Revue de statistique appliquée

P. Dagnelie
P. Florins

Étude du caractère aléatoire de la répartition de points dans des espaces à deux et à trois dimensions

Revue de statistique appliquée, tome 39, n° 1 (1991), p. 11-20
<http://www.numdam.org/item?id=RSA_1991__39_1_11_0>

© Société française de statistique, 1991, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (http://www.sfds.asso.fr/publicat/rsa.htm) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/legal.php). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

Numdam
Article numérisé dans le cadre du programme Numérisation de documents anciens mathématiques
http://www.numdam.org/

Rev. Statistique Appliquée, 1991, **XXXIX** (1), 11-20

ÉTUDE DU CARACTÈRE ALÉATOIRE DE LA RÉPARTITION DE POINTS DANS DES ESPACES A DEUX ET A TROIS DIMENSIONS

P. DAGNELIE, P. FLORINS

Faculté des Sciences Agronomiques, B-5800 Gembloux (Belgique)

RÉSUMÉ

Le problème de l'étude du caractère aléatoire de la répartition de points ou d'individus dans des espaces à deux dimensions peut se présenter également à trois dimensions.

Pour résoudre ce problème, les principales méthodes bidimensionnelles sont transposées à trois dimensions et une méthode nouvelle est introduite (paragraphes 2 et 3). Ces diverses méthodes sont comparées par simulation, en ce qui concerne tant leurs niveaux de signification que leurs puissances (paragraphe 4).

Une application pratique est également présentée (paragraphe 5).

Mots-clés : *tests du caractère aléatoire à deux et trois dimensions, répartition de points à deux et trois dimensions.*

ABSTRACT

The problem of testing the randomness of the distribution of points or individuals in two-dimensional spaces may also occur in three-dimensional spaces.

To solve this problem, the most important two-dimensional methods are extended to three-dimensional spaces and a new method is developed (sections 2 and 3). All these methods are compared by simulation, concerning both their significance levels and powers (section 4).

A practical application is also given (section 5).

Key-words : *two- and three-dimensional tests of randomness, two- and three-dimensional distribution of points.*

1. Introduction

L'étude du caractère aléatoire de la répartition de points ou d'individus (plantes, organismes, etc.) dans des espaces à deux dimensions est un problème classique, en écologie végétale notamment [GOUNOT, 1969; GREIG-SMITH, 1967; PIELOU, 1977; etc.]. Ce problème peut être traité selon diverses approches, comme le montre, entre autres, RIPLEY [1977].

Un problème semblable, beaucoup moins fréquent en pratique, peut se poser également à trois dimensions.

Au cours du paragraphe 2 de cette publication, nous présenterons brièvement les principales méthodes relatives au cas bidimensionnel. Nous envisagerons ensuite le cas tridimensionnel, en étendant les méthodes bidimensionnelles et en introduisant une méthode nouvelle (paragraphe 3), en exposant les principaux résultats d'une étude comparative réalisée par simulation (paragraphe 4) et en donnant un exemple concret d'application (paragraphe 5). Nous terminerons enfin par quelques conclusions (paragraphe 6).

2. Quelques tests du caractère aléatoire de la répartition de points dans des espaces à deux dimensions

Les tests classiques du caractère aléatoire de la répartition de points ou d'individus dans un espace à deux dimensions sont basés soit sur des dénombrements, soit sur des mesures de distances. Dans la première catégorie de méthodes, figurent le test d'ajustement à une distribution de POISSON et le test de l'indice de dispersion. Dans la deuxième catégorie, figurent notamment les méthodes basées sur l'étude des distances entre individus voisins (méthode des plus proches voisins) ou entre des points choisis au hasard et les individus les plus proches (méthode des plus proches individus).

2.1. Deux tests basés sur des dénombrements

Les méthodes basées sur des dénombrements supposent qu'on a observé, dans le domaine considéré, un certain nombre d'unités de dimension constante (par exemple des parcelles circulaires ou des parcelles carrées, appelées "quadrats"), réparties de manière complètement aléatoire, et qu'on a compté le nombre de points ou d'individus présents dans chacune de ces unités.

