Chapter 3

Bayesian Inference for Point Patterns

3.1. A general inference approach

Here, we consider inference exclusively within the Bayesian framework. As we have noted earlier, our general approach follows a simulation-based strategy predicated upon the ability to draw realizations (point patterns) under the model after model fitting. Customarily, model fitting is done using a Markov chain Monte Carlo algorithm. However, in Section 5.3 we discuss approximate Bayesian computation (ABC) model fitting which only requires simulation under the model and thus, fits in well with our general approach.

In particular, with posterior samples of model parameters, we can obviously do arbitrary inference on the model parameters. With posterior predictive samples of point patterns, we can perform arbitrary inference on any process features. With prior and/or predictive samples of point patterns, following Section 1.2, we can study model adequacy and model comparison, as well as make prior-posterior comparison. We adapt and extend familiar tools. In particular, for model adequacy, we propose examination of Bayesian residuals (drawing on the work of [9, 14]), both realized and predictive, as well as empirical coverage, and prior predictive checks through Monte Carlo tests. For model selection, we use predictive mean square error, empirical coverage, and ranked probability scores. We consider both in-sample and, when possible, out-of-sample approaches.

Specifically, at a high level, we consider the generic model form $[\mathcal{S}|\boldsymbol{\theta}][\boldsymbol{\theta}]$. We observe $\mathcal{S}_{obs}$. Upon model fitting, we obtain posterior samples $\boldsymbol{\theta}_l^*$ from $[\boldsymbol{\theta}|\mathcal{S}_{obs}]$. These enable inference about a function of $\boldsymbol{\theta}$, $b(\boldsymbol{\theta})$, assuming it is available explicitly. Next, using composition sampling, we create posterior predictive samples $\mathcal{S}_l^*$ from $[\mathcal{S}|\mathcal{S}_{obs}]$ by drawing $\mathcal{S}_l^*$ from $[\mathcal{S}|\boldsymbol{\theta}_l^*]$. Using these point pattern samples, we create posterior samples of any function h of $\mathcal{S}$ as $\{h(\mathcal{S}_l^*), l = 1, 2, \ldots, L\}$ from $[h(\mathcal{S})|\mathcal{S}_{obs}]$. *So, if we can fit and if we can sample, we can carry out arbitrary inference.*

Important Bayesian contributions for analyzing spatial point patterns have been made by Møller and colleagues [see, e.g., 142, 143] and references therein. In addition, there has been a recent strand which considers Poisson process models, focusing on a rich range of specifications for the intensity [see, e.g., 115, 195]. Some recent Bayesian work has employed integrated nested Laplace approximation (INLA) [176] for inference. Bayesian point pattern analysis using INLA can be found in [102, 104, 112]. Work that follows our inference paradigm is suggested in a conference address by Møller [139]. With regard to model adequacy, *posterior* predictive model checking has also been proposed. We can demonstrate that *prior* predictive model checking is needed for effective assessment of model adequacy. Finally, recent Bayesian model checking work, primarily validating intensities, is found in [194, 219, 220].

A primary feature we are trying to infer about is a (random) surface, i.e., an intensity. It is important to appreciate that $\lambda(\mathbf{s})$ informs about observed data points

but also about unobserved points. That is, a point pattern provides more information than just the locations of the observed points. *Absence* at other locations is informative [14]. This leads to a consequential challenge: we never observe a point on this surface. We can see the analogue with density estimation. In fact, we have empirical kernel intensity estimates, as shown in Section 2.2. However, unlike density estimation, here, the number of points is random. This implies that comparison of point patterns must be done with care.

To see the challenge in its simplest form, consider a homogeneous Poisson process (HPP) setting. Any observed point pattern will give a two-dimensional step surface which is not at all close to a flat surface and similarly a kernel intensity estimate which is not close to flat unless a very large bandwidth is employed (Section 2.2). How far from a constant surface can we be and still believe that a constant intensity is operating? In this regard, the null hypotheses, $H_o : \lambda(\mathbf{s}) = \lambda$ seems *silly*; it would never be operating in practice except perhaps under a simulation setting. Rather, it is more sensible to compare inference under an HPP model with that from other models.

In general, with spatial point patterns, it is hard to criticize models and to choose between models. For instance, a regression based intensity using spatially referenced covariates, say an NHPP, and a clustering-based intensity using say a Neyman Scott process, may equally well explain peaks in an intensity. Often, the choice will be made according to the nature of the process generating the observed data. There is not a large literature here (but we do review some below and the `spatstat` package [12] offers some tools). There is essentially no Bayesian work.

Returning to the general inference framework, let us elaborate posterior study of features. Posteriors of interest might include: for arbitrary sets A and B, $[N(A)|\mathcal{S}_{obs}]$, $[N(A), N(B)|\mathcal{S}_{obs}]$, $[N(A)|N(B), \mathcal{S}_{obs}]$, and $[\frac{N(A)}{N(D)}|\mathcal{S}_{obs}]$. We might also seek the posterior for the G and K functions under a given model. This is novel territory since the literature only considers empirical estimates of these function (Section 2.2). Comparison of the posterior distribution for G or K with the associated empirical estimate could be informative. Comparison, if appropriate, with the G or K functions under an HPP could also be informative. Here, comparison is through formal inference rather than exploratory comparison (again, as in Section 2.2).

Turning to model assessment, residuals are a commonly used tool. In particular, [9, 14] develop various notions of residuals for point patterns. For example, they define a *raw* residual, analogous to the standard residual from a regression model, as

$$R_{\hat{\boldsymbol{\theta}}}(B) \equiv N(B) - \int_B \hat{\lambda}(\mathbf{s}|\mathcal{S})d\mathbf{s} \tag{3.1}$$

for $B \subseteq D$, where $\hat{\lambda}(\mathbf{s}|\mathcal{S}) \equiv \lambda(\mathbf{s}|\mathcal{S}; \hat{\boldsymbol{\theta}})$ is the estimated Papangelou conditional intensity function. In the Bayesian setting, we would work with the *realized* residual, which considers the posterior of (3.1), employing $\lambda(\mathbf{s}|\mathcal{S}, \boldsymbol{\theta})$.

More generally, [14] define the h-weighted innovation measure as

$$I(B, h, \lambda) \equiv \sum_{\mathbf{s}_i \in \mathcal{S} \cap B} h(\mathbf{s}_i, \mathcal{S}\backslash\mathbf{s}_i) - \int_B h(\mathbf{s}, \mathcal{S}\backslash\mathbf{s})\lambda(\mathbf{s}|\mathcal{S})d\mathbf{s}. \tag{3.2}$$

These innovations have mean 0 under the true model, as can be seen using (3.7), developed below. Choices of h include $h(\mathbf{s}, \mathcal{S}\backslash\mathbf{s}) = 1/\lambda(\mathbf{s}|\mathcal{S})$ which defines the inverse λ residuals, in the spirit of [191]. With $h(\mathbf{s}, \mathcal{S}\backslash\mathbf{s}) = 1/\sqrt{\lambda(\mathbf{s}|\mathcal{S})}$, we obtain an

analogue of the Pearson residual from Poisson regression. Estimators are obtained by inserting an estimator of $\lambda(\mathbf{s}|\mathcal{S})$. A final residual, which we shall not consider here, is the pseudoscore residual, which sets $h(\mathbf{s}, \mathcal{S}\backslash\mathbf{s}) = \frac{\partial}{\partial\boldsymbol{\theta}} \log\{\lambda(\mathbf{s}|\mathcal{S})\}$, where $\boldsymbol{\theta}$ denotes the parameters of the intensity function λ.

Again, from a Bayesian perspective, the posterior distribution of $\int_B h(\mathbf{s}, \mathcal{S}\backslash\mathbf{s})$ $\lambda(\mathbf{s}|\mathcal{S})d\mathbf{s}$ and $I(B, h, \lambda)$ would be studied. In particular, these innovations are of the form $t(\mathcal{S}, \boldsymbol{\theta})$ and so their posteriors can be obtained through simulation as detailed below. We can use the posterior mean, $E(\int_B h(\mathbf{s}, \mathcal{S}\backslash\mathbf{s})\lambda(\mathbf{s}|\mathcal{S})d\mathbf{s} \,|\, \mathcal{S}_{obs})$, to obtain a point estimate and can also examine whether 0 falls in a given credible interval. [9, 14] provide formulas for the variance calculations of residuals and innovations.

With regard to validation, under a given model, consider credible intervals created from these innovation distributions developed over many sets. If the model is true, should we expect to achieve empirical coverage of 0 at roughly the nominal level? For the raw/realized innovations, the answer is clearly no. The raw innovations compare an observed count with the posterior distribution for the *expectation* of that count. Though we hope the expectations are close to the raw innovations, the credible intervals provide coverage for the expected counts rather than for the counts themselves. Thinking of the regression analogue, the raw innovations are akin to employing the distribution $[y - \mu_y|\text{Data}]$ when we should be employing the distribution for the *predictive* innovations, $[y - y_{\text{pred}}|\text{Data}]$.

