Les Cahiers de l'Analyse des Données
Vol. XXI - 1996 - n°3 - pp.275-288.

EXEMPLE DE PROCESSUS TENDANT À ÉTABLIR LA RÉPARTITION ALTERNÉE DES LETTRES DANS UNE SÉQUENCE FORMÉE DE 'A' ET DE 'B'

[RÉPARTITION ALTERNÉE]

A. ALAWIEH

1 Origine multidimensionnelle du problène de la répartition alternée

Le Traité de Physique de LANDAU et LIFSCHITZ est justement célèbre; et, du russe, il a été traduit dans plusieurs langues. Dans le volume consacré à la thermodynamique, les auteurs considèrent, comme offrant un exemple de transition de phase, la combinaison Cu Zn.

Dans la forme solide la plus régulière de Cu Zn, les atomes de chacun des deux métaux sont disposés suivant un réseau cubique dont chaque maille a pour centre un atome de l'autre réseau. Quand la température s'élève, l'ensemble des atomes des deux métaux reste disposé suivant deux réseaux cubiques mutuellement imbriqués; mais la répartition des atomes de Cu et de Zn devient aléatoire s'étendant aux deux réseaux.

En un sens obvie, cette répartition aléatoire détruit l'ordre du cristal; mais, d'autre part, et L.&L. insistent sur ce fait, il en résulte que le groupe de symétrie comprend, outre les transformations laissant invariant un réseau cubique, les composés de ces transformations avec celles échangeant les deux réseaux imbriqués: par exemple, les translations allant d'un sommet d'une maille cubique au centre de celle-ci. Quant à nous, dans le présent exposé, nous préférons dire que la forme dans laquelle Cu et Zn sont disposés suivant deux réseaux distincts est la forme régulière; tandis qu'avec la répartition aléatoire de Cu et Zn, on a la forme irrégulière.

Pour apprécier globalement la régularité, ainsi comprise, d'une structure, on peut, sur l'ensemble des segments joignant le centre d'une maille cubique à un sommet de celle-ci, (nous dirons alors qu'il s'agit de deux atomes voisins,) déterminer le taux, τ, de ceux dont les extrémités sont occupées par des atomes de métaux différents: la régularité parfaite sera caractérisée par $\tau=1$; l'irrégularité maxima par $\tau=1/2$; tandis que des valeurs de τ inférieures à 1/2 indiqueraient que la combinaison tend à se séparer en ses deux métaux constituants.

Les cahiers de l'analyse des données 0339-3097/96/03/$ 7.00/© Gauthier-Villars

Il est vraisemblable, en effet, que la mobilité des atomes interdit de distinguer, à grande échelle, entre deux réseaux, celui de Cu et celui de Zn; les irrégularités étant dénombrées comme le nombre des atomes de Zn occupant des places du réseau de Cu, et *vice versa.* Nous dirons plutôt qu'il y a une tendance locale, plus ou moins forte, à ce que des groupes de 8 atomes de l'un des métaux soient placés aux sommets d'un cube dont le centre est occupé par un atome de l'autre métal.

Le problème se pose de déterminer ce que peut être la loi locale des tendances à la permutation entre atomes voisins qui favorisent les formes régulières; et, en particulier, de concevoir l'efficacité d'un processus, qui, à partir d'une répartition aléatoire des deux métaux, engendrerait une répartition régulière. Dans les §§2 et 3, nous considérons un tel processus, dans le cas simple d'une distribution linéaire. Aux §4 et 5 , nous donnons, pour ce même cas, des formules combinatoires, telles que celles utilisées en thermodynamique pour caractériser la distribution des formes en fonction de la température; et considérons de ce point de vue la tendance à la régularité.

Le cas d'une distribution non linéaire, mais plane ou spatiale, comme pour les atomes de Cu et Zn, est seulement abordé au §6.

2 Processus de régularisation en dimension 1

Tel quel, le problème de la répartition des atomes nous paraît difficile. C'est pourquoi nous prenons un exemple analogue, mais beaucoup plus simple. Des lettres A et B sont disposées sur une ligne. On suppose, pour plus de précision, que le nombre des A est égal à celui des B; ou diffère de celui-ci d'une unité. Et on considère un processus propre à ranger les lettres dans l'ordre ABABABABABAB...

Partons d'une séquence quelconque de longueur *n*: l'irrégularité de la séquence peut être mesurée par le nombre des positions j ($1 \leq j \leq n-1$) telles que les lettres de rang j et j+1 soient identiques (i.e., toutes deux des A, ou toutes deux des B).

