Concentration inequalities for Poisson point processes with application to adaptive intensity estimation

M. Kroll

Institut für Mathematik, Universität Mannheim, Mannheim, Germany; Institut für Mathematik, Ruhr-Universität Bochum, Bochum, Germany

ABSTRACT
We derive concentration inequalities for maxima of empirical processes associated with Poisson point processes. The proofs are based on a careful application of Ledoux’s entropy method. We demonstrate the utility of the obtained concentration inequalities by application to adaptive intensity estimation.

ARTICLE HISTORY
Received 15 November 2020
Accepted 24 January 2022

KEYWORDS
Poisson point process; concentration inequality; entropy method; adaptive estimation

1. Introduction

Poisson point processes (PPPs) are of fundamental importance in probability theory and statistics, both from a theoretical and an applied point of view. For instance, they serve as the elementary building blocks for complex point process models which are used in stochastic geometry [14] and a wide range of applications including, amongst others, extreme value theory [32], finance [5], forestry [30], and queueing theory [12].

This paper is divided into two parts. In the first one, inspired by results in [22], we derive concentration inequalities for maxima of empirical processes associated with a PPP. In the second part, we demonstrate the potential applicability of our results to statistics: assuming that we observe i.i.d. PPPs with absolutely continuous intensity measure $\Lambda$, we consider the non-parametric estimation of the corresponding intensity function. We derive optimal rates of convergence in terms of the sample size $n$ and propose a fully data-driven estimator of the intensity. The theoretical study of the estimator is essentially based on the concentration inequalities derived in the first part.

Concentration inequalities provide upper bounds on the probability that a random variable deviates from its mean or median by a certain amount: the well-known inequalities by Markov, Hoeffding and Bernstein provide classical examples of such inequalities. Modern research on concentration inequalities goes back at least to the 1970s when concentration inequalities for deviations from the mean and median for Lipschitz continuous functions of multivariate Gaussian random variables were derived: the papers [9,38] consider deviations from the median, whereas [15] deals with deviations from the mean. The main advantage
of these results is that they do not depend on the dimension of the underlying Euclidean space and allow to control suprema of Gaussian processes.

In a series of papers, Talagrand [39,40] developed concentration inequalities for suprema of empirical processes associated with random variables on general product spaces. His original proof of the Talagrand inequality (see Theorem 1 in [26], for instance) is rather technical and essentially based on geometric arguments. Ledoux [25] proposed the entropy method as an essentially different and more accessible approach to prove Talagrand’s results. However, the paper [25] regained Talagrand’s result with different numerical constants and variance factors only. Massart’s [26] approach is also based on Ledoux’s method and a careful adaption of Gross’s logarithmic Sobolev inequality to a non-Gaussian framework and led to reasonable numerical constants in Talagrand type inequalities when the underlying random variables are independent but eventually non-identically distributed. The paper [11] proves Talagrand type inequalities for sub-additive functions using the entropy method. Finally, in the same setup as in [26], Klein and Rio [22] obtained concentration results by an elaborate application of the entropy method. In addition, they provided a brief comparison of the numerical constants occurring in different papers on the subject showing that their results are optimal in some sense. The main motivation of the first part of this paper is to transfer the results from [22] to the setup with PPPs.

There is already some amount of research papers dealing with concentration inequalities for point processes, see for instance [8,18,19,42]. As pointed out in [33] (cf. p. 109 therein), the main drawback of all these results is that they provide a variance term that is difficult to deal with in statistical applications. Recent concentration results have also been motivated by applications in stochastic geometry and [1,2].

The scope of the paper [33] is similar to the one of the present one: it is inspired by Talagrand type results from [26] and derives analogous results in a model with PPPs with exactly the same numerical constants as in the non-PPP setup. Similarly, the main intention of the present paper is to transfer the concentration inequalities from [22] to a PPP framework (again, by keeping exactly the same numerical constants as in [22]; see Remark 2.2 for a direct comparison of our result with the one of [33]).

Let us emphasize that it is not possible to apply the results from [22] immediately, that is, interpreting point processes as random variables in the space of locally finite measures, in order to obtain our results. The stochastic integral of a constant function with respect to a PPP provides an example that can be dealt with by our result but does not fit into the framework of [22] (see Remark 2.1). Furthermore, we do currently not see how one could derive the results of our paper via a simple Poissonization argument (however, also in the work [33], the proof of the concentration inequalities was not based on such an argument). To obtain our results, one key argument of the proof is borrowed from [33]: the infinite divisibility of PPPs is exploited and the underlying probability space split into two parts: on the first one, the proofs from [22] can be mimicked to a great extent (making a careful adaptation of the many auxiliary results in [22] necessary), whereas the probability of the second one can be neglected asymptotically. Based on our main results, we derive Proposition C.1 in the appendix as an auxiliary concentration result which represents a key tool for our statistical application: non-parametric adaptive intensity estimation.

