MATHÉMATIQUES ET SCIENCES HUMAINES

G. TH. GUILBAUD

# Exercices de calcul pour préparer à l'usage raisonnable du khi deux

*Mathématiques et sciences humaines*, tome 14 (1966), p. 31-40

<http://www.numdam.org/item?id=MSH_1966__14__31_0>

G. Th. GUILBAUD

# EXERCICES DE CALCUL
# POUR PREPARER A L'USAGE RAISONNABLE DU KHI DEUX

## 1. - OBSERVATION :

L'observation de 17 candidats (classés par le centre de sélection en trois catégories A, B ou C, et par le centre d'apprentissage en deux catégories X, Y) est résumée dans le tableau :

| | C | B | A | |
|---|---|---|---|---|
| X | 0 | 3 | 2 | 5 |
| Y | 6 | 5 | 1 | 12 |
| | 6 | 8 | 3 | 17 |

Un tel tableau extrait de: G.H. FREEMAN et J.H. HALTON, (Exact treatment of contingency and goodness of fit, Biometrika, Vol. 38, june 1951, pp. 143-145), est ici présenté à titre d'exemple d'une situation statistique tout à fait ordinaire.

## 2. - L'UNIVERS DES POSSIBLES

On établit d'abord la liste des tableaux possibles ayant mêmes marges; chacun d'eux sera désigné par un numéro.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (01) | 0 | 2 | 3 | (07) | 1 | 3 | 1 | (13) | 3 | 0 | 2 | (19) | 5 | 1 | 0 |
| | 6 | 6 | 0 | | 5 | 5 | 2 | | 3 | 8 | 1 | | 1 | 8 | 3 |
| (02) | 0 | 3 | 2 | (08) | 1 | 4 | 0 | (14) | 3 | 1 | 1 | (20) | 5 | 2 | 0 |
| | 6 | 5 | 1 | | 5 | 4 | 3 | | 3 | 7 | 2 | | 1 | 7 | 4 |
| (03) | 0 | 4 | 1 | (09) | 2 | 0 | 3 | (15) | 3 | 2 | 0 | (21) | 5 | 3 | 0 |
| | 6 | 4 | 2 | | 4 | 8 | 0 | | 3 | 6 | 3 | | 1 | 6 | 5 |
| (04) | 0 | 5 | 0 | (10) | 2 | 1 | 2 | (16) | 4 | 0 | 1 | (22) | 5 | 4 | 0 |
| | 6 | 3 | 3 | | 4 | 7 | 1 | | 2 | 8 | 2 | | 1 | 5 | 6 |
| (05) | 1 | 1 | 3 | (11) | 2 | 2 | 1 | (17) | 4 | 1 | 0 | | | | |
| | 5 | 7 | 0 | | 4 | 6 | 2 | | 2 | 7 | 3 | | | | |
| (06) | 1 | 2 | 2 | (12) | 2 | 3 | 0 | (18) | 5 | 0 | 0 | | | | |
| | 5 | 6 | 1 | | 4 | 5 | 3 | | 1 | 8 | 3 | | | | |

Le dispositif peut être représenté en plan (deux libertés) :

```
          01 .. 02 .. 03 .. 04
       05 .. 06 .. 07 .. 08
    09 .. 10 .. 11 .. 12
       13 .. 14 .. 15
          16 .. 17
             18
```

(tous les alignements: 01, 06, 11, 15 ou bien 05, 10, 14, 17 etc... sont significatifs - on voit aisément ce que les tableaux d'une pareille suite ont en commun).

## 3. - PROBABILISATION :

Dans l'hypothèse d'indépendance des deux classifications, il suffit de dénombrer les réalisations de chaque tableau.

Par exemple pour le cas observé n° (02): $\begin{matrix} 0 & 3 & 2 \\ 6 & 5 & 1 \end{matrix}$

le nombre des réalisations sera :

$$6!\ 8!\ 3!\ :\ 0!\ 3!\ 2!\ 6!\ 5!\ 1!$$

Quant au nombre total des réalisations, il est :