Dans l'hypothèse d'une répartition complètement aléatoire des individus, la distribution des nombres d'individus par unité est une distribution de POISSON et l'hypothèse émise peut être vérifiée par le *test χ^2 d'ajustement* [DAGNELIE, 1985-1986].

Dans les mêmes conditions, et en fonction des propriétés des distributions de POISSON, la moyenne et la variance du nombre d'individus par unité sont deux valeurs théoriquement égales. L'égalité de ces deux paramètres peut être vérifiée à l'aide du *test de l'indice de dispersion*, cet indice étant égal au rapport entre la somme des carrés des écarts observée et la moyenne observée. Pour n unités et lorsque la répartition des individus est complètement aléatoire, la distribution de l'indice de dispersion est une distribution χ^2 à $n-1$ degrés de liberté [DAGNELIE, 1985-1986].

On notera que, si le test d'ajustement est unilatéral (rejet de l'hypothèse nulle lorsque la valeur observée de la variable χ^2 est trop élevée), le test de l'indice de dispersion est au contraire bilatéral (rejet de l'hypothèse nulle lorsque la valeur observée de la variable χ^2 est soit trop faible, soit trop élevée).

2.2. Trois tests basés sur des mesures de distances

La *méthode des plus proches voisins* consiste à choisir au hasard un certain nombre d'individus et à mesurer, pour chacun de ceux-ci, la distance au plus proche voisin [CLARK et EVANS, 1954]. Si on désigne par λ le nombre moyen d'individus par unité de surface, la valeur attendue et la variance de la distance X au plus proche voisin sont :

$$m_X = 1/(2\sqrt{\lambda}) \quad \text{et} \quad \sigma_X^2 = (4-\pi)/(4\pi\lambda) = 0,068310/\lambda.$$

Pour des effectifs suffisamment élevés, on peut en déduire un test qui permet de vérifier la conformité de la moyenne observée des distances à la moyenne attendue.

L'emploi de ce test soulève cependant diverses difficultés, relatives au choix aléatoire d'individus, à l'estimation de la densité λ et, lorsque le nombre de distances mesurées n'est pas suffisamment élevé, au caractère non normal de la distribution de la distance moyenne.

La *méthode des plus proches individus* consiste à choisir au contraire, de manière complètement aléatoire également, un certain nombre de points du domaine considéré et à déterminer, pour chacun de ces points, la distance à l'individu le plus proche [MOORE, 1954; SKELLAM, 1952]. Dans l'hypothèse d'une répartition complètement aléatoire des individus, la quantité :

$$2\pi\lambda \sum_{i=1}^{n} Y_i^2,$$

c'est-à-dire, à une constante près, la somme des carrés des distances Y_i, possède une distribution χ^2 à $2n$ degrés de liberté. On peut en déduire facilement un test de l'hypothèse du caractère aléatoire de la répartition des individus.

L'emploi de cette méthode élimine le problème du choix aléatoire des individus, mais pas celui de l'estimation du paramètre λ.

Cette dernière difficulté, commune aux deux méthodes, peut être écartée en combinant les deux méthodes et en faisant alors appel aux distributions F de FISHER-SNEDECOR [HOPKINS, 1954]. On peut en effet démontrer que, pour n_1 distances X_i au plus proche voisin et n_2 distances Y_i au plus proche individu, obtenues indépendamment les unes des autres, la quantité :

$$\frac{n_2 \sum_{i=1}^{n_1} X_i^2}{n_1 \sum_{i=1}^{n_2} Y_i^2}$$

possède une distribution F de FISHER-SNEDECOR à $2n_1$ et $2n_2$ degrés de liberté.

Cette façon de procéder n'élimine cependant pas le problème du choix aléatoire des individus, propre au test de CLARK et EVANS.

