



ACADEMIC
PRESS

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Theoretical Population Biology 64 (2003) 81–87

**Theoretical
Population
Biology**

<http://www.elsevier.com/locate/ytptbi>

Covariances among join-count spatial autocorrelation measures

Bryan K. Epperson*

Department of Forestry, 126 Natural Resources Building, Michigan State University, East Lansing, MI 48824, USA

Received 30 April 2002

Abstract

Spatial distributions of biological variables are often well-characterized with pairwise measures of spatial autocorrelation. In this article, the probability theory for products and covariances of join-count spatial autocorrelation measures are developed for spatial distributions of multiple nominal (e.g. species or genotypes) types. This more fully describes the joint distributions of pairwise measures in spatial distributions of multiple (i.e. more than two) types. An example is given on how the covariances can be used for finding standard errors of weighted averages of join-counts in spatial autocorrelation analysis of more than two types, as is typical for genetic data for multiallelic loci.

© 2003 Elsevier Science (USA). All rights reserved.

Keywords: Join counts; Spatial autocorrelation; Spatial distribution; Nominal data; Genetics

1. Introduction

Spatial distributions of variation over scales ranging from landscape to geographic are of considerable interest in many fields, including ecology and genetics. Spatial patterns may be used to make inferences about generative processes, and spatial structure can affect the operation of processes. Much of the information content of spatial patterns is captured by pairwise measures, i.e. measures based on observations at paired locations, and most statistics used to analyse spatial patterns also fundamentally are based on pairs (Cliff and Ord, 1981). For example, most measures of ecological diversity are based on $1/\sum p_i^2$, where p_i is the probability or frequency of species i , or other functions of pairs of observations taken either with or without replacement (MacArthur, 1972). In population genetics, commonly used measures such as effective number of alleles ($1/\sum p_i^2$, where p_i is the probability or frequency of allele i), and genetic diversity or expected heterozygosity under Hardy-Weinberg proportions ($1-\sum p_i^2$), as well as measures of genetic distance (e.g. Nei, 1973) or genetic similarity (e.g. Moran's I -statistics), are also based on pairwise comparisons. How such measures are distributed over

space is part and parcel with the spatial distribution of variants.

In most spatial analyses pairs of locations are classified into groups that usually reflect some aspect of spatial proximity, often each group represents a range of Euclidean distances separating pairs of locations. A group of pairs defines “joins,” e.g. a join for two locations both having type i and separated by a distance that falls within a certain distance class. By defining joins, whether spatially or otherwise, constraints are placed on higher order combinations, and joins of different types are generally not independent. While the case where there are only two types in the spatial distribution has been fully characterized, in cases of three or more types there are previously uncharacterized dependencies among the different pairs of types. Because most statistics are based on pairwise comparisons of multiple types, there exist dependencies or correlations that make the sampling distributions of such statistics difficult to derive.

For example, there is a strong need for characterizing these correlations or covariances for studies of spatial patterns of genetic variation. Pairwise spatial autocorrelation measures used in genetics (Epperson, 1993) are directly related to genetical theory models (e.g. Malécot, 1948, 1972), which are also primarily based on pairwise measures of association. Experimental genetic distributions almost always have more than two types, and

*Fax: +517-432-1143.

E-mail address: epperson@pilot.msu.edu.

typically counts of joins are averaged, but covariance among join-counts has prevented summary statistical analysis, for example the finding of standard deviations of averages. A popular statistic used for analyzing spatial distributions of gene variation within populations is based on Moran's autocorrelation measure, I , for diploid genotypes converted into frequencies of an allele. This statistic is a weighted average of measures based on pairs of genotypes (Epperson, 1995). The values of I for different alleles are not independent, which has prevented characterization of the standard error of average values over alleles and loci. However, averaging is generally required in order to make precise inferences. By characterizing the covariances for different pairs of genotypes, it is possible to find the covariances of I statistics for different alleles and hence the standard error of averages. The covariances could be used in many other studies of the spatial distribution of nominal types. Moreover, the results are not necessarily limited to spatial distributions, and could be used in other contexts where sets of pairs are studied. Nonetheless, for sake of brevity and concreteness, here we will consider spatial distributions, and use notation that follows the spatial autocorrelation literature.