$$17!\ :\ 5!\ 12!$$

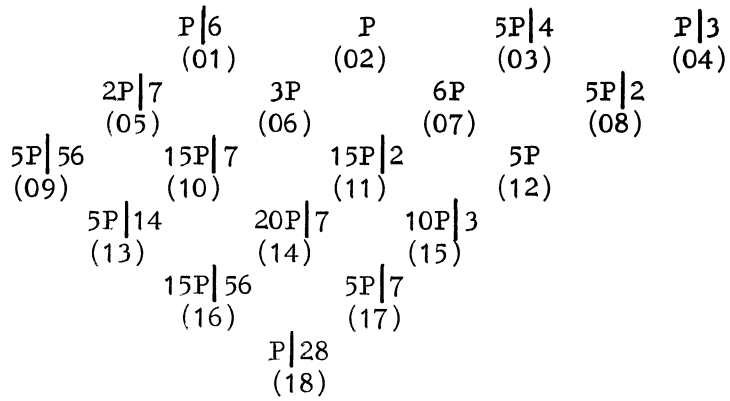
d'où la probabilité du tableau n° (02) :

$$P = 5!\ 12!\ 6!\ 8!\ 3!\ :\ 17!\ 0!\ 3!\ 2!\ 6!\ 5!\ 1!$$
$$= 12!\ 8!\ :\ 17!\ 2! = 6 : 221 = 0{,}027\ 149\ 3\ \ldots$$

Il n'est pas nécessaire de refaire le calcul pour les 17 autres cas: le calcul du rapport de deux probabilités est plus facile :

$$P_7 : P = 0!\ 3!\ 2!\ 6!\ 5!\ 1!\ :\ 1!\ 1!\ 3!\ 1!\ 5!\ 5!\ 2!$$
$$= 0!\ 2!\ 6!\ 1!\ :\ 1!\ 1!\ 5!\ 2! = 2.6 : 1.2 = 6$$

et ainsi de suite de proche en proche (car les deux expressions factorielles des probabilités de deux tableaux voisins ont beaucoup de facteurs communs).



En désignant par P la probabilité du cas n° 2 on obtient les résultats ci-dessus. On peut facilement calculer toutes les probabilités, et même les ranger d'abord en ordre croissant.

Liste des Probabilités :

| | |
|---|---|
| (01) | 0,0 0 4 5 |
| (02) | 0 2 7 1 |
| (03) | 0 3 3 9 |
| (04) | 0 0 9 0 |
| (05) | 0 0 7 8 |
| (06) | 0 8 1 5 |
| (07) | 1 6 2 9 |
| (08) | 0 6 7 9 |
| (09) | 0 0 2 4 |
| (10) | 0 5 8 2 |
| (11) | 2 0 3 7 |
| (12) | 1 3 5 7 |
| (13) | 0 0 9 7 |
| (14) | 0 7 7 6 |
| (15) | 0 9 0 5 |
| (16) | 0 0 7 3 |
| (17) | 0 1 9 4 |
| (18) | 0 0 1 0 |
| | 1,0 0 0 1 |

EN ORDRE CROISSANT :

| | |
|---|---|
| (18) | 0,0 0 1 0 |
| (9) | 0 0 2 4 |
| (1) | 0 0 4 5 |
| (16) | 0 0 7 3 |
| (5) | 0 0 7 8 |
| (4) | 0 0 9 0 |
| (13) | 0 0 9 7 |
| (17) | 0 1 9 4 |
| (2) | 0 2 7 1 |
| | 0,0 8 8 2 |

## 4. - DECISION :

La probabilité totale des cas, au moins aussi "exceptionnels" que le cas observé, est égale à 0,0882 soit presque 9 %; les auteurs auxquels l'exemple est emprunté considèrent que ce chiffre fournit une bonne base de jugement pour accepter ou rejeter l'hypothèse d'indépendance.

## 5. - PROCEDURES D'APPROXIMATION

On peut comparer un tel verdict avec celui du Khi Deux traditionnel.

Pour un tableau :

| | |
|---|---|
| x y z | d |
| u v w | e |
| a b c | n |

il est prescrit de calculer la somme des termes tels que :

$$\frac{(x - \frac{ad}{n})^2}{\frac{ad}{n}}$$

Une telle somme, que nous désignerons dans ce qui suit par $\Omega^2$, représente une mesure de l'éloignement entre le tableau :

$$\begin{matrix} x & y & z \\ u & v & w \end{matrix}$$

et la moyenne de tous les tableaux possibles :

$$\begin{matrix} \frac{ad}{n} & \frac{bd}{n} & \frac{cd}{n} \\ \frac{ae}{n} & \frac{be}{n} & \frac{ce}{n} \end{matrix}$$

(qu'on pourra appeler tableau moyen, bien que le plus souvent, les nombres qui y figurent n'étant pas entiers, ce ne soit pas un tableau "possible").

