Simulation stochastique et Méthode de Monte Carlo.

Pierre Del Moral – Caroline Hillairet – Benjamin Jourdain

Map 564 - PC 1 - 6 Janvier 2012

EXERCICE 1 - : vitesse de convergence de la moyenne empirique
Soit $(Z_j)_{j\geq 1}$ une suite de variables aléatoires i.i.d. suivant la loi de Pareto symétrique de paramètre $\alpha > 0$ de densité $\frac{\alpha}{2|z|^{\alpha+1}}1_{\{|z|\geq 1\}}$.

1. Lorsque $\alpha > 1$, que peut-on dire de la suite $\bar{Z}_n = \frac{1}{n}\sum_{j=1}^n Z_j$ lorsque $n \to +\infty$?
2. Dans le cas où $\alpha > 2$, montrer que $\sqrt{n}\bar{Z}_n$ converge en loi vers une limite à préciser.
3. Vérifier que la fonction caractéristique Φ commune aux Z_j est telle que
$$\Phi(u) - 1 = \alpha|u|^\alpha \int_{|u|}^{+\infty} \frac{\cos(t) - 1}{t^{\alpha+1}}dt$$
4. Dans le cas où $0 < \alpha < 2$, donner un équivalent de $\Phi(u) - 1$ pour u au voisinage de 0. En déduire que $n^{\frac{\alpha-1}{\alpha}}\bar{Z}_n$ converge en loi. À quelle vitesse la moyenne empirique $\bar{Z}_n$ converge t-elle vers 0 lorsque $1 < \alpha < 2$.
5. Dans le cas où $\alpha = 2$, montrer que $\Phi(u) - 1 - u^2 \ln|u| \sim_{u\to 0} Cu^2$. En déduire que $\sqrt{\frac{n}{\ln n}}\bar{Z}_n$ converge en loi vers une limite à préciser. On se place dans le cas $\alpha = 1$ et on pose $\zeta_j = |Z_j|$ et $\bar{\zeta}_n = \frac{1}{n}\sum_{j=1}^n \zeta_j$.
   (a) Vérifier que pour $n \geq 2$, $\mathbb{E}\left(\frac{1}{n\ln n}\sum_{j=1}^n \zeta_j 1_{\{\zeta_j\leq n\ln n\}}\right) = 1 + \frac{\ln\ln n}{\ln n}$.
   (b) Calculer $\mathbb{E}\left(\zeta_1^2 1_{\{\zeta_1\leq n\ln n\}}\right)$ et vérifier que Var $\left(\frac{1}{n\ln n}\sum_{j=1}^n \zeta_j 1_{\{\zeta_j\leq n\ln n\}}\right) \leq \frac{1}{\ln n}$. En déduire que $\frac{1}{n\ln n}\sum_{j=1}^n \zeta_j 1_{\{\zeta_j\leq n\ln n\}}$ converge dans L^2 vers 1.
   (c) Vérifier que $\mathbb{P}(\zeta_1 > n\ln n) = \frac{1}{n\ln n}$ et conclure que $\frac{\bar{\zeta}_n}{\ln n}$ converge en probabilité vers 1.

EXERCICE 2 -Approximation de quantiles par méthode de Monte Carlo.
Soit X une variable aléatoire réelle à densité continue strictement positive partout. Pour tout $0 < a < 1$ on note $\rho(a)$ son quantile d'ordre a : $F(\rho(a)) = a$, où F est la fonction de répartition de X.
On considère N réalisations indépendantes de la loi de X et on les range par ordre croissant :
$$X_1 \leq X_2 \leq \cdots \leq X_N.$$
On suppose que le nombre Na est non entier. On note μ la partie entière de Na et $\rho_N(a) := X_{\mu+1}$.

1. Pour tout x dans $\mathbb{R}$ et ϵ tendant vers 0 calculer $\mathbb{P}(\rho_N(a) \in [x, x+\epsilon])$. En déduire que la loi de $\rho_N(a)$ a pour densité
$$g_N(x) = C_N^\mu(N-\mu)(F(x))^\mu(1-F(x))^{N-\mu-1}F'(x).$$
2. On pose
$$Y := \frac{\sqrt{N}}{\sqrt{a(1-a)}}F'(\rho(a))(\rho_N(a) - \rho(a)).$$
   Montrer que la densité q_N de Y est
$$q_N(x) := \frac{\sqrt{a(1-a)}}{\sqrt{N}}\frac{1}{F'(\rho(a))}g_N(\rho(a) + \frac{\sqrt{a(1-a)}}{\sqrt{N}}\frac{x}{F'(\rho(a))}).$$

