REVUE DE STATISTIQUE APPLIQUÉE

F. CHARTIER

Le problème du collectionneur

Revue de statistique appliquée, tome 7, nº 1 (1959), p. 63-82

<http://www.numdam.org/item?id=RSA_1959__7_1_63_0>

LE PROBLÈME DU COLLECTIONNEUR

par F. CHARTIER

Administrateur à l'I. N. S. E. E.

Une forme de promotion des ventes assez répandue actuellement, notamment pour les produits tels que café, chocolat, ... consiste à mettre dans chaque paquet un objet (image, figurine de plastique, ...) pris dans une collection de n objets. L'acheteur est invité à conserver ces objets et à reconstituer la collection. Une règle fixe généralement les conditions d'échange des pièces en double, triple, ..., exemplaires contre une nouvelle que le collectionneur n'aurait pas encore. Une fois la collection réalisée, et après avoir répondu à quelques "questions subsidiaires", l'acheteur recevra un prix.

Le but de l'opération est évidemment, pour le vendeur, de conserver la clientèle et de l'inviter à acheter davantage (désir du client de poursuivre la collection et de l'achever le plus rapidement possible) et même d'accroître la clientèle (les personnes ayant entrepris une collection demandent à leurs connaissances de les aider).

Deux questions se posent pour l'organisateur : quelle importance (nombre de pièces) donner à la collection dont on demande la réalisation ? Quelle règle fixer pour l'échange des pièces en double, triple, ..., exemplaire ? Il convient, en effet, que les conditions du succès ne soient ni trop faciles (car très rapidement tous les prix seraient distribués et l'opération terminée, sans grand profit sur le plan commercial), ni trop difficiles (car les clients se lasseraient et abandonneraient).

En fait, le problème que l'on va traiter ci-après intéresse plutôt le client que le commerçant. En effet, on va déterminer quel est le nombre moyen de paquets à acheter pour avoir réalisé la collection ou quelle est la probabilité de l'avoir réalisée quand on a acheté un nombre donné, p, de paquets. Chaque client, connaissant ses habitudes d'achat, peut alors se faire une idée de la difficulté de l'opération pour lui-même. En revanche, le commerçant ne sait pas du tout comment se répartissent ses clients par fréquence d'achat. Il peut lui paraître raisonnable de supposer l'achat d'un paquet par semaine. Mais combien de clients agissent ainsi ? Combien de clients achètent un paquet tous les 4 ou 5 jours ou au contraire tous les 10 ou 15 jours ? Même le nombre de clients est ignoré du commerçant qui ne connaît que le nombre total de paquets vendus. Au reste, tous les acheteurs ne se transforment pas obligatoirement en collectionneurs. Beaucoup, par négligence ou ignorance, n'entrent pas dans le jeu. Ils distribuent les objets qu'ils trouvent, ou en font cadeau à d'autres, ce qui facilite la tâche de ces derniers, comme d'ailleurs les inévitables échanges entre collectionneurs. Echanges à un contre un, sans doute, et dont on peut seulement dire qu'ils réduisent la difficulté de réalisation de la collection de chacun sans pouvoir préciser dans quelle mesure.

On voit donc que les résultats auxquels on arrivera dans l'étude suivante *ne doivent être considérés que comme des ordres de grandeur quand on veut les appliquer aux problèmes concrets dont tant de données sont inconnues et ne semblent pas aisés à mesurer par enquête auprès d'un échantillon de ménages.*

Dans toute l'étude on supposera que les n types d'objets (images, figurines, ...) dont l'ensemble constitue la collection à réaliser sont également fréquents dans les paquets et qu'ils y sont répartis indépendamment les uns des autres, de sorte qu'il est strictement équivalent d'acheter un paquet chaque semaine pendant un an ou d'acheter cinquante deux paquets en une seule fois. On suppose encore qu'il n'y a aucun échange d'objets entre collectionneurs.

I - NOMBRE MOYEN DE PAQUETS A ACHETER POUR REALISER UNE COLLECTION DE n OBJETS

Il est évident que le premier paquet acheté fournit toujours le premier élément de la collection, tandis que les paquets suivants auront de plus en plus de chances de n'amener que des éléments dont on aura déjà au moins un exemplaire. Autrement dit, ce sont les derniers éléments de la collection qui seront les plus difficiles à obtenir, qui nécessitent en moyenne le plus grand nombre d'achats.

Précisons ces notions intuitives. Le nombre X de paquets tel que le $n^{ième}$ et dernier élément de la collection se trouve dans le Xième paquet peut être mis sous la forme :

$$X = X_1 + X_2 + X_3 + \dots + X_n$$

où X_i est le nombre de paquets achetés pour obtenir le $i^{ème}$ élément nouveau (quand on a déjà i - 1 éléments en au moins un exemplaire chacun).

$X_1 = 1$ de façon certaine ; quant à X_i pour $i = 2, 3, \dots, n$, c'est une variable aléatoire susceptible des valeurs

$$1 \quad 2 \quad 3 \quad 4 \quad \dots.$$

auxquelles sont attachées les probabilités

$$p_i \quad q_i p_i \quad q_i^2 p_i \quad q_i^3 p_i \quad \dots.$$

en posant

$p_i = \dfrac{n - (i-1)}{n}$ prob. d'avoir un élément nouveau dans un paquet

$q_i = 1 - p_i$ " " " " dont on possède déjà au moins un exemplaire.

La valeur moyenne (ou espérance mathématique) de X_i est

$$E(x_i) = p_i + 2q_ip_i + 3q_i^2 p_i + 4q_i^3 p_i + \dots.$$
$$= \frac{1}{p_i} = \frac{n}{n - (i-1)},$$

de sorte que la moyenne de X est :

$$E(X) = 1 + E(X_2) + E(X_3) + \ldots\ldots\ldots + E(X_n)$$
$$= 1 + \frac{n}{n-1} + \frac{n}{n-2} + \ldots\ldots\ldots + \frac{n}{1}$$

$$\boxed{E(X) = n\left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \ldots\ldots\ldots + \frac{1}{1}\right)}$$

Par exemple, pour n = 5, on a

X_1	$= 1$	, nombre certain d'achats pour avoir le 1er	élément
$E(X_2) = \frac{5}{4} =$	1,125	, nombre moyen - - -	2ème -
$E(X_3) = \frac{5}{3} =$	1,667	, - - - - -	3ème -
$E(X_4) = \frac{5}{2} =$	2,500	, - - - - -	4ème -
$E(X_5) = \frac{5}{1} =$	5,000	, - - - - -	5ème -
Total	= 11,292	, - - - - -	la collection complète.

Quelques autres valeurs numériques, à titre indicatif :

n =	E(X) =
5	11,42
10	29,29
20	71,95
50	225,0
100	518,7
200	1 176

Bien entendu, il ne s'agit là que d'un nombre moyen d'achats. Il est possible d'avoir réalisé la collection plus tôt, mais aussi de ne l'avoir achevée que plus tard et même beaucoup plus tard. Il est aisé de calculer l'écart-type, mesure de la variabilité de X ; mais, étant donnée la dissymétrie de la loi de probabilité de X on ne peut guère utiliser cet écart-type pour former un intervalle encadrant E(X) et dans lequel il y aurait une probabilité donnée de trouver X. Cet intervalle sera formé par une autre voie au § 4.

II - REPARTITION DE p OBJETS PAR FREQUENCE

Une autre façon d'envisager la question est de se demander, pour le collectionneur disposant de p objets, combien d'objets seront en 0 exemplaire, en 1 exemplaire, en 2 exemplaires, en 3, en 4, ...

On posera p_0 le nombre d'objets en 0 exemplaire
p_1 " " 1 "
p_2 " " 2 "
....
p_n " " n "

Ces nombres $p_0, p_1, p_2, \ldots, p_n$ sont évidemment nuls ou positifs et tels que

$p_0 + p_1 + p_2 + \ldots + p_n = n$, nombre de types d'objets

$p_1 + 2p_2 + \ldots. + np_n = p$, " d'objets possédés par le collectionneur.

De plus, ils sont aléatoires.

On trouve au tableau 1 les répartitions possibles de $p = 7$, 9 et 12 objets quand il en existe $n = 5$ types.

La 1ère répartition pour $p = 7$ signifie : 1 type d'objet en 7 exemplaires, 4 types d'objets en 0 exemplaire ($p_0 = 4$, $p_1 = p_2 = \ldots = p_6 = 0$, $p_7 = 1$). La seconde : 1 type d'objet en 6 exemplaires, 1 en 1 exemplaire, 3 en 0 exemplaire ($p_0 = 3$, $p_1 = 1$, $p_2 = p_3 = p_4 = p_5 = p_7 = 0$, $p_6 = 1$)...

