REVUE DE STATISTIQUE APPLIQUÉE

P. THIONET

Sur l'approximation d'une distribution par une loi limite

Revue de statistique appliquée, tome 16, n° 2 (1968), p. 5-20

<http://www.numdam.org/item?id=RSA_1968__16_2_5_0>

SUR L'APPROXIMATION D'UNE DISTRIBUTION PAR UNE LOI LIMITE

P. THIONET
Faculté des Sciences de Poitiers

En Statistique Appliquée, on a besoin de substituer à certaines distributions plus ou moins compliquées de probabilités, des distributions de type courant leur ressemblant suffisamment.

On suppose qu'on a affaire à une distribution D appartenant à une famille L_1, L_2, ..., L_n convergeant vers une loi limite L de type courant, quand n tend vers l'infini. Si D correspond à L_n, où n est assez grand, L pourra remplacer avantageusement D. Mais il arrive qu'en fait n soit trop petit : un cas intéressant est celui où une autre famille de lois P_1, P_2, ..., P_n, celles-là d'un type banal, converge vers une loi limite P identique à L ; on peut essayer alors de s'arranger pour que L_n soit très proche de P_n bien qu'encore éloignée de L. Alors P_n peut remplacer L_n dans les applications. Dans ce qui suit on suppose que la loi L (ou P) est de Poisson.

Exemple étudié en détail : Les P_n sont des lois de Poisson, de même moyenne que L_n. Ceci suppose que L_n a au moins ses deux premiers cumulants quasi-égaux.

Le cas où la variance est bien inférieure à la moyenne a été étudié par Katz, qui a pris pour P_n une famille de lois binomiales convergeant vers L (l'ajustement de P_n sur L_n assure l'égalité des moyennes et aussi celle des variances).

Dans le meme esprit, nous avons étudié une distribution du nombre de rencontres, donnée par Barton, et vérifié que l'approximation par une loi binomiale était meilleure que celle par la loi de Poisson indiquée alors.

Dans le cas inverse, il est proposé de prendre pour P_n des lois binomiales négatives (variances plus grandes que moyennes).

P. Thionet

La présente note n'a aucune prétention d'ordre théorique. Nous présenterons un fait observé à l'occasion d'un travail d'initiation à la recherche, puis nous essaierons de faire le point des explications (très classiques) qu'il a été possible de rassembler

1ere PARTIE : PRESENTATION D'UN CAS PEU CONNU DE CONVERGENCE

1/ Nous avions confié à M. FANNECHERE comme Diplôme d'Etudes Supérieures, la tache d'explorer quelques prolongements d'un récent travail sur les signes des différences premières d'une suite de N nombres distincts soumis à N! permutations [1]; une séquence de signes + encadrée de signes - ou l'inverse, s'appelle un train ou run (run up and down, plus précisément).

1.1 - En particulier, pour N encore petit, nous connaissons la distribution du nombre t_r de runs de longueur r (et disposons d'un procédé d'itération permettant de passer de N à N + 1).

Tableau 1

Cas de N = 10 et 11 : Distribution du nombre t_r de runs de longueur r = 1 ou 2 (Nombres de permutations)

	r = 1	r = 2	r = 1	r = 2
t = 10			353 792	
9	50 521		0	
8	0		1 995 181	
7	252 750		756 148	
6	94 175		4 178 411	
5	449 799		2 633 982	45 541
4	268 013	45 541	4 267 984	944 072
3	362 904	281 880	2 758 868	3 861 707
2	202 569	626 428	1 990 771	6 978 179
1	107 314	622 380	831 338	6 047 008
0	26 355	238 171	202 925	2 081 893
Total N! =	1 814 400	1 814 400	19 958 400	19 958 400
Espérances mathématiques	$Et_1 = 4,25$	$Et_2 = 1,60$	$Et_1 = 4,666$	$Et_2 = 1,783$

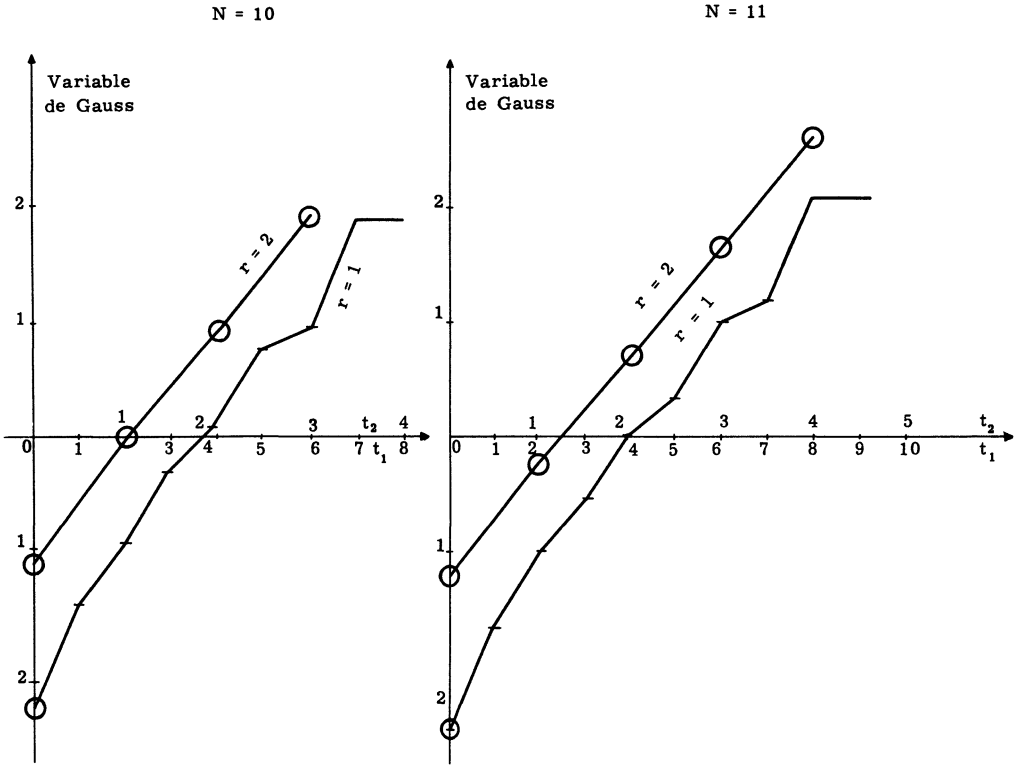
Le tableau 1 reproduit, pour N = 10 et N = 11, les distributions de t_1 et t_2 (celles t_r, r = 3, 4, 5, sont données un peu plus loin pour N = 11).

