STATISTIQUE. — *Valeurs caractéristiques du coefficient de corrélation par rang de Kendall dans le cas général.* Note de M. **Marcel-Paul Schützenberger**, présentée par M. Gaston Julia.

Kendall (1) a défini un *coefficient de corrélation* par rang

$$\tau = \frac{4}{n(n-1)} \sum_{(i,j)} X_{ij} - 1,$$

où X_{ij} est une indicatrice égale à 1 ou à 0 suivant que la paire (i, j) est *rangée ou non dans le même ordre* dans les deux séries S et S′ de n objets à comparer.

Bien que cette expression ne soit pas à proprement parler un véritable *coefficient de corrélation*, son usage se recommande pourtant dans de nombreux cas comme plus commode et plus *naturel* que le coefficient ρ de Spearman.

Pour étudier la répartition de τ quand sa valeur moyenne *n'est pas nulle*, on considérera le modèle probabiliste suivant :

Les classements dans S et S′ sont ceux des valeurs hypothétiques $a_1, a_2, \ldots, a_n$ et $a'_1, a'_2, \ldots, a'_n$ prises dans les deux séries de tirages indépendants par les variables aléatoires A_i, *réparties sur un même axe*, selon des fonctions F_i sur lesquelles diverses hypothèses peuvent être faites.

Dans le cas général, les moments d'ordre élevé de la répartition de τ ont une expression compliquée, cependant au moyen des coefficients

$$\delta_{ij} = \int_{-\infty}^{+\infty} F_i(t)\, dF_j(t) - F_j(t)\, dF_i(t),$$

$$\gamma_{ijk} = \int_{-\infty}^{+\infty} \int_{0}^{+\infty} [F_i(t+t') - F_i(t)][F_k(t+t') - F_k(t)]\, dF_j(t')\, dF_j(t),$$

[qui ne sont autres que les *déviations d'indépendance* définies dans une Note antérieure (2)], on peut calculer simplement les valeurs de la moyenne $\bar{\tau}_{11}$ et de la variance σ^2_{11} de τ_{11} entre deux classements aléatoires :

$$\bar{\tau}_{11} = \frac{2}{n(n-1)} \sum_{(i,j)} \delta^2_{ij},$$

$$\sigma^2_{11} = \frac{8}{n^2(n-1)^2} \sum_{(i,j,k)} S\gamma_{ijk}(\gamma_{ijk} - 2\,\delta_{ij}\,\delta_{jk}) + \frac{4}{n^2(n-1)^2} \sum_{(i,j)} (1 - \delta^4_{ij})$$

(S, sommation sur les 3 permutations : *ijk, jki, kij*).

---

(1) *A new measure of rank correlation-Biometrika*, 20, 1938, p. 81.

(2) M. P. Schützenberger, *Comptes rendus*, 225, 1947, p. 227.

Remarquons que si, pour m objets $a, b, \ldots, m$, la probabilité qu'ils soient observés dans cet ordre est désignée par $\mathrm{Prob}[a, b, c, \ldots, m]$, on a

$$\delta_{ij} = \mathrm{Prob}[a_i a_j] - \mathrm{Prob}[a_j a_i]$$
$$\gamma_{ijk} = \mathrm{Prob}[a_i a_j] \times \mathrm{Prob}[a_j a_k] - \mathrm{Prob}[a_i a_j a_k].$$

La théorie des *formes polaires* permet d'en déduire les expressions correspondantes $\bar{\tau}_{12}$ et σ_{12}^2 relatives au cas où les deux séries de tirage aléatoire S et S′ sont effectués dans deux *familles différentes* des fonctions F_i et F'_i et l'on vérifie l'inégalité de Schwartz $\bar{\tau}_{12}^2 \leq \bar{\tau}_{11} \times \bar{\tau}_{22}$.

Notamment la corrélation τ_{01} entre un classement aléatoire et le classement *a priori* où l'on pose $A_1 < A_2 < A_3 < \ldots < A_n$ a pour valeurs caractéristiques

$$\bar{\tau}_{01} = \frac{2}{n(n-1)} \sum_{(i,j)} \delta_{ij}$$
$$\sigma_{01}^2 = \frac{8}{n^2(n-1)^2} \sum_{(i,j,k)} (\gamma_{jki} + \gamma_{kij} - \gamma_{ijk}) + \frac{4}{n^2(n-1)^2} \sum_{(i,j)} (1 - \delta_{ij}^2).$$

Dans tous les cas si d est une *borne inférieure* des $|\delta_{ij}|$, σ^2 est majorée par

$$\frac{2}{3} \times \frac{2n-1}{n(n-1)} \times (1 - d^4).$$

Cette valeur est en général *plus grande* que celle donnée par Kendall

$$\frac{2}{9} \times \frac{2n+5}{n(n-1)},$$

mais est *toujours valable* et non pas seulement dans l'hypothèse restrictive où *tous* les classements de *tous* les triples d'objets ont la *même probabilité* 1/6.

(Extrait des *Comptes rendus des séances de l'Académie des Sciences*, t. 226, p. 2122-2123, séance du 28 juin 1948.)

GAUTHIER-VILLARS IMPRIMEUR-LIBRAIRE DES COMPTES RENDUS DES SÉANCES DE L'ACADÉMIE DES SCIENCES
130048-48 Paris. — Quai des Grands-Augustins, 55.