Multiscale Scanning in Higher Dimensions:

Limit theory, statistical consequences and an application in STED microscopy

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen

Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität zu Göttingen

im Promotionsprogramm

"PhD School of Mathematical Sciences (SMS)"

der Georg-August University School of Science (GAUSS)

vorgelegt von
Claudia Juliane König
aus Kiel

Göttingen, 2018

Betreuungsausschuss:

Dr. Frank Werner
Max Planck Institute for Biophysical Chemistry, Göttingen,
Institut für Mathematische Stochastik, Universität Göttingen

Prof. Dr. Axel Munk
Institut für Mathematische Stochastik, Universität Göttingen

Mitglieder der Prüfungskommission:

Referent:
Dr. Frank Werner
Max Planck Institute for Biophysical Chemistry, Göttingen
Institut für Mathematische Stochastik, Universität Göttingen

Korreferent:
Prof. Dr. Axel Munk
Institut für Mathematische Stochastik, Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Russell Luke
Institut für Numerische und Angewandte Mathematik, Universität Göttingen

Jun.-Prof. Dr. Daniel Rudolf
Institut für Mathematische Stochastik, Universität Göttingen

Prof. Dr. Dominic Schuhmacher
Institut für Mathematische Stochastik, Universität Göttingen

Prof. Dr. Ingo Witt
Mathematisches Institut, Universität Göttingen

Tag der mündlichen Prüfung: 26.06.2018

Acknowledgments

When years of work and effort converge into a single document like this dissertation it is time to reflect and be thankful to the many people that made it possible.
As I am going through the list below it becomes evident how many people have been involved in the process and helped, supported, guided & encouraged me: I am incredibly grateful for your support!

- I would like to express my sincere gratitude to my supervisor **Dr. Frank Werner**: His guidance and enthusiasm have been very encouraging throughout my PhD and have been tremendously helpful in developing this dissertation.

- Thank you to **Prof. Axel Munk** for providing the interesting and challenging topic of my dissertation and being my second supervisor. His statistical intuition was a great inspiration to me!

- Thank you **Prof. Russell Luke**, **Jun.-Prof. Daniel Rudolf**, **Prof. Dominic Schuhmacher**, and **Prof. Ingo Witt** for being part in my committee.

- I owe many thanks to **Asst. Prof. Johannes Schmidt-Hieber** and **Dr. Katharina Proksch** for the fruitful discussions about my work.

- I am thankful to **René Siegmund** from the Laser-Laboratorium Göttingen e.V. for providing the STED data.

- I gratefully acknowledge the financial support by the **German Research Foundation DFG** through subproject A04 of CRC 755.

- Special thanks go to my IMS colleagues for providing such a nice working environment. In particular, thank you, **Anne Hein**, **Carla Tameling**, **Dr. Katharina Proksch**, **Dr. Merle Behr**, **Dr. Manuel Diehn**, **Henning Höllwarth**, **Robin Richter**, **Laura Fee Schneider** and **Miguel Del Alamo**! Especially, I would like to thank **Dr. Merle Behr** and **Dr. Manuel Diehn** for creating a creative and enjoyable office atmosphere.

vi

Preface

Scan statistics have a broad area of applications ranging from astrophysics over genetic screening to fluorescence microscopy. Here, we consider a calibrated scan statistic based on local likelihood ratio tests of homogeneity against heterogeneity. The problem is to find anomalies in a d-dimensional field of independent random variables $\{Y_i\}_{i \in \{1,...,n\}^d}$, each distributed according to a one-dimensional natural exponential family $\mathcal{F} = \{F_\theta\}_{\theta \in \Theta}$. Further motivation and details on modelling assumptions are given in Chapter 1.

Chapter 2 is split into two parts: the setting of an known and unknown underlying baseline parameter. In a first step, in Section 2.1, given some baseline parameter $\theta_0 \in \Theta$, the field is scanned using local likelihood ratio tests to detect from a (large) given system of regions $\mathcal{R}$ those regions $R \subset \{1, ..., n\}^d$ with $\theta_i \neq \theta_0$ for some $i \in R$. We provide a unified methodology which controls the overall family wise error rate (FWER) to make a wrong detection at a given level. Fundamental to our method is a Gaussian approximation of the asymptotic distribution of the underlying multiscale scanning test statistic with explicit rate of convergence. From this, we obtain a weak limit theorem which can be seen as a generalized weak invariance principle to non-identically distributed data and is of independent interest. Furthermore, we give an asymptotic expansion of the procedure's power, which yields minimax optimality in case of Gaussian observations. In a second step, in Section 2.2, we consider the situation where the baseline parameter is unknown. This is motivated by the detection of hot spots in STED measurements with Poisson image data. The baseline parameter will be estimated by the sample mean and Section 2.2 gives the theoretical results for this setting (following the lines of Section 2.1).

The theoretical results are then applied in Chapter 3 in a simulation study and to real data of STED measurements of crimson beads.

A discussion and an outlook for further research can be found in Chapter 4. Appendix A gives more insight into the complexity of families of sets, which is an important tool for our proving techniques. The proofs are outsourced to Appendix B.

Large parts of Chapter 2.1 (and the corresponding proofs) are taken from the preprint

(König et al., 2018) with little modifications.

Contents

List of Symbols

$\mathbb{N}$	set of positive integers
$\mathbb{R}$	set of real numbers
$\mathbb{E}[X]$	expected value of a random variable X
$\overset{\mathbb{P}}{\to}$	convergence in probability
$\overset{\mathcal{D}}{\to}$	convergence in distribution
$\#\mathcal{R}$	cardinality of the set $\mathcal{R}$
$\lesssim, \gtrsim$	inequalities up to a (universal) constant
$a_n = o\,(b_n)$	$\limsup_{n\to\infty} \lvert a_n/b_n \rvert = 0$
$X_n = o_{\mathbb{P}}(Z_n)$	there exists a sequence of random vectors $(Y_n)_{n\in\mathbb{N}}$ such that $X_n = Y_n Z_n$ for all $n \in \mathbb{N}$ and $Y_n \overset{\mathbb{P}}{\to} 0$
$X_n = O_{\mathbb{P}}(Z_n)$	there exists a sequence of random vectors $(Y_n)_{n\in\mathbb{N}}$ such that $X_n = Y_n Z_n$ for all $n \in \mathbb{N}$, such that $\{Y_n\}_{n\in\mathbb{N}}$ is tight
$\mathrm{Poi}(\lambda)$	Poisson distribution with intensity $\lambda \in \mathbb{R}$
$\mathcal{N}(0,1)$	standard normal distribution
$\mathbb{I}_R$	indicator function of R with $\mathbb{I}_R(x) = 1$, if $x \in R$ and $\mathbb{I}_R(x) = 0$ otherwise

CHAPTER 1

Introduction

Suppose we observe an independent, d-dimensional field Y of random variables

$$Y_i \sim F_{\theta_i}, \qquad i \in I_n^d := \{1, ..., n\}^d, \qquad (1.1)$$

where each observation is drawn from the same given one-dimensional natural exponential family model $\mathcal{F} = \{F_\theta\}_{\theta \in \Theta}$, but with potentially different parameters θ_i. Our setting includes the important special cases of Gaussian with varying normal means μ_i (Arias-Castro et al., 2005; Sharpnack and Arias-Castro, 2016), Bernoulli (Walther, 2010), and Poisson (Zhang et al., 2016) random fields.

Given some baseline parameter $\theta_0 \in \Theta$ (e.g., all $\mu_i = 0$ for a Gaussian field), we consider the problem of finding anomalies (hot spots) in the field Y, i.e., we aim to identify those regions $R \subset I_n^d$ where $\theta_i \neq \theta_0$ for some $i \in R$. We will discuss both cases of known and unknown θ_0. Here R runs through a given family of candidate regions $R \in \mathcal{R}_n \subset 2^{I_n^d}$ where 2^X denotes the power set of some set X. For simplicity, we will suppress the subindex n whenever it is clear from the context, i.e., write $\mathcal{R} = \mathcal{R}_n$ in what follows. Figure 1.1 shows an example with Poisson observations, where the baseline intensity is $\theta_0 = 1$ and in the squares in the right plot the intensity is given by $\theta_i = 4$.
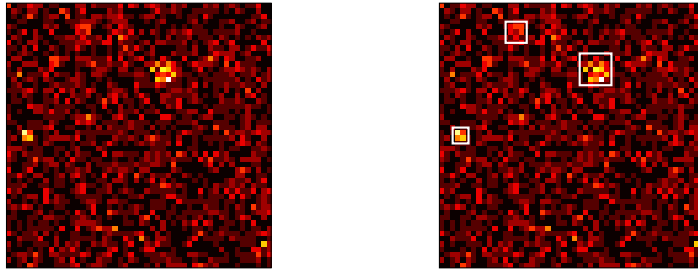

Figure 1.1: A simulated Poisson field $Y_i \sim \text{Poi}(\theta_i)$, $i \in \{1, \ldots, 64\}^2$, baseline intensity $\theta_0 = 1$ and the corresponding anomalies illustrated on the right ($\theta_i = 4$).

As in the above mentioned references (cf. Section 1.2), the problem of finding hot spots can be regarded as a multiple testing problem, thereby, many 'local' tests on

the regions $\mathcal{R}$ are performed simultaneously, while keeping the overall error of wrong detections controllable.

For a fixed region $R \in \mathcal{R}$ we consider the testing problem with hypothesis

$$\forall \, i \in R : \theta_i = \theta_0 \qquad\qquad (H_{R,n})$$

vs. alternative

$$\exists \, i \in R \text{ s.t. } \theta_i \neq \theta_0. \qquad\qquad (K_{R,n})$$

In Figure 1.2 the local testing problem is visualized. The Poisson field from Figure 1.1 is shown with a frame around a region R where the local test is applied.
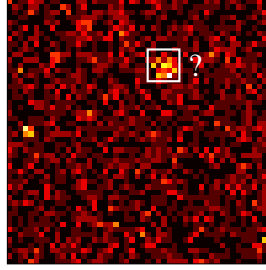
Figure 1.2: Poisson field from Figure 1.1 with one fixed region R (white squared box).

The likelihood ratio test (LRT) for this testing problem is a powerful test in general, and for a known θ_0 often known to have certain optimality properties, depending on the structure of R, see e.g., Lehmann and Romano (2005). Therefore, the LRT will always be considered throughout this work as the 'local' test. We stress, however, that our methodology could also be used for other systems of local tests, provided they admit a sufficiently well behaved asymptotic expansion. The LRT is based on the test statistic

$$T_R(Y, \theta_0) := \sqrt{2 \log \left(\frac{\sup_{\theta \in \Theta} \prod_{i \in R} f_\theta(Y_i)}{\prod_{i \in R} f_{\theta_0}(Y_i)} \right)}, \qquad\qquad (1.2)$$

where f_θ denotes the density of F_θ and $H_{R,n}$ is rejected when $T_R(Y, \theta_0)$ is too large. For example this yields for a Poisson random field Y with given baseline intensity $\lambda_0 > 0$

$$T_R(Y, \lambda_0) := \sqrt{2|R| \left[\overline{Y}_R \log(\overline{Y}_R / \lambda_0) - (\overline{Y}_R - \lambda_0) \right]}, \qquad\qquad (1.3)$$

where $|R|$ denotes the number of points in R and $\overline{Y}_R := |R|^{-1} \sum_{i \in R} Y_i$. To be well-defined, set $T_R(Y, \lambda_0) := \sqrt{2|R|\lambda_0}$ if $\overline{Y}_R = 0$. If θ_0 is unknown, $H_{R,n}$ is rejected when $T_R(Y, \overline{Y})$ is too large with $\overline{Y}$ denoting the sample mean.

Since it is not known a priori which regions R might contain anomalies, i.e., for which

$R \in \mathcal{R}$ the alternative ($K_{R,n}$) might be true, it is of great importance to control the family wise error arising from the multiple test decisions of the local tests based on $T_R(Y, \theta_0)$, $R \in \mathcal{R}$. In order to create a powerful test while controlling this error, further restrictions on the complexity of $\mathcal{R}$ are necessary. To this end, we will assume that the regions $\mathcal{R}$ can be represented as a sequence of discretized regions

$$\mathcal{R} = \mathcal{R}_n := \left\{ R \subset I_n^d \mid R = I_n^d \cap nR^* \text{ for some } R^* \in \mathcal{R}^* \right\} \tag{1.4}$$

for some system of subsets of the unit cube $\mathcal{R}^* \subset 2^{[0,1]^d}$ (e.g., all hypercubes), to be specified later. This gives rise to the sequence of multiple testing problems

$$H_{R,n} \text{ vs. } K_{R,n} \qquad \textbf{simultaneously over } R \in \mathcal{R}_n. \tag{1.5}$$

Figure 1.3 illustrates exemplarily this sequence of multiple testing problems.
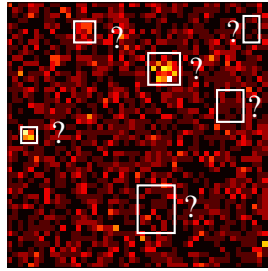
Figure 1.3: Poisson field from Figure 1.1 with several regions (white squares) over which the simultaneous test procedure is performed.

We want to provide a methodology to (asymptotically) control the family wise error rate (FWER) $\alpha \in (0, 1)$ when (1.5) is considered as a multiple testing problem, i.e., to provide a multiple test ϕ for (1.5) (see e.g., Dickhaus, 2014) such that

$$\sup_{R \in \mathcal{R}_n} \mathbb{P}_{H_{R,n}} \left[\text{"any false rejection by } \phi \text{"} \right] \le \alpha + o(1) \qquad \text{as} \qquad n \to \infty. \tag{1.6}$$

This ensures that the probability of making any wrong detection is controlled by a given error level α, as $n \to \infty$.

This task has been the focus of several papers during the last decades; for a detailed discussion see Section 1.2. We contribute to this field by providing a general theory for a unifying method in the model (1.1), which includes not only Gaussian (Arias-Castro et al., 2005; Sharpnack and Arias-Castro, 2016; Kou, 2017; Cheng and Schwartzman, 2017), but also Bernoulli (Walther, 2010) or Poissonian observations (Kulldorff et al., 2005; Rivera and Walther, 2013; Tu, 2013; Zhang et al., 2016). In contrast to (Arias-Castro et al., 2011), where also observations from exponential families as in (1.1) are

Figure 1.4: Test results for the setting of Figure 1.3.

discussed, but where the local tests are always as in the Gaussian setting, we emphasize that our local tests are of type (1.2), hence exploiting the likelihood in the exponential family, which results in improved power (see Frick et al. (2014) for $d = 1$).

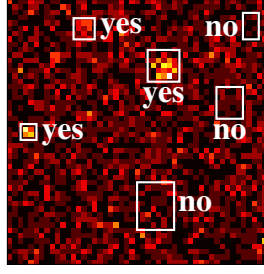**Multiscale testing for known baseline parameter** Our test in Chapter 2.1 will be of scanning-type, controlling the FWER by taking the maximum over all local LRT statistics in (1.2), i.e.,

$$T_n \equiv T_n(Y, \theta_0, \mathcal{R}_n, v) := \max_{R \in \mathcal{R}_n} \left[T_R(Y, \theta_0) - \text{pen}_v(|R|) \right]. \tag{1.7}$$

The values

$$\text{pen}_v(r) := \sqrt{2v \left(\log \left(n^d / r \right) + 1 \right)} \tag{1.8}$$

where log denotes the natural logarithm, act as a scale penalization, which is necessary to guarantee the optimal detection power on all scales simultaneously, as it prevents smaller regions from dominating the overall test statistic, as noticed by Dümbgen and Spokoiny (2001) and others (see e.g., Dümbgen and Walther, 2008; Walther, 2010; Frick et al., 2014). The constant v in (1.7) can be any upper bound $V_{\mathcal{R}^*}$ of the complexity of $\mathcal{R}^*$, measured in terms of the packing number (see Remark 2.1.2 below). If $\mathcal{R}^*$ has finite VC-dimension $v(\mathcal{R}^*)$, we may choose $V_{\mathcal{R}^*} = v(\mathcal{R}^*)$. However, the test will have better detection properties if v is as small as possible with this property (see Section 2.1.3). Hence, from this point of view it is advantageous to know exactly the complexity $V_{\mathcal{R}^*}$ of $\mathcal{R}^*$, a topic which has received less attention than computing VC-dimensions. Therefore, we compute the packing numbers for three important examples of $\mathcal{R}^*$, namely hyperrectangles, hypercubes, and halfspaces explicitly in Appendix A. To construct a test which controls the FWER (1.6), we have to find a sequence of universal global thresholds $q_{1-\alpha,n}$ such that

$$\mathbb{P}_0 \left[T_n > q_{1-\alpha,n} \right] \leq \alpha + o(1), \tag{1.9}$$

where $\mathbb{P}_0 := \mathbb{P}_{H_{I_n^d},n}$ corresponds to the case that no anomaly is present. Such a threshold

suffices, as it can be readily seen that

$$\sup_{R \in \mathcal{R}_n} \mathbb{P}_{H_{R,n}} \left[\text{"any false rejection in } R\text{"} \right] \leq \mathbb{P}_0 \left[\text{"any false rejection in } I_n^{d}\text{"} \right].$$

Given $q_{1-\alpha,n}$, the multiple test then will reject the null hypothesis whenever $T_n \geq q_{1-\alpha,n}$, and each local test rejects if $T_R(Y, \theta_0) \geq q_{1-\alpha,n} + \mathrm{pen}_v(|R|)$. Due to (1.7) and (1.9), any of these rejections is correct with probability $\geq 1 - \alpha$, asymptotically.

To obtain the thresholds $q_{1-\alpha,n}$ we provide a Gaussian approximation of the scan statistic (1.7) under $\mathbb{P}_0$ given by

$$M_n \equiv M_n(\mathcal{R}_n, v) := \max_{R \in \mathcal{R}_n} \left[|R|^{-1/2} \left| \sum_{i \in R} X_i \right| - \mathrm{pen}_v(|R|) \right] \tag{1.10}$$

with i.i.d. standard normal r.v.'s X_i, $i \in I_n^d$. We also give a rate of convergence of this approximation (Thm. 2.1.5), which is determined by the smallest scale in $\mathcal{R}_n$. Based on this, we obtain the $\mathbb{P}_0$-limiting distribution of T_n as that of

$$M \equiv M(\mathcal{R}^*, v) := \sup_{R^* \in \mathcal{R}^*} \left[\frac{|W(R^*)|}{\sqrt{|R^*|}} - \mathrm{pen}_v\left(n^d |R^*|\right) \right] < \infty \text{ a.s.}, \tag{1.11}$$

where W is white noise on $[0, 1]^d$ and (with a slight abuse of notation) $|R^*|$ denotes the Lebesgue measure of $R^* \in \mathcal{R}^*$. This is true as long as $\mathcal{R}^*$ has finite complexity, $\mathcal{R}^*$ consists of sets with a sufficiently regular boundary (see Assumption 1 below), and the smallest scales $|R_n|$ of the system $\mathcal{R}_n$ are suitably restricted, see (1.23) and the discussion there.

In the case of $\mathcal{R}^*$ being the subset of all hypercubes, we will also give an asymptotic expansion of the above test's power (see Theorem 2.1.9), which allows to determine the necessary average strength of an anomaly needed to be detected with asymptotic probability 1 (see Corollary 2.1.11). This is only possible due to the penalization in (1.7), as otherwise the asymptotic distribution is not a.s. finite. If the anomaly is sufficiently small, we show that the anomalies which can be detected with asymptotic power one by the described multiscale testing procedure are the same as those of the oracle single scale test, which knows the size (scale) of the anomaly in advance. This generalizes findings of Sharpnack and Arias-Castro (2016) to situations where not only the mean of the signal is allowed to change, but its whole distribution. Furthermore, if the observations are Gaussian, our test with properly chosen v achieves the asymptotic optimal detection boundary, i.e., no test can have larger power asymptotically in a minimax sense.

Note finally, that weak convergence of T_n to M as in (1.11) can be viewed as a gen-

eralized weak invariance principle, since the right hand side does not depend on any unknown quantity, and hence can be, e.g., simulated generically in advance for any given system $\mathcal{R}$ as soon as a bound for the complexity of $\mathcal{R}^*$ can be determined.

Multiscale testing for unknown baseline parameter In Chapter 2.2 we will restrict ourselves to observations given by an independent d-dimensional field Y of Poisson random variables

$$Y_i \sim \text{Poi}(\lambda_i), \quad i \in I_n^d := \{1, \ldots, n\}^d, \tag{1.12}$$

for some $\lambda_i > 0$. These appear in many applications ranging from astronomy and biophysics to genetics engineering. Specific examples include detection in radiographic images (Kazantsev et al., 2002), genome screening (Jiang et al., 2016), and object detection in astrophysical image analysis (Friedenberg and Genovese, 2013), just to mention a few. In astrophysics, this is relevant for the detection of galaxies or stars, and in fluorescence microscopy it can be employed to segment the image into parts which deserve further investigation and 'inactive' parts, a topic which will be discussed in more detail later.

As before, given some baseline parameter $\lambda_0 > 0$, we consider the problem of finding hot spots in the field Y, i.e., to detect regions $R \subset I_n^d$, where $\lambda_i \neq \lambda_0$ for some $i \in R$ out of a given family of candidate regions $R \subset \mathcal{R}_n \subset 2^{I_n^d}$. The hypothesis $H_{R,n}$ and alternative $K_{R,n}$ then become

$$\forall\, i \in R : \lambda_i = \lambda_0 \tag{$H_{R,n}$}$$

vs.

$$\exists\, i \in R \text{ s.th. } \lambda_i \neq \lambda_0, \tag{$K_{R,n}$}$$

Again, we use the LRT as local tests.

However, now, the underlying true baseline parameter λ_0 is assumed to be unknown and will be pre-estimated by the sample mean $\overline{Y}$. Due to the pre-estimation step one has to cut off the largest scales to derive a Gaussian approximation. This leads to a change of the penalty-term, since the penalty term (1.8) used before would result in a degenerate limit, see (Schmidt-Hieber et al., 2013).

Otherwise, the setup will be similar to the case of known baseline parameter, since we will consider a test of scanning-type, based on local LRT statistics (1.2). Therefore we will combine (1.3) with $\overline{Y}$ as a plug-in estimator for λ_0. More precisely, the local test for a fixed $R \subset I_n^d$ in this model is given by

$$T_R(Y) := \sqrt{2|R| \left[\overline{Y}_R \log(\overline{Y}_R/\overline{Y}) - (\overline{Y}_R - \overline{Y}) \right]}, \tag{1.13}$$

where $\overline{Y}_R = |R|^{-1} \sum_{i \in R} Y_i$. To be well-defined, we set $T_R(Y) := \sqrt{2\overline{Y}|R|}$ if $\overline{Y}_R = 0$.

No prior information on the location and size of the regions which contain an anomaly is assumed, and this results in the sequence of multiple testing problems (1.5).

In contrast to (1.7) in this case of an unknown baseline parameter, we will extend the setting to general scale calibration terms (which fulfil certain assumptions) to allow the user to choose the penalization of their choice. This means that the considered multiscale statistic has the following form

$$T_n^\omega \equiv T_n^\omega(Y, \mathcal{R}_n) := \max_{R \in \mathcal{R}_n} \widetilde{\omega}(|R|) \left[T_R(Y) - \omega(|R|) \right], \tag{1.14}$$

where the scale calibration terms $\widetilde{\omega}, \omega$ should be chosen in a way such that domination of small scales is avoided. The test setup follows the lines of the previous subsection. Again, the construction of the test is chosen such that the FWER is controlled, i.e., $\mathbb{P}_0 \left[T_n^\omega > q_{1-\alpha,n} \right] \leq \alpha + o(1)$, for a sequence of universal global thresholds $q_{1-\alpha,n}$. To obtain these thresholds $q_{1-\alpha,n}$, we establish a Gaussian approximation M_n^ω of the scan statistic (1.14) under $\mathbb{P}_0$, and based on this obtain the $\mathbb{P}_0$-limiting distribution M^ω of T_n^ω. Naturally, the Gaussian approximation M_n^ω and the $\mathbb{P}_0$-limit M^ω have to be adapted to the new scale calibration terms, that is,

$$M_n^\omega(\mathcal{R}_n) := \max_{R \in \mathcal{R}_n} \widetilde{\omega}(|R|) \left[|R|^{-1/2} \left| \sum_{i \in R} X_i \right| - \omega(|R|) \right] \tag{1.15}$$

with i.i.d. standard normal r.v.'s X_i, $i \in I_n^d$, and

$$M^\omega(\mathcal{R}^*) := \sup_{R^* \in \mathcal{R}^*} \widetilde{\omega}(n^d|R^*|) \left[\frac{|W(R^*)|}{\sqrt{|R^*|}} - \omega(n^d|R^*|) \right] \tag{1.16}$$

where W denotes white noise on $[0, 1]^d$.

This is true if certain complexity assumptions on $\mathcal{R}^*$ are satisfied and the range of scales are suitably restricted, see (1.24) and the discussion there.

1.1 Applications

Finding anomalies in a field Y is a widely occurring problem. Anomalies show up in numerous areas of application ranging from astronomy and biophysics to genetics engineering.

In this thesis, we will focus on finding hot spots in microscopic images of fluorescence scanning microscopes. The resolution level in microscopy has increased over the past years, and the resulting images now contain details on different scales and a
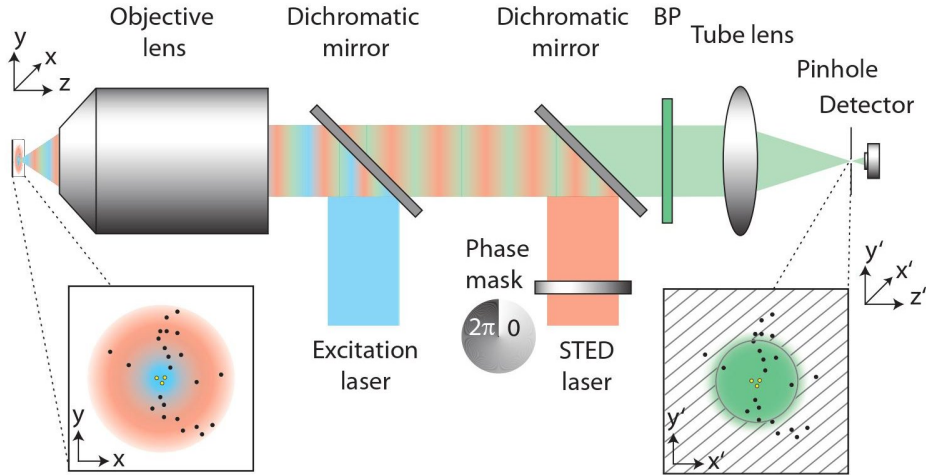
Figure 1.5: Figure of a STED microscope (taken from (Aspelmeier et al., 2015)). The excitation laser beam is superimposed with a STED laser beam. This results in a diffraction-limited excitation spot overimposed by a doughnut-shaped depletion intensity distribution that triggers stimulated emission. At the very center of the STED-doughnut, fluorophores are able to fluoresce. The signal is then mapped from this area onto a point-like detector. Therefore, it has to pass through a pinhole that allows only the central signal to pass through (right box) and then a computer records the photon count. The complete image is then produced by moving the laser and detection spots jointly pixel by pixel through the specimen.

reconstruction should adapt and preserve these details. Further, we will illustrate why our theoretical results are useful in the context of the Multiscale Nemirovskij-Dantzig Estimator (MIND) (Grasmair et al., 2018) and touch on the challenges brought up by considering higher dimensions.

1.1.1 STED

In fluorescence scanning microscopy, structures of interest (e.g., proteins) in a sample are labelled by fluorescent markers, which are afterwards visualized by the microscope. To this end, the prepared specimen is scanned spatially along a grid with a diffraction-limited spot centered at the current grid point. Whenever a marker is hit by the incoming light it is excited with a certain probability, and if so, afterwards light (of a different wavelength) is emitted and recorded by detectors (cf. Aspelmeier et al., 2015, for details). However, this procedure cannot be repeated too often, as each marker is only able to pass through the cycle of excitation and emission a limited number of times before the marker bleaches.

Figure 1.5 shows a sketch of a STimulated Emission Depletion (STED) microscope, which belongs to the fluorescence microscopes and is one of the best known concepts to surpass Abbe's diffraction limit (Hell, 2007). STED enhances the resolution by applying an additional doughnut-shaped depletion beam around the current scanning

position sending most of the surrounding markers to a dark state in which they cannot emit light. Then only the markers within a subdiffraction-sized spot remain in a bright state and will emit light. For further details we refer to (Aspelmeier et al., 2015).

Exposure to light always pre-stresses the marker and further reduces the possible number of cycles. Note that due to the nature of light, when centered at grid point, also the markers at neighbouring points are illuminated and hence pre-stressed. Therefore, by scanning over the whole image, only a very limited number of cycles per grid point can be performed (typically around 10^3). As the markers are typically concentrated in comparably small regions of interest, but are also stressed when scanning mostly empty regions in the image, this gives rise to a considerable loss of information.

As a remedy, methods have been developed to reduce the number of excitation cycles in mostly empty regions of the image. They are mostly based on real-time spatial control of the illumination and are collected under the name spatially-controlled illumination microscopy (SCIM), see e.g., the review by Krishnaswami et al. (2016). The specifically used techniques depend on the underlying fluorescent microscope. Here we will focus on STED microscopy (Hell, 2007).

In case of STED, the reduction of illumination cycles can be achieved by introducing a decision number N and a threshold $lTh \in \mathbb{N}$. On each grid position, only N laser pulses are applied, and if less then lTh photons were collected during these cycles, the laser foci are moved to the next grid position. The underlying idea is that in active regions, more photons are to be expected than in inactive regions, and hence the aforementioned method ensures that inactive regions are scanned less intensive. For details we refer to the original publication (Staudt et al., 2011) where the method was introduced as REduction of State transition Cycles (RESCue)-STED. Mathematically we model this as

$$Y_i \sim \text{Poi}(t \cdot \lambda_i), \qquad i \in I_n^d, \tag{1.17}$$

where $\lambda_i = I(x_i)$ and

x_i: sampling grid,

I: energy density, e.g., in microscopy $I = h * f$ where h is the effective point spread function (psf) of the microscope and f is the relative marker density,

t: (pixel) dwell time, limited by bleaching.