Although spatial autocorrelation measures of biological types, e.g. species or genotypes, were introduced into biology by Sokal and Oden (1978), the probability distributions of nominal types in space have been only partially characterized. Moran (1948) found the distribution for two types, say black (B) and white (W), under Mahalanobis' definition of "free sampling." Under free sampling, the type at each location is determined by the probability p_B or p_W , independently of other locations. Free sampling specifies one form of the null hypothesis that types are randomly distributed in space, in the distribution of spatial autocorrelation statistics known as join-count statistics. Join-counts are pairwise measures of spatial association, which seemingly utilize much of the information contained in spatial patterns. Krishna-Iyer (1950) found the probability distributions for multiple types (i.e. more than two types) under both free and "non-free sampling," but only for cases where the types are distributed among locations on regularly spaced lattices. Non-free sampling treats the numbers of types as being fixed, and then join-counts are based on sampling pairs without replacement. Non-free sampling provides a more useful form of the null hypothesis for join-counts and related spatial autocorrelation statistics. Cliff and Ord (1973) generalized join-counts for arbitrary spatial distributions of locations (i.e. not confined to lattices), but only partially described the joint distribution of various join-counts, and mostly for the two-type case. They found the expected values and variances of join-counts, but not most of the covariances or expected values of cross-products of join-counts, under the two different

probability models (free and non-free sampling) that specify the two alternative forms of null hypotheses for spatial autocorrelation statistics. This is adequate only when the distribution of each join-count is studied separately. There are many other applied problems in which it is necessary to characterize the covariances among different join-counts, and this is the primary objective of this article.

In a general spatial setting of the developments in this article there may be r different nominal types in a spatial distribution. At each different location there is a single stochastic variable that represents the type that occurs there. Each location has one and only one type. The probability that a location has type b is p_b . In spatial autocorrelation analysis all pairs of locations are classified in some way, usually in terms of a range of Euclidean distances between pairs of locations, and usually mutually exclusive and exhaustive (distance) classes are formed. The objective here is to specify the probability distribution of the numbers of pairs of types found in any particular (distance) class k .

2. Development of the means and variances of join-counts using new indicator variables

Let x_{ib} be an indicator variable such that $x_{ib} = 1$ if type b is found at location i , zero if not. This is a different formation of indicator variables than that used by Cliff and Ord (1973, 1981). Note that these indicator variables are not independent. Let $n_{bc}(k)$ denote the number of joins (pairs of locations) for types b and c for distance class k (i.e. the number of pairs of locations which have type b at one location and c at the other and are separated by distances that fall within distance class k). Let $\delta_{ij}(k)$ be an indicator variable such that when the pair of locations i and j is separated by a distance in class k , it equals one, zero otherwise. Then:

$$n_{bb}(k) = \frac{1}{2} \sum_{(2)} \delta_{ij}(k) x_{ib} x_{jb},$$

$$n_{bc}(k) = \sum_{(2)} \delta_{ij}(k) x_{ib} x_{jc} \quad (1)$$

for $b \neq c$. The factor $\frac{1}{2}$ for the join-counts $n_{bb}(k)$ corrects for double counting. The double summation is twice over all pairs of locations, excluding where the same location is indicated twice, without respect to ordering of locations, i.e.

$$\sum_{(2)} \equiv \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n .$$

By the definition of a join $\delta_{ij}(k) = \delta_{ji}(k)$.

There are two different ways to configure the probability distribution of $n_{bc}(k)$ under the null hypoth-

esis that there is no spatial structure, i.e. that the pairs of types are independent of their spatial locations. The first way is appropriate if the type probabilities, p_b , are posited and if the types at two locations represent two independent draws with replacement, as in an infinite urn model. This case corresponds to Mahalanobis' definition of "free sampling," and it corresponds to the "free-sampling" null hypothesis in the spatial autocorrelation literature (Cliff and Ord, 1981). The expected values $\mu_{bc}(k)$ and variances $\sigma_{bc}^2(k)$ under "free sampling" for joins between two locations with the same type b are

$$\mu_{bb}(k) = \frac{S_0(k)}{2} p_b^2 \tag{2}$$

$$\sigma_{bb}^2(k) = \frac{1}{4} [S_1(k) p_b^2 + (S_2(k) - 2S_1(k)) p_b^3 + (S_1(k) - S_2(k)) p_b^4], \tag{3}$$

where

$$S_0(k) = \sum_{(2)} \delta_{ij}(k) \tag{4}$$

or twice the number of joins total for distance class k ,

$$S_1(k) = \frac{1}{2} \sum_{(2)} (\delta_{ij}(k) + \delta_{ji}(k))^2 \tag{5}$$

which is four times the number of joins and

$$S_2(k) = \sum_{i=1}^n \left(\sum_{j=1}^n \delta_{ij}(k) + \sum_{j=1}^n \delta_{ji}(k) \right)^2 \tag{6}$$