Nous dirons qu'une permutation entre deux lettres consécutives XY de la séquence est acceptable si $X \neq Y$ et qu'est réalisée l'une des deux conditions suivantes: la lettre qui précède X est identique à X; ou la lettre qui suit Y est identique à Y (i.e.: XXY ou XYY). Si les deux conditions sont réalisées simultanément, on dira que la permutation est favorable (i.e.: XXYY). Il est clair qu'effectuer une permutation favorable augmente la régularité de 2; tandis qu'une permutation acceptable, mais non favorable, laisse l'irrégularité inchangée (mis à part le cas où la séquence commence par XYY).

On postule le processus de régularisation suivant: si la séquence admet des permutations favorables, se réalise l'une de celles-ci, choisie au hasard;

sinon est choisie, au hasard également, une permutation acceptable. En général, une suite de permutations satisfaisant aux conditions ainsi imposées au déroulement du processus sera appelée: Suite Compatible.

Dans le cas où, au terme d'une suite compatible, il n'y a pas de permutation acceptable, il est clair que, sous les hypothèses faites quant aux nombres des A et des B, le processus a abouti à une séquence régulière.

Pour démontrer que cette éventualité se réalise avec une probabilité arbitrairement proche de 1 au bout d'un temps suffisamment grand, dépendant de la longueur, L, de la séquence (par temps, on entend ici: nombre de permutations réalisées), il suffit d'établir qu'il existe, pour toute séquence S, une suite compatible de permutations bien choisies qui régularise S.

En effet, en bref, si pour tout L existe un nombre N tel que toute séquence de longueur L puisse être régularisée par une suite compatible de m permutations, avec $m \leq N$, il y a une probabilité π, fonction seulement de L, telle qu'une suite compatible de permutations choisies au hasard régularise S en un temps $\leq$N: la probabilité que le processus se prolonge sans succès au-delà du temps R.N est donc moindre que $(1-\pi)^R$.

3 Démonstration par récurrence de l'existence d'une régularisation

Reste donc à montrer qu'on peut choisir une suite compatible de permutations pour régulariser S. La proposition étant évidente si L=1 ou 2, on procédera par récurrence sur la longueur L de S.

De façon précise, on démontrera, par récurrence sur n, une suite de propositions (P.n) dont l'énoncé dépend de la parité de n; soit:

(P.2p) : Soit S une séquence formée de 2.p caractères, p fois A et p fois B, disposés dans un ordre quelconque: alors on peut régulariser S par une suite compatible de permutations dont chacune laisse fixe le premier caractère de S.

(P.2p+1) : Soit S une séquence formée de (2p)+1 caractères: p fois A, p fois B, et encore un caractère qui peut être A ou B, l'ordre initial étant quelconque; alors on peut régulariser S par une suite compatible de permutations.

Il est d'abord clair que (P.2p-1) implique (P.2p). En effet, soit S une séquence de 2p caractères, commençant, e.g., par un A; après cet A, on a une séquence, S', de 2p-1 caractères dont p fois B et (p-1) fois A. La séquence S' peut être régularisée par une suite compatible de permutations, et prend alors une forme qui, compte tenu du nombre des lettres, ne peut être que BABA...B: *ipso facto*, sans déplacer le premier caractère A, S est régularisée; et la suite des permutations effectuées, compatible pour S', l'est aussi pour S;

car la permutation du premier caractère avec le second ne peut jamais s'imposer comme favorable, faute d'un contexte à gauche.

Reste donc à démontrer (P.2p+1), sous l'hypothèse que sont déjà établies les propositions (P.r) de rang r≤2p. Pour fixer les notations, nous considérons une séquence S formée de (p+1) fois A et p fois B. La séquence S bien rangée doit prendre la forme ABABAB...ABA.

Si la séquence S commence par AB, on peut opérer sur la séquence S', formée des 2p caractères venant après le A initial. La séquence S', qui commence par un B, peut d'après (P.2p) être régularisée sous la forme BABABA...BA. Et l'on a ainsi une suite compatible pour régulariser S, car, dans le contexte de S tout entier, la permutation des deux premiers caractères ne peut jamais (cf. *supra*) s'imposer comme favorable.

Si la séquence S ne commence pas par AB, effectuons d'abord sur S, sans choix, des permutations favorables, aussi longtemps qu'il s'en trouve; (en nous arrêtant, toutefois, pour bénéficier de la récurrence, si apparaît, au début de S, la paire AB). Si le début de S est AAAAB..., avec r caractères A consécutifs suivi d'un B, il est acceptable de faire remonter le B jusqu'à la position r: AAABA...; et finalement jusqu'à la position 2; en effectuant, si elles s'imposent, des permutations favorables qui toutefois ne peuvent concerner le B, inclus entre deux caractères A, ni aucun de ces deux caractères. Ainsi transformée, la séquence S commence par AB: on entre dans la récurrence.

