Geographical Analysis ISSN 0016-7363

Beyond Mule Kicks: The Poisson Distribution in Geographical Analysis

Daniel A. Griffith[1], Robert Haining[2]

[1]Ashbel Smith Professor, School of Social Sciences, University of Texas at Dallas, Richardson, TX,
[2]Department of Geography, University of Cambridge, Cambridge, U.K.

The Poisson model, discovered nearly two centuries ago, is the basis for analyses of rare events. Its first applications included descriptions of deaths from mule kicks. More than half a century ago the Poisson model began being used in geographical analysis. Its initial descriptions of geographic distributions of points, disease maps, and spatial flows were accompanied by an assumption of independence. Today this unrealistic assumption is replaced by one allowing for the presence of spatial autocorrelation in georeferenced counts. Contemporary statistical theory has led to the creation of powerful Poisson-based modeling tools for geographically distributed count data.

Introduction

Although the 1711 draft version of de Moivre's *The Doctrine of Chance* indicates that he came very close to discovering the Poisson distribution (Hald 1998, p. 214), Siméon-Denis Poisson is credited with introducing the distribution in 1837 that bears his name, when he published rules of probability based on the incidence of fatalities in the French army because of mule kicks. Although the Poisson distribution constituted an important contribution to mathematical probability theory, wider appreciation of its importance did not occur until 1898, when von Bortkiewicz published data on the number of deaths by horse or mule kicks in 10 (of 14 reported) corps of the Prussian army, each observed over 20 years. Shortly thereafter, appearing in one of the oldest textbooks on statistics, Bowley (1901, pp. 301–02) fitted the Poisson model to deaths from anthrax during the years 1875–1894. One feature of the Poisson distribution is that it deals with the occurrence of discrete events whose individual probability is low, but whose number of opportunities for occurrence is very high (the law of large numbers). During the century that followed Bortkiewicz's paper, the Poisson probability model has been used to describe and analyze rare events (the law of small numbers; Falk etal. 1994), rang-

Correspondence: Prof. Daniel Griffith, School of Social Sciences, University of Texas at Dallas, PO Box 830688, GR31, Richardson, TX 75083-0688
e-mail: daniel.griffith@utdallas.edu

Submitted: June 10, 2004. Revised version accepted: February 4, 2005.

ing from accident frequencies (or other types of malfunctions/failures), through queuing situations, to zoological sampling involving catch–recatch strategies. The many different types of spatial problems that are reviewed in this article have stimulated quantitative geographers' interest in the Poisson model.

The description and explanation of map patterns of objects or events has been of continuing interest in geography since its quantitative revolution in the 1950s and 1960s. The Poisson distribution has played an important role in this work. This article briefly reviews how geographers who are engaged in spatial analysis have applied the Poisson distribution, and how its specification as a model has been adapted to meet the requirements of specific problems.

The Poisson distribution may be defined in the following way. Let Y denote a random variable whose values are constrained to be nonnegative integers that are counts of the number of times some event occurs. Let $P\{Y = y\}$ denote the probability that the random variable Y takes the particular value y. If

$$P\{Y = y\} = \frac{e^{-\lambda}\lambda^{y}}{y!}, \lambda > 0, \text{ and } y = 0, 1, 2, \ldots, \tag{1}$$

then Y is said to have a Poisson distribution with intensity parameter λ. Fig. 1 shows examples of the Poisson distribution for different values of λ. As λ increases, the Poisson distribution first becomes relatively more symmetric (Fig. 1b, c), and then relatively less peaked (Fig. 1d), approximating a normal distribution as λ goes to

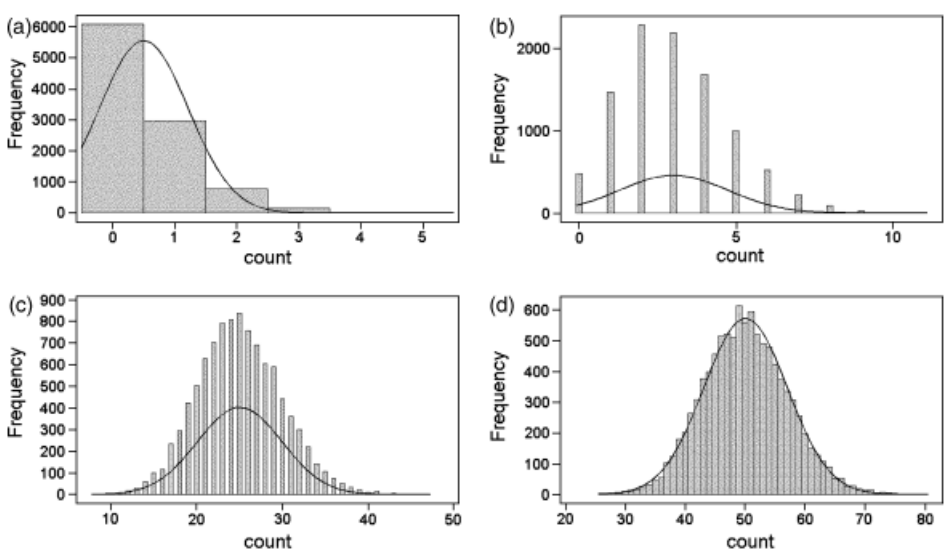

Figure 1. Histograms of 10,000 values randomly drawn from four Poisson distributions, upon which a normal curve with the same mean and variance has been superimposed. (a) Top left: $\lambda = 0.5$. (b) Top right: $\lambda = 3$. (c) Bottom left: $\lambda = 25$. (d) Bottom right: $\lambda = 50$. Note: The maximum frequency on the vertical axis decreases as λ increases.