(Cliff and Ord, 1981). For joins between different types b and c ($b \neq c$):

$$\mu_{bc}(k) = S_0(k) p_b p_c, \tag{7}$$

$$\sigma_{bc}^2(k) = \frac{1}{4} [2S_1(k) p_b p_c + (S_2(k) - 2S_1(k)) \times p_b p_c (p_b + p_c) + 4(S_1(k) - S_2(k)) p_b^2 p_c^2]. \tag{8}$$

Note that weighted join-counts can be configured by simply replacing the binary indicators, $\delta_{ij}(k)$, with weights $w_{ij}(k)$ in the above expressions as well as in what follows. However, these are rarely used in practice, and here we will use only the binary $\delta_{ij}(k)$ to simplify discussion.

Another commonly used probability distribution corresponds to sampling pairs without replacement from a finite urn. This distribution, called "non-free sampling," is even more useful than "free-sampling" as a null hypothesis because it does not require knowledge of the true p_b (Cliff and Ord 1981), which are generally unknown in the experimental setting. For "non-free" the expected number of joins for any distance class k is

$$\mu_{bb}(k) = \frac{S_0(k) n_b^{(2)}}{2n^{(2)}} \tag{9}$$

and

$$\mu_{bc}(k) = \frac{S_0(k) n_b n_c}{n^{(2)}} \tag{10}$$

for $b \neq c$. Here n_b is the number of times that type b occurs among all n locations, and $n_b^{(2)} = n_b(n_b - 1)$. The variances, $\sigma^2(k)$ under non-free sampling are:

$$\sigma_{bb}^2(k) = \frac{1}{4} \left[\frac{S_1(k) n_b^{(2)}}{n^{(2)}} + \frac{(S_2(k) - 2S_1(k)) n_b^{(3)}}{n^{(3)}} + \frac{(S_0(k)^2 + S_1(k) - S_2(k)) n_b^{(4)}}{n^{(4)}} \right] - \mu_{bb}^2(k), \tag{11}$$

where $\mu_{bb}(k)$ is given by Eq. (9), for joins between identical types, and

$$\sigma_{bc}^2(k) = \frac{1}{4} \left[\frac{2S_1(k) n_b n_c}{n^{(2)}} + \frac{(S_2(k) - 2S_1(k)) n_b n_c (n_b + n_c - 2)}{n^{(3)}} + \frac{4(S_0(k)^2 + S_1(k) - S_2(k)) n_b^{(2)} n_c^{(2)}}{n^{(4)}} \right] - \mu_{bc}^2(k) \tag{12}$$

for $b \neq c$, where $\mu_{bc}(k)$ is given for joins between different types (Eq. (10)), and $n^{(2)} = n(n - 1)$, $n^{(3)} = n(n - 1)(n - 2)$, $n^{(4)} = n(n - 1)(n - 2)(n - 3)$, and similarly for the $n_b^{(3)}$ etc. (Cliff and Ord, 1973, 1981).

The relationships among sets of joins are inherently complex. For example, in the simplest case where there are only two types, B and W , in a spatial distribution none of the numbers of types of joins, n_{BB} , n_{BW} , or n_{WW} contain all information, but any one of these is redundant with respect to the other two since they sum to $S_0(k)/2$. For example, the values of $n_{BB}(k)$ and $n_{BW}(k)$ joins are not independent, under either the free or non-free sampling null hypothesis. Although the variances have been found Cliff and Ord (1973), the probabilities of most other products of join-counts have not.

3. Probability distributions of other products of join-counts

Other products must be found in order to more fully characterize the joint distributions of join-counts for arbitrary numbers of types, and this will for example also allow specification of the covariances among different weighted averages of join-counts, under both free and non-free sampling. To do so requires consideration of distributions of only four types, B , W , C , D . To simplify exposition we make the minor assumption that there are enough of all types so that the $n_b^{(4)}$ are defined.

