithm with the following parameters:

- Sample size : $n = 5 \times 10^5$

- Burn time : $nburn = 1000$

- Variance for the proposal law of theta : $\theta_{sd} = 0.2$

- Variance for the proposal law of p : $p_{sd} = 0.2$

- Initialization of the law of our parameters : $\theta_{init} \sim \mathcal{U}[0,1]$, $p_{init} \sim \mathcal{U}[0,1]$

For the uniform prior, we obtain the following plots :
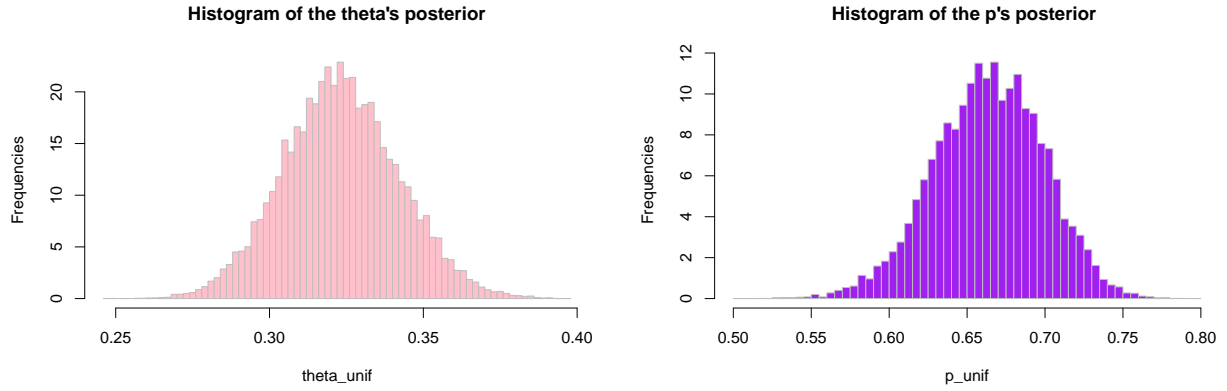
For the Beta prior, $(\alpha_p = 4, \beta_p = 2, \alpha_\theta = 2, \beta_\theta = 4)$, we obtain the following graphs :
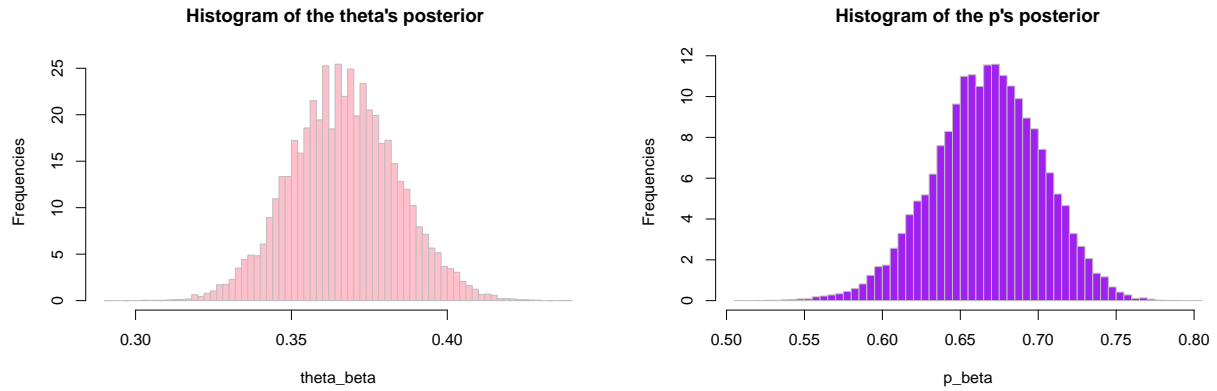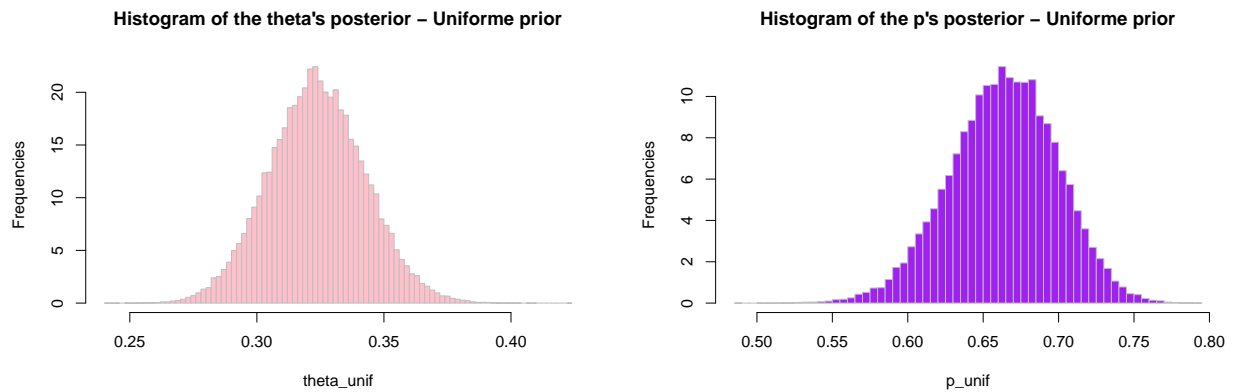

We can observe how the *posterior* distributions react as we change the parameters of the Beta *prior* :

For the Beta prior, $(\alpha_p = 100, \beta_p = 100, \alpha_\theta = 2, \beta_\theta = 3)$, we obtain the following graphs :

Histogram of the theta's posterior

Frequencies

theta_beta

Histogram of the p's posterior

Frequencies

p_beta
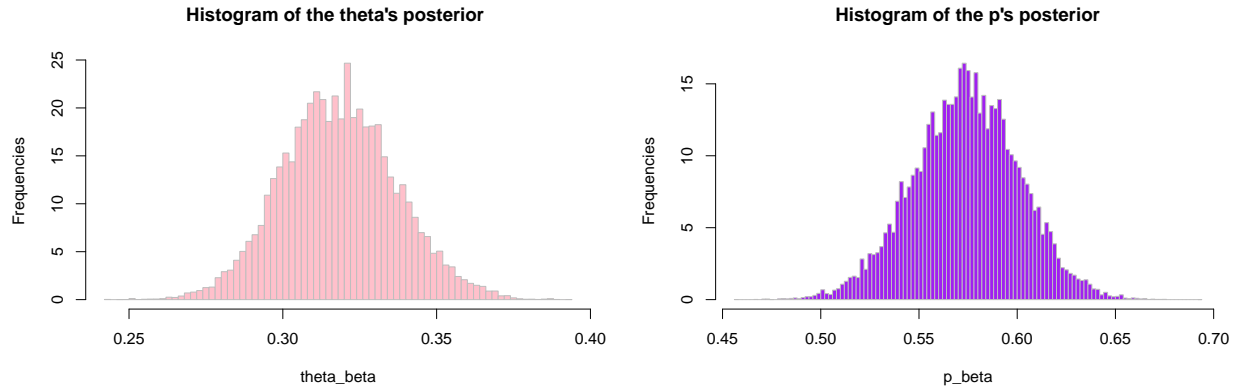
For the Beta prior, $(\alpha_p = 2, \beta_p = 3, \alpha_\theta = 100, \beta_\theta = 100)$, we obtain the following graphs :

