PROPERTIES OF RESIDUALS FOR SPATIAL POINT PROCESSES

A. BADDELEY,* *University of Western Australia*

J. MØLLER,** *University of Aalborg*

A.G. PAKES,* *University of Western Australia*

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, lack of correlation, and marginal distributions.

Keywords: Georgii-Nguyen-Zessin formula; Gibbs point process; set-indexed martingale; Papangelou conditional intensity; Pearson residuals; scan statistic; smoothed residual field

2000 Mathematics Subject Classification: Primary 62M30
Secondary 62J20

1. Introduction

The inspection of residuals is an important check on the appropriateness of a probability model fitted to data [3]. This paper defines residuals for spatial point processes, and describes their properties.

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 [14, Thm. 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 [15, 5, 7, 6] and survival analysis [2, 10, 13].

For point processes in higher dimensions, the lack of a natural ordering 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* [17] *conditional intensity* $\lambda(u, \mathbf{X})$ which conditions on the outcome of the process at all spatial locations other than u. In [4] we used the Papangelou conditional intensity to define residuals for finite point processes in $\mathbb{R}^2$,

* Postal address: School of Mathematics & Statistics M019, University of Western Australia, 35 Stirling Highway, Nedlands WA 6009, Australia
** Postal address: Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg Ø, Denmark

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 and residuals for spatial point processes. Section 4 obtains expressions for the variances of innovations and residuals. Section 5 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 [8]. 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. Furthermore, $\mathbf{X}$ 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 $\mathbb{R}^d \times \mathcal{N}$. Equation (2) is called the *Georgii-Nguyen-Zessin formula* [11, 16], 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

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

For instance, this becomes convenient in Section 3 when we define the Pearson and inverse-λ innovations/residuals.

In [4] 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 $0/0 = 0$.

In applications we often consider a Markov point process [19] of finite interaction range $R < \infty$. This is a Gibbs process where the conditional intensity is defined with respect to a so-called potential V as follows [12, 18, 20]. Suppose that $V(\mathbf{x}) \in [-\infty, \infty)$ is defined for all $\mathbf{x} \in \mathcal{N}$ such that $V(\mathbf{x}) = 0$ whenever $\mathbf{x}$ contains two points u, v with distance $\|u - v\| > R$. Then we say that a Gibbs point process $\mathbf{X}$ is Markov with potential V if

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

In other words, the point process is Markov if $\lambda(u, \mathbf{x})$ 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. 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

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

3. Innovations and residuals

This section defines innovations and residuals for (finite as well as infinite) spatial point processes $\mathbf{X}$ governed by a parameter θ and having Papangelou conditional intensity $\lambda = \lambda_\theta$. We assume that $\mathbf{X}$ is Markov with interaction range $R < \infty$ and that $\mathbf{X}$ is observed within a bounded window $W \subset \mathbb{R}^d$, with positive volume $|W|$, and let $\hat{\theta} = \hat{\theta}(\mathbf{X}_W)$ denote an estimator of θ based on $\mathbf{X}_W$. If $\mathbf{X}$ may have points outside W, 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). In fact many of our results remain true for general Gibbs point processes (defined as point processes satisfying the GNZ formula (2)), including non-Markovian point processes such as Cox processes, but for specificity we have here chosen to restrict attention to Markov point processes. For example, in Section 3.1.1, Proposition 3.1 but not Proposition 3.2 remains true for a general Gibbs point process.

Throughout this paper we let the set A be defined as follows. If the process $\mathbf{X}$ lives within W, let $A = W$ and $\partial A = \emptyset$. If the point process may have points outside W, let $A = V$.

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; we shall often let $h = h_\theta$ depend on the model parameter θ. Equation (8) rather than (2) is the relevant form of the Georgii-Nguyen-Zessin formula when inference is performed on the conditional point process $\mathbf{X}_A|\mathbf{X}_{\partial A}$. If $\mathbf{X}$ lives in W, then $\mathbf{X}_A|\mathbf{X}_{\partial A}$ 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 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. [4] 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 equation (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 noticed above, we find (13) to be the more relevant property when inference is based on $\mathbf{X}_A|\mathbf{X}_{\partial A}$.

3.1.1. *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 martingale.

