MATHÉMATIQUES ET SCIENCES HUMAINES

Problèmes d'enseignement

Mathématiques et sciences humaines, tome 4 (1963), p. 41-48
<http://www.numdam.org/item?id=MSH_1963__4__41_0>

G. MORLAT

NOMBRES FORTUITS ET TABLES DE NOMBRES FORTUITS (dits "au hasard")

1. - PEUT-ON DONNER UNE DEFINITION DU HASARD ?

On utilise souvent l'expression "au hasard" sans en donner une définition précise; cela semble sans inconvénient pratique, tant qu'on se borne à parler de tirages au hasard, expression parfaitement claire. Dans ce chapitre où nous allons étudier comment on fabrique et on utilise des "nombres au hasard", la question d'une définition préalable peut se poser.

L'expression "au hasard" est employée dans la littérature probabiliste et statistique avec une foule de sens différents. Il suffit d'en être prévenu. Un peu de réflexion permettra dans chaque cas d'expliciter le sens qu'il faut y attribuer.

Une seule chose est générale, et d'ailleurs fort essentielle: c'est que l'expression "au hasard" doit s'appliquer au mécanisme (plus précisément d'ailleurs, à la schématisation que nous faisons de ce mécanisme) et non pas au résultat. "Au hasard" signifiera: "dérivant d'un schéma probabiliste".

Un phénomène sera dit se produire au hasard si le mécanisme qui le produit est assimilable à une loi de probabilité (dans le sens le plus général). Exemple: les chiffres suivants ;

2, 7, 1, 0, 0, 9, 8, 4

sont-ils au hasard? - Question dépourvue de signification.

Par contre, une question qui a un sens est la suivante: la façon dont ces chiffres ont été déterminés peut-elle raisonnablement être schématisée par une loi de probabilité?

Il ne faut pas confondre un mécanisme très complexe avec un mécanisme assimilable à des tirages au hasard.

Pour ces raisons, il est préférable de désigner par nombres fortuits les nombres résultant d'un mécanisme probabiliste bien défini.

2. - UNE EXPERIENCE

Ainsi, lorsqu'un homme écrit des nombres tels qu'ils lui viennent à l'esprit ces nombres sont déterminés par un jeu complexe (nous supposerons) de ses circonvolutions cérébrales, de ses nerfs et des muscles de sa main (peut-être).

Si un homme fait effort pour écrire exprès des chiffres "sans suite" (c'est-à-dire ne correspondant à aucune suite d'idées) il peut penser qu'il écrit des chiffres au hasard, comme s'il les tirait dans une urne. Mais M. BOREL a insisté sur la difficulté, voire même l'impossibilité pour l'homme, d'imiter le hasard.

Nous avons pu constater que beaucoup de gens non prévenus pensent qu'il n'y a aucune différence entre des nombres tirés au hasard et des nombres écrits tels qu'ils leur viennent à l'esprit. Et il leur arrivera de considérer avec une ironie songeuse et pleine d'arrière-pensées le statisticien qui s'entoure de précautions un peu maniaques pour tirer ses nombres au hasard (Il se sert de tables fabriquées pour cela et encore il lui arrive de s'en méfier!).

Une expérience simple mettra en lumière les justifications de l'attitude du statisticien. Elle consiste à écrire des chiffres au fil de la plume, et à comparer leurs propriétés à celles de chiffres tirés au hasard. Ici, "au hasard" veut dire constituant des épreuves indépendantes de la loi de probabilité obtenue en affectant des probabilités égales aux diverses valeurs possibles.

Par raison de simplicité, l'expérience sera effectuée dans le système binaire et on demandera d'écrire 51 symboles prenant les valeurs 0 ou 1.

On comptera successivement :
1° - le nombre de zéros parmi les 50 premiers symboles écrits, soit X
2° - le nombre de séquences ("runs") auquel on retranchera 1, soit Y.

Chacune de ces quantités suit une loi binomiale lorsqu'il s'agit de nombres fortuits avec équiprobabilité. Cette loi a pour paramètres $p = \frac{1}{2}$, $n = 50$.

L'allure de la distribution de fréquence des X et des Y obtenue par un ensemble de 20 ou 30 auditeurs est très suggestive.

L'ensemble des échantillons obtenus pourra être testé par le χ^2. On constatera généralement que l'hypothèse d'une loi binomiale est assez admissible pour les x, mais tout à fait inadmissible pour les y. Par contre, ceux-ci sont en général assez bien expliqués par un schéma markovien, dans lequel la probabilité d'écrire "zéro" après avoir écrit "un" serait de l'ordre de 0,6. Une expérience plus prolongée, permettant l'étude d'un tel schéma pour chaque individu, offrirait sans doute quelque intérêt pour les psychologues.

