

Méthodes de Monte Carlo-Projet

Hind DADOUN
R. R. Aaraona

8 Janvier 2018

1 Algorithme de rejet

Soit $U \sim U[0, 1]$ et M telle que $\tilde{f} < M \times g$ ou $f = c \times \tilde{f}$ avec c la constante de normalisation de f . La condition d'acceptation de la simulation d'une variable aléatoire dans l'algorithme de rejet s'écrit :

$$U < \frac{f(Y)}{c \times M \times g(Y)} \iff U < \frac{\tilde{f}(Y)}{M \times g(Y)} \quad (1)$$

L'exécution de l'algorithme ne nécessite donc pas la connaissance de c .

Calcul des constantes M_1 et M_2 :

Dans notre cas les constantes M_1 et M_2 doivent vérifier $\tilde{f}_1(X) < M_1 \times g(X)$ et $\tilde{f}_2(X) < M_2 \times g(X)$ ou g est une densité. On a alors :

$$\tilde{f}_1(x_1, x_2) = \exp(-\frac{1}{2}(\frac{x_1^2}{4} + x_2^2)) \mathbf{1}_{|x_2| < 1} \leq \exp(-\frac{1}{2}(\frac{x_1^2}{4} + x_2^2)) \frac{1}{4\pi} \times 4\pi$$

$$\begin{aligned} \tilde{f}_2(x_1, x_2) &= (\cos(x_1)^2 + 0.5(\sin(3x_2)^2) \cos(y_1)^4) \exp(-\frac{1}{2}(\frac{x_1^2}{4} + x_2^2)) \\ &\leq \frac{3}{2} \times \exp(-\frac{1}{2}(\frac{x_1^2}{4} + x_2^2)) \frac{1}{4\pi} \times 4\pi \end{aligned}$$

Donc $M_1 = 4\pi$ et $M_2 = 6\pi$

Afin d'optimiser notre algorithme on a fait le choix de calculer la constante de normalisation, en effet la probabilité d'acceptation est de $p = E[T] = \frac{1}{c \times M}$ avec $T = \inf(n > 0, U < \frac{f(Y)}{c \times M \times g(Y)})$, il faut donc en moyenne $\frac{n}{p}$ simulations pour obtenir n réalisations suivant la densité f_1 .

Calcul de la constante de normalisation de f_1 :

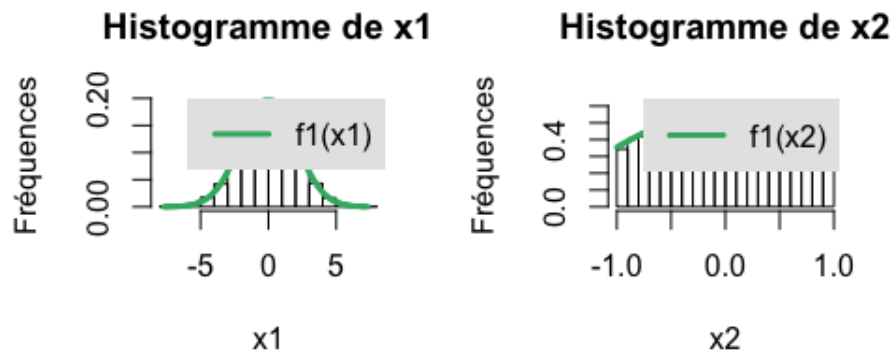
x_1 et x_2 étant indépendantes on sait que $\exp(-\frac{1}{2}(\frac{x_1^2}{4})) \frac{1}{4\pi}$ est une densité, on multiplie donc f_1 en haut et en bas par 4π ce qui nous laisse $\exp(-\frac{x_2^2}{2}) \times 4\pi \times \tilde{c}$ qui devrait s'intégrer à 1. Par un simple calcul on trouve alors $c = 4\pi \times \tilde{c} = 4\pi \times 2(\phi(1) - 1)$ ou ϕ est la fonction de répartition d'une $N(0, 1)$.