Les collectionneurs ayant les répartitions 11 et 13 ont terminé leur collection. A ceux ayant les répartitions 7, 10 et 12, il manque 1 élément de la collection. A ceux ayant les répartitions 4, 6, 8 et 9, il en manque 2, etc.

Toutes ces répartitions ne sont pas également probables. C'est ce qu'on voit en considérant les probabilités indiquées pour les répartitions du tableau 1.

D'une manière générale, la probabilité d'obtenir sur p objets la répartition p_0, p_1, p_2, ..., p_n est :

$$\frac{n!}{p_0!\, p_1!\, p_2! \ldots p_n!} \cdot \frac{p!}{(0!)^{p_0}\,(1!)^{p_1}\,(2!)^{p_2} \ldots\ldots (n!)^{p_n}} \cdot \frac{1}{n^p}$$

Les deux derniers facteurs représentent la probabilité d'obtenir
aucun exemplaire de p_0 types d'objets donnés
un - - p_1 - - -
...

(loi multimoniale : p tirages indépendants dans une urne à n catégories équiprobables).

Le premier facteur représente le nombre de façons dont on peut choisir parmi les n types d'objets les p_0 types, p_1 types, p_2 types, ... devant être représentés par 0, 1, 2, ... objets.

A partir du tableau des répartitions possibles pour p et n donnés, il est facile de répondre à des questions telles que : "quelle est la probabilité d'avoir une collection complète avec p objets" "une collection à laquelle il ne manque qu'un élément ?" "une double collection ?".

Réponses : pour $n = 5$, par totalisation des probabilités des lignes convenables du tableau 1 :

	p = 7		p = 9		p = 12	
	Lignes n°	Probabilité	Lignes n°	Probabilité	Lignes n°	Probabilité
Collection complète	11, 13	0,215 04	12, 17, 20 22, 23	0,427 069	12, 18, 25, 27, 32, 35, 36, 40, 42, 43, 44, 46, 47	0, 594 739
Collection sauf un élément	7, 10, 12	0, 537 60	7, 11, 15 16, 19, 21	0,477 389	7, 11, 16, 17, 22, 24, 26, 29, 31, 33, 34, 38, 39, 41, 45	0,383 828
Double collection	impossible		impossible		44, 47	0,033 116

Tableau 1

Répartition possibles de p objets quand il y a n = 5 types d'objets

Répartition N°	Nombre d'objets sur p en 0, 1, 2, ..., p exemplaires: 0	1	2	3	4	5	6	7(=p)	Probabilité de chaque répartition
1	4	.	.	.	.	.	.	1	0,000 064
2	3	1	.	.	.	.	1	.	1 792
3	3	.	1	.	.	1	.	.	5 376
4	2	2	.	.	.	1	.	.	16 128
5	3	.	.	1	1	.	.	.	8 960
6	2	1	1	.	1	.	.	.	80 640
7	1	3	.	.	1	.	.	.	53 760
8	2	1	.	2	.	.	.	.	53 760
9	2	.	2	1	.	.	.	.	80 640
10	1	2	1	1	.	.	.	.	322 560
11	.	4	.	1	.	.	.	.	53 760
12	1	1	3	.	.	.	.	.	161 280
13	.	3	2	.	.	.	.	.	161 280
									1,000 000

Répartition N°	Nombre d'objets sur p en 0, 1, 2, ..., p exemplaires: 0	1	2	3	4	5	6	7	8	9(=p)	Probabilité de chaque répartition
1	4	.	.	.	.	.	.	.	.	1	0,000 002 56
2	3	1		.	.	.	.	.	1	.	92 16
3	3	.	1	.	.	.	.	1	.	.	368 64
4	2	2	.	.	.	.	.	1	.	.	1 105 92
5	3	.	.	1	.	.	1	.	.	.	860 16
6	2	1	1	.	.	.	1	.	.	.	7 741 44
7	1	3	.	.	.	.	1	.	.	.	5 160 96
8	3	.	.	.	1	1	.	.	.	.	1 290 24
9	2	1	.	1	.	1	.	.	.	.	15 482 88
10	2	.	2	.	.	1	.	.	.	.	11 612 16
11	1	2	1	.	.	1	.	.	.	.	46 448 64
12	.	4	.	.	.	1	.	.	.	.	7 741 44
13	2	1	.	.	2	.	.	.	.	.	9 676 80
14	2	.	1	1	1	.	.	.	.	.	38 707 20
15	1	2	.	1	1	.	.	.	.	.	77 414 40
16	1	1	2	.	1	.	.	.	.	.	116 121 60
17	.	3	1	.	1	.	.	.	.	.	77 414 40
18	2	.	.	3	.	.	.	.	.	.	8 601 60
19	1	1	1	2	.	.	.	.	.	.	154 828 80
20	.	3	.	2	.	.	.	.	.	.	51 609 60
21	1	.	3	1	.	.	.	.	.	.	77 414 40
22	.	2	2	1	.	.	.	.	.	.	232 243 20
23	.	1	4	.	.	.	.	.	.	.	58 060 80
											1,000 000 00

Répar-tition N°	Nombre d'objets sur p en 0, 1, 2, ... , p exemplaires													Probabilité de chaque répartition
	0	1	2	3	4	5	6	7	8	9	10	11	12(=p)	
1	4	.	.	.	.	.	.	.	.	.	.	.	1	0,000 000 020 48
2	3	1	.	.	.	.	.	.	.	.	.	1	.	983 04
3	3	.	1	.	.	.	.	.	.	.	1	.	.	5 406 72
4	2	2	.	.	.	.	.	.	.	.	1	.	.	16 220 16
5	3	.	.	1	.	.	.	.	.	1	.	.	.	18 022 40
6	2	1	1	.	.	.	.	.	.	1	.	.	.	162 201 60
7	1	3	.	.	.	.	.	.	.	1	.	.	.	108 134 40
8	3	.	.	.	1	.	.	.	1	.	.	.	.	40 550 40
9	2	1	.	1	.	.	.	.	1	.	.	.	.	486 604 80
10	2	.	2	.	.	.	.	.	1	.	.	.	.	364 953 60
11	1	2	1	.	.	.	.	.	1	.	.	.	.	1 459 814 40
12	.	4	.	.	.	.	.	.	1	.	.	.	.	243 302 40
13	3	.	.	.	.	1	.	1	.	.	.	.	.	64 880 64
14	2	1	.	.	1	.	.	1	.	.	.	.	.	973 209 60
15	2	.	1	1	.	.	.	1	.	.	.	.	.	1 946 419 20
16	1	2	.	1	.	.	.	1	.	.	.	.	.	3 892 838 40
17	1	1	2	.	.	.	.	1	.	.	.	.	.	5 839 257 60
18	.	3	1	.	.	.	.	1	.	.	.	.	.	3 892 838 40
19	3	.	.	.	.	.	2	.	.	.	.	.	.	37 847 04
20	2	1	.	.	.	1	1	.	.	.	.	.	.	1 362 493 44
21	2	.	1	.	1	.	1	.	.	.	.	.	.	3 406 233 60
22	1	2	.	.	1	.	1	.	.	.	.	.	.	6 812 467 20
23	2	.	.	2	.	.	1	.	.	.	.	.	.	2 270 822 40
24	1	1	1	1	.	.	1	.	.	.	.	.	.	27 249 868 80
25	.	3	.	1	.	.	1	.	.	.	.	.	.	9 083 289 60
26	1	.	3	.	.	.	1	.	.	.	.	.	.	6 812 467 20
27	.	2	2	.	.	.	1	.	.	.	.	.	.	20 437 401 60
28	2	.	1	.	.	2	.	.	.	.	.	.	.	2 043 740 16
29	1	2	.	.	.	2	.	.	.	.	.	.	.	4 087 480 32
30	2	.	.	1	1	1	.	.	.	.	.	.	.	6 812 467 20
31	1	1	1	.	1	1	.	.	.	.	.	.	.	40 874 803 20
32	.	3	.	.	1	1	.	.	.	.	.	.	.	13 624 934 40
33	1	1	.	2	.	1	.	.	.	.	.	.	.	27 249 868 80
34	1	.	2	1	.	1	.	.	.	.	.	.	.	40 874 803 20
35	.	2	1	1	.	1	.	.	.	.	.	.	.	81 749 606 40
36	.	1	3	.	.	1	.	.	.	.	.	.	.	40 874 803 20
37	2	.	.	.	3	.	.	.	.	.	.	.	.	1 419 264 00
38	1	1	.	1	2	.	.	.	.	.	.	.	.	34 062 336 00
39	1	.	2	.	2	.	.	.	.	.	.	.	.	25 546 752 00
40	.	2	1	.	2	.	.	.	.	.	.	.	.	51 093 504 00
41	1	.	1	2	1	.	.	.	.	.	.	.	.	68 124 672 00
42	.	2	.	2	1	.	.	.	.	.	.	.	.	68 124 672 00
43	.	1	2	1	1	.	.	.	.	.	.	.	.	204 374 016 00
44	.	.	4	.	1	.	.	.	.	.	.	.	.	7 569 408 00
45	1	.	.	4	.	.	.	.	.	.	.	.	.	90 832 896 00
46	.	1	1	3	.	.	.	.	.	.	.	.	.	68 124 672 00
47	.	.	3	2	.	.	.	.	.	.	.	.	.	25 546 752 00
														1,000 000 000 00

