MATHÉMATIQUES ET SCIENCES HUMAINES

H. ROUANET

Note pédagogique. Un théorème de convergence probabiliste et son application à la justification de méthodes statistiques courantes

Mathématiques et sciences humaines, tome 42 (1973), p. 41-47
<http://www.numdam.org/item?id=MSH_1973__42__41_0>

Math. Sci. hum. (11e année, n° 42, 1973, p. 41-47)

NOTE PÉDAGOGIQUE

UN THÉORÈME DE CONVERGENCE PROBABILISTE ET SON APPLICATION A LA JUSTIFICATION DE MÉTHODES STATISTIQUES COURANTES

par
H. ROUANET [1]

RÉSUMÉ

Dans cette note on présente un théorème de convergence probabiliste relativement élémentaire et son application à la justification des méthodes statistiques suivantes (grands échantillons) : inférence sur une moyenne, inférence sur la différence de deux moyennes (groupes indépendants), comparaison partielle de deux fréquences, comparaison de deux fréquences (groupes appareillés et groupes indépendants).

SUMMARY

In this note we present a relatively elementary theorem of probabilistic convergence and its application to the justification of the following statistical methods (large samples) : inference on a mean, inference on the difference of two means (independent groups), partial comparison of two frequencies, and comparison of two frequencies (matched groups and independent groups).

1. *Introduction*

L'une des exigences d'un enseignement sérieux d'une discipline scientifique est de fournir, pour les méthodes présentées, des justifications adaptées au niveau des enseignés. Cette exigence est particulièrement critique en statistique lorsque l'enseignant s'adresse à un public de niveau mathématique peu élevé ; l'enseignant risque alors d'osciller entre deux écueils : donner de pures « recettes » sans justification aucune, ou énoncer des justifications au delà du niveau des auditeurs ; si pour l'enseignant, la deuxième méthode peut apporter une certaine sécurité, pour l'enseigné elle revient sans doute au même que la première.

Pour fixer les idées, donnons l'exemple qui est à l'origine de cette note. Il s'agit des tests usuels du χ^2 dans les cas courants (comparaison de fréquences pour deux groupes indépendants ou deux groupes appareillés, etc.). Dans les ouvrages de statistique élémentaire, ces tests ne sont pas réellement justifiés. Dans les traités de statistique mathématique, on trouve des résultats très généraux, à partir desquels on peut retrouver les tests usuels comme cas particuliers, le résultat fondamental étant le suivant (théorème de Fisher sur l'ajustement à une distribution paramétrique) : « Sous des conditions peu restrictives, la statistique d'ajustement Q^2 suit asymptotiquement, sous l'hypothèse nulle, la distribution d'un χ^2 à ν degrés de liberté, avec $\nu = L - 1 - S$, où L est le nombre de classes et S le

1. U.E.R. de Mathématique, Logique formelle et Informatique, Université René-Descartes, Paris.

nombre de paramètres estimés à partir des observations.» (On trouvera une démonstration de ce théorème dans Cramér, p. 426, ou encore dans Rao.) Les « conditions peu restrictives» concernent les fonctions qui expriment les fréquences parentes des classes en fonction des paramètres ; elles présentent un caractère technique assez rebutant pour un non-mathématicien.

C'est pourquoi il nous paraît utile de disposer également de justifications plus directes des tests courants. Recherchant une telle justification, nous avons trouvé un théorème de convergence probabiliste relativement élémentaire qui permet de justifier diverses méthodes statistiques courantes (pour de grands échantillons) lesquelles ne se limitent d'ailleurs pas à des tests du χ^2. Dans cette note, nous présenterons ce théorème ainsi que les justifications d'un certain nombre de ces méthodes statistiques courantes.

2. *Un théorème de convergence probabiliste ; conséquences immédiates*

Nous énoncerons le théorème fondamental de cette note sous la forme du lemme suivant :

Lemme : Soit (ξ_n) $n = 1,2...$ une suite de v.a. numériques convergeant (en distribution) [2] vers une distribution donnée ; soit η_n une autre suite de v.a. numériques convergeant en probabilité vers 1 ; alors la suite des v.a. $\frac{\xi_n}{\eta_n}$ converge en distribution vers la distribution limite de ξ_n.

