Posterior Contraction Rates for Gaussian Cox Processes with Non-identically Distributed Data

James A. Grant*[1,3] and David S. Leslie†[2,3]

[1]STOR-i Centre for Doctoral Training, Lancaster University, UK
[2]Department of Mathematics and Statistics, Lancaster University, UK
[3]PROWLER.io, Cambridge, UK

June 25, 2019

Abstract

This paper considers the posterior contraction of non-parametric Bayesian inference on non-homogeneous Poisson processes. We consider the quality of inference on a rate function λ, given non-identically distributed realisations, whose rates are transformations of λ. Such data arises frequently in practice due, for instance, to the challenges of making observations with limited resources or the effects of weather on detectability of events. We derive contraction rates for the posterior estimates arising from the Sigmoidal Gaussian Cox Process and Quadratic Gaussian Cox Process models. These are popular models where λ is modelled as a logistic and quadratic transformation of a Gaussian Process respectively. Our work extends beyond existing analyses in several regards. Firstly, we consider non-identically distributed data, previously unstudied in the Poisson process setting. Secondly, we consider the Quadratic Gaussian Cox Process model, of which there was previously little theoretical understanding. Thirdly, we provide rates on the shrinkage of both the width of balls around the true λ in which the posterior mass is concentrated and on the shrinkage of posterior mass outside these balls - usually only the former is explicitly given. Finally, our results hold for certain finite numbers of observations, rather than only asymptotically, and we relate particular choices of hyperparameter/prior to these results.

1 Introduction

The non-homogeneous Poisson process (NHPP) is the most widely used model for inference on point process data. It is parameterised by a non-negative rate function λ and satisfies the key property that the expected number of events in any area is equal to the integral of the rate function over that area. A Gaussian Cox process (GCP) is a nonparametric Bayesian version of the NHPP model where λ is modelled as a transformation of a Gaussian process (GP). In this paper we consider two classes of GCP, the Sigmoidal GCP (SGCP) of (Adams et al., 2009) and the Quadratic GCP (QGCP) of (Lloyd et al., 2015). In the SGCP the rate function is modelled as a multiple of a logistic transformation of a GP. In the QGCP the rate function is modelled as the square of a

*j.grant@lancaster.ac.uk; corresponding author
†d.leslie@lancaster.ac.uk

GP. This paper is concerned with the quality of posterior inference on λ arising from these models. Specifically we are interested in the rate at which the expected posterior mass the models assign to functions far from the true λ decreases.

The GCP is a model over functions and is defined on some space of non-negative functions Λ. Given a true rate function $\lambda_0 \in \Lambda$, observed data $X_{1:n}$ collected over $n \in \mathbb{N}$ timesteps, and a relevant distance $d_n(\lambda, \lambda')$ defined for all $\lambda, \lambda' \in \Lambda$, we look for results of the form

$$\mathbb{E}_{\lambda_0}\Big(\Pi\big(\lambda \in \Lambda : d_n(\lambda, \lambda_0) \geq \epsilon_n | X_{1:n}\big)\Big) \leq f_n \tag{1}$$

for decreasing sequences ϵ_n, f_n, where $\Pi(\cdot|X_{1:n})$ denotes the posterior probability mass and $\mathbb{E}_{\lambda_0}$ denotes expectation with respect to the probability measure implied by λ_0. The sequences ϵ_n, f_n define the rate of posterior contraction of a model. If such a bound holds for certain $\epsilon_n, f_n \to 0$ as $n \to \infty$ this displays that the model is consistent. However, we are also interested in the order of the sequences and for which n results of the form (1) can be identified.

Asymptotic consistency results of the form

$$\mathbb{E}_{\lambda_0}\Big(\Pi\big(\lambda \in \Lambda : d_n(\lambda, \lambda_0) \geq \epsilon_n | X_{1:n}\big)\Big) \to 0$$

as $n \to \infty$ are prevalent in the Bayesian nonparametrics literature; for example (Kirichenko and Van Zanten, 2015) gives such a result for i.i.d. $X_{1:n}$ under the SGCP and a broader family of GCPs which have a smooth and bounded link function. Such asymptotic results are undoubtedly useful contributions to the understanding Bayesian models and inference, however they provide limited support to finite-time analyses thereof. We extend beyond existing results in four important regards by

1. Providing results for independent non-identically distributed (i.n.i.d.) data,
2. Providing results for the QGCP model as well as the SGCP,
3. Providing a rate on the shrinkage of the posterior mass f_n (as well as on ϵ_n), and
4. Providing results for finite values of n, not only asymptotically, and relating specific choices of hyperpriors (and parameters) to these results.

Studying i.n.i.d. data is in contrast to the majority of previous studies of non-parametric inference on NHPPs. However, ours is an important, practically-relevant setting. Commonly when observing point process data, the detection of events may be imperfect. This may be due to visibility conditions, unreliable signals or the fallability of observation equipment. A result of this is that while events may occur independently and according to a stationary process, the distribution of *observed* events can vary as data is collected. Equally, different subsections of a region of interest may be observed at different rates by design. Data collectors may be more readily able to gather data in a particular region, resources may be too costly to gather the same quality of information everywhere or multiple sub-investigations may be combined to form a joint dataset. As the GCP models are typically used to model situations with underlying spatial smoothness and covariance structure, a unified analysis is still desirable, however existing contraction results only handle the setting where an entire region of interest has been observed uniformly. The results we obtain in this paper apply to the setting where (whether through design or imprecision) different rates of observation have been applied at different locations. Therefore, we present results that are more relevant to the practical settings in which GCP models are utilised than those which consider only identically distributed data.

The QGCP model has recently received attention in the literature (Lloyd et al., 2015; John and Hensman, 2018) as a model for NHPP inference, due to the ability to carry out fast and accurate inference. Previously however, there was little theoretical understanding of the model. We provide theoretical foundations for this new variant of the GCP model. This is non-trivial since the link function in the QGCP is not bounded, in contrast with the SGCP. Consequently, we find the rate of contraction to be lower for the QGCP than for the traditional SGCP.

Providing a rate f_n on the shrinkage of the posterior mass for finite values of n is also an important development. A trend in the existing literature is to focus on the asymptotic results and present that if the width of a ball around the true rate is chosen to decrease at the correct rate (with respect to the number of observations) then the probability of lying outside this ball tends to 0 as the number of observations goes to infinity. Such results are typically cleaner, and clearly demonstrate the consistency of a method, while the finite-time result can usually be extracted from the proofs provided for such results if desired. If the rate f_n is explicitly given or can be inferred, it is often only specified as holding for "sufficiently large" n. Inferring the rate f_n and determining the order of n that qualifies as sufficiently large, can be challenging to users of these results. By explicitly giving a form of f_n and quantifying the values of n (in terms of functions of the chosen hyperparameters) for which it is valid, we present a more informative set of results that are useful for end-users of this theory.

One use case of these results is in the theory of sequential decision making problems. In sequential sensor placement problems such as that studied in (Grant et al., 2019), decision makers adaptively select intervals over which to observe events subject to costs on the length of the interval, with an overall aim of maximising a cumulative reward. To do so optimally, balancing between exploring undersampled regions and "exploiting" - making repeated samples of areas where the rate is known to be high - is required. To understand the optimal balance of exploration and exploitation, one must understand the rate at which the inference model used contracts. Previous work on these problems has relied on assuming simpler inference models to obtain performance guarantees (Grant et al., 2018, 2019). Guarantees on the contraction of Cox process posteriors *with rates on the posterior mass* will be important in the design and analysis of more sophisticated approaches to these problems.

Another use for these results is in experimental design and resource planning problems. It is valuable for decision-makers to know the expected level of uncertainty in a rate function given a certain number of observations. They can then appropriately design sampling strategies or deploy resources to collect information in a way that is tailored to achieving a certain level of confidence in the inference.

In the remainder of this section, we discuss related work in GCPs and general contraction results for Bayesian models. In Section 2 we formally introduce our GCP models and notation. Section 3 includes all our main theoretical results and proofs, and in Section 4 we conclude with a discussion. Throughout we have aspired to make our assumptions transparent and demonstrate how they can be met. In Appendix H we verify that all assumed conditions can be satisfied for finite numbers of observations.

1.1 Related Literature

The Cox process (Cox, 1955) is a class of doubly stochastic process where the rate function of an non-homogeneous Poisson process (see e.g. (Møller and Waagepetersen, 2003)) is modelled as another stochastic process. The Gaussian Cox process (GCP), as mentioned above, is a particular subset of this class where the rate function of the NHPP is modelled via a transformation of a Gaussian process (Williams and Rasmussen, 2006). Three main transformations have been proposed yielding three main models. Firstly, the Log-Gaussian Cox Process (LGCP) of (Rathbun and Cressie, 1994) and (Møller et al., 1998) where λ is modelled as an exponential transformation of a GP. Secondly, the Sigmoidal-Gaussian Cox Process (SGCP) of (Adams et al., 2009) where λ is modelled as a multiple of a logistic transformation of a GP. Finally, the Quadratic-Gaussian Cox Process (QGCP) of (Lloyd et al., 2015) where λ is modelled as a quadratic transformation of a GP. We focus on the SGCP and QGCP models, as (for reasons discussed fully in Section 4) the LGCP model requires separate techniques to derive a contraction result.

General results for the contraction of posterior density estimates given i.i.d. data are available thanks to the seminal papers (Ghosal et al., 2000) and (Ghosal and Van Der Vaart, 2001). The link between density estimation and function estimation is exploited in (Belitser et al., 2015) to extend this work to show contraction rates for Bayesian Poisson process inference subject to appropriate prior conditions. Furthermore, (Belitser et al., 2015) proposes a spline based prior satisfying these conditions. The result of (Belitser et al., 2015) and GP concentration results of (van der Vaart and van Zanten, 2009) are used by (Kirichenko and Van Zanten, 2015) to show an asymptotic rate of posterior contraction for the SGCP - (Kirichenko and Van Zanten, 2015) is the existing work most similar to our contribution. However we are able to move beyond i.i.d. data to the independent non-identically distributed (i.n.i.d.) case, thanks to the work of (Ghosal and Van Der Vaart, 2007) in deriving contraction results for posterior density estimates under such data.

2 Model

In this section we introduce the data generating model and two prior models considered in the paper, along with other relevant notation required to understand our main results.

2.1 Likelihood

We consider an NHPP with bounded non-negative rate function λ_0 on $[0,1]^d$. We suppose that n independent realisations of the NHPP $\tilde{X}_1, ..., \tilde{X}_n$ are generated. Each realisation j consists of a collection m_j of points $\{\tilde{X}_j^1, ..., \tilde{X}_j^{m_j}\} \in [0,1]^d$. We write

$$\tilde{X}_j = \sum_{i=1}^{m_j} \delta_{\tilde{X}_j^i}, \quad j = 1, ..., n$$

where δ_x denotes the Dirac measure at x. By the definition of the NHPP model, each realisation j is distributed such that the number of points in any set $R \subseteq S$, denoted $\tilde{X}_j(R)$ follows a Poisson distribution with mean $\int_B \lambda(s)ds$. Furthermore $\tilde{X}_j(R_1), \tilde{X}_j(R_2)$ are independent if the sets $R_1, R_2 \subseteq S$ are disjoint.

Under our model, the realisations $\tilde{X}_1, ..., \tilde{X}_n$ are not directly observed. Instead, so-called *filtered realisations* $X_{1:n} = X_1, ..., X_n$ are observed. The events in a filtered realisation X_j are a subset of the events in the corresponding *raw realisation* $\tilde{X}_j$. The relationship between $X_{1:n}$ and $\tilde{X}_{1:n}$ is governed by a set of *filtering functions* $\gamma_{1:n} = \gamma_1, ..., \gamma_n$.

Each filtering function $\gamma_j : [0,1]^d \to [0,1]$ evaluated at a point $s \in S$ gives the probability of observing an event in $\tilde{X}_j$ given that it has occurred at location s. Every event that occurs in $\tilde{X}_j$ is observed or not independently according to these probabilities.

By standard results, X_j is distributed according to an NHPP with rate $\gamma_i \lambda_0$. That is to say, the n filtered realisations $X_{i:n}$ are then realisations of *independent, non-identically distributed* NHPPs with rates $\gamma_1\lambda_0, ..., \gamma_n\lambda_0$ respectively.

It follows that the likelihood of a particular set of observations $X_{1:n} = X_1, ..., X_n$ given a rate function λ and filtering functions $\gamma_{1:n}$ can be written

$$\mathcal{L}(X_{1:n}|\lambda_0, \gamma_{1:n}) = \prod_{j=1}^{n} \exp\left(\int_S \gamma_j(s)\lambda_0(s)dX_j(s) - \int_S (\gamma_j(s)\lambda_0(s) - 1)ds \right),$$

using the law of the realisation X_j as given by Proposition 6.1 of (Karr, 1986). We note that the case of i.i.d. data as considered in (Kirichenko and Van Zanten, 2015) and (Gugushvili et al., 2018) is a special case of this model, where $\gamma_j(x) = 1, \forall x \in [0,1]^d, \forall j = 1, \ldots, n$.

2.2 Prior Models

In this paper we consider two Bayesian models of the Poisson process where the rate function λ_0 is modelled a priori as a transformation of a Gaussian process. Under the SGCP model (Adams et al., 2009), the true rate function is modelled a priori as

$$\lambda(s) = \lambda^* \sigma(g(s)) = \lambda^*(1 + e^{-g(s)})^{-1} \quad s \in S \tag{2}$$

where $\lambda^* > 0$ is a scalar hyperparameter endowed with an independent Gamma prior and g is a zero-mean GP. The sigmoidal transformation σ is bounded in $[0,1]$ so the hyperparameter λ^* models the maximum of the rate function, $||\lambda_0||_\infty$. The QGCP model (Lloyd et al., 2015) uses a more straightforward transformation. The rate function is modelled a priori as

$$\lambda(s) = (g(s))^2 \quad s \in S \tag{3}$$

where again, g is a GP.

For both models, we specify certain additional properties of the GP to support our subsequent analyses. These conditions are standard in the posterior contraction literature (van der Vaart and van Zanten, 2008, 2009; Kirichenko and Van Zanten, 2015). We require that the covariance kernel f of the GP g, can be given in its spectral form by

$$Ef(s)f(s') = \int e^{-i<\xi, l(s'-s)>} \mu(\xi) d\xi, \quad s, s' \in S. \tag{4}$$

Here $l > 0$ is an (inverse) length scale parameter and μ is a spectral density on $\mathbb{R}^d$ such that the map $a \mapsto \mu(a\xi)$ on $(0, \infty)$ is decreasing for every $\xi \in \mathbb{R}^d$ and that satisfies

$$\int e^{\delta||\xi||} \mu(d\xi) < \infty$$

for some $\delta > 0$. Condition (4) is satisfied, for instance, by the squared exponential covariance function

$$Ef(s)f(s') = e^{-l^2||s-s'||^2}, \quad s, s' \in S$$

since it corresponds to a centred Gaussian spectral density.

