Tutorials in Quantitative Methods for Psychology
2012, Vol. 8(3), p. 151-162.
DOI: 10.20982/tqmp.08.3.p151

L'interprétation stochastique du sociogramme et le problème des choix exceptionnels
A statistical interpretation of the sociogram and the determination of exceptional choice

Louis Laurencelle
Université du Québec à Trois-Rivières

The mutual designation of a preferred member among a group, be it for presidency, a special mandate or even for psychosocial purposes, proceeds frequently by Moreno's sociogram technique: in its simpler form, each member votes for some other member, and the one who receives a majority of votes is elected. But, is this elected member really preferred, or may his election be ascribed to some random, haphazard voting? This article explores the statistical aspects of this electoral process and incidentally grazes the arcane concept of the partitions of an integer.

Article first published in Lettres Statistiques, *1993, vol. 9, p. 115-133.*

Supposons un groupe de personnes, une équipe, un collège électoral, et que ce groupe ait à désigner un porte-parole, un chef, un président. Dans un scrutin libre où chacun est candidat, chacun votera pour un autre, et l'on compilera les voix reçues par chaque membre. Celui ayant reçu le plus grand nombre de voix pourra être nommé porte-parole, chef, président.

Ce petit problème de choix électoral a, semble-t-il, échappé aux soins des mathématiciens. Il correspond à la technique du *sociogramme* de Moreno, et il renferme des développements virtuels d'une grande richesse. Dans un premier temps, nous définirons la classe de problèmes que nous admettons dans cette situation de collège électoral. Nous présenterons ensuite les mathématiques du choix d'une *personne désignée*, puis nous aborderons le problème beaucoup plus difficile du choix d'une *personne exceptionnelle*, ou du choix exceptionnel. Ensuite, la question connexe de la répartition des choix, ou des partitions d'un nombre, sera traitée. Nous terminerons avec un exemple pratique, qui d'ailleurs servit d'amorce de la présente étude.

La situation de collège électoral et les sociogrammes

Qu'ils aient des rapports habituels ou qu'ils se rencontrent pour l'occasion spécifique d'une élection, les membres d'un collège électoral sont admissibles au poste proposé et ont droit d'exprimer un vote, désignant l'un d'entre eux. Le *sociogramme* proposé par Moreno (1943) consiste lui aussi à exprimer les choix, positifs ou négatifs, des membres d'un groupe les uns envers les autres: l'analyse des relations de choix complexes entre les membres constitue à la fois un objet d'étude et un instrument d'intervention pour le psychologue. Réduit à sa plus simple expression, le sociogramme pourra représenter le choix exprimé par chaque membre pour le membre le plus « intéressant » du groupe, à l'exclusion de soi-même bien entendu. Or, au retour d'une incursion dans la littérature du domaine, il semble qu'on n'ait pas exploré sérieusement cette situation d'un point de vue de modèle stochastique; par exemple, rien ne nous permet de savoir si, dans un groupe de 8 membres, le choix de l'un d'eux par 4 membres représente ou non un événement remarquable ou si une telle situation est simplement imputable au hasard.

Tableau 1. Probabilités de choix du membre « 1 » dans un groupe de n = 8 membres où tous votent une fois selon le calcul binomial (B), l'approximation Gram-Charlier (B_1) et l'approximation Poisson (B_2)

c_1	0	1	2	3	4	5	6	7	8
$B(c_1)$	0,344	0,393	0,196	0,056	0,010	0,001	0,000+	0,000+	0,000+
$B_1(c_1)$	0,329	0,408	0,185	0,065	0,012	0,001	0,000+	0,000+	0,000+
$B_2(c_2)$	0,368	0,368	0,184	0,061	0,015	0,003	0,001	0,000+	0,000+

Conventions et notations

Soit un groupe de n membres, à qui l'on demande de se choisir mutuellement. Nous ne considérons pas ici les situations où quelqu'un se choisirait lui-même [ces situations sont admissibles en général dans les collèges électoraux, où le scrutin secret empêcherait de toutes façons le contrôle]. Le cas simple, ou sociogramme simple, consiste à ce que chacun choisisse un autre membre. La classe de cas à considérer inclut les groupes de n candidats dans lesquels n' votants expriment chacun r_i votes, selon $n' \leq n$ et $r_i \geq 1$. Nous étudierons surtout le cas simple, dans lequel $n' = n$ et $r_i = 1$, ainsi que quelques-uns des autres cas possibles.

Disons que l'élection va avoir lieu. Il y aura n' votants, exprimant chacun r_i voix. Parmi les n candidats, le candidat « 1 » recevra c_1 votes, le candidat « 2 » en recevra c_2, ainsi de suite jusqu'au candidat « n » qui en aura c_n. On a $c_i \geq 0$, $3c_i = 3\, r_j$ en général, et $3c_i = n'$ si chacun n'exprime qu'une voix. Le choix le plus nombreux est $c_{max} = \max[c_1, c_2,⌡, c_n]$ et il peut échoir à un candidat, ou à plus d'un si $c_{max} \leq ½n'$. Par *personne désignée*, on entend un membre désigné d'avance, par exemple le membre « 1 », et l'on s'intéressera au nombre de votes (c_1) qu'il a reçus. Le candidat exceptionnel, quant à lui, dépend de la variable c_{max}; a priori, il peut s'agir de n'importe lequel des n candidats en lice.

Le choix d'une personne désignée

Soit la personne « 1 », l'un des n membres du groupe. Si l'on suppose que le vote se fait au hasard et que chacun a une chance égale d'être choisi, alors les votes enregistrés par le membre « 1 », ou c_1, suivent une distribution binomiale. Cette question est la seule traitée dans Moreno (1943; voir aussi Bronfenbrenner 1945 et Bastin 1970).

Considérons les situations où chacun n'exprime qu'un vote. Il faut néanmoins distinguer deux cas. Dans le premier cas, le membre « 1 » fait partie des votants, que tous votent ou non: alors, comme pour chacun il y a n–1 candidats (c.-à-d. tous sauf soi-même) et que n'–1 sont capables d'élire le membre « 1 », la loi binomiale a pour paramètres $\pi = 1/(n-1)$ et $N = n'-1$ et pour statistiques l'espérance $\mu = (n'-1)/(n-1)$ et la variance $\sigma^2 = (n'-1)(n-2)/(n-1)^2$. Dans le second cas, le membre « 1 » ne fait pas partie des votants: n' personnes peuvent donc élire le membre « 1 », et les paramètres du binôme sont $\pi = 1/(n-1)$ et $N = n'$, avec $\mu = n'/(n-1)$ et $\sigma^2 = n'(n-2)(n-1)^2$.