On voit par exemple que la probabilité d'avoir $t_1 = 0$ est

1,4 % = 26 335/1 814 400 pour N = 10
1,0 % = 202 925/19 958 400 pour N = 11

1.2 - En 1944 [2] Wolfowitz a établi que, quand N tend vers l'infini, chaque variable t_r admet une distribution-limite de Laplace-Gauss. Il est assez clair que, pour N = 10 ou 11, on se trouve encore très loin de l'état limite. Et pourtant la méthode de la droite de Henry ne donne

pas des résultats trop mauvais (voir fig. 1) pour r = 1; et les résultats sont excellents pour r = 2.



1.3 - Dans le même article, Wolfowitz montre que, si l'on choisit r un peu plus grand chaque fois que N augmente, de façon à laisser constant le rapport $N/(r+1)!$, alors la variable t_r admet pour loi limite une loi de Poisson de paramètre $\lambda = \lim 2N/(r+1)!$

Ceci veut dire que, si N = 11 était assez grand, les distributions de t_1 et t_2 seraient déjà voisines de distributions de Poisson de paramètres respectifs :

$$\left|\begin{array}{l} \lambda\ (11\ ;\ 1) = 22/2! = 11 \\ \lambda\ (11\ ;\ 2) = 22/3! = 3,66.. \end{array}\right.$$

Il ne faudrait pas s'étonner, bien entendu, d'une double ressemblance (avec Laplace-Gauss et avec Poisson) qui est le cas d'autres distributions (binomiales entre autres). Mais il ne s'agit que de ressemblances très vagues. Pour r = 3, λ (11 ; 3) = 22/24 = 0,9, la comparaison donne ce qui suit :

N = 11 | t_3 =	0	1	2	3	4 et plus
Vraie probabilité	0,60662	0,33860	0,05301	0,00177	0
Probabilité (Loi de Poisson) λ = 0,9	0,40657	0,36591	0,16466	0,04939	0,01347

Il faut avouer que la ressemblance est médiocre : N est trop petit.

2. AJUSTEMENT D'UNE LOI DE POISSON MEILLEURE QUE LA LOI LIMITE

2.1 - Or il est facile de voir que la distribution de t_3 a l'allure d'une distribution de Poisson. Si on devait procéder à l'ajustement d'une loi de Poisson sur l'échantillon des N! observations distribuées comme t_3, on choisirait pour λ la moyenne des t_3 (ajustements par la méthode des moments et du maximum de vraisemblance).

Dans le cas présent, la moyenne Et_3 est 0,45, qu'on obtient :

soit en faisant N = 11, r = 3 dans la formule (cf. [1], page 14) :

$$Et_r = \frac{2}{(r+3)!}\,[(N-r)\,(r^2+3r+1)+2(r+2)], \qquad 1 < r \leqslant N-2$$

soit directement :

	N = 11, r = 3
$t_3 = 3$	35 380
2	1 058 221
1	6 758 698
0	12 106 101
Total : N! =	19 958 400

(C'est de cette répartition des N! combinaisons que sont tirées les valeurs décimales des probabilités données plus haut).

Les tables de la Loi de Poisson, pour $\lambda = 0,5$ et $\lambda = 0,4$ fournissent par interpolation la table pour $\lambda = 0,45$

	t_3 =	0	1	2	3	4 et plus
Loi de Poisson	$\lambda = 0,5$	0,60653	0,30327	0,7582	0,01264	0,00174
	$\lambda = 0,4$	0,67032	0,26813	0,05362	0,00715	0,00078
	$\lambda \# 0,45$	0,638	0,286	0,065	0,010	0,001
Loi de t_3		0,60662	0,33860	0,05301	0,00177	0

Cette fois, la ressemblance apparaît, entre la distribution de t_3 et celle de Poisson avec $\lambda = 0,45$.

2.2 - Recherches ultérieures

Nous empruntons au mémoire de M. Fannechère [3] les résultats de calculs plus poussés. Ceux-ci ont utilisé certaines tabulations préliminaires établies à l'ordinateur, mais la préparation des tables de distribution de t_2 n'a été ni programmée ni passée sur machine. S'agissant de calculs manuels, on ne s'étonnera pas du caractère très fragmentaire des résultats :

	r = 4	r = 5
N = 11	$\theta = 0,183$; $Et_r = 0,085$	$\theta = 0,030$; $Et_r = 0,0125$
12	$\theta = 0,20$; $Et_r = 0,09$	$\theta = 0,035$; $Et_r = 0,015$
16	$\theta = 0,266$; $Et_r = 0,143$	$\theta = 0,045$; $Et_r = 0,0236$

La comparaison entre lois de t_r et lois de Poisson donne ce qui suit :

