On Maximum Likelihood Estimation of a Spatial Poisson Cluster Process

By JOHN M. CASTELLOE

SAS Institute Inc.

And DALE L. ZIMMERMAN

University of Iowa

Summary

The likelihood function of a bivariate Poisson cluster process is derived in closed form. Although maximum likelihood estimation in this setting is quite computationally intensive, an example demonstrates the successful convergence of a maximum likelihood estimation procedure for a small data set.

Short Running Title: MLE of a Spatial Poisson Cluster Process

Keywords: Bivariate normal distribution, Likelihood, Maximum likelihood estimation, Nelder-Mead simplex, Poission cluster process

1. Introduction

Spatial cluster processes have been developed in several different forms (Neyman and Scott, 1958, 1972; Strauss, 1975; Kelly and Ripley, 1976; Ripley, 1977; Diggle, 1975, 1978, 1983). Neyman and Scott (1972) discuss several applications, including the spatial distribution of larvae on an experimental field and the clustering of galaxies. In this article we consider bivariate Poisson cluster processes given by the following three postulates:

PCP1 Parent events occur according to a homogeneous Poisson process in $\Re^2$ with intensity ρ.

PCP2 Each parent j produces a random number S_j of offspring, realized independently and identically for each parent according to a Poisson distribution with mean ν.

PCP3 The positions of the offspring relative to their parents are independently and identically distributed in $\Re^2$ according to a common bivariate density function $h(\cdot;\boldsymbol{\theta})$ with mean $(0,0)$.

Note that such a process is stationary.

A realization of this Poisson cluster process is taken to be the collection of all offspring events in a region $A \in \Re^2$. Perhaps an extra "Poisson" should be inserted into the name since there are two Poisson distributions involved (one of the parent process and the other of the offspring counts), but it is omitted for simplicity.

Previous authors have declared the likelihood function for a Poisson cluster process intractable. Baudin (1981) derives an overly general expression for the likelihood of a Poisson cluster process (see $L(Y,k,U)$ at the top of p. 884), describes it as "obviously unrealistic" due to its "prohibitive complexity" and recommends basing estimation techniques instead on nearest neighbor distances, Ripley's K-function, and the spherical contact distribution (see Stoyan, D., Kendall, W. S. and Mecke, 1987). Stoyan (1992) concurs that maximum likelihood estimation is intractable and develops estimation methods based on Baudin's suggestions.

In this paper we derive the likelihood function, simplify its form considerably, and illustrate its computation and (local, at least) maximization for the special case of a bivariate normal offspring displacement distribution. It is not our intention to provide a method for immediate practical applicability; with current computing power, the maximum likelihood estimation solution is only feasible for toy examples such as ours. Rather, we present the likelihood function itself in closed form and encourage research on alternatives to traditional likelihood maximization. Previous expressions for the likelihood (such as that in Baudin, 1981) lack readiness for implementation and clarification of computational complexity. The solution we develop addresses both of these issues for a special case, achieving implementation-readiness and characterizing computational complexity.

2. The Likelihood Function

Consider the Poisson cluster process we have described, observed in a region $A \in \Re^2$. The parameters of this process are $\Phi = \{\rho, \nu, \boldsymbol{\theta}\}$. The observed data are

$$\mathbf{Y} = \{\text{locations of offspring in } A\} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)', \quad \text{where } \mathbf{y}_i = (y_{i1}, y_{i2})'$$

There is a substantial amount of latent data in this model formulation. The latent data are as follows:

$$\begin{aligned}
k &= \#(\text{parents in } A) \\
\boldsymbol{\mu} &= \{\text{parent locations}\} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k)', \quad \text{where } \boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2})' \\
\mathbf{Z} &= \{\text{“allocations”}\} = \begin{bmatrix} z_{11} & \cdots & z_{1k} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nk} \end{bmatrix}
\end{aligned}$$

$$\text{where } z_{ji} = \begin{cases} 1, & \text{if offspring } j \text{ belongs to parent } i \\ 0, & \text{otherwise} \end{cases}$$