Later, in Chapter 3, we will present our idea to achieve similar results as RESCue-STED (i.e., saving photons) by identifying those regions where the intensity is differ-

ent from the baseline intensity by applying our approach on measurements with small dwell time. This data allows to locate interesting areas and investigate them further. The full dwell time can then just be used on the interesting spots. We stress that our method provides uniform statistical guarantees. With a chosen level α, we can state after performing our hypothesis testing procedure that with $\alpha \cdot 100\%$ confidence each detected region contains an anomaly.

1.1.2 Constrained Multiscale Optimization

The theoretical results in this thesis are also important in the context of image reconstruction. We will not reconstruct here, but still mention why the theory gives validation for algorithms based on the following reconstruction methods.

Instead of (1.1) let us consider more general observations (1.18), where u is the image to be reconstructed and K is a matrix modelling the image acquisition and sampling (e.g., convolution):

$$Y_i \sim F_{Ku\left(\frac{i}{n}\right)}, \qquad i \in I_n^d, \tag{1.18}$$

i.e., $Ku := Ku\left(\frac{i}{n}\right)$ are now considered as parameters in setting (1.1). Estimation procedures of the form

$$\min_u R(u) \qquad \text{s.t.} \quad T_n(Y, Ku) \leq q \tag{1.19}$$

where R is a variational functional (e.g., the TV-norm or the l^1-norm), T_n is the multiscale constraint from (1.7) adapted to this setting, Y are observations, have been suggested in the literature, see e.g., (Candes et al., 2007), (Grasmair et al., 2018), (Frick et al., 2012) . The Multiscale Nemirovskij-Dantzig Estimator (MIND) introduced in (Grasmair et al., 2018) results from the minimization problem (1.19). MIND can be seen as a generalization of the Dantzig selector, a particular data-fidelity constraint estimator (Candes et al., 2007). There, the observations are given by $Y_i = (Ku)_i + \sigma \epsilon_i$, where K is a linear, bounded operator and ϵ_i are i.i.d. standard normally distributed random variables. The Dantzig estimator is the solution to the l^1-regularization problem

$$\min_{u \in \mathbb{R}^d} \|u\|_{l^1} \quad \text{subject to} \quad \|K^*(Y - Ku)\|_{l^\infty} \leq q. \tag{1.20}$$

So the generalization of MIND is in the sense that other reconstructions than the l^1-norm (namely a general convex functional R) can be used, e.g., the $TV-$norm or a Sobolev norm, and it furthermore relaxes the $l^\infty-$constraint.

The numerical benefit of MIND has been studied in (Frick et al., 2012) and (Frick et al., 2013).

Theoretical benefit of the MIND estimator We show in Chapter 2 that the quantiles of T_n can be approximated uniformly in our setting by those of a Gaussian version M_n, which does not depend on any unknown quantity. Generalizing these to non-constant θ would give the possibility to compute $q_{1-\alpha}$ by simulations as the $(1-\alpha)$-quantile of M_n (see Chapter 4 for an idea) and get as in (1.9)

$$\lim_{n\to\infty} \mathbb{P}_0 \left[T_n \le q_{1-\alpha} \right] = 1 - \alpha.$$

Then the MIND estimator $\hat{u}_n(q)$ is at least as smooth - measured in R - as the true image with large probability of $(1-\alpha)$, i.e.,

$$\lim_{n\to\infty} \sup_u \mathbb{P} \left[R \left(\hat{u}_n(q_{1-\alpha}) \le R(u) \right) \right] \ge 1 - \alpha.$$

This illustrates why the asymptotic distribution of T_n is of interest in this context. The determination of q in (1.19) often requires rather complicated distributional theory. The limit theorems in this thesis provide an important step.

1-dimensional case Frick et al. (2014) investigated the 1-dimensional case in a change-point regression setting. They assume observations

$$Y_i \sim F_{\vartheta(i/n)}, \qquad i = 1, \dots, n \tag{1.21}$$

where F_θ is a one-dimensional exponential family and $\vartheta : [0,1) \to \Theta$ is a right-continuous change-point function with an unknown number of change points.
Their method "Simultaneous Multiscale Change-Point Estimator" (SMUCE) for the estimation of a step function is a hybrid method consisting of likelihood ratio test and the minimization of a cost functional, i.e., two simultaneously combined steps will be addressed: model selection (estimation of the number of change-points K)

$$\hat{K} := \inf_{\vartheta \in \mathcal{S}} \#J(\vartheta) \quad \text{s.t.} \quad T_n(Y, \vartheta) \le q,$$

and estimation of the change-point function given $\hat{K}$, where $\mathcal{S}$ denotes the class of right-continuous change-point functions and $J(\vartheta)$ is the ordered vector of change-points.
One aspect which the authors address is that for Poisson data $Y_i \overset{i.i.d.}{\sim} \text{Poi}(u_i)$ SMUCE works better if the LRT's are adapted to Poisson distribution instead of working with

a Gaussian approximation (e.g., Anscombe transformation). In our setting with higher dimensions, we illustrate a behaviour of this type as well trough a simulation (Section 3.2.1).

1.2   Literature review and connections to existing work

Scan statistics and scanning-type procedures based on the maximum over an ensemble of local tests have received much attention in the literature over the past decades.
To determine the quantile, a common option is to suitably approximate the tails of the asymptotic distribution, as done e.g., by Siegmund and Venkatraman (1995); Siegmund and Yakir (2000); Naus and Wallenstein (2004); Pozdnyakov et al. (2005); Fang and Siegmund (2016) for $d = 1$, by Haiman and Preda (2006) for $d = 2$, and by Jiang (2002) in arbitrary dimensions. If the random field is sufficiently smooth (in contrast to the setting here) the Gaussian kinematic formula can be employed, see e.g., Taylor and Worsley (2007), Adler (2000). We also mention Alm (1998), who considers the situation of a fixed rectangular scanning set in two and three dimensions. In all these papers, no penalization has been used, which automatically leads to a preference for small scales of order $\log(n)$ (see e.g., Kabluchko and Munk, 2009) and to an extreme value limit, in contrast to (1.11). Arias-Castro et al. (2017) study the case of an unknown null distribution and propose a permutation based approximation, which is shown to perform well in the natural exponential family setting (1.1), however, only for $d = 1$. Conceptually most related to our work are weak limit theorems for scale penalized scan statistics, which have e.g., been obtained by Frick et al. (2014) and Sharpnack and Arias-Castro (2016). However, these results are either limited to specific situations such as Gaussian observations, or to $d = 1$. If a limit exists, the quantiles of the finite sample statistic can be bounded the quantiles by limiting ones as e.g., done by Dümbgen and Spokoiny (2001); Rivera and Walther (2013).

1.3   Main results

The main technical contribution of this thesis is to prove a weak limit theorem for the asymptotic distribution of our test statistic for general exponential family models as in (1.1) and for arbitrary dimensions d. This can be viewed as a "multiscale" weak invariance principle for independent but not necessarily identically distributed r.v.'s. Further, we provide an asymptotic expansion of the test's power in the first setting of a known baseline parameter θ_0 which leads to minimax optimal detection of the test in specific

models, cf. 2.1.3. In total, we acquire a unified theory for (calibrated) scan statistics in exponential families, a topic in which many special cases have received attention before (see the literature review before). Furthermore we address the application side by providing a theory for unknown baseline parameters.

More precisely, our results are twofold as we give a Gaussian approximation of the scan statistic in (1.7) by (1.10), *and* that we obtain (1.11) as a weak limit, and correspondingly we give for the scan statistic (1.14) a Gaussian approximation (1.15) and as a weak limit (1.16).

Weak limit theorems for T_n as in (1.7) or (1.14) are closely connected to those for partial sum processes. Classical KMT-like approximations (see e.g., Komlós et al., 1976; Rio, 1993; Massart, 1989) provide, in fact, a strong coupling of the whole process $(T_R(Y, \theta_0))_{R \in \mathcal{R}_n}$ to a Gaussian version. Results of this form have been previously employed for $d = 1$ in (Schmidt-Hieber et al., 2013; Frick et al., 2014). Proceeding like this for general d, however, will restrict the system $\mathcal{R}_n$ to scales r_n s.t. $|R| \geq r_n$ where

$$n^{d-1} \log(n) = o(r_n) \tag{1.22}$$

as $n \to \infty$, which is unfeasibly large for $d \geq 2$. Therefore, we use a different approach and employ a coupling of the maxima in (1.7) and (1.10), which relies on recent results by Chernozhukov et al. (2014), see also (Proksch et al., 2018). However, in contrast to the consideration here, these authors do not consider the local LR statistic and require that the largest scale has to go to zero. This leads to an extreme value type limit in contrast to (1.11) which incorporates all (larger) scales. To make use of Chernozhukov et al.'s (2014) coupling results in our general setting, we provide a symmetrization-like upper bound for the expectation of the maximum of a partial sum process by a corresponding Gaussian version, cf. Lemma 2.1.14. This way we can approximate the distribution of T_n in (1.7) by (1.10) if we restrict ourselves to sets $R \in \mathcal{R}_n$ with $|R| \geq r_n$ where the smallest scales only need to satisfy

$$\log^\gamma(n) = o(r_n) \text{ as } n \to \infty, \tag{1.23}$$

for a suitable fixed γ (to be specified later), which compared to (1.22) allows for considerably smaller scales whenever $d \geq 2$. Note that the lower scale restriction (1.23) does not depend on d. However, as we consider sets in I_n^d here, the corresponding lower bound a_n for sets in $\mathcal{R}^* \subset 2^{[0,1]^d}$ is $n^{-d} \log^\gamma(n) = o(a_n)$, which in fact depends on d as now the volume of the largest possible set has been standardized to one (see (1.4) and Theorem 2.1.9) and coincides with the sampling rate n^{-d} up to a poly-log-factor.

In contrast, (1.22) gives $n^{-1} \log(n) = o(a_n)$, independent of d, which only for $d = 1$ achieves the sampling rate n^{-d}. Under (1.23) we also obtain $O_{\mathbb{P}}\left(\left(\log^{12}(n)/r_n \right)^{1/10} \right)$ as rate of convergence of this approximation (see (2.6) below).

To use the coupling result for the scan statistic (1.14) we restrict the sets $R \in \mathcal{R}_n$ further to allow for the possibility of a multiplicative scale calibration term $\widetilde{\omega}$. We may approximate the distribution of T_n^{ω} in (1.14) by (1.15) if we restrict ourselves to scales $R \in \mathcal{R}_n$ with upper and lower bounds, namely $m_n \geq |R| \geq r_n$, where r_n fulfils (1.23) with a γ based on the behaviours of the scale calibration terms to be specified later and

$$m_n = C \frac{n^d}{\log(n)} c_n^3, \tag{1.24}$$

where C denotes a positive constant and c_n a sequence tending to zero arbitrary slowly.

CHAPTER 2

Theory

In this chapter we give details on our theoretical findings. Section 2.1.1 contains an overview, details on our precise setting and the assumptions on the set of candidate regions $\mathcal{R}^*$. In Section 2.1.2 we provide the validity of the Gaussian approximation in (1.10) and determine the $\mathbb{P}_0$-limiting distribution of T_n. In Section 2.1.3 we derive an asymptotic expansion of the detection power. Afterwards in Section 2.2, extensions to unknown baseline parameters and to general scale calibration terms will be considered in the setting of a Poisson random field. In Section 2.2.1 we state the new setup and modified assumptions, and in Section 2.2.2 we give the limit theory for this setting.

2.1 Exponential Families

2.1.1 Setting and Assumptions

In the following we assume that $\mathcal{F} = \{F_\theta\}_{\theta \in \Theta}$ in (1.1) is a one-dimensional exponential family, which is regular and minimal, i.e., that the Lebesgue densities of F_θ are of the form $f_\theta(x) = h(x) \exp\left(\langle \theta, x \rangle - \psi(\theta)\right)$, that the natural parameter space

$$\mathcal{N} = \left\{ \theta \in \mathbb{R}^d : \int_{\mathbb{R}^d} \exp(\theta x) \, dx < \infty \right\}$$

is open, and that the cumulant transform ψ is strictly convex on $\mathcal{N}$. Then, the moment generating function exists and the random variables Y_i have sub-exponential tails, see (Casella and Berger, 2002) and (Brown, 1986) for details. Let $\phi(x) := \sup_{\theta \in \Theta} \left[\theta \cdot x - \psi(\theta) \right]$ be the Legendre-Fenchel conjugate of ψ and

$$J(x, \theta) := \phi(x) - \left[\theta \cdot x - \psi(\theta) \right].$$

Then the LRT statistic $T_R(Y, \theta_0)$ in (1.2) can be written as

$$T_R(Y, \theta_0) = \sqrt{2 \left(\sup_\theta \sum_{i \in R} (\theta \cdot Y_i - \psi(\theta)) - \sum_{i \in R} (\theta_0 \cdot Y_i - \psi(\theta_0)) \right)}$$

$$= \sqrt{2 |R| J \left(\bar{Y}_R, \theta_0 \right)} \tag{2.1}$$

with $\bar{Y}_R = |R|^{-1} \sum_{i \in R} Y_i$. Note that by definition we have that $J \left(\bar{Y}_R, \theta_0 \right) \geq 0$.

Remark 2.1.1. *If the observations are not drawn from an exponential family as in (1.1), or if $\theta_0 \in \Theta$ is replaced by a field $(\theta_i)_{i \in I_n^d}$ of known baseline intensities, the representation of the LRT statistic T_R as in (2.1) is not valid anymore. Our proofs rely on a third-order Taylor expansions of T_R and on the sub-exponential tails of the random variables Y_i (see Theorem 2.1.5 below), but not explicitly on the exponential family structure. Therefore, if in more general models correspondingly modified assumptions are posed (see also Arias-Castro et al., 2017, Sec. 2.2), our results do extend to this situation.*

To control the supremum in (1.11), we have to restrict the system of regions $\mathcal{R}^*$ suitably. To this end, we introduce some notation and definition. For a set $R^* \in \mathcal{R}^*$ and $x \in [0, 1]^d$ we define $d(x, \partial R^*) := \inf_{y \in \partial R^*} \|x - y\|_2$ where ∂R^* denotes the topological boundary of R^*, i.e., $\partial R^* = \overline{R^*} \setminus (R^*)^\circ$. Furthermore we define the $\epsilon-$annulus $R^*(\epsilon)$ around the boundary of R^* for some $\epsilon > 0$ as

$$R^*(\epsilon) := \left\{ x \in [0, 1]^d \ \big| \ d(x, \partial R^*) < \epsilon \right\}.$$

Further, the VC-Dimension of a family of sets $\mathcal{R}^*$ is defined as the largest integer n s.t. $S_{\mathcal{R}^*}(n) = 2^n$, where $S_{\mathcal{R}^*}(n)$ denotes the shatter coefficient, i.e., the maximal number of different subsets of a set of n points which can be obtained by intersecting it with elements of $\mathcal{R}^*$. If $S_{\mathcal{R}^*}(n) = 2^n$ for all n, we say that the $VC-$Dimension is infinity.

Assumption 1 (Complexity and regularity of $\mathcal{R}^*$)**.**

(a) *The VC-Dimension of $\mathcal{R}^*$ is bounded by $\nu(\mathcal{R}^*) < \infty$.*

(b) *There exists some constant $C > 0$ such that $|R^*(\epsilon)| \leq C\epsilon$ for all $\epsilon > 0$ and all $R^* \in \mathcal{R}^*$ where $|\cdot|$ denotes the Lebesgue measure.*

Let us briefly comment on the above assumption.

Remark 2.1.2.

- *Assumption 1(a) is a standard assumption to control the complexity of the set indexed process, see e.g., (Massart, 1989; van der Vaart and Wellner, 1996; Dümbgen and Spokoiny, 2001).*

- *Assumption 1(a) immediately implies an upper bound for the cardinality $\#(\mathcal{R}_n)$ of $\mathcal{R}_n$ in (1.4), namely, there exist constants $c_1, c_2 > 0$ such that*

$$\#(\mathcal{R}_n) \leq c_1 n^{c_2}. \tag{2.2}$$

  This allows us to apply recent results by Chernozhukov et al. (2014) to couple the process in (1.7) with a Gaussian version as in (1.10).

- *In the following, we also assume a bound on the complexity of $\mathcal{R}^*$ in terms of the packing number. The packing number $\mathcal{K}(\epsilon, \rho, \mathcal{W})$ of a subset $\mathcal{W}$ of $\mathcal{R}^*$ w.r.t. a metric ρ is given by the maximum number m of elements $W_1, \ldots, W_m \in \mathcal{W}$ such that $\rho(W_i, W_j) > \epsilon$ for all $i \neq j$, i.e., the largest number of disjoint $\epsilon/2$-balls w.r.t. ρ which can be packed inside $\mathcal{W}$, see e.g., (van der Vaart and Wellner, 1996, Def. 2.2.3). In the following we consider the symmetric difference*

$$R_1^* \triangle R_2^* := \left(R_1^* \cup R_2^* \right) \setminus \left(R_1^* \cap R_2^* \right), \qquad R_1^*, R_2^* \in \mathcal{R}^*$$

  and the corresponding metric

$$\rho^* \left(R_1^*, R_2^* \right) := \sqrt{\left| R_1^* \triangle R_2^* \right|}, \qquad \text{for} \qquad R_1^*, R_2^* \in \mathcal{R}^*. \tag{2.3}$$

  Suppose that there exists a positive number $V_{\mathcal{R}^*}$ and constants $k_1, k_2 > 0$ such that

$$\mathcal{K}\left((\delta u)^{1/2}, \rho^*, \{R \in \mathcal{R}^* : |R| \leq \delta\} \right) \leq k_1 u^{-k_2} \delta^{-V_{\mathcal{R}^*}} \tag{2.4}$$

  for all $u, \delta \in (0, 1]$. If Assumption 1(a) is satisfied, then (2.4) holds true with $V_{\mathcal{R}^*} = \nu(\mathcal{R}^*)$, which basically follows from the relationship between capacity and covering numbers and (van der Vaart and Wellner, 1996, Thm. 2.6.4). However, (2.4) might also be satisfied for considerably smaller numbers $V_{\mathcal{R}^*}$, as examples below show.

- *We stress that the assumption on the boundary smoothness (b) is satisfied whenever $\mathcal{R}^*$ consists of regular Borel sets R^* only, i.e., R^* is a Borel set and $|\partial R^*| = 0$ for all $R^* \in \mathcal{R}^*$.*

Example 2.1.3.

a) *Consider the set $\mathcal{S}^*$ of all hyperrectangles in $[0, 1]^d$, i.e., each $S^* \in \mathcal{S}^*$ is of the form $S^* = [s, t] := \left\{ x \in [0, 1]^d \mid s_i \leq x_i \leq t_i \text{ for } 1 \leq i \leq d \right\}$. According to (van der Vaart and Wellner, 1996, Ex. 2.6.1) we have $\nu(\mathcal{S}^*) = 2d$. More refined computations in the appendix (cf. Lemma A.1.1) show that $V_{\mathcal{S}^*}$ may be chosen as $V_{\mathcal{S}^*} = 2d - 1 + \epsilon$ with arbitrary $\epsilon > 0$. Obviously, the corresponding discretization $\mathcal{S}_n$ consists of hyperrectangles in I_n^d, which are determined by their upper left and lower right corners, i.e., $\#(\mathcal{S}_n) \leq n^{2d}$. As $\mathcal{S}^*$ consists only of regular Borel sets, the assumption on the boundary smoothness is also satisfied.*

b) *We may also consider the (smaller) set $\mathcal{Q}^*$ of all hypercubes in $[0, 1]^d$, i.e., each $Q^* \in \mathcal{Q}^*$ is of the form $[t, t + h]$ with $t \in [0, 1]^d$ and $0 < h \leq 1 - \max_{1 \leq i \leq d} t_i$. As $\mathcal{Q}^* \subset \mathcal{S}^*$, Assumption 1 is satisfied. More precisely, according to (Despres, 2014), $\nu(\mathcal{Q}^*) = \lfloor \frac{3d+1}{2} \rfloor$, and refined computations in the appendix (cf. Lemma A.1.2) show $V_{\mathcal{Q}^*} = 1$ independent of d (as opposed to the VC-dimension).*

c) *Let $\mathcal{H}^*$ be the set of all half-spaces in $[0, 1]^d$, that is*

$$\mathcal{H}^* := \left\{ H_{a,\alpha} \mid \alpha \in \mathbb{R}, a \in \mathbb{S}^{d-1} \right\}, \quad H_{a,\alpha} := \left\{ x \in [0, 1]^d \mid \langle x, a \rangle \geq \alpha \right\}.$$

The VC-dimension of $\mathcal{H}^*$ is $\leq d + 1$ (see e.g., Devroye and Lugosi, 2001, Cor. 4.2), which proves that Assumption 1 is satisfied. On the other hand, we prove that we may take $V_{\mathcal{H}^*} = 2$ (cf. Lemma A.1.3 the appendix).

2.1.2 Limit Theory

We now show that the quantiles of the multiscale statistic in (1.7) can be approximated uniformly by those of the Gaussian version in (1.10) and furthermore, that $M_n(\mathcal{R}_n, v)$ in (1.10) converges to a non-degenerate limit for $v \geq V_{\mathcal{R}^*}$. For the first fact we require a lower bound on the smallest scale as given in (1.23). Given a discretized set of candidate regions $\mathcal{R}_n \subset 2^{I_n^d}$ and $c > 0$ we introduce

$$\mathcal{R}_{n|c} := \left\{ R \in \mathcal{R}_n \mid |R| \geq c \right\}.$$

With this notation we can formulate our main theorem:

Theorem 2.1.4 (Weak $\mathbb{P}_0$ limit)**.** *Let Y_i, $i \in I_n^d$ be a field of random variables as in (1.1), $\mathcal{R}^*$ satisfy Assumption 1 and let $(r_n)_n$ be a sequence such that (1.23) holds with*

$\gamma = 12$. *Then we have under $\mathbb{P}_0$ that*

$$T_n\left(Y, \theta_0, \mathcal{R}_{n|r_n}, v\right) \xrightarrow{\mathcal{D}} M\left(\mathcal{R}^*, v\right) \qquad as \qquad n \to \infty, \tag{2.5}$$

with $M\left(\mathcal{R}^*, v\right)$ as in (1.11) for any fixed $v \in \mathbb{R}$. If furthermore $v \geq V_{\mathcal{R}^*}$ in (2.4), then $M(\mathcal{R}^*, v)$ is almost surely finite and non-degenerate.

Note that M does not depend on any unknown quantities and can e.g., be simulated. However, for practical purposes it is advantageous to use the finite sample Gaussian approximation in (1.10) to approximate quantiles for T_n as in (1.7) by simulated quantiles of M_n as in (1.10). This is justified by the following theorem:

Theorem 2.1.5 (Gaussian approximation). *Let Y_i, $i \in I_n^d$ be a field of random variables as in (1.1) and let $\mathcal{R}^*$ be a set of candidate regions satisfying Assumption 1(a) and let $(r_n)_n \subset (0, \infty)$ be a sequence such that (1.23) holds with $\gamma = 12$. Let $v \in \mathbb{R}$ be fixed.*

(a) *Then under $\mathbb{P}_0$*

$$T_n\left(Y, \theta_0, \mathcal{R}_{n|r_n}, v\right) - M_n\left(\mathcal{R}_{n|r_n}, v\right) = O_{\mathbb{P}}\left(\left(\frac{\log^{12}(n)}{r_n}\right)^{1/10}\right), \tag{2.6}$$

as $n \to \infty$, with M_n as in (1.10).

(b) *For all $q \in \mathbb{R}$ we have*

$$\lim_{n \to \infty} \left|\mathbb{P}_0\left[T_n\left(Y, \theta_0, \mathcal{R}_{n|r_n}, v\right) > q\right] - \mathbb{P}\left[M_n\left(\mathcal{R}_{n|r_n}, v\right) > q\right]\right| = 0. \tag{2.7}$$

Remark 2.1.6. *Theorems 2.1.4 and 2.1.5 are compatible in the sense that for any set of candidate regions satisfying Assumption 1, any $v \in \mathbb{R}$, and any sequence r_n such that $\frac{r_n}{n^d} \to 0$ satisfying (1.23) with $\gamma = 12$ we have convergence in distribution*

$$M_n\left(\mathcal{R}_{n|r_n}, v\right) \xrightarrow{\mathcal{D}} M\left(\mathcal{R}^*, v\right) \qquad as \qquad n \to \infty, \tag{2.8}$$

with $M\left(\mathcal{R}^*, v\right)$ as in (1.11).

Example 2.1.7. *Suppose that $\mathcal{R}^*$ is a set of candidate regions satisfying Assumption 1. Let us discuss three important examples of the model (1.1).*

1. *Gaussian fields: Let $Y_i \sim \mathcal{N}\left(\theta, \sigma^2\right)$ where the variance $\sigma^2 > 0$ is fixed. In this case, $\psi(\theta) = \frac{1}{2}\theta^2$, and*

$$T_R\left(Y, \theta_0\right) = \sqrt{|R|}\frac{\left|\overline{Y}_R - \theta_0\right|}{\sigma}$$

2. *Bernoulli fields: Let $Y_i \sim Bin\,(1, p)$ with $p \in (0, 1)$. The cases $p = 0$ and $p = 1$ have to be excluded to obtain a natural exponential family. However, also in these cases one would screen the field correctly, anyway. The natural parameter of this exponential family is $\theta = \log\,(p/(1 - p))$, and using $\psi\,(\theta) = \log\,\big(1 + \exp\,(\theta)\big)$ we find*

$$T_R\,(Y, \theta_0) = \sqrt{2|R| \left[\overline{Y}_R \log\left(\frac{\overline{Y}_R}{\frac{\exp(\theta_0)}{1+\exp(\theta_0)}} \right) - (1 - \overline{Y}_R) \log\left(\frac{1 - \overline{Y}_R}{\frac{1}{\exp(\theta_0)+1}} \right) \right]}.$$

3. *Poisson fields: Let $Y_i \sim Poi(\lambda)$ with $\lambda \in \mathbb{R}$. Again, $\lambda = 0$ has to be excluded, but this case is again trivial. The natural parameter of the exponential family is $\theta = \log\,(\lambda)$, and using $\psi\,(\theta) = \exp\,(\theta)$ we obtain*

$$T_R(Y, \theta_0) = \sqrt{2|R| \left[\overline{Y}_R \log\left(\frac{\overline{Y}_R}{\exp(\theta_0)} \right) - (\overline{Y}_R - \exp(\theta_0)) \right]}.$$

Example 2.1.8 (Gaussian approximation in the hyperrectangle / hypercube case)**.** *Recall Example 2.1.3 and let $\mathcal{S}^*$ be the set of all hyperrectangles and $\mathcal{Q}^*$ be the set of all hypercubes in $[0, 1]^d$. Then for $(r_n)_n$ as in (1.23) with $\gamma = 12$ it holds under $\mathbb{P}_0$ that*

$$T_n\left(Y, \theta_0, \mathcal{S}_{n|r_n}, v\right) - M_n\left(\mathcal{S}_{n|r_n}, v\right) = O_{\mathbb{P}} \left(\left(\frac{\log^{12}(n)}{r_n} \right)^{1/10} \right),$$

and

$$T_n\left(Y, \theta_0, \mathcal{Q}_{n|r_n}, v\right) - M_n\left(\mathcal{Q}_{n|r_n}, v\right) = O_{\mathbb{P}} \left(\left(\frac{\log^{12}(n)}{r_n} \right)^{1/10} \right),$$

as $n \to \infty$, with M_n as in (1.10). Monte-Carlo simulations (by means of (1.10) with $n = 96$ and $d = 2$) of the densities of M_n with different values of v are shown in Figure 2.1. The smallest possible values of v which we may choose according to Example 2.1.3 are given by the packing number estimate, i.e., $v = 3 + \epsilon$ and $v = 1$, respectively. The corresponding results are depicted in the top row of Figure 2.1 with $\epsilon = 0$ for simplicity. Alternatively, we may use the VC-dimension $v = 4$ and $v = 3$ respectively, which lead to the simulated distributions shown in the bottom row of Figure 2.1. This nicely illustrates that using a larger value of v will lead to larger quantiles and hence a loss of detection power: As the distributions of $M_n\,(\mathcal{S}^n, 4)$ and $M_n\,(\mathcal{Q}_n, 3)$ are extremely close, detecting in the system of squares is not easier than detecting in the system of rectangles, even though the latter is by far bigger and more complex. The explanation

for this is the penalization $\mathrm{pen}_v\,(|Q|)$, *which by appropriate choice of the parameter v can be tailored to the system* $\mathcal{R}^*$.
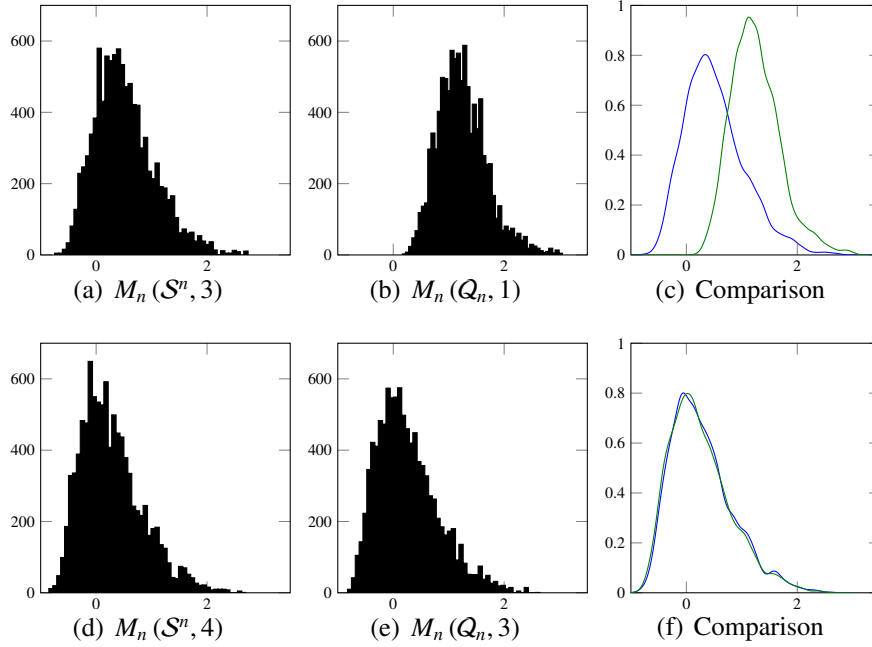

Figure 2.1: Simulated densities of the Gaussian approximations. The histograms are obtained from 10^4 runs of the test statistic (1.10). For the comparison, the corresponding densities have been estimated by a standard kernel estimator ($M_n\,(\mathcal{S}^n, v)$ (——), $M_n\,(Q_n, v)$ (——)). Top row: optimal calibration with the covering number, bottom row: alternative calibration using the VC-dimension.

