Ann Inst Stat Math (2008) 60:627–649
DOI 10.1007/s10463-007-0116-6

Properties of residuals for spatial point processes

A. Baddeley · J. Møller · A. G. Pakes

Received: 6 March 2006 / Revised: 27 September 2006 / Published online: 1 June 2007

Abstract For any point process in $\mathbb{R}^d$ that has a Papangelou conditional intensity λ, we define a random measure of 'innovations' which has mean zero. When the point process model parameters are estimated from data, there is an analogous random measure of 'residuals'. We analyse properties of the innovations and residuals, including first and second moments, conditional independence, a martingale property, and lack of correlation. Some large sample asymptotics are studied. We derive the marginal distribution of smoothed residuals by solving a distributional equivalence.

Keywords Distributional equivalence · Georgii-Nguyen-Zessin formula · Gibbs point process · Set-indexed martingale · Papangelou conditional intensity · Pearson residuals · Scan statistic · Smoothed residual field

1 Introduction

The inspection of residuals is an important check on the appropriateness of a probability model fitted to data (Atkinson 1985). This paper defines residuals for spatial point processes, and describes their properties.

A. Baddeley (✉) · A. G. Pakes
School of Mathematics and Statistics M019, University of Western Australia,
35 Stirling Highway, Nedlands WA 6009, Australia
e-mail: adrian@maths.uwa.edu.au

A. Baddeley
CSIRO Mathematical and Information Sciences, Private Bag 5, Wembley WA 6913, Australia

J. Møller
Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G,
9220 Aalborg Ø, Denmark

For a point process in one-dimensional time, residual analysis is well understood. Let N_t be the associated counting process, and assume it has a conditional intensity λ_t given the history up to time t. Informally $\lambda_t = \mathbb{E}[\mathrm{d}N(t) \mid N_s, s < t]/\mathrm{d}t$. Define the 'innovation' process $I_t = N_t - \int^t \lambda_s \, \mathrm{d}s$; this is a martingale with zero mean (Karr 1985, Theorem 2.14, p. 60). When a point process model is fitted to observed data, the 'residual' process is $R_t = N_t - \int^t \widehat{\lambda}_s \, \mathrm{d}s$ where $\widehat{\lambda}_s$ is the conditional intensity of the fitted model, i.e. with parameters determined by fitting the model to the process $(N_t, t > 0)$. If the model is correct and the parameter estimate is accurate, then $\mathbb{E}[R_t] \approx 0$. This fact enables us to check the validity of a point process model fitted to data. Such techniques are now familiar in signal processing (Brillinger 1978, 1994; Brillinger and Segundo 1979; Lewis 1972) and survival analysis (Andersen et al. 1993; Fleming and Harrington 1991; Kalbfleisch and Prentice 1980). They have also been extended to space-time point processes, with important application to earthquake modelling (Ogata 1988; Ogata et al. 2003; Zhuang et al. 2005).

For spatial point processes, residual analysis is more difficult. The lack of a natural ordering in higher dimensions implies that there is no natural generalisation of the conditional intensity of a temporal process given the "past" or "history" up to time t. Instead, the appropriate counterpart for a spatial point process is the *Papangelou conditional intensity* $\lambda(u, \mathbf{X})$ (Papangelou 1974) which conditions on the outcome of the process at all spatial locations other than u. In Baddeley et al. (2005) we used the Papangelou conditional intensity to define residuals for finite point processes in $\mathbb{R}^2$, and showed that they have practical utility for checking point process models fitted to spatial point pattern data.

In this paper, we give a more general definition of the innovations and residuals for finite or infinite point processes in $\mathbb{R}^d$, and study their properties, including first and second moments, variance deflation, conditional independence, a set-indexed martingale property, lack of correlation, and marginal distributions. Section 2 gives background details about the Papangelou conditional intensity. Section 3 defines innovations for spatial point processes, and Sect. 4 obtains expressions for their variances. Section 5 defines residuals for a spatial point process model fitted to a point pattern, and Sect. 6 obtains expressions for their variances, including large-sample asymptotics. Section 7 discusses the distribution of residuals in a special case.

2 Conditional intensities

We consider the general setting of a locally finite point process $\mathbf{X}$ on $\mathbb{R}^d$ with no multiple points. Let $\mathcal{N}$ denote the set of all locally finite point configurations in $\mathbb{R}^d$, that is, subsets $\mathbf{x} \subset \mathbb{R}^d$ with $n(\mathbf{x}_B) < \infty$ for all bounded $B \subset \mathbb{R}^d$, where $n(\mathbf{x}_B)$ denotes the number of points in $\mathbf{x}_B = \mathbf{x} \cap B$, the restriction of $\mathbf{x}$ to B. We view $\mathbf{X}$ as a random variable with values in $\mathcal{N}$, such that $N(B) \equiv n(\mathbf{X}_B)$ is a finite random variable whenever $B \subset \mathbb{R}^d$ is a bounded Borel set (Daley and Vere-Jones 1988). For simplicity, we assume that

$$\mathbb{P}(u \in \mathbf{X}) = 0 \quad \text{for any fixed point } u \in \mathbb{R}^d, \tag{1}$$

which is satisfied e.g. if $\mathbf{X}$ is stationary.

Throughout this paper, $\mathbf{X}$ is assumed to be a Gibbs point process with Papangelou conditional intensity λ, that is,

$$\mathbb{E}\left[\sum_{u\in\mathbf{X}} h(u, \mathbf{X} \setminus \{u\})\right] = \mathbb{E}\left[\int_{\mathbb{R}^d} h(u, \mathbf{X})\lambda(u, \mathbf{X})\, \mathrm{d}u\right] \tag{2}$$

for all nonnegative measurable functions $h(u, \mathbf{x})$ on $\mathbb{R}^d \times \mathcal{N}$. The class of Gibbs processes includes all Cox processes, Markov point processes, many cluster processes, and all finite point processes which have a density with respect to the Poisson process. Equation (2) is called the *Georgii–Nguyen–Zessin* (*GNZ*) *formula* (Georgii, 1976; Nguyen and Zessin, 1979), and it is one way of defining the Papangelou conditional intensity. Indeed the Papangelou conditional intensity is uniquely characterised by (2) up to null-sets: if both λ_1 and λ_2 satisfy (2), then

$$\mathbb{P}(\lambda_1(u, \mathbf{X}) = \lambda_2(u, \mathbf{X}) \text{ for Lebesgue almost all } u \in \mathbb{R}^d) = 1.$$

Combining this with (1) we can and do make the assumption that

$$\lambda(u, \mathbf{x}) = \lambda(u, \mathbf{x} \setminus \{u\}) \quad \text{for all } u \in \mathbb{R}^d \text{ and } \mathbf{x} \in \mathcal{N}. \tag{3}$$

In a more rigorous treatment of measurability properties, (3) would be replaced by the requirement that λ be an *exvisible* process (Daley and Vere-Jones 1988, Chap. 13). However, the notational form (3) is a more accessible way to communicate the key definitions (e.g. of the weighted innovations in Sect. 3), and is useful in the algebraic calculations in Sect. 4.

In Baddeley et al. (2005) we adopted a simpler setting, in which $\mathbf{X}$ was assumed to be a finite point process with an hereditary density f. Suppose that $\mathbf{X}$ lives within a bounded Borel set $W \subset \mathbb{R}^d$, and $\mathbf{X}$ has a density f with respect to the unit rate Poisson process on W such that f is hereditary, i.e. $f(\mathbf{x}) > 0$ implies $f(\mathbf{x} \setminus \{u\}) > 0$ for all $\mathbf{x} \in \mathcal{N}_W$ and all $u \in \mathbf{x}$, where $\mathcal{N}_W$ is the set of finite point configurations contained in W. It is then straightforward to verify that the definition

$$\lambda(u, \mathbf{x}) = f(\mathbf{x} \cup \{u\})/f(\mathbf{x} \setminus \{u\}), \quad \text{for all } u \in W,\ \mathbf{x} \in \mathcal{N}_W \tag{4}$$

satisfies (2) and (3) (when the point process is empty outside W). Here and throughout the paper we interpret $0/0 = 0$ when considering ratios of densities.