The length scale parameter should have a prior π_l on $[0, \infty)$ which satisfies

$$C_1 x^{q_1} \exp(-D_1 x^d \log^{q_2} x) \leq \pi_l(x) \leq C_2 x^{q_1} \exp(-D_2 x^d \log^{q_2} x) \tag{5}$$

for positive constants C_1, C_2, D_1, D_2, non-negative constants q_1, q_2, and every sufficiently large $x > 0$. In particular if l^d is endowed with a Gamma(a, b) prior, then

$$\pi_l(x) = \frac{b^a d}{\Gamma(a)} x^{da-1} \exp(-bx^d)$$

for $x > 0$, and thus (5) is satisfied with $C_1 = C_2 = \frac{b^a d}{\Gamma(a)}$, $D_1 = D_2 = b$, $q_1 = da - 1$, and $q_2 = 0$. We will assume a Gamma prior on l^d in the remainder of the paper for ease of analysis and presentation, but note that similar results are obtainable for other choices.

Finally, for the SGCP model we assume a positive, continuous prior $p_{\lambda*}$ for λ^* on $[0, \infty)$ satisfying

$$\int_{\lambda'}^{\infty} p_{\lambda^*}(x) dx \leq C_0 e^{-c_0 (\lambda')^{\kappa}} \tag{6}$$

for some constants $c_0, C_0, \kappa > 0$ and all $\lambda' > 0$. This condition is satisfied by, for instance, choosing a Gamma prior on λ^*.

2.3 Additional Notation

In the following section, we will derive results on the posterior distribution of $\lambda_0 | X_{1:n}$ under the two models. We will denote the prior distributions as $\Pi(\cdot)$ and the posteriors as $\Pi(\cdot | X_{1:n})$. Certain results will be valid for the class of all continuous functions on $[0,1]^d$, which will be denoted $\mathcal{C}([0,1]^d)$, and others will hold for the class of all α-Hölder continuous functions on $[0,1]^d$ denoted $\mathcal{C}^\alpha[0,1]^d$.

Contraction results will inevitably depend on the particular filtering functions $\gamma_{1:n}$, therefore it is convenient to define versions of standard distances averaged with respect to $\gamma_{1:n}$. We have the averaged infity norm

$$\Gamma_{n,\infty}(\lambda, \lambda') = \frac{1}{n} \sum_{i=1}^{n} ||\lambda\gamma_i - \lambda'\gamma_i||_\infty = \frac{1}{n} \sum_{i=1}^{n} \sup_{x \in [0,1]^d} |\lambda(x)\gamma_i(x) - \lambda'(x)\gamma_i(x)|,$$

averaged L_2 norm

$$\Gamma_{n,2}(\lambda, \lambda') = \frac{1}{n} \sum_{i=1}^{n} ||\lambda\gamma_i - \lambda'\gamma_i||_2 = \frac{1}{n} \sum_{i=1}^{n} \int_{[0,1]^d} (\lambda(x)\gamma_i(x) - \lambda'(x)\gamma_i(x))^2 dx,$$

and square rooted averaged L_2 norm

$$\Gamma_{n,2}^{1/2}(\lambda, \lambda') = \frac{1}{n} \sum_{i=1}^{n} ||\sqrt{\lambda\gamma_i} - \sqrt{\lambda'\gamma_i}||_2 = \frac{1}{n} \sum_{i=1}^{n} \int_{[0,1]^d} (\sqrt{\lambda(x)\gamma_i(x)} - \sqrt{\lambda'(x)\gamma_i(x)})^2 dx,$$

for rate functions $\lambda, \lambda \in \mathcal{C}([0,1]^d)$. Using these definitions we can guarantee a rate of convergence appropriate to the level of filtering.

Finally let $N(\epsilon, \mathcal{S}, l)$ denote the ϵ-covering number of a set $\mathcal{S}$ with respect to distance l.

3 Posterior Contraction Results

In this section we state our results on the finite-time contraction of the posterior of the QGCP and SGCP models. Our results assert that given n realisations of the NHPP, the expected posterior mass concentrated on functions outside a Hellinger-like ball of a given width will not exceed a transformation of the width of the ball. Theorem 1 gives the result for the QGCP, and Theorem 2 for the SGCP.

Theorem 1 *Suppose that $\lambda_0 \in \mathcal{C}^\alpha([0,1]^d)$ for some $\alpha > 0$ and $\lambda_0 : [0,1]^d \to [\lambda_{0,min}, \infty)$. Suppose that the filtering functions $\gamma_{1:n}$ are known. Then for all sufficiently large $M, n > 0$ the posterior under the QGCP satisfies*

$$E_{\lambda_0}\left(\Pi\left(\lambda : \frac{1}{n}\sum_{i=1}^{n} ||\sqrt{\lambda\gamma_i} - \sqrt{\lambda_0\gamma_i}||_2 \geq \sqrt{2}M\epsilon_n | X_{1:n}\right)\right) = \tilde{o}\left(n^{\frac{-d}{4\alpha+d}}\right) \tag{7}$$

for $\epsilon_n = 2\sqrt{||\lambda_0|_\infty n^{-\alpha/(4\alpha+d)}(\log(n))^{\rho+d+1} + n^{-2\alpha/(4\alpha+d)}(\log(n))^{2\rho+2d+2}}$ *with* $\rho = \frac{1+d}{4+d/\alpha}$.

Theorem 2 *Suppose that $\lambda_0 \in \mathcal{C}^\alpha([0,1]^d)$ for some $\alpha > 0$ and $\lambda_0 : [0,1]^d \to [\lambda_{0,min}, \lambda_{0,max}]$. Suppose that the filtering functions $\gamma_{1:n}$ are known. Then for all sufficiently large $M, n > 0$ the posterior under the SGCP satisfies*

$$E_{\lambda_0}\left(\Pi\left(\lambda : \frac{1}{n}\sum_{i=1}^{n} ||\sqrt{\lambda\gamma_i} - \sqrt{\lambda_0\gamma_i}||_2 \geq \sqrt{2}M\epsilon_n | X_{1:n}\right)\right) = \tilde{o}\left(n^{\frac{-d}{2\alpha+d}}\right) \tag{8}$$

for $\epsilon_n = n^{-\alpha/(2\alpha+d)}(\log(n))^{\rho+d+1}$ *with* $\rho = \frac{1+d}{2+d/\alpha}$.

In each case analytical results free from "little-o" notation and a specific value for the "sufficiently large" conditions on M and n are given in the proofs in Sections 3.2 and 3.3.

The key difference between the two results is that for the QGCP we can only guarantee convergence on larger ball widths ϵ_n and at a slower rate f_n. Notice that under the QGCP the ball width is $\tilde{o}(n^{-\alpha/(4\alpha+d)})$ and the contraction rate is $\tilde{o}(n^{-d/(4\alpha+d)})$, whereas for the SGCP the ball width is $\tilde{o}(n^{-\alpha/(2\alpha+d)})$ and the contraction rate is $\tilde{o}(n^{-d/(2\alpha+d)})$.

In the simplest setting where $\lambda_0 \in \mathcal{C}^1([0,1])$ - i.e. where we consider Lipschitz smooth functions on $d = 1$ - this means we have a contraction rate of $\tilde{o}(n^{-1/5})$ on balls of width $\tilde{o}(n^{-1/5})$ for the QGCP and a contraction rate of $\tilde{o}(n^{-1/3})$ on balls of width $\tilde{o}(n^{-1/3})$ for the SGCP. The result on the SGCP is therefore tighter in two senses, we are able to say that the posterior mass shrinks quicker than for the QGCP and on the probability of being in a larger subspace (since the ball width ϵ_n is smaller, the area outside the ball is larger).

The different results arise as a consequence of the different transformation functions. For the posterior to contract at a given rate, we must demonstrate that the prior model satisfies certain properties related to this rate. Both models are built upon a GP g, and by considering the properties of g, we can verify that the SGCP and QGCP prior models meet the necessary conditions.

The results of (van der Vaart and van Zanten, 2009) demonstrate that for g as described in Section 2, relevant properties of g can be shown, i.e. that the prior mass g assigns to certain parts of the function space is bounded by sequences of a known form. It follows that appropriate transformations of these sequences can be used to show that the SGCP and QGCP priors also assign their prior mass across the function space in the required manner. The transformed sequences give rise to our ball widths ϵ_n which in turn influence the contraction rate. Since the SGCP and QGCP involve different transformations of g, we also require different transformations of the sequences for which desirable properties of g hold, and therefore different results are obtained.

More informally, the issue is that by applying a quadratic transformation to the GP over a logistic one, prior mass is dispersed more across the function space and the resulting posterior takes longer to contract around the true λ_0.

3.1 Contraction of NHPP models under general priors

Before we prove Theorems 1 and 2 we introduce a third result which gives a sufficient set of conditions on prior models to attain posterior contraction at a known rate under i.n.i.d. observations. Theorem 3 extends Theorem 1 of (Ghosal and Van Der Vaart, 2007) to apply to for Poisson processes. The extension is in the same manner as the result of (Belitser et al., 2015) extends Theorem 2 of (Ghosal and Van Der Vaart, 2001) for i.i.d. Poisson process realisations. In addition we retain the rate on the shrinkage of the posterior mass, as well as on the ball width, unlike these earlier papers.

Theorem 3 *Assume that $\lambda_0 : [0,1]^d \to [\lambda_{0,min}, \infty)$ and that filtering functions $\gamma_{1:n}$ are known. Suppose that for positive sequences δ, $\bar{\delta}_n \to 0$, such that $n \min(\delta_n, \bar{\delta}_n)^2 \to \infty$ as $n \to \infty$, it holds that there exist subsets $\Lambda_n \subset \mathcal{C}(S)$, some $n_0 \in \mathbb{N}$, and constants $c_1, c_2, c_3 > 0$, $c_4 > 1$, and $c_5 > c_2 + 2$ such that*

$$\Pi_n\Big(\lambda : \Gamma_{n,\infty}(\lambda, \lambda_0) \leq \delta_n\Big) \geq c_1 e^{-c_2 n \delta_n^2} \tag{9}$$

$$\sup_{\delta > \bar{\delta}_n} \log N\left(\frac{\delta}{36\sqrt{2}}, \sqrt{\Lambda_{n,\delta}}, \Gamma_{n,2}\right) \leq c_3 n \bar{\delta}_n^2 \tag{10}$$

$$\Pi_n(\Lambda \setminus \Lambda_n) \leq c_4 e^{-c_5 n \delta_n^2}. \tag{11}$$

for all $n \geq n_0$ where $\Lambda_{n,\epsilon} = \Big\{\lambda \in \Lambda_n : h_n(p_\lambda, p_{\lambda_0}) \leq \epsilon\Big\}$, and $h_n(p_\lambda, p_{\lambda_0})$, is given by

$$h_n^2(p_\lambda, p_{\lambda'}) = \frac{1}{n}\sum_{i=1}^{n} 2\left(1 - E_{\lambda\gamma_i}\left(\sqrt{\frac{p(X^{(i)}|\lambda, \gamma_i)}{p(X^{(i)}|\lambda', \gamma_i)}}\right)\right).$$

Then for $\epsilon_n = \max(\delta_n, \bar{\delta}_n)$ and any $C > 0, J \geq 1, M \geq 2$,

$$E_{\lambda_0}\left[\Pi_n\left(\lambda : \Gamma_{n,2}^{1/2}(\lambda, \lambda_0) \geq \sqrt{2} J M \epsilon_n | X_{1:n}\right)\right] \leq \frac{1}{C^2 n \epsilon_n^2} + e^{-M^2 n \epsilon_n^2/4} + 2e^{-(M^2/2 - c_3) n \epsilon_n^2} + \frac{2}{c_1} e^{-(c_2 M^2 J^2/4 - C - 1) n \epsilon_n^2} \tag{12}$$

for $n \geq \max(n_0, n_1, n_2, n_3)$ where $n_1 = \arg\min\{n : \epsilon_n \leq \lambda_{min}\}$, $n_2 = \arg\min\{n : \epsilon_n \leq \frac{1}{\sqrt{2}M}\}$, and $n_3 = \arg\min\{n : e^{-n\epsilon_n^2 K M^2/4} \leq 1/2\}$.

We prove this theorem in Section 3.4. This establishes that given the prior model satisfies certain conditions, the expected posterior mass assigned to rate functions outside an order ϵ_n width ball around λ_0 (measured with respect to an averaged L_2 distance) decreases at rate $o((n\epsilon_n^2)^{-1})$

for sufficiently large n. The conditions on the prior model are standard and are inherited from the conditions of Theorem 4 of (Ghosal and Van Der Vaart, 2007) required to show posterior contraction in a density estimation setting. Condition (9), the *prior mass condition*, ensures that a sufficient proportion of the prior mass is assigned to functions close to λ_0. Condition (10), the *entropy condition*, and condition (11), the *remaining mass condition*, together prescribe that there exist subsets of the function space such that the entropy of these subsets is not too large, but the probability of lying outside these is also small.

Equipped with this general result, we are now in a position to prove Theorems 1 and 2 by demonstrating that the QGCP and SGCP models meet conditions (9), (10), and (11).

3.2 Proof of Theorem 1: Contraction of the QGCP model

To prove Theorem 1 we verify that the QGCP model described in Section 2 meets the conditions of Theorem 3. The following sections handles each condition in turn. Throughout we have

$$\delta_n = 2\sqrt{||\lambda_0||_\infty} n^{-\alpha/(4\alpha+d)} \log^{\rho}(n) + n^{-2\alpha/(4\alpha+d)} \log^{2\rho}(n), \tag{13}$$

$$\bar{\delta}_n = 2\sqrt{||\lambda_0||_\infty} n^{-\alpha/(4\alpha+d)} \log^{\rho+d+1}(n) + n^{-2\alpha/(4\alpha+d)} \log^{2\rho+2d+2}(n). \tag{14}$$

3.2.1 Prior Mass Condition

The first condition, the so-called prior mass condition (9) does not rely on the existence of particular subsets Λ_n, and can be verified by the following lemma, which we prove in Appendix A.

Lemma 4 *If $\lambda_0 = g_0^2$ where $g_0 \in \mathcal{C}^\alpha([0,1]^d)$ for some $\alpha > 0$ then under the QGCP model there exist constants $c_1, c_2 > 0$ for δ_n as defined in (13) such that the prior satisfies*

$$\Pi(\lambda : ||\lambda - \lambda_0||_\infty \le \delta_n) \ge c_1 e^{-c_2 n \delta_n^2}$$

for all $n \ge 3$.

Then consider that since $\gamma_i \in [0,1]$ for all $i = 1, ..., n$,

$$\Gamma_{n,\infty}(\lambda, \lambda_0) = \frac{1}{n} \sum_{i=1}^{n} ||\lambda\gamma_i - \lambda_0\gamma_i||_\infty \le ||\lambda - \lambda_0||_\infty. \tag{15}$$

Thus, by Lemma 4 we have that there exist constants $c_1, c_2 > 0$ such that

$$\Pi_n\Big(\lambda : \Gamma_{n,\infty}(\lambda, \lambda_0) \le \delta_n\Big) \ge \Pi_n(\lambda : ||\lambda - \lambda_0||_\infty \le \delta_n) \ge c_1 e^{-c_2 n \delta_n^2},$$

satisfying condition (9).

