Petite Classe 7 - MAP 311 - Groupes 1 et 12

Josselin Garnier

12 juin 2017 - salle PC n^o 16

Exercice 1. (Modèle auto-régressif d'ordre 1 AR(1)). Soit $(Y_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes et de même loi $\mathcal{N}(m,\sigma^2)$. Soient $a\in\mathbb{R}$ et X_0 une variable aléatoire de loi $\mathcal{N}(m_0,\sigma_0^2)$ indépendante de $(Y_n)_{n\geq 1}$. Pour tout $n\geq 1$, on définit la suite récurrente aléatoire $(X_n)_{n\geq 1}$ par

$$X_n = aX_{n-1} + Y_n.$$

Il s'agit d'un cas particulier de *séries temporelles*, utilisées pour modéliser l'évolution passée d'une quantité pour en prévoir le comportement futur.

1. Montrer que $(X_0,\dots,X_n)$ est un vecteur gaussien.
2. Déterminer la loi de X_n et exprimer $\mathrm{Cov}(X_k,X_n)$ en fonction de $\mathrm{Var}(X_k)$ pour $0\leq k\leq n$. Trouver la valeur de c telle que X_0+cX_1 soit indépendant de X_0.
3. A quelle condition la suite (X_n) converge-t-elle en loi ? Quelle est alors la loi limite ? Quelle est la loi de X_n si X_0 a cette loi limite ?
4. Montrer que si $a\in]-1,1[$, le vecteur (X_n,X_{n+1}) converge en loi vers un vecteur gaussien dont on déterminera les paramètres.
5. Pour $a\in]-1,1[$, montrer que la moyenne empirique $\overline{X}_n=\frac{1}{n}\sum_{i=1}^n X_i$ converge en probabilité vers $\frac{m}{1-a}$.

Solution.

1. Toute combinaison linéaire de $(X_0,\dots,X_n)$ est une combinaison linéaire de $(X_0,Y_1,\dots,Y_n)$, donc elle suit une loi normale. Ceci montre que $(X_0,\dots,X_n)$ est un vecteur gaussien.
2. X_n est une gaussienne de moyenne n_n et de variance σ_n^2. On a les relations de récurrence pour $n\geq 1$:
$$m_n = am_{n-1}+m, \quad \sigma_n^2 = a^2\sigma_{n-1}^2+\sigma^2$$
On trouve donc :
$$m_n = \begin{cases} mn+m_0 & \text{si } a=1 \\ a^n\left(m_0-\frac{m}{1-a}\right)+\frac{m}{1-a} & \text{si } a\neq 1, \end{cases}$$
$$\sigma_n^2 = \begin{cases} \sigma^2 n+\sigma_0^2 & \text{si } |a|=1 \\ a^{2n}\left(\sigma_0^2-\frac{\sigma^2}{1-a^2}\right)+\frac{\sigma^2}{1-a^2} & \text{si } |a|\neq 1. \end{cases}$$
Soit $0\leq k$. On a $\mathrm{Cov}(X_k,X_k)=\sigma_k^2$. Pour tout $n>k$,
$$\mathrm{Cov}(X_n,X_k) = a\mathrm{Cov}(X_{n-1},X_k)+\mathrm{Cov}(Y_n,X_k) = a\mathrm{Cov}(X_{n-1},X_k)$$

donc
$$\mathrm{Cov}(X_n, X_k) = a^{n-k}\sigma_k^2$$
$X_0 + cX_1$ et X_0 sont indépendants ssi $\mathrm{Cov}(X_0 + cX_1, X_0) = 0$ ssi $\sigma_0^2(1 + ac) = 0$ ssi $c = -1/a$.

3. La suite (X_n) converge en loi si et seulement si sa fonction caractéristique converge en tout point vers une fonction continue en 0.