In applications we often consider the special case of a Markov point process (Ripley and Kelly 1977) of finite interaction range $R < \infty$. This is a Gibbs process whose conditional intensity $\lambda(u, \mathbf{x})$ depends on $\mathbf{x}$ only through $\mathbf{x} \cap b(u, R)$,

where $b(u, R)$ is the closed ball in $\mathbb{R}^d$ with centre u and radius R. Equivalently (Georgii 1976; Preston 1976; Ruelle 1969)

$$\lambda(u, \mathbf{x}) = \exp\left(\sum_{\mathbf{y} \subseteq \mathbf{x}} V(\mathbf{y} \cup \{u\})\right) \quad \text{whenever } u \notin \mathbf{x} \tag{5}$$

where the 'potential' V is an extended-real-valued function $V(\mathbf{x}) \in [-\infty, \infty)$, defined for $\mathbf{x} \in \mathcal{N}$, with the property that

$$V(\mathbf{x}) = 0 \quad \text{whenever } \mathbf{x} \text{ contains two points } u, v \text{ with distance } \|u - v\| > R.$$

This local Markov property implies a spatial Markov property. For $B \subset \mathbb{R}^d$, let ∂B be its R-close neighbourhood, i.e. the set of all points in $B^c = \mathbb{R}^d \setminus B$ within distance R from some point in B. Then for bounded Borel sets $B \subset \mathbb{R}^d$, $\mathbf{X}_B$ conditional on $\mathbf{X}_{\partial B}$ is independent of $\mathbf{X}_{B^c \setminus \partial B}$, with conditional density

$$f_B(\mathbf{x}|\mathbf{x}_{\partial B}) \propto \exp\left(\sum_{\mathbf{y} \subseteq \mathbf{x}} V(\mathbf{y} \cup \mathbf{x}_{\partial B})\right), \quad \text{for all } \mathbf{x} \in \mathcal{N}_B, \ \mathbf{x}_{\partial B} \in \mathcal{N}_{\partial B} \tag{6}$$

with respect to the unit rate Poisson process on B, where the normalizing constant on the right side in (6) may depend on $\mathbf{X}_{\partial B}$. Combining (4) and (6) we see that the Papangelou conditional intensity $\lambda(\cdot, \cdot|\mathbf{x}_{\partial B})$ of the conditional point process $\mathbf{X}_B|\mathbf{X}_{\partial B} = \mathbf{x}_{\partial B}$ agrees with the conditional intensity of $\mathbf{X}$, meaning that we can take

$$\lambda(u, \mathbf{x}|\mathbf{x}_{\partial B}) = \lambda(u, \mathbf{x} \cup \mathbf{x}_{\partial B}), \quad \text{for all } u \in B, \ \mathbf{x} \in \mathcal{N}_B, \ \mathbf{x}_{\partial B} \in \mathcal{N}_{\partial B}. \tag{7}$$

Examples of Markov and non-Markov Gibbs point process models and their conditional intensities are presented in Baddeley et al. (2005), Møller and Waagepetersen (2003, 2007).

3 Innovations

This section defines innovations for (finite as well as infinite) spatial point processes $\mathbf{X}$ having Papangelou conditional intensity λ. The process is observed within a bounded window $W \subset \mathbb{R}^d$, with positive volume $|W|$. We assume that either

(i) $\mathbf{X}$ is a Gibbs process in W; or
(ii) $\mathbf{X}$ is a Markov point process in $\mathbb{R}^d$ with interaction range $R < \infty$, but is observed only within $W \subset \mathbb{R}^d$.

In case (ii), we account for edge effects by considering inference based on the conditional process $\mathbf{X}_V|\mathbf{X}_{\partial V}$, where $V = W \setminus \partial(W^c)$. Since $\partial V = \partial(W^c)$,

the point process $\mathbf{X}_V$ given $\mathbf{X}_{\partial V}$ is independent of $\mathbf{X}_{W^c}$ and has Papangelou conditional intensity $\lambda(u, \mathbf{X}_W)$ for $u \in V$, cf. (7).

So as to avoid duplicating our results for the two cases (i) and (ii), we state them in terms of the sets A and ∂A defined as follows. In case (i), define $A = W$ and $\partial A = \emptyset$. In case (ii), let $A = V$ and $\partial A = \partial V = \partial(W^c)$.

The GNZ formula corresponding to the conditional point process $\mathbf{X}_A|\mathbf{X}_{\partial A}$ is

$$\mathbb{E}\left[\sum_{u\in\mathbf{X}_A} h(u, \mathbf{X}_W \setminus \{u\}) \,\middle|\, \mathbf{X}_{\partial A}\right] = \mathbb{E}\left[\int_A h(u, \mathbf{X}_W)\lambda(u, \mathbf{X}_W)\, \mathrm{d}u \,\middle|\, \mathbf{X}_{\partial A}\right] \tag{8}$$

for nonnegative measurable functions h. Equation (8) rather than (2) is the relevant form of the GNZ formula when inference is performed on the conditional point process $\mathbf{X}_A|\mathbf{X}_{\partial A}$. In case (i), $\mathbf{X}_A|\mathbf{X}_{\partial A}$ is equivalent to the "marginal" process $\mathbf{X}_W$.

We shall exploit (8) intensively when studying the properties of innovations and residuals. For illustrative purposes we sometimes consider a Poisson process with intensity function $\lambda(u, \mathbf{x}) = \lambda(u)$, in which case we take $R = 0$ so that $A = W$ and $\partial A = \emptyset$, meaning that $\mathbf{X}_A|\mathbf{X}_{\partial A} \equiv \mathbf{X}_W$ and the expectations in (8) are with respect to the point process restricted to W.

In the sequel we always implicitly assume that means, variances, etc. exist whenever needed. For example, when we apply (8) we assume that the (conditional) expectations are finite. Finally, B always denotes a generic Borel set contained in A.

3.1 Innovations

The *h-weighted innovation* is the signed random measure defined by

$$I_h(B) = \sum_{u\in\mathbf{X}_B} h(u, \mathbf{X}_W \setminus \{u\}) - \int_B h(u, \mathbf{X}_W)\lambda(u, \mathbf{X}_W)\, \mathrm{d}u. \tag{9}$$

We allow infinite values of $h(u, \mathbf{X}_W)$ at points $u \notin \mathbf{X}_W$, setting $h(u, \mathbf{X}_W)$ $\lambda(u, \mathbf{X}_W) = 0$ if $\lambda(u, \mathbf{X}_W) = 0$. Baddeley et al. (2005) study in particular the *raw, inverse-λ,* and *Pearson* innovations given by $h = 1$, $1/\lambda$ and $1/\sqrt{\lambda}$ respectively. That is,

$$I(B) \equiv I_1(B) = N(B) - \int_B \lambda(u, \mathbf{X}_W)\, \mathrm{d}u \tag{10}$$

$$I_{1/\lambda}(B) = \sum_{u\in\mathbf{X}_B} \frac{1}{\lambda(u, \mathbf{X}_W)} - \int_B \mathbf{1}[\lambda(u, \mathbf{X}_W) > 0]\, \mathrm{d}u \tag{11}$$

$$I_{1/\sqrt{\lambda}}(B) = \sum_{u\in\mathbf{X}_B} \frac{1}{\sqrt{\lambda(u, \mathbf{X}_W)}} - \int_B \sqrt{\lambda(u, \mathbf{X}_W)}\, \mathrm{d}u \tag{12}$$

where $\mathbf{1}[\cdot]$ denotes the indicator function. By Eq. (8),

$$\mathbb{E}[I_h(B)|\mathbf{X}_{\partial A}] = 0 \tag{13}$$

and so the unconditional mean $\mathbb{E}[I_h(B)]$ is zero as well; as noted above, we find (13) to be the more relevant property when inference is based on $\mathbf{X}_A|\mathbf{X}_{\partial A}$.

3.2 Some martingale and independence properties

The definition (10) of the raw innovation is closely analogous to that for temporal processes, i.e. the martingale obtained by subtracting the compensator from the counting process, except for the use of the Papangelou conditional intensity in place of the conditional intensity given the past history. We now show that our raw innovation is indeed a set-indexed martingale.

Proposition 1 *If A_n is increasing in $\mathbb{R}^d$ (i.e. $A_n \subset A_{n+1}$, $n = 1,2,\ldots$), then $I_n = I(A_n)$ is a martingale.*

Proof To stress that the innovation is defined conditionally on $\mathbf{X}_{A^c}$ (or equivalently, conditionally on $\mathbf{X}_{\partial A}$) we write $I(A|\mathbf{X}_{A^c})$ for $I(A)$. Since $\lambda(u,\mathbf{X}) = \lambda(u,\mathbf{X}_{A\cup\partial A})$ if $u \in A$,

$$I(A|\mathbf{X}_{A^c}) = N(A) - \int_A \lambda(u,\mathbf{X})\, \mathrm{d}u$$

where by the GNZ formula (8)

$$\mathbb{E}[I(A|\mathbf{X}_{A^c})|\mathbf{X}_{A^c}] = 0. \tag{14}$$

Now

$$\begin{aligned}\mathbb{E}[I_{n+1}|\mathbf{X}_{A_n}] &= \mathbb{E}\left[\mathbb{E}\left(I_{n+1}\middle|\mathbf{X}_{A_n},\mathbf{X}_{A_{n+1}^c}\right)\middle|\mathbf{X}_{A_n}\right]\\ &= \mathbb{E}\left[I_n + \mathbb{E}\left(I(A_{n+1}\setminus A_n|\mathbf{X}_{(A_{n+1}\setminus A_n)^c})\middle|\mathbf{X}_{A_n},\mathbf{X}_{A_{n+1}^c}\right)\middle|\mathbf{X}_{A_n}\right]\end{aligned}$$

and so by (14), since $(A_{n+1}\setminus A_n)^c = A_n \cup A_{n+1}^c$, the inner expectation on the last line is zero, which implies the martingale property $\mathbb{E}[I_{n+1}|I_n, I_{n-1},\ldots] = I_n$.

