Le problème des deux fréquences et sa généralisation

Par Edouard Franckx, Bruxelles

Résumé

La note a pour objet de démontrer que, moyennant certaine condition de régularité, la différence des fréquences $\frac{\alpha}{m}$ et $\frac{\beta}{n}$, observées sur deux séries dépendantes d'épreuves obéit à la tendance centrale.

Le résultat est étendu à la moyenne d'épreuves relatives à des variables uniformément bornées.

Le problème de la prédiction est une des préoccupations de l'actuaire; le professeur Cramer l'a défini dans le sens le plus large en disant que la prédiction est la réponse à la question «Que va-t-il se produire sous telles ou telles conditions?».

La question de la stabilité des opérations d'assurance est directement liée au problème de la prédiction.

Dans ce domaine la loi faible des grands nombres et la tendance centrale, la théorie collective du risque sont des moyens puissants de prédétermination, le second étant beaucoup plus puissant que le premier puisque plus précis.

En pratique, tout problème asymptotique donne numériquement une méthode d'approximation valable pour n grand. A l'échelle des organismes d'assurance, le nombre n répond en général à cette condition.

Dans cette note pour étudier le comportement asymptotique nous ferons appel au théorème de la tendance centrale sous la forme de Levy-Lindberg.

Les variables X_k sont indépendantes et uniformément bornées. On désigne par

$$s_n = \sum_1^n x_k, \qquad m_{s_n} = \sum_1^n m_{x_k}, \qquad a_{s_n}^2 = \sum_1^n a_{x_k}^2,$$

respectivement la somme des n premières variables, le moment du 1er ordre et la fluctuation de cette somme

$$\lim_{n\to\infty} \text{prob.} \left\{ a \leqq \frac{s_n - m_{s_n}}{a_{s_n}} \leqq b \right\} = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \longleftrightarrow \lim_{n\to\infty} a_{s_n}^2 = \infty \text{ [1]}.$$

A. Le problème des deux fréquences dans le cas d'un seul événement

1. Nous considérons un événement A de probabilité p_A invariable et nous effectuons une série d'épreuves indépendantes relatives à l'arrivée ou la non-arrivée de cet événement A.

Si les épreuves en considération sont fixées à m, on observera une fréquence que l'on désigne par $\frac{\alpha}{m}$, si on prolonge la suite d'épreuves pour en obtenir n, nous constaterons une fréquence $\frac{\beta}{n}$.

Désignons par
$$\begin{cases} E_1 E_2 \ldots E_m \\ E_1 E_2 \ldots E_m E_{m+1} \ldots E_n \end{cases}$$

les deux suites d'épreuves, elles comportent en commun les épreuves $E_1 \ldots E_m$ et par suite les fréquences $\frac{\beta}{n}$ et $\frac{\alpha}{m}$ sont stochastiquement dépendantes.

Nous nous proposons d'étudier le comportement asymptotique, au point de vue de la tendance centrale, de la différence $\frac{\beta}{n} - \frac{\alpha}{m}$ sous certaines conditions de régularité à spécifier.

2. Nous pouvons observer que nous réduisons aisément cette étude à celle de deux fréquences indépendantes.

Car la suite des épreuves $(E_{m+1} \ldots E_n)$ est stochastiquement indépendante de la première suite $(E_1 E_2 \ldots E_m)$.

Désignons par
$$\begin{cases} p = m - n\,, \\ \gamma = \beta - \alpha\,, \end{cases}$$

---

[1]) Le signe $\longleftrightarrow$ désigne la condition nécessaire et suffisante.

la fréquence $\frac{\gamma}{p}$ définie sur la suite $(E_{m+1} \ldots E_n)$ est indépendante de la fréquence $\frac{\alpha}{m}$.

Or on a identiquement:

$$\frac{\beta}{n} - \frac{\alpha}{m} = \frac{\alpha + \varrho}{m + p} - \frac{\alpha}{m} = \frac{\frac{\varrho}{p} - \frac{\alpha}{m}}{\frac{m + p}{p}}$$

et cette transformation permet de ramener la question à l'étude de la différence de deux fréquences indépendantes.

3. Nous considérons les deux séries d'épreuves différentes

$$E_1 E_2 \ldots E_m, \qquad E'_1 E'_2 \ldots E'_p$$

et les fréquences correspondantes $\frac{\alpha}{m}$ et $\frac{\varrho}{p}$.