   Si $|a| < 1$ ou si $(m_0, \sigma_0^2) = (\frac{m}{1-a}, \frac{\sigma^2}{1-a^2})$, on voit que
$$\mathbb{E}(e^{itX_n}) \underset{n\to\infty}{\longrightarrow} e^{it\frac{m}{1-a} - \frac{\sigma^2}{1-a^2}\frac{t^2}{2}}.$$
   Ainsi,
$$X_n \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}\left(\frac{m}{1-a}, \frac{\sigma^2}{1-a^2}\right).$$
   Si $|a| = 1$, ou si $|a| > 1$ avec $\sigma_0^2 \neq \frac{\sigma^2}{1-a^2}$, alors $e^{-\sigma_n^2\frac{t^2}{2}} \to 0$ lorsque $t \neq 0$. Dans ce cas,
$$\mathbb{E}(e^{itX_n}) \underset{n\to\infty}{\longrightarrow} \mathbf{1}_{t=0},$$
   fonction qui n'est pas continue en 0. Donc X_n ne converge pas en loi.

   Finalement, le cas où $|a| > 1$ et $\sigma_0^2 = \frac{\sigma^2}{1-a^2}$ est plus délicat. Si $m_0 = \frac{m}{1-a}$ il y a convergence en loi comme expliqué plus haut. Si $m_0 > \frac{m}{1-a}$, alors $m_{2n} \to +\infty$. Supposons par l'absurde que X_n converge en loi vers X. L'ensemble des points de discontinuité de la fonction de répartition F_X de X est au plus dénombrable, il existe donc $A > 0$ suffisamment grand tel que $\mathbb{P}(X \geq A) < 1/4$ et tel que F_X soit continue en A. Comme m_{2n} tend vers l'infini, pour n suffisamment grand (de sorte que $m_{2n} > A$):
$$\mathbb{P}(X_{2n} \geq A) \geq \mathbb{P}(X_{2n} \geq m_{2n}) = \frac{1}{2},$$
   car X_{2n} est une gaussienne de moyenne m_{2n}. Or $\mathbb{P}(X_{2n} \geq A) \to \mathbb{P}(X \geq A) < 1/4$, absurde. Le cas où $m_0 < \frac{m}{1-a}$ se traite de manière similaire.

   Si $X_0 = \mathcal{N}\left(\frac{m}{1-a}, \frac{\sigma^2}{1-a}\right)$, on voit que X_n a la même loi que X_0 pour tout $n \geq 1$.

4. D'après la question précédente, comme X_n et Y_{n+1} sont indépendantes, (X_n, Y_{n+1}) converge en loi vers un vecteur gaussien (X, Y) avec $X = \mathcal{N}\left(\frac{m}{1-a}, \frac{\sigma^2}{1-a^2}\right)$, $Y = \mathcal{N}(m, \sigma^2)$ et X indépendant de Y. Comme $(X_n, X_{n+1}) = (X_n, aX_n + Y_{n+1})$, on en déduit que (X_n, X_{n+1}) converge en loi vers $(X, aX+Y)$ (par composition avec l'application continue $(x, y) \mapsto (x, ax + y)$). Tout d'abord, on a
$$\mathrm{Cov}(X, aX + Y) = \mathrm{Cov}(X, aX) = a\mathrm{Var}(X) = \frac{a\sigma^2}{1-a^2}.$$
   Le vecteur gaussien $(X, aX + Y)$ a donc pour moyenne $(\frac{m}{1-a}, \frac{am}{1-a} + m) = (\frac{m}{1-a}, \frac{m}{1-a})$ et pour matrice de covariance
$$\begin{pmatrix} \frac{\sigma^2}{1-a^2} & \frac{a\sigma^2}{1-a^2} \\ \frac{a\sigma^2}{1-a^2} & \frac{\sigma^2}{1-a^2} \end{pmatrix}$$

5. La variable aléatoire $\overline{X}_n$ suit une loi gaussienne car $(X_1, \ldots, X_n)$ est un vecteur gaussien. On va montrer que $\overline{X}_n$ converge dans L^2 vers $\frac{m}{1-a}$. Tout d'abord,

$$\mathbb{E}(\overline{X}_n) = \frac{\sum_{k=1}^n \mathbb{E}(X_k)}{n} \underset{n\to\infty}{\longrightarrow} \frac{m}{1-a}.$$