Using Eq. (1), the expected value of the product $n_{BB}(k)n_{BW}(k)$ is:

$$E\{n_{BB}(k)n_{BW}(k)\} = E\left\{\left(\frac{1}{2}\sum_{(2)}\delta_{ij}(k)x_{iB}x_{jB}\right)\left(\sum_{(2)}\delta_{kl}(k)x_{kB}x_{lW}\right)\right\}. \tag{13}$$

Because the $\delta_{ij}(k)$ are fixed by the spatial distribution of locations and the definition of distance class k , Eq. (13) equals

$$= \frac{1}{2}\sum_{(2)}\sum_{(2)}\delta_{ij}(k)(k)Ex_{iB}x_{jB}x_{kB}x_{lW} \tag{14}$$

(Please note that the index k is not to be considered as the distance class k). Eq. (14) may be written out more explicitly:

$$= \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n\delta_{ij}(k)\delta_{kl}(k)Ex_{iB}x_{jB}x_{kB}x_{lW} \tag{15}$$

It is helpful to separate the $[n^2(n-1)^2]$ terms in Eq. (15) into three components. The first component contains all terms in which both locations are indexed twice. One of the following two constraints applies: either $k = i, j = l$; or $k = j, l = i$. This component is:

$$\sum_{(2)}[\delta_{ij}^2(k)Ex_{iB}x_{jB}x_{iB}x_{jW} + \delta_{ij}(k)\delta_{ji}(k)Ex_{iB}x_{jB}x_{jB}x_{iW}]. \tag{16}$$

There are a total of $2n(n-1)$ terms in this component (spread equally among the two constraints), and by Eq. (5) the sum of the functions of $\delta_{ij}(k)$ (or weights, if the weights are symmetric, i.e. all $w_{ij}(k) = w_{ji}(k)$) is $S_1(k)/2$.

The second component of Eq. (15) contains all terms where one and only one of the locations is indexed twice. One of the following four constraints must apply: either: $i = k, j \neq l$; $i \neq l, j = k$; $i = l, j \neq k$; or $i \neq k, j = l$. The second component is:

$$\begin{aligned} & \sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n[\delta_{ij}(k)\delta_{il}(k)Ex_{iB}x_{jB}x_{iB}x_{lW} \\ & + \delta_{ij}(k)\delta_{jl}(k)Ex_{iB}x_{jB}x_{jB}x_{iW}] \\ & + \sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n[\delta_{ij}(k)\delta_{ki}(k)Ex_{iB}x_{jB}x_{kB}x_{iW} \\ & + \delta_{ij}(k)\delta_{kj}(k)Ex_{iB}x_{jB}x_{kB}x_{jW}] \end{aligned} \tag{17}$$

and

$$\sum_{(3)} \equiv \sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n \text{ where } i \neq j \neq k.$$

There are a total $4n(n-1)(n-2)$ terms in this component spread equally among the four constraints, and the binary variables (or weights) sum to $S_2(k) - 2S_1(k)$ (using Eqs. (5) and (6)).

The third component consists of all terms in which no location is indicated twice:

$$\sum_{(4)}\delta_{ij}(k)\delta_{kl}(k)Ex_{iB}x_{jB}x_{kB}x_{lW}, \tag{18}$$

where

$$\sum_{(3)} \equiv \sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n \text{ where } i \neq j \neq k \neq l.$$

In this component the total number of terms is $n(n-1)(n-2)(n-3)$ and the sum of the binary variables (or weights) is $S_0^2(k) - S_2(k) + S_1(k)$ (using Eqs. (4)–(6)).

As noted, all of the terms in the first component (Eq. (16)) are subject to a constraint that they have either $k = i, j = l$ or $k = j, l = i$. Both of these conditions give impossible assignments of both B and W to the same location. For example, the product $x_{iB}x_{jB}x_{iB}x_{jW}$ assigns both B and W to location j . Thus it is not possible for all four indicator variables to equal one in any of the terms, and the expected values of all terms are zero, as is the expected value for this component.

For the second component (Eq. (17)), the triple summation sums over i, j , and k , with the constraint that $i \neq j \neq k$. This means that one of the following four constraints applies with equal incidence: $i = k, j \neq l$; $i \neq l, j = k$; $i = l, j \neq k$; or $i \neq k, j = l$. Each of the latter two constraints requires assignment of two types (B and W) to one of the three locations, and hence add nothing to the expected value of the sum. In contrast, all of the terms (one-half of the terms in the second component) meeting the first two constraints are possible. Under free sampling they all have expected value equal to the probability of choosing two B types and one W type with replacement, i.e. $p_B^2p_W$.