Calcul de la constante de normalisation de f_2 :

Ce calcul n'est pas évident, une solution serait d'utiliser l'algorithme de rejet avec une boucle for et une boucle while imbriquées ou encore par Monte Carlo classique en estimant :

$E_f[(\cos(x_1)^2 + 0.5(\sin(3x_2)^2) \cos(y_1)^4)]$ ou f est la densité d'une normale bivariée. On a choisit de le faire avec la deuxième méthode, on trouve alors $c \approx 0.5929545$

Comparaison avec les densités théoriques de f_1



Temps d'exécution de l'algorithme de rejet :

```
> system.time(rejet_f1(10000))
  user  system elapsed 
0.120   0.026   0.149 
> |
```

2 Retour sur les méthodes de réduction de variance

2.1 Les méthodes

2.1.1 Variable antithétique

Soit $X \sim f_1$, on a :

- 1- $X = -X$ en loi (par parité de la densité de f_1).
- 2- $A : x \mapsto -x$ est décroissante.
- 3- $h(x, y) = 1_{\exp(x)+\exp(y)>5}$ est croissante et donc de monotonie opposée à A .

On prend donc $Y = -X$ comme variable antithétique.

Comparaison avec l'estimateur de Monte Carlo classique :

L'estimateur par variable antithétique $\hat{\delta}_n$ nécessite $2n$ évaluations de la fonction h contre n évaluations pour \bar{Y}_n mais par la **Proposition 3.1** du cours on sait que l'estimateur $\hat{\delta}_n$ est plus efficace (en terme de réduction de variance) que l'estimateur \bar{Y}_n . Nous vérifierons cette remarque dans la section "Comparaison des variances".

2.1.2 Variable de contrôle

En suivant l'indication on fait le choix de prendre :

$$h(x_1, x_2) = 1_{\exp(x_1)+\exp(x_2)>5} \quad h_0(x_1, x_2) = 1_{\sqrt{\exp(x_1+x_2)}>K}$$

En effet, en utilisant la méthode d'importance sampling on peut réécrire l'espérance :

$$\mathbb{E}_{f_1}[1_{\exp(x_1+x_2)>5}] = \mathbb{E}_{N(0,1) \times N(0,2)}\left[\frac{1_{\exp(x_1+x_2)>5} \times 1_{x_2<1}}{2(\phi(1)-1)}\right]$$

Par la **Proposition 3.4** on sait que l'estimateur de variance minimale $\hat{\delta}_n(b^*)$ a une variance inférieure à celle de l'estimateur classique \bar{Y}_n . On a donc choisit d'inclure dans notre méthode le calcul du b^* optimal définit par :

$$b^* = \frac{Cov[h_0(x_1, x_2), h(x_1, x_2)]}{Var[h_0(x_1, x_2)]}$$

Ce calcul se fait dans une période de chauffe (burn-in period) à partir des m premiers éléments de notre échantillon. On utilise ensuite les termes restants pour calculer $\hat{\delta}_{n-m}(b^*)$ qui sera alors sans biais.

On introduit le coefficient de corrélation entre $h_0(x_1, x_2)$ et $h(x_1, x_2)$ définit par :

$$\rho(h_0(x_1, x_2), h(x_1, x_2)) = \frac{Cov[h_0(x_1, x_2), h(x_1, x_2)]}{\sqrt{Var[h_0(x_1, x_2)] \times Var[h(x_1, x_2)]}}$$

La méthode est d'autant plus efficace que $\rho(h_0(x_1, x_2), h(x_1, x_2))^2$ est proche de 1.

Il faut donc choisir un bon candidat pour K dans la formule de $h_0(x_1, x_2)$.

En évaluant l'algorithme pour différentes valeurs de K , on remarque que $\rho(h_0(x_1, x_2), h(x_1, x_2))^2$ semble maximal pour $K=2$.

