Université de Montpellier
HMMA310 - Statistique bayésienne

Année 2019-2020

Examen terminal 8 novembre 2019
Durée 2h00 - Sans document, sans matériel

Exercice 1 (4 pts) Soit f une densité de probabilité sur $[0,1]$. On suppose que f est continue et strictement décroissante sur $[0,1]$.

Proposer une méthode de simulation suivant f basée l'algorithme d'acceptation-rejet. Écrire la fonction `R` associée admettant pour entrée le nombre N de simulations et comme sorties les N réalisations indépendantes. On suppose que la densité f est présente dans `R` et qu'elle porte le même nom.

Exercice 2 (4 pts) Soit X une variable aléatoire telle que, $\forall x \in \mathbb{N}$,

$$\mathbb{P}(X = x|\theta) = \theta^x(1-\theta)$$

avec $\theta \in]0,1[$. On souhaite estimer θ et on se place dans le paradigme bayésien.

1) (2 pts) Donner la loi non informative de Jeffreys pour le paramètre θ.

2) (2 pts) Soit $(x_1, \ldots, x_n)$ la réalisation d'un n-échantillon de X. Donner la loi a posteriori de θ pour la loi a priori obtenue à la question précédente.

Exercice 3 (4 pts) On souhaite approcher

$$I_l = \mathbb{P}_{Exp(1)}\left(X \in [l, l+1]\right) .$$

1 (1 pt) Proposer un estimateur $\hat{I}_l^1$ de I_l basé sur une méthode de Monte-Carlo faisant intervenir la loi exponentielle.

2 (1 pt) Proposer un estimateur $\hat{I}_l^2$ de I_l basé sur une méthode de Monte-Carlo faisant intervenir la loi uniforme.

3 (2 pts) Comparer $\hat{I}_l^1$ et $\hat{I}_l^2$ suivant les valeurs de l.

Exercice 4 (4 pts) On souhaite estimer l'intégrale

$$I = \int_0^{2\pi} \sqrt{2 + \sqrt{|\sin(x)|}} \mathrm{d}x\,.$$

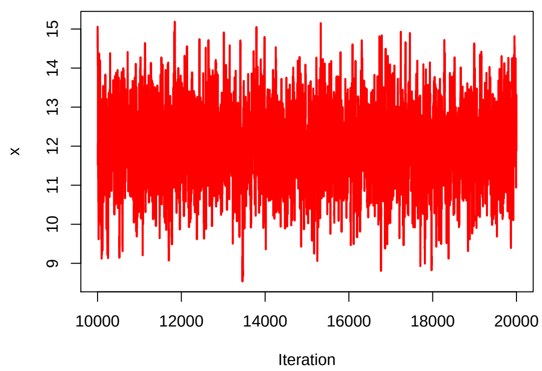
1 (1 pt) Proposer une méthode Monte-Carlo permettant d'estimer I.

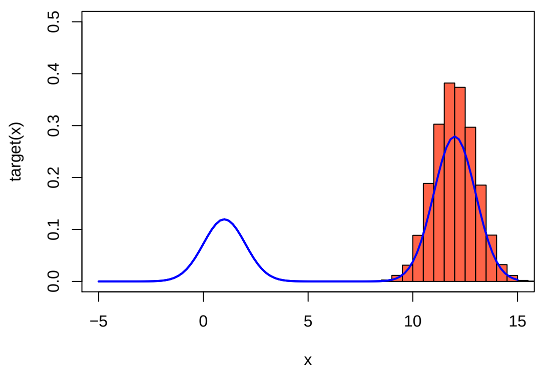
2 (3 pts) En utilisant l'inégalité de Bienaymé-Tchebychev, déterminer le nombre de simulations nécessaires N_0 pour que l'erreur relative associée à la méthode de Monte-Carlo proposée à la question précédente soit inférieure à 1% avec une probabilité supérieure ou égale à 99%.