On trouvera une démonstration de ce lemme dans Cramér, p. 254. L'enseignant de statistique pourra juger utile de présenter aux étudiants cette démonstration (qui est un exercice d'analyse) ; sinon il se trouvera tout de même dans une situation plus favorable qu'avec le théorème de Fisher énoncé plus haut : 1) l'énoncé du lemme ne nécessite pas de « restrictions mentales» comme celles évoquées par la formule « sous des conditions peu restrictives» ; 2) il est tout à fait intuitif pour qui a appréhendé même de façon rudimentaire les deux notions de convergence en distribution et de convergence en probabilité.

Du lemme précédent nous déduirons deux théorèmes que nous utiliserons directement dans les applications. Du point de vue des notations : nous représenterons la convergence en distribution par le symbole $\underset{\rightarrow}{\simeq}$. Ainsi si U_n est une suite de v.a. convergeant vers une distribution normale réduite nous écrirons $U_n \underset{\rightarrow}{\simeq} \mathrm{N}\ (0,1)$; la suite des carrés convergeant vers un χ^2 à 1 d.l. nous écrirons $U^2_n \underset{\rightarrow}{\simeq} \chi^2_{[1]}$. Enfin, si X_n est une suite de v.a. d'espérances $\mathrm{E}(X_n)$ et de variances $\sigma^2\ (X_n)$ telles que la suite des v.a. réduites associées converge vers une distribution normale réduite nous écrirons $X_n \simeq \mathrm{N}(\mathrm{E}(X_n), \sigma^2\ (X_n))$.

Théorème 1 : Soient X_n et Z_n deux suites de v.a. telles que

$$X_n \simeq \mathrm{N}\ (0, \sigma^2\ (X_n)) \text{ et } \frac{Z_n}{\sigma\ (X_n)} \xrightarrow{\text{Prob}} 1 ; \quad \text{alors } \frac{X_n}{Z_n} \underset{\rightarrow}{\simeq} \mathrm{N}(0,1).$$

En effet, posons $\xi_n = \frac{X_n}{\sigma\ (X_n)}$; par hypothèse, $\xi_n \underset{\rightarrow}{\simeq} \mathrm{N}(0,1)$. Posons $\eta_n = \frac{Z_n}{\sigma\ (X_n)}$; par hypothèse $\eta_n \xrightarrow{\text{Prob}} 1$. D'après le lemme, $\frac{\xi_n}{\eta_n}$ c'est-à-dire $\frac{X_n}{Z_n} \underset{\rightarrow}{\simeq} \mathrm{N}(0,1)$.

2. Ou : « en loi».

Théorème 2 : Soient X_n et Y_n deux suites de v.a. telles que

$$X_n \simeq \mathrm{N}(0, \sigma^2(X_n)) \text{ et } \frac{Y_n}{\sigma^2(X_n)} \xrightarrow{\text{Prob}} 1 \text{ ; alors } \frac{X^2{}_n}{Y_n} \underset{\to}{\simeq} \chi^2{}_{[1]}.$$

En effet, posons $\xi_n = \dfrac{X^2{}_n}{\sigma^2(X_n)}$; par hypothèse $\xi_n \underset{\to}{\simeq} \chi^2{}_{[1]}$. Posons $\eta_n = \dfrac{Y_n}{\sigma^2(X_n)}$; par hypothèse $\eta_n \xrightarrow{\text{Prob}} 1$. D'après le lemme, $\dfrac{\xi_n}{\eta_n}$, c'est-à-dire $\dfrac{X^2{}_n}{Y_n} \underset{\to}{\simeq} \chi^2{}_{[1]}$.