3.2.2 Definition of Sieves

We now define the subsets Λ_n for which the QGCP satisfies the constraints of Theorem 3. Let,

$$\Lambda_n = (\mathcal{G}_n)^2 \tag{16}$$

where

$$\mathcal{G}_n = \left[\beta_n \sqrt{\frac{\zeta_n}{\chi_n}} \mathbb{H}_1^{\zeta_n} + \kappa_n \mathbb{B}_1\right] \cup \left[\bigcup_{a \leq \chi_n} (\beta_n \mathbb{H}_1^a) + \kappa_n \mathbb{B}_1\right], \tag{17}$$

$\mathbb{B}_1$ is the unit ball in $\mathcal{C}([0,1]^d)$ with respect to the uniform norm, and $\mathbb{H}_1^l$ is the unit ball of the RKHS $\mathbb{H}^l$ of the GP g with covariance as given in (4). We define the sequences involved as follows,

$$\begin{aligned}
\zeta_n &= L_2 n^{\frac{1}{d}\frac{2\alpha+d}{4\alpha+d}} (\log(n))^{2\rho/d} + L_3 n^{\frac{1}{d}\frac{\alpha+d}{4\alpha+d}} (\log(n))^{3\rho/d} + L_4 n^{\frac{1}{d}\frac{d}{4\alpha+d}} (\log(n))^{4\rho/d} \\
\beta_n &= L_5 n^{\frac{1}{2}\frac{2\alpha+d}{4\alpha+d}} (\log(n))^{2\rho+\frac{d+1}{2}} + L_6 n^{\frac{1}{2}\frac{\alpha+d}{4\alpha+d}} (\log(n))^{3\rho+\frac{d+1}{2}} + L_7 n^{\frac{1}{2}\frac{d}{4\alpha+d}} (\log(n))^{4\rho+\frac{d+1}{2}} \\
\kappa_n &= \frac{1}{3}\bar{\delta}_n, \qquad \chi_n = \frac{\bar{\delta}_n}{6\tau\sqrt{d}\beta_n},
\end{aligned}$$

for constants

$$L_2 > (8c_5||\lambda_0||_\infty)/D_1, \qquad L_3 > (8c_5\sqrt{||\lambda_0||_\infty})/D_1, \qquad L_4 > 2c_5/D_1$$

such that $L_2 + L_3 + L_4 > \max(A, e)$ and

$$\begin{aligned}
L_5 &\geq \max\left(\sqrt{\frac{16K_5 L_2^d \mathcal{K}_1^{1+d}}{\log^{2\rho}(3)}}, \sqrt{32||\lambda_0||_\infty c_5}, L_2^{1/3}\Big(\frac{8\max(1, \sqrt{||\lambda_0||_\infty})}{(3/36\sqrt{2})^{3/2} d^{1/4}\sqrt{2\tau}}\Big)^{2/3}\right) \\
L_6 &\geq \max\left(\sqrt{\frac{16K_5 L_3^d \mathcal{K}_1^{1+d}}{\log^{3\rho}(3)}}, \sqrt{32\sqrt{||\lambda_0||_\infty} c_5}\right), \quad L_7 \geq \max\left(\sqrt{\frac{16K_5 L_4^d \mathcal{K}_1^{1+d}}{\log^{4\rho}(3)}}, \sqrt{8c_5}\right),
\end{aligned}$$

such that $L_5 + L_6 + L_7 > \frac{4L_1 \max(1, \sqrt{||\lambda_0||_\infty})}{3\sqrt{||\mu||}}$, and $L_2 L_5^3 > \left(\frac{8\max(1, ||g_0||_\infty)}{(3/L_1)^{3/2} d^{1/4}\sqrt{2\tau}}\right)^2$ where $L_1 = 1/(36\sqrt{2})$ and $\mathcal{K}_1 = \log\big(\frac{3(L_2+L_3+L_4)}{\min(1, 2\sqrt{||\lambda_0|_\infty})}\big) + \frac{2\alpha d + 2\alpha + d}{4\alpha d + d^2} + (4\rho - \rho/d - d - 1)$.

The definition of $\mathcal{G}_n$ and these sequences is important as it allows general GP results of (van der Vaart and van Zanten, 2009) to be applied. The extensive conditions on the constants are important to ensure that the results hold for finite values of n.

3.2.3 Entropy Condition

The following lemma allows us to verify condition (10) which stipulates that the log entropy of the subsets Λ_n is not too large. The proof of this lemma is provided in Appendix C. In particular it exploits an existing bound on the covering number of $\mathcal{G}_n$ with respect to the infinity norm from (van der Vaart and van Zanten, 2009).

Lemma 5 *For Λ_n defined as in (16), a constant $L_1 > 0$, and $\bar{\delta}_n$ as defined in (14), there exists a constant $c_3 > 0$ such that*

$$\log N(L_1\bar{\delta}_n, \sqrt{\Lambda_n}, ||\cdot||_2) \leq c_3 n \bar{\delta}_n^2,$$

for all n such that

$$4||\lambda_0||_\infty \log^{2d+2-2\rho}(n) \geq \frac{m\sum_{i=2}^4 L_i^d}{2^{1+d}}\left(\log\Big(\frac{27\tau\sqrt{d}(\sum_{i=5}^7 L_i)^3 \sum_{i=2}^4 L_i}{4||\lambda_0||_\infty^{3/2}}\Big) + \Big(4 + \frac{12 + d + d^2}{8\alpha d + 2d^2}\Big)\log(n)\right)^{1+d}$$

and

$$2\log\left(\frac{6\sqrt{||\mu||}(L_5+L_6+L_7)}{2L_1\sqrt{||\lambda_0||_\infty}+L_1}\right) \leq 4||\lambda_0||_\infty n^{(4\alpha+2d)/(8\alpha+2d)}\log^{2\rho+2d+2}(n) - \log\left(n^{(6\alpha+d)/(4\alpha+d)}\log^{6\rho}(n)\right).$$

To apply Lemma 5, notice that $\frac{1}{n}\sum_{i=1}^{n}||\gamma_i \times \cdot||_2 \leq ||\cdot||_2$ since the functions $\gamma_i \in [0,1]$ for all $i = 1, ..., n$. It follows that

$$N\left(\frac{\delta}{36\sqrt{2}}, \sqrt{\Lambda_{n,\delta}}, \frac{1}{n}\sum_{i=1}^{n}||\gamma_i \times \cdot||_2\right) \leq N\left(\frac{\delta}{36\sqrt{2}}, \sqrt{\Lambda_{n,\delta}}, ||\cdot||_2\right) \leq N\left(\frac{\delta}{36\sqrt{2}}, \sqrt{\Lambda_n}, ||\cdot||_2\right). \quad (18)$$

As any ϵ-covering number is decreasing in ϵ, it follows by Lemma 5 that

$$\sup_{\delta > \bar{\delta}_n} \log N\left(\frac{\delta}{36\sqrt{2}}, \sqrt{\Lambda_{n,\delta}}, \frac{1}{n}\sum_{i=1}^{n}||\gamma_i \times \cdot||_2\right) \leq c_3 n \bar{\delta}_n^2.$$

Thus we have satisfied constraint (10).

3.2.4 Remaining Mass Condition

Finally, Lemma 6 below is sufficient to validate condition (11) directly. Its proof is given in Appendix D.

Lemma 6 *Under the QGCP model, with Λ_n as defined in* (16)*, and δ_n as defined in* (13) *there exist constants $c_4 > 0, c_5 \geq c_2 + 4$ such that*

$$\Pi(\lambda : \lambda \notin \Lambda_n) \leq c_4 e^{-c_5 n \delta_n^2},$$

for all n such that

$$n^{(2\alpha+d)/(4\alpha+d)}\log^{2\rho}(n) \geq \frac{q_1}{4c_5||g_0||_\infty^2}\log\left((L_2+L_3+L_4)n^{(2\alpha+d)/(4\alpha d+d^2)}\log^{4\rho/d}(n)\right).$$

3.2.5 Conlcuding the Proof

By Lemmas 4, 5, and 6 and the definitions of δ_n and $\bar{\delta}_n$ therein we have that the conditions of Theorem 3 are satisfied. Thus, for the QGCP model

$$E_{\lambda_0}\left[\Pi_n\left(\lambda : \Gamma_{n,2}^{1/2}(\lambda,\lambda_0) \geq \sqrt{2}JM\epsilon_n|\tilde{X}_{1:n}\right)\right] \leq \frac{1}{C^2 n\epsilon_n^2} + e^{-M^2 n\epsilon_n^2/4} + 2e^{-(M^2/2-c_3)n\epsilon_n^2} + \frac{2}{c_1}e^{-(c_2M^2J^2/4-C-1)n\epsilon_n^2}$$

holds with $\epsilon_n = \max(\delta_n, \bar{\delta}_n) = \bar{\delta}_n$ for any $C > 0, J \geq 1, M \geq 2$. Specific values of the remaining constants can be extracted from Lemmas 4, 5, 6, and 11.

Then, so long as M and J are sufficiently large, the second, third and fourth terms on the RHS of equation (12) decay much more quickly than the first and the bound is $\tilde{o}(n^{\frac{-d}{4\alpha+d}})$ as stated, for all n such that the conditions of Theorem 3 and Lemmas 5 and 6 are met. □

3.3 Proof of Theorem 2: Contraction of the SGCP model

Like the proof of Theorem 1, the proof of Theorem 2 relies on demonstrating the the SGCP model described in Section 2 meets the conditions of Theorem 3. In (Kirichenko and Van Zanten, 2015) the conditions of Theorem 1 of (Belitser et al., 2015) - the asymptotic and i.i.d. analogue of Theorem 3 - are verified for the SGCP model. However certain asymptotic arguments are used in said proof. In the following sections we handle each condition of Theorem 3 in turn under our setting. Throughout we have

$$\delta_n = n^{-\alpha/(2\alpha+d)}(\log(n))^{(1+d)/(2+d/\alpha)} \tag{19}$$

$$\bar{\delta}_n = n^{-\alpha/(2\alpha+d)}(\log(n))^{(1+d)/(2+d/\alpha)+d+1} \tag{20}$$

3.3.1 Prior Mass Condition

For the SGCP model, the prior mass condition (9) can be verified by the following lemma which we prove in Appendix E.

Lemma 7 *If $\lambda_0 = ||\lambda_0||_\infty \sigma(g_0)$ where $g_0 \in \mathcal{C}^\alpha([0,1]^d)$ for some $\alpha > 0$ then under the SGCP model there exist constants c_1, c_2 for δ_n as defined in (19) such that the prior satisfies*

$$\Pi(\lambda : ||\lambda - \lambda_0||_\infty \le \delta_n) \ge c_1 e^{-c_2 n \delta_n^2}$$

for all $n \ge 3$.

Then, by (15) and Lemma 7 we have that there exist constants $c_1, c_2 > 0$ such that

$$\Pi\bigg(\lambda : \Gamma_{n,\infty}(\lambda, \lambda_0) \le \delta_n\bigg) \ge c_1 e^{-c_2 n \delta_n^2}$$

and we have shown condition (9) is satisfied under the SGCP model.

3.3.2 Definition of Sieves

We now define the sets Λ_n such that the remaining coniditions hold. Consider,

$$\Lambda_n = \bigcup_{\lambda \le \lambda_n} \lambda \sigma(\mathcal{G}_n) \tag{21}$$

where

$$\mathcal{G}_n = \left[\beta_n \sqrt{\frac{\zeta_n}{\chi_n}} \mathbb{H}_1^{\zeta_n} + \kappa_n \mathbb{B}_1\right] \cup \left[\bigcup_{a \le \chi_n} (\beta_n \mathbb{H}_1^a) + \kappa_n \mathbb{B}_1\right], \tag{22}$$

$\mathbb{B}_1$ is the unit ball in $\mathcal{C}([0,1]^d)$ with respect to the uniform norm, and $\mathbb{H}_1^l$ is the unit ball of the RKHS $\mathbb{H}^l$ of the GP g with covariance as given in (4). Though the structure of the sieves $\mathcal{G}_n$ is the same as in the proof of Theorem 1, the sequences are defined differently, as below

$$\zeta_n = L_8 n^{\frac{1}{2\alpha+d}} (\log(n))^{2\rho/d}, \quad \beta_n = L_9 n^{\frac{d}{2(2\alpha+d)}} (\log(n))^{d+1+2\rho},$$
$$\lambda_n = L_{10} n^{\frac{d}{\kappa(2\alpha+d)}} (\log(n))^{4\rho/\kappa}, \quad \kappa_n = \frac{1}{3}\bar{\delta}_n, \quad \chi_n = \frac{\kappa_n}{2\tau\sqrt{d}\beta_n},$$

for constants

$$L_8 > \max\left(A, 1, \left(\frac{2c_5}{D_1}\right)^{1/d}\right), \quad L_9 \geq \sqrt{8c_5}, \quad L_{10} > \left(\frac{c_5}{c_0}\right)^{1/\rho}$$

such that

$$L_8 L_9^3 L_{10}^{3/2} > \frac{2}{(6cL_1)^{3/2}\tau\sqrt{d}}, \quad L_9 L_{10}^{1/2} > \frac{1}{6cL_1\sqrt{||\mu||}},$$

where $L_1 = 1/(36\sqrt{2})$, $c = 2^{-5/2}$, and κ is a positive constant.

3.3.3 Entropy Condition

The following lemma will allow us to verify condition (10). We prove it in Appendix F.

Lemma 8 *For Λ_n as defined in (21), a constant $L_1 > 0$ and $\bar{\delta}_n$ as defined in (20), there exists a constant $c_3 > 0$ such that*

$$\log N(L_1\bar{\delta}_n, \sqrt{\Lambda_n}, ||\cdot||_2) \leq c_3 n\bar{\delta}_n^2,$$

for all n such that

$$\log^{2d+2}(n) > K_1 L_8^d \Bigg(\log(\sqrt{2\tau L_8 L_9^3} L_{10}^{3/4} d^{1/4}) + \frac{\kappa(6d+6\alpha+2)+3d}{4\kappa(2\alpha+d)} \log(n) + \log\left(\log^{3\rho/2+3\rho/\kappa+\rho/d-d-1}(n)\right)\Bigg)^{1+d} \tag{23}$$

$$n^{\frac{d}{2\alpha+d}} > \max\left(2\log(12cL_1L_9L_{10}^{1/2}), 2\log(L_1L_{10}^{1/2})\right) + 1. \tag{24}$$

Then, as in the proof of Theorem 1, using (18) and Lemma 8 we have

$$\sup_{\delta > \bar{\delta}_n} \log N\left(\frac{\delta}{36\sqrt{2}}, \sqrt{\Lambda_{n,\delta}}, \frac{1}{n}\sum_{i=1}^n ||\gamma_i \times \cdot||_2\right) \leq c_3 n\bar{\delta}_n^2,$$

verifying condition (10).

3.3.4 Remaining Mass Condition

Finally, Lemma 9 below is sufficient to validate condition (11) directly. Its proof is given in Appendix G.

