ANNALES DE LA FACULTÉ DES SCIENCES DE TOULOUSE

R. HURON
J. MÉRIC

Sur une application du schéma d'urnes de Poisson

Annales de la faculté des sciences de Toulouse 4e série, tome 17 (1953), p. 265-272
<http://www.numdam.org/item?id=AFST_1953_4_17__265_0>

SUR UNE APPLICATION DU SCHÉMA D'URNES DE POISSON

par R. HURON et J. MÉRIC

Résumé. — Étude de la distribution du nombre d'échantillons de composition déterminée apparaissant dans un prélèvement non exhaustif effectué sur les urnes d'un schéma de Poisson. Application à l'étude statistique du comportement des blattes dans le « problème du T ».

[1]. Désignons par :

$$U_1, U_2, \dots U_i, \dots U_k,$$

les urnes du schéma de Poisson, et par :

$$p_1, p_2, \dots p_i, \dots p_k$$

les proportions respectives des boules blanches dans ces urnes. La population de chaque urne est supposée suffisamment grande pour que les tirages puissent être considérés comme non exhaustifs.

On extrait de chaque urne λ échantillons de taille n. Soit X_r le nombre de ceux qui contiennent exactement r blanches; X_r est une variable aléatoire dont on se propose d'étudier la loi de probabilité.

[2]. Considérons l'urne U_i. La probabilité pour qu'un échantillon de taille n, extrait de cette urne, contienne r blanches est :

(1) $$P_{i,r} = C^r_n \, p^r_i \, (1 - p_i)^{n-r}$$

On extrait λ échantillons de taille n de l'urne U_i. Soit $X_{i,r}$ le nombre de ceux contenant r blanches. On a :

(2) $$Pr\,[X_{i,r} = m] = C_\lambda^m \, P^m_{i,r} \, (1 - P_{i,r})^{\lambda-m}$$

$X_{i,r}$ est une variable aléatoire distribuée suivant une loi binômiale de moyenne :

$$m_{i,r} = \lambda \, P_{i,r}$$

et de variance :

$$v_{i,r} = \lambda \, P_{i,r} \, (1 - P_{i,r}),$$

On a, d'autre part :

(3) $$X_r = \sum_{i=1}^{i=k} X_{i,r} \, ,$$

d'où :

(4) $$M_r = E\,(X_r) = \sum_{i=1}^{i=k} m_{i,r} \, ,$$

et, puisque les $X_{i,r}$ sont des variables aléatoires indépendantes :

(5) $$V_r = E(X_r - M_r)^2 = \sum_{i=1}^{i=k} v_{i,r}$$

[3]. Supposons que les points d'abscisses : $p_1, p_2 \ldots p_k$ divisent l'intervalle $(0 - 1)$ en k intervalles égaux, de longueur $\frac{1}{k}$, c'est-à-dire que :

(6) $$p_i = \frac{i}{k}$$

Nous traduirons cette hypothèse en disant que les p_i sont *également répartis* sur l'intervalle $(0 - 1)$. On a alors :

(7) $$\lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{i=k} p_i^r (1 - p_i)^{n-r} = \int_0^1 p^r (1 - p)^{n-r} \, dp$$

$$= B(r + 1, n - r + 1)$$

D'où, *pour k suffisamment grand :*

$$M_r = \lambda k \, C_n^r \, B(r + 1, n - r + 1)$$

c'est-à-dire

(8) $$M_r = \lambda \frac{k}{n+1} = M$$

puisque cette valeur moyenne ne dépend pas de r.

D'autre part, d'après (5) :

$$V_r = \lambda \, C_n^r \sum_{i=1}^{i=k} p_i^r (1 - p_i)^{n-r} - \lambda (C_n^r)^2 \sum_{i=1}^{i=k} p_i^{2r} (1 - p_i)^{2n-2r}$$

Si l'on introduit l'hypothèse (6), on montre comme ci-dessus que :

(9) $$V_r = \lambda k \left[\frac{1}{n+1} - (C_n^r)^2 \, B(2r + 1, 2n - 2r + 1)\right]$$

[4]. **Remarques.**

a) Si l'on suppose, en outre, n assez grand, l'emploi de la formule de Stirling conduit à la valeur asymptotique :

(10) $$V_r = \lambda k \left[\frac{1}{n+1} - \frac{1}{(2n+1)\sqrt{\pi}} \sqrt{\frac{n}{r(n-r)}}\right]$$

qui suppose $r \neq 0$ et $r \neq n$.

Pour $r = 0$, on a exactement :

$$V_0 = \lambda\, k \frac{n}{(n+1)(2n+1)} \tag{11}$$

b) En supposant toujours l'hypothèse (6) satisfaite, on montre que :

$$\underset{k \to \infty}{\text{limite}} \frac{1}{k} \sum_{i=1}^{i=k} P_{i,r}^{\mu} = (C_n^r)^{\mu}\, B\,[r\mu + 1, (n-r)\mu + 1], \tag{12}$$

formule qui nous sera utile dans la suite.