3. *Application à des méthodes statistiques courantes*

a) Inférence sur une moyenne, grands échantillons

Soit un échantillon de n observations extrait d'une distribution parente ayant une moyenne μ et une variance σ^2, et par ailleurs quelconque (normale ou non). Soit M la moyenne de l'échantillon, S^2 sa variance-corrigée (somme des carrés des écarts à la moyenne, divisée par $n-1$). Pour l'étude de l'inférence sur la moyenne μ, on part de la v.a. $M-\mu$: si on la rapporte à son écart-type $\dfrac{\sigma}{\sqrt{n}}$ on obtient une v.a. réduite asymptotiquement normale (théorème classique de convergence normale) : $\dfrac{M-\mu}{\sigma/\sqrt{n}} \underset{\to}{\simeq} \mathrm{N}(0,1)$.

En remplaçant $\sigma/\sqrt{n}$ par $S/\sqrt{n}$ on obtient la v.a. $\dfrac{M-\mu}{S/\sqrt{n}}$ dont nous allons montrer qu'elle est encore asymptotiquement N(0,1). Posons $M-\mu = X_n$, $S/\sqrt{n} = Z_n$. On a $X_n = M-\mu \simeq \mathrm{N}\left(0, \dfrac{\sigma^2}{n}\right)$. D'autre part, on sait que S est un estimateur convergent de σ, c'est-à-dire $S \xrightarrow{\text{Prob}} \sigma$, d'où $\dfrac{Z_n}{\sigma/\sqrt{n}} = \dfrac{S/\sqrt{n}}{\sigma/\sqrt{n}} \xrightarrow{\text{Prob}} 1$.

Par application du théorème 1 : $\dfrac{M-\mu}{S/\sqrt{n}} \underset{\to}{\simeq} \mathrm{N}(0,1)$. De la distribution asymptotique de cette v.a. on déduit le test de l'hypothèse nulle $\mu = \mu_0$ (comparaison à une norme) pour de grands échantillons : sous H_0, la statistique de test $\dfrac{M-\mu_0}{S/\sqrt{n}} \underset{\to}{\simeq} \mathrm{N}(0,1)$.

b) Inférence sur la différence de deux moyennes, groupes indépendants, grands échantillons

Soient deux échantillons indépendants extraits de distributions parentes de moyennes μ', μ'', de variances σ'^2, σ''^2, et par ailleurs quelconques (normales ou non). Soient n', n'' les tailles des échantillons, M', M'' leurs moyennes, S'^2, S''^2 leurs variances corrigées. Pour l'étude de l'inférence sur la différence $\mu'-\mu''$, on part de la v.a. $M'-M''-(\mu'-\mu'')$: si on la rapporte à son écart-type $\sqrt{\dfrac{\sigma'^2}{n'}+\dfrac{\sigma''^2}{n''}}$, on obtient une v.a. réduite asymptotiquement normale (d'après une extension du théorème de convergence normale) : $\dfrac{M'-M''-(\mu'-\mu'')}{\sqrt{\dfrac{\sigma'^2}{n'}+\dfrac{\sigma''^2}{n''}}} \underset{\to}{\simeq} \mathrm{N}(0,1)$. En remplaçant $\dfrac{\sigma'^2}{n'}+\dfrac{\sigma''^2}{n''}$ par $\dfrac{S'^2}{n'}+\dfrac{S''^2}{n''}$ on obtient la v.a. $\dfrac{M'-M''-(\mu'-\mu'')}{\sqrt{\dfrac{S'^2}{n'}+\dfrac{S''^2}{n''}}}$ dont nous allons montrer qu'elle est encore asymptotiquement N(0,1).

Posons $M' - M'' - (\mu' - \mu'') = X_n$, $\sqrt{\dfrac{S'^2}{n'} + \dfrac{S''^2}{n''}} = Z_n$.

On a $X_n = M' - M'' - (\mu' - \mu'') \underset{\rightarrow}{\simeq} N\left(0, \dfrac{\sigma'^2}{n'} + \dfrac{\sigma''^2}{n''}\right)$.