Lemma 1 *Assume case (ii) so that $\mathbf{X}$ is Markov. Suppose $h(u,\mathbf{x}_W)$ is a nonnegative measurable function such that $h(u,\mathbf{X}_W) = h(u,\mathbf{X}_W \cap b(u,R))$ for all $u \in A$. Then $I_h(B)$ depends on $\mathbf{X}_W$ only through $\mathbf{X}_{B\cup\partial B}$, and for any Borel set $C \subseteq \mathbb{R}^d$ such that $C \supseteq \partial B$,*

$$\mathbb{E}\left[I_h(B)|\mathbf{X}_C\right] = 0.$$

Proof The first property follows from the definition of innovations and the local Markov property, while the second property follows from a version of the GNZ formula [viz. (8) with A replaced by B] and the global Markov property (see Sect. 2).

Proposition 2 *Assume case (ii) so that* $\mathbf{X}$ *is Markov. Suppose* $B_1, B_2 \subset A$ *are Borel sets at least a distance* R *apart, i.e.* $\|u - v\| > R$ *for any* $u \in B_1$ *and* $v \in B_2$, *and that* $h(u, \mathbf{x}_W)$ *is a nonnegative measurable function such that* $h(u, \mathbf{X}_W) = h(u, \mathbf{X}_W \cap b(u, R))$ *for all* $u \in A$. *Let* $C \subseteq \mathbb{R}^d$ *be a Borel set such that* $C \supseteq \partial(B_1 \cup B_2)$. *Then* $I_h(B_1)$ *and* $I_h(B_2)$ *are uncorrelated, and are conditionally independent given* $\mathbf{X}_C$.

Proof Follows immediately from Lemma 1, the spatial Markov property (see Sect. 2) and basic properties of conditional moments.

As a result of these propositions, one may expect a strong law of large numbers and a central limit theorem to hold for the raw, inverse-λ and Pearson innovations as the sampling window W expands. However, we do not investigate this in the present paper.

4 Variances of innovations

4.1 General variance formulae

Formulae for the variances of innovations and residuals can be obtained using the *second-order* Papangelou conditional intensity $\lambda(u, v, \mathbf{X})$ of the point process $\mathbf{X}$. This is a random function satisfying the second-order counterpart of the GNZ formula

$$\mathbb{E}\left[\sum_{u,v\in\mathbf{X}:\, u\neq v} h(u, v, \mathbf{X} \setminus \{u, v\})\right] = \mathbb{E}\left[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h(u, v, \mathbf{X})\lambda(u, v, \mathbf{X})\, \mathrm{d}u\, \mathrm{d}v\right] \quad (15)$$

for any nonnegative measurable function $h(u, v, \mathbf{x})$ on $\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{N}$. Note that $\lambda(u, v, \mathbf{X})$ is symmetric in u and v [for Lebesgue almost all (u, v)]. It follows immediately from the first-order GNZ formula (2) that

$$\lambda(u, v, \mathbf{x}) = \lambda(u, \mathbf{x} \setminus \{v\})\lambda(v, \mathbf{x} \cup \{u\}), \quad u, v \in \mathbb{R}^d,\ \mathbf{x} \in \mathcal{N} \quad (16)$$

up to almost sure, almost everywhere equivalence. If $\mathbf{X}$ lives within W and has density f with respect to the unit rate Poisson process, we can take

$$\lambda(u, v, \mathbf{x}) = f(\mathbf{x} \cup \{u, v\})/f(\mathbf{x} \setminus \{u, v\}).$$

Below we use the fact that a Markov process with pairwise interaction only [i.e. whose potential $V(\mathbf{x})$ is zero whenever $n(\mathbf{x}) > 2$] has

$$\lambda(u, v, \mathbf{x}) = \lambda(u, \mathbf{x} \setminus \{v\})\ \lambda(v, \mathbf{x} \setminus \{u\})\ c(u, v) \quad (17)$$

where $\log c(u,v) = V(\{u,v\})$ is the second order potential.

By the same arguments as in Sect. 2, $\lambda(u,v,\mathbf{X}) = \lambda(u,v,\mathbf{X}_W)$ when u,v are points in A, and (16) also specifies the second-order Papangelou conditional intensity of the conditional process $\mathbf{X}_A$ given $\mathbf{X}_{\partial A}$. Moreover, the second-order GNZ formula for this conditional point process is

$$\mathbb{E}\left[\sum_{u,v\in\mathbf{X}_A:u\neq v} h(u,v,\mathbf{X}_W\setminus\{u,v\})\,\middle|\,\mathbf{X}_{\partial A}\right] = \mathbb{E}\left[\int_A\int_A h(u,v,\mathbf{X}_W)\lambda(u,v,\mathbf{X}_W)\,\mathrm{d}u\,\mathrm{d}v\,\middle|\,\mathbf{X}_{\partial A}\right]. \tag{18}$$

Proposition 3 *For any non-negative measurable function h,*

$$\mathrm{var}\left[\sum_{u\in\mathbf{X}_A} h(u,\mathbf{X}_W\setminus\{u\})\,\middle|\,\mathbf{X}_{\partial A}\right] = \int_A \mathbb{E}\left[h(u,\mathbf{X}_W)^2\,\lambda(u,\mathbf{X}_W)|\mathbf{X}_{\partial A}\right]\,\mathrm{d}u + \int_A\int_A T(u,v)\,\mathrm{d}u\,\mathrm{d}v \tag{19}$$

where

$$T(u,v) = \mathbb{E}\left[h(u,\mathbf{X}_W\cup\{v\})h(v,\mathbf{X}_W\cup\{u\})\lambda(u,v,\mathbf{X})|\mathbf{X}_{\partial A}\right] - \mathbb{E}\left[h(u,\mathbf{X}_W)\lambda(u,\mathbf{X}_W)|\mathbf{X}_{\partial A}\right]\mathbb{E}\left[h(v,\mathbf{X}_W)\lambda(v,\mathbf{X}_W)|\mathbf{X}_{\partial A}\right]. \tag{20}$$

Proof Follows immediately by expanding the square of the sum on the left side of (19) as a double sum, and using (8) and (18).

For example, for a Poisson process with intensity function $\lambda(u)$, (19) reduces to

$$\mathrm{var}\left[\sum_{u\in\mathbf{X}_W} h(u,\mathbf{X}_W\setminus\{u\})\right] = \int_W\int_W \mathbb{E}\left[h(u,\mathbf{X}_W\cup\{v\})h(v,\mathbf{X}_W\cup\{u\})\right]\lambda(u)\lambda(v)\,\mathrm{d}u\,\mathrm{d}v + \int_W \mathbb{E}\left[h(u,\mathbf{X}_W)^2\right]\lambda(u)\,\mathrm{d}u - \left(\int_W \mathbb{E}\left[h(u,\mathbf{X}_W)\right]\lambda(u)\,\mathrm{d}u\right)^2.$$

In the special case $h(u,\mathbf{x}_W) = h(u)$, this further reduces to

$$\mathrm{var}\left[\sum_{u\in\mathbf{X}_W} h(u)\right] = \int_W h(u)^2\lambda(u)\,\mathrm{d}u \tag{21}$$

as expected by the independence properties of the Poisson process.

Lemma 2 *For nonnegative measurable functions h and g,*

$$\begin{aligned}
&\operatorname{var}\left[\sum_{u\in\mathbf{X}_A} h(u,\mathbf{X}_W\setminus\{u\}) - \int_A g(u,\mathbf{X}_W)\,\mathrm{d}u \,\middle|\, \mathbf{X}_{\partial A}\right] \\
&\quad= \int_A \mathbb{E}\left[h(u,\mathbf{X}_W)^2\,\lambda(u,\mathbf{X}_W)|\mathbf{X}_{\partial A}\right]\,\mathrm{d}u \\
&\qquad+ \int_A\int_A \operatorname{cov}[\mathrm{g}(\mathrm{u},\mathbf{X}_{\mathrm{W}}),\mathrm{g}(\mathrm{v},\mathbf{X}_{\mathrm{W}})|\mathbf{X}_{\partial\mathrm{A}}]\,\mathrm{du}\,\mathrm{dv} \\
&\qquad+ \int_A\int_A T(u,v)\,\mathrm{d}u\,\mathrm{d}v - 2\int_A\int_A M(u,v)\,\mathrm{d}u\,\mathrm{d}v
\end{aligned} \tag{22}$$

where

$$\begin{aligned}
M(u,v) = {}& \mathbb{E}[h(u,\mathbf{X}_W)g(v,\mathbf{X}_W\cup\{u\})\lambda(u,\mathbf{X}_W)|\mathbf{X}_{\partial A}] \\
&- \mathbb{E}[h(u,\mathbf{X}_W)\lambda(u,\mathbf{X}_W)|\mathbf{X}_{\partial A}]\mathbb{E}[g(v,\mathbf{X}_W)|\mathbf{X}_{\partial A}].
\end{aligned}$$