Il ne reste plus à traiter que le cas où, le nombre maximum de permutations favorables ayant été effectué, S commence par un B; et à montrer que, dans un tel cas, un A peut être mis en position initiale en effectuant une suite compatible.

Afin de rejoindre l'hypothèse de récurrence, nous imposons aux suites de permutations que nous effectuons la condition de ne comprendre que des permutations qui sont soit des permutations favorables soit des permutations acceptables qui diminuent strictement la somme des rangs, comptés à partir de la gauche, des p+1 caractères A que comprend la séquence S.

Du fait de cette condition complémentaire, le nombre des permutations acceptables intercalées entre deux permutations favorables est borné par un nombre max ne dépendant que de p (e.g.: max = 4.p.p); et comme le nombre des permutations favorables est lui-même borné, (par celui des irrégularités que comporte la séquence,) il apparaît que la longueur des suites de permutations que laisse possible notre condition est bornée par un nombre Max (ne dépendant que de p).

Convenons d'appeler suite maximale une de ces suites, prolongée aussi longtemps que notre condition le permet. Pour entrer dans l'hypothèse de récurrence, il nous suffit de démontrer qu'en effectuant une suite maximale on aboutit, nécessairement, à avoir un A au début de S.

En effet, si la séquence S, ainsi transformée, commençait par un B, elle comprendrait, nécessairement, au moins un bloc BAA (car, dans la séquence, le nombre des A surpasse, par hypothèse, celui des B); il serait acceptable de permuter ce bloc en ABA, opération qui diminue la somme des rangs des caractères A dans la séquence. Et la suite considérée ne serait pas maximale.

On peut donc, finalement, dans tous les cas, entrer dans l'hypothèse de récurrence: la démonstration est ainsi achevée.

4 Dénombrement, par type, des configurations d'une suite

Dans le présent §, nous considérons, comme précédemment, des séquences linéaires formées d'un nombre déterminé de lettres, en supposant, pour plus de généralité, qu'il y a n_a lettres A et n_b lettres B; ce qu'on appellera, en bref, des séquences de type (n_a, n_b).

On notera respectivement k_a et k_b le nombre des contacts entre deux lettres A et entre deux lettres B; autrement dit, des contacts AA et BB. De façon précise, k_a est le nombre de lettres A qui, sont, dans la séquence, suivies immédiatement par un autre A; et de même, *mutatis mutandis*, pour k_b.

Du fait de la présence des lettres B, les lettres A sont réparties en blocs connexes successifs: nous noterons r_a le nombre de ces blocs; de même, le nombre des blocs connexes formés de lettres B sera noté r_b.

Au sein de l'ensemble des séquences de type (n_a, n_b), nous dénombrerons celles présentant des nombres déterminés, k_a et k_b, de contacts respectifs.

Afin d'éviter les accumulations d'indices, on notera C(n, r) l'ensemble des combinaisons de n objets distincts pris r à r; soit:

$$C(n, r) = n! \ / ((n-r)! \, . \, r!) \ .$$

En particulier, le nombre des suites ordonnées, $(u_1, u_2, \ldots, u_r)$ de r entiers strictement positifs dont la somme est n est C(n-1, r-1). En effet, en bref, une telle suite peut être donnée en partant d'une suite S de (n-1) points, parmi lesquels on choisit, arbitrairement, (r-1) points de séparation: en augmentant de 1 la longueur, éventuellement nulle de chacun des r intervalles successifs ainsi délimités, dans la suite S, on a la suite cherchée. [Une autre démonstration procède par récurrence sur n].

Dans le dénombrement des séquences suivant les valeurs de k_a et k_b, nous distinguerons, selon ce que sont la première et la dernière lettre, 4 classes de séquences notées:

$$(A\ldots A), (A\ldots B), (B\ldots A), (B\ldots B) \quad .$$

Pour chaque classe $(X\ldots Y)$, on notera:

$$K[(X\ldots Y), n_a, n_b; k_a, k_b] \quad ,$$

le nombre des séquences de type (n_a, n_b) présentant k_a contacts AA et k_b contacts BB.