Posons d'autre part $n' = k'\, n$, $n'' = k''\, n$; alors $\dfrac{\dfrac{S'^2}{n'} + \dfrac{S''^2}{n''}}{\dfrac{\sigma'^2}{n'} + \dfrac{\sigma''^2}{n''}} = \dfrac{\dfrac{S'^2}{k'} + \dfrac{S''^2}{k''}}{\dfrac{\sigma'^2}{k'} + \dfrac{\sigma''^2}{k''}}$.

Supposons que les facteurs k' et k'' restant fixes, $n \longrightarrow \infty$: S'^2 et S''^2 sont des estimateurs convergents de σ'^2 et σ''^2. Donc d'après les propriétés élémentaires de la convergence en probabilité :

$\dfrac{S'^2}{k'} + \dfrac{S''^2}{k''} \xrightarrow{\text{Prob}} \dfrac{\sigma'^2}{k'} + \dfrac{\sigma''^2}{k''}$, d'où

$\dfrac{\dfrac{S'^2}{k'} + \dfrac{S''^2}{k''}}{\dfrac{\sigma'^2}{k'} + \dfrac{\sigma''^2}{k''}} \xrightarrow{\text{Prob}} 1$, d'où $\dfrac{\sqrt{\dfrac{S'^2}{k'} + \dfrac{S''^2}{k''}}}{\sqrt{\dfrac{\sigma'^2}{k'} + \dfrac{\sigma''^2}{k''}}} \xrightarrow{\text{Prob}} 1$, c'est-à-dire

$\dfrac{Z_n}{\sqrt{\dfrac{\sigma'^2}{n'} + \dfrac{\sigma''^2}{n''}}} \xrightarrow{\text{Prob}} 1$. Par application du théorème 1 : $\dfrac{M' - M'' - (\mu' - \mu'')}{\sqrt{\dfrac{S'^2}{n'} + \dfrac{S''^2}{n''}}} \underset{\rightarrow}{\simeq} N(0,1)$. De la distribution asymptotique de cette v.a. on déduit le test de l'hypothèse nulle H_0 : $\mu' = \mu''$ pour de grands échantillons : sous H_0 : la statistique de test $\dfrac{M' - M''}{\sqrt{\dfrac{S'^2}{n'} + \dfrac{S''^2}{n''}}} \underset{\rightarrow}{\simeq} N(0,1)$.

c) Comparaison partielle de deux fréquences

Soit un échantillon de n observations à valeurs dans un ensemble de L classes et extrait d'une distribution de fréquences parentes $\varphi_1, \varphi_2, \ldots, \varphi_L$. Soient $N_1, N_2, \ldots, N_L$ les effectifs, $F_1, F_2, \ldots, F_L$ les fréquences des L classes. Cherchons à éprouver l'hypothèse nulle H_0 : $\varphi_1 = \varphi_2$ (comparaison partielle de deux fréquences). On est conduit à considérer la statistique $F_1 - F_2$, différence des fréquences dans l'échantillon. La distribution de $(F_1, F_2, \ldots, F_L)$ est multinomiale, d'où on déduit $E(F_1) = \varphi_1$, $E(F_2) = \varphi_2$, $\text{Var}(F_1) = \dfrac{\varphi_1\,(1 - \varphi_1)}{n}$, $\text{Var}(F_2) = \dfrac{\varphi_2\,(1 - \varphi_2)}{n}$, $\text{Cov}(F_1, F_2) = -\dfrac{\varphi_1\;\varphi_2}{n}$, d'où

$$\text{Var}\,(F_1 - F_2) = \frac{\varphi_1 - \varphi_1^2 + \varphi_2 - \varphi_2^2 + 2\,\varphi_1\,\varphi_2}{n} = \frac{\varphi_1 + \varphi_2 - (\varphi_1 - \varphi_2)^2}{n}.$$

Sous H_0, $\text{Var}\,(F_1 - F_2) = \dfrac{\varphi_1 + \varphi_2}{n}$, et asymptotiquement $F_1 - F_2 \simeq N\left(0, \dfrac{\varphi_1 + \varphi_2}{n}\right)$.