Lemma 9 *Under the SGCP model, with Λ_n as defined in (21), and δ_n as defined in (20) there exist constants $c_4 > 0$, $c_5 \geq c_2 + 4$ such that*

$$\Pi(\lambda : \lambda \notin \Lambda_n) \leq c_4 e^{-c_5 n\delta_n^2}$$

for all n such that

$$\log^{2\rho}(n) > \frac{16K_5D_1L_8^d}{L_9^2}\left(\frac{\log(\sqrt{L_{10}}L_8) + \log\left(n^{\frac{2\alpha\kappa+2\kappa+d}{2\kappa(2\alpha+d)}}\log^{\rho(2/\kappa+2/d-1)-d-1}(n)\right)}{\log(n)}\right)^{1+d}, \tag{25}$$

$$n^{\frac{d}{2\alpha+d}} > \frac{1}{c_5}\left(\log(L_8^{q_1-d+1}) + 1\right), \tag{26}$$

3.3.5 Concluding the Proof

Thus, the three conditions (9), (10), and (11) of Theorem 3 are satisfied by the SGCP model, and we have

$$E_{\lambda_0}\left[\Pi_n\left(\lambda : \Gamma_{n,2}^{1/2}(\lambda,\lambda_0) \geq \sqrt{2}JM\epsilon_n|\tilde{X}_{1:n}\right)\right] \leq \frac{1}{C^k(n\epsilon_n^2)^{k/2}} + e^{-M^2 n\epsilon_n^2/4} + 2e^{-(M^2/2-c_3')n\epsilon_n^2} + \frac{2}{c_1'}e^{-(c_2' M^2 J^2/4-C-1)n\epsilon_n^2}$$

with $\epsilon_n = \max(\delta_n, \bar{\delta}_n) = \bar{\delta}_n$. Then, so long as M and J are sufficiently large, the second, third and fourth terms on the RHS of equation (12) decay much more quickly than the first and the bound is $\tilde{o}(n^{\frac{-d}{2\alpha+d}})$ as stated, for all n such that the conditions of Theorem 3 are met and that (23), (24), (25), and (26) hold. □

3.4 Proof of Theorem 3: Generic contraction in NHPPs

The proof of Theorem 3 depends on a general result for convergence of posterior parameter estimation given i.n.i.d. observations. Such a result is given in (Ghosal and Van Der Vaart, 2007), but without a finite time rate on the probability. We restate their result below as Theorem 10 but with a rate included.

Consider as in Ghosal and Van Der Vaart (2007), a model in which a parameter $\theta_0 \in \Theta$ gives rise to a i.n.i.d sequence of data. The data at time i are drawn independently from the data at other times from a distribution P_i^θ, which we assume admits a density p_i^θ with respect to a dominating measure.

We define the following subsets of the parameter space for $n \geq 1$ and $k > 1$

$$B_n(\theta_0, \epsilon; k) = \left\{\theta \in \Theta : \frac{1}{n}\sum_{i=1}^n K_i(\theta_0,\theta) \leq \epsilon^2,\ \frac{1}{n}\sum_{i=1}^n V_{k,0;i}(\theta_0,\theta) \leq C_k\epsilon^k\right\}$$

where $K_i(\theta_0,\theta) = \int p_i^{\theta_0}\log(p_i^{\theta_0}/p_i^\theta)d\mu$ is the Kullback-Leibler divergence and $V_{k,0;i}(\theta_0,\theta) = \int p_i^{\theta_0}|\log(p_i^{\theta_0}/p_i^\theta) - K(\theta_0,\theta)|^k d\mu$ is a variance discrepancy measure. Furthermore, let d_n be the averaged Hellinger distance, defined by

$$d_n^2(\theta,\theta') = \frac{1}{n}\sum_{i=1}^n \int (\sqrt{p_{\theta,i}} - \sqrt{p_{\theta',i}})^2 d\mu_i.$$

Our modified version of Theorem 4 of (Ghosal and Van Der Vaart, 2007) is as below.

Theorem 10 *Suppose $Y_i \sim P_i^\theta$ independently for $i = 1, ..., n$ and let d_n be defined as the average Hellinger distance. Further, suppose that for a sequence $\epsilon_n \to 0$ such that $n\epsilon_n^2$ is bounded away from 0, some $k > 1$, all sufficiently large $j \in \mathbb{N}$, constants $c_1, c_2, c_3 > 0$, and sets $\Theta_n \subset \Theta$, the following conditions hold:*

$$\frac{\Pi_n(\theta \in \Theta_n : j\epsilon_n < d_n(\theta,\theta_0) \le 2j\epsilon_n)}{\Pi_n(B_n^*(\theta_0,\epsilon_n;k))} \le c_1 e^{\frac{c_2 n\epsilon_n^2 j^2}{4}} \tag{27}$$

$$\frac{\Pi_n(\Theta \setminus \Theta_n)}{\Pi_n(B_n^*(\theta_0,\epsilon_n;k))} = o(e^{-2n\epsilon_n^2}) \tag{28}$$

$$\sup_{\epsilon>\epsilon_n} \log N\left(\frac{\epsilon}{36}, \{\theta \in \Theta_n : d_n(\theta,\theta_0) < \epsilon\}, d_n\right) \le c_3 n\epsilon_n^2. \tag{29}$$

Then for any $C > 0, J \ge 1,$ and $M \ge 2$,

$$\mathbb{E}_{\theta_0}\Pi_n(\theta : d_n(\theta,\theta_0) \ge JM\epsilon_n | Y^{(n)}) \le \frac{1}{C^k(n\epsilon_n^2)^{k/2}} + e^{-M^2 n\epsilon_n^2/4} + 2e^{-(M^2/2-c_3)n\epsilon_n^2} + \frac{2}{c_1}e^{-(c_2M^2J^2/4-C-1)n\epsilon_n^2}$$

for all n such that $e^{-n\epsilon_n^2 M^2/4} \le 1/2$.

The proof of Theorem 10 is a modification of proof of Theorem 4 of (Ghosal and Van Der Vaart, 2007). We replace arguments that hold in the limit with finite-time versions and handle the introduction of the constants c_1, c_2, c_3, assumed to be 1 in (Ghosal and Van Der Vaart, 2007). We can set the constant K present in the original theorem to $1/2$ since we are dealing with the Hellinger distance.

Proof of Theorem 10: By Lemmas 9 and 10 of (Ghosal and Van Der Vaart, 2007) and given conditions (27), (28), and (29), we have for n such that $e^{-n\epsilon_n^2 M^2/4} \le 1/2$ any $M \ge 2$, $J \ge 1$ and $C > 0$,

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\Pi_n(\theta : d_n(\theta,\theta_0) \ge JM\epsilon_n | Y^{(n)}) \\
&\le \frac{1}{C^k(n\epsilon_n^2)^{k/2}} + e^{-M^2n\epsilon_n^2/4} + \frac{e^{-(M^2/2-c_3)n\epsilon_n^2}}{1-e^{-M^2n\epsilon_n^2/2}} + \sum_{j\ge J}\frac{1}{c_1}e^{-n\epsilon_n^2(c_2M^2j^2/4-C-1)} \\
&\le \frac{1}{C^k(n\epsilon_n^2)^{k/2}} + e^{-M^2n\epsilon_n^2/4} + 2e^{-(M^2/2-c_3)n\epsilon_n^2} + \frac{1}{c_1}e^{-n\epsilon_n^2(c_2M^2J^2/4-C-1)}\sum_{j=0}^{\infty}\left(e^{-n\epsilon_n^2(c_2M^2/4)}\right)^{j^2} \\
&\le \frac{1}{C^k(n\epsilon_n^2)^{k/2}} + e^{-M^2n\epsilon_n^2/4} + 2e^{-(M^2/2-c_3)n\epsilon_n^2} + \frac{1}{c_1}e^{-n\epsilon_n^2(c_2M^2J^2/4-C-1)}\sum_{j=0}^{\infty}\left(e^{-n\epsilon_n^2(c_2M^2/4)}\right)^{j^2} \\
&\le \frac{1}{C^k(n\epsilon_n^2)^{k/2}} + e^{-M^2n\epsilon_n^2/4} + 2e^{-(M^2/2-c_3)n\epsilon_n^2} + \frac{e^{-n\epsilon_n^2(c_2M^2J^2/4-C-1)}}{c_1(1-e^{-n\epsilon_n^2c_2M^2/4})} \\
&\le \frac{1}{C^k(n\epsilon_n^2)^{k/2}} + e^{-M^2n\epsilon_n^2/4} + 2e^{-(M^2/2-c_3)n\epsilon_n^2} + \frac{2}{c_1}e^{-(c_2M^2J^2/4-C-1)n\epsilon_n^2}. \ \square
\end{aligned}$$

To apply Theorem 10 we define averaged versions of the Hellinger distance, KL divergence and variance measure. Let $p_{\lambda\gamma_i}(N) = p(X^{(i)}|\lambda,\gamma_i)$ and we define the averaged Hellinger distance $h_n(p_\lambda, p_{\lambda'})$ by

$$h_n^2(p_\lambda, p_{\lambda'}) = \frac{1}{n}\sum_{i=1}^{n} 2\left(1 - E_{\lambda\gamma_i}\left(\sqrt{\frac{p_{\lambda\gamma_i}(N)}{p_{\lambda'\gamma_i}(N)}}\right)\right),$$

the averaged KL-divergence as

$$k_n(p_\lambda, p_{\lambda'}) = -\frac{1}{n}\sum_{i=1}^{n} E_{\lambda'\gamma_i}\Big(\log\Big(\frac{p_{\lambda\gamma_i}(N)}{p_{\lambda'\gamma_i}(N)}\Big)\Big),$$

and variance measure as

$$v_n(p_\lambda, p_{\lambda'}) = \frac{1}{n}\sum_{i=1}^{n} Var_{\lambda'\gamma_i}\Big(\log\Big(\frac{p_{\lambda\gamma_i}(N)}{p_{\lambda'\gamma_i}(N)}\Big)\Big).$$

Through component-wise application of the relations in Section A.1 of (Belitser et al., 2015) we have deterministic expressions for these quantities as

$$h_n^2(p_\lambda, p_{\lambda'}) = \frac{1}{n}\sum_{i=1}^{n} 2\Bigg(1 - \exp\Big\{-\frac{1}{2}\int_{R_i}\Big(\sqrt{\lambda(t)\gamma_i(t)} - \sqrt{\lambda'(t)\gamma_i(t)}\Big)^2 dt\Big\}\Bigg),$$

$$k_n(p_\lambda, p_{\lambda'}) = \frac{1}{n}\sum_{i=1}^{n}\Bigg(\int_{R_i}(\lambda(t) - \lambda'(t))\gamma_i(t)dt + \int_{R_i}\lambda'(t)\gamma_i(t)\log\Big(\frac{\lambda'(t)}{\lambda(t)}\Big)dt\Bigg),$$

$$v_n(p_\lambda, p_{\lambda'}) = \frac{1}{n}\sum_{i=1}^{n}\int_{R_i}\lambda'(t)\gamma_i(t)\log^2\Big(\frac{\lambda'(t)}{\lambda(t)}\Big)dt.$$

Lemma 1 of (Belitser et al., 2015) gives bounds on the non-averaged versions of these quantities, but as the bounds will hold for each component of the average, we can trivially extend these results to give the following inequalities:

$$\frac{1}{\sqrt{2}n}\sum_{i=1}^{n}\big(||\sqrt{\lambda\gamma_i} - \sqrt{\lambda'\gamma_i}||_2 \wedge 1\big) \le h_n(p_\lambda, p_{\lambda'}) \le \frac{\sqrt{2}}{n}\sum_{i=1}^{n}\big(||\sqrt{\lambda\gamma_i} - \sqrt{\lambda'\gamma_i}||_2 \wedge 1\big) \quad (30)$$

$$k_n(p_\lambda, p_{\lambda'}) \le \frac{3}{n}\sum_{i=1}^{n}||\sqrt{\lambda\gamma_i} - \sqrt{\lambda'\gamma_i}||_2^2 + v_n(p_\lambda, p_{\lambda'}) \quad (31)$$

$$\frac{1}{n}\sum_{i=1}^{n}||\sqrt{\lambda\gamma_i} - \sqrt{\lambda'\gamma_i}||_2^2 \le \frac{1}{4n}\sum_{i=1}^{n}\int_{R_i}\gamma_i(s)(\lambda(s) \vee \lambda'(s))\log^2\Big(\frac{\lambda(s)}{\lambda'(s)}\Big)ds \quad (32)$$

where for numbers x and y, the minimum is denoted $x \wedge y$ and the maximum is denoted $x \vee y$.

By assumption, λ_0 is bounded away from 0. It follows that any $\lambda \in \Lambda$ with $||\lambda_0 - \lambda||_\infty \le \lambda_{min}$ is also bounded away from 0, and that by the results (31) and (32) above $k_n(p_{\lambda_0}, p_\lambda)$ and $v_n(p_{\lambda_0}, p_\lambda)$ are both bounded by a constant times the averaged uniform norm $\frac{1}{n}\sum_{i=1}^{n}||\lambda_0\gamma_i - \lambda\gamma_i||_\infty$. Therefore for $n \ge n_1$ the ball

$$B_n^*(\epsilon_n) = \Big\{\lambda \in \Lambda : k_n(p_{\lambda_0}, p_\lambda) \le \epsilon_n^2, v_n(p_{\lambda_0}, p_\lambda) \le \epsilon_n^2\Big\}$$

is bounded by a multiple of the ball

$$\Big\{\lambda \in \Lambda : \frac{1}{n}\sum_{i=1}^{n}||\lambda_0\gamma_i - \lambda\gamma_i||_\infty \le \epsilon_n\Big\}$$

for $\epsilon_n \leq \lambda_{min}$. It follows that for $n \geq n_1$, the condition (9) implies

$$\Pi_n(B_n^*(\delta_n)) \geq c_1 e^{-c_2 n \delta_n^2}. \tag{33}$$

By (30) we have that

$$N\left(\frac{\epsilon}{36}, \Lambda_{n,\epsilon}, h_n\right) \leq N\left(\frac{\epsilon}{36\sqrt{2}}, \sqrt{\Lambda_{n,\epsilon}}, \frac{1}{n}\sum_{i=1}^{n} ||\cdot||_2\right)$$

where $\Lambda_{n,\epsilon} = \left\{\lambda \in \Lambda_n : h_n(p_\lambda, p_{\lambda_0}) \leq \epsilon\right\}$. Thus the condition (10) implies (29).

Combining these results we have

$$\frac{\Pi_n(\lambda \in \Lambda_n : j\delta_n < h_n(p_\lambda, p_{\lambda_0}) \leq 2j\delta_n)}{\Pi_n(B_n^*(\delta_n))} \leq \frac{1}{\Pi_n(B_n^*(\delta_n))} \leq c_1 e^{c_2 n \delta_n^2}$$

by (33) to satisfy condition (27), and

$$\frac{\Pi_n(\Lambda_n^c)}{\Pi_n(B_n^*(\delta_n))} \leq \frac{c_4 e^{-c_5 n \delta_n^2}}{c_1 e^{-c_2 n \delta_n^2}} = o(e^{-2n\delta_n^2})$$

by (11) and (33) for $c_5 - c_2 \geq 2$ to satisfy (28). Thus all the conditions of Theorem 10 are satisfied by the assumptions of Theorem 3 and the conclusion of Theorem 10 carries forward to Theorem 3 where we choose $k = 2$. □

4 Conclusion

We have derived finite time rates on the posterior contraction of the QGCP and SGCP models given i.n.i.d observations. This allows us to quantify the contraction of posterior estimates in the setting where events are not detected perfectly or the observation region is not sampled uniformly. As well as a new consistency result for the QGCP model, and the innovations of studying i.n.i.d data over i.i.d., the presentation of explicit rates on the posterior mass for the contraction of non-homogeneous Poisson process models is new. These results are of theoretical importance and practical interest in problems such as sequential decision making and experimental design.