N = 11	P (t_4)	Poisson 0,085	Poisson 0,20		P (t_5)	Poisson 0,0125	Poisson 0,03
$t_4 = 0$	0,91625	0,918512	0,81873	$t_5 = 0$	0,98712	0,987580	0,970446
1	0,08243	0,078074	0,16375	1	0,01286	0,012338	0,029113
2	0,00132	0,003318	0,01638	2	0,00002	0,000080	0,000437
$\geqslant 3$	0	0,000096	0,00114	$\geqslant 3$	0	0,000002	0,000004
	Accord :	Bon	Mauvais		Accord :	Bon	Médiocre
N = 12	P (t_4)	Poisson 0,09	Poisson 0,20		P (t_5)	Poisson 0,015	Poisson 0,035
$t_4 = 0$	0,905521	0,91391	0,81873	$t_5 = 0$	0,985099	0,985112	0,965605
1	0,09248	0,082254	0,16375	= 1	0,014875	0,014776	0,033797
2	0,001999	0,003701	0,01638	2	0,000252	0,000111	0,000591
$\geqslant 3$	0	0,000114	0,00114	$\geqslant 3$	0	0,000001	0,000007
	Accord :	Bon	Mauvais		Accord :	Bon	Médiocre
N = 16	P (t_4)	Poisson 0,15	Poisson 0,20		P (t_5)	Poisson 0,0225	Poisson 0,045
$t_4 = 0$	0,863113	0,86178	0,81873	$t_5 = 0$	0,9771	0,977754	0,955997
1	0,130991	0,12712	0,16375	1	0,0228	0,021993	0,043020
2	0,005829	0,01044	0,01637	2	0,0001	0,000250	0,000968
3	0,000066	0,00062	0,00109	$\geqslant 3$	0,0	0,000003	0,000015
$\geqslant 4$	0	0,00004	0,00006				
	Accord :	Bon	Mauvais		Accord :	Bon	Médiocre

Le phénomène observé a reçu un commencement d'explication théorique.

2.3 - Explication théorique

Pour qu'une variable soit une variable de Poisson, il est nécessaire que sa variance et son espérance soient égales. Ce n'est évidemment pas suffisant ; mais il semble que ce soit déjà une condition très forte. M. Fannechère a remarqué (sur les expressions théoriques) que $E(t_r)$ et $V(t_r)$ étaient pratiquement confondues pour des valeurs de N encore faibles :

D'une part on peut récrire $E(t_r)$ comme suit :

$$Et_r = \frac{2N(r^2 + 3r + 1)}{(r + 3)!} - \frac{2(r^3 + 3r^2 - r - 4)}{(r + 3)!}$$

D'autre part, l'expression exacte de $V(t_r)$, donnée par Levene et Wolfowitz [4] est :

$$Vt_r = \frac{2N(r^2 + 3r + 1)}{(r + 3)!} - \frac{2(r^3 + 3r^2 - r - 4)}{(r + 3)!}$$

$$+4\,\frac{(3r^6+24r^5+69r^4+90r^3+67r^2+42r+10)-N(2r^5+15r^4+41r^3+55r^2+48r+26)}{(r+3)!\,(r+3)!}$$

$$+4\,\frac{N(2r^2+9r+12)-2(2r^3+11r^2+19r+9)}{(2r+3)\,(2r+5)\,(r+1)!\,(r+3)!}$$

$$+4\,\frac{(16r^4+80r^3+116r^2+32r-19)-2N(4r^3+18r^2+23r+7)}{(2r+5)!}$$

$$+4\,\frac{N-2r}{(2r+1)r!\,r!}$$

De la forme : $V(t_r) = E(t_r)\,[1+\varphi(N,r)]$, avec $\varphi = \varphi_1+\varphi_2+\varphi_3+\varphi_4$

Pour des raisons variées, chacun des φ_i apparaît comme très petit pour r = 3 et plus :

$(r+3)! = 6! = 720$, par exemple, figure au dénominateur de φ_1.

Nota : On peut retrouver la même observation dans Olmstead [5] ; mais celui-ci a traité un problème différent : dresser la table (approximative) des $P(t_r)$ en assimilant sa fonction de distribution (probabilités cumulées) à celle d'une loi exponentielle.

CONCLUSION

La distribution de t_r (si r! est de l'ordre de grandeur de N) est très proche de la loi de Poisson de même espérance mathématique. Lorsque N tend vers l'infini (ainsi que r), le paramètre de cette loi tend vers $2N/(r+1)!$, et la loi de Poisson de paramètre $E(t_r)$ tend vers la loi de Poisson de paramètre $2N/(r+1)!$

2ème PARTIE : REMARQUE A PROPOS DE L'EXEMPLE PRECEDENT

Le lecteur a déjà certainement une expérience des phénomènes désignés sous le nom de loi des grands nombres et aussi sans doute de ceux qualifiés de "Poissonniens" ou loi des petits nombres. En statistique appliquée on emploie beaucoup plus la distribution de Laplace-Gauss et la distribution de Poisson (le processus de Poisson inclus) que toutes les autres distributions, sans doute pour la commodité des calculs mais aussi en raison de leur caractère de lois limites.

Ce qu'on appelle ainsi lois des grands nombres et des petits nombres, c'est avant tout, un aspect (et le plus courant) de la convergence en loi.

1.1 - Convergence en loi

On dit que la suite $\mathcal{L}_1\,\mathcal{L}_2 \ldots \mathcal{L}_n$ de lois de probabilités (ou la suite $X_1X_2 \ldots X_n$ de variables aléatoires) converge vers la loi $\mathcal{L}$ (ou la variable X) si la suite des fonctions caractéristiques $\varphi_1(t)\ \varphi_2(t) \ldots \varphi_n(t)$ correspondantes converge uniformément (au sens usuel de l'analyse) vers la fonction caractéristique $\varphi(t)$ de $\mathcal{L}$.

