8
L'enchevêtrement des chiffres d'une table de chiffres aléatoires

par Maurice GLAYMANN

1. Point de départ

r est un naturel supérieur à 1 et (x_i) une suite de chiffres écrits en *base* r. Proposons-nous d'étudier la notion d'*enchevêtrement* des chiffres de cette suite. Cette notion conduit à construire un *test* qui permet de juger si une suite de chiffres se *comporte comme une suite donnée par une machine aléatoire* (urne, roulette, etc...).

Il existe de nombreuses méthodes pour construire des suites de chiffres aléatoires. Cependant, la plus simple et la plus naturelle, consiste à placer dans une urne, r boules identiques, numérotées 0, 1, 2, ..., (r −1). On extrait, *avec remise*, successivement N boules de cette urne: la suite $x_1, x_2, ..., x_N$ des chiffres ainsi obtenues est une suite de *chiffres aléatoires*.

a désigne un chiffre écrit dans la base r. La répartition du chiffre a dans la suite (x_i) est a priori quelconque. Cependant, comme la suite (x_i) est *aléatoire*, les chiffres 0, 1, 2, ..., (r—1) sont *équirépartis*. Mais, inversement *l'équirépartition* des chiffres d'une suite (y_i), n'entraîne pas pour cette suite, d'être une suite de chiffres aléatoires. Ainsi, par exemple la suite formée par p *zéros*, suivis par p *un*, etc..., terminée par p chiffres (r—1), est une suite de chiffres équirépartis, sans être pour autant une table de chiffres aléatoires.

Une notion intéressante est *l'écart* entre deux chiffres a consécutifs dans la suite (x_i). Nous dirons que cet écart est n, lorsque deux chiffres a sont séparés par n chiffres b_1, b_2, ..., b_n, *tous différents* de a et que cet écart est *nul*, lorsque deux chiffres a se suivent dans la suite (x_i).

La notion d'écart va nous conduire tout naturellement dans le paragraphe suivant à la notion d'*enchevêtrement*.

2. Etude d'un aléa

(x_i) est une suite de chiffres en base r et a un élément de l'ensemble $R = \{0, 1, 2, ..., (r-1)\}$. Ω_a désigne l'ensemble des sous-suites de (x_i) telles que

$$a, b_1, b_2, ..., b_n, a \qquad \text{avec} \qquad \forall\, i \;, b_i \neq a$$

ou de la forme aa.

Posons
$$\Omega = \bigcup_{a \in R} \Omega_a$$

X est l'application de Ω dans l'ensemble des naturels **N**:

$$X : \Omega \longrightarrow \mathbf{N}$$

telle que
$$X(a, b_1, b_2, ..., b_n, a) = n$$
et
$$X(a, a) = 0$$

L'application X associe à tout élément de Ω un naturel appelé *écart*.

Si la suite (x_i) est aléatoire, la probabilité de l'événement (a, a) est $1/r$.

La probabilité de l'événement a, b_1, b_2, ..., b_n, a est $\left(1 - \frac{1}{r}\right)^n \times \frac{1}{r}$,

car après le chiffre donné a, il y a n chiffres différents de a, chacun d'eux ayant la probabilité $\left(1 - \frac{1}{r}\right)$ d'apparaître, puis le chiffre a.

L'application $X : \Omega \longrightarrow \mathbf{N}$ est un aléa avec pour probabilité

$$p_n^a = P_r(a, b_1, b_2, ..., b_n, a) = \frac{1}{r}\left(1 - \frac{1}{r}\right)^n$$

Cette probabilité est *indépendante* du chiffre a envisagé; nous écrirons dans ces conditions:

$$p_n^a = p_n = \frac{1}{r}\left(1-\frac{1}{r}\right)^n$$

La table suivante donne les valeurs de p_n pour n variant de 1 à 20 et la base r de 2 à 12. La dernière ligne de cette table donne