Il apparaît que, quelle que soit la classe $(X\ldots Y)$, on a toujours la relation combinatoire:

$$K[(X\ldots Y), n_a, n_b; k_a, k_b] \;=\; C(n_a-1, k_a) \,.\, C(n_b-1, k_b) \quad ;$$

mais la compatibilité des nombres (k_a, k_b) avec le type (n_a, n_b) dépend de la classe $(X\ldots Y)$ considérée.

En effet, considérons une séquence de type (n_a, n_b). Le nombre, k_a, des contacts AA, et le nombre, r_a, des blocs de A, sont liés à n_a par la relation:

$$k_a = n_a - r_a \quad ;$$

car chaque bloc renferme un nombre de contacts AA égal à sa longueur diminuée de 1; la somme des longueurs est n_a ; mais, relativement à n_a, le nombre des contacts perd autant de fois 1 qu'il y a de blocs. On a, de même, pour les lettres B:

$$k_b = n_b - r_b \quad .$$

Partons de la suite ordonnée des r_a longueurs des blocs de lettres A. D'après ce qu'on a rappelé plus haut, cette suite peut se présenter sous un nombre de formes distinctes donné par la formule:

$$C(n_a-1, r_a-1) = C(n_a-1, k_a) \quad .$$

Cherchons, à partir d'une telle suite, à reconstituer la séquence complète; ou, plus précisément, à déterminer de combien de façons la suite des blocs de A peut être complétée par l'insertion de blocs de B.

Le nombre, r_b, de ces blocs dépend de la classe de séquences que l'on considère; de façon précise, on a les relations de compatibilité:

$$(A\ldots A): \; r_a - 1 = r_b \quad ; \quad k_b = k_a + (n_b-n_a) + 1 \quad ;$$

$$(A\ldots B) \text{ et } (B\ldots A): r_a = r_b \quad ; \quad k_b = k_a + (n_b-n_a) \quad ;$$

$$(B\ldots B): \; r_a + 1 = r_b \quad ; \quad k_b = k_a + (n_b-n_a) - 1 \quad ;$$

il n'y a pas lieu d'épiloguer sur ces formules, enseignées à tous les enfants à propos d'arbres et d'intervalles.

Quant au nombre des formes que peut revêtir la suite ordonnée des r_b blocs de lettres B, il se calcule comme pour les blocs de A; et il vaut donc $C(n_b\text{-}1, k_b)$. D'où la formule annoncée:

$$K[(X\ldots Y), n_a, n_b; k_a, k_b] \;=\; C(n_a\text{-}1, k_a) \,.\, C(n_b\text{-}1, k_b) \quad ;$$

qui doit seulement être complétée par la relation de compatibilité entre k_a et k_b propre à chaque classe (X...Y).

5 Dénombrement des configurations et loi de régularité

Aux §§2 et 3, on a considéré un processus susceptible de donner forme régulière à toute séquence initiale formée de n lettres A et d'autant de lettres B (à une unité près). Moyennant quelques modifications, on peut donner de ce processus une interprétation mécanique, en terme d'énergie potentielle: l'échange entre X et Y a une probabilité forte dans le contexte XXYY, où il implique une forte diminution d'énergie; la probabilité est faible dans XXYX ou YXYY; nulle dans YXYX.

Mais, plus précisément, de tels processus sont considérés, en physique, dans le cadre de la thermodynamique. Et ils suivent des lois bien établies qui prennent en compte des dénombrements de configurations tels que ceux effectués au §4. Nous voulons ici considérer, de ce point de vue, la distribution probabiliste de la régularité d'une suite.

Revenant aux hypothèses des §§2 et 3, on considérera exclusivement des séquences pour lesquelles, selon les notations du §4: $n_a = n_b = n$.

La régularité de la séquence sera dite d'autant plus forte qu'est plus faible la somme $k_s = k_a + k_b$, des contacts AA et BB. On supposera que la tendance à la régularité résulte de ce que l'énergie d'une séquence est proportionnelle à k_s. Il serait facile de reprendre les calculs qui suivent sous des hypothèses plus générales, quant aux nombres des lettres et à l'expression de l'énergie.

On sait que, considéré comme fonction de Δ, $C(2m, m+\Delta)$ décroît, de part et d'autre de son maximum, atteint pour $\Delta = 0$, suivant le profil d'une loi normale de variance m/2. En effet, en bref, les $C(2m, m+\Delta)$ donnent le profil de la loi d'une somme de 2m termes aléatoires indépendants prenant équiprobablement les valeurs 1 et 0. On écrira:

$$C(2m, m+\Delta) \approx C(2m, m) \,.\, \exp(-\Delta^2/m) \quad ;$$

En supposant m grand, on peut se borner à des valeurs de Δ de l'ordre de $\sqrt{(m)}$; e.g.: $|\Delta| < 5.\sqrt{(m)}$. Dans ce cas le logarithme du rapport entre les deux termes équivalents, $C(2m, m+\Delta)$ et $(C(2m, m).\exp(-\Delta^2/m))$ est négligeable.