On peut développer chacun des carrés des différences et obtenir :

$$\Omega^2 = n \left(\frac{x^2}{ad} + \frac{y^2}{bd} + \ldots + \frac{w^2}{ce} - 1\right)$$

Dans le cas présent : a = 6, b = 8, c = 3, d = 5, e = 12; n = 17; et on calcule aisément les dix-huit valeurs, qui se répartissent ainsi :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | 9,78 (01) | | 4,76 (02) | | 4,15 (03) | 7,97 (04) |
| | 8,77 (05) | | 2,55 (06) | | 0,75 (07) | | 3,35 (08) |
| 10,58 (09) | | 3,15 (10) | | 0,14 (11) | | 1,55 (12) | |
| | 6,56 (13) | | 2,35 (14) | | 2,55 (15) | | |
| | | 7,37 (16) | | 6,36 (17) | | | |
| | | | 12,99 (18) | | | | |

Le cas n° 11 est le plus proche possible de la moyenne, comme on le vérifiera aisément en calculant cette moyenne.

Selon cette procédure on ne détermine plus le caractère plus ou moins "exceptionnel" d'un évènement par sa probabilité mais par $\Omega^2$, c'est-à-dire l'éloignement de la moyenne. L'ordre n'est plus tout à fait le même :

(proba.croiss.): 18, 9, 1, 16, 5, 4, 13, 17, 2, 3, 10, 8, 14, 6, 15, 12, 7, 11.

($\Omega^2$ décroiss.) : 18, 9, 1, 5, 4, 16, 13, 17, 2, 3, 8, 10, 15, 6, 14, 12, 7, 11.

Il se trouve ici, cependant, que la coupure faite par le cas observé (à savoir le n° 2) est la même pour les deux critères.

Pour classer tous les tableaux possibles, au moins approximativement par rang de probabilité, on peut aussi utiliser un autre indicateur que l'on nommera $\chi^2$.

Pour calculer l'indicateur d'éloignement $\Omega^2$ on compare chaque effectif du

tableau à sa valeur moyenne: x à $\frac{ad}{n}$. On calcule le carré de la différence et on fait la somme pondérée. Il y a, pour $\chi^2$, quelque chose d'analogue: mais on fait le quotient au lieu de la différence, et un produit au lieu d'une somme. Bien entendu on pourra parler le langage logarithmique.

On posera donc:

$$\frac{1}{2}\chi^2 = x \log x + y \log y + \ldots + w \log w - a \log a - b \log b - \ldots - e \log e + n \log n$$

pour le tableau :

| | | | |
|---|---|---|---|
| x | y | z | d |
| u | v | w | e |
| a | b | c | n |

ce qu'on peut écrire aussi bien :

$$\exp\left(\frac{1}{2}\chi^2\right) = \text{produit des facteurs } \left(x : \frac{ad}{n}\right)^x$$

Avec une table de logarithmes le calcul est rapide. Il existe d'ailleurs dans le commerce des tables qui donnent directement le produit n log n en fonction de n (ces tables ont souvent été calculées en base binaire, à cause de l'usage prévu; dit "théorie de l'information". Mais cette base n'a ici aucun intérêt, et si l'on veut choisir autre chose que les bons vieux logarithmes décimaux, on fera bien d'adopter les logarithmes népériens. Il est bon d'avoir sous la main une table de ces logarithmes: conseillons le recueil publié par LABORDE, chez DUNOD, intitulé Tables numériques de fonctions élémentaires. Avec de telles tables nous formons les calculs suivants):

| n | $n^n$ | $n \log n = \log (n^n)$ |
|---|---|---|
| 0 | 1 | 0 |
| 1 | 1 | 0 |
| 2 | 4 | 1, 3863 |
| 3 | 27 | 3, 2958 |
| 4 | 256 | 5, 5452 |
| 5 | 3125 | 8, 0472 |
| | | etc ... |

## Valeurs de $\chi^2$

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 11,6 (01) | | 6,19 (02) | | 5,69 (03) | | 10,0 (04) |
| | 9,16 (05) | | 2,37 (06) | | 0,78 (07) | | 4,10 (08) | |
| 12,9 (09) | | 3,11 (10) | | 0,14 (11) | | 2,36 (12) | | |
| | 8,45 (13) | | 2,43 (14) | | 3,28 (15) | | | |
| | | 9,13 (16) | | 6,91 (17) | | | | |
| | | | 15,2 (18) | | | | | |