Exercice 4 (4 pts) Décrire l'algorithme mis en oeuvre et les résultats obtenus à l'aide du code R ci-dessous.

```
Niter <- 100000
x <- rep(0,Niter)
x[1] <- 10
sigma2 <- 1
lulu <- 0
for (i in 2:Niter)
{
  xtilde <- rnorm(1,x[i-1],sqrt(sigma2))
  rho <- target(xtilde)/target(x[i-1])
  if (runif(1)<=rho) { x[i] <- xtilde ; lulu <- lulu+1 }
  else x[i] <- x[i-1]
}

plot(10^4:(2*10^4),x[10^4:(2*10^4)],xlab="Iteration",ylab="x",
type="l",lwd=2,col="red")
curve(target,-5,15,col="blue",lwd=2,ylim=c(0,0.5))
hist(x[-(1:500)],prob=TRUE,col="tomato",add=TRUE)
curve(target,-5,15,col="blue",lwd=2,ylim=c(0,0.5),add=TRUE)
```





(1)

Correction examen terminal
8/11/2019
#ΠΠΑ 310
Statistique bayésienne

Exercice 1

Nous allons utiliser la loi uniforme sur [0,1] comme loi instrumentale.
Comme f est strictement croissante sur [0,1] alors

$\forall x \in \mathbb{R}$

$f(x) \leq f(0)$

et $\forall x \in [0,1]$

$\frac{f(x)}{g(x)} \leq f(0)$ où $g(x)$ est la densité de la loi uniforme

Pour la loi instrumentale choisie la mise en oeuvre de l'algorithme acceptation-rejet consiste à

i] simuler $Y \sim \mathcal{U}_{[0,1]}$

ii] calculer $\Pi = \dfrac{f(Y)}{f(0)}$

iii] accepter Y si $\Pi \leq U$
où $U \sim \mathcal{U}_{[0,1]}$

la fonction R associée est

```
ar <- function(N) {
x <- rep(0,N)
for (i in 1:N) {bool <- T; while(bool){
y <- runif(1); Π <- f(y)/f(0)
if (Π <= runif(1)) {
bool <- F; x[i] <- y }}}}
```

Exercice 2

$$\mathbb{P}(X=x|\theta) = \theta^{x}(1-\theta) \;, \forall x \in \mathbb{N}$$

et $\theta \in]0,1[$

1) $\forall x \in \mathbb{N}$

$$\log(f(x|\theta)) = x\log(\theta) + \log(1-\theta)$$

$$\frac{\partial \log f}{\partial \theta}(x|\theta) = \frac{x}{\theta} - \frac{1}{1-\theta}$$

$$\frac{\partial^2 \log f}{(\partial \theta)^2}(x|\theta) = -\frac{x}{\theta^2} - \frac{1}{(1-\theta)^2}$$

$$I_X(\theta) = \frac{\mathbb{E}(x|\theta)}{\theta^2} + \frac{1}{(1-\theta)^2}$$

$$\mathbb{E}(x|\theta) = \sum_{x\in\mathbb{N}} x\theta^{x}(1-\theta)$$

$$\mathbb{E}(x|\theta) = (1-\theta)\sum_{x\in\mathbb{N}} x\theta^{x}$$

(4)

$$\mathbb{E}(x|\theta) = (1-\theta)\,\theta \sum_{x=1}^{\infty} x\,\theta^{x-1}$$

$$\mathbb{E}(x|\theta) = (1-\theta)\,\theta \left(\sum_{x=0}^{\infty} \theta^{x}\right)'$$

$$\mathbb{E}(x|\theta) = (1-\theta)\,\theta \left(\frac{1}{1-\theta}\right)'$$

$$\mathbb{E}(x|\theta) = \boxed{\frac{\theta}{1-\theta}}$$

Ainsi, $I_x(\theta) = \frac{1}{\theta(1-\theta)} + \frac{1}{(1-\theta)^2}$

$$I_x(\theta) = \frac{1-\theta+\theta}{\theta(1-\theta)^2} = \frac{1}{\theta(1-\theta)^2}$$

$$\pi^J(\theta) \propto \theta^{-1/2}(1-\theta)^{-1}\,\mathbb{I}_{]0,1[}(\theta)$$

loi impropre !