En effet, posons: $Q(2m, \Delta) = C(2m, m+\Delta)/C(2m, m)$; il vient:

$$Q(2m, \Delta) = (m/(m+1)) . ((m-1)/(m+2) \dots ((m+1-\Delta)/(m+\Delta))$$
$$= (1-(1/(m+1))) . (1-(3/(m+2))) \dots (1- ((2\Delta-1)/(m+\Delta))) \quad ;$$

pour rejoindre la formule exponentielle donnée ci-dessus, il suffit d'assimiler les facteurs $1-((2a-1)/(m+a))$ à $\exp(-(2a-1)/m)$; en bref, pour chaque terme, l'erreur relative est du second ordre, donc en $(a/m)^2$. L'erreur pour le produit est majorée par $(\Sigma a^2) /m^2$: elle est donc en Δ^3/m^2 ; ou encore en $1/\surd(m)$; donc négligeable. (En fait l'écart est en 1/m : voir §7, appendice, *in fine*).

Nous utiliserons la formule d'équivalence pour $C(n, (n/2)+\Delta)$ même si n n'est pas un entier pair, en supposant que C(n, p) est intrapolé linéairement ce qui n'introduit que des erreurs relatives négligeables.

Pour le nombre des séquences ayant, en tout, k_s contacts, on a (cf. §4) deux formules différentes, selon que k_s est pair ou impair:

$$k_s = 2.k \quad : \quad K[2n, 2k] \quad = 2 . C(n-1, k)^2 \quad ;$$
$$k_s = 2.k + 1 \quad : \quad K[2n, 2k+1] \quad = 2 . C(n-1, k) . C(n-1, k+1) \quad ;$$

(formules où, pour chacun des deux cas, sont cumulées les deux seules différentes classes (X...Y) de séquences possibles).

Mais, compte tenu des approximations adoptées ci-dessus, on peut écrire une formule exponentielle unique, avec les notations ci-après:

longueur de la séquence: $2.n = 4.m$;

nombre des contacts: $k_s = 2.k = 2.(m+\Delta) = n+2\Delta$;

distribution sur l'ensemble de toutes les séquences: $\exp(-2\Delta^2/m)$.

Il apparaît que, sur l'ensemble de toutes les séquences possibles, la différence $\Delta = (k_s - n) / 2 = k-m$, est distribuée normalement autour de zéro, avec pour variance m/4; l'écart type de k_s est donc $\surd(m) = \surd(n/2)$.

Du fait des fluctuations thermiques, cette distribution doit être corrigée en suivant BOLTZMANN et GIBBS: la probabilité relative d'un état sera donnée par l'exponentielle d'une expression où figure, l'énergie, W, de cet état; la température T; et une constante universelle, notée ici k_u, quotient de la constante R des gaz parfaits par le nombre d'AVOGADRO:

$$pr \approx \exp(-W/(k_u.T)) \quad .$$

On a supposé que l'énergie est proportionnelle au nombre des contacts; d'autant plus faible que ceux-ci sont moins nombreux; ce qui favorise la régularité des séquences. On peut écrire:

$$pr \approx \exp(-W/(k_u.T)) \approx \exp(-4ß.k) = \exp(-4ß.(m+\Delta)) \quad ;$$

où est introduit un coefficient ß dont il suffit de noter ici qu'il est inversement proportionnel à la température T.

Ainsi, la fréquence relative d'une valeur de Δ, fréquence qu'on a trouvée égale à $\exp(-2\Delta^2/m)$ sur l'ensemble de toutes les séquences, doit être pondérée par $\exp(-4ß.\Delta)$; de telle sorte qu'on a pour la suite des probabilités relatives, qu'il resterait à amener à avoir somme 1, en introduisant un coefficient convenable:

$$prob(\Delta) \approx \exp(-2(\Delta^2+2ßm\Delta)/m) \ ;$$

d'où il résulte que Δ+ß.m suit une loi normale centrée. La variance de Δ, indépendante de ß, est, comme avant pondération, égale à (m/4). La moyenne de Δ est -ßm: formule satisfaisante en ce que:

premièrement: la moyenne, (-ßm), de Δ, est négative, ce qui marque une tendance des séquences à la régularité; avec un nombre total de contacts, $k_s = 2.k = 2.(m+\Delta) = n+2\Delta$, inférieur à la moitié, n, du nombre des lettres;

deuxièmement: (-ß.m) est proportionnel à n, ce qui exprime que le taux de régularité est indépendant de la longueur (supposée grande) des séquences;

troisièmement: la régularité, étant liée à ß, tend à disparaître totalement quand la température s'élève: ce qui est un effet général des fluctuations thermiques.