Histogram of the theta's posterior

Frequencies

theta_beta

Histogram of the p's posterior

Frequencies

p_beta

When decreasing the parameters of the Beta laws, we obtain distributions similar to when we used a prior uniform, which makes sense given that $Beta(0,0)$ is a uniform on $[0,1]$. If we increase the parameters, we do not notice any significant differences except for a slight shift in the value around which the distributions are centred.
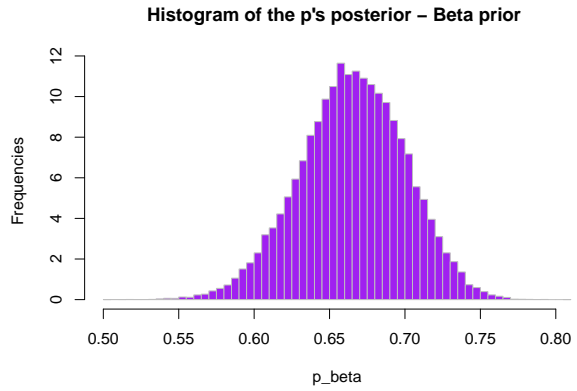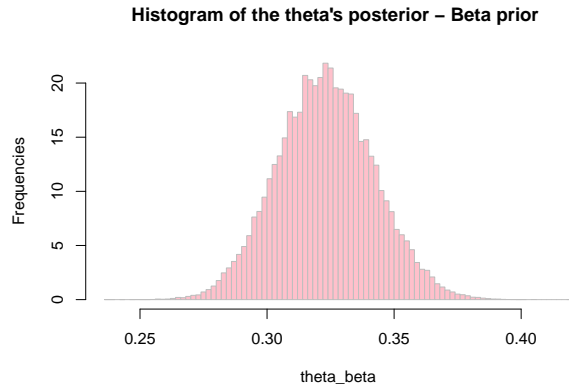
Question 5 :

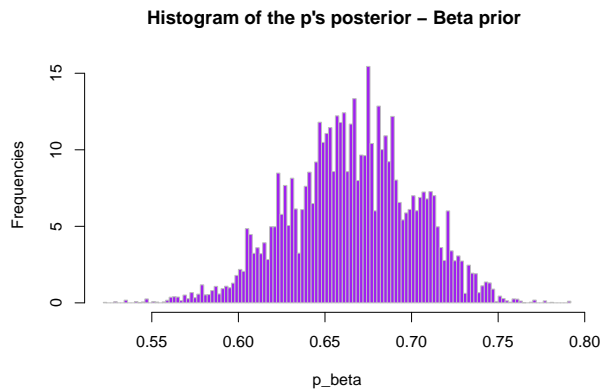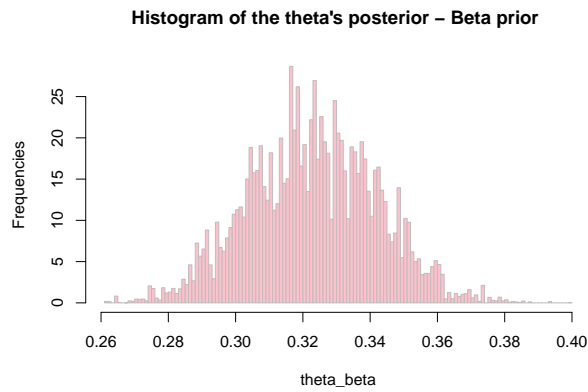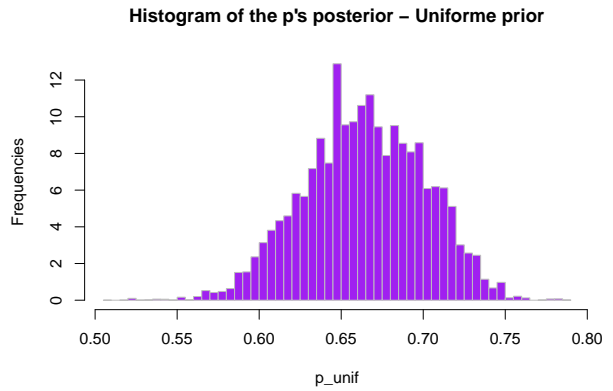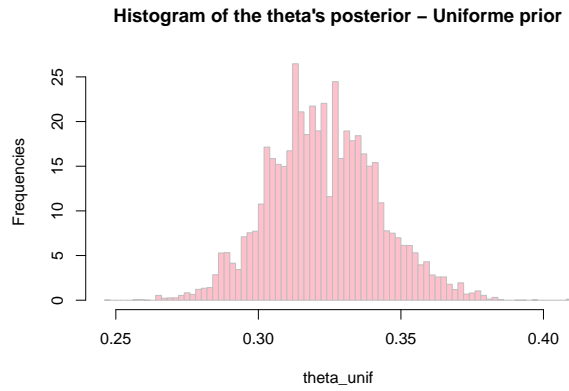Let's see how the method reacts when we change the variance of the proposal law :

For a variance equal to 0.1, we obtain the following graphs :

