X2016 – MAP 311
PC 7 – 12 juin 2017 – Estimateurs, TCL, vecteurs gaussiens
Corrigé des questions non abordées en PC

Igor Kortchemski – igor.kortchemski@cmap.polytechnique.fr

Corrigé des exercices non traités sur http://www.normalesup.org/~kortchem/MAP311 *un peu après la PC.*

1 Estimateur du maximum de vraisemblance

Exercice 3. **(Plus appliqué)** Une colonie de vampires a élu domicile dans un château des Carpates, et la comptesse D. souhaite estimer leur nombre. Pour cela, une nuit de pleine lune, la comptesse en capture 10, leur mord les oreilles, puis les relâche. La nuit suivante, elle en capture 10 au hasard. Trois ont une morsure aux oreilles.

Aidez la comptesse D. en utilisant un estimateur du maximum de vraisemblance.

Corrigé : 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

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

qui est strictement inférieur à 1 pour $n \leq 32$ et qui est strictement supérieur à 1 pour $n \geq 33$. Ainsi, p_n est maximal pour $n = 33$. □

2 Problème récapitulatif

Exercice 4. **(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.

Première partie.

(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) À quelle condition sur a 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 ?

Pour des questions, demande d'explications etc., n'hésitez pas à m'envoyer un mail.

(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 $a \in]-1,1[$, étudier la convergence en loi de la moyenne empirique $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

Deuxième partie. On suppose 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), que $X_0 = 0$ et que a est inconnu.

(6) Soit g la densité d'une loi gaussienne $\mathcal{N}(0,\sigma^2)$. Montrer qu'une densité f_n de $(X_1,\dots,X_n)$ au point $(x_1,\dots,x_n)$ est

$$f_n(x_1,\dots,x_n) = g(x_1)g(x_2-ax_1)\cdots g(x_n - ax_{n-1}).$$

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

Corrigé :

Première partie.

(1) On montre par récurrence sur n que toute combinaison linéaire de $(X_0,\dots,X_n)$ est gaussienne. Pour $n=0$ il n'y a rien à faire. Supposons acquis le résultat au rang n et montrons le au rang $n+1$. Pour des réels $(\lambda_i)_{0\leq i\leq n+1}$, on a

$$\sum_{i=0}^{n+1}\lambda_i X_i = \lambda_0 X_0 + \cdots + \lambda_{n-1}X_{n-1} + (a\lambda_{n+1}+\lambda_n)X_n + Y_{n+1}.$$

Par hypothèse de récurrence, $\lambda_0 X_0 + \cdots + \lambda_{n-1}X_{n-1} + (a\lambda_{n+1}+\lambda_n)X_n$ est une gaussienne, de plus indépendante de Y_{n+1}, d'où le résultat.

(2) Pour trouver la loi de X_n, il suffit de trouver son espérance et sa variance. On a

$$\mathbb{E}[X_n] = a\mathbb{E}[X_{n-1}] + \mathbb{E}[Y_n] = a\mathbb{E}[X_{n-1}] + m, \qquad \mathrm{Var}(X_n) = a^2\mathrm{Var}(X_{n-1}) + \sigma^2.$$

Ce sont deux suites arithmético-géométriques, et on trouve

$$\mathbb{E}[X_n] = \begin{cases} mn + m_0 & \text{si } a = 1 \\ a^n\left(m_0 - \frac{m}{1-a}\right) + \frac{m}{1-a} & \text{si } a \neq 1, \end{cases} \qquad \mathrm{Var}(X_n) = \begin{cases} \sigma^2 n + \sigma_0^2 & \text{si } a = 1 \\ a^{2n}\left(\sigma_0^2 - \frac{\sigma^2}{1-a}\right) + \frac{\sigma^2}{1-a} & \text{si } a \neq 1. \end{cases}$$

Pour calculer $\mathrm{Cov}(X_k, X_n)$, on remarque que

$$X_n = Y_n + aY_{n-1} + \cdots + a^{n-k-1}Y_{k+1} + a^{n-k}X_k,$$

d'où, par indépendance de X_k et des Y_i,

$$\mathrm{Cov}(X_k, X_n) = a^{n-k}\mathrm{Var}(X_k).$$

Comme (X_0, X_1) est un vecteur gaussien, $X_0 + cX_1$ est indépendant de X_0 si et seulement si $\mathrm{Cov}(X_0 + cX_1, X_0) = 0$. Or

$$\mathrm{Cov}(X_0 + cX_1, X_0) = \mathrm{Var}(X_0) + c\mathrm{Cov}(X_0, X_1) = \mathrm{Var}(X_0)(c+a).$$

Ainsi, $X_0 + cX_1$ est indépendant de X_0 si et seulement si $c = -a$.