On voit que les rangements par $Q^2$ et par $\chi^2$ ne sont pas tout à fait pareils:

$Q^2$ : 18, 9, 1, 5, 4, 16, 13, 17, 2, 3, 8, 10, 15, 6, 14, 12, 7, 11

$\chi^2$ : 18, 9, 1, 4, 5, 16, 13, 17, 2, 3, 8, 15, 10, 14, 6, 12, 7, 11

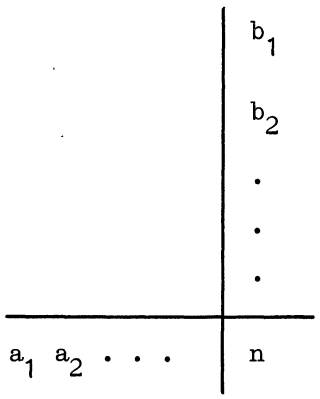
et l'un comme l'autre ne correspondent qu'approximativement au rangement par probabilités croissantes.

Mais on constate aussi que la correspondance entre $Q^2$ et $\chi^2$ est voisine de l'égalité, surtout pour les valeurs situées à l'intérieur de la configuration (les écarts sont notables sur les bords).

| | $Q^2$ | $\chi^2$ |
|---|---|---|
| n° 11 | 0,14 | 0,14 |
| n° 7 | 0,75 | 0,78 |
| n° 12 | 1,55 | 2,37 |
| n° 14 | 2,35 | 2,43 |
| n° 6 | 2,55 | 2,37 |
| n° 15 | 2,55 | 3,28 |
| n° 10 | 3,15 | 3,11 |
| n° 8 | 3,35 | 4,10 |
| n° 3 | 4,15 | 5,69 |
| n° 2 | 4,76 | 6,19 |
| etc... | | |

## 6. - DISTRIBUTION DE PROBABILITE

On comprendra, par l'exemple qui vient d'être esquissé, qu'on pourrait refaire tous les calculs pour chaque cas, c'est-à-dire pour chaque type de tableau statistique




et pour tout système de valeurs numériques des marges (les $a_i$, les $b_i$ et leur somme n).

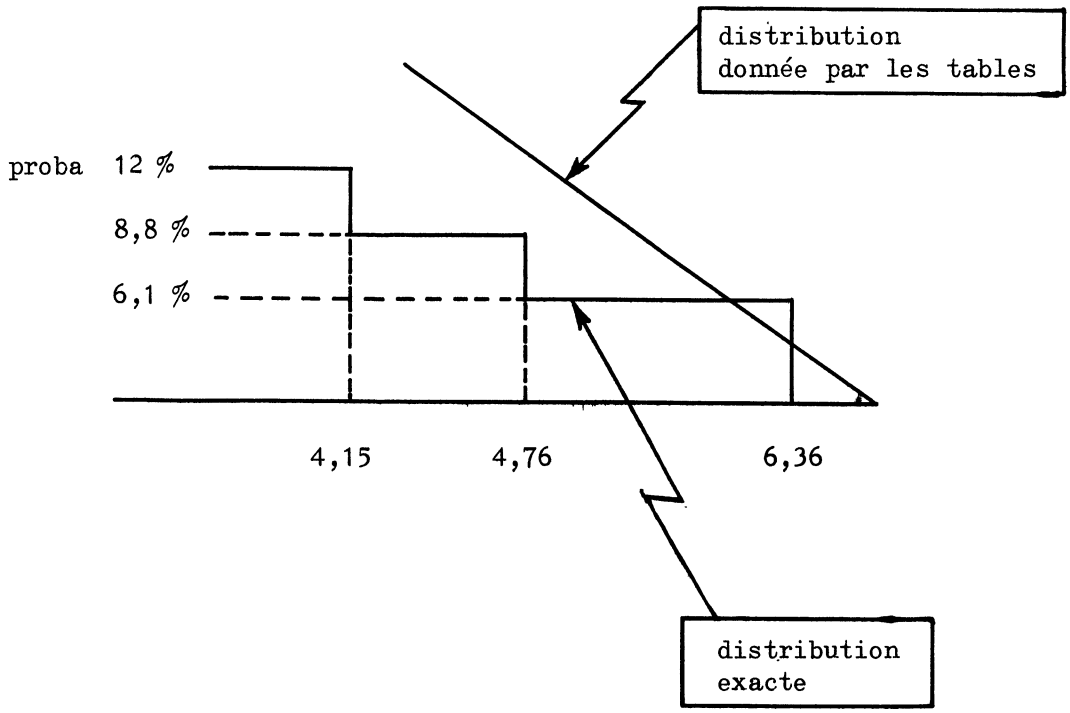
Dans l'exemple, les nombres présentés n'étaient pas bien grands (n = 17). Pour de plus grands nombres, on constaterait que les approximations sont meilleures ($Q^2$ et $\chi^2$ sont moins différents l'un de l'autre, et leur ordre croissant correspond mieux à celui des probabilités décroissantes).