Proof Using standard properties of variances, we expand the left side of (22) as

$$\begin{aligned}
&\operatorname{var}\left[\sum_{u\in\mathbf{X}_A} h(u,\mathbf{X}_W\setminus\{u\}) \,\middle|\, \mathbf{X}_{\partial A}\right] + \operatorname{var}\left[\int_A g(u,\mathbf{X}_W)\,\mathrm{d}u \,\middle|\, \mathbf{X}_{\partial A}\right] \\
&\quad- 2\operatorname{cov}\left(\sum_{\mathrm{u}\in\mathbf{X}_{\mathrm{A}}} \mathrm{h}(\mathrm{u},\mathbf{X}_{\mathrm{W}}\setminus\{\mathrm{u}\}), \int_{\mathrm{A}} \mathrm{g}(\mathrm{u},\mathbf{X}_{\mathrm{W}})\,\mathrm{du} \,\middle|\, \mathbf{X}_{\partial\mathrm{A}}\right) \\
&= \operatorname{var}\left[\sum_{u\in\mathbf{X}_A} h(u,\mathbf{X}_W\setminus\{u\}) \,\middle|\, \mathbf{X}_{\partial A}\right] \\
&\quad+ \int_A\int_A \operatorname{cov}\left(\mathrm{g}(\mathrm{u},\mathbf{X}_{\mathrm{W}}),\mathrm{g}(\mathrm{v},\mathbf{X}_{\mathrm{W}})|\mathbf{X}_{\partial\mathrm{A}}\right)\,\mathrm{du}\,\mathrm{dv} \\
&\quad- 2\mathbb{E}\left[\sum_{u\in\mathbf{X}_A} h(u,\mathbf{X}_W\setminus\{u\}) \int_A g(u,\mathbf{X}_W)\,\mathrm{d}u \,\middle|\, \mathbf{X}_{\partial A}\right] \\
&\quad+ 2\mathbb{E}\left[\sum_{u\in\mathbf{X}_A} h(u,\mathbf{X}_W\setminus\{u\}) \,\middle|\, \mathbf{X}_{\partial A}\right] \mathbb{E}\left[\int_A g(u,\mathbf{X}_W)\,\mathrm{d}u \,\middle|\, \mathbf{X}_{\partial A}\right].
\end{aligned} \tag{23}$$

Denote the four terms on the right-hand side of (23) by V, C, E_1 and E_2 respectively. The variance term V is now expanded using Proposition 3. The

first expectation E_1 is converted to an integral using (8). The second expectation E_2 is evaluated by putting

$$k(v,\mathbf{x}) = h(v,\mathbf{x}) \int_A g(u,\mathbf{x}\cup\{v\})\,\mathrm{d}u, \quad v \notin \mathbf{x},\ \mathbf{x} \in \mathcal{N}_W,$$

so that

$$k(v,\mathbf{x}\setminus\{v\}) = h(v,\mathbf{x}\setminus\{v\}) \int_A g(u,\mathbf{x})\,\mathrm{d}u, \quad v \in \mathbf{x},\ \mathbf{x} \in \mathcal{N}_W.$$

Applying (8) gives

$$\begin{aligned}
E_2 &= \mathbb{E}\left[\sum_{u\in\mathbf{X}_A} k(u,\mathbf{X}_W\setminus\{u\}) \,\middle|\, \mathbf{X}_{\partial A}\right] \\
&= \int_A \mathbb{E}\left[k(u,\mathbf{X}_W)\lambda(u,\mathbf{X}_W)|\mathbf{X}_{\partial A}\right]\,\mathrm{d}u \\
&= \int_A \mathbb{E}\left[h(u,\mathbf{X}_W)\lambda(u,\mathbf{X}_W)\int_A g(v,\mathbf{X}_W\cup\{u\})\,\mathrm{d}v \,\middle|\, \mathbf{X}_{\partial A}\right]\,\mathrm{d}u \\
&= \int_A\int_A \mathbb{E}\left[h(u,\mathbf{X}_W)g(v,\mathbf{X}_W\cup\{u\})\lambda(u,\mathbf{X}_W) \,\middle|\, \mathbf{X}_{\partial A}\right]\,\mathrm{d}u\,\mathrm{d}v.
\end{aligned}$$

Rearrangement yields the result (22).

Proposition 4 *The variance of the h-weighted innovation is*

$$\begin{aligned}
\mathrm{var}\,[I_h(B)|\mathbf{X}_{\partial A}] = &\int_B \mathbb{E}\left[h(u,\mathbf{X}_W)^2\,\lambda(u,\mathbf{X}_W)|\mathbf{X}_{\partial A}\right]\,\mathrm{d}u \\
&+ \int_B\int_B \mathbb{E}\left[S(u,v,\mathbf{X}_W)|\mathbf{X}_{\partial A}\right]\,\mathrm{d}u\,\mathrm{d}v
\end{aligned} \tag{24}$$

for Borel sets $B \subseteq A$, where

$$\begin{aligned}
S(u,v,\mathbf{x}) = &\ \lambda(u,\mathbf{x})\lambda(v,\mathbf{x})h(u,\mathbf{x})h(v,\mathbf{x}) \\
&+ \lambda(u,v,\mathbf{x})h(v,\mathbf{x}\cup\{u\})\left[h(u,\mathbf{x}\cup\{v\}) - 2h(u,\mathbf{x})\right].
\end{aligned} \tag{25}$$

Proof In Eq. (22), replace $h(u,\mathbf{x})$ by $h(u,\mathbf{x})\ \mathbf{1}\{u \in B\}$ and substitute $g(u,\mathbf{x}) = \lambda(u,\mathbf{x})h(u,\mathbf{x})\mathbf{1}\{u \in B\}$.

As a corollary, by combining (13) and (24) we obtain

$$\mathrm{var}\,[I_h(B)] = \int_B \mathbb{E}\left[h(u,\mathbf{X}_W)^2\,\lambda(u,\mathbf{X})\right]\,\mathrm{d}u\,\mathrm{d}v. \tag{26}$$

Again, the conditional variance (24) is a more relevant result for us than (26) when doing inference conditional on $\mathbf{X}_{\partial A}$.

4.2 Variance of innovations in particular cases

4.2.1 Raw innovations

For the raw innovations, taking $h \equiv 1$, Eq. (25) reduces to

$$S(u, v, \mathbf{x}) = \lambda(u, \mathbf{x})\lambda(v, \mathbf{x}) - \lambda(u, v, \mathbf{x})$$

so that (24) becomes

$$\begin{aligned} \operatorname{var}[I(B)|\mathbf{X}_{\partial A}] = &\int_B \mathbb{E}[\lambda(u, \mathbf{X}_W)|\mathbf{X}_{\partial A}]\ \mathrm{d}u \\ &+ \int_B \int_B \mathbb{E}[\lambda(u, \mathbf{X}_W)\lambda(v, \mathbf{X}_W) - \lambda(u, v, \mathbf{X}_W)|\mathbf{X}_{\partial A}]\ \mathrm{d}u\ \mathrm{d}v. \end{aligned} \tag{27}$$

For a Poisson process with intensity function $\lambda(u)$, the expression S in (25) is identically zero, and (27) reduces to (21) with $h = 1$.

4.2.2 Inverse-lambda innovations

Suppose for simplicity that $\lambda(\cdot, \cdot) > 0$. Applying (19) to $h(u, \mathbf{x}) = 1/\lambda(u, \mathbf{x})$, we find that

$$\begin{aligned} \operatorname{var}\left[I_{1/\lambda}(B)\middle|\mathbf{X}_{\partial A}\right] = &\int_B \int_B \mathbb{E}\left[\frac{\lambda(u, v, \mathbf{X}_W)}{\lambda(u, \mathbf{X}_W \cup \{v\})\lambda(v, \mathbf{X}_W \cup \{u\})}\middle|\mathbf{X}_{\partial A}\right] \mathrm{d}u\ \mathrm{d}v \\ &+ \int_B \mathbb{E}\left[\frac{1}{\lambda(u, \mathbf{X}_W)}\middle|\mathbf{X}_{\partial A}\right] \mathrm{d}u - |B|^2. \end{aligned} \tag{28}$$

For example, consider a pairwise interaction process with a finite potential [i.e. $\lambda(\cdot, \cdot) > 0$ and $c(\cdot, \cdot) > 0$). Then (17) and (28] yield

$$\operatorname{var}\left[I_{1/\lambda}(B)|\mathbf{X}_{\partial A}\right] = \int_B \int_B \frac{1}{c(u, v)}\ \mathrm{d}u\ \mathrm{d}v + \int_B \mathbb{E}\left[\frac{1}{\lambda(u, \mathbf{X}_W)}\middle|\mathbf{X}_{\partial A}\right] \mathrm{d}u - |B|^2. \tag{29}$$

This was derived in Stoyan and Grabarnik (1991) in the unconditional case, when the first and second order potentials are translation invariant [$V(\{u\}) \equiv \beta$, $c(u, v) = c(u-v)$]. For a Poisson process with intensity function $\lambda(\cdot) > 0$, Eq. (29) reduces to (21) with $h(u) = 1/\lambda(u)$. The general case, where λ may have zeroes, is derived similarly.