The latent data component $\mathbf{Z}$ can also be referred to as “cluster memberships” or “parentage identifiers.” The cluster counts can be represented as the column sums of $\mathbf{Z}$:

$$\mathbf{s} = \{\text{cluster counts}\} = (S_1, \ldots, S_k)', \quad \text{where } S_i = \textstyle\sum_{j=1}^{n} z_{ji}$$

Note that the “sample size” (n, the total number of offspring) is random. However, we proceed as is standard in statistical inference for spatial point processes and condition on the observed sample size (see e.g. Ripley, 1977, 1981, 1988; Diggle, 1983; Baddeley and Møller, 1989). Also, we point out that our analysis does not take “boundary effects” into account. It is possible for parents outside the region of study A to produce offspring within A and for parents in A to have offspring outside A. Bias may be introduced when boundary effects are not accounted for, especially when many parent events occur near the boundary. Possible remedies such as addition of a “buffer zone” around A, toroidal edge corrections, or re-definition of the likelihood to incorporate truncation, are beyond the scope of consideration of this paper.

Let the observed-data likelihood be represented as $p(\mathbf{Y}|\Phi, n)$. Our main result is the expression of this likelihood in closed form, which we accomplish by obtaining the complete-data likelihood $p(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\mu}, k|\Phi, n)$ and integrating over the latent data. We can write the complete-data likelihood as:

$$p(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\mu}, k|\Phi, n) = p(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\mu}, k, \Phi, n)p(\mathbf{Z}|\boldsymbol{\mu}, \mathbf{s}, k, \Phi, n)p(\boldsymbol{\mu}, \mathbf{s}|k, \Phi, n)p(k|\Phi, n) \tag{2·1}$$

Each factor on the right-hand-side of (2·1) is derived in what follows. First note that

$$p(k|\Phi, n) = \frac{p(n|k, \Phi)p(k|\Phi)}{p(n|\Phi)} \tag{2·2}$$

Clearly $n|\{k, \Phi\} \sim \text{Poiss}(k\nu)$ and $k|\Phi \sim \text{Poiss}(\rho|A|)$, and so we have

$$p(n|k, \Phi) = \frac{(k\nu)^n \exp(-k\nu)}{n!}$$

and

$$p(k|\Phi) = \frac{(\rho|A|)^k \exp(-\rho|A|)}{k!}$$

To obtain the denominator of (2·2), we first establish the following lemma.

Lemma 1 *Suppose* $X \sim \text{Poiss}(\lambda)$ *and* n *is a positive integer. Let* $\Psi(t)$ *denote the moment generating function of* X*,*

$$\Psi(t) = \exp\left\{\lambda\left[\exp(t) - 1\right]\right\}$$

and let $\Psi^{(n)}(t_0)$ *denote the* n^{th} *derivative of* $\Psi(t)$ *with respect to* t *evaluated at* t_0*,*

$$\Psi^{(n)}(t_0) = \left.\frac{\mathrm{d}^n\Psi(t)}{\mathrm{d}t^n}\right|_{t_0}$$

Then

$$\Psi^{(n)}(t) = \sum_{j=1}^{n} a_{n,j}\lambda^j \exp\left\{\lambda\left[\exp(t) - 1\right] + jt\right\}$$

where

$$a_{n,j} = \begin{cases} 1, & \textit{if } j = 1 \textit{ or } j = n \\ j(a_{n-1,j}) + a_{n-1,j-1}, & \textit{otherwise} \end{cases} \tag{2·3}$$

Proof: see Appendix 1.