D'autre part, la correspondance entre les probabilités et $Q^2$ (ou bien $\chi^2$) dépend peu de n (et d'autant moins que ce nombre est grand). Une seule table pourra suffire pour tous les tableaux de même format.

C'est pourquoi, dans la pratique courante, on ne s'astreint pas à calculer toutes les probabilités individuelles, mais qu'on estime la distribution des probabilités par l'approximation eulérienne (Khi Deux de Pearson, formules équivalentes aux lois de Poisson).

Il est intéressant de comparer la distribution de $Q^2$ (exacte) et celle de $\chi^2$ (approchée); dans la comparaison on n'oubliera pas que la seconde est continue alors que la première est en escalier.

Le lecteur est invité à tracer lui-même les graphiques en utilisant les données ci-après (et mieux encore en consultant des tables de Pearson ou de Poisson).


distribution
donnée par les tables
proba
12 %
8,8 %
6,1 %
4,15
4,76
6,36
distribution
exacte

| | $Q^2$ Proba. cumulées |
|---|---|
| (n° 18) | 12,99 |
| | .....0,00097 |
| | 10,58 |
| | .....0,0034 |
| | 9,78 |
| | .....0,0079 |
| | 8,77 |
| | .....0,0157 |
| | 7,97 |
| | .....0,0247 |
| | 7,37 |
| | .....0,0320 |
| | 6,56 |
| | .....0,0417 |
| | 6,36 |
| | .....0,0611 |
| (n° 2) | 4,76 |
| | .....0,0882 |
| | 4,15 |
| | .....0,1228 |
| | 3,35 |
| | .....0,1900 |
| | 3,15 |
| | .....0,2482 |
| | 2,55 |
| | .....0,4202 |
| | 2,35 |
| | .....0,4977 |
| | 1,55 |
| | .....0,6334 |
| | 0,75 |
| | .....0,7963 |
| (n° 11) | 0,14 |
| | .....1,0000 |

Khi Deux (2 libertés)

| Proba | Valeurs |
|---|---|
| 0,001 | 13,82 |
| 0,002 | 12,50 |
| 0,004 | 11,00 |
| 0,005 | 10,50 |
| 0,010 | 9,21 |
| 0,011 | 9,00 |
| 0,015 | 8,40 |
| 0,020 | 7,80 |
| 0,025 | 7,38 |
| 0,030 | 7,00 |
| 0,033 | 6,80 |
| 0,037 | 6,60 |
| 0,041 | 6,40 |
| 0,050 | 5,99 |
| 0,055 | 5,80 |
| 0,067 | 5,40 |
| 0,074 | 5,20 |
| 0,082 | 5,00 |
| 0,091 | 4,80 |
| 0,100 | 4,60 |
| 0,111 | 4,40 |
| 0,150 | 3,80 |
| etc... | |

On pourra représenter graphiquement de la même manière la distribution de probabilité concernant $\chi^2$.

| | $\chi^2$ | proba |
|---|---|---|
| | | 1,000 |
| (n° 11) | 0,14 | |
| | | 0,796 |
| (n° 7) | 0,75 | |
| | | 0,633 |
| (n° 6 et 12) | 2,37 | |
| | | 0,416 |
| (n° 14) | 2,43 | |
| | | 0,338 |
| (n° 10) | 3,11 | |
| | | 0,280 |
| (n° 15) | 3,28 | |
| | | 0,190 |
| (n° 8) | 4,10 | |
| | | 0,122 |
| (n° 3) | 5,69 | |
| | | 0,088 |
| (n° 2) | 6,19) | |
| | | 0,61 |
| | etc... | |