

# THÉORÈME DE COCHRAN TEST DU $\chi^2$ .

Théorème (de Cochran): Soit  $X \sim \mathcal{N}(0, I_d)$  un vecteur gaussien. On pose:

$$\mathbb{R}^d = E_1 \overset{\perp}{\oplus} \dots \overset{\perp}{\oplus} E_r$$

avec  $\dim E_i := d_i \quad \forall i=1, \dots, r$ . Alors, les projections orthogonales  $(P_i(X))_i$  de  $X$  sur  $E_i$  sont des vecteurs gaussiens indépendants. De plus,  $\forall i=1, \dots, r$ ,  
 $\|P_i(X)\|^2 \sim \chi^2(d_i)$ .

Preuve. Soit  $i=1, \dots, r$ , on pose  $B_i = (e_{i,1}, \dots, e_{i,d_i})$  une base orthonormée de  $E_i$ . On note  $B = B_1 \cup \dots \cup B_r$  qui est une base orthonormée de  $\mathbb{R}^d$ . On note  $Q \in \mathcal{O}_n(\mathbb{R})$  la matrice de passage entre  $B$  et  $B_{\text{can}}$ . On note

$$Y_{ij} = \langle X, e_{ij} \rangle \quad \begin{cases} \forall i=1, \dots, r \\ \forall j=1, \dots, d_i \end{cases}$$

Le vecteur colonne  $Y$  des  $(Y_{ij})_{i,j}$  vérifie alors:

$$Y = QX \sim \mathcal{N}(0, Q I_d {}^t Q) = \mathcal{N}(0, I_d)$$

par transformation linéaire du vecteur gaussien  $X$ . Comme la matrice de covariance de  $Y$  est diagonale, on a que les  $(Y_{ij})_{i,j}$  sont indépendantes mutuellement, et  $\forall i, \forall j=1, \dots, d_i, \quad Y_{ij} \sim \mathcal{N}(0, 1)$ . Par regroupement par paquets, on a que les

$$P_i(X) = \sum_{j=1}^{d_i} Y_{ij} e_{ij}$$

sont mutuellement indépendants. Enfin, en se souvenant que  $B$  est orthonormée,

$$\|P_i(X)\|^2 = \sum_{j=1}^{d_i} |Y_{ij}|^2 \sim \chi^2(d_i).$$

Remarque: Soit  $F$  un sev de  $\mathbb{R}^d$ , et  $M$  la matrice de projection orthogonale sur  $F$ . On note  $q = \dim F$ , et  $X \sim \mathcal{N}(0, I_d)$ . Alors,

$$\begin{aligned} X \sim \mathcal{N}(0, I_d) &\equiv \mathcal{N}(0, M^2) \quad \text{car } M \text{ est un projecteur} \\ &\equiv \mathcal{N}(0, M I_d {}^t M) \quad \text{car } M \text{ est un projecteur orthogonal} \\ &\equiv M U \quad \text{où } U \sim \mathcal{N}(0, I_d). \end{aligned}$$

Il vient alors par le théorème de Cochran que  $\|MU\|^2 \stackrel{\mathcal{L}}{=} \|X\|^2 \sim \chi^2(d-q)$ .

Test du  $\chi^2$  discret: Soit  $v$  une loi sur un espace fini  $\{1, \dots, q\}$  ( $q \geq 2$ )

$\forall j=1, \dots, q$ . Soit  $(X_i)_{i \geq 1}$  une suite de v.a. iid de loi  $v$ . On note

$$N_j^{(n)} := \sum_{i=1}^n \mathbb{1}_{\{X_i=j\}} \quad T_n = \sum_{j=1}^q \frac{(N_j^{(n)} - n v(j))^2}{n v(j)}$$

Alors,

$$T_n \xrightarrow{\mathcal{L}} \chi^2(q-1).$$

Preuve: Soit  $Y_i := (\mathbb{1}_{X_i=1}, \dots, \mathbb{1}_{X_i=q})$  de sorte que  $S_n := Y_1 + \dots + Y_n = (N_1^{(n)}, \dots, N_q^{(n)})$ .