En général, si chaque votant exprime r votes, avec $r \geq 1$ constant, seul le paramètre π est affecté et il devient $r/(n-1)$. Si chaque votant dépose un nombre variable de votes r_i, alors la quantité c_1 peut être envisagée comme le résultat d'une mixture de n' (ou n'–1) binomiales à paramètre π_i variable; nous n'entrerons pas dans ce sujet (voir cependant Johnson et Kotz 1969).

Les auteurs (Moreno 1943) abordent aussi la question des choix mutuels, etc. Ils remarquent que, dans un groupe réel, le nombre d'isolés (membres recevant c_i = 0) et la popularité des vedettes (valeur de c_{max}) sont plus élevés que selon les calculs binomiaux.

Pour des valeurs modestes de n', disons n' < 50, le calcul exhaustif des probabilités binomiales reste abordable. On peut aussi exploiter une loi normale pour approcher une binomiale fortement asymétrique: il s'agit de l'expansion Gram-Charlier de type A (Johnson et Kotz 1969). Soit la variable binomiale c de paramètres π et N, l'intégrale binomiale $B(c)$, où $B(c) = \text{Prob}(x \leq c)$, l'intégrale normale $\Phi(z)$ et sa densité $\phi(z)$. Il s'agit d'obtenir:

$$z = (c + ½ - N\pi)/\sigma, \quad \text{où } \sigma^2 = N\pi\cdot(1-\pi)\,, \qquad (1)$$

puis l'intégrale estimée est donnée par l'expansion incorporant l'indice d'asymétrie $\gamma_1 = (1-2\pi)/\sigma$, comme suit:

$$B_1(c) \approx \Phi(z) - (1 - 2\pi)(z^2 - 1)\phi\,(z)/(6\sigma)\,. \qquad (2)$$

Enfin, pour n élevé, le choix au hasard du membre désigné est un « événement rare », et l'on peut croire que la répartition du nombre de choix approche une loi de Poisson de paramètre $\mu = (n'-1)/(n-1)$, d'où:

$$B_2(c) \approx e^{-\mu}\,\mu^c/c!\,. \qquad (3)$$

L'une et l'autre méthode d'estimation se comparent avantageusement, en précision et en méthode, à la « table de

Tableau 2. Nombre d'arrangements des choix de n personnes produisant c_{max} choix

c_{max} / n	2	3	4	5	6	7	8	9
1	1	2	9	44	265	1854	14833	133496
2		6	60	720	10020	161070	2955624	61145280
3			12	240	4710	99960	2326800	59402448
4				20	600	15750	422800	12050640
5					30	1260	42336	1382976
6						42	2352	98784
7							56	4032
8								72
Total	1	8	81	1024	15625	279936	5764801	134217728

Salvosa » suggérée dans la littérature*.

Supposons un groupe de n = 9 personnes, dans lequel chacune vote une fois. Les paramètres binomiaux sont ici π = 1/8 et N = 8. La probabilité que le membre « 1 » ne reçoive des autres aucun vote est $(1 - 1/8)^8$ = 0,34361, etc.; les différentes valeurs sont présentées au Tableau 1, de même que celles obtenues par l'expansion Gram-Charlier (ligne B_1) et par la distribution de Poisson (ligne B_2). Le lecteur notera que, dans le cas du sociogramme simple, μ = 1 et alors les valeurs approchées de Poisson apparaissant au Tableau 1 sont universelles.

Le choix maximum

La quantité c_1, le nombre de votes en faveur du membre « 1 », se distribue comme une binomiale de paramètres π et N. Il en est de même pour c_2, etc., mais l'on s'intéresse à présent au choix le plus nombreux, c_{max}. Peut-on considérer que c_{max}, qui est le maximum de l'ensemble $\{c_1, c_2, \ldots, c_n\}$, représente aussi le maximum de n binomiales indépendantes? Si tel était le cas, on pourrait en estimer aisément les probabilités.

Soit b(i) et B(i), $0 \le i \le N$, les probabilités individuelles et cumulatives d'un processus binomial à paramètres π et N quelconques, et supposons qu'on observe n processus binomiaux semblables et indépendants (représentant nos n candidats). Dénotant par p(j) la probabilité que le maximum de « succès » (de votes reçus) parmi les n processus (candidats) soit j, on a $p(0) = B^n(0)$, et $p(j) = B^n(j)-B^n(j-1)$, pour $j > 0$.

* Le lecteur intéressé aux approximations de l'intégrale binomiale consultera avec profit l'article de F. Gebhardt (1969), qui discute les mérites d'une dizaine de techniques.

Toutefois les candidats ne constituent pas des objets binomiaux indépendants, puisque les n' votes sont forcément partagés entre eux et que $\Sigma c_i = n'$ (lorsque chaque votant n'exprime qu'une voix). Le problème est décidément plus complexe, et nous ne l'avons pas entièrement résolu.

Nous aborderons le difficile problème du choix maximum en nous référant au sociogramme simple, soit un groupe de n membres votant chacun une fois, un membre ne pouvant pas voter pour soi-même.

Le nombre total d'arrangements et le domaine de c_{max}. Chaque membre choisit son candidat parmi n–1 membres, puisqu'il s'exclut lui-même. Comme il y a n votants, le nombre total d'arrangements possibles (T) est:

$$T = (n-1)^n \,. \qquad (4)$$

Quant à c_{max}, puisqu'il y a au moins 1 vote exprimé et n–1 personnes pouvant voter pour un candidat, les valeurs possibles vont de 1 à n–1. Le Tableau 2, obtenu d'emblée par énumération complète, présente les fréquences d'arrangements conduisant à chaque valeur de c_{max}, pour n allant de 2 à 9.

Le nombre d'arrangements « dominants ». Le vote le plus élevé que peut recevoir un candidat est $c_{max} = n-1$. S'il reçoit r voix, il reste $n-r$ voix à partager, voix qui pourraient échoir à un seul autre membre ou à plusieurs. Par contre, lorsque $r > n-r$, alors le candidat domine le vote, chacun des autres candidats ayant forcément reçu moins que lui.

Désignons par Y le candidat ayant reçu $r > n-r$ voix. Outre ce consensus de r membres, il reste à Y et $n-r-1$ membres à voter; Y peut voter parmi n–1 membres, et les $n-r-1$ autres doivent choisir chacun parmi n–2, soit en excluant soi-même et Y. Comme il y a ${}_nC_r$ combinaisons possibles du groupe consensuel et que le consensus peut désigner Y parmi les autres $n-r$ membres, le nombre d'arrangements « dominants » est:

$$a(r) = \binom{n}{r}(n-r)(n-1)(n-2)^{n-r-1} \qquad (5)$$

Ainsi, la popularité la plus grande, correspondant à $c_{max} = n-1$, a pour fréquence $a(n-1) = n(n-1)$ et pour probabilité $n/(n-1)^{n-1}$.