Now,

$$\begin{aligned} p(n|\Phi) &= \text{pr}\left(\sum_{i=1}^{k} S_i = n \,\middle|\, \Phi\right) \\ &= \sum_{q=0}^{\infty} \text{pr}\left(\sum_{i=1}^{q} S_i = n \,\middle|\, k=q,\Phi\right) \text{pr}\left(k=q|\Phi\right) \\ &= \left(\frac{\nu^n \exp\left\{-\left[1-\exp(-\nu)\right]\rho|A|\right\}}{n!}\right) \cdot \\ &\quad \sum_{q=0}^{\infty} q^n \left[\frac{\left[\rho|A|\exp(-\nu)\right]^q \exp\left\{-\rho|A|\exp(-\nu)\right\}}{q!}\right] \\ &= \left(\frac{\nu^n \exp\left\{-\left[1-\exp(-\nu)\right]\rho|A|\right\}}{n!}\right) E\left(X^n\right) \\ &\quad \text{where } X \sim \text{Poiss}\left(\rho|A|\exp(-\nu)\right) \\ &= \left(\frac{\nu^n \exp\left\{-\left[1-\exp(-\nu)\right]\rho|A|\right\}}{n!}\right) \sum_{j=1}^{n} a_{n,j}\left[\rho|A|\exp(-\nu)\right]^j \\ &\quad \text{where } a_{n,j} = \begin{cases} 1, & \text{if } j=1 \text{ or } j=n \\ j(a_{n-1,j}) + a_{n-1,j-1}, & \text{otherwise} \end{cases} \end{aligned}$$

where the last equality follows from Lemma 1.

Next observe that $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k)|\{k, \Phi, n\}$ are independent and distributed uniformly on A, $\mathbf{s}|\{k, \Phi, n\} \sim$ Mult $\left(n, \frac{1}{k}\mathbf{1}\right)$, and $\boldsymbol{\mu}$ and $\mathbf{s}$ are independent, so that

$$p(\boldsymbol{\mu}, \mathbf{s}|k, \Phi, n) = \frac{1}{|A|^k} \binom{n}{S_1 \cdots S_k} \frac{1}{k^n}$$

Given the offspring counts, all possible allocations satisfying the offspring counts are clearly equally likely (marginally, not taking into account the offspring locations). Denote the set of all possible allocations as $\Omega(\mathbf{s})$. The cardinality of $\Omega(\mathbf{s})$ is given by

$$\#(\Omega(\mathbf{s})) = \binom{n}{S_1 \cdots S_k}$$

and so

$$p(\mathbf{Z}|\boldsymbol{\mu}, \mathbf{s}, k, \Phi, n) = \frac{1}{\binom{n}{S_1 \cdots S_k}}$$

Finally, since $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ are independent, we have

$$p(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\mu}, k, \Phi, n) = \prod_{i=1}^{k}\prod_{j=1}^{n}\left[h(\mathbf{y}_j - \boldsymbol{\mu}_i; \boldsymbol{\theta})\right]^{Z_{ji}}$$

Combining terms and simplifying, we arrive at the form of the complete-data likelihood:

$$p(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\mu}, k|\Phi, n) = \frac{\frac{\rho^k}{k!}\exp\left[-k\nu - \rho|A|\exp(-\nu)\right]}{\sum_{j=1}^{n} a_{n,j}\left[\rho|A|\exp(-\nu)\right]^j} \prod_{i=1}^{k}\prod_{j=1}^{n}\left[h(\mathbf{y}_j - \boldsymbol{\mu}_i; \boldsymbol{\theta})\right]^{Z_{ji}} \tag{2·4}$$

where $\{a_{n,j}\}$ are defined as in (2·3). It can be integrated to produce the observed-data likelihood:

$$\begin{aligned} p(y|\Phi, n) &= \int \cdots \int p(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\mu}, k|\Phi, n)\, \mathrm{d}\boldsymbol{\mu}\, \mathrm{d}\mathbf{Z}\, \mathrm{d}k \\ &= \sum_{k=0}^{\infty} \sum_{\mathbf{s}\in\Lambda_n(k)} \sum_{\mathbf{Z}\in\Omega(\mathbf{s})} \iint_A \cdots \iint_A p(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\mu}, k|\Phi, n)\, \mathrm{d}\boldsymbol{\mu} \qquad (2·5) \\ &\quad \text{where } \Lambda_n(k) = \text{all possible values of } \mathbf{s} \text{ given } k \text{ and } n. \end{aligned}$$

An important special case of a Poisson cluster process results from taking the offspring displacement distribution to be bivariate normal with mean $(0, 0)$ and positive definite covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$. In this case, $\boldsymbol{\theta} = (\sigma_{11}, \sigma_{22}, \sigma_{12})'$ and the integral over $\boldsymbol{\mu}$ in (2·5) can be simplified to a closed-form, easily computed expression, as shown in the following lemma.