Histogram of the theta's posterior – Uniforme prior

Frequencies

theta_unif

Histogram of the p's posterior – Uniforme prior

Frequencies

p_unif

Histogram of the theta's posterior – Beta prior

theta_beta

Histogram of the p's posterior – Beta prior

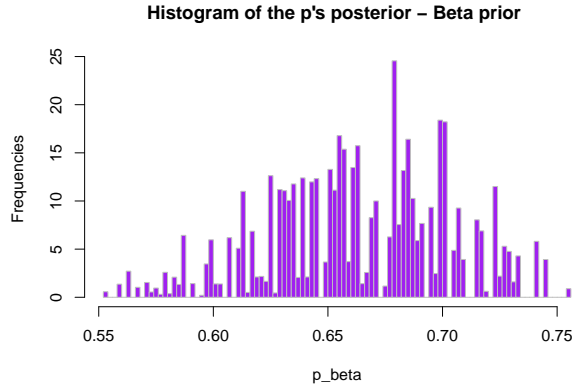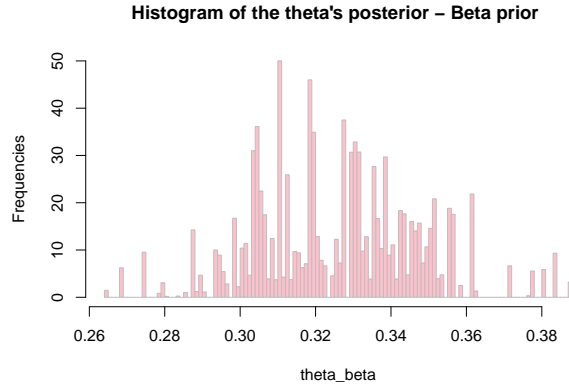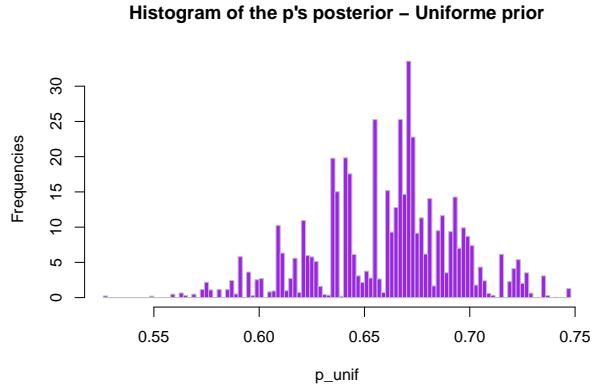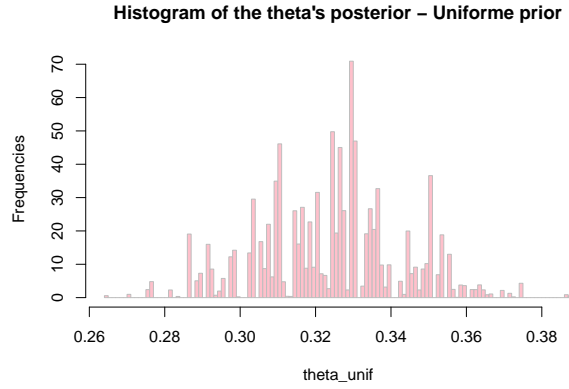
p_beta

For a variance equal to 0.5, we obtain the following graphs :

Histogram of the theta's posterior – Uniforme prior

theta_unif

Histogram of the p's posterior – Uniforme prior

p_unif

Histogram of the theta's posterior – Beta prior

theta_beta

Histogram of the p's posterior – Beta prior

p_beta

For a variance equal to 2, we obtain the following graphs :

Histogram of the theta's posterior – Uniforme prior

Histogram of the p's posterior – Uniforme prior

Histogram of the theta's posterior – Beta prior

Histogram of the p's posterior – Beta prior

We can see that increasing or decreasing the variance of the proposal law reciprocally increases or decreases the variance of the sample of the target law obtained with the algorithm.

Question 6 :

Using the latent representation from Question 1, we can compute the conditional distributions needed for the Gibbs sampler.

Conditional distribution of $z|(x_{obs}, p, \theta)$ **:** We have :

$$
\begin{aligned}
\mathbb{P}(Z_i = 1 | X_i = x_i, p, \theta) &= \frac{p(1-\theta)}{p(1-\theta) + (1-p)(1-\theta)^2} \quad \text{si} \quad x_i = AA \\
&= \frac{p\theta}{p\theta + (1-p)\theta^2} \quad \text{si} \quad x_i = aa \\
&= 0 \quad \text{si} \quad x_i = aA
\end{aligned}
$$

$$\mathbb{P}(Z_i = 0 | X_i = x_i, p, \theta) = \frac{(1-p)(1-\theta)^2}{p(1-\theta) + (1-p)(1-\theta)2} \quad \text{si} \quad x_i = AA$$

$$= \frac{(1-p)\theta^2}{p\theta + (1-p)\theta^2} \quad \text{si} \quad x_i = aa$$

$$= 1 \quad \text{si} \quad x_i = aA$$

Conditionnal distribution of $(p, \theta)|(z, x_{obs})$ **:** We know that :

$$\pi(p, \theta | z, x_{obs}) = \pi(p, \theta) l(p, \theta | z, x_{obs})$$

We can compute the likelihood, then by denoting :

- $N = \sum_{i=1}^{n} Z_i$
- $N_{aA} = \sum_{i=1}^{n} \mathbb{1}_{(x_{obs,i}=aA)}$
- $N_{aa}^1 = \sum_{i=1}^{n} \mathbb{1}_{(x_{obs,i}=aa, Z_i=1)}$
- $N_{AA}^1 = \sum_{i=1}^{n} \mathbb{1}_{(x_{obs,i}=AA, Z_i=1)}$
- $N_{aa}^0 = \sum_{i=1}^{n} \mathbb{1}_{(x_{obs,i}=aa, Z_i=0)}$
- $N_{AA}^0 = \sum_{i=1}^{n} \mathbb{1}_{(x_{obs,i}=AA, Z_i=0)}$

We have :

$$l(p, \theta | z, x_{obs}) = (p(1-\theta))^{N_{AA}^1}((1-p)(1-theta^2))^{N_{AA}^0}(p\theta)^{N_{aa}^1}((1-p)\theta^2)^{N_{aa}^0}(2(1-p)\theta(1-\theta))^{N_{aA}}$$

$$\propto p^N (1-p)^{n_{obs}-N} \theta^{N_{aa}^1 + 2N_{aa}^0 + N_{aA}} (1-\theta)^{N_{AA}^1 + 2N_{AA}^0 + N_{aA}}$$

Uniform prior : We suppose that $\theta \sim \mathcal{U}(0,1)$ et $p \sim \mathcal{U}(0,1)$.

$$\pi(p, \theta | z, x_{obs}) \propto p^N (1-p)^{n_{obs}-N} \theta^{N_{aa}^1 + 2N_{aa}^0 + N_{aA}} (1-\theta)^{N_{AA}^1 + 2N_{AA}^0 + N_{aA}} \mathbb{1}_{[0,1]}(p, \theta)$$

We recognize the product between a $Beta(N+1, n_{obs}-N+1)$ and a $Beta(N_{aa}^1 + 2N_{aa}^0 + N_{aA} + 1, N_{AA}^1 + 2N_{AA}^0 + N_{aA} + 1)$.