3. Extension à trois dimensions des différents tests

3.1. Tests basés sur des dénombrements

La transposition à trois dimensions des deux méthodes présentées au paragraphe 2.1 (*test d'ajustement* à une distribution de POISSON et *méthode de l'indice de dispersion*) ne soulève aucune difficulté particulière. La seule différence est que les unités dans lesquelles doivent être comptés les points ou les individus considérés sont ici des éléments de volumes, sphériques ou cubiques par exemple, au lieu d'être des éléments de surfaces, le plus souvent circulaires ou carrés.

3.2. Tests basés sur des mesures de distances

En ce qui concerne la *méthode des plus proches voisins* (méthode de CLARK et EVANS), on peut démontrer que la valeur attendue et la variance de la distance séparant un individu choisi au hasard de son plus proche voisin est, à trois dimensions :

$$m_X = (4\pi\lambda/3)^{-1/3}\Gamma(4/3) = 0,55396\lambda^{-1/3}$$

et $$\sigma_X^2 = (4\pi\lambda/3)^{-2/3}\{\Gamma(5/3) - [\Gamma(4/3)]^2\} = 0,040536\lambda^{-2/3}.$$

Comme à deux dimensions, un test du caractère aléatoire de la répartition des individus peut être réalisé par approximation normale, quand le nombre de distances observées est suffisamment élevé.

Quant à l'extension à trois dimensions de la *méthode des plus proches individus* (méthode de SKELLAM et MOORE), elle conduit au résultat suivant. Si Y_i sont les distances des points choisis au hasard aux individus les plus proches, la distribution de :

$$\frac{8}{3}\pi\lambda \sum_{i=1}^{n} Y_i^3,$$

est une distribution χ^2 à $2n$ degrés de liberté. Il en résulte un test simple de l'hypothèse d'une répartition complètement aléatoire des individus considérés, pour autant qu'on connaisse ou qu'on possède une bonne estimation de leur densité λ (nombre d'individus par unité de volume).

La méthode de HOPKINS peut aussi être étendue à trois dimensions, d'une manière tout à fait semblable, le quotient des sommes des carrés des distances devant être remplacé par le quotient des sommes des cubes des distances, d'une part entre les individus les plus proches (méthode de CLARK et EVANS) et d'autre part entre les points choisis au hasard et les individus les plus proches (méthode de SKELLAM et MOORE).

3.3. Une méthode nouvelle

En vue de traiter l'exemple dont il sera question au paragraphe 5, nous avons été amenés à proposer une méthode nouvelle, basée sur l'étude des distances des différents individus à leur centre de gravité. Dans le cas qui est considéré ci-dessous, la *moyenne quadratique des distances au centre de gravité* apparaît en effet, tout naturellement, comme un bon indice de concentration ou de dispersion éventuelle des individus.

Pour un ensemble de n points de coordonnées (X_i, Y_i, Z_i), cette moyenne quadratique est :

$$\overline{D}_q = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[(X_i-\overline{X})^2+(Y_i-\overline{Y})^2+(Z_i-\overline{Z})^2]} = \sqrt{S_X^2+S_Y^2+S_z^2},$$

$\overline{X}$, $\overline{Y}$ et $\overline{Z}$, et S_X^2, S_Y^2 et S_Z^2 désignant les moyennes et les variances des X_i, des Y_i et des Z_i.

Dans le cas d'un volume de forme parallélépipédique, dont les trois dimensions sont w_X, w_Y et w_Z, on peut démontrer que l'espérance mathématique et la variance de $\overline{D}_q$ sont approximativement :

$$\mathrm{E}(\overline{D}_q) = \frac{n-1,05}{n-0,5}\sqrt{(w_X^2+w_Y^2+w_Z^2)/12}$$

et

$$\sigma^2(\overline{D}_q) = \frac{n+1,5}{60n^2}\,\frac{w_X^4+w_Y^4+w_Z^4}{w_X^2+w_Y^2+w_Z^2},$$

ou encore :

$$\mathrm{E}(\overline{D}_q) = \frac{n-1,05}{n-0,5}\,\frac{w}{2} \quad \text{et} \quad \sigma^2(\overline{D}_q) = \frac{n+1,5}{60n^2}w^2,$$