Intensity estimation in parametric and non-parametric models has been dealt with in a wide range of monographs and research papers. For a general treatment of the subject, we
refer to [21] as a general introduction to the statistics of point processes, [24] for examples of intensity estimation in different parametric and non-parametric models, and [28] for estimation in general spatial models. Early approaches to non-parametric intensity estimation include kernel [24,35] and histogram estimators [35]. In addition, the paper [35] already discusses the adaptive estimation of the intensity. Baraud and Birgé [3] consider a Hellinger type loss function and propose a histogram estimator for intensity estimation. Other contributions focus on non-linear wavelet thresholding techniques, cf., for instance, the articles [23,34,36,41] and [6]. The paper [6] proposes a non-linear hard thresholding estimator for intensity estimation from indirect observations. Moreover, there exist other approaches to non-parametric intensity estimation in more specific models. Let us mention the paper [17] that proposes a minimum complexity estimator in the Aalen model and [29] that uses a wavelet approach to estimation in a multiplicative intensity model, without making a claim to be exhaustive.

In [33], the focus is on a model slightly different from the one we will consider: estimation of the intensity on the interval \([0, T]\) from only one observation and asymptotics for \(T \to \infty\) are considered whereas we assume the availability of an independent sample of realizations of the point process on the fixed interval \([0, 1]\). Common ground of our approach and the one taken in [33] is the use of projection estimators. Beyond that, the statistical methodology of the present paper is rather motivated by the procedure developed in [20] for circular deconvolution. We derive a minimax lower bound under abstract smoothness conditions and propose a projection estimator that can attain this lower bound. We note that the proof of the minimax lower bound given in [33] does not hold for ellipsoids defined in terms of the trigonometric basis (cf. the remark below Definition 4 and Proposition 3 in [33]), whereas we derive such a minimax lower bound. In return, the analysis in [33] is rather general with an emphasis on wavelet methods and not tailored to the trigonometric basis which we exclusively consider in this paper.

As usual in non-parametric statistics, the performance of our proposed projection estimator crucially depends on the appropriate selection of a dimension parameter. Since the optimal choice of this parameter depends on the unknown intensity and is thus unavailable in practise, we propose a fully data-driven selection of the dimension parameter leading to an adaptive estimator of the intensity. Our approach is based on model selection by minimization of a penalized contrast function. The combination of model selection techniques and concentration inequalities has attracted great attention in non-parametric statistics [4,16]. In particular, a considerable amount of research has been devoted to Gaussian regression and density estimation frameworks. We refer to the monograph [27] for results and further references concerning these two frameworks. Concentration inequalities have already been exploited in the context of non-parametric intensity estimation in the papers [3,33] mentioned above but their approaches are different from our one. Taking a different point of view, the paper [7] introduces an approach to model selection via hypothesis testing in the framework of intensity estimation. Following the model selection approach as in [4], we obtain an adaptive estimator that attains the optimal rate of convergence for intensities belonging either to some Sobolev space or some space of (generalized) analytic functions.

The paper is organized as follows. In Section 2, we derive concentration inequalities for deviations from the mean for empirical processes associated with PPPs. Section 3 deals with non-parametric intensity estimation: Subsection 3.1 introduces the model, in Subsection
3.2 we derive the minimax theory, and in Subsection 3.3, we study the adaptive estimator. Most of the proofs are deferred to the appendix.

2. Concentration inequalities for Poisson point processes

Let $N$ be a PPP with finite intensity measure $\Lambda$ on some Polish space $\mathbb{X}$. We denote the underlying probability space with $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{S}$ be a countable class of measurable functions from $\mathbb{X}$ to $[-1, 1]$. For $s \in \mathcal{S}$, let us define

$$I(s) = \int_{\mathbb{X}} s(x) (dN(x) - d\Lambda(x)). \quad (1)$$

The first aim of this paper is to establish concentration inequalities for $Z = \sup_{s \in \mathcal{S}} I(s)$. In the spirit of [22], concentration inequalities for the right-hand side and left-hand side deviations are established separately. As already sketched in the introduction, one key idea for the proofs of both theorems is borrowed from the article [33], namely to exploit the infinite divisibility of PPPs. This property is used to break the proof for the point process case into handy pieces such that the proof from [22] can be mimicked. However, adapting the variety of auxiliary results used in the proofs of [22] to our setup is non-trivial. The analogues of these auxiliary results are collected in Subsections A.1 and B.1 in the appendix.

Since the infinite divisibility is the essential ingredient for the proofs, it is not clear whether and if yes, how, concentration results as given in Theorems 2.1 and 2.3 can be transferred to point processes which are not infinitely divisible. For Cox processes (which represent a natural generalization of PPPs), the proof presented here fails and the development of tools to deal with this case might be worth further investigation. On the contrary, it should be possible to adapt the proof to frameworks with other infinitely divisible processes such as Poisson cluster processes (see [14], p. 151) under additional assumptions (for instance, that the number of daughter points is bounded).

The following theorem (the proof of which is given in Appendix A) provides concentration inequalities for right-hand side deviations of $Z$ from its mean.