2.1.3 Asymptotic Power

In this subsection we will analyze the power of our multiscale testing approach in the hypercube-case. The detection power clearly depends on the size and strength of the anomaly. To describe the latter, we will frequently employ the functions

$$m(\theta) := \psi'(\theta) = \mathbb{E}\,[Y]\,, \qquad v(\theta) := \psi''(\theta) = \mathbb{V}\,[Y] \tag{2.9}$$

for $Y \sim F_\theta$.

Heuristics The key point for the following power considerations is that the observations in (1.1) can be approximated as

$$\frac{Y_i - m\,(\theta_0)}{\sqrt{v\,(\theta_0)}} \sim \frac{m\,(\theta_i) - m\,(\theta_0)}{\sqrt{v\,(\theta_0)}} + \frac{\sqrt{v\,(\theta_i)}}{\sqrt{v\,(\theta_0)}} \frac{F_{\theta_i} - m\,(\theta_i)}{\sqrt{v\,(\theta_i)}}, \tag{2.10}$$

i.e., as 'signal' $v(\theta_0)^{-1/2}(m(\theta_i) - m(\theta_0))$, which is non zero on the anomaly only, plus a standardized noise component $(F_{\theta_i} - m(\theta_i))/\sqrt{v(\theta_i)}$ which is scaled by a factor $v_i := \sqrt{v(\theta_i)/v(\theta_0)}$. In case of Gaussian observations with variance 1, one has $v_i \equiv 1$ and recovers the situation considered by Sharpnack and Arias-Castro (2016). Whenever the 'signal' part in (2.10) is strong enough, the anomaly should be detected. In the following, we will make this statement mathematically precise and also compare the multiscale testing procedure with an oracle procedure.

Considered alternatives Consider a given family $\left(Q_n^*\right)_{n\in\mathbb{N}}$ of hypercube anomalies $Q_n^* \subset [0,1]^d$ with Lebesgue measure $\left|Q_n^*\right| = a_n \in (0,1)$. The corresponding discretized anomalies $Q_n := I_n^d \cap nQ_n^* \subset I_n^d$ have size $|Q_n| \sim n^d a_n$. We will consider alternatives $K_{i,n}$ in (1.5) where $\theta^n \in \Theta^{n^d}$ s.t.

$$\theta_i^n = \theta_1^n \mathbb{I}_{Q_n} + \theta_0 \mathbb{I}_{Q_n^c}. \tag{2.11}$$

The parameters θ_1^n determine the total strength of the anomaly, which is given by

$$\mu^n(Q_n) := \sqrt{|Q_n|}\frac{m(\theta_1^n) - m(\theta_0)}{\sqrt{v(\theta_0)}}.$$

Clearly, any anomaly with fixed size or strength can be detected with asymptotic probability 1. Therefore, we will consider vanishing anomalies in the sense that

$$a_n \searrow 0, \qquad \mu^n(Q_n) \nearrow \infty, \qquad \text{as} \qquad n \to \infty. \tag{2.12}$$

Furthermore, we will restrict ourselves to parameters θ_1^n in (2.11) which yield uniformly bounded variances and uniform sub-exponential tails, this is

$$\mathbb{E}\left[\exp\left(sY\right)\right] \leq C \text{ for all } 0 \leq s \leq t \text{ and } \theta \in \{\theta_0\} \cup \bigcup_{n\in\mathbb{N}}\left\{\theta_1^n\right\}, \tag{2.13}$$

$$\underline{v} \leq v_i = \sqrt{\frac{v(\theta_1^n)}{v(\theta_0)}} \leq \bar{v} \qquad \text{for all} \qquad i \in I_n^d, n \in \mathbb{N} \tag{2.14}$$

for $Y \sim F_\theta$ with constants $t > 0, C > 0$ and $0 < \underline{v} < \bar{v} < \infty$.

In the case of Gaussian observations with variance σ^2, (2.13) and (2.14) are obviously satisfied, for a Poisson field this means that the intensities are bounded away from zero and infinity.

Oracle and multiscale procedure Recall that Q^* is the set of all hypercubes in $[0,1]^d$ (cf. Example 2.1.3) and Q^n its discretization (cf. (1.4)).

If the size a_n of the anomaly is known, but its position is unknown, one would naturally restrict the set of candidate regions to $\mathcal{R}_{\mathrm{O}}^* := \{Q^* \in Q^* \mid |S^*| = a_n\}$ and consequently scan only over (cf. (1.4))

$$\mathcal{R}_n^{\mathrm{O}} := \{Q \subset I_n^d \mid Q = I_n^d \cap nQ^* \text{ for some } Q^* \in \mathcal{R}_{\mathrm{O}}^*\},$$

as for the true anomaly $Q^* \in \mathcal{R}_{\mathrm{O}}^*$ its discretized version Q_n also satisfies $Q_n \in \mathcal{R}_n^{\mathrm{O}}$. This gives rise to an oracle test, which rejects whenever $T_n\left(Y, \theta_0, \mathcal{R}_n^{\mathrm{O}}, v\right) > q_{1-\alpha,n}^{\mathrm{O}}$ where $q_{1-\alpha,n}^{\mathrm{O}}$ is the $1 - \alpha$ quantile of $M_n\left(R_n^{\mathrm{O}}, v\right)$ as in (1.10). Similar as in Theorem 2.1.5 one can show that this quantile sequence ensures the oracle test to have asymptotic level α. The asymptotic power of this oracle test can be seen as a benchmark for any multiscale test.

To obtain a competitive multiscale procedure, let us choose a sequence r_n satisfying (1.23) with $\gamma = 12$ and let us furthermore assume that $r_n = o\left(n^d a_n\right)$, as otherwise the multiscale procedure will never be able to detect the true anomaly (since it is not contained in the set of candidate regions which we scan over). Since now position and size of the anomaly are unknown, we consider all sets in $\mathcal{R}_{\mathrm{MS}}^* = Q^*$ as candidate regions and consequently scan over

$$\mathcal{R}_{n|r_n}^{\mathrm{MS}} := \{Q \subset I_n^d \mid Q = I_n^d \cap nQ^* \text{ for some } Q^* \in Q^* \text{ and } |Q| \geq r_n\}.$$

Clearly the true anomaly Q^* satisfies $Q^* \in \mathcal{R}_{\mathrm{MS}}^*$ and, as $r_n = o\left(n^d a_n\right)$ its discretized version Q_n also satisfies $Q_n \in \mathcal{R}_{n|r_n}^{\mathrm{MS}}$. This gives rise to a multiscale test which rejects whenever $T_n\left(Y, \theta_0, \mathcal{R}_{n|r_n}^{\mathrm{MS}}, v\right) > q_{1-\alpha,n}^{\mathrm{MS}}$, where $q_{1-\alpha,n}^{\mathrm{MS}} := q_{1-\alpha}^{M_n\left(\mathcal{R}_{n|r_n}^{\mathrm{MS}}, v\right)}$ is the $1 - \alpha$ quantile of $M_n\left(\mathcal{R}_{n|r_n}^{\mathrm{MS}}, v\right)$ as in (1.10). Theorem 2.1.5 ensures that the multiscale test has asymptotic level α.

Now, due to Theorem 2.1.4 $q_{1-\alpha}^* := q_{1-\alpha}^{M(Q^*, v)} < \infty$ whenever $v \geq V_{Q^*} = 1$, and we have

$$q_{1-\alpha,n}^{\mathrm{O}} \leq q_{1-\alpha,n}^{\mathrm{MS}} \leq q_{1-\alpha}^* < \infty$$

for all $n \in \mathbb{N}$ and $v \geq 1$.

Asymptotic power We now show that the multiscale procedure described above, which requires no a priori knowledge on the scale of the anomaly, asymptotically detects the same anomalies with power 1 as the oracle benchmark procedure for a known scale. Hence, the penalty choice to calibrate all scales as in (1.7) (where $\mathcal{R}^* = Q^*$), renders the adaptation to all scales for free, at least asymptotically. This can be seen as a structural generalization of (Sharpnack and Arias-Castro, 2016, Theorems 2 and

4), as under the alternative the whole distribution in (1.1) and not just its mean might change. Also the power considerations in (Proksch et al., 2018) restrict to this simpler case. We show the following theorem.

Theorem 2.1.9. *In the setting described above, let $a_n \searrow 0$ be a sequence of scales such that $(\log n)^{12} / n^d = o(a_n)$ as $n \to \infty$. Denote by*

$$F\left(x, \mu, \sigma^2\right) := \Phi\left(-\frac{x+\mu}{\sigma}\right) + \Phi\left(\frac{\mu-x}{\sigma}\right), \qquad x \geq 0$$

the survival function of a folded normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, where Φ is the cumulative distribution function of $\mathcal{N}(0,1)$. Let furthermore $v \geq V_{Q^*} = 1$. If (2.12) is satisfied, then the following is true:

(a) The single scale procedure has asymptotic power

$$\mathbb{P}_{\theta^n}\left[T_n\left(Y, \theta_0, \mathcal{R}_n^O, v\right) > q_{1-\alpha,n}^O\right]$$
$$= \alpha + (1-\alpha)F\left(q_{1-\alpha,n}^O + \sqrt{2v \log\left(\frac{1}{a_n}\right)}, n^{d/2}\sqrt{a_n}\frac{m\left(\theta_1^n\right) - m\left(\theta_0\right)}{\sqrt{v\left(\theta_0\right)}}, \frac{v\left(\theta_1^n\right)}{v\left(\theta_0\right)}\right) + o(1).$$

(b) If $a_n = o\left(n^{\alpha-d}\right)$ with $\alpha > 0$ sufficiently small, then the multiscale procedure has asymptotic power

$$\mathbb{P}_{\theta^n}\left[T_n\left(Y, \theta_0, \mathcal{R}_{n|r_n}^{MS}, v\right) > q_{1-\alpha,n}^{MS}\right]$$
$$\geq \alpha + (1-\alpha)F\left(q_{1-\alpha,n}^{MS} + \sqrt{2v \log\left(\frac{1}{a_n}\right)}, n^{d/2}\sqrt{a_n}\frac{m\left(\theta_1^n\right) - m\left(\theta_0\right)}{\sqrt{v\left(\theta_0\right)}}, \frac{v\left(\theta_1^n\right)}{v\left(\theta_0\right)}\right) + o(1).$$

Remark 2.1.10. *In (Sharpnack and Arias-Castro, 2016) a similar result in case of Gaussian observations is shown. We note that the condition that $a_n = o\left(n^{\alpha-d}\right)$ with $\alpha > 0$ sufficiently small is missing there. However, it is necessary for the proof to work. In (Proksch et al., 2018) it suffices to assume $a_n \searrow 0$, as the maximum can be explicitly controlled due to its Gumbel-limit.*

The above Theorem allows us to explicitly describe those anomalies which will be detected with asymptotic power 1:

Corollary 2.1.11. *Under the setting of this section and the Assumptions of Theorem 2.1.9 any such anomaly is detected with asymptotic power 1 either by the single scale*

or the multiscale testing procedure if and only if

$$\frac{\sqrt{2v \log\left(\frac{1}{a_n}\right) v\left(\theta_0\right)} - n^{d/2} \sqrt{a_n} \left|m\left(\theta_1^n\right) - m\left(\theta_0\right)\right|}{\sqrt{v\left(\theta_1^n\right)}} \to -\infty \qquad (2.15)$$

as $n \to \infty$.

Example 2.1.12.   *1. In case of Gaussian observations $Y_i \sim \mathcal{N}\left(\Delta_n \mathbb{I}_{Q_n}, \sigma^2\right)$ with variance σ^2, where the baseline mean is 0 and where Δ_n is the size of the anomaly, this yields detection if and only if*

$$|\Delta_n| \, n^{d/2} \sqrt{a_n} \gtrsim \sigma \sqrt{2v \log\left(\frac{1}{a_n}\right)} \qquad as \qquad n \to \infty.$$

If we calibrate the statistic with the packing number $v = V_{Q^*} = 1$ (cf. Example 2.1.3), then this coincides with the well known asymptotic detection boundary for hypercubes, see e.g., (Arias-Castro et al., 2005; Frick et al., 2014) for $d = 1$, (Butucea and Ingster, 2013) for $d = 2$, or (Kou, 2017) for general d.

2. For Bernoulli r.v.'s $Y_i \sim Ber\left(p_0 \mathbb{I}_{Q_n^c} + p_n \mathbb{I}_{Q_n}\right)$ with $p_0, p_n \in (0, 1)$ s.t. $p_0 + p_n \leq 1$, condition (2.15) reads as follows:

$$\frac{\sqrt{2v p_0(1 - p_0) \log\left(\frac{1}{a_n}\right)} - n^{d/2} \sqrt{a_n} \, |p_n - p_0|}{\sqrt{p_n(1 - p_n)}} \to -\infty.$$

The minimax detection rate is unknown in this case to best of our knowledge.

3. For a Poisson field $Y_i \sim Poi\left(\lambda_0 \mathbb{I}_{Q_n^c} + \lambda_n \mathbb{I}_{Q_n}\right)$ with $\lambda_0, \lambda_n > 0$, Theorem 2.1.9 and Corollary 2.1.11 can be applied only if λ_n is a bounded sequence. In this case, (2.15) reduces to

$$\frac{\sqrt{2v\lambda_0 \log\left(\frac{1}{a_n}\right)} - n^{d/2} \sqrt{a_n} \, |\lambda_n - \lambda_0|}{\sqrt{\lambda_n}} \to -\infty.$$

Again, the minimax detection rate is unknown in this case to best of our knowledge.

2.1.4   Auxiliary results

Our results rely heavily on a coupling result which allows us to replace the maximum over partial sums of standardized natural exponential family (NEF) r.v.'s by a maximum over a corresponding Gaussian version. This can be obtained from recent results by Chernozhukov et al. (2014) as soon as certain moments can be controlled, which is the purpose of the following two Lemmata. In the following, the letter $C > 0$ denotes some absolute constant, which might change from line to line.

We start with controlling the maximum of powers of uniformly sub-exponential random variables, which will be an essential tool in the proof of the Coupling result (Theorem 2.1.15).

Lemma 2.1.13. *Let W_i, $i = 1, 2, \ldots$ be independent sub-exponential random variables s.t. there exist $k_1 > 1$ and $k_2 > 0$ s.t.*

$$\mathbb{P}\left[|W_i| > t\right] \le k_1 \exp(-k_2 t) \tag{2.16}$$

for all i. Then for all $m \in \mathbb{N}$ there exists a constant C, s.t. for all $N \ge 2$

$$\mathbb{E}\left[\max_{1 \le i \le N} |W_i|^m\right] \le C \left(\log N\right)^m.$$

It is well-known that the above bound can be improved for sub-Gaussian random variables to

$$\mathbb{E}\left[\max_{1 \le i \le N} |X_i|\right] \le C \sqrt{\log N}. \tag{2.17}$$

Next we will show that the maximum over the partial sum process of independent random variables can be bounded by the maximum over the corresponding Gaussian version. The latter can be controlled as in (2.17) by exploiting the fact that a maximum over dependent Gaussian random variables is always bounded by a maximum over corresponding independent Gaussian random variables (see e.g., Šidák, 1967) and hence,

$$\mathbb{E}\left[\max_{I \in \mathcal{I}} \frac{|X_I|}{\sqrt{|I|}}\right] \le C \sqrt{\log\left(\#\left(\mathcal{I}\right)\right)} \tag{2.18}$$

with $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ and $X_I := \sum_{i \in I} X_i$. This allows us to prove the following:

Lemma 2.1.14. *Let $(Z_i)_{i=1,\ldots,N}$ be independent random variables with $\mathbb{E}[Z_i] = 0$ and denote $Z_I := \sum_{i \in I} Z_i$. If $\mathcal{I}$ is an arbitrary index set of sets $\{I\}_{I \in \mathcal{I}}$, then there exists a*

constant $C > 0$ independent of $\mathcal{I}$ s.t.

$$\mathbb{E}\left[\max_{I \in \mathcal{I}} \frac{|Z_I|}{\sqrt{|I|}}\right] \leq C \sqrt{\log\left(\#\left(\mathcal{I}\right)\right)} \, \mathbb{E}\left[\max_{1 \leq i \leq N} |Z_i|\right].$$

With the help of these two lemmata, the following coupling result can be shown:

Theorem 2.1.15 (Coupling). *Let $Z_i, i \in I_n^d$ be independent random variables with $\mathbb{E}[Z_i] = 0$ and $\mathbb{V}[Z_i] = 1$, such that (2.16) is satisfied for all i with absolute constants $k_1 > 1$ and $k_2 > 0$. Suppose further that $v_i \in [\underline{v}, \overline{v}], i \in I_n^d$ are given with $0 < \underline{v} \leq \overline{v} < \infty$ independent of i and n, and that $X_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1), i = 1, \dots, n^d$, and that $\mathcal{R}_n$ is a system of sets such that inequality (2.2) holds. Then*

$$\max_{\substack{R \in \mathcal{R}_n: \\ |R| \geq r_n}} |R|^{-1/2} \sum_{i \in R} v_i Z_i - \max_{\substack{R \in \mathcal{R}_n: \\ |R| \geq r_n}} |R|^{-1/2} \sum_{i \in R} v_i X_i = O_{\mathbb{P}}\left(\left(\frac{\log^{10}(n)}{r_n}\right)^{1/6}\right).$$

2.2 The case of unknown baseline parameter for Poisson observations

2.2.1 Setting and Assumptions

In this section we restrict the setting to a *d*-dimensional field of Poisson random variables as in (1.12). As seen in Example 2.1.7, the theory of the previous section may be applied to these random variables as well.

The new challenging part of considering the statistic T_n^ω as in (1.14) is the change of scale calibration terms, in particular, as the multiplicative factor $\widetilde{\omega}$ is concerned. To control the supremum of (1.16) we have to restrict our system of regions $\mathcal{R}^*$ suitably.

Assumption 2 (Cardinality of $\mathcal{R}_n$). *There exists constants $c_1, c_2 > 0$ such that*

$$\#\left(\mathcal{R}_n\right) \leq c_1 n^{c_2}. \tag{2.19}$$

where # denotes the number of elements.

This assumption on $\mathcal{R}^*$ is less restrictive than in the previous section due to the smaller range of scales (recall (2.2)). The scale calibration terms are assumed to be of a certain decreasing order.

Assumption 3 (Average Hölder condition)**.** *All regions R in the set $\mathcal{R}^*$ satisfy the average Hölder condition, that is*

$$\int |\mathbb{I}_R(t-z) - \mathbb{I}_R(s-z)|^2 \, dz \leq K\|t-s\|_{l^2}^{2\zeta},$$

for some $\zeta \in [1/2, 1]$ and some constant K.

Example 2.2.1. *Whenever $\mathbb{I}_R$ is Hölder continuous, Assumption 3 is fulfilled.*

Assumption 4. *Suppose $\omega, \widetilde{\omega} : [1, n^d] \to [1, \infty)$ are decreasing functions and that there exist $\alpha, \widetilde{\alpha} > 0$ and $\beta, \widetilde{\beta} \in \mathbb{R}$ s.t.*

$$\omega(r) \lesssim \left(\log \frac{n^d}{r}\right)^\alpha, \qquad |\omega'(r)| \lesssim \left(\log \frac{n^d}{r}\right)^\beta \frac{1}{r},$$

$$\widetilde{\omega}(r) \lesssim \left(\log \frac{n^d}{r}\right)^{\widetilde{\alpha}}, \qquad |\widetilde{\omega}'(r)| \lesssim \left(\log \frac{n^d}{r}\right)^{\widetilde{\beta}} \frac{1}{r}, \quad r \in [1, n^d].$$

Remark 2.2.2.   *1. For the Gaussian approximation [Theorem 2.2.4] Assumptions 2 and 4 are sufficient.*

2. *The average Hölder assumption 3 is needed to prove the weak limit theorem. The proof uses a different method which is applicable if the largest scales are cut off. For the same reason the VC-dimension Assumption 1(a) can be relaxed to the condition on the cardinality of $\mathcal{R}_n$ in Assumption 2.*

Example 2.2.3.   *1. Following (Dümbgen and Spokoiny, 2001), consider*

$$\widetilde{\omega} \equiv 1,$$
$$\omega(|R|) = \sqrt{2\nu(\log(n^d/|R|)) + 1}, \tag{2.20}$$

where ν depends on the complexity of the candidate region. These terms fulfil Assumption 4 with $\alpha = \frac{1}{2}, \beta = -\frac{1}{2}$, and $\widetilde{\alpha} = \widetilde{\beta} = 0$.

2. *Following (Proksch et al., 2018), consider*

$$\widetilde{\omega}(|R|) := \sqrt{2\log\left(\frac{(2\pi)^{-\frac{1}{2}}n^d}{|R|}\right)} + C_d \frac{\log\left(\sqrt{2\log\left(\frac{(2\pi)^{-\frac{1}{2}}n^d}{|R|}\right)}\right)}{\sqrt{2\log\left(\frac{(2\pi)^{-\frac{1}{2}}n^d}{|R|}\right)}},$$
$$\omega(|R|) \equiv \widetilde{\omega}(|R|), \tag{2.21}$$

where C_d depends on the dimension, the system of considered scales and the underlying smoothness of $\mathbb{I}_R$. These fulfil Assumption 4 with $\alpha = \widetilde{\alpha} = \frac{1}{2}$ and $\beta = \widetilde{\beta} = -\frac{1}{2}$.

2.2.2 Limit Theory

Theorem 2.2.4 (Gaussian approximation). *Let $Y_i, i \in I_n^d$ be a field of Poisson random variables and let $\mathcal{R}^*$ be a set of candidate regions satisfying Assumption 2 and let $(r_n)_n \subset (0, \infty)$ be a sequence such that (1.23) holds with*
$\gamma = 12 + 6\widetilde{\alpha} + 2\max\left(\frac{1}{2} + \max\left(\widetilde{\beta}, 0\right), \alpha + \max\left(\widetilde{\beta}, 0\right), \widetilde{\alpha} + \max(\beta, 0)\right)$ *and m_n s.t.*
(1.24) holds. Assume further that the scale calibrations $\omega, \widetilde{\omega}$ fulfil Assumption 4. Then under $\mathbb{P}_0$

$$T_n^\omega\left(Y, \mathcal{R}_n(r_n, m_n)\right) - M_n^\omega\left(\mathcal{R}_n(r_n, m_n)\right) = O_\mathbb{P}\left(\left(\frac{\log^\gamma(n)}{r_n \lambda_0^3}\right)^{1/12}\right) \tag{2.22}$$

as $n \to \infty$ with M_n^ω as in (1.15).

Theorem 2.2.5 (Weak $\mathbb{P}_0$ limit). *Let Y_i, $i \in I_n^d$ be a field of Poisson random variables, $\mathcal{R}^*$ satisfy Assumption 2, 3 and let $(r_n)_n, (m_n)_n$ be sequences such that (1.23) with γ as in Theorem 2.2.4 and (1.24) hold. Then we have under $\mathbb{P}_0$ that*

$$T_n^\omega\left(Y, \mathcal{R}_n(r_n, m_n)\right) \xrightarrow{\mathcal{D}} M^\omega\left(\mathcal{R}^*\right) \qquad as \qquad n \to \infty, \tag{2.23}$$

with M^ω as in (1.16).

The penalty term in (2.20), used in Section 2.1, would also result in an almost surely finite limit statistic, but due to the cut off of the largest scales, it would result in a degenerate limit, see (Schmidt-Hieber et al., 2013, Lemma C.1).Therefore the scale calibration as in (2.21) is more appealing, since we can then prove the following Theorem 2.2.7.

For this, we need the following definition.

Definition 2.2.6. *A continuous random variable X follows a* Gumbel distribution *with scaling parameter β and location parameter μ, if its cumulative distribution function is given by*

$$F(x; \mu, \beta) = \exp\left(-\exp\left(-\frac{x - \mu}{\beta}\right)\right).$$

Theorem 2.2.7. *Let $\mathcal{R}^*$ satisfy Assumptions 2 and 3. Let the scale calibration term be chosen as in (2.21).*

Then the statistic $M^\omega(\mathcal{R}^*)$ is almost surely finite. In particular, there exist $\underline{D}$ and $\overline{D}$ such that two Gumbel distributions with scaling parameter $\beta = 1$ and location parameter $\mu = \log(\underline{D})$ and $\mu = \log(\overline{D})$, respectively, bound the distribution from below and above, i.e.,

$$e^{-\underline{D}e^{-t}} \leq \lim_{n \to \infty} \mathbb{P}\left[M^\omega(\mathcal{R}^*) \leq t\right] \leq e^{-\overline{D}e^{-t}} \qquad \forall t \in \mathbb{R}.$$

Simulation study and real data example

This chapter contains a simulation study as well as a real data example with STED measurements (provided by René Siegmund from the Laser-Laboratorium Göttingen e.V.). Therefore, the focus is on Poisson random fields.

3.1 Implementation of our multiscale procedure

The implementation is based on a code used in (Proksch et al., 2018), which was kindly provided to me. In their notation we use $\Phi_i = \mathbb{I}_{B_i}$ with boxes B_i. For a run-time analysis see (Proksch et al., 2018). Using this code to calculate $\overline{Y}_R$, we implemented our hypothesis testing procedure.

To construct the test the dimension d, the size of the image, and the candidate regions, i.e., the scales we consider, need to be specified. We consider images in $d = 2$ dimensions of size 512×512. For the test $n = 96$ is the largest feasible, if all scales are included, but since we are just interested in a smaller range of scales, $n = 512$ is manageable. We consider as candidate regions all boxes of size $4 \times 4, 4 \times 5, \ldots, 4 \times 10, 5 \times 4, \ldots, 10 \times 10$ pixels (49 different scales). The implementation is then done using the Fast Fourier Transform (FFT). At first, out of the fixed test setup (d, n and candidate regions), Gaussian approximations were simulated. Performing the multiscale test on a data set now requires computing the quantiles of these Gaussian approximations and computing the statistic T_R and check for regions, where the statistic exceeds the threshold.

3.2 Simulation study

We simulated a Poisson random field with $a = 20$ anomalies.

First, we randomly selected the underlying signal (the so-called *groundtruth*) by choosing randomly 20 coordinate points in a 512×512 pixel image. Around these 20 points, we randomly chose marker positions with number of markers uniformly chosen in the

interval [1, 8]. This results in a field with value ≥ 1 at a marker position and zero elsewhere. An example is shown in Figure 3.1. Notice that, due to the randomness, there can be more than one marker on the same coordinate point.
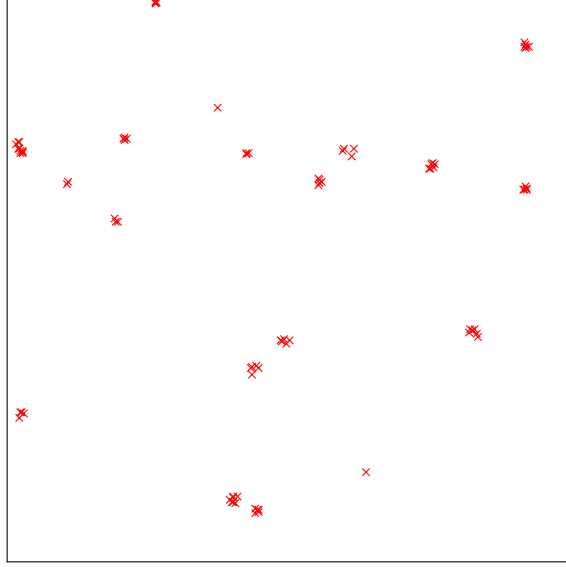

Figure 3.1: Example of simulated marker locations where each cross represents a marker.

Based on the marker locations (groundtruth) the signal λ in (1.12) is chosen as
$\lambda = t \cdot \big(\text{intensity} + l \cdot \text{ones}(512, 512)\big)$ with observations

$$Y = \text{Poi}\left(t \cdot \big(\text{intensity} + l \cdot \text{ones}(512, 512)\big)\right),$$

where t denotes the dwell time, l stands for the baseline intensity and intensity stands for the groundtruth convoluted with a Gaussian kernel. This is in good accordance with STED data since the point spread function of the microscope is often approximated by a Gaussian density for simplicity, see e.g., (Bertero et al., 2009). These parameters were chosen such that the resulting field visually coincides with the measurements (e.g., in Figure 3.2: $l = 0.015$ and $t = 5,\ 15,\ 50$).

3.2.1 Comparison to Gaussian Test (with pre-transformation of the Poisson data)

In this subsection we will illustrate the superior performance of our test tailored to the Poisson setting (from now on called *Poisson test*) in contrast to the test for Gaussian observations which uses as input the data pre-processed by Anscombe's transformation (from now on called *Anscombe-based test*).
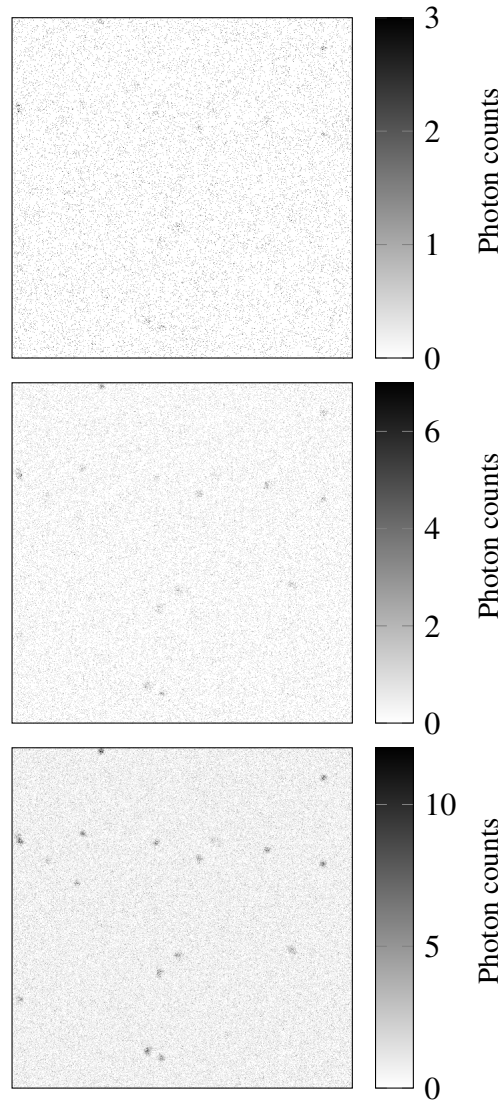
Figure 3.2: Simulated data Y. Top: with $l = 0.015$ and $t = 5$. Middle: with $l = 0.015$ and $t = 15$. Bottom: with $l = 0.015$ and $t = 50$.