[5]. Posons : $Q_{i,r} = 1 - P_{i,r}$; on a :

$$\varphi_i(t) = E\left[e^{(X_{i,r} - \lambda P_{i,r})\,t}\right] = \left[Q_{i,r}\, e^{-P_{i,r}\,t} + P_{i,r}\, e^{(1-P_{i,r})\,t}\right]^{\lambda}$$

Désignons la quantité entre crochets par $\Psi_i(t)$. Les variables aléatoires $X_{i,r}$ étant indépendantes, on aura pour la fonction caractéristique de X_r :

$$\varphi(t) = \left[\prod_{i=1}^{i=k} \Psi_i(t)\right]^{\lambda}$$

D'où :

$$\text{Log}\, \varphi(t) = \lambda \sum_{i=1}^{i=k} \text{Log}\, \Psi_i(t)$$

Mais :

$$\text{Log}\, \Psi_i(t) = P_{i,r}\, Q_{i,r} \frac{t^2}{2!} + P_{i,r}\, Q_{i,r}\,(Q_{i,r} - P_{i,r}) \frac{t^3}{3!} \cdots$$

Il en résulte que :

$$\text{Log}\, \varphi(t) = \lambda \sum_{i=1}^{k} P_{i,r}\, Q_{i,r} \frac{t^2}{2!} + \lambda \sum_{i=1}^{k} P_{i,r}\, Q_{i,r}\,(Q_{i,r} - P_{i,r}) \frac{t^3}{3!} \cdots$$

Pour passer à la variable réduite :

$$Z_r = \frac{X_r - \lambda \sum_{i=1}^{k} P_{i,r}}{\sqrt{\lambda \sum_{i=1}^{k} P_{i,r}\, Q_{i,r}}} \tag{13}$$

il suffit de poser : $$t = \frac{\theta}{\sqrt{\lambda \sum_{i=1}^{k} P_{i,r}\, Q_{i,r}}}$$

On obtient :

$$\overline{\varphi(\theta)} \sim e^{\left[\frac{\theta^2}{2!} + \frac{\overset{i}{\Sigma} P_{i,r} Q_{i,r} (Q_{i,r} - P_{i,r})}{\sqrt{\lambda} \left(\overset{i}{\Sigma} P_{i,r} Q_{i,r}\right)^{\frac{3}{2}}} \frac{\theta^3}{3!} \cdots\cdots\right]},$$

soit :

$$\varphi(\theta) \sim e^{\frac{\theta^2}{2}} \left[1 + h\,\theta^3\right] \tag{14}$$

avec :

$$h = \frac{\overset{i}{\Sigma} P_{i,r} Q_{i,r} (Q_{i,r} - P_{i,r})}{3! \sqrt{\lambda} \left(\overset{i}{\Sigma} P_{i,r} Q_{i,r}\right)^{\frac{3}{2}}} \tag{15}$$

[**6**]. La formule de réciprocité de Fourier permet de déduire de (13) la fonction de distribution de Z_r. On a :

$$f(z) \sim f_0(z) - h\, f_0'''(z)$$

avec :

$$f_0(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}$$

D'où :

$$f(z) \sim \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}} \left[1 - h\,(3z - z^3)\right] \tag{16}$$

[**7**]. Si l'on désigne par $E_{i,r}$ l'événement consistant en ce que l'échantillon de taille n extrait de l'urne U_i contienne exactement r blanches, l'événement $E_{i,r}$ a, pour r fixé, une probabilité constante $P_{i,r}$. Au schéma d'urnes envisagé dans cette note, on peut superposer un deuxième schéma d'urnes de Poisson de compositions respectives $P_{i,r}$, $Q_{i,r}$ $(i = 1, 2, \ldots\ldots k)$.

Les formules (13) et (14) se déduisent alors immédiatement de formules classiques [1].

[**8**]. **Ordre de grandeur de h pour k grand.**

On a :

$$\overset{i}{\Sigma} P_{i,r} Q_{i,r} (Q_{i,r} - P_{i,r}) = \overset{i}{\Sigma} (P_{i,r} - 3 P^2_{i,r} + 2 P^3_{i,r})$$

1. Poisson : « Recherches sur la probabilité des jugements ». (Bachelier, 1837.) Chapitre IV : « suite du calcul des probabilités dépendantes d'un très grand nombre », p. 246, n^{os} 94-95.