Beta prior : We suppose that $\theta \sim Beta(\alpha_\theta, \beta_\theta)$ et $p \sim Beta(\alpha_p, \beta_p)$.

We have :

$$\pi(p, \theta | z, x_{obs}) \propto p^{N-\alpha_p-1}(1-p)^{n_{obs}s-N+\beta_p-1} \theta^{N_{aa}^1 + 2N_{aa}^0 + N_{aA} + \alpha_\theta - 1}(1-\theta)^{N_{AA}^1 + 2N_{AA}^0 + N_{aA} + \beta_\theta - 1} \mathbb{1}_{[0,1]}(p, \theta)$$

We recognize the product between a $Beta(N+\alpha_p, n_{obs}-N+\beta_p)$ and a $Beta(N_{aa}^1 + 2N_{aa}^0 + N_{aA} + \alpha_\theta, N_{AA}^1 + 2N_{AA}^0 + N_{aA} + \beta_\theta)$.

Question 7 :

At each iteration t, we have the following updates :

$$z^{(t)} \sim \pi(z|p^{(t-1)}, \theta^{(t-1)}, x_{obs}) \quad (p^{(t)}, \theta^{(t)}) \sim \pi(p^{(t-1)}, \theta^{(t-1)}|z^{(t)}, x_{obs})$$

We can then compute the Gibbs Sampler and initialize our parameters.

- Sample size : $n = 1000$
- Burn time : $nburn = 300$
- Initialization of the law of our parameters : $\theta_{init} \sim \mathcal{U}[0,1]$, $p_{init} \sim \mathcal{U}[0,1]$
- Initialization of the vector z of size 500 : we sample each components of z from [0,1] according to the vector of probability $(p_{init}, 1 - p_{init})$

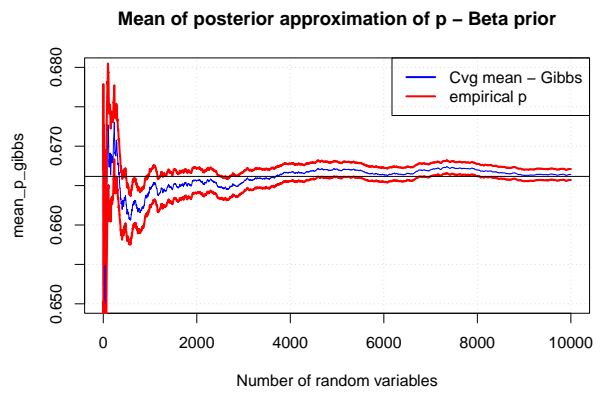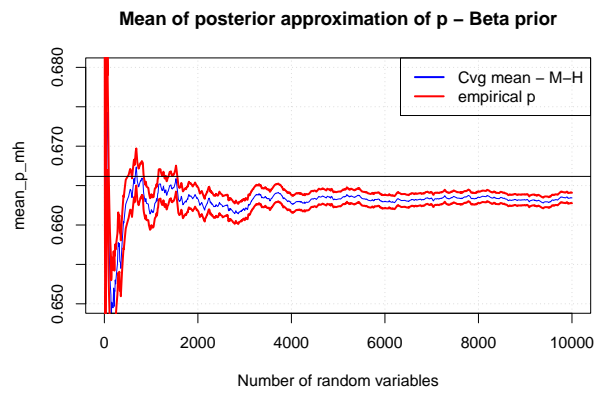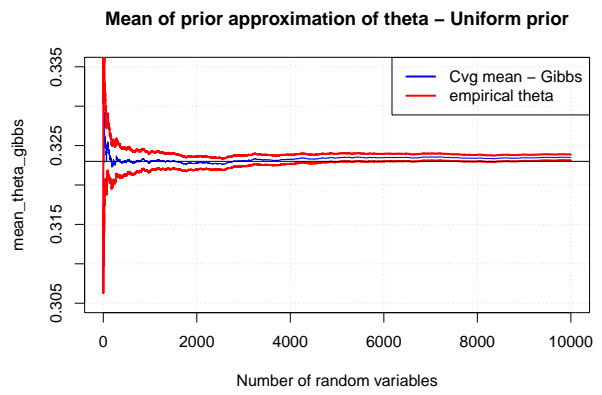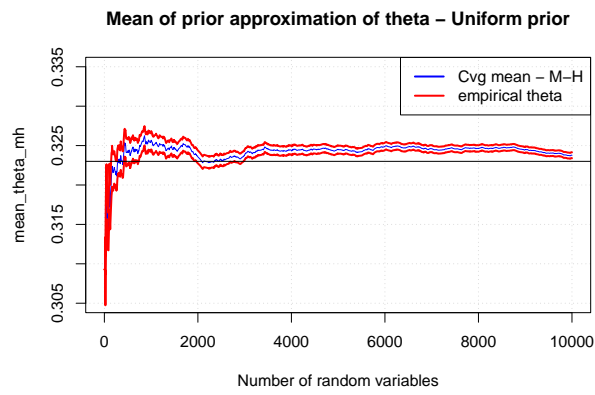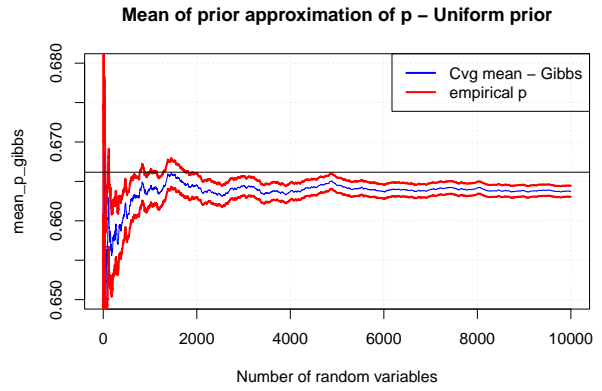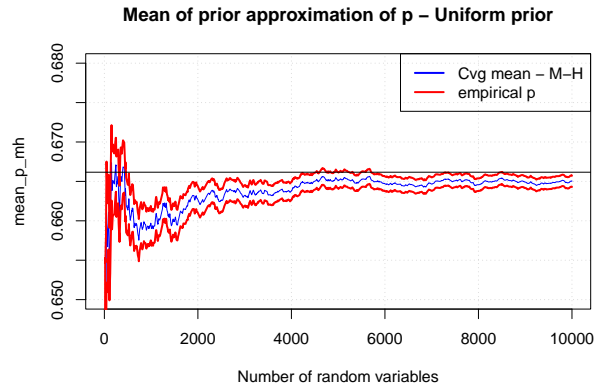We obtain the following graphs :
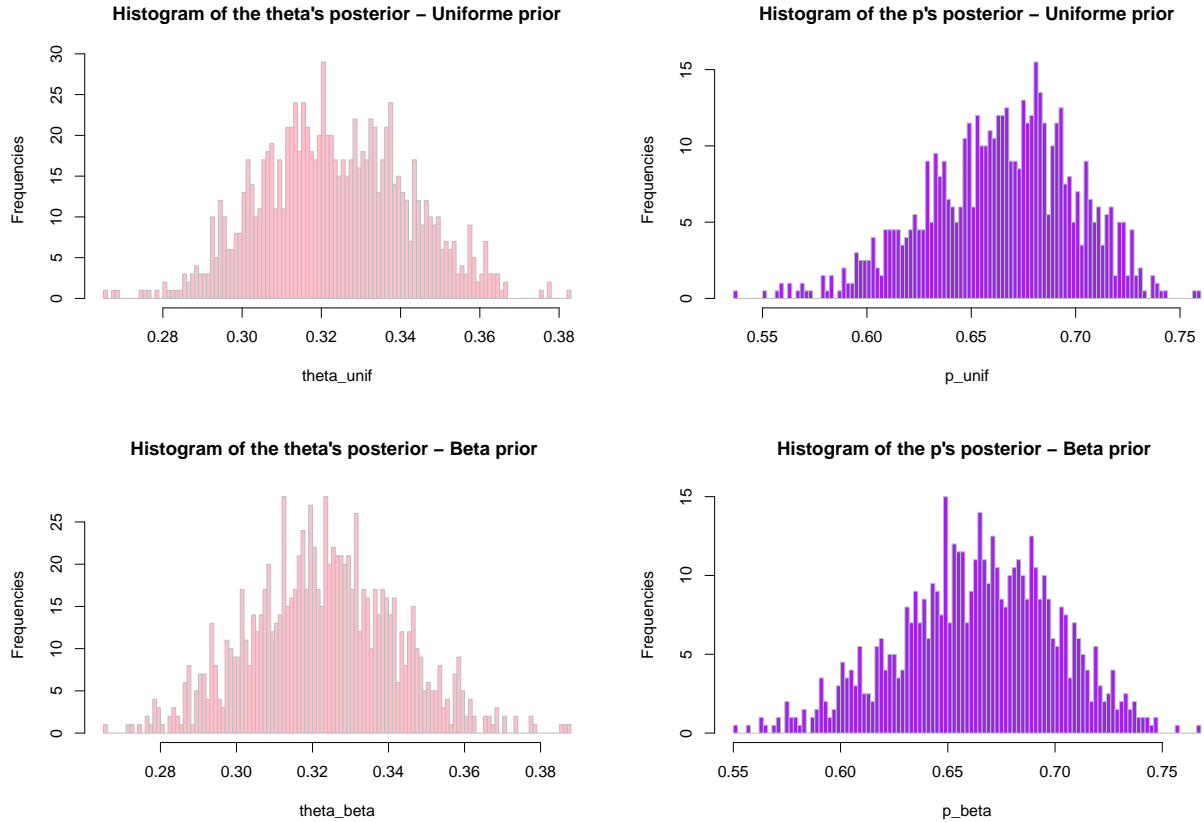
Question 8 :