Pour calculer la variance de $\overline{X}_n$, on peut remarquer que $X_k = a^k X_0 + \sum_{i=0}^k a^{k-i} Y_i$ pour $k \geq 0$. Ainsi,

$$X_1 + \cdots + X_n = \lambda_0 X_0 + \lambda_1 Y_1 + \cdots + \lambda_n Y_n$$

avec

$$\lambda_0 = a + a^2 + \cdots + a^n \leq \frac{1}{1-a} \qquad \text{et } \lambda_i = 1 + a + \cdots + a^{n-i} \leq \frac{1}{1-a} \text{ pour } 1 \leq i \leq n.$$

Ainsi,

$$\mathrm{Var}(X_1 + \cdots + X_n) = \lambda_0^2 \sigma_0^2 + \sum_{i=1}^n \lambda_i^2 \sigma^2 \leq Cn.$$

pour une certaine constante $C > 0$ qui ne dépend pas de n. Donc

$$\mathrm{Var}(\overline{X}_n) = \frac{\mathrm{Var}(X_1 + \cdots + X_n)}{n^2} \underset{n\to\infty}{\longrightarrow} 0.$$

D'après l'inégalité de Markov, pour tout $\epsilon > 0$

$$\begin{aligned} \mathbb{P}\Big(\Big|\overline{X}_n - \frac{m}{1-a}\Big| > \epsilon\Big) \leq \frac{1}{\epsilon^2}\mathbb{E}\Big(\Big(\overline{X}_n - \frac{m}{1-a}\Big)^2\Big) &= \frac{1}{\epsilon^2}\mathbb{E}\Big(\Big(\overline{X}_n - \mathbb{E}(\overline{X}_n) + \mathbb{E}(\overline{X}_n) - \frac{m}{1-a}\Big)^2\Big) \\ &= \frac{1}{\epsilon^2}\Big[\mathrm{Var}(\overline{X}_n) + \Big(\mathbb{E}(\overline{X}_n) - \frac{m}{1-a}\Big)^2\Big], \end{aligned}$$

qui tend vers 0.
Donc $\overline{X}_n$ converge en probabilité vers $\frac{m}{1-a}$.

Exercice 2. (Estimateurs linéaires) Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d. de carré intégrable. Trouver l'estimateur $\widehat{\theta}_n$ de la moyenne $\theta = \mathbb{E}(X_1)$ qui soit sans biais (c'est-à-dire $\mathbb{E}(\widehat{\theta}_n) = \theta$) et de variance minimale dans la classe des estimateurs linéaires $\widehat{\theta}_n = \sum_{k=1}^n a_k X_k$.

Solution. Comme $\widehat{\theta}_n$ est sans biais, on a $a_1 + \cdots + a_n = 1$. Comme les variables aléatoires sont indépendantes, on a $\mathrm{Var}(\widehat{\theta}_n) = \sum_{i=1}^n a_i^2 \sigma^2$ avec $\sigma^2 = \mathrm{Var}(X_1)$. D'après l'inégalité de Cauchy-Schwarz, on a

$$\left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n \frac{1}{n^2}\right) \geq \left(\sum_{i=1}^n \frac{a_i}{n}\right)^2.$$

Puisque $a_1 + \cdots + a_n = 1$, on en déduit que $\sum_{i=1}^n a_i^2 \geq \frac{1}{n}$ avec égalité si et seulement si $a_i = \frac{1}{n}$ pour tout $i \in \{1, 2, \ldots, n\}$. L'estimateur de variance minimale dans la classe des estimateurs linéaires et sans biais est donc la moyenne empirique $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

Exercice 3. On observe un échantillon $(X_1, \dots, X_n)$ de n variables aléatoires indépendantes et de même loi exponentielle de paramètre $\theta > 0$ inconnu.

1. Déterminer l'estimateur du maximum de vraisemblance $\widehat{\theta}_n = \theta_n(X_1, \dots, X_n)$ de θ et montrer qu'il converge presque sûrement vers θ lorsque $n \to \infty$.

2. Cet estimateur est-il sans biais ?
Indication. On rappelle que la densité de la loi $\Gamma(a, \lambda)$ est $\frac{1}{\Gamma(a)}\lambda^a x^{a-1} e^{-\lambda x} \mathbf{1}_{x>0}$.