Chercher à maximiser ρ rigoureusement est un problème de la meme difficulté que le calcul de $\mathbb{E}[h(x_1, x_2)]$. Pour $h_0(x_1, x_2) = h(x_1, x_2)$, on a $\rho(h_0(x_1, x_2), h(x_1, x_2))^2 = 1$. On va donc plutôt vérifier si pour $K=2$, $h_0(x_1, x_2) \approx h(x_1, x_2)$

$$\{(x_1, x_2) : \exp(x_1) + \exp(x_2) > 5\} \approx \{(x_1, x_2) : \exp((x_1 + x_2)/2) > 2\}$$

Cela se traduit par :

$$\exp(x_1) + \exp(x_2) > 5 \iff x_2 > \log(5 - \exp(x_1)) (:= f_1)$$

$$\exp((x_1 + x_2)/2) > K \iff x_2 > 2 \log(2 * \exp(-x_1/2)) (:= f_2)$$

En étudiant $f_1 - f_2$ on remarque que $f_1 - f_2 < 0.2$ sur $[0; 1.38]$ i.e pour x_1 positif proche de 0 et fixe, les bornes *inf* des ensembles :

$$\{x_2 : \exp((x_1 + x_2)/2) > K\}$$

$$\{x_2 : \exp(x_1) + \exp(x_2) > 5\}$$

sont proches au sens ou elles different d'au plus de 0.2 et par symétrie on a le meme résultat pour x_1 .

Remarque si on prend K tel que cette difference soit grande la valeur du coefficient p est alors plus petite. Nous donnons les chiffres suivant pour illustrer cette remarque, on note que ces chiffres sont obtenus car le coefficient de corrélation est une variable aléatoire bornée par 1 et qui admet donc un moment d'ordre deux fini, nous pouvons calculer un intervalle de confiance sur la moyenne des réalisations du coefficient p^2 , en utilisant les résultats de la loi forte des grands nombre et du théorème central limite. En itérant l'algorithme on obtient alors des réalisations de p^2 et en prenant leur moyenne empirique on obtient à un niveau de confiance 95% :

$$K = 2 \Rightarrow \mathbb{E}[\rho^2] = 0.426 \pm 0.002$$

$$K = 3.5 \Rightarrow \mathbb{E}[\rho^2] = 0.185 \pm 0.002$$

$$K = 5 \Rightarrow \mathbb{E}[\rho^2] \pm 0.002$$

On constate alors que le coefficient n'est pas proche de 1 comme voulu. On pourrait éventuellement prendre h_0 un développement limite de h pour avoir un coefficient de corrélation plus grand mais dans ce cas là l'estimation de la quantité $\mathbb{P}(h_0 > 5)$ qui est de meme difficulté que le problème initial. Au vu des résultats nous pensons que le choix de h_0 n'est pas adapté à la méthode, ce qui explique la faible baisse de la variance comparée à celle de l'estimateur de Monte Carlo classique que nous verrons plus tard dans la section 2.4.

2.1.3 Méthode de stratification

En utilisant la formule des probabilités totales on obtient :

$$\hat{p}_n = \frac{1}{L} \sum_{k=1}^L \mathbb{P}(X_1 \in D_k) \frac{1}{n_k} \sum_{j=1}^{n_k} h(X_j^k)$$

On prend $D_k = [\phi^{-1}(\frac{k}{L}); \phi^{-1}(\frac{k+1}{L})]$ avec $k \in [1; L]$

Comme nous avons choisit la méthode de stratification par allocation proportionnelle, on a $\mathbb{P}(X_1 \in D_k) = \frac{1}{L}$ et $n_k = \frac{1}{L}$. Le paramètre à déterminer est le nombre de strates L . On rappelle que l'estimateur stratifié élimine la variabilité inter-strates i.e la variance des moyennes conditionnelles et garde la variabilité intra-strates i.e la variance des moyennes conditionnelles. On a alors calculé la variance intra strate pour différentes valeurs de L et il nous semble que cette quantité soit croissante de L . En étudiant le modèle linéaire : $\log(\sigma^2) = c \log(L) + \epsilon$ (pour satisfaire les hypothèses du modèle on travaille avec le log des observations et des variables explicatives) avec $\epsilon \sim \mathcal{N}(0, \nu^2)$ On obtient :

```
Call:
lm(formula = log(res) ~ log(z))

Residuals:
    Min       1Q   Median       3Q      Max
-0.39740 -0.15782  0.01591  0.20104  0.27667

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -10.07850     0.30234  -33.33 7.16e-10 ***
log(z)        0.81916     0.05868   13.96 6.72e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2387 on 8 degrees of freedom
Multiple R-squared:  0.9606,    Adjusted R-squared:  0.9556
F-statistic: 194.9 on 1 and 8 DF,  p-value: 6.717e-07
```