Le nombre d'arrangements « non dominants ». Si un candidat Y reçoit r votes en sa faveur et $r \leq n-r$, les voix qui restent à assigner peuvent échoir à un ou quelques adversaires principaux. La majorité de Y peut alors être doublée par un concurrent ayant même majorité, voire éclipsée par un autre obtenant une majorité plus forte. Or l'expression (5) compte le nombre d'arrangements à r voix qu'un candidat peut obtenir, quelle que soit la répartition des autres voix.

Il apparaît donc que, pour les cas où la majorité r n'est pas dominante, la formule proposée surestime le nombre d'arrangements. En effet, dans les cas d'un arrangement $\{r, r', v < r\}$, soit une double majorité accompagnée de quelques votes minoritaires (v), on obtient $c_{max} = r$ et l'arrangement ne doit compter qu'une fois pour r ou r', alors que la formule (5) recombinera $r'+v$ pour produire davantage d'arrangements à r-majorité. Dans les cas d'un arrangement $\{r, V, v < r\}$, on a une majorité $V > r$ apparaissant à côté d'une moindre majorité r et de quelques votes minoritaires (v) : dans ce cas, $c_{max} = V$ et l'arrangement ne doit pas être enregistré pour $c_{max} = r$, comme le fait la formule (5).

Les considérations développées plus haut nous conduisent à une formule générale du nombre d'arrangements correspondant à chaque valeur de c_{max}. Cette formule incorpore la formule (5) pour $c_{max} = r$, dont on retranche ensuite les arrangements produisant des majorités supérieures à r et les arrangements conduisant à un surcompte d'une même r- majorité qnand celle-ci est répétée:

$$f(r) = a(r) - \sum\nolimits_{k=2}^{n/2}(k-1)h(r^k, v<r) - \sum\nolimits_{V=r+1}^{n-r} k \cdot h(V, r^k, v<V) \; ; \qquad (6)$$

dans cette expression, $h(a,b,c,\text{etc.})$ désigne le nombre d'arrangements du sociogramme simple produisant la partition majoritaire indiquée. Cette expression (6) tient plus d'une voie de recherche que d'une solution proprement dite, puisqu'en effet les quantités $h(\ )$ sont difficiles à évaluer. Notons cependant qu'on peut y aller récursivement, à partir de $r = n-1$ en descendant, certaines quantités $h(V, r^k, v<V)$ ou $h(r^k, v<r)$ pouvant être obtenues par soustraction. Les cas particuliers suivants sont présentés à toutes fins utiles, soit pour une répartition en deux majorités ($\Delta = 1$ si les arguments sont différents, 0 sinon):

$$h(a,b) = \frac{(a+b)!}{(a-1)!(b-1)!2^{1-\Delta(a,b)}} \; , \qquad (7a)$$

soit en une majorité concomitante à deux choix uniques :

$$h(a, 1, 1) = a \cdot h(a, 2) \; . \qquad (7b)$$

Prenons l'exemple d'un sociogramme simple, avec $n = 6$. Les deux majorités dominantes sont fournies par la formule (5), soit $a(5) = 30$ et $a(4) = 600$. Pour $r = 3$, la majorité peut être doublée sous la forme $h(3,3)$. Ici la formule (7a) fournit $h(3,3) = 90$. Par la formule générale, $f(3) = a(3) - 1 \cdot h(3,3) = 4800 - 90 = 4710$. Pour $c_{max} = r = 2$, la formule (6) nous demande d'évaluer:

$$f(2) = a(2) - 1 \cdot h(2,2,1,1) - 2 \cdot h(2,2,2) - 1 \cdot h(3,2,1) - 1 \cdot h(4,2) \; .$$

D'abord, $a(2) = 19200$, d'après (5). La quantité $h(4,2)$, correspondant à (7a), égale 120. Nous avons dû procéder par dénombrement pour obtenir $h(2,2,1,1) = 5580$, $h(2,2,2) = 600$ et $h(3,2,1) = 2280$. Le calcul fournit $f(2) = 10020$. Pour évaluer f (1), c'est-à-dire la situation où aucune majorité n'apparaît et chaque vote est différent, on pourrait utiliser la même méthode fastidieuse, mais il existe pour ce cas une solution directe et remarquable.

Le nombre d'arrangements à votes tous différents. Pour obtenir $c_{max} = 1$ dans le sociogramme simple, il faut que chaque membre vote pour un candidat différent. On peut analyser cette situation en considérant: 1) que l'électeur 1(=A) choisit parmi (n–1A) personnes, engendrant $(n-1)!$ arrangements; 2) que l'électeur 2(=B) choisit parmi (n–1B) personnes; que, s'il était considéré indépendamment, il engendrerait aussi $(n-1)!$ arrangements; que toutefois, son choix est potentiellement réduit par le choix de l'électeur 1, sauf pour les $(n-2)!$ arrangements dans lesquels l'électeur 1 aurait restreint son choix aux (n–1A–1B) personnes; etc. D'où l'on voit que $(n-1)!$ }$\leq D(n) < n!$, où $D(n) = f(1)$ et dénote le nombre d'arrangements différents produisant c_{max}=1.

Par induction empirique (voir Tableau 2), on constate que les $D(n)$ ont une structure récursive, satisfaisant la relation:

$$D(n) = (n-1)[D(n-1)+D(n-2)] \; , \qquad (8)$$

en posant $D(1) = 0$ et $D(2) = 1$.

La même relation récursive se retrouve dans un problème classique de probabilités, le « problème des rencontres » (Feller 1950). Ce problème, solutionné par Montmort au XVIII[e] siècle, consiste à donner la probabilité qu'il y ait X rencontres (X = 0, 1, 2, etc.), c.-à-d. appariements, en comparant deux séries numérotées commensurables, par exemple deux paquets de cartes à jouer, un groupe de personnes et leurs chapeaux, etc., la rencontre étant l'égalité des numéros (ou l'*appartenance*, ou la *correspondance*). Ce problème est l'illustration par excellence de la méthode dite d'inclusion et d'exclusion, d'un intérêt si profond en analyse combinatoire (voir par exemple Brualdi 1977). La situation de $X = 0$ rencontre dans deux séries de n correspond ici à notre cas $c_{\max} = 1$. Le nombre d'arrangements (dénommés « dérangements » dans le cas présent), soit $n!\Pr\{X = 0\}$, est:

$$\mathrm{D}(n) = n!\sum_{i=0}^{n} \frac{(-1)^i}{i!}\,, \tag{9}$$

et, approximativement dès que $n \geq 4$:

$$\mathrm{D}(n) \approx n!e^{-1}\,. \tag{10}$$

Cas où il y a n′ < n votants. Chacun des n' membres votants choisit un candidat parmi les n membres du groupe à l'exclusion de soi-même, d'où $T = (n-1)^{n'}$ est le nombre d'arrangements pour ce groupe†. L'ensemble des n candidats est divisé en un sous-ensemble de n' votants et un de $n - n'$ non-votants; chacun des votants peut piger quiconque parmi les non-votants, mais son choix dans les votants est évidemment restreint à $n' - 1$ membres. Le nombre d'arrangements « dominants », c.-à-d. pour lesquels la majorité de r voix domine selon $r > n' - r$, est:

$$a(r|n') = \binom{n'}{r} \times \Big((n'-r)(n-1)(n-2)^{n'-r-1} + (n-n')(n-2)^{n'-r}\Big) \tag{11}$$

l'expression additionne deux parts, la première pour laquelle le candidat favorisé est votant, la seconde pour laquelle il fait partie des non-votants. Par exemple, si tous votent pour le même candidat, alors $a(r\,|\,n') = n - n'$.