Il est encore possible de calculer le nombre moyen d'éléments en 0, 1, 2, ... exemplaires (valeur moyenne de p_0, p_1, p_2, ..., p_n). Toujours pour n = 5 on a :

	p = 7	p = 9	p = 12
$E(p_0)$ =	1,048 576	0,671 088 64	0,426 860 871 68
$E(p_1)$ =	1,835 008	1,509 949 44	1,008 083 927 04
$E(p_2)$ =	1,376 256	1,509 949 44	1,194 987 847 68
$E(p_3)$ =	0,573 440	0,880 803 84	1,360 889 446 40
$E(p_4)$ =	0,143 360	0,330 301 44	0,646 400 409 60
$E(p_5)$ =	0,021 504	0,082 575 36	0,265 751 101 44
$E(p_6)$ =	0,001 792	0,013 762 56	0,077 510 737 92
$E(p_7)$ =	0,000 064	0,001 474 56	0,016 609 443 84
$E(p_8)$ =	-	0,000 092 16	0,002 595 225 60
$E(p_9)$ =	-	0,000 002 56	0,000 288 358 40
$E(p_{10})$ =	-	-	0,000 021 626 88
$E(p_{11})$ =	-	-	0,000 000 983 04
$E(p_{12})$ =	-	-	0,000 000 020 48

L'inconvénient évident de la méthode est sa longueur. Le nombre des répartitions possibles croît très vite avec n et p. Aussi ne poursuivrons-nous pas plus avant la recherche dans cette voie. Au surplus, ce qui nous intéresse n'est pas l'ensemble des nombres p_0, p_1, p_2, ..., p_n, mais seulement p_0, le nombre d'éléments dont le collectionneur n'a aucun exemplaire. En effet, même si on envisage un échange des seconds, troisièmes, ... exemplaires, le nombre des éléments disponibles pour échange est :

$$\begin{aligned} d &= p_2 + 2p_3 + 3p_4 + \dots + (n-1)\,p_n \\ &= p - n + p_0 \end{aligned}$$

Seul p_0 apparaît en dehors de p et de n.

III - LOI DE PROBABILITE DU NOMBRE D'OBJETS MANQUANTS

Afin d'alléger l'écriture, on notera désormais x (au lieu de p_0) le nombre d'objets en zéro exemplaire.

x est une variable aléatoire qui, pour $p \geqslant n$ (nombre de paquets supérieur ou au moins égal au nombre d'objets constituant la collection complète), peut prendre les valeurs

$$0,\ 1,\ 2,\ \dots,\ n-1.$$

On se propose d'établir la probabilité de réalisation de chacune de ces valeurs.

Soit E_i l'événement "la pièce i ne figure pas parmi les p objets collectés". Sa probabilité est :

$$\Pr(E_i) = p_i = \left(\frac{n-1}{n}\right)^p, \text{ quel que soit } i = 1, 2, \dots, n.$$

Pour deux événements E_i et E_j définis de même on a :

$$Pr\ (E_i \text{ et } E_j) = p_{ij} = \left(\frac{n-2}{n}\right)^p \text{ , quels que soient i et j avec } i \neq j.$$

etc.

Si l'on pose maintenant :

$$S_1 = \sum_i p_i = C_n^1 \left(\frac{n-1}{n}\right)^p \qquad 1 \leqslant i \leqslant n$$

$$S_2 = \sum_{i,j} p_{ij} = C_n^2 \left(\frac{n-2}{n}\right)^p \qquad 1 \leqslant i < j \leqslant n$$

etc., la probabilité que x pièces manquent à la collection, c'est-à-dire que les p objets collectés ne comportent que n - x pièces différentes, est :

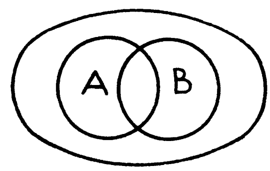
$$P\ (x) = S_x - C_{x+1}^x\ S_{x+1} + C_{x+2}^x\ S_{x+2} - \dots \pm C_n^x\ S_n$$

$$= C_n^x \left[C_{n-x}^0 \left(1-\frac{x}{n}\right)^p - C_{n-x}^1 \left(1-\frac{x+1}{n}\right)^p + C_{n-x}^2 \left(1-\frac{x+2}{n}\right)^p - \dots \pm C_{n-x}^{n-x} \left(1-\frac{n}{n}\right)^p\right] \qquad (1)$$

L'expression de P (x) en fonction des S_x, S_{x+1}, ..., S_n généralise le théorème dit des probabilités totales. On sait que pour deux événements A et B, par exemple, la probabilité que un et un seul soit réalisé est :

$$\begin{aligned} Pr\ (A\bar{B}) + Pr\ (\bar{A}B) &= \underbrace{Pr\ (A) + Pr\ (B)}_{S_1} - 2\ Pr\ (AB) \\ & - C_2^1\ S_2 \end{aligned}$$

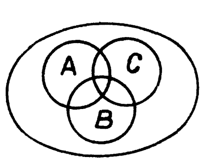
$A\bar{B}$ signifie A et non B, soit A seul
$\bar{A}B$ signifie B et non A, soit B seul



A se compose de (A et B) et de (A et $\bar{B}$)
B " " " (A et B) et de ($\bar{A}$ et B)

De même, pour trois événements A, B, C, la probabilité que un et un seul soit réalisé est :

$$\begin{aligned} Pr\ (A\bar{B}\bar{C}) + Pr\ (\bar{A}B\bar{C}) + Pr\ (\bar{A}\bar{B}C) &= Pr\ (A) + Pr\ (B) + Pr\ (C) \\ &\quad - 2\ [Pr\ (AB) + Pr\ (AC) + Pr\ (BC)] \\ &\quad + 3\ Pr\ (ABC) \\ &= S_1 - C_2^1\ S_2 + C_3^1\ S_3, \end{aligned}$$



tandis que la probabilité que deux et deux seulement soient réalisés est :

$$\begin{aligned} Pr\ (AB\bar{C}) + Pr\ (A\bar{B}C) + Pr\ (\bar{A}BC) &= Pr\ (AB) + Pr\ (AC) + Pr\ (BC) - 3\ Pr\ (ABC) \\ &= S_2 - C_3^2\ S_3 \end{aligned}$$

Grâce au coefficient $(-1)^r\ C_{x+r}^x = (-1)^r \frac{(x+r)!}{x!\,r!}$ de S_{x+r} on compte une fois et une seule la probabilité de réalisation de x événements simultanément.

Exemple : Pour n = 5 et p = 7, on a :

$$P(0) = 1 \quad [1-5(0,8)^7 + 10(0,6)^7 - 10(0,4)^7 + 5(0,2)^7] = 0,215\ 040$$
$$P(1) = \ \ 5 \ [\ (0,8)^7 - 4\ (0,6)^7 + 6\ (0,4)^7 - 4\ (0,2)^7\ ] = 0,537\ 600$$
$$P(2) = 10\ [\ (0,6)^7 - 3\ (0,4)^7 + 3\ (0,2)^7\ ] = 0,231\ 168$$
$$P(3) = 10\ [\ (0,4)^7 - 2\ (0,2)^7\ ] = 0,016\ 128$$
$$P(4) = \ \ 5\ [\ (0,2)^7\ ] = 0,000\ 064$$
$$1,000\ 000$$

Ce sont les probabilités que l'on obtient par récapitulation des lignes du tableau 1 par valeur de p_0.