Anscombe proposed $x \mapsto 2 \cdot \sqrt{x + \frac{3}{8}}$ in (Anscombe, 1948) as variance stabilizing transformation for Poisson data, and by Lemma 1 of (Brown et al., 2010) for $Y \sim$ Poi(λ), $2 \cdot \sqrt{Y + \frac{3}{8}} - 2 \cdot \sqrt{\lambda} - \frac{1}{8\sqrt{\lambda}}$ is approximately $\sim \mathcal{N}(0, 1)$, since then the mean is zero and variance one.

Figure 3.3 illustrates the performance of both tests in the before-mentioned settings (from Figure 3.2). For small dwell time $t = 5$, the Anscombe-based test finds less anomalies compared to the Poisson test. For $t = 15$, both tests find the same amount of anomalies, but the Anscombe-based test finds them on larger scales. This means that the Poisson test localizes much better. This "localization"-property will be also illustrated in the STED-data-example, see Figure 3.9. For $t = 50$ both test work well.

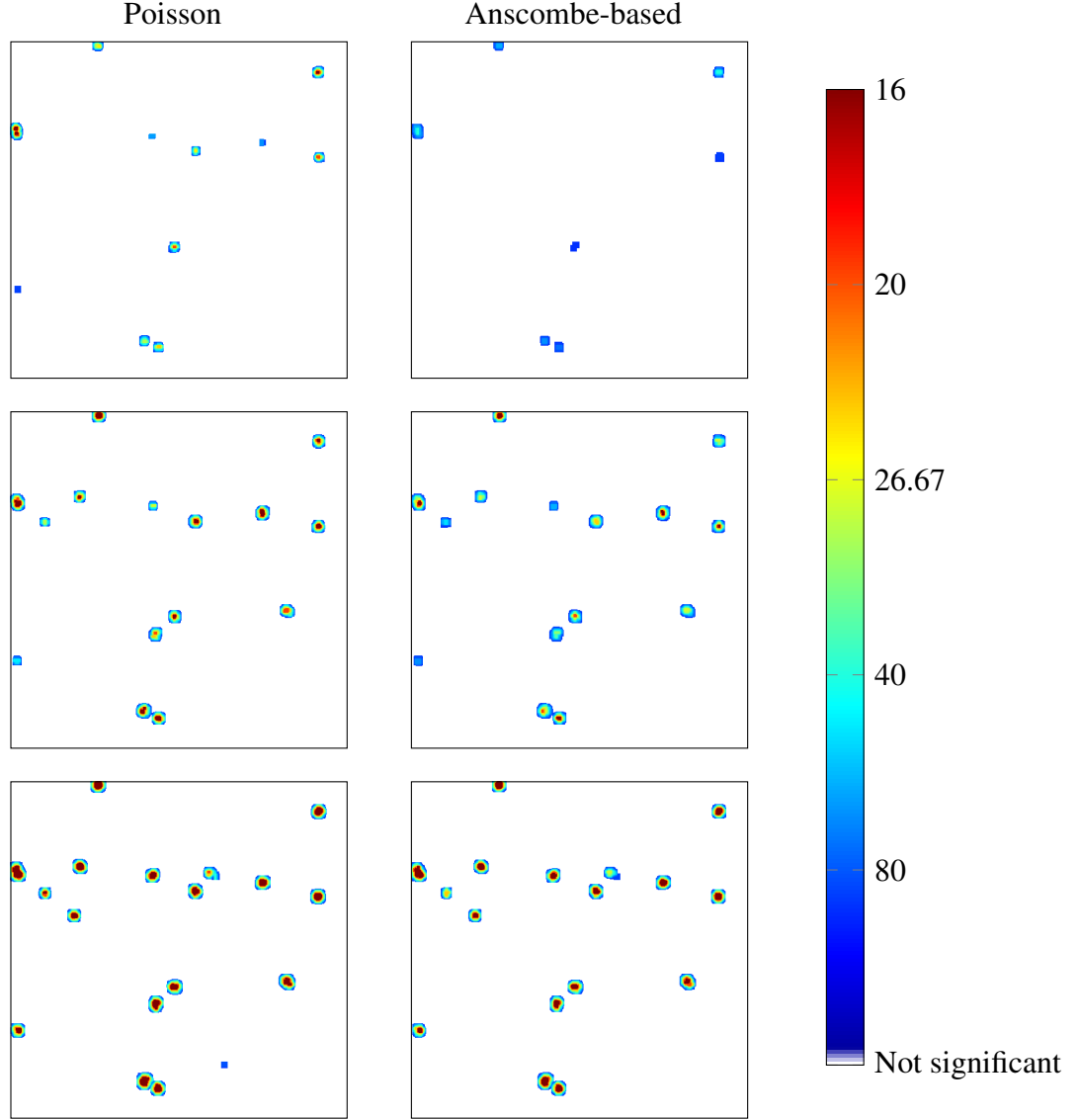Notice that the Poisson test also detects the anomaly, which just has one marker, see bottom right corner.

Figure 3.3: Test results. Top: $l = 0.015$, $t = 5$, middle: $l = 0.015$, $t = 15$, bottom: $l = 0.015$, $t = 50$.

3.2.2 Error analysis

This subsection contains the results of a simulation study to derive the type I error as well as the ratio of missed markers to all markers.

In hypothesis testing a type I error means the incorrect rejection of a true null hypothesis. In our setting this means that we put a box indicating an anomaly although there is no marker at that location. Keep in mind that due to the theory, the type I error is controlled, since the method controls the FWER asymptotically, see (1.6).

Type I error The worst case for the type I error is the constant field with no anomaly. Therefore, we simulated a Poisson field with no anomalies (just baseline intensity) and counted how many wrongly detected anomalies appeared. The type I error is depicted in Figure 3.4 for $t = 1$, different values of l and $\alpha = 0.1$ (horizontal line) coming from 1000 trials. It turns out, that for small values of l, the empirical level exceeds the nominal level, which might be due to a too small n to be in the asymptotic regime in this case. However, already for moderate values of l, the test keeps its level quite stable.
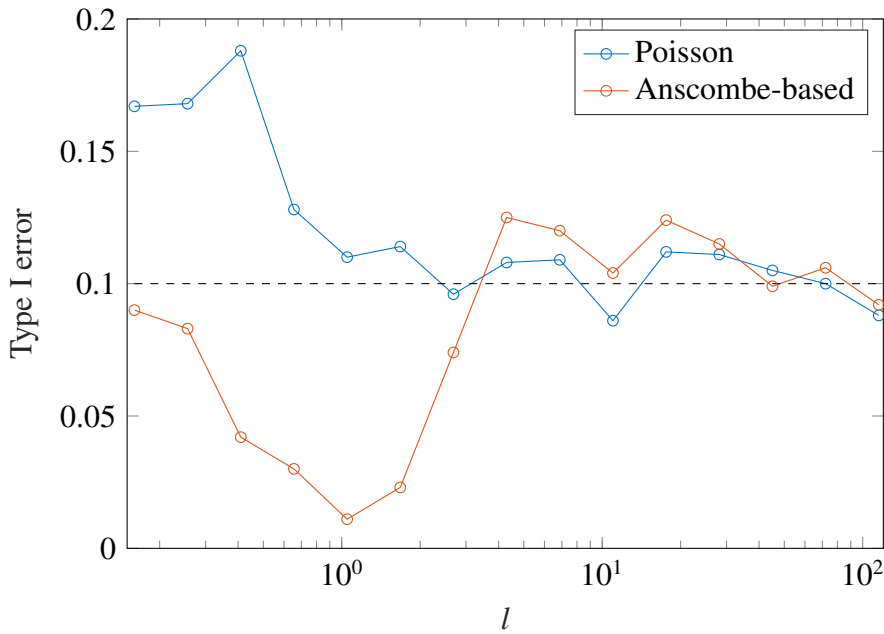

Figure 3.4: Type I error. Dashed line illustrates the nominal level.

Missed markers For the number of missed markers out of all markers we fixed a groundtruth and a baseline intensity (here: groundtruth from Figure 3.1 containing 86 markers and $l = 0.015$ as in Figure 3.2) and evaluated the ratio of missed markers and the total number of markers by running 10000 trials and dwell times t ranging from 1 to 10. In Figure 3.5 you see the superior performance of the Poisson test in terms of detection power. For both tests the error gets smaller with larger t, i.e., the larger the dwell time, the more anomalies are detected.

3.3 STED data example

The data in this section comes from a STED microscope and shows measurements of sparsely distributed 40nm large fluorescent crimson beads. Therefore, the hot spots
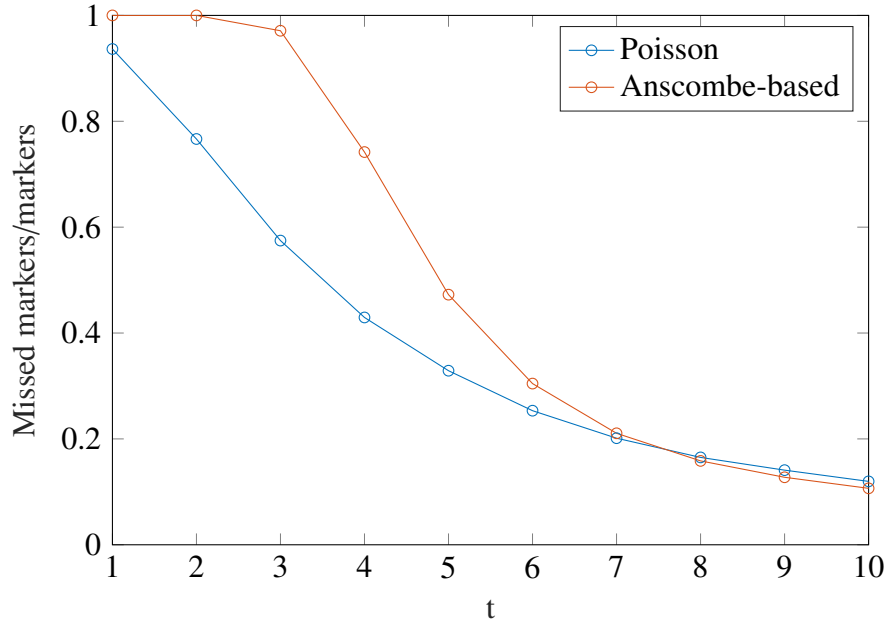
Figure 3.5: Ratio of the number of missed markers to the total amount of markers for different dwell times t.

will be around 120 nanometers. This results in a Poisson field with intensity proportional to the exposure time, recall Section 1.1.1 and equation (1.17).

There, we already sketched the idea of applying the hypothesis testing procedure to segment the image into regions which should be further investigated and those which are "uninteresting". Underlying this approach is that high energy of STED light could cause sample damage. RESCue-STED (Staudt et al., 2011) avoids unnecessary excitation and de-excitation cycles and thus reduces photobleaching of any fluorescent marker. In particular, RESCue-STED has two workings: shut-off the excitation and STED laser where there was no signal for the first, e.g., α percentage of the pixel dwell time and additionally where there has already been a sufficiently strong signal. We will tackle with our hypothesis testing procedure the first action. Applying the approach on an image with short dwell time $t_0 \leq t$, we will identify the interesting regions. After building segments around the detected regions, just these can be measured with the full dwell time.

The first data-set is measured with exposure time $t_0 = 5\mu s$, i.e., resulting in $\text{Poi}(t_0 \cdot \lambda_i)$, the so-called *limited data*. Another data-set is derived by using an exposure time of $t = 100\mu s$, the so-called *full data*. The full data can be used for validating our results. Following Section 2.2 we will pre-estimate the baseline parameter λ_0 by the sample mean (total average). The candidate regions will be the same as in the simulation section 3.2, i.e., boxes of size $4 \times 4, 4 \times 5, \ldots, 4 \times 10, 5 \times 4, \ldots, 10 \times 10$ pixels (1 pixel

$= 20 \times 20$ nm).

Performing the hypothesis testing procedure (Poisson test) using the limited data, we derive a spatial visualization of the rejected regions, i.e., by one-to-one correspondence of testing and confidence statements, regions which contain with 90% confidence an anomaly.

Figure 3.6 shows the experimental data. The top row displays the limited data as well as an enlargement of a specific region. The middle row depicts the test results, again with the same area enlarged. As a kind of justification, the last row shows the full data. The test results are in good correspondence with the full data, especially when one looks at the enlarged area. The left bottom corner of the limited data is not very informative, but despite this, the multiscale procedure indicates an anomaly, which also appears in the full data.

Figure 3.7 you see in the first row the test results for all regions, large and small, and in the second row the test results of the smallest significant regions, i.e., all large boxes are removed if they contained a smaller one. This nicely illustrates that our test locates the anomalies very accurately.

Figure 3.8 shows the Anscombe-based test for the STED-data as well as the test results where just the smallest significant regions are depicted. In contrast to Figure 3.7 the reduction does not yield a more accurate localization.

This benefit of the Poisson test is also illustrated in Figure 3.9. It nicely illustrates the additional information our test yields when compared to the Gaussian one. The Poisson test localizes much better which can be already seen from looking at the color code of both test results.
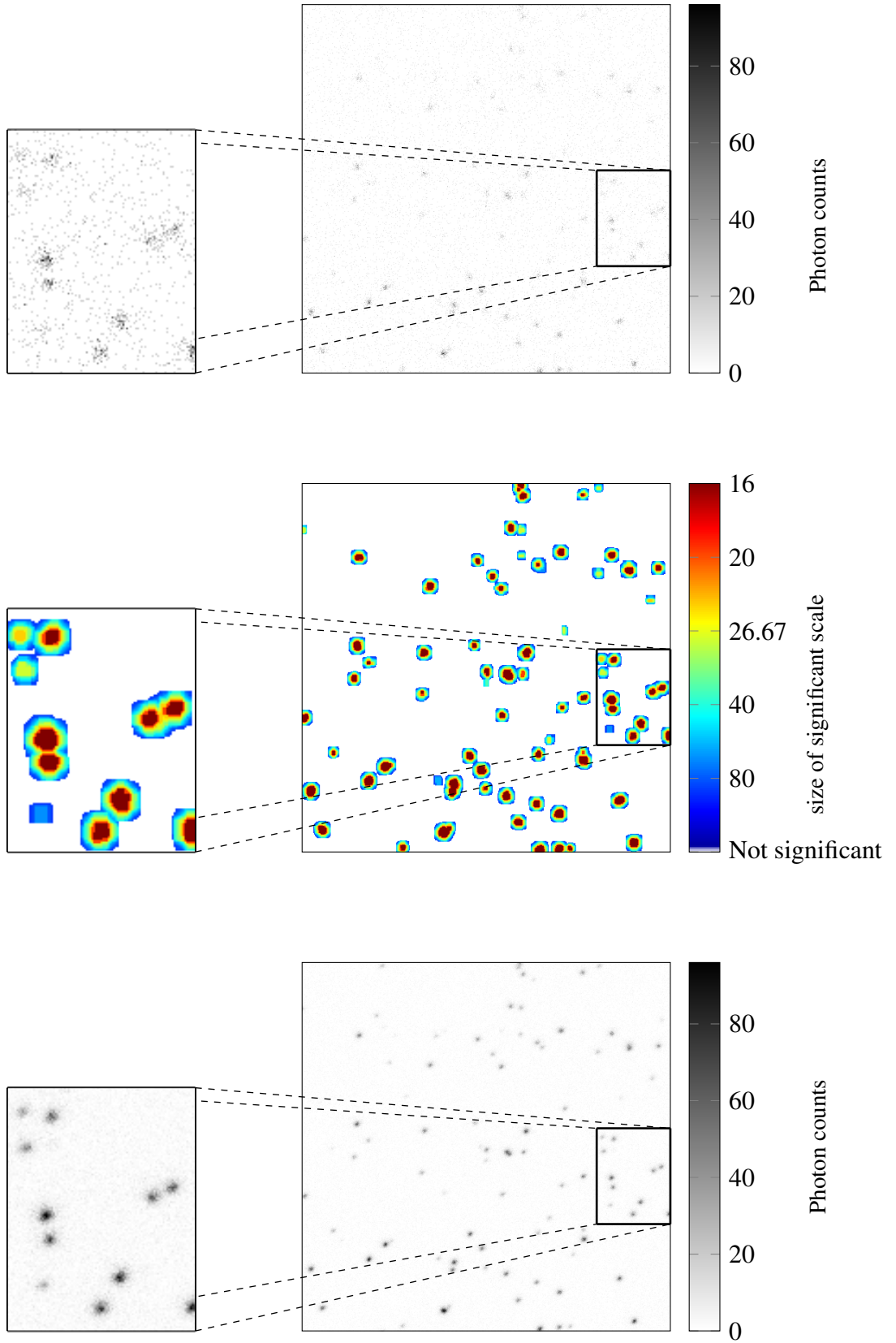
Figure 3.6: Top row: Limited Data ($t = 5$) and zoomed in region (102×132 pixels). Middle row: Spacial visualization of the rejected test (with 90% confidence each colored region contains an anomaly) and corresponding zoomed region. Bottom row: Full data ($t = 100$) and zoomed in region. The FWER control ensures that with asymptotic probability at least 90 percent among the selected boxes there is no wrong detection.
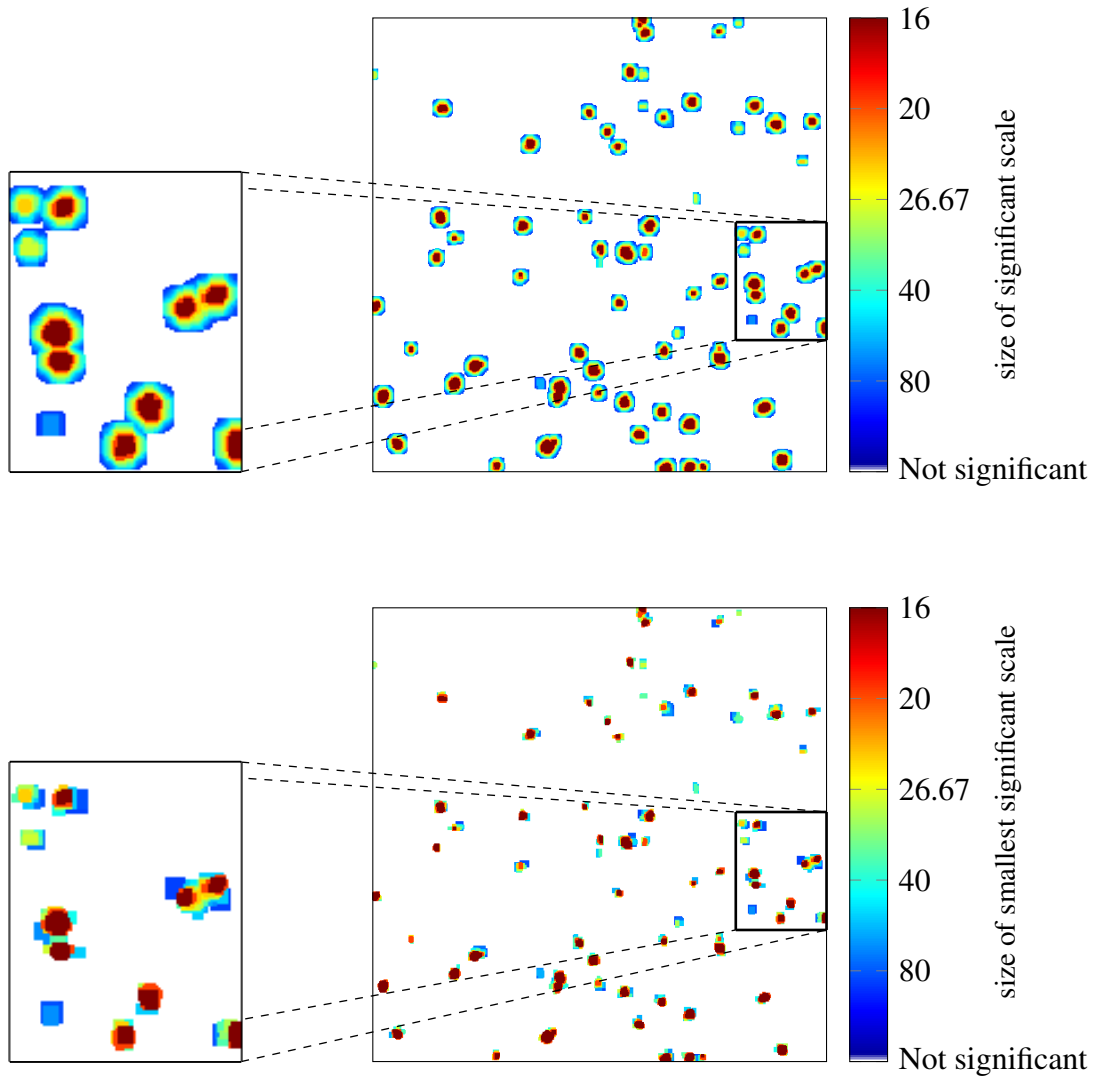
Figure 3.7: 90% significance maps for the data of Figure 3.6. In the first row the test results with all detected hot spots are depicted, below the test results after reduction of the boxes, i.e., where the regions are just the smallest ones which are significant. Reduced boxes: 3086 boxes left. Before there were 214420 boxes.
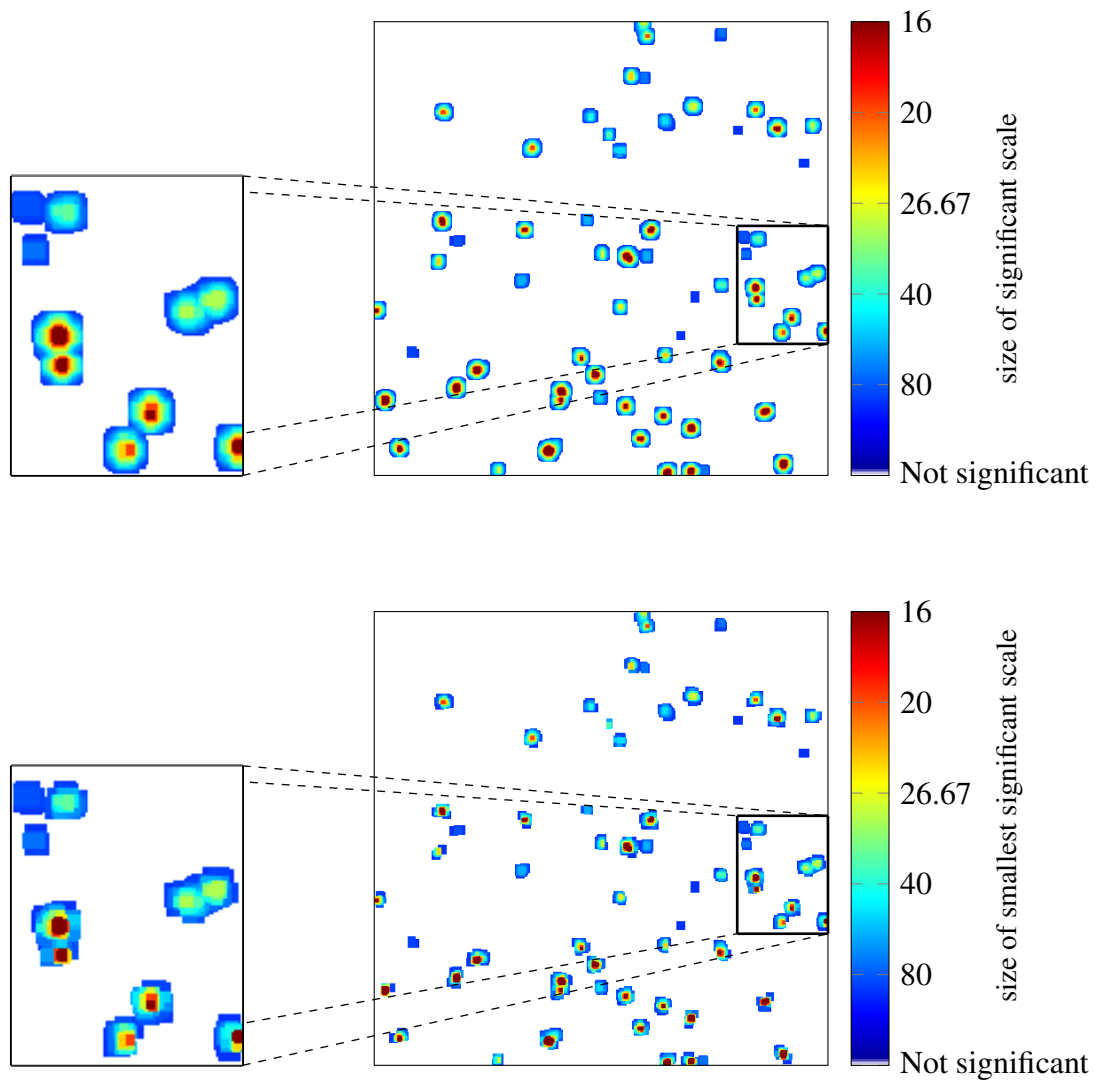
Figure 3.8: 90% significance maps for the data of Figure 3.6. In the first row you can see the test results coming from the Anscombe-based test with all detected hot spots. Below you can see the test, where the regions are just the smallest ones which are significant. Before 106693 significant boxes and afterwards 3333.
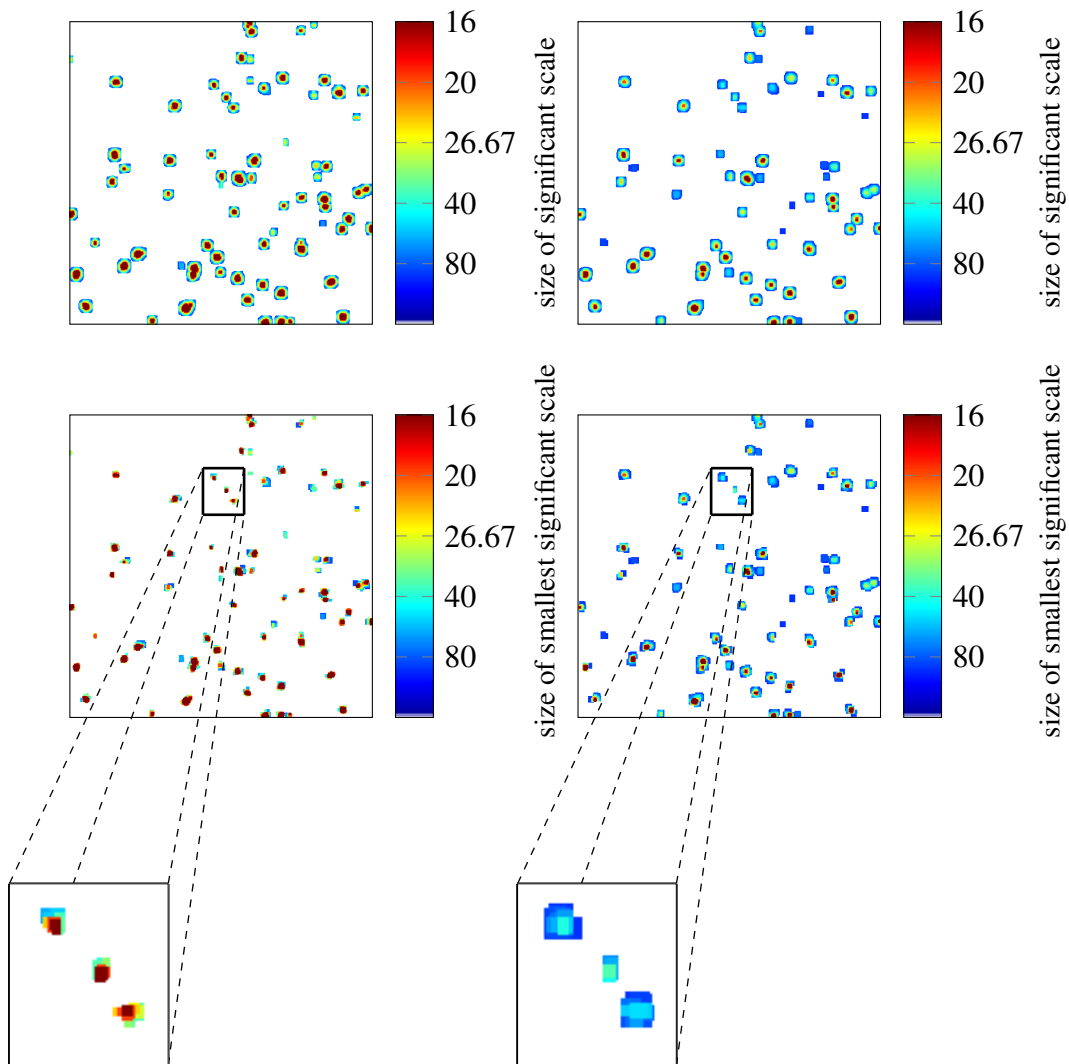
Figure 3.9: 90% significance maps for the data of Figure 3.6. In the top row: On the left you can see the results of the Poisson test and on the right the ones from the Gaussian test. Below you can see the corresponding results with reduced boxes. Compared to the Ancombe-based test, the Poisson test localizes much better.

CHAPTER 4

Discussion and Outlook

This thesis considered multiscale scanning in higher dimension in the setting of NEF. The first part was about finding anomalies in a d-dimensional field of independent random variables $\{Y_i\}_{i \in \{1,\ldots,n\}^d}$, each distributed according to a one-dimensional natural exponential family $\mathcal{F} = \{F_\theta\}_{\theta \in \Theta}$, where the underlying baseline parameter is known. We provide a unified methodology which controls the FWER. A Gaussian approximation (Theorem 2.1.5) as well as a weak limit theorem (Theorem 2.1.4) was derived. Further, we analyzed the power of our multiscale testing approach in the hypercube-case. As a result we showed that this procedure detects asymptotically the same anomalies with power 1 as the oracle benchmark procedure for a known scale. Theorem 2.1.9 can be seen as a generalization of Theorems 2 and 4 of (Sharpnack and Arias-Castro, 2016). Motivated by STED microscopy, we looked in a second step at a Poisson field with unknown baseline parameters. A challenging aspect was the consideration of arbitrary (up to some conditions) scale calibration terms $\omega, \widetilde{\omega}$. As before, we derived a Gaussian approximation (Theorem 2.2.4) and a weak limit theorem (Theorem 2.2.5).