Nous n'avons pas cherché d'expression analogue à (6) pour le cas général des majorités r non dominantes, soit $r \leq n'-r$. Cependant le cas particulier de $c_{\max} = 1$ admet encore la solution remarquable, basée sur la méthode d'inclusion et d'exclusion.

Soit deux séries commensurables de taille n ; cherchons la probabilité qu'il n'y ait aucune rencontre des n' premiers membres d'une série parmi les n membres de l'autre série, $P(X = 0)$ ou $P(0)$. Cette probabilité est le complément de la probabilité qu'il y ait une rencontre ou plus, soit pour chacun des n membres, $P(0) = 1 - P(\cup_{n'} \pi_i)$, où π_i est la probabilité qu'un membre de la première série en « rencontre » un de la seconde. Cette probabilité d'*union* est donnée par:

$$P(\cup_{n'} \pi_i) = \Sigma_1\, \pi_i - \Sigma_2\, \pi_{ij} + \Sigma_3\, \pi_{ijk} - \text{etc.}$$

Dans cette expression, désignons π_{ij} par $\pi_{(2)}$, π_{ijkl} par $\pi_{(4)}$, etc. La quantité $\pi_{(r)}$ est donc la probabilité qu'il y ait r rencontres, c.-à-d. que r membres de la première série rencontrent leurs pareils dans les n membres de la seconde. Pour 1 membre, π_i est manifestement $1/n$; pour 2 membres, elle est de $1/[n(n-1)] = 1/n_{(2)}$, et en général $\pi_{(r)} = 1/n_{(r)}$. Il faut ensuite sommer ces probabilités $\pi_{(r)}$ selon l'opérateur Σ_r. Pour π_i, soit $r = 1$, on doit sommer à travers les n' membres de la première série (il y a n' possibilités de rencontres simples), à travers $\mathrm{C}(n',2)$ pour les rencontres doubles à probabilité π_{ij}, et en général à travers $\mathrm{C}(n', r)$.

La probabilité qu'il n'y ait aucune rencontre dans le contexte cherché est donc:

$$P(0) = 1 - \frac{n'}{n} + \frac{n'(2)}{2!n_{(2)}} + \frac{n'(3)}{3!n_{(3)}} + \text{etc.}$$

Comme il y a seulement n' membres présents, le nombre d'arrangements effectifs correspond à $\mathrm{P}(n, n') = n!/(n-n')!$ permutations. La probabilité que dans chacune il y ait zéro rencontre étant $P(0)$ ci-dessus, nous obtenons par multiplication le nombre d'arrangements conduisant à $c_{\max} = 1$, soit:

$$\mathrm{D}(n\,|\,n') = \frac{n!}{(n-n')!}\left(\frac{n'}{n} + \frac{n'(2)}{2!n_{(2)}} + \frac{n'(3)}{3!n_{(3)}} + \text{etc.}\right), \tag{12}$$

la formule (9), pour $n' = n$, apparaissant comme un cas particulier de cette formule.

À titre illustratif, le Tableau 3 présente le nombre d'arrangements produisant les différentes valeurs de $c_{\max}$ pour des groupes de $n = 7$ membres dans lesquels seuls $n' \leq 7$ membres votent.

Les sociogrammes irréguliers. Tout est possible, en particulier les cas où chaque membre exprime un nombre variable de voix ($r_i \geq 0$) ou ceux où il peut voter pour lui-même. Les chemins des sociogrammes dits réguliers s'étant révélés ardus à défricher, nous avons ignoré les cas irréguliers. Ces derniers cependant restent d'étude légitime et correspondent occasionnellement à des cas réels, comme on le verra en épilogue. Pour ceux-là, une quantification approximative par la méthode Monte Carlo (voir p. ex. Laurencelle, 2001), en échantillonnant au hasard uniforme parmi les T arrangements possibles, permettra d'estimer la distribution exacte de $c_{\max}$.

Distributions de probabilité de $c_{\max}$. Au delà des problèmes liés aux nombres d'arrangements associés à chaque valeur de $c_{\max}$, nous avons voulu vérifier s'il existait des régularités utiles dans les distributions de probabilités de $c_{\max}$ selon n. Pour des tailles n dépassant 10, nous avons procédé par méthode Monte Carlo en échantillonnant 10.000.000 arrangements aléatoires de votes pour $n = 11$ à 100. Les données, correspondant à un sociogramme régulier où

† Il y a toutefois $\mathrm{C}(n,n')$ regroupements possibles de n' votants, d'où l'ensemble total d'arrangements est en réalité $C(n,n')(n-1)^{n'}$. Les $\mathrm{C}(n,n')$ regroupements engendrent cependant tous la même répartition aléatoire de $c_{\max}$; nous nous en tiendrons donc au « premier » regroupement.

Tableau 3. Nombre d'arrangements des choix mutuels de n' personnes dans un groupe de n = 7 produisant c_{max} choix

c_{max} / n'	7	6	5	4	3	2	1
1	1854	2119	1214	465	134	31	6
2	161070	31845	5380	744	78	5	
3	99960	11350	1100	84	4		
4	15750	1275	80	3			
5	1260	66	2				
6	42	1					
Total	279936	46656	7776	1296	216	36	6

chaque membre vote, sont partiellement transcrites au Tableau 4.

Il appert que la moyenne de c_{max} suit une progression logarithmique en fonction de n ; en excluant les tailles n inférieures à 10, nous obtenons:

$$\text{moy}\ (c_{max}) \approx 0{,}64 \log_e n + 1{,}29 \quad (R^2 \approx 0{,}999)\ . \qquad (13)$$

D'autre part, l'écart-type semble se stabiliser près de 0,75, ou bien évoluer très paresseusement; grâce à sa petitesse, l'écart-type indique que la distribution est massée dans seulement quelques valeurs principales de c_{max}. L'inspection des distributions exactes et Monte Carlo confirme en effet que, pour n allant de 6 à 19, 95 pour-cent des fréquences échoient dans l'intervalle c_{max} = (2, 4); pour n = 20 à 29, l'intervalle central est (2, 5), et au delà de n = 30 (jusqu'à 100), l'intervalle devient (3, 5). La distribution est positivement asymétrique et pointue (leptokurtique), comme on peut s'y attendre dans le cas d'un maximum.