For the third component (Eq. (18)), all of the terms in the quadruple summation (over $i \neq j \neq k \neq l$) are possible, because there are four different locations involved. For free sampling all terms have expectation $p_B^3p_W$. Thus, under free sampling:

$$E\{n_{BB}(k)n_{BW}(k)\} = \frac{1}{2}p_B^2p_W \times [(S_2(k) - 2S_1(k))/2 + (S_0^2(k) - S_2(k) + S_1(k))p_B]$$

$n_{BW}(k) \cdot n_{BC}(k)$ is

$$E\{n_{BW}(k)n_{BC}(k)\} = p_{BPWP_C}[\frac{1}{4}(S_2(k) - 2S_1(k)) + (S_0^2(k) - S_2(k) + S_1(k))p_B] \tag{29}$$

or

$$E\{n_{BW}(k)n_{BC}(k)\} = \frac{p_{BPWP_C}}{4} \times [(S_2(k) - 2S_1(k)) + 4(S_0^2(k) - S_2(k) + S_1(k))p_B]. \tag{30}$$

Again it can be shown that under non-free sampling the expected value can be obtained by first combining powers $p_B^i p_W^j p_C^k$ and substituting $n_B^{(i)} n_W^{(j)} n_C^{(k)} / n^{(i+j+k)}$ in Eq. (29) or Eq. (30).

The next case is for products of the form $BWCC$:

$$E\{n_{BW}(k)n_{CC}(k)\} = E\left\{ \left(\sum_{(2)} \delta_{ij}(k) x_{iB} x_{jW} \right) \left(\frac{1}{2} \sum_{(2)} \delta_{kl}(k) x_{kC} x_{lC} \right) \right\} \tag{31}$$

Again, Eq. (31) can be separated into three components. All terms in both the first and second components of Eq. (31) involve double assignments of types to location(s). The third component has all terms that are possible and under free-sampling each has expected value $p_{BPWP_C}^2$, and under free sampling the expected value is

$$E\{n_{BW}(k)n_{CC}(k)\} = \frac{1}{2} p_{BPWP_C}^2 (S_0(k)^2 - S_2(k) + S_1(k)). \tag{32}$$

Under non-free sampling the expected value can be obtained by substituting for $p_{BPWP_C}^2$ and substituting $n_B n_W n_C^{(2)} / n^{(4)}$ in Eq. (32).

The final type of product is $BWCD$, and

$$E\{n_{BW}(k)n_{CD}(k)\} = E\left\{ \left(\sum_{(2)} \delta_{ij}(k) x_{iB} x_{jW} \right) \left(\sum_{(2)} \delta_{kl}(k) x_{kC} x_{lD} \right) \right\}. \tag{33}$$

As in the preceding case, the first two components of Eq. (33) all involve impossible assignments and all of the third component terms do not. Under free sampling, each of the latter have expected value $p_{BPWP_C} p_D$. Hence, under free sampling the expected value is

$$E\{n_{BW}(k)n_{CD}(k)\} = p_{BPWP_C} p_D (S_0(k)^2 - S_2(k) + S_1(k)). \tag{34}$$

The expected value under non-free sampling is obtained by substituting $n_B n_W n_C n_D / n^{(4)}$ for $p_{BPWP_C} p_D$ in Eq. (34).

The above results can be used to obtain the cross-products for all pairs of types of joins, irregardless of the

number of types, simply by observing the appropriate case above and substituting the types accordingly. These can be used to obtain the covariances (in addition to the variances) and correlation coefficients among all types of joins for any multitype spatial distribution. Moreover, this can also be done for nonspatial problems that may involve the placement of pairs of multinomial data elements into a class, under free or non-free sampling. The results also allow the formulation of variances for any weighted sum of join-counts.

It is worth noting that under free-sampling the covariance between BW and CC is

$$\frac{1}{2} p_{BPWP_C}^2 (S_1(k) - 2S_2(k)). \tag{35}$$

The term within the parentheses in Eq. (35) is always negative, thus the covariance is always negative. The same is true for the non-free case, and it is also true for covariances of the type $BWCD$.

It is also worth noting that the covariances typically involve $S_2(k)$ and hence the constellation of sample locations. Thus the covariances are intimately connected to the actual distribution of sample locations.

4. Discussion

All covariances between join-counts can be found under both free and non-free sampling, using the results. The above equations (Eqs. (19), (24), (30), (32) and (34) and the equations for the variances—Eqs. (3), (8), (11) and (12)) are mostly expressed for the expected values of the products, explicitly for the free sampling case. But in all cases expected values for non-free sampling can be obtained by combining powers of $p_B^i p_W^j p_C^k p_D^l$ in the equation and substituting $n_B^{(i)} n_W^{(j)} n_C^{(k)} n_D^{(l)} / n^{(i+j+k+l)}$. Covariances are obtained by subtracting off the appropriate form of the product of expected values of join-counts (using Eqs. (2), (7), (9) or Eq. (10)).