4.2.3 Pearson innovations

For the Pearson innovations (11), we have $h(u,\mathbf{x}) = 1/\sqrt{\lambda(u,\mathbf{x})}$ so that $h(u,\mathbf{x})^2 \lambda(u,\mathbf{x}) = \mathbf{1}[\lambda(u,\mathbf{x}) > 0]$. Hence by (24)

$$\begin{aligned} \mathrm{var}\left[I_{1/\sqrt{\lambda}}(B)|\mathbf{X}_{\partial A}\right] &= \int_B \mathbb{P}(\lambda(u,\mathbf{X}_W) > 0|\mathbf{X}_{\partial A})\, \mathrm{d}u \\ &\quad + \int_B \int_B \mathbb{E}\left[S(u,v,\mathbf{X})\right]\, \mathrm{d}u\, \mathrm{d}v \end{aligned} \tag{30}$$

where (25) is now

$$\begin{aligned} S(u,v,\mathbf{x}) &= \sqrt{\lambda(u,\mathbf{x})}\sqrt{\lambda(v,\mathbf{x})} \\ &\quad + \frac{\lambda(u,v,\mathbf{x})}{\sqrt{\lambda(v,x\cup\{u\})}}\left[\frac{1}{\sqrt{\lambda(u,\mathbf{x}\cup\{v\})}} - \frac{2}{\sqrt{\lambda(u,\mathbf{x})}}\right]. \end{aligned} \tag{31}$$

For a Poisson process with intensity function $\lambda(u)$, S is identically zero and (30) reduces to

$$\mathrm{var}\left[I_{1/\sqrt{\lambda}}(B)\right] = \mathrm{var}\left[\sum_{u\in\mathbf{X}} \frac{1}{\sqrt{\lambda(u)}}\right] = \int_B \mathbf{1}[\lambda(u) > 0]\, \mathrm{d}u \tag{32}$$

in agreement with (21).

For a Markov point process with pairwise interaction only, (31) becomes

$$\begin{aligned} S(u,v,\mathbf{x}) &= \sqrt{\lambda(u,\mathbf{x})}\sqrt{\lambda(v,\mathbf{x})} \\ &\quad + \frac{\lambda(u,\mathbf{x}\setminus\{v\})\,\lambda(v,\mathbf{x}\setminus\{u\})\,c(u,v)}{\sqrt{\lambda(v,\mathbf{x}\cup\{u\})}}\left[\frac{1}{\sqrt{\lambda(u,\mathbf{x}\cup\{v\})}} - \frac{2}{\sqrt{\lambda(u,\mathbf{x})}}\right] \\ &= \sqrt{\lambda(u,\mathbf{x})}\sqrt{\lambda(v,\mathbf{x})} \\ &\quad + \frac{\lambda(u,\mathbf{x}\setminus\{v\})\sqrt{\lambda(v,\mathbf{x}\setminus\{u\})}\mathbf{1}[c(u,v) > 0]}{\sqrt{c(u,v)}}\left[\frac{1}{\sqrt{\lambda(u,\mathbf{x}\cup\{v\})}} - \frac{2}{\sqrt{\lambda(u,\mathbf{x})}}\right] \end{aligned}$$

by virtue of (17). For $u, v \notin \mathbf{x}$ this reduces to

$$\begin{aligned} S(u,v,\mathbf{x}) &= \sqrt{\lambda(u,\mathbf{x})}\sqrt{\lambda(v,\mathbf{x})} \\ &\quad + \frac{\lambda(u,\mathbf{x})\sqrt{\lambda(v,\mathbf{x})}\mathbf{1}[c(u,v) > 0]}{\sqrt{c(u,v)}}\left[\frac{1}{\sqrt{\lambda(u,\mathbf{x})c(u,v)}} - \frac{2}{\sqrt{\lambda(u,\mathbf{x})}}\right] \\ &= \sqrt{\lambda(u,\mathbf{x})}\sqrt{\lambda(v,\mathbf{x})} + \frac{\sqrt{\lambda(u,\mathbf{x})}\sqrt{\lambda(v,\mathbf{x})}\mathbf{1}[c(u,v) > 0]}{c(u,v)}\left[1 - 2\sqrt{c(u,v)}\right] \\ &= \sqrt{\lambda(u,\mathbf{x})}\sqrt{\lambda(v,\mathbf{x})}\left[1 + \left(\frac{1}{c(u,v)} - \frac{2}{\sqrt{c(u,v)}}\right)\mathbf{1}[c(u,v) > 0]\right]. \end{aligned}$$

The expression in brackets on the last line is nonnegative, and positive when $c(u,v) \neq 1$. Thus any nontrivial pairwise interaction gives rise to inflation of the

variance of the Pearson innovations, relative to any Poisson point process with an intensity function such that the support of the intensity function contains $\{u \in A : V(\{u\}) > -\infty\}$, the support of $\lambda(u, \emptyset)$.

5 Residuals

For the rest of the paper, we assume that a spatial point process *model*, governed by a parameter θ, is fitted to a realisation of the *true* point process $\mathbf{X}$. This scenario allows us to study the sensitivity of residuals to mis-specifications of the model. Both the 'true' point process $\mathbf{X}$ and the model process (for all values of θ) are assumed to satisfy the requirements stated in Sect. 3, namely, they are Gibbs processes observed in W under either scenario (i) or (ii).

Write λ_θ for the Papangelou conditional intensity of the model. As foreshadowed, we allow the weight function $h = h_\theta$ to depend on the parameter θ of the point process model. We assume θ is estimated by $\widehat{\theta} = \widehat{\theta}(\mathbf{X}_W)$ and plugged in to h, yielding $\hat{h} = h_{\hat{\theta}(\mathbf{X}_W)}$. Nothing is assumed about $\hat{\theta}$, other than its existence and uniqueness. The *h-weighted residual* (or more precisely the $\hat{h}$-weighted residual) is the signed random measure defined by

$$R_{\hat{h}}(B) = \sum_{u \in \mathbf{X}_B} h_{\hat{\theta}(\mathbf{X}_W)}(u, \mathbf{X}_W \setminus \{u\}) - \int_B h_{\hat{\theta}(\mathbf{X}_W)}(u, \mathbf{X}_W) \lambda_{\hat{\theta}(\mathbf{X}_W)}(u, \mathbf{X}_W) \, \mathrm{d}u. \quad (33)$$

In particular, the raw, inverse-λ and Pearson residuals are given by replacing $\lambda(u, \mathbf{X}_W)$ by $\lambda_{\hat{\theta}(\mathbf{X}_W)}(u, \mathbf{X}_W)$ on the right hand sides of (10)–(12); we denote these residuals by $R, R_{1/\hat{\lambda}}, R_{1/\sqrt{\hat{\lambda}}}$, respectively. In order that the Pearson and inverse-λ residuals be well defined, we require that $\lambda_{\hat{\theta}(\mathbf{X}_W)}(u, \mathbf{X}_W) > 0$ for all $u \in \mathbf{X}_A$, almost surely.

5.1 Homogeneous Poisson case

Consider the special case where $\mathbf{X}$ is a stationary Poisson process in $\mathbb{R}^d$ with intensity θ, i.e. $\lambda_\theta \equiv \theta$, and where the model is also a stationary Poisson process, fitted using the maximum likelihood estimator $\hat{\theta} = N(W)/|W|$. Recall that in this case, $A = W$ and $\partial A = \emptyset$. We have

$$\begin{aligned} R(B) &= N(B) - N(W)|B|/|W| \\ R_{1/\hat{\theta}}(B) &= |W|N(B)/N(W) - |B| \\ R_{1/\sqrt{\hat{\theta}}}(B) &= N(B)\sqrt{|W|/N(W)} - \sqrt{N(W)|W|} \end{aligned}$$

when $N(W) > 0$, and zero otherwise. It can be verified directly that these residuals have mean zero if the model is true. Notice also that when $B = W$ is the entire sampling window, we get

$$R(W) = R_{1/\hat{\theta}}(W) = R_{1/\sqrt{\hat{\theta}}}(W) = 0.$$

This is analogous to the fact that the raw residuals in simple linear regression sum to zero.

5.2 General expressions for mean of residuals

By (13) we hope that the (conditional) mean of the residual measure is approximately zero when the model is true and the parameter estimate is accurate. If $\mathbb{E}$ and λ denote the mean and the Papangelou conditional intensity for the true process $\mathbf{X}$, then the h-weighted residual (33) has true expectation

$$\begin{aligned}&\mathbb{E}[R_{\hat{h}}(B)|\mathbf{X}_{\partial A}]\\ &\quad= \int_B \mathbb{E}\left[h_{\hat{\theta}(\mathbf{X}\cup\{u\})}(u,\mathbf{X}_W)\lambda(u,\mathbf{X}_W) - h_{\hat{\theta}(\mathbf{X}_W)}(u,\mathbf{X})\lambda_{\hat{\theta}(\mathbf{X}_W)}(u,\mathbf{X}_W)\big|\,\mathbf{X}_{\partial A}\right]\,\mathrm{d}u\end{aligned}$$

Explicit results for the raw, inverse and Pearson residuals follow directly (Baddeley et al., 2005). Further analysis depends on the nature of the estimator $\widehat{\theta}$.

One case of interest is the inhomogeneous Poisson process with intensity $\lambda_\theta(u)$. The maximum likelihood estimate $\lambda_{\hat{\theta}(\mathbf{X}_W)}(u,\mathbf{X}_W)$ is in general a biased estimator of the true intensity $\lambda(u,\mathbf{X}_W)$, so the raw residuals do not in general have exactly zero mean.

5.3 Practical applicability

For Markov point processes, the form of the conditional intensity is known explicitly, by virtue of (5), so that the residuals are typically easy to evaluate. Examples were presented in Baddeley et al. (2005).