Donnons encore les valeurs des P (x) pour p = 9 et 12.

	p = 9	p = 12
x = 0	P (x) = 0,427 069 44	P (x) = 0,594 739 200 00
1	477 388 80	383 828 459 52
2	092 928 00	021 264 629 76
3	002 611 20	000 167 690 24
4	000 002 56	000 000 020 48
	1,000 000 00	1,000 000 000 00

Pour n et p grands le calcul de l'expression (1) de P(x) est assez lourd. On montre (*) que, pourvu que n et p soient entre eux dans un rapport fini, c'est-à-dire que le nombre moyen d'exemplaires de chacun des n objets, p/n, ne soit ni très petit ni très grand, on a pour P (x) la valeur approchée :

$$P(x) = e^{m} \frac{m^{x-m}}{x!} \quad \text{où } m = n\,e^{-p/n} \qquad (2)$$

Ainsi, x suit approximativement la loi de Poisson de paramètre m.

On peut se faire une idée de l'erreur introduite par l'emploi de l'expression approchée (2) plutôt que de l'expression exacte (1) de P (x) en comparant les valeurs présentées au tableau 2 pour n = 50 et diverses valeurs de p.

IV - QUELQUES RESULTATS NUMERIQUES

On trouve aux tableaux 2, 3 et 4 les probabilités cumulées, c'est-à-dire les probabilités d'avoir x ou moins de x pièces manquantes lors de la réalisation d'une collection de n = 50, 100, 200 pièces, quand on a acheté p paquets.

Les figures 1 à 3 construites à partir de ces tableaux montrent comment croît la probabilité d'avoir x ou moins de x pièces manquantes quand p augmente. Ainsi, on lit sur le graphique 1 établi pour n = 50, en suivant la courbe x = 0 que la probabilité de n'avoir aucune pièce manquante est seulement de 2/100 pour p = 130, de 1/2 pour p = 210 environ, de 9/10 pour p = 310. On voit de même, à l'aide de la courbe x = 3, que la probabilité de ne pas avoir plus de trois pièces manquantes sera de 1/2 si on a acheté p = 130 paquets.

Ces tableaux et graphiques ont été établis en utilisant l'approximation de Poisson mentionnée plus haut et en interpolant la table de la loi de Poisson figurant dans les Biometrika Tables for Statisticians, vol 1. On reviendra plus loin sur la précision de cette approximation.

(*) Voir Feller, Probability Theory and its Applications, p. 70 à 73.

V - RECHERCHE D'UNE REGLE D'ECHANGE DES EXEMPLAIRES SECONDS, TROISIEMES, ...

On voit sur les tableaux et graphiques que le nombre de paquets à acheter pour avoir des chances raisonnables de terminer la collection croît très vite quand n augmente. Ainsi, si l'on se contente d'une chance sur deux, il faut environ :

p = 210 paquets si n = 50
p = 500 - si n = 100
p = 1 150 - si n = 200

Il en faudrait beaucoup plus si l'on voulait une quasi certitude de 0, 90 ou 0, 95.

Heureusement pour les collectionneurs, cette difficulté est réduite par la possibilité d'échanger les exemplaires seconds, troisièmes, ... d'une même pièce contre une dont ils ne disposent pas encore. Il suffit alors d'un nombre beaucoup plus restreint de pièces.

Considérons, par exemple, le cas d'un collectionneur disposant de 80 pièces au total alors que la collection comprend 50 éléments dont 10 ne sont pas représentés parmi ses 80 : n = 50, p = 80, x = 10. Gardant un seul exemplaire des 40 éléments représentés dans les 80, le collectionneur dispose donc de 80 - 40 (= p - n + x) exemplaires seconds, troisièmes, ... dont il demande l'échange. Si cet échange se fait à raison de 1 pour 4, il a terminé sa collection.

On a repris dans les graphiques 4, 5, 6, les courbes des graphiques 1 à 3 pour les valeurs de p peu supérieures à n et indiqué sur ces courbes à partir de quelle valeur de p la collection était terminée suivant les modalités d'échange : 1 pour 2, 3, 4, 5, 6, 8, 10, 12, 15. Ainsi le 4 marqué sur la courbe x = 10, pour p = 80, de la figure 4 (n = 50), indique que le collectionneur a achevé sa collection. Pour p = 83, on peut se permettre x = 11 pièces manquantes. On aura, en effet, 83 + 11 - 50 = 44 pièces disponibles pour échange. Et ainsi de suite de 3 en 3 si le taux d'échange est 1 pour 4. On voit alors, en suivant la courbe des 4 que, avec p = 70 paquets, on n'a qu'une probabilité de 3 à 4/100 d'avoir terminé, avec 71 8 à 9/100, ..., avec 79 41/100, avec 80 57/100, ..., avec 90 96/100 environ... Plus le nombre de pièces à fournir pour en avoir une nouvelle est faible, plus grandes sont les chances d'avoir achevé la collection avec un nombre p donné de paquets.

A titre indicatif, si on veut avoir une chance sur deux de terminer, il faudra :

	si n = 50	si n = 100	si n = 200
pour 1 contre 2	p = 63 ou 4	p = 132	.....
3	73	146	.....
4	79 ou 80	160	p = 320
5	85 ou 6	171	343 ou 4
6	90	180	362
8	98 ou 9	197	395
10	104	208	417

Contrairement à ce qui se passait sans échange, les valeurs de p sont à peu près proportionnelles à n (à taux d'échange constant).

Si on veut avoir 95 chances sur 100 de terminer, il faudra :

	si n = 50		si n = 100	si n = 200
pour 1 contre 2	p = 68 ou 9		p = 143	...
3	80	(77 ou 8)	157	...
4	89	(86)	174 ou 5	p = 336
5	97	(93 ou 4)	187 ou 8	365
6	104	(99 ou 100)	199 ou 200	388
8	...		...	424

[Entre parenthèses, dans la collonne "si n = 50", les valeurs exactes de p, c'est-à-dire en utilisant la formule (1) et non l'approximation de Poisson]. A taux d'échange constant, le nombre minimum d'objets à réussir, p, croît légèrement moins vite que proportionnellement à n.

VI - VALEUR DE L'APPROXIMATION DE POISSON

Toutes les probabilités indiquées dans les tableaux 2 à 4 et dans les graphiques en découlant (fig. 1 à 6) ont été obtenues par l'approximation de Poisson [formule (2), page 71]. Quelle est l'erreur, en valeur et en signe, introduite par cette approximation ? Afin de le savoir, on a calculé les probabilités exactes (formule (1), page 70) pour n = 50 et p = 70, 80, 90 et 100, soit les valeurs intéressantes de ce paramètre.

La difficulté dans l'utilisation de la formule (1) réside dans la nécessité de calculer les puissances p-ièmes de $1 - (x + k)/n$, $k = 0, 1, 2, \ldots$, avec une grande précision : jusqu'à 12 chiffres significatifs dans certains cas. Voici, à titre d'exemple, le calcul de P (4) pour p = 100 :

	Coefficient binomial C^k_{n-x}	Puissance $\left(1 - \frac{x+k}{n}\right)^p$	Produit pris avec son signe
k = 0	1	10^{-4}. 2,39211 87465	0,00023 92118 75
1	46	10^{-5}. 2,65613 98888	- 122 18243 49
2	1035	(-6). 2,80716 03112	290 54109 22
3	15180	(-7). 2,81737 80347	- 427 67798 57
4	1 63185	(-8). 2,67872 79317	437 12821 75
5	13 70754	(-9). 2,40649 65221	- 329 87147 34
6	93 66819	(-10). 2,03703 59763	190 80547 29
7	535 24680	(-11). 1,61979 58019	- 86 69905 20
8	2609 32815	(-12). 1,20603 31021	31 46936 12
9	11017 16330	(-14). 8,37860 99490	- 9 23085 14
10	40763 50421	(-15). 5,41065 25116	2 20557 16
11	1 33407 83196	(-16). 3,23447 65096	- 43150 45
12	3 89106 17655	(-17). 1,78193 25894	6933 61
13	10 17662 30790	(-19). 9,00313 06848	- 916 21
14	23 98775 44005	(-20). 4,14951 55688	99 54
15	51 17387 60544	(-21). 1,73447 86157	- 8 88
16	99 14938 48554	(-23). 6,53318 62350	65
17	174 96950 26860	(-24). 2,22019089478	- 4
		Total	= 0,00000 03868 77
			× 230300 (soit C^4_{50})
			P (4) = 0,08910

Les coefficients binomiaux ont été calculés de proche en proche. Les puissances ont été calculées par logarithmes : lecture du log. de $1 - (x+k)/n$ avec 16 décimales dans "Logarithmetica Britannica", multiplication par p, arrondissement à 13 décimales et recherche de l'antilog. dans la table de Andoyer.