2] $\pi(\theta|x_1,\dots,x_n) \propto \left[\prod_{i=1}^{n} \theta^{x_i}(1-\theta)\right]\theta^{-1/2}(1-\theta)^{-1}$
$\forall \theta \in]0,1[$

$$\pi(\theta|x_1,\dots,x_n) \propto \theta^{\sum x_i - 1/2}(1-\theta)^{n-1}\,\mathbb{I}_{]0,1[}(\theta)$$

$$\Rightarrow \theta|x_1,\dots,x_n \sim \text{Beta}\left(\sum x_i + \tfrac{1}{2},\ n\right)$$

Exercice 3

$$I_\ell = \mathbb{P}(X \in [\ell, \ell+1]) \quad \ell \geq 0$$

avec $X \sim \mathcal{E}xp(1)$

1) $$\hat{I}^1_\ell = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{[\ell,\ell+1]}(X_i)$$

où $X_1, \dots, X_N \overset{iid}{\sim} \mathcal{E}xp(1)$

2) $$I_\ell = \mathbb{E}_{\mathcal{E}xp(1)}\left[\mathbb{1}_{[\ell,\ell+1]}(X)\right]$$

$$\iff I_\ell = \int_{-\infty}^{+\infty} \mathbb{1}_{[\ell,\ell+1]}(x)\, e^{-x}\, \mathbb{1}_{\{x \geq 0\}}\, dx$$

$$\iff I_\ell = \mathbb{E}_{U_{[\ell,\ell+1]}}\left(e^{-X}\right)$$

$$\hat{I}^2_\ell = \frac{1}{N} \sum_{i=1}^{N} e^{-X_i}$$

où $X_1, \dots, X_N \overset{iid}{\sim} U_{[\ell,\ell+1]}$

3] $\mathbb{E}(\hat{I}_l^1) = \mathbb{E}(\hat{I}_l^2) = I_l$

$$\mathbb{V}(\hat{I}_l^1) = \frac{1}{N^2}\sum_{i=1}^{N} \mathbb{V}_{\mathcal{E}(1)}\left(\mathbb{1}_{[l,l+1]}(X_i)\right)$$

$$\mathbb{V}(\hat{I}_l^1) = \frac{1}{N}\left(\mathbb{E}_{\mathcal{E}(1)}\left[\mathbb{1}_{[l,l+1]}(X)^2\right] - I_l^2\right)$$

$$\mathbb{V}(\hat{I}_l^1) = \frac{1}{N}\left(I_l - I_l^2\right)$$

$$\mathbb{V}(\hat{I}_l^2) = \frac{1}{N^2}\sum_{i=1}^{N} \mathbb{V}_{\mathcal{U}_{[l,l+1]}}\left(e^{-X_i}\right)$$

$$\mathbb{V}(\hat{I}_l^2) = \frac{1}{N}\left(\mathbb{E}_{\mathcal{U}_{[l,l+1]}}\left(e^{-2X}\right) - I_l^2\right)$$

Pour comparer $\hat{I}_l^1$ et $\hat{I}_l^2$, nous allons donc comparer

$\mathbb{E}_{\mathcal{U}_{[l,l+1]}}\left(e^{-2X}\right)$ et I_l

$$\mathbb{E}_{\mathcal{U}_{[l,l+1]}}\left(e^{-2X}\right) = \int_l^{l+1} e^{-2x}\,dx$$

$$I_\ell = \int_\ell^{\ell+1} e^{-x}\, dx$$

Comme $e^{-2x} < e^{-x}$

pour tout $x \in [\ell, \ell+1]$

avec $\ell > 0$

alors $\mathbb{E}_{\mathcal{U}_{[\ell,\ell+1]}}(e^{-2x}) < I_\ell$

et donc $\mathbb{V}(\hat{I}_\ell^2) < \mathbb{V}(\hat{I}_\ell^1)$

<u>Exercice 4</u>

$$I = \int_0^{2\pi} \sqrt{2 + \sqrt{|\sin(x)|}}\, dx$$

1] $I = (2\pi)\, \mathbb{E}_{\mathcal{U}_{[0,2\pi]}}\left(\sqrt{2 + \sqrt{|\sin(X)|}}\right)$