**Theorem 2.1:** Let $N$ be a PPP on a Polish space $\mathbb{X}$ with finite intensity measure $\Lambda$, and $\mathcal{S}$ be a countable class of measurable functions from $\mathbb{X}$ to $[-1, 1]$. For $s \in \mathcal{S}$, define $I(s)$ as in (1) and consider $Z = \sup_{s \in \mathcal{S}} I(s)$. Let $L_Z(t) = \log \mathbb{E} \{ \exp(tZ) \}$ denote the logarithm of the moment-generating function of $Z$ and $V = \sup_{s \in \mathcal{S}} \text{Var}(I(s))$. Then, for any non-negative $t$,

(a) $$L_Z(t) \leq t \mathbb{E}Z + \frac{t}{2} (2 \mathbb{E}Z + V)(\exp((e^{2t} - 1)/2) - 1).$$

Setting $\nu = 2 \mathbb{E}Z + V$, we obtain that, for any non-negative $x$,

(b) $$\mathbb{P}(Z \geq \mathbb{E}Z + x) \leq \exp \left( -\frac{x}{4} \log(1 + 2 \log(1 + x/\nu)) \right),$$

and,
(c) for any $x \geq 0$,
\[
\mathbb{P}(Z \geq \mathbb{E}Z + x) \leq \exp\left(-\frac{x^2}{2\upsilon + 3x}\right).
\]

**Remark 2.1:** We emphasize that Theorem 2.1 cannot be immediately deduced from Theorem 1.1 in [22]. For instance, for $s \equiv 1$ the stochastic integral $\int_X s(x) dN(x)$ is an unbounded function of $N$ but obviously $s \equiv 1$ fits into the framework of Theorem 2.1.

**Remark 2.2:** Let us compare our result with the concentration inequality derived in [33]. More precisely, we consider Statement (c) from Theorem 2.1, and compare it with Corollary 2 from [33] which is closely related to our results. First note that Corollary 2 in [33] is formulated for the quantity $Z = \sup_{s \in \mathcal{S}} |I(s)|$ instead of $Z = \sup_{s \in \mathcal{S}} I(s)$ (however, in our application we will consider a symmetric set $\mathcal{S}$ and the two definitions coincide). Then, Corollary 2 from [33], reads as follows:
\[
\mathbb{P}(Z \geq (1 + \epsilon)\mathbb{E}Z + x) \leq \exp\left(-\frac{x^2}{12\upsilon_0 + 2\kappa(\epsilon)x}\right)
\]
where $\epsilon > 0$ is arbitrary but fixed, $\kappa(\epsilon) = 5/4 + 32/\epsilon$, and
\[
\upsilon_0 = \sup_{s \in \mathcal{S}} \int s^2(x) d\Lambda(x).
\]
Note that our concentration result in Statement (c) of Theorem 2.1 holds for $\mathbb{P}(Z \geq \mathbb{E}Z + x)$ instead of $\mathbb{P}(Z \geq (1 + \epsilon)\mathbb{E}Z + x)$ for $\epsilon > 0$.

**Remark 2.3:** In analogy to Corollary 1.1 in [22], the assumptions of Theorem 2.1 imply that $\text{Var} Z \leq V + 2\mathbb{E}Z$.

The quantity $V$ is usually referred to as the *wimpy variance* (cf. [10], Chapter 11). In the proof of Proposition C.1 in Appendix C, a suitable bound for the quantity $\upsilon$ will be determined.

For the statistical applications we have in mind, we state in the following an immediate corollary of Theorem 2.1. For the formulation of the corollary, we consider $n$ independent PPPs on the Polish space $\mathbb{X}$ with finite intensity measures $\Lambda_1, \ldots, \Lambda_n$. Again we denote the common underlying probability space of the PPPs with $(\Omega, \mathcal{A}, \mathbb{P})$. Let further $\mathcal{S}$ be a countable class of measurable functions from $\mathbb{X}$ to $[-1, 1]^n$. For $s = (s^1, \ldots, s^n) \in \mathcal{S}$ and $k \in \{1, \ldots, n\}$, let us define
\[
I^k(s) = \int_{\mathbb{X}} s^k(x)(dN_k(x) - d\Lambda_k(x)) \quad \text{and} \quad S_n(s) = I^1(s) + \ldots + I^n(s). \quad (2)
\]

**Corollary 2.2:** Let $N_1, \ldots, N_n$ be independent PPPs on a Polish space $\mathbb{X}$ with finite intensity measures $\Lambda_1, \ldots, \Lambda_n$, and $\mathcal{S}$ be a countable class of measurable functions from $\mathbb{X}$ to $[-1, 1]^n$. For $s \in \mathcal{S}$, define $S_n(s)$ as in (2) and consider $Z = \sup_{s \in \mathcal{S}} S_n(s)$. Let $L_Z(t) = \log \mathbb{E}[\exp(tZ)]$
denote the logarithm of the moment-generating function of $Z$ and $V_n = \sup_{s \in S} \Var(S_n(s))$. Then, for any non-negative $t$,

(a) \[ L_Z(t) \leq t \EE Z + \frac{t}{2} (2 \EE Z + V_n) (\exp((e^{2t} - 1)/2) - 1). \]

Setting $\nu = 2 \EE Z + V_n$, we obtain that, for any non-negative $x$,

(b) \[ \mathbb{P}(Z \geq \EE Z + x) \leq \exp \left( -\frac{x}{4} \log(1 + 2 \log(1 + x/\nu)) \right), \]

and,

(c) for any $x \geq 0$,

\[ \mathbb{P}(Z \geq \EE Z + x) \leq \exp \left( -\frac{x^2}{\nu + \sqrt{\nu^2 + 3ux} + (3x/2)} \right) \leq \exp \left( -\frac{x^2}{2\nu + 3x} \right). \]

The following theorem (the proof of which is given in Appendix B) provides concentration inequalities for left-hand side deviations of $Z$ from its mean.