Les résultats sont présentés numériquement au tableau 5 où l'on a rappelé l'approximation de Poisson (entre parenthèses) et graphiquement (fig. 7). On voit que, quel que soit p (= 70, 80, 90 ou 100), la loi exacte de x et son approximation de Poisson ont sensiblement même valeur centrale (médiane) mais que l'approximation de Poisson est un peu plus dispersée surtout vers les grandes valeurs de x. Il en résulte que si l'on cherche la valeur de p telle que l'on ait une chance sur deux d'avoir achevée la collection par l'achat de p paquets l'approximation de Poisson est très bonne. Si l'on cherche la valeur de p donnant 90 ou 95 chances sur 100, l'approximation de Poisson donne une valeur trop grande de quelques unités. On a établi la figure 8 analogue à la figure 4, mais avec les valeurs exactes des probabilités et construit la figure 9 qui permet de comparer les valeurs de p suivant différents taux d'échange des seconds, troisièmes, ... exemplaires contre des pièces encore manquantes. On lit sur cette figure 9 que si l'on se contente d'une chance sur deux de terminer la collection ou si au contraire on veut une quasi-certitude de 0,95, il faut les valeurs suivantes de p :

	Prob. $\geqslant$ 1/2	Prob. $\geqslant$ 0,95
1 pour 3	p = 73	p = 78 (80)
1 " 4	80	86 (89)
1 " 5	86	94 (98)

(entre parenthèses, valeurs de p par approximation de Poisson).

En conclusion, eu égard à la facilité de calcul de l'approximation de Poisson et à la faible erreur qu'elle introduit, elle semble très acceptable.

VII - CONCLUSION

Les résultats obtenus intéressent le collectionneur. Ils lui permettent d'apprécier les chances qu'il a de terminer la collection compte-tenu de la fréquence de ses achats. Il est à noter qu'entre le nombre de paquets à partir duquel la probabilité de réussite cesse d'être pratiquement nulle et le nombre à partir duquel la réussite est quasi-certaine, la différence est assez faible, comparée au nombre moyen de paquets, d'autant plus faible que le taux d'échange est plus avantageux. Par exemple, pour n = 50 et un taux d'échange de 1 contre 4, avec 70 paquets on a moins d'une chance sur 100 -il est pratiquement impossible- de réussir, tandis qu'avec 89 paquets on a plus de 99 chances sur 100 -on est pratiquement certain- de réussir.

Tableau 2

Probabilité d'avoir x ou moins de x pièces manquentes
quand on dispose de p objets et que n = 50
(approximation de Poisson)

p	50	60	70	80	90	100	120	150	200	250	300
$m = n\,e^{-p/n}$	18,39	15,06	12,33	10,10	8,265	6,767	4,536	2,490	0,916	0,337	0,124
x = 0						0,001	0,011	0,083	0,41	0,71	0,88
1					0,002	009	059	29	77	95	99
2				0,003	011	035	17	55	94	99	
3			0,002	010	035	095	34	76	99		
4		0,001	006	029	085	20	53	89			
5		003	017	064	17	33	71	96			
6	0,001	007	038	12	28	49	83	99			
7	002	017	076	21	42	64	91				
8	006	036	13	32	56	76	96				
9	012	068	22	45	68	85	98				
10	025	12	31	57	79	92	99				
11	046	18	42	68	87	96					
12	078	26	54	78	92	98					
13	12	36	65	86	96	99					
14	18	46	74	91	98						
15	26	56	82	95	99						
16	34	66	88	97							
17	43	74	92	98							
18	52	82	95	99							
19	62	87	97								
20	70	91	98								

Tableau 3

Probabilité d'avoir x ou moins de x pièces manquantes quand on dispose de p objets et que n = 100 (approximation de Poisson)

p	100	120	140	160	180	200	240	300	400	500	600
$m = n\,\bar{e}^{p/n}$	36,79	30,12	24,66	20,19	16,53	13,53	9,072	4,979	1,832	0,674	0,248
x = 0								0,007	0,16	0,51	0,78
1							0,001	040	45	85	97
2							006	12	72	97	
3						0,001	020	27	88	99	
4						003	052	44	96		
5					0,001	008	11	62	99		
6					003	019	20	76			
7				0,001	007	041	31	87			
8				002	016	078	44	93			
9				004	034	13	58	97			
10			0,001	009	062	21	69	99			
11			002	020	10	30	79				
12			004	036	16	41	87				
13			008	061	24	52	92				
14		0,001	015	10	32	62	96				
15		002	026	15	42	72	98				
16		004	044	21	52	80	99				
17		007	069	28	61	86					
18		012	10	37	70	91					
19	0,001	021	15	45	77	94					
20	002	034	20	54	84	96					
21	003	052	27	63	89	98					
22	006	078	34	70	92	99					
23	010	11	42	77	95						
24	017	15	50	83	97						
25	026	20	58	88	98						
26	040	26	65	91	99						
27	062	32	72	94							
28	082	39	78	96							
29	11	47	84	97							
30	15	54	88	98							

Tableau 4

Probabilité d'avoir x ou moins de x pièces manquantes quand on dispose de p objets et que n = 200
(approbation de Poisson)

	280	320	360	400	480	600	800	1 000	1 200
$m = n\,\bar{e}^{p/n}$	49,3	40,4	33,1	27,1	18,1	9,96	3,66	1,35	0,50
x = 0							0,026	0,26	0,61
1						0,001	12	61	91
2						003	29	84	99
3						011	50	95	
4						030	69	99	
5						069	83		
6					0,001	13	92		
7					003	22	97		
8					006	34	99		
9					014	46			
10					028	59			
11					051	70			
12				0,001	083	80			
13				002	13	87			
14				004	20	92			
15				008	27	95			
16				014	36	97			
17			0,001	026	45	99			
18			003	043	55				
19			006	067	64				
20			010	10	72				
21			017	14	79				
22		0,001	027	19	84				
23		002	042	25	89				
24		004	063	32	92				
25		006	089	39	95				
26		010	12	47	97				
27		017	17	54	98				
28		026	22	62	99				
29	0,001	039	27	69					
30	002	055	33	75					
31	004	076	40	80					
32	005	11	47	85					
33	009	14	54	89					
34	013	17	61	92					

Tableau 5

Probabilité d'avoir x ou moins de x pièces manquantes
quand on dispose de p objets et que n = 50
Valeur exacte et approximation de Poisson entre parenthèses

	p = 70		p = 80		p = 90		p = 100	
x = 0	0,000		0,000		0,000		0,000	(0,001)
1	0		0		0	(0,002)	2	(0,009)
2	0		0	(0,003)	2	(0,011)	13	(0,035)
3	0	(0,002)	1	(0,010)	10	(0,035)	52	(0,095)
4	0	(0,006)	5	(0,029)	37	(0,085)	141	(0,20)
5	1	(0,017)	18	(0,064)	103	(0,17)	292	(0,33)
6	5	(0,038)	55	(0,12)	222	(0,28)	485	(0,49)
7	17	(0,076)	131	(0,21)	391	(0,42)	676	(0,64)
8	49	(0,13)	258	(0,32)	580	(0,56)	827	(0,76)
9	116	(0,22)	426	(0,45)	749	(0,68)	922	(0,85)
10	230	(0,31)	607	(0,57)	872	(0,79)	971	(0,92)
11	387	(0,42)	766	(0,68)	945	(0,87)	991	(0,96)
12	564	(0,54)	880	(0,78)	980	(0,92)	998	(0,98)
13	728	(0,65)	947	(0,86)	994	(0,96)	999	(0,99)
14	853	(0,74)	981	(0,91)	999	(0,98)	1,000	
15	932	(0,82)	994	(0,95)	1,000	(0,99)		
16	973	(0,88)	998	(0,97)				
17	991	(0,92)	1,000	(0,98)				
18	998	(0,95)		(0,99)				