On peut donc dire que le nombre de strate optimal en ce sens est $L=3$ qui est le nombre minimal de strate que l'on peut insérer dans l'algorithme. Ce résultat est cependant discutable car il introduit un biais dans l'estimation de la quantité d'intérêt ; en effet X_1 prend ces valeurs dans R et notre algorithme ne simule que des variables $X_1 | X_1 \in [a; b], a, b \in \mathbb{R}$. Ce biais disparaît bien évidemment quand L est grand car la partition est alors plus grande au sens de l'inclusion. On peut donc mettre en place une méthode de stratification mais elle n'est pas optimale au sens où on ne peut déterminer le L sans favoriser ou la variance ou le biais qui est dû à l'algorithme et non la méthode. On peut toujours cependant à une erreur fixée trouver le plus petit L et donc le L optimal pour la méthode.

Temps d'exécution des méthodes La fonction `system.time` qui prend en argument la méthode utilisée et le nombre de simulations permet d'avoir le temps d'exécution de chaque méthode. Ici on prend $n = 10000$.

```
> sys_time_va_contrôle
  user system elapsed
0.001  0.000  0.002
> sys_time_mc
  user system elapsed
0.002  0.000  0.002
> sys_time_antith
  user system elapsed
0.002  0.000  0.003
> sys_time_strat
  user system elapsed
0.119  0.005  0.126
```

Remarque : En terme de temps d'exécution la méthode de stratification est couteuse à cause de la boucle *for*.

2.2 Comparaison des variances

On utilise le résultat théoriques suivant pour construire les intervalles de confiance associées à un niveau de confiance de 95% pour chaque méthode :

$$\sqrt{n}(\bar{Y}_n - \mathbb{E}[\tilde{Y}]) \sim \mathcal{N}(0, \mathbb{V}(\tilde{Y}))$$

On estime la variance de la méthode d'estimation via la variance empirique associée aux réalisations de la variable aléatoire \tilde{Y} :

Monte Carlo classique : $\tilde{Y} = h(\mathbf{X})$

Variable antithétique : $\tilde{Y} = \frac{h(\mathbf{X}) + h(A(\mathbf{X}))}{2}$

Variable de controle : $\tilde{Y} = h(\mathbf{X})b(h_0(\mathbf{X}) - E[h_0(\mathbf{X})])$

Méthode de stratification :

L'intervalle est déduit de celui des variances intra strates car on a :

$$\sigma^2 = \sum_{k=1}^L p_k \sigma_k^2 \text{ ou } p_k = \frac{1}{L} \text{ et } \sigma_k^2 \text{ la variance intra-strate.}$$

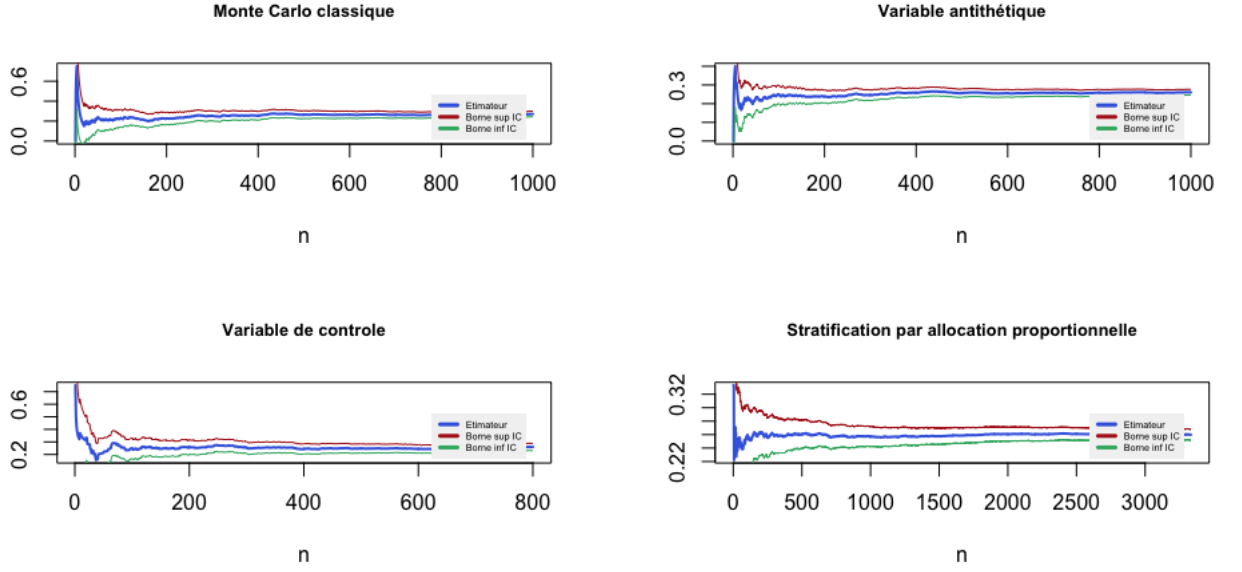
Résultats numériques :