We found that the SGCP model admitted a much tighter analysis than the QGCP model. For the simple setting of 1-dimensional 1-Hölder smooth rate functions, the SGCP model can be shown to have convergence of the near-optimal order $\tilde{o}(n^{-1/3})$. Our best result for the QGCP model only shows convergence of order $\tilde{o}(n^{-1/5})$. This discrepancy arises because of the different link functions used in the two models. In comparison to the bounded sigmoid function, the quadratic function induces a larger space of rate functions when the GP is transformed - meaning that wider sieves are required to give the desired results and the contraction guarantees are looser. Guarantees on the tightness of these bounds are currently unavailable, but this work provides some evidence to suggest that the SGCP model is superior to the QGCP in terms of rate of posterior contraction at least. This is an observation that would merit further empirical and analytical study in to the relationship between the models. We did not consider the LGCP model in this work as its exponential link function makes it very difficult to adapt the existing GP results of (van der Vaart and van Zanten, 2009) into meaningful results in the NHPP posterior contraction setting. In particular, the high probability bound $\{||g - g_0||_\infty \leq \eta_{\beta_n}\}$ on the GP model, does not imply a useful bound on $||e^g - e^{g_0}||_\infty$ - the distance to be bounded in the prior mass condition for the LGCP - that gives useful contraction results.

We have focussed on particular choices of smoothness class, the link function used within the GCP construction and the width of the balls used in the contraction rate statements. There is of course potential to expand on these results by studying other choices. We believe however that the choices we have are consistent with the most common modelling choices in implementation of GCPs and useful for relating our results to the existing literature on posterior contraction of Bayesian nonparametric models.

A Proof of Lemma 4

Proving Lemma 4 relies on a bound on the uniform norm in the GP space. The following lemma gives a particular prior mass result which holds for all $n > 0$ and uses a general term $\eta_{\beta,n}$ which fits with the analysis of both the SGCP and QGCP models.

Lemma 11 *If $g_0 \in \mathcal{C}^\alpha([0,1]^d)$ for some $\alpha > 0$, then there exist constants $c_1, c_2 > 0$ such that*

$$\Pi\big(||g - g_0||_\infty \leq \eta_{\beta,n}\big) \geq c_1 e^{-c_2 n \eta_{\beta,n}^\beta}$$

for $\eta_{\beta,n} = n^{-\alpha/(\beta\alpha+d)}(\log(n))^{\rho_\beta}$ and $\rho_\beta = \frac{1+d}{\beta+d/\alpha}$ for all $n \geq 3$ where $\beta > 1$.

Furthermore, we rely on the following simple result which allows us to move between probabilistic bounds on the uniform norm of the GP and the squared GP.

Lemma 12 *Let w_1 and w_2 be functions defined on $[0,1]^d$ such that $||w_2||_\infty$ is finite, and c be a positive constant. Given the standard definition of the uniform norm, we have the following relation:*

$$\big\{||w_1 - w_2||_\infty \leq c\big\} \Rightarrow \big\{||w_1^2 - w_2^2||_\infty \leq 2c||w_2||_\infty + c^2\big\}.$$

We prove Lemma 11 in Appendix B and prove Lemma 12 below.

Proof of Lemma 12:

We have:

$$\begin{aligned}
&||w_1 - w_2||_\infty \leq c \\
&\quad \Rightarrow w_1(x) \leq w_2(x) + c && \forall\ x \in S \\
&\quad \Rightarrow w_1^2(x) \leq w_2^2(x) + 2cw_2(x) + c^2 && \forall\ x \in S \\
&\Rightarrow ||w_1^2 - w_2^2|| \leq 2c||w_2||_\infty + c^2. && \square
\end{aligned}$$

Proof of Lemma 4:

Recall the defintion $\delta_n = 2\eta_n||g_0||_\infty + \eta_n^2$, with $\eta_n = \eta_{4,n}$. By definition we have:

$$\begin{aligned}
&\Pi\big(\lambda : ||\lambda - \lambda_0||_\infty \leq \delta_n\big) \\
&= \Pi\big(g : ||g^2 - g_0^2||_\infty \leq 2\eta_n||g_0||_\infty + \eta_n^2\big) \\
&\geq \Pi\big(g : ||g - g_0||_\infty \leq \eta_n\big) \\
&\geq c_1 e^{-c_2 n \eta_n^4} \geq c_1 e^{-c_2 n \delta_n^2},
\end{aligned}$$

Here, the first inequality is due to Lemma 12. The second is by application of Lemma 11 and the third is by definition of δ_n. $\square$

B Proof of Lemma 11

We will utilise the following result from Section 5.1 of (van der Vaart and van Zanten, 2009), which holds for a constant H depending only on g_0 and μ, a constant K_2 depending only on g_0, μ, α, d and D_1 and any $\epsilon > 0$

$$\Pi\big(||g - g_0||_\infty \leq 2\epsilon\big) \geq C_1 \exp\left\{ - K_2 \left(\frac{1}{\epsilon}\right)^{d/\alpha} \left(\log\left(\frac{1}{\epsilon}\right)\right)^{1+d} \right\} \left(\frac{H}{\epsilon}\right)^{\frac{q_1+1}{\alpha}}.$$

Recall that C_1, D_1, and q_1 are constants from the assumption (5) on the length scale of the GP prior.

Substituting the particular form of $\epsilon = \epsilon_n = n^{-\alpha/(\beta\alpha+d)}(\log(n))^\rho$ and $\rho = \frac{1+d}{\beta+d/\alpha}$ from Lemma 11 into the above we have:

$$\Pi\big(||g - g_0||_\infty \leq \epsilon_n\big) \geq C_1 \exp\left\{ - 2^{\frac{d}{\alpha}} K_2 n^{\frac{d}{\beta\alpha+d}} (\log(n))^{-\frac{d}{\alpha}\frac{1+d}{\beta+d/\alpha}} \left(\log\left(2n^{\frac{\alpha}{\beta\alpha+d}}(\log(n))^{-\frac{1+d}{\beta+d/\alpha}}\right)\right)^{1+d}\right\}$$
$$\times \left(2Hn^{\frac{\alpha}{\beta\alpha+d}}(\log(n))^{-\frac{1+d}{\beta+d/\alpha}}\right)^{\frac{q_1+1}{\alpha}},$$

defining $Z(n) = \left(Hn^{\frac{\alpha}{\beta\alpha+d}}(\log(n))^{-\frac{1+d}{\beta+d/\alpha}}\right)^{\frac{q_1+1}{\alpha}}$ and expanding the logarithm,

$$= C_1 Z(n) \exp\left\{ - 2^{\frac{d}{\alpha}} K_2 n^{\frac{d}{\beta\alpha+d}} (\log(n))^{-\frac{d}{\alpha}\frac{1+d}{\beta+d/\alpha}} \left(\frac{\alpha}{\beta\alpha+d}\log(2n) - (\log(n))^{\frac{1+d}{\beta+d/\alpha}}\right)^{1+d}\right\}$$
$$\geq C_1 Z(n) \exp\left\{ - 2^{\frac{d}{\alpha}} K_2 n^{\frac{d}{\beta\alpha+d}} (\log(n))^{-\frac{d}{\alpha}\frac{1+d}{\beta+d/\alpha}} \left(\frac{\alpha}{\beta\alpha+d}\log(2n)\right)^{1+d}\right\}$$

using $2\log(n) \geq \log(2n)$ for $n \geq 2$

$$\geq C_1 Z(n) \exp\left\{ - 2^{1+d/\alpha} K_2 n \cdot n^{-\frac{\beta\alpha}{\beta\alpha+d}} (\log(n))^{(1+d)\frac{\beta+d/\alpha-d/\alpha}{\beta+d/\alpha}}\right\}$$
$$= C_1 Z(n) \exp\left\{ - 2^{1+d/\alpha} K_2 n\epsilon_n^\beta\right\}$$

letting $K_3 = \min_{n\geq 3}(Z(n))$

$$\geq C_1 K_3 \exp\left\{ - 2^{1+d/\alpha} K_2 n\epsilon_n^\beta\right\} = c_1 e^{-c_2 n\epsilon_n^\beta},$$

where $c_1 = C_1 K_3$, and $c_2 = 2^{1+d/\alpha} K_2$. $\square$

C Proof of Lemma 5

As $\sqrt{\Lambda_n} = \mathcal{G}_n$, the covering numbers $N(L_1\bar{\delta}_n, \sqrt{\Lambda_n}, ||\cdot||_2)$ and $N(L_1\bar{\delta}_n, \mathcal{G}_n, ||\cdot||_2)$ are equivalent. It follows that

$$N(L_1\bar{\delta}_n, \sqrt{\Lambda_n}, ||\cdot||_2) \leq N(L_1\bar{\delta}_n, \mathcal{G}_n, ||\cdot||_\infty).$$

Defining $\mathcal{G}_n$ as in (17) allows us to use the following result, (5.4) of (van der Vaart and van Zanten, 2009):

$$\log N(L_1\bar{\delta}_n, \mathcal{G}_n, ||\cdot||_\infty) \leq m\zeta_n^d \left(\log \frac{3^{3/2} d^{1/4} \beta_n^{3/2} \sqrt{2\tau\zeta_n}}{(L_1\bar{\delta}_n)^{3/2}} \right)^{1+d} + 2\log \frac{6\beta_n \sqrt{||\mu||}}{L_1\bar{\delta}_n}$$

for $||\mu||$ the total mass of the spectral measure μ, τ^2 as the second moment of μ, positive constant m depending only on μ and d, and given

$$(3/L_1)^{3/2} d^{1/4} \beta_n^{3/2} \sqrt{2\tau\zeta_n} > 2\bar{\delta}_n^{3/2}, \qquad (3/L_1)\beta_n \sqrt{||\mu||} > \bar{\delta}_n.$$

By the definitions of β_n, and ζ_n we have that

$$m\zeta_n^d \left(\log \frac{3^{3/2} d^{1/4} \beta_n^{3/2} \sqrt{2\tau\zeta_n}}{(L_1\bar{\delta}_n)^{3/2}} \right)^{1+d} \leq n\bar{\delta}_n^2$$

$$2\log \frac{6\beta_n \sqrt{||\mu||}}{L_1\bar{\delta}_n} \leq n\bar{\delta}_n^2,$$

for the values of n specified in the statement of Lemma 5. It follows that the lemma is satisfied with $c_3 = 2$. □

D Proof of Lemma 6

Firstly note that $\Pi(\lambda \notin \Lambda_n) = \Pi(g \notin \mathcal{G}_n)$. By a simplification of (5.3) of (van der Vaart and van Zanten, 2009) to account for our assumption that $q_2 = 0$, we have

$$\Pi(g \notin \mathcal{G}_n) \leq C_1 \zeta_n^{q_1 - d + 1} e^{-D_1\zeta_n^d} + e^{-\beta_n^2/8}$$

$\bar{\delta}_n < \delta_0$ for small $\delta_0 > 0$, and β_n, ζ_n, and $\bar{\delta}_n$ satisfying

$$\beta_n^2 > 16K_5\zeta_n^d \left(\log \left(\frac{3\zeta_n}{\bar{\delta}_n} \right) \right)^{1+d}, \qquad \zeta_n > 1,$$

for a constant K_5 depending only on μ and g. The definitions of β_n, δ_n and ζ_n give us the following relations, for a constant $c_5 = c_2 + 4$

$$D_1\zeta_n^d \geq 2c_5 n\delta_n^2, \;\; \beta_n^2 \geq 8c_5 n\delta_n^2, \;\; \zeta_n^{q_1 - d + 1} \leq e^{c_5 n\delta_n^2},$$

with the final of these holding for values of n as specified in the statement of Lemma 6. Using these we can obtain the necessary result as follows:

$$\begin{aligned}\Pi(\lambda : \lambda \notin \Lambda_n) = \Pi(g : g \notin \mathcal{G}_n) &\leq C_1\zeta_n^{q_1-d+1} e^{-D_1\zeta_n^d} + e^{-\beta_n^2/8} \\ &\leq C_1 e^{c_5 n\delta_n^2} e^{-2c_5 n\delta_n^2} + e^{-c_5 n\delta_n^2} \\ &= \big(C_1 + 1\big) e^{-2c_5 n\delta_n^2} \\ &\leq c_4 e^{-(c_2+4)n\delta_n^2}\end{aligned}$$

for $c_4 = C_1 + 1$. □

E Proof of Lemma 7

Under the SGCP model we have

$$\Pi\bigg((\lambda^*, g) : ||\lambda^*\sigma(g) - \lambda_0||_\infty \le \delta_n\bigg) \ge \Pi\bigg(\lambda^* : |\lambda^* - 2||\lambda_0||_\infty| \le \frac{\delta_n}{2}\bigg)\Pi\bigg(g : ||\sigma(g) - \sigma(g_0)||_\infty \le \frac{\delta_n}{4||\lambda_0||_\infty}\bigg).$$

By the assumption that λ^* has a positive continuous density, the first term on the RHS of the inequality can be bounded below by a constant times δ_n, which can itself be lower bounded by a constant for finite n. The second term can be bounded below by $\Pi_n(||g - g_0||_\infty \le \delta_n/(16||\lambda_0|_\infty))$ since $1/4$ is the Lipschitz constant of the sigmoid transformation. Thus, by Lemma 11 (given in Appendix A) we have:

$$\Pi\Big(\lambda : \Gamma_{n,\infty}(\lambda, \lambda_0) \le \delta_n\Big) \ge c_0'\Pi\bigg(g : ||g - g_0||_\infty \le \frac{\delta_n}{16||\lambda_0||_\infty}\bigg) \ge c_1' e^{-nc_2'\delta_n^2}$$

for positive constants c_1', c_2', showing condition (9) is satisfied under the SGCP model.

F Proof of Lemma 8

Define $\psi_n = \bar{\delta}_n/(2L_1\sqrt{\lambda_n})$. We have

$$\begin{aligned}\log N(L_1\bar{\delta}_n, \sqrt{\Lambda_n}, ||\cdot||_2) &= \log N(2\psi_n\sqrt{\lambda_n}, \sqrt{\Lambda_n}, ||\cdot||_2)\\ &\le \log N(\psi_n\sqrt{\lambda_n}, [0, \lambda_n], \sqrt{|\cdot|}) + \log N(\psi_n/c, \mathcal{G}_n, ||\cdot||_\infty)\\ &\le \log\frac{1}{\psi_n} + \log N(\psi_n/c, \mathcal{G}_n, ||\cdot||_\infty) \qquad (34)\end{aligned}$$

for $c = 2^{-5/2}$, the Lipschitz constant of $\sqrt{\sigma}$.