The covariances appear to depend on some of the values of $S_0(k)$, $S_1(k)$, and $S_2(k)$. The value of $S_2(k)$ is particularly dependent on the specific spatial distribution of locations. However, in cases where the spatial distribution of locations is supported on a lattice, and distance classes are well-defined as spatial lags, $S_2(k)$ can be determined from the size of the lattice (e.g. Krishna-Iyer, 1950). Nonetheless, in the general case $S_2(k)$ is determined by computer in spatial autocorrelation analyses. This suggests an algorithm, which the author has put into place, for calculating covariances, for application to any data set where each spatial location has one type. The algorithm classifies the expected products of joins (and ultimately covariances) each into one of seven classes, while simultaneously counting the numbers of types. (As noted earlier non-free sampling is more useful, and it requires knowledge

of the number of each type.) The algorithm keeps track of the types involved, numbers of joins, and $S_0(k)$, $S_1(k)$, and $S_2(k)$. The seven classes of products are, generically speaking, typified by $BBBB$, $BWBW$, $BBBW$, $BBWW$, $BWBC$, $BBWC$, and $BWCD$. For example, the product of EE joins with FG joins (or vice versa) is in the same class as $BBWC$. The class defines the form of the equation for finding the products or covariances.

One application of the covariances is in finding standard errors of (weighted or unweighted) averages of join-counts. For example, as was noted in Section 1, in spatial autocorrelation analysis of genotypes, typical applications of Moran's I -statistics involve weighted averages of join-counts of pairs of diploid genotypes (Epperson 1995, 2003). Usually several to many Moran's I -statistics are calculated for the same distribution of single-locus genotypes, one I -statistic for each allele. In cases where there are only two alleles, the two I -statistics are identical and procedures are straightforward. For a locus with three or more alleles, per-allele I -statistics are not identical, but because they involve overlap of some joins they are dependent. By expressing the covariances of I -statistics in terms of expected values of products of join-counts, given by the results in this article, it becomes possible to find the covariances of I -statistics under the null hypothesis (Epperson, 2003). Because I -statistics are asymptotically normally distributed, it is further possible to find the distribution of average values of Moran's I (averaged over alleles), and to form proper test statistics, for any spatial distribution of sample locations. Further, because it

can usually be assumed that covariances among loci are near zero, these allow derivation of the first test statistics averaged over alleles and loci that have known distributions. Such summary statistics should have considerably greater statistical power than single-allele measures.

References

- Cliff, A.D., Ord, J.K., 1973. Spatial Autocorrelation. Pion, London.
- Cliff, A.D., Ord, J.K., 1981. Spatial Processes. Pion, London.
- Epperson, B.K., 1993. Recent advances in correlation studies of spatial patterns of genetic variation. *Evol. Biol.* 27, 95–155.
- Epperson, B.K., 1995. Fine-scale spatial structure: correlations for individual genotypes differ from those for local gene frequencies. *Evolution* 49, 1022–1026.
- Epperson, B.K., 2003. Geographical Genetics. Princeton University Press, Princeton, NJ (in press).
- Krishna-Iyer, P.V., 1950. The theory of probability distributions of points on a lattice. *Ann. Math. Stat.* 21, 198–217.
- MacArthur, R.H., 1972. Geographical Ecology: Patterns in the Distribution of Species. Princeton Univ. Press, Princeton, New Jersey.
- Malécot, G., 1948. Les Mathématiques de l'Hérédité. Masson, Paris.
- Malécot, G., 1972. Génétique des populations naturelles dans le cas d'un seul locus. II Étude du coefficient de parenté. *Ann. Génét. Sélection Animale*, 4, 385–409.
- Moran, P.A.P., 1948. The interpretation of statistical maps. *J. R. Stat. Soc. B* 10, 243–251.
- Nei, M., 1973. Analysis of gene diversity in subdivided populations. *Proc. Nat. Acad. Sci. USA* 70, 3321–3323.
- Sokal, R.R., Oden, N.L., 1978. Spatial autocorrelation in biology. 1. Methodology. *Biol. J. Linn. Soc.* 10, 199–228.