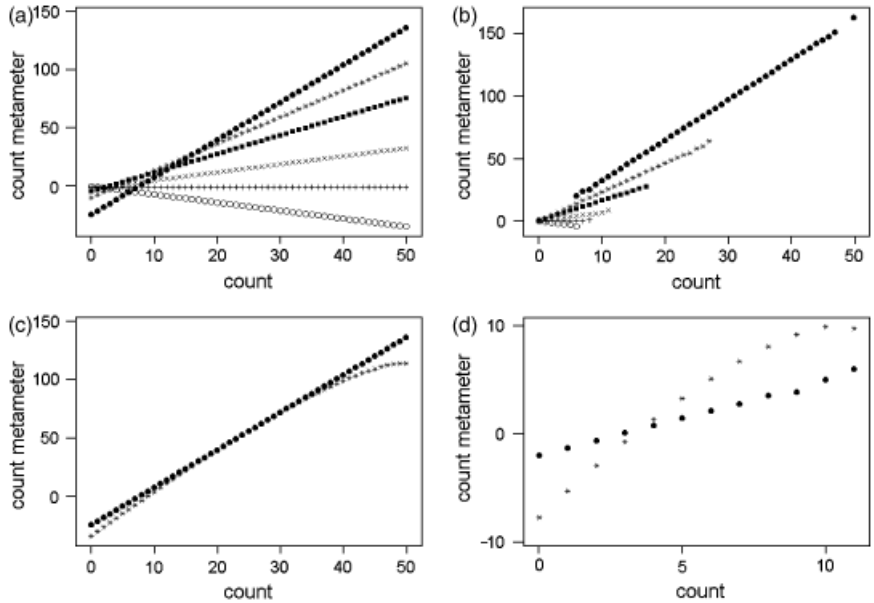

Figure 2. Poissonness plots, where count metameter = ln(theoretical frequency for count k)+ln[(count k)!]. Top: $\lambda = 0.5$ (open circle), 1 (plus), 2 (cross), 5 (solid square), 10 (asterisk), 25 (solid circle). (a) Top left: theoretical results. (b) Top right: simulated results based on 100,000 replications (note: many of the very rare extreme values do not get sampled). Bottom: Poisson variable results are denoted by a solid circle, and binomial variable results are denoted by an asterisk. (c) Bottom left: theoretical results for Poisson $\lambda = 25$, and binomial $n = 50$ and $p = 0.5$. (d) Bottom right: simulated results based on 100,000 replications for Poisson $\lambda = 2$, and binomial $n = 11$ and $p = 0.5$.

infinity. A Poissonness plot (Hoaglin 1980), which is comparable with a normal quantile plot, results in points along a straight line when the data follow a Poisson distribution. Fig. 2 shows examples of this plot for selected values of λ. Fig. 2b illustrates that as λ decreases from 25 to 0, the probability of observing counts even as large as 10, becomes negligible. The binomial example (Fig. 2c, d) illustrates that when data deviate from a Poisson distribution, the points form a curved line. An important property of the Poisson distribution is that both its mean and its variance equal λ; hence, straight lines appear as in Fig. 2.

The Poisson distribution may be used as an approximation to the binomial distribution when the probability of some phenomenon occurring is very small (e.g., $P<0.01$) but the number of trials (n) is very large (e.g., $n>1000$). Suppose n trials of an experiment are carried out so that the outcome for each trial is either a ''success'' (denoted by 1) or a ''failure'' (denoted by 0). Let P denote the probability of a ''success'' ($0<P<1$); hence, $(1-P)$ denotes the probability of a ''failure.'' The outcome of each trial is independent of all other trials, and the probabilities of success and failure remain the same throughout n trials. If Y denotes the random

variable that counts the number of successes in n trials, then

$$P\{Y = y\} = \frac{n!}{y!(n-y)!} P^y (1-P)^{n-y}, P > 0, \text{ and } y = 0, 1, 2, \ldots, n, \qquad (2)$$

is said to have a binomial distribution with parameters n and P. If n gets very large ($n \to \infty$) whereas P gets very small ($P \to 0$), such that the product $nP \to \lambda$ (a constant), then under these conditions the distribution of Y can be shown to converge in the limit to a Poisson distribution whose mean is given by nP. (A proof of this result is given, for example, in Meyer [1970, pp. 161–62].) Therefore, the Poisson distribution is used as a model to describe the outcome of a series of a large number, n, of independent trials where the probability, P, of a ''success'' in each trial is small (e.g., $P<0.01$).

Consider another situation in which a total of n point events are allocated to a spatially continuous geographic region A in such a way that:

1. the number of points found in nonoverlapping subregions in A are independent;
2. the probability of finding two or more points in a sufficiently small subregion (ΔA) is negligible (more formally, [probability of two or more events in ΔA]/ $|\Delta A| \to 0$ as $|\Delta A| \to 0$); and
3. the distribution of the number of points found in any subregion B of A is directly proportional to the area of B (denoted $|B|$), does not depend on the location of B, and follows a Poisson distribution with mean $\lambda|B|$.

Given these assumptions, the number of point events in any subregion B of A is a random variable that follows a Poisson distribution with intensity parameter λ, which defines the density of points per unit area. This situation defines a (homogeneous) Poisson process, also sometimes referred to as complete spatial randomness, the basic model for randomness in two-dimensional space (Cressie 1991, p. 586). Fig. 3 shows one realization of a Poisson process having $n = 250$ points. Note that even though the process is random, subregions can be identified from visual inspection that show apparent clusters of points, the reason for which is explained, for example, in Cressie (1991, p. 586).

Such events may be observed using randomly or contiguously placed quadrats in A, in which case data comprise counts. This aggregation results in a significant loss of spatial information, because the relative positions of events within each quadrat are discarded. In earlier studies, quadrat location was also ignored, thus, discarding still more spatial information (see Cressie 1991, pp. 588–94). Of note is that the associated likelihood function for statistical analysis purposes is different in the case of quadrat counts compared with when the original point locations are recorded (Cressie 1991, Chapter 8).