Les  $(Y_i)_i$  sont iid et  $\mathbb{E}(Y_1) = (v(1), \dots, v(q))$ . On note  $\Gamma := \text{cov}(Y_1)$ .

Par le TCL multivariée, il vient:

$$Z_n := \frac{S_n - n \mathbb{E}(Y_1)}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma)$$

On note  $Z \sim \mathcal{N}(0, \Gamma)$ . Soit  $(e_1, \dots, e_q)$  la base canonique de  $\mathbb{R}^q$ . On

note  $Z_{n,j} := \langle Z_n, e_j \rangle$ . Alors

$$T_n := \sum_{j=1}^q \frac{Z_{n,j}^2}{v(j)}$$

Par transformation continue, il vient que:

$$T_n \xrightarrow{\mathcal{L}} \sum_{j=1}^q \frac{Z_j^2}{v(j)} = \|DZ\|^2$$

avec  $D := \text{diag}\left(\frac{1}{\sqrt{v(1)}}, \dots, \frac{1}{\sqrt{v(q)}}\right)$ . Par transformation linéaire,

$DZ \sim \mathcal{N}(0, D\Gamma D)$ . Or,  $\forall i=1, \dots, q, \forall j=1, \dots, q$  on a:

$$\begin{aligned} (D\Gamma D)_{ij} &= \frac{1}{\sqrt{v(i)}} \Gamma_{ij} \frac{1}{\sqrt{v(j)}} = \frac{1}{\sqrt{v(i)}\sqrt{v(j)}} (\mathbb{E}(\mathbb{1}_{X=i} \mathbb{1}_{X=j}) - \mathbb{E}(\mathbb{1}_{X=i})\mathbb{E}(\mathbb{1}_{X=j})) \\ &= \frac{1}{\sqrt{v(i)}\sqrt{v(j)}} (\delta_{ij} v(i) - v(i)v(j)) \\ &= \delta_{ij} - \sqrt{v(i)v(j)} := c_{ij}. \end{aligned}$$

On note  $C := (c_{ij})_{i,j}$  et  $u = (\sqrt{v(1)}, \dots, \sqrt{v(q)})$ . On a alors  $\|u\|^2 = 1$  et  $I_q - C = u \cdot u^t$ .

On remarque que  $\forall x \in \mathbb{R}^q, (I_q - C)x = u \cdot \langle u, x \rangle$ . D'où,  $I_q - C = P_{\text{Vect}(u)}^\perp$ .

Donc,  $C = D\Gamma D = P_{\text{Vect}(u)}^\perp$ . Par la remarque suivant le thm de

Cochran,  $T_n \xrightarrow{\mathcal{L}} \|DZ\|^2 \sim \chi^2(q-1)$ .

## Remarques:

- ▷ Développement long. Il est inévitable à l'oral de ne pas réécrire les énoncés des théorèmes, qui permettent d'introduire les variables aléatoires en jeu. Je recommande 5 min pour le thm de Cochran, 10 min pour le reste. Je recommande aussi de ne pas prouver la remarque, mais de le mettre en corollaire du thm de Cochran dans le plan de la leçon. Également, j'admettais le calcul (long et fastidieux) de  $D^2$ .
- ▷ Je réserverais ce développement aux étudiants en option A, on a besoin de pas mal d'astuces techniques: vecteurs gaussiens, transformations affines de vecteurs gaussiens, TCL multivariée, regroupement par paquets. Au contraire, pour les options A, je recommande beaucoup ce développement: il est d'excellent niveau, et le comprendre éclaire beaucoup ~~la~~ les applications du théorème de Cochran. Faire des vecteurs gaussiens un point fort est bien utile pour l'oral d'option A...!
- ▷ Le test du  $\chi^2$  que l'on explique dans ce développement est le test d'adéquation du  $\chi^2$  discret: étant donné un échantillon  $(X_1, \dots, X_n)$ , on se demande si chaque  $X_i$  suit une loi  $\nu$ . On a donc affaire à un test bilatéral de l'hypothèse  $H_0: "X \sim \nu"$  contre  $H_1: "X \not\sim \nu"$ . Sous l'hypothèse  $H_0$ ,  $T_n \rightarrow \chi^2(q-1)$ , sinon,  $T_n \rightarrow +\infty$  ps.