3. On rappelle :
$$\lim_{N\to+\infty}\frac{\sqrt{a(1-a)}}{\sqrt{N}}C_N^{\mu}a^{\mu}(1-a)^{N-\mu}\frac{N-\mu}{1-a}=\frac{1}{\sqrt{2\pi}}.$$
Montrer que, quand N tend vers l'infini, la loi de $\rho_N(a)$ est proche d'une gaussienne centrée en $\rho(a)$ et d'écart type
$$\frac{\sqrt{a(1-a)}}{\sqrt{N}}\frac{1}{F'(\rho(a))}.$$
Qu'en concluez vous pour l'approximation de quantiles par méthode de Monte-Carlo ?

EXERCICE 3 - : stratification

On veut calculer $\mathbb{E}(f(X))$ où X est un vecteur aléatoire à valeurs dans $\mathbb{R}^d$ et $f:\mathbb{R}^d\to\mathbb{R}$ est une fonction telle que $f(X)$ est de carré intégrable. On suppose qu'il existe une partition $D_1,\dots,D_I$ de $\mathbb{R}^d$ telle que pour tout $i\in\{1,\dots,I\}$, $p_i=\mathbb{P}(X\in D_i)$ est strictement positif et connu et on sait simuler des variables aléatoires $(X_k^i)_{k\geq1}$ suivant la loi conditionnelle de X sachant $X\in D_i$ avec les variables $(X_k^i)_{1\leq i\leq I,k\geq1}$ indépendantes.

1. Pour $(n_1,\dots,n_I)\in\mathbb{N}^{*I}$, on pose $Y_{n_1,\dots,n_I}=\sum_{i=1}^I\frac{p_i}{n_i}\sum_{k=1}^{n_i}f(X_k^i)$. Calculer $\mathbb{E}(Y_{n_1,\dots,n_I})$. Quel est le comportement de $Y_{n_1,\dots,n_I}$ lorsque $\min_{1\leq i\leq I}n_i\to+\infty$?
2. On note $\sigma_i^2=\text{Var }(f(X_1^i))$. Montrer que $\text{Var }(Y_{n_1,\dots,n_I})=\frac{1}{\sum_{i=1}^I n_i}v\left(\frac{n_1}{\sum_i n_i},\dots,\frac{n_I}{\sum_i n_i}\right)$ où $v(q_1,\dots,q_I)=\sum_{i=1}^I\frac{(p_i\sigma_i)^2}{q_i}$.
3. Montrer que $\forall q=(q_1,\dots,q_I)$ probabilité sur $\{1,\dots,I\}$, $v(q)\geq\left(\sum_{i=1}^I p_i\sigma_i\right)^2$ où le minorant est atteint pour $q^\star=\left(\frac{p_1\sigma_1}{\sum_{i=1}^I p_i\sigma_i},\dots,\frac{p_I\sigma_I}{\sum_{i=1}^I p_i\sigma_i}\right)$. Pourquoi la clé de répartition optimale $q^\star$ des tirages dans les strates est-elle en général inconnue ?
   En remarquant que $\sigma_i^2=\mathbb{E}(f^2(X)|X\in D_i)-(\mathbb{E}(f(X)|X\in D_i))^2$, vérifier que $v(p_1,\dots,p_I)\leq\text{Var }(f(X))$ et interpréter ce résultat.

On suppose désormais que $X\sim\mathcal{N}_d(0,I_d)$ et que $D_i=\{y_{i-1}\leq u.X\leq y_i\}$ où $-\infty=y_0<y_1<\dots<y_{I-1}<y_I=+\infty$ et $u\in\mathbb{R}^d$ est tel que $|u|=1$. On note $N(y)=\int_{-\infty}^y e^{-\frac{z^2}{2}}\frac{dz}{\sqrt{2\pi}}$.

1. Quelle est la loi de $u.X$? Exprimer p_i à l'aide de la fonction N.
2. Montrer que $u.X$ et $X-(u.X)u$ sont indépendants.
3. Soit U une variable aléatoire uniforme sur $[0,1]$ indépendante de X. Vérifier que $Z_i=N^{-1}[N(y_{i-1})(1-U)+N(y_i)U]$ suit la loi conditionnelle de $u.X$ sachant $X\in D_i$. Conclure que $X+(Z_i-u.X)u$ suit la loi conditionnelle de X sachant $X\in D_i$.