Si on rapporte $(F_1 - F_2)^2$ à sa variance, on obtient une v.a. distribuée asymptotiquement comme un χ^2 à 1 d.l. : $\dfrac{(F_1 - F_2)^2}{(\varphi_1 + \varphi_2)/n} \underset{\rightarrow}{\simeq} \chi^2_{[1]}$. Remplaçons maintenant $\dfrac{\varphi_1 + \varphi_2}{n}$ par $\dfrac{F_1 + F_2}{n}$ dans cette v.a.: on obtient la statistique $Q^2 = \dfrac{(F_1 - F_2)^2}{(F_1 + F_2)/n}$ qu'on peut encore écrire $Q^2 = \dfrac{(N_1 - N_2)^2}{N_1 + N_2}$. Nous allons montrer que Q^2 est encore asymptotiquement $\chi^2_{[1]}$. Posons $F_1 - F_2 = X_n$, $\dfrac{F_1 + F_2}{n} = Y_n$.

Sous H_0, $X_n = F_1 - F_2 \simeq N\left(0, \dfrac{\varphi_1 + \varphi_2}{n}\right)$ et $\dfrac{Y_n}{\dfrac{\varphi_1 + \varphi_2}{n}} = \dfrac{\dfrac{F_1 + F_2}{n}}{\dfrac{\varphi_1 + \varphi_2}{n}} \xrightarrow{\text{Prob}} 1$. Donc par application du théorème 2 : $\dfrac{X^2_n}{Y_n} = \dfrac{(F_1 - F_2)^2}{(F_1 + F_2)/n} \underset{\rightarrow}{\sim} \chi^2_{[1]}$, ou encore $Q^2 = \dfrac{(N_1 - N_2)^2}{N_1 + N_2} \underset{\rightarrow}{\sim} \chi^2_{[1]}$.

On prendra donc $Q^2 = \dfrac{(N_1 - N_2)^2}{N_1 + N_2}$ pour statistique de test.

d) Comparaison de deux fréquences, groupes appareillés (test de McNemar)

Soit un échantillon de n observations à valeurs dans le carré d'un ensemble binaire $\{a_1, a_2\}$, les fréquences parentes étant données par le tableau :

	a_1	a_2	
a_1	φ_{11}	φ_{12}	φ'_1
a_2	φ_{21}	φ_{22}	φ'_2
	φ''_1	φ''_2	

et les effectifs correspondants par le tableau :

	a_1	a_2	
a_1	N_{11}	N_{12}	N'_1
a_2	N_{21}	N_{22}	N'_2
	N''_1	N''_2	n

Soit à tester l'hypothèse nulle H_0 : $\varphi'_1 = \varphi''_1$ (égalité des fréquences marginales correspondant aux deux groupes appareillés). Cette hypothèse est équivalente à l'hypothèse H'_0 : $\varphi_{12} = \varphi_{21}$. On est ainsi ramené au problème c précédent : comparaison partielle des deux fréquences φ_{12} et φ_{21}, d'où la statistique de test $\dfrac{(N_{12}-N_{21})^2}{N_{12}+N_{21}} \underset{\rightarrow}{\sim} \chi^2[1]$ sous H_0.

e) Comparaison de deux fréquences, groupes indépendants

Soient deux groupes d'observations considérés comme des échantillons indépendants à valeurs dans le même ensemble binaire $\{a_1, a_2\}$, les fréquences parentes de a_1 étant respectivement φ'_1, φ''_1 (et les fréquences complémentaires étant $\varphi_2' = 1 - \varphi'_1$, $\varphi''_2 = 1 - \varphi''_1$). Soient n' et n'' les tailles des deux groupes, N'_1, N'_2 et N''_1, N''_2 les effectifs correspondant à a_1 et a_2 dans les deux groupes, conformément au tableau :

	a_1	a_2	
groupe 1	N'_1	N'_2	n'
groupe 2	N''_1	N''_2	n''
	N_1	N_2	n

(on pose $N'_1 + N''_1 = N_1$, $N'_2 + N''_2 = N_2$, $n' + n'' = n$).

De plus on posera $F'_1 = \dfrac{N'_1}{n'}$, $F'_2 = \dfrac{N'_2}{n'}$, $F''_1 = \dfrac{N''_1}{n''}$, $F''_2 = \dfrac{N''_2}{n''}$.