In the following we discuss further areas which could be investigated.

Beyond constant intensities Considering more general function classes for the intensity is one interesting extension. Generalizing the limit law and the Gaussian approximation to a non-constant intensity field $\lambda_i = \lambda(i)$ would give the possibility to use these results in the context of image reconstruction, recall MIND in Section 1.1.2. First, the results were shown for a known, constant baseline parameter θ_0, second, the theory was extended to an unknown intensity λ_0, which was pre-estimated by the sample mean. Finally one would want to consider inhomogeneous λ_0. To this end, one can compare λ_0 on every box with $\overline{\lambda} = \frac{1}{|\text{box}|} \int_{\text{box}} \lambda_0 \, dx$, using the triangle inequality - similar as the comparison of $\overline{Y}$ and λ_0 in Proposition B.4.1. Thereby all estimates depend on the error between $\overline{\lambda}$ and λ_0, which scales with $\|\lambda_0'\|_{\infty,\text{box}} \times |\text{box}|$, using the mean value theorem. Therefore, this error can be controlled if the boxes are going to zero sufficiently fast.

Unknown parameter in the setting of NEF   As in the Section 2.2, where an unknown intensity for a Poisson field was assumed, one could extend this to NEF in a straightforward way. The arguments in the proofs of Section 2.2 do not rely on the Poisson setting, but on the coupling result which has already been shown for general NEF (Theorem 2.1.15).

Sample mean replaced by median   Another interesting aspect is the question if one can replace in the pre-estimation step the sample mean by the median for the unknown baseline parameter. This would require new techniques to prove a result equivalent to Proposition B.4.1.

Complexity of sets: VC-dimensions and related quantities

As explained in Remark 2.1.2 a finite VC-dimension is a standard way to control the complexity of a set-indexed process. Furthermore, the penalty-term (1.8) needs a bound on the complexity measured in terms of the packing number, namely the constant v. In the following the relationship between these quantities will be stated and covering numbers will be computed.

A.1 Calculation of packing numbers

Recall that the packing number $\mathcal{K}(\epsilon, \rho, \mathcal{W})$ was defined in Remark 2.1.2. In this appendix we estimate the packing numbers given in Example 2.1.3, which will be done by determining the covering number. The covering number $\mathcal{N}(\epsilon, \rho, \mathcal{W})$ of a subset $\mathcal{W} \subset \mathcal{R}^*$ w.r.t. a metric ρ is given by the minimal number of balls of radius $\epsilon > 0$ needed to cover $\mathcal{W}$ (cf. van der Vaart and Wellner, 1996, Def. 2.2.3). It is immediately clear that

$$\mathcal{N}(\epsilon, \rho, \mathcal{W}) \le \mathcal{K}(\epsilon, \rho, \mathcal{W}) \le \mathcal{N}\left(\frac{\epsilon}{2}, \rho, \mathcal{W}\right),$$

and hence it suffices to compute $\mathcal{N}\left((\delta u)^{1/2}, \rho^*, \{R \in \mathcal{R}^* \mid |R| \le \delta\}\right)$ with ρ^* as in (2.3) to show (2.4). By Theorem 2.6.7 of (van der Vaart and Wellner, 1996) one can bound the covering number of $\mathcal{W}$ using the VC-dimension of $\mathcal{W}$, that is

$$\mathcal{N}\left(\frac{\epsilon}{2}, \rho, \mathcal{W}\right) \le C\epsilon^{-v(\mathcal{W})},$$

where $v(\mathcal{W})$ is the VC-dimension of $\mathcal{W}$. But as pointed out in Remark 2.1.2 and Example 2.1.3, this bound can be improved, which will be illustrated now. In the following, we will use the notation used in Example 2.1.3.

In the case of hyperrectangles we have

Lemma A.1.1. *For any $\epsilon > 0$ there exists a constant C depending only on the dimen-*

sion d and on ϵ such that for all $u, \delta \in (0, 1]$ we have

$$\mathcal{K}\left((\delta u)^{1/2}, \rho^*, \{S \in \mathcal{S}^* : |S| \leq \delta\}\right) \leq Cu^{-(2d+\epsilon)}\delta^{-(2d-1+\epsilon)},$$

i.e., (2.4) is true with $k_1 = C, k_2 = 2d + \epsilon$ and $V_{\mathcal{S}^*} = 2d - 1 + \epsilon$.

Proof. We approximate the hyper-rectangles in $\mathcal{W} = \{S \in \mathcal{S}^* : |S| \leq \delta\}$ by hyper-rectangles with vertices in the lattice $\mathbb{L}_m := \left\{\frac{i}{m} \mid i = 0, \ldots, m\right\}^d$ where m will be specified later. The set of all hyper-rectangles with vertices in $\mathbb{L}_m$ and size $\leq \delta$ will be denoted by $\mathcal{W}'_m$. For $S \in \mathcal{W}$ denote the edge lengths by $k_1, \ldots, k_d$. Then $\prod_{j=1}^d k_j \leq \delta$ and $k_i \leq 1$. It is immediately clear that there exists an approximating hyper-rectangle $S' \in \mathcal{W}'_m$ such that

$$\begin{aligned}
\left(\rho^*\left(S, S'\right)\right)^2 &= |S \triangle S'| \\
&\leq 2\left(k_2 \cdot \ldots \cdot k_d + k_1 \cdot k_3 \cdot \ldots \cdot k_d + \ldots + k_1 \cdot \ldots \cdot k_{d-1}\right) \cdot \frac{1}{2m} \\
&\leq \frac{d}{m}.
\end{aligned} \tag{A.1}$$

Hence, we obtain $\rho^*(S, S') \leq (\delta u)^{1/2}$ if we choose $m := \frac{d}{\delta u}$. We now compute the cardinality of $\mathcal{W}'_{d/(\delta u)}$. First note that the number of possible left bottom vertices is bounded from above by $m^d = \#\mathbb{L}_m$. If we denote the edge lengths of $S' \in \mathcal{W}'_m$ by $l_1, \ldots, l_d$, we can find integers $i_1, \ldots, i_d$ such that $l_j = \frac{i_j}{m}$ and $\prod_{j=1}^d i_j \leq \delta m^d =: N$. Therefore, we obtain

$$\mathcal{N}\left((\delta u)^{1/2}, \rho^*, \{S \in \mathcal{S}^* : |S| \leq \delta\}\right) \leq \#\mathcal{W}'_m \leq m^d \cdot \#\mathcal{P}_N \tag{A.2}$$

with $\mathcal{P}_N := \left\{(i_1, \ldots, i_d) \in \mathbb{N}^d \mid \prod_{j=1}^d i_j \leq N\right\}$.
To compute $\#\mathcal{P}_N$, we employ Minkowski's theorem (cf. Cassels, 1997, Sec. III.2.2), which ensures that the Lebesgue volume $\Delta_d(N)$ of $\left\{(x_1, \ldots, x_d) \in [1, N]^d \mid x_1 \cdot \ldots \cdot x_d \leq N\right\}$ is comparable with $\#\mathcal{P}_N$ up to a factor of 2^d.
Claim:
We show by induction that

$$\Delta_d(N) = \frac{1}{(d-1)!}N\left(\log N\right)^{d-1}. \tag{A.3}$$

Proof. $d = 1$: $\Delta_1 = N$ ✓
$d \mapsto d + 1$: $x_1 * \ldots * x_{d+1} \leq N \Leftrightarrow x_1 * \ldots * x_d \leq \frac{N}{x_{d+1}}$, where $x_{d+1} \in [1, N]$. $\Rightarrow$

$$\Delta_{d+1}(N) = \int_1^N \Delta_d\left(\frac{N}{x_{d+1}}\right) dx_{d+1}$$

$$= \frac{1}{(d-1)!} \int_1^N \frac{N}{x_{d+1}} \left(\log \left(\frac{N}{x_{d+1}} \right) \right)^{d-1} dx_{d+1}, \quad y = \frac{N}{x_{d+1}}$$

$$= \frac{N}{(d-1)!} \left(- \int_N^1 y (\log y)^{d-1} \frac{1}{y^2} dy \right)$$

$$= \frac{N}{(d-1)!} \int_1^N (\log y)^{d-1} \frac{1}{y} dy$$

$$= \frac{N}{(d-1)!} \left[\frac{(\log y)^d}{d} \right]_1^N$$

$$= \frac{1}{d!} N (\log N)^d.$$

□

Inserting (A.3) into (A.2), we obtain up to constants depending on d only

$$\mathcal{N} \left((\delta u)^{1/2}, \rho^*, \{ S \in \mathcal{S}^* : |S| \leq \delta \} \right)$$

$$\leq m^d * \# \mathcal{P}_N$$

$$\leq 2^d m^d \Delta_d \left(\delta m^d \right)$$

$$\lesssim \delta m^{2d} \log \left(\delta m^d \right)^{d-1}$$

$$= \delta^{-(2d-1)} u^{-2d} \left[\log \left(d^d \delta^{-(d-1)} u^{-d} \right) \right]^{d-1}$$

$$\lesssim \delta^{-(2d-1)} \left(\log(1/\delta) \right)^{d-1} u^{-2d} \left(\log(1/u) \right)^{d-1}$$

where we used $(x + y)^{d-1} \leq c x^{d-1} y^{d-1}$ for $x, y \geq 1$. This proves the claim. □

In the case of hypercubes we have a stronger result.

Lemma A.1.2. *There exists a constant C depending only on the dimension d such that for all $u, \delta \in (0, 1]$ we have*

$$\mathcal{K} \left((\delta u)^{1/2}, \rho^*, \{ Q \in \mathcal{Q}^* : |Q| \leq \delta \} \right) \leq C \delta^{-1} u^{-(d+1)},$$

i.e., (2.4) is true with $k_1 = C, k_2 = d + 1$ and $V_{\mathcal{Q}^*} = 1$.

Proof. We proceed as in the Proof of Lemma A.1.1. In contrast to hyper-rectangles, we obtain here instead of (A.1) the better estimate

$$\left(\rho^* \left(Q, Q' \right) \right)^2 = |Q \triangle Q'| \leq \frac{d \delta^{\frac{d-1}{d}}}{m}.$$

Since all edges have the same length, and we can choose $m := \frac{d}{\delta^{1/d} u}$. Furthermore, the cardinality of $\mathcal{W}''_m$ is bounded by the number of lower left vertices times the number

of possibilities for an adjacent vertex, which gives

$$\#\mathcal{W}'_m \le m^d \cdot \left(\delta^{1/d} m\right) = m^{d+1}\delta^{1/d}.$$

Therefore we finally obtain

$$\mathcal{N}\left((\delta u)^{1/2}, \rho^*, \{S \in \mathcal{S}^* : |S| \le \delta\}\right) \le \#\mathcal{W}'_{d/(\delta^{1/d}u)}$$
$$\le \left(\frac{d}{\delta^{1/d}u}\right)^{d+1} \delta^{1/d} = d^{d+1}u^{-(d+1)}\delta^{-1},$$

which proves the claim. $\qquad\square$

In the case of half-spaces, we show

Lemma A.1.3. *There exists a constant C depending only on the dimension d such that for all $u, \delta \in (0, 1]$ it holds*

$$\mathcal{K}\left((\delta u)^{1/2}, \rho^*, \{H \in \mathcal{H}^* : |H| \le \delta\}\right) \le C\delta^{-2}u^{-2},$$

i.e., (2.4) holds with $k_1 = C, k_2 = 2$ and $V_{\mathcal{H}^*} = 2$.

Proof. Let $\mathcal{W}_{N,m} = \left\{H_{a_i,\alpha_j} \mid i = 1, ..., N, j = 1, ..., m\right\}$ with numbers $a_1, ..., a_N \in \mathbb{S}^{d-1}$ and $\alpha_1, ..., \alpha_m \in \left[0, \sqrt{d}\right]$ and H_{a_i,α_j} defined as in Example 2.1.3 c). Note that $H_{a,\alpha} = \emptyset$ for $\alpha > \sqrt{d}$ by definition and Pythagoras' theorem. It is convenient to choose $\alpha_1, ..., \alpha_m$ as equidistant, e.g.,

$$\alpha_i := \frac{i - \frac{1}{2}}{m} \sqrt{d}n, \quad i = 1, \dots, m.$$

Furthermore, let $a_1, ..., a_N$ be a maximal system of points in $\mathbb{S}^{d-1}$ such that $\sphericalangle(a_j, a_k) \ge \left(\frac{1}{m}\right)^{\frac{1}{d-1}}$ for all $j \ne k$. This implies that

$$\mathbb{S}^{d-1} \subset \bigcup_{j=1}^{N} S_{a_j}\left(\left(\frac{1}{m}\right)^{\frac{1}{d-1}}\right)$$

where $S_a(\theta_0)$ denotes the spherical cap $S_a(\theta_0) = \left\{e \in \mathbb{S}^{d-1} \mid \sphericalangle(a, e) \le \theta_0\right\}$. Note that

$$|S_a(\theta_0)| \sim \frac{\int_0^{\theta_0} (\sin t)^{d-2}\,dt}{\int_0^{\pi} (\sin t)^{d-2}\,dt} \sim \theta_0^{d-1}$$

for small values of θ_0, up to constants depending on d only. Now, for any given $a \in \mathbb{S}^{d-1}$

and $\alpha \in \left[0, \sqrt{d}\right]$, there are $1 \le i \le N$ and $1 \le j \le m$ such that

$$\sphericalangle(a, a_i) \le \left(\frac{1}{m}\right)^{\frac{1}{d-1}}, \qquad \left|\alpha - \alpha_j\right| \le \frac{\sqrt{d}}{m}.$$

Now we split

$$\left(\rho^*\left(H_{a,\alpha}, H_{a_i,\alpha_j}\right)\right)^2 \le \left|H_{a,\alpha} \triangle H_{a_i,\alpha}\right| + \left|H_{a_i,\alpha} \triangle H_{a_i,\alpha_j}\right|,$$

and since $H_{a_i,\alpha} \triangle H_{a_i,\alpha_j}$ is a d-dimensional stripe of width $\le \frac{\sqrt{d}}{m}$ and $H_{a,\alpha} \triangle H_{a_i,\alpha}$ is a union of hyperpyramids with opening angle $\le \left(\frac{1}{m}\right)^{\frac{1}{d-1}}$, we obtain

$$\left(\rho^*\left(H_{a,\alpha}, H_{a_i,\alpha_j}\right)\right)^2 \le \frac{C}{m},$$

where C is some generic constant depending only on d. Hence if we choose $m = C^{-1}\delta^{-1}u^{-1}$, then for each $H \in \{H \in \mathcal{H}^* : |H| \le \delta\}$ there exists $H' \in \mathcal{W}_{N,m}$ such that $\rho^*(S, S') \le (\delta u)^{1/2}$. We now estimate N. By elementary geometry it follows that

$$\bigcup_{j=1}^{N} S_{a_j}\left(\frac{1}{2}\left(\frac{1}{m}\right)^{\frac{1}{d-1}}\right) \subset \mathbb{S}^{d-1} \subset \bigcup_{j=1}^{N} S_{a_j}\left(\left(\frac{1}{m}\right)^{\frac{1}{d-1}}\right),$$

and furthermore up to boundary points, the sets on the left-hand side are disjoint. Therefore we obtain for the volumes that

$$N\left|S_{a_j}\left(\frac{1}{2}\left(\frac{1}{m}\right)^{\frac{1}{d-1}}\right)\right| \le \left|\mathbb{S}^{d-1}\right| \le N\left|S_{a_j}\left(\left(\frac{1}{m}\right)^{\frac{1}{d-1}}\right)\right|$$

which implies $N \sim m$. Consequently, $\#\mathcal{W}_{N,m} \sim m^2$ which proves the claim.

$\square$

APPENDIX B

Proofs

In this section we will give the proofs of our results. In the following we denote by p_n the cardinality of $\mathcal{R}_n$, i.e., $p_n := \#(\mathcal{R}_n)$, which by (2.2) satisfies $\log(p_n) \sim \log n$. Again, C will be a generic constant which might differ from line to line.

B.1   Proof of the auxiliary results

We start proving the auxiliary statements from section 2.1.4.

Proof of Lemma 2.1.13. Let $h(t) := k_1 \exp(-k_2 t)$. We may assume $h \le 1$. Then

$$\mathbb{P}\left[\max_{1 \le i \le N} |W_i| > t\right] = 1 - \mathbb{P}\left[\max_{1 \le i \le N} |W_i| \le t\right] \le 1 - (1 - h(t))^N \le N h(t).$$

Let $\bar{t} = h^{-1}(1/N) \sim C\log(N)$, then

$$\begin{aligned}
\mathbb{E}\left[\max_{1 \le i \le N} |W_i|^m\right] &= m\int_0^\infty t^{m-1}\mathbb{P}\left[\max_{1 \le i \le N} |W_i| > t\right]\,\mathrm{d}t \\
&\le m\int_0^{\bar{t}} t^{m-1}\,\mathrm{d}t + m\int_{\bar{t}}^\infty t^{m-1} N h(t)\,\mathrm{d}t \\
&\le \left(C\log(N)\right)^m + k_1 m N \int_{\bar{t}}^\infty t^{m-1}\exp(-k_2 t)\,\mathrm{d}t \\
&\le C\left(\log N\right)^m,
\end{aligned}$$

where the last inequality follows from repeated integration by parts, reducing $m-1$ to zero.

$\square$

Proof of Lemma 2.1.14. Let $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and r_i be i.i.d. Rademacher random variables, which take the values ± 1 with probability $1/2$.

Step (i): Since the variables X_i are symmetric,

$$\mathbb{E}\left[\max_I \frac{1}{\sqrt{|I|}}\left|\sum_{i\in I} X_i\right|\right] = \mathbb{E}_r\left[\mathbb{E}\left[\max_I \frac{1}{\sqrt{|I|}}\left|\sum_{i\in I} r_i X_i\right|\right]\right]. \qquad (B.1)$$

We use Lemma 4.5 of (Ledoux and Talagrand, 1991) and choose there $F(t) = t$, $\eta_i = X_i$ and $x_i := (c_{i,I})_I$, where $c_{i,I} := \frac{1}{\sqrt{|I|}}\mathbb{I}_{\{i\in I\}}$ is a scaled indicator function and as norm the maximum-norm. Then we obtain

$$\mathbb{E}_r\left[\max_I \frac{1}{\sqrt{|I|}}\left|\sum_{i\in I} r_i\right|\right] \le \sqrt{\frac{\pi}{2}}\mathbb{E}\left[\max_I \frac{1}{\sqrt{|I|}}\left|\sum_{i\in I} X_i\right|\right]. \qquad (B.2)$$

Step (ii): Let $(Z_i')_{1\le i\le N}$ be a sequence of independent copies of $(Z_i)_{1\le i\le N}$ and define the symmetrized version of Z_i by $\widetilde{Z}_i := Z_i - Z_i'$ and equally the symmetrized version of Z_I by $\widetilde{Z}_I := \sum_{i\in I}(Z_i - Z_i')$. By using the same argument as in (B.1) and Fubini's theorem, we derive

$$\begin{aligned}
\mathbb{E}\left[\max_I \frac{1}{\sqrt{|I|}} |Z_I|\right] &\le 2\mathbb{E}\left[\max_I \frac{1}{\sqrt{|I|}}\left|\widetilde{Z}_I\right|\right] \\
&= 2\mathbb{E}_r\left[\mathbb{E}\left[\max_I \frac{1}{\sqrt{|I|}}\left|\sum_{i\in I}\widetilde{Z}_i r_i\right|\right]\right] \\
&= 2\mathbb{E}\left[\mathbb{E}_r\left[\max_I \frac{1}{\sqrt{|I|}}\left|\sum_{i\in I}\widetilde{Z}_i r_i\right|\right]\right] \\
&= 2\mathbb{E}\left[\mathbb{E}_r\left[\max_I \frac{1}{\sqrt{|I|}}\left|\sum_{i\in I}\left|\widetilde{Z}_i\right| r_i\right|\right]\right],
\end{aligned}$$

where the last equality holds in view of the symmetry of r_i. Now we will use the contraction principle, i.e., Theorem 4.4 of (Ledoux and Talagrand, 1991) with $F(t) = t$ conditionally on $\alpha_i := \frac{|\widetilde{Z}_i(\omega)|}{\max_j |\widetilde{Z}_j(\omega)|}$, which is independent of (r_i). We get after multiplying both sides with $\max_j |\widetilde{Z}_j(\omega)|$

$$\mathbb{E}_r\left[\max_I \frac{1}{\sqrt{|I|}}\left|\sum_{i\in I}\left|\widetilde{Z}_i(\omega)\right| r_i\right|\right] \le \mathbb{E}\left[\max_{1\le i\le N}\left|\widetilde{Z}_i(\omega)\right|\right]\mathbb{E}_r\left[\max_I \frac{1}{\sqrt{I}}\left|\sum_{i\in I} r_i\right|\right].$$

Therefore

$$\mathbb{E}\left[\max_I \frac{1}{\sqrt{|I|}} |Z_I|\right] \le 2\mathbb{E}\left[\max_{1\le i\le N}\left|\widetilde{Z}_i(\omega)\right|\right]\mathbb{E}_r\left[\max_I \frac{1}{\sqrt{|I|}}\left|\sum_{i\in I} r_i\right|\right]$$

$$\leq 4\mathbb{E}\left[\max_{1\leq i\leq N}|Z_i|\right]\sqrt{\frac{\pi}{2}}\mathbb{E}\left[\max_{I}\frac{1}{\sqrt{|I|}}\left|\sum_{i\in I}X_i\right|\right]$$

$$=\sqrt{8\pi}\mathbb{E}\left[\max_{1\leq i\leq N}|Z_i|\right]\mathbb{E}\left[\max_{I}\frac{|X_I|}{\sqrt{|I|}}\right],$$

where we used (B.2) in the second inequality. Now the statement follows from (2.18).

$\square$

Proof of Theorem 2.1.15. Enumerate the regions R in $\mathcal{R}_n$ by j, $1\leq j\leq p_n$ and define

$$X_{ij}:=\frac{v_i}{\sqrt{|R_j|}}Z_i\mathbb{I}_{\{i\in R_j\}}\mathbb{I}_{\{|R_j|\geq r_n\}},$$

$$X_i:=\left(X_{ij}\right)_{j=1,\dots,p_n},\quad i=1,\dots,N=n^d,\tag{B.3}$$

for some sequence r_n. Then $Z:=\max\limits_{1\leq j\leq p_n}\sum_{i=1}^{N}X_{ij}$ satisfies

$$Z\overset{\mathcal{D}}{=}\max_{\substack{R\in\mathcal{R}_n:\\|R|\geq r_n}}\frac{1}{\sqrt{|R|}}\sum_{i\in R}v_iZ_i.$$

Recall that $\log(p_n)\sim\log(n)$. According to (Chernozhukov et al., 2014, Cor. 4.1) for every $\delta>0$ there exists a Gaussian version $\tilde{Z}:=\max\limits_{1\leq j\leq p_n}\sum_{i=1}^{N}v_iN_{ij}$ with independent random vectors $N_1,\dots,N_n$ in $\mathbb{R}^{p_n}$, $N_i\sim N(0,\mathbb{E}\left[X_iX_i^t\right])$, $1\leq i\leq N$, such that

$$\mathbb{P}\left[\left|Z-\tilde{Z}\right|>16\delta\right]\lesssim\delta^{-2}\left\{B_1+\delta^{-1}(B_2+B_4)\log(n)\right\}\log(n)+\frac{\log(n)}{n^d}$$

where

$$B_1:=\mathbb{E}\left[\max_{1\leq j,k\leq p_n}\left|\sum_{i=1}^{N}\left(X_{ij}X_{ik}-\mathbb{E}\left[X_{ij}X_{ik}\right]\right)\right|\right]$$

$$B_2:=\mathbb{E}\left[\max_{1\leq j\leq p_n}\sum_{i=1}^{N}\left|X_{ij}\right|^3\right]$$

$$B_4:=\sum_{i=1}^{N}\mathbb{E}\left[\max_{1\leq j\leq p_n}\left|X_{ij}\right|^3\mathbb{I}_{\{\max_{1\leq j\leq p_n}|X_{ij}|>\delta/\log(p_n\vee n)\}}\right].$$

We now give estimates for B_1, B_2 and B_4.

B_1 can be controlled as follows. With X_{ij} as above (see (B.3)) we find

$$B_1=\mathbb{E}\left[\max_{\substack{1\leq j,k\leq p_n:\\|R_j|,|R_k|\geq r_n}}\left|\sum_{i\in R_j\cap R_k}\frac{v_i^2(Z_i^2-1)}{\sqrt{|R_j||R_k|}}\right|\right]$$

$$= \mathbb{E}\left[\max_{\substack{1 \le j,k \le p_n: \\ |R_j|,|R_k| \ge r_n}} \frac{\sqrt{|R_j \cap R_k|}}{\sqrt{|R_j||R_k|}} \left| \frac{1}{\sqrt{|R_j \cap R_k|}} \sum_{i \in R_j \cap R_k} v_i^2(Z_i^2 - 1) \right| \right].$$

Using the restriction on the size of the rectangles we find:

$$\sqrt{\frac{|R_j \cap R_k|}{|R_j||R_k|}} \le \sqrt{\frac{|\min\{|R_j|,|R_k|\}|}{|R_j||R_k|}} \le \frac{1}{\sqrt{r_n}}.$$

Denote $V_i := v_i^2(Z_i^2 - 1)$, $I := R_j \cap R_k \in \mathcal{I} \subset I_n^d$ and $S_I := \sum_{i \in I} V_i$. Then

$$B_1 \le \frac{1}{\sqrt{r_n}} \mathbb{E}\left[\max_{I \in \mathcal{I}} \frac{|S_I|}{\sqrt{|I|}} \right].$$

Using Lemma 2.1.14 we obtain

$$B_1 \le \frac{C}{\sqrt{r_n}} \underbrace{\sqrt{\log(\#(\mathcal{I}))}}_{\sim \sqrt{\log(n)}} \mathbb{E}\left[\max_{1 \le i \le N} |v_i^2(Z_i^2 - 1)| \right]$$

It remains to estimate

$$\mathbb{E}\left[\max_{1 \le i \le N} |v_i^2(Z_i^2 - 1)| \right] = \mathbb{E}\left[\max_{1 \le i \le N} v_i^2|Z_i^2 - 1| \right]$$

$$\le \mathbb{E}\left[\max_{1 \le i \le N} \bar{v}^2|Z_i^2 - 1| \right]$$

$$\le \bar{v}^2 \mathbb{E}\left[\max_{1 \le i \le N} |Z_i|^2 \right] + \bar{v}^2.$$

Hence using Lemma 2.1.13 we get

$$B_1 \lesssim \frac{\sqrt{\log(n)}}{\sqrt{r_n}} \left(\bar{v}^2 C \log(N)^2 + \bar{v}^2 \right) \lesssim \left(\frac{\log^5(n)}{r_n} \right)^{1/2}.$$

To estimate B_2, we compute

$$B_2 \le \frac{1}{(r_n)^{1/2}} \mathbb{E}\left[\max_{1 \le i \le N} |v_i Z_i|^3 \right]$$

$$\le \frac{\bar{v}^3}{(r_n)^{1/2}} \mathbb{E}\left[\max_{1 \le i \le N} |Z_i|^3 \right] \lesssim \left(\frac{\log^6(n)}{r_n} \right)^{1/2},$$

where we again used Lemma 2.1.13. Now let $\delta > 0$ be fixed. Then

$$B_4 \leq \sum_{i=1}^{N} \mathbb{E}\left[\max_{\substack{1 \leq j \leq p_n: \\ |R_j| \geq r_n}} \frac{|v_i Z_i|^3}{|R_j|^{3/2}} \mathbb{I}_{\{\max_j \frac{|v_i Z_i|}{|R_j|^{1/2}} > \frac{\delta}{(\log p_n)}\}}\right]$$

$$\leq \frac{N}{r_n^{3/2}} \max_{1 \leq i \leq N} \mathbb{E}\left[|v_i Z_i|^3 \mathbb{I}_{\{|Z_i| > \frac{\delta r_n^{1/2}}{|v_i| \log p_n}\}}\right]$$

By the assumption on r_n, we have that $r_n > \left(\frac{2d|v_i|}{\delta}\right)^2 (\log n)^{2+2\gamma}$ for some $\gamma > 1$ and $n \geq n_o(\delta, d)$. Hence

$$B_4 \leq \frac{N}{r_n^{3/2}} \max_{1 \leq i \leq N} \mathbb{E}\left[|v_i|^3 |Z_i|^3 \mathbb{I}_{\{|Z_i| > (\log n)^\gamma\}}\right]$$

$$\leq \frac{N\bar{v}^3}{r_n^{3/2}} 3 \max_{1 \leq i \leq N} \int_{(\log n)^\gamma}^{\infty} t^2 \mathbb{P}\left[|Z_i| > t\right] \mathrm{d}t$$

$$\leq 3k_1 \frac{N\bar{v}^3}{r_n^{3/2}} \int_{(\log n)^\gamma}^{\infty} t^2 \exp(-k_2 t) \mathrm{d}t$$

$$= 3k_1 \frac{N\bar{v}^3}{r_n^{3/2}} \frac{1}{k_2^3} \int_{k_2(\log n)^\gamma}^{\infty} u^2 \exp(-u) \mathrm{d}u.$$

For $u \geq 10$ we have $u^2 \leq \exp(u/2)$. Choose $n_1(\delta, d, \gamma)$ such that for all $n \geq n_1(\delta, d, \gamma)$ we have $k_2 (\log n)^\gamma > 10$ and $\frac{k_2}{2} (\log n)^\gamma \geq d \log(n)$. Then

$$B_4 \leq \frac{3k_1\bar{v}^3}{k_2^3} \frac{N}{r_n^{3/2}} \int_{k_2(\log n)^\gamma}^{\infty} \exp\left(-\frac{u}{2}\right) \mathrm{d}u = \frac{3k_1\bar{v}^3}{k_2^3} \frac{N}{r_n^{3/2}} \exp\left(-\frac{k_2}{2} (\log n)^\gamma\right)$$

$$\leq \frac{3k_1\bar{v}^3}{k_2^3} \frac{n^d}{(r_n)^{3/2}} n^{-d} = \frac{3k_1\bar{v}^3}{k_2^3} \frac{1}{(r_n)^{3/2}}.$$

The estimates for B_1, B_2 and B_4 imply

$$\mathbb{P}\left[|Z - \tilde{Z}| > 16\delta\right] \lesssim \delta^{-2} \left(\frac{\log^7(n)}{r_n}\right)^{1/2} + \delta^{-3} \left(\frac{\log^{10}(n)}{r_n}\right)^{1/2}$$

$$+ \delta^{-3} \left(\frac{\log^4(n)}{r_n^3}\right)^{1/2} + \frac{\log(n)}{n^d}$$

$$\lesssim \delta^{-3} \left(\frac{\log^{10}(n)}{r_n}\right)^{1/2}.$$

which yields the claim. $\square$

B.2 Proofs of Section 2.1.2

Let us now prove the results from Section 2.1.2, in particular, Theorems 2.1.4 and 2.1.5. We start with a Taylor expansion of T_n, which will allows us to apply Theorem 2.1.15. Therefore recall the definition of the functions $m(\theta)$ and $v(\theta)$ from (2.9) as the mean and variance of $Y \sim F_\theta$.