Lemma 2 *Suppose the offspring displacement distribution of a Poisson cluster process is bivariate normal with mean* $(0, 0)$ *and covariance matrix* $\boldsymbol{\Sigma} = (\sigma_{ij})$. *Then* $\iint_A \cdots \iint_A p(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\mu}, k|\Phi, n)\, \mathrm{d}\boldsymbol{\mu}$ *can be expressed as a product of terms of the form* $cP(\mathbf{X} \in A)$, *where* c *is a simple algebraic expression, and* $\mathbf{X}$ *has a bivariate normal distribution with known parameters.*

Proof: see Appendix 2.

3. Computational Aspects

In the bivariate normal case, the integral in (2·5) can be calculated numerically using readily available techniques. For example, if A is a square region, then the integral can be calculated by the function `pmvnorm` in S-Plus version 4.5 for Windows (Mathsoft, Inc.).

The summation over k in (2·5) can be justifiably truncated, for example at n (otherwise it would not make sense to model the data as a cluster process in the first place). This reduces the observed-data likelihood to the summation of a finite number of computable terms, which can thus be maximized (in principle, at least) by an optimization procedure such as the Nelder-Mead simplex method (see Nelder and Mead, 1965; Olsson and Nelson, 1975; and Press et al., 1988, section 10.4).

The summations over $\mathbf{s}$ and $\mathbf{Z}$, however, pose serious problems for even moderately-sized data sets. The number of terms grows astronomically with n. The cardinality of $\Lambda_n(k)$ is difficult to calculate, but it can be shown by a simple combinatorial argument that the number of ways to choose a collection of non-zero counts is $\binom{n-1}{k-1}$, and so $\#\left(\Lambda_n(k)\right) > \binom{n-1}{k-1}$. Expressions describing the exact number of such terms are unwieldy, but it will suffice to demonstrate that for $k = 2$ and n odd, we have

$$\sum_{\mathbf{s}\in\Lambda_n(2)} \sum_{\mathbf{Z}\in\Omega(\mathbf{s})} 1 = 2^{n-1}$$

This result follows from the fact that there are 2^n ways to allocate n offspring to 2 ordered clusters, and each such possibility has a redundant duplicate (for n odd, at least) since order should not be counted. Thus, it would appear that computation of even one likelihood value (using a truncation of k) for a moderately-sized data set is infeasible in current computing environments. Since the computation of the likelihood itself is the bottleneck in parameter estimation for this situation, we encourage research on techniques to work with truncated versions of the likelihood function.

4. Example

For illustrative purposes, we implement a Nelder-Mead simplex algorithm to find a local maximum of the observed-data likelihood (2·5) for a very simple pattern, shown in Figure 1. This pattern is a realization of a Poisson cluster process on the unit square with bivariate normal offspring displacement, conditional on two clusters centered at $\boldsymbol{\mu}_1 = (.33, .67)'$ and $\boldsymbol{\mu}_2 = (.67, .33)'$, 7 events in each cluster, and dispersal parameters $\sigma_{11} = \sigma_{22} = 0.0025, \rho_{12} = 0.8$. Clearly this is only a toy example, but it is useful to demonstrate successful convergence of the Nelder-Mead simplex algorithm in our context.

Following the guidelines of Olsson and Nelson (1975) for bounded parameters, we transform $\Phi = \{\rho, \nu, \sigma_{11}, \sigma_{22}, \sigma_{12}\}$ to $\{\log\rho, \log\nu, \log\sigma_{11}, \log\sigma_{22}, 2z(\rho_{12})\}$, where ρ_{12} is the correlation and $z(\cdot)$ is Fisher's z-transformation, for use in the actual algorithm. The starting simplex (shown in Table 1) is chosen to be very close to the true value of the parameter vector (to represent a "best case" scenario). Only the values $\{1, 2, 3, 4\}$ are used for k in each likelihood computation. The