Convergence gaussienne : En particulier, si $\mathcal{L}$ désigne la loi normée de Laplace-Gauss $\mathcal{N}(0\ ;\ 1)$ la convergence en question est généralement celle d'une suite de lois elles-mêmes normées :

Posons

$$\mathcal{E}X_n = \mu_n,\ \mathcal{V}X_n = \sigma_n^2,\ Y_n = (X_n - \mu_n)/\sigma_n$$

la suite des lois des variables $Y_1\ Y_2 \ldots .\ Y_n$ (normées) est supposée converger vers la loi d'une variable η de Laplace-Gauss normée, $\mathcal{N}(0,1)$.

1.2 - Convergence et approximation (gaussienne)

Si n est assez grand (expression volontairement vague), la variable Y_n devrait avoir une distribution différant assez peu de $\mathcal{N}(0\ ;\ 1)$; de sorte que souvent on se croit autorisé à substituer l'une à l'autre dans les applications. En pareil cas, on substitue volontiers (parce que c'est commode) la variable de Laplace-Gauss non normée

$$\xi_n = \mu_n + \eta\ \sigma_n, \qquad \text{soit } \mathcal{N}(\mu_n\ ;\ \sigma_n^2) = \mathcal{N}_n$$

à la variable initiale X_n. On définit donc la suite de lois de Laplace-Gauss

$$\mathcal{N}_1, \mathcal{N}_2, \ldots \mathcal{N}_n$$

où $\mathcal{N}_n$ prend la place de $\mathcal{L}_n$ dans les applications.

C'est par exemple le cas (au delà de n = 30) pour la suite des lois de χ^2 à n degrés de liberté. C'est encore le cas pour des statistiques diverses utilisées dans les tests non-paramétriques usuels : leurs distributions n'ont été tabulées que pour d'assez faibles valeurs de n. A partir de (disons) n = 20 ou n = 50, on substitue des distributions approchées, qui (le plus souvent) sont de Laplace-Gauss [6] ch. 6.

1.3 - Remarque essentielle

La convergence étant supposée établie, la vitesse de convergence n'en est pas pour autant connue. Il est exact que, toutes choses égales d'ailleurs, les choses ne vont pas à la vitesse de n mais plutôt de $\sqrt{n}$; disons que les écarts ont (grosso-modo) des variances de l'ordre de n^{-1}, ils sont donc eux-mêmes de l'ordre de $n^{-1/2}$. Mais pour certaines convergences, n est assez "grand" s'il est égal à 10 ; pour d'autres, il faudrait qu'il soit de l'ordre de 10^2, ou de 10^3, ou de 10^6. Ainsi avec le "théorème-central-limite", la vitesse de convergence de la distribution d'une moyenne échantillon $\bar{x} = \Sigma x_i/n$ vers une distribution de Laplace-Gauss dépend de la valeur des moments (d'ordre 2, 3 et 4) de la distribution de chaque x_i. Il est bien connu enfin que cette vitesse de convergence peut être accélérée par certaines circonstances favorables :

- existence d'une symétrie : p = q = 1/2 dans le cas de la distribution de Bernoulli

- existence d'une distribution unique des x_i (théorème-central-limite)

- loi particulièrement favorable de distribution des x_i (le cas extrême étant celui d'une distribution de x_i de Laplace-Gauss, auquel cas $\bar{x}$ est d'emblée distribué suivant sa loi-limite, quelque soit n).

On n'est guère renseigné sur les divers cas où il y a convergence sans que le théorème-central-limite s'applique (notamment pour la moyenne de n valeurs corrélées d'une variable aléatoire) ; pour les convergences relatives aux tests non-paramétriques (cf. [6], Ch. 6) on ne semble guère s'être préoccupé des vitesses de convergence, qui semblent plus rapides qu'ailleurs.

2. CONVERGENCE POISSONNIENNE

Avec une loi de Poisson, les choses sont à la fois plus simples (puisqu'il s'agit d'une distribution à un seul paramètre) et plus compliquées, car il n'est plus question de normer la variable :

La loi de Poisson est (on le sait) une distribution sur l'ensemble {N} des valeurs entières (0, 1, 2 ...) de la variable. Nous ne nous intéressons donc ici qu'à des variables X définies aussi sur {N} ; une translation d'un nombre entier d'unités est la seule transformation autorisée.

Dans ces conditions, il est souvent commode d'utiliser, non la fonction caractéristique $E(e^{itx})$, mais la fonction génératrice $E(t^x) = g(t)$. La convergence en loi signifie que la suite $g_n(t)$ tend vers $g(t)$, $n \longrightarrow \infty$, $\forall t$.

2.1 - Un cas extrême est fourni par la convergence (bien connue) de la loi binomiale $B(p, n)$, dépendant des paramètres p et n, vers la loi de Poisson de paramètre λ, si l'on s'astreint à avoir constamment

$$pn = \lambda$$

Alors la fonction caractéristique :

$$\varphi_n(t) = (q + e^{it}p)^n = \left[1 + \frac{\lambda(e^{it} - 1)}{n}\right]^n, \text{ de } B\left(\frac{\lambda}{n}, n\right)$$

constitue une suite qui converge vers :

$$\varphi(t) = \exp\ \lambda(e^{it} - 1)$$

puisque la limite de $(1 + a/n)^n$ est $(\exp.\ a)$.