Lemma B.2.1. *Let $\mathcal{R}_n$ be a collection of sets s.t. (2.2) holds, $\epsilon > 0$ and $(r_n)_n \subset (0, \infty)$ be a sequence, s.t. $(\log n)^{3+\epsilon}/r_n \to 0$. Suppose $Y_i \sim F_{\theta_0} \in \mathcal{F}$, $i \in I_n^d$, are i.i.d. random variables, and recall that for $R \in \mathcal{R}_n$ we denote $\overline{Y}_R := |R|^{-1} \sum_{i \in R} Y_i$. Then it holds*

$$\max_{\substack{R \in \mathcal{R}_n: \\ |R| \geq r_n}} \left| T_R(Y, \theta_0) - |R|^{\frac{1}{2}} \frac{|\overline{Y}_R - m(\theta_0)|}{\sqrt{v(\theta_0)}} \right| = O_{\mathbb{P}} \left(\left(\frac{\log^3(n)}{r_n} \right)^{1/4} \right)$$

as $n \to \infty$.

Proof. For independent Gaussian random variables $X_i \sim \mathcal{N}(0, 1)$ it follows from (2.18) and (2.2) that

$$\mathbb{E} \left[\left| \max_{R \in \mathcal{R}_n} |R|^{-1/2} \sum_{i \in R} X_i \right| \right] \leq C \sqrt{\log n},$$

hence

$$\frac{1}{\sqrt{\log(n)}} \max_{\substack{R \in \mathcal{R}_n: \\ |R| \geq r_n}} |R|^{-1/2} \sum_{i \in R} X_i = o_{\mathbb{P}}(1).$$

Combining this result with Theorem 2.1.15 (with $Z_i = \frac{Y_i - m(\theta_0)}{\sqrt{v(\theta_0)}}$, $v_i = 1$ for all i) we obtain

$$\frac{1}{\sqrt{\log n}} \max_{\substack{R \in \mathcal{R}_n: \\ |R| \geq r_n}} \frac{1}{\sqrt{|R|}} \left| \sum_{i \in R} \frac{Y_i - m(\theta_0)}{\sqrt{v(\theta_0)}} \right| = o_{\mathbb{P}}(1). \tag{B.4}$$

Together with (2.2) it follows

$$\max_{\substack{R \in \mathcal{R}_n: \\ |R| \geq r_n}} \left| \overline{Y}_R - m(\theta_0) \right| \leq C \left(\frac{\log(n)v(\theta_0)}{r_n} \right)^{\frac{1}{2}} (1 + o_{\mathbb{P}}(1)) \to 0, \quad n \to \infty.$$

Therefore, $\overline{Y}_R > m(\theta_0)/\sqrt{2}$ in probability if n is large enough uniformly over those R with $|R| \geq r_n$. Now we are in position to analyze $J(\overline{Y}_R, \theta_0) = \phi\left(\overline{Y}_R\right) - \left[\theta_0 \overline{Y}_R - \psi(\theta_0)\right]$ in the definition of T_R, see (2.1). As the supremum $\sup_{\theta \in \Theta} \prod_{i \in R} p_\theta(Y_i)$ is attained at the θ for which $\psi'(\theta) = \overline{Y}_R$ holds, we find

$$\phi(\overline{Y}_R) = \langle m^{-1}(\overline{Y}_R), \overline{Y}_R \rangle - \psi\left(m^{-1}(\overline{Y}_R)\right),$$

and therefore

$$J(\overline{Y}_R, \theta_0) = \langle m^{-1}(\overline{Y}_R), \overline{Y}_R \rangle - \psi\left(m^{-1}(\overline{Y}_R)\right) - \left(\langle \theta_0, \overline{Y}_R \rangle - \psi(\theta_0)\right)$$
$$= \langle m^{-1}(\overline{Y}_R) - \theta_0, \overline{Y}_R \rangle - \left(\psi\left(m^{-1}(\overline{Y}_R)\right) - \psi(\theta_0)\right).$$

Note that $\overline{Y}_R$ is in the domain of m^{-1}, i.e., $\overline{Y}_R \in D(m^{-1})$ for large enough n, as the latter is an open set. A Taylor expansion of ψ at θ_0 and one of second order of m^{-1} at $m(\theta_0)$ yields

$$T_R(Y, \theta_0) = \left(|R| \left(\frac{\overline{Y}_R - m(\theta_0)}{\sqrt{v(\theta_0)}}\right)^2 + |R| s_n \left(\frac{\overline{Y}_R - m(\theta_0)}{\sqrt{v(\theta_0)}}\right)\right)^{1/2}, \qquad \text{(B.5)}$$

where s_n satisfies $|s_n(x)| \leq cx^3 (1 + o_{\mathbb{P}}(1))$ for some $c > 0$. Consequently

$$\max_{\substack{R \in \mathcal{R}_n: \\ |R| \geq r_n}} \left| T_R^2(Y, \theta_0) - |R| \frac{(\overline{Y}_R - m(\theta_0))^2}{v(\theta_0)} \right|$$

$$\lesssim \max_{\substack{R \in \mathcal{R}_n: \\ |R| \geq r_n}} |R| \frac{|\overline{Y}_R - m(\theta_0)|^3}{v(\theta_0)^{3/2}} (1 + o_{\mathbb{P}}(1))$$

$$= \max_{\substack{R \in \mathcal{R}_n: \\ |R| \geq r_n}} |R|^{-\frac{1}{2}} \left| \frac{\sum_{i \in R}(Y_i - m(\theta_0))}{\sqrt{|R|} \sqrt{v(\theta_0)}} \right|^3 (1 + o_{\mathbb{P}}(1))$$

$$\leq (\log^3(n) r_n^{-1})^{1/2} (1 + o_{\mathbb{P}}(1)),$$

where we again used (B.4). Now $|a - b| \leq |a^2 - b^2|^{\frac{1}{2}}$ yields the claim. $\qquad \square$

We now prove Theorem 2.1.5. So far we have shown that the maximum over the local likelihood ratio statistics can be approximated by Gaussian versions, but we did not include the scale penalization terms $\text{pen}_v(|R|)$ in (1.8). To include this in the approximation result, we will decompose the system of sets R into scales of sets, for which the penalty-term is almost constant. Then we bound the maximum over all scales by the sum of the maximum over theses families, at the expense of an additional $\log(n)$ factor on the smallest scale.

Proof of Theorem 2.1.5. (a) It follows from the triangle inequality

$$|\|x\|_\infty - \|y\|_\infty| \leq \|x - y\|_\infty,$$

Lemma B.2.1 and (1.23) with $\gamma = 12$ that

$$
\left| \max_{\substack{R \in \mathcal{R}_n: \\ |R| \geq r_n}} \left(T_R(Y, \theta_0) - \text{pen}_v(|R|) \right) - \right.
$$

$$
\left. \max_{\substack{R \in \mathcal{R}_n: \\ |R| \geq r_n}} \left(|R|^{1/2} \left| \frac{\overline{Y}_R - m(\theta_0)}{\sqrt{v(\theta_0)}} \right| - \text{pen}_v(|R|) \right) \right| = O_{\mathbb{P}} \left(\left(\frac{\log^3(n)}{r_n} \right)^{1/4} \right).
$$

Define

$$
Y^R := |R|^{-1/2} \sum_{i \in R} \left(\frac{Y_i - m(\theta_0)}{\sqrt{v(\theta_0)}} \right)
$$

$$
X^R := |R|^{-1/2} \sum_{i \in R} X_i, \quad X_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1).
$$

Using this notation and a symmetry argument, we get from the proof of Theorem 2.1.15 with $v_i \equiv 1$ that

$$
\mathbb{P} \left[\left| \max_{\substack{R \in \mathcal{R}_n: \\ |R| \geq r_n}} |Y^R| - \max_{\substack{R \in \mathcal{R}_n: \\ |R| \geq r_n}} |X^R| \right| > \delta \right] \lesssim \delta^{-3} \left(\frac{\log^{10}(n)}{r_n} \right)^{1/2}. \tag{B.6}
$$

Let $\delta_n := ((\log^{12}(n)/r_n)^{1/10} \searrow 0$ and define $\epsilon_j := j\delta_n, j \in \mathbb{N}$ as well as

$$
\mathcal{R}_{n,j} := \left\{ R \in \mathcal{R}_n \mid \exp(\epsilon_j) < |R| < \exp(\epsilon_{j+1}) \right\}.
$$

Then the set of candidate regions $\mathcal{R}_{n|r_n}$ can be written as

$$
\mathcal{R}_{n|r_n} = \bigsqcup_{j \in J} \mathcal{R}_{n,j}, \quad J := \left\{ \frac{1}{\delta_n} \log \left(\log^{12}(n) \right), \ldots, \frac{1}{\delta_n} \log(n^d) \right\}
$$

with $|J| \leq \frac{\log(n^d)}{\delta_n}$. If we abbreviate

$$
\text{pen}_j := \text{pen}_v \left(\exp(\epsilon_j) \right) = \sqrt{2v \left(\log \left(\frac{n^d}{\exp(\epsilon_j)} \right) + 1 \right)},
$$

then the above decomposition implies

$$
\text{pen}_{j+1} \leq \text{pen}_v(|R|) \leq \text{pen}_j, \quad \text{for all} \quad R \in \mathcal{R}_{n,j}.
$$

Using $\sqrt{a} - \sqrt{b} = (a - b)/\left(\sqrt{a} + \sqrt{b}\right)$, we get

$$0 \leq \mathrm{pen}_j - \mathrm{pen}_{j+1}$$
$$= \frac{2v(\epsilon_{j+1} - \epsilon_j)}{\sqrt{2v\left[\log(n^d) + 1 - \epsilon_j\right]} + \sqrt{2v\left[\log(n^d) + 1 - \epsilon_{j+1}\right]}}.$$

The largest index in J is $\frac{1}{\delta_n}\log(n^d)$ and therefore the maximal value of ϵ_i is given by $\bar{\epsilon} = \log(n^d)$ and $\log(n^d) + 1 - \bar{\epsilon} = 1$. Therefore,

$$0 \leq \mathrm{pen}_j - \mathrm{pen}_{j+1} \leq \frac{2v(\epsilon_{j+1} - \epsilon_j)}{2\sqrt{2v}} = \delta_n \sqrt{\frac{v}{2}}.$$

Hence for $n \to \infty$ the penalty terms $\mathrm{pen}_v(|R|)$ can be considered as constant on $\mathcal{R}_{n,j}$. Now straightforward computations show that

$$\left|\max_{\substack{R \in \mathcal{R}_n: \\ |R| \geq r_n}} \left(\left|Y^R\right| - \mathrm{pen}_v(|R|)\right) - \max_{\substack{R \in \mathcal{R}_n: \\ |R| \geq r_n}} \left(\left|X^R\right| - \mathrm{pen}_v(|R|)\right)\right|$$

$$\leq \max_{j \in J} \left|\max_{R \in \mathcal{R}_{n,j}} \left|Y^R\right| - \max_{R \in \mathcal{R}_{n,j}} \left|X^R\right|\right| + \delta_n \sqrt{\frac{v}{2}}.$$

Now the claim in (a) of Theorem 2.1.5 follows from (B.6) and $|J| \leq \frac{\log(n^d)}{\delta_n}$.

(b) This is a direct consequence of (a). $\qquad\qquad\square$

We now prove Theorem 2.1.4. Taking into account Theorem 2.1.5, we only have to prove Remark 2.1.6, which will be done in the following.

Lemma B.2.2. *Let $\mathcal{R}^*$ satisfy Assumption 1 and be equipped with the canonical metric ρ^* as in (2.3) and define $\mathcal{R}_n$ as in (1.4). Furthermore let W denote white noise on $[0, 1]^d$. For $X_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $i \in I_n^d$ define*

$$Z_n(R^*) := n^{-d/2} \sum_{i/n \in R^*} X_i \overset{\mathcal{D}}{=} n^{-d/2} \sum_{i \in \{1,\dots,n\}^d} |nR^* \cap A_i| X_i, \qquad R^* \in \mathcal{R}^*$$

where $A_i = (i_1 - 1, i_1] \times \dots \times (i_d - 1, i_d]$ is the unit cube with upper corner i. Then we have

$$Z_n \overset{\mathcal{D}}{\to} W, \quad n \to \infty.$$

Proof. Note that $\mathcal{R}^*$ is totally bounded w.r.t. ρ^*.
We verify the assumptions of (Kosorok, 2008, Thm. 2.1):

1. *Tightness:* The white noise W is tight.

2. *Totally boundedness:* By Markov's inequality and standard bounds on the modulus of continuity, we obtain using Assumption 1(a) that

$$
\mathbb{P}^* \left[\sup_{\substack{R_1^*, R_2^* \in \mathcal{R}^* \\ \rho(R_1^*, R_2^*) \leq \delta}} |Z_n(R_1^*) - Z_n(R_2^*)| > \epsilon \right]
$$

$$
\leq \frac{1}{\epsilon} \mathbb{E} \left[\sup_{\substack{R_1^*, R_2^* \in \mathcal{R}^* \\ \rho(R_1^*, R_2^*) \leq \delta}} \left| Z_n(R_1^*) - Z_n(R_2^*) \right| \right]
$$

$$
\lesssim \int_0^\delta \sqrt{2\nu(\mathcal{R}^*) \log\left(\frac{C}{u}\right)} \, \mathrm{d}u
$$

which tends to 0 as $\delta \searrow 0$.

3. *Finite dimensional convergence*: The convergence of the finite-dimensional laws is an application of the central limit theorem for random fields (Dedecker, 1998, Thm 2.2) and (Dedecker, 2001, Lemma 2), which implies that

$$
\frac{\left| nR^* \cap \mathbb{Z}^d \right|}{n^d} \to |R^*|
$$

for regular Borel sets $R^* \subset [0, 1]^d$ with $|R^*|) > 0$. Consequently, the central limit theorem yields for any fixed $R^* \in \mathcal{R}^*$ that

$$
Z_n(R^*) \xrightarrow{\mathcal{D}} N\left(0, |R^*|\right) \qquad \text{as} \qquad n \to \infty
$$

A similar computation shows that

$$
\mathrm{Cov}\left(Z_n(R_1^*), Z_n(R_2^*)\right) \to \left| R_1^* \cap R_2^* \right|
$$

for all $R_1^*, R_2^* \in \mathcal{R}$. This proves finite dimensional convergence.

The claim now follows by the Theorem 2.1 of (Kosorok, 2008) since the above statements 1.-3. are equivalent to $Z_n \xrightarrow{\mathcal{D}} W$. $\qquad\qquad\square$

Now we want to apply the generalized version of the continuous mapping theorem (see e.g., Billingsley, 2013, Thm. 5.5). For $c \geq 0$ and $x \in C(\mathcal{B}([0, 1]^d), \mathbb{R})$, where $\mathcal{B}([0, 1]^d)$

denote the Borel sets in $[0,1]^d$, define

$$h^c(x) := \sup_{\substack{R^* \in \mathcal{R}^*: \\ |R^*| > c^d}} \left(\frac{|x(R^*)|}{\sqrt{|R^*|}} - \text{pen}_v\left(n^d\, |R^*|\right) \right)$$

$$h_n^c(x) := \max_{\substack{R \in \mathcal{R}_n: \\ |R| > (cn)^d}} \left(\frac{|x(R/n)|}{\sqrt{|R|/n^d}} - \text{pen}_v\left(|R|\right) \right).$$

The necessary conditions to apply the continuous mapping theorem are shown in the following Lemma:

Lemma B.2.3. *Consider h^c, h_n^c as functions $\left(C(\mathcal{B}([0,1]^d), \mathbb{R}), \|\cdot\|_\infty \right) \to \mathbb{R}$.*

   i) h^c *is uniformly continuous and* $\left(h_n^c \right)_{n \in \mathbb{N}}$ *is a sequence of equi-continuous functions, (uniformly in n).*

   ii) For $(x_n)_n \in C(\mathcal{B}([0,1]^d), \mathbb{R})$, s.t. $x_n \to x$ we have

$$h_n^c(x_n) \to h^c(x), \quad n \to \infty.$$

Proof.   i) Let $\epsilon > 0$ and choose $\delta = \epsilon c^{d/2}$. Consider two functions $x, y \in C(\mathcal{B}([0,1]^d), \mathbb{R})$ s.t. $d(x,y) = \sup_{R^* \subset [0,1]^d} \||x(R^*)| - |y(R^*)|\| < \delta$. Using $|\max a_i - \max b_i| \le \max |a_i - b_i|$, we find

$$|h_n^c(x) - h_n^c(y)| \le \max_{\substack{R \in \mathcal{R}_n \\ |R| > (cn)^d}} \left| \frac{|x(R/n)| - |y(R/n)|}{\sqrt{|R|/n^d}} \right| \le \frac{\delta}{c^{d/2}} = \epsilon.$$

Similar arguments yield the uniform continuity of h^c.

  ii) Let $(x_n)_n$, $x \in C(\mathcal{B}([0,1]^d), \mathbb{R})$, s.t. $x_n \to x$. Since the functions $(h_m^c)_{m \in \mathbb{N}}$ are equi-continuous, for any $\epsilon > 0$ we can find an $N_1 \in \mathbb{N}$ s.t. $\forall n > N_1 \; \forall m$:

$$|h_m^c(x_n) - h_m^c(x)| < \frac{\epsilon}{2}.$$

Given ϵ and N_1 and $n > N_1$ with $|h_m^c(x_n) - h_m^c(x)| < \frac{\epsilon}{2}$, choose $m = n$. Then

$$|h_n^c(x_n) - h_n^c(x)| < \epsilon/2. \tag{B.7}$$

Now let us define

$$\mathcal{A} := \left\{ R^* \in \mathcal{R}^* : |R^*| \ge c^d \right\}, \quad \mathcal{B}_n := \left\{ R/n \in \mathcal{R}^* : R \in \mathcal{R}_n, |R| \ge (cn)^d \right\}.$$

The set $\mathcal{A}$ is a compact set with respect to the metric ρ^* defined in (2.3), with respect to which $\mathcal{R}^*$ is totally bounded. Furthermore $\mathcal{B}_n$ is a finite subset of $\mathcal{A}$. If we fix $x \in \mathcal{B}([0,1]^d)$ and introduce $g : \mathcal{A} \to \mathbb{R}$ by

$$g(R^*) := \left(\frac{|x(R^*)|}{\sqrt{|R^*|}} - \mathrm{pen}_v\left(|R^*|\right) \right), \qquad R^* \in \mathcal{R}^*,$$

then

$$h^c(x) = \sup_{R^* \in \mathcal{A}} g(R^*) \geq h_n^c(x) = \max_{R^* \in \mathcal{B}_n} g(R^*). \tag{B.8}$$

since $\mathcal{B}_n$ is a subset of $\mathcal{A}$. Straightforward computations show that g is continuous with respect to ρ^*, which implies by compactness of $\mathcal{A}$ that there exists an $\widetilde{R} \in \mathcal{A}$ such that $h^c(x) = g(\widetilde{R})$. Now let $R_n \in \mathcal{B}_n$ be a sequence such that $R_n \to \widetilde{R}, n \to \infty$ with respect to ρ. Then $g(R_n) \to g(\widetilde{R})$ as $n \to \infty$ and hence

$$h^c(x) \overset{\text{(B.8)}}{\geq} h_n^c(x) \geq g(R_n) \to g(\tilde{R}) = h^c(x).$$

Consequently there exists a $N_2 \in \mathbb{N}$ such that for all $n > N_2$ we have

$$|h^c(x) - h_n^c(x)| < \epsilon/2,$$

which together with (B.7) implies

$$|h_n^c(x_n) - h^c(x)| \leq \epsilon \qquad \text{for all} \qquad n > \max\{N_1, N_2\}.$$

$\square$

We are now able to prove Remark 2.1.6:

Proof of Remark 2.1.6. By Lemma B.2.2, Lemma B.2.3 and the generalized version of the continuous mapping theorem (see e.g., Billingsley, 2013, Thm. 5.5) we get

$$h_n^c(Z_n) \xrightarrow{\mathcal{D}} h^c(W), \quad n \to \infty.$$

The functions h_n^c and h^c have been defined such that

$$h_n^c(Z_n) = M_n \left(\mathcal{R}_{n|(cn)^d}, v \right), \quad \text{and} \quad h^c(W) = M \left(\mathcal{R}^*_{|c^d}, v \right).$$

Therefore for all $c > 0$ we have convergence in distribution

$$M_n \left(\mathcal{R}_{n|(cn)^d}, v \right) \xrightarrow{\mathcal{D}} M \left(\mathcal{R}^*_{|c^d}, v \right) \qquad \text{as} \qquad n \to \infty.$$

Since $M\left(\mathcal{R}^*_{|c^d}, v\right) \xrightarrow{\mathcal{D}} M(\mathcal{R}^*, v)$, $c \to 0$, we get

$$\lim_{c \to 0} \lim_{n \to \infty} M_n\left(\mathcal{R}_{n|(cn)^d}, v\right) = M(\mathcal{R}^*, v).$$

It can also readily be seen from the definition that

$$\liminf_{n \to \infty} \mathbb{P}\left[M_n\left(\mathcal{R}_{n|r_n}, v\right) \leq t\right] \geq \mathbb{P}\left[M(\mathcal{R}^*, v) \leq t\right].$$

Now fix $c > 0$ and assume that $r_n < (cn)^d$ for all $n \in \mathbb{N}$. Then we find that

$$\begin{aligned}
\mathbb{P}\left[M(\mathcal{R}^*, v) \leq t\right] &\leq \liminf_{n \to \infty} \mathbb{P}\left[M_n\left(\mathcal{R}_{n|r_n}, v\right) \leq t\right] \\
&\leq \limsup_{n \to \infty} \mathbb{P}\left[M_n\left(\mathcal{R}_{n|(cn)^d}, v\right) \leq t\right] \\
&\to \mathbb{P}\left[M(\mathcal{R}^*, v) \leq t\right] \qquad \text{as} \qquad c \searrow 0,
\end{aligned}$$

which proves the claim. $\qquad\square$

Proof of Theorem 2.1.4. The main statement follows from Theorem 2.1.5 together with Remark 2.1.6. It remains to show the a.s. boundedness and non-degenerateness of $M(\mathcal{R}^*, v)$. We apply (Dümbgen and Spokoiny, 2001, Thm. 6.1) with ρ^* as in (2.3) and

$$\sigma^2(R^*) := |R^*|, \qquad X(R^*) := W(R^*).$$

Let us check the three conditions from their theorem:

i) $\sigma^2(R_1^*) \leq \sigma^2(R_2^*) + \rho^*(R_1^*, R_2^*)^2$ for all $R_1^*, R_2^* \in \mathcal{R}^*$ is obviously fulfilled since $R_1^* \cap R_2^* \subset R_2^*$ and $R_1^* \setminus R_2^* \subset R_1^* \triangle R_2^*$. Since $\mathbb{V}\left[W(R^*)\right] = |R^*|$,

$$\mathbb{P}\left[X(R^*) > \sigma(R^*)\eta\right] = \mathbb{P}\left[W(R^*) > \eta\left(|R^*|\right)^{1/2}\right] \leq \frac{1}{2} \exp\left(-\frac{\eta^2}{2}\right). \qquad \text{(B.9)}$$

ii) For

$$\mathbb{P}\left[|X(R_1^*) - X(R_2^*)| > \rho(R_1^*, R_2^*)\eta\right] = \mathbb{P}\left[\left|W(R_1^*) - W(R_2^*)\right| > \left|R_1^* \triangle R_2^*\right|)^{1/2}\eta\right]$$

we compute and use that

$$\begin{aligned}
W(R_1^*) - W(R_2^*) &\sim \mathcal{N}(0, \sigma^2_{R_1^*, R_2^*}), \\
\sigma^2_{R_1^*, R_2^*} &= \left|R_1^*\right| + \left|R_2^*\right| - 2\,\text{Cov}(W(R_1^*), W(R_2^*)), \\
\left|R_1^* \triangle R_2^*\right| &= \left|R_1^*\right| + \left|R_2^*\right| - 2\left|R_1^* \cap R_2^*\right|.
\end{aligned}$$

Since $\mathrm{Cov}(W(R_1^*), W(R_2^*)) = \left| R_1^* \cap R_2^* \right|$, we consequently find

$$\mathbb{P}\left[|X(R_1^*) - X(R_2^*)| > \rho(R_1^*, R_2^*)\eta \right] \leq \exp\left(-\frac{\eta^2}{2} \frac{\left| R_1^* \bigtriangleup R_2^* \right|}{\sigma_{R_1^*, R_2^*}^2} \right) = \exp\left(-\frac{\eta^2}{2} \right).$$

iii) This is fulfilled by Assumption 1(a) (cf. Remark 2.1.2).

(B.9) holds with $-W(R^*)$ as well, hence we get that the statistic $M(\mathcal{R}^*, v)$ is finite a.s. Non-degenerateness is obvious, as M is always larger than the value of the local statistic on one fixed scale, which is non-degenerate. $\qquad\square$

B.3 Proofs of Section 2.1.3

Let us now prove the results in Section 2.1.3, namely Theorem 2.1.9 and Corollary 2.1.11. First we introduce some abbreviations to simplify the notation. Let

$$q^* := q_{1-\alpha,n}^{\mathrm{O}}, \qquad q := q_{1-\alpha,n}^{\mathrm{MS}},$$

and denote the total signal on $Q \in \mathcal{Q}^n$ by

$$\mu^n(Q) := |Q|^{-1/2} \sum_{i \in Q} \frac{m(\theta_i^n) - m(\theta_0)}{\sqrt{v(\theta_0)}} = \frac{|Q \cap Q_n|}{\sqrt{|Q|}} \frac{m(\theta_1^n) - m(\theta_0)}{\sqrt{v(\theta_0)}}, \tag{B.10}$$

since the signal $\frac{m(\theta_i^n) - m(\theta_0)}{\sqrt{v(\theta_0)}}$ is non zero on the anomalies only.
For brevity introduce the Gaussian process

$$\gamma(Q) := \left| \mu^n(Q) + |Q|^{-\frac{1}{2}} \sum_{i \in Q} v_i X_i \right| - \mathrm{pen}_v(|Q|), \qquad Q \in \mathcal{Q}^n$$

with $X_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $v_i = \sqrt{v(\theta_i)/v(\theta_0)}$.

Let us now analyse the oracle procedure. In a first step, we leave out a suitable subset of hypercubes close to the true anomaly Q_n. More precisely, choose a sequence ε_n such that $\varepsilon_n \searrow 0$ but $\varepsilon_n \mu^n(Q_n) \to \infty$ and denote the set of all hypercubes which are ϵ_n–close to the anomaly by

$$\mathcal{U}_n := \left\{ Q \in \mathcal{Q}^n(a_n) \mid \mu^n(Q) \geq \mu^n(Q_n)(1 - \varepsilon_n) \right\}.$$

Furthermore let us introduce an extended neighborhood $\mathcal{U}$ of the anomaly by

$$\mathcal{U} := \left\{ Q \in \mathcal{Q}^n(a_n) \mid Q \cap Q' \neq \emptyset \text{ for some } Q' \in \mathcal{U}_n \right\},$$

and its complement by $\mathcal{T} := Q^n(a_n) \setminus \mathcal{U}$. By definition, $\{\gamma(Q)\}_{Q \in \mathcal{T}}$ and $\{\gamma(Q)\}_{Q \in \mathcal{U}_n}$ are independent, which allows us to compute the asymptotic power of the single-scale procedure. For a sketch of $\mathcal{U}_n$ and $\mathcal{U}$ see Figure B.1.
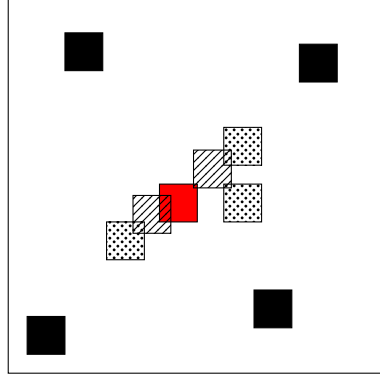

Figure B.1: Exemplary elements of the sets $\mathcal{U}_n, \mathcal{U}$ and $\mathcal{T}$ in dimension $d = 2$: The anomaly is shown in red, the hatched cubes belong to $\mathcal{U}_n$, the dotted cubes to $\mathcal{U}$, and all black cubes belong to $\mathcal{T}$. By definition, for all $Q \in \mathcal{U}_n$ and $Q' \in \mathcal{T}$ we have $Q \cap Q' = \emptyset$, which implies independence of $\{\gamma(Q)\}_{Q \in \mathcal{T}}$ and $\{\gamma(Q)\}_{Q \in \mathcal{U}_n}$.

We start by bounding the covering number $N(\mathcal{U}, \rho, \epsilon)$ with respect to the canonical metric $\rho(Q, Q')^2 = 2 - 2|Q \cap Q'| / \sqrt{|Q||Q'|}$.

Lemma B.3.1. *For any $\epsilon > 0$ we have*

$$N(\mathcal{U}, \rho, \epsilon) \leq C \left(\frac{6d}{\epsilon} \right)^d.$$

Proof.