G. Darmois : « Statistique Mathématique ». (Doin, 1928.) Chapitre III, §§ 8-10.

D'où, d'après (12)

$$\sum^{i} P_{i,r} Q_{i,r} (Q_{i,r} - P_{i,r}) \sim k \left[\frac{1}{n+1} - 3 (C^r_n)^2 B (2r+1, 2n-2r+1) \right.$$
$$\left. + 2 (C^r_n)^3 B (3r+1, 3n-3r+1) \right]$$

D'autre part :

$$\sum^{i} P_{i,r} Q_{i,r} \sim k \left[\frac{1}{n+1} - (C^r_n)^2 B (2r+1, 2n-2r+1) \right]$$

On en déduit, grâce à (15), l'expression de h, et l'on voit que h tend vers zéro avec $\frac{1}{\sqrt{k}}$ et $\frac{1}{\sqrt{\lambda}}$ Donc, pour k assez grand, $f(z)$ est assimilable à une loi normale réduite.

[9]. **Étude de deux cas particuliers.**

Nous supposons que :

$$\lambda = 1 \quad , \quad k = 30.$$

D'une manière générale, nous posons :

$$I = \int_{-\infty}^{-z} f(u)\, du \qquad I' = \int_{z}^{+\infty} f(u)\, du$$

où $f(z)$ est donnée par (16).

La table (I) donne I et I′ pour $n = 4$,

La table (II) donne I et I′ pour $n = 10$ (2).

Pour chaque couple de valeurs n et r, on a déterminé z_r et z'_r tels que

$$I = \int_{-\infty}^{-z_r} f(u)\, du = 2{,}5\,\%_0 \quad , \quad I' = \int_{z'_r}^{+\infty} f(u)\, du = 2{,}5\,\%_0$$

d'où d'après (13) les limites de variations de X_r, au seuil de probabilité de 5 % :

$$(17) \qquad \begin{cases} x_r = M - \sigma_r z_r \\ x'_r = M + \sigma_r z'_r \end{cases}$$

avec :

$$\sigma_r = \sqrt{\sum^{i} P_{i,r} Q_{i,r}}$$

La table III est relative au cas où $n = 4$, la table IV au cas où $n = 10$.

---

2. Si l'on pose $I_1 = \int_{-\infty}^{-z} f_0(u)\, du$, et $I_2 = f''_0(z)$, on a : $I = I_1 - hI_2$; $I' = I_1 + hI_2$; les valeurs de I_1 et I_2 ont été lues dans les « Tables numériques universelles » de Marcel Boll (Dunod, 1947); I_1 dans la table p. 21 : « Courbe de Galton », pp. 600-611; $I_2 = y''$ dans la table p. 26 : « Dérivées de la Courbe de Gauss », pp. 628-632.

[10]. **Application.**

Dans une « Étude statistique du comportement de Blatella Germanica dans le problème du T » (3), l'un de nous a montré qu'à une blatte donnée on pouvait, au moins pendant une période de quelques jours, attribuer un coefficient p, correspondant à la probabilité qu'a cette blatte d' « aller à droite ».

Pour les blattes étudiées, les p, se répartissent également entre 0 et 1. On peut donc assimiler l'ensemble des blattes étudiées *dans une journée* à un schéma d'urnes de Poisson pour lequel l'hypothèse du n° 3 serait approximativement vérifiée et où on aurait $\lambda = 1$.

k représente alors le nombre des blattes soumises à l'expérience dans une journée : nous avions $k = 30$;

n représente le nombre de parcours imposé à chaque animal; nous avions $n = 4$.

La table III montre qu'avec ces données :

$$M = 6$$

et qu'au seuil de 5 % le nombre de séries de 4 parcours où une blatte est allée $r = 0, 1, 2, 3, 4$ fois à droite peut varier entre les limites approximatives suivantes :

2 — 9	pour	$r = 0$ ou 4
2 — 10	—	$r = 1$ ou 3
2 — 10	—	$r = 2$

En se reportant au travail cité, on verra, sur les 20 jours d'expérience rapportés, que les résultats expérimentaux viennent se placer dans les zones de confiance indiquées ci-dessus, sauf pour les journées n^{os} 10-12-18-26. Sans doute, ces jours-là, et malgré toutes les précautions prises un facteur externe a-t-il imposé un déterminisme aux blattes.

Certains expérimentateurs prennent $n = 10$; c'est pour critiquer leurs résultats expérimentaux par la méthode indiquée ci-dessus, que nous avons établi la table IV.

3. *Bulletin de la Société d'Histoire Naturelle de Toulouse*, t. 88, fasc. 3-4, 1943, p. 346.

TABLE I.