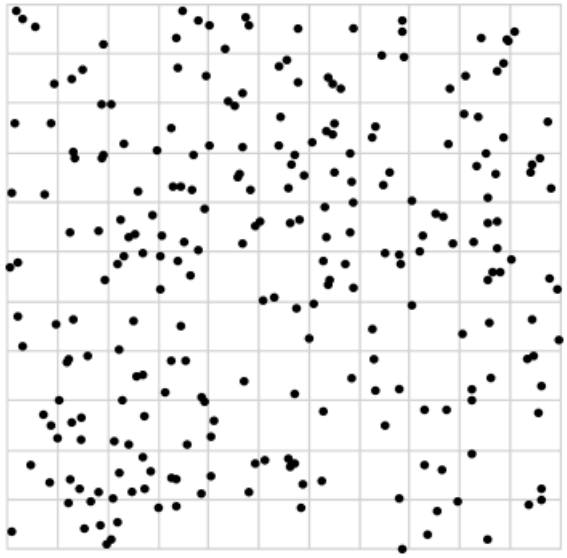

Figure 3. The geographic distribution of coordinates that are pairs of random drawings from a bivariate uniform distribution.

Early geographic applications of the Poisson distribution

Applications of the Poisson distribution in the 1950s and 1960s fell broadly into three categories: analysis of quadrat counts, analysis of point patterns, and identifying when the number of times an event has occurred is extreme. In the 1960s, fitting probability models became an important way of describing certain types of maps. Berry and Marble's (1968) edited collection provides an interesting range of papers that reflects this type of work, including a pioneering study by Matui (1932) of the distribution of scattered villages in two regions on the Tonami Plain in Japan. Both regions were partitioned into N contiguous quadrats—regular square spatial units of the same size, $|B|$. Let $Y(i)$ denote the number of events (villages) in quadrat i ($i = 1, \ldots, N$) for a selected region. Under the assumption of an underlying Poisson process for the distribution of villages, Equation (1) becomes

$$P\{Y(i) = y(i)\} = \frac{e^{-\lambda|B|}\lambda|B|^{y(i)}}{y(i)!}, y(i) = 0, 1, 2, \ldots, i = 1, 2, \ldots, N. \qquad (3)$$

The mean of the Poisson process, $\lambda|B|$, is estimated by $\Sigma_i y(i)/N$, which is its maximum likelihood estimator. The frequency distribution of observed counts can be compared with the theoretical frequency distribution that would arise if the $\Sigma_i y(i)$ villages followed a Poisson distribution with mean $\lambda|B| = \Sigma_i y(i)/N$. Matui shows that the distribution of village counts, by quadrats, for one of the regions on the Tonami Plain follows a Poisson distribution, while the other does not.

If data are available for the location of objects on a continuous surface, then tests can be undertaken that compare the average observed with the expected

distance between nearest neighbors under the assumption of a Poisson point process. King (1962) provides an illustration of this approach, analyzing the spatial pattern of settlements in different U.S. states. More recently, tests for randomness where the spatial information on point locations is retained are based on *K* function analysis, which allows tests to be performed simultaneously along a continuum of scales (Ripley 1981).

Another important application of the Poisson model is the identification of areas where the numbers of events of a particular type are unlikely to have occurred by chance alone. This is an early example of cluster detection. Choynowski (1959) proposed the following test for regions with different sizes of at-risk populations, and then applied a Poisson model to the incidence of brain tumors in a region of Poland subdivided into 17 counties. In his application, Equation (3) becomes

$$P\{Y(i) = y(i)\} = \frac{e^{-\lambda(i)}\lambda(i)^{y(i)}}{y(i)!}, y(i) = 0, 1, 2, \ldots, i = 1, 2, \ldots, 17, \qquad (4)$$

where $Y(i)$ is the number of brain tumors in county i, and the mean, $\lambda(i)$, depends on the size of the at-risk population in county i. The Poisson model is used here as an approximation to the binomial model. The number of trials in county i is its at-risk population, and the probability of having a brain tumor is very small. Under the assumption that brain tumors occur at random across a region, $\lambda(i)$ is estimated as the product of the assumed region-wide risk (p = the total number of cases divided by the total population at risk in a study region) times the size of the at-risk population in county i (i.e., $n(i)$). Choynowski's test, for each county, involves calculating $P\{Y(i) \geq y(i)\}$, where $y(i)$ is the observed number of brain tumors in county i. The null hypothesis that an observed count could be due to chance variation is rejected when a probability is less than or equal to α, the level of significance. This test amounts to establishing, for any county i, whether or not its observed count lies in the right-hand ($\alpha \times 100\%$) tail of the Poisson distribution with parameter $\lambda(i)$. Rejection of the null hypothesis might encourage an analyst to explore whether area-specific factors are responsible for a high count.

Some of the early applications of the Poisson model in locational analysis were undertaken in the belief that this was a first step in trying to identify the underlying process responsible for an observed spatial pattern of objects. Such an agenda had its limitations, as Matui (1932) points out. The ''mere coincidence of their superficial characters (does not imply) that the two phenomena have also the same causes'' (Berry and Marble 1968, p. 152). In other words, just because a map pattern can be described by a Poisson distribution does not necessarily imply that the underlying process generating it was a purely random one. Getis and Boots (1978) demonstrate this point with examples of different stochastic location processes that give rise to the same probability distribution of counts by quadrat. At best, fitting such models provides a description of a static situation that is a manifestation, at a given point in time, of some underlying space–time process.

Similarly, apparent clusters of cases can occur even in a completely spatially random pattern (see Fig. 3). As the number of tests increases (i.e., as the number of quadrats, counties, or subregions increases), the probability of finding at least one areal unit for which the null hypothesis is rejected converges to 1. This complication is known as the problem of multiple testing, and for independent tests, can be corrected in a conservative way that preserves an overall level of significance, α, by incorporating a Bonferroni adjustment to significance levels for the individual tests (i.e., α/(number of tests)).