Nous avons tenté de trouver un modèle approximatif de la distribution de c_{max}, en vain; les modèles binomiaux simples, de loi de Poisson, de loi Gamma, etc. ne conviennent aucunement, leurs propriétés jurant avec celles observées dans nos distributions. Parmi les modèles explorés, celui qui s'est montré le moins insatisfaisant est celui du maximum de n' binomiales, décrit plus haut : voir un exemple en annexe. Pour un sociogramme simple, où n' = n, les probabilités obtenues se concentrent à peu près dans l'intervalle c_{max} = (2, 5), et les valeurs de queue (c_{max} > 4) approchent remarquablement les probabilités réelles.

Ces données nous permettent d'établir un jeu de valeurs critiques de c_{max}, des valeurs très fiables grâce à la grande stabilité des distributions Monte Carlo obtenues; ces valeurs apparaissent au Tableau 5.

Les partitions de l'entier *n*

Les problèmes du sociogramme, que nous venons d'étudier avec un succès mitigé, nous ont mis face à une question qui paraissait d'abord petite, voire oiseuse: celle du nombre de partitions d'un entier n, ou p(n). Cette question apparaît dès qu'on explore les configurations « majoritaires » que peuvent endosser les votes d'un ensemble de n membres votant entre eux. Si l'on admet par exemple que chacun peut voter pour soi-même, et qu'on considère 5 votants, les différentes configurations majoritaires possibles sont: {5}, {41}, {32}, {311}, {221}, {2111} et {11111}; on considère ici les répartitions (41) et (14) comme une seule et même partition, qu'on décline de façon descendante sous forme {41}. Dans une configuration affichant k termes, on sous-entend n– k zéros, p. ex. {32} représente (32000), le candidat favorisé ayant obtenu 3 voix, le second en ayant 2, et les trois autres aucune. Les partitions de n sont aussi les différentes façons de fragmenter l'entier n en un ou plusieurs morceaux non nuls, quel que soit l'ordre des morceaux. Comme on voit, il y a 7 partitions de n = 5, ou p(5) = 7. Voici quelques valeurs de p(n):

n	1	2	3	4	5	6	7	8	9	10
p(n)	1	2	3	5	7	11	15	22	30	42

Excepté par leur énumération, comment trouver le nombre de partitions de n?

Bose et Manvel (1984) suggèrent la fonction génératrice suivante:

$$\prod_{k=1}^{\infty} (1-x^k)^{-1} = \frac{1}{(1-x)(1-x^2)(1-x^3)\cdots}\,; \qquad (14)$$

Tableau 4. Statistiques descriptives[a] de la distribution de c_{max} dans le contexte d'un sociogramme régulier, pour différentes tailles n

c_{max}	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$
5	2,230	0,550	0,760	1,093
10	2,731	0,709	0,742	0,539
15	3,025	0,742	0,627	0,932
20	3,226	0,741	0,691	1,239
25	3,374	0,741	0,790	1,328
30	3,491	0,743	0,842	1,227
40	3,668	0,754	0,836	0,995
50	3,804	0,764	0,770	0,901
60	3,917	0,768	0,710	0,921
70	4,013	0,767	0,677	1,008
80	4,095	0,763	0,668	1,114
90	4,168	0,759	0,676	1,196
100	4,232	0,755	0,698	1,270

[a] Les indices γ_1 et γ_2 décrivent respectivement l'asymétrie et la voussure (ou aplatissement) et sont tous deux nuls dans une distribution normale. Noter que les statistiques pour n = 5 et 10 sont exactes.

en développant cette expression comme un polynôme en x, le coefficient de x^n est p(n), la valeur cherchée. D'autres approches sont possibles. Riordan (1958) rapporte qu'Euler (1707-1783) proposait une autre fonction, plus pratique, pour obtenir les partitions de n en parties (ou éléments) désignées: par exemple, les partitions de 6 en parties égales à 1, 2, …, répétées ou non. Selon le même auteur (Riordan 1958, p. 120 et suivantes), c'est Gupta qui fournit la méthode de calcul la plus utile, à partir du nombre de partitions de n constituée d'une partie égale à m et d'autres parties égales ou supérieures à m. Notre méthode, plus bas, est quelque peu similaire.

Partitions de longueur k. On peut subdiviser p(n) selon la *longueur* des partitions engendrées, et alors p(n) = s(n,1)+s(n,2)+…+s(n, n); par exemple s(n,2) est le nombre de partitions de longueur 2, à forme $\{r_1, r_2\}$. On voit tout de suite que s(n,1) = 1 et s(n,n) = 1. Pour obtenir s(n, 2), il faut respecter les contraintes $r_1 + r_2 = n$ et $r_1 \geq r_2$, d'où $r_2 = 1,2,\ldots,\lfloor \frac{1}{2}n \rfloor$ et par conséquent s(n,2) = $\lfloor \frac{1}{2}n \rfloor$.[‡] Pour s($n$, 3), on doit compter $\{r_1,\{r_2, r_3\}\}$. La première position r_1 varie de n–2 (e.g. $\{n-2,1,1\}$) jusqu'à $\lceil n/3 \rceil$.[§] Une fois r_1 fixée, $n-r_1$ unités doivent être réparties en deux positions r_2, r_3, sous la condition supplémentaire que $r_1 \geq \{r_2,r_3\}$. La valeur de s(n,3) est la somme $\lfloor \frac{1}{2}(n-1) \rfloor + \lfloor \frac{1}{2}(n-4) \rfloor + \lfloor \frac{1}{2}(n-7) \rfloor + \ldots + 1$, soit $\lfloor (n+1)(n-1)/12+\frac{1}{2} \rfloor$. Une évaluation semblable de s(n,4) donne les résultats *approximatifs* suivants, selon que n = 4k, 4k+1, 4k+2 ou 4k+3:

$$s(n=4k+0,4) \approx [(n-4)/144][36 + 9n + n(n-2)] ;$$

$$s(n=4k+1,4) \approx [(n-1)/144][36 + 9(n-5) + n(n-5)] ;$$

$$s(n=4k+2,4) \approx [(n-1)/144][45 + 9(n-2) + (n-6)(n+2)] ;$$

$$s(n=4k+3,4) \approx [(n-3)/144][36 + 9(n-3) + (n-4)(n+1)].$$

En procédant ainsi pour les valeurs ultérieures de s(n, k), les approximations obtenues deviennent de moins en moins justes.

Le tableau 6 présente les nombres de partitions de toutes longueurs, pour n allant de 1 à 10. On y remarque que s(n, n–1) = 1 pour $n \geq 2$; s(n, n–2) = 2 pour $n \geq 4$; s(n, n–3) = 3 pour $n \geq 6$; voire, généralement, s(n, $n-r$) = p(r) pour $n \geq 2r$.