So instead, we adopt *predictive residuals*,

$$R_{\text{pred}}(B) = N_{\text{obs}}(B) - N_{\text{pred}}(B), \tag{3.3}$$

where, as above, posterior samples $\mathcal{S}_l^*$ supply the draws $N_l^*(B)$, hence the posterior predictive distribution of $N_{\text{pred}}(B)$ and, in turn, of $R_{\text{pred}}(B)$.

Finally, for an h-scaled innovation as in (3.2), [14] define the smoothed innovation field $r(\mathbf{u}; \boldsymbol{\theta})$ at location $\mathbf{u} \in D$ as

$$\begin{aligned} r(\mathbf{u}; \boldsymbol{\theta}) &= e(\mathbf{u}) \int_D k(\mathbf{u} - \mathbf{v}) dI(\mathbf{v}, \mathbf{h}, \boldsymbol{\theta}) \\ &= e(\mathbf{u}) \left[\sum_{\mathbf{s}_i \in \mathcal{S}} k(\mathbf{u} - \mathbf{s}_i)\mathbf{h}(\mathbf{s}_i, \mathcal{S}\backslash\{\mathbf{s}_i\}) - \int_D k(\mathbf{u} - \mathbf{v})\mathbf{h}(\mathbf{v}, \mathcal{S})\lambda(\mathbf{v}|\mathcal{S}; \boldsymbol{\theta}) d\mathbf{v} \right], \end{aligned} \tag{3.4}$$

where $k(\mathbf{s})$ is a probability density on $\mathbb{R}^2$ used as a smoothing kernel and $e(\mathbf{u}) \equiv 1/\int_D k(\mathbf{u}-\mathbf{v})d\mathbf{v}$ is an edge correction. This field puts positive atoms at each $\mathbf{s}_i \in \mathcal{S}$ and a negative value elsewhere and then smoothes using the kernel. So, positive values indicate locations where the empirical intensity was higher than the intensity of the fitted model while negative values indicate areas where the intensity of the fitted model was higher.

[14] estimate $\boldsymbol{\theta}$ to obtain a residual field, $r(\mathbf{u}; \hat{\boldsymbol{\theta}})$. For us, for the NHPP and LGCP models, with a posterior distribution for $\lambda(\mathbf{s}; \boldsymbol{\theta})$, we can obtain a posterior distribution for $r(\mathbf{u}; \boldsymbol{\theta})$. Additionally, we can create a plot showing those regions that have a credible interval (for the smoothed innovation) which contains 0, as well as those regions that have a credible interval above or below 0.

To summarize our overall inference strategy in light of the above, under the model, suppose we have interest in $b(\boldsymbol{\theta})$ using $[b(\boldsymbol{\theta})|\mathcal{S}_{obs}]$. With posterior samples $\{\boldsymbol{\theta}_l^*\}$, we obtain $\{b(\boldsymbol{\theta}_l^*)\}$. If interest is in $[h(\mathcal{S})|\mathcal{S}_{obs}]$, then for each $\boldsymbol{\theta}_l^*$, we generate $\mathcal{S}_l^*$ obtaining $\{\mathcal{S}_l^*\}$ and thus $\{h(\mathcal{S}_l^*)\}$.

Often, $b(\boldsymbol{\theta})$ is not available explicitly (see below). Then, we need to find an $h(\mathcal{S})$ such that $E(h(\mathcal{S})|\boldsymbol{\theta}) = b(\boldsymbol{\theta})$. As a result, in order to obtain a $b(\boldsymbol{\theta}_l^*)$, for each $\boldsymbol{\theta}_l^*$, we

need to generate replicates, the set $\{\mathcal{S}^*_{lb}, b = 1, 2, \ldots, B\}$. These replicates provide a Monte Carlo integration for $b(\boldsymbol{\theta}^*_l)$, i.e., $\frac{1}{B}\sum_b h(\mathcal{S}^*_{lb})$.

The most general objects of interest would take the form $t(\mathcal{S}, \boldsymbol{\theta})$, leading to the posterior, $[t(\mathcal{S}, \boldsymbol{\theta})|\mathcal{S}_{obs}]$. With t available explicitly (as it will be in practice), we can use the $\{\boldsymbol{\theta}^*_l, \mathcal{S}^*_l\}$ to create posterior draws.

Again, examples of $b(\boldsymbol{\theta})$ include $E(N(A)|\boldsymbol{\theta})$ and $E(N(A)N(B)|\boldsymbol{\theta})$. They also include $\lambda(\mathbf{s}; \boldsymbol{\theta})$, $\gamma(d; \boldsymbol{\theta})$, $\lambda(A; \boldsymbol{\theta})$, $g(d; \boldsymbol{\theta})$ (the pair correlation function), $G(d; \boldsymbol{\theta})$, and $K(d; \boldsymbol{\theta})$. To be more explicit, $E(N(A)|\mathcal{S}_{\text{obs}}) \approx \frac{1}{L}\sum_{l=1}^{L}\sum_{\mathbf{s}^*_{li}\in\mathcal{S}^*_l} 1(\mathbf{s}^*_{li} \in A)$. This suggest how we could create model-based Bayesian intensity estimates in the setting where we do not have an explicit form for $\lambda(\mathbf{s})$. (With an explicit form, we would directly obtain the posterior mean $E(\lambda(\mathbf{s})|\mathcal{S}_{obs})$ from posterior samples and plot this as a surface using a fine grid of $\mathbf{s}$.) Taking $A = \partial\mathbf{s}$ yields the Bayes estimator for $\lambda(\partial\mathbf{s}) \approx \lambda(\mathbf{s})|\partial\mathbf{s}|$, hence for $\lambda(\mathbf{s})$. Therefore, again with a fine grid of $\mathbf{s}$, a Bayesian estimator of the intensity surface results. The size of $\partial\mathbf{s}$ can be viewed as analogous to a bandwidth selection for a kernel intensity estimate. In fact, recall the usual *kernel* smoothing yields kernel intensity estimate, $\lambda_\tau(\mathbf{s}) = \frac{1}{\tau^2}\sum_{\mathbf{s}_i\in\mathcal{S}} h(||\mathbf{s} - \mathbf{s}_i||/\tau)$ discussed in Section 2.2. We can contrast the model-based posterior estimate of the intensity with the empirical kernel intensity estimate.

Turning things around, if we can write λ as a parametric function, $\lambda(\mathbf{s}; \boldsymbol{\theta})$ (say for an NHPP but not for a LGCP), posterior samples of $\boldsymbol{\theta}$ yield an estimate of $\lambda(\mathbf{s}; \boldsymbol{\theta})$. Then, numerical integration would enable a posterior estimate of $\lambda(A; \boldsymbol{\theta})$. Details for the G and K functions, as parametric functions, are presented below.

Examples of $h(\mathcal{S})$'s include: $N(A)$, $(N(A), N(B))$, $\frac{N(A)}{N(D)}$ along with the foregoing predictive residuals , e.g., the distribution $[N_{obs}(A) - N(A)|\mathcal{S}_{obs}]$. We could also estimate conditional events with distribution $[N(A)|N(B) = m; \mathcal{S}_{obs}]$. Examples of $t(\mathcal{S}, \boldsymbol{\theta})$ include: realized residuals, e.g., the distribution of $[N(A) - \lambda(A; \boldsymbol{\theta})|\mathcal{S}_{obs}]$ and the inhomogeneous K function $K_{inhom}(d; \boldsymbol{\theta})$ [15] discussed below.

3.1.1. Campbell's Theorem and the GNZ result

The main theoretical tool we employ here is Campbell's Theorem [101], which gives the expectation of the summation over $\mathcal{S} \cap D$ of a function $h(\mathbf{s}_i)$ (restriction to D ensures that realizations of $\mathcal{S}$ are finite so that expectations exist). It states that

$$E_{\mathcal{S}\cap D}\Big(\sum_{\mathbf{s}_i\,\in\,\mathcal{S}\cap D} h(\mathbf{s}_i)\Big) = \int_D h(\mathbf{s})\lambda(\mathbf{s})\,d\mathbf{s}. \tag{3.5}$$

For example, letting $h(\mathbf{s}) = 1(s \in A)$ for some set $A \subset D$, Campbell's Theorem says that $\sum_{\mathbf{s}_i\in\mathcal{S}} 1(\mathbf{s}_i \in A)$ is an unbiased estimator for $\int_D 1(s \in A)\lambda(\mathbf{s})\,d\mathbf{s} = \int_A \lambda(\mathbf{s})\,d\mathbf{s} = \lambda(A)$ based on $\mathcal{S}$, i.e., $N(A)$ is an unbiased estimator of $\lambda(A)$, as we already know. However, more generally, $\sum_{\mathbf{s}_i\in\mathcal{S}\cap D} h(\mathbf{s}_i)$ is an unbiased estimator of $\int_D h(\mathbf{s})\lambda(\mathbf{s})d\mathbf{s}$. More usefully, (3.5) suggests Monte Carlo integration to obtain the right side. With samples $\mathcal{S}^*_l$? from $[\mathcal{S} \cap D|\lambda(\mathbf{s})]$, or, more generally, $[\mathcal{S} \cap D|\boldsymbol{\theta}]$, averaging the sum on the left side of (3.5) over these replicates provides a Monte Carlo integration for the left side, hence for the right side. That is, we can compute integrations relative to $\lambda(\mathbf{s})$ even if we can not obtain $\lambda(\mathbf{s})$ explicitly!