Then, as in the proof of Lemma 5, by equation (5.4) of (van der Vaart and van Zanten, 2009), we have for $B_n > 0$,

$$\log N(B_n\bar{\delta}_n, \mathcal{G}_n, ||\cdot||_\infty) \le m\zeta_n^d\bigg(\log\frac{3^{3/2}d^{1/4}\beta_n^{3/2}\sqrt{2\tau\zeta_n}}{(B_n\bar{\delta}_n)^{3/2}}\bigg)^{1+d} + 2\log\frac{6\beta_n\sqrt{||\mu||}}{B_n\bar{\delta}_n} \qquad (35)$$

subject to the conditions

$$(3/B_n)^{3/2}d^{1/4}\beta_n^{3/2}\sqrt{2\tau\zeta_n} > 2\bar{\delta}_n^{3/2}, \quad (3/B_n)\beta_n\sqrt{||\mu||} > \bar{\delta}_n, \quad \zeta_n > A$$

for a constant $A > 0$. These conditions hold by defintion for n as specified by (23), with $B_n = (2cL_1\sqrt{\lambda_n})^{-1}$. Then, combining (34) and (35) we have

$$\begin{aligned}\log N\big(L_1\bar{\delta}_n, \sqrt{\Lambda_n}, ||\cdot||_2\big) \le \log\frac{2L_1\sqrt{\lambda_n}}{\bar{\delta}_n} + m\zeta_n^d\bigg(\log\frac{(6cL_1)^{3/2}d^{1/4}\beta_n^{3/2}\sqrt{2\tau\lambda_n\zeta_n}}{\bar{\delta}_n^{3/2}}\bigg)^{1+d}\\ + 2\log\frac{12cL_1\beta_n\sqrt{\lambda_n||\mu||}}{\bar{\delta}_n}.\end{aligned}$$

For n as specified by (24), we have

$$\log \frac{2L_1\sqrt{\lambda_n}}{\bar{\delta}_n} < n\bar{\delta}_n^2,$$

$$m\zeta_n^d\bigg(\log \frac{(6cL_1)^{3/2}d^{1/4}\beta_n^{3/2}\sqrt{2\tau\lambda_n\zeta_n}}{\bar{\delta}_n^{3/2}}\bigg)^{1+d} < n\bar{\delta}_n^2,$$

$$2\log \frac{12cL_1\beta_n\sqrt{\lambda_n||\mu||}}{\bar{\delta}_n} < n\bar{\delta}_n^2.$$

Thus for n satisfying (23) and (24) we have

$$\log N\Big(L_1\bar{\delta}_n, \sqrt{\Lambda_n}, ||\cdot||_2\Big) \leq 3n\bar{\delta}_n^2$$

proving Lemma 8 with $c_3 = 3$. □

G Proof of Lemma 9

As in (Kirichenko and Van Zanten, 2015), we may decompose the probability of interest

$$\begin{aligned}\Pi\Big(\lambda : \lambda \notin \Lambda_n\Big) &= \Pi\Big((\lambda^*, g) : \lambda^*\sigma(g) \notin \Lambda_n\Big)\\ &\leq \int_0^{\lambda_n} \Pi\Big((\lambda^*, g) : \lambda^*\sigma(g) \notin \Lambda_n\Big)p_{\lambda^*}(\lambda)d\lambda + \int_{\lambda_n}^{\infty} p_{\lambda^*}(\lambda)d\lambda\\ &\leq \Pi\Big(g : g \notin \mathcal{G}_n\Big) + C_0e^{-c_0\lambda_n^\rho},\end{aligned}$$

by the assumption (6). As utilised in the proof of Lemma 6, equation (5.3) of (van der Vaart and van Zanten, 2009) states that

$$\Pi(g \notin \mathcal{G}_n) \leq C_1\zeta_n^{q_1-d+1}e^{-D_1\zeta_n^d} + e^{-\beta_n^2/8}$$

given conditions

$$\beta_n^2 > 16K_5\zeta_n^d\bigg(\log\Big(\frac{\lambda_n^{1/2}\zeta_n}{\bar{\delta}_n}\Big)\bigg)^{1+d}, \quad \zeta_n > 1$$

which are satisfied by our earlier definitions, for a constant K_5 depending only on μ and g. Then for n as specified by equations (25) and (26), we have the following results

$$c_0\lambda_n^\rho > c_5n\delta_n^2, \quad D_1\zeta_n^d \geq 2c_5n\delta_n^2, \quad \zeta_n^{q_1-d+1} \leq e^{c_5n\delta_n^2}, \quad \beta_n^2 \geq 8c_5n\delta_n^2.$$

The required result then follows. □

H Verifying conditions on sieves

Throughout the analysis the sequences used in defining the sieves are subject to numerous conditions and assumptions, in order that we may demonstrate the conditions of Theorem 3 are met for the GCP models. By choosing $L_{2:10}$ as specified in the main body, these conditions are met by definition for values of n as specified. There are numerous such conditions to verify, and doing so can be non-trivial. In this section we show the link between the conditions and constraints on $L_{2:10}, n$ and demonstrate fully that the necessary results hold.

H.1 QGCP model

Recall, the definitions of the following sequences:

$$\begin{aligned}
\delta_n &= 2||g_0||_\infty n^{-\alpha/(4\alpha+d)} \log^{\rho}(n) + n^{-2\alpha/(4\alpha+d)} \log^{2\rho}(n), \\
\bar{\delta}_n &= 2||g_0||_\infty n^{-\alpha/(4\alpha+d)} \log^{\rho+d+1}(n) + n^{-2\alpha/(4\alpha+d)} \log^{2\rho+2d+2}(n) \\
\zeta_n &= L_2 n^{(2\alpha+d)/(4\alpha d+d^2)} \log^{2\rho/d}(n) + L_3 n^{(\alpha d+d^2)/(4\alpha d+d^2)} \log^{3\rho/d}(n) + L_4 n^{d/(4\alpha+d)} \log^{4\rho/d}(n) \\
\beta_n &= L_5 n^{(2\alpha+d)/(8\alpha+2d)} \log^{2\rho+(d+1)/2}(n) + L_6 n^{(\alpha+d)/(8\alpha+2d)} \log^{3\rho+(d+1)/2}(n) \\
&\quad + L_7 n^{d/(8\alpha+2d)} \log^{4\rho+(d+1)/2}(n)
\end{aligned}$$

with $L_2, ..., L_7$ satisfying

$$\begin{aligned}
L_2 + L_3 + L_4 &> \max(A, e) \\
L_2 L_5^3 &> \left(\frac{8 \max(1, ||g_0||_\infty)}{(3/L_1)^{3/2} d^{1/4} \sqrt{2\tau}} \right)^2 \\
L_5 + L_6 + L_7 &> \frac{4 L_1 \max(1, ||g_0||_\infty)}{3\sqrt{||\mu||}} \\
L_2 &\geq (8c_5 ||g_0||_\infty^2)/D_1 \\
L_3 &\geq (8c_5 ||g_0||_\infty)/D_1 \\
L_4 &\geq 2c_5/D_1 \\
L_5 &\geq \max\left(\sqrt{\frac{16 K_5 L_2^d \mathcal{K}_1^{1+d}}{\log^{2\rho}(3)}}, \sqrt{32||g_0||_\infty^2 c_5} \right) \\
L_6 &\geq \max\left(\sqrt{\frac{16 K_5 L_3^d \mathcal{K}_1^{1+d}}{\log^{3\rho}(3)}}, \sqrt{32||g_0||_\infty c_5} \right) \\
L_7 &\geq \max\left(\sqrt{\frac{16 K_5 L_4^d \mathcal{K}_1^{1+d}}{\log^{4\rho}(3)}}, \sqrt{8c_5} \right)
\end{aligned}$$

for $n \geq \max(3, n_3, n_4, n_5)$. Here n_3 is the smallest integer n such that

$$\begin{aligned}
4||g_0||_\infty^2 \log^{2d+2-2\rho}(n) \geq \frac{m \sum_{i=2}^4 L_i^d}{2^{1+d}} \Bigg(\log \Big(\frac{27\tau\sqrt{d}(\sum_{i=5}^7 L_i)^3 \sum_{i=2}^4 L_i}{4||g_0||_\infty^3} \Big) \\
+ \left(4 + \frac{12+d+d^2}{8\alpha d + 2d^2}\right) \log(n) \Bigg)^{1+d}
\end{aligned}$$

n_4 is the smallest integer n such that

$$2\log\left(\frac{6\sqrt{||\mu||}(L_5+L_6+L_7)}{2L_1||g_0||_\infty + L_1} \right) \leq 4||g_0||_\infty^2 n^{(4\alpha+2d)/(8\alpha+2d)} \log^{2\rho+2d+2}(n) - \log\left(n^{(6\alpha+d)/(4\alpha+d)} \log^{6\rho}(n) \right)$$

and n_5 is the smallest integer n such that

$$n^{(2\alpha+d)/(4\alpha+d)} \log^{2\rho}(n) \geq \frac{q_1}{4c_5||g_0||_\infty^2} \log\Big((L_2+L_3+L_4) n^{(2\alpha+d)/(4\alpha d+d^2)} \log^{4\rho/d}(n) \Big).$$

In the remainder of this subsection, we show that the following conditions, which are all restatements of required results in our main analysis, hold for the sequences described above.

$$\zeta_n > \max(A, 1) \tag{36}$$

$$(3/L_1)^{3/2} d^{1/4} \beta_n^{3/2} \sqrt{2\tau\zeta_n} > 2\bar{\delta}_n^{3/2} \tag{37}$$

$$(3/L_1)\beta_n \sqrt{||\mu||} > \bar{\delta}_n \tag{38}$$

$$m\zeta_n^d \Bigg(\log\left(\frac{(3/L_1)^{3/2} d^{1/4} \beta_n^{3/2} \sqrt{2\tau\zeta_n}}{\bar{\delta}_n^{3/2}} \right) \Bigg)^{1+d} \leq n\bar{\delta}_n^2 \tag{39}$$

$$2\log\left(\frac{6\beta_n \sqrt{||\mu||}}{L_1 \bar{\delta}_n} \right) \leq n\bar{\delta}_n^2 \tag{40}$$

$$\beta_n^2 > 16K_5\zeta_n^d \Bigg(\log\left(\frac{3\zeta_n}{\bar{\delta}_n} \right) \Bigg)^{1+d} \tag{41}$$

$$D_1\zeta_n^d \Bigg(\log^{q_2}(\zeta_n) \Bigg) \geq 2c_5 n\delta_n^2 \tag{42}$$

$$\beta_n^2 \geq 8c_5 n\delta_n^2 \tag{43}$$

$$\zeta_n^{q_1 - d + 1} \leq \exp(c_5 n\delta_n^2), \tag{44}$$

H.1.1 Verifying (36)

For $n = 3$, $\log(n) > 1$ thus $\zeta_n > L_2 + L_3 + L_4$ for all $\alpha \in [0, 1]$, and $d \geq 1$. It follows that (36) is satisfied for $n = 3$ given $L_2 + L_3 + L_4 > \max(A, e)$. To show it holds for all $n > 3$ we simply note that ζ_n is an increasing function.

H.1.2 Verifying (37)

First consider,

$$\begin{aligned}
\bar{\delta}_n &= 2||g_0||_\infty n^{-\alpha/(4\alpha+d)} \log^{\rho+d+1}(n) + n^{-2\alpha/(4\alpha+d)} \log^{2\rho+2d+2}(n) \\
&\leq 4\max(1, ||g_0||_\infty) n^{-\alpha/(4\alpha+d)} \log^{2\rho+2d+2}(n) \\
\Rightarrow \quad 2\bar{\delta}_n^{3/2} &\leq (4\max(1, ||g_0||_\infty))^{3/2} n^{-3\alpha/(8\alpha+2d)} \log^{3\rho+3d+3}(n))^{3/2} \\
&= (4\max(1, ||g_0||_\infty))^{3/2} n^{-3\alpha/(8\alpha+2d)} \log^{3\rho+3(d+1)/4}(n) \log^{9(d+1)/4}(n)
\end{aligned}$$

Let $z_1 = (3/L_1)^{3/2} d^{1/4} \sqrt{2\tau}$,

$$\begin{aligned}
z_1\sqrt{\beta_n^3\zeta_n} &\geq z_1 \log^{4\rho+3(d+1)/4}(n) \sqrt{\left(L_5 n^{\frac{2\alpha+d}{8\alpha+2d}} + L_6 n^{\frac{\alpha+d}{8\alpha+2d}} + L_7 n^{\frac{d}{8\alpha+2d}}\right)^3 \left(L_2 n^{\frac{2\alpha+d}{4\alpha d+d^2}} + L_3 n^{\frac{\alpha+d}{4\alpha d+d^2}} + L_4 n^{\frac{d}{4\alpha d+d^2}}\right)} \\
&\geq z_1 \log^{3\rho+3(d+1)/4}(n) \sqrt{L_5^3 L_2 n^{\frac{6\alpha+3d}{8\alpha+2d}}}.
\end{aligned}$$

Thus values of L_2, L_5 such that

$$z_1\sqrt{L_2 L_5^3} n^{\frac{6\alpha+3d}{16\alpha+4d} + \frac{3\alpha}{8\alpha+2d}} > 8\max(1, ||g_0||_\infty) \log^{9(d+1)/4}(n)$$

are sufficient to verify (37). For $n \geq 3$, $d > 1$ and $\alpha \in [0,1]$ $n^{\frac{12\alpha+3d}{16\alpha+4d}} > \log^{9(d+1)/4}(n)$ so

$$L_2 L_5^3 > \left(\frac{8 \max(1, ||g_0||_\infty)}{(3/L_1)^{3/2} d^{1/4} \sqrt{2\tau}} \right)^2$$

is a sufficient condition to verify (37).

H.1.3 Verifying (38)

First consider,

$$L_1 \bar{\delta}_n \leq 4 L_1 \max(1, ||g_0||_\infty) n^{-\alpha/(4\alpha+d)} \log^{2\rho+2d+2}(n),$$

and

$$3\sqrt{||\mu||}\beta_n \geq 3\sqrt{||\mu||}(L_5 + L_6 + L_7) n^{(2\alpha+d)/(8\alpha+2d)} \log^{2\rho+(d+1)/2}(n).$$

Plainly $n^{(2\alpha+d)/(8\alpha+2d)} \log^{2\rho+(d+1)/2}(n) > n^{-\alpha/(4\alpha+d)} \log^{2\rho+2d+2}(n)$ for $n \geq 3$, so

$$L_5 + L_6 + L_7 > \frac{4 L_1 \max(1, ||g_0||_\infty)}{3\sqrt{||\mu||}}$$

is a sufficient condition to verify (38).