$$p = P_r \; (20 < n < \infty)$$

n \ r	2	3	4	5	6	7	8	9	10	11	12
0	0,500	0,333	0,250	0,200	0,167	0,143	0,125	0,111	0,100	0,091	0,083
1	0,250	0,222	0,188	0,160	0,139	0,122	0,109	0,099	0,090	0,083	0,076
2	0,125	0,148	0,141	0,128	0,116	0,105	0,096	0,088	0,081	0,075	0,070
3	0,063	0,099	0,105	0,102	0,096	0,090	0,084	0,078	0,073	0,068	0,064
4	0,031	0,066	0,079	0,082	0,080	0,077	0,073	0,069	0,066	0,062	0,059
5	0,016	0,044	0,059	0,066	0,067	0,066	0,064	0,062	0,059	0,056	0,054
6	0,008	0,029	0,044	0,052	0,056	0,057	0,056	0,055	0,053	0,051	0,049
7	0,004	0,020	0,033	0,042	0,047	0,049	0,049	0,049	0,048	0,047	0,045
8	0,002	0,013	0,025	0,034	0,039	0,042	0,043	0,043	0,043	0,042	0,042
9	0,001	0,009	0,019	0,027	0,032	0,036	0,038	0,038	0,039	0,039	0,038
10	0	0,006	0,014	0,021	0,027	0,031	0,033	0,034	0,035	0,035	0,035
11	0	0,004	0,011	0,017	0,022	0,026	0,029	0,030	0,031	0,032	0,032
12	0	0,003	0,008	0,014	0,019	0,022	0,025	0,027	0,028	0,029	0,029
13	0	0,002	0,006	0,011	0,016	0,019	0,022	0,024	0,025	0,026	0,027
14	0	0,001	0,004	0,009	0,013	0,017	0,019	0,021	0,023	0,024	0,025
15	0	0,001	0,003	0,007	0,011	0,014	0,017	0,019	0,021	0,022	0,023
16	0	0,001	0,003	0,006	0,009	0,012	0,015	0,017	0,019	0,020	0,021
17	0	0	0,002	0,005	0,008	0,010	0,013	0,015	0,017	0,018	0,019
18	0	0	0,001	0,004	0,006	0,009	0,011	0,013	0,015	0,016	0,017
19	0	0	0,001	0,003	0,005	0,008	0,010	0,012	0,014	0,015	0,016
20	0	0	0,001	0,002	0,004	0,007	0,009	0,011	0,012	0,014	0,015
>20	0	0	0,002	0,009	0,022	0,039	0,061	0,084	0,109	0,135	0,161

Notons que pour une valeur donnée de n l'application

$$x \longmapsto x(1-x)^n$$

admet un maximum pour $x = \frac{1}{n+1}$. Ainsi, pour n donné, la probabilité p_n est *maximale* pour $r = n + 1$.

Cas particuliers

a) $r = 2$

Ce modèle est équivalent à N jets indépendants d'une pièce bien équilibrée (l'événement pile est noté 0 et l'événement face est noté 1).

La probabilité pour que deux "faces" consécutifs soient séparées par n "piles" est

$$p_n = \frac{1}{2^{n+1}}$$

b) $r = 6$

Ce modèle est équivalent à N jets indépendants d'un dé bien équilibré (il faut convenir de noter "0" l'événement qui correspond à la venue de la face "6").

La probabilité pour que deux as, par exemple, soient séparés par n autres faces différentes d'un as est

$$p_n = \frac{1}{6}\left(\frac{5}{6}\right)^n$$

c) $r = 10$

Ce modèle correspond à une table ordinaire de chiffres aléatoires, en base dix, avec

$$p_n = \frac{1}{10}\left(\frac{9}{10}\right)^n$$

C'est la probabilité d'avoir deux chiffres a séparés par n chiffres différents de a.

Fonction de répartition

La fonction de répartition de l'aléa X s'écrit

$$F(n) = \sum_{k=0}^{n} P_r(X = k) = P_r(X \leqslant n)$$

$$= \frac{1}{r} \sum_{k=0}^{n} \left(\frac{r-1}{r}\right)^{k} = 1 - \left(\frac{r-1}{r}\right)^{n+1}$$

comme $$r\,p_n = \left(\frac{r-1}{r}\right)^{n}$$

il vient $$F(n) = 1 - r\,p_{n+1}$$

En particulier, $$P_r(X \geqslant n+1) = 1 - F(n)$$

et $$P_r(X \geqslant n+1) = r\,p_{n+1} = \left(1 - \frac{1}{r}\right)^{n+1}$$

C'est ce résultat que nous avons utilisé pour calculer les valeurs de la dernière ligne de la table précédente.