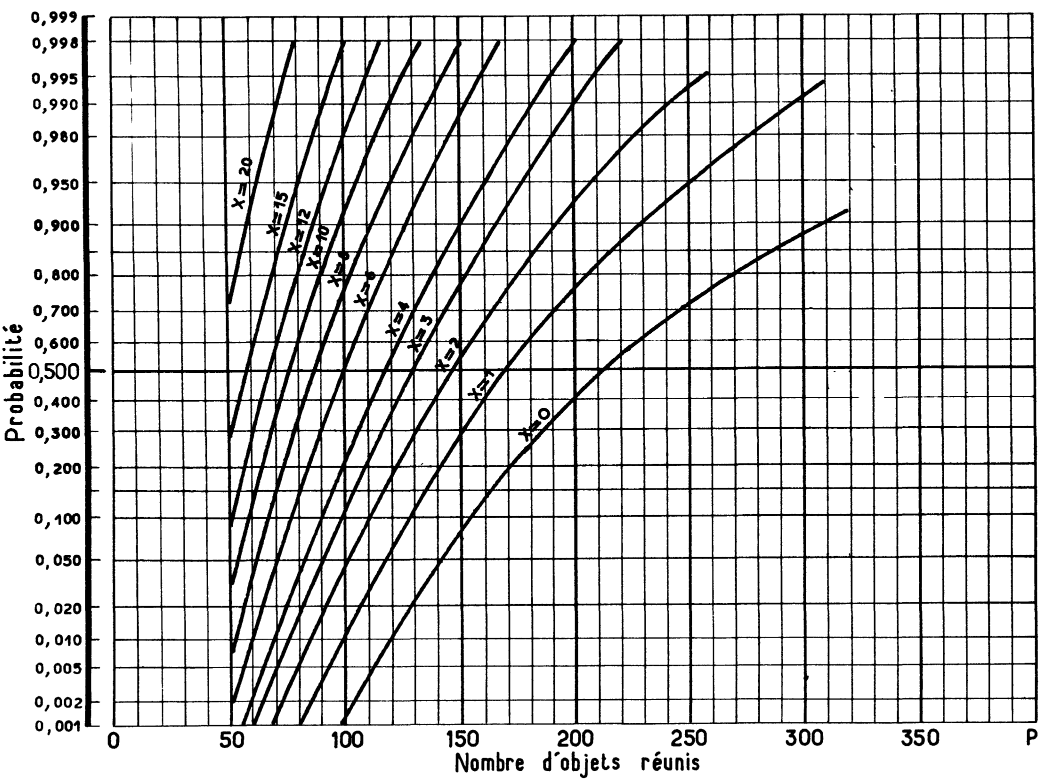


Fig. 1 - Probabilité de x ou moins de x pièces manquantes en fonction de p pour n = 50. (approximation de Poisson - données du tableau 2)

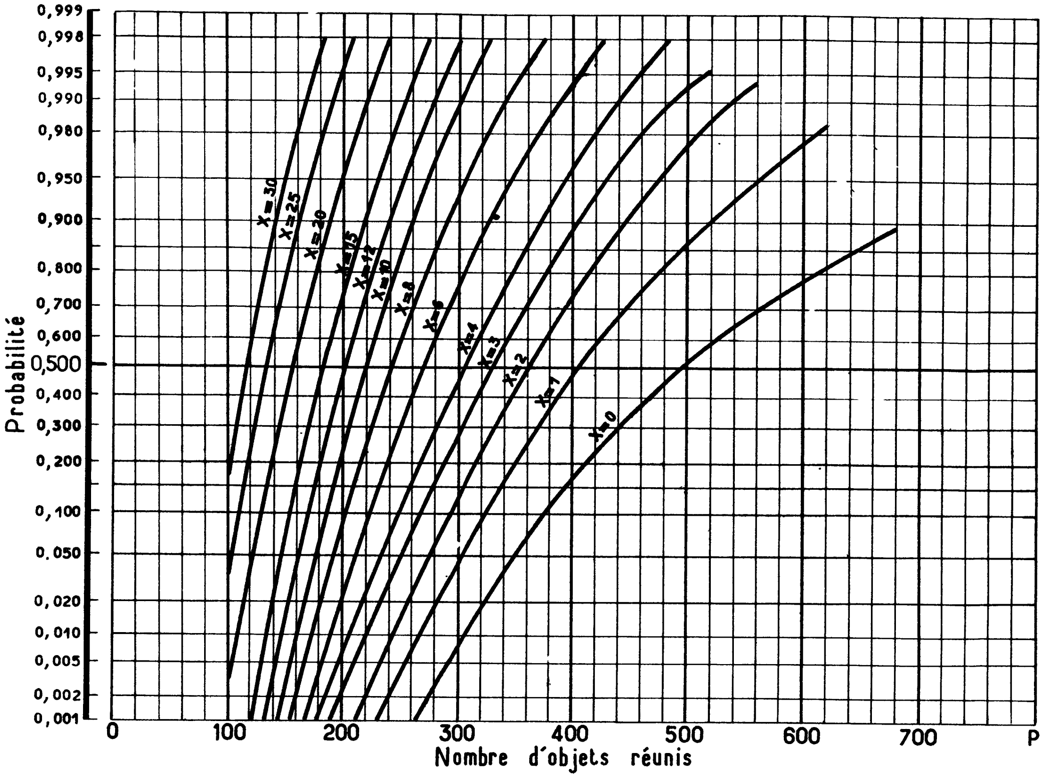


Fig. 2 - Probabilité de x ou moins de x pièces manquantes en fonction de p pour n = 100. (approximation de Poisson - données du tableau 3)

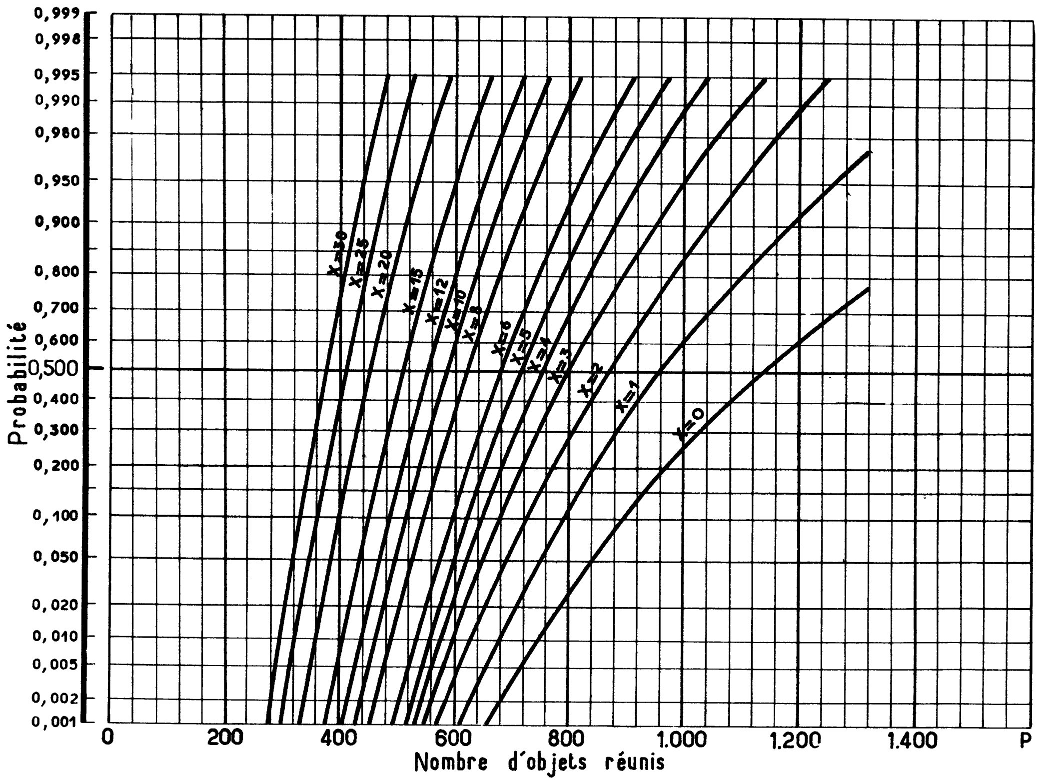


Fig. 3 - Probabilité de x ou moins de x pièces manquantes en fonction de p pour n = 200. (approximation de Poisson - données du tableau 4)