The histograms of the samples obtained with Metropolis-Hastings and Gibbs seem to be distributed according to the same law.

We can also compare the evolution of the mean a posteriori :

Mean of prior approximation of p – Uniform prior

Mean of prior approximation of p – Uniform prior

Mean of prior approximation of theta – Uniform prior

Mean of prior approximation of theta – Uniform prior

Mean of posterior approximation of p – Beta prior

Mean of posterior approximation of p – Beta prior

Mean of posterior approximation of theta – Beta prior
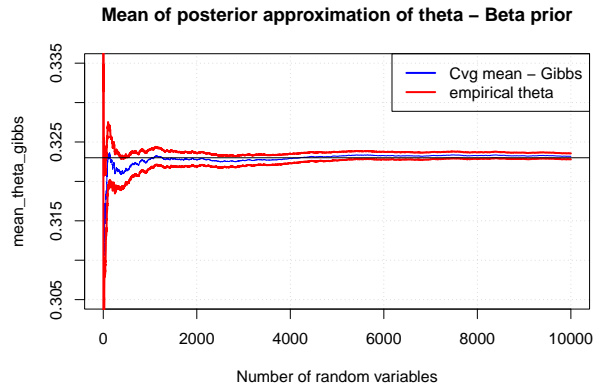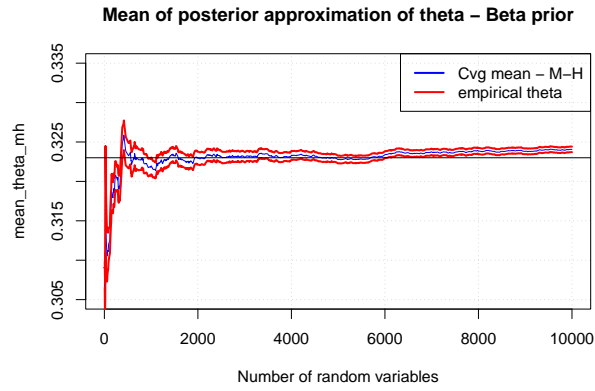
It can be seen by comparing the evolution of the a posteriori means for samples generated by Metropolis-Hastings and Gibbs that overall Gibbs obtains better results although the difference is not obvious.

Moreover, in our case, Gibbs is more advantageous because we could recognize the conditional distributions and sample according to them exactly.

Question 9 :

To implement an ABC algorithm on our problem, we need to establish a summary statistic. This will allow us to compare our observed sample to the samples we will generate based on the parameters obtained from the priors.