Fonction génératrice, espérance, variance et écart-type

t est un réel; la fonction génératrice de l'aléa X s'écrit

$$g(t) = \sum_{n=0}^{+\infty} p_n\, t^n = \frac{1}{r} \sum_{n=0}^{+\infty} \left[\frac{(r-1)\,t}{r}\right]^{n}$$

pour $|t| < \dfrac{r}{r-1}$, il vient

$$g(t) = \frac{1}{r-(r-1)\,t}$$

En particulier, pour $t = 1$, on a bien

$$g(1) = \sum_{n=0}^{+\infty} p_n = 1$$

D'autre part, $$\frac{dg}{dt} = \frac{r-1}{[r-(r-1)t]^2}$$

et $$\frac{d^2g}{dt^2} = 2\,\frac{(r-1)^2}{[r-(r-1)t]^3}$$

D'où $$g'(1) = r-1$$

et $$g''(1) = 2(r-1)^2$$

Il en résulte que l'espérance de l'aléa X est

$$E(X) = r - 1$$

le moment factoriel d'ordre 2 est

$$E[X(X-1)] = 2(r-1)^2$$

l'espérance de l'aléa X^2 est

$$\begin{aligned} E(X^2) &= E[X(X-1)] + E(X) \\ &= (r-1)(2r-1) \end{aligned}$$

D'où la variance de l'aléa X :

$$V(X) = r(r-1)$$

et son écart-type $\sigma(X) = \sqrt{r(r-1)}$

Voici les valeurs dans trois cas particuliers:

r	2	6	10
E(X)	1	5	9
$\sigma(X)$	1,414	5,478	9,487

3. Test de "bon enchevêtrement"

(x_i) est une suite de chiffres écrits en base r.

Associons à cette suite l'application

$$X : \Omega \longrightarrow N$$

Nous dirons que les chiffres de la suite (x_i) sont *bien enchevêtrés*, si X suit une loi de probabilité *proche* de la loi théorique p_n. Cette *proximité* sera testée par le χ^2. En cas de non rejet de l'hypothèse, les chiffres de la suite (x_i) sont dits *bien enchevêtrés* et constituent une table de chiffres aléatoires valable.

Exemples et contre-exemple

a) Nous avons appliqué ce test aux 1 000 premiers chiffres proposés par M.G. KENDALL et B.B. SMITH dans "*Tables of Random*

Sampling Numbers" et reproduits dans "*Tables for Statisticians*" de ARKIN et COLTON:

	1-4	5-8	9-12	13-16	17-20	21-24	25-28	29-32	33-36	37-40
1	23 15	75 48	59 01	83 72	59 93	76 24	97 08	86 95	23 03	67 44
2	05 54	55 50	43 10	53 74	35 08	90 61	18 37	44 10	96 22	13 43
3	14 87	16 03	50 32	40 43	62 23	50 05	10 03	22 11	54 38	08 34
4	38 97	67 49	51 94	05 17	58 53	78 80	59 01	94 32	42 87	16 95
5	97 31	26 17	18 99	75 53	08 70	94 25	12 58	41 54	88 21	05 13
6	11 74	26 93	81 44	33 93	08 72	32 79	73 31	18 22	64 70	68 50
7	43 36	12 88	59 11	01 64	56 23	93 00	90 04	99 43	64 07	40 36
8	93 80	62 04	78 38	26 80	44 91	55 75	11 89	32 58	47 55	25 71
9	49 54	01 31	81 08	42 98	41 87	69 53	82 96	61 77	73 80	95 27
10	36 76	87 26	33 37	94 82	15 69	41 95	96 86	70 45	27 48	38 80
11	07 09	25 23	92 24	62 71	26 07	06 55	84 53	44 67	33 84	53 20
12	43 31	00 10	81 44	86 38	03 07	52 55	51 61	48 89	74 29	46 47
13	61 57	00 63	60 06	17 36	37 75	63 14	89 51	23 35	01 74	69 93
14	31 35	28 37	99 10	77 91	89 41	31 57	97 64	48 62	58 48	69 19
15	57 04	88 65	26 27	79 59	36 82	90 52	95 65	46 35	06 53	22 54
16	09 24	34 42	00 68	72 10	71 37	30 72	97 57	56 09	29 82	76 50
17	97 95	53 50	18 40	89 48	83 29	52 23	08 25	21 22	53 26	15 87
18	93 73	25 95	70 43	78 19	88 85	56 67	16 68	26 95	99 64	45 69
19	72 62	11 12	25 00	92 26	82 64	35 66	65 94	34 71	68 75	18 67
20	61 02	07 44	18 45	37 12	07 94	95 91	73 78	66 99	53 61	93 78
21	97 83	98 54	74 33	05 59	17 18	45 47	35 41	44 22	03 42	30 00
22	89 16	09 71	92 22	23 29	06 37	35 05	54 54	89 88	43 81	63 61
23	25 96	68 82	20 62	87 17	92 65	02 82	35 28	62 84	91 95	48 83
24	81 44	33 17	19 05	04 95	48 06	74 69	00 75	67 65	01 71	65 45
25	11 32	25 49	31 42	36 23	43 86	08 62	49 76	67 42	24 52	32 45