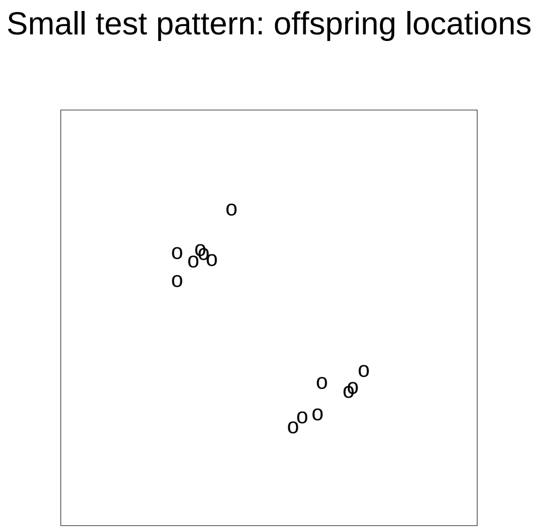

Figure 1: Small test pattern to demonstrate Nelder-Mead simplex maximum likelihood estimation procedure.

clusters in the test pattern Figure 1 were intentionally located far from the boundary so that the term $\prod_{i=1}^{k}(P(\mathbf{x}_i \in A))^{\oplus}$ in (2·5) (see Lemma 2) is extremely close to 1 and need not be computed at each iteration.

The Nelder-Mead simplex algorithm was run until the relative difference between likelihood values at successive iterations was less than 0.0001 (i.e., with a fractional tolerance of 0.0001). The simplex converged in 84 iterations and required 18.27 hours of computation time. Virtually all of the computation time was spent in calculation of the likelihood. Table 2 shows the resulting parameter estimates, along with true values and also values computed separately using the true allocations $\mathbf{Z}$ and the usual sample correlation coefficient and sample variance. The source code was written in C++ using matrix and random number libraries authored by Davies (1997). Simulations were run on a Hewlett Packard workstation with a 132 MHz CPU and 128 Mb RAM..

The Nelder-Mead simplex estimates of σ_{11}, σ_{22} and ρ_{12} are very close to the estimates obtained with knowledge of $\mathbf{Z}$. This is not too surprising since the pattern has clear structure, and the Nelder-Mead simplex starting values are close to the true values. Experimentation with other starting

ρ	ν	σ_{11}	σ_{22}	ρ_{12}
3	8	0.003	0.002	0.75
2	8	0.003	0.002	0.75
3	7	0.003	0.002	0.75
3	8	0.002	0.002	0.75
3	8	0.003	0.003	0.75
3	8	0.003	0.002	0.85

Table 1: Starting simplex used in Nelder-Mead simplex algorithm for small test pattern.

	ρ	ν	σ_{11}	σ_{22}	ρ_{12}
Nelder-Mead simplex	2.29651	7.65753	0.003154	0.002622	0.827185
Truth	(k =2)	($S_1 = S_2$ =7)	0.0025	0.0025	0.8
Estimates given **Z**			0.003150	0.002464	0.824253

Table 2: Parameter estimates from Nelder-Mead simplex algorithm implemented for small test pattern, along with true values and estimates computed using knowledge of **Z**.

simplex values suggests that the algorithm converges to many different local maxima, and many more iterations of the algorithm are usually required.

5. Conclusions and Future Directions

We have derived in closed form the likelihood function of the two-dimensional Poisson cluster process with biviariate normal dispersal. Although the ensuing maximum likelihood estimation method is not currently practical for reasonably sized data sets, the approach has potential utility in the future: the enumerative nature of the computation of a likelihood value makes it suitable for parallel computing. Future research can characterize the error induced by ignoring boundary effects and derive variance estimates for the MLE's. Experimentation with multiple starting values and assessment of the convexity of the likelihood surface can improve confidence in the results.