De même : $g_n(t) = (q + pt)^n = [1 + (t - 1)p]$, converge vers

$$g(t) = \exp\ [(t - 1)\lambda]$$

Alors, quel que soit n, l'espérance mathématique μ_n de la loi $\mathcal{L}_n$ est $np = \lambda$, c'est-à-dire coïncide avec celle de la loi limite $\mathcal{L}$ ou $\mathcal{P}(\lambda)$.

2.2 - La convergence vers la loi de Poisson $\mathcal{P}(\lambda)$ subsiste si $EX_n = \mu_n$ tend vers λ sans lui être constamment égale ;

par exemple si $\mu_n = pn = \lambda + \frac{1}{\sqrt{n}}$ à partir de $n \geqslant n_o$.

On peut définir une suite de lois de Poisson de paramètre $\lambda_n = \mu_n$,

soit : $\mathcal{P}_1 \mathcal{P}_2 \ldots . \mathcal{P}_n$ avec $\mathcal{P}_n = \mathcal{P}(\lambda_n)$;

et cette suite $\mathcal{P}_n$ converge vers $\mathcal{P} = \mathcal{P}(\lambda)$, puisque exp $\lambda_n(e^{it} - 1)$ converge vers $\varphi(t)$. Dans quelle mesure est-il préférable de substituer, dans les applications, la loi de Poisson $\mathcal{P}_n$ et non la loi de Poisson $\mathcal{P}$ à la loi binomiale B(p,n), c'est-à-dire :

$$B\left(p = \frac{\lambda}{n} + \frac{1}{n\sqrt{n}},\ n\right) ?$$

La théorie de la convergence en loi ne répond pas à pareille question.

2.3 - Plus généralement, si une suite de lois (quelconques)

$$\mathcal{L}_1 \mathcal{L}_2 \ldots \mathcal{L}_n \text{ converge vers } \mathcal{L} = \mathcal{P}(\lambda = \mu)$$

posons $EX_n = \mu_n$; considérons la suite de lois de Poisson

$$\mathcal{P}_1 \mathcal{P}_2 \ldots \mathcal{P}_n, \text{ avec } \mathcal{P}_n = \mathcal{P}(\lambda = \mu_n)$$

qui converge aussi vers $\mathcal{P}$, avec $\mu_n = \mu + \varepsilon(n)$

On rencontre couramment des cas où $\mathcal{P}_n$ est pour $\mathcal{L}_n$ une meilleure approximation que $\mathcal{L}$. On notera d'ailleurs que, si l'on dispose des Tables des Lois de Poisson pour des valeurs de λ assez rapprochées (avec les possibilités d'interpolation que cela implique), il n'est pas plus difficile d'en extraire la Table de $\mathcal{P}_n$ que celle de $\mathcal{P}$.

2.4 - Discusion

On peut se demander s'il est bien exact que EX_n converge vers EX : la convergence en loi implique-t-elle la convergence des moments ?

Prenons quelques précautions : Supposons pour $\mathcal{L}_n$ et $\mathcal{L}$ l'existence des moments (aussi loin qu'utile). Les fonctions $\varphi_n(t)$ et $\varphi(t)$ sont développables (en série entière ou en développement limité) et les coefficients de $(it)^j/j!$ sont respectivement les moments $(M_j)_n$ et M_j. On ne voit guère comment $\varphi_n(t)$ pourrait converger, $\forall t$, vers $\varphi(t)$ si leurs deux développements ne se composaient pas de termes convergeant l'un vers l'autre (1).

Les mêmes considérations s'appliquent aux secondes fonctions caractéristiques

$$\psi_n(t) = \text{Log } \varphi_n(t)\ ,\ \psi(t) = \text{Log } \varphi(t)$$

dont les développements (en série etc...) sont analogues, sauf que les cumulants $(K_j)_n$ et K_j prennent la place des moments dans les coefficients des puissances de t. On sait que : $K_1 = M_1$, $K_2 = M_2 - M_1 = \sigma^2$; etc...

Enfin il en sera de même pour les fonctions génératrices $g_n(t)$ et g(t), dont les développements suivant les puissances de t font intervenir les moments factoriels : $M_1, M_2 - M_1, M_3 - 3M_2 + 2M_1$, etc...

En résumé il paraît légitime de supposer au moins l'existence de $E(X_n)$, EX, $V(X_n)$, VX ainsi que la convergence de $E(X_n)$ vers EX, et de $V(X_n)$

---

(1) Si les moments de tous les ordres existent, il y a à ce sujet un théorème de Fréchet et Shobat [6], page 208 dont l'énoncé correct demande diverses précautions (comme tout ce qui touche au problème des moments).

vers VX, c'est-à-dire $(M_1)_n \longrightarrow M_1$; $(K_2)_n \longrightarrow K_2$. Bien entendu si $\mathcal{L}_n$ converge vers $\mathcal{L} = \mathcal{R}$ (Loi de Poisson) on a VX = EX (c'est-à-dire $K_2 = M_1$ puisque K_2 = VX et M_1 = EX).

2.5 - Cas favorable

Il arrive que $V(X_n)$ soit encore éloignée de VX (pour n petit), alors que déjà $E(X_n)$ et $V(X_n)$ sont très proches, c'est-à-dire que $(K_1)_n$ et $(K_2)_n$ sont très voisins. [En toute rigueur, on devrait dire : tous les $(K_h)_n$, $\forall n$; exiger que tous les cumulants soient voisins les uns des autres]. Dans ce cas $\mathcal{L}_n$ et $\mathcal{R}_n$ sont pratiquement confondus, alors que la loi-limite $\mathcal{R}$ est bien loin d'être atteinte. Tel était le cas dans l'exemple de la 1ère partie.