Subsequent attention has focused on this as well as on other aspects of the problem of analyzing map patterns. The following observations apply to fitting the Poisson distribution or, indeed, any other probability model to a spatial data set:

- *Scale effects*. The goodness of fit of a Poisson model may depend on the choice of quadrat size, implying that different sized quadrats should be investigated. In the case of distance-based tests, not only first, but also second, third, and higher order nearest neighbor statistics must be examined. Ripley (1981) developed the *K* function, which provides a description and hypothesis test for point patterns at a continuum of scales. See Haining (1982) for a *K* function application to settlement geography. The identification of clusters depends on the relationship between the geographical extent of any real cluster and the size of the spatial partitioning unit used in an empirical analysis. Openshaw et al. (1987) developed the geographical analysis machine (GAM) to test for clusters at different scales. Subsequent developments by, among others, Besag and Newell (1991) and Kulldorff (1997, 1998), have addressed different aspects of the multiple testing problem.
- *Partition effects*. At any given scale, the goodness of fit of a model could depend on the specific choice of surface partitioning, which is important because of the arbitrary nature of partitionings such as census tracts. The combination of scale and partitioning effects has become known as the modifiable areal unit problem (MAUP; Holt et al. 1996a, b; Tranmer and Steel 1998).
- *Boundary (i.e., edge) and frame effects*. The choice of study area, as well as its areal extent and shape, can influence results. Detection of a cluster of points depends on the size of a study region, whereas the fit of a model to describe spatial variation can be affected by the shape of a study area (Haining 1990, p. 373), as well as by the distribution of locations within a region to which attribute values are tagged (Unwin and Wrigley 1987). At the boundary, nearest neighbors lying outside a region may be unknown, and statistics need to be adapted to reflect this lack of data (Ripley 1981; also see Griffith 1983, 1985; Griffith and Amrhein 1983). (See Gatrell et al. (1996) for an application in medical geography using Ripley's *K* function with an edge correction.)
- *Spatial heterogeneity*. Spatial variation may arise because the appropriate model formulation varies across a map—subregional map descriptions require different model forms and/or varying parameters of a single model form. In

other words, the underlying structure or generation mechanism is not the same across the entire map. In the case of a Poisson process, not all parts of a study region may be equally likely to contain the object or event of interest (e.g., not all parts of an area may be equally appropriate for settlement), resulting in λ varying across a geographic landscape (see ''Models assuming independence between neighboring counts''). So-called ''local statistics'' (e.g., LISA statistics) based on analyzing spatially defined subsets of data may be used to investigate for the presence of these situations (Getis and Ord 1992; Anselin 1995; Haining 2003, pp. 250–63).

- *Location errors*. Errors in recording the location of an event can have implications for detecting the presence of clusters, as well as for testing whether a map pattern shows evidence of clustering (Kulldorff 1998).

Evolution of Poisson-based forms of spatial analysis

Developments in the use and specification of the Poisson model within spatial analysis have followed three basic thrusts, namely: modified Poisson process models for describing spatial variation, explaining variation in Poisson counts, and Poisson modeling of spatial flows.

Modified Poisson process models

Models assuming independence between neighboring counts

In the study by Matui (1932), one of the two regions of the Tonami Plain had a count distribution, by quadrats, that did not conform to a Poisson distribution. Dacey (1964) revisited the data set using a modified Poisson model. He superimposed a second process on the Poisson process defined by Equation (3). In this second process, a second set of points is distributed at random under the constraint that no quadrat could receive more than one point. This is a Bernoulli process. The combination of Poisson and Bernoulli processes yields a ''point pattern more regular than random.'' This was a modification Dacey was to use in other applications (see, e.g., Dacey 1966b). It is an example of an additive mixed Poisson process. Other Poisson mixtures deal with the problem of spatial heterogeneity by combining the weighted sum of two spatially separated Poisson processes that have different intensity parameters. These models are summarized and discussed in Getis and Boots (1978, pp. 45–48), who also discuss the class of multiplicative mixed models. This latter class includes *compound* models in which the Poisson intensity parameter (λ) is itself a random variable. When λ varies from quadrat to quadrat as an independent Γ random variable, the resulting distribution for the counts is negative binomial (see, e.g., Dacey 1966a). Multiplicative mixed models also include generalized models that involve a two-stage process. The location of, say, a settlement is the outcome of a Poisson location process; the number of houses in the settlement (the ''offspring'') also is the outcome of a Poisson process (leading to a Neyman Type A distribution). Alternatively, the second-stage process is the outcome of a truncated Poisson process that prohibits the occurrence of a zero count,

thus ensuring that settlements always have at least one house (a Thomas process). Getis and Boots (1978, pp. 48–70) discuss these and other models at length.

Models allowing for spatial correlation between neighboring counts

All of the preceding models assume that area counts are independent. In the case of the generalized models described in ''Early geographic applications of the Poisson distribution,'' the ''offspring'' are simply counts with no spatial distribution. There is an implicit assumption that, in a quadrat-based analysis, any spatial dispersion of offspring is very small relative to the scale of quadrats.

In many generic contexts the assumption of independence between neighboring counts cannot be expected to hold. One source of spatial dependence could be the nature of the underlying process generating the data. Haining (2003, pp. 15–22), for example, discusses different types of generic spatial processes that would invalidate the independence assumption, including dispersal processes in which the dispersion range is moderate to large relative to the scale of the spatial units in terms of which the process is observed and measured. Retaining information on the location of the spatial units now becomes important. To model such situations—that is, to model both the aspatial counts and the spatial variation in those counts—Besag (1974) defined a general class of models that could be used to describe different types of spatially correlated data. He called these ''auto-models,'' and he included in his paper the ''auto-Poisson'' model. Let $P\{Y(i) = y(i)|\{Y(j)\}, j \in N(i)\}$ denote the conditional probability that the random variable $Y(i)$, defined at location i, takes on the particular value $y(i)$, given the values of Y at the sites that are the neighbors of i, say $N(i)$. If

$$P\{Y(i) = y(i)|\{Y(j)\}, j \in N(i)\} = \frac{e^{-\lambda(i)}\lambda(i)^{y(i)}}{y(i)!}, y(i) = 0, 1, 2, \ldots, \qquad (5)$$