H.1.4 Verifying (39)

Consider,

$$
\begin{aligned}
& m\zeta_n^d\Bigg(\log\left(\frac{3^{3/2} d^{1/4} \beta_n^{3/2} \sqrt{2\tau\zeta_n}}{(L_1\bar{\delta}_n)^{3/2}} \right) \Bigg)^{1+d} \\
\leq\ & m(L_2^d + L_3^d + L_4^d) n^{\frac{2\alpha+d}{4\alpha+d}} \log^{4\rho}(n) \Bigg(\frac{3}{2}\log\left(\frac{3\beta_n}{L_1\bar{\delta}_n} \right) + \frac{1}{2}\log(2\tau d^{1/2}\zeta_n) \Bigg)^{1+d}, \\
\leq\ & m(L_2^d + L_3^d + L_4^d) n^{\frac{2\alpha+d}{4\alpha+d}} \log^{4\rho}(n) \Bigg(\frac{3}{2}\log\left(\frac{3(L_5+L_6+L_7)}{2||g_0||_\infty} n^{\frac{6\alpha+d}{8\alpha+2d}} \log^{\rho}(n) \right) + \\
& \qquad \frac{1}{2}\log(2\tau\sqrt{d}(L_2+L_3+L_4) n^{\frac{2\alpha+d}{4\alpha d+d^2}} \log^{4\rho/d}(n)) \Bigg)^{1+d} \\
\leq\ & \frac{m(L_2^d + L_3^d + L_4^d)}{2^{1+d}} \Bigg(3\log\left(\frac{3(L_5+L_6+L_7)}{2||g_0||_\infty} \right) + \frac{18\alpha+3d}{8\alpha+2d}\log(n) + 3\rho\log(\log(n)) \\
& \qquad + \log(2\tau\sqrt{d}(L_2+L_3+L_4)) + \frac{2\alpha+d}{4\alpha d+d^2}\log(n) + \frac{4\rho}{d}\log(\log(n)) \Bigg)^{1+d} n^{\frac{2\alpha+d}{4\alpha+d}} \log^{4\rho}(n) \\
=\ & \frac{m(L_2^d + L_3^d + L_4^d)}{2^{1+d}} \Bigg(\log\left(\frac{27\tau\sqrt{d}(L_5+L_6+L_7)^3}{4||g_0||_\infty^3} (L_2+L_3+L_4) \right) + \left(\frac{18\alpha d + 4\alpha + 2d + 3d^2}{8\alpha d + 2d^2} \right)\log(n) \\
& \qquad + \left(3\rho + \frac{4\rho}{d} \right)\log(\log(n)) \Bigg)^{1+d} n^{\frac{2\alpha+d}{4\alpha+d}} \log^{4\rho}(n) \\
\leq\ & \frac{m\sum_{i=2}^4 L_i^d}{2^{1+d}} \Bigg(\log\left(\frac{27\tau\sqrt{d}(\sum_{i=5}^7 L_i)^3}{4||g_0||_\infty^3} \sum_{i=2}^4 L_i \right) + \left(\frac{18\alpha d + 4\alpha + 2d + 3d^2}{8\alpha d + 2d^2} + \frac{3\rho d + 4\rho}{d} \right)\log(n) \Bigg)^{1+d} n^{\frac{2\alpha+d}{4\alpha+d}} \log^{4\rho}(n) \\
\leq\ & \frac{m\sum_{i=2}^4 L_i^d}{2^{1+d}} \Bigg(\log\left(\frac{27\tau\sqrt{d}(\sum_{i=5}^7 L_i)^3}{4||g_0||_\infty^3} \sum_{i=2}^4 L_i \right) + \frac{32\alpha d + 12\alpha + 2d + 3d^2 + 6\alpha d^2}{8\alpha d + 2d^2}\log(n) \Bigg)^{1+d} n^{\frac{2\alpha+d}{4\alpha+d}} \log^{4\rho}(n) \\
\leq\ & \frac{m\sum_{i=2}^4 L_i^d}{2^{1+d}} \Bigg(\log\left(\frac{27\tau\sqrt{d}(\sum_{i=5}^7 L_i)^3}{4||g_0||_\infty^3} \sum_{i=2}^4 L_i \right) + \left(4 + \frac{12 + d + d^2}{8\alpha d + 2d^2} \right)\log(n) \Bigg)^{1+d} n^{\frac{2\alpha+d}{4\alpha+d}} \log^{4\rho}(n)
\end{aligned}
$$

and

$$
n\bar{\delta}_n^2 \geq 4||g_0||_\infty^2 n^{\frac{2\alpha+d}{4\alpha+d}} \log^{2\rho+2(d+1)}(n)
$$

Condition (39) is then satisfied for all n such that

$$
\begin{aligned}
4||g_0||_\infty^2 \log^{2d+2-2\rho}(n) \geq \frac{m\sum_{i=2}^4 L_i^d}{2^{1+d}} \Bigg(& \log\left(\frac{27\tau\sqrt{d}(\sum_{i=5}^7 L_i)^3 \sum_{i=2}^4 L_i}{4||g_0||_\infty^3} \right) \\
& + \left(4 + \frac{12 + d + d^2}{8\alpha d + 2d^2} \right)\log(n) \Bigg)^{1+d}
\end{aligned}
$$

H.1.5 Verifying (40)

Consider,

$$
\begin{aligned}
2\log\left(\frac{6\beta_n\sqrt{||\mu||}}{L_1\bar{\delta}_n}\right) &\le 2\log\left(\frac{6\sqrt{||\mu||}(L_5+L_6+L_7)n^{(2\alpha+d)/(8\alpha+2d)}\log^{4\rho+(d+1)/2}(n)}{(2||g_0||_\infty+1)L_1 n^{-2\alpha/(4\alpha+d)}\log^{\rho+(d+1)/2}(n)}\right)\\
&= 2\log\left(\frac{6\sqrt{||\mu||}(L_5+L_6+L_7)}{2L_1||g_0||_\infty+L_1}n^{(6\alpha+d)/(8\alpha+2d)}\log^{3\rho}(n)\right)\\
&= 2\log\left(\frac{6\sqrt{||\mu||}(L_5+L_6+L_7)}{2L_1||g_0||_\infty+L_1}\right)+\log\left(n^{(6\alpha+d)/(4\alpha+d)}\log^{6\rho}(n)\right)
\end{aligned}
$$

and

$$
\begin{aligned}
n\bar{\delta}_n^2 &= 4||g_0||_\infty^2 n^{(2\alpha+d)/(4\alpha+d)}\log^{2\rho+2d+2}(n)+4||g_0||_\infty n^{(\alpha+d)/(4\alpha+d)}\log^{3\rho+3d+3}(n)+n^{d/(4\alpha+d)}\log^{4\rho+4d+4}(n)\\
&\ge 4||g_0||_\infty^2 n^{(4\alpha+2d)/(8\alpha+2d)}\log^{2\rho+2d+2}(n)
\end{aligned}
$$

Therefore, condition (40) holds for all n such that

$$
2\log\left(\frac{6\sqrt{||\mu||}(L_5+L_6+L_7)}{2L_1||g_0||_\infty+L_1}\right)\le 4||g_0||_\infty^2 n^{(4\alpha+2d)/(8\alpha+2d)}\log^{2\rho+2d+2}(n)-\log\left(n^{(6\alpha+d)/(4\alpha+d)}\log^{6\rho}(n)\right)
$$

H.1.6 Verifying (41)

Consider

$$
\beta_n^2 = L_5^2 n^{\frac{2\alpha+d}{4\alpha+d}}\log^{4\rho+d+1}(n)+L_6^2 n^{\frac{\alpha+d}{4\alpha+d}}\log^{6\rho+d+1}(n)+L_7^2 n^{\frac{d}{4\alpha+d}}\log^{8\rho+d+1}(n)
$$

and

$$
\begin{aligned}
\zeta_n^d\Bigg(\log\left(\frac{3\zeta_n}{\bar{\delta}_n}\right)\Bigg)^{1+d} &= \left(L_2^d n^{\frac{2\alpha+d}{4\alpha+d}}\log^{2\rho}(n)+L_3^d n^{\frac{\alpha+d}{4\alpha+d}}\log^{3\rho}(n)+L_4^d n^{\frac{d}{4\alpha+d}}\log^{4\rho}(n)\right)\\
&\quad\times\Bigg(\log\Bigg(\frac{3L_2 n^{(2\alpha+d)/(4\alpha d+d^2)}\log^{2\rho/d}(n)+3L_3 n^{(\alpha+d)/(4\alpha d+d^2)}\log^{3\rho/d}(n)+3L_4 n^{d/(4\alpha d+d^2)}\log}{2||g_0||_\infty n^{-\alpha/(4\alpha+d)}\log^{\rho+d+1}(n)+n^{-2\alpha/(4\alpha+d)}\log^{2\rho+2d+2}(n)}\\
&\le \left(L_2^d n^{\frac{2\alpha+d}{4\alpha+d}}\log^{2\rho}(n)+L_3^d n^{\frac{\alpha+d}{4\alpha+d}}\log^{3\rho}(n)+L_4^d n^{\frac{d}{4\alpha+d}}\log^{4\rho}(n)\right)\\
&\quad\times\left(\log\left(\frac{3(L_2+L_3+L_4)n^{(2\alpha+d)/(4\alpha d+d^2)}\log^{4\rho/d}(n)}{\min(1,2||g_0||_\infty)n^{-2\alpha/(4\alpha+d)}\log^{\rho+d+1)}(n)}\right)\right)^{1+d}\\
&\le \left(L_2^d n^{\frac{2\alpha+d}{4\alpha+d}}\log^{2\rho}(n)+L_3^d n^{\frac{\alpha+d}{4\alpha+d}}\log^{3\rho}(n)+L_4^d n^{\frac{d}{4\alpha+d}}\log^{4\rho}(n)\right)\\
&\quad\times\left[\log\left(\frac{3(L_2+L_3+L_4)}{\min(1,2||g_0||_\infty)}\right)+\log\left(n^{\frac{2\alpha+d}{4\alpha d+d^2}+\frac{2\alpha}{4\alpha+d}}\log^{4\rho-\rho/d-d-1}(n)\right)\right]^{1+d}.
\end{aligned}
$$

Define

$$
\begin{aligned}
\mathcal{K}(n) &= \left[\log\left(\frac{3(L_2+L_3+L_4)}{\min(1,2||g_0||_\infty)}\right)+\frac{2\alpha d+2\alpha+d}{4\alpha d+d^2}\log(n)+(4\rho-\rho/d-d-1)\log(\log(n))\right]^{1+d}\\
&\le \log^{1+d}(n)\Bigg(\log\left(\frac{3(L_2+L_3+L_4)}{\min(1,2||g_0||_\infty)}\right)+\frac{2\alpha d+2\alpha+d}{4\alpha d+d^2}+(4\rho-\rho/d-d-1)\Bigg)^{1+d},
\end{aligned}
$$

for $n \geq 3$. Let $\mathcal{K}_1 = \left(\log\left(\frac{3(L_2+L_3+L_4)}{\min(1,2||g_0||_\infty)}\right) + \frac{2\alpha d+2\alpha+d}{4\alpha d+d^2} + (4\rho - \rho/d - d - 1)\right)$. Grouping terms of the same order we require the following for all sufficiently large n

$$L_5^2 \log^{2\rho}(n) \geq 16K_5 L_2^d \mathcal{K}_1^{1+d},$$
$$L_6^2 \log^{3\rho}(n) \geq 16K_5 L_3^d \mathcal{K}_1^{1+d},$$
$$L_7^2 \log^{4\rho}(n) \geq 16K_5 L_4^3 \mathcal{K}_1^{1+d},$$

to satisfy (41). Thus, the following are sufficient conditions to satisfy (41) for all $n \geq 3$

$$L_5 \geq \sqrt{\frac{16K_5 L_2^d \mathcal{K}_1^{1+d}}{\log^{2\rho}(3)}}, \quad L_6 \geq \sqrt{\frac{16K_5 L_3^d \mathcal{K}_1^{1+d}}{\log^{3\rho}(3)}}, \quad L_7 \geq \sqrt{\frac{16K_5 L_4^d \mathcal{K}_1^{1+d}}{\log^{4\rho}(3)}}.$$

H.1.7 Verifying (42)

Consider,

$$2c_5 n\delta_n^2 = 2c_5\left(4||g_0||_\infty^2 n^{\frac{2\alpha+d}{4\alpha+d}} \log^{2\rho}(n) + 4||g_0||_\infty n^{\frac{\alpha+d}{4\alpha+d}} \log^{3\rho}(n) + n^{\frac{d}{4\alpha+d}} \log^{4\rho}(n)\right),$$

and

$$D_1\zeta_n^d \log^{q_2}(\zeta_n) \geq D_1\left(L_2 n^{\frac{2\alpha+d}{4\alpha+d}} \log^{2\rho}(n) + L_3 n^{\frac{\alpha+d}{4\alpha+d}} \log^{3\rho}(n) + L_4 n^{\frac{d}{4\alpha+d}} \log^{4\rho}(n)\right)$$

for $\zeta_n > e$ - i.e. such that $\log(\zeta_n) \geq 1$. Then condition (36) and $L_2 \geq (8c_5||g_0||_\infty^2)/D_1$, $L_3 \geq (8c_5||g_0||_\infty)/D_1$ and $L_4 \geq 2c_5/D_1$ are sufficient conditions to verify (42).

H.1.8 Verifying (43)

Consider,

$$8c_5 n\delta_n^2 = 8c_5\left(4||g_0||_\infty^2 n^{\frac{2\alpha+d}{4\alpha+d}} \log^{2\rho}(n) + 4||g_0||_\infty n^{\frac{\alpha+d}{4\alpha+d}} \log^{3\rho}(n) + n^{\frac{d}{4\alpha+d}} \log^{4\rho}(n)\right),$$

and

$$\beta_n^2 \geq L_5^2 n^{\frac{2\alpha+d}{4\alpha+d}} \log^{2\rho}(n) + L_6^2 n^{\frac{\alpha+d}{4\alpha+d}} \log^{3\rho}(n) + L_7^2 n^{\frac{d}{4\alpha+d}} \log^{4\rho}(n).$$

Then (42) is satisfied with $L_5^2 > 32||g_0||_\infty^2 c_5$, $L_6^2 > 32||g_0||_\infty c_5$, and $L_7^2 > 8c_5$.

H.1.9 Verifying (44)

Consider

$$\exp(c_5 n\delta_n^2) = \exp\left(c_5\left(4||g_0||_\infty^2 n^{\frac{2\alpha+d}{4\alpha+d}} \log^{2\rho}(n) + 4||g_0||_\infty n^{\frac{\alpha+d}{4\alpha+d}} \log^{3\rho}(n) + n^{\frac{d}{4\alpha+d}} \log^{4\rho}(n)\right)\right),$$
$$\geq \exp\left(4c_5||g_0||_\infty^2 n^{(2\alpha+d)/(4\alpha+d)} \log^{2\rho}(n)\right)$$

and

$$
\begin{aligned}
\zeta_n^{q_1-d+1} &\leq \left(L_2 n^{(2\alpha+d)/(4\alpha d+d^2)} \log^{2\rho/d}(n) + L_3 n^{(\alpha+d)/(4\alpha d+d^2)} \log^{3\rho/d}(n) + L_4 n^{d/(4\alpha d+d^2)} \log^{4\rho/d}(n)\right)^{q_1}, \\
&\leq \left((L_2+L_3+L_4) n^{(2\alpha+d)/(4\alpha d+d^2)} \log^{4\rho/d}(n)\right)^{q_1}
\end{aligned}
$$

The condition is then satisfied for all n such that

$$
n^{(2\alpha+d)/(4\alpha+d)} \log^{2\rho}(n) \geq \frac{q_1}{4c_5||g_0||_\infty^2} \log\Big((L_2+L_3+L_4) n^{(2\alpha+d)/(4\alpha d+d^2)} \log^{4\rho/d}(n)\Big).
$$