Our log-Likelihood is:

$$l(x|p,\theta) = ln(p_{AA}) \sum_{i=1}^{n} \mathbb{1}_{x_i=AA} + ln(p_{aa}) \sum_{i=1}^{n} \mathbb{1}_{x_i=aa} + ln(p_{aA}) \sum_{i=1}^{n} \mathbb{1}_{x_i=aA}$$

$$= ln(p_{AA})N_{AA} + ln(p_{aa})N_{aa} + ln(p_{aA})(n - N_{AA} - N_{aa})$$

Our likelihood is therefore entirely characterized by the statistics :

$$\begin{cases} N_{AA} & = & \sum_{i=1}^{n} \mathbb{1}_{x_i=AA} \\ N_{aa} & = & \sum_{i=1}^{n} \mathbb{1}_{x_i=aa} \end{cases}$$

An iteration of the ABC algorithm will have the following form:

Algorithm 2: ABC i-th iteration

input: π, $(X_k^{obs})_{1 \leq k \leq n}$, ϵ

1 $(p_i, \theta_i) \sim \pi$

2 $p_{AA,i} = p_i(1 - \theta_i) + (1 - p_i)(1 - \theta_i)^2$
3 $p_{aa,i} = p_i\theta_i + (1 - p_i)\theta_i^2$
4 $p_{aA,i} = 1 - p_{AA,i} - p_{aa,i}$

5 $(X_{i,k}^{new})_{1 \leq k \leq n} \sim \mathcal{L}(p_{AA,i}, p_{aa,i}, p_{aA,i})$

6 $N_i^{new} := (N_{AA,i}^{new}, N_{aa,i}^{new})$

7 if $d(N_i^{new}, N^{obs}) \leq \epsilon$ **then**
8 $\quad$ **return** (p_i, θ_i) ;
**9 else**
10 $\quad$ **return** $\emptyset$;

We have two possibilities:
_Decide in advance the number of iterations and our algorithm will return an approximation of a posterior sample with a size depending on the rejection rate
_Decide on the desired sample size and run our algorithm until we have accepted the desired number of random variables

The problems associated with this choice will be discussed later.

We will run our algorithm once with: $N = 10^5$ iterations, uniform prior, the distance L^2 and $\epsilon = 15$. The following distributions are obtained:

Simulation of p with ABC

Simulation of theta with ABC
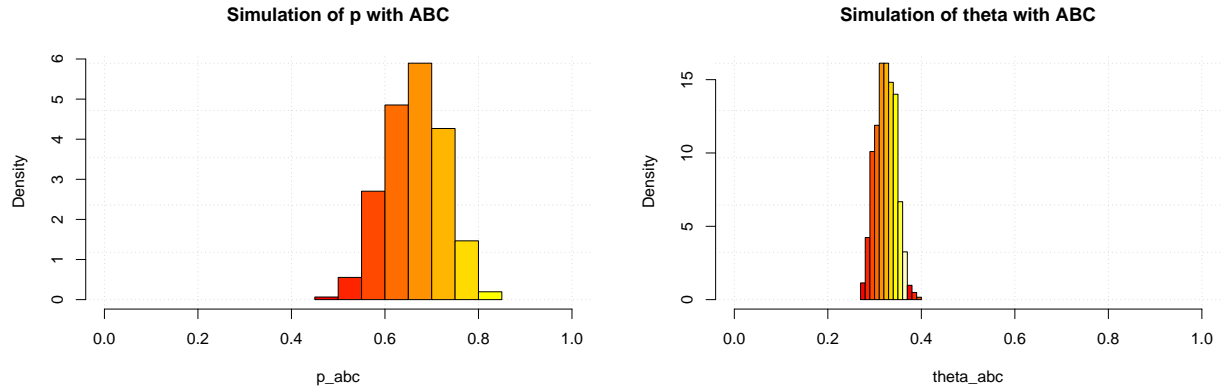
We will now study the impact of the size of the neighbourhood around which we will accept or not the parameters generated from the priors as an estimate of the posterior.


Acceptance rate in function of epsilon (N= 10000)
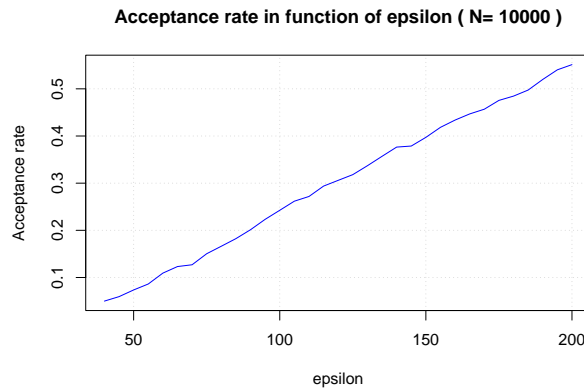
Naturally, the smaller the size of the neighborhood around x_{obs}, the more accurate our estimates will be. However, we must take into account the computational counterpart. Indeed, the more we reduce the size of the neighborhood, the more the probability that a prior is accepted is reduced (see graph above). Thus, there is a trade-off between the desired precision and the desired number of random variables.

From a more theoretical point of view, we understand that the more epsilon increases, the more we will tend to accept the simulation priors. So when ϵ tends to $+\infty$, we will accept all our priors, and our estimate of the posterior will tend towards the prior.

We can indeed observe this phenomenon on the graph below, where the means of our posteriors for the parameter p tend towards 0.5, the mean of a uniform [0,1], i.e., our prior.

On the other hand, if we reduce epsilon to 0, we will only accept THE distribution satisfying the zero distance condition, so we will have switched to a frequentist method, i.e., the method of moments.

posterior's mean in function of epsilon (N= 10000)
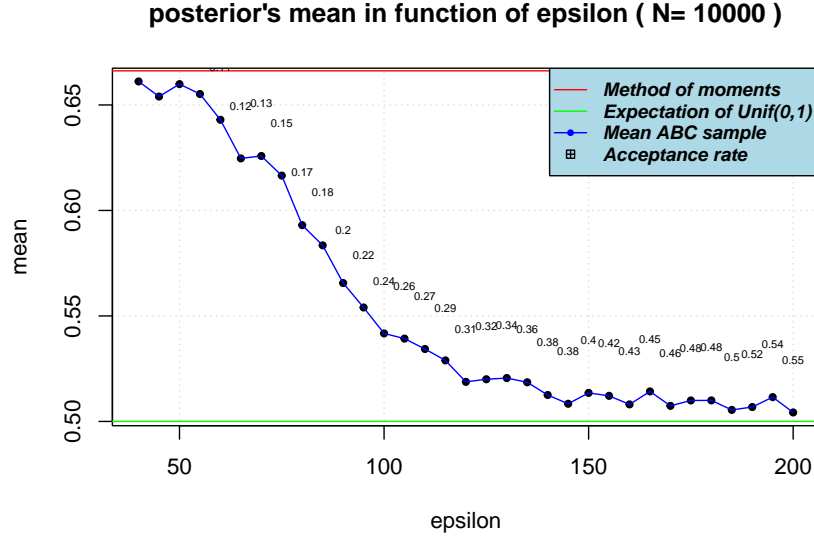
So far, we have used uniform laws [0,1] for the priors of p and theta. Now we will use $Beta(\alpha, \beta)$, and study the sensitivity of our algorithms to $\alpha > 0, \beta > 0$.