where $\ln[\lambda(i)] = \alpha(i) + \sum_{j \in N(i)} b(i,j)y(j)$, then the $\{Y(i)\}$ are said to have an auto-Poisson distribution with intensity parameter $\lambda(i)$, through which spatial dependency is introduced. The parameter $\alpha(i)$ is an area-specific effect. The term $b(i,j) = b(j,i)$, following customary practice in this area of modeling, might be specified as follows: $b(i,j) = \gamma\ w(i,j)$, where γ is a spatial autoregressive parameter. The term $w(i,j)$ equals 0 or 1; this term represents the neighborhood structure of the areas under study. Thus, if $N(i)$ denotes the set of neighbors of area i, then $w(i,j) = 1$ if j is a neighbor of i [$j \in N(i)$]; otherwise, $w(i,j) = 0$. In addition, $w(i,i) = 0$ must be assumed. However, there is a problem with this model. When consistency conditions are enforced (e.g., the probability of all possible outcomes must sum to 1), then $\gamma \leq 0$. Therefore, the auto-Poisson model can only describe map patterns with negative spatial correlation—that is, where the underlying processes are competitive. Because negative spatial autocorrelation rarely is encountered in practice, this restriction severely limits the applicability of this model. Kaiser and Cressie (1997) circumvent this negativity constraint by Winsorizing counts to a finite set of integers, which sets an upper limit on the largest count that can occur. Winsorizing

involves systematically replacing extremely high counts with the value of some cut-off criterion (after Barnet and Lewis 1978), and is similar to truncating extreme counts. The infinite sum of a Poisson probability model is replaced with a finite sum, whose upper limit is an acceptable most extreme count, which utilizes all of the Poisson type counts information in a data set.

Although the Kaiser–Cressie modification of the auto-Poisson model allows positive spatial autocorrelation to be captured with a single parameter, Griffith (2002, 2003) proposes a spatial filtering approach to modeling spatially correlated Poisson counts that usually requires far more than one parameter. In his approach a map represents the realization of n independent Poisson samples, one for each areal unit, where the mean, $\lambda(i)$—and hence, the variance—varies across the map and is spatially structured. In this model

$$\ln[\lambda(i)] = \alpha(i) + \sum_{k=1}^{K} \eta(k) E(i,k),$$

where $E(i,k)$ is the value for areal unit i associated with the kth eigenvector of the matrix $(\mathbf{I}–\mathbf{D})\mathbf{W}(\mathbf{I}–\mathbf{D})$ that appears in the numerator of the Moran Coefficient. $\mathbf{W}$ is the matrix of $w(i,j)$ values describing an assumed neighborhood structure (see the preceding discussion), $\mathbf{I}$ is the identity matrix, and $\mathbf{D}$ is a matrix containing $1/n$ in each of its cells. $K < n$ is the number of eigenvectors used to construct a spatial filter, and $\eta(k)$ and $\alpha(i)$ are parameters to be estimated, where $\alpha(i)$ denotes area-specific effects. Specification of spatial structure derives from assumptions about which areal units are neighbors. This model formulation specifies spatial heterogeneity in the mean and thus has similarities with the mixture models discussed in ''Models assuming independence between neighboring counts.'' It differs from that for the auto-Poisson model, in which Poisson samples are only conditionally independent. Furthermore, Griffith's specification allows counts to be positively autocorrelated—containing anywhere from weak to strong levels—because spatial correlation is embedded into the mean response of a Poisson model.

Yet another way to capture positive or negative spatial autocorrelation in a Poisson model is to define the mean response as

$$\ln[\lambda(i)] = \alpha(i) + S(i), \tag{6}$$

where $S(i)$ is a random variable (or random effect). When $\mathbf{S} = \{S(1), S(2), \ldots, S(n)\}$ is defined by the conditional autoregressive (CAR) model (Besag 1974)—whose multivariate normal covariance is specified in matrix form as $(\mathbf{I} - \tau\mathbf{W})^{-1}\ \sigma^2$—positive or negative spatial autocorrelation can be introduced into the counts $Y(i)$ through the spatial autoregressive parameter (τ) of the CAR model. Unlike the auto-Poisson model given by Equation (5), each $Y(i)$ value is an independent sample from a Poisson distribution, given $\mathbf{S}$. As with Griffith's model specification, spatial autocorrelation is captured by the spatial structure of the intensity parameter across a map.

(See Wakefield et al. (2000) for a discussion of applications of these models in the context of disease mapping.)

Explaining variation in Poisson counts

Regression is the basic workhorse of statistics for analyzing relationships. Poisson counts are modeled using the log-linear function

$$\ln\{E[Y(i)]\} = \ln[\lambda(i)] = \beta_0 + \beta_1 X_1(i) + \ldots.. + \beta_k X_k(i), \tag{7}$$

where $E[Y(i)]$ denotes the expected value of the Poisson variable $Y(i)$, $X_1, \ldots, X_k$ are a set of k predictors or explanatory variables, β_0 is the intercept parameter, and β_1, β_k are the k regression coefficient parameters. If size differences exist (e.g., areal units have varying population sizes), then an offset variable (i.e., a variable whose coefficient is set equal to one, rather than estimated) needs to be included (see, e.g., McCullagh and Nelder (1989), for an introduction to this class of models; see, e.g., Lovett and Flowerdew (1989) and Flowerdew and Geddes (1999) for geographical applications of these models).

The Poisson model assumes that the mean and variance of a random variable are equal. In the case of area data, because of within- and between-area heterogeneity, overdispersion is likely to be present; that is, $E[Y(i)] = \lambda(i)$ but $\mathrm{Var}[Y(i)] > \lambda(i)$. For example, when modeling the distribution of disease counts by area, overdispersion should be anticipated because: (1) population heterogeneity occurs within spatial units (e.g., in terms of genetic composition and lifestyle differences) and (2) important covariates are missing from the model specification, either because they have been overlooked or because measurements for them are unavailable. Spatial autocorrelation induces overdispersion. Fig. 4 illustrates the effects of increasing levels of positive spatial autocorrelation (the Moran Coefficient (MC) and the Geary Ratio (GR)) on a Poisson distribution ($\lambda = 5$) generated with a spatial filtering model specification. Considerable overdispersion coincides with the presence of strong positive spatial autocorrelation. Fig. 5 portrays the geographic distributions associated with the histograms appearing in Fig. 4. Fig. 5a exhibits a