H.2 SGCP model

Recall the definitions of the following sequences:

$$
\begin{aligned}
\delta_n &= n^{-\alpha/(2\alpha+d} \log^{\rho}(n) \\
\bar{\delta}_n &= n^{-\alpha/(2\alpha+d} \log^{\rho+d+1}(n) \\
\zeta_n &= L_8 n^{\frac{1}{2\alpha+d}} (\log(n))^{2\rho/d}, \\
\beta_n &= L_9 n^{\frac{d}{2(2\alpha+d)}} (\log(n))^{d+1+2\rho}, \\
\lambda_n &= L_{10} n^{\frac{d}{\kappa(2\alpha+d)}} (\log(n))^{4\rho/\kappa}
\end{aligned}
$$

with L_8, L_9, L_{10} satisfying

$$
\begin{aligned}
L_8 &> \max\left(A, 1, \left(\frac{2c_5}{D_1}\right)^{1/d}\right) \\
L_9 &\geq \sqrt{8c_5} \\
L_{10} &> \left(\frac{c_5}{c_0}\right)^{1/\rho} \\
L_8 L_9^3 L_{10}^{3/2} &> \frac{2}{()6cL_1)^{3/2}\tau\sqrt{d}} \\
L_9 L_{10}^{1/2} &> \frac{1}{6cL_1\sqrt{||\mu||}}
\end{aligned}
$$

for $n \geq \max(3, n_6, n_7, n_8)$. Here n_6 is the smallest integer such that

$$
n^{\frac{d}{2\alpha+d}} > \max\left(2\log(12cL_1L_9L_{10}^{1/2}) + 1, \log(2L_1L_{10}^{1/2}) + 1, \frac{1}{c_5}\big(\log(L_8^{q_1-d+1}) + 1\big)\right),
$$

n_7 is the smallest integer such that

$$
\log^{2d+2}(n) > mL_8^d\Big(\log((6cL_1)^{3/2}\sqrt{2\tau L_8 L_9^3} L_{10}^{3/4} d^{1/4}) + \frac{\kappa(6d+6\alpha+2)+3d}{4\kappa(2\alpha+d)}\log(n) + \log\Big(\log^{3\rho/2+3\rho/\kappa+\rho/d-d-1}(n)\Big)\Big)
$$

and n_8 is the smallest integer such that

$$
\log^{2\rho}(n) > \frac{16K_5D_1L_8^d}{L_9^2}\left(\frac{\log(\sqrt{L_{10}}L_8) + \log\big(n^{\frac{2\alpha\kappa+2\kappa+d}{2\kappa(2\alpha+d)}} \log^{\rho(2/\kappa+2/d-1)-d-1}(n)\big)}{\log(n)}\right)^{1+d}.
$$

In the remainder of this subsection, we show that the following conditions, which are all restatements of required results in our main analysis, hold for the sequences described above.

$$(6cL_1)^{3/2}d^{1/4}\beta_n^{3/2}\lambda_n^{3/4}\sqrt{2\tau\zeta_n} > 2\bar{\delta}_n^{3/2} \tag{45}$$

$$6cL_1\beta_n\sqrt{\lambda_n||\mu||} > \bar{\delta}_n \tag{46}$$

$$\zeta_n > \max(A, 1) \tag{47}$$

$$m\zeta_n^d\Bigg(\log\Big(\frac{(6cL_1)^{3/2}\lambda_n^{3/4}\beta_n^{3/2}d^{1/4}\sqrt{2\tau\zeta_n}}{\bar{\delta}_n^{3/2}}\Big)\Bigg)^{1+d} < K_3 n\bar{\delta}_n^2 \tag{48}$$

$$2\log\Big(\frac{12cL_1\beta_n\sqrt{\lambda_n||\mu||}}{\bar{\delta}_n}\Big) < K_4 n\bar{\delta}_n^2 \tag{49}$$

$$\log\Big(\frac{2L_1\lambda_n^{1/2}}{\bar{\delta}_n}\Big) < K_5 n\bar{\delta}_n^2 \tag{50}$$

$$\beta_n^2 > 16K_5\zeta_n^d\Bigg(\log\Big(\frac{\lambda_n^{1/2}\zeta_n}{\bar{\delta}_n}\Big)\Bigg)^{1+d} \tag{51}$$

$$c_0\lambda_n^{\rho} > c_5 n\delta_n^2 \tag{52}$$

$$D_1\zeta_n^d \geq 2c_5 n\delta_n^2 \tag{53}$$

$$\zeta_n^{q_1-d+1} \leq e^{c_5 n\delta_n^2} \tag{54}$$

$$\beta_n^2 \geq 8c_5 n\delta_n^2 \tag{55}$$

In turn we demonstrate that each of the conditions (45) through (55) hold.

H.2.1 Verifying (45)

Consider

$$\begin{aligned}
&(6cL_1)^{3/2}d^{1/4}\beta_n^{3/2}\lambda_n^{3/4}\sqrt{2\tau\zeta_n}\\
&= (6cL_1)^{3/2}d^{1/4}L_9^{3/2}n^{\frac{3d}{4(2\alpha+d)}}\log^{\frac{3d+3}{4}+3\rho}(n)L_{10}^{3/4}n^{\frac{3d}{4\kappa(2\alpha+d)}}\log^{3\rho/\kappa}(n)\sqrt{2\tau L_8 n^{\frac{1}{2\alpha+d}}(\log(n))^{2\rho/d}}\\
&= (6cL_1)^{3/2}\sqrt{2\tau L_8 L_9^3}L_{10}^{3/4}d^{1/4}n^{\frac{3d}{4(2\alpha+d)}+\frac{3d}{4\kappa(2\alpha+d)}+\frac{1}{2(2\alpha+d)}}\log^{\frac{3d+3}{4}+3\rho+3\rho/\kappa+\rho/d}(n)
\end{aligned}$$

and

$$2\bar{\delta}_n^{3/2} = 2n^{\frac{-3\alpha}{2(2\alpha+d)}}\log^{3\rho/2+3(d+1)/2}(n)$$

So (45) can be rewritten:

$$(6cL_1)^{3/2}\sqrt{2\tau L_8 L_9^3}L_{10}^{3/4}d^{1/4}n^{\frac{3d}{4(2\alpha+d)}+\frac{3d}{4\kappa(2\alpha+d)}+\frac{3\alpha+1}{2(2\alpha+d)}}\log^{3\rho/2+3\rho/\kappa+\rho/d-\frac{3d-3}{4}}(n) > 2,$$

which holds for L_8, L_9, L_{10} such that $L_8L_9^3L_{10}^{3/2} > \frac{2}{(6cL_1)^{3/2}\tau\sqrt{d}}$.

H.2.2 Verifying (46)

We may rewrite (46) as

$$6cL_1\sqrt{||\mu||}L_9L_{10}^{1/2}n^{\frac{1}{2}}\log^{\rho+2\rho/\kappa-d-1}(n) > 1$$

which holds for all L_9, L_{10} such that $L_9L_{10}^{1/2} > 1/(6cL_1\sqrt{||\mu||})$.

H.2.3 Verifying (47)

If $n \geq 3$ then ζ_n holds for all $L_8 \geq \max(A, 1)$.

H.2.4 Verifying (48)

Consider

$$m\zeta_n^d \bigg(\log \Big(\frac{(6cL_1)^{3/2} \lambda_n^{3/4} \beta_n^{3/2} d^{1/4} \sqrt{2\tau\zeta_n}}{\bar{\delta}_n^{3/2}} \Big) \bigg)^{1+d}$$

$$= mL_8^d n^{\frac{d}{2\alpha+d}} \log^{2\rho}(n) \log \left((6cL_1)^{3/2} \sqrt{2\tau L_8 L_9^3} L_{10}^{3/4} d^{1/4} n^{\frac{3d}{4(2\alpha+d)} + \frac{3d}{4\kappa(2\alpha+d)} + \frac{3\alpha+1}{2(2\alpha+d)}} \log^{3\rho/2+3\rho/\kappa+\rho/d-d-1}(n) \right)^{1+d}$$

and

$$n\bar{\delta}_n^2 = n^{\frac{d}{2\alpha+d}} \log^{2\rho+2d+2}(n)$$

Thus (48) holds for all n such that

$$\log^{2d+2}(n) > mL_8^d \bigg(\log((6cL_1)^{3/2} \sqrt{2\tau L_8 L_9^3} L_{10}^{3/4} d^{1/4}) + \frac{\kappa(6d+6\alpha+2)+3d}{4\kappa(2\alpha+d)} \log(n) + \log \Big(\log^{3\rho/2+3\rho/\kappa+\rho/d-d-1}(n)$$

H.2.5 Verifying (49)

We may rewrite (49) as

$$2\log \left(12cL_1 L_9 L_{10}^{1/2} n^{\frac{1}{2}+\frac{d}{2\kappa(2\alpha+d)}} \log^{2\rho/\kappa+\rho-d-1}(n) \right) < n^{\frac{d}{2\alpha+d}} \log^{2\rho+d+1}(n)$$

which holds for all n such that

$$n^{\frac{d}{2\alpha+d}} > 2\log(12cL_1 L_9 L_{10}^{1/2}) + 1.$$

H.2.6 Verifying (50)

We may rewrite (50) as

$$\log \left(2L_1 L_{10}^{1/2} n^{\frac{d}{2\kappa(2\alpha+d)}+\frac{\alpha}{2\alpha+d}} \log^{2\rho/\kappa-\rho-d-1}(n) \right) < n^{\frac{d}{2\alpha+d}} \log^{2\rho+d+1}(n)$$

which holds for all n such that

$$n^{\frac{d}{2\alpha+d}} > \log(2L_1 L_{10}^{1/2}) + 1.$$

H.2.7 Verifying (51)

Consider

$$\beta_n^2 = L_9^2 n^{\frac{d}{2\alpha+d}} \log^{d+1+4\rho}(n)$$

and

$$16K_5\zeta_n^d\bigg(\log\Big(\frac{\lambda_n^{1/2}\zeta_n}{\bar{\delta}_n}\Big)\bigg)^{1+d} = 16K_5D_1L_8^d n^{\frac{d}{2\alpha+d}}\log^{2\rho}(n)\bigg(\log\Big(\frac{\sqrt{L_{10}}L_8 n^{\frac{2\kappa+d}{2\kappa(2\alpha+d)}}\log^{2\rho/\kappa 2+2\rho/d}(n)}{n^{\frac{-\alpha}{2\alpha+d}}\log^{\rho+d+1}(n)}\Big)\bigg)^{1+d}$$

$$= 16K_5D_1L_8^d n^{\frac{d}{2\alpha+d}}\log^{2\rho}(n)\bigg(\log(\sqrt{L_{10}}L_8)+\log\Big(n^{\frac{2\alpha\kappa+2\kappa+d}{2\kappa(2\alpha+d)}}\log^{\rho(2/\kappa+2/d-1)-d-1}(n)\Big)\bigg)^{1+}$$

So (51) can then be rewritten as

$$L_9^2\log^{d+1+2\rho}(n) > 16K_5D_1L_8^d\bigg(\log(\sqrt{L_{10}}L_8)+\log\Big(n^{\frac{2\alpha\kappa+2\kappa+d}{2\kappa(2\alpha+d)}}\log^{\rho(2/\kappa+2/d-1)-d-1}(n)\Big)\bigg)^{1+d}$$

which holds for all n such that

$$\log^{2\rho}(n) > \frac{16K_5D_1L_8^d}{L_9^2}\left(\frac{\log(\sqrt{L_{10}}L_8)+\log\big(n^{\frac{2\alpha\kappa+2\kappa+d}{2\kappa(2\alpha+d)}}\log^{\rho(2/\kappa+2/d-1)-d-1}(n)\big)}{\log(n)}\right)^{1+d}$$

H.2.8 Verifying (52)

We may rewrite (52) as

$$c_0L_{10}^{\rho}n^{\frac{d\rho}{\kappa(2\alpha+d)}}\log^{4\rho^2/\kappa}(n) > c_5 n^{\frac{d}{2\alpha+d}}\log^{2\rho}(n)$$

If $\rho/\kappa > 1$ this holds for all $L_{10} > (c_5/c_0)^{1/\rho}$.

H.2.9 Verifying (53)

We may rewrite (53) as

$$D_1L_8^d n^{\frac{d}{2\alpha+d}}\log^{2\rho}(n) > 2c_5 n^{\frac{d}{2\alpha+d}}\log^{\frac{2+2d}{2+d/\alpha}}(n)$$

which is satisfied for all $L_8 > (2c_5/D_1)^{1/d}$.

H.2.10 Verifying (54)

Consider

$$\zeta_n^{q_1-d+1} = L_8^{q_1-d+1}n^{\frac{q_1-d+1}{2\alpha+d}}\log^{\frac{2\rho(q_1-d+1)}{d}}(n)$$

and

$$\exp(c_5 n\delta_n^2) = \exp\big(c_5 n^{\frac{d}{2\alpha+d}}\log^{2\rho}(n)\big).$$

Then (54) holds for all n such that

$$n^{\frac{d}{2\alpha+d}} > \frac{1}{c_5}\big(\log(L_8^{q_1-d+1})+1\big).$$

H.2.11 Verifying (55)

We may rewrite (55) as

$$L_9^2 n^{\frac{d}{2\alpha+d}} \log^{d+1+4\rho}(n) \geq 8c_5 n^{\frac{d}{2\alpha+d}} \log^{2\rho}(n).$$

which is satisfied for all $L_9 \geq \sqrt{8c_5}$.

Acknowledgements

The authors thank Marco Battison for helpful discussions during the preparation of this paper.

References

Adams, R. P., Murray, I., and MacKay, D. J. (2009). Tractable nonparametric Bayesian inference in poisson processes with gaussian process intensities. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 9–16. ACM.

Belitser, E., Serra, P., and van Zanten, H. (2015). Rate-optimal Bayesian intensity smoothing for inhomogeneous poisson processes. *Journal of statistical planning and inference*, 166:24–35.

Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society, B*, 17(2):129–157.

Ghosal, S., Ghosh, J. K., and Van Der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Annals of Statistics*, 28(2):500–531.

Ghosal, S. and Van Der Vaart, A. (2007). Convergence rates of posterior distributions for non-i.i.d. observations. *Annals of Statistics*, 35(1):192–223.

Ghosal, S. and Van Der Vaart, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Annals of Statistics*, 29(5):1233–1263.

Grant, J. A., Boukouvalas, A., Griffiths, R., Leslie, D., Vakili, S., and Munoz de Coté, E. (2019). Adaptive sensor placement for continuous spaces. In *Proceedings of the 36th Annual International Conference on Machine Learning*. ACM.

Grant, J. A., Leslie, D. S., Glazebrook, K., Szechtman, R., and Letchford, A. (2018). Adapative policies for perimeter surveillence problems. *arXiv preprint arXiv:1810.02176*.

Gugushvili, S., van der Meulen, F., Schauer, M., and Spreij, P. (2018). Fast and scalable non-parametric Bayesian inference for Poisson point processes. *arXiv preprint arXiv:1804.03616*.

John, S. and Hensman, J. (2018). Large-scale Cox process inference using variational Fourier features. *International Conference on Machine Learning*.

Karr, A. F. (1986). Inference for stationary random fields given Poisson samples. *Advances in Applied Probability*, 18(2):406–422.

Kirichenko, A. and Van Zanten, J. H. (2015). Optimality of Poisson processes intensity learning with Gaussian processes. *The Journal of Machine Learning Research*, 16(1):2909–2919.

Lloyd, C., Gunter, T., Osborne, M., and Roberts, S. (2015). Variational inference for Gaussian process modulated poisson processes. In *International Conference on Machine Learning*, pages 1814–1822.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25(3):451–482.

Møller, J. and Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC.

Rathbun, S. L. and Cressie, N. (1994). Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson point processes. *Advances in Applied Probability*, 26(1):122–154.

van der Vaart, A. W. and van Zanten, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Annals of Statistics*, 36(3):1435–1463.

van der Vaart, A. W. and van Zanten, J. H. (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *Annals of Statistics*, 37(5B):2655–2675.

Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian processes for machine learning*. Number 3. MIT Press, Cambridge, MA.