Méthode de Stratification : $V(\tilde{Y}) = 0.060 + / - 0.003$

Variable antithétique : $V(\tilde{Y}) = 0.062 + / - 0.003$

Variable de contrôle : $V(\tilde{Y}) = 0.106 + / - 0.006$

Monte Carlo classique : $V(\tilde{Y}) = 0.185 + / - 0.004$



Remarque : Le calcul de la variance pour la méthode antithétique est faite avec $\frac{n}{2}$ échantillons.

2.3 Estimation d'une intégrale

On voudrait estimer la quantité suivante en utilisant la méthode de Monte Carlo classique ainsi que celle de la variable antithétique :

$$\varrho = \int_{\mathbb{R}^2} \cos(y_1 y_2) \sin(y_1) \exp(\sin(y_1 + y_2)) f_2(y_1, y_2) dy_1 dy_2$$

On cherche donc un estimateur de : $E[h(y_1, y_2)]$ avec $h(y_1, y_2) = \cos(y_1 y_2) \sin(y_1) \exp(\sin(y_1 + y_2))$ et $\mathbf{X} = (y_1, y_2) \sim f_2$.

Remarque : Pour la méthode de la variable antithétique on choisit de prendre $Y = -X$ comme variable antithétique car la densité est paire et donc ces deux variables sont égales en loi.

On obtient les résultats numérique suivant pour les variances à un niveau de confiance de 95% :

Monte Carlo classique : $\sigma^2 = 0.73 \pm 0.02$ **Variable antithétique :** $\sigma^2 = 0.040 \pm 0.001$

La méthode de la variable antithétique est plus efficace comme prévu. En effet cette réduction de la variance s'explique par le fait que :

$$\mathbb{C}(h(X), h(-X)) = \mathbb{E}[h(-X)h(X)] - \mathbb{E}[h(-X)]^2$$

Car $X = -X$ en loi et on a donc :

$$\mathbb{C}(h(X), h(-X)) = -\mathbb{E}[\cos(xy)^2 \sin(x)^2] - \mathbb{E}[h(-X)]^2 < 0$$

3 Recyclage de l'algorithme de rejet

3.1 Calcul préliminaires

3.1.1 Loi de Z conditionnellement a T

On a que la densité de Z est donnée par :

$$Z \sim \frac{Mg(Z) - f(Z)}{M - 1} \quad (2)$$

En effet en reprenant les notations du cours : posons

$$\tilde{T} = \inf(n > 0, U_n > \alpha(X_n))$$

Alors :

$$\mathbb{E}[h(Z_{\tilde{T}})] = \mathbb{E}\left[\sum_{i=1}^{\infty} h(Z_i) \mathbb{1}_{\tilde{T}=i}\right] = \sum_{i=1}^{\infty} \mathbb{E}[h(Z_i) \mathbb{1}_{\tilde{T}=i}] \mathbb{E}[\mathbb{1}_{\tilde{T}=i}]^{i-1} = \sum_{i=1}^{\infty} \int_{\mathbb{R}^d} h(z) g(z) \left(1 - \frac{f(z)}{Mg(z)}\right) dz \mathbb{E}[\mathbb{1}_{\tilde{T}=i}]^{i-1}$$