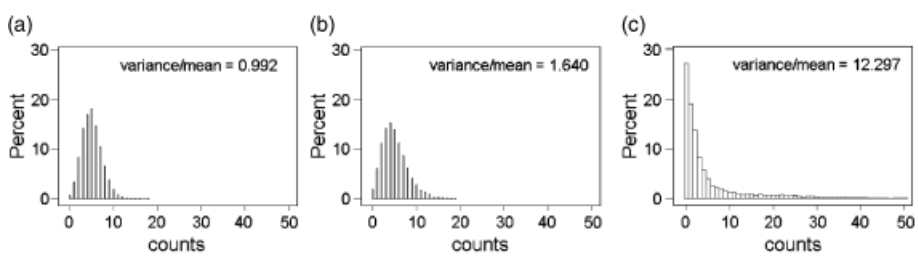

Figure 4. Histograms (in percentages) of 10,000 values—geographically distributed across a 100 × 100 regular square tessellation—randomly drawn from three Poisson distributions for which $\lambda = 5$. (a) Left: a negligible spatial autocorrelation case (MC = 0.011, GR = 0.989). (b) Middle: a moderate positive spatial autocorrelation case (MC = 0.414, GR = 0.591). (c) Right: a strong positive spatial autocorrelation case (MC = 0.982, GR = 0.081).

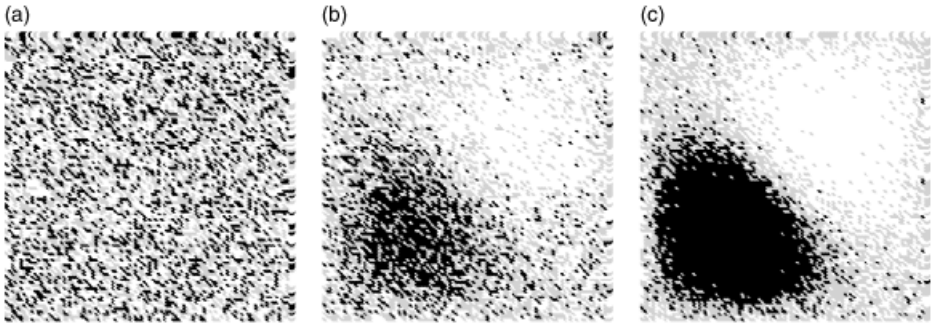

Figure 5. The geographic distributions of the Poisson variables portrayed in Fig. 4, with counts classified by approximate tertiles: white denotes 0–4, gray denotes 5–7, and black denotes 8–y_{max}. (a) Left: negligible spatial autocorrelation. (b) Middle: moderate positive spatial autocorrelation. (c) Right: strong positive spatial autocorrelation.

random distribution across the map of counts between 0 and 18. Fig. 5b exhibits a clustering of counts between 0 and 4 in the upper right-hand quadrant, and counts between 8 and 19 in the lower left-hand quadrant of the map. Fig. 5c exhibits a more pronounced clustering with the same geographic pattern contained in Fig. 5b; but in this case counts between 8 and 51 cluster in the lower left-hand quadrant of the map. Spatial autocorrelation induces a centripetal effect, amalgamating many of the smaller nonzero counts with larger counts, creating more 0 and very small counts and a few excessively large counts.

The ratio of the residual deviance to the corresponding residual degrees of freedom indexes the extent of overdispersion present after fitting a model; overdispersion is indicated when this ratio is significantly greater than 1 (note: 1.2 often is used as a threshold). Overdispersion can be handled by recognizing that the frequency distribution of a dependent variable resembles a particular probability distribution, even though establishing the precise form of the statistical model for the data may not be possible. Of course, without a proper statistical model maximum likelihood estimates of parameters cannot be computed because no likelihood function exists to maximize. In this situation a researcher can resort to quasi- (or pseudo-) maximum likelihood techniques. Here, inference is valid if the conditional mean is correctly specified: if the link function correctly relates the mean to covariates and/or if the variance is correctly related to the mean. Most log-likelihood asymptotic theory transfers to the quasi-likelihood case, which justifies computing standard errors and likelihood ratio statistics (McCullagh and Nelder 1989, Chapter 9). The quasi-likelihood approach involves the introduction of an artificial parameter that is a scaling factor, η, that accounts for overdispersion while relating the mean and variance. For Poisson-like data, $\mathrm{VAR}[Y(i)] = \eta\lambda$. This correction for overdispersion, for which $\eta > 1$, is equivalent to multiplying a covariance matrix by a scalar dispersion parameter to account for extra-Poisson variation. Some statistical packages (e.g., STATA and SAS) enable a researcher to compute

estimates of η. But overdispersion in Poisson data often is accounted for by resorting to use of the negative binomial model. One commonly used specification of the mean–variance relationship in this case is $\sigma^2 = \lambda + \eta\lambda^2$; η is the extra-Poisson variation parameter to be estimated, with $\eta = 0$ for a Poisson random variable.

If overdispersion is due to factors that generate spatial structure, such as missing covariates that are spatially autocorrelated, then a researcher might choose a random effects model of the form

$$\ln\{E[Y(i)]\} = \ln[\lambda(i)] = \beta_0 + \beta_1 X_1(i) + \ldots + \beta_k X_k(i) + S(i), \tag{8}$$

where $S(i)$ denotes the normal CAR model with spatial autocorrelation parameter τ (see ''Models allowing for spatial correlation between neighboring counts''). Parameters of Equation (8) can be estimated using the Bayesian software package WinBUGS (see, e.g., Wakefield et al. 2000; Spiegelhalter et al. 2002).