**Theorem 2.3:** Under the assumptions of Theorem 2.1, for any non-negative $t$, it holds

(a) \[ L_Z(-t) \leq -t \EE Z + \frac{\nu}{9} (e^{3t} - 3t - 1). \]

Consequently, for any non-negative $x$,

(b) \[ \mathbb{P}(Z \leq \EE Z - x) \leq \exp \left( -\frac{\nu}{9} h \left( \frac{3x}{\nu} \right) \right), \]

where $h(x) = (1 + x) \log(1 + x) - x$, and

(c) for any $x \geq 0$,

\[ \mathbb{P}(Z \leq \EE Z - x) \leq \exp \left( -\frac{x^2}{\nu + \sqrt{\nu^2 + 2ux} + x} \right) \leq \exp \left( -\frac{x^2}{2\nu + 2x} \right). \]
Remark 2.4: The concentration inequalities in Theorems 2.1 and 2.3 translate literally (that is, with exact coincidence of the numerical constants) the ones obtained in [22] to our framework with PPPs. This observation is in line with the remark made in [33] where the derived concentration inequalities translate literally previous results due to [26].

Remark 2.5: In many situations of interest, it is possible to apply the concentration inequalities proved in this section (and the one proved in Appendix C) to non-countable classes of measurable functions. A rigorous foundation for this practice can be based on density arguments (see Remarque 2.1 in [13]).

3. Non-parametric intensity estimation

3.1. Model assumptions

Let \( N_1, \ldots, N_n \) be i.i.d. realizations of a PPP on \([0,1]\) with square-integrable intensity function \( \lambda \in L^2(\mathbb{R}, dx) \). We denote \( \mathbb{N}_0 = \{0,1,2,\ldots\} \). Recall that the intensity function is the density of the intensity measure \( \lambda \) with respect to the standard Lebesgue measure. The \( N_i \) can be interpreted as \( \mathbb{N}_0 \)-valued random measures which motivates the notation \( N_i = \sum_j \delta_{x_i} \) (here, \( \delta_x \) denotes the Dirac measure with mass concentrated at \( x \)).

Our aim is to estimate the intensity function \( \lambda \) from the sample \( N_1, \ldots, N_n \). We consider the orthonormal basis \( \{\varphi_j\}_{j \in \mathbb{Z}} \) of \( L^2 \) which is given by

\[
\varphi_0(t) = 1, \quad \text{resp. } \varphi_{-j}(t) = \sqrt{2} \sin(2\pi jt)
\]

for \( j = 1, 2, \ldots \). Define the sequence \( \{\beta_j\}_{j \in \mathbb{Z}} \) of Fourier coefficients via

\[
\beta_j = \int_0^1 \lambda(t) \varphi_j(t) dt
\]

which yields the \( L^2 \)-convergent representation

\[
\lambda = \sum_{j \in \mathbb{Z}} \beta_j \varphi_j. \tag{3}
\]

In order to evaluate the performance of an arbitrary estimator \( \hat{\lambda} \) of \( \lambda \), we consider the mean integrated squared error \( \mathbb{E}[||\hat{\lambda} - \lambda||^2] \) (where, as usual, the expectation is taken under the true intensity function \( \lambda \) and \( || \cdot || \) denotes the \( L^2 \)-norm). We hold the minimax point of view and consider the maximum risk defined by \( \sup_{\lambda \in \Lambda} \mathbb{E}[||\hat{\lambda} - \lambda||^2] \) for some smoothness class \( \Lambda \) of potential intensity functions. The corresponding minimax risk is defined by

\[
\inf_{\lambda} \sup_{\lambda \in \Lambda} \mathbb{E}[||\hat{\lambda} - \lambda||^2]
\]

where the infimum is taken over all potential estimators \( \hat{\lambda} \) of \( \lambda \) based on the sample \( N_1, \ldots, N_n \). An estimator \( \lambda^* \) is called rate optimal if \( \sup_{\lambda \in \Lambda} \mathbb{E}[||\lambda^* - \lambda||^2] \lesssim \inf_{\lambda} \sup_{\lambda \in \Lambda} \mathbb{E}[||\hat{\lambda} - \lambda||^2] \), where the notation \( a_n \lesssim b_n \) means that \( a_n \leq C b_n \) for some numerical constant that does not depend on \( n \). The specific form of the class \( \Lambda \) will be introduced in the following Subsection 3.2.

For the moment, let us introduce the general type of projection estimator we will consider throughout this work: Since \( \mathbb{E}[\int_0^1 \varphi_j(t) dN_i(t)] = \beta_j \) for all \( j \in \mathbb{Z} \) by Campbell’s theorem (cf., for instance, [37], Chapter 2), \( \hat{\beta}_j = \frac{1}{n} \sum_{i=1}^n \int_0^1 \varphi_j(t) dN_i(t) \) is an unbiased estimator of \( \beta_j \), Equation (3) strongly suggests to consider orthogonal series estimators of the
form

\[ \hat{\lambda}_k = \sum_{0 \leq |j| \leq k} \hat{\beta}_j \phi_j, \quad (4) \]

where the dimension parameter \( k \in \mathbb{N}_0 \) has to be chosen appropriately.