2.6 - Approximation par une distribution binomiale (asympt. de Poisson)

Le cas se présente, où $\mathcal{L}_n$ diffère beaucoup de $\mathcal{R}_n$, ceci se traduisant par une valeur de $V(X_n)$ nettement différente de $E(X_n)$. On en trouve un bon exemple dans Katz [7]

Exemple de Katz : Distribution du nombre d'isolés dans un groupe social

Une certaine variable aléatoire X (appelée nombre d'isolés), dépendant de 2 paramètres entiers n et d, $d < n$, admet une distribution très compliquée ; Feller a établi que, d étant fixe et n tendant vers l'infini, une telle distribution admet une loi limite de Poisson.

Dans cet exemple, il se trouve que λ est infini, de sorte qu'il n'y a plus de loi $\mathcal{R}$. La loi $\mathcal{R}_n$ a pour paramètre :

$$\lambda_n = n\left(1 - \frac{d}{n-1}\right)^{n-1}$$

d'où

$$\text{Log } \lambda_n \# \text{Log } n - d \ , \ \lambda_n \sim n e^{-d} \ ;$$

on vérifie bien que λ_n tend avec n vers l'infini. Bien entendu λ_n est l'expression de $E(X_n)$. La variance de $\mathcal{L}_n$ est (d'après Katz) :

$$V(X_n) = \lambda_n \left[1 - (d+1)\left(1 - \frac{d}{n-2}\right)^{n-2} + \varepsilon \right]$$

où ε désigne le terme correctif

$$n\left[\left(1 - \frac{d}{n-2}\right)^{n-2} - \left(1 - \frac{d}{n-1}\right)^{n-1} \right]$$

qui modifie peu les résultats.

Valeurs réalistes des paramètres pour les applications

Le paramètre d peut être de l'ordre de 3, alors que n serait de l'ordre de 25.

Alors $V(X_n)/E(X_n)$ est systématiquement inférieur à 1 ; pour n = 26 ; d = 3, ce rapport est 0,8283 = 1 - 0,1717, beaucoup trop éloigné de 1.

Dans ces conditions l'approximation Poissonnienne est forcément mauvaise ; c'est ce qu'on vérifie ;

Distribution	i = 0	1	2	3	4	etc
($\mathcal{L}_{26}$) exacte	p_i = 0,3098	0,4026	0,2143	0,0615	0,0106	.....
($\mathcal{P}$) approchée	0,3450	0,3675	0,1954	0,0693	0,0184	.....

Katz a donné un procédé très satisfaisant dans des cas analogues ; il assimile la distribution $\mathcal{L}_{26}$ a une distribution binomiale de paramètres (π, ν).

Ces paramètres sont estimés par la méthode des moments. Ainsi ν n'est pas entier

$$B(\pi, \nu) \quad \begin{cases} \text{moyenne} : \nu\pi \# E(X_n) = M, \\ \text{variance} : \nu\pi(1-\pi) \# V(X_n) = K_2 \end{cases}$$

On peut : soit retenir $[\nu]$, entier le plus voisin, soit généraliser la distribution binomiale avec ν non entier. C'est une difficulté purement technique, donc mineure. Nous accepterons ν non entier.

Ce choix est finalement acceptable parce que $V(X_n)$ est inférieure à $E(X_n)$. Soit

$$V(X_n) = (1-\alpha)\,E(X_n) \text{ , c'est-à-dire } K_2 = (1-\alpha)M_1 ,$$

on choisit :

$$\pi = \alpha, \quad \nu = M_1/\alpha$$

Remarque : Si l'on envisageait ensuite des valeurs de n plus grandes, d tendrait vers 0 et ν vers l'infini ; la loi "binomiale" $B(\pi, \nu)$ ajustée formerait la suite $B_\nu, B_{\nu+1}, \ldots$ convergeant aussi vers $\mathcal{P}_n$

$$\text{avec } \lambda_n \sim ne^{-d} \longrightarrow \infty$$

Détails de calcul : $E(X_n) = \lambda_n = 26\ (1 - 3/25)^{25}$

(calcul logarithmique) $= 1,064 = \mu_{[1]}$

$V(X_n) = \mu_{[2]} - \mu^2_{[1]} + \mu_{[1]}$ (il s'agit des moments factoriels)

En se servant d'un résultat général de M. Fréchet sur les événements compatibles et dépendants, on trouve :

$$V(X_n) = \mu_{[1]}\ [1 - (d+1)\ (1 - d/(n-2))^{n-2} + \varepsilon\ ]$$

avec un terme correctif ε souvent négligeable, mais ce n'est pas le cas pour n = 26, nombre encore petit. L'expression exacte de ε est :

$$\varepsilon = n\left[\left(1 - \frac{d}{N-2}\right)^{n-2} - \left(1 - \frac{d}{n-1}\right)^{n-1}\right]$$

Avec n = 26 et d = 3, la table de logarithmes nous conduit à :

$$\varepsilon = 0,0092$$

$$V(X_n) = \mu_{[I]} (1 - 0,1623 - 0,0092) = \mu_{[I]} (1 - 0,1715)$$

Ajustement de la loi "binomiale" :

On adoptera donc $\pi = 0,1715$, $\nu\pi = \mu_{[I]} = 1,064$, d'où $\nu = 6,2$. Katz, qui utilise une table de log. à 7 décimales, trouve $\pi = 0,1717247$ et $\nu = 6,197378$;

ce qui donne finalement :

i =	0	1	2	3	4	
p'_i =	0,3111	0,3997	0,2153	0,0624	0,0104	....
pour	0,3098	0,4026	0,2143	0,0615	0,0106	....

l'accord étant bien meilleur ainsi qu'avec la loi de Poisson. Il s'agit d'un développement en série convergente, car on n'a pas arrondi ν à l'entier le plus voisin 6. La vraie variable X ne peut dépasser la valeur n-d-1 = 22, mais n'a qu'une probabilité infime de dépasser 7 ; et il en est de même de la variable "binomiale".