Equation (8) is used for estimating the relative risk of, say, disease by area, and identifying significant covariates that explain spatial variation in relative risk. In Equation (4), by definition, $\lambda(i) = E(i)r(i)$. $E(i)$ is the expected number of cases of a disease given the age and sex composition of the population in area i, using the age- and sex-specific rates from some larger population; $r(i)$ is the relative risk for area i—that is, the area-specific risk that cannot be attributed to the composition of the population, and, therefore, may be the result of factors that are unique to area i. The maximum likelihood estimate of $r(i)$ is $Y(i)/E(i)$, that is, the observed count divided by the expected count. Several problems are associated with this estimator. First, if $Y(i)$ is missing, then $r(i)$ cannot be computed. Second, in the case of rare diseases and where counts are accumulated over a short period of time, $Y(i)$ often is 0, and hence $r(i) = 0$, which most likely is a serious underestimate. Third, the standard error of $r(i)$, given by $Y(i)^{1/2}/E(i)$, is a decreasing function of the size of the population at risk in area i. Extreme values of $r(i)$ tend to be associated with areas having small populations; statistically significant values of $r(i)$—that is, values of $r(i)$ that are significantly different from 1—tend to be associated with areas having large populations. One solution to these types of problems is to adopt a principal from kriging and spatial interpolation, namely, to use information contained in data for neighboring areas, and a spatially structured $S(i)$ fulfills this role. Model [8] does this type of borrowing through the $S(i)$ term. In this case the log-mean response in Equation (8) takes the form

$$\ln[\lambda(i)] = \ln[E(i)] + \ln[r(i)],$$

where $\ln[E(i)]$ is an offset variable, and the right-hand side of Equation (8) is the log-relative risk for area i that explains the variation in risk as well as the spatially structured random effects term.

Poisson modeling of spatial flows

Flowerdew and Aitkin (1982) argue that the normal log-linear regression model provides an incorrect specification for spatial interaction models. They specify the

following model for flows, $T(i,j)$, between an origin i and a destination j:

$$P[T(i,j) = t(i,j)] = e^{-\lambda(i,j)}\lambda(i,j)^{t(i,j)}/t(i,j)!, \tag{9}$$

where the expected value is given by $E[T(i,j)] = \lambda(i,j) = e^{\alpha}\, m(i)^{\delta} m(j)^{\beta}\, d(i,j)^{\gamma}$; $m(i)$ and $m(j)$, respectively, denote the ''masses'' (i.e., sizes) of origin i and destination j; and α, β, δ, and γ are parameters. Equation (9) models flows as interpoint movement counts; hence, there is the specification of a discrete distribution. Flowerdew and Aitkin show the superiority of Poisson regression over ordinary least squares log-linear regression. Flowerdew and Lovett (1988) extend Equation (9) to singly and doubly constrained spatial interaction models, again assuming independent ''mass'' terms. For these constrained models, the Poisson regression formulation allows goodness-of-fit assessment (i.e., model diagnostics) and the inclusion of additional covariates to the basic ''mass'' terms and distance metric. Unfortunately, this formulation does not incorporate spatial dependencies in the origin and destination ''mass'' terms. Consequences of overlooking such spatial structure effects are conceptualized in Curry (1972), with their presence empirically demonstrated by Griffith and Jones (1980). One possible reformulation to account for them was posited by Bolduc et al. (1989, 1992, 1995). A more recent discussion of this topic appears in Tiefelsdorf (2003).

Concluding comments

This article provides a brief historical overview of the Poisson model and the evolution of its use in spatial analysis. This history traces a time path that begins with researchers ignoring spatial dependence in georeferenced count data, and ends with more contemporary analyses that offer a number of different ways of representing spatial dependency in such counts. Modern statistical theory, especially generalized linear model theory, and modern software, such as WinBUGS and MCMC algorithms (also see Wolpert and Ickstadt 1998), make implementation of models for counts that handle spatial dependence a feasible undertaking.

The use of log-normal approximations, with or without spatial model elements, should be a practice of the past. Not only has the application of the Poisson model come a long way from the days of deaths from mule kicks, but new statistical theory also has led to the creation of a powerful modeling tool for geographically distributed count data.

Acknowledgements

This research was completed during a visiting professorship residency in the Department of Geography, University of Cambridge. Support from the Leverhulme Trust and Fitzwilliam College is gratefully acknowledged.

References

Anselin, L. (1995). ''Local Indicators of Spatial Association—LISA.'' *Geographical Analysis* 27, 93–115.

Barnet, V., and T. Lewis. (1978). *Outliers in Statistical Data*. New York: Wiley.

Berry, B., and D. Marble. (1968). *Spatial Analysis: A Reader in Statistical Geography*. Englewood Cliffs, NJ: Prentice Hall.

Besag, J. (1974). ''Spatial Interaction and the Statistical Analysis of Lattice Systems.'' *Journal of Royal Statistical Society* 36B, 192–225.

Besag, J., and J. Newell. (1991). ''The Detection of Clusters in Rare Diseases.'' *Journal of the Royal Statistical Society* 154A, 143–55.

Bolduc, D., M. Dagenais, and M. Gaudry. (1989). ''Spatially Autocorrelated Errors in Origin–Destination Models: A New Specification Applied to Aggregate Mode Choice.'' *Transportation Research* 23B, 361–72.

Bolduc, D., R. Laferrière, and G. Santarossa. (1992). ''Spatial Autoregressive Error Components in Travel Flow Models.'' *Regional Science and Urban Economics* 22, 371–85.

Bolduc, D., R. Laferrière, and G. Santarossa. (1995). ''Spatial Autoregressive Error Components in Travel Flow Models: An Application to Aggregate Mode Choice.'' In *New Directions in Spatial Econometrics*, 96–108, edited by L. Anselin and R. Florax. Berlin, Germany: Springer-Verlag.

Bowley, A. (1901). *Elements of Statistics*. London: King & Son.

Choynowski, M. (1959). ''Maps Based on Probabilities.'' *Journal of the American Statistical Association* 54, 385–88.

Cressie, N. (1991). *Statistics for Spatial Data*. New York: Wiley.

Curry, L. (1972). ''Spatial Analysis of Gravity Flows.'' *Regional Studies* 6, 131–47.

Dacey, M. (1964). ''Modified Poisson Probability Law for Point Pattern More Regular Than Random.'' *Annals, Association of American Geographers* 54, 559–65.

Dacey, M. F. (1966a). ''A Compound Probability Law for a Pattern More Dispersed than Random and With Areal Inhomogeneity.'' *Economic Geography* 42, 172–79.

Dacey, M. (1966b). ''A County Seat Model for the Areal Pattern of an Urban System.'' *Geographical Review* 56, 527–42.