Ainsi

$$\hat{I} = (2\pi)\, \frac{1}{N} \sum_{i=1}^{N} \sqrt{2 + \sqrt{|\sin(X_i)|}}$$

avec $X_1, \dots, X_N \overset{iid}{\sim} \mathcal{U}_{[0,2\pi]}$

8

2) D'après l'inégalité de Bienaymé-Tchebyshev, nous avons

$$\forall \varepsilon > 0, \quad \mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathbb{V}(X)}{\varepsilon^2}$$

Par ailleurs, l'erreur relative de $\widehat{I}$ est définie par

$$ER(\widehat{I}) = \frac{|I - \widehat{I}|}{|I|} = \frac{|\widehat{I} - I|}{|I|}$$

Par construction, $\mathbb{E}(\widehat{I}) = I$

$$ER(\widehat{I}) = \frac{|\widehat{I} - \mathbb{E}(\widehat{I})|}{|I|}$$

$$\mathbb{P}\left(\frac{|\widehat{I} - I|}{|I|} \leq \frac{1}{100} \right) = \mathbb{P}\left(|\widehat{I} - I| \leq \frac{|I|}{100} \right)$$

$$\geq 1 - \frac{\mathbb{V}(\widehat{I})}{I^2} (100)^2$$

$$\geq 1 - \frac{10^4}{I^2} \mathbb{V}(\widehat{I})$$

$$V(\hat{I}) = \frac{1}{N}\left[(2\pi)^2 \mathbb{E}_{U_{[0,2\pi]}}\left(2+\sqrt{|\sin(x)|}\right) - \hat{I}^2\right]$$

$$\mathbb{P}\left(ER(\hat{I}) \leq 10^{-2}\right) \geq \left[1 - 10^4 \frac{\mathbb{E}_{U_{[0,2\pi]}}\left(2+\sqrt{|\sin(x)|}\right)}{N \hat{I}^2}(2\pi)^2 + \frac{10^4}{N}\right]$$

Or nous avons,

$$2+\sqrt{|\sin(x)|} \leq 3$$

$$\Rightarrow \frac{10^4}{N} \mathbb{E}_{U_{[0,2\pi]}}\left(2+\sqrt{|\sin(x)|}\right) \leq \frac{3\times 10^4}{N}$$

$$\sqrt{2+\sqrt{|\sin(x)|}} \geq \sqrt{2}$$

$$\Rightarrow \hat{I} \geq \sqrt{2}$$

$$\Rightarrow \hat{I}^2 \geq 2$$

$$\Rightarrow \frac{1}{\hat{I}^2} \leq \frac{1}{2}$$

Ainsi,

$$\mathbb{P}\left(ER(\hat{I}) \leq 10^{-2}\right) \geq 1 + \frac{10^4}{N} - \frac{3}{2} \times \frac{10^4}{N}(2\pi)^2$$

$$\mathbb{P}\left(ER(\hat{I}) \leq 10^{-2}\right) \geq 1 - \frac{10^4}{N}\left(6\pi^2 - 1\right)$$

Nous cherchons la plus petite valeur de N telle que $1 - \frac{10^4}{N}(6\pi^2 - 1) \geq \frac{99}{100}$

$$\Leftrightarrow \frac{10^4}{N}(6\pi^2 - 1) \leq \frac{1}{100}$$

$$\Leftrightarrow N \geq 10^6 (6\pi^2 - 1)$$

et donc $N_0 =$

Exercice 5

C'est l'algorithme de Hastings-Metropolis qui est mis en oeuvre, les résultats obtenus ne sont pas satisfaisants dans la mesure où seul l'un des modes de la cible est visité.