D'ou

$$\mathbb{E}[h(Z_{\tilde{T}})] = \int_{\mathbb{R}^d} h(z) \frac{Mg(z) - f(z)}{M - 1} dz$$

3.1.2 Biais des estimateurs

On travaille a T fixe i.e. $T=t$, on a :

$$\mathbb{E}[\delta_1] = \frac{1}{t} \sum_{i=1}^t \mathbb{E}[h(X_i)]$$

Or les X_i sont *iid* de loi f donc :

$$\mathbb{E}[\delta_1] = \mathbb{E}[h(X)]$$

Donc δ_1 est sans biais.

De la meme manière :

$$\mathbb{E}[\delta_2] = \frac{1}{N-t} \sum_{i=1}^{N-t} h(Z_i) \frac{(M-1)f(Z_i)}{Mg(Z_i) - f(Z_i)} = \mathbb{E}[h(X)]$$

3.1.3 Indépendance de δ_1 et δ_2

Comme X_i et Z_i sont issues d'un échantillon *iid* et que l'on travaille $\tilde{A} \quad T=t$ fixe on a que δ_1 et δ_2 sont indépendants.

$$\mathbb{V}(\alpha\delta_1 + (1-\alpha)\delta_2) = \alpha^2\mathbb{V}(\delta_1) + (1-\alpha)^2\mathbb{V}(\delta_2)$$

3.2 Détermination de alpha

On cherche alpha qui minimise le terme suivant :

$$Var(\alpha\delta_1 + (1 - \alpha)\delta_2) = \alpha^2 Var(\delta_1) + (1 - \alpha)^2 Var(\delta_2) \quad (3)$$

On dérivant on obtient que :

$$\alpha = Var(\delta_2) / (Var(\delta_2) - Var(\delta_1)) \quad (4)$$

On remarque alors que α est une fonction de T , cependant la formule qui nous permet de déterminer α utilise l'indépendance de $\alpha\delta_1$ et $(1 - \alpha)\delta_2$. Une idée serait d'estimer α sur une partie de l'échantillon (burn-in period) et effectuer l'estimation sur une autre partie de manière à ce qu'on puisse conserver 3 et 4. Une autre solution serait aussi d'estimer alpha sur autre échantillon et donc pour une autre valeur de T . On remarque que comme $0 \leq \alpha \leq 1$ alpha admet un moment d'ordre 2, on peut appliquer le TCL et la LGN pour trouver l'espérance de α et un intervalle de confiance; on trouve :

$$\mathbb{E}[\alpha(T)] = 0.5602 \pm 0.0001$$

On constate que cette solution permet d'avoir un algorithme plus rapide que précédemment de plus en estimant la variance de $\alpha(T)$ pour 200 réalisations on trouve :

$$Var(\alpha T) = 3.57 \times \pm 10^{-7}$$

Où l'intervalle de confiance est donné par l'approximation $n\widehat{\sigma^2}/\sigma^2 \sim \chi^2(n - 1)$. Les valeurs de $\alpha(T)$ sont donc concentrées autour de la moyenne, et on peut remplacer des lors par l'espérance estimée. Cette dernière remarque justifie les deux méthodes précédentes si n est grand car des lors comme : $T \sim Bin(n, 1/M)$, on a :

$$\mathbb{E}[T] = n/M$$

Donc T est croissante en moyenne en n , donc si n grand on peut déterminer α par la première méthode. De même avec la seconde méthode comme $\alpha(T)$ varie peu d'un échantillon à un autre, on peut l'estimer sur un autre échantillon.

