Eléments de statistique

Théorème de Cochran et applications

A. Godichon-Baggioni

I. Théorèmes asymptotiques multivariés

Théorème de continuité

Théorème (Théorème de continuité)

Soit $g : D \longrightarrow \mathbb{R}^{d'}$ une fonction continue sur un ouvert $D \subset \mathbb{R}^d$ et soit X un vecteur aléatoire à valeurs dans D. Soit (X_n) une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$. Si (X_n) converge vers X, alors $g(X_n)$ hérite du mode de convergence de (X_n), i.e

$$X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X \Longrightarrow g(X_n) \xrightarrow[n\to+\infty]{\mathcal{L}} g(X)$$

$$X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X \Longrightarrow g(X_n) \xrightarrow[n\to+\infty]{\mathbb{P}} g(X)$$

$$X_n \xrightarrow[n\to+\infty]{p.s} X \Longrightarrow g(X_n) \xrightarrow[n\to+\infty]{p.s} g(X)$$

Théorème de Slutsky

Théorème

Soit (X_n) une suite de vecteurs aléatoires convergeant en loi vers X et (Y_n) une suite de vecteurs aléatoires convergeant en probabilité vers un vecteur constant c. Alors

$$X_n + Y_n \xrightarrow[n \to +\infty]{\mathcal{L}} X + c$$

et

$$X_n Y_n \xrightarrow[n \to +\infty]{\mathcal{L}} Xc \quad \text{et} \quad Y_n X_n \xrightarrow[n \to +\infty]{\mathcal{L}} cX$$

Remarque : Les résultats précédents sous entendent bien évidemment que les dimensions de X et c sont "adaptées".

Théorèmes limites

Théorème (Loi des grands nombres)

Soient $X_1, \ldots, X_n$ une suite de vecteurs alétoires indépendants et identiquement distribués tel que $\mathbb{E}\left[\|X_1\|\right] < +\infty$. Alors

$$\overline{X}_n \xrightarrow[n\to+\infty]{p.s} \mathbb{E}\left[X_1\right].$$

Théorème (Théorème Central Limite)

Soient $X_1, \ldots, X_n$ une suite de vecteurs aléatoires indépendants et identiquement distribuées tel que $\mathbb{E}\left[\|X_1\|^2\right] < +\infty$. Alors

$$\sqrt{n}\left(\overline{X}_n - \mathbb{E}\left[X_1\right]\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, Var\left[X_1\right]\right).$$

Delta méthode

Théorème

Soit (X_n) *une suite de vecteurs aléatoires,* X *un vecteur aléatoire,* v_n *une suite de réels telle que* $\lim_{n\to+\infty} v_n = +\infty$ *et* m *un vecteur tel que*

$$v_n\left(X_n - m\right) \xrightarrow[n\to+\infty]{\mathcal{L}} X.$$

Soit $g = (g_1, \ldots, g_{d'}) : D \longrightarrow \mathbb{R}^{d'}$ *avec* $m \in D \subset \mathbb{R}^d$ *une fonction continuement différentiable, alors*

$$v_n\left(g\left(X_n\right) - g(m)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} J_g(m)X$$

où $J_g(m) = \left(\frac{\partial}{\partial m_i} g_j(m)\right)_{i,j}$ *est la matrice Jacobienne de* g *en* m*.*

CAS PARTICULIER

Soit (X_n) une suite d'estimateurs de m asymptotiquement normaux, i.e il existe une matrice $\Sigma \in \mathcal{M}_d(\mathbb{R})$ tel que

$$\sqrt{n}\,(X_n - m) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\,(0, \Sigma)\,.$$

Soit $g = (g_1, \ldots, g_{d'}) : D \longrightarrow \mathbb{R}^{d'}$ avec $m \in D \subset \mathbb{R}^d$ une fonction continuement différentiable, alors

$$\sqrt{n}\,(g\,(X_n) - g(m)) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, J_g(m)\Sigma J_g(m)^T\right).$$

II. Théorème de Cochran et applications

Théorème de Cochran

Théorème (Théorème de Cochran)

Soit X une vecteur aléaoire de $\mathbb{R}^n$ suivant une loi $\mathcal{N}\left(m, \sigma^2 I_n\right)$ avec $\sigma^2 > 0$. Soit $\mathbb{R}^n = E_1 \oplus E_2 \oplus \ldots \oplus E_p$ une décomposition de $\mathbb{R}^n$ en somme directe de p sous-espaces vectoriels orthogonaux de dimensions respectives $d_1, \ldots, d_p$. Soit P_k le projecteur orthogonal sur E_k et $Y_k = P_k X$ la projection orthogonale de X sur E_k.

1. *Les projections $Y_1, \ldots, Y_p$ sont des vecteurs gaussiens indépendants et $Y_k \sim \mathcal{N}\left(P_k m, \sigma^2 P_k\right)$.*
2. *Les variables aléatoires $\left\|Y_1 - P_1 m\right\|^2, \ldots, \left\|Y_p - P_p m\right\|^2$ sont indépendantes et*

$$\frac{\left\|Y_k - P_k m\right\|^2}{\sigma^2} \sim \chi^2_{d_k}$$

COROLLAIRE

Corollaire

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes et identiquement distribuées avec $X_1 \sim \mathcal{N}\left(m, \sigma^2\right)$. On a alors

1. $\overline{X}_n \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$.
2. $\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2_{n-1}$.
3. $\overline{X}_n$ *et* S_n^2 *sont indépendantes.*

Autres résultats

Proposition

Soient Z, Z_p, Z_q trois variables aléatoires telles que

1. $Z = Z_p + Z_q$
2. $Z_p \sim \chi_p^2$ *et* $Z_q \sim \chi_q^2$
3. Z_p *et* Z_q *sont indépendantes*

alors $Z \sim \chi_{p+q}$.

TEST DU KHI DEUX

Corollaire (Test du Khi deux d'ajustement)

Soit X une variable aléatoire à valeurs dans $\{a_1, \ldots, a_K\}$ avec $\mathbb{P}[X = a_k] = p_k > 0$. Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d de même loi que X. Soit $E_k = \sum_{i=1}^n \mathbf{1}_{X_i = a_k}$. On a alors

$$\sum_{k=1}^{K} \frac{(E_k - np_k)^2}{np_k} \xrightarrow[n \to +\infty]{\mathcal{L}} \chi^2_{K-1}.$$