Let $Q(3)$ denote the the cube of side length 3 times the side length of Q_n centered at the midpoint of Q_n. Let $0 < \delta \leq \frac{\epsilon}{2d}$, $\frac{1}{\delta} \in \mathbb{N}$. Choose equidistant points in $Q(3)$ of distance $\delta|Q_n|^{1/n}$ in each coordinate, which requires $\left(\frac{3}{\delta} \right)^d$ points. As a covering M for $\mathcal{U}$, consider the cubes of side length $|Q_n|^{1/n}$ which have vertices in the net of equidistant points. To approximate $Q \in \mathcal{U}$ with $Q \cap Q' \neq \emptyset$, where $Q' \in \mathcal{U}_n$, i.e., $|Q_n \cap Q'| \geq \left(1 - \frac{\delta}{2} \right) |Q_n|$, by elements of this net, note that Q is essentially contained in $Q(3)$ -up to distance $\frac{\delta}{2}$- and therefore there is a cube $\widetilde{Q} \in M$ in the covering such that the volume (or number of points) in $Q \triangle \widetilde{Q}$ is bounded by $\leq d \left(\delta|Q_n|^{1/d} \right) |Q_n|^{(d-1)/d}$, since the complements $Q \setminus \widetilde{Q}$ and $\widetilde{Q} \setminus Q$ in each fixed dimension have at most width $\leq \delta|Q_n|^{1/d}$ and extension $|Q_n|^{1/d}$ in the remaining $(n-1)$ dimensions. It is bounded by $d\delta|Q_n|$, $\left| Q \triangle \widetilde{Q} \right| \geq (1 - d\delta)|Q_n| \geq \left(1 - \frac{\epsilon}{2} \right) |Q_n|$. Therefore $N(\mathcal{U}, \rho, \epsilon) \leq \left(\frac{3}{\delta} \right)^d \leq C \left(\frac{6d}{\epsilon} \right)^d$. $\qquad \square$

Lemma B.3.2. *Consider the setup of Section 2.1.3 and recall that q^* is the $(1 - \alpha)$–quantile of $M_n(Q^n(a_n))$ as in (1.10). Then*

(a) $\max\limits_{Q \in \mathcal{U}} |Q|^{-\frac{1}{2}} \left| \sum_{i \in Q} v_i X_i \right| = O_{\mathbb{P}}(1) \text{ as } n \to \infty$

(b) $\lim\limits_{n \to \infty} \mathbb{P} \left[\max\limits_{Q \in \mathcal{T}} \gamma(Q) \leq q^* \right] = 1 - \alpha$

Proof. (a) It follows from Dudley's entropy integral (see e.g., Marcus and Rosen, 2006, Thm. 6.1.2) with any fixed $Q' \in \mathcal{U}$ that

$$
\mathbb{E} \left[\max\limits_{Q \in \mathcal{U}} |Q|^{-1/2} \left| \sum_{i \in Q} v_i X_i \right| \right]
$$

$$
\leq \mathbb{E} \left[|Q'|^{-1/2} \left| \sum_{i \in Q'} v_i X_i \right| \right] + \mathbb{E} \left[\max\limits_{Q, Q' \in \mathcal{U}} \left| |Q|^{-1/2} \left| \sum_{i \in Q} v_i X_i \right| - |Q'|^{-1/2} \left| \sum_{i \in Q'} v_i X_i \right| \right| \right]
$$

$$
\leq \sqrt{\frac{2\overline{v^2}}{\pi}} + C \int_0^2 \sqrt{\log \mathcal{N}(\mathcal{U}, \rho, \epsilon)} \, d\epsilon \leq \sqrt{\frac{2\overline{v^2}}{\pi}} + C \int_0^2 \sqrt{d} \sqrt{\log \left(\frac{6d}{\epsilon} \right)} \, d\epsilon < \infty
$$

which by Markov's inequality proves the claim.

(b) A direct consequence of (a) is that

$$
\max\limits_{Q \in \mathcal{U}} \left[|Q|^{-\frac{1}{2}} \left| \sum_{i \in Q} v_i X_i \right| - \mathrm{pen}_v(|Q|) \right] = o_{\mathbb{P}}(1).
$$

Furthermore note that $\mu^n(Q) = 0$ and $v_i \equiv 1$, $i \in Q$ for $Q \in \mathcal{T}$. Consequently

$$
\mathbb{P} \left[\max\limits_{Q \in \mathcal{T}} \gamma(Q) \leq q^* \right] = \mathbb{P} \left[\max\limits_{Q \in \mathcal{T}} \left[|Q|^{-\frac{1}{2}} \left| \sum_{i \in Q} X_i \right| - \mathrm{pen}_v(|Q|) \right] \leq q^* \right]
$$

$$
= \mathbb{P} \left[\max\limits_{Q \in \mathcal{Q}^n(a_n)} \left[|Q|^{-\frac{1}{2}} \left| \sum_{i \in Q} X_i \right| - \mathrm{pen}_v(|Q|) \right] \leq q^* \right] + o(1)
$$

$$
= \mathbb{P} \left[M_n(\mathcal{Q}^n(a_n)) \leq q^* \right] + o(1)
$$

which yields the claim. □

With this Lemma at hand, we are now in position to find the asymptotic power of the oracle procedure:

Proof of Theorem 2.1.9(a). To analyze $\mathbb{P}_{\theta^n} \left[T_n(Y, \theta_0, \mathcal{Q}^n(a_n)) > q^* \right]$, we start by showing the lower estimate $\geq$ in the statement of Theorem 2.1.9(a). By Lemma B.2.1 and the triangle inequality we can replace $T_n(Y, \theta_0, \mathcal{Q}^n(a_n))$ by

$$
\max\limits_{Q \in \mathcal{Q}^n(a_n)} \left[|Q|^{-\frac{1}{2}} \left| \sum_{i \in Q} \frac{Y_i - m(\theta_0)}{\sqrt{v(\theta_0)}} \right| - \mathrm{pen}_v(|Q|) \right]
$$

up to $o_{\mathbb{P}}(1)$. Moreover (2.10) and Theorem 2.1.15 allow us to approximate the latter sum by a Gaussian version, i.e.,

$$\mathbb{P}_{\theta^n}\left[T_n(Y,\theta_0,Q^n(a_n)) > q^*\right] = \mathbb{P}\left[\max_{Q \in Q^n(a_n)} \gamma(Q) > q^*\right] + o(1).$$

Now we estimate

$$\begin{aligned}
&\mathbb{P}\left[\max_{Q \in Q^n(a_n)} \gamma(Q) > q^*\right]\\
={}&\mathbb{P}\left[\left\{\max_{Q \in Q^n(a_n)} \gamma(Q) > q^*\right\} \cap \left\{\max_{Q \in \mathcal{T}} \gamma(Q) \le q^*\right\}\right]\\
&\quad + \mathbb{P}\left[\left\{\max_{Q \in Q^n(a_n)} \gamma(Q) > q^*\right\} \cap \left\{\max_{Q \in \mathcal{T}} \gamma(Q) > q^*\right\}\right]\\
={}&\mathbb{P}\left[\left\{\max_{Q \in \mathcal{T}} \gamma(Q) \le q^*\right\} \cap \left\{\max_{Q \in \mathcal{U}} \gamma(Q) > q^*\right\}\right] + \mathbb{P}\left[\max_{Q \in \mathcal{T}} \gamma(Q) > q^*\right]\\
\ge{}&\mathbb{P}\left[\left\{\max_{Q \in \mathcal{T}} \gamma(Q) \le q^*\right\} \cap \{\gamma(Q_n) > q^*\}\right] + \mathbb{P}\left[\max_{Q \in \mathcal{T}} \gamma(Q) > q^*\right]\\
={}&\mathbb{P}\left[\max_{Q \in \mathcal{T}} \gamma(Q) \le q^*\right]\mathbb{P}\left[\gamma(Q_n) > q^*\right] + \mathbb{P}\left[\max_{Q \in \mathcal{T}} \gamma(Q) > q^*\right]
\end{aligned}$$

where we exploited $Q_n \in \mathcal{U}$ and the independence of $\{\gamma(Q)\}_{Q \in \mathcal{T}}$ and $\gamma(Q_n)$. Lemma B.3.2(b) states that $\mathbb{P}\left[\max_{Q \in \mathcal{T}} \gamma(Q) \le q^*\right] = 1 - \alpha + o(1)$ and hence

$$\mathbb{P}_{\theta^n}\left[T_n(Y,\theta_0,Q^n(a_n)) > q^*\right] \ge \alpha + (1-\alpha)\mathbb{P}\left[\gamma(Q_n) > q^*\right] + o(1).$$

Furthermore note that $\gamma(Q_n) + \mathrm{pen}_v(|Q_n|)$ follows a folded normal distribution with parameters $\mu = \mu^n(Q_n)$ and $\sigma^2 = |Q_n|^{-1}\sum_{i \in Q_n} v_i^2$, i.e.,

$$\gamma(Q_n) \sim \left|\mathcal{N}(\mu,\sigma^2)\right| - \mathrm{pen}_v(|Q_n|).$$

We calculate

$$\mu^n(Q_n) = \sqrt{n^d a_n}\frac{5}{2}C\frac{m(\theta_1^n) - m(\theta_0)}{\sqrt{v(\theta_0)}}(1 + o(1)), \tag{B.11}$$

$$|Q_n|^{-1}\sum_{i \in Q_n} v_i^2 = \frac{v(\theta_1^n)}{v(\theta_0)}, \tag{B.12}$$

$$\mathrm{pen}_v(|Q|) = \sqrt{2v\log(a_n^{-1})} + o(1) \quad \text{for all} \quad Q \in Q^n(a_n), \tag{B.13}$$

which yields the claimed lower bound by using the continuity of F and the fact that $Q_n \in Q^n(a_n)$ satisfies the proposed lower bound. For the upper bound $\le$ in the state-

ment of Theorem 2.1.9(a) we proceed as before and obtain

$$\mathbb{P}_{\theta^n}\left[T_n(Y,\theta_0,Q^n(a_n)) > q^*\right]$$

$$=\alpha + \mathbb{P}\left[\left\{\max_{Q\in\mathcal{T}}\gamma(Q) \le q^*\right\} \cap \left\{\max_{Q\in\mathcal{U}}\gamma(Q) > q^*\right\}\right] + o(1)$$

$$=\alpha + \mathbb{P}\left[\left\{\max_{Q\in\mathcal{T}}\gamma(Q) \le q^*\right\} \cap \left\{\max_{Q\in\mathcal{U}}\gamma(Q) > q^*\right\} \cap \left\{\max_{Q\in\mathcal{U}_n}\gamma(Q) > q^*\right\}\right]$$

$$+ \mathbb{P}\left[\left\{\max_{Q\in\mathcal{T}}\gamma(Q) \le q^*\right\} \cap \left\{\max_{Q\in\mathcal{U}}\gamma(Q) > q^*\right\} \cap \left\{\max_{Q\in\mathcal{U}_n}\gamma(Q) \le q^*\right\}\right] + o(1)$$

$$\le\alpha + \mathbb{P}\left[\left\{\max_{Q\in\mathcal{T}}\gamma(Q) \le q^*\right\} \cap \left\{\max_{Q\in\mathcal{U}_n}\gamma(Q) > q^*\right\}\right] + \mathbb{P}\left[\max_{Q\in\mathcal{U}\setminus\mathcal{U}_n}\gamma(Q) > q^*\right] + o(1)$$

$$=\alpha + (1-\alpha)\mathbb{P}\left[\max_{Q\in\mathcal{U}_n}\gamma(Q) > q^*\right] + \mathbb{P}\left[\max_{Q\in\mathcal{U}\setminus\mathcal{U}_n}\gamma(Q) > q^*\right] + o(1)$$

where we used the independence of $\{\gamma(Q)\}_{Q\in\mathcal{T}}$ and $\{\gamma(Q)\}_{Q\in\mathcal{U}_n}$. Lemma B.3.2(a) yields

$$\max_{Q\in\mathcal{U}\setminus\mathcal{U}_n}\left||Q|^{-\frac{1}{2}}\sum_{i\in Q}v_iX_i\right| \le \max_{Q\in\mathcal{U}}\left||Q|^{-\frac{1}{2}}\sum_{i\in Q}v_iX_i\right| = O_{\mathbb{P}}(1).$$

Further we have by definition of $\mathcal{U}_n$ that $\mu^n(Q) \le (1-\varepsilon_n)\mu^n(Q_n)$ for all $Q \in \mathcal{U}\setminus\mathcal{U}_n$. Exploiting (B.13) this implies

$$\mathbb{P}\left[\max_{Q\in\mathcal{U}\setminus\mathcal{U}_n}\gamma(Q) > q^*\right]$$

$$\le\mathbb{P}\left[(1-\varepsilon_n)\mu^n(Q_n) + O_{\mathbb{P}}(1) - \sqrt{2\log\left(a_n^{-1}\right)} > q^*\right]$$

$$=\mathbb{P}\left[\mu^n(Q_n) - \sqrt{2\log\left(a_n^{-1}\right)} - \varepsilon_n\mu^n(Q_n) + O_{\mathbb{P}}(1) > q^*\right] = o(1)$$

if $\mu^n(Q_n) - \sqrt{2\log\left(a_n^{-1}\right)} \to C \in [-\infty,\infty)$ since $\varepsilon_n\mu^n(Q_n) \to \infty$ by construction. If $\mu^n(Q_n) - \sqrt{2\log\left(a_n^{-1}\right)} \to \infty$, then nothing has to be shown. Altogether this gives

$$\mathbb{P}_{\theta^n}\left[T_n(Y,\theta_0,Q^n(a_n)) > q^*\right] \le \alpha + (1-\alpha)\mathbb{P}\left[\max_{Q\in\mathcal{U}_n}\gamma(Q) > q^*\right] + o(1).$$

Using similar arguments as in Lemma B.3.2 we obtain from $\varepsilon_n \searrow 0$ that

$$\mathbb{P}\left[\max_{Q\in\mathcal{U}_n}\gamma(Q) > q^*\right] = \mathbb{P}\left[\gamma(Q_n) + o_{\mathbb{P}}(1) > q^*\right]$$

and hence the theorem is proven. $\qquad\square$

Now we turn to the multiscale procedure. Since different scales are considered there, the set $\mathcal{U}$ is not large enough any more. In particular, we cannot construct a subset $\mathcal{V}$ such that $\{\gamma(Q)\}_{\mathcal{V}^c}$ and $\gamma(Q_n)$ are independent and that $\max_{Q \in \mathcal{V}} \gamma(Q)$ is still negligible. Due to this problem, the corresponding proof in (Sharpnack and Arias-Castro, 2016) is incomplete. To overcome this difficulty, we employ the idea to distinguish the two cases that the anomaly Q_n has asymptotically an effect on $\gamma(Q)$ or not. Whenever Q is sufficiently large compared to Q_n, the impact will be asymptotically negligible.

For some sequence $\epsilon_n \searrow 0$ we introduce

$$\delta_n := \epsilon_n \left(\max \left\{ \mu^n(Q_n), \log(n) \sqrt{\frac{|Q_n|}{r_n}} \right\} \right)^{-1}, \tag{B.14}$$

$$\mathcal{V} := \left\{ Q \in Q_{n|r_n}^{\mathrm{MS}} \mid \mu^n(Q) \geq \delta_n \mu^n(Q_n) \right\}$$

and its complement $\mathcal{T}' := Q_{n|r_n}^{\mathrm{MS}} \setminus \mathcal{V}$. For a sketch see Figure B.2.
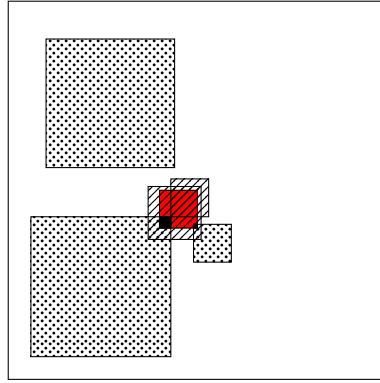

Figure B.2: Exemplary elements of the sets $\mathcal{V}$ and $\mathcal{T}'$ in $d = 2$: The anomaly is shown in red, the hatched cubes belong to $\mathcal{V}$ and the dotted cubes to $\mathcal{T}'$. However, the intersections marked in black are small enough such that they have asymptotically no influence on $\gamma(Q)$.

Contrary to the oracle procedure, we do not have independence of $\{\gamma(Q)\}_{Q \in \mathcal{T}'}$ and $\gamma(Q_n)$. However, asymptotically a similar property is true as shown in the following Lemma B.3.4.

Let us again start with bounding the covering number $N(\mathcal{V}, \rho, \epsilon)$ w.r.t. the canonical metric $\rho(Q, Q')^2 = 2 - 2|Q \cap Q'| / \sqrt{|Q||Q'|}$.

Lemma B.3.3. *There exists a constant C such that for any $\epsilon > 0$ we have*

$$N(\mathcal{V}, \rho, \epsilon) \leq C \left(\frac{6d}{\epsilon} \right)^d \frac{|Q_n|^{d+1}}{\delta_n^{2(d+1)}}.$$

Proof. For all $Q \in \mathcal{V}$ defined in (B.14) it holds $\mu^n(Q) \geq \delta_n \mu^n(Q_n)$, which implies

$$\delta_n \sqrt{|Q_n|} \leq \frac{|Q \cap Q_n|}{\sqrt{|Q|}} \leq \frac{|Q_n|}{\sqrt{|Q|}}.$$

Consequently, $\mathcal{V}$ contains only cubes Q with $r_n \leq |Q| \leq \delta_n^{-2} |Q_n|$. For a fixed scale k, $\mathcal{V}$ contains at most $(C|Q_n|)$ Q's with $|Q| = k$, and for the set of such Q's an $\sqrt{\epsilon}$-covering can be constructed as in the proof of Lemma B.3.1 with at most $C\left(\frac{6d}{\epsilon}k\right)^d$ elements, which gives

$$N(\mathcal{V}, \rho, \epsilon) \leq C \sum_{k=r_n}^{\lfloor \delta_n^{-2}|Q_n| \rfloor} \left(\frac{6d}{\epsilon}k\right)^d \leq C\left(\frac{6d}{\epsilon}\right)^d \frac{|Q_n|^{d+1}}{\delta_n^{2(d+1)}}.$$

$\square$

Lemma B.3.4. *Consider the setting from Section 2.1.3 and recall that q is the $(1 - \alpha)$-quantile of $M_n\left(Q_{n|r_n}^{\mathrm{MS}}\right)$ as in (1.10). Furthermore, assume that $|Q_n| = o(n^\alpha)$, for α sufficiently small. Then the following statements are true as $n \to \infty$:*

(a) $\displaystyle \max_{Q \in \mathcal{V}} \left| |Q|^{-\frac{1}{2}} \sum_{i \in Q} v_i X_i \right|$

$= O_{\mathbb{P}}\left(\sqrt{\ln\left(\frac{|Q_n|}{\delta_n^2}\right)}\right) = O_{\mathbb{P}}\left(\sqrt{\ln(|Q_n|)} + \sqrt{\ln\left(\left|m\left(\theta_1^n\right) - m(\theta_0)\right|\right)}\right)$

(b) $\displaystyle \max_{Q \in \mathcal{T}'} \left| |Q|^{-\frac{1}{2}} \sum_{i \in Q \cap Q_n} v_i X_i \right| = o_{\mathbb{P}}(1)$

(c) $\displaystyle \mathbb{P}\left[\max_{Q \in \mathcal{T}'} \gamma(Q) \leq q\right] = 1 - \alpha + o(1)$

Proof. (a) Again with the help of Dudley's entropy integral we find

$$\mathbb{E}\left[\max_{Q \in \mathcal{V}} |Q|^{-1/2} \left|\sum_{i \in Q} v_i X_i\right|\right]$$

$$\leq \sqrt{\frac{2\overline{v^2}}{\pi}} + C_1 \int_0^2 \sqrt{\log N(\mathcal{V}, \rho, \epsilon)}\, d\epsilon$$

$$\leq C_2 \left(\sqrt{\ln\left(\frac{|Q_n|}{\delta_n^2}\right)}\right)$$

$$\leq C_3 \left(\sqrt{\ln(|Q_n|)} + \sqrt{\ln\left(\left|m\left(\theta_1^n\right) - m(\theta_0)\right|\right)}\right).$$

Now Markov's inequality yields the claim.

(b) For $Q \in \mathcal{T}'$ we have $\mu^n(Q) < \delta_n \mu^n(Q^n)$ and hence $|Q \cap Q_n| \leq \delta_n \sqrt{|Q||Q_n|}$. Conse-

quently

$$\mathbb{E}\left[\max_{Q \in \mathcal{T}'} |Q|^{-1/2} \left| \sum_{i \in Q \cap Q_n} v_i X_i \right| \right]$$

$$\leq \sqrt{\delta_n} \left(\frac{|Q_n|}{r_n} \right)^{\frac{1}{4}} \mathbb{E}\left[\max_{Q \in Q_n | r_n} |Q \cap Q_n|^{-1/2} \left| \sum_{i \in Q \cap Q_n} v_i X_i \right| \right]$$

$$\leq C \bar{v} \sqrt{\delta_n} \left(\frac{|Q_n|}{r_n} \right)^{\frac{1}{4}} \sqrt{\log(n)}$$

where we used (2.18). As the right-hand side converges to 0 by (B.14), this proves the claim.

(c) We show that (c) can be deduced from (a) and (b) as follows. For all $Q \in \mathcal{T}'$ we have $\mu^n(Q) \leq \delta_n \mu^n(Q_n)$ and hence

$$\max_{Q \in \mathcal{T}'} \gamma(Q) - \max_{Q \in \mathcal{T}'} \left[\left| |Q|^{-\frac{1}{2}} \sum_{i \in Q \setminus Q_n} v_i X_i \right| - \text{pen}_v(|Q|) \right]$$

$$\leq \max_{Q \in \mathcal{T}'} \left[\mu^n(Q) + \left| |Q|^{-\frac{1}{2}} \sum_{i \in Q \cap Q_n} v_i X_i \right| \right]$$

$$\leq \delta_n \mu^n(Q_n) + \max_{Q \in \mathcal{T}'} \left| |Q|^{-\frac{1}{2}} \sum_{i \in Q \cap Q_n} v_i X_i \right| \overset{(b)}{=} o_{\mathbb{P}}(1) \qquad (B.15)$$

where the last estimate follows from $\delta_n \mu^n(Q_n) \searrow 0$. Furthermore, as $\mathcal{V}$ contains only scales of order $\leq \delta_n^{-2} |Q_n|$ we obtain that

$$\max_{Q \in \mathcal{V}} \left[\left| |Q|^{-\frac{1}{2}} \sum_{i \in Q} v_i X_i \right| - \text{pen}_v(|Q|) \right]$$

$$\leq \max_{Q \in \mathcal{V}} \left[\left| |Q|^{-\frac{1}{2}} \sum_{i \in Q} v_i X_i \right| - \text{pen}_v(\delta_n^{-2} |Q_n|) \right]$$

$$\overset{(a)}{=} O_{\mathbb{P}}\left(\sqrt{\ln\left(\frac{|Q_n|}{\delta_n^2} \right)} \right) - \text{pen}_v(\delta_n^{-2} |Q_n|)$$

$$= O_{\mathbb{P}}\left(\sqrt{\ln\left(\frac{|Q_n|}{\delta_n^2} \right)} - \sqrt{2v \left(\log\left(\frac{n^d}{|Q_n|} \right) \delta_n^2 + 1 \right)} \right)$$

$$= o_{\mathbb{P}}(1), \qquad (B.16)$$

if $|Q_n| \leq n^{d \frac{2v}{2v+1}} \delta_n^2$. Since δ_n^2 is up to $\log$ -terms and constants of order $\frac{\epsilon_n^2}{|Q_n|}$, thus is

satisfied if $|Q_n| \lesssim n^{d\frac{v}{2v+1}} \epsilon_n$ up to log $-$terms and constants, i.e., if $|Q_n| = o(n^\alpha)$ for sufficiently small $\alpha > 0$, i.e., $\alpha < d\frac{v}{2v+1}$. Consequently

$$
\mathbb{P}\left[\max_{Q\in\mathcal{T}'} \gamma(Q) \le q\right]
$$

$$
\stackrel{(B.15)}{=} \mathbb{P}\left[\max_{Q\in\mathcal{T}'}\left[\left||Q|^{-\frac{1}{2}}\sum_{i\in Q\setminus Q_n} v_i X_i\right| - \text{pen}_v(|Q|)\right] \le q\right] + o(1)
$$

$$
\stackrel{(b)}{=} \mathbb{P}\left[\max_{Q\in\mathcal{T}'}\left[\left||Q|^{-\frac{1}{2}}\sum_{i\in Q} v_i X_i\right| - \text{pen}_v(|Q|)\right] \le q\right] + o(1)
$$

$$
\stackrel{(B.16)}{=} \mathbb{P}\left[\max_{Q\in Q_{n|r_n}^{\text{MS}}}\left[\left||Q|^{-\frac{1}{2}}\sum_{i\in Q} v_i X_i\right| - \text{pen}_v(|Q|)\right] \le q\right] + o(1)
$$

$$
= \mathbb{P}\left[M_n\left(Q_{n|r_n}^{\text{MS}}\right) \le q\right] + o(1)
$$

which proves (c). $\qquad\square$

Proof of Theorem 2.1.9(b). For the multiscale procedure we have to prove a lower bound for $\mathbb{P}_{\theta^n}\left[T_n(Y, \theta_0, Q_{n|r_n}^{\text{MS}}) > q\right]$. Similar as in the proof of Theorem 2.1.9(a) we obtain

$$
\mathbb{P}_{\theta^n}\left[T_n(Y, \theta_0, Q_{n|r_n}^{\text{MS}}) > q\right]
$$

$$
\ge \mathbb{P}\left[\left\{\max_{Q\in\mathcal{T}'}\gamma(Q) \le q\right\} \cap \{\gamma(Q_n) > q\}\right] + \mathbb{P}\left[\max_{Q\in\mathcal{T}'}\gamma(Q) > q\right] + o(1).
$$

Lemma B.3.4(b) implies

$$
\mathbb{P}\left[\max_{Q\in\mathcal{T}'}\gamma(Q) \le q\right]
$$

$$
= \mathbb{P}\left[\max_{Q\in\mathcal{T}'}\left[\left|\mu^n(Q) + \frac{1}{\sqrt{|Q|}}\sum_{i\in Q} v_i X_i\right| - \text{pen}_v(|Q|)\right] \le q\right]
$$

$$
= \mathbb{P}\left[\max_{Q\in\mathcal{T}'}\left[\left|\mu^n(Q) + \frac{1}{\sqrt{|Q|}}\sum_{i\in Q\setminus Q_n} v_i X_i\right| - \text{pen}_v(|Q|)\right] \le q\right] + o(1).
$$

Therefore independence yields

$$
\mathbb{P}\left[\left\{\max_{Q\in\mathcal{T}'}\gamma(Q) \le q\right\} \cap \{\gamma(Q_n) > q\}\right]
$$

$$
= \mathbb{P}\left[\max_{Q\in\mathcal{T}'}\gamma(Q) \le q\right] \mathbb{P}\left[\gamma(Q_n) > q\right] + o(1).
$$

Now the proof can be completed as the one of Theorem 2.1.9(a). $\qquad\square$

Proof of Corollary 2.1.11. The procedures have asymptotic power 1 if and only if

$$F\left(\bar{q} + \sqrt{-2v\log(a_n)}, n^{d/2}\sqrt{a_n}\frac{m\left(\theta_1^n\right) - m\left(\theta_0\right)}{\sqrt{v\left(\theta_0\right)}}, \frac{v\left(\theta_1^n\right)}{v\left(\theta_0\right)}\right) \to 1$$

as $n \to \infty$, with $\bar{q} \in \{q^*, q\}$ respectively, see Theorem 2.1.9. The straightforward estimate

$$F\left(x, \mu, \sigma^2\right) \geq \max\left\{\Phi\left(\frac{-x-\mu}{\sigma}\right), \Phi\left(\frac{\mu-x}{\sigma}\right)\right\}$$

shows that this is the case if and only if

$$\frac{x+\mu}{\sigma} \to -\infty \qquad \text{or} \qquad \frac{x-\mu}{\sigma} \to -\infty.$$

Inserting the values for x, μ and σ and noting that q^*, q are uniformly bounded by the $(1-\alpha)$−quantile of $M\left(Q^*, v\right)$ gives the claim. □

B.4   Proofs of Section 2.2.2

In this section the notation is the one of Section 2.2, i.e., $T_R(Y)$ is now the local Poisson-LRT from (1.13). To prove the Gaussian approximation result, we first use a Taylor approximation and the coupling result (Theorem 2.1.15) to show that the local likelihood ratio statistic T_R can be approximated by a sum of Gaussian ones. The statement of the following Theorem is similar to the one of Theorem 2.1.5. But we stress that the difference of including the sample mean instead of the true parameter into the local statistic T_R requires an additional assumption on the range of scales. We not only assume a lower bound on the size of regions but also an upper one.