Proposition 3.1. *If $A = A_n$ 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 $(A_{n+1} \setminus A_n)^c = A_n \cup A_{n+1}^c$,

$$\mathbb{E}[I_{n+1}|\mathbf{X}_{A_n}] = \mathbb{E}[I_n + 0|\mathbf{X}_{A_n}] = I_n$$

which implies the martingale property

$$\mathbb{E}[I_{n+1}|I_n, I_{n-1}, \ldots] = I_n.$$

Lemma 1. *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 Section 2).

Proposition 3.2. *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 conditionally independent given $\mathbf{X}_C$. Moreover, conditional on $\mathbf{X}_C$, $I_h(B_1)$ and $I_h(B_2)$ are uncorrelated; and also without conditioning, $I_h(B_1)$ and $I_h(B_2)$ are uncorrelated.*

Proof. Follows immediately from Lemma 1, the spatial Markov property (see Section 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 asymptotic behaviour of the raw, inverse-λ and Pearson innovations and residuals as the sampling window W expands. However, we do not investigate this in the present paper.

3.2. Residuals

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 $\hat{h} = h_{\hat{\theta}(\mathbf{X}_W)}$. The *h-weighted residual* (or more precisely the $\hat{h}$-weighted residual) is the signed random measure defined by

$$R_{\hat{h}}(B) = \sum_{u \in \mathbf{X}_B} h_{\hat{\theta}(\mathbf{X}_W)}(u, \mathbf{X}_W \setminus \{u\}) - \int_B h_{\hat{\theta}(\mathbf{X}_W)}(u, \mathbf{X}_W) \lambda_{\hat{\theta}(\mathbf{X}_W)}(u, \mathbf{X}_W)\, \mathrm{d}u. \quad (15)$$

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 $u \in \mathbf{X}_A$.

Methods of visualising raw, Pearson and inverse-λ residuals are proposed in [4].

3.2.1. *Homogeneous Poisson case* Assume that $\mathbf{X}$ is a stationary Poisson process in $\mathbb{R}^d$ with intensity θ, i.e. $\lambda_\theta \equiv \theta$, and we use 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.

3.2.2. *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 model of $\mathbf{X}$, then the h-weighted residual (15) has true expectation

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

Explicit results for the raw, inverse and Pearson residuals follow directly [4]. Further analysis depends on the nature of the estimator $\widehat{\theta}$.

4. Variances

In this section we give details for deriving variances of residuals and innovations.

4.1. Variance formulae

Let $\mathbf{X} \otimes \mathbf{X}$ be the point process on $\mathbb{R}^d \times \mathbb{R}^d$ consisting of all pairs (u, v) of *distinct* points of $\mathbf{X}$. It follows immediately from the GNZ formula (2) that $\mathbf{X} \otimes \mathbf{X}$ is a Gibbs point process with (two-point) Papangelou conditional intensity

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

meaning that the GNZ formula in the form

$$\mathbb{E}\left[\sum_{u,v \in \mathbf{X}:\, u \neq v} h(u, v, \mathbf{X} \setminus \{u, v\})\right] = \mathbb{E}\left[\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(u, v, \mathbf{X})\lambda(u, v, \mathbf{X}) \, \mathrm{d}u \, \mathrm{d}v\right] \tag{17}$$

is satisfied 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 (more precisely for Lebesgue almost all (u, v)). 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. when the potential $V(\mathbf{x})$ is zero whenever $n(\mathbf{x}) > 2$) has

$$\lambda(u, v, \mathbf{x}) = \lambda(u, \mathbf{x} \setminus \{v\}) \ \lambda(v, \mathbf{x} \setminus \{u\}) \ c(u, v) \tag{18}$$

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

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

$$\begin{aligned} &\mathbb{E}\left[\sum_{u,v \in \mathbf{X}_A:\, u \neq v} h(u, v, \mathbf{X}_W \setminus \{u, v\}) \middle| \mathbf{X}_{\partial A}\right] \\ &= \mathbb{E}\left[\int_A \int_A h(u, v, \mathbf{X}_W)\lambda(u, v, \mathbf{X}_W)\lambda(u, v, \mathbf{X}_W) \, \mathrm{d}u \, \mathrm{d}v \middle| \mathbf{X}_{\partial A}\right]. \end{aligned} \tag{19}$$

Proposition 4.1. *For any nonegative measurable function h,*

$$\begin{aligned} &\operatorname{var}\left[\sum_{u \in \mathbf{X}_A} h(u, \mathbf{X}_W \setminus \{u\}) \middle| \mathbf{X}_{\partial A}\right] \\ &= \int_A \mathbb{E}\left[h(u, \mathbf{X}_W)^2 \, \lambda(u, \mathbf{X}_W) | \mathbf{X}_{\partial A}\right] \, \mathrm{d}u + \int_A \int_A T(u, v) \, \mathrm{d}u \, \mathrm{d}v \end{aligned} \tag{20}$$

where

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

Proof. Follows immediately by expanding the square of the sum on the left side of (20) as a double sum, and using (8) and (19).

