Test du χ^2

Ref:
- Probabilités et statistiques, M.-L. Chabanol.

Leçons: 261, 262, 264

Proposition

Soient $n \in \mathbb{N}^*$ et $m \in \mathbb{N}^*$. Soit $(X_k)_{1 \leq k \leq n}$ une suite de v.a. indépendantes identiquement distribuées, à valeurs dans $\{1, \dots, m\}$.
Notons pour tout $i \in [\![1, m]\!]$

$$N_i(n) := \sum_{k=1}^{n} \mathbb{1}_{\{X_k = i\}} \quad , \quad \pi_i := \mathbb{P}(X_1 = i)$$

Alors on a

$$D_n(\pi) \xrightarrow[n \to +\infty]{\text{Loi}} \chi^2(m-1)$$

où $D_n(\pi) := \sum_{i=1}^{m} \frac{(N_i(n) - n\pi_i)^2}{n\pi_i}$.

Preuve:

Notons $N(n)$ le vecteur $(N_1(n), \dots, N_m(n))$.
Alors $N(n)$ suit la loi multinomiale $\mathcal{M}_m(n, \pi_1, \dots, \pi_m)$.
En effet, soient $i_1, \dots, i_m$ entiers dans $[\![0, n]\!]$ tels que $\sum_{j=1}^{m} i_j = n$.
Alors on a

$$\mathbb{P}(N_1(n) = i_1, \dots, N_m(n) = i_m) = \binom{n}{i_1}\binom{n-i_1}{i_2} \cdots \binom{n-i_1-\dots-i_{m-1}}{i_m} \pi_1^{i_1} \cdots \pi_m^{i_m}$$

par indépendance des $(X_i)_{1 \leq i \leq n}$

Rappel: Soit $(X_k)_{k \in \mathbb{N}^*}$ une suite de vecteurs aléatoires de $\mathbb{R}^N$ indépendants identiquement distribués tels que $\mathbb{E}[X_1] = \mu \in \mathbb{R}^N$ et $\Gamma = \mathbb{E}[(X_1 - \mu)^t (X_1 - \mu)]$ la matrice de covariance.
Alors

$$\frac{\sum_{j=1}^{n} X_j - n\mu}{\sqrt{n}} \xrightarrow[n \to +\infty]{\text{Loi}} \mathcal{N}_N(0, \Gamma)$$

Par ce théorème central limite, on obtient

$$\frac{N(n) - n\pi}{\sqrt{n}} \xrightarrow[n \to +\infty]{\text{Loi}} \mathcal{N}(0, \Delta_\pi - \pi^t\pi)$$

où $\pi = (\pi_1, \dots, \pi_m)$ et $\Delta_\pi = \text{diag}(\pi)$.

En effet, on a

$$\forall i \in [\![1, m]\!] \quad \mathbb{E}[N_i(n)] = \mathbb{E}\Big[\sum_{k=1}^{n} \mathbb{1}_{\{X_k = i\}}\Big] \overset{(X_k) \text{ même loi}}{=} n\,\mathbb{E}[\mathbb{1}_{\{X_1 = i\}}]$$

$$" \qquad = n\pi_i$$

d'où $\mathbb{E}[N(n)] = n\pi$.

De plus, on a pour tous i, j dans $[\![1, m]\!]$ et tout $k \in [\![1, n]\!]$

$$\text{Cov}(\mathbb{1}_{\{X_k = i\}}, \mathbb{1}_{\{X_k = j\}}) = \mathbb{E}[\mathbb{1}_{\{X_k = i\}}\mathbb{1}_{\{X_k = j\}}] - \pi_i\pi_j = \delta_{i,j}\pi_i - \pi_i\pi_j$$

Notons Y_k le vecteur aléatoire $\begin{pmatrix} \mathbb{1}_{\{X_k = 1\}} \\ \cdots \\ \mathbb{1}_{\{X_k = m\}} \end{pmatrix}$. Alors comme (X_k) sont indépendantes, les vecteurs (Y_k) sont aussi indépendants.

Par ailleurs on en déduit

$$\mathbb{E}[Y_k] = \pi \quad \text{et} \quad \underbrace{\text{Cov}(Y_k, Y_k)}_{\text{matrice de covariance}} = \Delta_\pi - \pi^t\pi$$

Comme $N(n) = \sum_{k=1}^{n} Y_k$, on obtient bien la convergence annoncée

<u>Etape 2</u> : convergence de $D_n(\pi)$.

Notons f l'application continue (car polynomiale)

$$\begin{aligned} \mathbb{R}^m &\longrightarrow \mathbb{R} \\ x &\longmapsto \sum_{i=1}^{m} \frac{x_i^2}{\pi_i} \end{aligned}$$

On a alors $D_n(\pi) = f\left(\frac{N(n) - \pi n}{\sqrt{n}}\right)$. D'après la convergence en loi on obtient

$$D_n(\pi) \xrightarrow[n \to +\infty]{\text{Loi}} f(\mathcal{N}_m(0, \Delta_\pi - \pi^t\pi))$$

Il reste à déterminer la loi de $f(Z)$ avec $Z \sim \mathcal{N}_m(0, \Delta_\pi - \pi^t\pi)$.

<u>Etape 3</u>: loi de $f(\mathcal{N}_m(0, \Delta_\pi - \pi^t\pi))$

Soit Z un vecteur aléatoire de loi $\mathcal{N}_m(0, \Delta_\pi - \pi^t\pi)$.

On a alors

$$f(Z) = \sum_{i=1}^{m} \frac{Z_i^2}{\pi_i} = \|U\|_2^2$$

où $U = (U_1, \dots, U_m)$ et $U_i = \frac{Z_i}{\sqrt{\pi_i}}$.

Le vecteur U suit la loi $\mathcal{N}(0, I_m - \sqrt{\pi}^t\sqrt{\pi})$.

<u>Rappel</u>: soit $A \in \mathcal{O}_m(\mathbb{R})$. Alors $V = AU$ est un vecteur de loi

$$\mathcal{N}(0, I_m - A\sqrt{\pi}^t\sqrt{\pi}^tA)$$

tel que $\|V\|_2 = \|U\|_2$.

Comme $\|\sqrt{\pi}\|_2 = \left(\sum_{i=1}^{m} (\sqrt{\pi_i})^2\right)^{\frac{1}{2}} = 1$, il existe $A \in \mathcal{O}_m(\mathbb{R})$ tel que $A\sqrt{\pi} = e_m$, vecteur de la base canonique de $\mathbb{R}^m$.

On a alors

$$I_m - A\sqrt{\pi}^t\sqrt{\pi}^tA = I_m - e_m{}^te_m = \begin{pmatrix} 1 & & & (0) \\ & \ddots & & \\ & & 1 & \\ (0) & & & 0 \end{pmatrix}$$

On en déduit que $f(Z)$ suit la loi de $\|V\|_2^2 = \sum_{i=1}^{m-1} V_i^2$ où $V_i \sim \mathcal{N}(0,1)$, c-à-d la loi $\chi^2(m-1)$.

Cela conclut ∎