3. Démontrer que $\sqrt{n}(\widehat{\theta}_n - \theta)$ converge en loi vers une loi qu'on déterminera.
Indication. On pourra utiliser la méthode Delta.

Solution.

1. Une densité au point $(x_1, \dots, x_n)$ est $p(\theta) = \theta^n e^{-\theta(x_1+\cdots+x_n)}$. Maximisons la fonction f: sa dérivée est $\theta^{n-1}e^{-\theta(x_1+\cdots+x_n)}(n-\theta(x_1+\cdots+x_n))$. Ainsi, p est maximale au point $\theta = \frac{n}{x_1+\cdots+x_n}$, de sorte que l'EMV est

$$\widehat{\theta}_n = \frac{n}{X_1 + \cdots + X_n}.$$

D'après la loi forte des grands nombres, $\widehat{\theta}_n$ converge presque sûrement vers $\frac{1}{\mathbb{E}(X_1})=\theta$.

2. La variable aléatoire $X_1 + \cdots + X_n$ suit une loi $\Gamma(n, \theta)$. Ainsi, d'après le théorème de transfert,

$$\mathbb{E}(\widehat{\theta}_n) = \int_0^\infty \frac{n}{x}\frac{1}{\Gamma(n)}\theta^n x^{n-1}e^{-\theta x}dx = \frac{n\theta^n}{\Gamma(n)}\int_0^\infty x^{n-2}e^{-\theta x}dx = \frac{n\theta^n}{\Gamma(n)}\frac{\Gamma(n-1)}{\theta^{n-1}} = \theta\frac{n}{n-1}.$$

L'estimateur n'est donc pas sans biais.

3. D'après le théorème central limite,

$$\sqrt{n}\left(\frac{1}{\widehat{\theta}_n} - \frac{1}{\theta}\right) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, \frac{1}{\theta^2}).$$

On applique la méthode Delta avec la fonction $g(x) = \frac{1}{x}$, dérivable au point $1/\theta$. Comme $g'(1/\theta)^2 = \theta^4$, on a

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, \theta^2).$$

Exercice 4. On reprend le modèle AR(1) de l'exercice 1 en supposant maintenant que $(Y_n)_{n\geq 1}$ est une suite de variables aléatoires indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$ (avec σ connu). On fixe $\theta \in \mathbb{R}$ (inconnu) et on définit la suite récurrente aléatoire $(X_n)_{n\geq 1}$ par

$$X_0 = 0, \qquad X_i = \theta X_{i-1} + Y_i, \quad i \geq 1.$$

Déterminer l'estimateur du maximum de vraisemblance $\widehat{\theta}_n = \theta_n(X_1, \dots, X_n)$ de θ.

Solution. On commence par déterminer une densité p_n de $(X_1, \ldots, X_n)$ en utilisant la méthode de la fonction muette et en raisonnant par récurrence sur n. En notant g la densité d'une loi gaussienne $\mathcal{N}(0, \sigma^2)$, montrons que

$$p_n(x_1, \ldots, x_n) = g(x_1)g(x_2 - \theta x_1) \cdots g(x_n - \theta x_{n-1}).$$

Pour $n = 1$, le résultat est vrai. Supposons le résultat acquis au rang $n - 1$ et démontrons-le au rang n. Par indépendance de $(X_1, \ldots, X_{n-1})$ avec Y_n, pour toute fonction $F : \mathbb{R}^n \to \mathbb{R}$ continue bornée on a d'après le théorème de transfert

$$\begin{aligned} \mathbb{E}\big(F(X_1, \ldots, X_n)\big) &= \mathbb{E}\big(F(X_1, \ldots, X_{n-1}, \theta X_{n-1} + Y_n)\big) \\ &= \int_{\mathbb{R}^{n-1}} dx_1 \cdots dx_{n-1} \int_{\mathbb{R}} dy F(x_1, \ldots, x_{n-1}, \theta x_{n-1} + y) f_{n-1}(x_1, \ldots, x_n) g(y). \end{aligned}$$

En faisant le changement de variable $u = \theta x_{n-1} + y$, on a donc

$$\mathbb{E}\big(F(X_1, \ldots, X_n)\big) = \int_{\mathbb{R}^n} dx_1 \cdots dx_n F(x_1, \ldots, x_{n-1}, x_n) f_{n-1}(x_1, \ldots, x_{n-1}) g(x_n - \theta x_{n-1}).$$

Donc

$$p_n(x_1, \ldots, x_n) = f_{n-1}(x_1, \ldots, x_{n-1}) g(x_n - \theta x_{n-1})$$

et le résultat désiré s'ensuit.