EXERCICE 4 - : Inégalité de Hoeffding et concentration

1. Sur $(\Omega,\mathcal{F},\mathbb{P})$, soit X une variable aléatoire réelle à valeurs dans $[a,b]$ avec $-\infty<a<b<+\infty$. Pour $t\in\mathbb{R}$, on note $\mathbb{P}_t$ la probabilité de densité $\frac{d\mathbb{P}_t}{d\mathbb{P}}=\frac{e^{tX}}{\mathbb{E}(e^{tX})}$ et $\mathbb{E}_t$ l'espérance correspondantes.
   (a) Que vaut $\sup_{x\in[a,b]}(x-\frac{a+b}{2})^2$?
   En déduire que $\forall t\in\mathbb{R},\ \mathbb{E}_t((X-\mathbb{E}_t(X))^2)\leq\frac{(b-a)^2}{4}$.
   (b) Soit $f(t)=\ln\left(\mathbb{E}(e^{t(X-\mathbb{E}(X))})\right)$. Calculer $f''(t)$ et en déduire que
   $$\forall t\in\mathbb{R},\ \mathbb{E}(e^{t(X-\mathbb{E}(X))})\leq e^{\frac{(b-a)^2t^2}{8}}.$$

(c) Dans le cas où $\mathbb{P}(X=1)=\mathbb{P}(X=-1)=\frac{1}{2}$, calculer $\lim_{t\to 0}\frac{1}{t^2}(\mathbb{E}(e^{tX})-1)$ et en déduire que la constante $\frac{1}{8}$ dans l'inégalité précédente est optimale.

2. On suppose que les $(X_i)_{1\le i\le n}$ sont des variables aléatoires indépendantes à valeurs respectives dans $[a_i,b_i]$ avec $-\infty<a_i<b_i<+\infty$. Soit $\varphi:\mathbb{R}\to\mathbb{R}$ une fonction lischitzienne de constante de Lipschitz C.

   (a) Montrer que

$$\forall r>0,\ \mathbb{P}\left(\sum_{i=1}^{n}(\varphi(X_i)-\mathbb{E}(\varphi(X_i)))\ge r\right)\le e^{-\frac{2r^2}{C^2\sum_{i=1}^n(b_i-a_i)^2}}.$$

   (b) Dans le cas où les X_i sont i.i.d. suivant la loi de X, conclure que

$$\forall \varepsilon>0,\ \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\varphi(X_i)-\mathbb{E}(\varphi(X))\right|\ge\varepsilon\right)\le 2e^{-\frac{2n\varepsilon^2}{C^2(b-a)^2}}.$$

   Quel est l'intérêt de cet intervalle de confiance par rapport à celui obtenu avec l'inégalité de Bienaimé-Chebychev ? et celui obtenu par le théorème de la limite centrale ?

3. On suppose maintenant que $(M_k)_{0\le k\le n}$ est une $\mathcal{F}_k$-martingale t.q. M_0 est constante égale à m et pour tout $k\in\{1,\dots,n\}$, l'accroissement M_k-M_{k-1} prend ses valeurs dans un intervalle borné $[a_k,b_k]$.

   (a) Soit $k\in\{1,\dots,n\}$ et $A\in\mathcal{F}_{k-1}$ tel que $\mathbb{P}(A)>0$. En étudiant $f(t)\overset{\text{def}}{=}\ln\left(\mathbb{E}(e^{t(M_k-M_{k-1})}1_A)\right)$ montrer que $\mathbb{E}(e^{t(M_k-M_{k-1})}1_A)\le e^{\frac{(b_k-a_k)^2t^2}{8}}\times\mathbb{P}(A)$ pour tout $t\in\mathbb{R}$. En déduire que $\mathbb{P}\left(\mathbb{E}(e^{t(M_k-M_{k-1})}|\mathcal{F}_{k-1})\le e^{\frac{(b_k-a_k)^2t^2}{8}}\right)=1$.

   (b) Montrer que $\mathbb{E}(e^{t(M_n-m)})\le e^{\frac{t^2\sum_{k=1}^n(b_k-a_k)^2}{8}}$. Comparer avec l'inégalité obtenue en appliquant directement le résultat de la question 1b à la variable aléatoire M_n-m.

   (c) En déduire que pour $r>0$, $\mathbb{P}(|M_n-m|\ge r)\le 2e^{-\frac{2r^2}{\sum_{k=1}^n(b_k-a_k)^2}}$