lorsque, en particulier :

$$w_X = w_Y = w_Z = w.$$

Ces relations ont été établies en considérant que $\overline{D}_q^2$ est la somme des variances de trois échantillons aléatoires et indépendants, extraits de populations possédant des distributions uniformes d'amplitudes respectives w_X, w_Y et w_Z. Les termes $-1{,}05$ et $-0{,}5$ ont été déterminés de manière empirique, dans le but d'assurer une bonne adéquation des formules pour des effectifs aussi réduits que possible, et cela quelles que soient les valeurs relatives des amplitudes. A titre d'indication, l'erreur maximum relative à $\mathrm{E}(\overline{D}_q)$ est de l'ordre de 1% pour 5 observations ($n = 5$) et 0,5% pour 10 observations ($n = 10$), l'approximation étant donc très satisfaisante, même pour des effectifs très réduits.

En outre, une étude réalisée par simulation montre que, pour des effectifs égaux ou supérieurs à 10, la distribution de $\overline{D}_q$ peut être considérée sans aucun inconvénient comme approximativement normale.

4. Comparaison des différentes méthodes relatives aux espaces à trois dimensions

4.1. Principes

Les diverses méthodes décrites au paragraphe 3 ont été comparées par simulation, pour différents types de répartitions théoriques.

La première répartition étudiée est une répartition complètement aléatoire de points dans l'ensemble du domaine considéré. L'étude de cette répartition particulière a comme objectif de vérifier si les niveaux de signification observés pour les différentes méthodes (erreurs de première espèce) sont bien conformes aux valeurs attendues.

L'étude des autres répartitions, de type régulier et de type agrégatif, a par contre pour objectif de comparer la puissance des différentes méthodes, face à diverses alternatives.

Des répartitions de type régulier ont été simulées en partant d'un ensemble initial de points disposés selon un maillage tout à fait systématique, et en imposant ensuite, aux différents points, des déplacements aléatoires plus ou moins importants par rapport à ce maillage.

Quant aux répartitions de type agrégatif, elles ont été réalisées selon trois principes différents.

Le premier type agrégatif est constitué de groupes de points d'effectifs égaux répartis de manière aléatoire dans le domaine considéré. Le deuxième type de répartition agrégative, qui s'apparente au modèle de NEYMAN type A [NEYMAN, 1939], consiste en une répartition aléatoire de groupes de points d'effectifs inégaux distribués selon une loi de POISSON. Et le troisième type de répartition agrégative correspond au modèle POISSON de POISSON présenté par THOMAS [1949] et adapté par DIGGLE *et al.* [1976].

4.2. Résultats

En ce qui concerne le risque de première espèce, les résultats obtenus se sont révélés tout à fait satisfaisants pour les différentes méthodes, au niveau de probabilité habituel 0,05. Pour un total de près de 2.000 simulations, réalisées dans différentes situations (différentes densités des populations notamment), aucune différence significative n'est en effet apparue entre les pourcentages observés de rejet de l'hypothèse du caractère aléatoire et la valeur attendue de 5%.

En ce qui concerne les répartitions non aléatoires, de type régulier et de type agrégatif, plus de 22.000 simulations ont été réalisées. Globalement, les conclusions obtenues peuvent être résumées de la manière suivante.

Des deux méthodes basées sur des dénombrements (paragraphe 3.1), la méthode de l'indice de dispersion s'avère toujours plus puissante que le test d'ajustement à une distribution de POISSON et, parmi les trois méthodes basées sur des mesures de distances (paragraphe 3.2), les méthodes de SKELLAM et MOORE et de HOPKINS s'avèrent pratiquement équivalentes l'une à l'autre et toujours plus puissantes que la méthode de CLARK et EVANS. En outre, les méthodes de SKELLAM et MOORE et de HOPKINS sont aussi plus puissantes que la méthode de l'indice de dispersion.