	$r = 0 — 4$		$r = 1 — 3$		$r = 2$	
z	I	I′	I	I′	I	I′
0,0	0,50940	0,40960	0,51310	0,48690	0,51450	0,48550
0,5	0,32480	0,31230	0,32720	0,30990	0,30810	0,30890
1,0	0,15870	0,15870	0,15870	0,15870	0,15870	0,15870
1,2	0,11310	0,11710	0,11230	0,11790	0,11200	0,11820
1,4	0,07738	0,08414	0,07605	0,08505	0,07553	0,08599
1,5	0,06300	0,07162	0,06151	0,07211	0,06092	0,07270
1,6	0,05072	0,05888	0,04913	0,06047	0,04851	0,06109
1,7	0,04038	0,04876	0,03875	0,05039	0,03810	0,05104
1,8	0,03177	0,04009	0,03014	0,04172	0,02950	0,04236
1,9	0,02468	0,03276	0,02311	0,03433	0,02249	0,03495
2,0	0,01893	0,02657	0,01744	0,02806	0,01686	0,02864
2,1	0,01433	0,02139	0,01295	0,02277	0,01240	0,02332
2,2	0,01069	0,01711	0,00944	0,01836	0,00894	0,01815
2,3	0,00786	0,01358	0,00674	0,01470	0,00630	0,01514
2,4	0,00569	0,01070	0,00471	0,01169	0,00432	0,01208
2,5	0,00404	0,00838	0,00320	0,00922	0,00286	0,00956
2,6	0,00282	0,00650	0,00210	0,00722	0,00182	0,00750
2,7	0,00193	0,00501	0,00133	0,00561	0,00109	0,00585
2,8	0,00139	0,00393	0,00089	0,00443	0,00069	0,00463
3,0	0,00051	0,00219	0,00019	0,00251	0,00006	0,00264
3,2	0,00017	0,00121				
3,4	0,00013	0,00086				

TABLE II.

	$r = 0 — 10$		$r = 1 — 9$		$r = 2 — 8$		$r = 37$		$r = 4 — 6$		$r = 5$	
z	I	I′	I	I′	I	′	I	I′	I	I′	I	I′
0	0,51660	0,48340	0,52270	0,47730	0,52640	0,47360	0,52800	0,47200	0,52870	0,47130	0,52890	0,47110
0,5	0,32980	0,30730	0,33360	0,30350	0,33600	0,30110	0,33700	0,30000	0,33750	0,29960	0,33770	0,29940
1	0,15870	0,15870	0,15870	0,15870	0,15870	0,15870	0,15870	0,15870	0,15870	0,15870	0,15870	0,15870
1,2	0,11140	0,11870	0,11020	0,11990	0,10940	0,12070	0,10920	0,10100	0,10890	0,12120	0,10890	0,12130
1,4	0,07466	0,08686	0,07258	0,08894	0,07126	0,09026	0,07069	0,09083	0,07043	0,09109	0,07035	0,09117
1,5	0,05994	0,07368	0,05759	0,07603	0,05611	0,07751	0,05546	0,07816	0,05517	0,07845	0,05508	0,07854
1,6	0,04745	0,06215	0,04495	0,06465	0,04336	0,06624	0,04268	0,06692	0,04236	0,06724	0,04226	0,06734
1,7	0,03702	0,05212	0,03445	0,05469	0,03282	0,05632	0,03211	0,05703	0,03178	0,05736	0,03169	0,05745
1,8	0,02842	0,04344	0,02586	0,04600	0,02424	0,04762	0,02354	0,04832	0,02322	0,04864	0,02312	0,04874
1,9	0,02145	0,03599	0,01353	0,03847	0,01740	0,04004	0,01672	0,04072	0,01640	0,04104	0,01631	0,04113
2	0,01587	0,02963	0,00932	0,03197	0,01204	0,03346	0,01140	0,03410	0,01110	0,03440	0,01101	0,03449
2,1	0,01149	0,02423	0,00615	0,02640	0,00794	0,02778	0,00735	0,02837	0,00707	0,02865	0,00699	0,02873
2,2	0,00812	0,01968	0,00380	0,02165	0,00490	0,02290	0,00436	0,02344	0,00411	0,02369	0,00403	0,02377
2,3	0,00556		0,00213		0,00269		0,00215		0,00198		0,00192	
2,4	0,00367		0,00097		0,00115		0,00073		0,00053		0,00048	
2,5	0,00230		0,00021		0,00013							
2,6	0,00134											
2,7	0,00069											
2,8	0,00035											

TABLE III.

r	M	z_r	z'_r	x_r	x'_r
0 ou 4	6	1,895	2,030	2,905	9,32
1 ou 3	6	1,873	2,056	2,310	10,16
2	6	1,864	2,067	2,140	10,28

TABLE IV.

r	M	z_r	z'_r	x_r	x'_r
0 ou 10	2,727	1,847	2,090	0,622	5,11
1 ou 9	2,727	1,811	2,128	0,182	5,72
2 ou 8	2,727	1,790	2,155	0,115	5,87
3 ou 7	2,727	1,781	2,167	0,088	5,94
4 ou 6	2,727	1.778	2,170	0,074	5,97
5	2,727	1.778	2,172	0,067	5,98