Remarquons que le problème posé possède une solution immédiate pour $p = m$. Car si on attache à chaque couple d'épreuves de même rang $E_i E'_i$ la variable

$$Z_i = X_i - Y_i,$$

X_i et Y_i étant les indicateurs des épreuves E_i et E'_i, on constate que les variables Z_j qui sont bornées, indépendantes, identiques pour tout i et telles que la fluctuation

$$a^2_{z_i} = 2p_A(1 - p_A)$$

reste invariable par rapport à i; par suite

$$a^2_{\sum_1^m z_i} = 2m\, p_A(1 - p_A) \quad \text{et} \quad \lim_{m\to\infty} a^2_{\sum_1^m z_i} = \infty,$$

et le théorème de Lindberg-Levy s'applique.

On a, dès lors, puisque

$$m_{\sum_1^m z_i} = \sum_{i=1}^n m_{z_i} = \sum_1^n (m_{x_i} - m_{y_i}) = 0,$$

$$\lim_{m\to\infty} \text{prob.}\left\{\alpha \leqq \frac{\sum z_i}{\sqrt{2n\, p_A(1-p_A)}} \leqq \beta\right\} = \int_\alpha^\beta \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$$= \lim_{m\to\infty} \text{prob.}\left\{\alpha \leqq \frac{\frac{\sum z_i}{m}}{\sqrt{\frac{2p_A(1-p_A)}{m}}} \leqq \beta\right\} = \int_\alpha^\beta \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx,$$

or

$$\frac{\sum z_i}{m} = \frac{\varrho}{m} - \frac{\alpha}{m} = 2\left[\frac{\beta}{2m} - \frac{\alpha}{m}\right].$$

D'où le résultat final

$$\lim_{m\to\infty} \text{prob.}\left\{\alpha \leqq \frac{\frac{\beta}{2m} - \frac{\alpha}{m}}{\sqrt{\frac{p_A(1-p_A)}{2m}}} \leqq \beta\right\} = \int_\alpha^\beta \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx,$$

et le résultat asymptotique est atteint.

4. Si les deux séries d'épreuves n'ont pas le même nombre d'expériences, c'est-à-dire $p \neq m$, la méthode précédente tombe en défaut, parce que les variables z_i ne peuvent plus être définies, sur l'ensemble des expériences effectuées. Pour rétablir cette possibilité nous utilisons la méthode des épreuves fictives.

Supposons pour fixer les idées que: $p < m$.

Nous complétons la série $E'_1 \ldots E'_p$ par les épreuves fictives $E'_{p+1} \ldots E'_m$ et à chacune de ces épreuves nous adjoignons la variable U_i, $j = p+1, \ldots, m$, qui ne prend qu'une seule valeur 0 avec la probabilité 1 (c'est l'échelon d'Heaviside d'abscisse nulle).

Nous admettrons de plus comme condition de régularité que m et p sont du même ordre de grandeur, c'est-à-dire pour p tendant vers l'infini

$$\lim_{\substack{p\to\infty \\ m>p}} \frac{m}{p} = k.$$

Cette condition de régularité n'est pas la plus large possible, mais elle suffit.

Cette fois-ci nous attachons à chaque couple d'épreuves de même rang

$$\begin{cases} \text{pour} \quad 0 \leqq i \leqq p & Z_i = \dfrac{m}{p} X_i - Y_i , \\ \text{pour} \quad p+1 \leqq j \leqq m & Z_j = U_j - Y_j . \end{cases}$$

Nous constatons que pour tout i ou j les variables z_i sont bornées, indépendantes, mais elles ne sont plus identiques.

En particulier:

$$\text{pour} \quad 0 \leqq i \leqq p \qquad m_{z_i} = \left(\frac{m}{p} - 1\right) p_A ,$$

$$a^2_{z_i} = p_A(1 - p_A) \left[\frac{m^2}{p^2} + 1\right] ;$$

$$\text{pour} \quad p+1 \leqq j \leqq m \qquad m_{z_j} = - p_A ,$$

$$a^2_{z_i} = p_A(1 - p_A) .$$

On en déduit pour:

$$s_m = \sum_{i=1}^{p} z_i + \sum_{j=p}^{m} z_j ,$$

$$m_{s_m} = p_A \left[\sum_{1}^{p} \left(\frac{m}{p} - 1\right) - \sum_{p+1}^{m} 1\right] = 0 ,$$

$$a^2_{s_m} = p_A(1 - p_A) \left[p \left[\frac{m^2}{p^2} + 1\right] + (m - p)\right] = m \left[\frac{m}{p} + 1\right] p_A(1 - p_A) .$$

Il en résulte:

$$\lim_{\substack{p \to \infty \\ m > p}} a^2_{s_m} = \lim_{p \to \infty} p\, k [k+1]\, p_A(1 - p_A) = \infty$$

et le théorème de Levy-Lindberg est d'application; il peut être mis sous la forme

$$\lim_{\substack{p \to \infty \\ m > p}} \text{prob.} \left\{ a \leqq \frac{\dfrac{\sum z_i}{m}}{\sqrt{p_A(1 - p_A)\left[\dfrac{1}{m} + \dfrac{1}{p}\right]}} \leqq b \right\} = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx .$$