La table suivante donne les effectifs des différents écarts pour l'ensemble des chiffres 0, 1, ..., 9. Dans la colonne F, nous avons calculé la somme des effectifs et dans la colonne F_t nous avons reproduit les effectifs théoriques: comme nous avons à tester 1 000 chiffres, ces effectifs théoriques valent $10^3 p_n$.

Pour appliquer le test χ^2, nous avons regroupé dans une seule classe tous les écarts supérieurs ou égaux à 21.

n	F	F_t	$\lvert F - F_t \rvert$	$(F - F_t)^2$	$\frac{(F-F_t)^2}{F_t}$
0	108	100	8	64	0,640
1	87	90	3	9	0,100
2	86	81	5	25	0,309
3	67	73	6	36	0,493
4	70	66	4	16	0,242
5	60	59	1	1	0,017
6	48	53	5	25	0,472
7	44	48	4	16	0,333
8	44	43	1	1	0,023
9	42	39	3	9	0,231
10	27	35	8	64	1,828
11	32	31	1	1	0,032
12	31	28	3	9	0,321
13	24	25	1	1	0,004
14	25	23	2	4	0,174
15	19	21	2	4	0,190
16	25	19	6	36	1,895
17	11	17	8	64	3,765
18	12	15	3	9	0,600
19	8	13	5	25	1,923
20	6	12	6	36	3,000
n>20	112	109	3	9	0,082
				Total	16,674

soit $\chi^2 = 16,674$

Nous avons 22 classes, donc un χ^2 à 21 degrés de liberté. La table du χ^2 donne (pour 21 degrés de liberté):

$$P_r\,(\chi^2 > 15,445) = 0,8$$

$$P_r\,(\chi^2 > 17,182) = 0,7$$

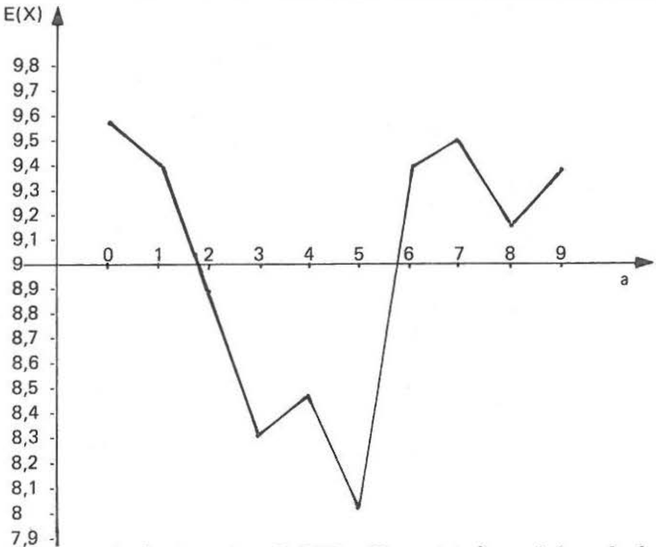
il vient par interpolation linéaire:

$$P_r\,(\chi^2 > 16{,}674) = 0{,}73$$

résultat qui conduit à accepter l'hypothèse de *proximité* de la répartition théorique et de la répartition étudiée.

Voici par ailleurs les moyennes des écarts pour chacun des chiffres 0, 1, ..., 9:

a	0	1	2	3	4	5	6	7	8	9
E(X)	9,543	9,398	8,871	8,308	8,448	7,973	9,398	9,479	9,146	9,383



La moyenne générale vaut 8,958; elle est très voisine de la moyenne théorique 9.