Soit à tester l'hypothèse nulle d'égalité des fréquences parentes $H_0 : \varphi'_1 = \varphi''_1$. On est conduit à considérer la statistique $F'_1 - F''_1$, différence des fréquences de a_1 dans les deux groupes. D'après les propriétés de la distribution binomiale, $E(F'_1) = \varphi'_1$, $E(F''_1) = \varphi''_1$ et

$$\operatorname{var}(F'_1) = \frac{\varphi'_1\ (1 - \varphi'_1)}{n'}, \qquad \operatorname{var}(F''_1) = \frac{\varphi''_1\ (1 - \varphi''_1)}{n''},$$

d'où (les v.a. F'_1 et F''_1 étant indépendantes)

$$\operatorname{var}(F'_1 - F''_1) = \frac{\varphi'_1\ (1 - \varphi'_1)}{n'} + \frac{\varphi''_1\ (1 - \varphi''_1)}{n''}.$$

Sous H_0, $F'_1 - F''_1 \simeq N\left(0, \varphi_1\ (1 - \varphi_1)\left(\dfrac{1}{n'} + \dfrac{1}{n''}\right)\right)$ (en appelant φ_1 la valeur commune de φ'_1 et de φ''_1). Si on rapporte $(F_1' - F''_1)^2$ à sa variance on obtient une v.a. distribuée asymptotiquement comme un χ^2 à 1 d.l. : $\dfrac{(F'_1 - F''_1)^2}{\varphi_1\ (1 - \varphi_1)\left(\dfrac{1}{n'} + \dfrac{1}{n''}\right)} \underset{\rightarrow}{\sim} \chi^2[1]$. Sous H_0, on a un échantillon de taille n avec les effectifs N_1 et N_2 : φ_1 peut donc être estimé par $\dfrac{N_1}{n}$ qu'on posera égal à F_1 et $1 - \varphi_1$ par $\dfrac{N_2}{n}$ qu'on posera égal à F_2, d'où l'estimation de $\varphi_1\ (1 - \varphi_1)$: F_1F_2. Remplaçant φ_1 et $1 - \varphi_1$ par leurs estimations dans la v.a. précédente, on obtient la statistique $Q^2 = \dfrac{(F'_1 - F''_2)^2}{F_1 F_2 \left(\dfrac{1}{n'} + \dfrac{1}{n''}\right)}$, laquelle peut être mise sous la forme $Q^2 = \dfrac{n\,(N'_1\,N''_2 - N'_2\,N''_1)^2}{n'\,n''\,N_1\,N_2}$ (sous cette dernière forme on reconnaît la statistique classique Q^2 associée à un tableau d'effectifs 2×2). Montrons que sous H_0, Q^2 est asymptotiquement $\chi^2_{[1]}$. Posons $F_1 - F_2 = X_n$, $F_1\,F_2\left(\dfrac{1}{n'} + \dfrac{1}{n''}\right) = Y_n$. Sous H_0, $X_n = F_1 - F_2 \simeq N\left(0,\ \varphi_1\,(1 - \varphi_1)\left(\dfrac{1}{n'} + \dfrac{1}{n''}\right)\right)$ et $\dfrac{Y_n}{\varphi_1\,(1 - \varphi_1)\left(\dfrac{1}{n'} + \dfrac{1}{n''}\right)} = \dfrac{F_1\,(1 - F_1)}{\varphi_1\,(1 - \varphi_1)} \xrightarrow{\text{Prob}} 1$. Donc par application du théorème 2, $\dfrac{X^2_n}{Y_n} \underset{\rightarrow}{\simeq} \chi^2_{[1]}$, c'est-à-dire sous H_0 : $Q^2 \underset{\rightarrow}{\simeq} \chi^2_{[1]}$. On prendra donc Q^2 pour statistique de test.

RÉFÉRENCE

CRAMÉR, H., *Mathematical methods of statistics*, Princeton, NJ, Princeton University Press, 1946.