Mais on a identiquement

$$\frac{\sum_{i=1}^{m} z_i}{m} = \frac{\sum_{i=1}^{p}\left(\frac{m}{p} x_i - y_i\right)}{m} + \frac{\sum_{p+1}^{m}(-y_j)}{m} = \frac{\sum_{i=1}^{p} x_i}{p} - \frac{\sum_{j=1}^{m} y_j}{m} = \frac{\varrho}{p} - \frac{\alpha}{m}$$

et le problème du comportement asymptotique de deux fréquences calculées sur des séries d'épreuves indépendantes est terminé, sous les conditions de régularité que nous reprenons.

1º Les nombres des épreuves de deux séries restent tels que p reste invariablement plus petit que m.

2º Pour p suivant vers l'infini, $m = 0(p)$, c'est-à-dire m reste du même ordre de grandeur de p.

5. Revenons pour finir au problème initial posé.

En vertu de la transformation

$$\frac{\sum z_i}{m} = \frac{\gamma}{p} - \frac{\alpha}{m} = \sum = \left(\frac{\beta}{m} - \frac{\alpha}{m}\right)\frac{n}{n-m}$$

on a

$$\frac{\frac{\sum z_i}{m}}{\sqrt{p_A(1-p_A)\left(\frac{1}{m}+\frac{1}{p}\right)}} = \frac{\frac{\beta}{n} - \frac{\alpha}{m}}{\sqrt{p_A(1-p_A)\left[\frac{1}{m}+\frac{1}{n-m}\right]\left[\frac{n-m}{n}\right]^2}} =$$

$$= \frac{\frac{\beta}{n} - \frac{\alpha}{m}}{\sqrt{p_A(1-p_A)\left[\frac{1}{m} - \frac{1}{n}\right]}},$$

et on obtient le résultat

$$\lim_{\substack{n>m \\ n-m\to\infty}} \text{prob.}\left\{a < \frac{\frac{\beta}{n} - \frac{\alpha}{m}}{\sqrt{p_A(1-p_A)\left[\frac{1}{m} - \frac{1}{n}\right]}} < b\right\} = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

$$\lim_{n\to\infty} \frac{n}{m} \equiv k$$

B. Généralisation

1. L'examen de la démonstration révèle immédiatement que la méthode est susceptible de généralisation:

a) Les X_i et Y_j sont des indicateurs, les variables U_j presque identiquement nulles; la propriété commune est d'être bornée, on généralisera en prenant des variables bornées dans leur ensemble.

b) Pour arriver à la condition $m_{s_n} = 0$, on a fait appel à la propriété que le moment $m_{x_i} = m_{y_j} = p_A$, mais cette condition peut s'obtenir tout aussi simplement, en ne considérant que des variables dont les moments de 1er ordre sont nuls.

On considère dès lors un ensemble de variables $\{x_i\}$ telles

1° les moments du 1er ordre $m_{x_i} = 0$ quel que soit i,

2° les variables sont uniformément bornées $|x_i| \leq L$ quel que soit i.

On effectue une série d'épreuves $E_1 \ldots E_m \ldots E_n$, l'épreuve E_i étant relative à l'arrivée de la variable X_i.

On se propose d'étudier le comportement asymptotique de la différence

$$\Delta = \frac{\sum\limits_1^n x_i}{n} - \frac{\sum\limits_{j=1}^m x_j}{m}.$$

2. Remarquons que, tout comme dans le premier cas, le problème peut être ramené au cas de la différence de deux moyennes calculées sur deux séries d'épreuves différentes et indépendantes car

$$\frac{\sum\limits_{j=1}^n x_i}{n} - \frac{\sum\limits_{j=1}^m x_j}{m} = \frac{\sum\limits_{i=1}^m x_i + \sum\limits_{k=m+1}^n x_k}{m+p} - \frac{\sum\limits_{j=1}^m x_j}{m} = \frac{m\sum\limits_{k=m+1}^n x_k - p\sum\limits_{j=1}^m x_j}{m(m+p)} =$$

$$= \frac{\dfrac{\sum\limits_{k=m+1}^n x_k}{p} - \dfrac{\sum\limits_{j=1}^m x_j}{m}}{\dfrac{m+p}{p}}.$$