The ABC algorithm being a rejection algorithm which accepts priors only if they return distributions sufficiently close to our observed distribution, we will consider that the higher the frequency of acceptance of a prior, the more efficient this prior is.

We therefore naturally deduce that the most efficient prior, for the parameter p for example, will be the beta law centered around $\bar{p_n}$ (estimation of p by the method of moments) and with the lowest variance. In practice, the second point is not necessarily desired, because if we go so far as to nullify the variance, we will lose the Bayesian character of the method and we will fall back on a frequentist method.

To illustrate our previous assertion, we will represent the acceptance frequency of our ABC algorithms ($\epsilon = 30$) for different couples (α, β) such that $\mathbb{E}[Beta(\alpha, \beta)] = \dfrac{\alpha}{\alpha + \beta} = \bar{p_n}$.
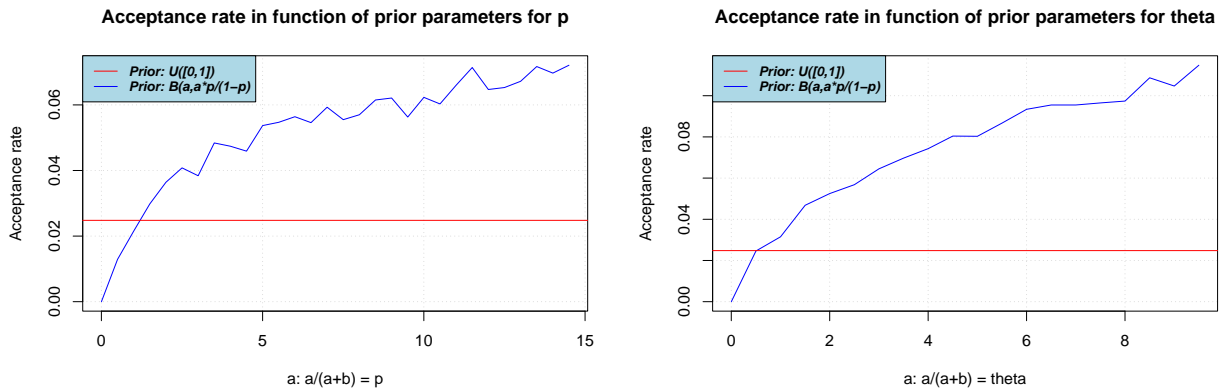
We can see that, for the parameter p that, past $\alpha = 0.5$, $\beta = \alpha \frac{\bar{p_n}}{1-\bar{p_n}}$, all the couples (α, β) are more efficient than the couple $(\alpha, \beta) = (1, 1)$ which corresponds to a $\mathcal{U}([0,1])$.

On the other hand, we observe a growth in terms of efficiency. Indeed, for $\beta := \alpha \frac{\bar{p_n}}{1-\bar{p_n}}$, the variance of our beta priors are decreasing in α, this corroborates our assertion: "The most efficient beta prior will be the one centred in $\bar{p_n}$ and with the lowest variance". The same tendency can be observed for theta.
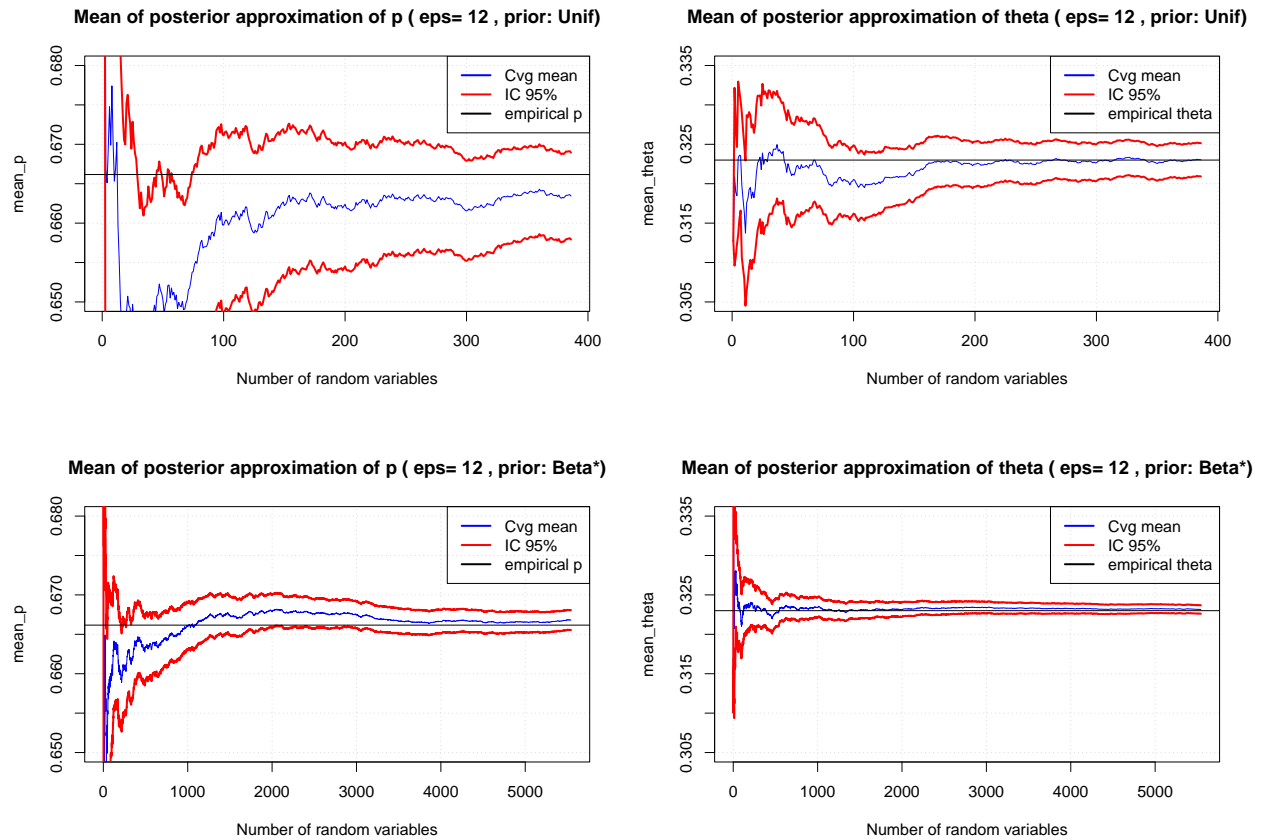
17

The distributions of the priors are relatively similar to those of our posterior estimates (observable on the first histogram of this section).