3. - LES TABLES DE NOMBRES FORTUITS (ou tables de nombres au hasard)

Les résultats de l'expérience précédente justifieront sans doute les efforts déployés pour mettre à la disposition des statisticiens des séries de nombres obtenus par un mécanisme aussi voisin que possible du "hasard parfait", c'est-à-dire susceptible d'être bien représentés par des tirages dans une urne de BERNOULLI contenant 10 boules marquées 0, 1, 2, ..., 9.

Historiquement il semble que SLUTZKY fut un des premiers à utiliser à des fins de pédagogie et de recherche, les résultats des tirages de la loterie de Moscou. En France, les lecteurs de nos quotidiens peuvent trouver chaque semaine des nombres fortuits, tirés dans des urnes quelque peu sophistiquées.

Mais les tables les plus usuelles sont les suivantes :

- FISHER et YATES ont donné (dans "Statistical Tables for Biological, agricultural and Medical Research") 15.000 nombres obtenus à partir des décimales de rang 15 à 19 de tables de logarithmes (on a modifié ces décimales, dans lesquelles il y avait trop de 6).

Nota: Ces nombres sont donc obtenus par un mécanisme pour lequel une schématisation probabiliste n'est pas très convaincante. On peut seulement dire que la plupart des propriétés de fréquence qu'on attend de nombres obtenus par un tel mécanisme sont convenablement vérifiées par les tables de FISHER et YATES. Cependant, il est établi qu'avec un assez grand nombre de chiffres obtenus par le procédé de FISHER et YATES, on mettrait en évidence des caractères inacceptables (Il existe un théorème montrant que dans la suite des chiffres obtenus en prenant la $k^{\text{ième}}$ décimale des logarithmes décimaux des entiers naturels, la fréquence d'un chiffre donné ne tend pas vers une limite quand la longueur de la suite augmente indéfiniment).

- Les tables de TIPPETT (Tracts for computers, n° 15) contiennent 41.600 chiffres extraits de tables de recensement. Le danger d'un tel procédé tient à la possibilité d'erreurs systématiques de la part des enquêteurs (préférences inconscientes pour certains chiffres) qui donneraient naissance à la longue à des fréquences anormales.

- Les tables de KENDALL et BABINGTON SMITH (Tracts for computers, n° 24) contiennent 100.000 chiffres construits par un procédé beaucoup plus satisfaisant, puisque leurs auteurs ont construit une machine conçue à cet effet (Un disque divisé en 10 sections, numérotées de 0 à 9, et tournant à grande vitesse, était éclairé de temps à autre pendant des durées très courtes de façon à permettre à un observateur de noter un chiffre).

- Les tables construites tout récemment par la RAND CORPORATION (un million de nombres) ont été construites par un procédé physique (écrètement d'un bruit de fond, comptage des particules reçues par un compteur électronique, parité des nombres obtenus donnant des chiffres au hasard dans le système binaire, qui furent ensuite traduits dans le système décimal).

Certaines difficultés rencontrées par ce procédé, en apparence exempt de critiques (fréquences anormales pour certains couples de 2 chiffres) montrent à quel point il est difficile de réaliser un mécanisme parfaitement schématisable par le calcul des probabilités.

Lorsqu'on ne possède pas de table de nombres fortuits, mais qu'on se trouve dans une grande ville et qu'on dispose d'une demi-heure pour se promener dans les rues, il existe un procédé fort commode pour fabriquer des suites de nombres dont le caractère aléatoire est très satisfaisant: cela consiste à relever les tranches finales (2 ou 3 chiffres par exemple) des numéros minéralogiques des voitures automobiles que l'on rencontre. Il sera bien facile pour le lecteur d'expliciter les raisons pour lesquelles ce procédé est susceptible d'une excellente représentation par une urne de BERNOULLI, ainsi que les quelques précautions qu'il convient de prendre selon le cas.

L'emploi des calculatrices électroniques a donné lieu au développement de procédés dits "arithmétiques", donnant naissance à des suites de nombres très irrégulières, mais parfaitement déterminées, que l'on traite comme s'il s'agissait de nombres fortuits. Ces procédés consistent, par exemple, à partir d'un nombre de 20 chiffres par exemple, l'élever au carré, conserver la tranche de 20 chiffres "centrale" dans le nombre obtenu, et itérer l'opération. Une analyse sérieuse des propriétés des suites de nombres obtenus de cette façon - analyse qui semble avoir été seulement esquissée - pourrait sans doute justifier l'emploi de ces procédés pour la solution de certains problèmes particuliers dans lesquels on a besoin de "suites uniformes" et non pas de suites "au hasard". Mais c'est certainement un abus grave et dangereux, que de prétendre que l'on obtient ainsi des nombres fortuits.