The utility of this approach for Bayesian inference is immediately evident. Now, $\lambda(\mathbf{s})$ is random and so, for a given h as above, the right side of (3.5) is random. However, with posterior (prior) predictive samples of point patterns, we can directly create a Monte Carlo integration for the expectation on the left side, hence a Monte

Carlo integration for the posterior (prior) mean of the integral on the right side. Again, the right side of (3.5) is viewed as $b(\boldsymbol{\theta})$ which typically will not be available explicitly. So, for example, to obtain $E(\int_D h(\mathbf{s})\lambda(\mathbf{s})d\mathbf{s}|\mathcal{S}_{obs})$, posterior draws of $\lambda_l(\mathbf{s})$ would provide posterior predictive draws, $\mathcal{S}_l^*$?, to which we apply the sum on the left side of (3.5). Averaging over these draws produces the posterior mean for the right side.

Similarly, Campbell's Theorem has a bivariate form for h, a function of two points in $\mathcal{S}$:

$$E_{\mathcal{S}\cap D}\Big(\sum_{\substack{\mathbf{s}_i,\,\mathbf{s}_j\,\in\,\mathcal{S}\cap D\\ i\neq j}} h(\mathbf{s}_i,\mathbf{s}_j)\Big) = \int_D\int_D h(\mathbf{s},\mathbf{s}')\gamma(\mathbf{s},\mathbf{s}')\,d\mathbf{s}\,d\mathbf{s}'. \tag{3.6}$$

(3.6) is useful for exploring second-order properties of a point process, e.g., the second order intensity which is defined (Section 2.1.2) through $h(\mathbf{s},\mathbf{s}') = 1(\mathbf{s}\in A, \mathbf{s}'\in B)$. Another application arises for the Strauss process (Section 2.5) where we consider the "close pairs" function (Section 4.3), $s_R(\mathcal{S}) = \sum_{\mathbf{s}_i,\mathbf{s}_j\in\mathcal{S}\subset D} 1(||\mathbf{s}_i-\mathbf{s}_j||\le R)$. Applying Campbell's Theorem, we find $E_{\mathcal{S}\cap D}(s_R(\mathcal{S})) = \int_D\int_D 1(||\mathbf{u}-\mathbf{v}||\le R)\gamma(\mathbf{u},\mathbf{v})d\mathbf{u}d\mathbf{v} = \int\int_{||\mathbf{u}-\mathbf{v}||\le d}\gamma(\mathbf{u},\mathbf{v})d\mathbf{u}d\mathbf{v}$. Therefore, it enables similar Monte Carlo integration for the posterior mean of the right side.

A more general result is the Georgii-Nguyen-Zessin (GNZ) formula [80, 218], which applies to h of the form $h(\mathbf{s};\mathcal{S}\backslash\{\mathbf{s}\})$ and gives the equality

$$E_{\mathcal{S}\cap D}\big(\sum_{\mathbf{s}_i\in\mathcal{S}} h(\mathbf{s}_i,\mathcal{S}\backslash\{\mathbf{s}_i\})\big) = E_{\mathcal{S}\cap D}\big(\int_D h(\mathbf{s},\mathcal{S}\backslash\mathbf{s})\lambda(\mathbf{s}|\mathcal{S})d\mathbf{s}\big), \tag{3.7}$$

where $\lambda(\mathbf{s}|\mathcal{S})$ is the Papangelou conditional intensity. Again, Monte Carlo integration enables a posterior mean for the right side. Here, a choice for $h(\mathbf{s}_i|\mathcal{S}\backslash\mathbf{s}_i)$ might be $1(\min_{\mathbf{s}_j\in\mathcal{S}\backslash\mathbf{s}_i}||\mathbf{s}_i-\mathbf{s}_j||\le d)$ for a given d which connects to the G function (Section 2.2) shown below.

Returning to Campbell's Theorem, it was noted above that by summing over the indicator function $1(\mathbf{s}_i\in A)$, we provide an unbiased estimator for $E(N(A)) = \lambda(A;\boldsymbol{\theta})$ whose usual Bayes estimate is $E(\lambda(A;\boldsymbol{\theta})|\mathcal{S}_{obs})$. If $\lambda(A;\boldsymbol{\theta})$ is available explicitly, a Monte Carlo integration for $E(\lambda(A;\boldsymbol{\theta})|\mathcal{S}_{obs})$ is $\frac{1}{L}\sum_l \lambda(A;\boldsymbol{\theta}_l^*)$. When we cannot calculate $\lambda(A;\boldsymbol{\theta})$, we note that $E(\lambda(A;\boldsymbol{\theta})|\mathcal{S}_{obs}) = E(N(A)|\mathcal{S}_{obs}) \approx \frac{1}{L}\sum_{l=1}^L\sum_{\mathbf{s}_{li}^*\in\mathcal{S}_l^*} 1(\mathbf{s}_{li}^*\in A)$, providing the desired Monte Carlo integration as shown at the end of Section 3.1. Of course, the elements of the set $\{\sum_{\mathbf{s}_{li}^*\in\mathcal{S}_l^*} 1(\mathbf{s}_{li}^*\in A)\}_{l=1}^L$ provide posterior samples of $N(A)$.

Summarizing, we may be interested in inference on $b(\boldsymbol{\theta})$ based upon $[b(\boldsymbol{\theta})|\mathcal{S}_{obs}]$. With posterior samples, $\{\boldsymbol{\theta}_l^*\}$, we obtain $\{b(\boldsymbol{\theta}_l^*)\}$ for such inference, as usual. If interest is in $[h(\mathcal{S})|\mathcal{S}_{obs}]$, then the set $\{\mathcal{S}_l^*\}$ provides the set $\{h(\mathcal{S}_l^*)\}$ for inference. For $[t(\mathcal{S},\boldsymbol{\theta})|\mathcal{S}_{obs}]$ with t available explicitly, we can use $\{\boldsymbol{\theta}_l^*,\mathcal{S}_l^*\}$. Again, if $b(\boldsymbol{\theta})$ is not available explicitly, the strategy is then to find $h(\mathcal{S})$ such that $E(h(\mathcal{S})|\boldsymbol{\theta}) = b(\boldsymbol{\theta})$.

With regard to the G and K functions, they are parametric functions of the form $G(d;\boldsymbol{\theta})$ and $K(d;\boldsymbol{\theta})$, respectively. Except in special cases, they are not available in closed form, leading to the use of empirical, rather than model-based, estimation. However, we have expressions (2.2) and (2.3) from Section 2.2. For each of these, the summation on the left side is over terms of the form $h(\mathbf{s},\mathcal{S}\backslash\mathbf{s})$. Hence the GNZ result in (3.7) applies enabling Monte Carlo integrations with respect to the Papangelou conditional intensity. In other words, these Monte Carlow integrations can provide the posterior mean of the G and K functions under a specified model given $\mathcal{S}_{obs}$.

That is, we obtain model based estimates for G and K rather than the empirical estimates supplied in Section 2.2. Comparison between the model based estimate and the empirical estimate can be enlightening with regard to the adequacy of the model. The same argument can be applied to the inhomogeneous K function [15] defined above. However, there we have a form $t(\mathcal{S}, \boldsymbol{\theta})$ to which we would apply the results as well.