Falk, M., J. Hüsler, and R.-D. Reiss. (1994). *Laws of Small Numbers: Extremes and Rare Events*. Basel, Switzerland: Birkhäuser Verlag.

Flowerdew, R., and M. Aitkin. (1982). ''A Method of Fitting the Gravity Model Based on the Poisson Distribution.'' *Journal of Regional Science* 22, 191–202.

Flowerdew, R., and A. Geddes. (1999). ''Poisson Regression Analysis of Limiting Long-Term Illness Data for Parts of North East England.'' *Geographical & Environmental Modelling* 3(1), 63–82.

Flowerdew, R., and A. Lovett. (1988). ''Fitting Constrained Poisson Regression Models to Interurban Migration Flows.'' *Geographical Analysis* 20, 297–307.

Gatrell, A., T. Bailey, P. Diggle, and B. Rowlingson. (1996). ''Spatial Point Pattern Analysis and its Application in Geographical Epidemiology.'' *Transactions, Institute of British Geographers* 21, 256–74.

Getis, A., and B. Boots. (1978). *Models of Spatial Processes*. Cambridge, UK: Cambridge University Press.

Getis, A., and J. K. Ord. (1992). ''The Analysis of Spatial Association by Use of Distance Statistics.'' *Geographical Analysis* 24, 189–206 (with correction, 1993, 25, p. 276).

Griffith, D. (1983). ''The Boundary Value Problem in Spatial Statistical Analysis.'' *Journal of Regional Science* 23, 377–87.

Griffith, D. (1985). ''An Evaluation of Correction Techniques for Boundary Effects in Spatial Statistical Analysis: Contemporary Methods.'' *Geographical Analysis* 17, 81–88.

Griffith, D. (2002). ''A Spatial Filtering Specification for the Auto-Poisson Model.'' *Statistics and Probability Letters* 58, 245–51.

Griffith, D. (2003). *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*. Berlin, Germany: Springer-Verlag.

Griffith, D., and C. Amrhein. (1983). ''An Evaluation of Correction Techniques for Boundary Effects in Spatial Statistical Analysis: Traditional Methods.'' *Geographical Analysis* 15, 352–60.

Griffith, D., and K. Jones. (1980). ''Explorations into the Relationship Between Spatial Structure and Spatial Interaction.'' *Environment and Planning A* 12, 187–201.

Hald, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. New York: Wiley.

Haining, R. (1982). ''Describing and Modeling Rural Settlement Maps.'' *Annals, Association of American Geographers* 27, 211–23.

Haining, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge, UK: Cambridge University Press.

Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge, UK: Cambridge University Press.

Hoaglin, V. (1980). ''A Poissonness Plot.'' *The American Statistician* 34, 146–49.

Holt, D., D. Steel, and M. Tranmer. (1996a). ''Area Homogeneity and the Modifiable Areal Unit Problem.'' *Geographical Systems* 3, 181–200.

Holt, D., D. Steel, M. Tranmer, and N. Wrigley. (1996b). ''Aggregation and Ecological Effects in Geographically Based Data.'' *Geographical Analysis* 28, 244–61.

Kaiser, M., and N. Cressie. (1997). ''Modeling Poisson Variables with Positive Spatial Dependence.'' *Statistics and Probability Letters* 35, 423–32.

King, L. (1962). ''A Quantitative Expression of the Pattern of Urban Settlements in Selected Areas of the United States.'' *Tijdschrift Voor Economische en Sociale Geographie* 53, 1–7.

Kulldorff, M. (1997). ''A Spatial Scan Statistic.'' *Communications in Statistics: Theory and Methods* 26, 1481–96.

Kulldorff, M. (1998). ''Statistical Methods for Spatial Epidemiology: Tests for Randomness.'' In *GIS and Health*, 49–62, edited by A. Gatrell and M. Löytönen. London: Taylor & Francis.

Lovett, A., and R. Flowerdew. (1989). ''Analysis of Count Data Using Poisson Regression.'' *Professional Geographer* 41, 190–98.

Matui, I. (1932). ''Statistical Study of the Distribution of Scattered Villages in Two Regions of the Tonami Plain, Toyama Prefecture.'' *Japan Journal of Geology and Geography* 9, 251–66.

McCullagh, P., and J. Nelder. (1989). *Generalized Linear Models*. London: Chapman & Hall.

Meyer, P. (1970). *Introductory Probability and Statistical Applications*, 2nd ed. Reading, MA: Addison Wesley.

Openshaw, S., M. Charlton, C. Wymer, and A. Craft. (1987). ''A Mark 1 Geographical Analysis Machine for the Automated Analysis of Point Datasets.'' *International Journal of Geographical Information Systems* 1, 335–58.

Ripley, B. (1981). *Spatial Statistics*. New York: Wiley.

Spiegelhalter, D., N. Best, B. Carlin, and A. Van der Linde. (2002). ''Bayesian Measures of Model Complexity and Fit (with Discussion).'' *Journal of the Royal Statistical Society* 64B, 583–640.

Tiefelsdorf, M. (2003). ''Misspecifications in Interaction Model Distance Decay Relations: A Spatial Structure Effect.'' *Journal of Geographical Systems* 5, 25–50.

Tranmer, M., and D. Steel. (1998). ''Using Census Data to Investigate the Causes of the Ecological Fallacy.'' *Environment and Planning A* 30, 817–31.

Unwin, D., and N. Wrigley. (1987). ''Control Point Distribution in Trend Surface Modelling Revisited: An Application of the Concept of Leverage.'' *Transactions, Institute of British Geographers* 12, 147–60.

Wakefield, J., N. Best, and L. Waller. (2000). ''Bayesian Approaches to Disease Mapping.'' In *Spatial Epidemiology: Methods and Applications*, 104–27, edited by P. Elliott, J. Wakefield, N. Best, and D. Briggs. Oxford, UK: Oxford University Press.

Wolpert, R., and K. Ickstadt. (1998). ''Poisson/Gamma Random Field Models for Spatial Statistics.'' *Biometrika* 85, 251–67.