X2015 – MAP 311

PC 8 – 20 juin 2016 – Vecteurs gaussiens, estimateurs du maximum de vraisemblance, tests

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

1 Vecteurs gaussiens

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) 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 ?

(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 $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

2 Estimateur du maximum de vraisemblance

Exercice 2. 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 ?
On rappelle que la densité de la loi $\Gamma(a,\lambda)$ *est* $\frac{1}{\Gamma(a)}\lambda^a x^{a-1}e^{-\lambda x}\mathbb{1}_{x>0}$.

(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 ».

(4) Construire un intervalle de confiance asymptotique à 95% pour θ.

Exercice 3. 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

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

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

3 Tests

Exercice 4. Sur un échantillon représentatif de 1000 personnes, on rapporte les avis favorables pour un homme politique. En novembre, il y avait 38% d'avis favorables, et 36% en décembre. Un éditorialiste dans son journal prend très au sérieux cette chute de 2 points. Le but de cet exercice est de confirmer ou d'infirmer la position du journaliste.

On note p la proportion d'avis favorables en novembre, et q cette proportion en décembre, et on se propose de test $H_0 : p-q = 0$ contre $H_1 : p-q \neq 0$ au niveau 5%. On note $\widehat{p}_n$ la proportion d'avis favorables dans un échantillon représentatif de n personnes en novembre (et de même $\widehat{q}_n$ pour décembre).

(1) Démontrer que $\sqrt{n}(\widehat{p}_n - p, \widehat{q}_n - q)$ converge en loi vers un vecteur gaussien dont on précisera la matrice de covariance. En déduire que

$$\sqrt{n}(\widehat{p}_n - \widehat{q}_n) \quad \xrightarrow[n\to\infty]{\text{loi}} \quad \mathcal{N}(0, p(1-p) + q(1-q)).$$

(2) Conclure en prenant pour statistique de test

$$T_n = \frac{\sqrt{n}(\widehat{p}_n - \widehat{q}_n)}{\sqrt{\widehat{p}_n(1-\widehat{p}_n) + \widehat{q}_n(1-\widehat{q}_n)}}$$

et une région de rejet une région de la forme $\{|T_n| > c\}$.

4 À chercher pour la prochaine fois

Exercice 5. Soit $(X_k)_{k\geq 1}$ des variables aléatoires indépendantes de même loi uniforme sur $[0,\theta]$ (avec θ inconnu).

(1) Montrer que l'estimateur du maximum de vraisemblance $\widehat{\theta}_n = \widehat{\theta}_n(X_1,\dots,X_n)$ de θ est $\widehat{\theta}_n = \max(X_1,\dots,X_n)$.

(2) Montrer que $W_n = n\left(1 - \frac{\widehat{\theta}_n}{\theta}\right)$ converge en loi quand $n \to \infty$ et déterminer sa limite.

(3) Construire un intervalle de confiance asymptotique à 95% pour θ.

5 Plus appliqué

Rappel (théorème de Cochran, extension de la proposition 6.2.8 du poly). Soit X un vecteur colonne aléatoire de $\mathbb{R}^n$ de loi $\mathcal{N}(m, \sigma^2 I_n)$ (avec $m \in \mathbb{R}^n$, $\sigma > 0$) et $\mathbb{R}^n = E_1 \oplus \cdots \oplus E_p$ une décomposition de $\mathbb{R}^n$ en somme directe de p sous-espaces vectoriels orthogonaux de dimensions $d_1, \ldots, d_p$ avec $d_1 + \cdots + d_p = n$. Soit $\mathbf{P}_k$ la matrice du projecteur orthogonal sur E_k et $Y_k = \mathbf{P}_k X$ la projection orthogonale de X sur E_k. Alors :

(1) les vecteurs aléatoires $(Y_1, \ldots, Y_p)$ sont indépendants et Y_k suit la loi $\mathcal{N}(\mathbf{P}_k m, \sigma^2 \mathbf{P}_k)$;

(2) les variables aléatoires réelles $(\|Y_i - \mathbf{P}_i m\|^2)_{1 \leq i \leq p}$ sont indépendantes et $\|Y_k - \mathbf{P}_k m\|^2/\sigma^2$ suit la loi $\chi^2(d_k)$.