The most important class of non-Markov point processes is that of Cox processes. Expressions for the conditional intensity of a Cox process involve the conditional distribution of the driving intensity, (cf. Baddeley et al. 2005; Møller and Waagepetersen 2003; Møller and Waagepetersen 2007) and must typically be evaluated by Monte Carlo. In practice it may be preferable to replace $\lambda(u,\mathbf{X}_W)$ by the intensity function $\lambda(u) = \mathbb{E}[\lambda(u,\mathbf{X}_W)]$ and to use only the raw residuals, as mooted in Møller and Waagepetersen (2007). This reduces to an application of Campbell's Theorem rather than the GNZ formula.

6 Variance of residuals

6.1 General formula

The raw, Pearson, and inverse-lambda residuals can be written in the common form

$$R^p(B) = \sum_{u \in \mathbf{X}_B} \lambda_{\widehat{\theta}(\mathbf{X}_W)}(u, \mathbf{X}_W \setminus \{u\})^p - \int_B \lambda_{\widehat{\theta}(\mathbf{X}_W)}(u, \mathbf{X}_W)^{p+1}\, \mathrm{d}u \tag{34}$$

where $p = 0, -\frac{1}{2}, -1$ respectively, and we interpret $0^0 = 0$.

Proposition 5 *The raw, Pearson and inverse-lambda residuals have true variance*

$$\begin{aligned} \mathrm{var} R^p(B) \;=\; & \int_B \mathbb{E}\left[\lambda_{\widehat{\theta}(\mathbf{X}_W \cup \{u\})}(u, \mathbf{X}_W)^{2p}\; \lambda(u, \mathbf{X}_W) \,\middle|\, \mathbf{X}_{\partial A} \right] \mathrm{d}u \\ & + \int_B \int_B \mathrm{cov}\left[\lambda_{\widehat{\theta}(\mathbf{X}_W)}(u, \mathbf{X}_W)^{p+1},\; \lambda_{\widehat{\theta}(\mathbf{X}_W)}(v, \mathbf{X}_W)^{p+1} \,\middle|\, \mathbf{X}_{\partial A} \right] \mathrm{d}u\, \mathrm{d}v \\ & + \int_B \int_B T(u,v)\, \mathrm{d}u\, \mathrm{d}v - 2 \int_B \int_B M(u,v)\, \mathrm{d}u\, \mathrm{d}v \end{aligned} \tag{35}$$

where

$$\begin{aligned} M(u,v) = \; & \mathbb{E}\left[\lambda_{\widehat{\theta}(\mathbf{X}_W \cup \{u\})}(u, \mathbf{X}_W)^{p}\; \lambda_{\widehat{\theta}(\mathbf{X}_W \cup \{u\})}(v, \mathbf{X}_W)^{p+1}\; \lambda(u, \mathbf{X}_W) \,\middle|\, \mathbf{X}_{\partial A} \right] \\ & - \mathbb{E}\left[\lambda_{\widehat{\theta}(\mathbf{X}_W \cup \{u\})}(u, \mathbf{X}_W)^{p}\; \lambda(u, \mathbf{X}_W) \,\middle|\, \mathbf{X}_{\partial A} \right] \cdot \mathbb{E}\left[\lambda_{\widehat{\theta}(\mathbf{X}_W)}(v, \mathbf{X}_W)^{p+1} \,\middle|\, \mathbf{X}_{\partial A} \right] \end{aligned}$$

and

$$\begin{aligned} T(u,v) = \; & \mathbb{E}\left[\lambda_{\widehat{\theta}(\mathbf{X}_W \cup \{u,v\})}(u, \mathbf{X}_W \cup \{v\})^{p}\; \lambda_{\widehat{\theta}(\mathbf{X}_W \cup \{u,v\})}(v, \mathbf{X}_W \cup \{u\})^{p}\; \lambda(u, v, \mathbf{X}_W) \,\middle|\, \mathbf{X}_{\partial A} \right] \\ & - \mathbb{E}\left[\lambda_{\widehat{\theta}(\mathbf{X}_W \cup \{u\})}(u, \mathbf{X}_W)^{p}\; \lambda(u, \mathbf{X}_W) \,\middle|\, \mathbf{X}_{\partial A} \right] \cdot \mathbb{E}\left[\lambda_{\widehat{\theta}(\mathbf{X}_W \cup \{v\})}(v, \mathbf{X}_W)^{p}\; \lambda(v, \mathbf{X}_W) \,\middle|\, \mathbf{X}_{\partial A} \right] \end{aligned}$$

where the expectations and variances are with respect to the true distribution of the point process $\mathbf{X}$*, and* $\lambda(u, \mathbf{X}_W)$ *and* $\lambda(u, v, \mathbf{X}_W)$ *are the first and second order Papangelou conditional intensities of the true process* $\mathbf{X}$*.*

This result is obtained from Lemma 2 by substituting $h(u, \mathbf{x}) = \lambda_{\widehat{\theta}(\mathbf{x} \cup \{u\})}(u, \mathbf{x})^p$ and $g(u, \mathbf{x}) = \lambda_{\widehat{\theta}(\mathbf{x})}(u, \mathbf{x})^{p+1}$.

Note carefully that this result does not assume the point process $\mathbf{X}$ actually conforms to the model. The variance formula (35) involves characteristics of both the fitted model and the true underlying point process. On the right hand side of (35), the first term is likely to be the dominant term, since it is the variance of the sum in (34) when $\mathbf{X}$ is Poisson. The second term is the variance of the

integral in (34); it arises from variability in the parameter estimate $\widehat{\theta}$ and should be of smaller order than the first term. The cross-terms $T(u,v)$ and $M(u,v)$ are 'influence' terms that measure the effect on the *estimated* conditional intensity $\lambda_{\widehat{\theta}(\mathbf{X})}$ of adding extra points to the realisation of $\mathbf{X}$, weighted by second-order properties of the true process. For the raw residuals, T is closely related to the pair correlation function of X, and is identically zero for Poisson processes, while M is a kind of influence function for the effect of adding one extra point.

6.2 Variance deflation

In the very special case where a homogeneous Poisson process model with intensity θ is fitted to a realisation of a homogeneous Poisson process with intensity β, the residual variances are

$$\begin{aligned}
\mathrm{var}R(B) &= \beta\,|B|\left(1-\frac{|B|}{|W|}\right)\\
\mathrm{var}R_{1/\hat{\theta}}(B) &= |B|(|W|-|B|)\mathbb{E}\left(\frac{\mathbf{1}[N(W)>0]}{N(W)}\right)\\
\mathrm{var}R_{1/\sqrt{\hat{\theta}}}(B) &= |B|\left(1-\frac{|B|}{|W|}\right).
\end{aligned}$$

Note that the residual variances are smaller than the corresponding innovation variances

$$\mathrm{var}I(B) = \theta\,|B|, \quad \mathrm{var}I_{1/\theta}(B) = |B|/\theta, \quad \mathrm{var}I_{1/\sqrt{\theta}}(B) = |B|.$$

This is analogous to the deflation of residual variance in the linear model cf. Atkinson 1985; Baddeley et al. 2005.

6.3 Large-sample asymptotics for inhomogeneous Poisson case

Suppose an inhomogeneous Poisson process model with intensity function $\lambda_\theta(u)$ is fitted to a realisation of an inhomogeneous Poisson process with true intensity $\beta(u)$. Then by Proposition 5 the raw residuals have exact variance

$$\begin{aligned}
\mathrm{var}R(B) = &\int_B \beta(u)\,\mathrm{d}u + \int_B\int_B \mathrm{cov}\left[\lambda_{\hat{\theta}}(X_W)(u),\, \lambda_{\hat{\theta}(X_W)}(v)\right]\,\mathrm{d}u\,\mathrm{d}v\\
&-2\int_B\int_B \mathbb{E}\left[\lambda_{\hat{\theta}(X_W\cup\{u\})}(v) - \lambda_{\hat{\theta}(X_W)}(v)\right]\,\beta(u)\,\mathrm{d}v\,\mathrm{d}u \qquad (36)
\end{aligned}$$

where the expectation is with respect to the true model.