Proposition B.4.1. *Let $\epsilon > 0$ and $(r_n)_n \subset (0, \infty)$ be a sequence, s.t. $(\log n)^{3+\epsilon}/r_n \to 0$. Furthermore let m_n be a sequence s.t. (1.24) holds. For $i \in \mathbb{N}^d$ and i.i.d. r.v.'s $Y_i \sim Pois(\lambda_0)$, where $\lambda_0 > 0$ is estimated by the sample mean $\overline{Y}$, there exists an array of i.i.d. r.v.'s $X_i \sim \mathcal{N}(0, 1)$, s.t.*

$$\max_{R \in \mathcal{R}_n(r_n, m_n)} \left| T_R(Y) - |R|^{1/2}\frac{|\overline{Y}_R - \lambda_0|}{\sqrt{\lambda_0}} \right| = O_{\mathbb{P}}\left(\left(\frac{\log^3(n)}{r_n\lambda_0}\right)^{1/4}\right).$$

Proof of Proposition B.4.1. By the triangle inequality

$$\max_{R \in \mathcal{R}_n(r_n, m_n)} \left| T_R^2(Y) - |R|\frac{\left(\overline{Y}_R - \lambda_0\right)^2}{\lambda_0} \right|$$

$$\leq \max_{R \in \mathcal{R}_n(r_n, m_n)} \left| T_R^2(Y, \lambda_0) - |R| \frac{\left(\overline{Y}_R - \lambda_0\right)^2}{\lambda_0} \right| + \max_{R \in \mathcal{R}_n(r_n, m_n)} \left| T_R^2(Y) - T_R^2(Y, \lambda_0) \right|$$

with $T_R(Y, \lambda_0)$ as in (1.3). The first part can be estimated from above as in Lemma B.2.1 by $\log(n)^3 / (r_n \lambda_0) \to 0$, $n \to \infty$, which tends to zero for $n \to \infty$.
Introducing the function f by $f(y) := \overline{Y}_R \left(\log \overline{Y}_R - \log y \right) - (\overline{Y}_R - y)$, $y > 0$, we write

$$\max_{R \in \mathcal{R}_n(r_n, m_n)} \left| T_R^2(Y) - T_R^2(Y, \lambda_0) \right| = \max_{R \in \mathcal{R}_n(r_n, m_n)} 2|R| \left| f(\lambda_0) - f\left(\overline{Y}\right) \right|.$$

A Taylor expansion of f at λ_0 yields

$$f\left(\overline{Y}\right) = f(\lambda_0) + \left(\overline{Y} - \lambda_0\right) \left(1 - \frac{\overline{Y}_R}{\zeta} \right),$$

with $\zeta > 0$ between λ_0 and $\overline{Y}$. This implies, in particular,

$$|\lambda_0 - \zeta| \leq |\lambda_0 - \overline{Y}| \tag{B.17}$$

and therefore

$$
\begin{aligned}
\left| f(\overline{Y}) - f(\lambda_0) \right| &\leq |\overline{Y} - \lambda_0| \frac{1}{|\zeta|} |\zeta - \overline{Y}_R| \\
&\leq \frac{1}{|\zeta|} |\overline{Y} - \lambda_0| \left(|\overline{Y}_R - \lambda_0| + |\lambda_0 - \zeta| \right) \\
&\leq \frac{1}{|\zeta|} |\overline{Y} - \lambda_0| |\overline{Y}_R - \lambda_0| + \frac{1}{|\zeta|} |\overline{Y} - \lambda|^2
\end{aligned}
\tag{B.18}
$$

By the C.L.T. we have

$$n^{d/2} \left(\overline{Y} - \lambda_0\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \lambda_0).$$

Let $c_n \searrow 0$, which may decay arbitrarily slowly. Then

$$c_n n^{d/2} \left(\overline{Y} - \lambda_0\right) = o_{\mathbb{P}}(1).$$

Therefore

$$\left| \overline{Y} - \lambda_0 \right| \leq \frac{1}{c_n n^{d/2}} \searrow 0.$$

Hence we conclude

$$|\zeta| \geq \lambda_0 - |\zeta - \lambda_0| \overset{(B.17)}{\geq} \lambda_0 - |\lambda_0 - \overline{Y}| \geq \frac{\lambda_0}{2} + o_{\mathbb{P}}(1).$$

With (B.18) we get

$$\max_{R \in \mathcal{R}_n(r_n, m_n)} 2|R| \left| f(\overline{Y}) - f(\lambda_0) \right|$$

$$\leq \left(\frac{4}{\lambda_0} |\overline{Y} - \lambda_0| \max_{R \in \mathcal{R}_n(r_n, m_n)} |R| \left| \overline{Y}_R - \lambda_0 \right| + \frac{4}{\lambda_0} \max_{R \in \mathcal{R}_n(r_n, m_n)} |R| \left| \overline{Y} - \lambda_0 \right|^2 \right) (1 + o_{\mathbb{P}}(1))$$

$$\leq \frac{4}{\sqrt{\lambda_0}} \frac{\sqrt{m_n}}{c_n n^{d/2}} \max_{R \in \mathcal{R}_n(r_n, m_n)} \frac{1}{\sqrt{|R|}} \left| \sum_{i \in R} \frac{Y_i - \lambda_0}{\sqrt{\lambda_0}} \right| (1 + o_{\mathbb{P}}(1)) + \frac{4}{\lambda_0} m_n \frac{1}{(c_n n^{d/2})^2} (1 + o_{\mathbb{P}}(1))$$

$$\leq c \left(\frac{\sqrt{m_n} \sqrt{\log(n)}}{c_n n^{d/2}} + \frac{m_n}{c_n^2 n^d} \right) (1 + o_{\mathbb{P}}(1))$$

since $\max_{R \in \mathcal{R}_n(r_n, m_n)} \frac{1}{\sqrt{|R|}} \left| \sum_{i \in R} \frac{Y_i - \lambda_0}{\sqrt{\lambda_0}} \right|$ can be estimated by $\max_{R \in \mathcal{R}_n(r_n, m_n)} \frac{1}{\sqrt{|R|}} \left| \sum_{i \in R} X_i \right|$ using Theorem 2.1.15. By our choice of m_n this expression tends to zero. Using $|a - b| \leq |a^2 - b^2|^{1/2}$, Proposition B.4.1 follows. $\square$

We are now ready to prove the Gaussian approximation result Theorem 2.2.4. To show a corresponding approximation result which includes the new penalty term we will use the decomposition technique. The ideas is to consider the scales, where the penalty-term is almost constant as one "set" or "family of sets" and decompose all scales into these families. We then bound the maximum over all scales by the sum of the maximum over theses families. These maxima then can be treated similarly as before by cancelling the (almost constant) penalty term and use Corollary 4.1 of (Chernozhukov et al., 2014).

Proof of Theorem 2.2.4. Define

$$Y^R := |R|^{-1/2} \sum_{i \in R} \left(\frac{Y_i - \lambda_0}{\sqrt{\lambda_0}} \right),$$

$$G^R := |R|^{-1/2} \sum_{i \in R} X_i, \quad X_i \overset{i.i.d.}{\sim} N(0, 1).$$

It follows from Proposition B.4.1, $\left| \|x\|_\infty - \|y\|_\infty \right| \leq \|x - y\|_\infty$ and Assumption 4 that

$$\left| \max_{R \in \mathcal{R}_n(r_n, m_n)} \widetilde{\omega}(|R|) \left(T_R(Y) - \omega(|R|) \right) - \right.$$

$$\left. \max_{R \in \mathcal{R}_n(r_n, m_n)} \widetilde{\omega}(|R|) \left(|Y^R| - \omega(|R|) \right) \right| = O_{\mathbb{P}} \left(\left(\frac{\log^{3+4\widetilde{\alpha}}(n)}{r_n \lambda_0} \right)^{1/4} \right)$$

A slight refinement of the proof of Theorem 2.1.15 and a symmetrization argument yields that

$$\mathbb{P}\left[\left|\max_{R\in\mathcal{R}_n(r_n,m_n)}|Y^R| - \max_{R\in\mathcal{R}_n(r_n,m_n)}|G^R|\right| > \delta\right] \lesssim \delta^{-3}\left(\frac{\log^{10}n}{r_n\lambda_0^3}\right)^{1/2}. \qquad \text{(B.19)}$$

Choose $\delta_n := \left(\frac{\log^\gamma n}{r_n\lambda_0^3}\right)^{1/6}$ and define for $j\in\mathbb{N}$

$$\epsilon_j := j\frac{\delta_n}{\left(\log(n^d)\right)^\eta},$$

$$\mathcal{R}_{n,j} := \left\{R\in\mathcal{R}_n \mid \exp(\epsilon_j) < |R| \le \exp(\epsilon_{j+1})\right\}$$

where $\eta = \max\left(\frac{1}{2} + \max(\widetilde{\beta},0), \alpha + \max(\widetilde{\beta},0), \widetilde{\alpha} + \max(\beta,0)\right)$. Let $\theta_n := \left(\frac{\log^\gamma n}{r_n\lambda_0^3}\right)^{1/12}$. We then decompose

$$\mathcal{R}_n(r_n,m_n) = \bigcup_{j\in J}\mathcal{R}_{n,j},$$

$$J := \left\{\frac{\left(\log n^d\right)^\eta\log(\log^\gamma(n))}{\delta_n}, \ldots, \frac{\left(\log n^d\right)^\eta}{\delta_n}\log\left(\frac{n^d}{\left(\log n^d\right)}\right)\right\},$$

$$|J| \le \frac{\left(\log n^d\right)^{\eta+1}}{\delta_n}.$$

We introduce the approximating scale calibration terms by

$$\widetilde{\omega}_j := \widetilde{\omega}\left(\exp(\epsilon_j)\right) \quad \text{and} \quad \omega_j := \omega\left(\exp(\epsilon_j)\right).$$

Then $\omega_{j+1} \le \omega(|R|) \le \omega_j$ and $\widetilde{\omega}_{j+1} \le \widetilde{\omega}(|R|) \le \widetilde{\omega}_j$ for $R\in\mathcal{R}_{n,j}$. First of all,

$$\max_{R\in\mathcal{R}_{n,j}}\widetilde{\omega}(|R|)\left(|Y^R| - \omega(|R|)\right) - \max_{R\in\mathcal{R}_{n,j}}\widetilde{\omega}(|R|)\left(|G^R| - \omega(|R|)\right)$$

$$\le \left(\widetilde{\omega}_j\max_{R\in\mathcal{R}_{n,j}}|Y^R| - \widetilde{\omega}_{j+1}\max_{R\in\mathcal{R}_{n,j}}|G^R|\right) + \left(\widetilde{\omega}_j\omega_j - \widetilde{\omega}_{j+1}\omega_{j+1}\right)$$

$$= \widetilde{\omega}_j\left(\max_{R\in\mathcal{R}_{n,j}}|Y^R| - \max_{R\in\mathcal{R}_{n,j}}|G^R|\right) + \left(\widetilde{\omega}_j - \widetilde{\omega}_{j+1}\right)\max_{R\in\mathcal{R}_{n,j}}|G^R| + \left(\widetilde{\omega}_j\omega_j - \widetilde{\omega}_{j+1}\omega_{j+1}\right),$$

and similarly if G^R and Y^R are interchanged. We estimate these terms by

$$\widetilde{\omega}_j = \widetilde{\omega}\left(\exp(\epsilon_j)\right) \lesssim \left(\log\left(\frac{n^d}{\exp(\epsilon_j)}\right)\right)^{\widetilde{\alpha}} \le \left(\log(n^d)\right)^{\widetilde{\alpha}}$$

$$\widetilde{\omega}_j - \widetilde{\omega}_{j+1} = \widetilde{\omega}\left(\exp(\epsilon_{j+1})\right) - \widetilde{\omega}\left(\exp(\epsilon_j)\right)$$

$$= \left(\exp(\epsilon_{j+1}) - \exp(\epsilon_j) \right) \widetilde{\omega}' \left(\exp(\zeta_j) \right), \qquad \zeta_j \in (\epsilon_j, \epsilon_{j+1})$$

$$\lesssim \left(\exp(\epsilon_{j+1}) - \exp(\epsilon_j) \right) \left(\log \left(\frac{n^d}{\exp(\zeta_j)} \right) \right)^{\widetilde{\beta}} \frac{1}{\exp(\zeta_j)}$$

$$\leq \left(\log(n^d) \right)^{\max(\widetilde{\beta}, 0)} \frac{\exp(\epsilon_{j+1}) - \exp(\epsilon_j)}{\exp(\zeta_j)}$$

$$\leq \left(\log(n^d) \right)^{\max(\widetilde{\beta}, 0)} \frac{\exp(\epsilon_{j+1}) - \exp(\epsilon_j)}{\exp(\epsilon_j)}$$

$$\leq \left(\log(n^d) \right)^{\max(\widetilde{\beta}, 0)} \left(\exp(\epsilon_{j+1} - \epsilon_j) - 1 \right)$$

$$\leq \left(\log(n^d) \right)^{\max(\widetilde{\beta}, 0)} \left(\exp \left(\frac{\delta_n}{\left(\log(n^d) \right)^{\eta}} \right) - 1 \right).$$

For small x we have $\exp(x) - 1 \lesssim 2x$ and therefore

$$\widetilde{\omega}_j - \widetilde{\omega}_{j+1} \lesssim 2 \frac{\delta_n}{\left(\log(n^d) \right)^{\eta}} \left(\log(n^d) \right)^{\max(\widetilde{\beta}, 0)} \leq \frac{2\delta_n}{\sqrt{\log(n^d)}}.$$

Similar statements hold for ω_j and $\omega_j - \omega_{j+1}$. Using those, we also find

$$\left| \widetilde{\omega}_j \omega_j - \widetilde{\omega}_{j+1} \omega_{j+1} \right|$$

$$= \left| \widetilde{\omega}_j (\omega_j - \omega_{j+1}) - \omega_{j+1} \left(\widetilde{\omega}_j - \widetilde{\omega}_{j+1} \right) \right|$$

$$\lesssim \frac{2\delta_n}{\left(\log(n^d) \right)^{\eta}} \left(\left(\log(n^d) \right)^{\widetilde{\alpha}} \left(\log(n^d) \right)^{\max(\beta, 0)} + \left(\log(n^d) \right)^{\alpha} \left(\log(n^d) \right)^{\max(\widetilde{\beta}, 0)} \right)$$

$$= 2\delta_n \left(\left(\log(n^d) \right)^{\widetilde{\alpha} + \max(\beta, 0) - \eta} + \left(\log(n^d) \right)^{\alpha + \max(\widetilde{\beta}, 0) - \eta} \right)$$

$$\leq 2\delta_n,$$

by the choice of η.

We therefore find

$$\left| \max_{R \in \mathcal{R}_n} \widetilde{\omega}(|R|) \left(|Y^R| - \omega(|R|) \right) - \max_{R \in \mathcal{R}_n} \widetilde{\omega}(|R|) \left(|G^R| - \omega(|R|) \right) \right|$$

$$\lesssim \left(\log(n^d) \right)^{\widetilde{\alpha}} \max_{j \in J} \left| \max_{R \in \mathcal{R}_{n,j}} |Y^R| - \max_{R \in \mathcal{R}_{n,j}} |G^R| \right| + \frac{2\delta_n}{\sqrt{\log(n^d)}} \left(\max_{R \in \mathcal{R}_n} |G^R| \right) + 2\delta_n.$$

Hence for any $0 < \delta < 1$

$$\mathbb{P} \left[\left| \max_{R \in \mathcal{R}_n} \omega(|R|) \left(|Y^R| - \omega(|R|) \right) - \max_{R \in \mathcal{R}_n} \omega(|R|) \left(|G^R| - \omega(|R|) \right) \right| \geq \delta \theta_n \right]$$

$$\leq \mathbb{P} \left[\left(\log(n^d) \right)^{\widetilde{\alpha}} \max_{j \in J} \left| \max_{R \in \mathcal{R}_{n,j}} |Y^R| - \max_{R \in \mathcal{R}_{n,j}} |G^R| \right| > \delta \theta_n / 3 \right]$$

$$+ \mathbb{P}\left[\frac{2\delta_n}{\sqrt{\log(n^d)}} \left(\max_{R \in \mathcal{R}_n} |G^R|\right) > \delta\theta_n/3\right] + \mathbb{P}\left[2\delta_n \geq \theta_n\delta/3\right]$$

$$\overset{\text{Markov}}{\leq} \mathbb{P}\left[\max_{j \in J}\left|\max_{R \in \mathcal{R}_{n,j}} |Y^R| - \max_{R \in \mathcal{R}_{n,j}} |G^R|\right| > \frac{\delta\theta_n}{3\left(\log(n^d)\right)^{\widetilde{\alpha}}}\right] + \frac{\mathbb{E}\left[\max_{R \in \mathcal{R}_n} |G^R|\right]}{\left(\frac{\delta\theta_n \sqrt{\log(n^d)}}{6\delta_n}\right)} + \mathbb{P}\left[\frac{\delta_n}{\theta_n} \geq \frac{\delta}{6}\right]$$

$$\overset{\text{(B.19)}}{\leq} |J|\frac{3^3(\log n^d)^{3\widetilde{\alpha}}}{\delta^3\theta_n^3}\left(\frac{\log^{10}(n)}{r_n\lambda_0^3}\right)^{1/2} + C\frac{6\delta_n}{\delta\theta_n} + o(1)$$

$$\leq \frac{\left(\log(n^d)\right)^{\eta+1}}{\delta_n}\frac{3^3(\log n^d)^{3\widetilde{\alpha}}}{\delta^3\theta_n^3}\left(\frac{\log^{10}(n)}{r_n\lambda_0^3}\right)^{1/2} + C\frac{6\delta_n}{\delta\theta_n} + o(1)$$

$$\leq \left(\frac{\left(\log(n)\right)^{\gamma}}{r_n\lambda_0^3}\right)^{1/2}\frac{1}{\delta_n\delta^3\theta_n^3} + 6C\frac{\theta_n}{\delta} + o(1)$$

$$\lesssim \frac{1}{\delta^3}\left(\left(\frac{\left(\log(n)\right)^{\gamma}}{r_n\lambda_0^3}\right)^{1/4}\frac{1}{\delta_n} + \theta_n\right) + o(1)$$

$$= \frac{2}{\delta^3}\theta_n + o(1).$$

□

Proof of Theorem 2.2.5. By applying Lemma 6 of (Proksch et al., 2018) with indicator functions over our set R out of the large sets $\mathcal{R}_n(r_n, m_n)$ as test functions, we derive

$$\max_{1 \leq i \leq p_n}\left\{|R_i|^{-1/2}\sum_{j \in R_i} X_j - \frac{1}{|R_i^*|^{1/2}}W(R_i^*)\right\} = o_{\mathbb{P}}\left(\frac{1}{\sqrt{\log(n)}}\right),$$

where $p_n = \text{card}(\mathcal{R}_n(r_n, m_n))$ and $\cdot^*$ denotes the embedded version into $[0, 1]^d$. Thereby we exploited the structure of our considered regions as a wavelet structure. (This result can be seen as a multivariate version of the Donsker argument).

Then it follows by the triangle-inequality for the supremum-norm and Assumption 4 that

$$\left|\max_{R \in \mathcal{R}_n(r_n, m_n)} \widetilde{\omega}(|R|)\left(T_R(Y) - \omega(|R|)\right) -\right.$$

$$\left.\sup_{R^* \in \mathcal{R}^*(r_n/n^d, m_n/n^d)} \widetilde{\omega}(n^d|R^*|)\left(\frac{W(R^*)}{\sqrt{|R^*|}} - \omega(n^d|R^*|)\right)\right| \overset{\mathbb{P}}{\to} 0, \quad n \to \infty.$$

□

Proof of Theorem 2.2.7. By an application of Theorem 5 of (Proksch et al., 2018), we will prove a lower and an upper bound for our statistic by Gumbel-distributions. Apply their Theorem 5 with indicator functions of our set R for the large sets $\mathcal{R}_n$ as

test functions, i.e., $\Xi\left(\frac{t_i - x_j}{h_i}\right) = \mathbb{I}_{\tilde{R}_i}(x_j), 1 \le i \le p_n$. Since the correlation function of $Z_t := \int \mathbb{I}_{R_t}(z)dW_z$ is the convolution of $\mathbb{I}_{R_t}$ with itself, we get the following expression for the correlation function by using $\mathbb{E}\left[|Z_t - Z_s|^2\right] = 2 - 2r_\Xi(s - t)$:

$$r_\Xi(s - t) = 1 - \frac{1}{2}\lambda^d\left(R_s \triangle R_t\right).$$

It can be shown that

$$r_\Xi(0) = 1 - \sum_{i=1}^d |t_i| + o\left(\sum_{i=1}^d |t_i|\right), \quad \|t\| \to 0.$$

Therefore our scale calibrations are the result and we get that there exists a constant $\underline{D}$ such that

$$\exp\left(-\underline{D}\exp(-t)\right) \le \lim_{n\to\infty} \mathbb{P}\left[M^\omega \le t\right] \quad \text{for any fixed } t \in \mathbb{R}.$$

Furthermore there exists a function F (independent of n) with $\lim_{t\to\infty} F(t) = 0$ such that

$$\mathbb{P}\left[M^\omega > t\right] \le F(t).$$

This means that our statistic $M(\mathcal{R}^*)$ is almost surely bounded.
The upper bound by another Gumbel distribution follows in a similar manner. There exists a constant $\underline{D}$ such that

$$\exp\left(-\overline{D}\exp(-t)\right) \ge \lim_{n\to\infty} \mathbb{P}\left[M^\omega \le t\right] \quad \text{for any fixed } t \in \mathbb{R}.$$

$\square$

Bibliography

Adler, R. J. (2000). On excursion sets, tube formulas and maxima of random fields. *The Annals of Applied Probability*, pages 1–74.

Alm, S. E. (1998). Approximation and simulation of the distributions of scan statistics for Poisson processes in higher dimensions. *Extremes*, 1(1):111–126.

Anscombe, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 35(3/4):246–254.

Arias-Castro, E., Candès, E. J., and Durand, A. (2011). Detection of an anomalous cluster in a network. *The Annals of Statistics*, 39(1):278–304.

Arias-Castro, E., Castro, R. M., Tánczos, E., and Wang, M. (2017). Distribution-free detection of structured anomalies: Permutation and rank-based scans. *Journal of the American Statistical Association*, (to appear).

Arias-Castro, E., Donoho, D. L., and Huo, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Transactions on Information Theory*, 51(7):2402–2425.

Aspelmeier, T., Egner, A., and Munk, A. (2015). Modern Statistical Challenges in High-Resolution Fluorescence Microscopy. *Annual Review of Statistics and Its Application*, 2:163–202.

Bertero, M., Boccacci, P., Desiderà, G., and Vicidomini, G. (2009). Image deblurring with Poisson data: from cells to galaxies. *Inverse Problems*, 25(12):123006.

Billingsley, P. (2013). *Convergence of Probability Measures*. John Wiley & Sons.

Brown, L., Cai, T., Zhang, R., Zhao, L., and Zhou, H. (2010). The root–unroot algorithm for density estimation as implemented via wavelet block thresholding. *Probability theory and related fields*, 146(3-4):401.

Brown, L. D. (1986). Fundamentals of statistical exponential families with applications in statistical decision theory. *Instiute of Mathematical Statistics, Lecture Notes-Monograph Series*.

Butucea, C. and Ingster, Y. I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli*, 19(5B):2652–2688.

Candes, E., Tao, T., et al. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, 35(6):2313–2351.

Casella, G. and Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

Cassels, J. W. S. (1997). *An introduction to the geometry of numbers*. Classics in Mathematics. Springer, Berlin. Corrected reprint of the 1971 edition.

Cheng, D. and Schwartzman, A. (2017). Multiple testing of local maxima for detection of peaks in random fields. *The Annals of Statistics*, 45(2):529–556.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics*, 42(4):1564–1597.

Dedecker, J. (1998). A central limit theorem for stationary random fields. *Probability Theory and Related Fields*, 110(3):397–426.

Dedecker, J. (2001). Exponential inequalities and functional central limit theorems for random fields. *ESAIM: Probability and Statistics*, 5:77–104.

Despres, C. J. (2014). The Vapnik-Chervonenkis Dimension of Norms on $\mathbb{R}^d$. *arXiv preprint arXiv:1412.6612*.

Devroye, L. and Lugosi, G. (2001). *Combinatorial Methods in Density Estimation*. Springer Series in Statistics. Springer, New York.

Dickhaus, T. (2014). *Simultaneous Statistical Inference*. Springer, Heidelberg.

Dümbgen, L. and Spokoiny, V. G. (2001). Multiscale testing of qualitative hypotheses. *The Annals of Statistics*, 29(1):124–152.

Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *The Annals of Statistics*, 36(4):1758–1785.

Fang, X. and Siegmund, D. (2016). Poisson approximation for two scan statistics with rates of convergence. *The Annals of Applied Probability*, 26(4):2384–2418.

Frick, K., Marnitz, P., and Munk, A. (2013). Statistical multiresolution estimation for variational imaging: With an application in Poisson-biophotonics. *Journal of Mathematical Imaging and Vision*, 46(3):370–387.

Frick, K., Marnitz, P., Munk, A., et al. (2012). Statistical multiresolution Dantzig estimation in imaging: fundamental concepts and algorithmic framework. *Electronic Journal of Statistics*, 6:231–268.

Frick, K., Munk, A., and Sieling, H. (2014). Multiscale change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):495–580. With 32 discussions by 47 authors and a rejoinder by the authors.

Friedenberg, D. A. and Genovese, C. R. (2013). Straight to the source: Detecting aggregate objects in astronomical images with proper error control. *Journal of the American Statistical Association*, 108(502):456–468.

Grasmair, M., Li, H., Munk, A., et al. (2018). Variational multiscale nonparametric regression: smooth functions. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 54, pages 1058–1097. Institut Henri Poincaré.

Haiman, G. and Preda, C. (2006). Estimation for the distribution of two-dimensional discrete scan statistics. *Methodology and Computing in Applied Probability*, 8(3):373–382.

Hell, S. W. (2007). Far-field optical nanoscopy. *Science*, 316(5828):1153–1158.

Jiang, T. (2002). Maxima of partial sums indexed by geometrical structures. *The Annals of Probability*, 30(4):1854–1892.

Jiang, Y., Qiu, Y., Minn, A. J., and Zhang, N. R. (2016). Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing. *Proceedings of the National Academy of Sciences*, 113(37):E5528–E5537.

Kabluchko, Z. and Munk, A. (2009). Shao's theorem on the maximum of standardized random walk increments for multidimensional arrays. *ESAIM: Probability and Statistics*, 13:409–416.

Kazantsev, I., Lemahieu, I., Salov, G., and Denys, R. (2002). Statistical detection of defects in radiographic images in nondestructive testing. *Signal Processing*, 82(5):791 – 801.

Komlós, J., Major, P., and Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 34(1):33–58.

König, C., Munk, A., and Werner, F. (2018). Multidimensional multiscale scanning in Exponential Families: Limit theory and statistical consequences. *arXiv preprint arXiv:1802.07995*.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. Springer, New York.

Kou, J. (2017). Identifying the Support of Rectangular Signals in Gaussian Noise. *arXiv preprint arXiv:1703.06226*.

Krishnaswami, V., Van Noorden, C. J., Manders, E. M., and Hoebe, R. A. (2016). Spatially-controlled illumination microscopy: For prolonged live-cell and live-tissue imaging with extended dynamic range. *Quarterly Reviews of Biophysics*, 49.

Kulldorff, M., Heffernan, R., Hartman, J., Assunção, R., and Mostashari, F. (2005). A Space–Time Permutation Scan Statistic for Disease Outbreak Detection. *PLoS medicine*, 2(3):e59.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: isoperimetry and processes*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer, Berlin.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer, New York, third edition.

Marcus, M. B. and Rosen, J. (2006). *Markov processes, Gaussian processes, and local times*, volume 100 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

Massart, P. (1989). Strong approximation for multivariate empirical and related processes, via KMT constructions. *The Annals of Probability*, 17(1):266–291.

Naus, J. I. and Wallenstein, S. (2004). Multiple window and cluster size scan procedures. *Methodology and Computing in Applied Probability*, 6(4):389–400.

Pozdnyakov, V., Glaz, J., Kulldorff, M., and Steele, J. M. (2005). A martingale approach to scan statistics. *Annals of the Institute of Statistical Mathematics*, 57(1):21–37.

Proksch, K., Werner, F., and Munk, A. (2018). Multiscale scanning in inverse problems. *arXiv preprint arXiv:1611.04537*, (to appear in The Annals of Statistics).

Rio, E. (1993). Strong approximation for set-indexed partial-sum processes, via KMT constructions. II. *The Annals of Probability*, 21(3):1706–1727.

Rivera, C. and Walther, G. (2013). Optimal detection of a jump in the intensity of a Poisson process or in a density with likelihood ratio statistics. *Scandinavian Journal of Statistics*, 40(4):752–769.

Schmidt-Hieber, J., Munk, A., and Dümbgen, L. (2013). Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features. *The Annals of Statistics*, 41(3):1299–1328.

Sharpnack, J. and Arias-Castro, E. (2016). Exact asymptotics for the scan statistic and fast alternatives. *Electronic Journal of Statistics*, 10(2):2641–2684.

Siegmund, D. and Venkatraman, E. S. (1995). Using the generalized likelihood ratio statistic for sequential detection of a change-point. *The Annals of Statistics*, 23(1):255–271.

Siegmund, D. and Yakir, B. (2000). Tail probabilities for the null distribution of scanning statistics. *Bernoulli*, 6(2):191–213.

Staudt, T., Engler, A., Rittweger, E., Harke, B., Engelhardt, J., and Hell, S. W. (2011). Far-field optical nanoscopy with reduced number of state transition cycles. *Optics express*, 19(6):5644–5657.

Taylor, J. E. and Worsley, K. J. (2007). Detecting sparse signals in random fields, with an application to brain mapping. *Journal of the American Statistical Association*, 102(479):913–928.

Tu, I.-P. (2013). The maximum of a ratchet scanning process over a Poisson random field. *Statistica Sinica*, 23(4):1541–1551.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer, New York.

Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633.

Walther, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *The Annals of Statistics*, 38(2):1010–1033.

Zhang, N. R., Yakir, B., Xia, L. C., and Siegmund, D. (2016). Scan statistics on Poisson random fields with applications in genomics. *The Annals of Applied Statistics*, 10(2):726–755.

Curriculum Vitae

Claudia König

Born 5th July 1989 in Kiel, Germany.

▬ Education

since **PhD student**, *Georg-August-Universität Göttingen.*
2014 (Supervisors: Dr. F. Werner, Prof. Dr. A. Munk)

2014 **M.Sc. in Mathematics**, *Georg-August-Universität Göttingen, with distinction.*
Thesis: *Convergence Rates for Exponentially Ill-Posed Inverse Problems with Impulsive Noise* (Supervisor: Prof. Dr. T. Hohage)

2012 **B.Sc. in Mathematics**, *Georg-August-Universität Göttingen.*
Thesis: *Bestimmung der Menge aller Qualifikationen des Showalter-Verfahrens und verwandter Verfahren* (Supervisor: Prof. Dr. T. Hohage)

2009 **Abitur**, *Max-Planck-Schule*, Kiel.

▬ Academic grants & funding

since **Member of the CRC 755 "Nanoscale Photonic Imaging"**.
2014 Project A04

2017 **Oberwolfach-Leibniz-Graduate-Student**.
Grant for travel expenses concerning Oberwolfach Workshop ID 1712.

▬ Publications & Preprints

[1] **C. König**, F. Werner and T. Hohage, *Convergence rates for exponentially ill-posed inverse problems with impulsive noise*, SIAM Journal on Numerical Analysis, 54(1): 341-360 (2016)

[2] **C. König**, A. Munk and F. Werner, *Multidimensional multiscale scanning in Exponential Families: Limit theory and statistical consequences*, submitted. arXiv 1802.07995 (2018)

▬ Invited Talks & Seminars

2017 **31st European Meeting of Statisticians**, Helsinki.

2017 **Applied Inverse Problems conference**, Hangzhou.

2016 **3rd Conference of the ISNPS**, Avignon.

▬ Contributed Talks

2017 **Oberwolfach Workshop ID 1712**,
Statistical Recovery of Discrete, Geometric and Invariant Structures,
Oberwolfach.

2016 **IFIP Workshop "Optimal Control Meets Inverse Problems"**, Essen.

2016 **12. Doktorandentreffen Stochastik**, Bielefeld.

2015 **Workshop on Inverse Problems at the WIAS**, Berlin.

2015 **Applied Inverse Problems conference**, Helsinki.