**Remark 3.1:** The estimator \( \hat{\lambda}_k \) is by definition not guaranteed to attain only non-negative values (which holds for the true intensity \( \lambda \)). In practice, this undesirable feature can be avoided by considering the estimator \( \hat{\lambda}_{k+} \) defined via \( \hat{\lambda}_{k+}(t) = \hat{\lambda}_k(t) \vee 0 \), the risk of which is evidently bounded from above by the one of \( \hat{\lambda}_k \).

### 3.2. Minimax theory

In order to define the class of admissible intensity functions in the definition of the minimax risk, let \( \gamma = (\gamma_j)_{j \in \mathbb{Z}} \) be a strictly positive symmetric sequence of weights and \( r > 0 \). Set

\[ \Lambda = \Lambda(\gamma, L) = \left\{ \lambda \in \mathbb{L}^2 : \lambda \geq 0 \text{ and } \sum_{j \in \mathbb{Z}} \gamma_j^2 \beta_j^2 =: \|\lambda\|_\gamma^2 \leq L^2 \right\}. \]

In the following, our aim is to study the minimax risk with respect to the function class \( \Lambda \). Our results will be obtained under the following mild regularity assumptions on the sequence \( \gamma \).

**Assumption 3.1:** \( \gamma = (\gamma_j)_{j \in \mathbb{Z}} \) is a strictly positive symmetric sequence such that \( \gamma_0 = 1 \) and \( (\gamma_n)_{n \in \mathbb{N}_0} \) is non-decreasing.

The following proposition provides an upper risk bound for the estimator \( \hat{\lambda}_k \) defined in (4) under an appropriate choice of the dimension parameter \( k \).

**Proposition 3.1:** Let Assumption 3.1 hold. Consider the estimator \( \hat{\lambda}_{k^*_n} \) with dimension parameter \( k^*_n = \arg\min_{k \in \mathbb{N}_0} \max\{\gamma_k^{-2}, \frac{2k+1}{n}\} \). Then, for any \( n \in \mathbb{N} \),

\[ \sup_{\lambda \in \Lambda} \mathbb{E}[\|\hat{\lambda}_{k^*_n} - \lambda\|_2^2] \lesssim \Psi_n := \max\left\{ \gamma_k^{-2}, \frac{2k^*_n+1}{n} \right\} \]

where the constant hidden in \( \lesssim \) depends only on \( L \).

**Proof:** Introduce the function \( \lambda_{k^*_n} := \sum_{0 \leq |j| \leq k^*_n} \beta_j \phi_j \) which suggests the decomposition

\[ \mathbb{E}[\|\hat{\lambda}_{k^*_n} - \lambda\|_2^2] = \|\lambda - \lambda_{k^*_n}\|_2^2 + \mathbb{E}[\|\hat{\lambda}_{k^*_n} - \lambda_{k^*_n}\|_2^2] \]

of the considered risk into squared bias and variance. Using the smoothness assumption \( \lambda \in \Lambda \), it is easy to see that \( \|\lambda - \lambda_{k^*_n}\|_2^2 \leq L^2 \gamma_k^{-2} \) and \( \mathbb{E}[\|\hat{\lambda}_{k^*_n} - \lambda_{k^*_n}\|_2^2] \leq L \cdot \frac{2k^*_n+1}{n} \) and the statement of the theorem follows. \( \blacksquare \)

The rate-optimality of the estimator \( \hat{\lambda}_{k^*_n} \) considered in Proposition 3.1 is demonstrated by means of the following theorem which is valid under mild additional assumptions. The
proof makes use of an adaptation of standard techniques in non-parametric statistics for the derivation of minimax lower bounds to our point process framework and is deferred to Appendix D.

**Theorem 3.2:** Let Assumption 3.1 hold and further assume that

(C1) \( \sum_{j \in \mathbb{Z}} y_j^{-2} < \infty \), and
(C2) \( 0 < \eta^{-1} = \inf_{n \in \mathbb{N}} (\Psi_n)^{-1} \min\{y_{k_n^*}^{-2}, \frac{2k_n^*+1}{n}\} \) for some \( \eta \geq 1 \),

where the quantities \( k_n^* \) and \( \Psi_n \) are defined in Proposition 3.1. Then, for any \( n \in \mathbb{N} \),

\[
\inf_{\hat{\lambda}} \sup_{\lambda \in \Lambda} \mathbb{E}[\|\hat{\lambda} - \lambda\|^2] \geq \Psi_n
\]

where the infimum is taken over all estimators of \( \hat{\lambda} \) of \( \lambda \) based on the sample \( N_1, \ldots, N_n \) and the constant hidden in \( \geq \) depends only on \( \eta, L \) and \( \Gamma \).