This yields asymptotic expressions for residual variance using the delta method. For example, suppose the model is of the form $\lambda_\theta(u) = \exp(\theta^{\mathrm{T}} S(u))$ where θ is a p-dimensional vector and $S : W \to \mathbb{R}^p$ is a known function. Assume

the model is correctly specified, i.e. $\beta(u) = \lambda_{\theta_0}(u)$ for some $\theta_0 \in \mathbb{R}^p$. Let $\hat{\theta}$ be the maximum likelihood estimator based on $\mathbf{X}_W$. Under regularity conditions, $\hat{\theta}$ is the unique solution of the MLE normal equations

$$\sum_{u \in \mathbf{X}_W} S(u) = \int_W S(u) e^{\theta^{\mathrm{T}} S(u)} \, \mathrm{d}u.$$

Consider a limiting context in which the usual asymptotic results for maximum likelihood apply. That is, assume $\hat{\theta}$ is asymptotically multivariate Normal with mean vector θ_0 and variance-covariance matrix $\mathcal{I}^{-1}$, where

$$\mathcal{I} = \mathcal{I}(\theta_0) = \int_W \left[S(u) S(u)^{\mathrm{T}} \right] \, \beta(u) \, \mathrm{d}u$$

is the Fisher information matrix. Then on the right hand side of (36) we have asymptotically

$$\mathrm{cov}\left[\lambda_{\hat{\theta}(X_W)}(u), \, \lambda_{\hat{\theta}(X_W)}(v) \right] \sim \beta(u) \beta(v) S(u)^{\mathrm{T}} \mathcal{I}^{-1} \, S(v).$$

The estimating equation for $\hat{\theta}(\mathbf{X}_W \cup \{u\})$ is

$$S(u) + \sum_{v \in \mathbf{X}_W} S(v) = \int_W S(v) e^{\theta^{\mathrm{T}} S(v)} \, \mathrm{d}v$$

so that, to first order,

$$\hat{\theta}(\mathbf{X}_W \cup \{u\}) - \hat{\theta}(\mathbf{X}_W) \sim \mathcal{I}^{-1} \, S(u)$$

yielding

$$M(u, v) \sim \beta(u) \beta(v) \, S(u)^{\mathrm{T}} \mathcal{I}^{-1} \, S(v).$$

Substituting in (36) gives an expression for the asymptotic variance of the raw residuals. By similar arguments for $p = -\frac{1}{2}, -1$ we get the asymptotic variance of the raw, Pearson and inverse-lambda residuals

$$\begin{aligned} \mathrm{var} R^p(B) &\sim \int_B \beta(u)^{2p+1} \, \mathrm{d}u - \int_B \int_B \beta(u)^{p+1} \beta(v)^{p+1} \, S(u)^{\mathrm{T}} \mathcal{I}^{-1} \, S(v) \, \mathrm{d}u \, \mathrm{d}v \\ &= \int_B \beta(u) \, \mathrm{d}u - G^{\mathrm{T}} \mathcal{I}^{-1} \, G \end{aligned} \tag{37}$$

where $G = G_p(B) = \int_B \beta(u)^{p+1} S(u) \, \mathrm{d}u$. This asymptotic expression also exhibits its variance deflation of the residuals compared to the innovations.

7 Null distribution of smoothed residual field

In practice it is useful to smooth the residual measure (Baddeley et al. 2005). Let the smoothing kernel $k(\cdot)$ be a probability density on $\mathbb{R}^d$. The *smoothed residual field* is the random function

$$s(u) = e(u) \int_A k(u - v)\, \mathrm{d}R_{\hat{h}}(v)$$

for $u \in A$, where $e(u)$ is a correction for edge effects in the window W given by $e(u)^{-1} = \int_W k(u - v)\, \mathrm{d}v$, see Baddeley et al. (2005). An important question for applications is to determine the distribution of $S = s(u)$ at a fixed location $u \in W$ under a null hypothesis, especially under the hypothesis of a stationary Poisson process. This is related to the distribution of the scan statistic (Alm 1988) as explained in Baddeley et al. 2005, p. 643.

In this section we assume $\mathbf{X}$ is a stationary Poisson process with intensity λ in $\mathbb{R}^2$, and that the fitted model is also a stationary Poisson process. We calculate the distribution of $S = s(u)$ at a fixed $u \in W$ when $h = 1$. Note that for the stationary Poisson process model, the raw, inverse-λ and Pearson innovations/residuals are all proportional to each other. We ignore the effect of parameter estimation, that is, we consider the kernel-smoothed innovation measure, rather than the kernel-smoothed residual measure. Edge effects will also be ignored, and edge correction is not applied.

Letting $\mathbf{X} = \{x_i, i = 1, 2, \ldots\}$ denote the points of the process, we consider the uncorrected, smoothed, raw innovation field

$$s(u) = \sum_i k(u - x_i) - \lambda$$

where the kernel is the isotropic Gaussian density

$$k(u) = \frac{1}{2\pi\sigma^2} \exp(-||u||^2/(2\sigma^2))$$

so that

$$S = s(u) = \frac{1}{2\pi\sigma^2} \sum_i \exp(-||u - x_i||^2/(2\sigma^2)) - \lambda.$$

The ordered values $||u - x_i||^2$ are the event times T_i of a homogeneous Poisson process of intensity $\lambda\pi$ on $\mathbb{R}_+$. Since the inter-event times $V_i = T_i - T_{i-1}$ are exponentially distributed with rate $\lambda\pi$ we can represent S as

$$S = \frac{\lambda}{\mu} \sum_i \left(\prod_{j=1}^{i} U_j \right)^{1/\mu} - \lambda \tag{38}$$

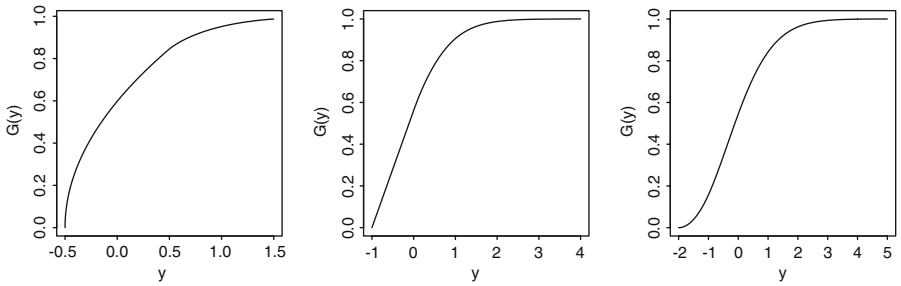

Fig. 1 Cumulative distribution function of $Y = 2\pi\sigma^2 S$ for the cases $\mu = 0.5,\ 1,\ 2$ (left to right), where $S = s(0)$ is a typical value of the kernel-smoothed raw innovation field for a homogeneous Poisson process of rate λ, smoothed by an isotropic Gaussian kernel with standard deviation σ, and $\mu = 2\lambda\pi\sigma^2$

where U_j are i.i.d. uniform $[0,1]$ r.v.'s and $\mu = 2\lambda\pi\sigma^2$.

Let $X = \mu(1 + S/\lambda)$ be the sum in (38). Then X satisfies the distributional equivalence

$$X \equiv U^{1/\mu}(1 + X) \tag{39}$$

where U is a uniform $[0,1]$ random variable independent of X. This equivalence is discussed by Vervaat (1979) with references to its prior occurrence. As shown in Appendix A, Eq. (39) leads to an integral equation for the c.d.f. $F(x)$ of X,

$$F(x) = \mu x^\mu \int_{(x-1)_+}^{\infty} \frac{F(z)}{(1+z)^{1+\mu}}\,\mathrm{d}z = x^\mu \left[C - \mu \int_0^{(x-1)_+} \frac{F(z)}{(1+z)^{1+\mu}}\,\mathrm{d}z \right] \tag{40}$$

where

$$C = \mu \int_0^{\infty} \frac{F(z)}{(1+z)^{1+\mu}}\,\mathrm{d}z = \mathbb{E}[(1+X)^{-\mu}] = e^{-\gamma\mu}/\,\Gamma(1+\mu)$$

where $\gamma \approx 0.5772$ is Euler's constant. For $x \in [0,1]$ the integral in (40) is zero and

$$F(x) = Cx^\mu, \quad 0 \le x \le 1.$$

One may then apply (40) recursively to obtain the values of F on successive intervals $[n, n+1]$ for $n = 1, 2, \ldots$, see Appendix A. We have no analytic form for the solution, but it may be computed numerically.

For any given value of μ, these recursive computations yield the distribution of $X = \mu(1 + S/\lambda)$, so the c.d.f. of $Y = (\mu/\lambda)S = 2\pi\sigma^2 S$ is $G(y) = F(y + \mu)$ for $-\mu \le y \le \infty$. Figure 1 shows the computed G for the cases $\mu = 0.5$, 1 and 2.

Acknowledgments This paper was prepared in conjunction with Baddeley et al. (2005); we thank our coauthors Martin Hazelton and Rolf Turner for their collaboration. We thank David Brillinger, Michael Buckley, David Vere-Jones, Rick Vitale and Rasmus Waagepetersen for illuminating comments. This research was supported by the Australian Research Council (Large Grant A69941083

Extrapolating and interpolating spatial patterns) and by The Danish Natural Science Research Council.

Appendix A: Study of the distributional equivalence

Here we consider the distributional equivalence (39) where X is a positive continuous random variable X with c.d.f. F. This gives the following integral equation for F:

$$F(x) = \int_0^1 F(u^{-1/\mu}x - 1)\, \mathrm{d}u = \mu x^\mu \int_{x-1}^\infty F(z) \frac{\mathrm{d}z}{(1+z)^{1+\mu}}.$$

Since $F(z) = 0$ if $z < 0$, we have

$$F(x) = \mu x^\mu \int_{(x-1)^+}^\infty F(z) \frac{\mathrm{d}z}{(1+z)^{1+\mu}} \tag{41}$$

whereby (40) is verified.