Appendix 1

Proof of Lemma 1

We prove by induction. First note that $\Psi^{(1)}(t) = \lambda \exp\{\lambda[\exp(t)-1]+t\}$, and so the lemma holds for $n = 1$. Suppose that the lemma holds for $n = m$, where $m \geq 1$. Then

$$\begin{aligned}
\Psi^{(m+1)}(t) &= \sum_{j=1}^{m} \left(a_{m,j}\lambda^j \exp\{\lambda[\exp(t)-1]+jt\}\right)[\lambda\exp(t)+j] \\
&= \sum_{j=1}^{m} \Big(a_{m,j}\lambda^{j+1} \exp\{\lambda[\exp(t)-1]+(j+1)t\} \;+ \\
&\qquad j a_{m,j}\lambda^{j} \exp\{\lambda[\exp(t)-1]+jt\}\Big) \\
&= a_{m,1}\lambda \exp\{\lambda[\exp(t)-1]+t\} \quad + \\
&\qquad \sum_{j=2}^{m} (a_{m,j-1}+j a_{m,j})\lambda^j \exp\{\lambda[\exp(t)-1]+jt\} \quad + \\
&\qquad a_{m,m}\lambda^{m+1} \exp\{\lambda[\exp(t)-1]+(m+1)t\} \\
&= a_{m+1,1}\lambda \exp\{\lambda[\exp(t)-1]+t\} \quad + \\
&\qquad \sum_{j=2}^{m} a_{m+1,j}\lambda^j \exp\{\lambda[\exp(t)-1]+jt\} \quad + \\
&\qquad a_{m+1,m+1}\lambda^{m+1} \exp\{\lambda[\exp(t)-1]+(m+1)t\} \\
&= \sum_{j=1}^{m+1} a_{m+1,j}\lambda^j \exp\{\lambda[\exp(t)-1]+jt\}
\end{aligned}$$

Thus the lemma is satisfied for any integer $n \geq 1$, and the proof is complete. □

Appendix 2

Proof of Lemma 2

First, we have

$$\begin{aligned}
&\iint_A \cdots \iint_A p(\mathbf{Y},\mathbf{Z},\boldsymbol{\mu},k|\Phi,n)\,\mathrm{d}\boldsymbol{\mu} \\
&\quad = \quad p(k|\Phi,n)p(\boldsymbol{\mu},\mathbf{s}|k,\Phi,n)p(\mathbf{Z}|\boldsymbol{\mu},\mathbf{s},k,\Phi,n)\iint_A \cdots \iint_A p(\mathbf{Y}|\mathbf{Z},\boldsymbol{\mu},k,\Phi,n)\,\mathrm{d}\boldsymbol{\mu}
\end{aligned}$$

since $p(k|\Phi,n)p(\boldsymbol{\mu},\mathbf{s}|k,\Phi,n)p(\mathbf{Z}|\boldsymbol{\mu},\mathbf{s},k,\Phi,n)$ is constant in $\boldsymbol{\mu}$ (see (2·1) and (2·4)). Define the notation

$$[X_i]^{\oplus} = \begin{cases} X_i, & \text{if } S_i > 0 \\ 1, & \text{otherwise} \end{cases} \qquad \text{(for any expression } X_i \text{ depending on } i\text{)}$$

and

$$[X_i]^{\circledcirc} = \begin{cases} X_i, & \text{if } S_i > 0 \\ 0, & \text{otherwise} \end{cases} \qquad \text{(for any expression } X_i \text{ depending on } i\text{)}$$

and an alternative indexing scheme for the elements of $\mathbf{Y}$:

$$\begin{aligned} \mathbf{y}_{i1},\ldots,\mathbf{y}_{ik} &\equiv (y_{i1;1},y_{i1;2})',\ldots,(y_{ik;1},y_{ik;2})' \\ &\equiv \text{locations of offspring from parent } i. \end{aligned}$$

The remaining integral can be re-written as follows:

$$\begin{aligned}
&\iint_A \cdots \iint_A p(\mathbf{Y}|\mathbf{Z},\boldsymbol{\mu},k,\Phi,n)\,\mathrm{d}\boldsymbol{\mu} \\
&= \iint_A \cdots \iint_A \prod_{i=1}^{k}\prod_{j=1}^{n} [h(\mathbf{y}_j - \boldsymbol{\mu}_i;\boldsymbol{\theta})]^{z_{ji}}\,\mathrm{d}\boldsymbol{\mu} \\
&= \iint_A \cdots \iint_A \prod_{i=1}^{k} \left[\prod_{j=1}^{S_i} h(\mathbf{y}_{ij} - \boldsymbol{\mu}_i;\boldsymbol{\theta})\right]^{\oplus} \mathrm{d}\boldsymbol{\mu} \\
&= \iint_A \cdots \iint_A \prod_{i=1}^{k} \left[\prod_{j=1}^{S_i} \left(\frac{1}{[2\pi(\sigma_{11}\sigma_{22}-\sigma_{12}^2)]^{\frac{1}{2}}} \exp\left\{\frac{-1}{2(\sigma_{11}\sigma_{22}-\sigma_{12}^2)}\cdot\right.\right.\right. \\
&\qquad \left.\left.\left.\left[\sigma_{22}\left(y_{ij;1}-\mu_{i1}\right)^2 + \sigma_{11}\left(y_{ij;2}-\mu_{i2}\right)^2 - 2\sigma_{12}\left(y_{ij;1}-\mu_{i1}\right)\left(y_{ij;2}-\mu_{i2}\right)\right]\right\}\right)\right]^{\oplus} \mathrm{d}\boldsymbol{\mu} \\
&= \frac{1}{[2\pi(\sigma_{11}\sigma_{22}-\sigma_{12}^2)]^{\frac{n}{2}}} \prod_{i=1}^{k} \left[\exp\left(\frac{-1}{2(\sigma_{11}\sigma_{22}-\sigma_{12}^2)}\cdot\right.\right. \\
&\qquad \left\{\sigma_{22}\left[\sum_{j=1}^{S_i} y_{ij;1}^2 - \frac{1}{S_i}\left(\sum_{j=1}^{S_i} y_{ij;1}\right)^2\right] + \sigma_{11}\left[\sum_{j=1}^{S_i} y_{ij;2}^2 - \frac{1}{S_i}\left(\sum_{j=1}^{S_i} y_{ij;2}\right)^2\right] - \right. \\
&\qquad \left.\left. 2\sigma_{12}\left[\sum_{j=1}^{S_i} y_{ij;1}y_{ij;2} - \frac{1}{S_i}\left(\sum_{j=1}^{S_i} y_{ij;1}\right)\left(\sum_{j=1}^{S_i} y_{ij;2}\right)\right]\right\}\right)\cdot \\
&\qquad \left[\frac{2\pi}{S_i}\left(\sigma_{11}\sigma_{22}-\sigma_{12}^2\right)^{\frac{1}{2}}\right] \iint_A \frac{1}{2\pi\left[\left(\frac{\sigma_{11}}{S_i}\right)\left(\frac{\sigma_{22}}{S_i}\right)-\left(\frac{\sigma_{12}}{S_i}\right)^2\right]^{\frac{1}{2}}} \exp\left(\frac{-1}{2(\sigma_{11}\sigma_{22}-\sigma_{12}^2)}\cdot\right.
\end{aligned}$$

$$\left\{\sigma_{22}\left[S_i\left(\mu_{i1}-\frac{1}{S_i}\sum_{j=1}^{S_i}y_{ij;1}\right)^2\right]+\sigma_{11}\left[S_i\left(\mu_{i2}-\frac{1}{S_i}\sum_{j=1}^{S_i}y_{ij;2}\right)^2\right]-\right.$$

$$\left.\left.\left.2\sigma_{12}\left[S_i\left(\mu_{i1}-\frac{1}{S_i}\sum_{j=1}^{S_i}y_{ij;1}\right)\left(\mu_{i2}-\frac{1}{S_i}\sum_{j=1}^{S_i}y_{ij;2}\right)\right]\right\}\right)\ \mathrm{d}\boldsymbol{\mu}\right]^{\oplus}$$

$$= \left\{\frac{1}{[2\pi(\sigma_{11}\sigma_{22}-\sigma_{12}^2)]^{\frac{n}{2}}}\right\}\exp\left[\frac{-1}{2(\sigma_{11}\sigma_{22}-\sigma_{12}^2)}\cdot\right.$$

$$\left.\left(\sigma_{22}\sum_{j=1}^{n}y_{j1}^2+\sigma_{11}\sum_{j=1}^{n}y_{j2}^2-2\sigma_{12}\sum_{j=1}^{n}y_{j1}y_{j2}\right)\right]\left\{\prod_{i=1}^{k}\left[\frac{2\pi}{S_i}\left(\sigma_{11}\sigma_{22}-\sigma_{12}^2\right)^{\frac{1}{2}}\right]^{\oplus}\right\}\cdot$$