Le résultat obtenu ici diffère grandement de la régularité parfaite à laquelle conduit le processus étudié aux §§2, 3. Car, afin d'appliquer nos approximations des calculs combinatoires, on a supposé une température élevée qui maintient proche de n le nombre k_s des contacts. La régularité parfaite, au contraire, ne peut régner qu'à très basse température.

On peut reprendre les calculs ci-dessus (en supposant comme ci-dessus que m = 2n) de telle sorte qu'ils s'appliquent au cas où la valeur moyenne de k n'est pas voisine de m; mais où la fraction (k/m) a une valeur quelconque, par exemple (1/10); ce qui correspond à une plus grande régularité moyenne des séquences.

Si l'on néglige des variations relatives de l'ordre de (1/n), le profil des probabilités de la suite des valeurs de k est celui de la suite des nombres $\pi(k)$, donnés par la formule:

$$\pi(k) \approx \exp(-4ß.k).\ C(n, k)^2 \ = \ (\exp(-2ß.k).\ C(n, k))^2 \quad .$$

Or la suite des nombres exp(-2ß.k). C(n, k) donne la loi d'une variable entière, f=k, qui est la somme de n=2m variables aléatoires indépendantes prenant chacune les valeurs 1 et 0 avec des probabilités dont le rapport est exp(-2ß); soit: prob(1) = exp(-2ß) / (1+exp(-2ß)) = exp(-ß) / (2.ch(ß)) . D'où, pour la loi de f :

$$\mathrm{moy}(f) = n \,.\, \exp(-\beta) / (2.\mathrm{ch}(\beta)) \quad ;$$

$$\mathrm{var}(f) = n \,.\, (2.\mathrm{ch}(\beta))^{-2} \quad ;$$

$$\exp(-2\beta.k).\ C(n, k) \approx q \,.\, \exp(-(k-\mathrm{moy}(f))^2 / (2.\mathrm{var}(f)) \quad ;$$

avec, dans la dernière formule, une constante q qu'il n'y a pas lieu de préciser.

On peut ainsi estimer les $\pi(k)$; et, par suite, la loi de k dans le processus thermodynamique de régularisation (avec, dans la formule, une autre constante Q, d'ailleurs facile à calculer en fonction de la variance de k):

$$\mathrm{prob}(k) \approx Q \,.\, \exp(-(k-\mathrm{moy}(f))^2 / (\mathrm{var}(f)) \quad ;$$

$$\mathrm{Esp}(k) = \mathrm{moy}(f) = m \,.\, \exp(-\beta) / \mathrm{ch}(\beta) \quad ;$$

$$\mathrm{Var}(k) = (1/2).\mathrm{var}(f) = m \,.\, (2.\mathrm{ch}(\beta))^{-2} \quad .$$

Si ß est petit, on retrouve les résultats précédemment obtenus pour une température élevée: Esp(k) = m - ß.m ; Var(k) = m/4 .

Quand la température, T, tend vers zéro, ß, qui est inversement proportionnel à T, tend vers l'infini. On peut négliger exp(-2ß) relativement à 1; d'où, pour Esp(k) et Var(k), les estimations:

$$\mathrm{Esp}(k) \approx 2m \,.\, \exp(-2\beta) \quad ;$$

$$\mathrm{Var}(k) \approx m \,.\, \exp(-2\beta) \quad ;$$

le logarithme de Esp(k)/n tend donc vers $(-\infty)$ comme la quantité (-1/T).

On peut reprendre ici les commentaires faits ci-dessus à propos de la loi de k_s , dans le cas particulier où ce nombre diffère peu de m=(n/2): il y a une tendance à la régularité, d'autant plus forte que la température est plus basse; et le rapport entre k_s et m est indépendant de la longueur de la séquence.

6 Une méthode de calcul de la distribution du nombre des contats applicable aux réseaux multidimensionnels

Les calculs du présent §, présentés ici en détail dans le cas d'une distribution linéaire, ne donnent pas tous les résultats obtenus au §5; mais ils présentent l'intérêt de se généraliser à des réseaux; ainsi qu'on l'exposera brièvement en conclusion.