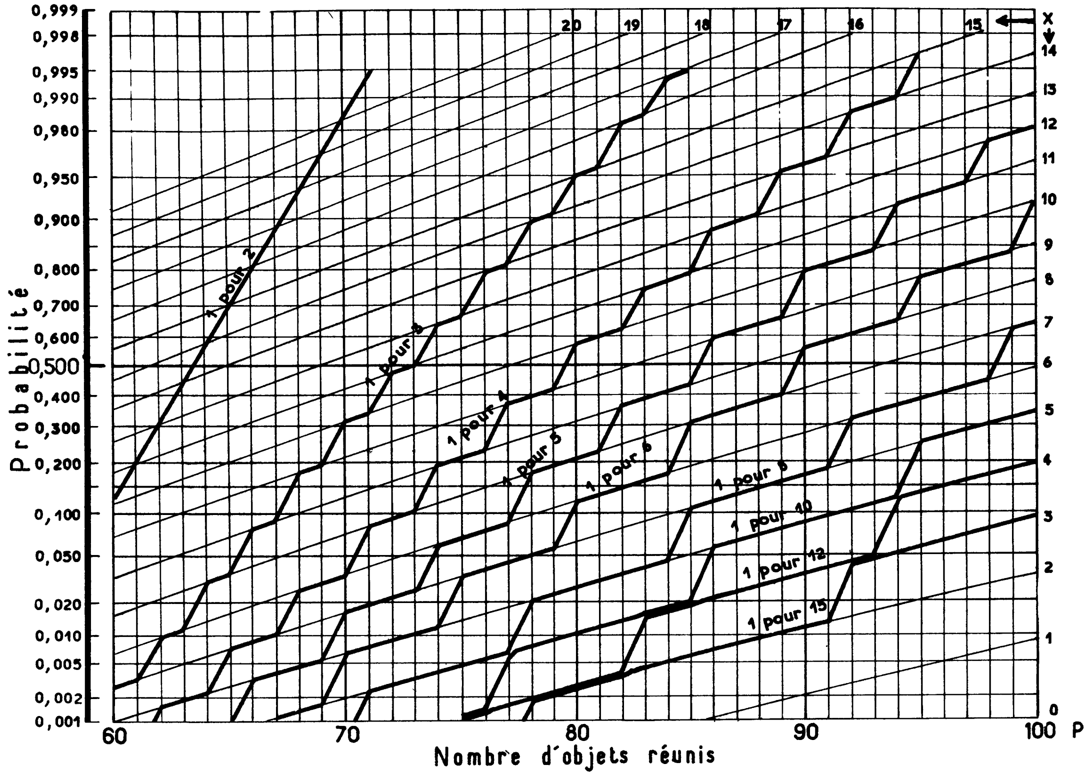


Fig. 4 - Probabilité d'avoir achevé la collection par échange de pièces à raison de 1 pour 2, 3, 4, 5, 6, 8, 10, 12 ou 15 en fonction de p quand n = 50. (approximation de Poisson comme Fig. 1).

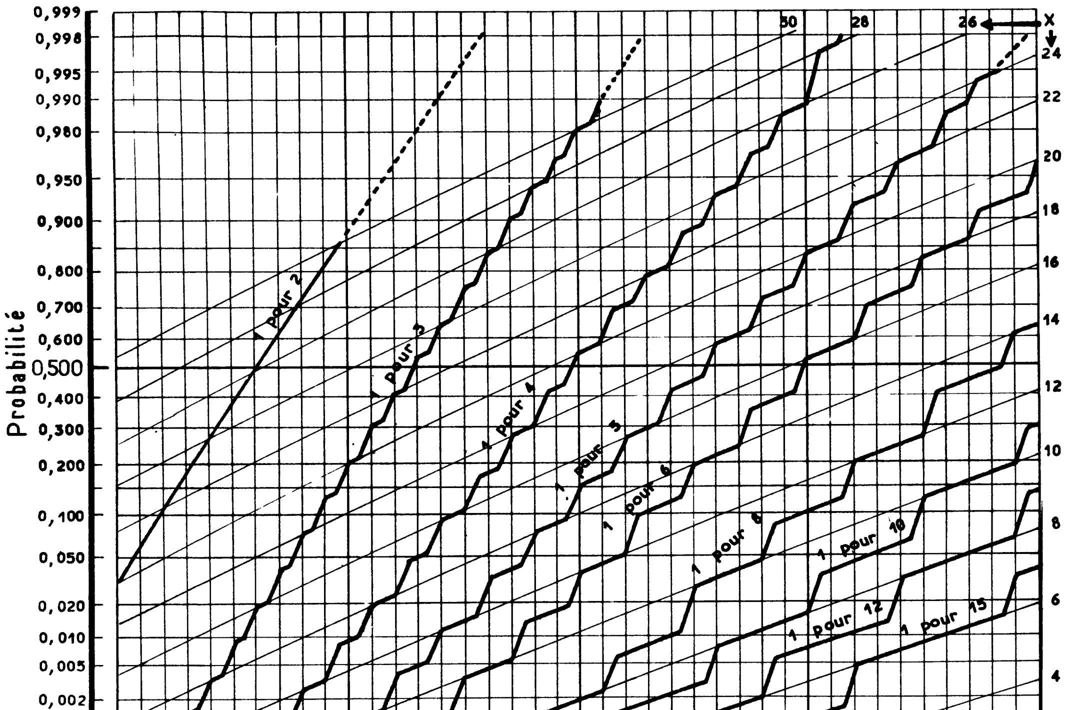


Fig. 5 - Probabilité d'avoir achevé la collection par échange de pièces à raison de 1 pour 2, 3, 4, 5, 6, 8, 10, 12 ou 15 en fonction de p quand n = 100. (approximation de Poisson comme Fig. 2).

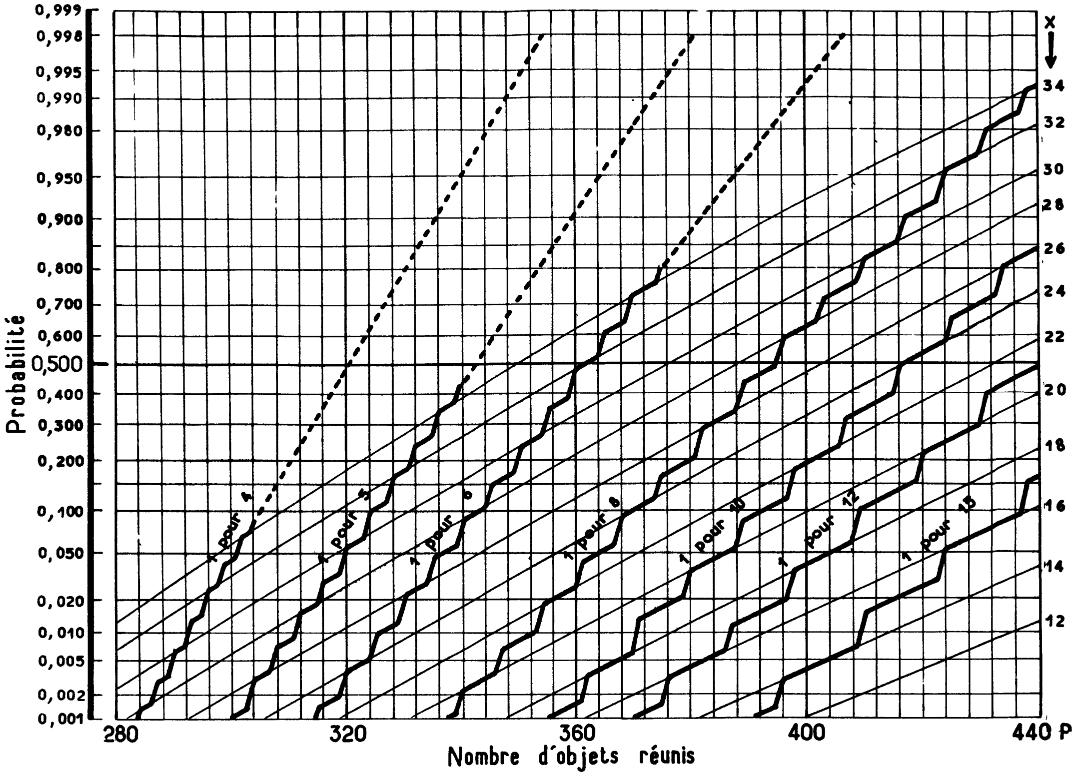


Fig. 6 - Probabilité d'avoir achevé la collection par échange de pièces à raison de 1 pour 4, 5, 6, 8, 10 ou 15 en fonction de p quand n = 200. (approximation de Poisson comme Fig. 3).