2.7 - Autre exemple :

Dans l'exemple qui suit, ν est entier et on peut ajuster une vraie loi binomiale (à un nombre fini de termes).

Barton a étudié, après de Montmort (1708) et bien d'autres, le problème des Rencontres, [8] (1958) et montré la convergence vers la loi de Poisson des distributions des nombres de "rencontres" dans des problèmes de Rencontres généralisés.

Il est également question dans cet article d'autres ajustements de ces distributions. Un tableau II (p. 80) permet la comparaison de 5 méthodes différentes d'ajustement (dont une binomiale). Nous nous en tiendrons au cas beaucoup plus simple du Tableau I (p. 76) où le seul ajustement fait est celui de Poisson.

On compare deux jeux de cartes, comprenant chacun 20 cartes, à savoir : 2 as, 2 deux, 2 trois,... 2 dix. L'un des jeux est rangé n'importe comment, l'autre est soigneusement battu. Le nombre de rencontres (dans la comparaison des deux jeux) peut être 20, 18, 17 ... 2, 1, 0. Soit r ce nombre.

C'est une variable aléatoire ; on démontre que $E(r) = 2$, $V(r) = 36/19$. (plus généralement, si $N = 2k$ est le nombre de cartes d'un jeu de ce type, $E(r) = 2$, $V(r) = (2N - 4)/(N - 1)$).

Ajustons une loi binomiale : $\nu\pi = 2$, $\nu\pi(1 - \pi) = 36/19$;

d'où :

$$\pi = 1/19, \quad \pi/(1 - \pi) = 1/18 \ ; \quad \nu = 38.$$

Calculons (par logarithmes) les premiers termes de la distribution et comparons (tableau ci-dessous) les 3 distributions :

1/ vraie (d'après Barton)

2/ de Poisson, $\theta = 2$;

3/ Binomiale $\pi = 1/19$, $\nu = 38$

On constate que cette dernière est excellente.

i =	0	1	2	3	4	5	6	7	8
1/	0,128	0,270	0,278	0,186	0,090	0,034	0,010	0,003	0,001
2/	0,135	0,271	0,271	0,180	0,090	0,036	0,012	0,003	0,001
3/	0,12818	0,27056	0,27808	0,18539	0,09012	0,034045	0,01040	0,00264	0,00057

Remarque :

1/ Ici encore, la variance était inférieure à la moyenne. Sans quoi, l'ajustement d'une loi binomiale serait impossible.

2/ On trouve dans David et Barton [9] p. 252, quatre modèles d'occupation, où la variance est inférieure à la moyenne ; il en est de même pour les runs de boules de même couleur (Ch. VI) ; etc... La méthode s'appliquerait en pareils cas.

2.8 - Cas d'une distribution dont la variance est plus grande que la moyenne :

On rencontre aussi les distributions pour lesquelles la variance dépasse la moyenne. On sait que le "chi-carré" χ^2 à N degrés de liberté a pour moyenne N et pour variance 2N ; bien entendu, il n'est pas question ici de telles variables, puisque :

X_n est supposée discrète, définie sur $\{N\}$.

X_n a une loi qui converge vers la loi de Poisson.

Nous supposerons donc que la variance $V(X_n)$ est équivalente à $E(X_n)$ pour n grand, mais en dépassant toujours $E(X_n)$.

Alors, une distribution binomiale négative, asymptotique à une loi de Poisson, serait d'un emploi aussi commode que la distribution binomiale de Katz, comme approximation de la loi de X_n lorsque n n'est pas très grand.

Rappels sur la distribution binomiale négative :

Soit $g_n(t) = q^n(1 - pt)^{-n} = E(X^j)$ la fonction génératrice ; celle de Poisson est exp $\lambda(t - 1) = g(t)$

On a

$$g_n'(1) = np/q = m_1 \; ; \; g_n''(1) = n(n + 1)p^2/q^2 = m_2 - m_1$$

d'où la Variance

$$\sigma^2 = m_2 - m_1^2 = m_1 + m_1^2/n = m_1(1 + m_1/n)$$

Lorsque n tend vers l'infini, σ^2 tend vers m_1. Si $m_1 = np/q = \lambda$ reste constant, d'où

$$\frac{n}{q} = \frac{\lambda}{P} = \frac{n + \lambda}{1} ,$$

on a :

$$\text{Log } g_n(t) = n \text{ Log } [(1 - p)\ (1 - pt)^{-1}] = - n \text{ Log} \left[1 + \frac{\lambda}{n}(1 - t) \right] \longrightarrow - \lambda(1 - t)$$

donc :

$$\text{Log } g_n(t) \longrightarrow \text{Log } g(t) \ ; \ \text{donc } g_n(t) \longrightarrow g(t)$$

C'est la transposition exacte du cas classique de la loi $\mathcal{L}_n$ binomiale avec $\lambda = np$.

Application éventuelle : Soit X_n une variable aléatoire, de loi plus ou moins compliquée, dont la convergence vers la loi de Poisson aurait été établie déjà. Si n n'est pas très grand, on constaterait

$$VX_n = EX_n(1 + \alpha_n) \qquad \text{avec } \alpha_n > 0.$$

α converge vers 0 mais en est encore loin pour les valeurs de n qui nous intéressent. Dans ces conditions on représenterait $\mathcal{L}_n$ par une distribution binomiale négative (π, ν) ajustée par les deux premiers moments (ou cumulants) :

$$\alpha_n = \frac{m_1}{\nu} = \frac{\pi}{1 - \pi} \Longrightarrow \pi = \alpha_n/(\alpha_n + 1) \ ; \ \nu = m_1/\alpha_n$$

(même difficulté avec ν non entier).