The discussion of residuals above (Section 3.1) provides another set of quantities of interest to which we can apply the foregoing. For example, suppose $h(\mathbf{u}; \mathcal{S}\backslash\mathbf{u}) = 1(\mathbf{u} \in B)$. This yields $E_{\mathcal{S}} N(\mathcal{S} \cap B) = \int_B E_{\mathcal{S}} \lambda(\mathbf{u}|\mathcal{S}) d\mathbf{u}$. In turn, this suggests $N(\mathcal{S} \cap B) - \int_B \lambda(\mathbf{s}|\mathcal{S}) d\mathbf{s}$, the realized *innovation* residuals, which have mean 0 [14]. Suppose $h(\mathbf{u}; \mathcal{S}\backslash\mathbf{u}) = 1(\mathbf{u} \in B)/\lambda(\mathbf{u}|\mathcal{S})$. This yields $E_{\mathcal{S}}(\sum_{\mathbf{s}_i \in \mathcal{S}} 1(\mathbf{s}_i \in B)/\lambda(\mathbf{s}_i|\mathcal{S}\backslash\mathbf{s}_i)) = |B|$ [191], the so-called "inverse" residuals. Application to other *scaled* residuals is clear.

Consider the GNZ applied to a LGCP with random intensity $\log\lambda(\mathbf{s}) = z(\mathbf{s})$ where $z(\mathbf{s})$ is a GP. Then, given D, $E_{\mathcal{S}\cap D}(\sum_{\mathbf{s}_i \in \mathcal{S}\cap D} h(\mathbf{s}_i; (\mathcal{S} \cap D)\backslash\mathbf{s}_i) = \int_D h(\mathbf{s}) E(\lambda(\mathbf{s})) d\mathbf{s}$ if h depends only on $\mathbf{s}$. Again, with restriction to D, a finite point pattern and posterior samples immediately provide a Monte Carlo integration yielding a Bayes estimate of the right side.

Suppose $h(\mathbf{u}; \mathcal{S}\backslash\mathbf{u}) = 1(\mathbf{u} \in \partial\mathbf{s})$ yielding $E_{\mathcal{S}} N(\mathcal{S} \cap \partial\mathbf{s}) = \int_{\partial\mathbf{s}} E_{\mathcal{S}} \lambda(\mathbf{u}|\mathcal{S}) d\mathbf{u}$. The left side is $\int_{\partial\mathbf{s}} \lambda(\mathbf{u}) d\mathbf{u} \approx \lambda(\mathbf{s})|\partial\mathbf{s}|$. The right side is $E_{\mathcal{S}}(\int_{\partial\mathbf{s}} \lambda(\mathbf{u}|\mathcal{S}) d\mathbf{u}) = E_{\mathcal{S}}(\lambda(\partial\mathbf{s}|\mathcal{S})) \approx E_{\mathcal{S}}(\lambda(\mathbf{s}|\mathcal{S}))|\partial\mathbf{s}|$. So, marginally, $\lambda(\mathbf{s}) \approx E_{\mathcal{S}}(\lambda(\partial\mathbf{s})|\mathcal{S})$.

Continuing with the GNZ result, now suppose $h_D(\mathcal{S}) \equiv \sum_{\mathbf{s}_i \in \mathcal{S}\cap D} h(\mathbf{s}_i; (\mathcal{S} \cap D)\backslash\mathbf{s}_i)$. Then, $E_{\mathcal{S}\cap D|\boldsymbol{\theta}}(h(\mathcal{S})) = E_{\mathcal{S}\cap D|\boldsymbol{\theta}}(\int_D h(\mathbf{s}; (\mathcal{S} \cap D)\backslash\mathbf{s}) \lambda(\mathbf{s}|\mathcal{S}) d\mathbf{s}) \equiv b_{h_D}(\boldsymbol{\theta})$. In order to achieve a normalization, we propose to work with $\bar{h}_D(\mathcal{S}) \equiv h_D(\mathcal{S})/N(\mathcal{S} \cap D)$. If $N(\mathcal{S} \cap D) = 0$, then $h_D(\mathcal{S}) = 0$ and we define $\frac{0}{0} = 1$. So, we consider $E_{\mathcal{S}\cap D|\boldsymbol{\theta}}(\bar{h}(\mathcal{S})) \equiv b_{\bar{h}_D}(\boldsymbol{\theta})$. Evidently, we need a different version of the GNZ result which we offer in the next subsection.

3.1.2. An iterated expectation version

We turn to a different way of calculating expectations which suggests a different way of developing Monte Carlo integrations. We can imagine that our model can provide a realization $\mathcal{S}$ over $\mathbb{R}^2$ which induces $\mathcal{S} \cap D$ over D with an associated finite $N(\mathcal{S} \cap D)$. Alternatively, suppose, given D, we first generate $N(\mathcal{S} \cap D) = n$ and then we locate $\mathcal{S}$ over D given $N(\mathcal{S} \cap D) = n$, assuming the $\mathbf{s}_i$ are exchangeable. This is the *generative* view that we have noted before for, e.g., a NHPP or a cluster process as opposed to a *modeling* or mechanistic view, e.g., a Gibbs process.

Regardless, there is a joint distribution $[\mathcal{S} \cap D, N(\mathcal{S} \cap D)]$, hence a conditional times marginal version $[\mathcal{S} \cap D|N(\mathcal{S} \cap D)][N(\mathcal{S} \cap D)]$. We may not be able to write down these densities explicitly but, formally, we can calculate the expectation iteratively (and we can obtain expectations explicitly in certain cases as we show below).

That is,

$$E_{\mathcal{S}\cap D}\Big(\sum_{\mathbf{s}_i \in \mathcal{S}\cap D} h(\mathbf{s}_i; (\mathcal{S} \cap D)\backslash\mathbf{s}_i)\Big) = E_{N(\mathcal{S}\cap D)} E_{\mathcal{S}\cap D|N(\mathcal{S}\cap D)} \sum_{\mathbf{s}_i \in \mathcal{S}\cap D} h(\mathbf{s}_i; (\mathcal{S} \cap D)\backslash\mathbf{s}_i)$$
$$= E_{N(\mathcal{S}\cap D)}(N(\mathbf{s} \cap D) E_{\mathcal{S}\cap D|N(\mathcal{S}\cap D}(h(\mathbf{s}, (\mathcal{S} \cap D)\backslash\mathbf{s})).$$

And, using the normalized form (and defining $0/0 = 1$), $E_{\mathcal{S}\cap D}(\sum_{\mathbf{s}_i \in \mathcal{S}\cap D} h(\mathbf{s}_i; (\mathcal{S} \cap D)\backslash\mathbf{s}_i))/N(\mathcal{S} \cap D) = E_{\mathcal{S}\cap D} h(\mathbf{s}; (\mathcal{S} \cap D)\backslash\mathbf{s})$. Attractively, we remove $N(\mathcal{S} \cap D)$ from the right side. With this notation, $E_{\mathcal{S}\cap D|\boldsymbol{\theta}} \bar{h}_D(\mathcal{S}) = b_{\bar{h}_D}(\boldsymbol{\theta})$ where $b_{\bar{h}_D}(\boldsymbol{\theta}) =$

$E_{\mathcal{S}\cap D|\boldsymbol{\theta}}h(\mathbf{s};(\mathcal{S}\cap D)\backslash\mathbf{s})$. Below, we propose choices for $h(\mathbf{s};(\mathcal{S}\cap D)\backslash\mathbf{s})$ which are of interest. A usual Bayes estimate for $b_{\bar{h}_D}(\boldsymbol{\theta})$ is $E(b_{\bar{h}_D}(\boldsymbol{\theta})|\mathcal{S}_{obs})$. With posterior samples, $\{\boldsymbol{\theta}_l^*\}$, a Monte Carlo integration for the posterior mean is $\frac{1}{L}\sum_l b_{h_D}(\boldsymbol{\theta}_l^*)$. Of course, typically, we can not calculate $b_{\bar{h}_D}(\boldsymbol{\theta})$ explicitly. However, from above, $E_{\mathcal{S}\cap D|\mathcal{S}_{obs}}\bar{h}(\mathcal{S}) = E_{\boldsymbol{\theta}|\mathcal{S}_{obs}}E_{\mathcal{S}\cap D|\boldsymbol{\theta}}\bar{h}(\mathcal{S}) = E_{\boldsymbol{\theta}|\mathcal{S}_{obs}}b_{\bar{h}}(\boldsymbol{\theta})$. So, a direct Monte Carlo integration becomes $\frac{1}{L}\bar{h}(\mathcal{S}_l^*)$.

Suppose we want posterior samples of $b_{\bar{h}_D}(\boldsymbol{\theta})$. Obviously, they are $\{b_{\bar{h}_D}(\boldsymbol{\theta}_l^*)\}$. But again, we can't calculate $b_{\bar{h}_D}(\boldsymbol{\theta}_l^*)$. With a sample of point patterns $\{\mathcal{S}_{lb}^*, b = 1,2,\ldots,B\}$ from $[\mathcal{S}|\boldsymbol{\theta}_l^*]$, we have Monte Carlo integrations for $\{b_{\bar{h}_D}(\boldsymbol{\theta}_l^*)\}$. Altogether, we need a *nested* sampling of point patterns to obtain the desired posterior samples.