For example, for a Poisson process with intensity function $\lambda(u)$, (20) reduces to

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

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

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

as expected by the independence properties of the Poisson process.

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

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

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

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

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 4.1. 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 4.2. *The variance of the h-weighted innovation is*

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

where

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

Proof. Substitute $g(u,\mathbf{x}) = \lambda(u,\mathbf{x})h(u,\mathbf{x})$ in Lemma 2.

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

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

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 $h \equiv 1$, equation (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

$$\mathrm{var}\left[I(B)|\mathbf{X}_{\partial A}\right] = \int_B \mathbb{E}\left[\lambda(u,\mathbf{X}_W)|\mathbf{X}_{\partial A}\right]\,\mathrm{d}u + \int_B\int_B \mathbb{E}\left[\lambda(u,\mathbf{X}_W)\lambda(v,\mathbf{X}_W) - \lambda(u,v,\mathbf{X}_W)|\mathbf{X}_{\partial A}\right]\,\mathrm{d}u\,\mathrm{d}v. \quad (27)$$

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 (20) to $h(u,\mathbf{x}) = 1/\lambda(u,\mathbf{x})$, we find that

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

For example, consider a pairwise interaction process with a finite potential (i.e. $\lambda(\cdot,\cdot)) > 0$ and $c(\cdot,\cdot) > 0$). Then (18) and (28) yield

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

This was derived in [21] 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 $\lambda(\cdot) > 0$, equation (29) reduces to (21) with $h(u) = 1/\lambda(u)$.

4.2.3. *Pearson innovations* For the Pearson innovations (11), $h(u,\mathbf{x}) = 1/\sqrt{\lambda(u,\mathbf{x})}$ 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

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

by virtue of (18). 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 $\lambda(u,\emptyset)$.

4.3. Variance of residuals

General formulae for the variances of the residuals can also be obtained from Proposition 2. These formulae are cumbersome, involving characteristics of both the fitted model and the underlying point process.

For example, suppose a Poisson process model with intensity function $\lambda_\theta(u)$ is fitted to a realisation of a Poisson process with true intensity $\lambda(u)$. Then the raw residuals have variance

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

where the expectation is with respect to the true model.

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

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

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

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

This is analogous to the deflation of residual variance in the linear model; cf. [4].

5. Null distribution of smoothed residual field

In practice it is useful to smooth the residual measure [4]. 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 [4]. 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 [1] as explained in [4, 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

$$S = \frac{\lambda}{\mu} \sum_i \left(\prod_{j=1}^{i} U_j \right)^{1/\mu} - \lambda \tag{33}$$

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 (33). Then X satisfies the distributional equivalence

$$X \equiv U^{1/\mu}(1 + X) \tag{34}$$

where U is a uniform $[0, 1]$ random variable independent of X. This equivalence is discussed by Vervaat [23] with references to its prior occurrence. As shown in Appendix A, equation (34) leads to an integral equation for the c.d.f. $F(x)$ of X,

$$F(x) = \mu x^\mu \int_{(x-1)_+}^{\infty} \frac{F(z)}{(1+z)^{1+\mu}} \, \mathrm{d}z = x^\mu \left[C - \mu \int_0^{(x-1)_+} \frac{F(z)}{(1+z)^{1+\mu}} \, \mathrm{d}z \right] \tag{35}$$

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 (35) is zero and

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

One may then apply (35) 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.