**Remark 3.2:** The mild assumption (C1) on the convergence of the series \( \sum_{j \in \mathbb{Z}} y_j^{-2} \) is needed only in order to guarantee the non-negativity of the candidate intensities considered in the proof. On the whole, the proof is very much in line with the proof of Theorem 2.1 in [20] expanded with the essential ingredient that the Hellinger distance between two PPPs is bounded by the Hellinger distance of the corresponding intensity measures (see Theorem 3.2.1 in [31]).

**Remark 3.3:** Note that the lower bound proof given in [33] is based on a specific property called localization and is not valid for ellipsoids expressed in terms of the trigonometric basis.

**Example 3.3 (Sobolev ellipsoids):** Let \( \gamma_0 = 1, \gamma_j = |j|^p \) for \( j \neq 0 \). This setting corresponds to \( \lambda \) belonging to a Sobolev ellipsoid. Then, Assumption 3.1 is satisfied and elementary computations show that \( k_n^* \asymp n^{1/(2p+1)} \) as well as \( \Psi_n \asymp n^{-2p/(2p+1)} \). Furthermore, the additional conditions of Theorem 3.2 are satisfied if \( p > 1 \) holds.

**Example 3.4 (Analytic functions):** Let \( \gamma_j = \exp(\rho |j|) \) for \( j \in \mathbb{Z} \) for some \( \rho > 0 \). This setting corresponds to \( \lambda \) belonging to a class of analytic functions. Assumption 3.1 is also fulfilled in this case and we obtain \( k_n^* \asymp \log n \) and \( \Psi_n \asymp \log n/n \). The additional assumption of Theorem 3.2 does not impose any additional restriction on \( \rho \).

**Example 3.5 (Generalized analytic functions):** Let \( \gamma_j = \exp(2\rho |j|^p) \) for \( \rho, p > 0 \). Note that in this case the Fourier coefficients of \( \hat{\lambda} \) obey a power exponential decay and \( \lambda \) belongs to a class of generalized analytic functions. Assumption 3.1 is satisfied in this case and there are no additional restrictions on \( p \) (and \( \rho \)) due to Theorem 3.2. We have \( k_n^* \asymp (\log n)^{1/p} \) resulting in the rate \( \Psi_n \asymp (\log n)^{1/p}/n \).

### 3.3. Adaptive estimation

The optimal choice of the dimension parameter \( k \) stated in Proposition 3.1 depends on the smoothness characteristics of the intensity via the sequence \( \gamma \). However, such an a priori
knowledge is a strong assumption and is not available in practice. Thus, there is a demand for a data-driven choice of the dimension parameter which hopefully does not deteriorate the quality of the upper risk bound or at least only leads to worse numerical constants, merely.

This data-driven choice of the dimension parameter and the resulting upper risk bound are investigated now. For this purpose, we follow a model selection approach which has been successfully applied to a wide range of estimation problems in non-parametric statistics (cf., for instance, [4,16] for general accounts to this model selection paradigm).

For \( s, t \in \mathbb{L}^2 \), introduce the notation \( \langle s, t \rangle = \int_0^1 s(x)t(x)\,dx \) and consider the contrast function

\[
\Upsilon_n(t) = \|t\|^2 - 2 \langle \hat{\lambda}_n, t \rangle, \quad t \in \mathbb{L}^2.
\]

Define the random sequence of penalties \((\text{pen}_k)_{k \in \mathbb{N}}\) via

\[
\text{pen}_k = 24 \cdot (\beta_0 \lor 1) \cdot \frac{2k + 1}{n}.
\]

Building on the definitions made until now, we define the data-driven selection \( \hat{k}_n \) of the dimension parameter as the minimizer of the penalized contrast

\[
\hat{k}_n := \arg\min_{0 \leq k \leq n} \{ \Upsilon_n(\hat{\lambda}_k) + \text{pen}_k \}.
\]

The following theorem provides an upper bound for the risk of the estimator \( \hat{\lambda}_{\hat{k}_n} \). Its proof is given in Appendix E.

**Theorem 3.6:** Let Assumption 3.1 hold. Then, for any \( n \in \mathbb{N} \), we have

\[
\sup_{\lambda \in \Lambda} \mathbb{E} [\| \hat{\lambda}_{\hat{k}_n} - \lambda \|^2] \lesssim \min_{0 \leq k \leq n} \max \left\{ \gamma_k^{-2}, \frac{2k + 1}{n} \right\} + \frac{1}{n} + \exp(-\kappa \sqrt{n})
\]

where \( \kappa > 0 \) is a numerical constant and the constant hidden in \( \lesssim \) depends only on \( L \).

**Remark 3.4:** The penalty term used in the definition of \( \hat{k}_n \) is random which is in contrast to penalty terms occurring, for instance, in density estimation or deconvolution problems. The need for randomization is due to the quantity \( \beta_0 \) in the definition of \( H \) in Lemma E.1.

If \( L \) (but not \( \gamma \)) was known, one could proceed without randomization by choosing the penalty proportional to \( \sqrt{L} (2k + 1)/n \). However, the factor \( L \) in this definition cannot be replaced by an estimate of \( L \) because a reasonable estimator of \( L \) is not reachable from the data. Note that the penalty terms considered in [33] in a point process framework similar to ours are also non-deterministic.