3.2. Model adequacy and model comparison

Model assessment using a fitting/training sample and an independent validation/test sample is now becoming standard practice. With point pattern data, such an approach may not be available. With a conditionally independent location distribution, as with NHPPs and LGCPs, the answer is yes. However, with an inhibition model, holding out points will alter the geometry of the point pattern. It will change the nature of the inter-point distances, hence the interaction structure. This will be true in general for a point pattern model where there is dependence between the locations of the points as with a Gibbs process.

For models with conditionally independent locations, we develop training and test datasets. Suppose we decide to administer 20% holdout. We can not simply remove 20% of the data at random. This will *fix* the size of the point pattern rather than allowing it to be random. Rather, the p-thinning approach, as in [101], can be applied to create appropriate training and test data. Letting p denote the retention probability, p-thinning proceeds by independently, point-by-point, removing $\mathbf{s}_i \in \mathcal{S}$ with probability $1-p$. This produces a training point pattern $\mathcal{S}_{train}$ and test point pattern $\mathcal{S}_{test}$, which are independent, conditional on $\lambda(\mathbf{s})$. In fact, $\mathcal{S}_{train}$ has intensity $p\lambda(\mathbf{s})$, $\mathcal{S}_{test}$ has intensity $(1-p)\lambda(\mathbf{s})$, and the revised validation intensity compared with the fitting intensity is $\lambda_{test}(\mathbf{s}) = \left(\frac{1-p}{p}\right)\lambda_{train}(\mathbf{s})$.

3.2.1. Model adequacy through empirical coverage

When cross-validation is possible, for a given set B, posterior predictive point patterns will supply the posterior predictive distribution of $N(B)$. The predictive residuals discussed in Section 3.1 should be centered around zero for an adequate model. If we look at many subregions B_k, we expect the empirical coverage to be roughly the nominal level of coverage if the model is adequate. How shall we create a set $\{B_k\}$? [14, section 11.1] propose to analyze a set of residuals over disjoint partitions B_k of the domain, similar to quadrat counting [54]. With an irregular domain D, division into disjoint subregions of similar size can be time-consuming and is, in fact, unnecessary. We suggest to draw random subregions uniformly over D and then evaluate the residuals or innovations in each subregion. Moreover, there is no reason to require the B_k be disjoint. If not, this allows us to draw as many B_k as desired, subject to the requirement that each B_k has the same area. Denote the area of each B_k by $q|D|$ where $q \in (0,1)$ such that q represents the size of each B_k relative to D. For various q's we can evaluate the innovation or residual measures on

each of the B_k's and obtain the observed empirical coverage of 0. Intuition suggests that larger regions will validate better than smaller ones.

In the sequel, we take the shape of each B_k to be a square but, depending upon D, there may be some reason to choose the shape more carefully. The use of squares sometimes limits the placement of the B_k when q is large and also access to the edges of D. Furthermore, with randomly placed, overlapping B_k, it can be hard to identify regions where the model fits poorly. Disjoint B_k alleviate this problem but, with regard to empirical coverage, the success rate of Bernoulli trials based upon random B_k's will suffice.

In-sample model adequacy

When we can not develop a test data set, how can we investigate model adequacy? Evidently, the foregoing empirical coverage approach can be implemented in-sample. However, we would not expect it to criticize the model well. This leads to the work on posterior model checks [79, henceforth GMS] and prior model checks [50, henceforth DGSV] discussed in Section 1.2. As noted there, GMS is more common and easier to do. However, it doesn't criticize the model well enough and uses the data twice (once to fit, once to check). DGSV is more computationally demanding but is formally more coherent and uses the data only once.

We elaborate both the GMS and DGSV approaches in the context of spatial point patterns. Both employ Monte Carlo tests in looking at discrepancy measures, $D(\mathcal{S};\boldsymbol{\theta})$ which, for instance, might take the form of a realized residual, $N(A) - \lambda(A;\boldsymbol{\theta})$. GMS looks at draws from $[D(\mathcal{S};\boldsymbol{\theta})|\mathcal{S}_{obs}]$ and compares with draws from $[D(\mathcal{S}_{obs};\boldsymbol{\theta})|\mathcal{S}_{obs}]$. The problem is evident; the data is used twice. Draws of $\mathcal{S}_l^*$ from $[\mathcal{S};\boldsymbol{\theta}|\mathcal{S}_{obs}]$ will look too much like $\mathcal{S}_{obs}$; discrepancies, $D(\mathcal{S}_l^*,\boldsymbol{\theta})$, will look too much like $D(\mathcal{S}_{obs};\boldsymbol{\theta})$. The model checking will not be critical enough.

DGSV create draws from $[D(\mathcal{S},\boldsymbol{\theta})|\mathcal{S}]$ by sampling $\boldsymbol{\theta}$ from the prior, then sampling $\mathcal{S}$ under the model given $\boldsymbol{\theta}$. Fitting the model enables draws from $[\boldsymbol{\theta}|\mathcal{S}]$ and, hence, with a collection $\mathcal{S}_l^*$, draws from $[D(\mathcal{S}_l^*,\boldsymbol{\theta})|\mathcal{S}_l^*]$. Then, comparison is made between $[D(\mathcal{S}_l^*,\boldsymbol{\theta})|\mathcal{S}_l^*]$ and $[D(\mathcal{S}_{obs};\boldsymbol{\theta})|\mathcal{S}_{obs}]$. This is an "apples vs. apples" comparison which uses the data only once. That is, DGSV compare the observed discrepancy with the discrepancies you expect under the model; GMS compare the observed discrepancy with what you expect under the model **and** the observed data. The computational demand required for DGSV is evident; one must fit and sample for every $\mathcal{S}_l^*$.

In-sample, our empirical coverage model adequacy check also suffers the GMS problem; it will not be critical enough. Again, for a collection of B_k's, we look at the set $\{[N_{obs}(B_k) - N(B_k)|\mathcal{S}_{obs}]\}$ and check empirical coverage relative to nominal coverage. We see that the $\mathcal{S}_l^*$'s will be too similar to $\mathcal{S}_{obs}$ (using noninformative priors) so the $N(B_k)$ that we generate given $\mathcal{S}_{obs}$ will tend to look too much like $N_{obs}(B_k)$, since the latter is a function of $\mathcal{S}_{obs}$. So, we assert that there is no role for empirical coverage here unless we can do it out-of-sample. In-sample empirical coverage will be inadequate to criticize the model.

As an alternative, it is better to generate $N_l^*(B)$ through $\mathcal{S}_l^*$'s from the marginal distribution rather than from the posterior distribution. Now, we can run a Monte Carlo test comparison between $[N_{obs}(B_k) - N(B_k)|\mathcal{S}_{obs}]$ and $\{[N_l^*(B_k) - N(B_k)|\mathcal{S}_l^*]\}$. Unfortunately, this demands a lot of comparison. For each B_k, we compare an "observed" posterior distribution vs. say 99 generated posterior distributions, say using quantiles. This presents the challenge of lots of simultaneous inference.

Let's consider a simpler checking function approach which can be expected to supply model criticism through the prior predictive framework without requiring

the computation associated with DGSV. Consider $h(\mathcal{S})$, a function only of the point pattern. For instance, in assessing the adequacy of an HPP or a Strauss process, given a radius R, suppose we consider the statistic, $\mathbf{s}_R(\mathcal{S})$ discussed in Section 2.5. We can implement a Monte Carlo test for $\mathbf{s}_R(\mathcal{S}_{obs})$ and the set $\{\mathbf{s}_R(\mathcal{S}_l^*), l = 1, 2, \ldots, L\}$ where the $\mathcal{S}_l^*$'s are generated under the model. If there is interaction between the points in $\mathcal{S}$, then, as we run through a set of R's (motivated by the size of the region), these Monte Carlo tests should criticize the HPP model but potentially support Strauss process models in the vicinity of a suitable R. Unfortunately, we may need a large number of points in the point pattern to extract criticism.

3.2.2. Model comparison

Model selection tools are notably lacking for point pattern models. A typical analysis uses ad hoc tests of the homogeneity and independence assumptions of CSR but, having decided which assumption to relax, there is no clear procedure for comparing models. Often model comparison is not even considered; as we noted earlier, a model is adopted on mechanistic or behavioral grounds. Lack of fit using the methods described above can eliminate some models but will not help when choosing among adequately fitting models.

Again, following Section 1.2, we argue that model comparison should be done in predictive space since parameters have no meaning across models. So, then the question is, "What would we be predicting?" A natural choice would focus on the distribution, $[N(A)|\mathcal{S}_{obs}]$ for $A \subset D$. In particular, we would compare $N_{obs}(A)$ with $[N(A)|\mathcal{S}_{obs}; M_j]$ for each model, $j = 1, 2, \ldots, J$. Here, for model j with parameters $\boldsymbol{\theta}_j$, we obtain posterior samples, $\boldsymbol{\theta}_{j,l}^*$ and then $\mathcal{S}_{j,l}^*$. Again, we would want to do this out of sample through p-thinning, as with NHPPs, LGCPs, and cluster processes, which are superpositions of NHPPs.