3. Dès lors on considère les deux séries d'épreuves indépendantes

$E_1\, E_2 \ldots E_j \ldots E_m$ à l'épreuve E_j est rattachée la variable X_j,

$E'_1\, E'_2 \ldots E'_k \ldots E'_p$ à l'épreuve E'_k est rattachée la variable Y_k, et nous supposons $p < m$ avec $\lim\limits_{p \to \infty} \dfrac{m}{p} = k$,

et tout comme dans le premier cas nous posons:

$$0 \leqq j \leqq p \qquad z_j = \frac{m}{p} x_j - y_i ,$$

$$p+1 \leqq k \leqq m \qquad z_k = u_k - y_j .$$

On en déduit

$$m_{z_j} = m_{z_k} = 0 ,$$

quels que soient j et k

$$a^2_{z_i} = \frac{m^2}{p^2} a^2_{x_j} + a^2_{y_j} , \qquad 1 \leqq j \leqq p ,$$

$$a^2_{z_k} = a^2_{y_k} , \qquad p+1 \leqq k \leqq m ,$$

et ensuite pour

$$s_m = \sum_{j=1}^{p} z_j + \sum_{k=p+1}^{m} z_k ,$$

$$m_{s_m} = 0 ,$$

$$a^2_{s_m} = \frac{m^2}{p^2} \sum_{j=1}^{p} a^2_{x_j} + \sum_{j=1}^{m} a^2_{y_j} ,$$

et

$$\lim_{p \to \infty} a^2_{s_m} = \infty ,$$

dès lors le théorème de Lindberg-Levy s'applique, mais on a identiquement

$$\frac{s_m}{m} = \frac{\sum\limits_{j=1}^{p} x_j}{p} - \frac{\sum\limits_{j=1}^{m} y_j}{m} , \qquad a^2_{\frac{s_m}{m}} = \frac{\sum\limits_{j=1}^{p} a^2_{x_j}}{p^2} + \frac{\sum\limits_{j=1}^{m} a^2_{y_j}}{m^2}$$

et

$$\lim_{\substack{p \to \infty \\ \lim\limits_{p \to \infty} \frac{m}{p} = k}} \text{prob.} \left\{ a \leqq \frac{\dfrac{\sum\limits_{j=1}^{p} x_j}{p} - \dfrac{\sum\limits_{j=1}^{m} y_j}{m}}{\sqrt{\dfrac{\sum\limits_{j=1}^{p} a^2_{x_j}}{p^2} - \dfrac{\sum\limits_{j=1}^{m} a^2_{y_j}}{m^2}}} \leqq b \right\} = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx .$$

4. Revenant au problème posé, on arrive après transformation, la variable x_i étant rapportée à l'épreuve E_i

$$\lim_{p\to\infty} \text{prob.} \left\{ a \leqq \frac{\dfrac{\sum_{i=1}^{n} x_i}{n} - \dfrac{\sum_{j=1}^{m} x_j}{m}}{\sqrt{\dfrac{m^2 \sum_{k=m+1}^{n} a_{x_k}^2 + (n-m)^2 \sum_{j=1}^{m} a_{x_j}^2}{m^2 n^2}}} \leqq b \right\} = \int \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$$\lim_{p\to\infty} \frac{m}{p} = b$$

on remarquera que l'expression sous le signe radical peut s'écrire:

$$\left(\frac{1}{m} - \frac{1}{n}\right) \left(\frac{m}{n} \frac{\sum_{k=m+1}^{n} a_{x_k}^2}{m-n} + \frac{n-m}{n} \frac{\sum_{i=1}^{m} a_{x_i}^2}{m} \right)$$

ce qui permet de retrouver une forme semblable à celle donnée dans le premier cas et en particulier si à chaque épreuve, on considère la même variable de fluctuation a^2, on aura

$$\left(\frac{1}{m} - \frac{1}{n}\right) a^2.$$

Zusammenfassung

Der Artikel zeigt, dass für wachsendes m und n die Differenz der Häufigkeiten $\frac{\alpha}{m}$ und $\frac{\beta}{n}$, welche in zwei abhängigen Folgen beobachtet werden, gegen eine Gaußsche Variable strebt.

Das Resultat wird dann verallgemeinert auf das Mittel von allgemeinen Beobachtungen gleichmässig beschränkter Variabeln.

Riassunto

Si dimostra che, assunte qualchi condizioni di regularità, la differenza fra le frequenze $\frac{\alpha}{m}$ e $\frac{\beta}{n}$ osservate su due serie dipendenti ubidisce alla convergenza centrale.

Il resultato è poi generalizzato per la media di fenomeni basati su variabili uniformamente limitate.

Summary

In this paper the author shows the central convergence to hold true also for the difference of frequencies $\frac{\alpha}{m}$ and $\frac{\beta}{n}$ observed on two dependent series provided that they satisfy certain regularity conditions.

The result is then extended to the mean of more general observations on uniformly bounded variables.