$$\exp\left(\frac{-1}{2(\sigma_{11}\sigma_{22}-\sigma_{12}^2)}\sum_{i=1}^{k}\left\{\frac{1}{S_i}\left[\sigma_{22}\left(\sum_{j=1}^{S_i}y_{ij;1}\right)^2+\sigma_{11}\left(\sum_{j=1}^{S_i}y_{ij;2}\right)^2-\right.\right.\right.$$

$$\left.\left.\left.2\sigma_{12}\left(\sum_{j=1}^{S_i}y_{ij;1}\right)\left(\sum_{j=1}^{S_i}y_{ij;2}\right)\right]\right\}^{\circledcirc}\right)\left\{\prod_{i=1}^{k}[\mathrm{pr}(\mathbf{x}_i\in A)]^{\oplus}\right\}$$

where

$$\mathbf{x}_i \sim N\left(\left(\frac{1}{S_i}\sum_{j=1}^{S_i}y_{ij;1},\ \frac{1}{S_i}\sum_{j=1}^{S_i}y_{ij;2}\right)',\frac{1}{S_i}\boldsymbol{\Sigma}\right)\quad\square$$

References

Baddeley, A. and Møller, J., 1989: Nearest-Neighbor Markov Point Processes and Random Sets. *International Statistical Review* **57**, 89–121.

Baudin, M., 1981: Likelihood and Nearest-Neighbor Distance Properties of Multidimensional Poisson Cluster Processes. *Journal of Applied Probability* **18**, 879–888.

Davies, R., 1997: NEWMAT/NEWRAN Matrix and Random Number Generator C++ Libraries. New Zealand.

Diggle, P. J., 1975: Robust Density Estimation Using Distance Methods. *Biometrika* **62**, 39–48.

Diggle, P. J., 1978: On Parameter Estimation for Spatial Point Processes. *Journal of the Royal Statistical Society, Series B, Methodological* **40**, 178–181.

Diggle, P. J., 1983: *Statistical Analysis of Spatial Point Patterns.* Academic Press, New York.

Kelly, F. P. and Ripley, B. D., 1976: A Note on Strauss's Model for Clustering. *Biometrika* **63**, 357–360.

Nelder, J. A. and Mead, R., 1965: A Simplex Method for Function Minimization. *The Computer Journal* **7**, 308–313.

Neyman, J. and Scott, E. L., 1958: Statistical Approach to Problems of Cosmology. *Journal of the Royal Statistical Society, Series B, Methodological* **20**, 1–43.

Neyman, J. and Scott, E. L., 1972: Processes of Clustering and Applications. In: P.A.W. Lewis (ed.): *Stochastic Point Processes: Statistical Analysis, Theory, and Applications.* Wiley-Interscience, New York, 646–681.

Olsson, D. M. and Nelson, L. S., 1975: The Nelder-Mead Simplex Procedure for Function Minimization. *Technometrics* **7**, 45–51.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T., 1988: *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, New York.

Ripley, B. D., 1977: Modelling Spatial Patterns. *Journal of the Royal Statistical Society, Series B, Methodological* **39**, 172–212.

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

Ripley, B. D., 1988: *Statistical Inference for Spatial Processes.* Cambridge: Cambridge University Press.

Stoyan, D., 1992: Statistical Estimation of Model Parameters of Planar Neyman-Scott Cluster Processes. *Metrika* **39**, 67–74.

Stoyan, D., Kendall, W. S. and Mecke, J., 1987: *Stochastic Geometry and Its Applications.* Wiley, Chichester.

Strauss, D. J., 1975: A Model for Clustering. *Biometrika* **62**, 467–475.

Contact Information

JOHN M. CASTELLOE

SAS Institute Inc., SAS Campus Dr., Cary, North Carolina, 27513, U.S.A.

john.castelloe@sas.com

DALE L. ZIMMERMAN

Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa, 52240, U.S.A.

dzimmer@stat.uiowa.edu