As for criteria, we can look at predictive mean square error (PMSE), perhaps normalized by the expected number (the usual loss function for Poisson counts) and we can look at ranked probability scores (RPS) [85]. The RPS arises from proper scoring rules and offers a useful metric for assessing the performance of a predictive distribution. That is, it compares an entire distribution (in this case a posterior predictive distribution) to the observed value. The more concentrated the distribution is around the observation, the smaller the RPS. That is, RPS prefers models which provide predictions that are concentrated around the observed value. For count data, the RPS is appropriate [63]. Specifically, the RPS compares the posterior predictive distribution for a cell count with the degenerate distribution associated with the observed cell count using a sum of squares over the set of support values $\{0, 1, 2, \ldots\}$.

We propose choosing subregions B_k uniformly over D, with each B_k having the same size and potentially overlapping other $B_{k'}$. In fact, we can use the same B_k as in the Monte Carlo assessment above. We obtain $N(B_k)$ from the hold-out dataset and compare with $[N(B_k|\mathcal{S}_{fitted})]$ using posterior predictive point patterns. For any B_k, we can write the RPS as $\text{RPS}(B_k) = \sum_{n=0}^{\infty}[F_{N(B_k)|\mathcal{S}_{fitted}}(n) - 1[n \geq N_{obs}(B_k)]]^2$. We would average over k to compare models. That is, model selection would choose the model with the smallest average RPS.

If cross-validation is available, we would employ the RPS with our hold-out data, comparing observed counts in subsets to posterior predictive distributions for these counts. If holding out data is not possible, we would examine these same metrics in-sample.

Finally, we remark that we can't use diagnostics like G, F, K, and K_{inhom} to compare models. For example, we can't say that G for one model is "better" than G for another model? That is, posterior distributions, e.g., $[G(d : \boldsymbol{\theta}_j)|\mathcal{S}_{obs}; M_j]$, can criticize say CSR which has known distance functions when CSR is nested within the fitted model. With a set of models say $\{M_j\}$, using the G function, we can compare, e.g., $[G(d : \boldsymbol{\theta}_j)|\mathcal{S}_{obs}; M_j]$ with the empirical estimate $\hat{G}(d)$. Since the latter is a *nonparametric* estimate, such comparison could be used to criticize M_j. However, with regard to the K function, since it involves parameters, the empirical estimate will be semiparametric with parameter estimates based upon some model. In any event, plotted across values of d, we see an analogy with theoretical Q-Q plots.

3.2.3. Bayesian edge correction

We conclude this chapter with a brief section on Bayesian edge correction. As with all of the other approaches we have discussed, the edge correction we propose here will also be simulation based. We start with the G function.

Recall the definition of $N_D(\mathbf{s}_i, d, \mathbf{S}) \equiv N(\mathbf{s}_i, d, \mathbf{S} \cap D)$, the number of points in $\mathcal{S} \cap D$ within distance d of location $\mathbf{s}_i \in \mathcal{S}$. There can be points in $\mathcal{S}$ (viewed over $\mathbb{R}^2$) that are within d of $\mathbf{s}_i$. The number of these points is $N(\mathbf{s}_i, d, \mathcal{S})$. We are interested in events associated with the latter but we only observe the former, $N_D(\mathbf{s}_i, d, \mathbf{S}) \leq N(\mathbf{s}_i, d, \mathcal{S})$. This is the challenge of edge correction.

From our previous notation, for a stationary process, $G(d) = P(N(\mathbf{s}, d, \mathbf{S}) > 0)$. In Section 2.2 we considered the estimator

$$\bar{h}_{G,d}(\mathbf{S}) = \sum_{\mathbf{s}_i \in \mathbf{S} \cap D} \frac{1(N_D(\mathbf{s}_i, d, \mathbf{S}) > 0)}{N(\mathbf{S} \cap D)} \tag{3.8}$$

which has expected value $P(N_D(\mathbf{s}, d, \mathbf{S}) > 0)$. So, here, we are estimating $P(N_D(\mathbf{s}, d, \mathbf{S}) > 0)$, yet we want to estimate $P(N(\mathbf{s}, d, \mathbf{S}) > 0)$. From the inequality above, $G(d) = P(N(\mathbf{s}, d, \mathbf{S}) > 0) \geq P(N_D(\mathbf{s}, d, \mathbf{S}) > 0)$, again confirming the need for edge correction.

Suppose we adjust $\bar{h}_{G,d}(\mathcal{S})$ to

$$\bar{h}_{G,c_d}(\mathcal{S}) = \sum_{\mathbf{s}_i \in \mathcal{S} \cap D} \frac{1(N_D(\mathbf{s}_i, d, \mathcal{S}) > 0,\ c_d(\mathbf{s}_i) \subset D)}{N(\mathcal{S} \cap D)}. \tag{3.9}$$

Here, $c_d(\mathbf{s}_i)$ is a circle of radius d around $\mathbf{s}_i$. Then, $E_{\mathcal{S} \cap D} \bar{h}_{G,c_d}(\mathcal{S}) = P(N_D(\mathbf{s}, d, \mathcal{S}) > 0, c_d(\mathbf{s}) \subset D)$. The right side is the probability that, for a random $\mathcal{S}$ and $\mathbf{s} \in \mathcal{S}$, $c_d(\mathbf{s}) \subset D$ and at least one $\mathbf{s}' \in (\mathcal{S} \backslash \{\mathbf{s}\}) \cap D$ is in $c_d(\mathbf{s})$. Define $b_{G,c_d}(\boldsymbol{\theta}) = P(N_D(\mathbf{s}, d, \mathcal{S}) > 0,\ c_d(\mathbf{s}) \subset D|\boldsymbol{\theta})$ and let $\tilde{G}(d) = P(N_D(\mathbf{s}, d, \mathcal{S}) > 0 \,|\, c_d(\mathbf{s}) \subset D)$. With $P(c_d(\mathbf{s}) \subset D) \equiv b_{c_d}(\boldsymbol{\theta})$, $\tilde{G}(d) = \frac{b_{G,c_d}(\boldsymbol{\theta})}{b_{c_d}(\boldsymbol{\theta})}$.

The Bayes estimator for $\tilde{G}(d)$ is $E(\tilde{G}(d)|\mathcal{S}_{obs}) = E\left(\frac{b_{G,c_d}(\boldsymbol{\theta})}{b_{c_d}(\boldsymbol{\theta})}|\mathcal{S}_{obs}\right)$. With posterior samples, Monte Carlo integration becomes $\frac{1}{L}\sum_{l=1}^{L} \frac{b_{G,c_d}(\boldsymbol{\theta}_l^*)}{b_{c_d}(\boldsymbol{\theta}_l^*)}$. As in examples above, we can not compute $b_{G,c_d}(\boldsymbol{\theta}_l^*)$ or $b_{c_d}(\boldsymbol{\theta}_l^*)$ explicitly. So, again, we turn to the nested sampling idea. $\frac{1}{B}\sum_b \bar{h}_{G,c_d}(\mathcal{S}_{lb}^*)$ provides a posterior estimate for $b_{G,c_d}(\boldsymbol{\theta}_l^*)$ and $\frac{1}{B}\sum_b \bar{h}_{c_d}(\mathcal{S}_{lb}^*)$ provides an estimate for $b_{c_d}(\boldsymbol{\theta}_l^*)$. ($\bar{h}_{c_d}(\mathcal{S}) = \sum_{\mathbf{s}_i \in \mathcal{S} \cap D} 1(c_d(\mathbf{s}_i) \subset$

$D)/N(\mathcal{S}\cap D)$). Putting the pieces together, the edge corrected model based Bayes estimate for $\hat{G}(d)$ becomes

$$\frac{1}{L}\sum_{l}\frac{\frac{1}{B}\sum_b \bar{h}_{G,c_d}(\mathcal{S}^*_{lb})}{\frac{1}{B}\sum_b \bar{h}_{c_d}(\mathcal{S}^*_{lb})}. \tag{3.10}$$

Now, we return to the K function. Here, edge correction is a bit more complicated. Again, for a model with constant first order intensity, λ,

$$E_{\mathcal{S}\cap D}\left(\sum_{\mathbf{s}_i\in\mathcal{S}\cap D}\frac{N_D(\mathbf{s}_i,d,\mathcal{S}\backslash\mathbf{s}_i)}{N(\mathcal{S}\cap D)}\right)=E_{\mathcal{S}\cap D}N_D(\mathbf{s},d,\mathcal{S}\backslash\mathbf{s}).$$

As with $G(d)$, $K(d)\equiv EN(s,d,\mathcal{S}\backslash\mathbf{s})/\lambda$. However, what we can create, since we can only observe $N_D(\mathbf{s},d,\mathcal{S})$ is $K_D(d)\equiv E_{\mathcal{S}\cap D}N_D(s,d,\backslash\mathbf{s})/\lambda$. So the uncorrected estimator is based on $\bar{h}_{K,d}(\mathcal{S})=\sum_{\mathbf{s}_i\in\mathcal{S}\cap D}\frac{N_D(\mathbf{s}_i,d,\mathcal{S}\backslash\mathbf{s}_i)}{N(\mathcal{S}\cap D)\lambda}$ whose expectation is $K_D(d)$. Again, we see the need for edge correction. We are estimating $K_D(d)$ rather than $K(d)$. Similar to $G(d)$, since $N_D(\mathbf{s},d,\mathcal{S})\le N(\mathbf{s},d,\mathcal{S})$, $K_D(d)\le K(d)$.