A.1 Solutions

In principle we can solve (41) section-wise. For the case $0 \le x \le 1$,

$$F(x) = C_0\, x^\mu$$

where

$$C_0 = \mu \int_0^\infty F(z) \frac{\mathrm{d}z}{(1+z)^{1+\mu}} = \mathbb{E}\left[(1+X)^{-\mu}\right] < 1.$$

It can be shown that

$$C_0 = \frac{1}{\Gamma(\mu)} \int_0^\infty v^{\mu-1} \exp\left(-v - \mu \int_0^1 \frac{1-e^{-vy}}{y}\, \mathrm{d}y\right) \mathrm{d}v.$$

Now consider the case $1 \le x \le 2$. We have

$$\begin{aligned} F(x) &= \mu x^\mu \left[\int_0^\infty F(z) \frac{\mathrm{d}z}{(1+z)^{1+\mu}} - \int_0^{x-1} C_0 z^\mu \frac{\mathrm{d}z}{(1+z)^{1+\mu}}\right] \\ &= C_0 x^\mu \left[1 - \mu \int_0^{x-1} \frac{z^\mu}{(1+z)^{1+\mu}}\, \mathrm{d}z\right]. \end{aligned}$$

The last integral transforms to an incomplete beta integral:

$$\int_0^{x-1} \frac{z^\mu}{(1+z)^{1+\mu}}\,\mathrm{d}z = \int_{1/x}^1 u^{-1}(1-u)^\mu\,\mathrm{d}u = H_\mu(x),\ \text{say.}$$

So

$$F(x) = C_0 x^\mu\left[1-\mu H_\mu(x)\right].$$

For example, if $\mu = 1$, we have $H_1(x) = \log x - 1 + 1/x$, giving $F(x) = C_0[2x - x\log x - 1]$ for $1 \le x \le 2$, and $F(2) = (3 - 2\log 2)C_0$. If instead $\mu = 2$, then $F(x) = C_0 x^2$ for $0 \le x \le 1$ and $F(x) = C_0((2x-1)^2 - 2x^2\log x)$ for $1 \le x \le 2$ with $F(2) = (9 - 8\log 2)C_0$.

A.2 Evaluation of constant C_0

We now prove that $C_0 = e^{-\gamma\mu}/\Gamma(1+\mu)$ where γ is Euler's constant. Write $\phi(\theta) = \mathbb{E}[\mathrm{e}^{-\theta X}]$ as

$$\phi(\theta) = \exp\left\{-\mu\int_0^1 \frac{1-\mathrm{e}^{-\theta x}}{x}\,\mathrm{d}x\right\} = \exp\left\{-\mu\int_0^\theta \frac{1-\mathrm{e}^{-y}}{y}\,\mathrm{d}y\right\}.$$

For $\theta > 1$ the integral above is

$$\int_0^1 \frac{1-\mathrm{e}^{-y}}{y}\,\mathrm{d}y + \log\theta - \int_1^\theta \frac{\mathrm{e}^{-y}}{y}\,\mathrm{d}y$$

whence, as $\theta \to \infty$,

$$\begin{aligned}\phi(\theta) &\sim \theta^{-\mu}\exp\left[-\mu\left(\int_0^1 \frac{1-\mathrm{e}^{-y}}{y}\,\mathrm{d}y - \int_1^\infty \frac{\mathrm{e}^{-y}}{y}\,\mathrm{d}y\right)\right]\\ &= \theta^{-\mu}\exp(-\gamma\mu).\end{aligned}$$

A.3 Further notes on F

The Tauberian theorem for Laplace–Stieltjes transforms (Feller 1971, p. 445) implies that

$$F(x) \sim x^\mu \mathrm{e}^{-\gamma\mu}/\Gamma(1+\mu), \quad x \to 0.$$

This comes effectively from Takács (1955, p. 376). He observes that

$$\phi(\theta) = \theta^{-\mu}\mathrm{e}^{-\gamma\mu}\exp(-\mu E_1(\theta))$$

where

$$E_1(\theta) = \int_\theta^\infty \frac{\mathrm{e}^{-y}}{y}\,\mathrm{d}y = \int_1^\infty \frac{\mathrm{e}^{-\theta z}}{z}\,\mathrm{d}z.$$

Since clearly

$$\theta^{-\mu} = \frac{1}{\Gamma(\mu)} \int_0^\infty x^{\mu-1}\mathrm{e}^{-\theta x}\,\mathrm{d}x,$$

the p.d.f. of X can be expressed as

$$f(x) = \frac{\mathrm{e}^{-\gamma\mu}}{\Gamma(\mu)} \left[x^{\mu-1} + \sum_{n\geq 1} \frac{(-\mu)^n}{n!} H_n(x) \right] \tag{42}$$

where

$$H_n(x) = \int_0^x (x-y)^{\mu-1} f_n(y)\,\mathrm{d}y$$

and f_n is the n-fold convolution of $y^{-1}1_{(1,\infty)}(y)$, i.e. $\widehat{f}_n(\theta) = (E_1(\theta))^n$. Obviously $f_n(y) = 0$ if $y < n$, and hence for a given x the series at (42) has only finitely many nonzero terms. Similarly

$$F(x) = \frac{\mathrm{e}^{-\gamma\mu}}{\Gamma(1+\mu)} \left[x^{\mu} + \sum_{n\geq 1} \frac{(-\mu)^n}{n!} J_n(x) \right]$$

where

$$J_n(x) = \int_0^x (x-y)^{\mu} f_n(y)\,\mathrm{d}y.$$

For example, if $\mu = 1$, since $f_1(y) = \frac{1}{y}1_{(1,\infty)}(y)$ we get $H_1(x) = (\log x)\,1_{(1,\infty)}(x)$. Since $H_n(x) = 0$ if $1 \leq x \leq 2$ for all $n \geq 2$, we find that

$$f(x) = e^{-\gamma}[1 - \log x], \quad 1 \leq x \leq 2,$$

which agrees with the expression found for F in this case. For $2 \leq x \leq 3$ it becomes more difficult to study f and F analytically although they can still be evaluated.

References

Alm, S. (1988). Approximation and simulation of the distributions of scan statistics for Poisson processes in higher dimensions. *Extremes*, *1*(1), 111–126.

Andersen, P., Borgan, Ø., Gill, R., Keiding, N. (1993). *Statistical Models based on Counting Processes*. New York: Springer.

Atkinson, A. (1985). *Plots, Transformations and Regression*. Number 1 in Oxford Statistical Science Series. Clarendon: Oxford University Press.

Baddeley, A., Turner, R., Møller, J., Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society, Series B*, *67*(5), 617–666.

Brillinger, D. (1978). Comparative aspects of the study of ordinary time series and of point processes. In P. Krishnaiah (Ed.), *Developments in Statistics* pp. 33–133 NewYork: Academic Press.

Brillinger, D. (1994). Time series, point processes, and hybrids. *Canadian Journal of Statistics*, *22*, 177–206.

Brillinger, D., Segundo, J. (1979). Empirical examination of the threshold model of neuron firing. *Biological Cybernetics*, *35*, 213–220.

Daley, D., Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*. New York: Springer.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, volume 2. New York: John Wiley and Sons, second edition.

Fleming, T., Harrington, D. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

Georgii, H.-O. (1976). Canonical and grand canonical Gibbs states for continuum systems. *Communications of Mathematical Physics*, *48*, 31–51.

Kalbfleisch, J., Prentice, R. (1980). *The Statistical Analysis of Failure Time Data*. NewYork: Wiley.

Karr, A. (1985). *Point Processes and Their Statistical Inference*. New York: Dekker.

Lewis, P. (1972). Recent results in the statistical analysis of univariate point processes. In: P. Lewis (Ed.) *Stochastic Point Processes* (pp. 1–54) New York: Wiley.

Møller, J. Waagepetersen, R. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman and Hall/CRC.

Møller, J. Waagepetersen, R. (2007). Modern spatial point process modelling and inference. *Scandinavian Journal of Statistics*. (To appear).

Nguyen, X, Zessin, H. (1979). Integral and differential characterizations of Gibbs processes. *Mathematische Nachrichten*, *88*, 105–115

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*, 9–27

Ogata, Y., Katsura, K., Tanemura, M. (2003). Modelling heterogeneous space-time occurrences of earthquakes and its residual analysis. *Applied Statistics*, *52*(4), 499–509.

Papangelou, F. (1974). The conditional intensity of general point processes and an application to line processes. *Zeitschrift fuer Wahscheinlichkeitstheorie und verwandte Gebiete*, *28*, 207–226.

Preston, C, (1976). *Random Fields*. New York: Springer.

Ripley, B. Kelly, F. (1977). Markov point processes. *Journal of the London Mathematical Society*, *15*, 188–192.

Ruelle, D. (1969). *Statistical Mechanics*. New York: Wiley.

Stoyan, D. Grabarnik, P. (1991). Second-order characteristics for stochastic structures connected with Gibbs point processes. *Mathematische Nachrichten*, *151*, 95–100.

Takács, L. (1955). On stochastic processes connected with certain physical recording apparatuses. *Acta Mathematica Academiae Scientiarum Hungaricae*, *6*, 363–374.

Vervaat, W. (1979). On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Advances in Applied Probability*, *11*, 750–783.

Zhuang, J., Ogata, Y., Vere-Jones, D. (2005). Diagnostic analysis of space-time branching processes for earthquakes. In A. Baddeley , P. Gregori, J. Mateu, R. Stoica, D. Stoyan, (Eds.) *Case Studies in Spatial Point Process Modelling*, number 185 in Lecture Notes in Statistics, chapter 15. New York: Springer.