Pour déterminer l'estimateur du maximum de vraisemblance, posons $h(\theta) = g(x_1)g(x_2 - \theta x_1) \cdots g(x_n - \theta x_{n-1})$ et maximisons h en maximisant plutôt $L(\theta) = \ln h(\theta)$. On a

$$L(\theta) = -\frac{n \ln(\sigma\sqrt{2\pi})}{2} - \frac{x_1^2}{2\sigma^2} + \sum_{i=1}^{n-1} \frac{(x_{i+1} - \theta x_i)^2}{2\sigma^2},$$

d'où

$$L'(\theta) = \sum_{i=1}^{n-1} \frac{-x_i(x_{i+1} - \theta x_i)}{\sigma^2} = \sigma^{-2} \left(\theta \sum_{i=1}^{n-1} x_i^2 - \sum_{i=1}^{n-1} x_i x_{i-1} \right).$$

On a donc

$$\widehat{\theta}_n = \frac{\sum_{i=1}^{n-1} X_i X_{i-1}}{\sum_{i=1}^{n-1} X_i^2}.$$

Exercice 5. Une colonie de vampires a élu domicile dans un château des Carpates, et le comte D. souhaite estimer leur nombre. Pour cela, une nuit de pleine lune, le comte en capture 10, leur mord les oreilles, puis les relâche. La nuit suivante, il en capture 10 au hasard. Il constate que trois ont une morsure aux oreilles.
Aidez le comte D. en utilisant un estimateur du maximum de vraisemblance.

Solution. Notons n le nombre de vampires dans le château. Alors la probabilité que parmi 10 vampires capturés au hasard la nuit suivante trois vampires ont des morsures aux oreilles est

$$p_n = \frac{\binom{10}{3}\binom{n-10}{7}}{\binom{n}{10}}.$$

Pour trouver la valeur de n maximisant cette quantité, on remarque que

$$\frac{p_n}{p_{n+1}} = \frac{n^2 - 15n - 16}{n^2 - 18n + 81}.$$

On a $p_n/p_{n+1} \leq 1$ ssi $n^2 - 15n - 16 \leq n^2 - 18n + 81$ ssi $3n \leq 97$. Donc p_n/p_{n+1} est strictement inférieur à 1 pour $n \leq 32$ et est strictement supérieur à 1 pour $n \geq 33$. Ainsi, p_n est maximal pour $n = 33$.

Exercice 7. Soit $g : [0,1] \to [0,1]$ une fonction (mesurable) bornée. On souhaite calculer $m = \int_0^1 g(x)dx$. On pose $\sigma^2 = \int_0^1 g(x)^2 dx - m^2$. Soient X et Y des variables aléatoires indépendantes de loi uniforme sur $[0,1]$. On pose

$$U = \mathbf{1}_{Y \leq g(X)}, \qquad V = g(X), \qquad W = \frac{g(X) + g(1-X)}{2}.$$

1. Calculer l'espérance et la variance de U, V, W. Comparer les variances de U et V.
2. Proposer trois méthodes de type Monte-Carlo pour calculer m.

On suppose dans la suite que g est monotone.

3. Vérifier que $\mathbb{E}(g(X)g(1-X)) \leq m^2$ et comparer les variances de V et W.

   Indication. On pourra montrer que $(g(x) - g(y))(g(1-x) - g(1-y)) \leq 0$ pour tout $x, y \in [0,1]$.

Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes de loi uniforme sur $[0,1]$.