Consider $\bar{h}_{K,c_d}(\mathcal{S})=$

$$\frac{\sum_{\mathbf{s}_i\in\mathcal{S}\cap D}N_D(\mathbf{s}_i,d,\mathcal{S})}{N(\mathcal{S}\cap D)}=\frac{\sum_{\mathbf{s}_i\in\mathcal{S}\cap D}\sum_{j\neq i}1(\mathbf{s}_j\in c_d(\mathbf{s})\cap D)}{N(\mathcal{S}\cap D)}. \tag{3.11}$$

Given $\mathbf{s}_i\in D$, $E\,1(\mathbf{s}_j\in c_d(\mathbf{s}_i)\cap D)=P(c_d(\mathbf{s}_i)\cap D)$. However, we want $P(c_d(\mathbf{s}_i))$. But $P(c_d(\mathbf{s}_i))=P(c_d(\mathbf{s}_i)\cap D)/P(D|c_d(\mathbf{s}_i))$. The denominator provides the appropriate inflation of the probability. Each of these probabilities is a function of $\boldsymbol{\theta}$.

In the literature (Section 2.2), $P(D|c_d(\mathbf{s}_i))$ is considered in terms of the empirical estimate for the K function, that is, for a given $\mathbf{s}_j$ in terms of $d_{ij}=||\mathbf{s}_i-\mathbf{s}_j||$. It is denoted by $w(\mathbf{s}_i,\mathbf{s}_j)$, and is approximated as the proportion of the circumference of $c_{d_{ij}}(\mathbf{s}_i)$ contained in D. This estimate is not model-based. For us, given $\mathbf{s}_i\in\mathcal{S}\cap D$, we can create a Monte Carlo integration, i.e., draw $\mathbf{s}$'s in $c_d(\mathbf{s}_i)$ and obtain the proportion which fall in D. So, the edge-adjusted $\bar{h}$ becomes

$$\bar{h}_{K,adj}(\mathcal{S})=\sum_{\mathbf{s}_i\in\mathcal{S}\cap D}\frac{\sum_{j\neq i}1(\mathbf{s}_j\in c_d(\mathbf{s}_i)\cap D)}{N(\mathcal{S}\cap D)P(D|c_d(\mathbf{s}_i))} \tag{3.12}$$

which has expectation $E_{\mathcal{S}\cap D|\boldsymbol{\theta}}\left(\frac{N_D(\mathbf{s},d,\mathcal{S})}{P(D\,|\,c_d(\mathbf{s}))}\right)\equiv b_K(\boldsymbol{\theta})$.

What is $b_K(\boldsymbol{\theta})$?

$$\begin{aligned}E_{\mathcal{S}\cap D}\left(\sum_{\mathbf{s}_i\in\mathcal{S}\cap D}\frac{\sum_{j\neq i}1(\mathbf{s}_j\in c_d(\mathbf{s}_i)\cap D)}{N(\mathcal{S}\cap D)P(D|c_d(\mathbf{s}_i))}\right)&=E_{\mathcal{S}\cap D}\left(\sum_{\mathbf{s}_i\in\mathcal{S}\cap D}\frac{N_D(\mathbf{s}_i,d,\mathcal{S})}{N(\mathcal{S}\cap D)P(D|c_d(\mathbf{s}_i))}\right)\\&=E_{\mathcal{S}\cap D}\left(\frac{N_D(\mathbf{s},d,\mathcal{S})}{P(D\,|\,c_d(\mathbf{s}))}\right).\end{aligned} \tag{3.13}$$

Now, given $\mathbf{s}$, $N_D(\mathbf{s},d,\mathcal{S})$ is the number of successes in $N(\mathbf{s},d,\mathcal{S})$ Bernoulli trials with success probability $P(D\,|\,c_d(\mathbf{s}))$. So, $E_{\mathcal{S}\cap D}\big(N_D(\mathbf{s},d,\mathcal{S})\,|\,N(\mathbf{s},d,\mathcal{S}),P(D\,|\,c_d(\mathbf{s}))\big)$ $=$ $N(\mathbf{s},d,\mathcal{S})\,P(D\,|\,c_d(\mathbf{s}))$ i.e., $E_{\mathcal{S}\cap D}\left(\frac{N_D(\mathbf{s},d,\mathcal{S})}{P(D\,|\,c_d(\mathbf{s}))}\,\middle|\,N(\mathbf{s},d,\mathcal{S}),P(D\,|\,c_d(\mathbf{s}))\right)$ $=$ $N(\mathbf{s},d,\mathcal{S})$. Hence, conditioning and unconditioning, $E_{\mathcal{S}\cap D}\left(\frac{N_D(\mathbf{s},d,\mathcal{S})}{P(D\,|\,c_d(\mathbf{s}))}\right)$ $=$ $E_{\mathcal{S}\cap D}N(\mathbf{s},d,\mathcal{S})$, which is exactly what we want.

Finally, we can create the edge corrected estimator for $K(d)$. We want $E\left(\frac{b_K(\boldsymbol{\theta})}{\lambda}|\mathcal{S}_{obs}\right)$ where $b_K(\boldsymbol{\theta}) \equiv E_{\mathcal{S}\cap D|\boldsymbol{\theta}}\left(\frac{N_D(\mathbf{s},d,\mathcal{S})}{P(D\mid c_d(\mathbf{s}))}\right)$. Again, with posterior samples $\boldsymbol{\theta}_l^*$, a Monte Carlo integration becomes $\frac{1}{L}\sum_l \frac{b_K(\boldsymbol{\theta}_l^*)}{\lambda_l^*}$. As above, $b_K(\boldsymbol{\theta}_l^*)$ can not be calculated explicitly. So, again, we need the nesting strategy, a Monte Carlo integration for each $\boldsymbol{\theta}_l^*$. With $\mathcal{S}_{bl}^*$, we calculate $\frac{1}{B}\sum_b \bar{h}_{K,adj}(\mathcal{S}_{bl}^*)$ as the approximation. Here, for a given $\mathcal{S}_{bl}^*$, we need an internal Monte Carlo integration for each of the desired conditional probabilities, i.e., for each $\mathbf{s}_m \in \mathcal{S}_{lb}^* \cap D$. Altogether, it is clear that this effort is computationally demanding.

For $K(d)$ in general, we need an extension. We remember that $K(d)$ is designed to be parameter-free under a constant first order intensity. Therefore, the foregoing works when λ is a parameter in the model with $E[N(\mathcal{S}\cap D)] = \lambda|D|$. However, for say, a stationary Gibbs process model, the Gibbs density is only known up to a normalizing constant that is not computable. We have a conceptual $\lambda(\boldsymbol{\theta})$ but it does not appear in the model, it is not calculable, and scaling by it, as above, is precluded.

However, $K(d)$ is still well defined and we would like a model based estimate for it, particularly to compare with $K(d)$ for the CSR specification. We offer a strategy to accomplish this. First, consider the special case where we can obtain λ explicitly as a function of $\boldsymbol{\theta}$. Then, in the foregoing, we simply replace λ_l^* with $\lambda(\boldsymbol{\theta}_l^*)$.