(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. On rappelle que

$$\mathbb{E}[X_n] = \begin{cases} mn + m_0 & \text{si } a = 1 \\ a^n\left(m_0 - \frac{m}{1-a}\right) + \frac{m}{1-a} & \text{si } a \neq 1, \end{cases} \qquad \mathrm{Var}(X_n) = \begin{cases} \sigma^2 n + \sigma_0^2 & \text{si } |a| = 1 \\ a^{2n}\left(\sigma_0^2 - \frac{\sigma^2}{1-a^2}\right) + \frac{\sigma^2}{1-a^2} & \text{si } |a| \neq 1. \end{cases}$$

On a

$$\mathbb{E}\left[e^{itX_n}\right] = e^{i\mathbb{E}[X_n]t - \frac{\mathrm{Var}(X_n)t^2}{2}}$$

Si $|a| < 1$ ou si $(m_0, \sigma_0^2) = (\frac{m}{1-a}, \frac{\sigma^2}{1-a^2})$, on voit que

$$\mathbb{E}\left[e^{itX_n}\right] \underset{n\to\infty}{\longrightarrow} e^{it\frac{m}{1-a} - \frac{\sigma^2}{1-a^2}\cdot\frac{t^2}{2}}.$$

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,

$$\mathbb{E}\left[e^{itX_n}\right] \underset{n\to\infty}{\longrightarrow} \mathbb{1}_{t=0},$$

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 beacoup plus délicat. 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.

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 calcule ainsi la fonction caractéristique de $\overline{X}_n$:

$$\mathbb{E}\left[e^{it\overline{X}_n}\right] = e^{i\mathbb{E}[\overline{X}_n]t - \frac{\mathrm{Var}(\overline{X}_n)}{2}t^2}.$$

Il s'agit donc d'étudier le comportement asymptotique de $\mathbb{E}\left[\overline{X}_n\right]$ et de $\mathrm{Var}(\overline{X}_n)$ lorsque $n \to \infty$.

Tout d'abord,

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

Pour calculer la variance de $\overline{X}_n$, on peut remarquer que $X_k = a^k X_0 + \sum_{i=1}^k a^{k-i} Y_i$ 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

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

Ainsi,

$$\mathbb{E}\left[e^{it\overline{X}_n}\right] \underset{n\to\infty}{\longrightarrow} e^{i\frac{m}{1-a}t}.$$

Donc $\overline{X}_n$ converge en loi vers la variable aléatoire constante $\frac{m}{1-a}$.

Remarque. Une fois qu'on sait que $\mathbb{E}\left[\overline{X}_n\right] \to \frac{m}{1-a}$ et que $\mathrm{Var}(\overline{X}_n) \to 0$, il est possible d'en déduire que $\overline{X}_n$ converge en loi vers $\frac{m}{1-a}$ en toute généralité en utilisant l'inégalité de Markov. En effet, pour tout $\epsilon > 0$

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

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

Deuxième solution.

Deuxième partie.

(6) On raisonne par récurence sur n en utilisant la méthode de la fonction muette. Pour $n = 1$, le résultat est vrai. Supposons le 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

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

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

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

Donc

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

et le résultat désiré s'ensuit.

(è) Pour déterminer l'estimateur du maximum de vraisemblance, posons $h(a) = g(x_1)g(x_2 - ax_1)\cdots g(x_n - ax_{n-1})$ et maximisons h en maximisant plutôt $L(a) = \ln h(a)$. On a

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

d'où

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

On a donc

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

□

4 Pour aller plus loin

Exercice 6. **(Estimateurs linéaires)** Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes de même loi et 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 $\mathbb{E}\left[\widehat{\theta}_n\right] = \theta$ et de variance minimale dans la classe des estimateurs linéaires $\widehat{\theta}_n = \sum_{k=1}^n a_k X_k$.

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

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

Puisque $a_1 + \dots + a_n = 1$, 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 $i \in \{1, 2, \dots, n\}$. 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 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

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

(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 $\mathbb{E}[g(X)g(1-W)] \leq m^2$ 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 :

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

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 $\frac{\mathsf{Var}(A_n)}{m^2}, \frac{\mathsf{Var}(B_n)}{m^2}$).

Corrigé :

(1) Tout d'abord, U suit une loi de Bernoulli de paramètre $\mathbb{P}(Y \leq g(X))$. On calcule cette quantité avec le théorème de transfert :

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

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

D'après le théorème de transfert,

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

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,

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

On a

$$\mathsf{Var}(U) - \mathrm{V} = \int_0^1 g(x)(1-g(x))dx.$$

(2) Par la loi forte des grands nombres, presque sûrement,

$$m = \lim_{n\to\infty} \frac{\mathrm{Card}(\{1 \le i \le n : Y_i \le g(X_i)\})}{n} = \lim_{n\to\infty} \sum_{i=1}^{n} g(X_i) = \lim_{n\to\infty} \frac{1}{2n}\sum_{i=1}^{n}(g(X_i)+g(1-X_i)).$$

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

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

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

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

La variance de A_n est donc moindre, mais A_n nécessite un échantillon deux fois plus gros que B_n.

(5) $\frac{\mathsf{Var}(A_n)}{m^2} = \frac{1}{100}$ 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 $n \simeq 40$ avec A_n, et en fait plus de 16 fois moins avec B_n car il s'avère que $\mathsf{Var}(W) = \frac{1}{180}$.

□