On trouvera dans les Tables du Centre de Formation* quelques pages de nombres au hasard (4.000 chiffres) pouvant satisfaire aux besoins les plus courants.

4.- EMPLOI DES TABLES DE NOMBRES AU HASARD POUR FABRIQUER UN ECHANTILLON DERIVANT D'UNE LOI QUELCONQUE

Pour les besoins pratiques, le statisticien désire en général des échantillons provenant de lois de probabilité qui peuvent être quelconques. Mais il est facile de transformer les nombres d'une table de nombres fortuits (nombres équiprobables) en un échantillon provenant d'une loi de probabilité F (x) quelconque. Le principe consiste dans l'anamorphose rectangulaire, c'est-à-dire dans une transformation de variable associant à X, de loi F (x), une variable uniforme sur le segment 0,1: il suffit de prendre $Y = F(X)$. En pratique, on remplace parfois la loi envisagée par une loi discrète, dont les sauts valent 10^{-a}, a représentant le nombre de chiffres que l'on conviendra d'associer pour former des nombres fortuits répartis uniformément entre 0 et 1. (On prendra la plupart du temps $a = 3$ ou 4).

Exemple:

Les quantités de pluie recueillies annuellement au Parc Saint-Maur ont une distribution qui peut être considérée comme normale. Sur 72 années d'observations (1874-1945), on a trouvé une moyenne de 605 mm., un écart-type égal à 106 mm. On constitue sans difficulté un échantillon fictif suivant cette loi.

Nous utiliserons les tables reproduites d'après KENDALL et BABINGTON SMITH.

TABLEAU I

Années	Probabilité	Variable réduite	Pluie
1	0,231	- 0,7356	527
2	0,055	- 1,5982	435
3	0,148	- 1,0450	494
4	0,389	- 0,2819	575
5	0,973	+ 1,9268	809
.	.	.	.
.	.	.	.
.	.	.	.

On remarquera que le procédé très simple que nous venons d'indiquer ne permet d'obtenir comme observations que les valeurs égales, ou proches, des quantiles correspondant aux sauts de la fonction de répartition discrète par laquelle on a remplacé F (x); dans l'exemple ci-dessus, ce sont les milliles. Dans l'échelle naturelle (millimètres de pluie) les milliles ne sont pas équidistants. Ce phénomène donne une mauvaise image de la précision des mesures physiques de précipitation. Cela sera mieux compris par une illustration numérique.

---

*"Centre de Formation des Ingénieurs et Cadres, Institut de Statistique de l'Université de Paris, 9 quai Saint Bernard, Paris 5°".

Si nous trouvons dans la table de nombres fortuits les valeurs (adjacentes) 500, 501, etc..., nous y associerons les quantités de pluie calculées (tableau XII).

TABLEAU II

Nombre au hasard	Variable normale réduite	Pluie en mm.
0,500	0	605
0,501	0,0025	605
0,502	0,0050	606
0,503	0,0075	606
0,504	0,0100	606
etc...		

Chaque nombre en millimètres correspond à plusieurs milliles et le procédé peut être jugé acceptable.

Mais si nous trouvons dans la table les valeurs 001, 002, etc..; nous obtiendrons les résultats du tableau III.

TABLEAU III

Nombre au hasard	Variable normale réduite	Pluie en mm.
0,001	- 3,0902	277
0,002	- 2,8782	300
0,003	- 2,7478	314

Ainsi, les valeurs inférieures à 277, ou comprises entre 278 et 299 ou comprises entre 301 et 313, etc.. ne sauraient être obtenues par ce procédé.

D'ailleurs ce mode de construction ne donnera jamais lieu qu'à la possibilité d'obtenir les $10^a - 1$ quantiles d'ordre 10^a (c'est-à-dire ici 999 valeurs possibles) tandis que le nombre des valeurs réellement observables, pour une grandeur non bornée, même mesurée avec une erreur finie, est infini, ou du moins indéterminé.

Il reste que le procédé en question, quelque grossier qu'il soit, peut être employé, dans certains cas, en raison de sa commodité. Mais il est facile d'établir un procédé bien meilleur, et c'est indispensable si la construction d'un échantillon vise à la solution d'un problème touchant aux valeurs très rares, par exemple.