[‡] La notation $\lfloor x \rfloor$ indique la valeur entière la plus grande mais non supérieure à x, soit la partie entière de x.

[§] De même, la notation $\lceil x \rceil$ indique la valeur entière la plus petite mais non inférieure à x.

Tableau 5. Valeurs critiques de c_{max} pour un sociogramme simple à n membres votants, à différents rangs centiles

N	$1-P \le 0{,}1$	$1-P \le 0{,}05$	$1-P \le 0{,}01$	$1-P \le 0{,}001$
2 – 4	–	–	–	–
5	4	4	–	–
6	4	4	5	–
7 – 8	4	5	5	6
9	5	5	6	6
10 – 19	5	5	6	7
20 – 24	5	6	6	7
25 – 26	5	6	6	8
27 – 34	5	6	7	8
35 – 96	6	6	7	8
97 – 117	6	7	7	8
118 – 135	6	7	7	9
136 – 150(+)	6	7	8	8

Chaque colonne du tableau est ainsi une reconstitution de p(n), de zéro en montant, en supposant p(0)=1. Pour obtenir p(n), il reste donc à trouver s(n,1), s(n,2), s(n,3), ..., s(n, $\lfloor n/2 \rfloor$), ensemble dont nous avons fixé les premiers éléments plus haut.

Chaque partition dans s(n, k) comporte k éléments $r_i \ge 1$; soustrayant 1 à chacun, nous avons un arrangement d'éléments $q_i = r_i - 1$ correspondant à présent au nombre total $n-k$. Soit alors s(n, k), le nombre de pseudo partitions admettant un ou plusieurs éléments nuls. Ignorant les éléments nuls, on voit que nous obtenons un sous-ensemble des partitions de $n-k$, excluant celles de longueurs supérieures à k. Ainsi:

$$\mathrm{s}(n, k) = \mathrm{p}(n-k) - \mathrm{s}(n-k, k+1) - \mathrm{s}(n-k, k+2) - \ldots - \mathrm{s}(n-k, n-k) \; ;$$

Par soustraction, on arrive enfin à:

$$\mathrm{s}(n, k) = \mathrm{s}(n-k,1) + \mathrm{s}(n-k,2) + \ldots + \mathrm{s}(n-k, k), \qquad (15)$$

expression récursive facile (dans laquelle s(n,1) = 1 pour tout $n \ge 0$), qui permet d'obtenir aussi p(n).

Partitions terminées par i. Une autre façon d'énumérer les partitions est de les catégoriser selon leur nombre terminal. Soit les partitions de n, et $\mathrm{h}_i(n)$, le nombre de partitions terminées par « i », $i \ge 1$. On voit que $\mathrm{h}_n(n) = 1$, puisque {n} est une partition. Outre $i = n$, les autres valeurs possibles de i sont 1, 2, …, $\lfloor ½\, n \rfloor$. Le nombre total de partitions peut donc s'établir selon:

$$\mathrm{p}(n) = 1 + \mathrm{h}_1(n) + \mathrm{h}_2(n) + \ldots + \mathrm{h}_r(n), \; r = \lfloor ½\, n \rfloor . \qquad (16)$$

Tableau 6. Nombre de partitions de n de longueur k, s(n, k)

k / n	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	1	1	1
2		1	1	2	2	3	3	4	4	5
3			1	1	2	3	4	5	7	8
4				1	1	2	3	5	6	9
5					1	1	2	3	5	7
6						1	1	2	3	5
7							1	1	2	3
8								1	1	2
9									1	1
10										1
p(n)	1	2	3	5	7	11	15	22	30	42

Peut-on déterminer $\mathrm{h}_i(n)$?

Pour $\mathrm{h}_1(n)$, on voit tout de go que les partitions désignées ont la structure {{n–1},1}; le sous-ensemble identifié {n–1} contient les partitions de n–1 éléments, toutes admissibles avec le successeur « 1 », d'où l'on a $\mathrm{h}_1(n) = \mathrm{p}(n-1)$.

Pour $\mathrm{h}_2(n)$, la structure correspondante est {{n–2},2}. Ici le sous-ensemble dénoté {n–2} contient des partitions non admissibles, celles terminées par « 1 ». On inclura donc seulement { {n–2} – {{n–3},1} }, soit au total p(n–2) – p(n–3).

La relation générale est récursive, et elle a la forme suivante:

$$\mathrm{h}_i(n) = \mathrm{p}(n-i) - \mathrm{h}_1(n) - \mathrm{h}_2(n) - \ldots - \mathrm{h}_{i-1}(n) \, . \qquad (17)$$

Le calcul se complique rapidement. Les premiers termes sont les suivants: $\mathbf{h_1(n)}$= p(n–1); $\mathbf{h_2(n)}$= p(n–2) – p(n–3); $\mathbf{h_3(n)}$= p(n–3) – p(n-4) – p(n–5) + p(n–6); $\mathbf{h_4(n)}$= p(n–4) – p(n–5) – p(n–6) + p(n–8) + p(n–9) – p(n–10); $\mathbf{h_5(n)}$ = p(n–5) – p(n–6) – p(n–7) + 2p(n–10) – p(n–13) – p(n–14) + p(n–15); $\mathbf{h_6(n)}$ = p(n–6) – p(n–7) – p(n–8) + p(n–11) + p(n–12) + p(n–13) – p(n–14) – p(n–15) –p(n–16) + p(n–19) + p(n–20) – p(n–21); etc. Noter que partout, p(0)=1 et p(x)=0 pour x < 0.

On peut combiner ces résultats sous une forme plus utile, soit les sommes $\mathrm{Q}_i = 1 + \mathrm{h}_1(n) + \mathrm{h}_2(n) + \ldots + \mathrm{h}_i(n)$. On obtient ainsi:

$$\mathrm{Q}_1 = 1 + \mathrm{p}(n-1)$$

$$\mathrm{Q}_2 = 1 + \mathrm{p}(n-1) + \mathrm{p}(n-2) - \mathrm{p}(n-3) \qquad (18)$$

```
program partitions; var n : integer;
    function p(n: integer): integer; var s,i: integer;
    function h(i,n: integer): integer; var s,j: integer;
    begin
        s:=p(n-i); for j:=1 to i=1 do s:=s-h(j,n-i); h:=s
    end;

    begin {fonction p}
        if n<0 then p:=0 else if n<2 then p:=1 else
        begin
            s:=1; for i:=1 to (n div 2) do s:=s+h(i,n); p:=s
        end
    end;

begin {programme partitions}
    write(' n:'); readln(n); writeln('p(n) =', p(n))
end.
```

Figure 1. Programme de calcul récursif du nombre de partitions, p(n)

$$Q_3 = 1 + p(n-1) + p(n-2) - p(n-4) - p(n-5) + p(n-6)$$

$$Q_4 = 1 + p(n-1) + p(n-2) - 2p(n-5) + p(n-8) + p(n-9) - p(n-10)$$

$$Q_5 = 1 + p(n-1) + p(n-2) - p(n-5) - p(n-6) - p(n-7) + p(n-8) + 2p(n-10) - p(n-13) - p(n-14) + p(n-15)$$

$$Q_6 = 1 + p(n-1) + p(n-2) - p(n-5) - 2p(n-7) + 2p(n-10) + p(n-11) + p(n-12) - 2p(n-14) - p(n-16) + p(n-19) + p(n-20) - p(n-21)$$

Par exemple, pour trouver p(7), il faut additionner $1+h_1(7)+h_2(7)+h_3(7)$, c.-à-d. $p(7) = Q_3(7) = 1 + p(6) + p(5) - p(3) - p(2) + p(1) = 1+11+7-3-2+1 = 15$. Il y a peut-être une structure (rigoureuse) dans les Q_i; je ne les ai pas développées assez pour la découvrir.