On aurait aimé donné ici une application numérique réelle : on suppose qu'il doit s'en trouver à propos des phénomènes d'attente, par exemple

Nous n'avons rien trouvé dans Feller, où l'on donne seulement des distributions limite (de Poisson, binomiale négative, de Kolmogoroff, etc...) ; mais nous ne disposions pas d'une documentation spécialisée, ou l'on ait des chances de trouver les vraies distributions correspondantes.

3 - CONCLUSION

Lorsque la loi limite $\mathcal{R}$ est encore éloignée de la loi $\mathcal{L}$ il existe quelques moyens d'approcher $\mathcal{L}$ par une loi $\mathcal{L}_n^*$; de Poisson, Binomiale, Bin. nég. suivant les cas, convergeant aussi vers $\mathcal{R}$ et ayant mêmes moments d'ordres 1 et 2 que $\mathcal{L}_n$. Ceci n'exclut pas bien entendu les autres méthodes.

Signalons qu'on trouve dans Katz [7] référence à des travaux déjà anciens de Guldberg (1931) et Ragnar Frish (1932) sur certains critères numériques permettant d'apprécier les écarts entre distributions.

Tout ceci n'est pas (on en a conscience) entièrement satisfaisant ; on espère cependant rendre service dans les applications et aussi suggérer un axe de recherches. Mais terminons en ruinant une illusion commode.

4 - CONVERGENCE EN LOI OU CONVERGENCE STOCHASTIQUE

On aurait pu se demander si le fait pour une variable X_n d'être plus ou moins rapidement assimilable à une variable X de Poisson (ou autre) ne mettait pas en jeu les concepts de convergence stochastique

plus ou moins forte. Il semble bien pourtant qu'il s'agisse purement de convergence en loi, c'est-à-dire entre distributions. On veut seulement savoir si la fonction de répartition $F_n(x)$ de X_n est de plus en plus voisine de celle, $F(x)$, d'une loi limite. Ceci est assuré s'il y a convergence de $F_n(x)$ vers $F(x)$, sauf aux points de discontinuité (définition de la convergence en loi).

Lorsqu'on définit les concepts de convergence en probabilité, en moyenne quadratique, presque sure, etc... on a affaire à une suite de variables aléatoires $X_1 X_2 \ldots$ qui (au moins à partir d'un certain rang) sont corrélées entre elles et avec la variable limite X (voir [10] p. 376). On devrait tirer au sort par exemple un groupe social dont l'effectif ν irait grandissant, et obtenir chaque fois le nombre d'isolés du groupe en refaisant la même épreuve, avec le même paramètre d). Cette suite (du nombre d'isolés : $X_{\nu-1} X_\nu X_{\nu+1} \ldots$) ne paraît pas utile dans le type de problèmes envisagé ici, même si (comme c'est souvent le cas) la distribution de X_n est (en fait) obtenue par récurrence, par exemple par un processus en chaîne (une suite markovienne). C'est en somme une complication que d'envisager une distribution à ν dimensions ($\nu \longrightarrow \infty$) alors qu'on n'a besoin que de sa projection sur un seul axe de coordonnées. On imagine que $X(n = \nu)$ fait partie d'une suite de variables X_ν, mais on ne s'intéresse finalement qu'à la distribution de cette seule variable X_n ; on ne définit pas (par rapport à X_n) les autres variables X_ν ni même la variable limite X.

Certains professeurs pensent d'ailleurs que la convergence en loi n'est pas une "vraie convergence" et qu'on devrait la désigner par un autre vocable. L'ennui est que des théorèmes prouvant l'existence d'une certaine loi-limite peuvent trouver leur origine dans des considérations de convergence stochastique ; c'est au moins un procédé de démonstration et on ne peut risquer de s'en priver par purisme.

REFERENCES BIBLIOGRAPHIQUES

[1] THIONET P. - Sur certains tests non paramétriques bien connus - Revue de l'Institut International de Statistique - 34 - 1 - 1966 - p. 13-26.

[2] WOLFOWITZ J. - Asymptotic distribution of runs up and down - Annals of Mathematical Statistics - 15 - 1944 - p. 163-72.

[3] FANNECHERE G. - Contribution à l'étude des suites des signes des différences premières d'une série chronologique (Mémoire pour le D. E. S.) - Université de Poitiers 1967.

[4] LEVENE H. et WOLFOWITZ J. - The covariance matrix of runs up and down - Annals of Mathematical Statistics 15 - 1944 - p. 58-69.

[5] OLMSTEAD P. S. - Distribution of sample arrangements for runs up and down-Annals of Mathematical Statistics - 17 - 1946 - p. 24-33.

[6] FRASER D. A. S. - Non-parametric methods in statistics - John Wiley - 1957.

[7] KATZ L. - The distribution of the number of isolates in a social group. Annals of Mathematical Statistics - 23 - 1952 - p. 271-276.

[8] BARTON (D.E.) - The matching distribution : Poisson limiting forms and derived methods of approximation, Jour. Royal Statist. Society B 20 (1958) p. 73-92.

[9] DAVID F.N. & BARTON D.E. - Combinatorial Chance, Ch. Griffin (Londres) 1962.

[10] FORTET R. - Eléments de la théorie des probabilités - I - C.N.R.S. 1965.