On peut supposer qu'une observation de x mm. provient, par le jeu des appareils de mesure, d'une quantité réellement comprise entre $x - \frac{1}{2}$ et $x + \frac{1}{2}$ mm. On affectera donc la valeur x à un nombre (fortuit) compris entre $F(x - \frac{1}{2})$ et $F(x + \frac{1}{2})$, les nombres fortuits étant lus dans la table avec un nombre de chiffres ad libitum (en fait, le nombre minimum de chiffres nécessaire pour situer à coup sûr un nombre dans l'un des intervalles $F(x - \frac{1}{2}), F(x + \frac{1}{2})$. Avec ce procédé, la seule limite à l'exactitude d'un échantillon est constituée par la précision des tables de la loi normale dont on dispose. En principe, si l'on doit construire un échantillon d'effectif un peu grand, il convient d'établir d'abord la table des valeurs F $(x + \frac{1}{2})$, x entier, avec le nombre de décimales dont on dispose pour la loi normale. Toujours pour notre exemple de la pluie annuelle à PARIS, cette table serait construite ainsi, pour les valeurs voisines de 450 mm.

TABLEAU IV

x	$x + \frac{1}{2}$	u (variable réduite)	F (u)
447	447,5	- 1,49	0,0681
448	448,5	- 1,48	0,0694
449	449,5	- 1,47	0,0708
450	450,5	- 1,46	0,0721
451	451,5	- 1,45	0,0735

Si l'on doit construire un échantillon d'effectif assez modéré, une partie importante d'une telle table peut s'avérer inutilisée et il est possible d'éviter sa construction complète; supposons que nous ayons trouvé dans la table de nombres fortuits les chiffres 002, nous avons vu que dans cette zone la distance entre 2 milliles consécutifs est de plusieurs millimètres. Si les chiffres suivants de la table sont 2 4 .... nous prendrons 00224 - la variable réduite correspondante est - 2,84, et nous trouvons :

$$x = 304$$

En résumé, le procédé le meilleur consiste dans l'utilisation d'une table de loi normale donnant la fonction de répartition pour des valeurs équidistantes de la variable réduite, la table de nombres fortuits étant utilisée par groupements de longueur variable.

Bien entendu, le même procédé peut être appliqué à la construction d'un échantillon provenant d'une loi quelconque, continue ou discrète. Il suffit pour cela d'avoir une table de la fonction de répartition. Dans le cas d'une loi discrète, on peut assez souvent construire un échantillon n'impliquant aucune approximation (cas de la loi de Poisson par exemple).

On imaginera sans peine comment, disposant d'une table classique de nombres fortuits (ou bien observant des automobiles) et usant de plus d'une table de logarithmes, on peut construire un échantillon dérivant d'une loi gamma d'indice entier. On peut même se passer pour cela d'une table de logarithmes.

Par ailleurs, il est tout aussi aisé de construire des épreuves de variables liées - par exemple des épreuves tirées d'une loi normale à deux variables, avec un coefficient de corrélation arbitrairement choisi. De tels échantillons ont d'ailleurs été publiés (3.000 épreuves) pour des coefficients de corrélation variant par dixième (Tract for Computers n° 26).

Enfin, on verra sans peine comment peut être construit un échantillon d'un processus stochastique déterminé - par exemple pour une chaîne de Laplace-Markov. De telles expériences seront riches d'enseignements pour quiconque aura la patience de s'y livrer.

EXERCICE SUR LES PERMUTATIONS (2ème Suite)

REPARTITION DES PERMUTATIONS DE n OBJETS SELON LA LONGUEUR l DE LA PLUS LONGUE CHAINE MONOTONE

(suite à l'Exercice proposé au N° 2, pp. 37-43 et N° 3, p. 43)

n =	1	2	3	4
l = 1	1	–	–	–
2	–	2	4	4
3	–	–	2	18
4	–	–	–	2

n =	5	6	7	8	9
l = 3	86	306	882	1 764	1 764
4	32	362	3 242	24 564	163 872
5	2	50	842	12 210	161 158
6	–	2	72	1 682	32 930
7	–	–	2	98	3 026
8	–	–	–	2	128
9	–	–	–	–	2
Total:	120	720	5 040	40 320	362 880

	n = 10	n = 11	n = 12
l = 4	985 032	5 323 032	25 038 288
5	1 969 348	22 493 900	241 860 366
6	592 652	10 182 250	169 394 678
7	76 562	1 749 772	37 886 686
8	5 042	159 722	4 501 774
9	162	7 922	306 682
10	2	200	11 882
11	–	2	242
12	–	–	2
	3 628 800	39 916 800	479 001 600

en pourcentages :

	n = 10	n = 11	n = 12
l = 4	27,1	13,3	5,2 %
5	54,3	56,3	50,5
6	16,3	25,5	35,4
7	2,1	4,4	7,9
8	0,1	0,4	0,9

ce qui permet de parier raisonnablement sur les longueurs des chaînes monotones quand on tire au hasard une permutation.

On expliquera (n° 5) comment se font les calculs.

G. Th. GUILBAUD