La définition récursive de p(n) élaborée ici permet facilement leur évaluation numérique. Le programme suivant (voir Figure 1), de langage Pascal, réalise ce calcul[**].

La décomposition d'un entier en ses parties a donné lieu au problème des partitions, mais aussi à d'autres problèmes intéressants, tels celui des compositions de n (dans lesquelles l'ordre des parties non égales importe), celui des partitions à parties toutes distinctes, des partitions à parties toutes inférieures à i, etc. Il est amusant de voir qu'un problème aussi beau, profond et abstrait ait surgi de considérations sur le choix d'un membre exceptionnel dans un groupe!

Épilogue: un choix concret dans des sociogrammes irréguliers

Le problème du sociogramme me fut posé au printemps 1992 par Mmes Nicole Briand et Marie Gendron, Ph.D.

[**] Pour un programme d'exécution efficace, il vaudrait mieux mémoriser dans un vecteur les valeurs antécédentes de p(n) afin de n'avoir pas à les recalculer chaque fois en effectuant la fonction "h(i, n)".

Tableau 7 - Données de 5 sociogrammes expérimentaux, d'après Nicole Briand*

Centre de Soins	Membres (n)	Votants (n')	Votants doubles**	Candidat	Votes (c_{max}/n')	Probabilité extrême***
C.S. #1	29	17	1	1	6/17	< 0,001
				2	4/17	0,077
C.S. #2	24	5	1	2	2/5	0,472
C.S. #3	28	5	2	1	3/5	0,027
C.S. #4	31	12	1	1	4/12	0,018
C.S. #5	219	80	2	1	8/80	< 0,001
				2-4	5/80	0,009
				5-10	3/80	0,788

* Lettre du 28 mai 1992 à l'auteur.

** Sous « Votants doubles », sont dénombrés ceux des "Votants" ayant présenté un vote surnuméraire. Par exemple, en C.S. #1, 17 membres ont voté, dont 16 ont donné une voix et un en a donné deux.

*** Les probabilités estimées sont justes à mieux que ±0,01, au niveau de confiance de 0,95.

(sciences infirmières), qui se voyaient confrontées à des choix difficiles. Afin d'identifier quelques personnes éminemment typiques de leur profession, elles avaient demandé à des infirmières, infirmiers et soignants, dans certains centres de soins gérontologiques, de désigner « le membre de mon groupe aux soins de qui je préférerais être confié(e) dans ma vieillesse ». Tel était le sens de la question, et chaque personne devait en pointer une autre *dans le groupe*. Par la suite, il s'agissait de compiler les votes, ou désignations exprimées, et de déterminer d'après eux le ou les candidats exceptionnels. Un portrait professionnel et personnel de ces candidats d'exception compléterait l'étude.

La première phase de l'étude consistait donc à trouver les candidats exceptionnels, désignés par voie élective par leurs collègues de travail. Cinq centres de soins furent investis; les données d'enquête obtenues[††] apparaissent au Tableau 7. Ainsi, le centre #1 comportait 29 intervenants qualifiés. De ces 29, 17 exprimèrent leur préférence, l'un des 17 ayant désigné deux personnes. Dans ce centre, le candidat préféré obtint les suffrages de 6 des 17 votants. C'est le candidat le plus favorisé; est-il « exceptionnel » pour autant?

Le modèle d'interprétation stochastique que nous avons élaboré plus haut permet de déterminer si le candidat favorisé est exceptionnel, et permet même d'en déterminer le degré. Ce modèle est toutefois assorti de sévères conditions d'interprétation, qui peuvent en réduire la portée. La première condition est que, à moins qu'un membre soit vraiment exceptionnel, le vote à son endroit se fait au hasard, comme on pige dans un chapeau. La seconde condition stipule que le choix au hasard s'exerce également parmi tous les membres éligibles (sauf soi-même). La troisième condition a trait à l'indépendance des votes, à savoir que chacun vote à l'insu et sans relation avec le vote d'autrui.

Dans un groupe réel, particulièrement dans un centre professionnel réunissant des collègues accoutumés les uns aux autres à divers degrés et domiciliés pour la plupart dans le même environnement urbain, les conditions 1 et 2 sont peu plausibles. Vis-à-vis de la condition 2, en particulier, il faut opposer l'argument que les intervenants peuvent travailler dans des équipes spécialisées ou dans des régimes horaires favorisant ou non leurs rapprochements. En raison de ces rapprochements, un membre choisira plutôt entre les collègues qu'il connaît que parmi ceux, salariés de la même institution, qu'il ne fréquente presque jamais.

De meilleurs modèles restent donc désirables, mais ces modèles exigeraient davantage d'information sur les structures sociales auxquelles on souhaite les appliquer.

À toutes fins pratiques, et comme un modèle même imparfait est utile puisqu'il permet d'interpréter la réalité, nous avons appliqué notre modèle « égalitaire » aux données du Tableau 7. Comme il s'agissait de données émanant de sociogrammes à tailles fortes et, pour la plupart, de type irrégulier, nous avons procédé à l'évaluation par

[††] Ces données anonymes sont reproduites avec la permission de Mme Nicole Briand.

méthode Monte Carlo: chaque centre, caractérisé par ses membres (n) et ses membres votants (n'), incluant les doublets, fut constitué en modèle stochastique et échantillonné 10.000 fois, les votes maximaux étant enregistrés.

Les probabilités indiquées en dernière colonne, au Tableau 7, révèlent le degré d'exceptionnalité des candidats favorisés, numérotés à la colonne « Candidat ».

