L'idée du test du χ^2 d'adéquation est la suivante : comme $N_j^n = \sum_{i=1}^n \mathbb{1}_{\{X_i=j\}}$, on a alors convergence presque sûre de N_j^n/n vers ν_j par la loi des grands nombres. On peut donc espérer que T_n ne deviendra pas trop grand pour $n \to +\infty$ (on va voir en fait que T_n converge en loi vers une variable aléatoire finie).

Théorème 12.8.4. *Soit ν une loi sur $\{1,\dots,q\}$, $(X_n)_{n\ge1}$ une suite de variables aléatoires indépendantes de loi ν. Alors les T_n convergent en loi vers une variable aléatoire de loi χ^2_{q-1} à $q-1$ degrés de liberté.*

Démonstration. Notons $(e_1,\dots,e_q)$ la base canonique de $\mathbb{R}^q$. On définit la variable q-dimensionnelle dont la $i-$ième composante vaut 1 si $X_j = i$ et 0 sinon. Autrement dit, on a $Y_j = e_{X_j}$. Bien sûr, les variables Y_j sont indépendantes et de même loi, avec $\mathbb{P}(Y_j = e_i) = \nu_i$. Posons $S_n = Y_1 + \dots + Y_n$. Soient m le vecteur espérance de Y_1 et K sa matrice de covariance. Par le théorème central limite, les variables $V_n = (S_n - nm)/\sqrt{n}$ convergent en loi vers un vecteur gaussien centré de covariance K. Si $V_{n,i}$ désigne la $i-$ième composante de V_n, comme $m = \sum_{i=1}^n \nu_i e_i$, on a aisément

$$T_n = \sum_{i=1}^q \frac{V_{n,i}^2}{\nu_i}.$$

Donc, si $U = (U_1,\dots,U_q)$ est un vecteur gaussien centré de covariance K, alors les variables T_n convergent en loi vers

$$T = \sum_{i=1}^q \frac{U_i^2}{\nu_i}.$$

On peut noter que $T = \|\operatorname{diag}(\frac{1}{\sqrt{\nu_1}},\dots,\frac{1}{\sqrt{\nu_q}})U\|_2^2$. Calculons la matrice de covariance K. On a

$$\begin{aligned} K_{i,j} &= \operatorname{Covar}(\langle Y_1, e_i\rangle, \langle Y_1, e_j\rangle) \\ &= \operatorname{Covar}(\mathbb{1}_{\{X_1=i\}}, \mathbb{1}_{\{X_1=j\}}) \\ &= \delta_{i,j}\nu_i - \nu_i\nu_j \\ &= \sqrt{\nu_i}(\delta_{i,j} - \sqrt{\nu_i}\sqrt{\nu_j})\sqrt{\nu_j}. \end{aligned}$$

On en déduit que la matrice de covariance de $\operatorname{diag}\left(\frac{1}{\sqrt{\nu_1}},\dots,\frac{1}{\sqrt{\nu_q}}\right)U$ a comme matrice la matrice C donnée par $c_{i,j} = \delta_{i,j} - \sqrt{\nu_i}\sqrt{\nu_j}$. Or on remarque que

$$\langle (\mathrm{Id}_q - C)e_i, e_j\rangle = (\mathrm{Id}_q - C)_{i,j} = \sqrt{\nu_i}\sqrt{\nu_j} = \langle \langle v, e_i\rangle v, e_j\rangle,$$