Nous considérons l'ensemble de toutes les séquences que l'on peut former en permutant, de toutes les façons possibles, 2.n lettres, dont n sont des A et n des B. La séquence comprend donc 2n-1 paires successives (ou segments de longueur 2) avec contact (de la forme AA ou BB), ou sans contact (AB ou BA). La probabilité qu'une paire présente un contact est:

$$(n-1)/(2n-1) = (1/2)\ (1 + (1/(2n)))\ ;$$

en effet, considérons, e.g, les lettres de rang r et r+1: si, en bref, celle-là est un A, les probabilités que celle-ci soit un A ou un B sont, respectivement, proportionnelles à (n-1), nombre de A restant après choix du r-ème caractère, et à n, nombre de B restant.

Donc, sur l'ensemble des (2n-1) paires, l'espérance, M1, du nombre total, k_s, des contacts est: n-1. Ce total k_s peut être considéré comme une somme de (2n-1) termes aléatoires, prenant les seules valeurs 1 ou 0. En calculant les comoments de ces termes, pris deux à deux, et faisant la somme, on aura l'espérance mathématique, M2, de k_s^2, donc finalement, la variance de k_s.

De façon précise, il y a $(2n-1)^2$ couples ordonnés de termes; que l'on distinguera selon que les paires sont identiques, empiétantes, ou séparées.

identiques: 2n-1 couples: pour un tel couple, la covariance n'est autre que la probabilité cumulée des cas AA et BB, soit: (n-1)/(2n-1);

empiétantes: 4n-4 couples: le comoment est le cumul des probabilités des cas pour lesquels il y a deux contacts, soit: AAA et BBB; ou encore, la probabilité conditionnelle que, la première lettre étant A, la séquence soit AAA, d'où la formule:

$$((n-1)/(2n-1)).((n-2)/(2n-2)) = (n-2)/((2n-1).2)\ ;$$

séparées: le nombre de couples de ce type est:

$$(2n-1)^2 - ((2n-1)+(4n-4)) = 6 + 4.n^2 - 10.n = (2n-3)(2n-2)\ ;$$

le comoment peut se calculer en faisant le produit de la probabilité que l'une des paires offre un contact, e.g. AA, par la somme des probabilités conditionnelles que l'autre paire soit AA ou BB; soit pour le comoment:

$$((n-1)/(2n-1))\ .\ ((n)(n-1) + (n-2)(n-3))\ /\ ((2n-2)(2n-3))$$

$$= ((n)(n-1) + (n-2)(n-3))\ /\ ((2n-1).(2n-3).2)$$

$$= (2(n^2)+6-6n)\ /\ ((2n-1).(2n-3).2) = (n^2 + 3 -3n)/((2n-1).(2n-3))\ ;$$

On calcule maintenant l'espérance mathématique, M2, de k_s^2 en faisant la somme des comoments afférents aux trois classes de paires, identiques, séparées et empiétantes, d'où la formule:

$$M2 = (n-1) + ((2n-2)(n-2)/(2n-1)) + ((2n-2)(n^2 + 3 -3n)/(2n-1))\ ;$$

soit une expression dont le dénominateur est (2n-1) et le numérateur:

numérateur = ((n-1)(2n-1)) + ((2n-2)(n-2)) + ((2n-2)(n^2 + 3 -3n))

= (n-1) ((2n-1) + (2n-4) + 2(n^2 + 3 -3n))

= (n-1) (1 + $2n^2$ - 2n) / (2n - 1) ;

d'où, après calculs, pour la variance:

variance = M2 - $(M1)^2$ = (n-1)(n) / (2n -1) ≈ n/2 ;

c'est bien la valeur déjà trouvée, au §5.

Ce calcul suffit à établir la loi normale asymptotique de k_s, donc aussi la distribution corrigée, comme au début du §5, en suivant BOLTZMANN et GIBBS, pour tenir compte des fluctuations thermiques. Mais on sait que cette méthode ne s'applique que dans le cas d'une température T élevée. Aux basses températures, cas où l'ordre prédomine dans les séquences, la méthode suivie dans la deuxième partie du §5 utilise, de façon essentielle, les formules combinatoires plus précises, obtenues au §4 pour la loi du nombre des contacts; formules qui valent même quand k_s est très loin de M1 = n-1.

Mais le calcul du présent § offre l'intérêt de se généraliser à toute forme de réseau. Prenons l'exemple d'un carré (2.n) × (2.n) sur lequel on dispose $2.n^2$ lettres A et autant de lettres B. On peut noter: N = $2.n^2$.

Le nombre total des paires de lettres contiguës, horizontalement ou verticalement est: P = (4n).(2n - 1).