Prior of p maximizing acceptance frequency

Prior of theta maximizing acceptance frequency
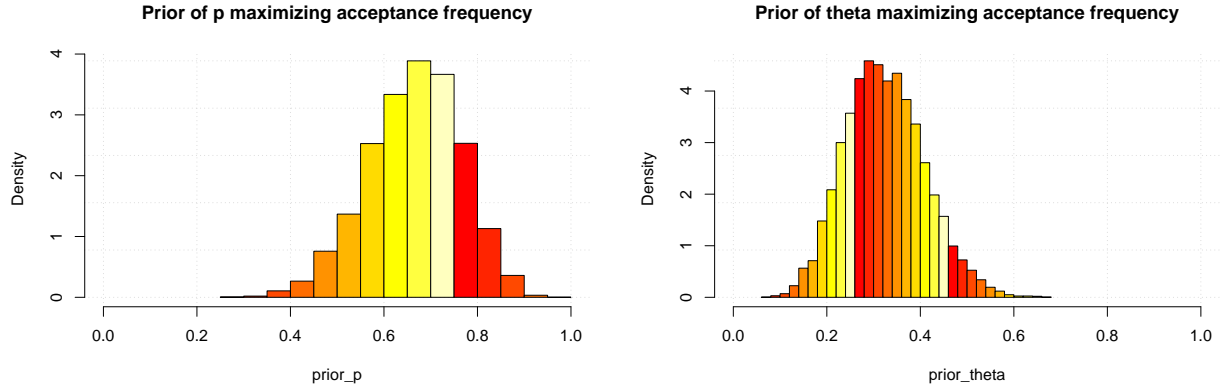
Question 10 :

Here, our distributions, contrary to the MCMC methods, do not depend solely on the choice of the prior. Indeed, it is the value of the parameter ϵ which will condition the variance of our law. It therefore seems delicate to compare the distributions, even using the same priors.

Regarding the posterior means, we observe a convergence towards the right values for p and theta. As for the amplitude of the confidence interval, it is also determined by the value of ϵ.

Mean of posterior approximation of p (eps= 12 , prior: Unif)

Mean of posterior approximation of theta (eps= 12 , prior: Unif)

Mean of posterior approximation of p (eps= 12 , prior: Beta*)

Mean of posterior approximation of theta (eps= 12 , prior: Beta*)

Question 11 :

The prior of the random variable representing the model will be a discrete uniform on {0,1}. As for the parameters associated with each model, they will be drawn from continuous uniforms on [0,1] in order not to favor any model.
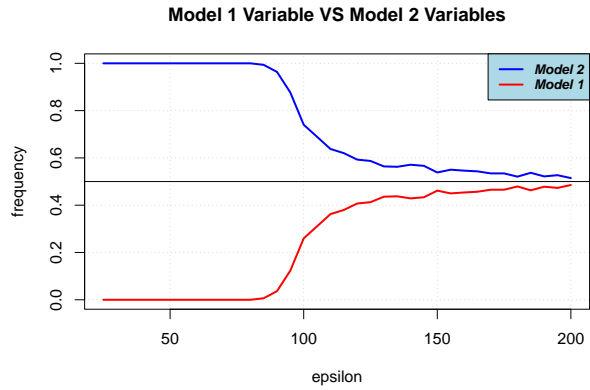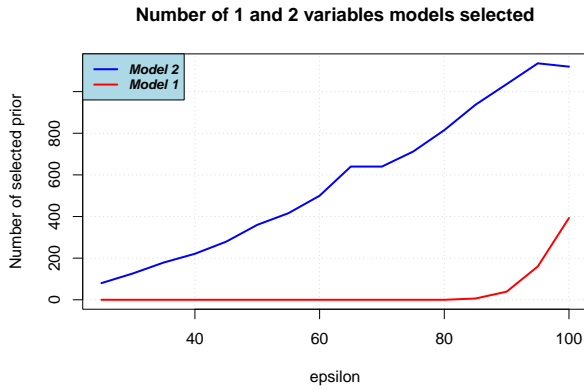

Algorithm 3: ABC model selection i-th iteration

$\quad$ **input:** $(X_k^{obs})_{1 \leq k \leq n}$, ϵ

1 $M_i \sim \mathcal{U}(\{1,2\})$

2 **if** $M_i = 1$ **then**
3 $\quad$ $\theta_i \sim \mathcal{U}([0,1])$

4 $\quad$ $p_{AA,i} = (1-\theta_i)^2$
5 $\quad$ $p_{aa,i} = \theta_i^2$
6 $\quad$ $p_{aA,i} = 1 - p_{AA,i} - p_{aa,i}$
7 **else**
8 $\quad$ $(p_i, \theta_i) \sim \mathcal{U}^2([0,1])$

9 $\quad$ $p_{AA,i} = p_i(1-\theta_i) + (1-p_i)(1-\theta_i)^2$
10 $\quad$ $p_{aa,i} = p_i\theta_i + (1-p_i)\theta_i^2$
11 $\quad$ $p_{aA,i} = 1 - p_{AA,i} - p_{aa,i}$

12 $(X_{i,k}^{new})_{1 \leq k \leq n} \sim \mathcal{L}(p_{AA,i}, p_{aa,i}, p_{aA,i})$

13 $N_i^{new} := (N_{AA,i}^{new}, N_{aa,i}^{new})$

14 **if** $d(N_i^{new}, N^{obs}) \leq \epsilon$ **then**
15 $\quad$ **return** M_i ;
16 **else**
17 $\quad$ **return** $\emptyset$;

We perform 10^4 iterations and vary ϵ:

For a precision of $\epsilon = 20$, it can be seen that no 1-variable model was selected, whereas about 100 2-variable models were selected. It was therefore reasonable to have considered this 2-variables model. It will be necessary to increase the tolerance to $\epsilon = 85$, in order to observe 1-variable models being selected.