4. On considère les estimateurs suivants de m:

$$A_n = \frac{1}{2n} \sum_{k=1}^{2n} g(X_k) \qquad \text{et} \qquad B_n = \frac{1}{2n} \sum_{k=1}^{n} (g(X_k) + g(1 - X_k)).$$

   Montrer qu'ils sont sans biais. Lequel possède la plus petite variance ?

5. Dans le cas où $g(x) = x^2$, déterminer le nombre n de simulations nécessaires garantissant une précision relative de 1% sur le calcul de m en erreur quadratique avec A_n et B_n (la précision relative étant $\frac{\mathrm{Var}(A_n)}{m^2}, \frac{\mathrm{Var}(B_n)}{m^2}$).

Solution.

1. Tout d'abord, U suit une loi de Bernoulli de paramètre $\mathbb{P}(Y \leq g(X))$. On a :

$$\mathbb{P}(Y \leq g(X)) = \mathbb{E}(\mathbf{1}_{Y \leq g(X)}) = \int_{[0,1]^2} \mathbf{1}_{y \leq g(x)} dx dy = \int_0^1 g(x) dx = m.$$

   Ainsi, $\mathbb{E}(U) = m$ et $\mathrm{Var}(U) = m(1-m)$.

   On a :

$$\mathbb{E}(V) = \int_0^1 g(x)dx = m \qquad \text{et} \qquad \mathbb{E}(V^2) = \int_0^1 g(x)^2 dx, \quad \text{donc } \mathrm{Var}(V) = \sigma^2.$$

Comme X et $1-X$ ont même loi, on a par linéarité de l'espérance

$$\mathbb{E}(W) = m \quad \text{et} \quad \mathbb{E}(W^2) = \frac{(\sigma^2+m^2) + \int_0^1 g(x)g(1-x)dx}{2}.$$

Ainsi,

$$\mathrm{Var}(W) = \frac{\sigma^2 - m^2 + \int_0^1 g(x)g(1-x)dx}{2}.$$

2. Par la loi forte des grands nombres, presque sûrement,

$$\begin{aligned} m &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{Y_i \leq g(X_i)} = \lim_{n\to\infty} \frac{\mathrm{Card}(\{1 \leq i \leq n : Y_i \leq g(X_i)\})}{n} \\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} g(X_i) \\ &= \lim_{n\to\infty} \frac{1}{2n}\sum_{i=1}^{n} (g(X_i) + g(1-X_i)). \end{aligned}$$

3. Comme g est monotone, on a $(g(x)-g(y))(g(1-x)-g(1-y)) \leq 0$ pour tous $x, y \in [0,1]$. On en déduit que $\mathbb{E}(g(X)g(1-X)) + \mathbb{E}(g(Y)g(1-Y)) - 2\mathbb{E}(g(X))\mathbb{E}(g(Y)) \leq 0$, c'est-à-dire que $\mathbb{E}(g(X)g(1-X)) \leq m^2$.

   Ainsi, $\mathrm{Var}(W) \leq \frac{\sigma^2}{2} = \frac{\mathrm{Var}(V)}{2}$.

4. On a bien $\mathbb{E}(A_n) = \mathbb{E}(B_n) = m$. Par ailleurs, compte tenu de ce qui précède,

$$\mathrm{Var}(A_n) = \frac{\sigma^2}{2n}, \qquad \mathrm{Var}(B_n) = \frac{\mathrm{Var}(W)}{n} \leq \frac{\sigma^2}{2n}.$$

   La variance de B_n est donc moindre, tout en ayant besoin d'appeler g en un même nombre de points que A_n.

5. $\frac{\mathrm{Var}(A_n)}{m^2} = \frac{1}{100}$ donne $n = \frac{50\sigma^2}{m^2}$. Pour $g(x) = x^2$, il vient $m = \frac{1}{3}$ et $\sigma^2 = \frac{4}{45}$ et donc $n = \frac{1800}{45} = 40$ avec A_n. Or ici $\mathrm{Var}(W) = \frac{1}{180}$, donc $\frac{\mathrm{Var}(B_n)}{m^2} = \frac{1}{100}$ donne $n = \frac{100\mathrm{Var}(W)}{m^2} = \frac{900}{180} = 5$. Il faut 8 fois moins d'appels à g avec B_n qu'avec A_n pour la même précision.