Exercice 6. On considère que la réponse d'un appareil de mesure à un signal déterministe ξ est égale à $a\xi$ plus un bruit gaussien centré de variance b, où $(a,b) \in \mathbb{R} \times \mathbb{R}_+^*$. On se propose d'étalonner l'appareil (c'est-à-dire estimer les valeurs de a et b) en envoyant une suite $x = (x_1, x_2, \ldots, x_n)$ de signaux connus. On note $Y_i = a x_i + \sqrt{b} U_i$ la réponse au i-ième signal où on suppose que les coordonnées du vecteur $U = (U_1, U_2, \ldots, U_n)$ sont des variables aléatoires gaussiennes centrées réduites indépendantes. On note $Y = (Y_1, Y_2, \ldots, Y_n)$,

$$\widehat{A}_n = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2} \qquad \text{et} \qquad \widehat{B}_n = \frac{\sum_{i=1}^n (Y_i - x_i \widehat{A}_n)^2}{n-1}.$$

(1) Donner la loi de $\widehat{A}_n$. À quelle condition sur la suite $(x_i)_{i \geq 1}$ a-t-on $\mathbb{E}\left[(\widehat{A}_n - a)^2\right] \to 0$ lorsque $n \to \infty$?

On complète $e_1 = \frac{x}{\|x\|}$ en une base orthonormée $(e_1, \ldots, e_n)$ de $\mathbb{R}^n$. Notons P la projection orthogonale sur $E_1 = \mathrm{Vect}(e_1)$ et Q la projection orthogonale sur $E_2 = \mathrm{Vect}(e_2, \ldots, e_n)$.

(2) Déterminer PY et QY. En déduire que $\widehat{A}_n$ et $\widehat{B}_n$ sont des variables aléatoires indépendantes. Donner l'espérance et la variance de $\widehat{B}_n$.

(3) Montrer qu'il existe une constante c (dépendant de x) telle que la variable aléatoire $c\frac{\widehat{A}_n - a}{\sqrt{\widehat{B}_n}}$ suive une loi de Student.

(4) Donner un intervalle de confiance à 95% pour le paramètre a, suivant que l'on connaît la valeur de b ou non.

6 Pour aller plus loin

Exercice 7. Soient $(X_i)_{1 \leq i \leq n}$ des variables aléatoires réelles indépendantes de même loi, de carré intégrable, d'espérance m et de variance σ^2. On suppose que la moyenne empirique $\overline{X}_n$ et la variance empiriques V_n, définies par

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \qquad V_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2,$$

sont indépendantes. Le but de cet exercice est de démontrer que la loi de X_i est une loi gaussienne $\mathcal{N}(m, \sigma^2)$.

(1) Calculer $\mathbb{E}[(n-1)V_n]$ en fonction de σ^2. Montrer que pour tout $t \in \mathbb{R}$:

$$\mathbb{E}\left[(n-1)V_n e^{itn\overline{X}_n}\right] = (n-1)\psi(t)^n \sigma^2.$$

(2) En développant V_n dans l'égalité précédente, vérifier que :

$$\mathbb{E}\left[(n-1)V_n e^{itn\overline{X}_n}\right] = -(n-1)\psi''(t)\psi(t)^{n-1} + (n-1)\psi'(t)^2\psi(t)^{n-2}.$$

(3) En déduire que, sur un voisinage ouvert de 0, ψ est solution de l'équation différentielle :

$$\frac{\psi''}{\psi} - \left(\frac{\psi'}{\psi}\right)^2 = -\sigma^2, \qquad \psi(0) = 1, \qquad \psi'(0) = 0.$$

(4) En déduire que la loi des variables X_i est la loi gaussienne $\mathcal{N}(0, \sigma^2)$.

(5) Que peut-on dire si l'on ne suppose plus $m = 0$?