L'écart-type de cette distribution vaut 9,645, valeur peu différente de la valeur théorique 9,487.

b) *Comportement des décimales de π*

Nous avons testé de même les 10 000 premières décimales de π, données par F. GENUYS (revue CHIFFRES, Volume 1, 1958 ; p.18 à 22). Nous avons regroupé dans la même classe les écarts supérieurs ou égaux à 30:

n	F	F_t	$\lvert F - F_t \rvert$	$(F - F_t)^2$	$\frac{(F - F_t)^2}{F_t}$
0	941	1 000	59	3 481	3,481
1	892	900	8	64	0,071
2	847	810	37	1 369	1,690
3	714	729	15	225	0,309
4	658	656	2	4	0,006
5	586	590	4	16	0,027
6	521	531	10	100	0,188
7	478	478	0	0	0,000
8	394	430	36	1 296	3,014
9	376	387	11	121	0,313
10	341	349	8	64	0,183
11	324	314	10	100	0,318
12	289	282	7	49	0,174
13	248	254	6	36	0,142
14	231	229	2	4	0,017
15	193	206	13	169	0,820
16	202	185	17	289	1,562
17	161	167	6	36	0,215
18	149	150	1	1	0,007
19	145	135	10	100	0,741
20	124	122	2	4	0,033
21	106	109	3	9	0,082
22	103	98	5	25	0,255
23	80	89	9	81	0,910
24	72	80	8	64	0,800
25	72	72	0	0	0,000
26	69	65	4	16	0,246
27	62	58	4	16	0,276
28	51	52	1	1	0,019
29	57	47	10	100	2,128
30	435	428	7	49	0,114
				Total	18,141

soit $\chi^2 = 18{,}141$

Nous avons 31 classes, donc un χ^2 à 30 degrés de liberté. La table du χ^2 donne (pour 30 degrés de liberté):

$$P_r(\chi^2 > 16{,}306) = 0{,}98$$
$$P_r(\chi^2 > 18{,}493) = 0{,}95$$

il vient par interpolation linéaire:

$$P_r(\chi^2 > 18{,}141) = 0{,}9548$$

La répartition coïncide fortement avec la répartition théorique.

Voici par ailleurs les différentes moyennes:

a	0	1	2	3	4	5	6	7	8	9
E(X)	9,151	8,826	8,910	9,327	8,913	8,625	8,958	9,286	9,732	9,083

La moyenne générale vaut 9,081; elle est très voisine de la moyenne théorique 9.
L'écart-type vaut 9,400, valeur peu différente de la valeur théorique 9,487.
Ainsi, nous pouvons affirmer que les décimales de π sont bien enchevêtrées; tout se passe, de ce point de vue, comme si elles avaient été tirées au hasard ! .... (en fait, on peut se demander si cela n'est pas trop beau pour être vrai!).

c) Pour terminer donnons un contre-exemple :
Considérons la suite formée par k fois la répétition de la suite (0, 1, 2, 3, 4, 5, 6, 7, 8, 9).
Cette suite n'est évidemment pas une suite de chiffres aléatoires.
Etudions pour cette suite le test de "bon enchevêtrement".

Ici, quel que soit a, $a \in \{0, 1, 2, ..., 9\}$, la variable X est *certaine* et vaut 9. On a donc

$$P_r(X = n) = \begin{cases} 0 & \text{pour } n \neq 9 \\ 1 & \text{pour } n = 9 \end{cases}$$

La moyenne de cette distribution vaut 9 et coïncide avec la moyenne théorique! Cependant, l'écart-type est *nul*, il diffère totalement de la valeur théorique 9,487.

Pour calculer le χ^2, considérons trois classes:

$$X < 9 \quad , \quad X = 9 \quad \text{et} \quad X > 9$$

D'où la table:

	F	F_t	$\frac{(F - F_t)^2}{F_t}$
$X < 9$	0	6,1258 k	6,1258 k
$X = 9$	10(k – 1)	0,3874 k	u
$X > 9$	0	3,4868 k	3,4868 k

avec

$$u = \frac{[10(k-1) - 0{,}3874\,k]^2}{0{,}3874\,k}$$

Après quelques simplifications, il vient:

$$\chi^2 = \frac{1}{k}(248\,k^2 - 496\,k + 258)$$

Voici quelques valeurs de χ^2 en fonction de k:

k	χ^2
1	10
2	129
3	337
10	2 009,8
100	24 306,58

La table du χ^2 à deux degrés de liberté donne

$$P_r(\chi^2 > 13{,}815) = 0{,}001$$

Comme à partir de k supérieur ou égal à 2, χ^2 est supérieur à 13,815, on en déduit que la suite envisagée *n'est pas bien enchevêtrée* et par conséquent n'est pas aléatoire (comme il fallait s'y attendre!) ...

Samarkand, juillet 1975