3.2.1 Choix de M et g

Comme $Z \sim \frac{Mg-f}{M-1}$ on remarque si $f = Mg$ n'est pas de mesure nulle, Z n'est pas définie sur un ensemble de mesure non nulle. Pour pallier à ce problème on prend donc M tel que : $f < Mg$
Or :

$$\tilde{f}_1(x_1, x_2) \leq 4\pi g$$

$$\tilde{f}_3(x_1, x_2) \leq 6\pi * g$$

Avec g la densité d'une $X \sim \mathcal{N}_2(\beta, \Sigma^2)$ avec $\beta = (0, 0)$ et $\Sigma^2 = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$ Alors on prend $M = 2(4\pi + c4\pi)$ avec $c = 0$ ou $1/2$ selon qu'on considère \tilde{f}_1 ou \tilde{f}_3

3.3 Variance des méthodes

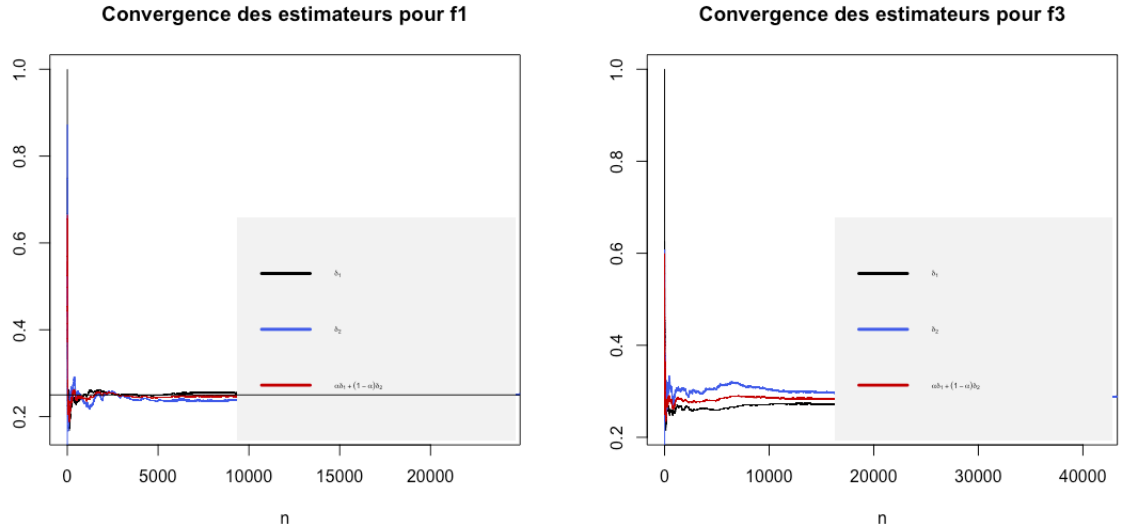
On prend $t = \mathbb{E}[(T)]$ alors on a les variances associées aux méthodes suivantes :

$$\sigma^2(\delta_1) = 0.180 \pm 0.01$$

$$\sigma^2(\delta_2) = 0.24 \pm 0.03$$

$$\sigma^2(\delta_3) = 0.075 \pm 0.008$$

On constate donc que la variance de δ_3 est meilleure que celle de l'estimateur de Monte Carlo classique de plus elle nous indique également qu'une manière de choisir M est de le prendre de manière à ce qu'il y est autant de variable acceptes que de rejettes ainsi la variance de δ_3 peut être minimisée.



4 Algorithme de Metropolis-Hasting

L'algorithme de Metropolis-Hasting permet de simuler une chaîne de Markov $(X_t)_{t \geq 0}$ (resp. $(Y_t)_{t \geq 0}$) qui a comme distribution limite f_1 (resp. f_2). En effet cet algorithme utilise la théorie des chaînes de Markov pour valider la convergence de la chaîne vers la distribution d'intérêt et la stabilisation des moyennes empiriques. Par le théorème ergodique on sait que :

$$\frac{1}{T} \sum_{k=1}^T X_k \longrightarrow f_1 \quad \frac{1}{T} \sum_{k=1}^T Y_k \longrightarrow f_2$$

Donc en calculant l'estimateur de Monte Carlo classique pour T assez grand, on remarque :

$$\frac{1}{T} \sum_{k=1}^T 1_{\exp(x_{1,k}) + \exp(x_{2,k}) > 5} \longrightarrow \rho \quad \frac{1}{T} \sum_{k=1}^T \cos(y_{1,k} y_{2,k}) \sin(y_{1,k}) \exp(\sin(y_{1,k} + y_{2,k})) \longrightarrow \varrho$$