Quant à la méthode proposée au paragraphe 3.3, elle possède une puissance équivalente à celles des méthodes de SKELLAM et MOORE et de HOPKINS, en présence de répartitions de type régulier, et elle se situe entre les méthodes, d'une part, de SKELLAM et MOORE et de HOPKINS, et d'autre part, de CLARK et EVANS et de l'indice de dispersion, en présence de répartitions de type agrégatif.

La possibilité d'utiliser simultanément deux méthodes différentes a également été envisagée, mais les résultats obtenus dans ce domaine ne se sont pas révélés fort intéressants. La principale conclusion de cette étude complémentaire est en effet relative à l'intérêt que peut présenter la combinaison des méthodes de CLARK et EVANS et de SKELLAM et MOORE. Toutefois, l'utilisation conjointe de ces deux méthodes ne fournit pas de conclusions plus favorables que la méthode de HOPKINS, qui est déjà, elle-même, une combinaison des deux méthodes considérées.

5. Exemple d'application

5.1. Matériel et méthodes

L'étude dont la présente publication reprend certains résultats partiels a été entreprise initialement en vue de résoudre notamment le problème suivant.

Des chercheurs s'intéressant au comportement de poissons cavernicoles, vivant dans une obscurité totale, se demandaient si la répartition de ces poissons correspond à un modèle complètement aléatoire ou, au contraire, à un modèle de type régulier ou agrégatif, et cela par comparaison avec des poissons épigés, dont le comportement peut être étudié en présence ou en absence de lumière [JANKOWSKA et THINES, 1982][1].

Dans ce but, ces chercheurs ont réalisé un montage qui assure, le moment venu, un éclairement latéral de l'aquarium dans lequel les poissons sont placés et qui permet de procéder à intervalles réguliers à une photographie verticale et une photographie transversale de l'aquarium, ces photographies permettant elles-mêmes de déterminer dans chaque cas les coordonnées des différents poissons.

---

(1) Les données reprises ici sont utilisées avec l'aimable autorisation de ces auteurs, que nous remercions.

Les nombres de poissons cavernicoles disponibles étant très limités (de 5 à 10 individus pour chacune des espèces considérées), les méthodes basées sur des comptages (paragraphe 3.1) étaient inapplicables et les méthodes basées sur des mesures de distances (paragraphe 3.2) étaient elles-mêmes difficilement utilisables. Seule la méthode présentée au paragraphe 3.3 est en conséquence considérée dans la suite à titre d'exemple.

5.2. Résultats

Le tableau qui figure ci-dessous présente les résultats obtenus pour quatre genre ou espèces cavernicoles (*Astyanax* gen., *Barbopsis devecchii, Caecobarbus geertsi* et *Uegitglanis zammaranoi*), par comparaison avec deux espèces épigées (*Barbus conchonius* et *Hyphessobricon scholzei*), observées d'une part dans l'obscurité et d'autre part en lumière permanente.

Pour chacun des cas considérés, ce tableau donne le nombre d'individus observés n, la distance moyenne quadratique observée $\bar{d}_q$, la valeur attendue de la distance moyenne quadratique $\mathrm{E}(\overline{D}_q)$, la différence entre ces valeurs, observée et attendue, l'erreur-standard de la distance moyenne $\sigma(\overline{D}_q)$, et le quotient de la différence entre les distances, observée et attendue, et l'erreur-standard. Ces différentes informations sont relatives à un volume de 98 × 38 × 38 cm.

Poissons	n	$\bar{d}_q$	$\mathrm{E}(\overline{D}_q)$	$\bar{d}_q - \mathrm{E}(\overline{D}_q)$	$\sigma(\overline{D}_q)$	$\frac{\bar{d}_q - \mathrm{E}(\overline{D}_q)}{\sigma(\overline{D}_q)}$
cavernicoles						
Astyanax gen.	10	26,2	30,4	−4,2	3,8	−1,1
B. devecchii	10	34,1	30,4	3,7	3,8	1,0
C. geertsi	5	22,4	28,3	−5,9	5,8	−1,0
U. zammaranoi	9	28,3	30,2	−1,9	4,1	−0,5
épigés						
dans l'obscurité						
B. conchonius	10	22,7	30,4	−7,7	3,8	−2,0
H. scholzei	10	17,5	30,4	−12,9	3,8	−3,4
à la lumière						
B. conchonius	10	23,3	30,4	−7,1	3,8	−1,9
H. scholzei	10	17,6	30,4	−12,8	3,8	−3,4