In general, with $K(d) = EN(\mathbf{s}, d, \mathcal{S}\backslash\mathbf{s})/\lambda$ what does λ mean? Clearly, $\lambda|D| = EN(\mathcal{S}\cap D)$ so $\lambda = EN(\mathcal{S}\cap D)/|D|$, i.e., $\lambda(\boldsymbol{\theta}) = E_{\mathcal{S}\cap D|\boldsymbol{\theta}}N(\mathcal{S}\cap D)/|D|$. So, posterior draws $\boldsymbol{\theta}_l^*$ yield posterior draws $\mathcal{S}_{lb}^*$ which yield posterior draws $N(\mathcal{S}_{lb}^*\cap D)$. As a result, given $\boldsymbol{\theta}_l^*$, we can learn about $E_{\mathcal{S}\cap D|\boldsymbol{\theta}_l^*}(N(\mathcal{S}\cap D)$, hence about $\lambda(\boldsymbol{\theta}_l^*)$. Explicitly, change $\frac{b_K(\boldsymbol{\theta})}{\lambda}$ to $|D|\frac{b_K(\boldsymbol{\theta})}{b_\lambda(\boldsymbol{\theta})}$ where $b_\lambda(\boldsymbol{\theta}) = E_{\mathcal{S}\cap D}N(\mathcal{S}\cap D)$. After the Monte Carlo approximation to the posterior expectation, we will need $b_\lambda(\boldsymbol{\theta}_l^*)$. We can use the $\mathcal{S}_{lb}^*$'s to obtain the approximation.

We conclude with a look at the inhomogeneous K function [15].. The inhomogeneous K function is associated with a nonstationary process so that we have $\lambda(\mathbf{s})$ rather than a constant λ. $K_{inhom}(d)$ is defined with regard to a bounded set, D, i.e.,

$$K_{inhom}(d) = \frac{1}{|D|}E_{\mathcal{S}\cap D|\boldsymbol{\theta}}\sum_{\mathbf{s}_i\in\mathcal{S}\cap D}\sum_{\mathbf{s}_j\in\mathcal{S}/\mathbf{s}_i}\frac{1(||\mathbf{s}_i-\mathbf{s}_j||\le d)}{\lambda(\mathbf{s}_i)\lambda(\mathbf{s}_j)}. \tag{3.14}$$

This form is ideally suited for us to create a Monte Carlo integration with replicates from $[\mathcal{S}|\mathcal{S}_{obs}]$. However, the second sum is over $\mathbf{s}_j \in \mathcal{S}/\mathbf{s}_i$ but we only observe $\mathbf{s}_j \in (\mathcal{S}/\mathbf{s}_i \cap D)$. That is, we have

$$K_{D,inhom}(d) = \frac{1}{|D|}E_{\mathcal{S}\cap D|\boldsymbol{\theta}}\sum_{\mathbf{s}_i\in\mathcal{S}\cap D}\sum_{\mathbf{s}_j\in(\mathcal{S}/\mathbf{s}_i\cap D)}\frac{1(||\mathbf{s}_i-\mathbf{s}_j||\le d)}{\lambda(\mathbf{s}_i)\lambda(\mathbf{s}_j)} \le K_{inhom}(d). \tag{3.15}$$

So, again we need edge correction. With $\lambda(\mathbf{s}_j)$ in the denominator, now we find that we divide the inner sum by $|D\cap c_d(\mathbf{s}_i)||c_d(\mathbf{s}_i)|$.

The edge corrected unbiased estimator of K_{inhom} offered by [15] is

$$\hat{K}_{inhom}(d) = \frac{1}{|D|}\sum_{\mathbf{s}_i\,\in\,\mathcal{S}\cap D}\sum_{\mathbf{s}_j\in\mathcal{S}\cap D\backslash\{\mathbf{s}_i\}}\frac{1(||\mathbf{s}_i-\mathbf{s}_j||\le d)}{w(\mathbf{s}_i,\mathbf{s}_j)\lambda(\mathbf{s}_i)\lambda(\mathbf{s}_j)} \tag{3.16}$$

where $w(\mathbf{s}_i,\mathbf{s}_j)$ is earlier edge-correction. Though $\hat{K}_{inhom}(d)$ is unbiased, it is not

an estimator since $\lambda(\mathbf{s})$ is unknown. In practice, we would replace $\lambda(\mathbf{s})$ with an empirical kernel intensity estimate, sacrificing unbiasedness.

Following our path, let

$$\begin{aligned}\tilde{h}_{K_{inhom}}(\mathcal{S}) &= \frac{1}{|D|N(\mathcal{S}\cap D)}\sum_{\mathbf{s}_i\in\mathcal{S}\cap D}\sum_{\mathbf{s}_j\in(\mathcal{S}\cap D)\setminus\mathbf{s}_i}\frac{1(||\mathbf{s}_i-\mathbf{s}_j||\le d)}{\lambda(\mathbf{s}_i)\lambda(\mathbf{s}_j)}\\ &= \frac{1}{|D|N(\mathcal{S}\cap D)}\sum_{\mathbf{s}_i\in\mathcal{S}\cap D}\frac{1}{\lambda(\mathbf{s}_i)}a_D(\mathbf{s}_i;\mathcal{S}\setminus\mathbf{s}_i)\end{aligned} \tag{3.17}$$

where $a_D(\mathbf{s}_i;\mathcal{S}\setminus\mathbf{s}_i)=\left[\sum_{\mathbf{s}_j\in\mathcal{S}}\frac{1(\mathbf{s}_j\in c_d(\mathbf{s}_i)\cap D)}{\lambda(\mathbf{s}_j)}\right]$. Then,

$$E_{\mathcal{S}\cap D|\boldsymbol{\theta}}\tilde{h}_{K_{inhom}}(\mathcal{S})=\frac{1}{|D|}E_{\mathcal{S}\cap D|\boldsymbol{\theta}}\frac{a_D(\mathbf{s};\mathcal{S}\setminus\mathbf{s})}{\lambda(\mathbf{s})}\equiv b_{K_{inhom}}(\boldsymbol{\theta}).$$

The right side cannot be collapsed into a form involving $N_D(\mathbf{s}_i,d,\mathcal{S})$ as in the homogeneous case. What we want to do is to modify $a_D(\mathbf{s}_i;\mathcal{S}\setminus\mathbf{s}_i)$ to $a(\mathbf{s}_i;\mathcal{S}\setminus\mathbf{s}_i)=\left[\sum_{\mathbf{s}_j\in\mathcal{S}}\frac{1(\mathbf{s}_j\in c_d(\mathbf{s}_i))}{\lambda(\mathbf{s}_j)}\right]$. We want to provide edge correction so that $b_{K_{inhom,corr}}(\boldsymbol{\theta})=\frac{1}{|D|}E_{\mathcal{S}\cap D}\frac{a(\mathbf{s};\mathcal{S}\setminus\mathbf{s})}{\lambda(\mathbf{s})}$ which is $K_{inhom}(d)$.

To make the correction, given $\mathbf{s}_i$,

$$\begin{aligned}&\frac{E\left(\dfrac{1(\mathbf{s}_j\in c_d(\mathbf{s}_i)\cap D)}{\lambda(\mathbf{s}_j)}\right)}{E\left(\dfrac{1(\mathbf{s}_j\in c_d(\mathbf{s}_i))}{\lambda(\mathbf{s}_j)}\right)}\\ &=\frac{\displaystyle\int_D\frac{1(\mathbf{s}\in c_d(\mathbf{s}_i)\cap D)}{\lambda(\mathbf{s})}\lambda(\mathbf{s})\,d\mathbf{s}}{\displaystyle\int_D\frac{1(\mathbf{s}\in c_d(\mathbf{s}_i))}{\lambda(\mathbf{s})}\lambda(\mathbf{s})\,d\mathbf{s}}\\ &=\frac{|c_d(\mathbf{s}_i)\cap D|}{|c_d(\mathbf{s}_i)|}\\ &=\frac{|c_d(\mathbf{s}_i)\cap D|}{\pi d^2}\equiv w_D(\mathbf{s}_i).\end{aligned}$$

We note that $w_D(\mathbf{s}_i)$ is not model dependent. It is a ratio of areas and can be obtained for any $\mathbf{s}_i$ by uniformly sampling over $c_d(\mathbf{s}_i)$ and recording the proportion of points falling in D. Hence, we revise $\tilde{h}_{Kinhom}(\mathcal{S})$ to $\tilde{h}_{Kinhom,corr}(\mathcal{S})=$

$$\frac{1}{|D|N(\mathcal{S}\cap D)}\sum_{\mathbf{s}_i\in\mathcal{S}\cap D}\frac{1}{\lambda(\mathbf{s}_i)w_D(\mathbf{s}_i)}a_D(\mathbf{s}_i;\mathcal{S}\setminus\mathbf{s}_i). \tag{3.18}$$

Again, the Bayes estimator for $K_{inhom,corr}(d)$ is the posterior mean, $E(b_{Kinhom,corr}(\boldsymbol{\theta})|\mathcal{S}_{obs})\approx\frac{1}{L}\sum_l b_{Kinhom}(\boldsymbol{\theta}_l^*)$. With posterior replications $\mathcal{S}_{lb}^*$, we obtain a Monte Carlo integration for each $b_{Kinhom,corr}(\boldsymbol{\theta}_l^*)$.