The adaptive estimator \( \hat{\lambda}_{\hat{k}_n} \) attains the rate \( \Psi_n \) if and only if

\[
\Psi_n \asymp \min_{0 \leq k \leq n} \max \left\{ \gamma_k^{-2}, \frac{2k + 1}{n} \right\}.
\]

Since under Assumption 3.1 it holds that \( k^*_n \lesssim n \), we immediately obtain the following result.
**Corollary 3.7:** Under Assumption 3.1, the estimator \( \hat{\lambda}_{k_n^*} \) is rate optimal over the class \( \Lambda \).

In particular, the estimator \( \hat{\lambda}_{k_n^*} \) is rate optimal in the framework of Examples 3.3, 3.4, and 3.5 where \( k_n^* \asymp n^{1/(2p+1)} \), \( k_n^* \asymp \log n \), and \( k_n^* \asymp (\log n)^{1/p} \), respectively.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

Financial support by the Deutsche Forschungsgemeinschaft (DFG) through the Research Training Group RTG 1953 is gratefully acknowledged. The final preparation of the manuscript has also been supported by the research grant DFG DE 502/27-1 of the German Research Foundation (DFG).

**References**

[1] S. Bachmann, *Concentration for Poisson functionals: Component counts in random geometric graphs*, Stoch. Process. Appl. 126 (2016), pp. 1306–1330. Available at [https://doi.org/10.1016/j.spa.2015.11.004]. MR 3473096

[2] S. Bachmann and G. Peccati, *Concentration bounds for geometric Poisson functionals: Logarithmic Sobolev inequalities revisited*, Electron. J. Probab. 21 (2016), pp. 1–44. Paper No. 6. Available at [https://doi.org/10.1214/16-EJP4235]. MR 3485348

[3] Y. Baraud and L. Birgé, *Estimating the intensity of a random measure by histogram type estimators*, Probab. Theory Related Fields 143 (2009), pp. 239–284. Available at [https://doi.org/10.1007/s00440-007-0126-6]. MR 2449129

[4] A. Barron, L. Birgé, and P. Massart, *Risk bounds for model selection via penalization*, Probab. Theory Related Fields 113 (1999), pp. 301–413. Available at [https://doi.org/10.1007/s004400050210]. MR 1679028

[5] L. Bauwens and N. Hautsch, *Modelling financial high frequency data using point processes*, in *Handbook of Financial Time Series*, T. Mikosch, J.P. Kreiß, R.A. Davis, and T.G. Andersen, eds., Springer, 2009, pp. 953–979.

[6] J. Bigot, S. Gadat, T. Klein, and C. Marteau, *Intensity estimation of non-homogeneous Poisson processes from shifted trajectories*, Electron. J. Stat. 7 (2013), pp. 881–931. Available at [https://doi.org/10.1214/13-EJS794]. MR 3044503

[7] L. Birgé, *Model selection for Poisson processes*, in *Asymptotics: Particles, Processes and Inverse Problems*, IMS Lecture Notes Monogr. Ser. Vol. 55, Inst. Math. Statist., Beachwood, OH, 2007, pp. 32–64. Available at [https://doi.org/10.1214/074921707000000265]. MR 2459930.

[8] S.G. Bobkov and M. Ledoux, *On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures*, J. Funct. Anal. 156 (1998), pp. 347–365. Available at [https://doi.org/10.1006/jfan.1997.3187]. MR 1636948

[9] C. Borell, *The Brunn-Minkowski inequality in Gauss space*, Invent. Math. 30 (1975), pp. 207–216. Available at [https://doi.org/10.1007/BF01425510]. MR 0399402

[10] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, 2016.

[11] O. Bousquet, *Concentration inequalities for sub-additive functions using the entropy method*, in *Stochastic Inequalities and Applications*, Progr. Probab. Vol. 56, Birkhäuser, Basel, 2003, pp. 213–247. MR 2073435.

[12] P. Brémaud, *Point Processes and Queues, Martingale Dynamics*, Springer, 1981.

[13] G. Chagny, *Estimation adaptative avec des données transformées ou incomplètes. Application à des modèles de survie*, Ph.D. diss., 2013. Available at [https://tel.archives-ouvertes.fr/tel-00863141].
[14] S.N. Chiu, D. Stoyan, W.S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*, 3rd ed., Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester, 2013. Available at https://doi.org/10.1002/9781118658222. MR 3236788

[15] B.S. Cirelson, I.A. Ibragimov, and V.N. Sudakov, *Norms of Gaussian Sample Functions*. Lecture Notes in Math., Vol. 550, 1976, pp. 20–41. MR 0458556.

[16] E. Comte, *Estimation Non-Paramétrique*, Spartacus, 2015.