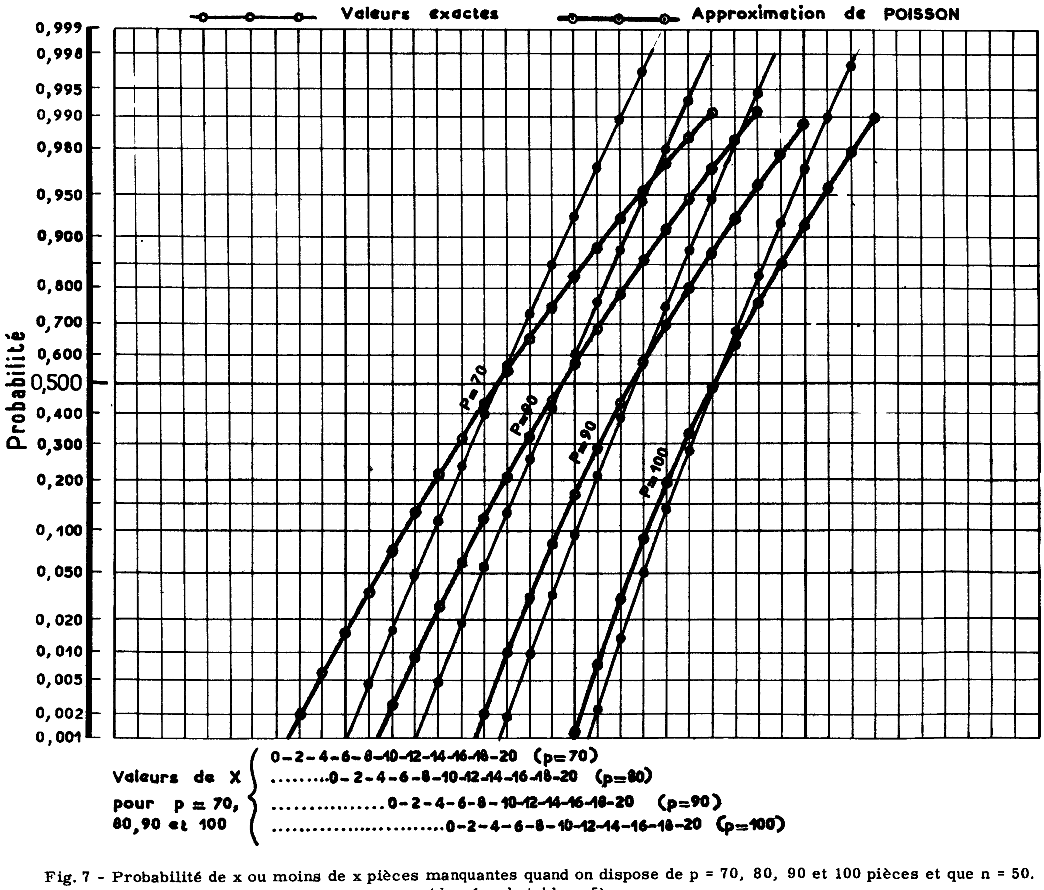


Fig. 7 - Probabilité de x ou moins de x pièces manquantes quand on dispose de p = 70, 80, 90 et 100 pièces et que n = 50. (données du tableau 5)

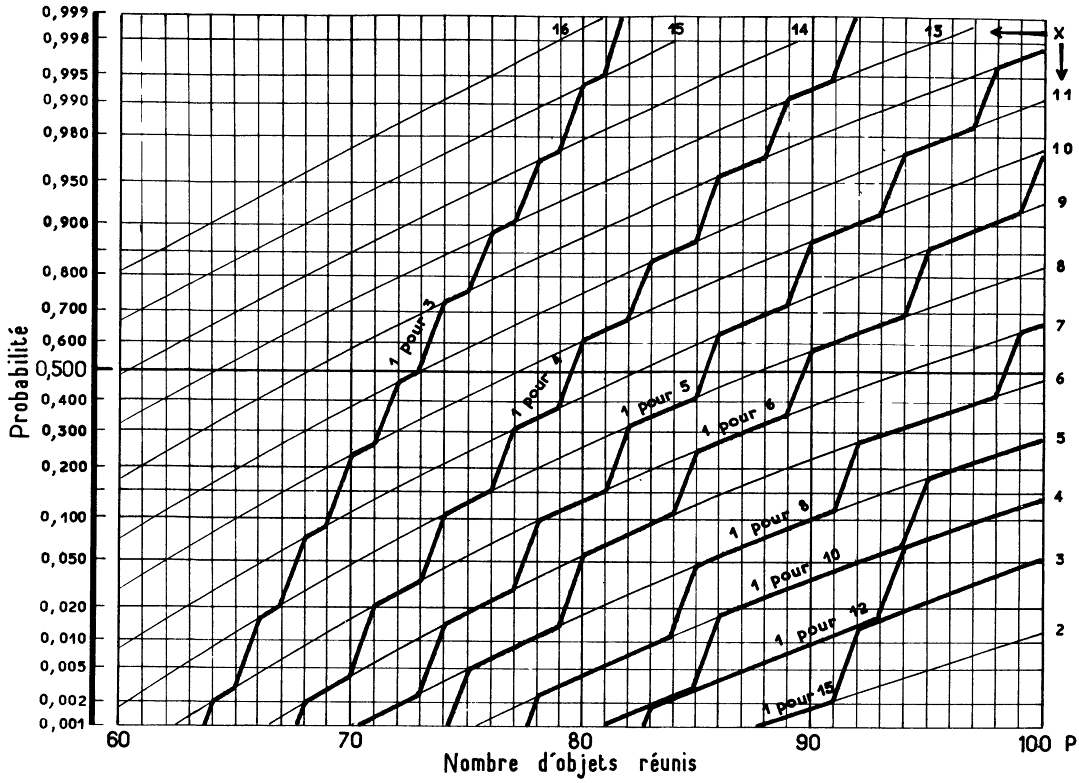


Fig. 8 - Probabilité d'avoir achevé la collection par échange de pièces à raison de 1 pour 3, 4, 5, 6, 8, 10, 12 ou 15 en fonction de p quand n = 50 (valeurs exactes) (Figure à rapprocher de la Fig. 4 qui montre l'approximation de Poisson).
N. B. - Pour p inférieur à 70 les courbes ont été tracées à vue.

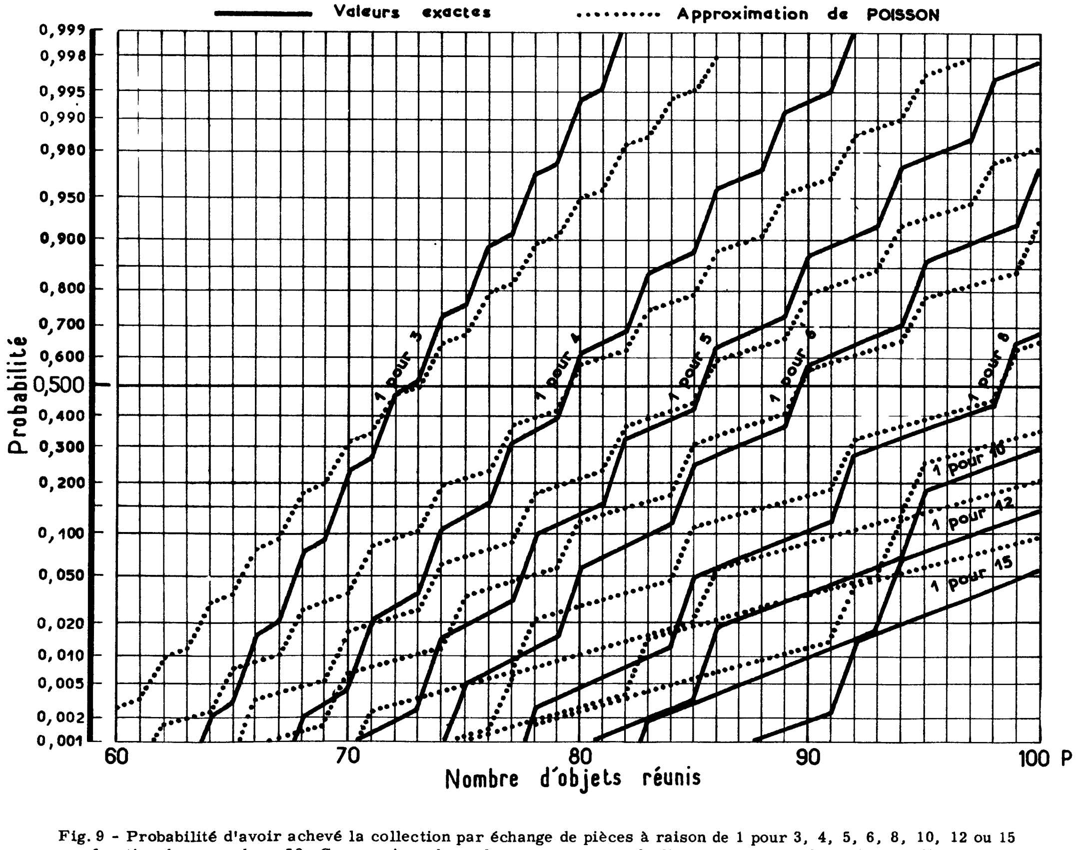


Fig. 9 - Probabilité d'avoir achevé la collection par échange de pièces à raison de 1 pour 3, 4, 5, 6, 8, 10, 12 ou 15 en fonction de p quand n = 50. Comparaison des valeurs exactes et de l'approximation de Poisson (Fig. 4 et 8).