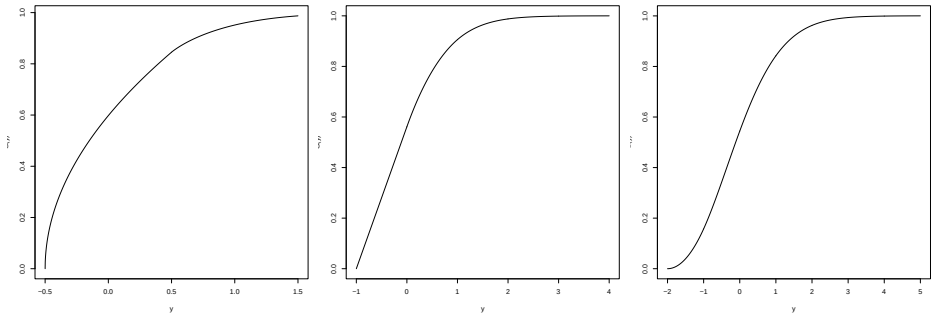

FIGURE 1: Cumulative distribution function of $Y = 2\pi\sigma^2 S$ for the cases $\mu = 0.5,\ 1,\ 2$ (left to right), where $S = s(0)$ is a typical value of the kernel-smoothed raw innovation field for a homogeneous Poisson process of rate λ, smoothed by an isotropic Gaussian kernel with standard deviation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

Acknowledgements

This paper was prepared in conjunction with [4]; we thank our coauthors Martin Hazelton and Rolf Turner for their collaboration. We thank David Brillinger, David Vere-Jones and Rick Vitale 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.

References

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

[2] ANDERSEN, P., BORGAN, Ø., GILL, R. AND KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

[3] ATKINSON, A. (1985). *Plots, Transformations and Regression*. No. 1 in Oxford Statistical Science Series. Oxford University Press/ Clarendon.

[4] BADDELEY, A., TURNER, R., MØLLER, J. AND HAZELTON, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society, series B* **67,** 617–666.

[5] BRILLINGER, D. (1978). Comparative aspects of the study of ordinary time series and of point processes. In *Developments in Statistics*. ed. P. Krishnaiah. Academic Press pp. 33–133.

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

[7] BRILLINGER, D. AND SEGUNDO, J. (1979). Empirical examination of the threshold model of neuron firing. *Biological Cybernetics* **35,** 213–220.

[8] DALEY, D. AND VERE-JONES, D. (1988). *An Introduction to the Theory of Point Processes*. Springer Verlag, New York.

[9] Feller, W. (1971). *An introduction to probability theory and its applications* second ed. vol. 2. John Wiley and Sons.

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

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

[12] Georgii, H.-O. (1988). *Gibbs Measures and Phase Transitions*. Walter de Gruyter, Berlin.

[13] Kalbfleisch, J. and Prentice, R. (1980). *The Statistical Analysis of Failure Time Data*. Wiley.

[14] Karr, A. (1985). *Point Processes and their Statistical Inference*. Dekker, New York.

[15] Lewis, P. (1972). Recent results in the statistical analysis of univariate point processes. In *Stochastic point processes*. ed. P. Lewis. Wiley, New York pp. 1–54.

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

[17] 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.

[18] Preston, C. (1976). *Random Fields*. Springer Verlag, Berlin-Heidelberg-New York.

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

[20] Ruelle, D. (1969). *Statistical mechanics*. John Wiley and Sons, New York.

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

[22] Takács, L. (1955). On stochastic processes connected with certain physical recording apparatuses. *Acta Math. Acad. Sci. Hung.* **6,** 363–374.

[23] Vervaat, W. (1979). On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Adv. Appl. Prob.* **11,** 750–783.

Appendix A. Study of the distributional equivalence

Here we consider the distributional equivalence (34) 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

$$F(x) = \mu x^\mu \int_{(x-1)^+}^\infty F(z)\frac{\mathrm{d}z}{(1+z)^{1+\mu}} \tag{36}$$

whereby (35) is verified.

A.1. Solutions

In principle we can solve (36) 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

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

For $\theta > 1$ the integral above is

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

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

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

A.3. Further notes on F

The Tauberian theorem for Laplace-Stieltjes transforms [9, p. 445] implies that

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

This comes effectively from Takacs [22, p. 376]. He observes that

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

where

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

Since clearly

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

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

$$f(x) = \frac{e^{-\gamma\mu}}{\Gamma(\mu)}\left[x^{\mu-1} + \sum_{n\ge 1}\frac{(-\mu)^n}{n!}H_n(x)\right] \tag{37}$$

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 (37) has only finitely many nonzero terms. Similarly

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

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 $1 \le x \le 2$ for all $n \ge 2$, we find that

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

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