[17] G. Grégoire and J. Nembé, *Convergence rates for the minimum complexity estimator of counting process intensities*, J. Nonparamet. Statist. 12 (2000), pp. 611–643. Available at https://doi.org/10.1080/10485250008832825. MR 1784801

[18] C. Houdré, *Remarks on deviation inequalities for functions of infinitely divisible random vectors*, Ann. Probab. 30 (2002), pp. 1223–1237. Available at https://doi.org/10.1214/aop/1029867126. MR 1920106

[19] C. Houdré and N. Privault, *Concentration and deviation inequalities in infinite dimensions via covariance representations*, Bernoulli 8 (2002), pp. 697–720. MR 1962538

[20] J. Johannes and M. Schwarz, *Adaptive circular deconvolution by model selection under unknown error distribution*, Bernoulli 19 (2013), pp. 1576–1611. Available at https://doi.org/10.3150/12-BEJ422. MR 3129026

[21] A.F. Karr, *Point Processes and Their Statistical Inference*, 2nd ed., Probability: Pure and Applied Vol. 7, Marcel Dekker, Inc., New York, 1991. MR 1113698

[22] T. Klein and E. Rio, *Concentration around the mean for maxima of empirical processes*, Ann. Probab. 33 (2005), pp. 1060–1077. Available at https://doi.org/10.1214/009117905000000044. MR 2135312

[23] E.D. Kolaczyk, *Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds*, Statist. Sinica 9 (1999), pp. 119–135. MR 1678884

[24] Y.A. Kutoyants, *Statistical Inference for Spatial Poisson Processes*, Lecture Notes in Statistics Vol. 134, Springer-Verlag, New York, 1998. Available at https://doi.org/10.1007/978-1-4612-1706-0. MR 1644620

[25] M. Ledoux, *On Talagrand's deviation inequalities for product measures*, ESAIM Probab. Statist. 1 (1995/97), pp. 63–87. Available at https://doi.org/10.1051/ps:1997103. MR 1399224

[26] P. Massart, *About the constants in Talagrand's concentration inequalities for empirical processes*, Ann. Probab. 28 (2000), pp. 863–884. Available at https://doi.org/10.1214/aop/1019160263. MR 1782276

[27] P. Massart, *Concentration Inequalities and Model Selection*, Lecture Notes in Mathematics Vol. 1896, Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard. MR 2319879

[28] J. Møller and R.P. Waagepetersen, *Statistical Inference and Simulation for Spatial Point Processes*, Monograph on Statistics and Applied Probability Vol. 100, Chapman & Hall/CRC, Boca Raton, FL, 2004. MR 2004226

[29] P.N. Patil and A.T.A. Wood, *Counting process intensity estimation by orthogonal wavelet methods*, Bernoulli 10 (2004), pp. 1–24. Available at https://doi.org/10.3150/bj/1077544601. MR 2044591

[30] A. Penttinen and D. Stoyan, *Recent applications of point process methods in forestry statistics*, Statist. Sci. 15 (2000), pp. 61–78. Available at https://doi.org/10.1214/ss/1009212674. MR 1842237

[31] R.D. Reiss, *A Course on Point Processes*, Springer Series in Statistics, Springer-Verlag, New York, 1993. Available at https://doi.org/10.1007/978-1-4613-9308-5. MR 1199815

[32] S.I. Resnick, *Extreme Values, Regular Variation, and Point Processes*, Applied Probability. A Series of the Applied Probability Trust Vol. 4, Springer-Verlag, New York, 1987. Available at https://doi.org/10.1007/978-0-387-75953-1. MR 900810

[33] P. Reynaud-Bouret, *Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities*, Probab. Theory Related Fields 126 (2003), pp. 103–153. Available at https://doi.org/10.1007/s00440-003-0259-1. MR 1981635
[34] P. Reynaud-Bouret and V. Rivoirard, *Near optimal thresholding estimation of a Poisson intensity on the real line*, Electron. J. Stat. 4 (2010), pp. 172–238. Available at https://doi.org/10.1214/08-EJS319. MR 2645482

[35] M. Rudemo, *Empirical choice of histograms and kernel density estimators*, Scand. J. Statist. 9 (1982), pp. 65–78. MR 668683

[36] L. Sansonnet, *Wavelet thresholding estimation in a Poissonian interactions model with application to genomic data*, Scand. J. Stat. 41 (2014), pp. 200–226. Available at https://doi.org/10.1111/sjos.12009. MR 3181140

[37] R.L. Streit, *Poisson Point Processes*, Springer, New York, 2010. Available at https://doi.org/10.1007/978-1-4419-6923-1, Imaging, tracking, and sensing. MR 3014399.

[38] V.N. Sudakov and B.S. Cirelson, *Extremal properties of half-spaces for spherically invariant measures*, Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) 41 (1974), pp. 14–24. 165. Problems in the theory of probability distributions, II. MR 0365680

[39] M. Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. 81 (1995), pp. 73–205. Available at http://www.numdam.org/item?id=PMIHES_1995__81__73_0. MR 1361756

[40] M. Talagrand, *New concentration inequalities in product spaces*, Invent. Math. 126 (1996), pp. 505–563. Available at https://doi.org/10.1007/s002220050108. MR 1419006

[41] R.M. Willett and R.D. Nowak, *Multiscale Poisson intensity and density estimation*, IEEE Trans. Inform. Theory 53 (2007), pp. 3171–3187. Available at https://doi.org/10.1109/TIT.2007.903139. MR 2417680

[42] L. Wu, *A new modified logarithmic Sobolev inequality for Poisson point processes and several applications*, Probab. Theory Related Fields 118 (2000), pp. 427–438. Available at https://doi.org/10.1007/PL00008749. MR 1800540