Il est apparent que la puissance de cette technique d'évaluation, sa capacité à déclarer exceptionnel le choix d'un candidat, dépend du nombre de votants (n'), tandis que la probabilité extrême de c_{max}, l'exceptionnalité, est davantage influencée par le *nombre* plutôt que par la *proportion* de voix reçues. Nous avons déjà remarqué que, dans les sociogrammes simples et pour n allant de 6 à 90 ou plus, les votes se massent aux valeurs c_{max} de 2 à 5, et que la probabilité extrême d'une *proportion* donnée de voix chute rapidement vers zéro quand n augmente. Remarquons par exemple que la vedette du C.S. #3, avec 3/5 ou 60 % des voix, a une exceptionnalité moins prononcée (P_{ext} = 0,027) que celle du C.S. #4, avec 4/12 ou 25 % (P_{ext} = 0,018). À un seuil d'exceptionnalité de 5 pour-cent, 7 candidats seraient retenus, soit les candidats 1 des centres #1, #3, #4 et #5, ainsi que les candidats 2, 3 et 4 du centre #5. Dans le cas de ce centre de grande taille (n = 219), on a vraisemblablement affaire à des sous-groupes ou équipes; le cas échéant, les probabilités fournies seraient négativement biaisées et devraient être plus fortes, mais l'importance des c_{max} (= 5) laisse croire qu'il s'agit là tout de même de candidats exceptionnels.

Références

Bastin, G. (1970). *Les techniques sociométriques*. Paris, Presses Universitaires de France (collection S.U.P.).

Bose, R.C., Manvel, B. (1984). *Introduction to combinatorial theory*. New York, Wiley.

Bronfenbrenner, V. (1945). The measurement of sociometric status, structure and developpment. *Sociometry Monographs*, *6*.

Brualdi, R.A. (1977). *Introductory combinatorics*. New York, North-Holland.

Feller, W. (1950). *An introduction to probability and its applications* (2 vol.). New York, Wiley.

Gebhardt, F. (1969). Some numerical comparisons of several approximations to the binomial distribution. *Journal of the American Statistical Association*, *64*, 1638-1648.

Johnson, N.L., Kotz, S. (1969). *Distributions in statistics. Discrete distributions*. Boston, Houghton-Mifflin.

Laurencelle, L. (2001). *Hasard, nombres aléatoires et méthode Monte Carlo*. Québec : Presses de l'Université du Québec.

Moreno, J.L. (1943). *Fondements de la sociométrie* (traduction 1970). Paris, Presses Universitaires de France.

Riordan, J. (1958). *An introduction to combinatorial analysis*. New York, Wiley.

L'annexe suit

Annexe : Approximation de la distribution de c_{max} par le maximum de n binomiales

Les données rapportées au tableau 8 détaillent les calculs requis pour la distribution de probabilités du maximum de n binomiales, comme approximation à la distribution exacte de c_{max}, cette dernière étant obtenue par énumération.

Rappelons d'abord que, pour l'énumération, la combinatoire des choix dans un sociogramme régulier de n membres, chacun exprimant un vote pour quiconque sauf lui-même, a une taille de $(n-1)^n$; ici, pour n = 10, nous comptons $(10-1)^9$ = 3.486.784.401 arrangements. Pour $x = c_{max} = 1$, l'approximation (10) donne $e^{-1}\times 10! \approx 1334961$, soit une fraction de 0,0003829 ou $,0^3383$. Le lecteur pourra vérifier l'utilité pour cet exemple des autres formules suggérées plus haut.

L'approximation par le maximum de n binomiales débute par l'obtention des probabilités binomiales individuelles, $p_x = {}_{N}C_x\pi^x(1-\pi)^{N-x}$, pour une loi à paramètres $\pi = 1/(n-1)$ et $N = n-1$. Pour n = 10, π = 1/9 et N = 9, nous avons p. ex. $p_2 = {}_9C_2\times(1/9)^2(8/9)^7 \approx 36 \times 0{,}0123 \times 0{,}4385 \approx 0{,}1949$. La colonne suivante, titrée « p_cum », présente les valeurs cumulatives de ces probabilités, depuis 0 jusqu'à $N = n-1$. P. ex., p_cum pour x = 3 s'obtient par $p_1 + p_2 + p_3$ = 0,3464 + 0,3897 + 0,1949 = 0,9310, ou ,9311 en conservant plus de précision. La probabilité qu'une variable binomiale X ait pour maximum $x = t$ est égale à p_cum_t, et la probabilité que n binomiales indépendantes de même sorte ait pour maximum $x = t$ est alors $[p\_cum_t]^n$. C'est la série de valeurs apparaissant à la quatrième colonne du tableau 8. P. ex., pour le maximum $x = t = 3$, nous calculons $(0{,}9879)^{10} \approx 0{,}8853$. Les probabilités individuelles (non cumulatives) de x, qui approcheraient celles de c_{max}, sont ensuite obtenues par soustraction. Ainsi, pour $x = c_{max} = 5$, nous faisons $[p\_cum_5]^{10} - [p\_cum_4]^{10} = 0{,}9^2883 - 0{,}9856 \approx 0{,}0132$.

Le lecteur constatera par lui-même que notre variable d'approximation, le maximum de n binomiales, constitue un modèle boiteux pour la distribution de c_{max}, ce particulièrement dans les basses valeurs de x. En fait, c_{max} n'a pas de zéro, et les basses probabilités du maximum binomial sont trop fortes, abaissant ainsi la moyenne de la distribution. Pour ce cas précis, le maximum binomial a pour paramètres globaux μ = 2,594 et σ = 0,793, alors que les valeurs correspondantes de c_{max} sont μ = 2,731 et σ = 0,709.

Cependant, comme le tableau le montre aussi, les valeurs hautes des deux distributions se rejoignent et paraissent s'épouser étroitement, et ce dès $x = c_{max} = 4$. Cette conjugaison numérique a pour effet avantageux de permettre l'établissement de valeurs critiques aux rangs centiles supérieurs, et cela, sans craindre la fluctuation aléatoire attachée aux estimations Monte Carlo.

Tableau 8. Distribution du maximum de n binomiales $B(\pi = 1/(n-1)\ ;\ n-1)$ vs la distribution exacte de c_{max} obtenue par énumération, pour un sociogramme régulier avec n = 10

x (c_{max})	p(Bin)	p_cum (Bin)	$[p\_cum\ (Bin)]^{10}$	$\Delta\, p$	p exacte
0	,3464	,3464	$,0^4249$	$,0^4249$	–
1	,3897	,7362	,0468	,0467	$,0^3384$
2	,1949	,9311	,4895	,4427	,4044
3	,0568	,9879	,8853	,3958	,4751
4	,0107	$,9^2855$	,9856	,1003	,1057
5	$,0^2133$	$,9^3883$	$,9^2883$	,0132	,0133
6	$,0^3111$	$,9^5386$	$,9^4386$	$,0^2111$	$,0^2111$
7	$,0^5595$	$,9^6812$	$,9^5812$	$,0^4595$	$,0^4595$
8	$,0^6185$	$,9^8742$	$,9^7742$	$,0^5186$	$,0^5186$
9	$,0^8258$	1	1	$,0^7258$	$,0^7258$