Le volume réel de l'aquarium considéré était en fait de 100 × 40 × 40 cm, mais en fonction de la taille moyenne des poissons, ce volume a été conventionnellement réduit d'un centimètre sur toute la surface extérieure, de manière à tenir compte de ce que les poissons (ou, plus exactement, les points de repère de leurs corps qui servaient à les localiser) ne pouvaient pas, en réalité, être observés jusqu'aux limites extrêmes de l'aquarium.

Ce tableau montre que le comportement des poissons cavernicoles est nettement plus aléatoire que le comportement des poissons épigés, beaucoup plus grégaires, et aussi que le comportement de ces derniers n'est pas influencé de façon sensible par l'existence ou non d'un éclairement permanent. Pour les poissons cavernicoles, les écarts entre les distances moyennes observées et attendues sont en effet, toutes, au maximum de l'ordre de grandeur de l'erreur-standard, tandis que ces écarts sont de l'ordre de deux fois l'erreur-standard pour *B. conchonius* et trois fois l'erreur-standard pour *H. scholzei*, et cela indépendamment des conditions d'éclairement.

6. Conclusions

Le problème de l'étude du caractère aléatoire de la répartition de points ou d'individus dans des espaces à deux dimensions peut être étendu à des espaces à trois dimensions.

Les méthodes bidimensionnelles classiques (test d'ajustement à une distribution de POISSON, test de l'indice de dispersion, méthode de CLARK et EVANS, méthode de SKELLAM et MOORE, et méthode de HOPKINS) peuvent être transposées à trois dimensions. Les méthodes de SKELLAM et MOORE et de HOPKINS, et dans une moindre mesure la méthode de l'indice de dispersion, s'avèrent les plus puissantes.

En outre, une méthode nouvelle, basée sur l'étude de la moyenne quadratique des distances des individus observés à leur centre de gravité, a été introduite. Cette méthode est particulièrement bien adaptée pour traiter le cas d'un nombre réduit d'individus observés dans une enceinte de forme parallélépipédique.

7. Bibliographie

CLARK P.J. et EVANS F.C. (1954). Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology* 35, 445-453.

DAGNELIE P. (1985-1986). *Théorie et méthodes statistiques* (2 vol.). Gembloux, Presses agronomiques, 378 + 463 p.

DIGGLE P.J., BESAG J. et GLEAVES J.T. (1976). Statistical analysis of spatial point pattern by means of distance methods. *Biometrics* 32, 659-667.

GOUNOT M. (1969). *Méthodes d'étude quantitative de la végétation.* Paris, Masson, 314 p.

GREIG-SMITH P. (1967). *Quantitative plant ecology.* London, Butterworths, 256 p.

HOPKINS B. (1954). A new method of determining the type of distribution of plant individuals. *Ann. Bot.* 18, 213-226.

JANKOWSKA M. et THINES G. (1982). Etude comparative de la densité de groupes de poissons cavernicoles et épigés (*Characidae, Cyprinidae, Clariidae*). *Behav. Processes* 7, 281-294.

MOORE P.G. (1954). Spacing in plant populations. *Ecology* 35, 222-227.

NEYMAN J. (1939). On a new class of contagious distributions, applicable in entomology and bacteriology. *Ann. Math. Stat.* 10, 35-57.

PIELOU E.C. (1977). *Mathematical ecology*. New York, Wiley, 384 p.

RIPLEY B.D. (1977). Modelling spatial patterns (*with discussion*). *J.R. Stat. Soc. Ser.* B, 179-212.

SKELLAM J.G. (1952). Studies in statistical ecology : spatial pattern. *Biometrika* 39, 346-362.

THOMAS M. (1949). A generalizing of POISSON's binomial limit for use in ecology. *Biometrika* 36, 18-25.