Pour chacune de ces paires, la probabilité qu'il y ait contact est:

(N-1)/(2N-1) .

Comme dans le cas linéaire, on peut distinguer des couples de paires identiques, empiétantes ou séparées. Le comoment afférent à une paire de chaque classe est donné par les mêmes formules que précédemment, pourvu que l'on y remplace n par N. Seul le dénombrement des classes de paires est à reprendre; plus précisément, celui des couples de paires empiétantes; le dénombrement des couples de paires séparées s'obtenant ensuite par différence.

À titre d'exercice combinatoire, considérons les couples empiétants. On en distingue quatre types, qu'on peut, en bref, noter HH, VV, HV et VH; selon qu'il s'agit de deux paires horizontales, de deux verticales, ou de deux paires différant quant à l'orientation. On montre, que, comme précédemment, il y a, sur chaque ligne, (4n-4) couples HH; soit, sur tout le carré, 2n.(4n-4) couples HH. Le nombre des couples VV est le même. Pour les couples HV, il faut distinguer selon que la paire H est sur les côtés inférieur ou supérieur du carré, ou sur une ligne intermédiaire. Dans ce dernier cas, un paire H rentre dans 4 couples HV; mais, dans le premier cas, dans deux seulement. D'où pour le nombre des couples HV:

$$4.(2n-2).(2n-1) + 2.2.(2n-1) = 4.(2n-1)^2 \quad ;$$

et le nombre des couples VH est le même. Soit, au total:

$$8.(2n-1)^2 + 16.n.(n-1) = 8.(1 + 6.n^2 - 6.n) \quad ,$$

pour le nombre des couples ordonnés de paires empiétantes.

7 Appendice: application de la formule de STIRLING à un calcul d'approximation

Au §5, on a posé la formule d'approximation:

$$C(2m, m+\Delta) \approx C(2m, m) . \exp(-\Delta^2/m) \quad ;$$

utilisée pour des valeurs de Δ de l'ordre de $\sqrt{(m)}$; e.g.: $|\Delta| < 5.\sqrt{(m)}$.

On vérifie ici que la formule classique de STIRLING, pour n!, permet d'améliorer la majoration trouvée au §5. On a:

$$n! = (n/e)^n . \sqrt{(2\pi.n)} . (1 + (1/(12n)) + \ldots) \quad ;$$

d'où pour le quotient considéré:

$$C(2m, m) / C(2m, m+\Delta) = (m+\Delta)! \, (m-\Delta)! \;/\; (m! \, m!)$$

$$= ((m+\Delta)/m)^{m+\Delta} . ((m-\Delta)/m)^{m-\Delta} . \sqrt{(((m+\Delta)/m)((m-\Delta)/m))} . (1+ O(1/m))$$

$$= ((1 - (\Delta/m)^2)^{m+(1/2)} . ((1+(\Delta/m))/(1-(\Delta/m)))^{\Delta} . (1+ O(1/m)) \quad ;$$

Notre but est de montrer que, sous l'hypothèse faite quant à la valeur de Δ, le logarithme du quotient ci-dessus ne diffère de Δ^2/m que par un terme de l'ordre de (1/m). On a:

$$\log (1 - (\Delta/m)^2) = - ((\Delta/m)^2 + (1/2)(\Delta/m)^4 + (1/3)(\Delta/m)^6 + \ldots) \quad ;$$

d'où, pour le premier facteur du quotient considéré:

$$\log((1 - (\Delta/m)^2)^{m+(1/2)} = (-\Delta^2/m) + O(1/m) \quad ;$$

formule où on a majoré Δ^2 par O(m).

Pour le second facteur on prend de même:

$$\log((1+(\Delta/m))/(1-(\Delta/m))) = 2 ((\Delta/m) + (1/3)(\Delta/m)^3 + \ldots) \quad ;$$

$$\log(((1+(\Delta/m))/(1-(\Delta/m)))^{\Delta}) = 2(\Delta^2/m) + (2/3)(\Delta^4/m^3) + \ldots \quad ;$$

dans cette formule, le second terme est de la forme O(1/m) du fait de la majoration posée pour Δ; d'où pour logarithme de l'ensemble des termes du quotient considéré:

$$(-\Delta^2/m) + 2(\Delta^2/m) + O(1/m) = (\Delta^2/m) + O(1/m) ;$$

comme on l'avait annoncé.

Remerciements

L'auteur du présent article est heureux de remercier Mr. R. ROUSSEAU, Vice-Recteur de l'Université Catholique d'Angers, qui a revu les calculs et les démonstrations et inspiré de nombreuses corrections.